UTokyo, Institute for Future Initiatives(IFI), SDGs Collaborative Research Unit
JSPS Grant Research Project
"The nexus of international politics in climate change and water resource, from the perspective of security studies and SDGs"
FY2020 Working Paper Series No. 9

Geographic Distribution of Water Conflicts Worldwide: A Comparative Analysis of Four Databases

The University of Tokyo
Takeshi Wada

Abstract

Is it possible to identify "hotspots" of disputes over water? The Nexus of International Politics in Climate Change and Water Resource is a collaborative research project where members conduct studies on specific cases of water conflicts around the world. The purpose of this article is to see these cases from a larger perspective. To this end, four databases have been selected and their characteristics compared in order to visualize a distribution of water conflicts around the world.

1. Necessities for grasping geographical distribution of water conflicts around the world

Disputes over water are rampant in many parts of the world. In the Nexus of International Politics in Climate Change and Water Resource[1], a collaborative research project launched by the University of Tokyo Institute for Future Initiatives, a focus of the research is water conflicts that are fought in developing countries, also known as the Global South. The chief aim here is to determine what factors will help enhance climate change resilience[2] and in turn lead the society in the direction of achieving the Sustainable Development Goals, by means of conducting in-depth analyses of the processes of conflicts at local levels.

Such micro-level analyses of cases, in addition to being a key element of this collaborative project, are also important for the purpose of untangling a complex web of elements including natural and environmental factors caused by climate change, socio-economic structural issues, historical and cultural background behind water resources, and fluid political processes of conflicts fought by different organizations and groups, and finding out ways to resolve the issues. At the same time, it is also necessary to see water conflicts from a larger perspective and put regional cases in a global context, and these are the aim of the present article. What is a geographical distribution of disputes over water like? Where do they occur most frequently, and where are they rarely seen? Is it possible to identify so-called hotspots, or conflict-prone areas? How are the cases that are handled in this collaborative research project positioned in a global context? Does information required for grasping a global or geographical distribution of water conflicts exist at all? It is the primary purpose of the present article to address these questions.

Specifically, the article will provide a brief overview of four databases that the author considers useful at present, and discuss findings obtained from them. For this reason, it should be noted that this article is of a descriptive nature. As it will be mentioned below, what motivated many of the researchers who created the databases on water conflicts that are discussed in the present article to do so was a pressing need for data; despite an abundance of case studies on disputes over water, few quantitative data were available to allow one to analyze them from a larger perspective (Yoffe and Kelli Larson 2001:7–8; Bernauer et al. 2012:529). This is a point that should be kept in mind by the members of this collaborative research project, which is focused on studies of cases from around the world. This is also where the present article may be of value in this project.

This article provides an overview of what a conflict database is and what their advantages and challenges are in the next section. Understanding these matters will help gain a perspective necessary in comparing the databases associated with water conflicts that are discussed later, and making a more proper interpretation of its results. In the four sections that follow, four databases on water conflicts will be profiled one at a time, and findings obtained from them will be discussed. The four databases that are discussed are as follows: (1) International Water Event Database, (2) Water-Related Intrastate Conflict and Cooperation, (3) Pacific Institute World Water Conflict Chronology, and (4) Global Database on Events, Language, and Tone (GDELT). The final section will discuss future directions for research in this field, as to what step should be taken next.

2. Challenges in building a water conflict database

Water problems may be one of the important nodes that link climate change and conflicts. This is because global warming affects the hydrological cycle and causes persistent water scarcity across the world, which may in turn increase the likelihood of dispute over scarce water resources (Denton 2015). When studying the relationship between climate change, water problems and conflict risk using quantitative analysis[3], one common approach is to apply water-related factors such as levels of water scarcity and precipitation as explanatory variables to an existing theoretical or empirical model (Koubi 2019; Kuzma et al. 2020). In doing so, researchers often use publicly available data on violent conflicts, which is popular in the fields of political science and sociology.

Bernauer et al. (2012:530–31) argued that this approach was limited by the possibilities that a violent conflict used as a dependent variable might not necessarily have been caused by a water problem. Even if a marginal effect of a water-related variable was observed in a regression analysis, one could hardly reason that water scarcity causes a conflict that is unrelated to water issues; at best, one could merely infer that some of the violent conflicts included in the data may have contained those caused by water issues. In order to further advance the research on climate change and water conflicts, it is of urgent importance to build a database on conflicts caused by water problems, and the subjects of study in the present article are such databases[4]. The analyses presented in the following will be focused on how the creators of the databases attempted to collect information on conflicts that had stemmed from water problems to build their databases.

When building a new database on water conflicts, there are a variety of challenges that need to be addressed, including what information sources to use (e.g., using existing conflict databases and extract data that involve water issues, or creating an entirely new database based on information obtained from media?), and how to define a "conflict concerning a water problem" (e.g., who decides it by what criteria?). Where the databases discussed below are concerned, (1) International Water Event Database and (3) Pacific Institute World Water Conflict Chronology started with collecting information from existing databases in political science, as opposed to (2) Water-Related Intrastate Conflict and Cooperation and (4) GDELT which involved collecting and coding of media articles. Each approach has its advantages and disadvantages its users should keep in mind, and these will be mentioned in the discussion of individual databases below.

The biggest challenge here is cost. Developing and maintaining a database is an expensive undertaking, for it requires manpower to perform the tasks of reading newspaper articles and other sources and coding the information into the database. People who undertake these tasks are called coders or annotators. Researchers themselves may perform these tasks, or they may hire students or other research assistants to do so. The more they aim to cover broader geographical areas, include a greater variety of behavioral patterns other than violent conflicts such as protests and cooperative activities in their data, use a greater number of media available in a greater variety of languages as their information sources, and update their data by incorporating newer information, the greater the cost of building a database will become.

Many research projects do not have enough funds to cover all regions of the world. It is also rare for a project to be able to continue paying for manpower after they have accomplished their research objectives established at the onset of their project. All this means that updating a database is fraught with difficulties. With the databases discussed in the present article, one could see that their creators made compromises, caught between ideal and reality. As the nature of such compromise inevitably affects the interpretation of the results obtained from the database, it is important to understand these cost issues when discussing the characteristics of individual databases. With this in mind, the following sections will look at the characteristics of the databases and findings obtained from them, starting with the International Water Event Database.

3. International Water Event Database (IWED)

The International Water Event Database (IWED) is a database provided as a part of the Program in Water Conflict Management and Transformation at Oregon State University, and it contains information concerning transboundary water-related incidents (Program in Water Conflict Management and Transformation 2021). It documents 7128 incidents/events from a period spanning 1948 through 2008. IWED was created because despite the number of case studies analyzing water conflict in international river basins, there was little quantitative, global-scale evidence (Yoffe and Kelli Larson 2001:7–8). By building this database, this project aimed to create historical indicators of international water conflicts and, from these indicators, identify river basins at risk for future conflict.

The information was sourced from a combination of existing political science datasets and searchable news databases. Because the existing datasets did not specifically concern water problems, they could not be used unless they allowed one to extract water-related events. For this reason, the following three databases were used to build IWED: the International Crisis Behavior Project (ICB), the Conflict and Peace Data Bank (COPDAB), and the Global Event Data System (GEDS). As for news databases, the Foreign Broadcast Information Service (FBIS), the World News Connection (WNC), and Lexis-Nexis were used by the researchers who conducted keyword and subject searches, read the articles retrieved, and coded relevant articles. In addition to the above, with respect to such events as concluding an international freshwater treaty, they also used the Transboundary Freshwater Dispute Database (TFDD), which was also created as a part of Oregon State University's Program in Water Conflict Management and Transformation.

What types of events are contained in this database? For the purpose of this project, “water events” were defined as “instances of conflict and cooperation that occur within an international river basin, that involve the nations riparian to that basin, and that concern freshwater as a scarce or consumable resource (e.g., water quality, water quantity) or as a quantity to be managed (e.g., flooding or flood control, managing water levels for navigational purposes)” (Yoffe and Kelli Larson 2001:8–9). While this may be somewhat abstract for a criterion, a look at the frequency distribution of a variable “Issue Areas (ISSUE_TYPE CATEGORIES)” provides insights into the specific types of events that are documented in the database. The most common type of water events was concerning Water Quantity (32.5% of all events), followed by Infrastructure/Development (21.2%), Joint Management (9.5%), Hydro-power/Hydro-electricity (8.7%), and Technical Cooperation/Assistance (8.2%).

For each event, the database documents information such as the date of the event, the countries involved, the main issue area, and a summary describing the event. There may be two key advantages with IWED. One is that it indicates the basin codes (BCode). By using these codes, IWED data may be consolidated with those of the above-mentioned TFDD, which allows one to visualize conflicts by basin or analyze associations with different variables used in TFDD for each basin (e.g., population density, number of dams in basin, precipitation in basin, hydro-political risk) (College of Earth, Ocean, and Atmospheric Sciences, Oregon State University. 2018).

The second advantage with IWED is that it documents not only international conflicts but their cooperative relationships as well. For each event, the international relationship was given a score from the most conflictive (-7) to the most cooperative (+7) on a scale called the Water Event Intensity Scale. Table 1 shows the specific types of events that are expected in each of the scores on the Water Event Intensity Scale. This was developed by Yoffe and Larson who worked on the Basins at Risk Project, based on which IWED was launched (Yoffe and Kelli Larson 2001:25–27).

Based on this database we will now look at the distribution of international water conflicts for each basin. Conflictive events that occurred in the current century (2000–2008) have been selected from the database, plotted on a world map, and displayed as a red gradient according to the event frequency, as shown in Figure 1A. As mentioned earlier, one of the characteristics of this database is that it allows one to categorize the locations of events according to the basin. This may provide useful information to professionals and specialists who are engaged in water resource management or aiming for a sustainable conflict solution.

Meanwhile, Figure 1B shows information on events categorized as cooperative international actions. Basins shown in darker green are those where cooperative actions occur more frequently. What is interesting is that the basins shown in darker colors here nearly match those in Figure 1A. In other words, basins are not divided into conflictive and cooperative ones, but rather basins that are home to a number of conflicts are also where cooperative actions take place frequently. This suggests that, where there is a seed of conflict, there is also effort being made to resolve it in a peaceful manner. Many of the databases on political actions that we use only handle data on conflicts such as violence, civil wars, and protests. When one is looking only at such data for analysis, one may run a risk of missing cooperative efforts for conflict resolution that are being made in parallel. This is a significant point indicated by IWED that one should keep in mind. It is also an indication that one should not necessarily have to be altogether pessimistic but may also find hope even in the face of a number of conflicts.

Where our cooperative research project is concerned, it would be valuable to pay attention to international conflicts at the periphery of India, as the area forms hotspots. Project members Takenaka and Chotani have been working on these areas, and the results are awaited.

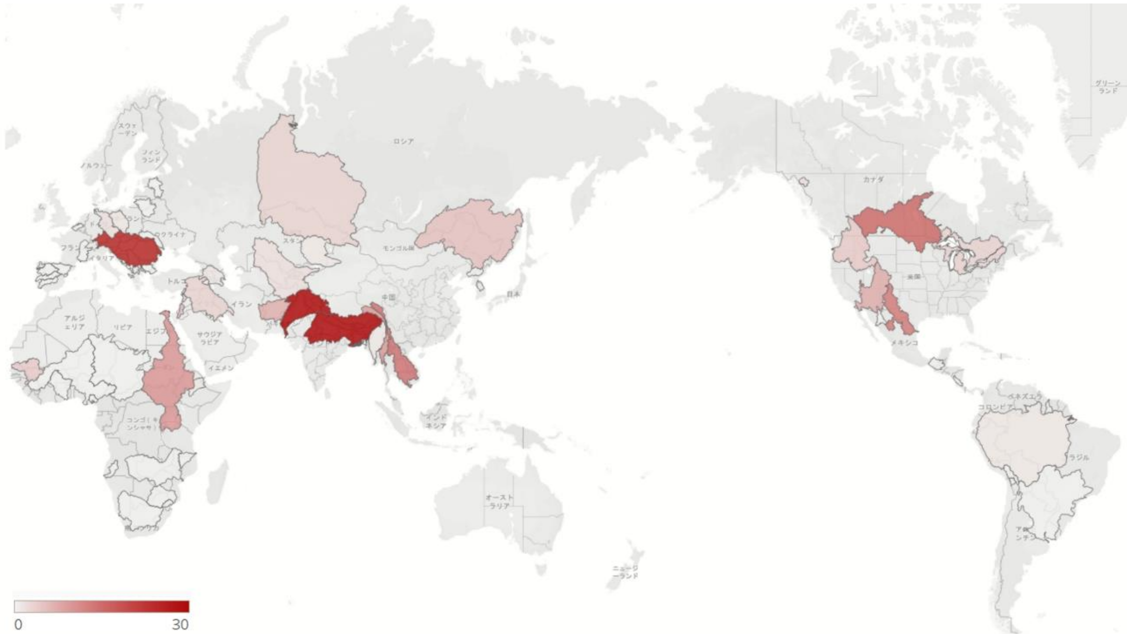
Table 1: IWED and GDELT scales for conflictive and cooperative actions

Score	IWED (Water Event Intensity Scale, WEIS)	GDELT (Goldstein Scale)
-10		Military attack, clash, assault
-9		Seize position or possessions (-9.2)
-8		Armed force mobilization, exercise, display, military buildup (-7.6)
-7	Formal declaration of war	Break diplomatic relations
-6	Extensive war acts (causing deaths, dislocation, or high strategic cost)	Threat with specific negative nonmilitary sanction (-5.8)
-5	Small scale military acts	Order person or personnel out of country
-4	Political-military hostile actions	Refuse, oppose
-3	Diplomatic-economic hostile actions	Give warning
-2	Strong verbal expressions displaying hostility in interaction	Make informal complaint (-1.9)
-1	Mild verbal expressions displaying discord in interaction	Deny an accusation (0.9)
0	Neutral or non-significant acts for the inter-nation situation	Explain or state policy, state future position
1	Minor official exchanges, talks or policy expressions in a form of mild verbal support	Meet with, send note
2	Official verbal support of goals, values, or regime	Admit wrongdoing, apologize, retract statement
3	Cultural or scientific agreement or support (non-strategic)	Agree to future action or procedure, to meet, or to negotiate
4	Non-military economic, technological, or industrial agreement	Endorse other's policy or position, give verbal support (3.6)
5	Military, economic, or strategic support	Promise material support (5.2)
6	International freshwater treaty: major strategic alliance (regional or international)	Make substantive agreement (6.5)
7	Voluntary unification into one nation	Extend economic aid (7.4)
8		Extend military assistance (8.3)
9		(No specific examples provided)
10		(No specific examples provided)

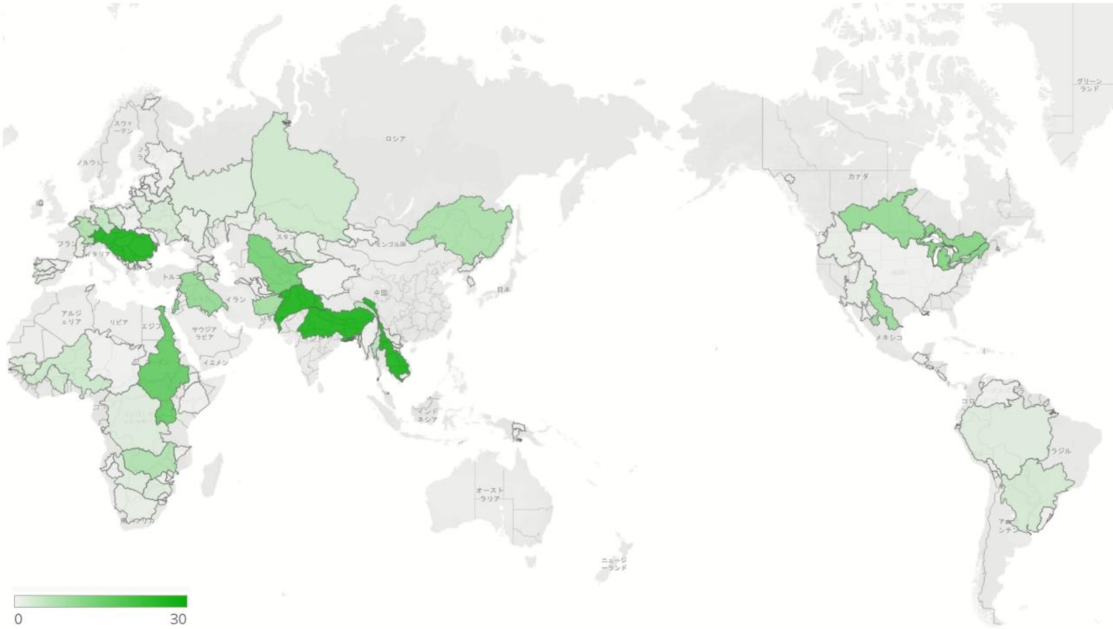
IWED Water Event Intensity Scale (WEIS) adapted from Yoffe and Kelli Larson (2001:25–27); GDELT's Goldstein Scale adapted from Goldstein (1992).

Figure 1: Distribution of international conflictive and cooperative water events by international Basin (2000-2008)

A) Conflictive



B) Cooperative



Created in Tableau Public with source data from IWED.

Lastly, limitations of IWED include the fact that the data are only available for up to 2008, and that it has no information on intrastate disputes over water. What databases are available that tell us about global distribution of water conflicts fought by not only states but a variety of social forces, and what are their advantages and limitations? Fortunately, there is a database that may be regarded as an intrastate version of IWED. It is called the Water-Related Intrastate Conflict and Cooperation (WARICC), and it is discussed in the next section.

4. Water-Related Intrastate Conflict and Cooperation

The Water-Related Intrastate Conflict and Cooperation (WARICC) is a dataset that may be considered an intrastate version of IWED (Bernauer et al. 2013). The researchers who built WARICC themselves pointed out that the poor data on domestic conflicts over water had been what had motivated them to attempt to create the database (Bernauer et al. 2012:529). WARICC covers 35 countries in the Mediterranean, the Middle East, and the Sahel, for a period from 1997 to 2009. It contains 10,352 events that concern domestic water problems.

As a source of information, BBC Monitoring was used. It monitors news articles from regional and international press and radio and TV broadcasting stations, translates them into English, and makes them accessible to the public. The researchers conducted searches using a string of keywords shown below, and coded data on water conflicts based on approximately 78,000 items retrieved (Bernauer et al. 2012:533).

> water OR lake OR river OR canal OR dam OR stream OR tributary OR dike OR dyke OR purification OR sewage OR effluence OR drought OR irrigation OR rain OR fish OR flood OR precipitation

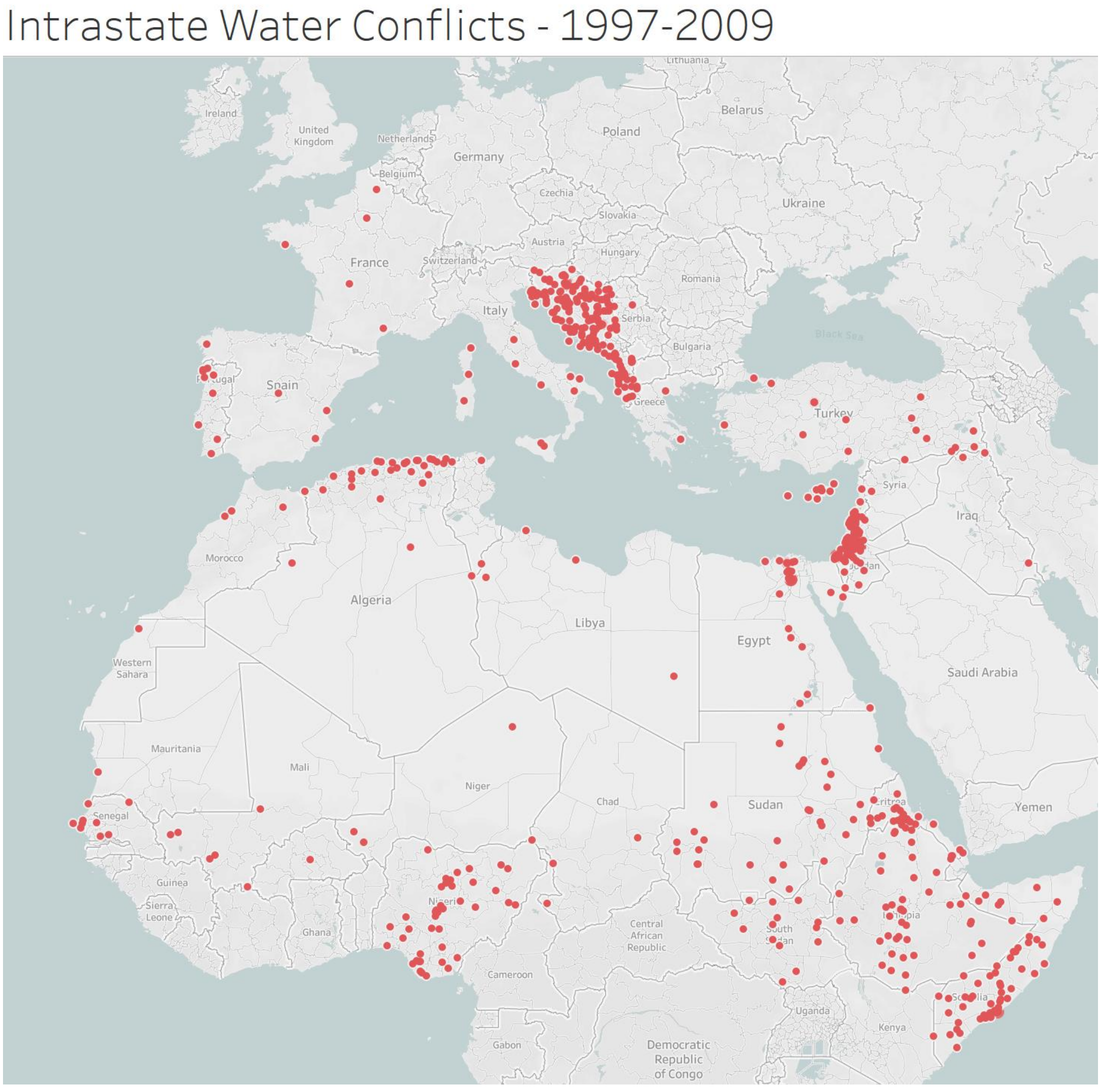
They arrived at the search string above through trial and error to improve efficiency (i.e., avoiding excessive amounts of irrelevant news items being retrieved) and precision (i.e., capturing news items that contain relevant events), building on keywords used in other databases such as IWED as reference.

One characteristics of the process of coding news items retrieved for WARICC is that data on not only conflictive but cooperative actions were also included, as it had been done with the IWED as an international conflict database. The key component of WARICC is Water Events Scale. It was used to score the intensity of conflictive or cooperative nature of the action for each event, ranging from the most conflictive (-5) to the most cooperative (+5) (Bernauer et al. 2012:537).

Figure 2 shows a map that visualizes the locations of the events in the database that indicate conflict (i.e., those with a negative score). This clearly shows areas with high conflict density. Firstly, a large number of conflicts have been located in and around the area where the former Yugoslavia was, which contains Slovenia, Croatia, Bosnia and Herzegovina, Montenegro, and Albania, among other countries. The density is also high in North Africa, along the coast of Algeria. Unfortunately, none of our cooperative research project members is specialized in these areas.

The map also shows a large number of conflicts located in the Middle East, especially in and

Figure 2: Distribution of violent water conflicts according to WARICC data (1977–2009)



Created in Tableau Public with source data from WARICC

around Israel and Jordan, as well as in a belt of countries across the Sahel region of Africa. Nishikida and Hanai have been tasked to analyze local cases from these two areas, respectively. This is significant in that it confirms that our team covers water conflict hotspots in the project.

One advantage of WARICC is its consistency with IWED, which allows one to compare data with those from the above-discussed database on international conflicts. It appears that WARICC contains information on key conflicts with fairly good precision. It has its limitations, however. One of them is that this database documents conflictive and cooperative events from a period up to only 2009, which leaves one with no information on more recent developments from 2010 onward. Another is its limited geographical coverage. One cannot analyze data on conflictive or cooperative events in other areas, including Latin America, based on WARICC.

The fact that both WARICC and IWED provide no latest information on conflicts over water may represent a structural problem faced by projects that deal with political event data. As mentioned earlier, it is too costly to develop and update such a database. In that regard, the two databases profiled in the following are valuable ones in that they provide information on relatively recent water conflicts from all regions of the world. We will first look at the Pacific Institute's database to see what it reveals about the world's disputes over water in recent years.

5. Pacific Institute World Water Conflict Chronology (PI)

The Pacific Institute (PI) is a U.S. think tank that involves itself in regional, domestic, and international initiatives with an aim to promote sustainable water policies. PI has published the World Water Conflict Chronology, which documents 926 violent conflicts concerning water resources and water systems that occurred between 3000 BC and August 2019 (Pacific Institute 2019)[5]. One of the benefits of using these data is that the database is updated continuously.

Updating data is very costly, and to overcome this problem, PI took an approach of pulling information on water-related issues from data on conflicts contained in existing databases that are publicly available, rather than amassing data on its own by using news media as information sources. PI referred to the following databases: Global Terrorism Database (GTD), Armed Conflict Location and Events Dataset (ACLED), Uppsala Conflict Data Program/Peace Research Institute Oslo Armed Conflict Datasets, Social Conflict Analysis Database (SCAD), Issue Correlates of War (ICOW) dataset, and the RAND Database of Worldwide Terrorism Incidents. It would be a rational approach to capitalize on existing databases which have been built at a large cost to collect information on conflicts. Now what can one learn from this database about the distribution of water conflicts?

Figure 3 shows a plot of events from the current century alone, 689 in total, on a map. What is instantly obvious is a strikingly uneven distribution of violent water conflicts. The conflict density is particularly high in the following areas: Yemen, which is located at the southern end of the Arabian Peninsula; Israel, Iraq, and their neighboring countries in the Middle East; the Sahel and the northern part of Algeria on the Mediterranean coast in Africa; India and Pakistan in South Asia; and Colombia in South America. According to the PI database, these areas represent hotspots of water-related conflicts in this century, and they correspond exactly to the areas covered by our research team.

The greatest advantage of the PI database is its categorization of events according to the nature of conflict, namely "Casualty," "Weapon," and "Trigger." These three categories are not mutually exclusive and many events involve more than one category. Figure 4 is a proportional area chart that

Figure 3: Distribution of violent water conflicts according to Pacific Institute data (2000-2019; N = 689)

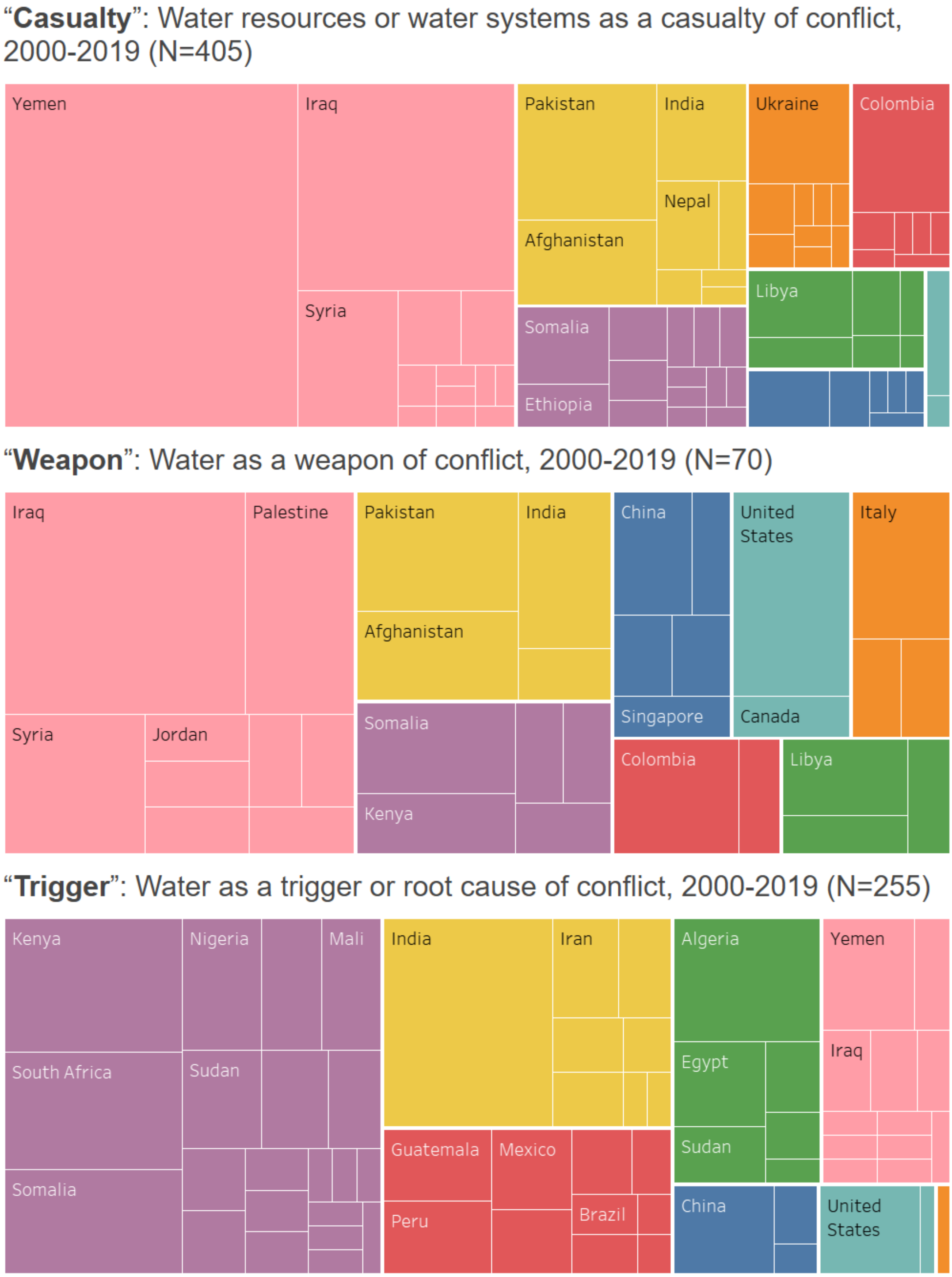
Created in Tableau Public with source data from PI database.

shows the frequency of events in each category by region (each shown in different color) and country (as square or rectangle).

Let us first have a look at "Casualty" in the top chart. This category represents events in which water resources or water systems are casualties of conflict, whether it is intentional or incidental. By region, it is evident that the Middle East (pink) accounts for roughly a half of all events, followed by South Asia (yellow) and Sub-Saharan Africa (purple). Yemen is home to by far the largest number of events, a trend that can be seen in Figure 3's map which plots all violent water conflicts. Considering that countries such as Yemen, Iraq, Syria, Pakistan, Afghanistan, Ukraine, and Columbia are prominently represented, one may infer that civil-war-affected countries may be seeing their water-related resources and assets becoming casualties as collateral damage in military action. These are, however, not necessarily a casualty resulting from a problem attributed to water, such as water scarcity.

Next, we look at the "Weapon" category, shown in the middle chart. This category includes events in which water is used as a weapon in a violent conflict. In these cases, water is not necessarily an objective of mobilization or a cause of hostile relationship; rather, water resources serve as a strategic tool to accomplish some other purpose. While the regional representation is similar to the Casualty category in that the Middle East, South Asia, and Sub-Saharan Africa are the most prominent in that order, the Weapon category has overall a more even distribution across all the regions of the world. In other words, use of water as a strategic weapon is a practice that is also observed in other regions, especially the United States and Italy.

Figure 4: Distribution of violent water conflicts by region and country (in 3 categories)



Created in Tableau Public with source data from PI database.

Lastly, "Trigger," which is shown in the bottom chart, is a category that includes events in which water itself is a root cause of conflict. When a dispute over the control of scarce water resources or water systems or a dispute over economic or physical access to water turns into violent conflict, it is documented as being in this category. As water is the principal cause of dispute for these events, this is the category of most interest to our research project team. Compared to the Casualty and Weapon categories, where the Middle East is represented most prominently, Sub-Saharan Africa and South Asia account for a vast majority of events in the Trigger category, followed by Latin America, northern Africa, and the Middle East. It is also evident that violent conflicts are not common in Asia, North America, and Europe.

From the standpoint of our collaborative research project, it was significant that the regions studied by many of the members are hotspots of conflicts in the Trigger category, including the Sahel (Hanai), South Asia (Hussain, Takenaka, Nagano, and Chotani), Latin America (Wada), and the Middle East (Nishikida).

The PI database is valuable in that its information is updated with the last update being relatively recent. The three categories of water conflicts presented by PI also represent a valuable concept that gives us a perspective on what should constitute a water conflict. Limitations of this database include the fact that it is dedicated to violent conflicts. Its data do not provide information on forms of dispute that do not necessarily involve a violent means, such as protests, or cooperative actions. While the PI database is useful in identifying hotspots, in order to answer the question “What should be done to solve water conflicts in a manner that may lead to sustainable development?” which is a research goal of our project, we need information on who the actors involved in cooperative actions are, and the PI database does not provide us that information.

6. GDELT

The last database to be discussed in this article is GDELT. As mentioned earlier, building an event database is highly costly as it requires a great deal of manpower, and some of the databases discussed above used existing conflict databases as a starting point for the gathering of information. The creators of GDELT took a different approach to build the event database; rather than hiring manpower for coding, they turned to natural language processing and big data analysis, and introduced a method of building a database without having humans read news articles.

Building on the Kansas Event Data System (KEDS) developed by Schrodt and colleagues (2006), GDELT handles a vast amount of information on news articles published in 100 languages and distributed by news media across the world, and auto-converts it into event data in real time. While one may go as far back as January 1, 1979, for event information, “theme” information, which is required to extract water-related event information, is only available with data from April 1, 2013 onward. Consequently, if one is to use GDELT on a theme of water conflicts, the period covered will necessarily be from April 1, 2013 or later[6].

Specifically, one will use the GDELT 1.0 Event Database (alternatively, its updated version, GDELT 2.0), which contains core information of events including date and time, location, action, and information source, with the GDELT 1.0 Global Knowledge Graph (GKG) (alternatively, GDELT 2.0 GKG), which documents theme information. A primary difference between GDELT 1.0 and GDELT 2.0 is that the latter has a new feature of incorporating articles published in 65 live-translated languages into the database.

What GDELT recognizes and records as events are over 300 categories of actions from protest and military assault to appeal for peacekeeping and diplomatic cooperation[7]. While the number of events contained is constantly growing, from 1979 onward a total of 664,506,044 events were recorded in GDELT 1.0 as of March 4, 2021[8].

We will now look at a global distribution of water conflicts using GDELT. Considering possibilities that the size of the datasets is simply too large for a regular personal computer to handle, GDELT 2.0 datasets are available in Google BigQuery for data searches. A user is only to enter a query to extract events they need, and they will be able to download a text file that contains such events instantaneously. We used this feature to obtain event data concerning water conflicts that were available as of February 21, 2021.

We began by performing searches in GDELT 2.0 GKG to retrieve only the water-related events. GDELT 2.0 GKG has a theme field called "V2Themes,", and we retrieved only the items that matched the following criteria: "Water (WB_137_WATER)" AND "Conflict and Violence (WB_2433_CONFLICT_AND_ VIOLENCE)" OR "Political Violence and War (WB_2462_POLITICAL_VIOLENCE_ AND_WAR)" OR "Political Violence and Civil War (WB_739_POLITICAL_ VIOLENCE_AND_CIVIL_WAR)." The theme codes all begin with the letters "WB" because GDELT 2.0 GKG uses a theme classification system based on the World Bank Group Topical Taxonomy. The World Bank Group's theme classification system is useful in that the classification is based on climate change and the Sustainable Development Goals, which means that it contains a number of codes that are in line with issues covered by the present cooperative research projects[9].

Next, we extracted key information of the events that matched the above-mentioned criteria, such as the location and actors, by creating a query that links GDELT 2.0 GKG data with GDELT 2.0 data and running it on Google BigQuery. To prevent the data file size from becoming too large, we limited our data to those from the period between January 1, 2020 and February 21, 2021, for the purpose of this article. The dataset downloaded contained a large number of entries with the same ID (GlobalEventID), and such duplicates were eliminated using R.

Furthermore, since not all the events in the downloaded dataset necessarily represent conflict, events that involved protest or violent action were selected[10]. There were 45,967 protest events and 208,292 violent action events. Clearly, GDELT records significantly larger numbers of events compared to the other databases that have been discussed earlier.

The two charts in Figure 5 show a geographical distribution of protest and violent action events, respectively. While large numbers of protest and violent action events were observed in areas identified as hotspots in earlier analyses such as parts of the Middle East in and around Israel, the Sahel in Africa, and India and its neighboring countries, the data are characterized by even larger numbers of events located in the United States as well as European countries such as Belarus. Does this mean that the Western countries are experiencing more water conflicts than in other regions?

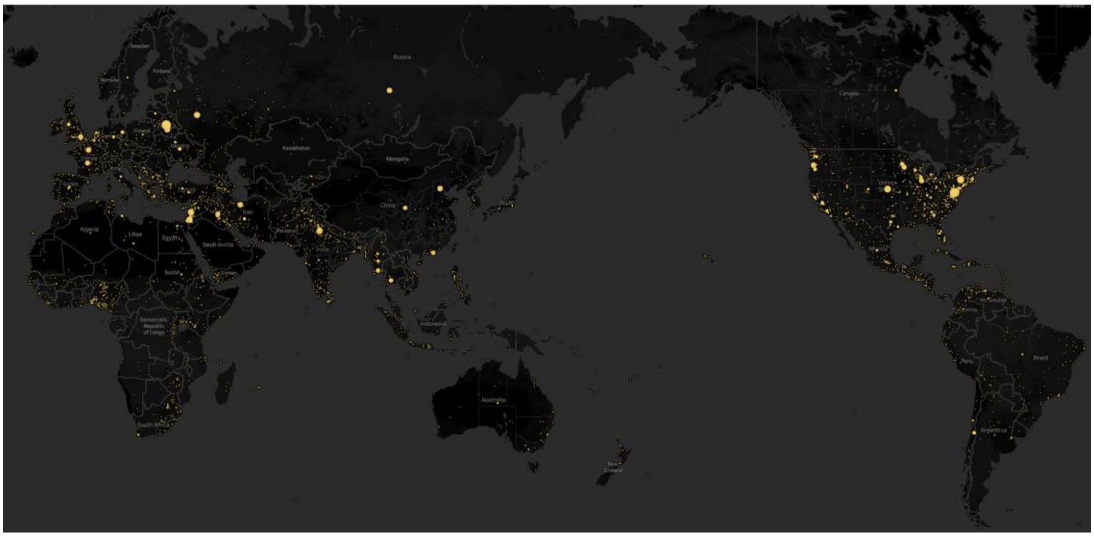
This finding should rather be interpreted as having resulted from the fact that the news media monitored by GDELT have higher volume of coverage of the Western world. Compared to other regions of the world, the West has a larger number of media and presumably they are more likely to distribute news online. This may have resulted in a larger number of events in the Western world being represented in the data. In addition, there will be a number of instances where a single event is reported by multiple media or articles. GDELT is aware of this, and has introduced a feature called "mentions" which automatically eliminates duplicate events as far as possible. The event deduplication is not complete, however, and a significant number of duplicates likely remained. In short, it is highly likely that the distribution of protest and violent action events shown in Figure 5 reflected both the number of events and that of news articles. Since the number of media and that of news articles captured by the GDELT system increase exponentially over time, GDELT recommends that one should "normalize" such data by dividing their count by the total number of news articles for a given unit of time (e.g., year, month, week, or day) when conducting time series analysis[11]. Judging by the charts above, the same seems to apply to comparative socio-spatial analysis.

For the purpose of the present article, we attempted to solve this problem by using the Goldstein Scale field in the GDELT event data, rather than by normalization. The Goldstein Scale is a scale that assigns a numeric score to a political action according to its nature, from being cooperative (highest score of +10) to conflictive (lowest score of -10), and it has been traditionally used for event analysis in political science (Goldstein 1992). It is a scale based on an idea similar to that behind IWED discussed earlier, and its details are shown in Table 1. As it assigns numeric scores to a wide variety of political actions, not all scores are integers; specific scores are indicated in parentheses in the table. The scale was first introduced for the analysis of international relations. As such, the items in the table are described in a manner that applies to action of a national government, although the descriptions have since been modified so that they may also apply to political actions of social forces.

Figure 5: Distribution of water conflicts according to GDELT 2.0 data (January 1, 2020–February 21, 2021)

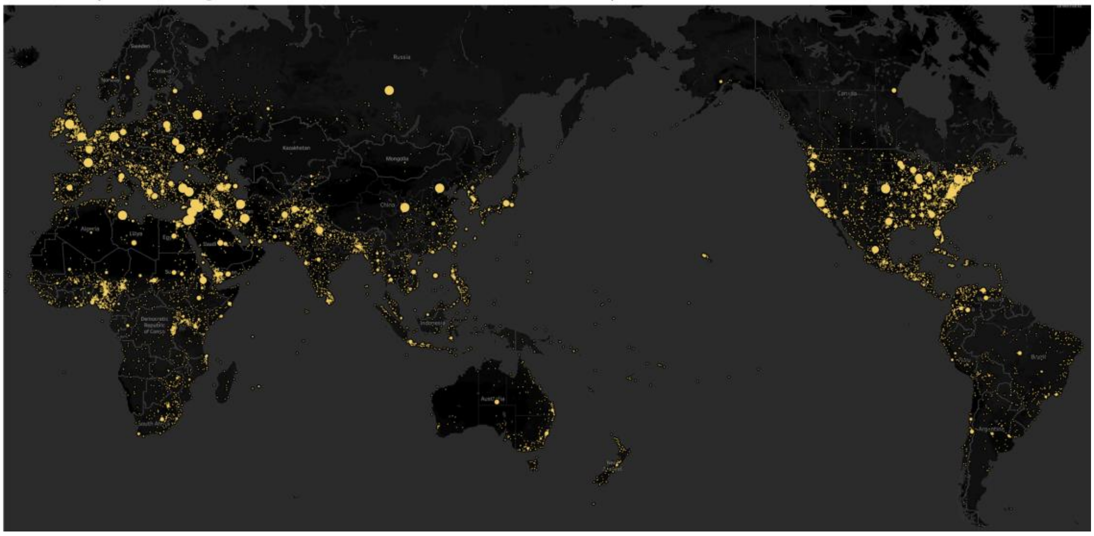
A) Protest (N = 45,967)

Protests -



B) Violent action (N = 208,292)

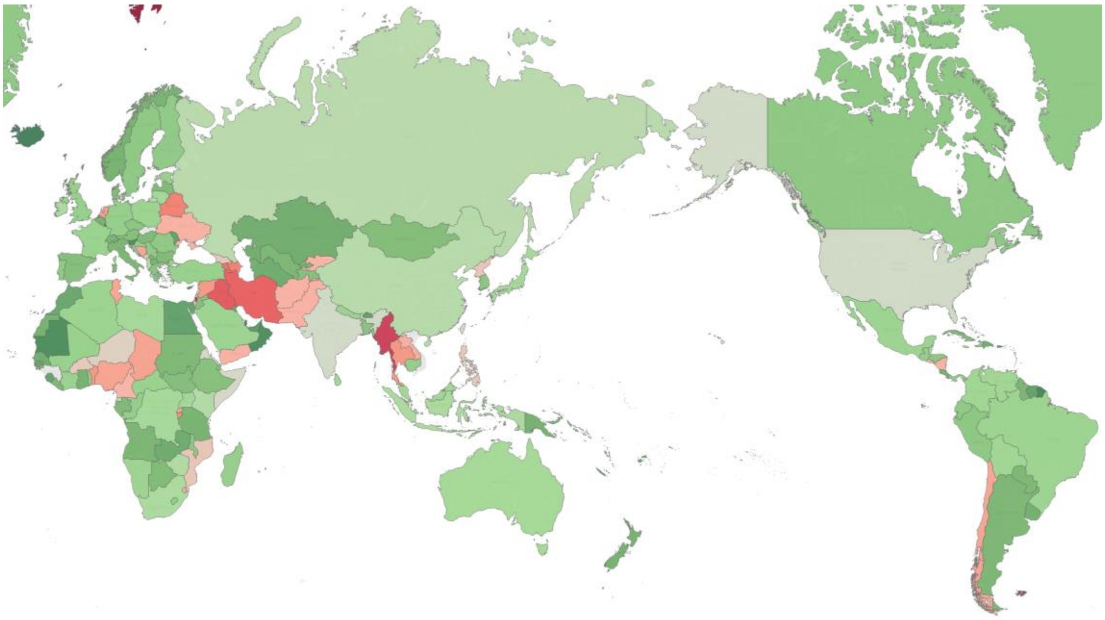
Violence (Assault, Fight, Use of Unconventional Mass Violence) -



Created in Tableau Public with source data from GDELT 2.0.

As is the case with the analysis of the IWED data, it was expected that the countries and regions for which a large number of conflictive events are reported on by the media would also see a large number of cooperative events reported (including cases where it was a result of a large number of news articles reporting such events). We reasoned that, by calculating the overall average, it might be possible to distinguish "water conflict hotspots" where conflictive actions account for greater percentages from areas with greater percentages of cooperative actions.

Figure 6: Conflictive and cooperative relationships involved in water events according to GDELT 2.0 data
(January 1, 2020–February 21, 2021)



Plots of national average Goldstein Scale scores in Tableau Public with source data from GDELT 2.0 event database.

The average score on the Goldstein Scale was calculated for each country and displayed as a color gradient on a map, as shown in Figure 6, where the areas shown in darker green have more events that involve cooperative actions, while those in darker red have more events that involve conflictive actions. In addition to the prominent dark red areas in Myanmar, Iran, Iraq, and the West Bank, the map shows clearly areas with high conflict density, including the Sahel in Africa, Israel and Syria in the Middle East, Belarus and Ukraine in Eastern Europe, Vietnam, Thailand, and the Philippines in Southeast Asia, Chile in South America, and El Salvador and Nicaragua in Central America. Many of these have also been identified as hotspots based on the analyses of the other databases, and it shows a certain degree of consistency.

GDELT is a favorable database that allows one to monitor conflictive and cooperative events around the world in real time. It may be highly useful in research that aims at predicting events that might occur in the future. With its system which codes articles published in a number of non-English languages using Google Translate, it has even more possibilities. That being said, one may find the process of obtaining data and using them for analysis too complicated at least until one becomes familiar with them, and the results of analysis can be difficult to interpret, as to whether they reflect the actual number of events or the number of articles that report such events. The coding precision of the natural language processing will also need to undergo scrutiny.

7. Next step

We have reviewed four databases that allow one to visualize where dispute over water is happening in the world. We have also pointed out that each has its advantages and disadvantages, and that care should be exercised when interpreting the results. This article has described global distribution of water conflicts, and visualized it in maps and other charts. A possible next step that should follow may be to provide an explanation of this distribution. What factors are behind the geographical distribution of water conflicts? What makes confrontations over water such common occurrences in certain areas of the world? Providing clear answers to these questions will help formulate a policy to avoid conflict, and having an accurate picture of the conflict distribution may be an important first step to do so.

For the purpose of our cooperative research project, it was useful to recognize how the case studies conducted by the members could be positioned in the global context. Nevertheless, our principal concern is to explore how such conflicts may be resolved at micro level, and what should be done to ensure that conflicts are resolved in a manner that helps reinforce climate change resilience. Fortunately, it has been confirmed that our case studies cover many of the hotspots of water conflicts. It is our intention to conduct comparison of the individual cases without losing a larger perspective that has been highlighted in the present article.

---

[1] An international cooperative research project undertaken by the SDGs Collaborative Research Unit at the University of Tokyo Institute for Future Initiatives. This project is supported by JSPS Kakenhi Grants-in-Aid for Scientific Research Nos. JP20K04995, JP19H00577, and JP18H00921.

[2] See Adger, Brown, and Waters (2011) for the concept of resilience in the context of climate change.

[3] Hsiang, Burke, and Miguel (2013) conducted a meta-analysis of a number of studies of this kind. A work by Koubi (2019) is also useful as it provides an overview of recent trends.

[4] One such database developed based on this idea is the Social Conflict Analysis Database (SCAD) by the Robert Strauss Center for International Security and Law at the University of Texas at Austin (Salehyan and Cullen Hendrix 2017). This is a database that provides highly interesting data, as it contains information on protests, riots, strikes, inter-communal conflicts, government violence against civilians, and other forms of social conflicts from a period between 1990 and 2017, covering Africa, Mexico, Central America, and the Caribbean. It even records the point of dispute for each event. It is, however, coded in a manner it puts "food, water, subsistence" all together, which renders impossible to extract water problems only. For this reason, this database is not discussed in the present article.

[5] As of the writing of this article, March 2021.

[6] As of the writing of this article; it is possible that news items from older dates may become available for analysis based on themes, etc. in future.

[7] An event classification system called the CAMEO Event Codes is used (Schrodt 2012).

[8] The total number of events was determined by performing calculations based on the "Yearly" file, available for downloading under the GDELT 1.0 Event Database Normalization Files section, which is found at the bottom of the Data page of the GDELT project website (https://www.gdeltproject.org/data.html). This file tabulates the yearly number of events from 1979 through the current year.

[9] For details on the World Bank Group Topical Taxonomy, see the website

(http://vocabulary.worldbank.org/taxonomy/1737.html).

[10] Such events were identified based on their value in the EventRootCode field: 14 (protest) for protest; and any of 18 (assault), 19 (fight), or 20 (use unconventional mass violence) for violent action.

[11] See the GDELT 1.0 Event Database Normalization Files section of the GDELT project website (https://www.gdeltproject.org/data.html) (Accessed March 4, 2021).

References

Adger, W. Neil, Katrina Brown, and James Waters. 2011. "Resilience" in *The Oxford handbook of climate change and society*, edited by John S. Dryzek, Richard B. Norgaard, and David Schlosberg. Oxford: Oxford University Press.

Bernauer, Thomas, Tobias Böhmelt, Halvard Buhaug, Nils Petter Gleditsch, Theresa Tribaldos, Eivind Berg Weibust, and Gerdis Wischnath. 2012. "Water-Related Intrastate Conflict and Cooperation (WARICC): A New Event Dataset." *International Interactions* 38 (4):529–45. doi: 10.1080/03050629.2012.697428.

Bernauer, Thomas, Tobias Böhmelt, Halvard Buhaug, Nils Petter Gleditsch, Theresa Tribaldos, Elvind Berg Weibust, and Gerdis Wischnath. 2013. "Water-Related Intrastate Conflict and Cooperation (WARICC): A New Event Dataset." Harvard Dataverse, V1. Retrieved February 15, 2021 (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/YT42X1).

College of Earth, Ocean, and Atmospheric Sciences, Oregon State University. 2018. "Draft Codebook for 2018 TFDD Spatial Update: Transboundary Freshwater Dispute Database." Retrieved February 26, 2021 (http://transboundarywaters.science.oregonstate.edu).

Denton, Fatima. 2015. "Climate Change and Conflict." *UN Chronicle* 52 (4):7–9.

Goldstein, Joshua S. 1992. "A conflict-cooperation scale for WEIS events data." *Journal of Conflict Resolution* 36 (2):369–85.

Hsiang, Solomon M., Marshall Burke, and Edward Miguel. 2013. "Quantifying the Influence of Climate on Human Conflict." *Science* 341 (6151):1235367.

Koubi, Vally. 2019. "Climate Change and Conflict." *Annual Review of Political Science* 22 (1):343–60. doi: 10.1146/annurev-polisci-050317-070830.

Kuzma, S., P. Kerins, E. Saccoccia, C. Whiteside, H. Roos, and C. Iceland. 2020. "Leveraging Water Data in a Machine Learning-Based Model for Forecasting Violent Conflict." Technical Note. Retrieved February 16, 2021 (www.wri.org/).

Pacific Institute. 2019. "Water Conflict Chronology." Retrieved March 4, 2021 (https://www.worldwater.org/water-conflict/).

Program in Water Conflict Management and Transformation. 2021. "International Water Event Database." Retrieved February 3, 2021 (https://transboundarywaters.science.oregonstate.edu/content/international-water-event-database).

Salehyan, Idean and Cullen Hendrix. 2017. “Social Conflict Analysis Database (SCAD) Version 3.3: Codebook and coding procedures.” Retrieved February 10, 2021.

Schrodt, Philip A. 2006. “Twenty years of the Kansas Event Data System Project.” *Political Methodologist* 14 (1):2–6.

Schrodt, Philip A. 2012. “CAMEO Conflict and Mediation Event Observations: Event and Actor Codebook.” Retrieved March 4, 2021 (http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf).

Yoffe, Shira and Kelli Larson. 2001. “Basins at Risk: Water Event Database Methodology.” Retrieved February 17, 2021 (https://transboundarywaters.science.oregonstate.edu/sites/transboundarywaters.science.oregonstate.edu/files/Database/Data/Events/Yoffe%20%26%20Larson-Event%20Coding.pdf).