We present the Malawi Active Fault Database (MAFD), an open-access (https://doi.org/10.5281/zenodo.5507190) geospatial database of 113 fault traces in Malawi and neighboring Tanzania and Mozambique. Malawi is located within the East African Rift's (EAR) Western Branch where active fault identification is challenging because chronostratigraphic data are rare, and/or faults are buried and so do not have a surface expression. The MAFD therefore includes any fault that has evidence for displacement during Cenozoic East African rifting or is buried beneath the rift valley and is favorably oriented to the regional stresses. To identify such faults, we consider a multidisciplinary data set: high-resolution digital elevation models, previous geological mapping, field observations, seismic reflection surveys from offshore Lake Malawi, and aeromagnetic and gravity data. The MAFD includes faults throughout Malawi, where seismic risk is increasing because of population growth and its seismically vulnerable building stock. We also investigate the database as a sample of the normal fault population in an incipient continental rift. We cannot reject the null hypothesis that the distribution of fault lengths in the MAFD is described by a power law, which is consistent with Malawi's relatively thick seismogenic layer (30–40 km), low (<8%) regional extensional strain, and regional deformation localization (50%–75%) across relatively long hard-linked border faults. Cumulatively, we highlight the importance of integrating onshore and offshore geological and geophysical data to develop active fault databases along the EAR and similar continental settings both to understand the regional seismic hazard and tectonic evolution.

Plain Language Summary Earthquakes represent the occurrence of slip along cracks in the Earth's crust. Therefore, mapping these cracks, or “faults,” is important when assessing earthquake hazards. However, faults are challenging to identify as they may not be visible at the surface. Fault mapping also requires recognizing which faults have slipped in earthquakes in the recent geologic past, as these “active” faults are the most likely faults to have future earthquakes. Here, we describe how we identified active faults in Malawi, which is located along the tectonically active East African Rift. Faults under Lake Malawi were mapped using images of lake sediments that were generated from sound waves. Onshore, some faults were mapped from their expression in the landscape. Other faults, not visible at the surface, were identified from aeromagnetic data, which image the spatial distributions of magnetic minerals in the Earth's crust. Faults are considered active if that show evidence for slip during East African rifting in Malawi. We combined the active faults identified from these analyses into the Malawi Active Fault Database, a freely available geospatial database. We suggest that this database will be useful for seismic hazard planning in Malawi, where population growth and vulnerable buildings are increasing earthquake risk.

1. Introduction Systematically mapping active faults and collating their geomorphic attributes into an active fault database provide an important tool for assessing regional seismic hazard and tectonic evolution (Faure Walker et al., 2021; Langridge et al., 2016; Maldonado et al., 2021; Styron & Pagani, 2020; Williams, Mdala, et al., 2021). In...
particular, there is a critical need to develop active fault databases along the Western Branch of the East African Rift (EAR), as population growth and the development of seismically vulnerable building stock are raising seismic risks in this region (Gođa et al., 2016, 2021; Meghraoui et al., 2016; Ngoma et al., 2019; Novelli et al., 2019; World-Bank, 2019). Furthermore, assessment of the Western Branch's seismic hazard typically considers the ~70-year-long instrumental seismic record alone (Midzi et al., 1999; Poggi et al., 2017; Tuluka et al., 2020); however, low EAR extension rates (0.5–2 mm/yr; Stamps et al., 2021; Wedmore et al., 2021) imply long earthquake repeat times, and so this is likely an incomplete record of earthquakes. Hence, seismic hazard assessment in the Western Branch should also consider other data, such as geodesy and active fault mapping (Hartnady, 2002; Hodge et al., 2015; Meghraoui et al., 2016; Williams, Mdala, et al., 2021).

By defining active fault databases in the EAR Western Branch to include all faults that have been active during East African rifting, these databases can also be used to investigate continental rift normal fault populations. Such studies are important because fault lengths in continental rifts are thought to evolve from a power law to an exponential distribution with an increasing regional extensional strain (>8%–12%) as relatively short faults link together or become inactive (A. Gupta & Scholz, 2000; Cowie et al., 1995; Hardacre & Cowie, 2003; Meyer et al., 2002; Michas et al., 2015; Shmela et al., 2021). However, this transition may be affected by preexisting crustal heterogeneities and the thickness of the seismogenic crust (Ackermann et al., 2001; Hardacre & Cowie, 2003; Soliva & Schulz, 2008; Walsh et al., 2002). The EAR Western Branch can place important constraints on this fault-length distribution evolution as (a) it has accommodated relatively small regional extensional strains during East African rifting (<15%; Accardo et al., 2020; Scholz et al., 2020; Wright et al., 2020) and so provides information about fault populations at a relatively early stage of rift evolution; (b) fault geometry may have been modulated by favorably oriented crustal fabrics that were imparted by successive Proterozoic orogenic events (Hodge, Fagereng, Biggs, et al., 2018; Katumwehe et al., 2015; Morley, 2010; Ring, 1994; Wedmore, Williams, et al., 2020; Williams et al., 2019); and (c) compared to the ~10- to 20-km-thick seismogenic layer in typical continental lithosphere (e.g., J. Jackson et al., 2021) earthquakes may nucleate through the 30- to 40-km-thick crust in the Western Branch (Craig & Jackson, 2021; Ebinger et al., 2019; Fagereng, 2013; Foster & Jackson, 1998; Lavayssière et al., 2019; Lindenfeld et al., 2012; Nyblade & Langston, 1995; Stevens et al., 2021), and possibly even into the upper lithospheric mantle (Yang & Chen, 2010).

There are many intrinsic challenges in locating and mapping active faults because of processes, such as scarp degradation, sedimentation, or because faults are buried or offshore (Nicol et al., 2016; Perea et al., 2006; Wallace, 1980), and these challenges are particularly pertinent in the EAR Western Branch (Williams, Mdala, et al., 2021). For example, the thick seismogenic layer means that earthquakes on active faults are less likely to propagate to the surface as demonstrated by examples of et al., 2021). For example, the thick seismogenic layer means that earthquakes on active faults are less likely to propagate to the surface as demonstrated by examples of (e.g., J. Jackson et al., 2021) earthquakes may nucleate through the 30- to 40-km-thick crust in the Western Branch (Craig & Jackson, 2021; Ebinger et al., 2019; Fagereng, 2013; Foster & Jackson, 1998; Lavayssière et al., 2019; Lindenfeld et al., 2012; Nyblade & Langston, 1995; Stevens et al., 2021), and possibly even into the upper lithospheric mantle (Yang & Chen, 2010).

Extension in the EAR Western Branch combined with a favorable hydroclimate has also resulted in the formation of several rift-axial lakes that have flooded the rift valleys and obscured the surface traces of active faults (Figure 1). In active fault databases from other offshore regions, seismic reflection and/or high-resolution (spatial accuracy <10 m) bathymetric data have been used to identify and map offshore active faults (Gracia et al., 2003; Langridge et al., 2016; Marlow et al., 2000; Pondard & Barnes, 2010; Styron et al., 2020). However, Lake Kivu, which represents <2% of the total area of the EAR Lakes, is the only lake with modern high-resolution bathymetric data (Ross et al., 2014; Wood et al., 2017) and although many of the lakes are covered by seismic reflection surveys (Karp et al., 2012; McGlue et al., 2006; Muirhead et al., 2019; Scholz et al., 2020), faults identified in these surveys are not typically incorporated into seismic hazard assessment. Nevertheless, the inclusion of offshore faults into active fault databases is critical as in addition to ground shaking, they also present secondary seismic hazards, such as earthquake-triggered landslides and near-field tsunami (Bardet et al., 2003; Masson et al., 2006).

In this study, we present the Malawi Active Fault Database (MAFD), which attempts to address the challenges of mapping active faults in the Western Branch of the EAR. The MAFD combines: (a) the South Malawi Active Fault Database (SMAFD; Williams, Mdala, et al., 2021), (b) offshore faults below Lake Malawi, which were mapped from available 2D seismic reflection surveys (Scholz et al., 2020; Shillington et al., 2020), (c) onshore faults that
have been previously identified in central and northern Malawi (Crossley, 1984; Ebinger et al., 1987; Harrison & Chapusa, 1975; Peters, 1975; Ray, 1975) and that have been remapped here using high-resolution digital elevation models (DEMs), and (d) buried intrarift faults inferred from aeromagnetic (Kolawole et al., 2018a, 2021) or gravity data (Chisenga et al., 2019) that are favorably oriented to the regional stresses.

Except for the Kivu Rift (Delvaux et al., 2017; Wood et al., 2017), onshore-offshore active fault databases have not been developed within the Western Branch. The strategies employed to map faults in the MAFD are therefore relevant elsewhere along the rift system and in other regions with onshore and offshore active faults. Furthermore, the systematic compilation of 113 EAR fault traces in the MAFD provides a data set to assess the population of normal faults in a low-strain continental rift that broadly follows preexisting crustal weaknesses and is hosted in a thick seismogenic layer.

2. Malawi Seismotectonics

Malawi is located near the southern end of the Western Branch of the EAR, where the rift accommodates 0.5–2 mm/yr of ENE-WSW extension between the San and Rovuma plates (Figure 1; Wedmore et al., 2021). The EAR in Malawi has mainly developed within Proterozoic greenschist to granulite facies metamorphic terranes and shear zones that bound Archean cratons (Figure 2), and that formed and evolved during the incremental assemblage of the African continent (Fritz et al., 2013; Lenoir et al., 1994; Manda et al., 2019; Ring, 1993). Cumulatively, these events imparted gently to steeply dipping, NE- to NW-striking metamorphic fabrics in a range of granulite, schistose, paragneiss, pelite, and calc-silicate lithologies (Figure 2b; e.g., Bloomfield, 1965; Dulanya, 2017; Fullgraf et al., 2017; Ray, 1975; Ring, 1994). Locally, these fabrics are well-oriented for reactivation under the region's ENE-trending minimum principal compressive stress, and so they may have guided the trajectory of some EAR faults in Malawi (Hodge, Fagereng, Biggs, et al., 2018; Kolawole et al., 2018a; Ring, 1994; Scholz et al., 2020; S. M. Dawson et al., 2018; Wedmore, Williams et al., 2020; Williams et al., 2019). These metamorphic fabrics may have also influenced regional Upper Permian to Lower Jurassic “Karoo” rifting (Catuneanu et al., 2005; Key et al., 2007; Wopfner, 2002). In central and northern Malawi, the EAR cuts across Karoo-age NW-SE trending basins (Accardo et al., 2018; Ring, 1994; Versfelt & Rosendahl, 1989), while in southern Malawi, Karoo-age
faults in the NW-SE trending Shire Rift Zone have likely been reactivated during EAR deformation (Figure 2; Castaing, 1991; Habgood, 1963; Kolawole et al., 2021; Wedmore, Williams, et al., 2020).

The late Oligocene/early Miocene age of the Rungwe Volcanic Province in southern Tanzania provides an upper estimate for the onset of EAR activity in northern Malawi (Mesko, 2020; Mortimer et al., 2016; Rasskazov et al., 2001; Roberts et al., 2012). To the south, the age of the EAR is poorly constrained with a Late Miocene-Pliocene age proposed for the central and southern basins of Lake Malawi from extrapolating modern depositional rates (Delvaux, 1995; McCartney & Scholz, 2016; Scholz et al., 2020). A southward propagation of the EAR in Malawi is also consistent with the thinner sedimentary cover and smaller escarpment heights in southern Malawi (Laõ-Dávila et al., 2015; Wedmore, Biggs, et al., 2020). South of the Rungwe Volcanic Province, there has been no reported surface volcanism in the EAR, and only negligible amounts of melt are inferred in Malawi’s lower crust and lithospheric mantle (Accardo et al., 2017, 2020; Hopper et al., 2020; Njinju et al., 2019; Wang et al., 2019).

In Malawi, the EAR can be divided along strike into several 50- to 200-km-long basins. Each basin contains one or more rift-bounding border faults that are commonly asymmetric and are linked by high-relief accommodation zones (Figure 2a; Accardo et al., 2018; Ebinger et al., 1987; Laõ-Dávila et al., 2015; McCartney & Scholz, 2016; Scholz, 1989; Scholz et al., 2020; Wedmore, Williams, et al., 2020; Williams, Mdala, et al., 2021). Lake Malawi
has flooded the three most northern EAR basins in Malawi (Scholz et al., 2020), while to the south, the rift valley is onshore and channels the Shire River, Lake Malawi’s only outlet, toward its confluence with the Zambezi River (Dixey, 1926; Dulanya, 2017; Ivory et al., 2016; Williams, Mdala, et al., 2021).

Historical and instrumental records of seismicity in Malawi extend at least until the 1920s (Ambraseys & Adams, 1991; Bloomfield, 1958; Dixey, 1926; Poggi et al., 2017). To the best of our knowledge, only one active fault in Malawi has exhibited coseismic surface rupture in this period, the St Mary Fault during the 2009 Karonga Earthquake sequence (Biggs et al., 2010; Gaherty et al., 2019; Hamiel et al., 2012; Kolawole et al., 2018a, 2018b; Macheyeki et al., 2015). The instrumental record for Malawi is complete for events $M > 4.5$ since 1965 with the largest event in this record being the 1989 $M_w$ 6.3 Salima Earthquake (Hodge et al., 2015; J. Jackson & Blenkinsop, 1991; Lemenkova, 2021; Poggi et al., 2017). However, the length and downdip width of faults in Malawi have been previously used to infer that they may host infrequent $M_w$ 6.5–7.8 earthquakes (Hodge et al., 2020; J. Jackson & Blenkinsop, 1997; Shillington et al., 2020; Wedmore, Biggs, et al., 2020; Williams, Mdala, et al., 2021), and earthquakes of these magnitudes have occurred in regions adjacent to Malawi, such as the 1910 M 7.4 Rukwa Earthquake in southern Tanzania (Ambraseys, 1991; Kervyn et al., 2006; Vittori et al., 1997) and 2006 $M_w$ 7.0 Machaze Earthquake in Mozambique (Copley et al., 2012; Fenton & Bommer, 2006).

3. The Malawi Active Fault Database

3.1. The MAFD Fault Mapping Strategy

The MAFD is a geospatial database of fault traces that we interpret as active. Seismic hazard planning is typically considered at the national level. Therefore, the MAFD is intended to cover all active faults within Malawi and those adjacent to its borders in Mozambique and Tanzania that may also contribute to seismic hazards. This definition closely follows the geological region of the “Malawi Rift” or “Nyasa Rift”; however, in some studies, the Shire Rift Zone in southern Malawi (Figure 2) is considered to represent an EAR section distinct from the Malawi Rift (Castaing, 1991; Kolawole et al., 2021). Therefore, to avoid confusion, we do not consider these geological subdivisions further. Possible active faults within 20 km of Malawi in the Luangwa Rift in eastern Zambia (Figure 2; Daly et al., 2020) and the Rukwa Rift in southern Tanzania (Delvaux et al., 2012; Fontijn et al., 2010; Kervyn et al., 2006) are not included in the MAFD.

Following the template used in the Global Earthquake Model Global Active Fault Database (Styron & Pagani, 2020), faults in the MAFD, including those that show branching geometry, are mapped as a single geographic information system (GIS) feature. For each fault, a number of attributes are assigned that detail its geomorphic attributes and confidence that it is active (Table 1). No assessment of the seismogenic properties of faults (e.g., earthquake magnitudes, recurrence intervals, segmentation, and degree of seismic coupling) is made in the MAFD. We do so as these parameters are often based on subjective judgments and simplified fault geometries, and so they should be compiled separately from the observational data contained within an active fault database (Faure Walker et al., 2021; Styron et al., 2020).

The criteria used to determine if a fault is “active” vary between regional active fault databases (Styron & Pagani, 2020). For example, only faults with evidence for displacement within the last 125 ka are included in New Zealand Active Fault Database (Langridge et al., 2016), while in the USA, the equivalent criteria are evidence for displacement in the Quaternary (i.e., the last 2.6 million years; Machette et al., 2004), and in Australia, which is a stable continental region, displacements within the last 5 million years (D. Clark et al., 2012). Using an age-based criteria for determining fault activity is further complicated as (a) faults capable of hosting future earthquakes will not necessarily preserve chronostratigraphic evidence for relatively recent earthquakes (Cox et al., 2012; King et al., 2019; Nicol et al., 2016; Perea et al., 2006), and (b) even in regions with well-developed active fault databases, in situ chronostratigraphic data are not available for most faults (T. E. Dawson et al., 2013); instead, their activity is inferred from correlating offset surfaces with surrounding chronostratigraphic surfaces (e.g., marine terraces and glacial outwash surfaces). For these reasons, some active fault compilations now assign varying degrees of confidence to whether a fault is active and/or include faults based on the factors, such as displacement within the current tectonic regime, large total displacements, orientation to the regional stresses, or proximity to historical earthquakes (Faure Walker et al., 2021; Jomard et al., 2017; Maldonado et al., 2021; Van Dissen et al., 2021).
With regard to active fault databases in the Western Branch of the EAR, the above discussion indicates that although there are limited chronostratigraphic data, this is not necessarily a barrier to defining and mapping “active” faults (Delvaux et al., 2017; Meghraoui et al., 2016; Williams, Mdala, et al., 2021). In the MAFD, we therefore first define faults as active if they have accommodated displacement in the current tectonic regime. In the context of Malawi, this implies activity since the onset of EAR-related ENE-WSW extension (Williams, Mdala, et al., 2021). Hence, it is not the absolute ages of fault displacements or rifting that are the key criteria for developing the MAFD, but the age of fault displacements relative to the onset of East African rifting in Malawi (Figure 3a). The evidence that we use to determine if fault displacements meet this criterion include steep scarps, offset alluvial fans, incised footwall drainage channels, and/or the offset or accumulation of EAR sediment in a fault’s hanging wall (Figures 4 and 5; Hodge et al., 2019, 2020; J. Jackson & Blenkinsop, 1997; Shillington et al., 2020; Wedmore, Biggs, et al., 2020; Wedmore, Williams, et al., 2020; Williams, Mdala, et al., 2021). In these cases, a fault in the MAFD is assigned a confidence level of 1 (Figure 3a, Table 1).

Faults buried beneath rift sediments in Malawi have also been previously inferred from gravity and aeromagnetic data (Chisenga et al., 2019; Kolawole et al., 2018a, 2021). Where these faults are favorably oriented to the regional stresses (Williams et al., 2019), we interpret that they are active and so also include them in the MAFD, but the age of fault displacements relative to the onset of East African rifting in Malawi (Figure 3a). The evidence that we use to determine if fault displacements meet this criterion include steep scarps, offset alluvial fans, incised footwall drainage channels, and/or the offset or accumulation of EAR sediment in a fault’s hanging wall (Figures 4 and 5; Hodge et al., 2019, 2020; J. Jackson & Blenkinsop, 1997; Shillington et al., 2020; Wedmore, Biggs, et al., 2020; Wedmore, Williams, et al., 2020; Williams, Mdala, et al., 2021). In these cases, a fault in the MAFD is assigned a confidence level of 1 (Figure 3a, Table 1).

3.2. MAFD Availability
The MAFD is a freely available open-source database issued under a Creative Commons CC-BY-4.0 license. The database is available on the Zenodo Data Archive (https://doi.org/10.5281/zenodo.5507190) and on GitHub.
The Global Earthquake Model Global Active Faults Database (Styron & Pagani, 2020) established the principle that fault databases should be published but also updated as new information becomes available. Future releases of the MAFD will appear on both Zenodo and GitHub with the redundant versions also archived. On GitHub, previous versions can also be tracked, and pull requests can be made by other users. We therefore encourage users to consult these pages for future updates to the MAFD. Further information is provided in the Data Availability section.

Figure 3. (a) Decision tree applied to determine if faults mapped from the various available data sets in Malawi are included in the Malawi Active Fault Database (MAFD) or Malawi Other Faults Database (MOFD). (b) Venn diagram to represent the relationship between the MAFD and MOFD, and faults in Malawi that are truly active and inactive. Note that the size of circles and overlap is schematic and does not represent the true (unknown) degree of overlap between these databases and faults in Malawi.
3.3. Data Sets Used for Mapping Faults in Malawi

3.3.1. Geological Maps and High-Resolution Digital Elevation Models

In south Malawi, active faults identified from fieldwork, high-resolution DEMs, and geological maps were previously compiled into the SMAFD by Williams, Mdala, et al. (2021). These traces are directly incorporated into the MAFD although in some cases, their length has been extended following new mapping from aeromagnetic data (Section 3.3.3). In central and northern Malawi, onshore EAR faults have been mapped previously (Crossley, 1984; Ebinger, 1989; Ebinger et al., 1987; Kolawole et al., 2018a; Macgregor, 2015; Ring, 1994; Ring et al., 1992; S. M. Dawson et al., 2018; Wheeler & Karson, 1989), most notably in geological maps compiled between the 1950s and 1970s (Harrison & Chapusa, 1975; Hopkins, 1973; Peters, 1975; Ray, 1975; Thatcher, 1975). However, these studies mapped faults at relatively coarse scales (typically 1:100,000 or coarser); it is not always clear whether the mapped faults are “active,” and the fault traces were not stored in a GIS environment. In the MAFD, we use a TanDEM-X DEM with a horizontal resolution of 12.5 m and an absolute vertical mean error of 0.2 m (Wessel et al., 2018) to remap these previously documented faults. This allows us to map the fault's topographic expression at a much finer scale (1:10,000) than previous studies, and we use the “reference” attribute in the MAFD (Table 1) to indicate where a fault has been previously described.

Faults were mapped from the TanDEM-X DEM following the base of the fault's scarp or escarpment. Fault tips are defined by where the fault's topographic expression is no longer visible and/or combine these observations with constraints from aeromagnetic data (Section 3.3.3). In this context, closely spaced en échelon faults with...
no evidence for a physical linkage (i.e., soft-linked) are mapped as discrete faults (e.g., Sani, Nkhotakota, and Chombo faults, Figure 4b). Alternatively, where the fault’s topographic expression is continuous between different along-strike geometrical segments (i.e., hard-linked), and scarps may form hard links across perpendicular bends in Malawi (Hodge, Fagereng, Biggs, et al., 2018; Hodge et al., 2019; J. Jackson & Blenkinsop, 1997; Wedmore, Biggs, et al., 2020, Wedmore, Williams, et al., 2020), it is mapped as a single fault.

3.3.2. Seismic Reflection Data

Approximately 3,500 km of 2D multichannel seismic reflection data across Lake Malawi were acquired between 1985 and 1987 through Project PROBE (Figure 5a; Flannery & Rosendahl, 1990; Scholz & Rosendahl, 1988; Specht & Rosendahl, 1989). This survey extended over the entire lake with a line spacing of 10–20 km and provided the first generation of maps detailing the structure and stratigraphy of Lake Malawi. Basin structure was subsequently revised in parts of the basin following the collection of single-channel high-resolution data between 1992 and 1995 (T. C. Johnson et al., 1995; McCartney & Scholz, 2016; Mortimer et al., 2007; Scholz, 1995) and revised again following reprocessing of the Project PROBE data and its integration with 2,000 km of 2D multichannel seismic reflection data from Lake Malawi’s Central and North basins acquired through the Study of Extension aNd magmatism in Malawi aNd Tanzania (SEGMeNT) project (Figure 5a; Scholz et al., 2020; Shillington et al., 2016, 2020). The SEGMeNT survey was acquired in an orthogonal grid with an average spacing of 8 km. In addition, the SEGMeNT project deployed lake-bottom seismometers and collected wide-angle seismic refraction data (Accardo et al., 2018; Shillington et al., 2020), which were used for assessments of the deeper crustal structure and depth migration of the seismic reflection data. Further details on data acquisition and processing are available in Shillington et al. (2016, 2020) and Scholz et al. (2020).

Faults within Lake Malawi are incorporated into the MAFD from previously mapped offsets on the syn-rift basement surface (Scholz et al., 2020). This surface was generated by interpolating all available interpreted seismic reflection data using a least-squares algorithm with a 750 × 750 m cell size. Faults that offset this basement surface were mapped as 2D heave polygons (Figures 6 and 7); however, for inclusion in the MAFD, in which faults are mapped as 1D traces, only the footwall cutoffs of these heave polygons are utilized. Faults in Lake
Malawi could alternatively be mapped on a megadrought horizon located near the top of the sedimentary package and dated through drill core to 75 ka (Scholz et al., 2007; Shillington et al., 2020). However, by only incorporating basement-rooted faults, we avoid the risk of omitting active faults that do not offset the near-surface reflectors and of including basement faults that splay in Lake Malawi’s sedimentary package (McCartney & Scholz, 2016; Mortimer et al., 2016; Scholz et al., 2020; Shillington et al., 2020) as several distinct faults.

3.3.3. Aeromagnetic and Gravity Data

Aeromagnetic data consist of a grid of total magnetic intensity (TMI) anomalies, which depends on the magnetic susceptibility of the sources and the depth to the top of the magnetic sources (e.g., Grauch & Hudson, 2007). A fault with vertical offset that laterally juxtaposes two layers of differing magnetic properties is observable in the aeromagnetic grids, depending on the lateral magnetic contrast between the layers, the burial depth, vertical extent, and dip of the contrast boundary (fault dip and depth extent; Grauch & Hudson, 2007, 2011). In Malawi, the magnetic sources are composed of syn-rift siliciclastic sequences and the pre-rift gneissic basement. Where the latter is exposed or only shallowly buried, it exhibits prominent banding of alternating high- and low-magnetic lineaments (magnetic foliation) that is a characteristic of its mafic-felsic-mafic mineralogical banding (Kolawole et al., 2018a, 2018b, 2021). Hence, abrupt linear magnetic gradients or linear discontinuities that offset the lateral continuity of the basement magnetic fabrics in vertical and horizontal derivative aeromagnetic maps may be inferred as a basement-rooted normal fault (Kolawole et al., 2018a, 2018b, 2021).

**Figure 6.** Use of aeromagnetic data and seismic reflection surveys to identify onshore to offshore active faults in northern Malawi (Kolawole et al., 2018a; Scholz et al., 2020; Shillington et al., 2020). (a) Location map. (b) Active fault map with syn-rift basement heaves mapped from seismic reflection surveys (white polygons; Scholz et al., 2020), and their extrapolation using offshore aeromagnetic data in yellow. Also show are foliation orientation surface measurements (Kemp, 1975; Ray, 1975; Thatcher, 1975), and the 2009 Karonga Earthquake sequence surface ruptures (Kolawole et al., 2018a; Macheyeki et al., 2015) and Global Centroid Moment Tensor catalog earthquake locations (Ekström et al., 2012; Gaherty et al., 2019). Map underlain by aeromagnetic image created from the first vertical derivative of the 2013 aeromagnetic grid (Kolawole et al., 2018a) and TanDEM-X 12 m DEM. (c) Surface rupture along the St Mary fault following the 2009 Karonga earthquake sequence.
The fault-related magnetic gradient commonly separates two magnetic domains, one that is characterized by longer wavelength and lower frequency basement fabrics and the other characterized by shorter wavelength, higher frequency, and higher amplitude basement fabrics (Kolawole et al., 2018a, 2018b). The former represents the hanging wall of the fault where the magnetic basement is relatively deeper and covered by thicker sedimentary cover, and the latter represents the footwall of the fault in which the magnetic basement is shallower (“closer” to the airborne sensor) and above which the sedimentary cover is thinner or absent. Based on this structure, the basement-rooting normal fault's dip direction can also be inferred (Kinabo et al., 2007, 2008; Kolawole et al., 2018a, 2018b, 2021).

Figure 7. Use of aeromagnetic data and TanDEM-X DEM to map faults in southern Malawi. (a) The South Malawi Active Fault Database (SMAFD; Williams, Mdala, et al., 2021) in comparison to the Malawi Active Fault Database (MAFD) where aeromagnetic data have been used to revise the length of previously mapped faults and to infer faults with no surface expression (Kolawole et al., 2021). Revised and newly inferred faults in the MAFD highlighted in yellow. Map underlain by vertical derivative of 2013 aeromagnetic grid and surface measurements of foliation (A. L. Dawson & Kirkpatrick, 1968; Bloomfield, 1958; Bloomfield & Garson, 1965; Habgood et al., 1973; Walshaw, 1965). The “Malawi Other Faults” database, which represents faults in Malawi that have no evidence for East African Rift activity or have low reactivation potential, is also shown. Examples of active faults in the MAFD from (b and c) the Zomba Graben and (d and e) Makanjira Graben and southwestern arm of Lake Malawi. In both examples, maps are shown with and without aeromagnetic data to highlight the faults in the MAFD that have and do not have a surface expression.
In the MAFD, the aeromagnetic TMI grid used for fault mapping consists of high-resolution (~62 m spatial resolution) country-wide aeromagnetic data, acquired in 2013 by the Geological Survey Department of Malawi using a line spacing of 250 m and a flight altitude of 80 m (Kolawole et al., 2018a; Laõ-Dávila et al., 2015; S. M. Dawson et al., 2018). Except for the Nsanje Basin (Figure 2a), the survey covers all onshore parts of Malawi and extends up to 10 km offshore into Lake Malawi. Prior to fault interpretation, the TMI grid is first pole-reduced to correct for latitude-dependent skewness of the magnetic intensity data (Arkani-Hamed, 1988; Baranov, 1957). Afterward, the derivative filters are applied to the pole-reduced grids to better resolve magnetic gradients from which the fault traces can be inferred. Specifically, we map faults along the edges of the abrupt linear gradients in the vertical derivative maps or along the 0° tilt-angle derivative contour of the tilt derivative maps, which are interpreted to represent the footwall cutoff of a basement-rooted normal fault (Kolawole et al., 2018a).

In the Lower Shire Valley, gravity surveys can resolve density contrasts between thick sequences of Karoo and EAR-age sedimentary rocks and the adjacent metamorphic basement and/or Karoo magmatic intrusions (Chisenga et al., 2019). Following the interpretation of Chisenga et al. (2019) that these density contrasts reflect buried fault displacement, lineaments identified from vertical and horizontal derivative maps of the gravity data are also incorporated into the MAFD.

We use fault traces interpreted from the aeromagnetic and gravity data in two ways. In cases where the fault's expression in these data coincides with and extends beyond their expression in a DEM or the Lake Malawi syn-rift basement surface, we use the aeromagnetic or gravity fault trace to revise the position of the fault tips (Figures 6 and 7). We do so as we interpret that the fault tips in these data may represent the buried displacements that would be expected at the tip of an elliptical fault plane (Kim & Sanderson, 2005). If, however, a fault trace in these data cannot be collocated with a previously identified EAR fault, further analysis is required as aeromagnetic or gravity data alone cannot be used to differentiate if a fault has been active during East African rifting.

In these cases, we first assess whether the trace coincides with a linear topographic feature (e.g., a Karoo fault escarpment) that does not show evidence for EAR activity (as defined in Section 3.1). If so, we interpret these as an inactive trace and include them in the “Malawi Other Faults Database” (Figures 3a and 7). Alternatively, if the fault trace lacks a surface expression, but is located within the rift valley, it may represent an active fault buried under rift sediments. In these cases, we use fault reactivation analysis (Leclère & Fabbri, 2013; Lisle & Srivastava, 2004; Morris et al., 1996; Sibson, 1985) to infer if a fault is active. Specifically, we calculate these faults' effective coefficient of friction ($\mu'_f$), which uses the Mohr-Coulomb theory to quantify the minimum required friction coefficient or minimum pore fluid pressure that faults reactivate without also inducing failure along a plane that is optimally oriented in the regional stress state (Muluneh et al., 2018; Sibson, 1985; Williams et al., 2019). This is calculated by

$$\mu'_f = \frac{\sqrt{Q^2C - 2QC + C - \frac{c_f}{\sigma_1}}}{QA + B}$$  

(1)

where $Q = \sigma_3/\sigma_1$, $\sigma_1$, and $\sigma_3$ are the maximum and minimum principal compressive stress, respectively, $c_f$ is the fault cohesion (20 MPa), and $A$, $B$, and $C$ are functions that describe the fault's orientation in the stress state (Equations S6–S8 in Supporting Information S1). To perform this analysis, we consider the stress tensor derived from Malawi from a focal mechanism inversion by Williams et al. (2019). The principal stress orientations derived from this inversion are consistent with other local and regional focal mechanisms' stress inversions in Malawi (Delvaux & Barth, 2010; Ebinger et al., 2019) and the regional extension direction inferred from geodetic data (Stamps et al., 2021; Wedmore et al., 2021; see also Supporting Information S1).

From fault zone compositional analysis and basement rock deformation experiments, it is inferred that faults in Malawi are frictionally strong (coefficient of friction ~0.55–0.7) and the surrounding crust is relatively dry (Hellebrekers et al., 2019; Wedmore, Williams, et al., 2020; Williams et al., 2019). Given these points, we therefore only include buried faults inferred from the aeromagnetic and gravity surveys in the MAFD with a $\mu'_f \geq 0.55$ (Figures 3a and 8c). A dip must be assigned to a fault for reactivation analysis; however, aeromagnetic and gravity data cannot be used to measure fault dip angles. We therefore infer that these are vertical faults in cases where they exhibit strike-slip offsets, and otherwise that they are normal faults dipping at 53°, which is consistent with measurements of normal fault dip elsewhere in Malawi (Table S2 in Supporting Information S1). Faults inferred from these data that are not included in the MAFD because of their topographic expression or low $\mu'_f$ are
incorporated into the “Malawi Other Faults” database (Figures 2b, 3a and 7). Further details on the fault reactivation analysis, and its sensitivity to our assumptions, are provided in Supporting Information S1.

3.4. Fault Length Distribution

We use the distribution of fault lengths in the MAFD to investigate the hypothesis that as rift extension proceeds, the distribution of normal fault lengths evolves from a power law to an exponential trend (Ackermann et al., 2001; A. Gupta & Scholz, 2000; Michas et al., 2015). We first consider the length of each distinct continuous fault trace in the MAFD. Where faults splay in map view, only the length of the longest branch is considered, so that the full extent of fault lengthening is assessed.

As the transition in fault length distribution is thought to arise from previously distinct faults linking, we also assess fault lengths under a “multi-fault” scenario. In this case, we identify en échelon faults that are currently mapped as distinct structures in the MAFD, but which may represent a single soft-linked structure that could coalesce into a hard-linked fault as rift extension proceeds. Empirical observations and Coulomb stress modeling indicate that two en échelon normal faults may behave as a single soft-linked structure through coseismic stress transfer when the across-strike distance between the two fault tips is <20% of the participating faults' total length up to a maximum across-strike distance of 10 km (Biasi & Wesnousky, 2016; Hodge, Fagereng, & Biggs, 2018). We therefore use this as a criterion to determine if two en échelon faults in Malawi may be part of a multi-fault system.

We then test whether the distributions of fault and multi-fault lengths in the MAFD are best described by a power law or exponential distribution function through a two-sample Kolmogorov-Smirnov (KS) test (Clauset et al., 2009; Massey, 1951). We first use a Maximum Likelihood Estimator to fit power law and exponential functions to the fault length data. The complementary cumulative distribution function (ccCDF; i.e., survival function), which is defined as the probability that the continuous variable \( L \geq l \), can then be expressed for a power law distribution as
\[ P(L \geq l) = \begin{cases} \left( \frac{l}{l_{\min}} \right)^{1-\alpha}, & \text{for } l \geq l_{\min}; \\ 1, & \text{for } l < l_{\min} \end{cases} \] (2)

where \( \alpha \) is the power law exponent and \( l_{\min} \) is the lower bound of fault length below which fault mapping is considered incomplete. For an exponential distribution, the equivalent expression is:

\[ P(L \geq l) = \begin{cases} e^{\lambda(l-l_{\min})}, & \text{for } l \geq l_{\min}; \\ 1, & \text{for } l < l_{\min} \end{cases} \] (3)

where \( \lambda \) is the rate parameter (Clauset et al., 2009). We then test the null hypothesis that the empirical cCDF function of fault lengths in the MAFD represents samples from these continuous theoretical functions. This is achieved by determining the maximum difference between the empirical and theoretical cumulative trends, \( D^* \), and the probability (\( p \)) that \( D^* \) would have been observed, given the null hypothesis and the number of available samples. In this analysis, the null hypothesis that the observed lengths are samples from a theoretical distribution is rejected if \( p < 0.1 \) (Clauset et al., 2009).

The completeness of fault mapping in the MAFD is likely to be lowest in Lake Malawi due to the 5–20 km grid spacing of the 2D seismic surveys (Figure 5). We therefore consider a range of \( l_{\min} \) values for the MAFD of between 5 and 30 km and apply the two-sample KS test at 1 km increments of \( l_{\min} \) in this range. As with investigations of all natural fault populations, this analysis has limitations, such as the relatively small range of fault lengths considered (typically 1–2 orders of magnitude; Ackermann et al., 2001; A. Gupta & Scholz, 2000; Bonnet et al., 2001; R. M. Clark et al., 1999), and whether the mapped trace of a fault represents its true length (Ackermann et al., 2001; R. M. Clark et al., 1999). This latter point is especially important for offshore faults where uncertainties in fault lengths and potential linkages are constrained by the line spacing of 2D seismic reflection surveys (Michas et al., 2015). Results may also be affected by differences in fault growth between synsedimentary and basement-involved faults due to variations in the surrounding lithology’s Young Modulus (Cowie & Scholz, 1992). However, since faults in the MAFD were mapped from bedrock scarps or escarpments in DEMs, metamorphic aeromagnetic fabrics, crustal density contrasts, or syn-rift basement surface offsets in seismic reflection surveys, all faults are basement-involved and so this uncertainty should not influence our analysis.

4. Results

4.1. Overview of the MAFD

The MAFD contains geospatial and geomorphic data on 113 fault traces in Malawi and its surrounding regions (Figure 2, Table 1). We interpret that these faults are active since they have hosted displacement during East African rifting or are buried within the rift valley and favorably oriented to the regional stresses (Section 3.1, Figure 3a). Malawi’s national borders broadly coincide with the trajectory of the EAR and hence, active faults are found along its length. There are however areas in western Malawi that may be up to 100 km from a mapped active fault (Figure 2). The MAFD, along with the “Malawi Other Faults” database, is freely available at https://doi.org/10.5281/zenodo.5507190.

The MAFD is compiled from a multidisciplinary range of data sets: 40 faults from previous geological mapping and high-resolution DEMs, 19 from aeromagnetic data (Kolawole et al., 2018a, 2021), 4 inferred from gravity data (Chisenga et al., 2019), and 50 offshore faults in Lake Malawi from 2D seismic reflection surveys (Scholz et al., 2020; Shillington et al., 2020). Further descriptions of these faults, and Malawi’s tectonic evolution, are provided in the previously referenced studies. The key innovations of the MAFD are that these fault traces are now stored in one place, criteria have been laid out as to how faults are classified as “active” and tips are mapped (Section 3), and geomorphic and confidence attributes have been associated with each fault (Table 1).

In south Malawi, the MAFD represents an update of the SMAFD as new fault mapping from aeromagnetic data (Kolawole et al., 2021) has been used to revise the length of 9 faults and identify 15 faults with no surface expression that were not included in the SMAFD (Figure 7b). All faults mapped from seismic reflection surveys in
Lake Malawi are interpreted as active, given that these faults inherently offset EAR-age sediments. The accuracy of fault mapping in seismic reflection data is, however, relatively low as their position could only be constrained within the 5- to 20-km-spaced 2D survey lines. In all cases that an offshore fault in the MAFD was mapped from aeromagnetic data and crossed the 2D seismic survey, it could be identified in the seismic reflection data for at least one of the lines it crossed (Figures 6 and 7).

For onshore faults, variations in exposure quality were observed between faults that exhibited steep scarps in DEMs (e.g., Figure 4c) and those that are expressed by degraded escarpments. Between the DEMs and aeromagnetic data, it is not always clear where the tips of some faults are, and in these cases, the epistemic quality of fault mapping (Table 1) is accordingly reduced. Low confidence for activity is assigned to the 23 intrarift faults inferred from gravity or aeromagnetic data alone as interpretation of the data is nonunique and their inclusion in the MAFD is based on their effective coefficient of friction ($\mu'$) being ≥0.55 (Section 3.3.3, Figure 8c). Furthermore, this calculation is made uncertain by poor constraints on fault dip and cohesion, and additional faults may also be reactivated in Malawi if frictionally weak materials or high fluid pressures are present (see Supporting Information S1 for further discussion).

4.2. Onshore Faults in Central Malawi

We highlight onshore faults in central Malawi as these faults are not typically considered in the region's tectonic evolution or seismic hazard. Evidence of recent activity on faults in this region is particularly apparent from antecedent and consequential drainage patterns (Figure 4b). For example, the west-dipping Sani and Chilangali fault scarps impede streams flowing eastward into Lake Malawi, and in the case of the latter has resulted in the formation of Lake Chilangali (Figure 4b; Harrison & Chapusa, 1975). Furthermore, the Liwaladzi scarp has diverted the Bua river (Figure 4b; Harrison & Chapusa, 1975; Peters, 1975). The TanDEM-X data did not reveal any previously undocumented active onshore faults in this region. However, some previously mapped “Neogene” faults (Crossley, 1984; Ebinger et al., 1987) have basement mapped in their hanging wall (Harrison & Chapusa, 1975), and there is no evidence for interactions between displacement along these faults and surrounding rivers (Figure 4b). Since this means they do not meet the criteria for inclusion in the MAFD (Figure 3a), they are included in the Malawi Other Faults database.

4.3. Probability Distribution of Fault Lengths in the MAFD

Assuming that each fault in the MAFD represents a distinct structure, we can reject the null hypothesis that the distribution of their lengths is drawn from an exponential function (i.e., $p < 0.1$) for cases with a lower bound of fault length ($l_{\text{min}}$) >6 km (Figure 9b). However, we cannot reject the null hypothesis that the distribution of lengths may form a power law relationship with an exponent ($\alpha$ in Equation 2) of 1.9 ± 0.2 when $l_{\text{min}}$ >10 km (Figures 9b and 9d and Figure S2 in Supporting Information S1). With increasing $l_{\text{min}}$, $\alpha$ increases to 2.7 ± 0.5 (Figures 9b and 9d and Figure S2 in Supporting Information S1).

Assuming the “multi-fault” case, in which closely spaced en échelon faults are considered to represent a single coherent structure, it is less clear which trend best describes the fault population (Figures 9e and 9f). When $l_{\text{min}}$ <14 km, a power law hypothesis can be rejected, but neither hypothesis can be rejected if $l_{\text{min}}$ >14 km (Figure 9b). For the multi-fault power law trend, $\alpha$ is 1.7 ± 0.2 for $l_{\text{min}}$ = 14 km and 2.2 ± 0.4 for $l_{\text{min}}$ = 30 km (Figure S2 in Supporting Information S1). For an exponential trend, the characteristic length scale ($1/\lambda$, where $\lambda$ is the rate parameter defined in Equation 3) ranges from 58 ± 16 km to 81 ± 30 km with increasing $l_{\text{min}}$ (Figure S2 in Supporting Information S1).

5. Discussion

5.1. Completeness of the MAFD

The MAFD represents a compilation of all known fault traces in Malawi that show evidence for activity, or potential for activity, related to EAR extension. It does not, however, represent a database of every active fault in Malawi (Figure 3b). For example, up to 30% of the extension within Lake Malawi could be accommodated by faults that are below the resolution of seismic reflection data or not covered by the 5- to 20-km-spaced seismic survey grid (Marrett & Allmendinger, 1992; Shillington et al., 2020). Additionally, offshore faults with basement
displacements less than $\sim 100$ ms were not mapped by Scholz et al. (2020) as these were generally too short to correlate between multiple seismic profiles. Also, since aeromagnetic grids are potential fields’ geophysical data, they may not image deep-seated basement-confined short-wavelength high-frequency anomalies of deeply buried small offset faults (Kolawole et al., 2017). The limitations of these data sets in resolving relative short faults may be seen in the break in the empirical fault length distributions at lengths <10 km (Figure 9a).

Any NW-SE striking, moderately dipping Proterozoic or Karoo age faults in Malawi will be favorably oriented for reactivation in its current stress state. We cannot definitively determine that these faults are inactive; however, we

Figure 9. Analyses of fault length distributions in the Malawi Active Fault Database. (a) The empirical (complementary) cumulative frequency of the lengths of all faults documented in the Malawi Active Fault Database (MAFD). This plot considers cases where each fault in the MAFD represents a distinct fault, and where closely spaced en échelon faults represent a single structure (the “multi-fault case”). (b) Results from two sample Kolmogorov-Smirnov (KS) tests for the fit between empirical and theoretical survival functions of MAFD fault lengths for lower bounds of fault length ($l_{\text{min}}$) between 5 and 30 km. Both power law (Equation 2) and exponential (Equation 3) survival functions are considered in the KS tests with the power law exponent and exponential rate parameter estimated via maximum likelihood. The $p$-value (0.1) below which the KS test rejects the null hypothesis is also highlighted. (c and d) Empirical and theoretical survival functions of fault lengths in the MAFD for representative values of $l_{\text{min}}$ of 10 and 30 km, and assuming that each fault represents a distinct structure. The equation for the theoretical trend and its fit to the empirical trend (i.e., the $p$-value from a KS test) are also reported. Panels (e and f) equivalent to panels (c and d), but assuming the multi-fault case.
consider that their location outside of the EAR and their lack of EAR surface offsets are a more reliable guide for fault activity than their reactivation potential; hence, they are included in the Malawi Other Faults database. For some faults, there is also evidence from seismic reflection data that they have not been active since EAR-related sediment accumulation began (e.g., the Mbamba Fault; Accardo et al., 2018; McCartney & Scholz, 2016).

Geomorphic and geologic offsets indicate that all fault included in the MAFD are normal faults (Table 1; Kolawole et al., 2018a; Scholz et al., 2020; Shillington et al., 2020; Wedmore, Biggs, et al., 2020). We highlight this is not because the MAFD excludes other types of faults; instead, this is indicative of normal faulting being the dominant type of deformation in Malawi (Delvaux & Barth, 2010; Ebinger et al., 2019; Hodge et al., 2015; Williams et al., 2019). Nevertheless, because lateral offsets are difficult to identify in the geomorphic record (J. Jackson, 2001; McCalpin, 2009), we cannot exclude the possibility that there are active strike- or oblique-slip faults in Malawi. If identified in future, they should be incorporated into the MAFD.

It is also possible that some EAR-age faults included in the MAFD are now inactive (Figure 3b). For example, Cowie (1998) used numerical models to demonstrate how faults in continental rifts may be abandoned depending on their position around neighboring faults. However, this process mainly occurs at a stage when fault coalescence starts to dominate over fault nucleation (Ackermann et al., 2001; Hardacre & Cowie, 2003), and the power law distribution of fault lengths we document suggests this is not yet occurring in Malawi (Figure 9). In Lake Malawi, additional evidence that basement-rooted faults are active is provided by lake floor scarps and/or their offset of a 75 ka seismic reflector (Crow & Eccles, 1980; Scholz et al., 2020; Shillington et al., 2020), while onshore, the steep scarps and interactions between faults and rivers we observe (Figures 4 and 5) are similar to those observed in other tectonically active regions (e.g., Goldsworthy & Jackson, 2000; J. Jackson et al., 1996; Morell et al., 2020). Local closely spaced temporary seismic deployments have also been able to resolve microseismicity across the projected location of some faults in the MAFD (Ebinger et al., 2019; Gaherty et al., 2019; Stevens et al., 2021).

Faults under EAR sediments with no surface expression are particularly challenging to classify. We consider it prudent that where inferred from gravity and aeromagnetic data, and if favorably oriented to the regional stresses, such faults should be included in the MAFD, given that the source of the 1989 $M_w$ 6.3 Salima earthquake is buried (H. K. Gupta & Malomo, 1995; J. Jackson & Blenkinsop, 1993) and the St Mary Fault had no surface expression prior to the 2009 Karonga earthquake sequence (Kolawole et al., 2018a; Macheyeki et al., 2015). Indeed, with regard to the St Mary Fault, there is a close correlation between its position as constrained by the Karonga Earthquake locations, Interferometric Synthetic Aperture Radar, coseismic surface rupture (Figure 6b; Biggs et al., 2010; Gaherty et al., 2019; Macheyeki et al., 2015), and as inferred from aeromagnetic data (Kolawole et al., 2018a). Hence, although there is a nonuniqueness to interpreting aeromagnetic anomalies, the inference that they are faults is consistent with available independent geological and geophysical data sets in Malawi (see also Figures 6 and 7). The high reactivation potential of active faults with known orientations in Malawi (Figure 8a) also provides confidence that this analysis can differentiate which buried faults are and are not active. Nevertheless, with the “ActivConf” parameter (Table 1), a MAFD user can readily distinguish between faults with and without evidence of EAR offsets.

New paleoseismic and chronostratigraphic data would help resolve which onshore faults in the MAFD have been active in the recent (i.e., Quaternary) geologic past. However, in low strain rate regions like Malawi, faults can go through very long periods (~50–100 ka) of quiescence (D. Clark et al., 2012; Pérouse & Wernicke, 2017; Taylor-Silva et al., 2020), during which evidence for past earthquakes may be buried or eroded away (Hodge et al., 2020; Nicol et al., 2016). Alternatively, a fault may have hosted earthquakes that did not rupture to the surface (Hecker et al., 2013; Wells & Coppersmith, 1993), a possibility that is increased in Malawi because of its thick seismicogenic layer. Hence, although we encourage new chronostratigraphic data to be incorporated into future updates of the MAFD, these data should not necessarily be used to revise the criteria for defining fault activity. The challenges described above are inherent to any active fault database, and although the MAFD is incomplete, this does not preclude from the fact that systematically mapping all known “active” faults in Malawi is still an informative tool for investigating regional seismic hazard.

### 5.2. The MAFD and Seismic Hazard in Malawi and Elsewhere in the East African Rift

The MAFD can be used as a primary data source for investigating seismic hazards in Malawi. Most commonly, this hazard is considered in terms of ground shaking through probabilistic seismic hazard analysis. However, the
MAFD can also be used to assess other earthquake hazards in Malawi, such as liquefaction and earthquake-triggered landslides, which occurred following the 2009 Karonga earthquakes (Kolawole et al., 2018b) and 1989 Salima earthquakes (H. K. Gupta & Malomo, 1995), respectively. There are also reports of seiches at the northern end of Lake Malawi following the 1910 M 7.4 Rukwa Earthquake (Ambraseys, 1991).

A more detailed assessment of the seismogenic properties of faults in Malawi is contained in the Malawi Seismogenic Source Database, which is currently in development (Williams, Wedmore, et al., 2021). Nevertheless, it is clear that given low regional extensional rates (0.5–2 mm/yr; Saria et al., 2014; Stamps et al., 2018; Wedmore et al., 2021), large magnitude earthquakes are rare in Malawi (fault recurrence intervals ~1,000–20,000 years, Hodge et al., 2015; Shillington et al., 2020; Williams, Mdala, et al., 2021), and the occurrence of these events may be further reduced if faults slip aseismically as was observed following an $M_w$ 5.2 earthquake in northern Malawi in 2014 (Zheng et al., 2020).

Similar data sets (e.g., seismic reflection and aeromagnetic data) to those used in the MAFD have been collected elsewhere in the Western Branch of the EAR (Heilman et al., 2019; Karp et al., 2012; Katumwehe et al., 2015; Kolawole et al., 2017, 2021; McGlue et al., 2006; Muirhead et al., 2019; Wright et al., 2020), and we suggest that new active fault databases could be developed from applying the MAFD framework to these data sets. This framework could also be applied to other EAR branches (Figure 1a); however, these do present additional challenges in defining and mapping active faults, since faults in more evolved EAR branches may now be inactive and/or they may have formed in response to dike intrusions (Agostini et al., 2011; Casey et al., 2006; Ebinger & Casey, 2001; Keir et al., 2006; Sieburg et al., 2020).

5.3. The Distribution of Fault Lengths in the MAFD and Tectonic Evolution of the East African Rift in Malawi

A power law distribution of fault lengths is favored in continental rifts when (a) fault growth occurs within a mechanically unconfined layer (Ackermann et al., 2001; Soliva & Schulz, 2008) and/or (b) total regional extension is low (<8%–12% extension; A. Gupta & Scholz, 2000; Michas et al., 2015). The power law distribution of fault lengths in the MAFD for $l_{\text{min}} > 10$ km (Figure 9b) is therefore consistent with unconfined fault growth in Malawi's thick seismogenic layer (30–40 km; Craig & Jackson, 2021; Ebinger et al., 2019; J. Jackson & Blenkinsop, 1993) and low total extension in Malawi (<8%; Scholz et al., 2020).

The exponent ($\alpha$ ∼ 2) of the power law distribution of fault lengths in Malawi is relatively high compared to other fault length distributions ($\alpha$ ∼ 1.5; R. M. Clark et al., 1999; Scholz & Cowie, 1990). This indicates a relatively high number of short faults in Malawi. This is despite our method of mapping faults as single, linear traces regardless of geometrical complexities (Section 3.1), which may have increased the number of relatively long faults. It has been suggested before that a power law distribution of fault lengths reflects localization of regional strain onto a small number of relatively long faults (Scholz & Cowie, 1990; Soliva & Schulz, 2008). This is broadly consistent with observations in Malawi that its longest faults (>100 km) tend to be hard-linked rift-bounding “border” faults, which have accommodated 50%–75% of rift extension (Accardo et al., 2018; Ebinger et al., 1987; Shillington et al., 2020; Wedmore, Biggs, et al., 2020; Wedmore, Williams, et al., 2020).

Following the “multi-fault” case to map faults in Malawi (Section 3.4), we identified 55 intrarift faults in the MAFD that may coalesce into 23 distinct structures as they accumulate displacement. In this case, we cannot distinguish whether the length distribution follows an exponential or a power law distribution (Figure 9). Hence, although fault coalescence promotes a shift toward an exponential distribution of fault lengths, this distribution can still be fitted with a power law. Some shorter faults may therefore need to become inactive in Malawi for a distinct exponential distribution of faults to form (Hardacre & Cowie, 2003; Meyer et al., 2002).

6. Conclusions

We present the MAFD, an open-access geospatial database that contains geomorphic data of 113 faults. We infer that these faults are active as they have hosted displacement since the onset of East African rifting in Malawi or because they are buried beneath the rift valley and are favorably oriented to the regional stresses. To address the challenges of mapping such faults in the Western Branch of the EAR's, the MAFD has been compiled from a multidisciplinary data set that includes fieldwork, existing geological maps, high-resolution DEMs,
seismic reflection surveys (Scholz et al., 2020; Shillington et al., 2016, 2020), and aeromagnetic (Kolawole et al., 2018a, 2021) and gravity data (Chisenga et al., 2019). We consider that the MAFD is currently the most complete compilation of active faults across Malawi and will be of use for future regional seismic hazard studies. We also encourage updates to the MAFD as and when new data become available.

Through the MAFD, we have also explored how active fault databases can be used to investigate regional geological evolution. We find that the distribution of fault lengths in the EAR in Malawi is consistent with a power law. This supports previous observations from Malawi and elsewhere that during early stages of continental rifting, regional extensional strain is localized along relatively long fully linked border fault systems. As the EAR in Malawi accumulates more extension over time, we anticipate that shorter faults will coalesce or become inactive to form an exponential distribution of fault lengths.

**Data Availability Statement**

The version of the Malawi Active Fault Database and Malawi Other Fault database associated with this study is available under the Creative Commons Attribution-ShareAlike (CC-BY-SA 4.0) at Zenodo (https://doi.org/10.5281/zenodo.5507190). The authors ask that users cite this version of the database alongside this manuscript. All future versions will be released on Zenodo and GitHub (https://github.com/LukeWedmore/malawi_active_fault_database). These databases are saved in several different file formats to enable compatibility with different software and seismic hazard codes. The version of record is the GeoJSON format, which is human readable, plain text, and is therefore subject to version control within Git. Future changes to the database will be made to the GeoJSON file and then converted to other formats using the GDAL tool ogr2ogr. The database is also available in ESRI Shapefile, GeoPackage, KML, and GMT formats, which are all available on Zenodo and GitHub. Seismic reflection data acquired through the Project PROBE and SEGMeNT Project are available through the Marine Geoscience Data System at https://www.marine-geo.org/tools/search/entry.php?id=Malawi_PROBE (date last accessed at 05/25/21) and https://www.marine-geo.org/tools/search/EARSSEGMENT (date last accessed 05/25/21), respectively. Aeromagnetic data for the Karonga region and southern Malawi are also archived on the Marine Geoscience Data System at https://www.marine-geo.org/tools/search/Files.php?data_set_uid=24314 (https://doi.org/10.1594/IEDA/324314, date last accessed 05/25/21) and https://www.marine-geo.org/tools/search/Files.php?data_set_uid=24860 (https://doi.org/10.1594/IEDA/124860, date last accessed 05/25/21), respectively. TanDEM-X data were provided through German Aerospace Centre (DLR) proposal DEM GEOLO686 DEM_ and GEOLO2881 and can be obtained from the DLR at https://tandemx-science.dlr.de/cgi-bin/wcm.pl?page=TDM-Proposal-Submission-Procedure (date last accessed 15/02/22).

**Acknowledgments**

This work is supported by the EPSRC-Global Challenges Research Fund PREPARE (EP/P028233/1) and SAFER-PREPARED (part of the “Innovative data services for aquaculture, seismic resilience and drought adaptation in East Africa” grant; EP/T015462/1) projects. The authors thank the editor, associate editor, Damien Delvaux, and three anonymous reviewers for their constructive comments. The Geological Survey Department of Malawi kindly gave us access to the 2013 aeromagnetic data across Malawi.

**References**


