A geospatial inventory dataset of study sites in a Korean Quaternary paleoecology database

Soo Hyun Kim[1] and Eunji Byun[2]

[1]*Institute of Sustainable Earth and Environmental Dynamics, Pukyong National University, Pusan, 48513, South Korea*

[2]*Department of Earth System Sciences, Yonsei University, Seoul, 03722, South Korea*

Correspondence to: Eunji Byun (eb@yonsei.ac.kr)

Abstract. Ecological insights beyond human-observable time scales are derived from records preserved in geological sediments around the world. Nonetheless, significant regional data gaps persist in global syntheses of these records as open data practices are still emerging around the world. Korean Quaternary paleoecological data remain underrepresented in global research efforts, despite a growing body of site-level research. Here, we organize an inventory of 328 paleoecological study sites (72 paleo-sites with paleoproxy-based records from sediments and 256 surface sites with modern pollen samples) in South Korea, compiled from 66 research articles published from 2003 to 2023. We have built three datasets related to this inventory: (1) Publication Metadata, which provides citation details of the 66 articles; (2) Site Inventory, which contains metadata about geospatial coordinates, depositional environments, chronological ranges, age coverage, and proxies; and (3) Geochronology Data, which includes chronological details (dating methods, ages, and depth points) for each site. The sites span from 33.2508° to 33.4808°N and from 126.1486° to 129.2132°E, with elevations from −156 to 1867.5 m. Sediment samples were collected by coring or trenching from six depositional environments: Open Coastal Zone, Estuary, Lagoon, River, Volcanic Cone, and Other. A total of 812 chronological controls were analyzed, and most are based on radiocarbon dating. Pollen, diatoms, and grain size have been widely applied, with multi-proxy approaches increasingly common in recent studies. To enhance accessibility, we have developed GeoEcoKorea, an open-access platform featuring interactive maps where each site marker displays site-level metadata and links to bibliographic information and uncalibrated geochronological datasets, when available. In addition to promoting FAIR (Findable, Accessible, Interoperable, and Reusable) data practices, our plantform aims to foster more collaborative and inclusive data-sharing cultures, enable regional syntheses of long-term ecosystem dynamics, and contribute Korean paleoecological data to global-scale reconstruction of past environmental changes.

1. Introduction

Quaternary paleoecologists and paleoclimatologists investigate environmental changes over the last ~2.5 million years by reconstructing long-term variability in ecosystems, climate, and other components of the Earth system. These reconstructions are based on proxy records from geological archives, particularly sediments in terrestrial depositional settings such as lakes, mires, and coastal marine systems. Proxies serve as indirect indicators of past environmental conditions—for example, fossil pollen for vegetation history and charcoal for fire history (Prentice, 1988; Whitlock and Larsen, 2002). A wide variety of proxies, including biological (e.g., fossil pollen and diatom), chemical (stable isotopes and elemental concentrations), and physical indicators (grain size and magnetic susceptibility) have been used to infer different aspects of past environments

34   (Bradley, 2015).

35   Yet no single proxy alone can resolve the complexity of environmental conditions, particularly when shaped by networks of
36   interacting processes within the Earth system (Birks and Birks, 2006; Mann, 2002). To improve interpretability, Quaternary
37   studies use multi-proxy strategies that combine proxies reflecting different environmental drivers acting at varying
38   magnitudes. Some proxies respond to a narrow range of conditions, while others are sensitive to multiple extrinsic (e.g.,
39   climate variability and ecological disturbances) and intrinsic factors (proxy-specific sensitivity) and may respond to either
40   their combined effects or a single dominant driver, depending on which prevailed at a given time (Birks and Birks, 2006;
41   Mann et al., 2002). For example, during the Holocene (the recent ~11,700 years), fossil pollen records used to reconstruct
42   past vegetation changes require different interpretive approaches depending on whether the study area is known to have had
43   early, intensive human settlement activity or whether natural vegetation communities can be assumed to have responded
44   primarily to climatic changes. In such a case, pairing pollen-based reconstructions with charcoal or archaeological records
45   has often allowed for disentangling anthropogenic land-use impacts from climatic signals at the study site (e.g., Abraham et
46   al., 2023; Kozáková et al., 2015; Yang et al., 2020).

47   In tandem with multi-proxy approaches, multi-site syntheses have led to the large-scale compilation of proxy-based datasets
48   over recent decades. These efforts have underpinned global-scale synthesis projects—for example, reconstructions of
49   Holocene mean surface temperature and land-cover and land-use changes (Gaillard et al., 2010; Kaufman et al., 2020).
50   Broad-scale paleodata syntheses have also enabled the quantification of ecological parameters that are unmeasurable through
51   direct observation—such as biome-specific residence and recovery times and pollen taxon-specific climate fidelity in North
52   America and subcontinental-scale rates of vegetation changes during the past ~20,000 years (Mottl et al., 2021; Wang et al.,
53   2020; Wang et al., 2023). In addition, growing efforts to integrate high-volume paleodata have driven the development of
54   open-source computational tools, such as neotoma2 for acquiring data from open-access repositories, R-Fossilpol for
55   processing pollen datasets into standardized and reproducible formats, and R-Ratepol for estimating the rate of compositional
56   change within palynological assemblages (Flantua et al., 2023; Mottl et al., 2021;Vidaña and Goring, 2023). These advances
57   have been made possible by a rich set of high-resolution proxy records that enable effective analyses of environmental
58   variability in the recent geological past (Farley et al., 2018; Williams et al., 2018).

59   Such multi-proxy and multi-site approaches underscore the importance of systematically preserving and archiving well-
60   recovered proxy data to enable immediate and broad reuse. For this purpose, the FAIR (findable, accessible, interoperable,
61   and reusable) guiding principles have been proposed as a model for open data sharing across the geosciences (Wilkinson et
62   al., 2016). It is encouraging to see the growing practice of open access data in many scientific communities, and the Neotoma
63   Paleoecology Database (https://www.neotomadb.org/) is one of the most remarkable contributions to the field (Williams et
64   al., 2018). Research community-based support is available throughout Neotoma's suite of services, from providing servers
65   and curating data uploads to providing a user-friendly interface for global data search and downloads. The Neotoma
66   community also advises regional communities building and governing their own data while leveraging and affiliating with
67   Neotoma for global networking. With this possibility in mind, this paper is the first step towards developing a Korea-based
68   research community database for regional Quaternary paleoenvironmental data. The goal is to build a new data system that

69 supports detailed hypothesis testing at regional to global scales by incorporating geological and ecological characteristics of
70 previously investigated sites while also accounting for societal and cultural factors in conducting research.

71 Late Quaternary studies in Korea have been multidisciplinary from an early stage, led by pioneering researchers across fields
72 such as archaeology, biology, earth-science education, geochemistry, paleoclimatology, palynology, physical geography, and
73 sedimentology. Since the 2000s, a few review articles have been published about progress in related topics within Korea
74 (Park, 2008; Nahm, 2018; Kim et al., 2024). Nonetheless, there has been little interest or practical support for publishing data
75 associated with the original publications in open and FAIR data resources, which is consistent with the low representation of
76 data from Korea in Neotoma (Fig. 1) and in a recent global synthesis (Herzschuh et al., 2023).

77 In this paper, we present the first public inventory of site-level metadata from paleoecological and paleoenvironmental
78 research in South Korea (33˚to 38˚ N, 124˚ to 132˚ E), comprising a total of 328 study sites compiled from 66 peer-reviewed
79 publications over the past 20 years (2003–2023). Our dataset provides metadata on study sites and sediment and modern
80 surface samples, including latitude, longitude, elevation, site description, proxy type, and chronological control. We compare
81 the spatial coverage of the current dataset with that of the global-scale database and explore the significance of compiling
82 data into a regional database (Fig. 1), highlighting the importance of understanding regional contexts through collaboration
83 with domestic researchers who have knowledge of the local data and study sites (following the approach of Flantua et al.,
84 2015). Using this inventory dataset as a reference, we continue our efforts to collaborate with the authors of the referenced
85 publications and deposit the raw data into a relational database, which will be made accessible through GeoEcoKorea (GEK,
86 https://geoecokorea.org), an online platform offering open access to the database in both English and Korean.

87 **2. Data Collection**

88 We compiled 66 research papers published between 2003 and 2023 that presented proxy data from sediments and surface
89 pollen samples in South Korean territory. The literature search followed four steps:

90   **Step 1. Initial journal search**: We searched three peer-reviewed Quaternary research journals (The Holocene;
91   Palaeogeography, Palaeoclimatology, Palaeoecology; and Quaternary Science Reviews) using four keywords,
92   "Korea," "sediments," "proxy," and "pollen."

93   **Step 2. Author search**: We compiled the names of all coauthors from the articles identified in the first step.

94   **Step 3. Expanded journal search**: Using Google Scholar and DBpia (a Korean academic journal database,
95   https://www.dbpia.co.kr/), we examined additional journal articles by the authors identified in the second step.

96   **Step 4. Compilation of research articles**: We gathered articles related to Korean Quaternary proxy records
97   published up to 2023 in the journals discovered in the first and third steps.

98 To evaluate the relevance of the compiled articles, we reviewed their abstracts and result sections. For proxy-based records,
99 we selected articles that provided reliable age-dating information, such as the names of Accelerator Mass Spectrometry

100 (AMS) laboratories where dating was conducted. This emphasis on accurate and precise geochronological reporting led to the
101 exclusion of articles published before 2003, resulting in 66 publications.

102 From the selected publications, we developed three datasets: **I. Publication Metadata**, **II. Site Inventory, and III.**
103 **Geochronology Data**. The first two datasets are linked through Publication ID, and the last two datasets through Site ID
104 (Fig. 2). The Site Inventory dataset serves as the core of the other two datasets, including key information on study site and
105 sample metadata (Table 1). The Publication Metadata contains citation details of the 66 articles covering study sites in the
106 inventory dataset. The Geochronology Data comprises chronological information used for sediment records of the sites. Each
107 dataset is accompanied by a supplementary worksheet titled *Column description*, which defines the structure and format of
108 the data fields (see Section 4. Data Availability).

109 **3. Data Processing and Description**

110 **3.1. Dataset I: Publication Metadata**

111 The Publication Metadata dataset contains information about authors, publication year, title, journal, numbers of volume,
112 issue, and pages, DOI, and language (either English or Korean). The annual number of papers has increased rapidly since
113 2010, driven by the growing number of English-written papers published in international journals (Fig. 3).

114 **3.2. Dataset II: Site Inventory**

115 This section outlines the data processing procedures for the Site Inventory dataset, with relevant details also summarized in
116 the Supplementary Document.

117 **3.2.1. Type of study site**

118 The Site Inventory dataset comprises 328 sites that consist of 72 paleo-sites with sedimentary records and 256 surface sites
119 with surface pollen samples (Fig. 4). The paleo-sites were derived from 64 publications that presented paleorecords
120 reconstructed from sediment sequences to investigate the temporal sequence of the targeted proxy and past ecosystem
121 elements (e.g., fossil pollen and past vegetation-climate interactions). The surface sites were presented in two articles that
122 investigated modern surface samples that provide spatial information on the modern ecosystems, serving as the basis for the
123 space-for-time substitution method (e.g., surface pollen and modern vegetation-climate associations) (Chevalier et al., 2020).

124 **3.2.2. Geospatial coverage**

125 Each site is georeferenced using geographic coordinates and elevation. We used coordinates and elevations reported in the
126 papers when available. If not, we estimated these data using site maps from the articles, cross-referencing with Google Earth.
127 Site locations were recorded in decimal degrees (to four-digit precision) with elevations in meters above mean sea level (unit:
128 m).

129 The distribution of paleo-sites and surface sites shows no significant variation in longitude but distinct patterns in elevation
130 (Fig. 5) and latitude (Fig. 6). In the geographic extent (33.2508° to 38.4808° N, 126.1486° to 128.9719° E) of paleo-sites,

131   their elevations (range: −156 to 1,305 m; median: 5 m) are consistent with the tendency of depositional environments for

132   stratigraphic sediments typically found at low elevations (Fig. 5a), although their latitudinal distribution exhibits a broad

133   coverage across South Korea except for the central eastern region (Fig. 6a). In contrast, the surface pollen sites (33.2916˚ to

134   35.6913˚ N, 126.3090˚ to 129.2132˚ E) span a wider range of elevation (range: 95 to 1,867 m; median: 605 m) compared to

135   the paleo-sites (Fig. 5b) but are concentrated in the southern regions (Fig. 6b). These surface pollen sites were primarily

136   contributed by two pioneering studies: Park and Park (2015), who investigated 39 elevational transects along mountain slopes

137   in Jeju Island to reflect temperature gradients, and Lee et al. (2022), who studied 37 transects in the southern part of the

138   Korean Peninsula for modern pollen-climate references (Fig. 6b). It is interesting to note the high elevations of Jeju's

139   volcanic-cone wetland sites (Fig. 7), which locally shift the statistical distribution of the island paleo-sites (range: 51 to 1,305

140   m; median: 692 m)

141   **3.2.3. Sample type**

142   Three sample types (Surface Pollen, Core, and Trench) for the Site Inventory dataset represent the study approaches

143   commonly used in South Korean Quaternary paleo-reconstruction research (Kim et al., 2024). Surface pollen samples were

144   systematically targeted and obtained from the soil surface on mountain slopes in Jeju Island and the southern region of the

145   mainland (Lee et al., 2022; Park and Park, 2015). For sediment samples, either coring or trenching was employed to obtain

146   the sediments for various proxy analyses, preserving the original stratigraphy of the layered, aged sediments. Coring

147   generally involves inserting empty tubes into sediments to extract vertically long and continuous sediments embedded in the

148   tubes. Trenching involves collecting subsamples directly from stratigraphic sequences of exposed outcrops. Since trenching

149   was unsuitable for submerged sites, coring was used more frequently at paleo-sites than trenching (Fig. 4).

150   **3.2.4. Depositional setting**

151   The depositional environments of paleo-sites are categorized into six groups: Open Coastal Zone, Estuary, Lagoon, River,

152   Volcanic Cone, and Other (Table 2). The Open Coastal Zone encompasses a set of environments dominantly influenced by

153   oceanic depositional processes, such as beaches, nearshore, and continental shelves. Estuary and Lagoon are classified

154   separately based on energy conditions within the transitional zone between terrestrial and oceanic sedimentation, with

155   Estuary characterized by high-energy conditions and Lagoon by low-energy conditions. The River category includes

156   depositional environments such as floodplains, backswamps, and other fluvial settings associated with rivers and streams.

157   The Volcanic Cone refers to wetland systems such as lakes, bogs, or swamps in the craters of cinder cones scattered across

158   Jeju Island, the largest island of South Korea, located to the south of the mainland. The Other comprises sites with

159   depositional settings that do not fit within the other five groups, and because surface pollen samples were all from the

160   montane soil surface, they were also classified under the Other category. The paleo-sites can be broadly categorized into two

161   groups, coastal and terrestrial sites. The coastal sites ($n = 41$), including Open Coastal Zone, Estuary, and Lagoon, outnumber

162   non-coastal sites ($n = 31$). River is a more commonly sampled site type than the other non-coastal categories—Volcanic Cone

163   and Other (Fig. 7b).

164   **3.2.5. Geochronology**

165   The chronological data are organized into eight data fields: Number of all age controls, Number of non-absolute age controls,
166   Absolute dating methods, Number of C-14 dates, Number of other absolute dates, Oldest age, Youngest age, Age scale, and
167   Total depth range (Table 1).

168   A total of 812 age control points compiled from 72 paleo-sites are classified into non-absolute and absolute types (Fig. 8a).
169   The non-absolute age controls ($n = 28$) consist of core top ($n = 22$) and estimated ages ($n = 6$). Core top ages are assigned to
170   the sediment surface as reference ages corresponding to the year of sampling. Estimated ages are extrapolated from age-depth
171   models. The absolute age controls ($n = 784$) are obtained by three dating methods, radiocarbon (^{14}C dates; $n = 623$), optically
172   stimulated luminescence (OSL dates, $n = 157$), and uranium-thorium (U-Th dates; $n = 4$) dating. Among these absolute age
173   controls, ^{14}C dating has been exclusively used to construct the geochronology of sediments at 55 out of 72 paleo-sites (Fig.
174   8b).

175   Oldest and Youngest ages are reported in calendar years before present (BP), using 1950 CE as the reference point for
176   "Present." For paleo-sites, these two endpoints correspond to ages of the youngest and oldest age control points used in age-
177   depth models in the original studies. This age selection is made not based on depth, as age reversals frequently occur within
178   sediment sequences, obscuring reliable geochronological boundaries of paleorecords. If either age is based on a radiocarbon
179   date, the lower (Oldest) or upper (Youngest) bound of the 95% calibrated range is documented, using the IntCal20 or
180   Marine20 calibration curves through the IntCal R package (Blaauw et al., 2022; Heaton et al., 2020; Reimer et al., 2020).
181   When the core top age reflects the youngest age, it is calculated by subtracting the year of coring or trenching from 1950 CE.
182   If the collection year is unspecified, a value of 0 is assigned for the youngest, following the same approach used for post-
183   bomb ^{14}C dates. For surface samples, the oldest and youngest ages are derived by subtracting the sampling year from 1950
184   CE. The publication year is used as the sampling year when it is unavailable in the original article. For OSL and U-Th dates,
185   the lower (Oldest) and upper (Youngest) bounds of the 1σ error ranges are also selected. Neither OSL nor U-Th ages were
186   calibrated. The U-Th ages were not normalized, as their reference year is defined as 1950 CE (Dutton et al., 2017). OSL age
187   limits, initially expressed as calendar years before the measurement year, are converted to calendar years before present by
188   assigning the measurement year as the sampling year, or the earliest publication year if unavailable. The calibration methods,
189   normalization approaches, and dating sources for Oldest and Youngest ages are provided in the Age scale field.

190   The difference between the oldest and youngest ages represents the temporal extent of each site. ~67% of paleo-sites ($n = 48$)
191   record only the Holocene, ~32% span from the Pleistocene to the Holocene ($n = 23$), and ~1% contain records exclusively
192   from the Pleistocene ($n = 1$) (Fig. 9a and 9b). The Holocene-only and the Pleistocene–Holocene sites are broadly distributed
193   across five types of depositional settings, although the Lagoon contains no Pleistocene record (Fig. 9c).

194   In addition, the chronological data are used to calculate two types of dating density: the number of age controls per time
195   interval (Flantua et al., 2015; Lacourse and Gajewski, 2020) and per depth interval. The temporal dating density is defined as
196   the total number of age controls divided by the time span between oldest and youngest ages, and the spatial dating density is
197   the total number divided by the total depth range. These metrics help assess the temporal and depth-spatial resolutions of age-

198 depth models (Lacourse and Gajewski, 2020). Both types of dating density are higher in Holocene-only sites compared to the

199 other two sites covering the Pleistocene (Fig. 9d and 9e).

3.2.6. Proxy

201 Information about proxies available at each site is organized in four data fields: Pollen, Diatom, Grain size, and Other

202 proxies. The first three fields indicate whether each site contains a proxy related to pollen and diatom assemblages and grain-

203 size distribution parameters (e.g., mean, median, or sorting and relative proportions of sand, silt, or clay), marking the

204 presence of each proxy as "Yes" and its absence as "NA." These three types of proxies are documented because they are the

205 most frequently investigated proxies, with some sites showing mono-proxy reconstructions that rely on only one of them

206 (Fig. 10b).

207 In the field of Other proxies, a total of 79 proxy types are listed, with abbreviations when necessary. The explanation of each

208 abbreviation is documented in a supplementary worksheet (see Section 4). In the list of the Other proxies, organic

209 geochemical indicators (e.g., C/N ratios and δ^{13}C) are more commonly available than inorganic proxies such as mineralogical

210 compositions, trace element content, and magnetic susceptibility.

211 Overall, for surface samples, our dataset includes only pollen, which is widely used to construct taxon-based transfer

212 functions for paleotemperatures estimates (Lee et al., 2022; Park and Park, 2015). While non-palynological organisms, such

213 as ostracods from brine and freshwater swamps and subterranean caves, have been studied in South Korea (McKenzie, 1972;

214 Smith et al., 2015) (Fig. 1), they were not included in our datasets because those studies focused on taxonomical descriptions

215 rather than quantitative reconstructions of past environments such as those produced by the modern analog techniques. For

216 paleo-sites, approximately 79% sites have multi-proxy records (Fig. 10a) and these sites are more widely distributed than

217 mono-proxy records. (Fig. 10c). Grain-size and pollen proxies are more frequently used than other proxies, with pollen being

218 the most utilized in mono-proxy sites (Fig. 10b). The extensive spatial coverage suggests that multi-proxy approaches have

219 become standard practice in the Quaternary paleoenvironmental research in South Korea (Fig. 10d).

3.3. Dataset III: Geochronology Data

221 The chronological dataset corresponding to 72 paleo-sites is documented separately. This dataset consists of nine categories:

222 SiteID, dating method (Non-absolute dating: Core top or estimated; Absolute dating: C-14 AMS, OSL, or U-Th dating),

223 sample ID, dating material, mean and 1σ-error of age (years before present), depth (cm), mean and 1σ-error of δ^{13}C (‰), and

224 Chronological age type. In these fields, most chronological data from the original publications are documented without

225 calibration, except in two standardized cases: core top ages are converted from AD to calendar years BP using 1950 CE as

226 the reference point, and post-bomb radiocarbon dates are expressed with "pMC." The field of Chronological age type

227 indicates timescale in which the date is reported, including uncalibrated ^{14}C years before present (1950 CE). Further details

228 on other data fields are provided in the Supplementary Document as well as the Column description worksheet within the

229 dataset file (see Section 4.).

230 Most absolute dates fall within the Holocene (Fig. 11). OSL dates often constrain the ages of sediment sequences older than

231   ~55,000 years BP, the upper limit of ^{14}C dating (Fig. 11b). Although U-Th dating has been applied to sediments exceeding

232   the maximum range of OSL ages, its usage for terrestrial sediments in South Korea remains limited (Fig. 11c).

4. Data Availability

234   Our three datasets are accessible on two platforms: Figshare and the GeoEcoKorea website. Figshare hosts tabular datasets in

235   Excel file (.xlsx) format at https://doi.org/10.6084/m9.figshare.28236596 (contact email: geoecokorea@gmail.com). The first

236   dataset file, **Publication Metadata.xlsx**, is composed of two worksheets: a) Publication dataset and b) Column description.

237   The second file, **Site Inventory.xlsx** contains three worksheets: a) Inventory dataset, b) Column description (identical to

238   Table 1), and c) OtherProxy_Abbreviation. The third file, **Geochronology Data.xlsx** has two worksheets, a)

239   Geochronological data and b) Column description

240   Concurrently, the GeoEcoKorea webpage (http://geoecokorea.org/) presents site locations on an interactive map at

241   https://geoecokorea.org/2025_Site_Inventory. The bibliographic and site inventory datasets are integrated, converted to

242   GeoJSON format, and visualized on OpenStreetMap-based interactive maps customized with the "Leaflet" JavaScript library.

243   Each site is represented by an interactive marker that opens a pop-up window displaying site-specific metadata, along with

244   direct links to the corresponding bibliographic information and a downloadable **Geochronology Data** (three worksheets:

245   Readme, Geochronological data, and Column description) in .xlsx format.

5. Summary and Dataset Reuse Potential

247   Our datasets provide a comprehensive overview of South Korean Quaternary paleoecological data published in peer-

248   reviewed journals from 2003 to 2023. The datasets highlight the high spatial density of paleorecords available from various

249   depositional environments, including wetlands, coastal deposits, lagoons, estuaries, rivers, and volcanic crater wetlands in

250   Jeju Island. While ~67% of records are covered only the Holocene, the remaining 33% extend into the Pleistocene. The

251   quality of these records has improved over the past decades, particularly with the increasing dominance of multiproxy

252   approaches and improved chronological constraints (Fig. 12), suggesting continued progress in the future.

253   Our site inventory offers practical reference for future site selection. Korea's compact geography and clearly differentiated

254   depositional settings (Fig. 7a) allow for temporal variability of sediment accumulation rates—defined as "deposition time" in

255   similar studies (e.g., Crann et al., 2015; Goring et al., 2012)—across the settings without the profound effects of large-scale

256   continental gradients. In addition, when the Korean paleorecords in the inventory are integrated, they can provide

257   foundational datasets for testing new hypotheses regarding the drivers of local ecosystem turnover in Northeast Asia. The

258   records span both human-cultivated lowlands and relatively undisturbed upland settings (Fig. 6a), enabling comparisons that

259   help separate climate- from human-induced impacts on past landscape changes during the Holocene. The interpretability of

260   past vegetation dynamics is enhanced by extensive surface pollen samples collected across elevation gradients, which serve

261   as modern analogues of climate−vegetation relationships (Fig. 6b). These features can contribute to understanding regional-

262   scale ecosystem dynamics in response to climate and anthropogenic forcings.

263   Although the site-specific findings can be explored through the main text of related publications in international peer-

264 reviewed journals, the 'interoperability' and 'reusability' of original data (and thus citation of the paper) has been relatively
265 rare in international collaborative research. A recent global pollen data compilation study for the Northern Hemisphere, for
266 example, included only a single dataset from Korea (Herzschuh et al., 2023). Our dataset framework can serve as not only to
267 characterize the inventory of paleo-studies in South Korea, but also a template for organizing similar inventories in other
268 underrepresented regions.

269 Moving forward, our database (GEK, GeoEcoKorea), integrated with the open-access web platform, aims to enhance the
270 'findability' and 'accessibility' of Korean paleoecological data, contributing to the FAIR data sharing and the advance of this
271 research field both regionally and globally. GEK currently curates two categories of datasets: those published from Korean-
272 language publications and those from international publications. Korean-language datasets are curated for their values in site
273 revisits, regional syntheses, and legacy preservation. As a part of our strategy for global integration, internationally published
274 datasets from nine Korean paleo-sites have been submitted to and are now stored in the Neotoma Paleoecology Database (as
275 of June 2025). Furthermore, GEK plans to become a Neotoma's constituent database (e.g., The Indo-Pacific Pollen Database)
276 (Herbert et al., 2024). Affiliation with Neotoma will be critical for increasing the availability of Korean paleoecological data
277 and fostering trust among both international and Korean researchers (Thomer et al., 2025; Yoon, 2017). These trust-building
278 efforts will help our regional initiatives to encourage and support Korean Quaternary researchers to contribute their data,
279 thereby fostering a more collaborative and inclusive data-sharing culture.

280 **Supplement**

281 A supplementary document is available online at http://doi.org/10.5194/essd-XXX.

282 **Author contributions**

283 SHK: Conceptualization, Data curation, Formal analysis, Investigation, Resources, Software, Visualization, Writing - Original
284 Draft, Review & Editing. EB: Conceptualization, Supervision, Project administration, Funding acquisition, Investigation,
285 Validation, Resources, Writing - Original Draft, Review & Editing.

286 **Competing interests**

287 The authors have no conflicts of interest to declare.

288 **Acknowledgements**

293 **Financial support**

References

299   Birks, H. H. and Birks, H. J. B.: Multi-proxy studies in palaeolimnology, Vegetation History and Archaeobotany, 15, 235–
300   251, https://doi.org/10.1007/s00334-006-0066-6, 2006.

301   Bradley, R. S.: Paleoclimatology: Reconstructing Climates of the Quaternary, 3rd ed., Academic Press, Oxford, 2015.

302   Blaauw, M.: IntCal: Radiocarbon Calibration Curves. R package version 0.3.1. https://CRAN.R-
303   project.org/package=IntCal, 2022.

304   Center for International Earth Science Information Network - CIESIN - Columbia University.: Gridded Population of the
305   World, Version 4 (GPWv4): Land and Water Area, Revision 11. Palisades, NY: NASA Socioeconomic Data and
306   Applications Center (SEDAC). https://doi.org/10.7927/h4z60m4z, 2018.

307   Crann, C. A., Patterson, R. T., Macumber, A. L., Galloway, J. M., Roe, H. M., Blaauw, M., Swindles, G. T., and Falck, H.:
308   Sediment accumulation rates in subarctic lakes: Insights into age-depth modeling from 22 dated lake records from the
309   Northwest Territories, Canada, Quaternary Geochronology, 27, 131–144, https://doi.org/10.1016/j.quageo.2015.02.001,
310   2015.

311   Dutton, A., Rubin, K. H., Mclean, N., Bowring, J., Bard, E., Edwards, R. L., Henderson, G. M., Reid, M. R., Richards, D.
312   A., Sims, K. W. W., Walker, J. D., and Yokoyama, Y.: Quaternary Geochronology Data reporting standards for
313   publication of U-series data for geochronology and timescale assessment in the earth sciences, Quaternary
314   Geochronology, 39, 142–149, https://doi.org/10.1016/j.quageo.2017.03.001, 2017.

315   Farley, S. S., Dawson, A., Goring, S. J., and Williams, J. W.: Situating ecology as a big-data science: Current advances,
316   challenges, and solutions, BioScience, 68, 563–576, https://doi.org/10.1093/biosci/biy068, 2018.

317   Flantua, S. G. A., Mottl, O., Felde, V. A., Bhatta, K. P., Birks, H. H., Grytnes, J. A., Seddon, A. W. R., and Birks, H. J. B.:
318   A guide to the processing and standardization of global palaeoecological data for large-scale syntheses using fossil
319   pollen, Global Ecology and Biogeography, 32, 1377–1394, https://doi.org/10.1111/geb.13693, 2023.

320   Flantua, S. G. A., Hooghiemstra, H., Grimm, E. C., Behling, H., Bush, M. B., González-Arango, C., Gosling, W. D., Ledru,
321   M. P., Lozano-García, S., Maldonado, A., Prieto, A. R., Rull, V., and Van Boxel, J. H.: Updated site compilation of the
322   Latin American Pollen Database, Review of Palaeobotany and Palynology, 223, 104–115,
323   https://doi.org/10.1016/j.revpalbo.2015.09.008, 2015.

324   Gaillard, M. J., Sugita, S., Mazier, F., Trondman, A. K., Broström, A., Hickler, T., Kaplan, J. O., Kjellström, E., Kokfelt,
325   U., Kuneš, P., Lemmen, C., Miller, P., Olofsson, J., Poska, A., Rundgren, M., Smith, B., Strandberg, G., Fyfe, R.,
326   Nielsen, A. B., Alenius, T., Balakauskas, L., Barnekow, L., Birks, H. J. B., Bjune, A., Björkman, L., Giesecke, T.,
327   Hjelle, K., Kalnina, L., Kangur, M., Van Der Knaap, W. O., Koff, T., Lageras, P., Latałowa, M., Leydet, M.,
328   Lechterbeck, J., Lindbladh, M., Odgaard, B., Peglar, S., Segerström, U., Von Stedingk, H., and Seppä, H.: Holocene
329   land-cover reconstructions for studies on land cover-climate feedbacks, Climate of the Past, 6, 483–499,
330   https://doi.org/10.5194/cp-6-483-2010, 2010.

331   Goring, S., Williams, J. W., Blois, J. L., Jackson, S. T., Paciorek, C. J., Booth, R. K., Marlon, J. R., Blaauw, M., and
332   Christen, J. A.: Deposition times in the northeastern United States during the Holocene: Establishing valid priors for
333   Bayesian age models, Quaternary Science Reviews, 48, 54–60, https://doi.org/10.1016/j.quascirev.2012.05.019, 2012.

334 Heaton, T. J., Köhler, P., Butzin, M., Bard, E., Reimer, R. W., Austin, W. E. N., Bronk Ramsey, C., Grootes, P. M.,
335   Hughen, K. A., Kromer, B., Reimer, P. J., Adkins, J., Burke, A., Cook, M. S., Olsen, J., and Skinner, L. C.: Marine20 -
336   The Marine Radiocarbon Age Calibration Curve (0-55,000 cal BP), Radiocarbon, 62, 779–820,
337   https://doi.org/10.1017/RDC.2020.68, 2020.

338 Herbert, A. V., Haberle, S. G., Flantua, S. G. A., Mottl, O., Blois, J. L., Williams, J. W., George, A., and Hope, G. S.: The
339   Indo-Pacific Pollen Database-a Neotoma constituent database, Climate of the Past, 20, 2473–2485,
340   https://doi.org/10.5194/cp-20-2473-2024, 2024.

341 Herzschuh, U., Böhmer, T., Li, C., Chevalier, M., Hébert, R., Dallmeyer, A., Cao, X., Bigelow, N. H., Nazarova, L.,
342   Novenko, E. Y., Park, J., Peyron, O., Rudaya, N. A., Schlütz, F., Shumilovskikh, L. S., Tarasov, P. E., Wang, Y., Wen,
343   R., Xu, Q., and Zheng, Z.: LegacyClimate 1.0: a dataset of pollen-based climate reconstructions from 2594 Northern
344   Hemisphere sites covering the last 30 kyr and beyond, Earth Syst. Sci. Data, 15, 2235–2258,
345   https://doi.org/10.5194/essd-15-2235-2023, 2023.

346 Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., Jaccard, S., Tierney, J., Dätwyler, C., Axford, Y.,
347   Brussel, T., Cartapanis, O., Chase, B., Dawson, A., de Vernal, A., Engels, S., Jonkers, L., Marsicek, J., Moffa-Sánchez,
348   P., Morrill, C., Orsi, A., Rehfeld, K., Saunders, K., Sommer, P. S., Thomas, E., Tonello, M., Tóth, M., Vachula, R.,
349   Andreev, A., Bertrand, S., Biskaborn, B., Bringué, M., Brooks, S., Caniupán, M., Chevalier, M., Cwynar, L., Emile-
350   Geay, J., Fegyveresi, J., Feurdean, A., Finsinger, W., Fortin, M. C., Foster, L., Fox, M., Gajewski, K., Grosjean, M.,
351   Hausmann, S., Heinrichs, M., Holmes, N., Ilyashuk, B., Ilyashuk, E., Juggins, S., Khider, D., Koinig, K., Langdon, P.,
352   Larocque-Tobler, I., Li, J., Lotter, A., Luoto, T., Mackay, A., Magyari, E., Malevich, S., Mark, B., Massaferro, J.,
353   Montade, V., Nazarova, L., Novenko, E., Pařil, P., Pearson, E., Peros, M., Pienitz, R., Płóciennik, M., Porinchu, D.,
354   Potito, A., Rees, A., Reinemann, S., Roberts, S., Rolland, N., Salonen, S., Self, A., Seppä, H., Shala, S., St-Jacques, J.
355   M., Stenni, B., Syrykh, L., Tarrats, P., Taylor, K., van den Bos, V., Velle, G., Wahl, E., Walker, I., Wilmshurst, J.,
356   Zhang, E., and Zhilich, S.: A global database of Holocene paleotemperature records, Scientific Data, 7, 1–34,
357   https://doi.org/10.1038/s41597-020-0445-3, 2020.

358 Kim, J. C., Han, M., Ahn, H.-S., Yoon, H. H., Lee, J.-Y., Park, S., Cho, A., Kim, J. Y., Nahm, W.-H., Choi, H.-W., Lim, J.,
359   Yang, D.-Y., Hong, S.-S., and Yi, S.: Quaternary environmental studies in South Korea, Episodes, 47, 511–535,
360   https://doi.org/10.18814/epiiugs/2024/02403s09, 2024.

361 Lacourse, T. and Gajewski, K.: Current practices in building and reporting age-depth models. Quaternary Research 96: 28-
362   38, https://doi.org/10.1017/qua.2020.47, 2020.

363 Lee, J., Jun, C.-P., Yi, S., Kim, Y., Lee, E., and Kim, D.: Modern pollen–climate relationships and their application for
364   pollen-based quantitative climate reconstruction of the mid-Holocene on the southern Korean Peninsula, The Holocene,
365   32, 127–136, https://doi.org/10.1177/09596836211060493, 2022.

366 McKenzie, K.G.: Results of the speleological survey in South Korea 1966; XXII, Subterranean Ostracoda from South
367   Korea. Bulletin of the National Science Museum, Tokyo 15: 155-166, 1972.

368 Mottl, O., Flantua, S. G. A., Bhatta, K. P., Felde, V. A., Giesecke, T., Goring, S., Grimm, E. C., Haberle, S., Hooghiemstra,
369   H., Ivory, S., Kuneš, P., Wolters, S., Seddon, A. W. R., and Williams, J. W.: Global acceleration in rates of vegetation
370   change over the past 18,000 years, Science, 372, 860–864, https://doi.org/10.1126/science.abg1685, 2021.

371 Nahm, W.H.: Present situation of research of Quaternary terrestrial unconsolidated sediments, in Korea. Journal of the
372   Geological Society of Korea 54: 107-119, 2018.

373 Park, J. and Park, J.: Pollen-based temperature reconstructions from Jeju island, South Korea and its implication for coastal
374   climate of East Asia during the late Pleistocene and early Holocene, Palaeogeogr. Palaeoclimatol. Palaeoecol., 417, 445–
375   457, https://doi.org/10.1016/j.palaeo.2014.10.005, 2015.

376 Park, J.: The Applicability of Stable Isotope Analyses on Sediments to Reconstruct Korean Paleoclimate. Journal of the
377   Korean Geographical Society 43: 477-494, 2008.

Prentice, C.: Records of Vegetation in Time and Space: the Principles of Pollen Analysis, in: Vegetation History. Handbook of Vegetation Science, Vol 7, edited by: Huntley, B. and Webb, T., Springer Netherlands, Dordrecht, 17–42, https://doi.org/10.1007/978-94-009-3081-0_2, 1988.

Reimer, P. J., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Hajdas, I., Heaton, T. J., Hogg, A. G., Hughen, K. A., Kromer, B., Manning, S. W., Muscheler, R., Palmer, J. G., Pearson, C., van der Plicht, J., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Turney, C. S. M., Wacker, L., Adolphi, F., Büntgen, U., Capano, M., Fahrni, S. M., Fogtmann-Schulz, A., Friedrich, R., Köhler, P., Kudsk, S., Miyake, F., Olsen, J., Reinig, F., Sakamoto, M., Sookdeo, A., and Talamo, S.: The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP), Radiocarbon, 62, 725–757, https://doi.org/10.1017/RDC.2020.41, 2020.

Smith, R. J., Lee, J., and Chang, C. Y.: Nonmarine Ostracoda (Crustacea) from Jeju Island, South Korea, including descriptions of two new species, J. Nat. Hist., 49, 37–76, https://doi.org/10.1080/00222933.2014.946110, 2015.

Flantua, S. G. A., Hooghiemstra, H., Grimm, E. C., Behling, H., Bush, M. B., González-Arango, C., Gosling, W. D., Ledru, M.-P., Lozano-García, S., Maldonado, A., Prieto, A. R., Rull, V., and Van Boxel, J. H.: Updated site compilation of the Latin American Pollen Database, Rev. Palaeobot. Palynol., 223, 104–115, https://doi.org/10.1016/j.revpalbo.2015.09.008, 2015.

Thomer, A., Williams, J., Goring, S., and Blois, J.: The Valuable, Vulnerable, Long Tail of Earth Science Databases, Eos, 106, https://doi.org/10.1029/2025EO250107, 2025.

Vidaña, S. D. and Goring, S. J.: neotoma2: An R package to access data from the Neotoma Paleoecology Database, Journal of Open Source Software, 8, 5561, https://doi.org/10.21105/joss.05561, 2023.

Wang, Y., Pineda-Munoz, S., and McGuire, J. L.: Plants maintain climate fidelity in the face of dynamic climate change, Proc. Natl. Acad. Sci., 120, https://doi.org/10.1073/pnas.2201946119, 2023.

Wang, Y., Shipley, B. R., Lauer, D. A., Pineau, R. M., and McGuire, J. L.: Plant biomes demonstrate that landscape resilience today is the lowest it has been since end-Pleistocene megafaunal extinctions, Glob. Chang. Biol., 26, 5914–5927, https://doi.org/10.1111/gcb.15299, 2020.

Whitlock, C. and Larsen, C.: Charcoal as a Fire Proxy, in: Tracking Environmental Change Using Lake Sediments. Developments in Paleoenvironmental Research, vol. 3, edited by: Smol, J. P., Birks, H. J. B., Last, W. M., Bradley, R. S., and Alverson, K., Springer, Dordrecht, 75–97, https://doi.org/10.1007/0-306-47668-1_5, 2002.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. 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. G., 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.

Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., Graham, R. W., Smith, A. J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A. C., Betancourt, J. L., Bills, B. W., Booth, R. K., Buckland, P. I., Curry, B. B., Giesecke, T., Jackson, S. T., Latorre, C., Nichols, J., Purdum, T., Roth, R. E., Stryker, M., and Takahara, H.: The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource, Quat. Res., 89, 156–177, https://doi.org/10.1017/qua.2017.105, 2018.

Yoon, A.: Data reusers' trust development, Journal of the Association for Information Science and Technology, 68, 946–956, https://doi.org/10.1002/asi.23730, 2017.

Table 1 Description of data types in the Site Inventory dataset.

Section	Field name	Description
Site	SiteID	Unique identifier for the site for paleorecord or surface pollen; formatted as Site-###.
	Site name	Name of the site.
	Type of study site	Classified either paleo-site or surface site based on the primary sampling objective of the study.
Geospatial information	Longitude	Longitude of the site location in four-digit decimal format (e.g., 123.1234).
	Latitude	Latitude of the site location in four-digit decimal format.
	Elevation (m)	Elevation of site in meters above mean sea level
Depositional setting	Type of depositional environment	Categorized as six groups: 1) Open Coastal Zone, 2) Estuary, 3) Lagoon, 4) River, 5) Volcanic Cone, or 6) Other.
	Site description	Detailed description of depositional conditions (e.g., blanket peat sediment in hilly district)
Sample	Sample type	Categorized as three groups: 1) Core, 2) Trench, or 3) Surface pollen
	SampleID	Unique identifier for each sample, formatted as Core-###, Trench-###, or Surface-###.
	Sample name	Name of the core, trench, or surface pollen.
Geochronology	Number of all age controls	Number of all age controls of each core or trench.
	Number of non-absolute dates	Number of non-direct dated age controls (e.g., the core top ages using the coring years or ages extrapolated from age-depth models).
	Absolute dating methods	Method used for dating materials from each core or trench.
	Number of ^{14}C dates	Number of radiocarbon dates.
	Number of dates from other direct methods	Number of dates from direct dating methods other than radiocarbon.
	Oldest age	Oldest age of chronological control points
	Youngest age	Youngest age of chronological control points
	Age scale	Age scales of Oldest and Youngest ages (e.g., Calendar_BP: calendar years before 1950 CE, OSL_yrs BP, and U-Th_yrs_BP)
	Total depth range	Vertical distance between the maximum and minimum depth points of age control points (unit: cm).
Proxy	Pollen	Indicates whether pollen analysis data are available (Yes/NA).
	Diatom	Indicates whether diatom analysis data are available (Yes/NA).
	Grain size	Indicates whether granulometric data are available (Yes/NA).
	Other proxies	Lists additional proxy types available (or NA if none are available).
Publication	PublicationID	Identifier for the publication, formatted as Pub-###, linking to PublicationID in the Publication metadata.

Table 2 Number of sites (total $n = 72$ with multi-proxy $n = 57$ and mono-proxy $n = 15$) for paleo-records by type of sample, depositional environment, and proxy.

Category	Type	Number of paleo-sites
Sample type	Core	54
	Trench	18
Depositional environment	Open Coastal Zone	15
	Estuary	15
	Lagoon	11
	River	18
	Volcanic Cone	6
	Other	7
Proxy	Pollen	42
	Diatom	14
	Grain size	45
	Other proxies	42
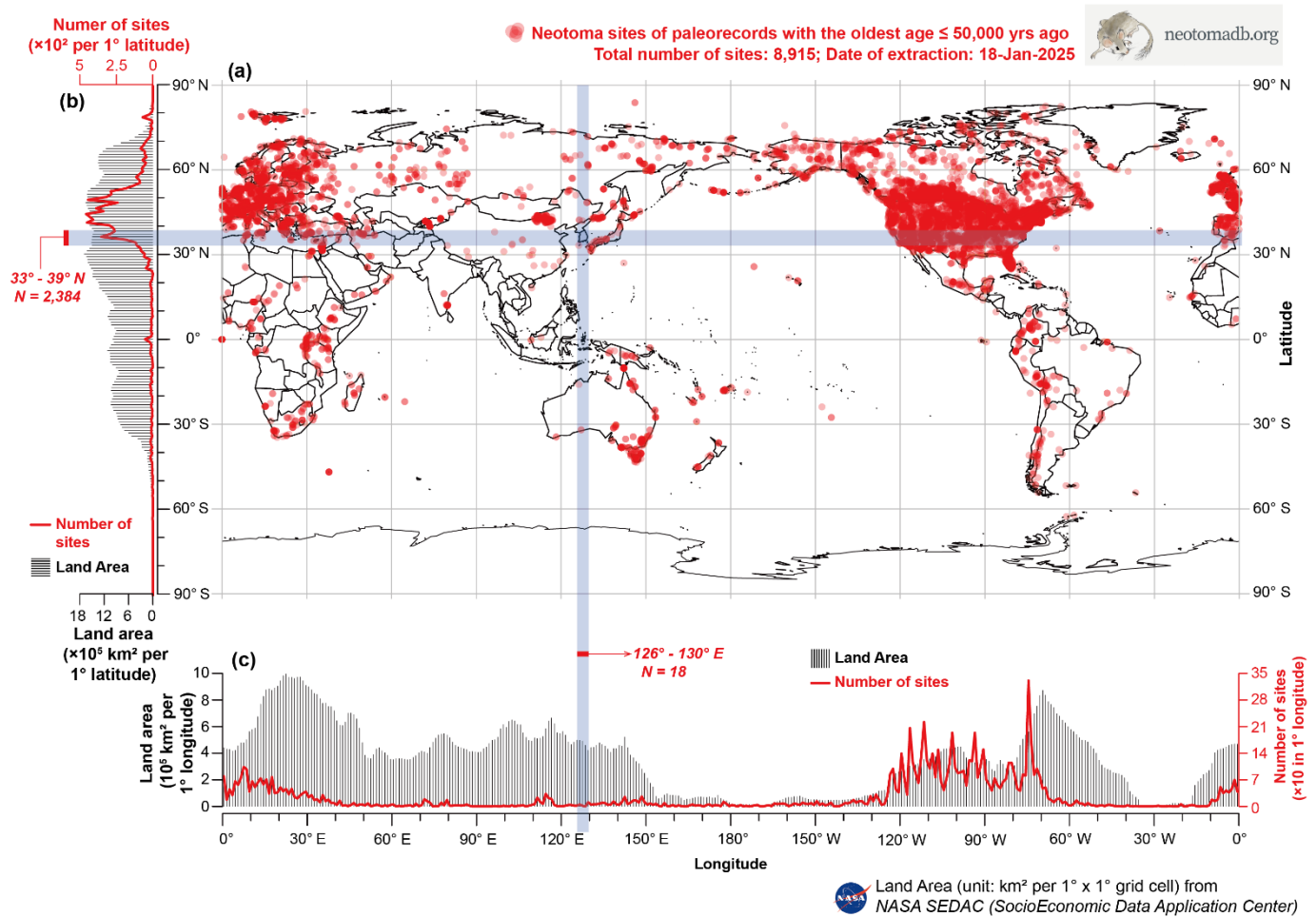
Figure 1 (a) Global distribution of sites with paleorecords (age range: ≤ 50,000 years ago) uploaded to the Neotoma Paleoecology Database (Total sites: 7,548; Extraction date: 2025/01/18). These records were selected based on the availability of data within the last 50,000 years, excluding sites having only modern data–surface samples, pollen trap, and datasets with the oldest age being after 1950 CE. *Blue band: The blue band indicates the Korean Peninsula and spatial gaps in site coverage. **(b)** Number of sites and total land area (10^5 km^2) per 1° latitude. Land area represents the sum of land areas (unit: 10^5 km^2) across all longitudes within each 1° latitudinal band. **(c)** Number of sites and total land area (unit: 10^5 km^2) per 1° longitude. The land area estimates in **(b)** & **(c)** are derived from 1-degree resolution datasets of the Grided Population of the World, Version 4 (GPWv4): Land and Water area, Revision 11 (https://doi.org/10.7927/H4Z60M4Z), produced by NASA's Socioeconomic Data and Applications Center (SEDAC).
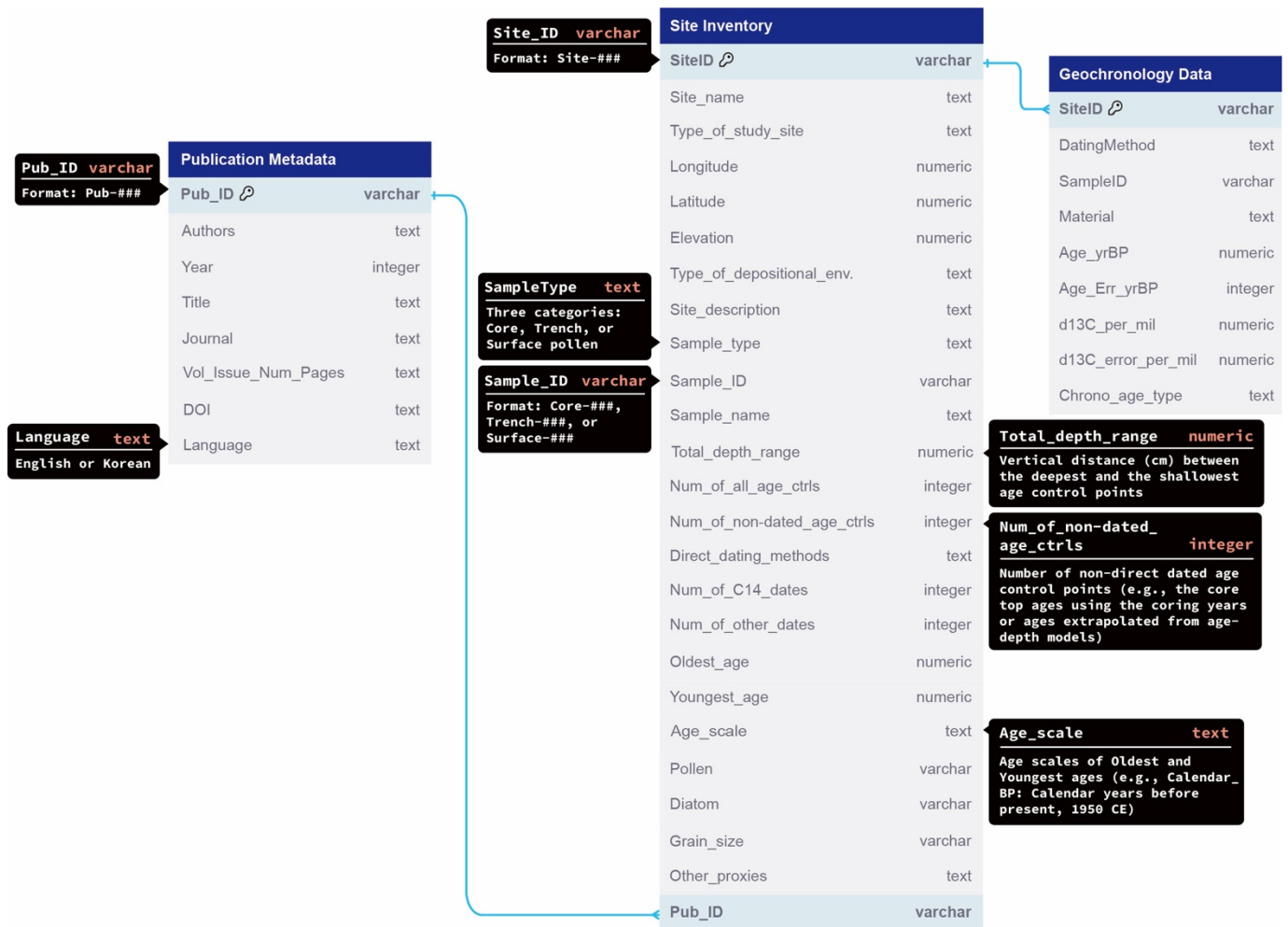
Figure 2 Entity-Relationship Diagram (ERD) illustrating the three datasets built in this study. This ERD was created using dbdiagram.io. Data types, following standard conventions of SQL (Standard Query Language), include varchar (variable-length character strings), text (long-form text), numeric (decimal values), and integer (whole number). Table 1 provides details of data fields.
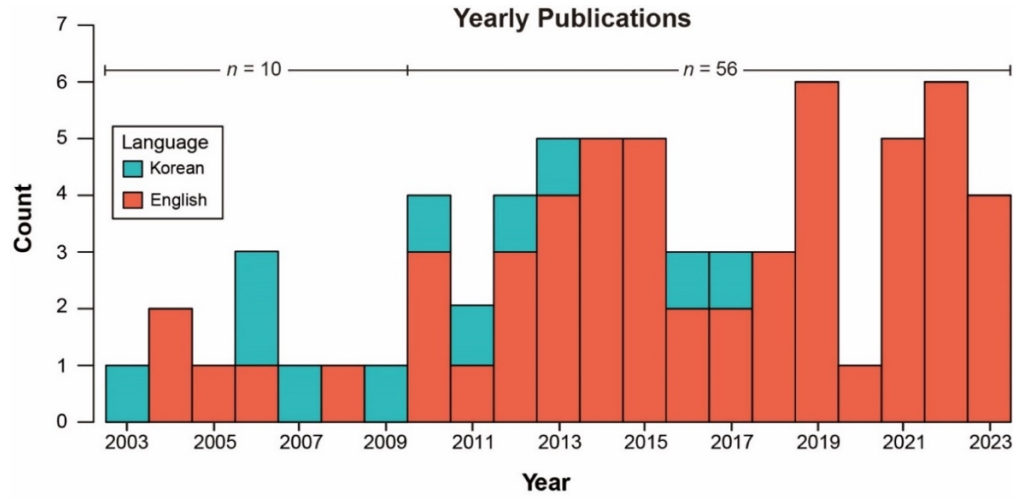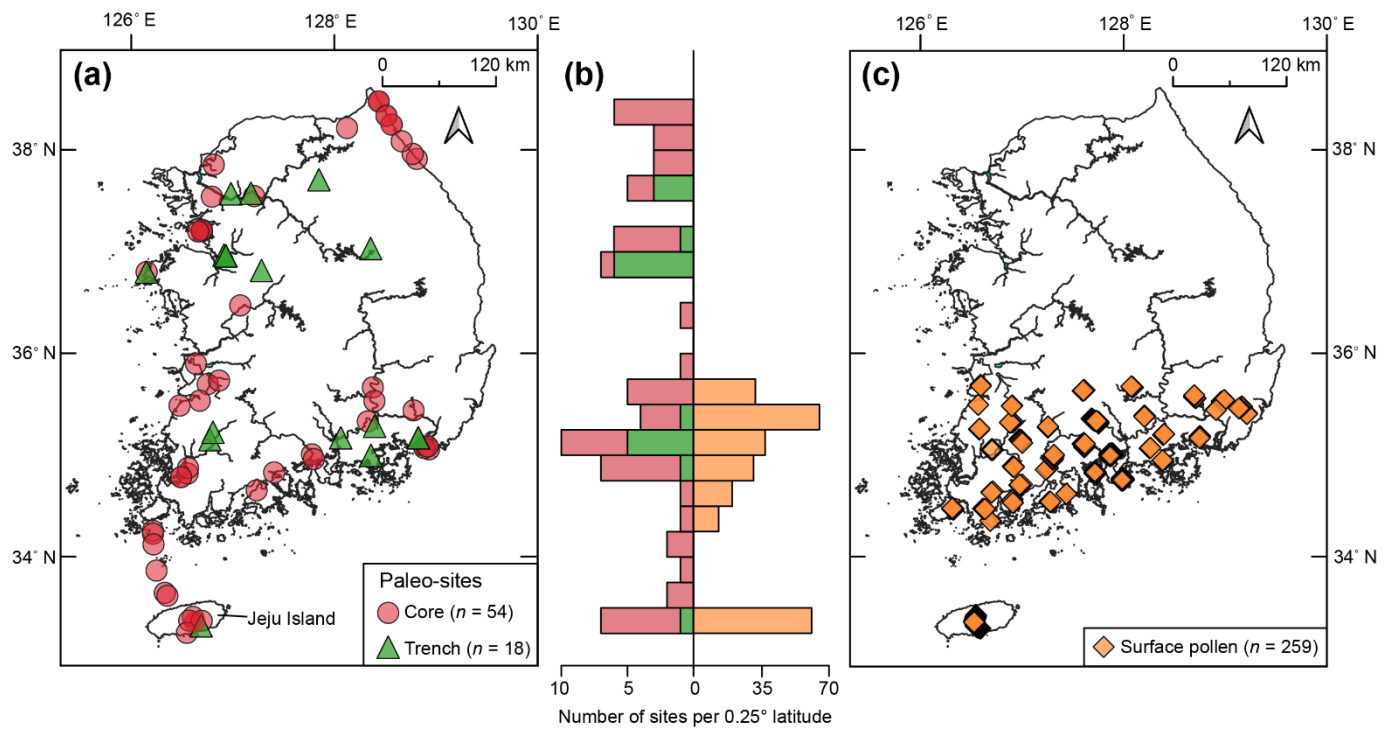
Figure 3 Barplot of annual publications from 2003 to 2023.

Figure 4 Geographic distributions of sites. **(a)** Paleo-sites. **(b)** Surface pollen sites. Site markers may appear overlapped in areas with high site density. Elevational distributions of the sites are shown in Fig. 6.
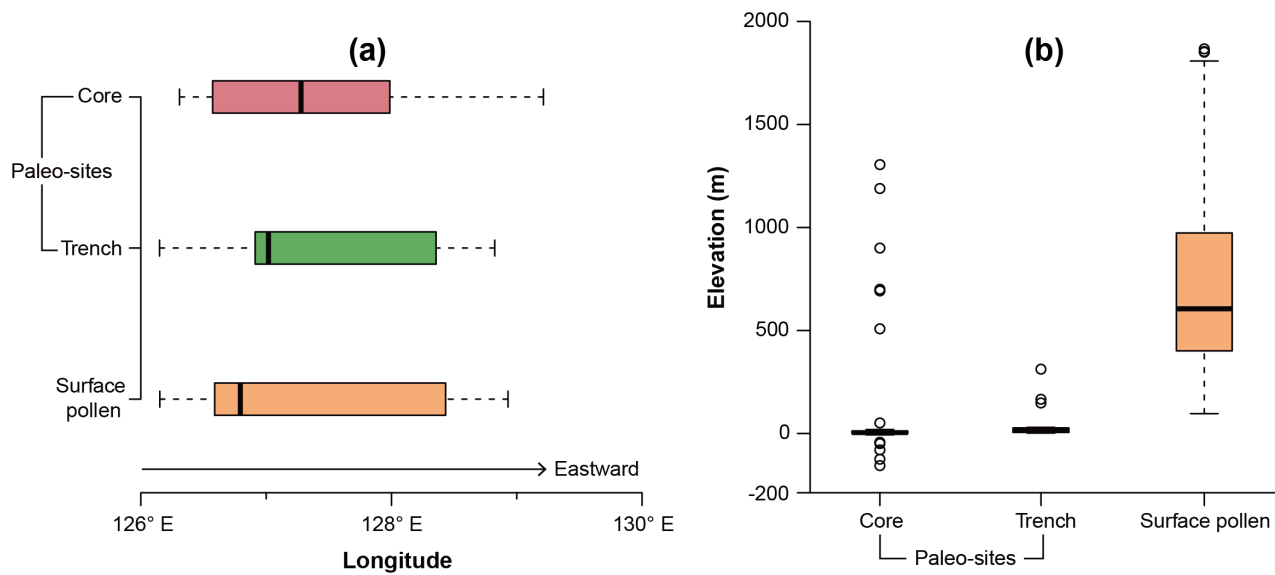
Figure 5 Boxplots of site locations and elevations. **(a)** Longitudinal distribution of sites. **(b)** Elevational distribution of sites.
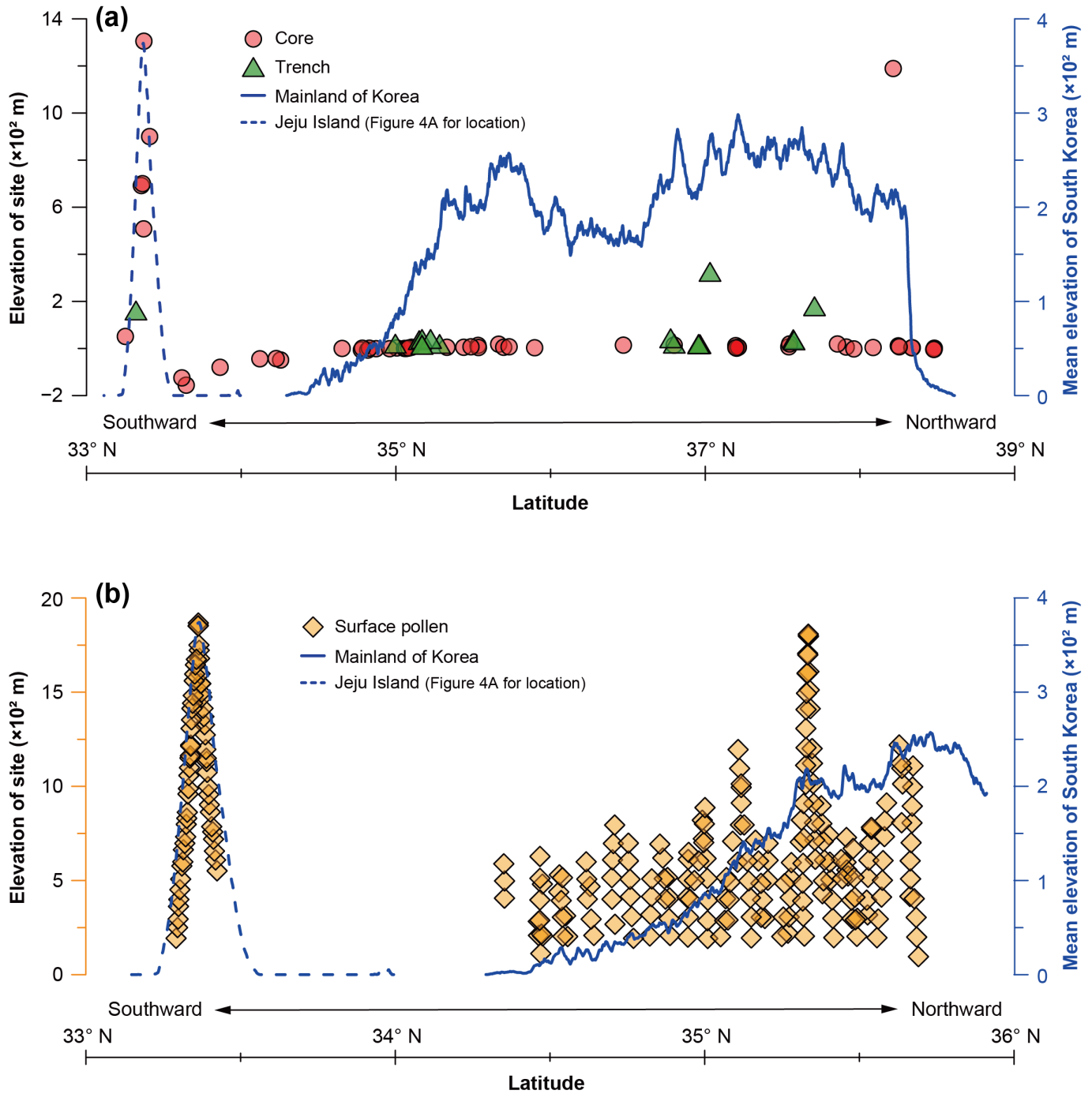
Figure 6 Elevational distribution of sites. **(a)** Paleo-sites. **(b)** Surface pollen sites. These plots have different latitudinal ranges (A: 33° N to 39° N, B: 33° N to 36° N). The blue lines display mean elevations along a latitudinal gradient, averaged elevations across all longitudes within the mainland (solid line) and Jeju Island (dashed line) (Source of Digital Elevation Model: NASA SRTM Void Filled, https://doi.org/10.5066/F7F76B1X). The location of Jeju Island is shown in Figure 4a.
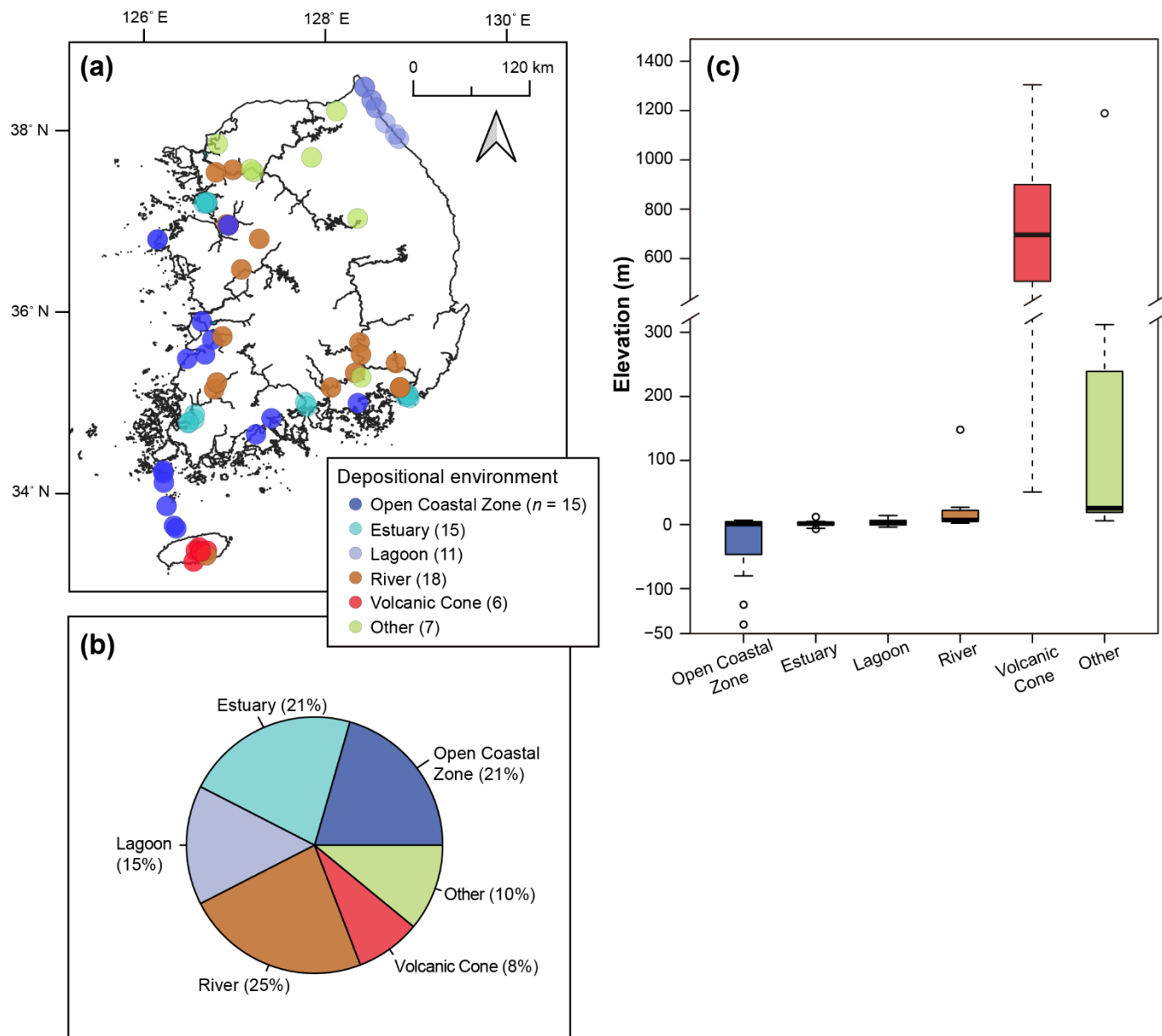
Figure 7 Paleo-sites (n = 72) categorized by six types of depositional environments. **(a)** Geographic locations of sites. **(b)** Proportion of sites by depositional environment. **(c)** Boxplots of elevations. The median elevations for each environmental type are as follows: Open Coastal Zone (1 m), Estuary (1.5 m), Lagoon (3 m), River (10 m), Volcanic Cone (696 m), and Other (21 m).
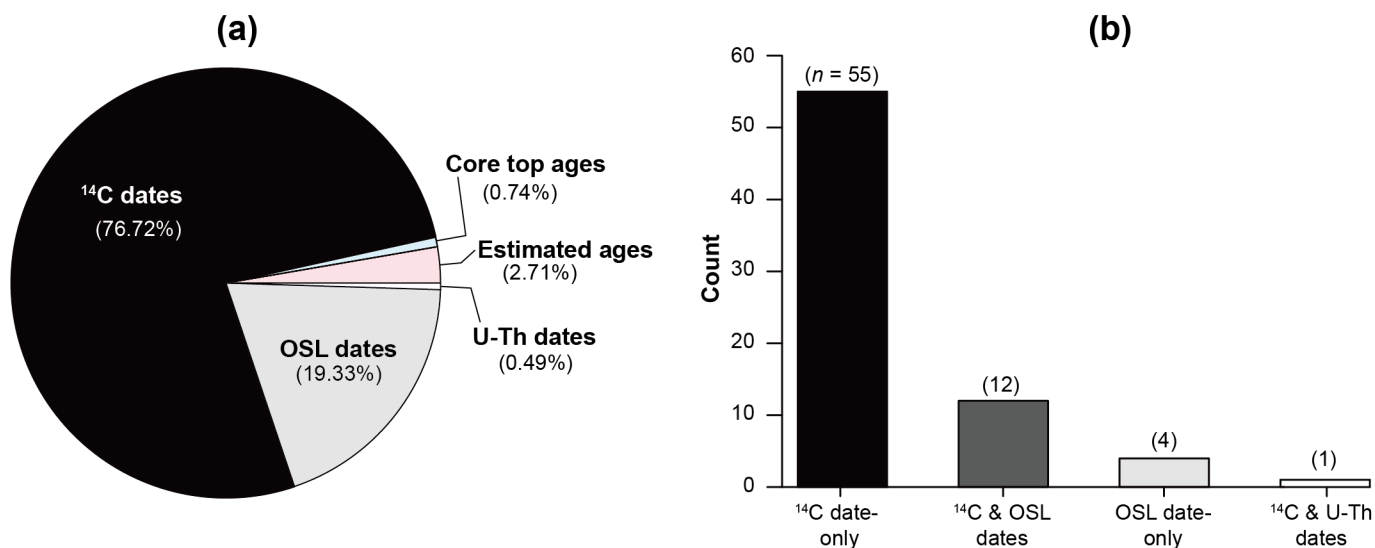
Figure 8 (a) Pie chart showing the proportional distribution of age control types, including three direct dating methods, ^{14}C, OSL, and U-Th, and two indirect controls, Core top ages and Estimated ages derived from age-depth modeling. Percentages may not total 100% due to rounding. **(b)** Bar plot showing the number of paleo-sites grouped by direct dating methods: sites dated only by ^{14}C, by both ^{14}C and OSL, by OSL only, and by ^{14}C and U-Th.
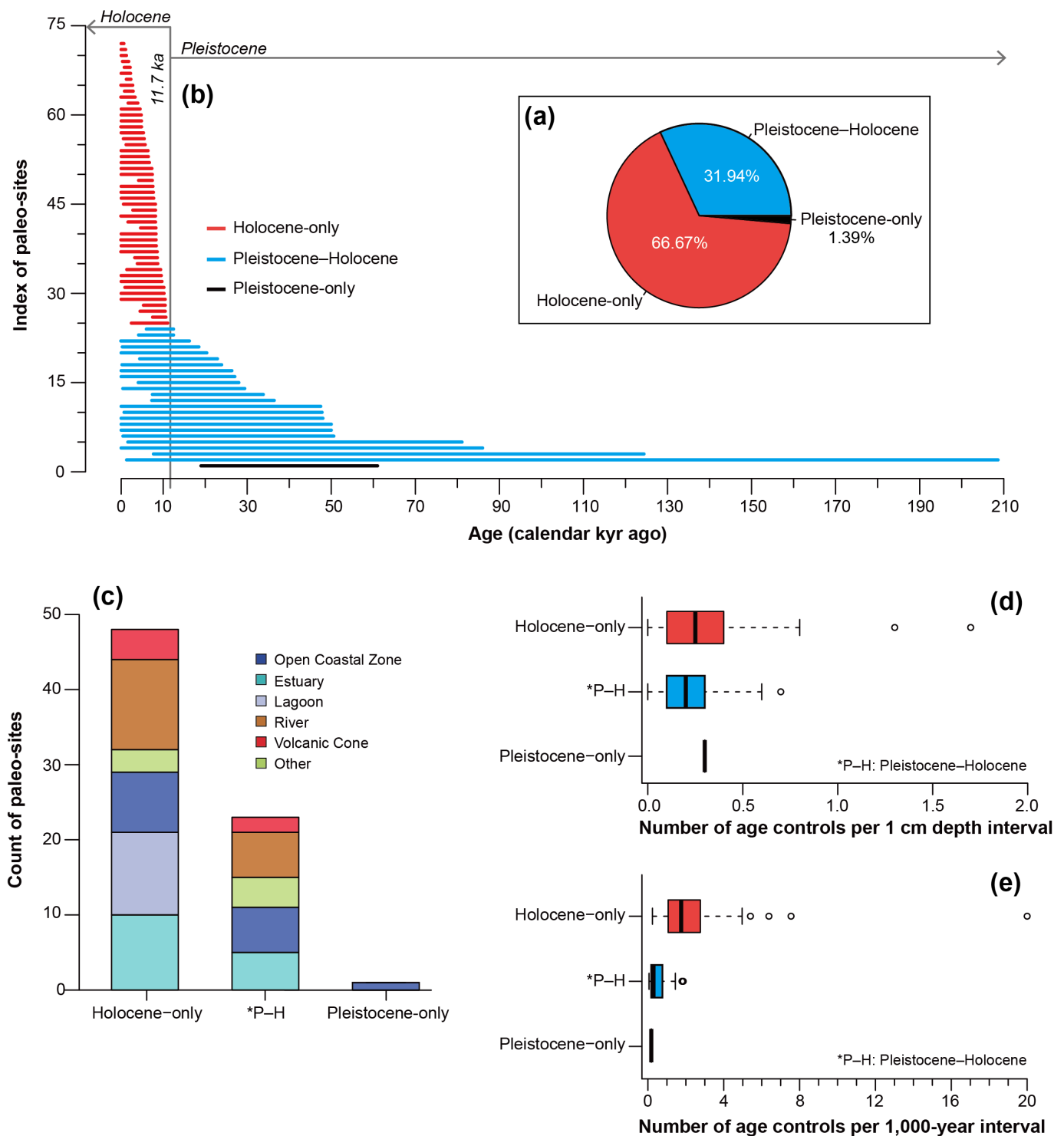
Figure 9 (a) Distribution of paleo-sites by three groups of age-coverage (kyr = 1,000 yea): Holocene-only (*n* = 48), Pleistocene−Holocene (23), and Pleistocene-only (1). **(b)** Chronological extent of records from paleo-sites, ordered by descending oldest ages within each age-coverage group. **(c)** Counts of depositional environments by age-coverage group. **(d)** Boxplot of the number of age controls per 1 cm depth interval. Medians of age control density per 1cm-depth for Holocene-only: 0.025, Pleistocene−Holocene: 0.020, and Pleistocene-only: 0.030. **(e)** Boxplot of the number of age contorls per 1,000-year interval. Medians of age control density per millennium for Holocene-only: 1.752, Pleistocene−Holocene: 0.292, and Pleistocene-only: 0.167.
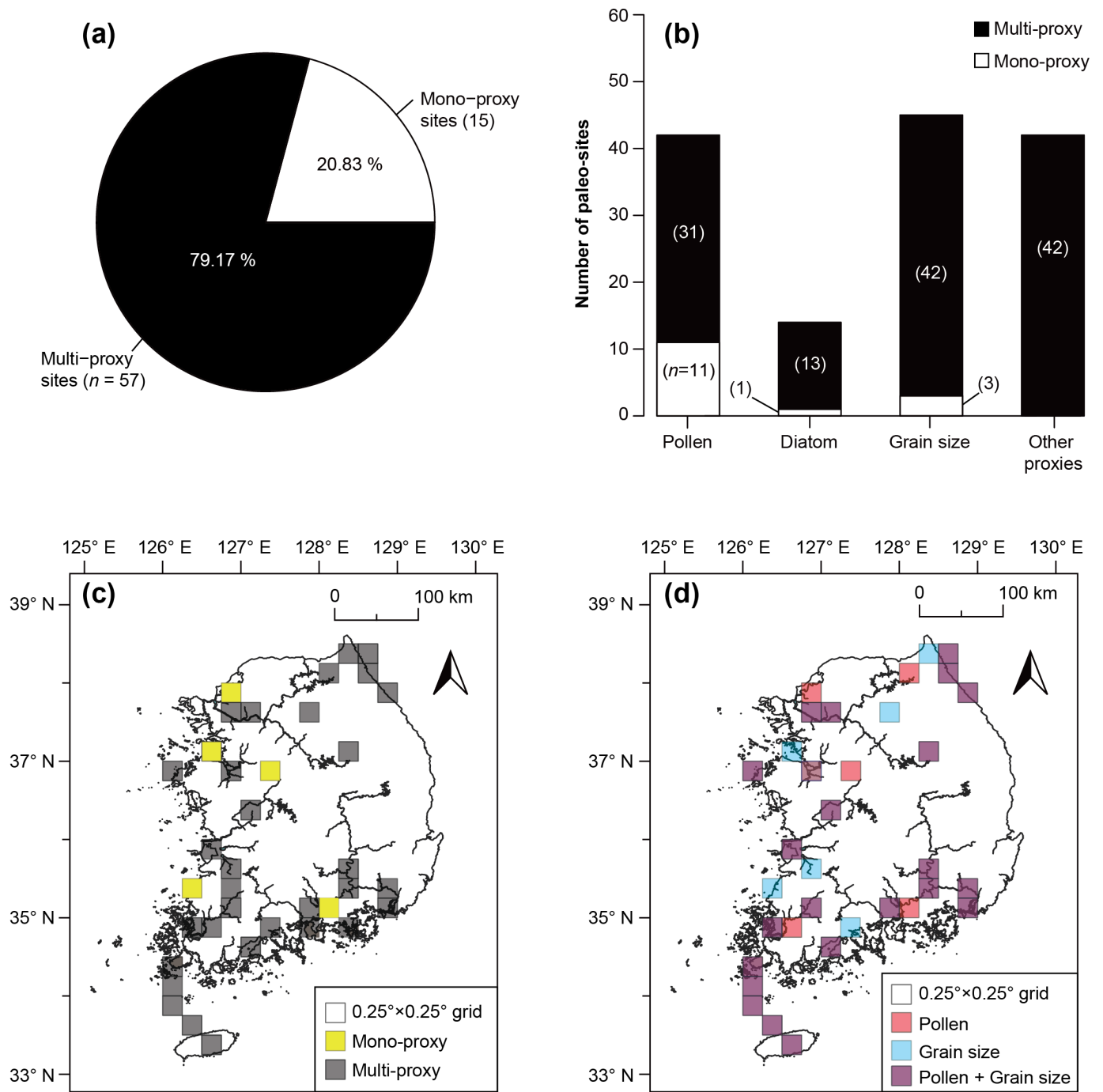
Figure 10 (a) Proportion of mono-proxy and multi-proxy sites. **(b)** Number of paleo-sites by four proxy types. **(c & d)** Geospatial distribution of proxy-based records for a 0.25° × 0.25° grid. These maps display grids containing proxy-based records from paleo-sites in South Korea. In **(c)**, grids are classified as either mono-proxy or multi-proxy. **(d)** shows grids containing records of pollen and grain size, the most widely used proxies in South Korea.
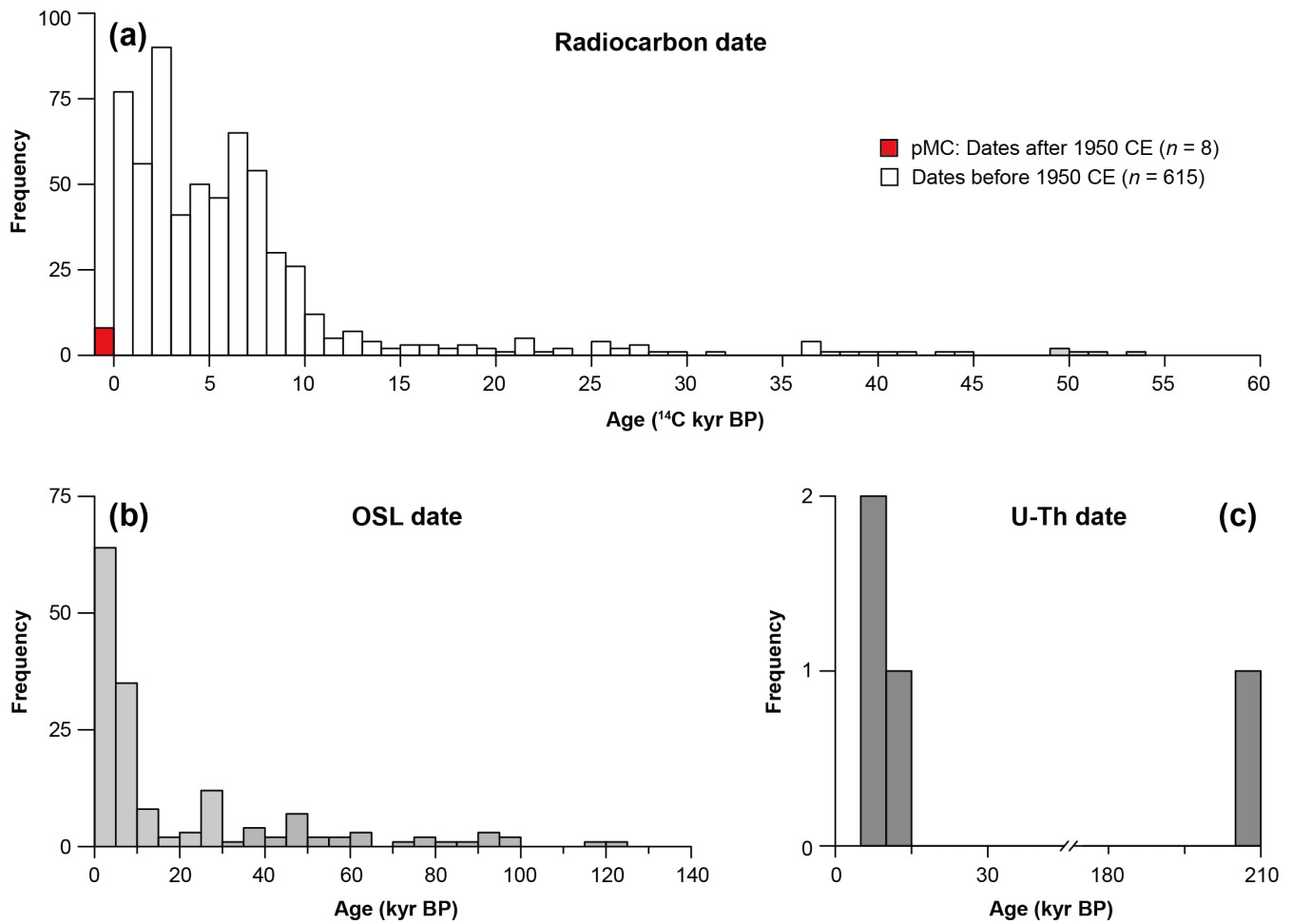
Figure 11 Histograms of dates over ages (unit: kyr BP, where means 1,000 years before present). **(a)** Radiocarbon dates (range of pre-1950 dates: >53,671 to 27 ^{14}C years BP; median: 4,790 ^{14}C years BP). **(b)** OSL dates (range: 124,400 to 140 years BP; median: 7,114 years BP). **(c)** U-Th dates (range: 208,000 to 6,900 years BP; median: 10,000 years BP). *Note: The reference year for "present" is 1950 CE for radiocarbon and U–Th dates, and the measurement year for OSL dates.
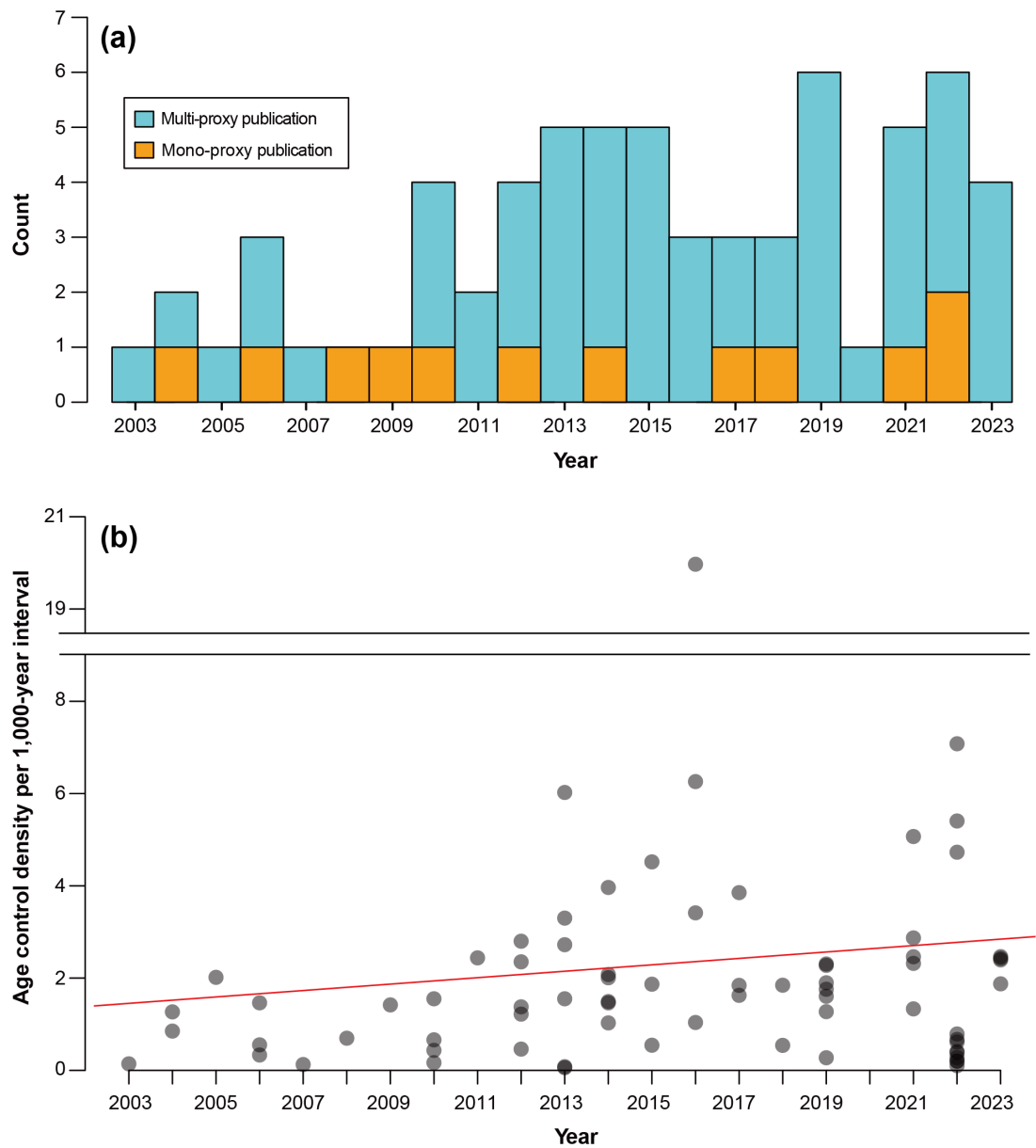
Figure 12 Temporal trends in **(a)** the number of multi-proxy and mono-proxy publications and **(b)** age control density per 1,000-year interval. In **(a)**, 64 publications are shown because two articles studying surface pollen samples are excluded. In **(b)**, the age control density, defined as the total number of age controls by the difference between the oldest and youngest ages (unit: 1,000 years), is calculated for sediments from 72 paleo-sites. The year corresponds to when each record from a paleo-site was published. When more than two publications reported data from the same site, the earlier one was selected. The red line represents a linear regression ($y = 0.06x – 112.27$; $R^2 = 0.02$; *p-value* = 0.30).