

The GEM Global Active Faults Database

Earthquake Spectra

I-21

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/8755293020944182

journals.sagepub.com/home/eqs**Richard Styron**, and **Marco Pagani**

Abstract

The GEM Global Active Faults Database (GAF-DB) is the first public, comprehensive database of active faults with worldwide coverage. The GAF-DB is a compilation of many regional datasets. The GAF-DB contains ~13,500 faults, each with associated attributes that describe the geometry, kinematics, slip rate, references, and other characteristics, as the information is available. Spatial completeness is high, and about 77% of the faults have slip rate information. The GAF-DB is built from its constituent datasets algorithmically and is designed to fluidly incorporate changes to or addition of any of the underlying datasets. This process reflects a philosophy of easily incorporating a change to avoid obsolescence and to quickly provide the most up-to-date information possible to the users. The database is licensed under a free and open-source license (CC-BY-SA 4.0) and is available at <https://github.com/GEMScienceTools/gem-global-active-faults>.

Keywords

Earthquake, active fault

Date received: 25 June 2020; accepted: 29 June 2020

Global Earthquake Model Foundation, Pavia, Italy

Corresponding author:

Richard Styron, Global Earthquake Model Foundation, Via Adolfo Ferrata 1, 27100 Pavia, Italy.
Email: richard.styron@globalquakemodel.org

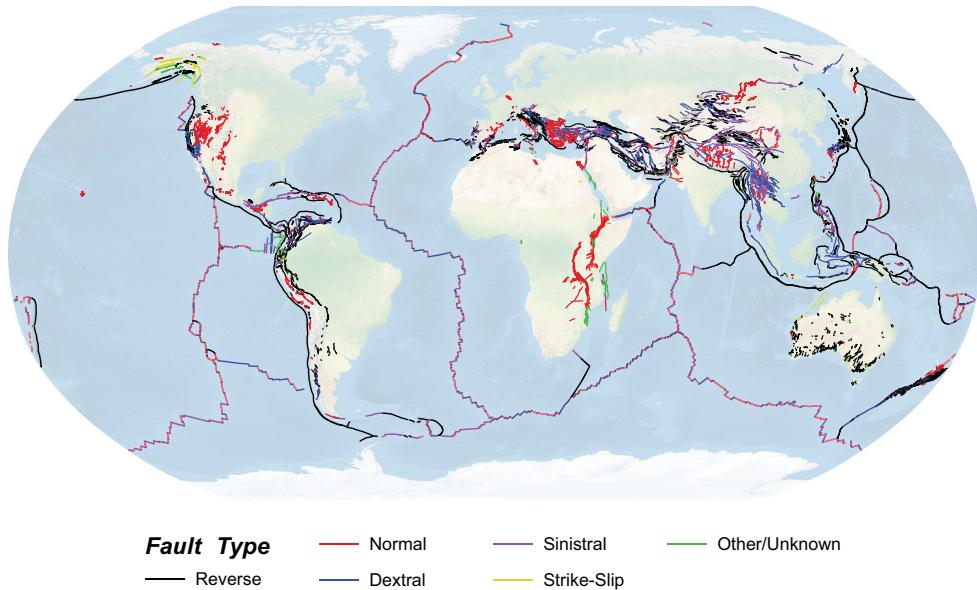


Figure 1. GEM Global Active Faults Database (v. 2019.0). Fault colors are based on kinematics; “Reverse” includes both continental reverse and thrust faults as well as subduction megathrusts; similarly, “Normal” includes both normal faults and spreading ridges, and “Dextral” and “Sinistral” include plate boundary transform faults. “Strike-slip” refers to undifferentiated strike-slip faults, where the direction of relative motion is uncertain. Note that in the database, more fault types are present (such as oblique slip types), and plate boundary and intraplate fault types are differentiated though combined here for clarity.

Introduction

The creation of a global dataset of active faults has been a major goal in seismic hazard and tectonic research for decades. Such a dataset allows for the characterization of faulting and of fault-source seismic hazard analysis throughout the globe, as well as facilitating comparisons between regions or between other global datasets (e.g. instrumental seismicity or geodesy). However, for decades, this goal has been out of reach due to both the breadth and diversity of faulting throughout the world, including many remote or poorly explored regions, as well as the technical complexity of the task and the need for multilateral international collaboration.

The Global Earthquake Model Foundation (GEM) has recently completed the first release (version 2019.0) of a global compilation of active faults, called the GEM Global Active Faults Database (GAF-DB) (Figure 1). The GAF-DB builds off of many existing regional and global fault databases (Figure 2), including GEM’s previous endeavor, the Faulted Earth database (Christophersen et al., 2015b). However, it differs substantially from many previous databases in its philosophy and technical design. The GAF-DB aims to have the most simple structure possible, while containing all of the information necessary to construct a fault source model for seismic hazard analysis (including uncertainties), and to accommodate additional information for individual faults as well. Furthermore, the GAF-DB is created to be constantly updated (rather than having a few episodic, large revisions) and is built from its constituent datasets through a build script, so that any modifications to the constituent datasets or the addition or removal of any datasets can be quickly

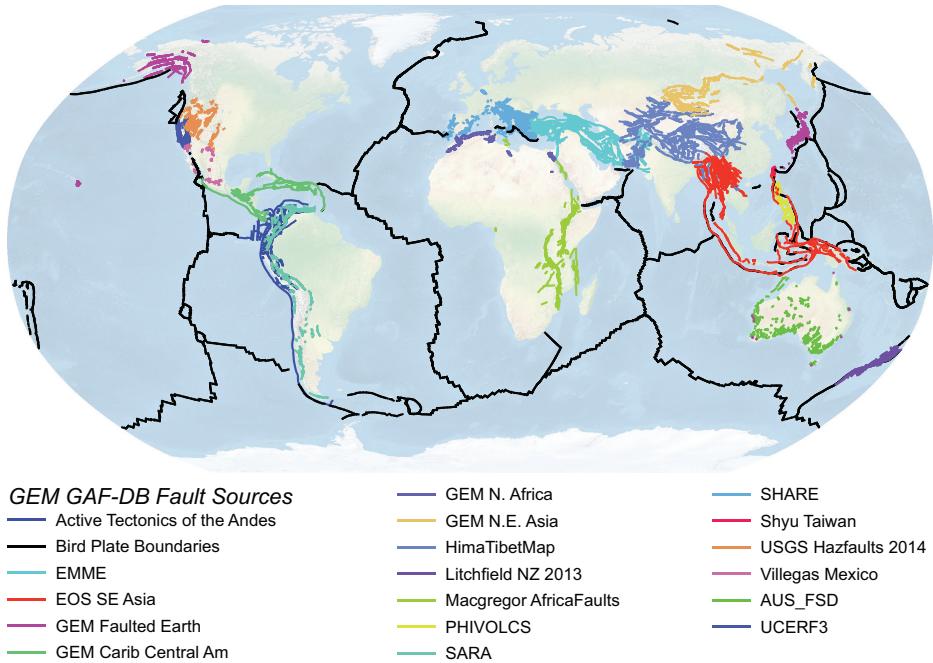


Figure 2. GEM GAF-DB, colored by source catalogs. “Active Tectonics of the Andes” from Veloza et al. (2012). “AUS_FSD” from Allen et al. (2018). “Bird Plate Boundaries” from Bird (2003). “EMME” from Danciu et al. (2018). “EOS SE Asia” from Chan et al. (2017). “GEM Faulted Earth” from Christophersen et al. (2015a). “GEM Carib Central Am” from Styron et al. (2020). “GEM N. Africa” from Styron and Poggi (2018). “GEM N.E. Asia” from Styron et al. (2018). “HimaTibetMap” from Styron et al. (2010). “Litchfield NZ 2013” from Litchfield et al. (2014). “Macgregor AfricaFaults” from Macgregor (2015). “PHIVOLCS” from Peñarubia et al. (2020). “SARA” from Alvarado et al. (2017). “SHARE” from Woessner et al. (2015). “Shyu Taiwan” from Shyu et al. (2016). “USGS Hazfaults 2014” from Petersen et al. (2014). “Villegas Mexico” from Villegas et al. (2017). “UCERF3” from Dawson and Weldon (2013).

incorporated into the GAF-DB. The GAF-DB is licensed through an open Creative Commons Attribution license (CC-BY-SA 4.0) and available on GitHub (<https://github.com/GEMScienceTools/gem-global-active-faults/>) in a variety of vector Geographic Information Systems (GIS) file formats. An interactive webmap is also available at <https://blogs.openquake.org/hazard/global-active-fault-viewer/>.

A brief but important note on nomenclature: here, we use the term “active fault” to denote a fault that is considered to be capable of producing moderate-to-large magnitude seismicity in the current tectonic conditions. Most of these faults also show geologic evidence of recent deformation (or within a specific time window such as the Quaternary), historic earthquake activity, or measurable geodetic strain accumulation. However, the evaluation of the activity of each fault is performed during the construction of each of the constituent datasets, and is not revised by us in the compilation of the GAF-DB. Therefore, there will be differences in the criteria used by the authors of each dataset, so that we may not offer one uniform, precise definition of fault activity that applies to all of the faults presented here. With this in mind, the continuity of fault representation across dataset boundaries demonstrates that, for practical purposes, the outcomes of these evaluations are quite similar regardless of the specific criteria used by the original fault mappers.

Additionally, the criteria for considering a fault “active” and/or including it in this database should be based primarily on regional considerations, rather than global consistency; the relative contribution of any structure to local or regional seismic hazard depends primarily on the tectonic context, that is, the hazard posed by other regional structures. A crustal fault with a relatively low slip rate may be the primary source of hazard in an intra-plate setting with low regional strain rates and few faults, but be of minor significance in the upper plate of a subduction zone (e.g. Halchuk et al., 2019). Therefore, fault data collection and seismic hazard analysis are most efficiently performed using regionally adapted criteria rather than global criteria (e.g. global cutoffs for a minimum slip rate or date of last surface-breaking rupture). In the GAF-DB, we prioritize regional accuracy over global consistency and trust the regional experts who assemble the constituent fault datasets to make the optimal decisions in their areas of study.

History of efforts toward a Global Active Fault compilation

The recognition of a need for active fault compilations for both geologic investigation and hazard analysis extend at least to the mid- to late-twentieth century, when the disciplines and techniques of neotectonics and paleoseismology were developing (e.g. Carlson, 1973; Sieh, 1978; Wallace, 1949; Wallace et al., 1984; Wilson et al., 1979). Work on concatenating and comparing disparate regional datasets began in the 1980s; much of the brief summary below draws from Robert Yeats, who was deeply involved. Readers interested in more detail are encouraged to read the Preface to *Active Faults of the World* (Yeats, 2012).

In the early 1980s, the first formal, international project aimed at the compilation and comparison of active faults began with the UNESCO-funded project of the International Geological Correlation Programme’s project *A Worldwide Comparison of the Characteristics of Major Active Faults*, lead by Robert Bucknam, Ding Guoyu, and Zhang Yuming. The project included meetings, field trips, and publications in various forms of fault studies and comparisons (Bucknam and Hancock, 1992), although no single fault compilation was produced.

The next major working group was Task Group II-2 of the International Lithosphere Program, initiated in 1990 under the direction of Vladimir Trifonov of the Soviet Academy of Sciences. Work in the early 1990s was hindered by geopolitical upheaval including the dissolution of the Soviet Union. However, after several years, progress began on an Eastern hemisphere compilation lead by Trifonov, a Western hemisphere compilation lead by Michael Machette of the US Geological Survey (USGS), and an oceanic compilation by G.B. Uditsev. The initial documentation of the project (Trifonov and Machette, 1993) displays an optimism that neotectonic research over the previous decade or two had produced sufficient datasets to support the assembly of the compilation by 1995 (aided by the recent developments in Geographic Information Systems (GIS) and personal computing technology), and defined a semi-technical database design or schema.

The resulting products of the next decade, however, were high-quality local to regional (national or physiographic province) compilations containing a great amount of new fault data, rather than orogen-scale or larger compilations. This highlights both the enormity of the task and the challenges and delays that are inherent in large international collaborations. The prototypical data product was the USGS *Quaternary Faults and Folds of the United States* or *QFaults* database (Haller et al., 2004; Machette et al., 2004), made available online and exportable in GIS vector formats, which proved highly influential in the style and scale of active fault mapping as well as the format and content of metadata

included. Unfortunately, the influence on subsequent work did not extend to the data format, and many of the forthcoming datasets were released as static maps, PDFs, images, or other formats that are prohibitive for automated data concatenation (Christophersen et al., 2015a) or quantitative analysis.

Around the turn of the century, the gradual improvements in personal computers, GIS technology, and high-quality regional or global digital datasets for seismicity (such as catalogs from the USGS National Earthquake Information Center, or the Global Centroid Moment Tensor project; Ekström et al., 2012), topography (e.g. Rosen et al., 2000), and GNSS geodesy allowed individual or small groups of academic researchers to produce fault data directly from digital data in a GIS program. GIS mapping is potentially much faster than with traditional methods of mapping on paper and then digitizing the results at a fixed map scale, although GIS mapping is frequently informed by traditional geologic observations. Notable contributions include the global plate boundary dataset by Bird (2003), orogen-scale active fault datasets from Michael Taylor's research group (Styron et al., 2010; Taylor and Yin, 2009; Veloza et al., 2012), and an atlas of African fault activity at different times in the Cenozoic (Macgregor, 2015). Despite being created for research into tectonic processes, these datasets are all of great utility for seismic hazard characterization in areas otherwise without data, although they may lack some of the required attributes for the creation of fault source models such as dip direction, seismogenic width, or slip rates.

With the founding of GEM, another effort toward the realization of a global collection of active faults began as the GEM Faulted Earth project (Christophersen et al., 2015b), one of GEM's Global Component data compilations. The GEM Faulted Earth project was carried out by GNS Science in New Zealand, and was ambitious and multifaceted, encompassing a sophisticated web portal and a dual database format consisting of a high-resolution neotectonic fault trace database with attributes for virtually any relevant geologic observable, and a linked, lower-resolution 3D seismic source database.

GEM also sponsored and orchestrated data collection efforts, in some instances continuing initiatives and collaborations begun under the previous projects mentioned above. For example, a major component of the South America Risk Assessment (SARA) project was the compilation of a continent-wide active fault dataset. This effort was led by Carlos Costa, who had been coordinating active fault data compilation through the earliest efforts.

Other national and international hazard projects produced active fault datasets covering large swaths of the deforming world. The Seismic Hazard Harmonization in Europe (SHARE) (Woessner et al., 2015) and Earthquake Model of the Middle East (EMME) (Danciu et al., 2018) projects characterized faulting in a consistent manner for the vast Alpine-Tethyan belt of southern Eurasia and northern Africa and Arabia, from Spain and Morocco in the west through Pakistan and Afghanistan in the east (Figure 3). Similar, smaller-scale projects by government and academic researchers contributed greatly to their respective regions (e.g. Peñarubia et al., 2020; Shyu et al., 2016; Villegas et al., 2017).

Most recently, in late 2016, GEM (now in Phase II) again attempted the completion of the global fault compilation, funded by the US Agency for International Development (USAID), leveraging the accumulating data produced by many investigators over the decades and supplemented by new mapping in regions where existing coverage was sparse or unavailable (Styron et al., 2018; Styron and Poggi, 2018). This iteration is to be released in conjunction with the GEM Global Hazard Mosaic (Pagani et al., this issue). The

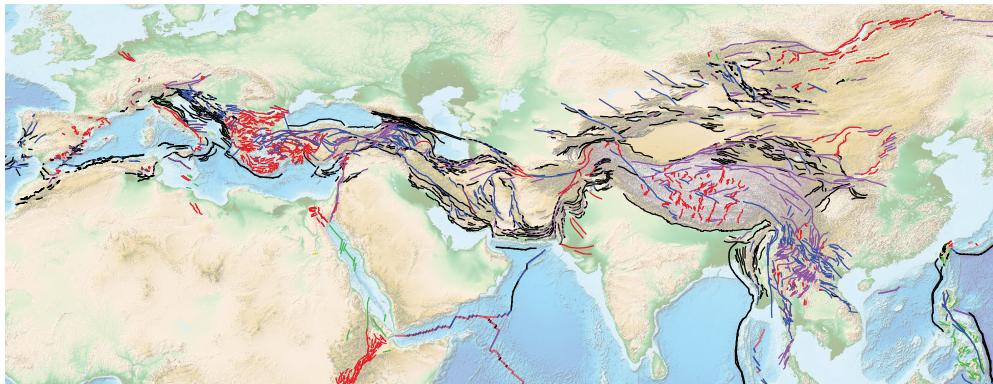


Figure 3. Faults of the southern margin of Eurasia and vicinity. Faults are colored by their kinematics, as in Figure 1.

present GEM GAF-DB is the result of this project, and represents the first attainment of the goal pursued by so many for so long. Although the objective of global coverage has been met, much of the data incorporated could (and will) be improved, and it is likely that improvements to the data concatenation, harmonization, and formatting will occur with some regularity, driven by the requirements and vision of its users.

We are extremely grateful to the many scientists before us for their countless thousands of hours of logically and technically challenging work.

Notes on the creation of this database

Due in part to our relatively attenuated timeline, interpretation of the successes of previous endeavors, as well as appreciation of the wisdom in open-source software development practices, we have chosen a strategy to build and implement the database that has substantial philosophical and technical differences from many of its predecessors.

The GEM GAF-DB is designed under the dictum, “Start simple, and add complexity as needed.” As discussed more thoroughly below, we have chosen to include just enough metadata to allow for the construction of a fault source for seismic hazard modeling (given the availability of this information for a fault) and to provide some reference information. Nonetheless, a small fraction of the faults in the database have even this set of attributes. The minimum criterion for inclusion of a fault is basically that there is a fault trace and no better representation of it or another collocated fault in other available datasets. This decision allowed us to rapidly incorporate as much data as were available, and focus later data collection efforts on the data gaps. It also keeps the resulting file sizes small, so that dissemination is easy. Current builds are about 15 MB, depending on file type. (Please note that the vast majority of faults have much more information than just the trace.)

A second guiding principle is, “Release early, release often.” This is a software design guideline that acknowledges that the structure, contents, and requirements of a body of work will change once it is in the hands of the users, and may continue to change as the needs and circumstances of the users evolve. The best response then is to continually release improvements to the product, which requires a framework that allows for updates with minimal friction. In our case, the most likely (and favorable) reasons for updating

Table I. Fault parameters in the GAF-DB

Attribute	Data type	Description (unit)	Example
Dip	Tuple	Dip (degrees)	(40,30,50)
dip_dir	String	Dip direction	W
downtthrown_side_id	String	Direction of downthrown side	NE
average_rake	Tuple	Slip rake of fault (degrees)	(45,25,55)
slip_type	String	Kinematic type	Sinistral
strike_slip_rate	Tuple	Strike-slip rate on fault (mm/year)	(1.5,0.5,2.5)
vert_sep_rate	Tuple	Vertical separation rate (mm/year)	(1.5,0.5,2.5)
shortening_rate	Tuple	Horizontal shortening rate (mm/year)	(1.5,0.5,2.5)
upper_seis_depth	Tuple	Upper limit of seismicity (km)	(0)
lower_seis_depth	Tuple	Lower limit of seismicity (km)	(15,10,25)
Accuracy	Integer	Denominator of map scale	40,000
activity_confidence	Integer	Certainty of neotectonic activity	1
exposure_quality	Integer	How well exposed (visible) fault is	2
epistemic_quality	Integer	Certainty that fault exists here	1
last_movement	String	Date of last significant earthquake	1865
name	String	Name of fault or segment	Polochic
fs_name	String	Name of fault system	Motagua-Polochic
reference	String	Paper used	Rogers and Mann (2007)
notes	String	Any relevant info	May be creeping
ogc_fid	Integer	ID used by GIS	8
catalog_id	String	Global ID	CCARA_8

GIS: Geographic Information Systems; GAF-DB: Global Active Faults Database.

See text for more explanation.

are either an update to one of the incorporated datasets, or the introduction of a new dataset that is more suited to the GAF-DB. Both scenarios are handled fluidly.

Our philosophy of database construction has informed many of the technical decisions. They are discussed in the following section.

Database structure and creation

Database schema

Each fault in the database is geometrically represented as a single, continuous fault trace with associated attributes, or metadata that characterize the geometry, kinematics, slip rate, and other features of the fault; “feature” refers to a single item in the database (Table 1). (Note that the data are primarily faults, but there are a few geologic folds as well. These are included if present in a constituent dataset, based on the criteria and classification of the dataset authors. Nonetheless, we will use the terms “fault” and “geologic structure” more or less interchangeably in this article.)

The GIS files and their associated metadata may be used to create fault sources for a probabilistic seismic hazard analysis model. Individual fault sources may be created using tools such as *FiSH* (Pace et al., 2016) or *SHERIFS* (Chartier et al., 2019). Regional fault source models, containing many faults, may be created for GEM’s OpenQuake software using utilities such as the OpenQuake Model Building Toolkit (<https://github.com/GEMScienceTools/oq-mbtk>) though documentation for the latter is currently being written.

Given the heterogeneity and incompleteness of the data, as well as issues such as the segmentation of plate boundary faults from Bird (2003), all discussed below, we do not

recommend that an earthquake occurrence or seismic hazard model be built directly from the GAF-DB at any large scale (i.e. a scale that incorporates multiple constituent datasets) without some efforts by the modeler to estimate missing values (particularly, for fault geometry and slip rate values) if this information is incomplete in the region of interest, and inspect both the harmonized and unharmonized versions of the GAF-DB to evaluate potential conflicts in the representation of faults in areas of overlapping constituent datasets. Nonetheless, in many regions of the Earth, all of the required information is present and conflicts are minimal. Furthermore, we expect continuous improvements in the completeness and integration of data in the future.

Data types. The GAF-DB uses three data types. The first is a string type, which is simply a character string, used for textual data. The second is an integer type, which is used as a database index, as the denominator of the map resolution, and as a ranked categorical variable for the fields representing qualitative epistemic uncertainty (activity_confidence, exposure_quality, epistemic_quality). In this final role, a value of 1 indicates high quality or confidence (i.e. the fault is well-exposed, or has received detailed investigation), while a value of 2 indicates low quality or confidence (the fault is buried under vegetation or has not received much study), in a regional context (i.e. these numbers cannot be compared across constituent datasets).

Continuous random variables are represented by a tuple type that has the format (most likely, minimum, maximum), or (most likely), if no meaningful uncertainty is present, or is simply not reported in the datasets as provided to us. The three-tuple format is used in order to simplify data entry during GIS mapping and to keep the overall size of the database small. Rake is in the Aki and Richards convention (Aki and Richards, 1980). Negative values for strike_slip_rate indicate the sinistral slip. vert_sep_rate values indicate the rate of vertical separation (i.e. fault throw) and are all positive; some combination of dip_dir, downthrown_side_id, or kinematics (rake or slip_type) may be used to infer whether the slip is contractional or extensional, depending on what information is present for a given structure.

The values for the upper and lower seismogenic depths may not share the same datum between different datasets; some may be relative to the sea level and some may be relative to the topographic or bathymetric surface. Users performing quantitative analysis where this may be a concern are encouraged to consult the individual sources for the datasets (and perhaps for each fault in the dataset as the datasets themselves may be inconsistent).

The last_movement column indicates the date of the last significant earthquake. This can be from the instrumental, historical, or paleoseismic record, and the criteria used by different dataset creators or maintainers may not be consistent from fault to fault or dataset to dataset.

Database construction

The GAF-DB is assembled through a Python script that reads in the constituent datasets, selects the relevant attributes for each fault and formats them for the final database, and then joins the results. Some error checking (and correction) of the fault attributes is then performed, and is followed by a fault harmonization process (described below). The initial build process is quite rapid, taking a number of seconds. The error checking and harmonization are a bit slower, but the entire process only takes a few minutes. The code to

assemble the database is in a separate repository than the database itself, but is also free and open source. It is found at <https://github.com/GEMScienceTools/gaf-processing/>. Readers with specific questions on the algorithms used in database construction and harmonization are encouraged to read the source code found there.

The build process is designed to be very configurable and modular, with a particular focus on the ease of adding new datasets and the accommodation of changes in the data within the datasets (such as the addition of new faults). The datasets may be in any vector GIS format. The Python build script is configured with a file that lists the datasets to be included, some metadata for each (e.g. the name of the dataset), and a boolean “flag” that indicates whether any formatting of the attributes will be necessary to incorporate the data into the master GAF-DB. An additional configuration file maps the column names of each dataset to the column names of the GAF-DB. For each dataset that requires some processing or formatting of the data, a Python module (a collection of functions in a single file) is written with the required functions to format the data, including a master function for each database that calls the individual processing functions. The master function is registered in the configuration file and called by the build script. A typical processing function may take maximum and minimum values for a parameter (such as fault dip), calculate the mean, and make a single tuple in the appropriate format for the GAF-DB.

Error checking is performed after the fault database is assembled. Most of the error checking is simply formatting, that is, making sure that all values in an integer column are indeed integers, and that the sorting of tuple columns is correct. Where it is possible, offending data are automatically fixed: tuples can be sorted, and minor misspellings of categorical variables (i.e. `slip_type`) can be corrected using a function that selects the best replacement value through an edit distance metric. Some “sanity checks” are performed for numerical data where appropriate, such as by checking that dip and rake values are within the ranges (0° – 90°) and (-180° to 180°), respectively. No changes are made to the underlying datasets, only to the final GAF-DB. Values that are erroneous but cannot be fixed raise warnings during the build process, but are still included in the GAF-DB (they may be optionally removed).

Database harmonization. The data undergo a geographic “harmonization” process in order to reconcile geographic conflicts between datasets; these conflicts fall under two categories, duplications and overlaps of individual faults, and overlaps between datasets. Both types of conflicts are resolved by selecting faults coming from a higher-priority fault catalog (as determined by characteristics such as map resolution or slip rate completeness) above those from a lower-priority catalog. The priority hierarchy is determined pair-wise, at the catalog level (meaning that all faults from a higher-priority dataset will supersede those from a lower-priority dataset). Individual fault duplications or overlaps are determined by selecting faults whose traces cross; the faults from the lower-priority catalog are removed. In the case of catalog-scale overlaps, such as where a more complete, higher-resolution catalog overlaps a global-scale, low-resolution catalog, all faults from the lower-priority catalog that fall even partially within a convex hull defined by the higher-priority catalog will be removed, regardless of whether individual faults overlap (a convex hull is, mathematically, the smallest polygon that encapsulates a set of objects).

Please note that both the harmonized and unharmonized versions of the GAF-DB are available in the repository and are identical except for the resolution of spatially conflicting faults in the harmonized version.

Table 2. Datasets used in the GEM GAF-DB

Coverage region	References	Dataset name (if any)
New Zealand	Litchfield et al. (2014)	None
Australia	Allen et al. (2018)	None
East Africa	Macgregor (2015)	None
Middle East	Danciu et al. (2018)	EMME
South America	Alvarado et al. (2017)	SARA
Europe	Woessner et al. (2015)	SHARE
Northern Andes	Veloza et al. (2012)	Active Tectonics of the Andes
Indo-Asian Collision Zone	Styron et al. (2010)	HimaTibetMap
Philippines	Peñarubia et al. (2020)	None
US mainland	Petersen et al. (2014)	None
California	Dawson and Weldon (2013)	UCERF3
Taiwan	Shyu et al. (2016)	None
Mexico	Villegas et al. (2017)	None
Southeast Asia	Chan et al. (2017)	None
Northeast Asia	Styron et al. (2018)	None
North Africa	Styron and Poggi (2018)	None
Central America and Caribbean	Styron et al. (2020)	CCAF
Global (various regions)	Christophersen et al. (2015a)	Faulted Earth
Global (plate boundaries)	Bird (2003)	None

EMME: Earthquake Model of the Middle East; SARA: South America Risk Assessment; SHARE: Seismic Hazard Harmonization in Europe; UCERF3: Third Uniform California Earthquake Rupture Forecast; CCAF: Caribbean and Central American Active Fault; GEM: Global Earthquake Model Foundation; GAF-DB: Global Active Faults Database.

Constituent datasets

The GAF-DB is currently assembled from 19 constituent datasets. These are all regional or national datasets with two exceptions, the global plate boundary dataset of Bird (2003) and some of the GEM Faulted Earth database (Christophersen et al., 2015a), which is itself a compilation. The datasets, their coverage regions, and references are given in Table 2.

The constituent datasets vary in the purposes for which they were created, their mapping style and resolution, and the amount and quality of metadata that are included.

Many of the datasets were created primarily for “pure” geoscience research, that is, to characterize the tectonics of regions and relate this to the forces that drive Earth deformation or the evolution of orogens; examples here include the HimaTibetMap (Styron et al., 2010; Taylor and Yin, 2009) dataset of the Indo-Asian Collision Zone (i.e. the Himalaya, Tibet, Pamir and Tien Shan mountain systems), the *Active Tectonics of the Andes* dataset (Veloza et al., 2012), data from Africa (Macgregor, 2015), and the global plate boundary dataset by Bird (2003). These datasets generally present moderate- to high-resolution fault trace mapping and have fault kinematics, although slip rates, upper and lower seismogenic depths, and other parameters of necessity for seismic hazard modeling may not be included. Nonetheless, these datasets present valuable information for much of the world.

Other datasets were created for seismic hazard analysis, and the style and metadata reflect this. The datasets from New Zealand (Litchfield et al., 2014), the Middle East (Danciu et al., 2018), South America (Alvarado et al., 2017), Europe (Woessner et al., 2015), and the US contiguous 48 states (Petersen et al., 2014) all were created as sources for the seismic hazard models of these regions. Three regional datasets were mapped by GEM to serve as the fault sources for GEM’s regional hazard models in their respective

regions, as well. These are the Caribbean and Central American Fault Database (Styron et al., 2020), the GEM North African Faults Database (Styron and Poggi, 2018), and the GEM Northeastern Asia Active Faults Database (Styron et al., 2018).

Some well-known datasets are not included in the GAF-DB because the mapping style or metadata is poorly suited for our purposes. Foremost here is the US Geological Survey's Quaternary Faults and Folds (QFaults) database, which is one of the oldest national datasets of active or potentially active fault trace, is heavily used within the geoscience community, and is the most well-cited fault database. However, the QFaults dataset is best considered to be a map of Quaternary surface deformation resulting from seismic activity, rather than a database of active faults. The database features tens of thousands of features (i.e. individual fault traces) that do not characterize their causative, bedrock seismogenic fault sources. For example, the segment of the Dixie Valley fault (Nevada, USA), which in 1954 ruptured a distance of 45 km along strike on what is thought to be a continuous fault plane (Caskey et al., 1996), is characterized as ~ 275 traces of mean length ~ 300 m. As such, an individual fault trace does not represent an individual seismic source, and therefore, the data are not ideal for hazard modeling or for inclusion in the GAF-DB. This is not to say that maps of Quaternary surface strain are not correct; they are the most accurate representations of the discontinuous and distributed nature of surface breaking over a few earthquake cycles. Nonetheless, this is a different purpose than fault source databases such as the GAF-DB, and the reason that both the GAF-DB and the US National Seismic Hazard Model (Petersen et al., 2014) use a dataset with fewer, longer fault traces.

Several datasets are present in the GAF-DB that were incorporated into the GEM Faulted Earth dataset, although not all of the faults from the Faulted Earth project are used. Those that are retained are from Japan (Active fault database of Japan, https://gbank.gsj.jp/activefault/index_e_gmap.html) and Alaska (Koehler, 2013; Koehler et al., 2012).

Global characteristics

The harmonized GAF-DB has 13,628 features (fault or fold traces) as of the writing of this article (though this number is subject to change). Both continental and oceanic faults are present. Because oceanic plates tend to be quite rigid with deformation localized at the plate boundaries, while continents typically show broad zones of deformation; therefore, oceanic faults in the GAF-DB are generally plate boundary faults, while continental faults are generally from distributed plate boundary or intraplate settings.

Spatial coverage

The GEM GAF-DB covers essentially all regions of active crustal deformation, although the spatial resolution and metadata completeness are variable. Coverage for major, known seismogenic faults can be considered complete for most of Europe, southern Asia, the United States, Central and South America, Australia, and New Zealand.

The areas with the worst coverage are Madagascar, Canada, and eastern and northern China. Very few active fault mapping projects have been performed in Madagascar or Canada. Some evidence exists for active faulting in Madagascar (Kusky et al., 2010), which is thought to be a segment of the Rovuma-Somalia plate boundary [Stamps et al., 2018]. In Canada, recent work has confirmed a Quaternary slip on the Leech River Fault

on Vancouver Island (Morell et al., 2018), and geodesy and instrumental seismicity indicate active deformation on the dextral Denali and Tintina faults as well as shortening in the Mackenzie Mountains in the Yukon and Northwest Territories (e.g. Leonard et al., 2007, 2008). A few major faults in eastern and northern China, the Tanlu Fault (e.g. Huang et al., 1996) and the Yilan-Yitong Fault (e.g. Yu et al., 2018), are suspected to be active but are not present in any of the compilations we have assembled.

Although fault coverage in Sub-Saharan Africa exists due to mapping by Macgregor (2015), the map style and metadata are sufficiently different from fault datasets assembled from hazard mapping (e.g. dip directions and slip rate are not given) that this represents a priority area for future fault mapping. GNSS geodetic coverage of the East African Rift, the source of most of the continent's seismicity, is good and many high-quality studies have characterized segments of the rift (e.g. Birhanu et al., 2016).

The final area lacking in coverage is not a single geographic region but instead a plate tectonic category: the interior of oceanic plates, including the outer rise regions adjacent to subduction zones (e.g. Naliboff et al., 2013), may rupture in very large earthquakes. For example, the 2012 M 8.6 Wharton Basin earthquake is the largest intraplate earthquake in the instrumental record (Hill et al., 2015). Although normal faults on the outer rise of the downgoing slab near the trench are evident in high-resolution bathymetry, systematic mapping of these faults would require great effort and accurate characterization of their geometries and slip rates is difficult if not impossible given available datasets. Furthermore, some large strike-slip oceanic intraplate earthquakes such as the Wharton Basin event and the 2018 Gulf of Alaska earthquake (Lay et al., 2018) seem to occur on orthogonal networks of faults, and there is no strong evidence that these faults are pre-existing or otherwise mappable. Nonetheless, the ocean basins are *terra incognita* and future generations, armed with much higher-quality bathymetry and other datasets than currently exist, may be able to make wonderful discoveries about oceanic intraplate fault networks.

Fault attributes and metadata completeness

Fault metadata included in the GAF-DB are primarily meant to characterize the geometry, kinematics, and slip rates of faults. These characteristics are informative for understanding regional deformation, and necessary for quantitative seismic hazard analysis. In this section, we describe the completeness of the metadata as it relates to the construction of a fault source model for probabilistic seismic hazard analysis (PSHA).

Kinematics. Most of the features in the database have a kinematic classification; 319 (2%) do not. Although fault kinematics categorized into the coarse “Normal,” “Sinistral,” “Dextral,” and “Reverse” groupings may be used to describe regional deformation patterns moderately effectively, hazard modeling often makes no distinctions between dextral and sinistral faults. Of the 4155 strike-slip faults, 574 (14%) are only described as “Strike-Slip” with no additional directional information.

Geometry. The surface area of a fault is a first-order control on the moment accumulation and release rate of a fault (given invariant shear modulus and slip rate), and in probabilistic seismic hazard analysis, is frequently used to constrain the maximum magnitude of earthquakes on a fault. The surface area of a fault is typically calculated as the product of the length of the fault and its down-dip width. This width cannot be directly estimated

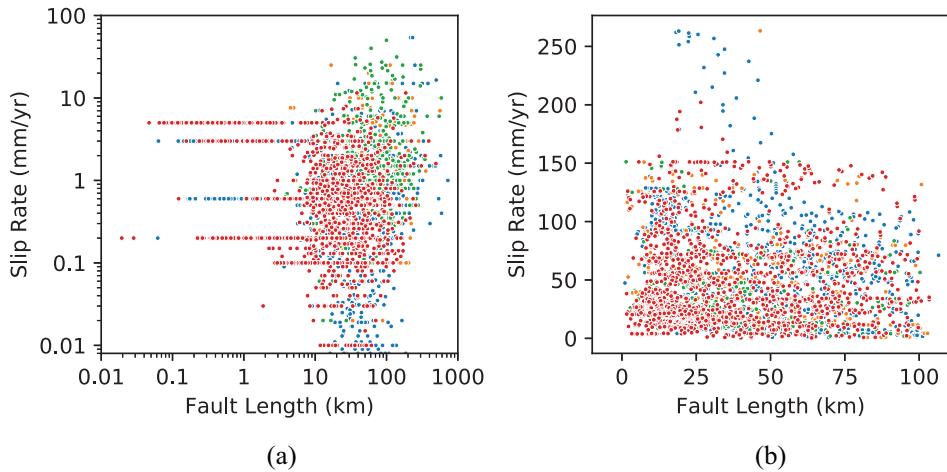


Figure 4. Slip rates and fault lengths for (a) continental faults and (b) oceanic plate boundary faults. Red = normal/spreading ridge; blue = reverse/subduction thrust; green = dextral; orange = sinistral. Note that subduction thrusts, which come from Bird (2003), are broken into segments not more than ~100 km to accommodate along-strike changes in geometry and slip rate.

without high-quality geophysical imagery or the construction of geologic cross-sections, and is therefore more commonly derived trigonometrically from the estimated values for fault dip and seismogenic thickness. Fault lengths may be measured directly from the trace under the assumption that the length is maintained at depth.

The dip of a fault may be assumed from the kinematics if it is not reported. A total of 5416 faults (40%) have dip information provided. The dip values for the rest may be considered consistent with “Andersonian” fault geometries (Anderson, 1951) or other models grounded in fault mechanics such as critical wedge theory (e.g. Dahlen, 1990). These yield vertical dips for strike-slip faults, dips of 50°–70° for normal faults, and dips of 10°–40° for reverse faults. It should be noted that large thrust decollements beneath thrust wedges may have extremely low dips and even horizontal “flats”; these values are not generally represented in the constituent datasets making up the GAF-DB.

Seismogenic thickness generally is in the range of 10–30 km for continental crust, and is thought to be a function of the geotherm in the region (e.g. Watts and Burov, 2003). The lower end of this range is in regions of elevated heat flow such as the Tibetan Plateau (e.g. Elliott et al., 2010). The upper end is in areas of colder crust such as cratons.

Fault lengths range from about 10 to 1000 km for most datasets (Figure 4) though subduction thrusts, if mapped as continuous traces, can be several times the upper length, and a few datasets that show Quaternary surface deformation rather than fault sources have faults with traces as short as 10 m. Although subduction thrusts can be thousands of kilometers in length, in the GAF-DB most of these are taken from Bird (2003), who limits the lengths of individual sections to about 100 km, so that fault dip, rake, and slip rates may vary accurately along strike.

Slip rates. Of the 13,628 features in the harmonized database, 10,526 features (77%) have some slip rate information (Figure 4).

Slip rates range over 4–5 orders of magnitude. Overall, the variation in slip rate is a function of tectonic environment. The lowest rates are about 0.01 mm/year, primarily for intraplate reverse faults in Australia (Clark et al., 2012). The highest slip rates are just above 250 mm/year, for oceanic plate boundary faults in the southeast Pacific near Samoa. Continental faults have median slip rates of 0.6 mm/year, although the highest rates are about 30–50 mm/year for continental plate boundary faults such as the San Andreas, Alpine, and Ramu Markham faults. Oceanic faults (which are all plate boundary faults in this dataset) have a low end well below 1 mm/year, but the median is ~30 mm/year, and the highest almost an order of magnitude higher. There is little doubt that if the catalog had the same degree of completeness for oceanic intraplate faults as for continental intraplate faults, the median would be lower, but as faulting is less distributed in oceanic plates and deformation is more concentrated at plate margins than in continental crust, it may not be that the median slip rates would be quite as low for an oceanic fault catalog with comparable completeness to a continental fault catalog.

Within continental or oceanic tectonic environments, the distribution of slip rates is broadly similar for all major kinematic types of faults (Figure 4), although normal faults and spreading ridges do not have maximum rates as high as other types.

It should also be noted that the slip rate completeness is much lower for continental faults than it is for oceanic plate boundary faults. The horizontal component of the relative motion of plates at plate boundary faults can be calculated directly from global plate motion circuits (Bird, 2003), assuming that the slip is taken up on a single fault, which is appropriate for our purposes. However, it is much more challenging to calculate how individual faults contribute to the total strain in zones of distributed deformation, and in many instances, each fault may require independent assessment through paleoseismological, neotectonic, or geodetic methods.

The completeness of rates in our dataset primarily varies with region and source catalog, but for those catalogs without full slip rate completeness, the completeness likely depends on the slip rate, as well, in a similar manner to completeness thresholds of seismic catalogs as a function of magnitude. Faster-slipping faults are generally more apparent in seismic, geodetic, and geomorphologic datasets and are therefore more likely to be studied. Although it is challenging to ascertain, we think it is unlikely that we are missing more than a handful of slip rates for continental faults that slip >10 mm/year, while there are surely tens to hundreds of faults that slip >1 mm/year that have yet to receive quantification of their slip rates.

Conclusion

The GEM Foundation has created the GEM GAF-DB, a compilation of existing active fault databases and new fault mapping in regions where no suitable datasets were available. The GAF-DB is the first comprehensive active fault database with worldwide coverage. The GAF-DB is public and open source with a permissive Creative Commons (CC-BY-4.0) license, and available in a variety of GIS vector formats. The GAF-DB is designed to be updated easily and regularly, as updates to the constituent datasets or new regional datasets are available.

The database contains about 13,500 individual fault traces, covering essentially all of the deforming world. The metadata for the faults provide information on the geometry, kinematics, and slip rates of the faults such that they may be used to create sources for

seismic hazard analysis, plus additional information on the origin of the data and any related notes. The completeness of the metadata is good but not total; most of the faults have kinematic and geometric information, while about two-thirds of the faults have slip rates. Because “larger” (i.e. longer, higher-displacement, or more topographically prominent) and faster-slipping faults are more obvious targets of research, we believe that the completeness of data for the most tectonically important and hazardous structures is much greater than for smaller, less tectonically important and less hazardous structures.

It is hoped that the GEM GAF-DB will be an important resource for seismic hazard modeling, geoscientific research, and education. Although the database may have a variety of applications, the primary application is seismic hazard and risk analysis. Many of the faults in the database were mapped and studied for this purpose; the mapping style and metadata were collected so that they could be used as fault sources in seismic hazard models. These data may form the template for future fault data collection and refinement, and may encourage the use of more widespread fault-based seismic hazard modeling.

The fault database may also serve as a resource for tectonics or other Earth scientific research. The data serve as a good characterization of the styles and, in many places, rates of current deformation. Researchers can build off of this for investigations into a variety of topics such as the processes that fundamentally control the style and distribution of faulting, the forces that drive deformation (e.g. Bird et al., 2008), or the interaction between faulting and Earth surface processes (e.g. Burbank and Pinter, 1999). The data may also be of interest to those outside of the fault and earthquake research communities, such as archeologists or ecologists who need to consider the effects of earthquakes and related phenomena (tsunamis or coseismic landslides, for example) on the populations that they study (e.g. Rodríguez-Pascua et al., 2013; Yang et al., 2018) and to identify potential sources for the earthquakes.

Finally, the relatively simple spatial and geometric nature of the fault data make the GAF-DB well suited for use in education. It is easy to look at the data on a map (such as the web viewer noted in the introduction, or in Google Earth) and note the spatial relationship between active faults and related features such as volcanoes and mountain belts; no expert understanding of fault data or quantitative skills are necessary for this. From our informal observations, many people’s first interaction with data is to zoom the web-map or GIS with the data to their home community to see the closest active faults; this is a natural and important use of the data that illustrates a personal connection to the data, perhaps rare in the sciences.

The open nature of the database is intended both to maximize its application in a number of domains, and to encourage user contributions. Ideally, this work will serve as a base for modification and addition by future mappers and paleoseismologists, who will no longer have to duplicate existing (but inaccessible or prohibitively licensed) datasets. A centralized but open database such as this should maximize the impact of contributions by individual scientists, by quickly and easily integrating and disseminating the contributed data.

The GEM GAF-DB is a product that is meant to be useful, the means to many ends. It is impossible for us to predict (much less prescribe) exactly what the data will be used for, and through which methods. Although we have made efforts to optimize the data schema and content of the database (particularly, for seismic hazard analysis), it is by no means final. However, we cannot improve upon this without feedback from other users, so any comments and suggestions are highly encouraged. Similarly, it is beyond our capabilities

and resources to map all faults and estimate their metadata everywhere in the world, so we are dependent on the work of others for refinements or new contributions to the existing data.

Future directions

This inaugural version of the GAF-DB is mostly spatially complete, and contains a minimal but functional metadata schema. However, as fault-based PSHA and tectonics research evolve, and the required data accumulate, the GAF-DB should evolve to support newer objectives, while still holding to the guidelines prioritizing simplicity.

For example, more complex fault geometries may be beneficial in accurately predicting seismic hazard, particularly at a local resolution (an individual site or city). Along-strike and down-dip changes in geometry could be accommodated through three-dimensional representations. Fault branching and splay faulting could also be accounted for.

More complex slip rate, earthquake recurrence, and magnitude–frequency distribution information may also prove to be useful at a global level, particularly if scientific advances lead to wide-spread and relatively homogeneous data to populate the GAF-DB (such as global geodetic block modeling; Graham et al., 2018).

Fault connectivity is a major topic in cutting-edge PSHA and tectonics research, and includes issues such as multisegment or multifault ruptures (e.g. Field et al., 2017), the transfer of slip, strain or stress between faults (e.g. Stein, 1999), and strain partitioning (e.g. Murphy et al., 2014). Efforts are being made at understanding and quantifying or parameterizing aspects of fault connectivity in a way that is suitable for inclusion in a fault database. However, this information is more properly treated as a graph (i.e. a network) from a computer science perspective, rather than the relational (i.e. tabular) data format (with each row representing a fault) that is currently used by all GIS-based fault databases, and may require a separate implementation outside of the GAF-DB.

Finally, a representation of zones of distributed strain and seismicity that is somehow more closely integrated with faults, as opposed to area sources in traditional PSHA that are often developed independently of fault data, may be beneficial. For example, these may represent areas of high strain rate (and therefore potentially high seismic hazard) but are not associated with known faults, or with faults that slip too slowly to produce the observed geodetic strain (e.g. Gold et al., 2013).

Contributions

Contributions in the form of data or constructive criticism and other feedback are quite welcome. A detailed guide to contributions is given in the GitHub repository for the database (<https://github.com/GEMScienceTools/gem-global-active-faults>). We may summarize here by stating that data contributions may be in the form of additional faults added to the GAF-DB, additions or modifications to constituent datasets maintained by GEM (data for South America, the Caribbean and Central America, North Africa, and Northeastern Asia) or by R. Styron (northern South America, the Himalaya, and Tibet), or replacement regional datasets. Modifications to regional datasets maintained by organizations other than GEM will not be considered; contributors are encouraged to contact the creators of those datasets.

Data contributed will be evaluated for inclusion based on the quality and style of mapping, the completeness of the metadata, and compatibility with existing regional data. Contributions are not guaranteed to be accepted. However, the license of the GAF-DB ensures that users may copy and modify the data to suit their purposes, provided the specific conditions of the license are met.

Any questions or contributions should be directed to hazard@globalquakemodel.org.

Acknowledgements

We are extremely grateful to the data providers who have contributed to the GEM GAF-DB. This work would not be possible without decades of research by these scientists. We also thank Peter Bird, two anonymous reviewers and the Associate Editor for their reviews, and Editor Vitor Silva for his editorial handling.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by USAID under grant AID-OFDA-G-16-00253 as well as support from GEM sponsors.

ORCID iDs

Richard Styron  <https://orcid.org/0000-0002-2374-9431>
Marco Pagani  <https://orcid.org/0000-0001-8125-1925>

References

- Aki K and Richards PG (1980) *Quantitative Seismology: Theory and Methods*. San Francisco, CA: W.H. Freeman.
- Allen TI, Griffin J, Ghasemi H, Leonard M, Clark D and Geoscience Australia (2018) *The 2018 National Seismic Hazard Assessment for Australia: Model Overview*. Canberra, ACT, Australia: Geoscience Australia.
- Alvarado A, Audemard F, Benavente Escobar C, Santibanez Boric I, Cembrano Perasso J, Costa C, Delgado Madera GF, García-Pelaez JA, Masquelin E, Minaya E, López MC, Paolini M and Perez I, Grupo de Neotectónica de SEGEMAR and Styron R (2017) The South American risk assessment active fault database. Available at: <https://github.com/GEMScienceTools/SARA-Active-Faults> (accessed 1 February 2020).
- Anderson EM (1951) *The Dynamics of Faulting and Dyke Formation with Applications to Britain*. New York: Hafner Publishing Company.
- Bird P (2003) An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems* 4(3): 1027.
- Bird P, Liu Z and Rucker WK (2008) Stresses that drive the plates from below: Definitions, computational path, model optimization, and error analysis. *Journal of Geophysical Research* 113(B11): B11406.
- Birhanu Y, Bendick R, Fisseha S, Lewi E, Floyd M, King R and Reilinger R (2016) GPS constraints on broad scale extension in the Ethiopian Highlands and Main Ethiopian Rift: Ethiopian Highlands GPS. *Geophysical Research Letters* 43(13): 6844–6851.

- Bucknam RC and Hancock PL (1992) *Major Active Faults of the World: Results of IGCP Project 206*. Rome: Consiglio Nazionale delle Ricerche.
- Burbank DW and Pinter N (1999) Landscape evolution: The interactions of tectonics and surface processes. *Basin Research* 11(1): 1–6.
- Carlson R (1973) Holocene faulting, southeastern Olympic Peninsula, Washington. *Geological Society of America* 5: 568–569.
- Caskey SJ, Wesnousky SG, Zhang P and Slemmons DB (1996) Surface faulting of the 1954 Fairview Peak (MS 7.2) and Dixie Valley (MS 6.8) earthquakes, central Nevada. *Bulletin of the Seismological Society of America* 86(3): 761–787.
- Chan CH, Wang Y, Shi X, Ornthammarath T, Warnitchai P, Kosuwan S, Thant M, Nguyen PH, Nguyen LM, Solidum R Jr, Irsyam M, Hidayati S and Sieh K (2017) Toward uniform probabilistic seismic hazard assessments for Southeast Asia. In: *AGU fall meeting abstracts*, New Orleans, LA, 11–15 December.
- Chartier T, Scotti O and Lyon-Caen H (2019) SHERIFS: Open-source code for computing earthquake rates in fault systems and constructing hazard models. *Seismological Research Letters* 90(4): 1678–1688. Available at: <https://pubs.geoscienceworld.org/ssa/srl/article-abstract/90/4/1678/570497/sherifs-open-source-code-for-computing-earthquake> (accessed 1 February 2020).
- Christophersen A, Berryman K and Litchfield NJ (2015a) The GEM faulted earth project. Technical report, GEM Foundation, Pavia.
- Christophersen A, Litchfield N, Berryman K, Thomas R, Basili R, Wallace L, Ries W, Hayes GP, Haller KM, Yoshioka T, Koehler RD, Clark D, Wolfson-Schwehr M, Boettcher MS, Villamor P, Horspool N, Ornthammarath T, Zuñiga R, Langridge RM, Stirling MW, Goded T, Costa C and Yeats R (2015b) Development of the Global Earthquake Model's neotectonic fault database. *Natural Hazards* 79(1): 111–135.
- Clark D, McPherson A and Van Dissen R (2012) Long-term behaviour of Australian stable continental region (SCR) faults. *Tectonophysics* 566–567: 1–30.
- Dahlen FA (1990) Critical taper model of fold-and-thrust belts and accretionary wedges. *Annual Review of Earth and Planetary Sciences* 18(1): 55–99.
- Danciu L, Sesetyan K, Demircioglu M, Gülen L, Zare M, Basili R, Elias A, Adamia S, Tsereteli N, Yalçın H, Utkucu M, Khan MA, Sayab M, Hessami K, Rovida AN, Stucchi M, Burg JP, Karakhanian A, Babayan H, Avanesyan M, Mammadli T, Al-Qaryouti M, Kalafat D, Varazanashvili O, Erdik M and Giardini D (2018) The 2014 earthquake model of the middle east: Seismogenic sources. *Bulletin of Earthquake Engineering* 16(8): 3465–3496.
- Dawson T and Weldon R (2013) *Geologic-slip-rate data and geologic deformation model*. Uniform California earthquake rupture forecast, version 3 (UCERF3), number 228 in California Geological Survey special report. Washington, DC: U.S. Department of the Interior.
- Ekström G, Nettles M and Dziewoński AM (2012) The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors* 200–201: 1–9.
- Elliott JR, Walters RJ, England PC, Jackson JA, Li Z and Parsons B (2010) Extension on the Tibetan plateau: Recent normal faulting measured by InSAR and body wave seismology: Extension on the Tibetan plateau. *Geophysical Journal International* 183(2): 503–535.
- Field EH, Milner KR, Hardebeck JL, Page MT, van der Elst N, Jordan TH, Michael AJ, Shaw BE and Werner MJ (2017) A Spatiotemporal clustering model for the Third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an operational earthquake forecast. *Bulletin of the Seismological Society of America* 107(3): 1049–1081.
- Gold RD, Stephenson WJ, Odum JK, Briggs RW, Crone AJ and Angster SJ (2013) Concealed quaternary strike-slip fault resolved with airborne lidar and seismic reflection: The Grizzly Valley fault system, northern Walker Lane, California: GRIZZLY Valley fault system. *Journal of Geophysical Research: Solid Earth* 118(7): 3753–3766.
- Graham SE, Loveless JP and Meade BJ (2018) Global plate motions and earthquake cycle effects. *Geochemistry, Geophysics, Geosystems* 19(7): 2032–2048.
- Halchuk S, Allen T, Adams J and Onur T (2019) Contribution of the Leech River Valley—Devil's Mountain fault system to seismic hazard in Victoria, BC. In: *12th Canadian conference on earthquake engineering*, Quebec City, QC, Canada, 17–20 June.

- Haller KM, Machette MN, Dart RL and Rhea BS (2004) U.S. Quaternary fault and fold database released. *Eos: Transactions American Geophysical Union* 85(22): 218–218.
- Hill EM, Yue H, Barbot S, Lay T, Tapponnier P, Hermawan I, Hubbard J, Banerjee P, Feng L, Natawidjaja D and Sieh K (2015) The 2012 Mw 8.6 Wharton Basin sequence: A cascade of great earthquakes generated by near-orthogonal, young, oceanic mantle faults. *Journal of Geophysical Research: Solid Earth* 120(5): 3723–3747.
- Huang W, Gao W and Ding G (1996) Neogene volcanism and Holocene earthquakes in the Tanlu fault zone, eastern China. *Tectonophysics* 260(4): 259–270.
- Koehler RD (2013) *Quaternary Faults and Folds (QFF)*. Technical report DDS 3, Fairbanks, AK: Alaska Division of Geological & Geophysical Surveys. Available at: <http://www.dggs.alaska.gov/pubs/id/24956> (accessed 1 February 2020).
- Koehler RD, Farrell RE, Burns PAC and Combellick RA (2012) *Quaternary faults and folds in Alaska: A digital database*. Technical report MP 141. Fairbanks, AK: Alaska Division of Geological & Geophysical Surveys. Available at: <http://www.dggs.alaska.gov/pubs/id/23944> (accessed 1 February 2020).
- Kusky TM, Toraman E, Raharimahefa T and Rasoazanamparany C (2010) Active tectonics of the Alaotra–Ankay Graben System, Madagascar: Possible extension of Somalian–African diffusive plate boundary? *Gondwana Research* 18(2): 274–294.
- Lay T, Ye L, Bai Y, Cheung KF and Kanamori H (2018) The 2018 MW 7.9 Gulf of Alaska earthquake: Multiple fault rupture in the Pacific plate. *Geophysical Research Letters* 45(18): 9542–9551.
- Leonard LJ, Hyndman RD, Mazzotti S, Nykolaishen L, Schmidt M and Hippchen S (2007) Current deformation in the northern Canadian Cordillera inferred from GPS measurements. *Journal of Geophysical Research* 112(B11): B11401.
- Leonard LJ, Mazzotti S and Hyndman RD (2008) Deformation rates estimated from earthquakes in the northern Cordillera of Canada and eastern Alaska: NORTHERN Cordillera seismic deformation. *Journal of Geophysical Research: Solid Earth* 113(B8): B08406.
- Litchfield NJ, Dissen RV, Sutherland R, Barnes PM, Cox SC, Norris R, Beavan RJ, Langridge R, Villamor P, Berryman K, Stirling M, Nicol A, Nodder S, Lamarche G, Barrell DJA, Pettinga JR, Little T, Pondard N, Mountjoy JJ and Clark K (2014) A model of active faulting in New Zealand. *New Zealand Journal of Geology and Geophysics* 57(1): 32–56.
- Macgregor D (2015) History of the development of the East African Rift System: A series of interpreted maps through time. *Journal of African Earth Sciences* 101: 232–252.
- Machette MN, Haller KM and Wald LA (2004) *Quaternary Fault and Fold Database for the Nation*. Washington, DC: US Department of the Interior, US Geological Survey.
- Morell KD, Regalla C, Amos C, Bennett S, Leonard L, Graham A, Reedy T, Levson V and Telka A (2018) Holocene surface rupture history of an active forearc fault redefines seismic hazard in Southwestern British Columbia, Canada. *Geophysical Research Letters* 45(21): 11,605–11,611.
- Murphy MA, Taylor MH, Gosse J, Silver CRP, Whipp DM and Beaumont C (2014) Limit of strain partitioning in the Himalaya marked by large earthquakes in western Nepal. *Nature Geoscience* 7(1): 38–42.
- Naliboff JB, Billen MI, Gerya T and Saunders J (2013) Dynamics of outer-rise faulting in oceanic-continental subduction systems. *Geochemistry, Geophysics, Geosystems* 14(7): 2310–2327.
- Pace B, Visini F and Peruzza L (2016) FiSH: MATLAB tools to turn fault data into seismic-hazard models. *Seismological Research Letters* 87(2A): 374–386.
- Peñarubia C, Johnson KL, Styron RH, Bacolcol TC, Sevilla WIG, Perez JS, Bonita JD, Narag IC, Solidum RU Jr, Pagani MM and Allen TI (2020) Probabilistic seismic hazard analysis model for the Philippines. *Earthquake Spectra*. Epub ahead of print 5 March. DOI: 10.1177/8755293019900521.
- Petersen MD, Moschetti MP, Powers PM, Mueller CS, Haller KM, Frankel AD, Zeng Y, Rezaeian S, Harmsen S, Boyd O, Field EH, Chen R, Rukstales KS, Luco N, Wheeler RL, Williams RA and Olsen AH (2014) *Documentation for the 2014 update of the United States national seismic hazard maps*. Open-file report 2014-1091. Reston, VA: U.S. Geological Survey.

- Rodríguez-Pascua M, Abad Casal L, Pérez-López R, Gamo Parra B, Silva P, Garduño-Monroy V, Giner-Robles J, Perucha M, Israde-Alcántara I, Bischoff J and Calvo Serando JP (2013) Roman, Visigothic and Islamic evidence of earthquakes recorded in the archaeological site of “El Tolmo de Minateda” (Prebetic Zone, southeast of Spain). *Cuaternario y geomorfología* 27(3–4): 5–32.
- Rogers RD and Paul M (2007) Transtensional deformation of the western Caribbean-North America plate boundary zone. *SPECIAL PAPERS-GEOLOGICAL SOCIETY OF AMERICA* 428: 37.
- Rosen PA, Hensley S, Joughin IR, Li FK, Madsen SN, Rodriguez E and Goldstein RM (2000) Synthetic aperture radar interferometry. *Proceedings of the IEEE* 88(3): 333–382.
- Shyu JBH, Chuang YR, Chen YL, Lee YR and Cheng CT (2016) A new on-land seismogenic structure source database from the Taiwan Earthquake Model (TEM) project for seismic hazard analysis of Taiwan. *Terrestrial, Atmospheric and Oceanic Sciences* 27(3): 311–323. Available at: <http://tao.cgu.org.tw/index.php/articles/archive/geophysics/item/1376> (accessed 1 February 2020).
- Sieh KE (1978) Prehistoric large earthquakes produced by slip on the San Andreas Fault at Pallett Creek, California. *Journal of Geophysical Research: Solid Earth* 83(B8): 3907–3939.
- Stamps DS, Saria E and Kreemer C (2018) A geodetic strain rate model for the East African Rift System. *Scientific reports* 8(1): 1–8.
- Stein RS (1999) The role of stress transfer in earthquake occurrence. *Nature* 402(6762): 605–609. Available at: <https://pdfs.semanticscholar.org/b380/0d6f107fdfa7011a9b9b66e39c94f150437c.pdf> (accessed 1 February 2020).
- Styron R and Poggi V (2018) GEM North Africa active fault database. Available at: https://github.com/GEMScienceTools/n_africa_active_faults (accessed 1 February 2020).
- Styron R, Garcia-Pelaez J and Pagani M (2020) CCAF-DB: The Caribbean and Central American active fault database. *Natural Hazards and Earth System Science* 20: 831–857.
- Styron R, Poggi V and Lunina OV (2018) GEM Northeastern Asia active fault database. Available at: <https://github.com/GEMScienceTools/ne-asia-active-faults> (accessed 1 February 2020).
- Styron R, Taylor M and Okoronkwo K (2010) Database of active structures from the Indo-Asian collision. *Eos: Transactions American Geophysical Union* 91(20): 181–182.
- Taylor M and Yin A (2009) Active structures of the Himalayan-Tibetan orogen and their relationships to earthquake distribution, contemporary strain field, and Cenozoic volcanism. *Geosphere* 5(3): 199–214.
- Trifunov VG and Machette MN (1993) The world map of major active faults project. *Annals of Geophysics* 36(3–4): 225–236.
- Veloza G, Styron R, Taylor M and Mora A (2012) Open-source archive of active faults for northwest South America. *GSA Today* 22(10): 4–10.
- Villegas GC, Mendoza C and Ferrari L (2017) Mexico quaternary fault database. *Terra Digitalis* 1(1): 1–9.
- Wallace RE (1949) Structure of a portion of the San Andreas rift in Southern California. *GSA Bulletin* 60(4): 781–806. Available at: <http://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/60/4/781/4295/structure-of-a-portion-of-the-san-andreas-rift-in> (accessed 1 February 2020).
- Wallace RE, Bonilla MG and Villalobos H (1984) *Faulting related to the 1915 earthquakes in Pleasant Valley, Nevada*. USGS numbered series 1274-A-B. Washington, DC: U.S. Government Publishing Office. Available at: <http://pubs.er.usgs.gov/publication/pp1274AB> (accessed 1 February 2020).
- Watts AB and Burov EB (2003) Lithospheric strength and its relationship to the elastic and seismogenic layer thickness. *Earth and Planetary Science Letters* 213(1): 113–131.
- Wilson JR, Bartholomew MJ and Carson RJ (1979) Late Quaternary faults and their relationship to tectonism in the Olympic Peninsula, Washington. *Geology* 7(5): 235–239. Available at: <http://geology.gsapubs.org/content/7/5/235.short> (accessed 1 February 2020).
- Woessner J, Laurentiu D, Giardini D, Crowley H, Cotton F, Grünthal G, Valensise G, Arvidsson R, Basili R, Demircioglu MB, Hiemer S, Meletti C, Musson RW, Rovida AN, Sesetyan K and

- Stucchi M and The SHARE Consortium (2015) The 2013 European seismic hazard model: Key components and results. *Bulletin of Earthquake Engineering* 13(12): 3553–3596.
- Yang W, Qi W and Zhou J (2018) Decreased post-seismic landslides linked to vegetation recovery after the 2008 Wenchuan earthquake. *Ecological Indicators* 89: 438–444.
- Yeats R (2012) *Active Faults of the World*. Cambridge: Cambridge University Press.
- Yu Z, Zhang P, Min W, Wei Q and Liu Y (2018) Late Holocene slip rate and average recurrence interval of great earthquakes on the Shangzhi segment of the Yilan-Yitong Fault Zone, northeastern China: Constraints from paleo-earthquakes and historical written records. *Island Arc* 27(1): e12231.