Deep-Time Marine Sedimentary Element Database

Jiankang Lai[1], Haijun Song[1*], Daoliang Chu[1], Jacopo Dal Corso[1], Erik A. Sperling[2], Yuyang Wu[1*], Xiaokang Liu[1], Lai Wei[3], Mingtao Li[4], Hanchen Song[1], Yong Du[1], Enhao Jia[1], Yan Feng[1], Huyue Song[1], Wenchao Yu[1], Qingzhong Liang[5], Xinchuan Li[5], Hong Yao[5]

[1]State Key Laboratory of Biogeology and Environmental Geology, School of Earth Sciences, China University of Geosciences, Wuhan 430074, China

[2]Department of Earth and Planetary Sciences, Stanford University, Stanford, CA 94305, USA.

[3]School of Future Technology, China University of Geosciences, Wuhan 430074, China

[4]School of Resources and Environment, Linyi University, Linyi 276000, China

[5]School of Computer Science, China University of Geosciences, Wuhan 430074, China

Corresponding authors: Haijun Song (haijunsong@cug.edu.cn), Yuyang Wu (wuyuyang@cug.edu.cn)

Abstract. Geochemical data from ancient marine sediments are crucial to study palaeoenvironments, palaeoclimates, and elemental cycles. With increased accessibility to geochemical data, many databases have emerged. However, there remains a need for a more comprehensive database that focuses on deep-time marine sediment records. Here, we introduce the "Deep-Time Marine Sedimentary Element Database" (DM-SED). The DM-SED has been built upon the "Sedimentary Geochemistry and Paleoenvironments Project" (SGP) database with the new compilation of 34,874 data entries from 433 studies, totalling 63,627 entries. The DM-SED contains 2,522,255 discrete marine sedimentary data points, including major and trace elements and some stable isotopes. It includes 9,207 entries from the Precambrian and 54,420 entries from the Phanerozoic, thus providing significant references for reconstructing deep-time Earth system evolution. The data files described in this paper are available at https://doi.org/10.5281/zenodo.14771859 (Lai et al., 2025).

1 Introduction

Geochemical data from deep-time marine sediments are fundamental for reconstructing the evolution of the Earth system. By analysing the content of chemical elements in sediments and their isotopic compositions, we can reconstruct the past cycling of elements in the Earth's surface systems and reveal the evolution of Earth's surface systems through time (Large et al., 2015; Reinhard et al., 2017; Farrell et al., 2021; Planavsky et al., 2023). For instance, total organic carbon (TOC), phosphorus (P), biogenic barium (Ba$_{bio}$), copper (Cu), zinc (Zn), nickel (Ni), etc., enable reconstructing marine primary productivity and carbon cycle perturbations, thereby revealing mechanisms driving past climate fluctuations (Scott et al., 2013; Schoepfer et al., 2015; Shen et al., 2015; Schoepfer et al., 2016; Xiang et al., 2018; Jin et al., 2020; Tribovillard, 2021; Wang et al., 2022; Zhang et al., 2022; Li et al., 2023; Sweere et al., 2023; Zhao et al., 2023). Elements such as uranium (U), vanadium (V), and molybdenum (Mo) can reveal how marine redox conditions changed during critical periods in animal evolution, including mass extinctions and evolutionary radiations (Algeo and Liu, 2020; Schobben et al., 2020; Stockey et al., 2024). Oxygen isotopes ($\delta^{18}O$) from fossilized marine organisms can reveal oceanic palaeo-temperature changes (Veizer and Prokoph, 2015; Song et al., 2019; Grossman and Joachimski, 2020; Scotese et al., 2021; Judd et al., 2022). However, many geochemical studies have focused on high-resolution research of limited time intervals and/or regions, and there is little comprehensive exploration across large-scale geological time and globally.

Fortunately, with more journals and institutions adopting strict data archiving rules and promoting adherence to FAIR (Findability, Accessibility, Interoperability, and Reusability) principles (Wilkinson et al., 2016; "FAIR Play in Geoscience Data," 2019), a large amount of geochemical data has become accessible, and sample meta-data records are more detailed. Several geochemical databases of varying scales and foci have emerged, such as the following:

- EarthChem, which covers igneous, sedimentary, and metamorphic rocks and comprises numerous joint databases (https://www.earthchem.org/, last accessed: 16 July 2024).
- Petrological Database of the Ocean Floor (PetDB), which includes elemental chemical, isotopic, and mineralogical data of global ocean floor igneous rocks, metamorphic rocks, minerals, and inclusions (https://www.earthchem.org/petdb, last accessed: 16 July 2024).
- Geochemistry of Rocks of the Oceans and Continents (GEOROC), a comprehensive compilation

of chemical, isotopic, and other data on igneous rock samples, including whole rock, glass, mineral, and inclusion analyses and metadata (http://georoc.mpch-mainz.gwdg.de, last access: 16 July 2024).

- Data Publisher for Earth & Environmental Science (PANGAEA), which is used for archiving, publishing, and disseminating georeferenced data from earth, environmental, and biodiversity sciences and includes a large number of sediment core data (https://www.pangaea.de, last accessed: 16 July 2024).

- Stable Isotope Database for Earth System Research (StabisoDB) containing $\delta^{18}O$ and $\delta^{13}C$ data for more than 67,000 macrofossil and microfossil samples, including benthic and planktonic foraminifera, benthic and nektonic mollusks, brachiopods, fish teeth, and conodonts (https://cnidaria.nat.uni-erlangen.de/stabisodb/, last accessed: 16 July 2024).

- Sedimentary Geochemistry and Paleoenvironments Project (SGP), which collects multi-proxy sedimentary geochemical data with an emphasis on Neoproterozoic-Palaeozoic shale data in its first data release (https://sgp-search.io/, last accessed: 12 June 2024).

- NOAA and MMS Marine Minerals Geochemical Database, which contains geochemical analyses and auxiliary information on present-day marine deposits of primarily ferromanganese nodules and crusts, as well as some data for heavy minerals and phosphorites (https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.geology:G01323, last accessed: 7 January 2025).

- An International Study of the Marine Biogeochemical Cycles of Trace Elements and Isotopes (GEOTRACES), which provides hydrographical and marine geochemical data acquired over the past decade (https://www.geotraces.org/, last accessed: 7 January 2025).

Many other government initiatives also host databases:

- The United States Geological Survey (USGS) National Geochemical Database, an archive of geochemical information and related metadata from USGS research (https://www.usgs.gov/energy-and-minerals/mineral-resources-program/science/national-geochemical-database, last accessed: 16 July 2024).

- The British Geological Survey (BGS), which provides data and information on UK geology, boreholes, geomagnetism, groundwater, rocks, etc. (http://www.bgs.ac.uk/, last accessed: 16 July 2024).

89   ●   The Australian National Whole Rock Geochemistry Database (OZCHEM), including chemical

90   compositions of rock, soil, and sediment samples (https://ecat.ga.gov.au/geonetwork/srv/, last

91   accessed: 16 July 2024).

92   Although some of these databases (Table 1) include data on ancient marine sediments, they have

93 shortcomings such as limited spatial coverage, the lack of age data and coarse age resolution, the

94 absence of recent publications, and missing information from original publications. Thus, we propose

95 the Deep-Time Marine Sedimentary Element Database (DM-SED), which focuses on the elemental

96 content changes in marine sediments across geological history. The current version of the DM-SED

97 database contains 63,627 entries, enabling research on a series of scientific issues related to

98 palaeoenvironmental, palaeoclimatic, and elemental cycles in deep-time Earth history.

99   DM-SED version 0.0.1 is presented in table (.csv) format. Dynamic versions of the most recent

100 release can be found on Zenodo (https://doi.org/10.5281/zenodo.14771859, last accessed: 30 January

101 2025) (Lai et al., 2025), and a static copy of Version 0.0.1 is archived in the Geobiology database

102 (http://geobiologydata.cug.edu.cn/, last accessed: 25 September 2024). In the following sections, we

103 provide a brief overview of the database, information on the data sources and selection criteria, and a

104 review of the definitions and decisions behind the metadata fields associated with each proxy

105 measurement. We explore the spatial and temporal distribution trends of the compiled data and discuss

106 future uses and limitations of the database.

107

108

109

110

111

112

113

114

115

116

117

118

119 **Table 1. Overview of different databases (Note: not all databases have a clear number of records).**

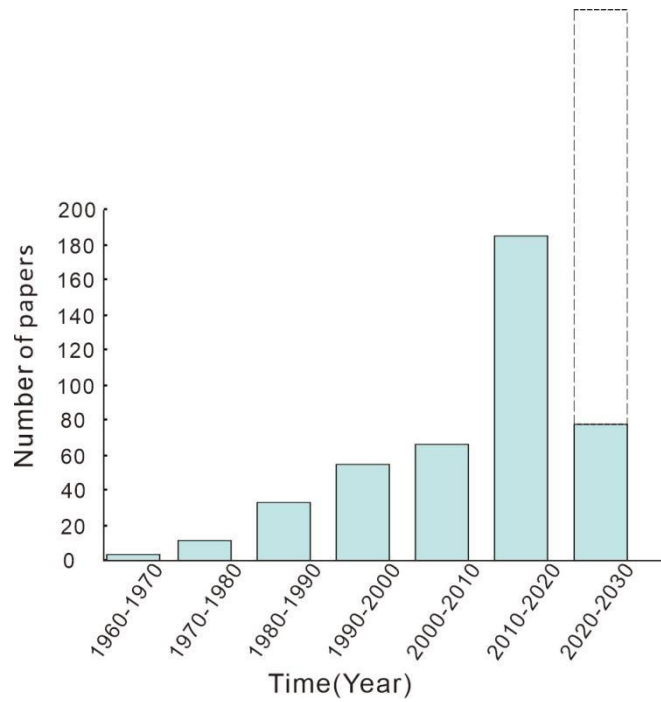
Database name	Content	Website information	Number of records	Data regions
EarthChem	Igneous, sedimentary, and metamorphic rocks; various joint databases	https://www.earthchem.org/, last accessed: 16 July 2024	Over 2,596 digital content files in EarthChem Library	Global
PetDB	Elemental chemical, isotopic, and mineralogical data of global ocean floor rocks	https://www.earthchem.org/petdb, last accessed: 16 July 2024	Over 6,000,000 samples	Global
GEOROC	Chemical, isotopic, and other data on igneous rock samples	http://georoc.mpch-mainz.gwdg.de, last access: 16 July 2024	672,990 samples	Global
PANGAEA	Georeferenced data from earth, environmental, and biodiversity sciences	https://www.pangaea.de, last accessed: 16 July 2024	Extensive dataset	Global
StabisoDB	$\delta^{18}O$ and $\delta^{13}C$ data for macrofossil and microfossil samples	https://cnidaria.nat.uni-erlangen.de/stabisodb/, last accessed: 16 July 2024	Over 67,000 samples	Global
SGP	Multi-proxy sedimentary geochemical data from the Palaeozoic and Neoproterozoic	https://sgp-search.io/, last accessed: 12 June 2024	82,578 samples	Global
NOAA and MMS Marine Minerals Geochemical Database	Geochemical analyses on ferromanganese nodules and crusts, as well as some heavy minerals and phosphorites	https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.geology:G01323, last accessed: 7 January 2025	Over 140,000 element/oxide analyses	Global
GEOTRACES	Hydrographical and marine geochemical data	https://www.geotraces.org/, last accessed: 7 January 2025	77 cruises and more than 800 hydrographic and geochemical parameters	Global
USGS	Geochemical information and related metadata from USGS research	https://www.usgs.gov/energy-and-minerals/mineral-resources-program/science/national-geochemical-database, last accessed: 16 July 2024	Extensive dataset	United States
BGS	Data on UK geology, boreholes, geomagnetism, groundwater, rocks, etc.	http://www.bgs.ac.uk/, last accessed: 16 July 2024	Extensive dataset	United Kingdom
OZCHEM	Chemical compositions of rock, soil, and sediment samples	https://ecat.ga.gov.au/geonetwork/srv/, last accessed: 16 July 2024	Extensive dataset	Australia

2 Dataset overview

The DM-SED aims at collecting geochemical data from deep-time marine sediments. The database is primarily sourced from the SGP database, supplemented with additional 34,874 newly compiled entries. The SGP contains a total of 82,578 entries, from which we selected 28,753 entries specifically related to marine sedimentary geochemical data, and is comprised of three parts: two parts from the U.S. Geological Survey (USGS), i.e. the National Geochemical Database (USGS NGDB, https://mrdata.usgs.gov/ngdb/rock, last accessed: 9 September 2024) and the Global Geochemical Database for Critical Metals in Black Shales project (USGS CMIBS, Granitto et al., 2017), with samples mainly from North America and Phanerozoic shales from various continents, respectively (Farrell et al., 2021). The third part comprises direct inputs by SGP members. The direct inputs in the Phase 1 SGP data release primarily focused on Neoproterozoic–Palaeozoic shales, although there are other lithologies and other time periods represented (Farrell et al., 2021). Our DM-SED database, built upon the SGP, includes a new compilation of 34,874 entries from 433 studies, spanning approximately 3800 Ma and including entries from North America, Europe, Asia, Africa, South America, Oceania, Pacific and Atlantic. This supplements the temporal and spatial distribution gaps in the SGP database, thereby creating a more comprehensive sedimentary marine geochemical database. The new compiled literatures span the time range from 1965 to 2023, with the number of papers per decade gradually increasing (Fig. 1). It should be noted that the top of the DM-SED version 0.0.1 data is the new compilation, and the bottom contains data imported from SGP.

Table 2. Summary of data entries and points in the DM-SED.

	Entries	Data points
New compilation	34,874	1,454,400
SGP	28,753	1,067,855
DM-SED	63,627	2,522,255

141

142 **Figure 1. The distribution of publication years for newly compiled literature (the dashed line denotes the**

143 **predicted literature from 2023 to 2030).**

144

145   The DM-SED database comprises 63,627 entries with 2,522,255 discrete data points (Table 2),

146 each including location (SampleID, SampleName, SiteName, Region, Elevation, SampleDepth,

147 ModLat, ModLon, PalaeoLat, PalaeoLon), age (Age, Period, Stage, Biozone), stratigraphic information

148 (LithName, LithType, Formation, Facies), carbon element (total carbon (Total C), inorganic carbon

149 (C_{inorg}), TOC, in wt%), isotopic values ($\delta^{18}O_{carb}$, $\delta^{13}C_{Ker}$, $\delta^{13}C_{TOC}$, $\delta^{13}C_{carb}$, $\delta^{34}S_{CAS}$, $\delta^{34}S_{pyr}$, $\delta^{15}N_{total}$,

150 $\delta^{15}N_{org}$, in ‰), major element (P, Al, Si, Ti, Fe, Ca, Mg, Na, K, S, N, in wt%), trace element (Ag, Ar,

151 As, B, Ba, Be, Bi, Br, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Ga, Gd, Ge, Hf, Hg, Ho, In, La, Li, Lu, Mn,

152 Mo, Nb, Nd, Ni, Pb, Pr, Rb, Re, Sb, Sc, Se, Sm, Sn, Sr, Ta, Tb, Te, Th, Tl, Tm, U, V, W, Y, Yb, Zn, Zr,

153 in ppm), methodology (TOC methods, Major elements methods, Trace elements methods), and data

154 sources (Reference, Project). The specific names and descriptions of each field in the database are

155 shown in Table 3. The standards and descriptions of isotope ratios in the database are shown in Table

156 4.

157

158

159

160 **Table 3. Field names and descriptions.**

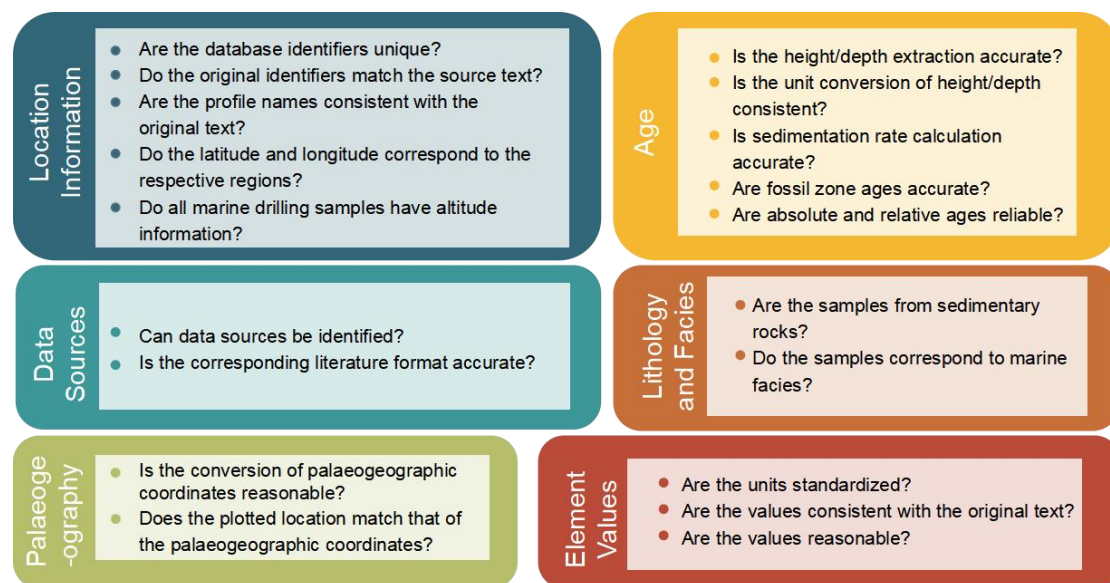
Field name	Description of field (units)
Location fields	
SampleID	Unique sample identification code
SampleName	Author denoted title for the sample (often non-unique)
SiteName	Name of the drill core site or section
Region	Country or ocean of the data collection site
Elevation	Distance between sampling location and sea level (m)
SampleDepth	Stratigraphic height or depth (m)
ModLat	Modern latitude of collection site rounded to two decimals; negative values indicate the Southern Hemisphere (decimal degrees)
ModLon	Modern longitude of the collection site rounded to two decimals; negative values indicate the Western Hemisphere (decimal degrees)
PalaeoLat	Palaeolatitude of collection site rounded to two decimals; negative values indicate the Southern Hemisphere (decimal degrees)
PalaeoLon	Palaeolongitude of the collection site rounded to two decimals; negative values indicate the Western Hemisphere (decimal degrees)
Age fields	
Age	Absolute Age, in reference to GTS2020 (Ma)
Period	The geologic period
Stage	The geologic stage (i.e. geochronologic age)
Biozone	Conodont, graptolite, ammonite biozone, etc
Stratigraphy	
LithName	Lithological name of the sample, as originally published
LithType	Lithology type of sample (e.g. carbonate, siliciclastic)
Formation	Geologic formation name
Facies	Depositional environment (e.g. mid-shelf, ramp)
Proxy fields	
Carbon	The content of carbon, including Total C, C_{inorg}, TOC, rounded to two decimals (wt%)
Isotopes	The isotope value, rounded to two decimals (‰)
Major elements	The content of major elements such as P, Al, and Si, rounded to two decimals (wt%)
Trace elements	The content of trace elements such as Ag, Ar, As, B, and Ba, rounded to two decimals (ppm)
Methodology	
TOC methods	A brief description of the testing methods for TOC
Major elements methods	A brief description of the testing methods for major elements
Trace elements methods	A brief description of the testing methods for trace elements
Data sources	
Reference	Data sources, including published literature or other databases
Project	Two parts: new compilation and SGP

161

Table 4. Standards and descriptions of isotope ratios in the DM-SED.

Symbol	Standard	Description
$\delta^{18}O_{carb}$	Vienna Pee Dee Belemnite (VPDB)	Oxygen isotope ratio of carbonate minerals, used in palaeoclimate studies.
$\delta^{13}C_{Ker}$	VPDB	Carbon isotope ratio of kerogen, used to study the source and depositional environment of organic matter.
$\delta^{13}C_{TOC}$	VPDB	Carbon isotope ratio of total organic carbon, used to analyse the source of organic matter and biogeochemical cycles in sediments.
$\delta^{13}C_{carb}$	VPDB	Carbon isotope ratio of carbonate minerals, used in palaeoclimate and carbon cycle research.
$\delta^{34}S_{CAS}$	Vienna Canyon Diablo Troilite (VCDT)	Sulfur isotope ratio of carbonate-associated sulfate, used to study the sulfur cycle and redox conditions.
$\delta^{34}S_{pyr}$	VCDT	Sulfur isotope ratio of pyrite, typically used to investigate the sulfur cycle and redox conditions in ancient oceans.
$\delta^{15}N_{total}$	Atmospheric Nitrogen (air N_2)	Nitrogen isotope ratio of total nitrogen, used to study the nitrogen cycle and nutrient sources.
$\delta^{15}N_{org}$	air N_2	Nitrogen isotope ratio of organic nitrogen, often used to analyse the source of organic matter and the nitrogen cycle.

3 Dataset screening and processing

This section details the screening and processing criteria for sample location, age, lithology and facies, specific geochemical values, and data source information (Fig. 2).

168 **Figure 2. The data filtering and processing criteria for DM-SED.**

169  For sample location, the database includes SampleID, SampleName, SiteName, Region, Elevation,

170 SampleDepth, ModLat, ModLon, PalaeoLat, and PalaeoLon. A unique SampleID is assigned to each

171 sample in the DM-SED. The SampleName corresponds to the identifier given in each referenced

172 publication, facilitating cross-referencing with the original data. The SiteName includes well name or

173 outcrop information, representing the smallest unit of location information. The Region indicates the

174 country or ocean area where the sample has been collected and represents a broader geographical range.

175 The Elevation data are mainly related to samples from the Deep Sea Drilling Project (DSDP) and the

176 Ocean Drilling Program (ODP) collected from post-Cretaceous sediments and indicate whether the

177 samples originate from deep or shallow marine environments. SampleDepth refers to the relative

178 position (in metres) of the sample within the well or outcrop, which is crucial for calculating sample

179 age. In some publications, specific heights are not provided directly but are given as relative heights

180 through figures. We manually extracted these heights using WebPlotDigitizer, rounding to two decimal

181 places (Drevon et al., 2017). For publications in which heights are expressed in feet or centimetres, we

182 converted the units to metres. Modern latitude and longitude (ModLat and ModLon) information are

183 the most precise location data. Although some publications provide exact coordinates, many offer only

184 section names (i.e. SiteName) and regions or merely a map marking the location of the section. For

185 publications providing section names, we determined accurate coordinates by consulting other studies

186 carried out in the same section. For those providing only a map marking the location of the section, we

187 used Google Maps to estimate relative coordinates. To ensure consistency, we recorded sample

188 coordinates in decimal degrees, rounded to two decimal places, with positive values indicating north

189 latitude and east longitude and negative values indicating south latitude and west longitude. The

190 coordinate reference system is WGS 84 (World Geodetic System 1984). For palaeo-coordinates, we

191 reconstructed palaeo-latitude and palaeo-longitude (PalaeoLat, and PalaeoLon) using the sample age

192 and modern coordinates, employing the PointTracker v7 rotation files from the PALEOMAP project,

193 which are based on current geographic reference data and global tectonic history models (Scotese,

194 2008). It is important to note that we only generated palaeogeographic locations for samples from the

195 Phanerozoic, as the geological records from this time are more complete and abundant compared to

196 those from the Precambrian, making the reconstruction of geographic features (such as ancient oceans,

197 mountains, plains, etc.) relatively more reliable and accurate (Scotese and Wright 2018). We plotted

198 the sample points on palaeogeographic maps based on Scotese's data using QGIS 3.16 (Scotese and

199 Wright 2018).

200 To assign specific ages to each sample in the database, we assumed a constant sedimentation

201 rate within the same formation or group of section. If the original studies provided numerical ages for

202 two or more samples, we calculated the precise age for each sample based on the sedimentation rate

203 and assigned it accordingly. If absolute ages were not provided in the original literature, we assigned

204 approximate ages based on corresponding fossil zones or the general age of the same lithostratigraphic

205 unit in the same region (Farrell et al., 2021; Judd et al., 2022). For samples with completely missing

206 height information in the original text, we assigned the same age to all samples within the section based

207 on lithostratigraphic information. However, the primary age constraints for these samples (mainly from

208 USGS NGDB and USGS CMIBS) remain derived from SGP age calls. Once each sample had a

209 specific age, we assigned it to a specific Period and Stage according to its age. We attempted to

210 incorporate the most recent age models; however, due to the extensive size of the data compilation, it

211 was not feasible to update all of them. All ages were based on the timescale provided by the Geologic

212 Time Scale 2020 (Gradstein et al., 2020). Although GTS 2020 is accurate, readers are advised to

213 consult the incremental updates of the International Chronostratigraphic Chart (ICS) for the most

214 accurate stratigraphic intervals.

215 For lithology and facies, the lithologies include shale, mudstone, sandstone, limestone, dolostone,

216 and others. We classified these into two types of rocks: siliciclastic sedimentary rocks (88.7%) and

217 carbonate rocks (11.3%). For outcrop sections, the lithostratigraphic unit was generally available;

218 however, for data from marine drilling sites, there was often no corresponding lithostratigraphic unit

219 information. Regarding facies classification, before the Cretaceous, the primary depositional

220 environment was marine settings on continental crust, including specific facies such as inner shelves,

221 outer shelves, and basinal. However, after the Cretaceous, with most samples coming from the DSDP

222 and ODP, shallow marine depositional environments still existed and were sampled, but deep-sea

223 pelagic settings began to be sampled as well.

224 For specific geochemical values in the DM-SED database, we standardized the units, converting

225 oxides to elements (e.g. P (ppm) to P (wt%), P_2O_5 (wt%) to P (wt%)). If a sample was analysed

226 multiple times, we averaged the value. For literature before 2000, some data were preserved as images,

227 requiring manual extraction of values, and some images were slightly blurry, potentially leading to

228 minor human error. We excluded data that were beyond detection limits (e.g. the trace element content

229 is too low and the value provided in the text represents the minimum detection limit) or unreasonable

230 (e.g. negative values for major and trace elements).

231 For the geochemical methodology, we briefly documented them based on the descriptions in the

232 original text, focusing primarily on the testing methods and instrument models used for TOC, major

233 elements, and trace elements. Methods for stable isotopes were not documented, as the testing methods

234 vary for different isotopes, and due to the limited amount of isotope data, recording them holds little

235 significance.

236 Regarding data sources, we ensured that each corresponding reference was collected and listed in

237 full citation format, including authors, title, publication date, journal, page numbers, and DOI. Most

238 data in the SGP database came directly from USGS NGDB and USGS CMIBS, without corresponding

239 literature sources, so we marked them individually. The entire database for this Project was divided

240 into two parts: new compilation and SGP. We used keyword searches in Google Scholar to identify

241 missing references and made efforts to target literature for data-scarce regions (e.g. South America)

242 and time intervals (e.g. Silurian, Jurassic).

243 **4 Data distribution**

244 The elemental data content distribution for the entire database is shown in Fig. 3. Overall, major

245 elements have the highest data quantity, followed by trace elements and carbon elements, with isotope

246   data having the lowest quantity. Among the major elements, N has the fewest entries, with 3,164

247   records, whereas the other major elements all have more than 10,000 entries. Al has the highest

248   quantity, with 51,906 records. Among trace elements, Mn has the largest record (42,499 records),

249   followed by Ba (41,471 records). Ar and Br have the fewest records, with 9 and 176 records,

250   respectively. Other elements such as Ag, B, Bi, Ge, Hg, In, Re, Se, Sn, Te, Tl, Tm, and W have data

251   quantities ranging from 1,000 to 10,000. Elements such as As, Be, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu,

252   Ga, Gd, Hf, Ho, La, Li, Lu, Mo, Nb, Nd, Ni, Pb, Pr, Rb, Sb, Sc, Sm, Sr, Ta, Tb, Th, U, V, Y, Yb, Zn,

253   and Zr all have more than 10,000 records each. For carbon elements, TOC has the largest record

254   (32,906 entries), followed by Total C (9,386 entries), while C_{inorg} has the lowest record (7,215 entries).

255   Isotope data are overall less abundant, with none exceeding 10,000 entries; the most abundant is

256    $\delta^{13}C_{TOC}$, with 8,166 records, and the least abundant is $\delta^{13}C_{Ker}$, with only 112 records.

257

258

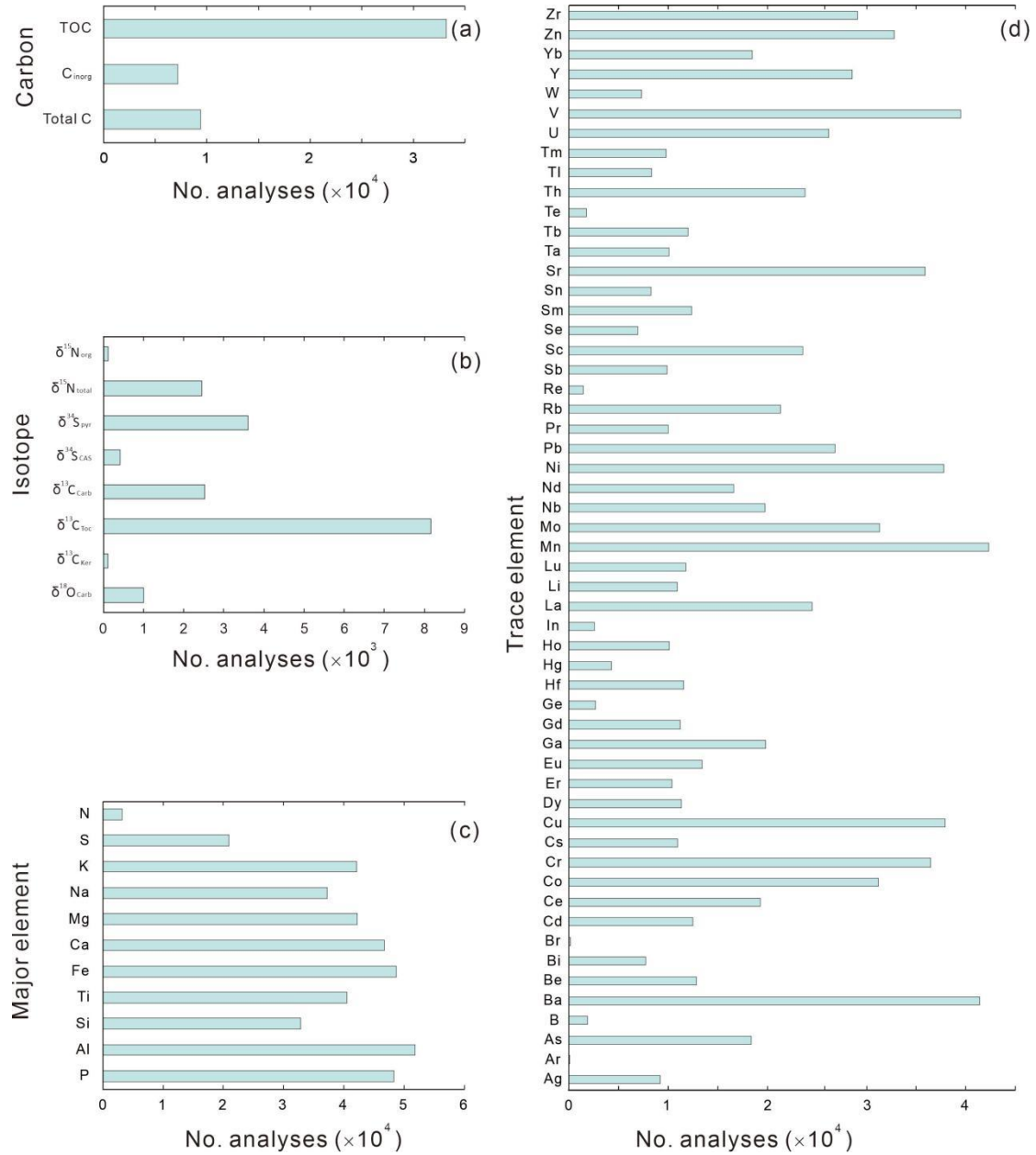



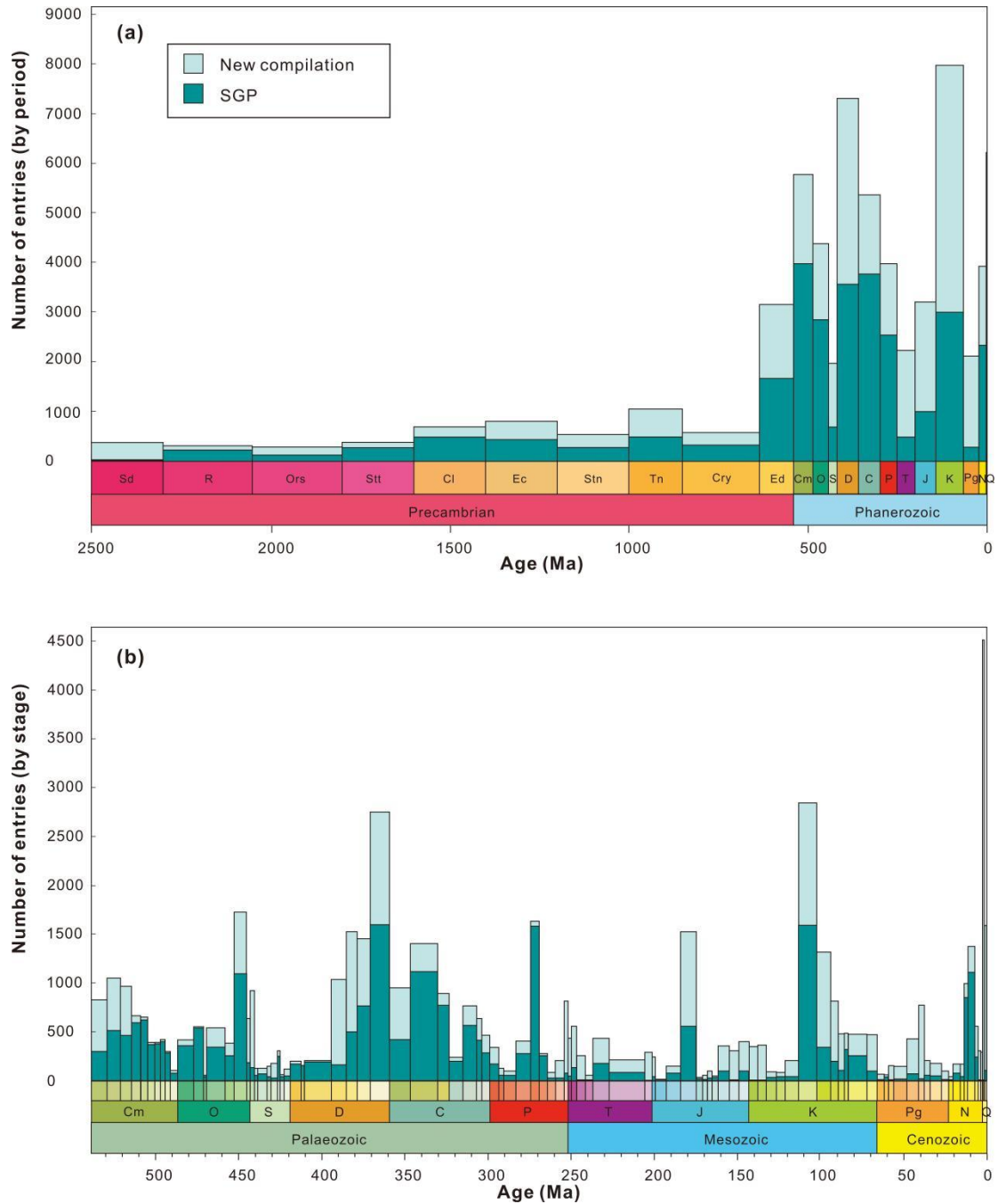
Figure 3. Histogram distribution of different subsets. (a) Carbon elements. (b) Isotopes. (c) Major elements.

(d) Trace elements.

The temporal trend of data density in the entire database, shown in Fig. 4a, indicates that the data are primarily distributed in the Phanerozoic Eon, which accounts for 85% of the entire database. Within the Phanerozoic, the Cenozoic Era accounts for 19% of the database, the Mesozoic Era accounts for 21%, and the Palaeozoic Era accounts for 45%. Precambrian data account for only 15% of the entire database. The SGP data are most concentrated in the Palaeozoic Era, in which they make up 27% of the total database, with the new compiled data contributing only 18%. In other eras, the new compiled data outnumber the SGP data: 4% versus 15% in the Cenozoic, 7% versus 14% in the Mesozoic, and 7%

268 versus 8% in the Precambrian. This is mainly the case because the SGP data in the first phase were

269 primarily from the Neoproterozoic and Palaeozoic eras (Farrell et al., 2021).

270

271 **Figure 4. The age distribution of samples in the database. (a) Age distribution of samples (excluding a small**

272 **number of samples with ages >2500 Ma from the figure, a total of 1258 samples). (b) Age distribution of**

273 **Phanerozoic samples at the stage level. Sd, Siderian; R, Rhyacian; Ors, Orosirian; Stt, Statherian; Cl,**

274 **Calymmian; Ec, Ectasian; Stn, Stenian; Tn, Tonian; Cry, Cryogenian; Ed, Ediacaran; Cm, Cambrian; O,**

275 **Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; T, Triassic; J, Jurassic; K, Cretaceous;**

276 **Pg, Paleogene; N, Neogene; Q, Quaternary.**

277   For the distribution of sample ages within the Phanerozoic, we divided the samples by stage, as

278   shown in Fig. 4b. For the Quaternary Period, due to its short duration, the data were not subdivided by

279   Stage but were instead divided into the Holocene and Pleistocene Series. Data distribution is not

280   uniform, with the highest concentration in the Quaternary Period. These data mainly come from DSDP

281   and ODP, which are characterised by a high number of core samples and high resolution. There are

282   fewer data for the Upper Permian, Lower Triassic, and Lower to Middle Jurassic, possibly because of

283   the existence of Pangaea at that time, which reduced the area of continental margins and inhibited

284   marine transgressions, resulting in fewer preserved marine environments in comparison to those of

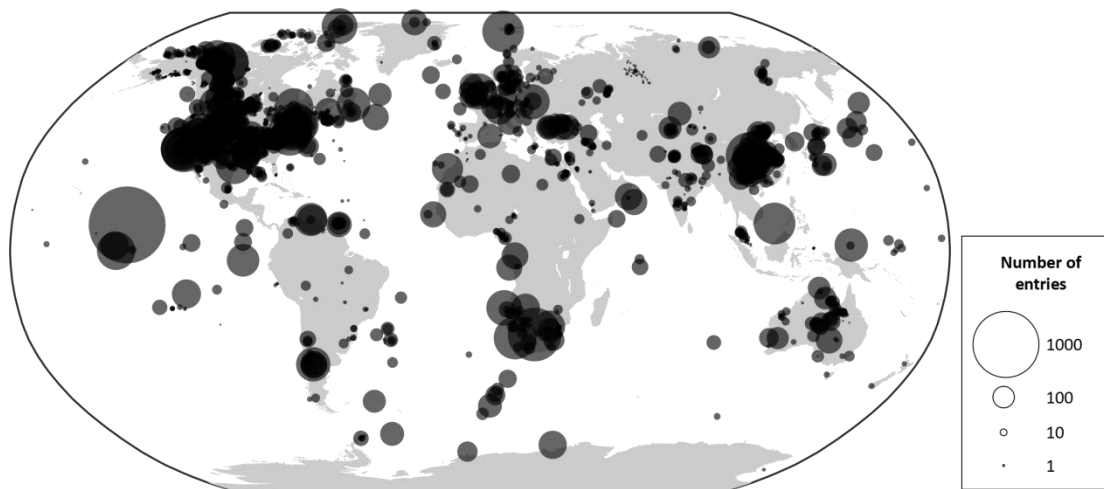
285   other geological periods (Mackenzie and Pigott, 1981; Walker et al., 2002). The distribution of sample

286   quantities in other periods fluctuates, often corresponding to periods of significant research interest,

287   such as the end-Ordovician, end-Devonian, end-Permian, Early Jurassic Toarcian and Early Cretaceous

288   Albian, which had peaks in sample numbers due to their association with major mass extinction events

289   and oceanic anoxic events (Fan et al., 2020).

290

291   **Figure 5. Bubble chart of modern geographical distribution and sample quantities in the database.**

292   In terms of spatial trends, the spatial distribution of sampling points in the DM-SED database is

293   inherently uneven, both in modern and palaeogeographic locations. Modern locations are primarily

294   concentrated in North America, Europe, South Africa, and China (Fig. 5). When modern coordinates

295   are converted to palaeogeographic coordinates and projected onto palaeogeographic maps, Cambrian to

296   Jurassic data come predominantly from continental margin environments, as subduction of oceanic

297   crust before the Cretaceous resulted in preservation of very few deep-sea environments (Fig. 6).

298   Cambrian and Ordovician data are distributed mainly on the Laurentia, Baltica, and South China plates,

299 with a few along the Gondwana margin. Silurian data occur mainly on Laurentia, South China, and

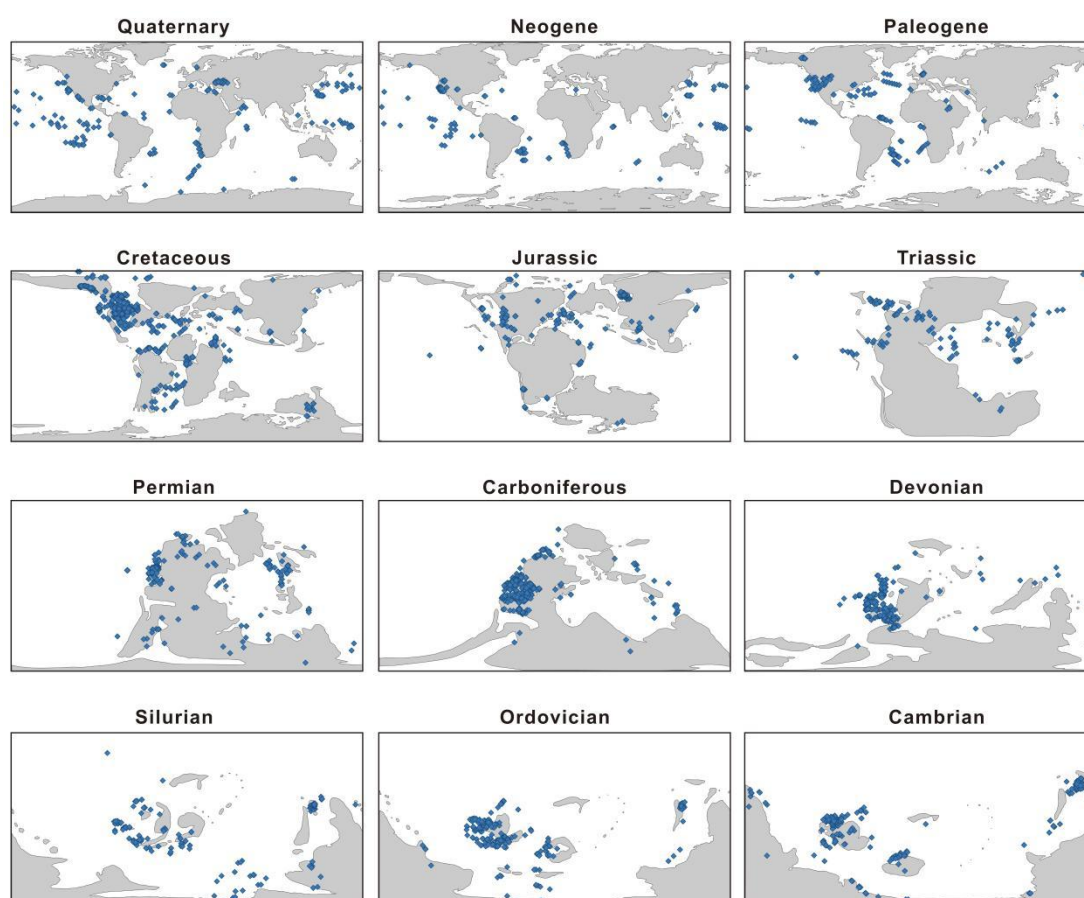
300 eastern Gondwana. Devonian and Carboniferous data are primarily on the Laurussia plate, with sparse

301 distribution in South China and Gondwana. Permian and Triassic data are mainly on the Laurussia and

302 South China plates, with sparse distribution in Gondwana. Jurassic data are primarily on the North

303 American, European shelf, with sparse distribution on other plates. From the Cretaceous to the

304 Quaternary, sample locations, dominated by data from the DSDP, ODP, and USGS NGDB projects, are

305 mainly located in the deep oceans and North America.

306
307 **Figure 6. The palaeogeographic distribution of sample sites in the DM-SED.**

308   When averaging all Phanerozoic data by stage and spatially averaging them into 15°

309 palaeolatitude bins (Fig. 7), Palaeozoic data records are mainly biased toward tropical regions.

310 Cambrian data are concentrated between 15° S and 30° N, Ordovician to Carboniferous data are

311 concentrated between 45° S and 15° N, and Permian data are concentrated between 0° N and 30° N,

312 with data mainly fluctuating around the equator. As continents migrated northward through the

313 Mesozoic and into the Cenozoic, records began to show bias toward mid-latitudes in the Northern

314 Hemisphere. From the Triassic to the Cretaceous, data are mainly concentrated between 0° N and 60°

N. Paleogene to Quaternary data are concentrated between 45° S and 45° N.

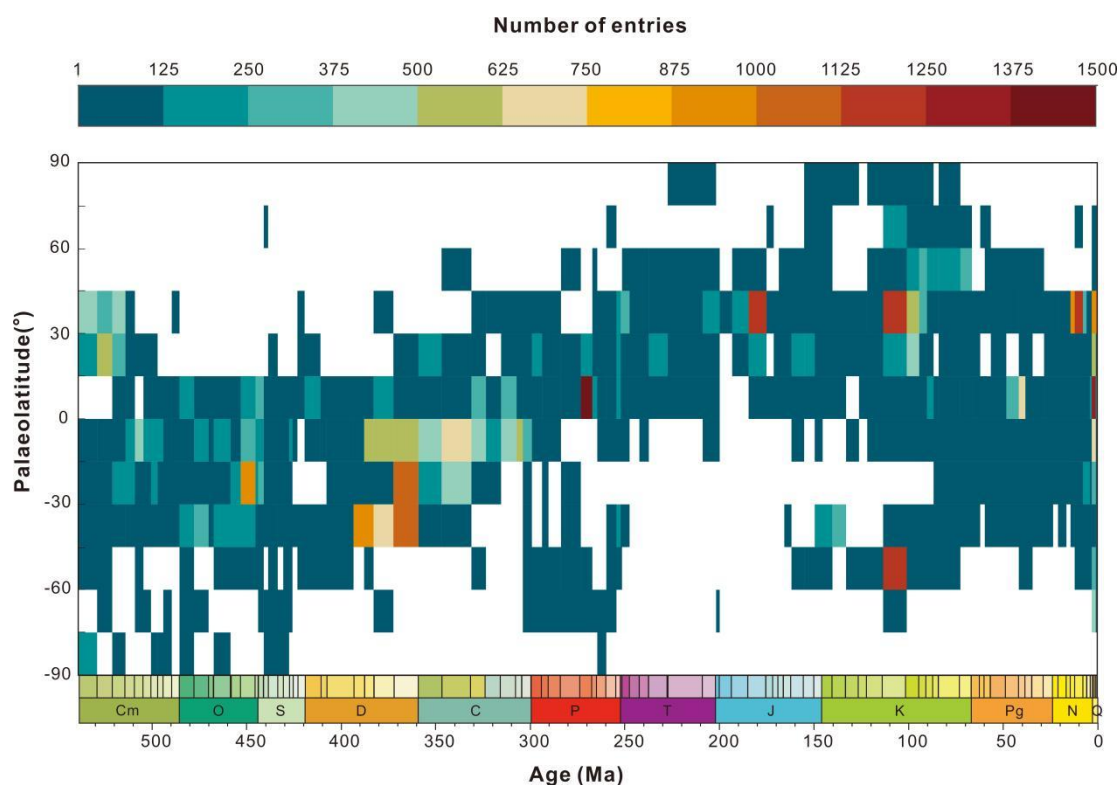



Figure 7. The spatiotemporal distributions of sample quantities (categorized temporally by stage and spatially by palaeolatitude intervals of 15°).

5 Usage instructions

The ultimate goal of the DM-SED database is to provide the geoscience community with a valuable resource of knowledge and geographic information. By deriving meaningful conclusions from a large marine sediment geochemistry database, we aim to enhance our understanding of Earth's environmental changes over time and space. All entries in DM-SED contain the source of original proxy values, ensuring traceability between DM-SED and the original datasets from which the data were extracted.

However, our database has some limitations. The criteria for age determination, relying variously on fossil zones and lithostratigraphic unit information, are not entirely uniform. Some age determinations are still coarse, with samples from a single section all assigned the same age. Additionally, the data quantity for some elements is still low. There may be significant differences in methodological precision between older and newer literature. Currently, these issues remain largely unresolved. Despite our best efforts to identify data from the literature and process quality control for

each entry, the sheer volume of data in DM-SED means that some errors or omissions are inevitable. Prompt corrections and continuous updates are expected to ensure the credibility of this database.

Finally, it is important to recognize that DM-SED merely compiles these various datasets and cannot impose any requirements on their generation. When using the data (and where practicable), we recommend citing both DM-SED and the original data sources to ensure proper attribution.

6 Data availability

Version controlled releases of the DM-SED can be found on Zenodo (https://doi.org/10.5281/zenodo.14771859, last accessed: 30 January 2025) (Lai et al., 2025). A static copy of DM-SED version 0.0.1 is archived in the Geobiology database (http://geobiologydata.cug.edu.cn/, last accessed: 25 September 2024). We plan to supplement and improve the database continuously and hope to collaborate with existing compilation authors to assist in adding new content.

7 Code availability

The software tools used in this study are available at the following links: WebPlotDigitizer can be downloaded from https://github.com/automeris-io/WebPlotDigitizer/releases (last accessed: 20 July 2024); the PointTracker v7 tool can be found at http://www.paleogis.com (last accessed: 20 July 2024); QGIS 3.16 can be downloaded from the https://qgis.org/project/overview/ (last accessed: 20 July 2024).

Author contributions. JKL: Writing – original draft, Visualization, Data collection, Investigation. HJS: Writing – review & editing, Supervision, Investigation, Funding acquisition. DLC: Writing – review & editing, Investigation. JDC: Writing – review & editing, Investigation. EAS: Writing – review & editing, Investigation. YYW: Writing – review & editing, Supervision, Investigation, Data collection. XKL: Writing – review & editing, Investigation. LW: Writing – review & editing, Data collection, Investigation. MTL: Writing – review & editing, Investigation. HCS: Writing – review &

359 editing, Investigation. YD: Writing– review & editing, Investigation. EHJ: Writing– review & editing,

360 Investigation. YF: Writing – review & editing, Investigation. HYS: Writing – review & editing,

361 Investigation. WCY: Writing – review & editing, Investigation. QZL: Writing – review & editing,

362 Investigation. XCL: Writing – review & editing, Investigation. HY: Writing – review & editing,

363 Investigation.

364

365 **Competing interests.** The authors declare that they have no conflicts of interest.

366

367 **Acknowledgements.**

368 This paper benefited greatly from comments from Thierry Adatte and an anonymous reviewer. We also

369 thank Jan Peter (Research Scientist, Geological Survey of Canada) for providing his literature data.

370

371 **Funding:**

372 This study was supported by the National Natural Science Foundation of China grant 42325202, the

373 State Key R&D Project of China (2023YFF0804000), 111 Project grant B08030, Natural Science

374 Foundation of Hubei (2023AFA006), and Graduate Student Project from the Hubei Research Center

375 for Basic Disciplines of Earth Sciences (NO. HRCES-202413). E.A.S. is supported by United States

376 National Science Foundation grant EAR-2143164.

377 **REFERENCES**

378 Algeo, T. J. and Liu, J.: A re-assessment of elemental proxies for paleoredox analysis, Chem. Geol.,

379   540, 119549, https://doi.org/10.1016/j.chemgeo.2020.119549, 2020.

380 Drevon, D., Fursa, S. R., and Malcolm, A. L.: Intercoder Reliability and Validity of WebPlotDigitizer

381   in Extracting Graphed Data, Behav. Modif., 41, 323-339,

382   https://doi.org/10.1177/0145445516673998, 2017.

383 FAIR: FAIR Play in geoscience data, Nat. Geosci., 12, 961, https://doi.org/10.1038/s41561-019-0506-4,

384   2019.

385 Fan, J., Shen, S., Erwin, D. H., Sadler, P. M., MacLeod, N., Cheng, Q., Hou, X., Yang, J., Wang, X.,

386   Wang, Y., Zhang, H., Chen, X., Li, G., Zhang, Y., Shi, Y., Yuan, D., Chen, Q., Zhang, L., Li, C.,

387   and Zhao, Y.: A high-resolution summary of Cambrian to Early Triassic marine invertebrate

388   biodiversity, Science, 367, 272-277, https://doi.org/10.1126/science.aax4953, 2020.

389 Farrell, U. C., Samawi, R., Anjanappa, S., Klykov, R., Adeboye, O. O., Agic, H., Ahm, A. C., Boag, T.

390   H., Bowyer, F., Brocks, J. J., Brunoir, T. N., Canfield, D. E., Chen, X., Cheng, M., Clarkson, M.

391   O., Cole, D. B., Cordie, D. R., Crockford, P. W., Cui, H., Dahl, T. W., Mouro, L. D., Dewing, K.,

392   Dornbos, S. Q., Drabon, N., Dumoulin, J. A., Emmings, J. F., Endriga, C. R., Fraser, T. A.,

393   Gaines, R. R., Gaschnig, R. M., Gibson, T. M., Gilleaudeau, G. J., Gill, B. C., Goldberg, K.,

394   Guilbaud, R., Halverson, G. P., Hammarlund, E. U., Hantsoo, K. G., Henderson, M. A.,

395   Hodgskiss, M. S. W., Horner, T. J., Husson, J. M., Johnson, B., Kabanov, P., Brenhin Keller, C.,

396   Kimmig, J., Kipp, M. A., Knoll, A. H., Kreitsmann, T., Kunzmann, M., Kurzweil, F., LeRoy, M.

397   A., Li, C., Lipp, A. G., Loydell, D. K., Lu, X., Macdonald, F. A., Magnall, J. M., Mand, K.,

398   Mehra, A., Melchin, M. J., Miller, A. J., Mills, N. T., Mwinde, C. N., O'Connell, B., Och, L. M.,

399   Ossa Ossa, F., Pages, A., Paiste, K., Partin, C. A., Peters, S. E., Petrov, P., Playter, T. L.,

400   Plaza-Torres, S., Porter, S. M., Poulton, S. W., Pruss, S. B., Richoz, S., Ritzer, S. R., Rooney, A.

401   D., Sahoo, S. K., Schoepfer, S. D., Sclafani, J. A., Shen, Y., Shorttle, O., Slotznick, S. P., Smith,

402   E. F., Spinks, S., Stockey, R. G., Strauss, J. V., Stueken, E. E., Tecklenburg, S., Thomson, D.,

403   Tosca, N. J., Uhlein, G. J., Vizcaino, M. N., Wang, H., White, T., Wilby, P. R., Woltz, C. R.,

404   Wood, R. A., Xiang, L., Yurchenko, I. A., Zhang, T., Planavsky, N. J., Lau, K. V., Johnston, D. T.,

405   and Sperling, E. A.: The Sedimentary Geochemistry and Paleoenvironments Project, Geobiology,

406   19, 545-556, https://doi.org/10.1111/gbi.12462, 2021.

407 Gradstein, F. M., Ogg, J. G., Schmitz, M. D., and Ogg, G. M.: Geologic time scale 2020, Elsevier,

408   2020.

409 Granitto, M., Giles, S. A., and Kelley, K. D.: Global Geochemical Database for Critical Metals in

410   Black Shales, U.S. Geological Survey data release [data set], https://doi.org/10.5066/F71G0K7X,

411   2017.

412 Grossman, E. L. and Joachimski, M. M.: Oxygen Isotope Stratigraphy, in: Geologic Time Scale 2020,

413   279-307, https://doi.org/10.1016/b978-0-12-824360-2.00010-3, 2020.

414 Jin, C., Li, C., Algeo, T. J., Wu, S., Cheng, M., Zhang, Z., and Shi, W.: Controls on organic matter

415   accumulation on the early-Cambrian western Yangtze Platform, South China, Mar. Pet. Geol.,

416   111, 75-87, https://doi.org/10.1016/j.marpetgeo.2019.08.005, 2020.

417   Judd, E. J., Tierney, J. E., Huber, B. T., Wing, S. L., Lunt, D. J., Ford, H. L., Inglis, G. N., McClymont,

418   E. L., O'Brien, C. L., Rattanasriampaipong, R., Si, W., Staitis, M. L., Thirumalai, K., Anagnostou,

419   E., Cramwinckel, M. J., Dawson, R. R., Evans, D., Gray, W. R., Grossman, E. L., Henehan, M. J.,

420   Hupp, B. N., MacLeod, K. G., O'Connor, L. K., Sanchez Montes, M. L., Song, H., and Zhang, Y.

421   G.: The PhanSST global database of Phanerozoic sea surface temperature proxy data, Sci. Data, 9,

422   753, https://doi.org/10.1038/s41597-022-01826-0, 2022.

423   Lai, J., Song, H., Chu, D., Dal Corso, J., Sperling, E. A., Wu, Y., Liu, X., Wei, L., Li, M., Song, H., Du,

424   Y., Jia, E., Feng, Y., Song, H., Yu, W., Liang, Q., Li, X., and Yao, H.: Deep-Time Marine

425   Sedimentary Element Database [data set], Zenodo. https://doi.org/10.5281/zenodo.14771859,

426   2025.

427   Large, R. R., Halpin, J. A., Lounejeva, E., Danyushevsky, L. V., Maslennikov, V. V., Gregory, D.,

428   Sack, P. J., Haines, P. W., Long, J. A., Makoundi, C., and Stepanov, A. S.: Cycles of nutrient

429   trace   elements   in   the   Phanerozoic   ocean,   Gondwana   Res.,   28,   1282-1293,

430   https://doi.org/10.1016/j.gr.2015.06.004, 2015.

431   Li, Z., Zhang, Y. G., Torres, M., and Mills, B. J. W.: Neogene burial of organic carbon in the global

432   ocean, Nature, 613, 90-95, https://doi.org/10.1038/s41586-022-05413-6, 2023.

433   Mackenzie, F. T. and Pigott, J. D.: Tectonic controls of Phanerozoic sedimentary rock cycling, J. Geol.

434   Soc., 138, 183-196, https://doi.org/10.1144/gsjgs.138.2.0183, 1981.

435   Planavsky, N. J., Asael, D., Rooney, A. D., Robbins, L. J., Gill, B. C., Dehler, C. M., Cole, D. B.,

436   Porter, S. M., Love, G. D., Konhauser, K. O., and Reinhard, C. T.: A sedimentary record of the

437   evolution   of   the   global   marine   phosphorus   cycle,   Geobiology,   21,   168-174,

438   https://doi.org/10.1111/gbi.12536, 2023.

439   Reinhard, C. T., Planavsky, N. J., Gill, B. C., Ozaki, K., Robbins, L. J., Lyons, T. W., Fischer, W. W.,

440   Wang, C., Cole, D. B., and Konhauser, K. O.: Evolution of the global phosphorus cycle, Nature,

441   541, 386-389, https://doi.org/10.1038/nature20772, 2017.

442   Schobben, M., Foster, W. J., Sleveland, A. R. N., Zuchuat, V., Svensen, H. H., Planke, S., Bond, D. P.

443   G., Marcelis, F., Newton, R. J., Wignall, P. B., and Poulton, S. W.: A nutrient control on marine

444   anoxia   during   the   end-Permian   mass   extinction,   Nat.   Geosci.,   13,   640-646,

445   https://doi.org/10.1038/s41561-020-0622-1, 2020.

446   Schoepfer, S. D., Algeo, T. J., Ward, P. D., Williford, K. H., and Haggart, J. W.: Testing the limits in a

447   greenhouse ocean: Did low nitrogen availability limit marine productivity during the end-Triassic

448   mass extinction?, Earth Planet. Sci. Lett., 451, 138-148, https://doi.org/10.1016/j.epsl.2016.06.050,

449   2016.

450   Schoepfer, S. D., Shen, J., Wei, H., Tyson, R. V., Ingall, E., and Algeo, T. J.: Total organic carbon,

451   organic phosphorus, and biogenic barium fluxes as proxies for paleomarine productivity, Earth

452   Sci. Rev., 149, 23-52, https://doi.org/10.1016/j.earscirev.2014.08.017, 2015.

453   Scotese, C.: The PALEOMAP Project PaleoAtlas for ArcGIS, version 1, 2, 16-31, 2008.

454   Scotese, C. R., Song, H., Mills, B. J. W., and van der Meer, D. G.: Phanerozoic paleotemperatures: The

455   earth's changing climate during the last 540 million years, Earth Sci. Rev., 215, 103503,

456   https://doi.org/10.1016/j.earscirev.2021.103503, 2021.

457   Scotese, C. R. and Wright, N.: PALEOMAP Paleodigital Elevation Models (PaleoDEMS) for the

458   Phanerozoic   PALEOMAP   Project,   Paleomap   Proj,

459   https://www.earthbyte.org/paleodem-resourcescotese-and-wright-2018/, 2018.

460   Scott, C., Planavsky, N. J., Dupont, C. L., Kendall, B., Gill, B. C., Robbins, L. J., Husband, K. F.,

461   Arnold, G. L., Wing, B. A., Poulton, S. W., Bekker, A., Anbar, A. D., Konhauser, K. O., and

462   Lyons, T. W.: Bioavailability of zinc in marine systems through time, Nat. Geosci., 6, 125-128,

463   https://doi.org/10.1038/ngeo1679, 2012.

464   Shen, J., Schoepfer, S. D., Feng, Q., Zhou, L., Yu, J., Song, H., Wei, H., and Algeo, T. J.: Marine

465   productivity changes during the end-Permian crisis and Early Triassic recovery, Earth Sci. Rev.,

466   149, 136-162, https://doi.org/10.1016/j.earscirev.2014.11.002, 2015.

467   Song, H., Wignall, P. B., Song, H., Dai, X., and Chu, D.: Seawater Temperature and Dissolved Oxygen

468   over   the   Past   500   Million   Years,   J.   Earth   Sci.,   30,   236-243,

469   https://doi.org/10.1007/s12583-018-1002-2, 2019.

470   Stockey, R. G., Cole, D. B., Farrell, U. C., Agić, H., Boag, T. H., Brocks, J. J., Canfield, D. E., Cheng,

471   M., Crockford, P. W., Cui, H., Dahl, T. W., Del Mouro, L., Dewing, K., Dornbos, S. Q., Emmings,

472   J. F., Gaines, R. R., Gibson, T. M., Gill, B. C., Gilleaudeau, G. J., Goldberg, K., Guilbaud, R.,

473   Halverson, G., Hammarlund, E. U., Hantsoo, K., Henderson, M. A., Henderson, C. M., Hodgskiss,

474   M. S. W., Jarrett, A. J. M., Johnston, D. T., Kabanov, P., Kimmig, J., Knoll, A. H., Kunzmann, M.,

475   LeRoy, M. A., Li, C., Loydell, D. K., Macdonald, F. A., Magnall, J. M., Mills, N. T., Och, L. M.,

476   O'Connell, B., Pagès, A., Peters, S. E., Porter, S. M., Poulton, S. W., Ritzer, S. R., Rooney, A. D.,

Schoepfer, S., Smith, E. F., Strauss, J. V., Uhlein, G. J., White, T., Wood, R. A., Woltz, C. R., Yurchenko, I., Planavsky, N. J., and Sperling, E. A.: Sustained increases in atmospheric oxygen and marine productivity in the Neoproterozoic and Palaeozoic eras, Nat. Geosci., 17, 667-674, https://doi.org/10.1038/s41561-024-01479-1, 2024.

Sweere, T. C., Dickson, A. J., and Vance, D.: Nickel and zinc micronutrient availability in Phanerozoic oceans, Geobiology, 21, 310-322, https://doi.org/10.1111/gbi.12541, 2023.

Tribovillard, N.: Re-assessing copper and nickel enrichments as paleo-productivity proxies, BSGF - Earth Sciences Bulletin, 192, 54, https://doi.org/10.1051/bsgf/2021047, 2021.

Veizer, J. and Prokoph, A.: Temperatures and oxygen isotopic composition of Phanerozoic oceans, Earth Sci. Rev., 146, 92-104, https://doi.org/10.1016/j.earscirev.2015.03.008, 2015.

Walker, L. J., Wilkinson, B. H., and Ivany, L. C.: Continental drift and Phanerozoic carbonate accumulation in shallow-shelf and deep-marine settings, J. Geol., 110, 75-87, https://doi.org/10.1086/324318, 2002.

Wang, D., Liu, Y., Zhang, J., Lang, Y., Li, Z., Tong, Z., Xu, L., Su, Z., and Niu, J.: Controls on marine primary productivity variation and organic matter accumulation during the Late Ordovician－Early Silurian transition, Mar. Pet. Geol., 142, 105742, https://doi.org/10.1016/j.marpetgeo.2022.105742, 2022.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., t Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.

Xiang, L., Schoepfer, S. D., Zhang, H., Cao, C., and Shen, S.: Evolution of primary producers and productivity across the Ediacaran-Cambrian transition, Precambrian Res., 313, 68-77, https://doi.org/10.1016/j.precamres.2018.05.023, 2018.

507 Zhang, Q., Bendif, E. M., Zhou, Y., Nevado, B., Shafiee, R., and Rickaby, R. E. M.: Declining metal

508   availability in the Mesozoic seawater reflected in phytoplankton succession, Nat. Geosci., 15,

509   932-941, https://doi.org/10.1038/s41561-022-01053-7, 2022.

510 Zhao, K., Zhu, G., Li, T., Chen, Z., and Li, S.: Fluctuations of continental chemical weathering control

511   primary productivity and redox conditions during the Earliest Cambrian, Geol. J., 58, 3659-3672,

512   https://doi.org/3659-3672, 10.1002/gj.4778, 2023.