

# Geocoded Disasters (GDIS) dataset

## Codebook

Elisabeth L. Rosvold, [elisabeth.rosvold@ekohist.su.se](mailto:elisabeth.rosvold@ekohist.su.se)

April 2021

### When using the dataset, please always cite

#### *The GDIS data descriptor*

Rosvold, E.L., Buhaug, H. GDIS, a global dataset of geocoded disaster locations. *Sci Data* 8, 61 (2021). <https://doi.org/10.1038/s41597-021-00846-6>

#### *The GDIS dataset*

Rosvold, E. and H. Buhaug. 2021. Geocoded disaster (GDIS) dataset. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/zz3b-8y61>. Accessed DAY MONTH YEAR.

#### *The EM-DAT dataset*

Guha-Sapir, Debarati, Below, Regina, & Hoyois, Philippe (2014). EM-DAT: International disaster database. *Centre for Research on the Epidemiology of Disasters (CRED)*.

### Data files

pend-gdis-1960-2018-disasterlocations.rdata  
pend-gdis-1960-2018-disasterlocations.gpkg  
pend-gdis-1960-2018-disasterlocations.gdb  
pend-gdis-1960-2018-disasterlocations.csv  
pend-gdis-1960-2018-priogrid-key.csv  
pend-gdis-1960-2018-replicationcode.R

This research has been conducted with financial support from the European Research Council, grant no. 648291. The dataset would not be available without the help of a number of people. Karim Bahgat wrote the automated script that was used to georeference a vast majority of the observations. Ingrid Vik Bakken, Nina Maureen Cadorin, Jens Kristoffer Haug and Bintu Zahara Sakor provided excellent research assistance and coding help. Responsibility for any errors remains my own. The data are available for download from [SEDAC](https://sedac.ciesin.columbia.edu/data/sets/0000012).

## What the dataset contains

This dataset is a geocoded extension of a selection of disasters from the Centre for Research on the Epidemiology of Disasters' (CRED) emergency events database (EM-DAT) (Guha-Sapir, Below, and Hoyois, 2014). The dataset encompasses 39,953 locations for 9,924 disasters that occurred worldwide in the years 1960 to 2018. All floods, storms (typhoons, monsoons etc.), earthquakes, landslides (wet and dry), droughts, volcanic activity and extreme temperatures that were recorded in EM-DAT during these 58 years are included in the dataset.<sup>1</sup>

Figure 1 shows the distribution of locations for the different types of disasters. Due to resource constraints and inherent challenges with ascribing precise geographical information to certain disasters, we decided to limit the geocoding to events occurring in this time period, and further exclude technological and biological disasters, as well as a limited number of smaller disaster types.<sup>2</sup> An overview of the disaster locations across countries is provided in Tables A.1-A.3 in the appendix.

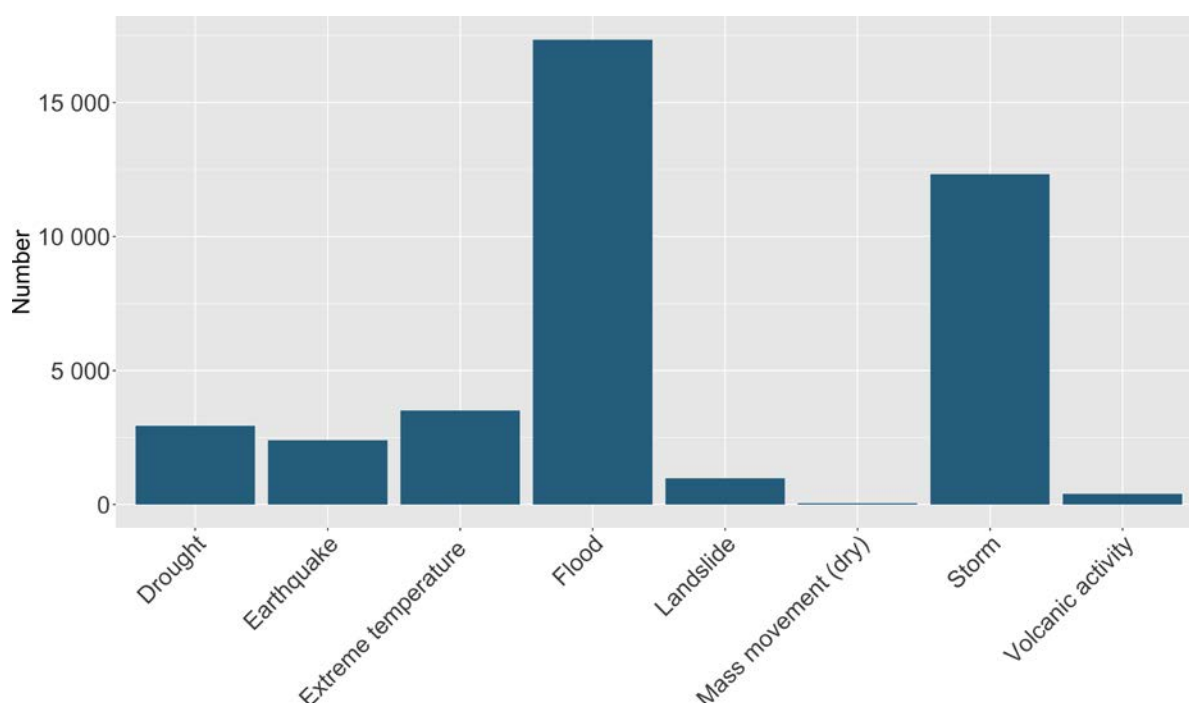


Figure 1: Disaster locations by disaster type, 1960–2018

The highest spatial resolution in the dataset corresponds to administrative level 3 (usually district/ commune/village) in the Global Administrative Areas database (GADM, 2018). Figure 2 shows the distribution of the disaster locations across these three levels, and reveals that the vast majority of the locations are on level 1 (typically state/province/region).

<sup>1</sup> The inclusion criteria for the EM-DAT database are that either (1) ten or more people must be reported killed, (2) hundred or more people reported affected, (3) a state of emergency declared or (4) a call for international assistance must be made. This means that only disasters of a certain magnitude are listed in the data. For more information and descriptions of the variables included in the disaster data, see the [EM-DAT documentation](#).

<sup>2</sup> These are: fog, wave action, glacial lake outburst, and wildfire.

## Coding procedure

The geocoding is based on the location variable in the original disaster data from EM-DAT. The list of locations for each disaster was split up, and the data reshaped so that each location got its own observation. From this, just short of 40,000 locations were validated for 9,924 disasters (out of a total of 11,801 disaster events). The list of locations was then sent through an automated script that matched on location name to return polygon geometries from the GADM database. Many locations did not match, and consequently this list was cleaned up manually and then automatically geocoded using Google API.<sup>3</sup> The remaining locations were manually matched with the GADM lists.

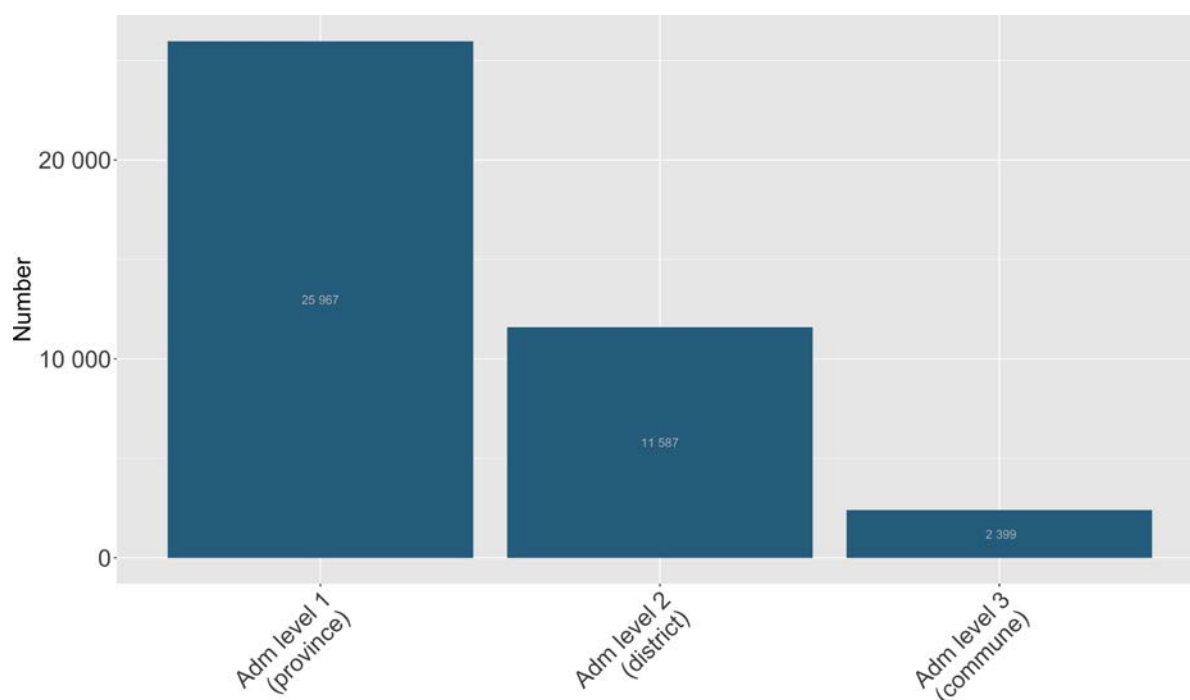


Figure 2: Disaster locations per administrative level, 1960–2018

The manual coding consisted of going through all observations that did not match in the automated queries and either 1) update the spelling of the location to match the spelling used in the GADM files, or 2) look up the location to identify which administrative level it was located in. In the latter case, the location was looked up in Google Maps and/or searched for in Google. The procedure here was to record which administrative level (1 or 2) the location lies within. The justification for this was both to save time for the coder, and because the majority of the locations listed are at these levels. Furthermore, this meant giving preference to more certain locations at a higher level of aggregation over less certain locations in higher disaggregation. In cases where the location was not found, unspecific (for example a river flowing through large parts of a country) or several locations within one country had the same name, the observation was dropped.

For several of the disasters, the regions mentioned in the location column in the original data were (large) regions encompassing several administrative level 1 units. In these cases, observations were expanded to the corresponding number of adm 1-locations. This was the case for many of the observations in the Philippines and Sri Lanka that both have large regions comprising several provinces on the first administrative level. In the instances when such localities had

---

<sup>3</sup> This procedure did not return polygon geometries, but rather coordinates for the locations plus info on all available administrative units.

more unclear boundaries however (for instance Hindu Kush, Aegian region (Turkey) or River Valley), the location was dropped.

Furthermore, several countries and regions have changed over time. On the country level this is most prominent for the former Soviet states, as well as Yemen, Sudan and the Balkans. The geocoding corresponds regions, administrative boundaries and countries as given in GADM version 3.6.<sup>4</sup> In order to allow users to identify which disasters happened in a country that no longer exists/has changed, these observations are marked by the *historical* variable (=1).

Regions change more frequently than countries, and are time again both split up in new regions or collapsed into larger entities. Locations that could easily be identified as lying within current (GADM) adm 1 (or 2) units were updated accordingly. However, in the cases where this identification was not straight-forward, the locations were dropped. It is worth noting that for some of the most recent reconfigurations of administrative boundaries GADM does not always corresponds to the most recent configuration (for example the case for Nepalese regions). In such cases, the coding of these observations went in the reverse direction, always corresponding to the boundaries of GADM 3.6.

The total number of disasters that were retrieved from EM-DAT was 11,801, meaning that for 1,157 disasters no location could be identified from the *location* column. In addition to the reasons mentioned above, some of these are also due to lacking geographical information from the GADM. This is most prevalent for small island states. Neither Kiribati, Maldives or Marshall Islands are in the GADM list, and consequently none of the disasters occurring here are geocoded, even if specific locations are given in EM-DAT.

In order to avoid inflating the number of locations, only the observation at the highest level of precision is kept in cases where the locations given for one specific disaster are overlapping. This means that the user has to scale up the unit of analysis for the observations that are on levels 3 and 2 in the event that the desired unit of observation at a higher level.

## Overview of variables

All variables in the dataset are presented and explained in Table 1 below. To get information on the specific disaster, such as date, type and magnitude, the user must consult the EM-DAT. This information can then be joined with the location data via the *disasterno* identifier.

---

<sup>4</sup> See the [Pshapes project](#) for an exciting project about geocoding historical regions

Table 1. Variable overview and description

Variable Name	Description
<b>disasterno</b>	ID-variable from EM-DAT (Guha-Sapir et al., 2014), use this to join the geocoded data with EM-DAT records in order to obtain information on the specific disasters *
<b>id</b>	ID-variable identifying each disaster in the geocoded dataset. Contrary to <i>disasterno</i> each disaster in each country has a unique id number
<b>geo_id</b>	Unique ID-variable for each location
<b>country</b>	Name of the country within which the location is
<b>iso3</b>	Three-letter country code, ISO 3166-1
<b>gwno</b>	Gleditsch and Ward country code (Gleditsch & Ward, 1999)
<b>geolocation</b>	Name of the location of the observation, which corresponds to the highest (most disaggregated) level available. For instance, observations at the third administrative level will have geolocation values identical to the <i>adm3</i> variable
<b>level</b>	The administrative level of the observation, ranges from 1-3 where 3 is the most disaggregated
<b>adm1</b>	Name of administrative level 1 for the given location
<b>adm2</b>	Name of administrative level 2 for the given location
<b>adm3</b>	Name of administrative level 3 for the given location
<b>location</b>	The location as it was extracted from the original dataset. This is the string on which the geocoding was based
<b>historical</b>	Marks whether the disaster happened in a country that has since changed, takes the value 1 if the disaster happened in a country that has since changed, and 0 if not
<b>hist_country</b>	Name of country at the time of the disaster, if the observation takes the value 1 on the <i>historical</i> variable, this is different from the <i>country</i> variable
<b>disastertype</b>	Type of disaster as defined by EM-DAT (Guha-Sapir et al., 2014): flood, storm, earthquake, extreme temperature, landslide, volcanic activity, drought or mass movement (dry)
<b>geometry</b>	Geometry for the observation's most disaggregated known administrative level. The polygon is extracted from the GADM database (GADM, 2018) (this variable is not included in the csv-file)
<b>centroid</b>	Centroid of the administrative level, longitude-latitude coordinates

\* Note that after the release of this dataset (February 2021) EM-DAT changed the code for *disasterno* to include a dash followed by the ISO3 code.

## Final note

All disasters locations are contingent on the data that were available at the time of download. The first disasters were downloaded from the EM-DAT already in primo 2016, with subsequent downloads of additional disasters in 2018 and 2019. As EM-DAT is continuously updated, this means that the list of disasters will not be identical to the original database as potential backlogged entries both on disaster events and their locations will not be included.

## **References**

GADM (2018). Database of Global Administrative Areas, Version 3.6.

Gleditsch, K. S. and M. D. Ward (1999). A revised list of independent states since the congress of Vienna. *International Interactions* 25 (4), 393–413.

Guha-Sapir, D., R. Below, and P. Hoyois (2014). EM-DAT: International disaster database. *Centre for Research on the Epidemiology of Disasters (CRED)*.

# Appendix

Table A.1: Disaster locations by country, 1960–2018.

Country	Drought	Earthquake	Extreme temperature	Flood	Landslide	Mass movement	Storm	Volcanic activity	Total no. locations	Unique disasters
Afghanistan	48	89	26	261	38	0	29	0	491	145
Albania	0	5	35	74	0	0	12	0	126	26
Algeria	0	30	21	186	1	0	13	0	251	68
American Samoa	0	2	0	2	0	0	1	0	5	3
Angola	31	0	0	141	1	0	0	0	173	45
Antigua and Barbuda	0	0	0	0	0	0	10	0	10	3
Argentina	18	5	56	226	10	0	30	7	352	96
Armenia	7	3	1	7	2	0	2	0	22	10
Australia	16	4	8	297	2	0	154	0	481	173
Austria	0	1	32	30	3	0	24	0	90	26
Azerbaijan	10	14	1	46	1	0	0	0	72	15
Bahamas	0	0	0	1	0	0	107	0	108	14
Bangladesh	9	26	150	376	11	0	358	0	930	220
Barbados	11	1	0	11	0	0	8	0	31	5
Belarus	0	0	63	20	0	0	14	0	97	11
Belgium	0	1	17	27	0	0	18	0	63	32
Belize	0	0	0	7	0	0	17	0	24	10
Benin	4	0	0	77	0	0	1	0	82	21
Bhutan	0	4	0	3	0	0	0	0	7	4
Bolivia	35	2	16	173	27	0	6	1	260	67
Bosnia and Herzegovina	4	1	22	76	1	0	5	0	109	23
Botswana	0	0	0	35	0	0	1	0	36	10
Brazil	52	3	8	382	28	0	24	0	497	179
British Virgin Islands	0	0	0	0	0	0	3	0	3	1
Bulgaria	27	5	82	120	1	0	20	0	255	39
Burkina Faso	15	0	0	83	0	0	0	0	98	20
Burundi	47	1	0	61	6	0	6	0	121	39
Cambodia	35	0	0	115	0	0	1	0	151	22
Cameroon	2	0	0	29	2	0	0	4	37	20
Canada	8	0	8	76	1	1	50	0	144	81
Cape Verde	6	0	0	2	0	0	35	1	44	7
Cayman Islands	0	0	0	0	0	0	5	0	5	3
Central African Republic	0	0	0	31	0	0	24	0	55	25
Chad	28	0	0	51	0	0	3	0	82	20
Chile	0	59	67	85	4	0	10	13	238	66
China	117	204	38	926	102	7	641	0	2,035	799
Colombia	16	50	0	391	41	3	15	41	557	135
Comoros	1	1	0	6	0	0	13	5	26	15
Costa Rica	9	37	0	118	2	0	22	3	191	52
Cote D'Ivoire	4	0	3	15	1	0	0	0	23	21
Croatia	20	2	86	31	0	0	8	0	147	21
Cuba	29	4	0	59	0	0	168	0	260	51
Cyprus	5	2	6	0	0	0	4	0	17	7
Czech Republic	0	0	22	30	0	0	4	0	56	18
Democratic Republic of the Congo	3	6	0	34	7	0	6	4	60	49
Denmark	0	0	0	0	0	0	7	0	7	4
Djibouti	25	0	0	9	0	0	0	0	34	13
Dominica	0	0	0	0	0	0	18	0	18	3
Dominican Republic	0	4	0	75	0	0	175	0	254	40
Ecuador	20	34	0	151	16	1	0	58	280	69
Egypt	0	4	6	24	0	2	18	0	54	22
El Salvador	29	29	1	57	2	0	26	6	150	36
Eritrea	7	0	0	2	0	0	1	0	10	5
Estonia	0	0	20	0	0	0	1	0	21	3
Ethiopia	72	2	0	105	4	1	0	3	187	69
Fiji	2	1	0	26	0	0	74	0	103	41
Finland	0	0	0	1	0	0	0	0	1	1
France	6	1	95	101	15	0	118	0	336	109
French Guiana	0	0	0	0	1	0	0	0	1	1
French Polynesia	0	0	0	1	4	0	6	0	11	7
Gabon	0	0	0	4	0	0	7	0	11	5
Gambia	8	0	0	9	0	0	5	0	22	11
Georgia	6	8	0	38	0	0	5	0	57	26
Germany	0	5	87	55	0	0	68	0	215	59
Ghana	0	0	0	74	0	0	1	0	75	21
Greece	0	45	24	72	0	0	6	0	147	58
Grenada	2	0	0	0	0	0	5	0	7	3
Guadeloupe	0	1	0	0	0	0	0	0	1	1
Guam	0	1	0	0	0	0	8	0	9	3
Guatemala	31	38	45	115	17	3	47	17	313	76
Guinea	0	1	0	33	1	0	1	0	36	14
Guinea-Bissau	11	0	0	4	0	0	7	0	22	7
Guyana	6	0	0	27	0	0	0	0	33	9

Table A.2: Disaster locations by country, 1960–2018, cont.

Country	Drought	Earthquake	Extreme temperature	Flood	Landslide	Mass movement	Storm	Volcanic activity	Total no. locations	Unique disasters
Haiti	39	9	0	125	1	0	133	0	307	94
Honduras	68	19	0	131	1	0	60	0	279	55
Hong Kong	0	0	0	23	6	0	50	0	79	39
Hungary	20	1	81	23	0	0	5	0	130	23
Iceland	0	4	0	1	3	0	0	3	11	9
India	315	99	206	1,172	66	1	394	0	2,253	533
Indonesia	20	250	0	663	119	2	27	97	1,178	425
Iran	8	151	0	227	5	0	38	0	429	183
Iraq	3	6	0	24	0	0	0	0	33	14
Ireland	0	0	0	20	0	0	14	0	34	9
Israel	0	0	3	4	1	0	11	0	19	9
Italy	41	70	44	123	12	0	56	3	349	114
Jamaica	19	0	0	36	0	0	112	0	167	25
Japan	1	207	206	90	9	0	835	13	1,361	206
Jordan	12	0	1	7	0	0	0	0	20	6
Kazakhstan	0	2	4	52	1	0	2	0	61	19
Kenya	127	2	0	276	7	0	0	0	412	63
Kosovo	0	2	24	11	0	0	0	0	37	9
Kuwait	0	0	0	1	0	0	0	0	1	1
Kyrgyzstan	7	20	5	16	15	0	2	0	65	27
Laos	0	0	0	91	0	0	27	0	118	22
Latvia	0	0	52	0	0	0	0	0	52	4
Lebanon	0	0	0	3	0	1	11	0	15	8
Lesotho	29	0	0	8	0	0	5	0	42	11
Liberia	0	0	0	10	0	1	2	0	13	9
Libya	0	1	0	3	0	0	0	0	4	3
Lithuania	10	0	40	5	0	0	0	0	55	7
Luxembourg	0	0	3	0	0	0	0	0	3	1
Macao	0	0	0	0	0	0	1	0	1	1
Macedonia	0	38	70	80	0	0	0	0	188	21
Madagascar	35	0	0	18	0	0	316	0	369	60
Malawi	35	3	0	180	0	0	2	0	220	44
Malaysia	6	2	0	202	4	1	15	0	230	62
Mali	27	0	0	84	0	0	0	0	111	29
Martinique	0	1	0	0	0	0	6	0	7	2
Mauritania	42	0	0	46	0	0	4	0	92	23
Mauritius	0	0	0	1	0	0	28	0	29	14
Mexico	27	71	56	153	12	0	244	7	570	207
Micronesia	5	0	0	1	0	0	5	0	11	8
Moldova	25	0	18	32	0	0	7	0	82	12
Mongolia	12	0	40	27	2	0	87	0	168	19
Montenegro	0	1	20	25	0	0	14	0	60	11
Montserrat	0	0	0	0	0	0	0	3	3	3
Morocco	1	5	27	79	1	1	6	0	120	36
Mozambique	59	4	0	224	2	0	65	0	354	67
Myanmar	0	17	0	115	10	1	52	0	195	60
Namibia	19	0	0	51	0	0	0	0	70	18
Nepal	6	41	48	385	72	1	8	0	561	83
Netherlands	0	1	49	3	0	0	8	0	61	12
New Caledonia	0	0	0	0	0	0	11	0	11	7
New Zealand	9	15	5	65	1	0	25	0	120	60
Nicaragua	47	13	0	69	2	0	53	14	198	48
Niger	54	0	0	118	0	0	1	0	173	33
Nigeria	0	0	1	192	3	0	16	0	212	51
North Korea	13	0	4	93	0	0	20	0	130	33
Northern Mariana Islands	0	0	0	0	0	0	6	1	7	5
Norway	0	0	0	5	0	0	1	0	6	4
Oman	0	0	0	6	0	0	10	0	16	9
Pakistan	6	66	26	255	29	2	47	0	431	165
Palau	0	0	0	0	0	0	11	0	11	1
Palestina	0	0	0	13	0	0	6	0	19	7
Panama	17	7	0	90	0	0	15	0	129	47
Papua New Guinea	23	16	0	34	11	3	7	17	111	59
Paraguay	42	0	42	58	0	0	30	0	172	36
Peru	49	85	109	297	42	2	23	15	622	138
Philippines	19	72	0	844	54	5	2,475	56	3,525	479
Poland	0	1	215	34	0	0	22	0	272	34
Portugal	1	4	78	34	1	0	9	0	127	30
Puerto Rico	0	0	0	17	0	0	22	0	39	9
Qatar	0	0	0	1	0	0	0	0	1	1
Republic of Congo	0	0	0	12	0	0	0	0	12	9
Reunion	0	0	0	0	1	0	2	1	4	4
Romania	8	6	190	235	0	0	31	0	470	65
Russia	52	17	81	171	9	0	45	0	375	125
Rwanda	21	3	0	46	8	0	10	0	88	27
Saint Helena	0	0	0	0	0	0	1	0	1	1



Table A.3: Disaster locations by country, 1960–2018, cont.

Country	Drought	Earthquake	Extreme temperature	Flood	Landslide	Mass movement	Storm	Volcanic activity	Total no. locations	Unique disasters
Saint Lucia	0	3	0	5	0	0	8	0	16	5
Saint Vincent and the Grenadines	1	0	0	9	0	0	2	0	12	7
Samoa	0	0	0	2	0	0	61	0	63	7
Saudi Arabia	0	0	0	43	0	0	1	0	44	20
Senegal	29	0	0	64	0	0	7	0	100	27
Serbia	0	4	117	86	0	0	5	0	212	32
Seychelles	0	16	0	32	0	0	29	0	77	5
Sierra Leone	0	0	0	10	2	0	0	0	12	7
Slovakia	0	0	29	16	0	0	4	0	49	16
Slovenia	0	7	13	5	0	0	4	0	29	7
Solomon Islands	5	15	0	16	0	0	28	1	65	27
Somalia	78	5	0	89	0	0	8	0	180	52
South Africa	12	5	4	67	1	0	52	0	141	71
South Korea	15	2	15	77	6	0	67	0	182	62
South Sudan	14	1	0	29	0	0	0	0	44	16
Spain	10	2	19	70	1	0	25	0	127	50
Sri Lanka	30	14	0	336	3	0	25	0	408	72
Sudan	11	1	1	158	0	0	2	0	173	44
Suriname	0	0	0	15	0	0	0	0	15	2
Swaziland	41	0	0	4	0	0	10	0	55	7
Sweden	0	0	15	3	1	0	0	0	19	3
Switzerland	0	0	60	18	9	2	34	0	123	30
Syria	6	0	0	5	1	0	14	0	26	9
Taiwan	0	18	6	23	1	0	105	0	153	61
Tajikistan	11	20	5	108	33	0	2	0	179	62
Tanzania	28	12	0	131	1	0	6	0	178	62
Thailand	174	23	10	616	2	0	206	0	1,031	114
Timor-Leste	10	1	0	29	0	0	7	0	47	9
Togo	2	0	0	35	0	0	2	0	39	12
Tokelau	0	0	0	0	0	0	11	0	11	4
Tonga	4	2	0	0	0	0	29	0	35	20
Trinidad and Tobago	15	0	0	5	1	0	2	1	24	6
Tunisia	0	0	0	67	0	0	0	0	67	14
Turkey	0	97	95	78	18	8	95	0	391	111
Turkmenistan	0	2	0	0	0	0	0	0	2	2
Turks and Caicos Islands	0	0	0	0	0	0	8	0	8	4
Tuvalu	1	0	0	0	0	0	16	0	17	7
Uganda	20	7	0	89	17	0	4	0	137	39
Ukraine	1	0	97	55	0	0	20	0	173	30
United Arab Emirates	0	0	0	3	0	0	0	0	3	1
United Kingdom	0	4	39	228	1	0	45	0	317	53
United States	74	50	155	683	7	0	2,848	2	3,819	731
Uruguay	19	0	32	89	0	0	12	0	152	25
Uzbekistan	2	5	0	2	0	0	0	0	9	7
Vanuatu	0	8	0	2	0	0	48	7	65	35
Venezuela	26	9	0	93	5	0	3	0	136	39
Vietnam	41	0	0	423	6	0	370	0	840	176
Virgin Islands, U.S.	0	0	0	0	0	0	8	0	8	4
Yemen	1	2	0	80	3	0	16	1	103	40
Zambia	3	0	0	71	1	0	0	0	75	22
Zimbabwe	33	0	0	34	0	0	47	0	114	20
Total	2,938	2,403	3,506	17,347	982	49	12,323	405	39,953	9,928