A global zircon U–Th–Pb geochronological database

Yujing Wu [1], Xianjun Fang [1], Jianqing Ji [1, *]

[1]School of Earth and Space Sciences, Peking University, Beijing 100871, China

Correspondence to: Jianqing Ji (grsange@pku.edu.cn)

Abstract. Since the start of the 21st century, the widespread application of ion probes has promoted the mass output of high-precision and high-accuracy U–Th–Pb geochronological data. Zircon, as a commonly used mineral for U–Th–Pb dating, widely exists in the continental crust and records a variety of geological activities. Due to the universality and stability of zircons and the long half-lives of U and Th isotopes, zircon U–Th–Pb geochronology can provide nearly continuous records for almost the entirety of Earth's history and is thus essential to studying the growth and evolution of the continental crust and Earth system evolution. Here, we present a database of zircon U–Th–Pb geochronological that samples the global continental crust and spans nearly all of Earth's history. This database collects ~2,000,000 geochronology records from ~12,000 papers and dissertations and is by far the largest geochronological database to our knowledge. This paper describes the complied raw data, presents the relationship between dating error and zircon age, compares the error levels of different dating instruments, and discusses the impact of sampling bias on data analysis as well as how to evaluate and minimize this impact. In addition, we provide an overview of the temporal and spatial distribution of global zircon ages and provide key insights into the potential research value of zircon ages for Earth system science, such as crustal evolution, plate tectonics, and paleoclimate changes, as well as commercial use in mining and energy exploration. Overall, this data collection provides us with a comprehensive platform with which to study zircon chronological data in deep time and space.

1. Introduction

Zircon U–Th–Pb geochronology has been made much more practical since Krogh (1973) invented isotope dilution-thermal ionization mass spectrometry (ID-TIMS) (Song, 2015; Davis et al., 2003). With the widespread use of ion probe mass spectrometers, in situ microanalysis can be performed efficiently and precisely, promoting the rapid development of zircon U–Th–Pb dating (Gehrels, 2014; Carrapa, 2010). The main dating instruments at present include (Becker, 2007) laser ablation inductively coupled plasma–mass spectrometry (LA–ICP–MS), secondary ion mass spectrometry (SIMS), sensitive high resolution ion microprobe (SHRIMP), and thermal ionization mass spectrometry (TIMS) instruments.

The U–Th–Pb decay system plays a key role in geochronology. The half-lives of ^{238}U, ^{235}U, and ^{232}Th isotopes are long enough to date Earth's entire history but short enough to allow for the accurate measurement of both parent and daughter isotopes. Based on the decay of ^{238}U$\rightarrow^{206}$Pb (half-life: 4.47 billion years (Gyr)), ^{235}U$\rightarrow^{207}$Pb (0.70-Gyr half-life), and ^{232}Th$\rightarrow^{208}$Pb (14.01-Gyr half-life) (Jaffey et al., 1971), we can obtain three ages, i.e., ^{206}Pb/^{238}U, ^{207}Pb/^{235}U, and ^{208}Pb/^{232}Th. Using the abundance ratio of natural U isotopes, ^{238}U/^{235}U$\approx$137.8, another ^{207}Pb/^{206}Pb age can be derived (Spencer et al., 2016; Hiess et al., 2012). If these ages are consistent with one another, the decay system is closed, verifying the reliability of the measured ages. This also represents an advantage of U–Th–Pb dating over other isotope dating methods.

Zircon is a common accessory mineral that can stably exist in various kinds of rocks and is distributed throughout the global continental crust of all ages (Hanchar and Hoskin, 2018; Hawkesworth et al., 2010). It is not unusual for zircons to survive through multiphase complex magmatism and metamorphism as a result of their physical and chemical stability (Hawkesworth et al., 2010). Due to its low original Pb content, rich Th and U contents, and high closure temperature for trace element diffusion, zircon is one of the most widely used minerals for U–Th–Pb isotopic dating (Williams, 2015). In addition, the zircon age distribution can span nearly all of the geological history. The oldest zircon discovered thus far is 4.4 Gyr old (Wilde et al., 2001). Thus, zircon U–Th–Pb geochronology provides an excellent means to explore the deep-time evolution of the continental crust (Voice et al., 2011).

With the widespread application of ion probe mass spectrometers, many zircon U–Th–Pb ages have been measured by

various chronology laboratories worldwide in the past two decades (Puetz and Condie, 2019; Wu et al., 2022). These zircons are sampled in the global continental crust, with ages nearly continuously spanning from 4.4 billion years ago (Ga) to the present (Puetz et al., 2017). Here, we collected zircon U–Th–Pb dating records for the past decades and built a global zircon database, which could provide a means for enhanced academic and commercial geological analyses.

Many scholars have previously collected zircon data to explore the evolution of the solid Earth and Earth's geodynamic history. Voice et al. (2011) compiled ~5100 individually dated detrital zircon samples with ~200,000 dating records and suggested that the temporal distribution of zircon ages could indicate episodic crustal recycling. This evidence provided by zircons is consistent with plate tectonics. By analyzing the temporal distribution of zircon U–Pb ages, combined with craton collisions and crustal cycles, Condie (2013) explored the evolution of the Proterozoic crust from the Nuna supercontinent to the Rodinia supercontinent. Mckenzie et al. (2016) used ~120,000 detrital zircon U–Pb age data from areas worldwide to explore the spatial distribution of continental magmatic arc systems in the Cryogenian period. Puetz and Condie (2019) collected U–Pb age data for ~610,000 detrital zircons and ~212,000 igneous zircons sampled from worldwide, as well as 5 other isotope databases, showing the geochemical cycles of mantle evolution. Wu et al. (2022) analyzed the zircon production series of the global continental crust, compared it to astronomical driving factors, and proposed that the evolution of the continental crust and even the Earth system may be affected by the astronomical environment.

However, insufficient data with limited global coverage can affect results, which in turn can contribute to misleading interpretations (Wu et al., 2020; Puetz et al., 2017). Here, we present a database of ~2,000,000 zircon U–Th–Pb dating data points sampled from the global continental crust (Wu et al., 2023). The dating instruments used include LA–ICP–MS, SHRIMP, SIMS, and TIMS instruments; zircon host rocks include igneous, metamorphic, and sedimentary rocks. This database may provide a more comprehensive and objective chronology data source both the temporally and spatially for future earth system science research. From this database, scholars can not only obtain an overview of global zircon production throughout Earth's entire history but also study the evolution of zircon production in each period and region. If these data are combined with other geological and astronomical events, it is also possible to deeply explore the energy source of Earth's dynamics and the mechanisms behind it. In the future, this database may even provide constraints for astronomical parameters and their evolution, expanding the deep-time dimension of astronomical parameters. In addition, from the perspective of data science, the large data volume and global sampling range of this database give us a good experimental platform for analyzing and solving biased sampling issues.

2. Database

Here, we collected ~2,000,000 zircon U–Th–Pb age data points sampled from the global continental crust from ~12,000 references, with zircon ages spanning all of Earth's history. This database is based on the original Chinese zircon U–Th–Pb dating database (Fang et al., 2018; Wu et al., 2019), with data sampled only in the Chinese continental crust and published in 2017 and earlier. Specifically, for the zircon records used in the study of Fang et al. (2018), the number of data points is 154,768, and the corresponding original references were published by the academic publisher of China National Knowledge Infrastructure (CNKI). The zircon data in the study of Wu et al. (2019) contained ~410,000 dating records, which were collected from Elsevier. The new database in this paper, however, expanded the sampling sites to the global continental crust and updated data published from 2018 to 2020. The new original references were collected from the academic publishers of Elsevier (ScienceDirect full-text database), Cambridge, Geological Society of London, Oxford, Springer, Taylor & Francis, Wiley, and CNKI. This new global database was used for time series analysis in the work of Wu et al. (2022), and we subsequently carried out more data cleaning work.

The compilation of zircon data is divided into two categories: 1) the "Database" files, which include ~2,000,000 records of zircon U–Th–Pb isotope ratios, ages, sample information, dating instruments, parent rock lithology, sampling locations, and reference numbers, and 2) the "References" file, which provides information on ~12,000 original references corresponding to the zircon geochronological data, including author, year of publication, and detailed publication data. For each record in the "Database" files, its reference can be found in the "References" file from its reference number. This database is available at https://doi.org/10.5281/zenodo.8040079.

Notably, we improved the codes for extracting GPS information, so the GPS data used here are slightly different from the sampling map given in Wu et al. (2022) (see Figure 1). The details of the improvement are as follows: 1) in the original references, the symbols of GPS "degrees, minutes and seconds" are varied, it is difficult for a computer to completely distinguish them, and sometimes they are not even recognized or are recognized incorrectly; the new modification expands the recognition range of the codes for "degrees, minutes and seconds" symbols and manually normalizes some abnormally identified GPS data according to the original references. 2) The initial GPS extraction codes were aimed at the Chinese continental crust, and the GPS range was restricted in the screening step. Later, when the code was updated to extract the global GPS data, some details were omitted, resulting in the loss of some GPS data, especially for Europe.

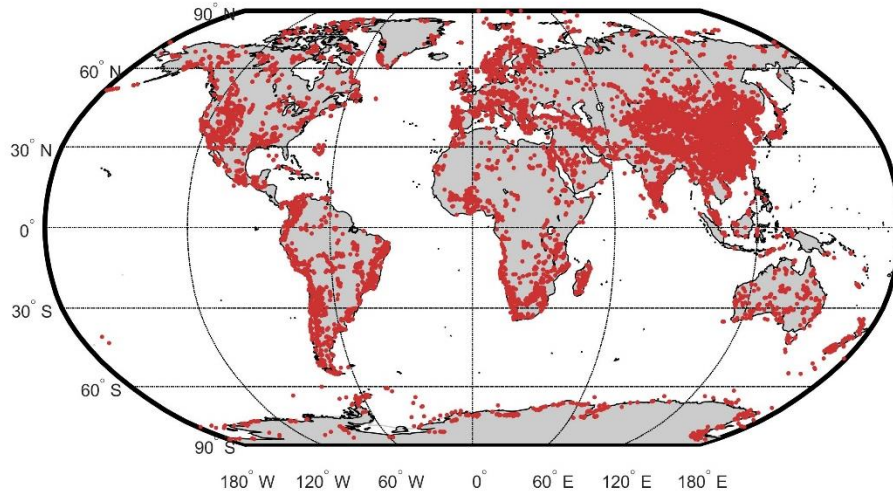



Figure 1. Map of the distribution of zircon sampling locations. Each red dot indicates a sampling location (adapted from Wu et al. (2022)).

2.1 Geochronological database

For the "Database" files, the fields of each record are detailed in Table 1. The database contains indices, isotope ratios and uncertainties of $^{206}Pb/^{238}U$, $^{207}Pb/^{235}U$, $^{207}Pb/^{206}Pb$, and $^{208}Pb/^{232}Th$; ages and uncertainties derived from the above isotope ratios; dating instruments; reference materials for deriving ages; lithology of the host rock; and GPS coordinates of the sampling location. *Data_number* is the unique index for each chronological record. For each record in the "Database," the source references can be found in the "References" file by *Ref_number*. The dating instruments include LA–ICP–MS, SIMS, SHRIMP, and TIMS. We also provide reference materials that were used to derive ages for further exploration. Notably, there are some null values in the database because the original papers did not publish them. Despite these deficiencies, the database still has considerable potential for geological research. The percentages of the records that are not null are listed in Table 1. Most fields have null values of less than 25%, except for reference materials (68.88%) and data on $^{232}Th–^{208}Pb$ decay (11-14%), which is used less than other decay systems.

Table 1. Data specifications of the zircon U–Th–Pb database.

Field name	Field description	Proportion	Data details
Ref_number	Reference number	100.00%	The number of the reference corresponding to each chronological record.
Data_number	Record number	100.00%	The unique index of each chronological record.
Sample_number	Sample number	99.9973%	The number of each rock sample or zircon sample corresponding to each

			chronological record.
isotope206Pb/238U	Isotope ratio of ^{206}Pb/^{238}U	86.52%	^{238}U–^{206}Pb decay.
isotope206Pb/238U_σ	Standard deviation of ^{206}Pb/^{238}U ratio	86.04%	
isotope207Pb/235U	Isotope ratio of ^{207}Pb/^{235}U	81.01%	^{235}U–^{207}Pb decay.
isotope207Pb/235U_σ	Standard deviation of ^{207}Pb/^{235}U ratio	80.58%	
isotope207Pb/206Pb	Isotope ratio of ^{207}Pb/^{206}Pb	77.07%	^{238}U–^{206}Pb and ^{235}U–^{207}Pb decays.
isotope207Pb/206Pb_σ	Standard deviation of ^{207}Pb/^{206}Pb ratio	76.57%	
isotope208Pb/232Th	Isotope ratio of ^{208}Pb/^{232}Th	13.57%	^{232}Th–^{208}Pb decay.
isotope208Pb/232Th_σ	Standard deviation of ^{208}Pb/^{232}Th ratio	13.47%	
age206Pb/238U	Age derived from ^{206}Pb/^{238}U	96.37%	Unit: Myr
age206Pb/238U_σ	Standard deviation of ^{206}Pb/^{238}U age	95.44%	Unit: Myr
age207Pb/235U	Age derived from ^{207}Pb/^{235}U	77.70%	Unit: Myr
age207Pb/235U_σ	Standard deviation of ^{207}Pb/^{235}U age	76.92%	Unit: Myr
age207Pb/206Pb	Age derived from ^{207}Pb/^{206}Pb	84.68%	Unit: Myr
age207Pb/206Pb_σ	Standard deviation of ^{207}Pb/^{206}Pb age	83.72%	Unit: Myr
age208Pb/232Th	Age derived from ^{208}Pb/^{232}Th	11.19%	Unit: Myr
age208Pb/232Th_σ	Standard deviation of ^{208}Pb/^{232}Th age	11.17%	Unit: Myr
Instrument	Dating instrument	98.59%	Including LA–ICP–MS, SHRIMP, SIMS, and TIMS.
Reference_material	Reference materials used to derive zircon U–Th–Pb ages.	68.88%	Including 91500, GJ-1, Temora 1, Temora 2, QGNG, Plešovice, SL13, M257, AS3, CZ3, etc.
Lithology	The lithology of the host rock	90.30%	Including igneous, sedimentary, and metamorphic rocks.
Longitude	The longitude of the sampling location	76.95%	Unit: degrees; range: -180 to 180.
Latitude	The latitude of the sampling location	76.95%	Unit: degrees; range -90 to 90.

Note: "Myr" indicates "million years." The "Proportion" column lists the proportion of nonnull values.

2.2 Data references

The chronological data used in this study were collected from ~12,000 references of the following academic publishers: Elsevier (ScienceDirect full-text database), Cambridge, Geological Society of London, Oxford, Springer, Taylor & Francis, Wiley, and China National Knowledge Infrastructure (CNKI). Original reference information for the chronological data is included in the "References" file (see the data repository on Zenodo). The fields of each reference record are detailed in Table 2. The field "*Ref_number,*" namely, the reference number, cites the unique index in the "References" file, corresponding to multiple chronological records in the "Database" files. The field "*DOI*" provides the unique Digital Object Identifier to access the original reference. However, not all references have a DOI, especially Chinese dissertations and old papers. Instead, we supplement a "URL" field, which contains web addresses to the references without a DOI. There is 82.93% of the literature with a DOI and 4.12% of the literature with a URL. For the remaining 13.09% of the literature, the unique identification of the data source can still be guaranteed by other information we provided: the first author, title, journal, volume, issue, first page, and last page.

Table 2. Data specifications of the References file.

Field name	Field description
Ref_number	Reference number
Author_surname	The surname of the first author
Author_given_name	The given name of the first author

Year_publication	Year of paper publication
Title	The title of the paper
Journal	The journal in which the paper was published
Volume	The volume of the paper
Issue	The issue of the paper
First_page	The first page of the paper.
Last_page	The last page of the paper.
DOI	The unique Digital Object Identifier (DOI).
URL	A web address to the literature.

Note: For references in Chinese, the "*Author_surname*" field is the full name of the first author owing to the Chinese citation format.

3. Data cleaning

3.1 Uncertainty normalization

We unified the uncertainty form of isotope ratios and ages into standard deviation for the convenience of conducting research. There were various uncertainty forms published in the original literature. The specific processing procedures for each form were as follows.

1) The most ideal form of uncertainty was standard deviation (e.g., 1 sd, 1 σ, and 1 sigma), which we collected directly from the literature.

2) If the uncertainty form was two standard deviations (e.g., 2 sd, 2 σ, and 2 sigma), we divided the original uncertainty by 2 and then compiled the new uncertainty into our database.

3) The uncertainty, which appeared as a percentage, represented a relative error. First, we multiplied this percentage by its corresponding isotope ratio or age to obtain an absolute error. Then, we repeated procedures 1) or 2) to normalize this absolute error into standard deviation.

3.2 Lithology classification

We categorized the lithology of the host rocks into three groups, namely, sedimentary, igneous, and metamorphic rocks. During the initial collection, we directly input the lithologies that were published along with the geochronological data in the original literature. However, there were more than 4000 phrases used in the literature to describe lithology, which brought great difficulties to the statistical analysis of data grouped by the host rock lithology. In this case, we extracted the original detailed lithologies and classified them into three major types, referring to the national standards of the Classification and Nomenclature Schemes of Igneous Rocks (GB/T 17412.1-1998) (Qin et al., 1998), Sedimentary Rocks (GB-T17412.2-1998) (Yang et al., 1998), and Metamorphic Rocks (GB/T 17412.3-1998) (Chong et al., 1998). We left the *Lithology* field empty if the original literature did not provide any lithology information. The null-value percentage was 9.70%.

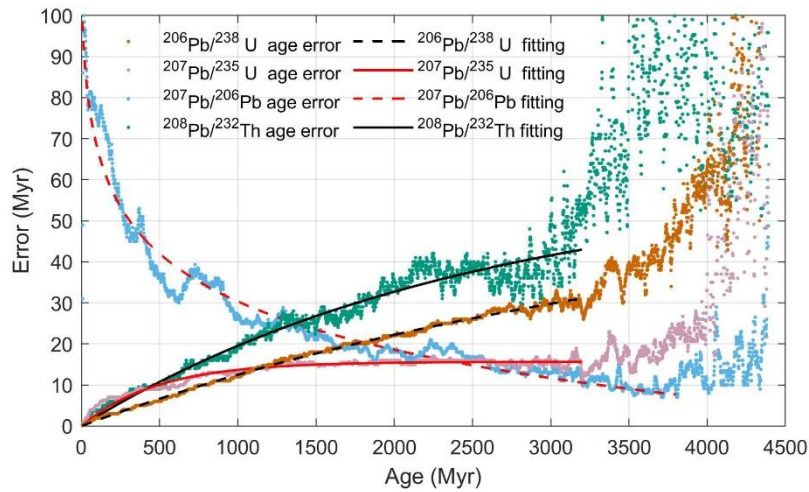
3.3 Grouping instruments

We cleaned the *Instrument* field and grouped them into four major groups: LA–ICP–MS, SHRIMP, SIMS, and TIMS. There were more than 300 instruments provided in the original literature. For research convenience, we categorized similar instruments into 4 major categories. The LA–ICP–MS instrument category includes LA–ICP–MS, LA–Q–ICP–MS, LA–SF–ICP–MS, LA–MC–ICP–MS, LAM–ICP–MS, etc. The SHRIMP instrument category includes SHRIMP, SHRIMP II, SHRIMP RG, etc. The SIMS instrument category contains SIMS, Cameca IMS 1270, Cameca IMS 1280, NORDSIM, etc. The TIMS instrument category includes TIMS, VG354 mass spectrometer, MAT 262 mass spectrometer, TRITON spectrometer, etc.

4. Data characteristics

4.1 Dating uncertainty

155 The zircon U–Th–Pb dating error is related to the zircon age and dating instruments. In this study, the age error within a certain time interval is obtained using the moving average method (Wu et al., 2022) (see Figure 2 and supplementary materials). When the sample is too old, the dating uncertainty of all age types increases significantly, which may be attributed to sample preservation. The error points for various ages were fitted to obtain error curves (see the equations in Table S1). For dating instruments, although TIMS is more precise, instruments such as LA–ICP–MS are more cost-effective and thus are more

160 widely used (Gehrels, 2014). This paper gives the relationship between the error and age of the four dating instruments of LA–ICP–MS, SHRIMP, SIMS, and TIMS. The numbers of chronological records for the instruments above are 1.6×10^6, 2.6×10^5, 8.1×10^4, and 3.3×10^4, respectively. The relationships between the various age errors of these four dating instruments are similar, but the specific intersection points of the curves are different (see Figure **3**, Figures S1-4 and Table S2).

165 **Figure 2. Age errors of four age dating types and their fitting curves. See Table S1 for more fitting information.**

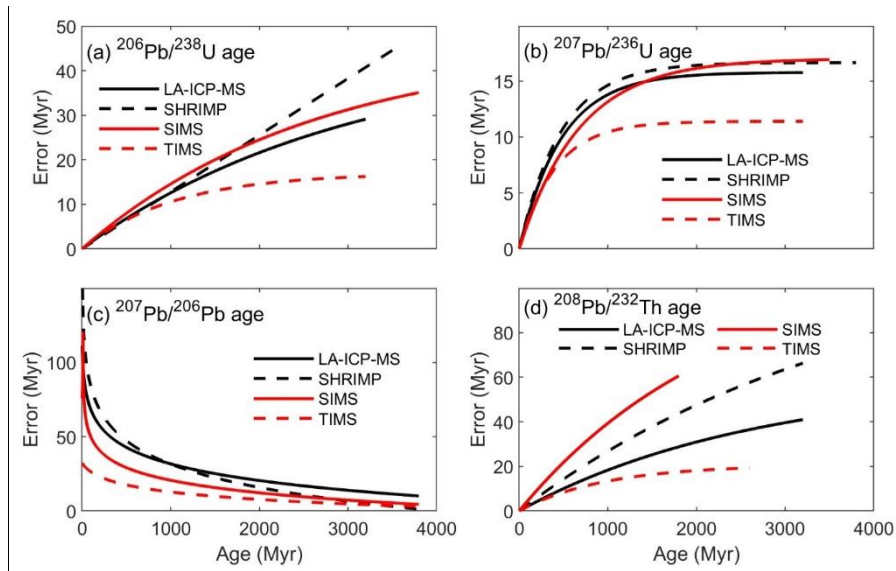



Figure 3. **Comparison of age errors of different dating instruments. (a)** ^{206}Pb/^{238}U **age; (b)** ^{207}Pb/^{235}U **age; (c)** ^{207}Pb/^{206}Pb **age; (d)** ^{208}Pb/^{232}Th **age.**

4.2 Temporal and spatial characteristics

170 Zircon production increases with magmatic and metamorphic activities. Therefore, the amount of zircon production can be used to understand the past intensity of geological activity (Arndt and Davaille, 2013). A simple and direct proxy is the number

of zircon age records for different geological times, which can indicate the intensity of magmatic and metamorphic activities (Wu et al., 2022). First, we used the noniterative probability model to derive a recommended age for each dating record (Puetz et al., 2021; Puetz and Spencer, 2023). Then, we conducted bootstrap resampling and Monte Carlo simulation with weights based on the spatiotemporal density to minimize the impact of biased sampling (Mehra et al., 2021; Keller and Schoene, 2012) (see the supplementary materials for detailed procedures). Figure 4 shows that the zircon age series of all lithologies is more similar to that of sedimentary rocks, which may be explained by zircons in sedimentary rocks being composed of a natural mixture of igneous and metamorphic zircons. Nevertheless, the zircon production peak periods reflected in the age series of sedimentary, igneous, and metamorphic rocks are basically consistent, i.e., ca. 800, 1000, 1850, 2500, 2700, 3200, and 3400 Ma in the Precambrian. Figure 5 is presented to more clearly show zircon production in the Phanerozoic, during which prominent zircon age peaks occurred ca. 50, 130, 250, 300, and 440 Ma, reflected in the zircons of all three lithologies. However, sedimentary rocks at 100 and 160 Ma; igneous rocks at 220 Ma; and metamorphic rocks at 220, 240, and 500 Ma have different zircon peaks. Specific geological meanings can be analyzed in depth with the help of this database as well as other geological evidence. Furthermore, Figure 4 and Figure 5 clearly show that the time series of zircon production is multiscale and periodic. Wu et al. (2022) systematically analyzed the periodicity of the zircon age series and finally gave the following cycles: ca. 800, 360, 220, 160, 69, 57, 44, 30, 20, and 17 Myr.

The database here supports the globally heterogeneous distribution of magmatic U–Pb ages (Hawkesworth et al., 2010; Puetz et al., 2017). The spatial evolution of zircon production can be obtained by extracting the GPS coordinates of zircon sampling sites in certain age intervals. Although the present geographical locations differ from those of the past, this spatial distribution still has indicative significance. According to the zircon production peak periods given above, the spatial distribution of zircons for these periods can be plotted. Because age error varies with age, we use a 50-Myr age interval for zircon peaks in the Phanerozoic and a 100-Myr interval for the Precambrian data. Because of the similarity in the spatial distribution of zircons with similar ages, this paper presents only the spatial pattern of the main peak periods, i.e., 50 ± 25, 130 ± 25, 250 ± 25, 440 ± 25, 1000 ± 50, 1850 ± 50, 2500 ± 50, and 3400 ± 50 Ma (Figure 6). Inevitably, some areas were oversampled, such as China and Europe, and overly old ages are sparse due to preservation. In this case, we should pay more attention to relative rather than absolute zircon densities when comparing regions.

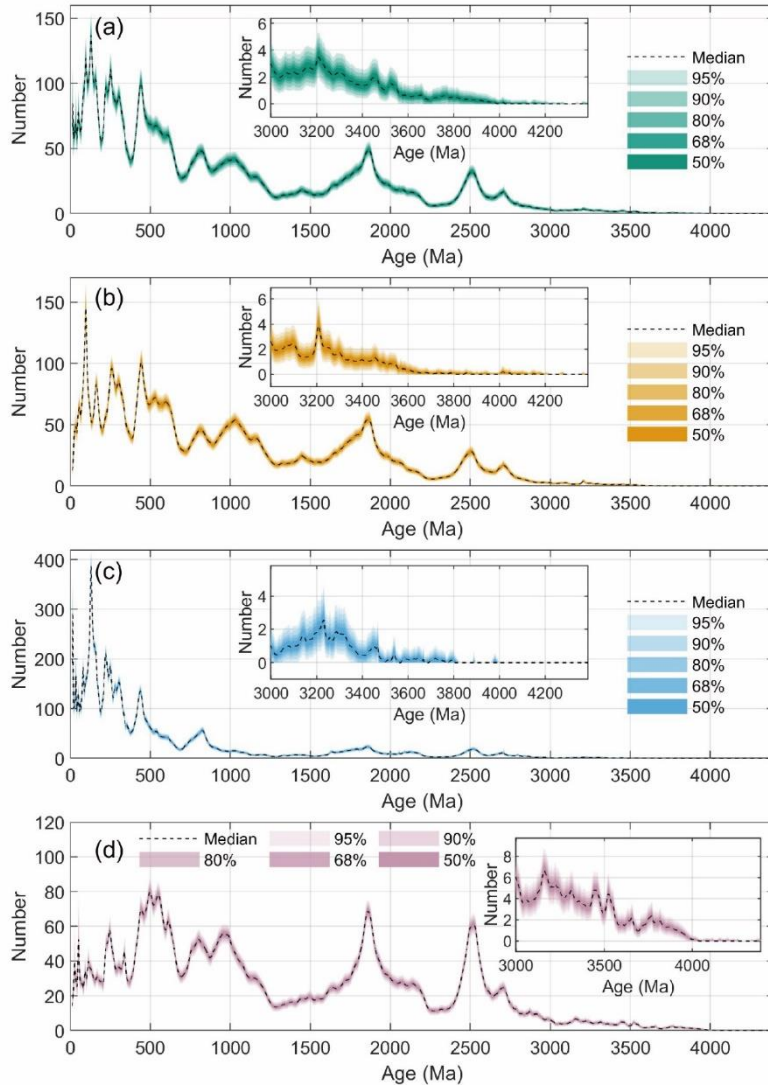



Figure 4. Zircon production series since 4.4 Ga. The host rocks of each zircon series are (a) of all lithologies, (b) sedimentary rocks, (c) igneous rocks, and (d) metamorphic rocks. The insets of panels (a-d) focus on the data before 3 Ga. Inverse proximity weighting, bootstrap resampling, and Monte Carlo simulation were applied to solve the biased sampling issue. The filled zones represent 10,000 Monte Carlo simulations of the zircon series with different transparencies, indicating different distribution probabilities, as shown in the legend. The dashed line indicates the median of the simulations. In each simulation, simulated zircon ages are selected based on a normal distribution. For more methodological details, see the supplementary materials.

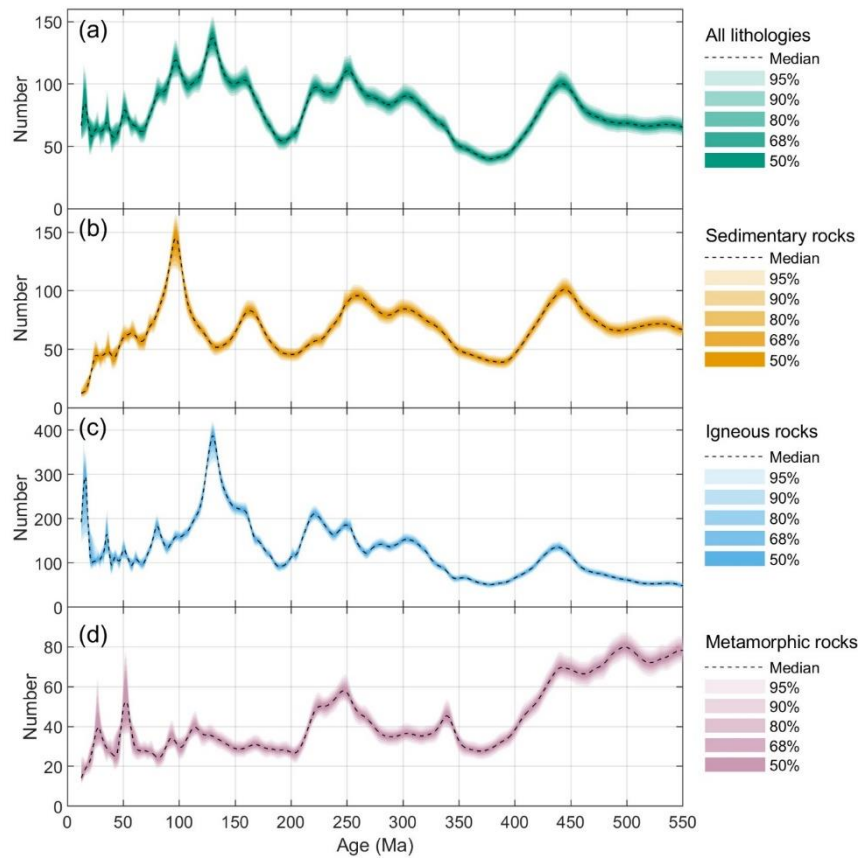



Figure 5. Zircon production series in the Phanerozoic. The host rocks of each zircon series are (a) of all lithologies, (b) sedimentary rocks, (c) igneous rocks, and (d) metamorphic rocks. Inverse proximity weighting, bootstrap resampling, and Monte Carlo simulation were applied to solve the biased sampling issue. The filled zones represent 10,000 Monte Carlo simulations of the zircon series with different transparencies, indicating different distribution probabilities, as shown in the legend. The dashed line indicates the median of the simulations. In each simulation, simulated zircon ages are selected based on a normal distribution. For more methodological details, see the supplementary materials.

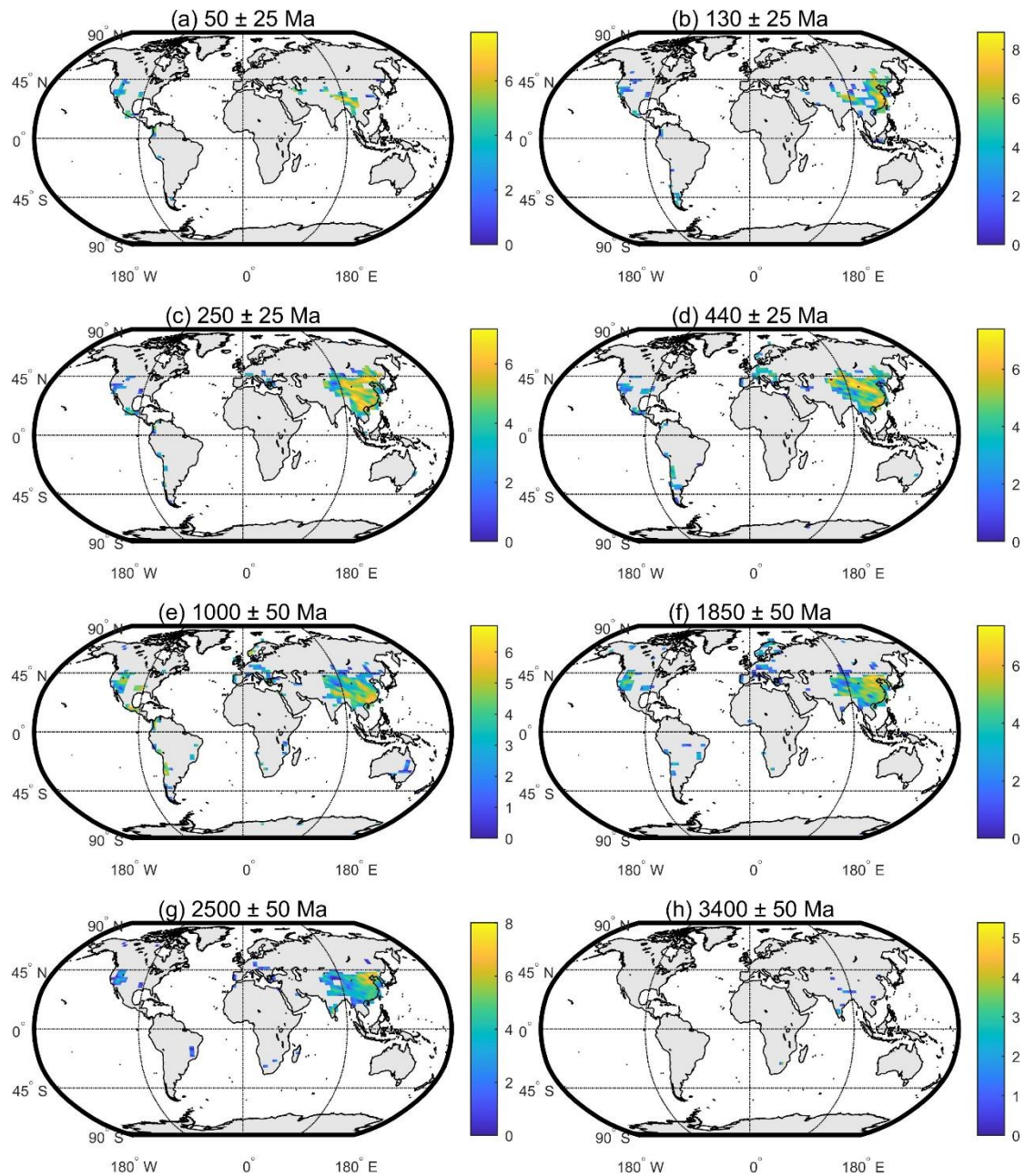



Figure 6. Spatial distribution of zircon production peak periods of all lithologies. Since some areas were oversampled and old ages
are sparse, we should pay more attention to relative rather than absolute zircon densities when comparing regions.

5. Discussion

5.1 Temporal and spatial evolution

The time series of zircon production indicates the evolution not only of the crust but also of the Earth system. The amount of zircon production can indicate the intensity of geological activity (Hawkesworth et al., 2010). However, using zircon production as a proxy involves making certain assumptions, which can be minimized if the dataset used is large enough. Fang et al. (2018) and Wu et al. (2019) used independent zircon datasets from the publishers CNKI and Elsevier to study the Chinese continental crust. The conclusions they obtained are surprisingly similar, which not only supports the reliability of big data statistics but also confirms that these geological events (periods of intense geological activity) are related to much zircon production. Apart from internal dynamic events (such as orogeny and plate movement), zircon production series can also indicate various associated surface processes, such as enhanced climate denudation during orogeny (Yu et al., 2011; Tu et al., 2015), the melting of ice sheets (Kim and Zhang, 2022; Mitrovica et al., 2001), weathering processes and atmospheric

composition (Kidder and Worsley, 2004; Campbell and Allen, 2008). Possible extraterrestrial drivers of geodynamics might also be explored by time series analysis of zircon production, providing information on plate tectonics, paleoclimate change, biological extinctions, etc. (Wu et al., 2022; Isley and Abbott, 2002; Puetz and Condie, 2019; Prokoph et al., 2013).

230   The spatial distribution of zircon production peaks of all lithologies is presented in Figure 6. Obviously, at different geological times, the regional intensity of geological activity varied as well as the amount of zircon production. For the spatial evolution of detrital, igneous, and metamorphic zircons, please see the supplementary materials (Figure S8-10). Fang et al. (2020) carried out a spatial classification of the Chinese continental crust by applying the grid clustering algorithm to a zircon dataset. Intriguingly, these crustal units identified by zircons are basically consistent with those based on tectonics (Yang and

235   Yu, 2015), verifying the scientific nature of zircon big data research and providing a new means of studying the spatial evolution of the crust. In the future, this database can also be used to study the spatial distribution of zircons on a global scale or in various regions to obtain the spatial evolution of the continental crust. With the help of zircon spatial distribution, the spatial evolution of global tectonic zones can also be further studied to explore the formation and storage of oil, gas, and minerals, which will be of great help to commercial mining. Furthermore, the implications of zircon production for geodynamic

240   processes may elucidate new energy sources.

5.2 Biased sampling

When using the number of zircon ages to indicate the intensity of geological activity, two conditions must be met: random exposure of the crust and random sampling of zircons. However, sampling bias, which occurs when zircons are oversampled or undersampled from a certain area or time interval, is inevitable (Wu et al., 2020; Puetz et al., 2017; Puetz et al., 2018). In

245   addition, considering the convenience and feasibility of fieldwork, sampling sites must be places that humans can reach. As shown in Figure 1, the sampling sites are dense in Europe and China but sparse in the Sahara and Siberia. Since zircon production is inherently unevenly distributed in the crust in both the spatial and temporal domains, the identification of biased sampled data is more complicated. Nevertheless, it is difficult to determine whether a large amount of zircon data is caused by artificial oversampling or crustal conditions that are suitable for zircon growth.

250   Puetz et al. (2017) proposed the methods of grid-area and modern-sediment sampling using the surface area to weigh zircon data. Although some important magmatic and metamorphic activities for crustal evolution (such as orogenic belts and subduction zones) were concentrated in certain regions and periods, the distributions of zircon ages were remarkably similar regardless of depth or height (Parman, 2015). Thus, this approach provides a simple and standard way to solve spatial sampling bias by weighing the records inversely proportional to sampling densities.

255   Alternatively, Wu et al. (2020) proposed the W and Y indices to measure the impact of biased sampling in the time domain. The W index should be used when artificial oversampling leads to a pseudo-peak without geological meaning in the zircon production series. Conversely, a zircon peak that should be prominent might be smoothed out due to undersampling around peaks or oversampling around troughs. This "homogeneous" effect can be evaluated by the Y index. In addition, Wu et al. (2022) proposed eliminating the influence of biased sampled data from the result based on coherence. The frequency signals

260   caused by the biased sampled data should exist only in the time series derived from the biased sampled region, suggesting little frequency similarity (low coherence) between the age series obtained from the normal sampled region and the biased sampled region. However, this method cannot quantitatively identify regions that tend to contain biased sampled data and needs to be further improved.

Keller and Schoene (2012) proposed an inverse spatiotemporal distance weighting method and applied it in conjunction

265   with bootstrap resampling and Monte Carlo simulation. This weighting approach calculated the spatiotemporal proximity for each sampling record, allowing us to balance the optimal sampling in space and through time at the same time (Mehra et al., 2021). Combined with bootstrap resampling and Monte Carlo simulation, the dating uncertainty could also be minimized. For comparison, we also obtained zircon production series data using only Monte Carlo simulation but without weighting, as in Wu et al. (2022) (see Figures S6 and S7 in the supplementary materials). After weighting using spatiotemporal density (Figure

270   S5, Figure 4 and Figure 5), however, the trends of the zircon age series did not present much change, which affirmed the statistical reliability of big data to some degree. New approaches can be explored to address biased sampling issues, and this zircon database provides great experimental materials.

6. Data availability

The database described in this manuscript can be accessed at the Zenodo repository at https://doi.org/10.5281/zenodo.8040079
275 (Wu et al., 2023).

7. Conclusions

Here, we introduce the largest known zircon U–Th–Pb geochronological database of the global continental crust. This database provides comprehensive research materials for Earth system science due to its large amount of data (~2 million records), wide sampling range (global continental crust), comprehensive samples (detrital, igneous, and metamorphic zircons), and various
280 dating instruments (LA–ICP–MS, SHRIMP, SIMS, and TIMS).

Based on this database, we described the characteristics of zircon dating errors, compared different dating instruments, and discussed biased sampling issues and possible solutions. Since the age errors vary over time and with dating instruments, the recommended ages should be selected with caution. Sampling bias would affect the objectivity of statistics to some degree. To solve this issue, we can weigh the data based on spatiotemporal sampling density in conjunction with bootstrapping
285 resampling and Monte Carlo simulation. However, we can always seek better solutions. The rich spatiotemporal information of this database provides a good experimental platform for exploring more potential methods.

This zircon database provides excellent materials for multiple fields of research, including but not limited to crustal growth and evolution, plate tectonics, and paleoclimate changes. The amount of zircon production can indicate the intensity of geological activities and be used to study the evolution of the continental crust and Earth system, whether from the global
290 or regional scale and whether for Earth's entire history or a specific period. In addition, this database has potential applications in the commercial mining of oil, gas, and minerals if associating structural geology with the temporal and spatial distribution of zircon production.

Supplement

295 **Author contributions**

YW, XF, and JJ compiled the data.
YW and JJ merged the data, formatted the data, performed the analyses, standardized the reference materials, organized the database, managed the publication of the database on the Zenodo repository, and drafted and revised the manuscript.
JJ initiated and supported data compilation.

Competing interests
300

The author for correspondence declares that neither they nor their coauthors have any competing interests to report.

Acknowledgments

We would like to thank the following people for their help in collecting the data: Muyuan Zhu, Sisi Liao, Lizhi Xue, Zhe Chen, Jiangnan Yang, Yamin Lu, Kun Ling, Shengyi Hu, Shuyuan Kong, Yiwei Xiong, Huacheng Li, Xiuqi Shang, Rui Ji, Xueyun
305 Lu, Biao Song, and Lei Zhang.

Financial support

None.

References

Arndt, N. and Davaille, A.: Episodic Earth evolution, Tectonophysics, 609, 661-674, 10.1016/j.tecto.2013.07.002, 2013.

310   Becker, J. S.: Inorganic Mass Spectrometry: Principles and Applications, John Wiley & Sons Ltd, Chichester, West Sussex, England2007.

Campbell, I. H. and Allen, C. M.: Formation of supercontinents linked to increases in atmospheric oxygen, Nature geoscience, 1, 554-558, 10.1038/ngeo259, 2008.

Carrapa, B.: Resolving tectonic problems by dating detrital minerals, Geology, 38, 191-192, 10.1130/focus022010.1, 2010.

315   Chong, R., Wang, M., Yang, L., and Qin, B.: Classification and namenclature schemes of metamorphic rocks: GB/T 17412.3-1998, in: Classification and namenclature schemes of the rocks, Standards Press of China, Beijing, 796-822, 1998.

Condie, K. C.: Preservation and recycling of crust during accretionary and collisional phases of proterozoic orogens: A bumpy road from nuna to rodinia, Geosciences (Basel), 3, 240-261, 10.3390/geosciences3020240, 2013.

Davis, D. W., Krogh, T. E., and Williams, I. S.: Historical development of zircon geochronology, Reviews in mineralogy and

320   geochemistry, 53, 145-181, 10.2113/0530145, 2003.

Fang, X., Wu, Y., Liao, S., Xue, L., Chen, Z., Song, B., Zhang, L., and Ji, J.: Preliminary analysis of the Chinese sub-database of the Chinese crust single-grain zircon U-Pb geochronology database, Acta Petrologica Sinica, 34, 3253-3265, 2018.

Fang, X., Wu, Y., Liao, S., Xue, L., Chen, Z., Yang, J., Lu, Y., Ling, K., Hu, S., Kong, S., Xiong, Y., Li, H., Shang, X., Ji, R., Lu, X., Song, B., Zhang, L., and Ji, J.: Division of crustal units in China using grid-based clustering and a zircon U–Pb

325   geochronology database, Computers & geosciences, 145, 104570, 10.1016/j.cageo.2020.104570, 2020.

Gehrels, G.: Detrital zircon U-Pb geochronology applied to tectonics, Annu Rev Earth Pl Sc, 42, 127-149, 10.1146/annurev-earth-050212-124012, 2014.

Hanchar, J. M. and Hoskin, P. W. O.: Zircon, Mineralogical Society of America; Geochemical Society, Chantilly, VA2018.

Hawkesworth, C. J., Dhuime, B., Pietranik, A. B., Cawood, P. A., Kemp, A. I. S., and Storey, C. D.: The generation and

330   evolution of the continental crust, J Geol Soc London, 167, 229-248, doi:10.1144/0016-76492009-072, 2010.

Hiess, J., Condon, D. J., McLean, N., and Noble, S. R.: 238U/235U Systematics in Terrestrial Uranium-Bearing Minerals, Science, 335, 1610-1614, doi:10.1126/science.1215507, 2012.

Isley, A. E. and Abbott, D. H.: Implications of the Temporal Distribution of High‐Mg Magmas for Mantle Plume Volcanism through Time, The Journal of geology, 110, 141-158, 10.1086/338553, 2002.

335   Jaffey, A., Flynn, K., Glendenin, L., Bentley, W. t., and Essling, A.: Precision measurement of half-lives and specific activities of U 235 and U 238, Physical review C, 4, 1889, 10.1103/PhysRevC.4.1889, 1971.

Keller, C. B. and Schoene, B.: Statistical geochemistry reveals disruption in secular lithospheric evolution about 2.5 Gyr ago, Nature, 485, 490-493, 10.1038/nature11024, 2012.

Kidder, D. L. and Worsley, T. R.: Causes and consequences of extreme Permo-Triassic warming to globally equable climate

340   and relation to the Permo-Triassic extinction and recovery, Palaeogeogr Palaeocl, 203, 207-237, 10.1016/S0031-0182(03)00667-9, 2004.

Kim, B. and Zhang, Y. G.: Methane hydrate dissociation across the Oligocene-Miocene boundary, Nature Geoscience, 15, 203-+, 10.1038/s41561-022-00895-5, 2022.

Krogh, T. E.: A low-contamination method for hydrothermal decomposition of zircon and extraction of U and Pb for isotopic

345   age determinations, Geochimica et Cosmochimica Acta, 37, 485-494, 10.1016/0016-7037(73)90213-5, 1973.

McKenzie, N. R., Horton, B. K., Loomis, S. E., Stockli, D. F., Planavsky, N. J., and Lee, C.-T. A.: Continental arc volcanism as the principal driver of icehouse-greenhouse variability, Science (American Association for the Advancement of Science), 352, 444-447, 10.1126/science.aad5787, 2016.

Mehra, A., Keller, C., Zhang, T., Tosca, N., Mclennan, S., Sperling, E., Farrell, U., Brocks, J., Canfield, D., and Cole, D.:

350   Curation and analysis of global sedimentary geochemical data to inform earth history, GSA Today, 10.1130/GSATG484A.1, 2021.

Mitrovica, J. X., Tamisiea, M. E., Davis, J. L., and Milne, G. A.: Recent mass balance of polar ice sheets inferred from patterns of global sea-level change, Nature, 409, 1026-1029, Doi 10.1038/35059054, 2001.

Parman, S.: Time-lapse zirconography: Imaging punctuated continental evolution, Geochemical Perspectives Letters, 1, 43-
355   52, 10.7185/geochemlet.1505, 2015.

Prokoph, A., El Bilali, H., and Ernst, R.: Periodicities in the emplacement of large igneous provinces through the Phanerozoic:Relations to ocean chemistry and marine biodiversity evolution, Di xue qian yuan., 4, 263-276, 10.1016/j.gsf.2012.08.001, 2013.

Puetz, S. J. and Condie, K. C.: Time series analysis of mantle cycles Part I: Periodicities and correlations among seven global
360   isotopic databases, Geoscience Frontiers, 10, 1305-1326, 10.1016/j.gsf.2019.04.002, 2019.

Puetz, S. J. and Spencer, C. J.: Evaluating U-Pb accuracy and precision by comparing zircon ages from 12 standards using TIMS and LA-ICP-MS methods, Geosystems and Geoenvironment, 2, 100177, 10.1016/j.geogeo.2022.100177, 2023.

Puetz, S. J., Spencer, C. J., and Ganade, C. E.: Analyses from a validated global U-Pb detrital zircon database: Enhanced methods for filtering discordant U-Pb zircon analyses and optimizing crystallization age estimates, Earth-Science Reviews,
365   220, 10.1016/j.earscirev.2021.103745, 2021.

Puetz, S. J., Ganade, C. E., Zimmermann, U., and Borchardt, G.: Statistical analyses of global U-Pb database 2017, Geoscience Frontiers, 9, 121-145, 10.1016/j.gsf.2017.06.001, 2018.

Puetz, S. J., Condie, K. C., Pisarevsky, S., Davaille, A., Schwarz, C. J., and Ganade, C. E.: Quantifying the evolution of the continental and oceanic crust, Earth-science reviews, 164, 63-83, 10.1016/j.earscirev.2016.10.011, 2017.

370   Qin, B., Wang, M., Chong, R., and Yang, L.: Classification and namenclature schemes of igneous rocks: GB/T 17412.1-1998, in: Classification and namenclature schemes of the rocks, Standards Press of China, Beijing, 741-775, 1998.

Song, B.: SHRIMP zircon U-Pb age measurement: Sample preparation, measurement, data processing and explanation, Geological Bulletin of China, 34, 1777-1788, 2015.

Spencer, C. J., Kirkland, C. L., and Taylor, R. J. M.: Strategies towards statistically robust interpretations of in situ U-Pb zircon
375   geochronology, Geoscience Frontiers, 7, 581-589, 10.1016/j.gsf.2015.11.006, 2016.

Tu, J.-Y., Ji, J.-Q., Sun, D.-X., Gong, J.-F., Zhong, D.-L., and Han, B.-F.: Thermal structure, rock exhumation, and glacial erosion of the Namche Barwa Peak, constraints from thermochronological data, Journal of Asian earth sciences, 105, 223-233, 10.1016/j.jseaes.2015.03.035, 2015.

Voice, P. J., Kowalewski, M., and Eriksson, K. A.: Quantifying the Timing and Rate of Crustal Evolution: Global Compilation
380   of Radiometrically Dated Detrital Zircon Grains, The Journal of geology, 119, 109-126, 10.1086/658295, 2011.

Wilde, S. A., Valley, J. W., Peck, W. H., and Graham, C. M.: Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago, Nature (London), 409, 175-178, 10.1038/35051550, 2001.

Williams, I. S.: Zircon, in: Encyclopedia of Astrobiology, edited by: M., G., Springer, 1790–1791, 2015.

Wu, Y., Fang, X., and Ji, J.: Global zircon U-Th-Pb geochronology database (v2), Zenodo [dataset], 10.5281/zenodo.8040079,
385   2023.

Wu, Y., Fang, X., Jiang, L., Song, B., Han, B., Li, M., and Ji, J.: Very long-term periodicity of episodic zircon production and Earth system evolution, Earth-Science Reviews, 104164, 10.1016/j.earscirev.2022.104164, 2022.

Wu, Y., Fang, X., Liao, S., Xue, L., Chen, Z., Yang, J., Lu, Y., Ling, K., Hu, S., Kong, S., Xiong, Y., Li, H., Shang, X., Ji, R., Lu, X., Song, B., Zhang, L., and Ji, J.: Zircon U-Pb geochronology of the Chinese continental crust: a preliminary analysis
390   of the Elsevier science database, Big earth data, 3, 26-44, 10.1080/20964471.2019.1576261, 2019.

Wu, Y., Fang, X., Liao, S., Xue, L., Chen, Z., Yang, J., Lu, Y., Ling, K., Hu, S., Kong, S., Xiong, Y., Li, H., Shang, X., Ji, R., Lu, X., Song, B., Zhang, L., and Ji, J.: Crustal evolution events in the Chinese continent: evidence from a zircon U-Pb database, International journal of digital earth, 13, 1532-1552, 10.1080/17538947.2020.1739152, 2020.

Yang, L., Han, G., Wang, M., Qin, B., and Chong, R.: Classification and namenclature schemes of sedimentary rocks: GB/T
395   17412.2-1998, in: Classification and namenclature schemes of the rocks, Standards Press of China, Beijing, 776-795, 1998.

Yang, W. C. and Yu, C. Q.: Tectonic divisions of the Chinese continental lithosphere based on forming tectonic processes,

Geol. Rev., 61, 709-716, 2015.

Yu, X., Ji, J., Gong, J., Sun, D., Qing, J., Wang, L., Zhong, D., and Zhang, Z.: Evidences of rapid erosion driven by climate in the Yarlung Zangbo (Tsangpo) Great Canyon, the eastern Himalayan syntaxis, Chinese science bulletin, 56, 1123-1130, 10.1007/s11434-011-4419-x, 2011.

400