[DATA PAPER]

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, South Korea*

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

~~*Correspondence to*: Eunji Byun (~~**Author correspondence: 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea, E-mail: eb@yonsei.ac.kr*~~)~~

Abstract~~.~~

Ecological insights beyond human-observable time scales are derived from ~~geologically~~records preserved ~~records~~ in ~~lake and wetland~~geological sediments around the world. Nonetheless, significant regional data gaps persist in global syntheses of these records as ~~regional~~ open data practices are still emerging~~. South~~ around the world. Korean Quaternary ~~paleoecology~~ paleoecological data remain underrepresented in ~~these~~ global research efforts, despite a growing body of ~~the relevant~~site-level research. Here, we organize an inventory of 328 paleoecological study sites (72 paleo-sites ~~for sediment~~with paleoproxy-based records from sediments and 256 surface sites ~~for surface~~with modern pollen samples) in South Korea, compiled from 66 research articles published ~~between~~from 2003 ~~and~~to 2023. We have ~~structured~~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 indexed publications~~; and (3) ~~Chron-Depth Collection~~Geochronology Data, which includes chronological details (dating methods, ~~age~~ages, and depth points) for each site. The sites span ~~latitudes~~ from 33.2508°~~N~~ to 33.4808°N and ~~longitudes~~ from 126.1486°~~E~~ to 129.2132°E, with elevations from ~~-~~–156~~ m~~ to 1867.5 m. Sediment samples were collected by coring or trenching from six depositional environments: Open~~-coastal zone~~ Coastal Zone, Estuary, Lagoon, River, Volcanic ~~cone~~Cone, and ~~Others~~Other. A total of ~~784~~812 chronological controls ~~(14C, OSL, and U-Th)~~ were analyzed~~ from 72 sediment records~~, and ~~the majority~~most are based on radiocarbon dating. Pollen, diatoms, and grain~~-~~ size ~~analysis, and geochemical markers~~ have been ~~extensively used as paleoenvironmental proxies~~widely applied, with ~~multiproxy analyses becoming~~multi-proxy approaches increasingly common in recent studies. To enhance accessibility, we have developed GeoEcoKorea, an ~~online platform archiving raw data of the compiled studies or linking to it through our metadata, site inventory, and chron-depth datasets if the data is made available elsewhere. This initiative seeks to establish more data sharing agreements with domestic researchers by promoting the collaborative benefits of findable, accessible, interoperable, and reusable (FAIR) data~~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.

Keywords: Data compilation, Chronological control, Regional dataset, Paleoenvironmental proxies, Geospatial data, Paleoecology, Korea, Quaternary

DATASET DETAILS

Identifier: Figshare DOI after publication with embargo
Creator: Soo Hyun Kim
Dataset correspondence: soohyunkim001@gmail.com
Title: A geospatial inventory dataset of study sites in a Korean Quaternary paleoecology database
Publisher: Figshare
Publication year: 2025

---

1. Introduction

~~Paleoecologists~~Quaternary paleoecologists and paleoclimatologists ~~in Quaternary (2.58 Ma to present) research area~~ investigate environmental changes over the ~~past~~ last ~2.5 million years by reconstructing long-term variability in ecosystems, ~~paleoclimate, and paleoenvironmental changes during the current geologic time period, making use~~ climate, and other components of ~~abundant well-preserved~~the Earth system. These reconstructions are based on proxy records from geological archives, particularly sediments in ~~certain~~terrestrial depositional ~~environments. While direct observations~~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 ~~are typically limited to~~ (Bradley, 2015).

Yet no single proxy alone can resolve the ~~past few decades, analysis~~complexity of environmental ~~proxies preserved in~~conditions, particularly when shaped by networks of interacting processes within the Earth system (Birks and Birks, 2006; Mann, 2002). To improve interpretability, Quaternary ~~sediments allows reconstructions over much longer time periods. For example, the past 10,000 years (or late Quaternary) have been the focus of interdisciplinary research aimed at understanding natural and anthropogenic factors and their interactions in the Earth System. Studies using fossil pollen records~~studies use multi-proxy strategies that combine proxies reflecting different environmental drivers acting at varying magnitudes. Some proxies respond to a narrow range of conditions, while others are sensitive to multiple extrinsic (e.g., climate variability and ecological disturbances) and intrinsic factors (proxyspecific sensitivity) and may respond to either their combined effects or a single dominant driver, depending on which prevailed at a given time (Birks and Birks, 2006; Mann et al., 2002). For example, during the Holocene (the recent ~11,700 years), fossil pollen records used to reconstruct past vegetation changes require different ~~assumptions~~interpretive approaches depending on whether the study area is known to have had early, intensive human settlement activity or whether natural vegetation communities can be assumed to have responded primarily to climatic ~~conditions. In this case, additional proxies are useful for defining the paleoenvironmental context of the study area and developing effective research hypotheses based on appropriate assumptions about key environmental attributes.~~changes. In such a case, pairing pollen-based reconstructions with charcoal or archaeological records has often allowed for disentangling anthropogenic land-use impacts from climatic signals at the study site (e.g., Abraham et al., 2023; Kozáková et al., 2015; Yang et al., 2020).

~~There are clear benefits to accumulating paleoenvironmental information by obtaining additional proxy records for the study site or by integrating results from adjacent sites. Processing sediment samples to extract fossil pollen grains, spores, diatom frustules, charcoal fragments, and other micro particles commonly requires careful handling and is time-consuming, not to mention the long hours required to examine thousands of microscopic particles. Additionally, labor-intensive and expensive laboratory procedures are required for analyses of grain size distributions, geochemical compositions, and age dating of the sediments. Although these efforts are not unique to this field, data reuse can be considered beneficial and highly effective in this field. For example, once a complete sedimentary record is established for a particular proxy, it can be readily reused and compared in future studies as new hypotheses or proxies are developed. In this way, future research can benefit from referencing well-established information about study sites of interest, and pioneering researchers can benefit from the long-term legacy effects of their original research data.~~

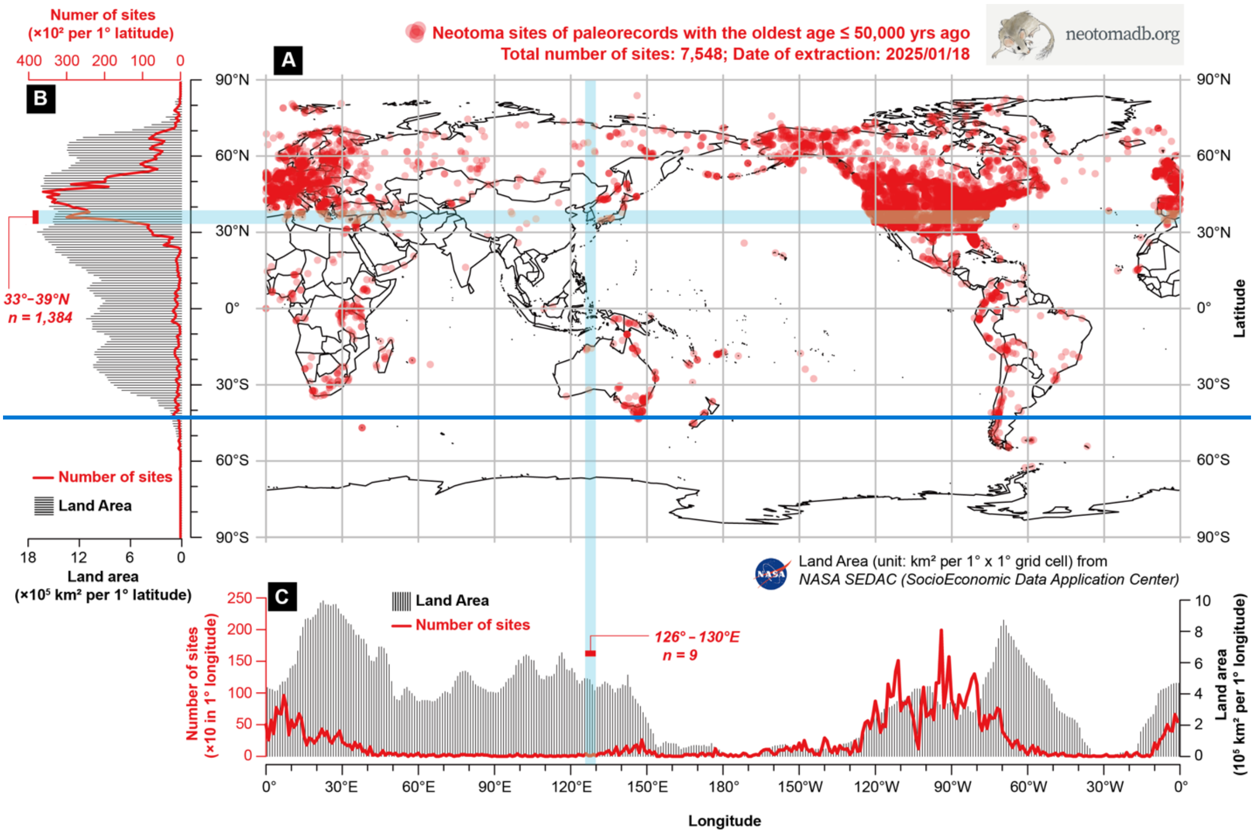
~~Recent advances in large-scale integration of palaeoecological data (e.g., development of 'ecological big-data systems' proposed by Farley et al., 2018; a methodological guide for global pollen data synthesis by Flantua et al., 2023) have provided new insights into global climate change and ecosystem interactions (e.g., Wang et al., 2023; Wang et al., 2020). These advances have been made possible by real-time observations on accelerating climate and landscape changes in the modern era, as well as by a rich set of high-resolution proxy records that allow effective assumptions about environmental factors in the recent geological past. This highlights the importance of systematically preserving and archiving well-recovered proxy data, so that interested researchers can have immediate access to the data.~~In tandem with multi-proxy approaches, multi-site syntheses have led to the large-scale compilation of proxy-based datasets over recent decades. These efforts have underpinned global-scale synthesis projects—for example, reconstructions of Holocene mean surface temperature and land-cover and land-use changes (Gaillard et al., 2010; Kaufman et al., 2020). Broad-scale paleodata syntheses have also enabled the quantification of ecological parameters that are unmeasurable through direct observation—such as biome-specific residence and recovery times and pollen taxon-specific climate fidelity in North America and subcontinental-scale rates of vegetation changes during the past ~20,000 years (Mottl et al., 2021; Wang et al., 2020; Wang et al., 2023). In addition, growing efforts to integrate high-volume paleodata have driven the development of open-source computational tools, such as neotoma2 for acquiring data from open-access repositories, R-Fossilpol for processing pollen datasets into standardized and reproducible formats, and R-Ratepol for estimating the rate of compositional change within palynological assemblages (Flantua et al., 2023; Mottl et al., 2021;Vidaña and Goring, 2023). These advances have been made possible by a rich set of high-resolution proxy records that enable effective analyses of environmental variability in the recent geological past (Farley et al., 2018; Williams et al., 2018).

Such multi-proxy and multi-site approaches underscore the importance of systematically preserving and archiving well-recovered proxy data to enable immediate and broad reuse. For this purpose, the FAIR (findable, accessible, interoperable, and reusable) guiding principles have been proposed as a model for open data sharing across the geosciences (Wilkinson et al~~.~~., 2016).

It is encouraging to see the growing practice of open access data in many scientific communities, and the Neotoma Paleoecology Database ~~(https://www.neotomadb.org/)~~(https://www.neotomadb.org/) is one of the most remarkable contributions to the field (Williams et al., 2018). Research community-based support is available throughout ~~the database use~~Neotoma's suite of services, from ~~securing~~providing servers and curating data uploads to providing a user-friendly interface for global data search and downloads. The Neotoma community also ~~helps~~advises regional ~~databases independently manage~~communities building and governing their own ~~servers~~data while leveraging and ~~affiliate~~affiliating with Neotoma for global networking. With this possibility in mind, this ~~work~~paper is ~~an initiative effort and aims to develop~~the first step towards developing a Korea-based research community database for ~~domestic~~regional Quaternary paleoenvironmental data. The goal is to ~~establish~~build a new data system that supports detailed hypothesis testing at regional ~~context~~to global scales by incorporating geological and ecological characteristics of previously investigated sites while also accounting for societal and cultural factors in conducting research.

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

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



~~Figure 1. (A) Global distribution of sites having paleorecords (age range: ≤ 50,000 years ago) uploaded to the Neotoma database (Total sites: 7,548; Extraction date: 2025/01/18). These paleo-sites 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 below zero.~~ ~~**(B) Number of sites and total land area (10⁵ km²) per 1° latitude.**~~ ~~The total land~~ ~~**area represents the sum of land areas (unit: 10⁵ km²) across all longitudes within each 1° latitudinal band. (C) Number of sites and total land area (unit: 10⁵ km²) per 1° longitude.**~~ ~~The total land area is calculated similarly to (B) but with latitude and longitude exchanged.~~ ~~**The land area estimates in (B) & (C) are derived from 1-degree resolution datasets of the Gridded 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).**~~

2. ~~2~~ Data Collection

We compiled 66 research papers published between 2003 and 2023~~, which~~ that presented proxy data from sediments and surface pollen samples in South Korean territory. The literature search ~~was conducted in the following~~followed four steps:

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

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

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

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

To evaluate the relevance of the compiled articles, we reviewed their abstracts and result sections. For proxy-based records, we selected articles that provided reliable age-dating information, such as the names of Accelerator Mass Spectrometry (AMS) laboratories where dating was conducted. This emphasis on accurate and precise geochronological reporting led to the exclusion of articles published before 2003, resulting in 66 publications.

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

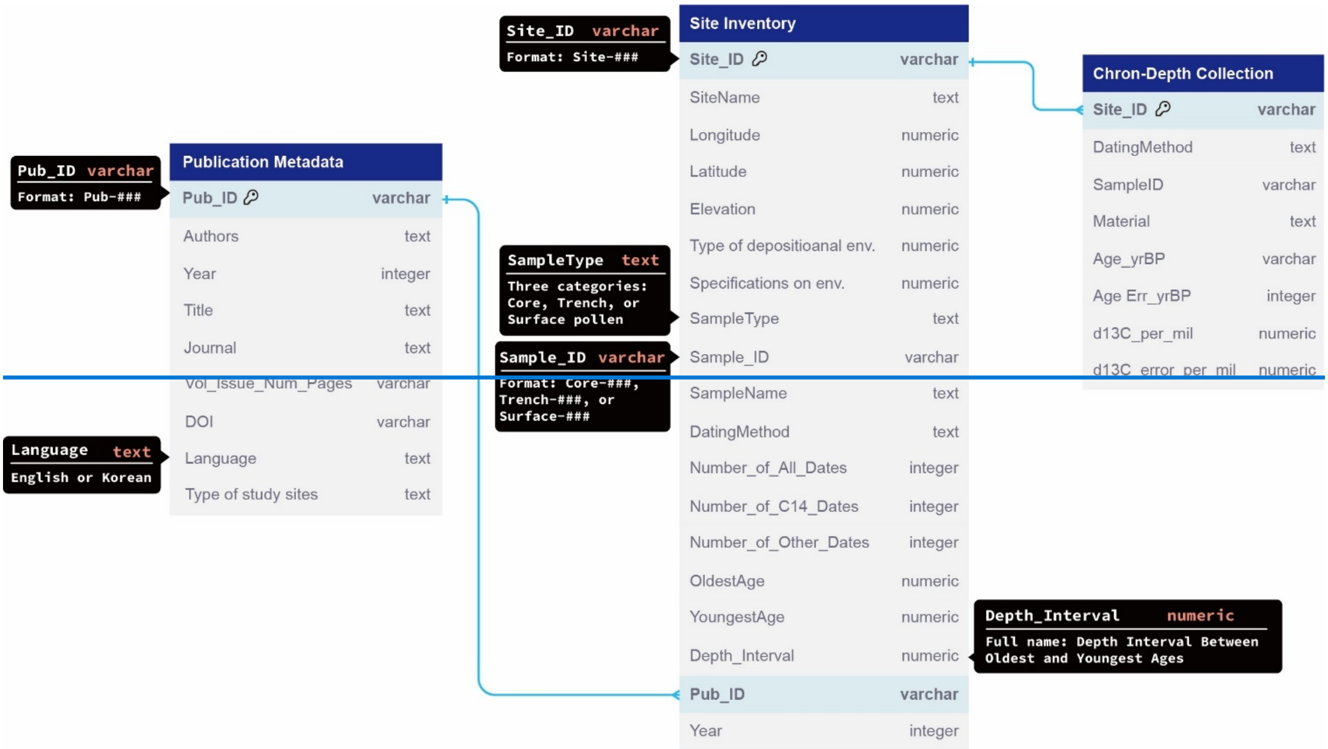


Figure 2. Entity-Relationship Diagram (ERD) illustrating the linking of three datasets in this study. The ERD was created using dbdiagram.io. The detailed description of data fields in the Site Inventory dataset is in Table 1.

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

Section	Field name	Description
Site	SiteID	Unique identifier for the site for each proxy-based records or surface-pollen; formatted as Site-###.
	Site name	Name of the site.
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.
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) Others.
	Specifications on environment	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 sample.
Geochronology	Dating methods	Methods used for dating materials from each core or trench.
	Total number of all dates	Number of all dates obtained from each core or trench.

	Number of ^{14}C dates	Number of radiocarbon dates.
	Number of dates from other methods	Number of dates from dating methods other than radiocarbon.
	Oldest age	Oldest age obtained from dating (unit: year before present).
	Youngest age	Youngest age obtained from dating (unit: years before present).
	Depth interval (cm) between oldest and youngest ages	Interval (cm) between depth points of oldest and youngest ages.

3. DATA PROCESSING AND DESCRIPTION

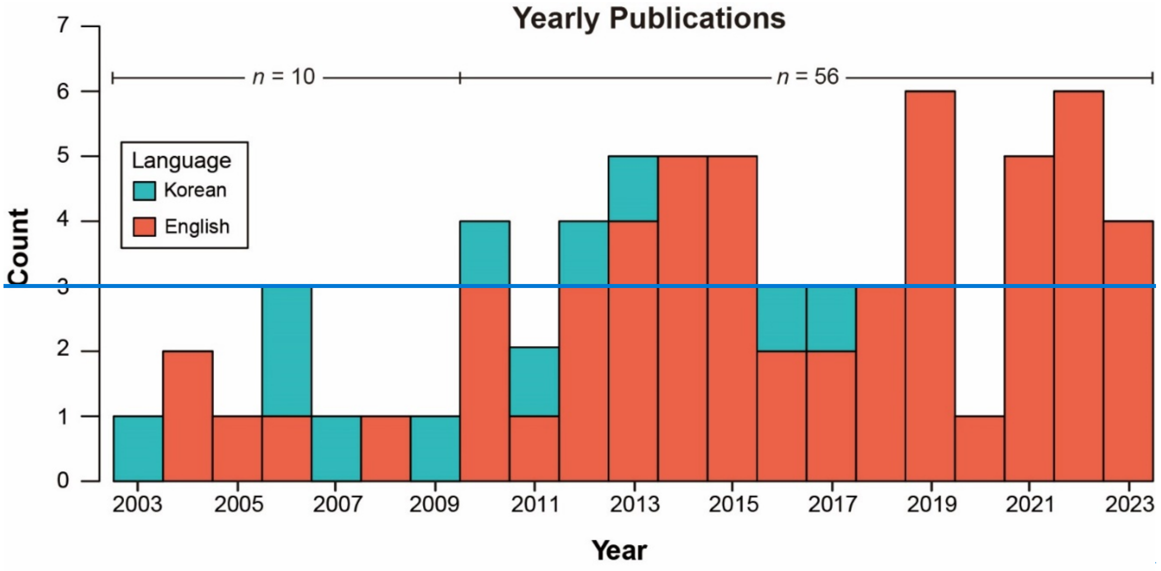
Proxy	Pollen	Indicates whether pollen analysis data are available (Yes/NA).
	Diatom	Indicates whether diatom analysis data are available (Yes/NA).
	Grain-size	Indicates whether granulometric data are available (Yes/NA).
	Other proxies	Lists additional proxy types available (or NA if none are available).
Publication	PublicationID	Identifier for the publication, formatted as Pub-###, linking to PublicationID in the meta dataset.
	Year	Year of publication.
	Written language	Language in which the publication is written.

3 Data Processing and Description

3.1. 3.1 Dataset I: Publication Metadata

The Publication Metadata dataset contains 66 papers published in peer-reviewed scientific journals in the recent 20 years (2003–2023). Extracted metadata includeinformation about authors, publication year, title, journal, numbers of volume, issue, and pages, DOI, and written language (either English or Korean), and type of study site.). The annual number of papers has shown a sharp increasing trendincreased rapidly since 2010, driven by the growing number of English-written papers published in international journals (Fig.Figure 3).

3.2. Dataset II: Site Inventory



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

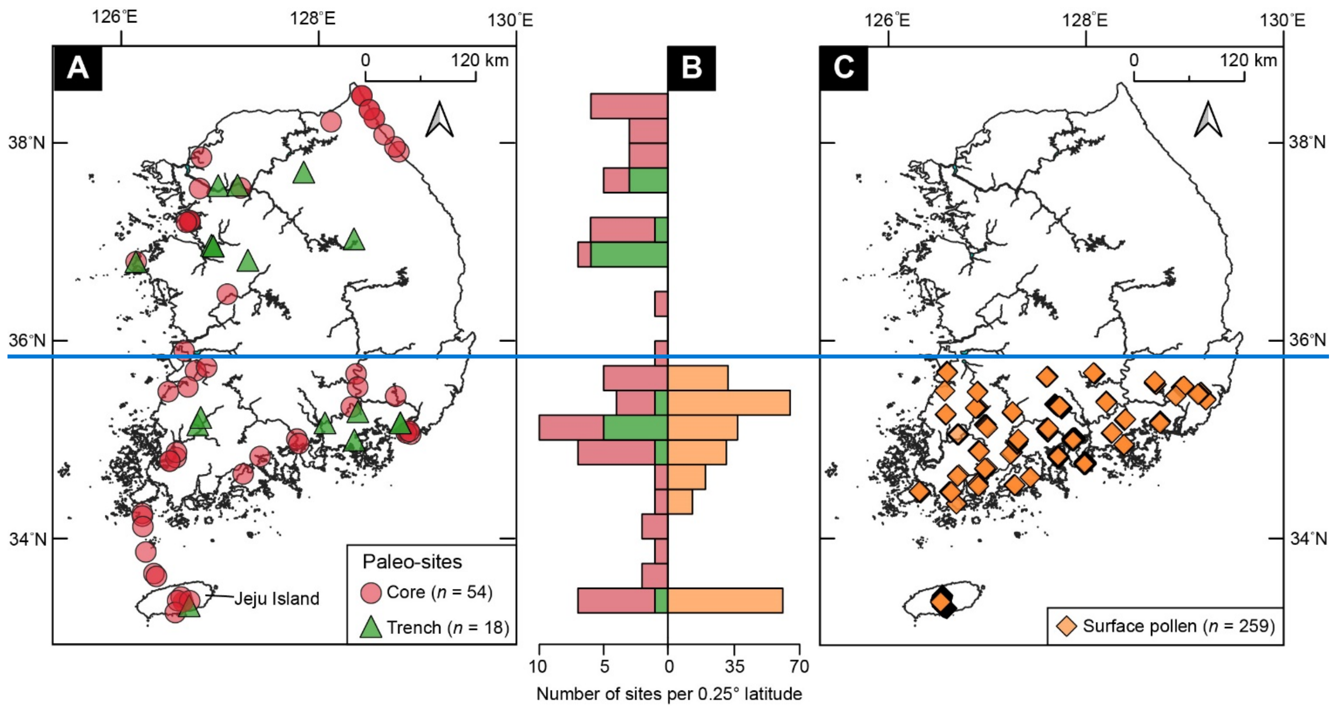
3. ~~**Barplot of annual publications between 2003 and 2023.**~~

~~Regarding type~~*2.1. Type of study site*

The Site Inventory dataset comprises 328 sites~~, the publications are broadly categorized into two groups based on the primary goals of the sampling: paleo-site and surface site. 64 publications focused on~~ that consist of 72 paleo-sites ~~where stratigraphic sediments were collected~~with sedimentary records and 256 surface sites with surface pollen samples (Figure 4). The paleo-sites were derived from 64 publications that presented paleorecords reconstructed from sediment sequences to investigate the temporal sequence of the targeted proxy and past ecosystem elements ~~to reconstruct~~ (e.g., fossil pollen and past vegetation-climate interactions~~), while 2~~). The surface sites were presented in two articles ~~covered~~that investigated modern surface ~~sites~~samples that provide spatial information on the modern ecosystems, serving as the basis for the space-for-time substitution method (e.g., surface pollen and modern vegetation-climate associations~~).~~) (Chevalier et al., 2020).

~~**1.1.**~~ 3.2 ~~Dataset II: Site Inventory~~

~~The Site Inventory dataset includes a total of 328 sites that consist of 72 paleo-sites with sedimentary records and 256 sites with surface pollen samples (Fig. 4).~~



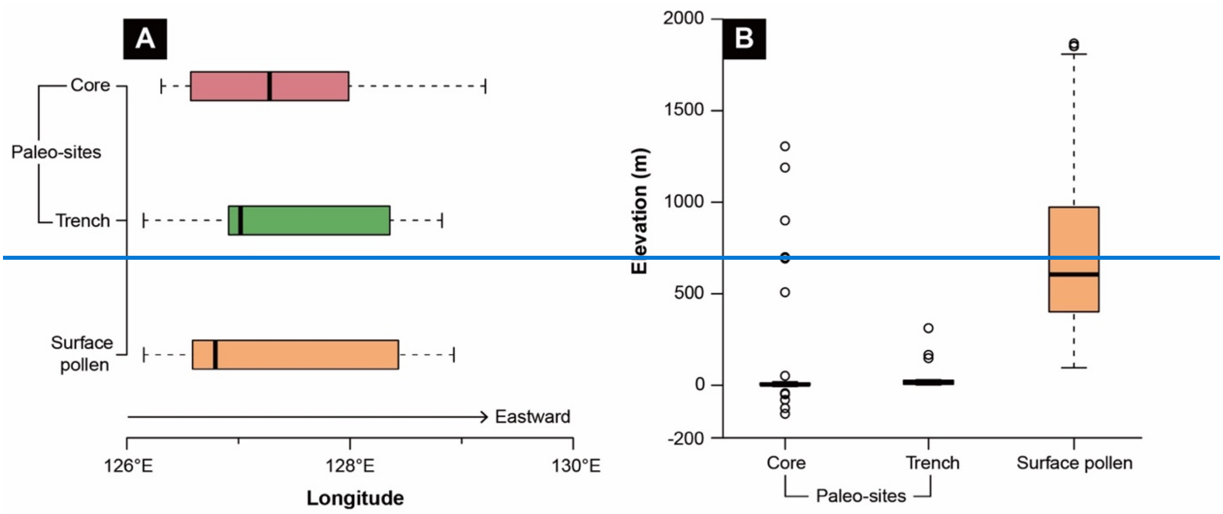
~~Figure 4. Geographic distributions of sites. (A) Paleo-sites. (B) Surface pollen sites.~~

~~3~~.2.~~1~~ Geospatial coverage

~~Sites with distinct~~Each site is georeferenced using geographic coordinates ~~were treated as separate entries.~~and elevation. We ~~prioritized~~used coordinates and elevations reported in the papers when available. If not, we estimated ~~them~~these data using site maps from the articles, cross-referencing with Google Earth. Site locations were recorded in decimal degrees (to four-digit ~~decimal degree coordinates~~precision) with elevations in meters~~.~~ above mean sea level (unit: m).

The distribution of paleo-sites and surface sites shows no significant variation in longitude but distinct patterns in elevation (~~Fig.~~Figure 5) and ~~in~~ latitude (~~Fig.~~Figure 6). In the geographic extent (33.2508° ~~–~~to 38.4808°N, 126.1486° ~~–~~to 128.9719°E) of paleo-sites, their elevations (range: ~~-~~–156 ~~–~~to 1,305 m; median: 5 m) are consistent with the tendency of depositional environments for stratigraphic sediments typically found at low elevations (~~Fig.~~Figure 5A), although their latitudinal distribution exhibits a broad coverage across South Korea except for the central eastern region (~~Fig.~~Figure 6A). In contrast, the surface pollen sites (33.2916° ~~–~~to 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 the paleo-sites (~~Fig.~~Figure 5B) but are concentrated in the southern regions (~~Fig.~~Figure 6B). These surface pollen sites were primarily contributed by two pioneering studies: Park and Park (2015), who investigated 39 elevational transects along mountain slopes in Jeju Island to reflect temperature gradients, and Lee et al. (2022), who studied 37 transects in the southern part of the Korean Peninsula for modern pollen-climate references (~~Fig.~~Figure 6B). It is interesting to note the high elevations of Jeju's volcanic-cone wetland sites (~~Fig.~~Figure 7), which locally shift the statistical distribution of the island paleo-sites (range: 51~~-~~ to 1,305 m; median: 692

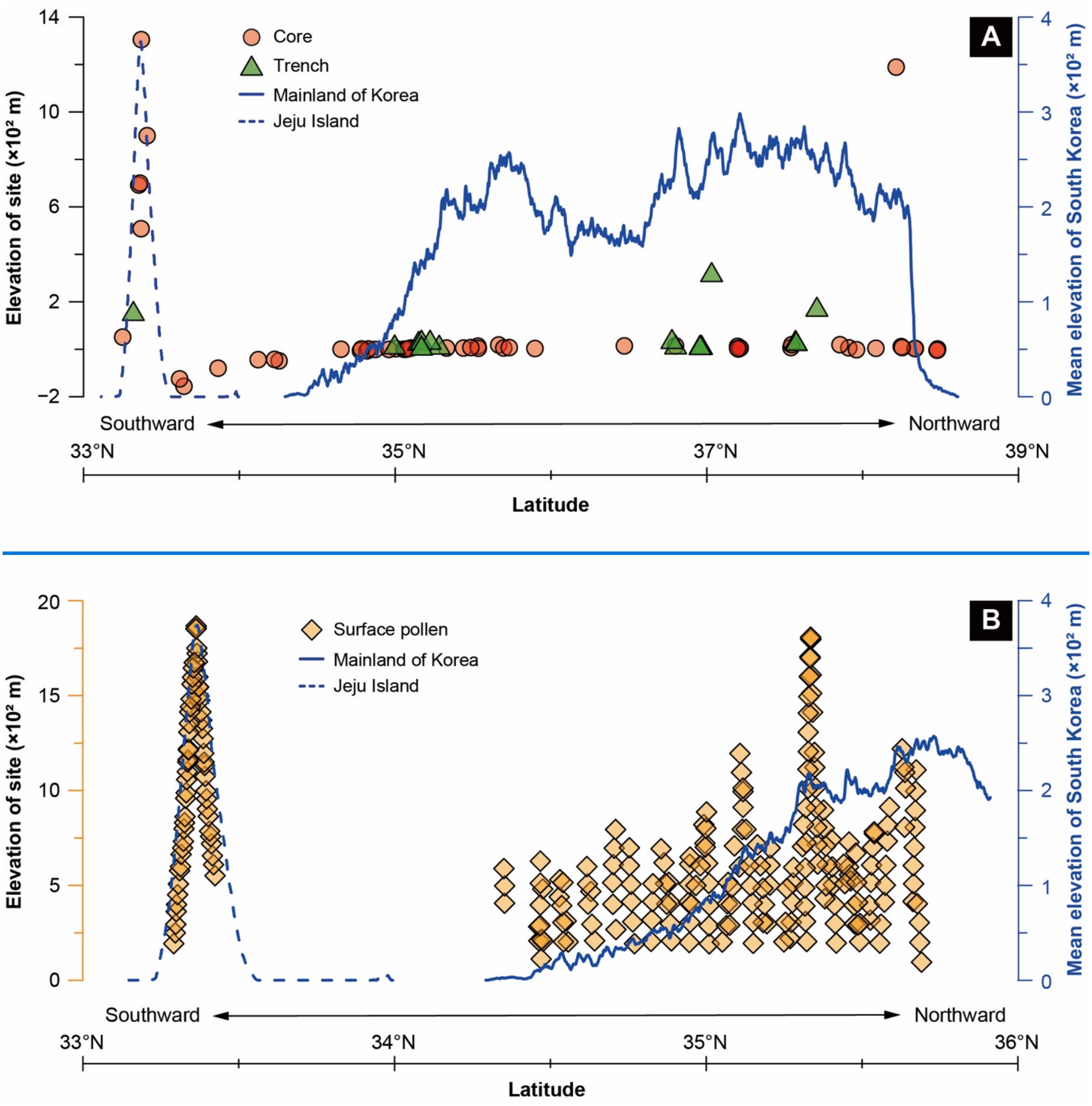
m~~).~~)



~~Figure 5. Boxplots of site locations and elevations. (A) Longitudinal distribution of sites. **(B) Vertical distribution of sites.**~~

3.2.~~2~~3. Sample type

Three sample types (Surface Pollen, Core, and Trench) for the Site Inventory dataset represent the study approaches commonly used in South Korean Quaternary paleo-reconstruction research (Kim et al., 2024). Surface pollen samples were systematically targeted and obtained from the soil surface on mountain slopes in Jeju Island and the southern region of the mainland (Lee et al., 2022; Park and Park, 2015). For sediment samples, either coring or trenching was employed to obtain the sediments for various proxy ~~analysis~~analyses, preserving the original stratigraphy of the layered, aged sediments. Coring generally involves inserting empty tubes into sediments to extract vertically long and continuous sediments embedded in the tubes~~, while trenching~~. Trenching involves collecting subsamples directly from stratigraphic sequences of exposed outcrops. Since trenching was unsuitable for submerged sites, coring was used more frequently at paleo-sites than trenching (~~Fig.~~Figure 4).



~~Figure 6. Elevational distribution of sites. (A) Paleo-sites. (B) Surface pollen sites. **These two** 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 **3A.**~~

3.2.~~3~~4. Depositional setting

The depositional environments of paleo-sites are categorized into six groups: Open~~-coastal zone~~ Coastal Zone, Estuary, Lagoon, River, Volcanic ~~cone~~Cone, and ~~Others~~Other (Table 2). The ~~"~~Open~~-coastal zone"~~ Coastal Zone encompasses a set of environments dominantly influenced by oceanic depositional processes, such as ~~beach~~beaches, nearshore, and continental shelves. ~~"~~Estuary~~"~~ and ~~"~~Lagoon~~"~~ are classified separately based on energy conditions within the transitional zone between terrestrial and oceanic sedimentation, with ~~"~~Estuary~~"~~ characterized by high-energy conditions and

"Lagoon" by low-energy conditions. The "River" category includes depositional environments such as floodplains, backswamps, and other fluvial ~~setting~~settings associated with rivers and streams. The "Volcanic ~~cone"~~Cone refers to ~~the~~ wetland systems such as lakes, bogs, or swamps in the craters of cinder cones scattered across Jeju Island, the largest island of South Korea, located to the south of the mainland. The ~~"Others"~~Other comprises sites with depositional settings that do not fit within the other five groups, and because surface pollen samples were all from the montane soil surface, they were also classified under the ~~"Others"~~Other category. The paleo-sites can be broadly categorized into two groups, coastal and terrestrial sites. The coastal sites (n = 41), including Open~~-coastal zone~~ Coastal Zone, Estuary, and Lagoon, outnumber non-coastal ~~(terrestrial)~~ sites (n = 31~~), including~~). River~~,~~ is a more commonly sampled site type than the other non-coastal categories— Volcanic ~~cone,~~Cone and ~~Others, with River being the popular sampling sites (Fig.~~Other (Figure 7B).

~~**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~~
	~~Others~~	~~7~~
~~Proxy~~	~~Pollen~~	~~42~~
	~~Diatom~~	~~14~~
	~~Grain-size~~	~~45~~
	~~Other proxies~~	~~42~~

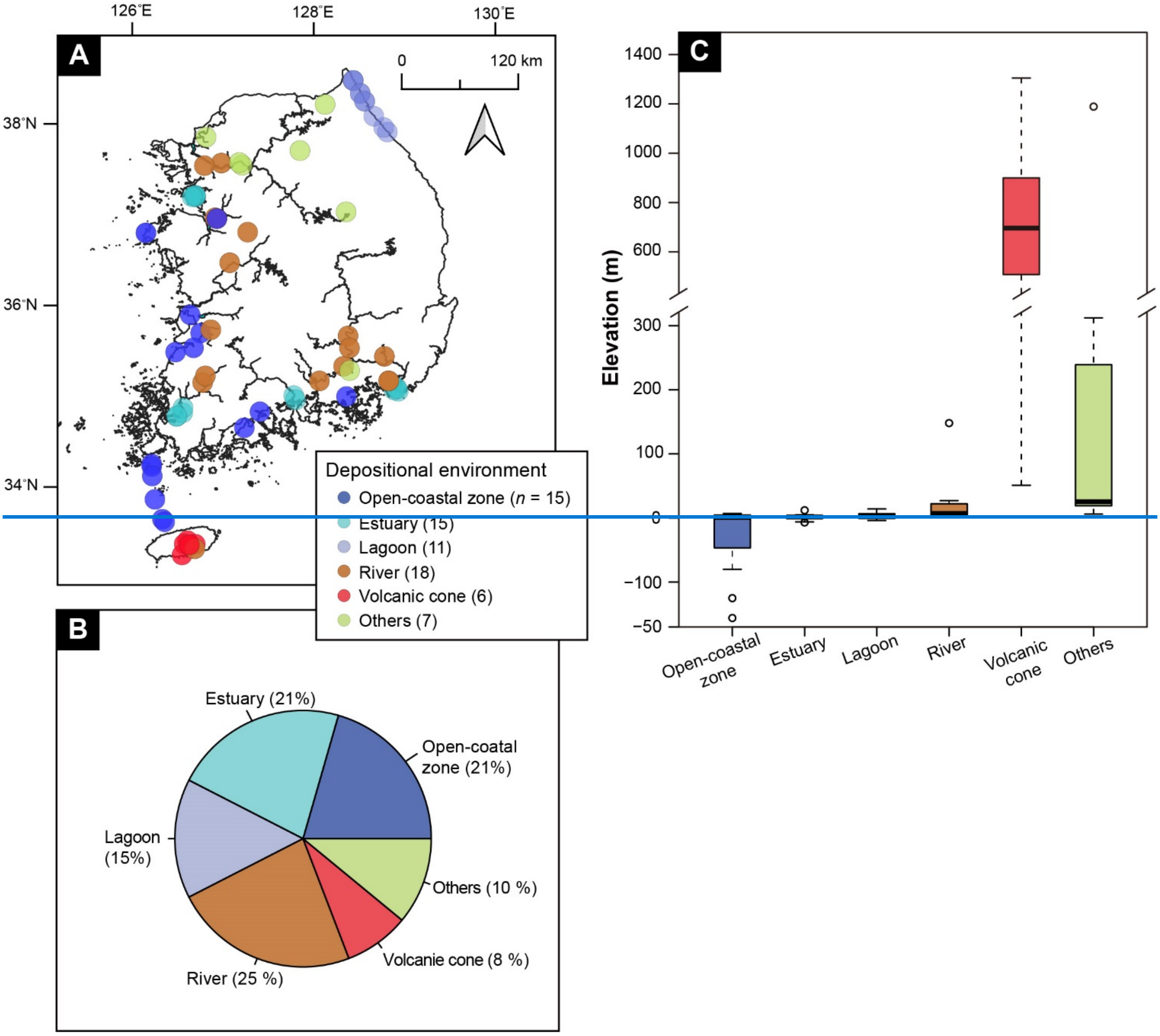
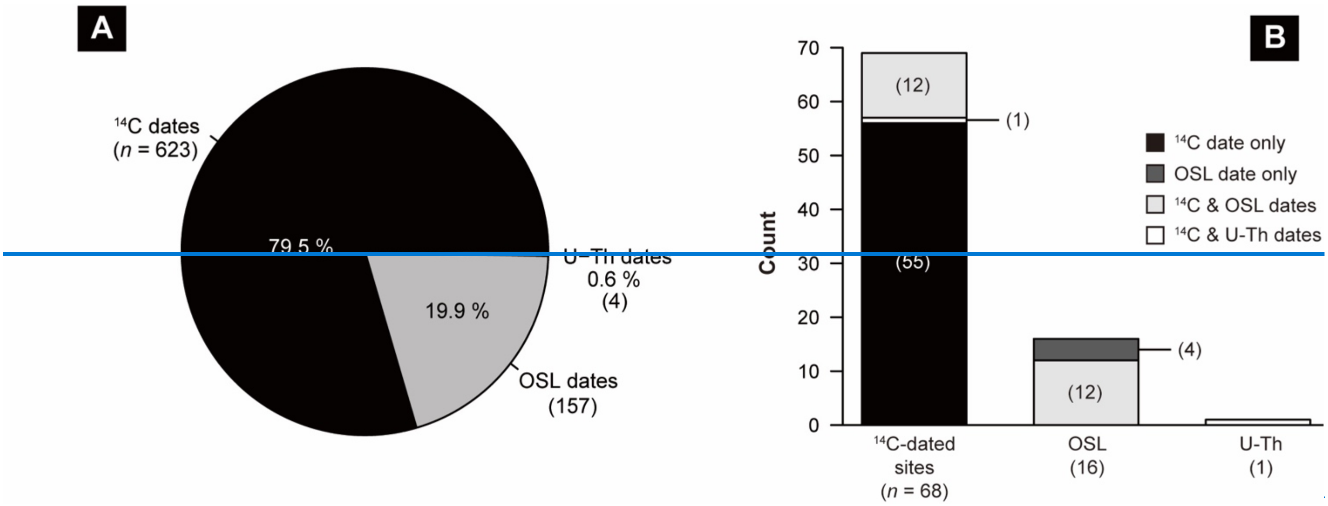


Figure 7. Paleo-sites (n = 72) categorized by six types of depositional environments. (A) Geographic locations of sites. (B) Proportion of sites by depositional environment. (C) Boxplots of elevations. **The median elevations for each environmental type are as follows: Open-coastal zone (1 m), Estuary (1.5 m), Lagoon (3 m), River (10 m), Volcanic cone (696 m), and Others (21 m).**

3.2.45. Geochronology

We compiled chronological data in the following seven categories: Dating methods, Total number of all dates, Number of radiocarbon dates, Number of dates from other methods, Oldest age, Youngest age, and Depth interval between oldest and youngest ages. For surface pollen samples, chronological dating was unnecessary, therefore, the field of "Dating methods" is filled with "NA," and all date numbers and depth intervals are entered as zero. The oldest and youngest ages are converted to radiocarbon years based on the sampling year. If the sampling year was not specified in the article, the publication year was used instead. For example, if the year of collection or publication is 2019, the oldest and youngest ages are both documented as -69 14C years BP. For paleo-sites, a total of 784 dates are obtained using three methods, radiocarbon dating, optically stimulated luminescence (OSL), and U-Th dating (Fig. 8A). Most paleo-sites (96 %) rely solely on radiocarbon dating for chronological control, although some sites incorporate OSL or U-Th dating (Fig. 8B).



~~**Figure 8. (A) Pie chart of absolute dates by dating methods, and (B) Bar plot of paleo-sites by dating methods.**~~

~~In "Oldest age" and "Youngest age" fields, the two end ages used for age-depth modeling in publications are recorded when they are explicitly stated or identifiable in the age-depth models. When this relevant information is unclearly provided, the oldest age is considered the maximum number and the youngest age the minimum number (only if the age is ≥ -73 years BP). The difference between the oldest and youngest ages represents the temporal extent of each site. Seventy percent of paleo-sites record only the Holocene, about 24 % extend through the last glacial-interglacial cycle, and 6 % are limited to the Pleistocene (Fig. 9A and 9B). The Holocene-only sites show greater variability in depositional settings compared to other sites containing older records (Fig. 9C).~~

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

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

Oldest and Youngest ages are reported in calendar years before present (BP), using 1950 CE as the reference point for "Present." For paleo-sites, these two endpoints correspond to ages of the youngest and oldest age control points used in age-depth models in the original studies. This age selection is made not based on depth, as age reversals frequently occur within sediment sequences, obscuring reliable geochronological boundaries of paleorecords. If either age is based on a radiocarbon date, the lower (Oldest) or upper (Youngest) bound of the 95% calibrated range is documented, using the

IntCal20 or Marine20 calibration curves through the IntCal R package (Blaauw et al., 2022; Heaton et al., 2020; Reimer et al., 2020). When the core top age reflects the youngest age, it is calculated by subtracting the year of coring or trenching from 1950 CE. If the collection year is unspecified, a value of 0 is assigned for the youngest, following the same approach used for post-bomb ^{14}C dates. For surface samples, the oldest and youngest ages are derived by subtracting the sampling year from 1950 CE. The publication year is used as the sampling year when it is unavailable in the original article. For OSL and U-Th dates, the lower (Oldest) and upper (Youngest) bounds of the 1σ error ranges are also selected. Neither OSL nor U-Th ages were calibrated. The U-Th ages were not normalized, as their reference year is defined as 1950 CE (Dutton et al., 2017). OSL age limits, initially expressed as calendar years before the measurement year, are converted to calendar years before present by assigning the measurement year as the sampling year, or the earliest publication year if unavailable. The calibration methods, normalization approaches, and dating sources for Oldest and Youngest ages are provided in the Age scale field.

The difference between the oldest and youngest ages represents the temporal extent of each site. ~67% of paleo-sites (*n* = 48) record only the Holocene, ~32% span from the Pleistocene to the Holocene (*n* = 23), and ~1% contain records exclusively from the Pleistocene (*n* = 1) (Figure 9A and 9B). The Holocene-only and the Pleistocene–Holocene sites are broadly distributed across five types of depositional settings, although the Lagoon contains no Pleistocene record (Figure 9C).

In addition, ~~our~~the chronological data are ~~useful for deriving the~~ used to calculate two types of dating density~~, which indicates~~: the number of ~~dates~~age controls per time interval (Flantua et al., 2015; Lacourse and Gajewski, 2020~~).~~) and per depth interval. The temporal dating density ~~(represented~~is defined as the ~~inverse of the age control density suggested in Flantua et al., 2015) is calculated by dividing the~~ total number of ~~dates~~age controls divided by the ~~gap~~time span between oldest and youngest ages~~. This metric is used to~~, and the spatial dating density is the total number divided by the total depth range. These metrics help assess the ~~accuracy~~temporal and depth-spatial resolutions of age-depth models ~~by the extent of age coverage~~ (Lacourse and Gajewski, 2020). ~~Although the number of dates per depth interval overlaps in their interquartile, with similar medians (Fig. 9D), most Holocene-only sites have a higher~~ Both types of dating density ~~than~~are higher in Holocene-only sites compared to the other two sites covering the Pleistocene~~-only~~ (Figure 9D and ~~Last glacial-interglacial cycle sites (Fig. 9E).~~ E).

3.2.~~5~~6. Proxy

~~The proxy-related data are~~ Information about proxies available at each site is organized in four data fields: ~~"~~Pollen,~~" "~~, Diatom,~~" "~~, Grain~~-~~ size,~~"~~, and ~~"~~Other proxies~~."~~. The first three fields indicate whether each site contains a proxy related to pollen and diatom assemblages and grain-size distribution parameters (e.g., mean, median, or sorting and relative proportions of sand, silt, or clay), marking the presence of each proxy as "Yes" and its absence as "NA." These three types of proxies are documented ~~separately as~~because they are the most frequently investigated proxies, with some sites ~~even~~ showing mono-proxy reconstructions that rely on only one of them (~~Fig.~~Figure 10B).

In the field of “Other proxies~~,~~”, a total of 79 proxy types are listed, with abbreviations when necessary. The explanation of each abbreviation is documented in a supplementary worksheet (see Section 4~~. Data access for further information on accessibility.).~~). In the list of the “Other proxies~~,~~”, organic geochemical indicators (e.g., C/N ratios and δ^{13}C) are more commonly ~~used~~available than inorganic proxies such as mineralogical compositions, trace element content, and magnetic susceptibility.

Overall, for surface samples, our dataset includes only pollen~~ as the primary indicator~~, which is widely used to construct taxon-based transfer functions for paleotemperatures estimates (Lee et al., 2022; Park and Park, 2015). While non-palynological organisms, such as ostracods from brine and freshwater swamps and subterranean caves, have been studied in South Korea (McKenzie, 1972; Smith et al., 2015) (~~Fig.~~Figure 1), they were not included in our datasets because those studies focused on taxonomical descriptions rather than quantitative ~~analyses~~reconstructions of past environments such as those produced by the modern analog techniques. For paleo-sites, approximately 79~~ ~~% sites have multi-proxy records (~~Fig.~~Figure 10A) ~~that~~and these sites are more widely distributed than mono-proxy records. (~~Fig.~~Figure 10C). Grain-size and pollen proxies are more frequently used than other proxies, with pollen being the most utilized in mono-proxy sites (~~Fig.~~Figure 10B). The extensive spatial coverage ~~of pollen with grain-size records suggest~~suggests that multi-proxy approaches have ~~been considered useful~~become standard practice in the Quaternary paleoenvironmental research in South Korea (~~Fig.~~Figure 10D).

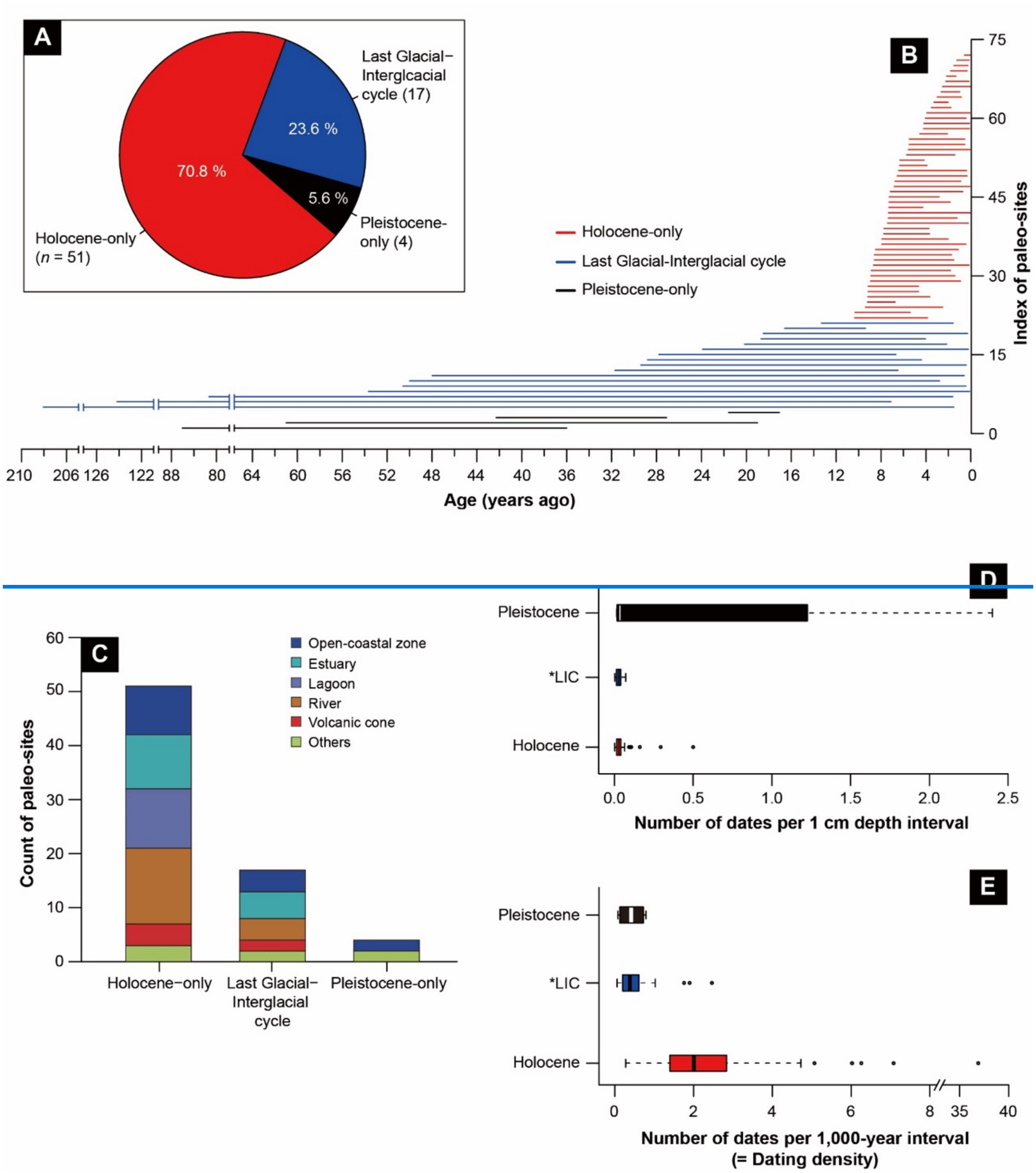


Figure 9. **(A) Distribution of paleo-sites by three groups of age-coverage: Holocene-only, Last Glacial-Interglacial cycle (LIC), and Pleistocene-only.** (B) Chronological extent of records from paleo-sites, ordered by descending oldest ages within each age-coverage group. (C) Counts of depositional environments by age-coverage group. **(D) Boxplot of the number of dates per 1 cm depth interval. Medians of Holocene-only sites: 0.03, LIC sites: 0.03, and Pleistocene-only sites: 0.04. (E) Boxplot of the dating density per 1,000-year interval. Medians of Holocene-only: 2.01, LIC: 0.38, and Pleistocene-only: 0.41.**

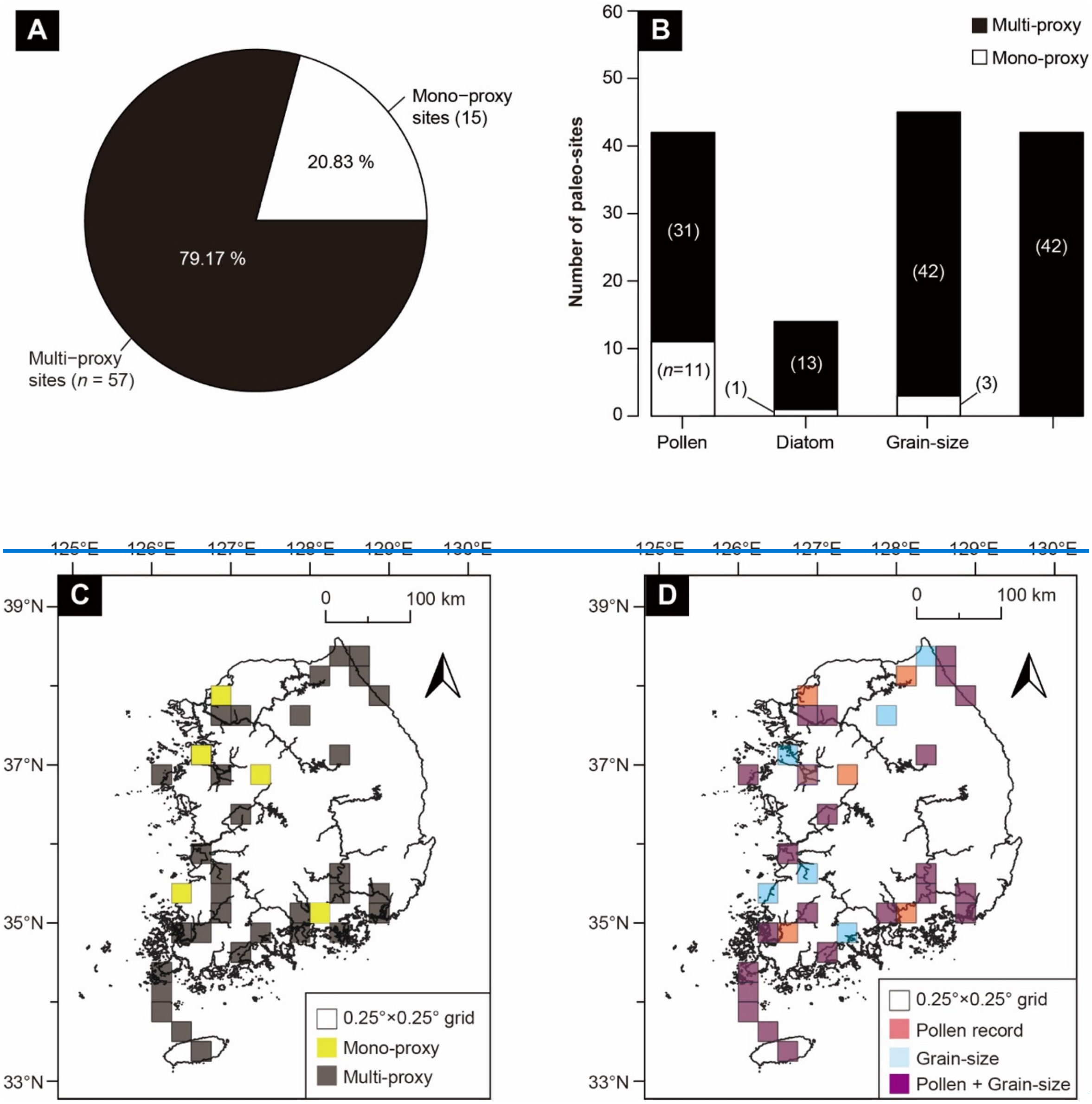
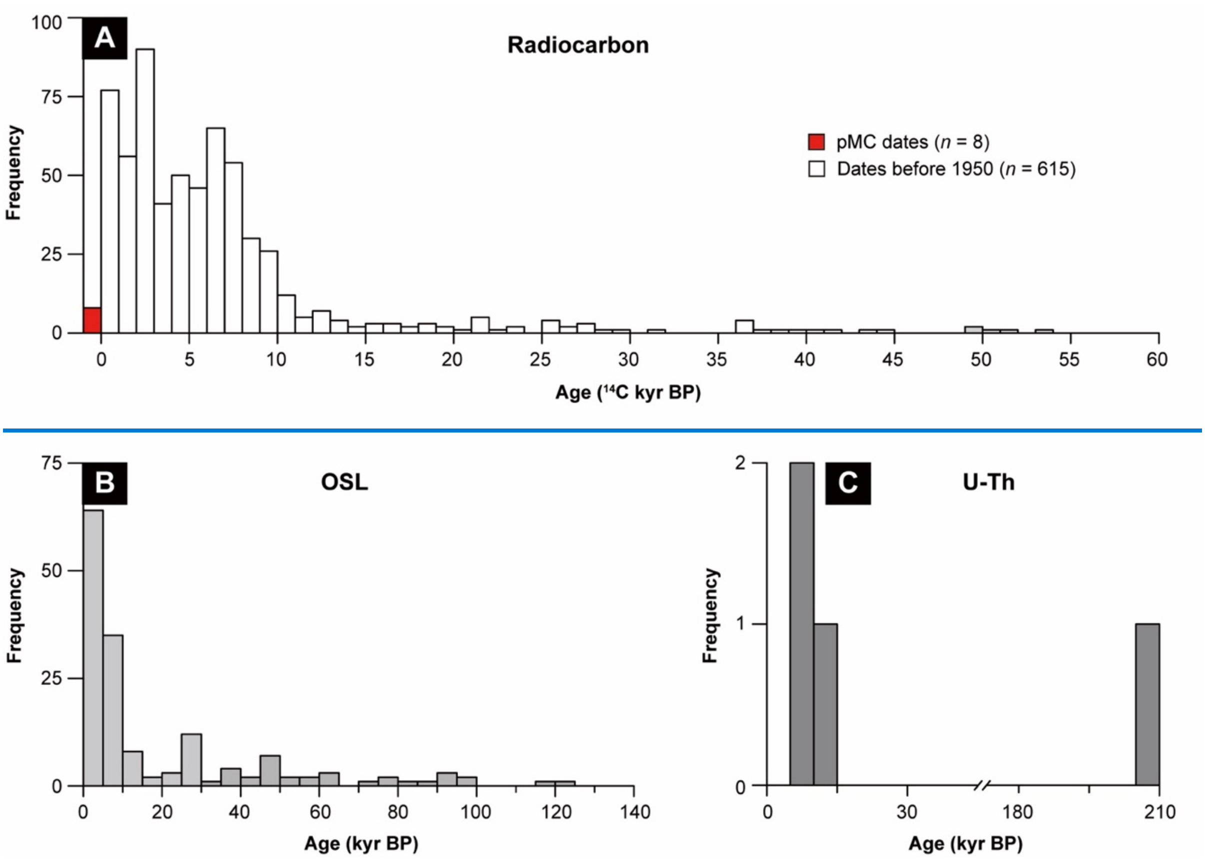


Figure 10. (A) Proportion of mono-proxy and multi-proxy sites. (B) Number of paleo-sites by four proxy types. (C & D) Geospatial distribution of proxy-based records **in 0.25° × 0.25° grids.** These maps display grids containing proxy-based records from paleo-sites in South Korea. In (C), grids are classified as either **only mono-proxy or multi-proxy. (D) shows grids containing records of pollen and grain-size, the most widely used proxies in South Korea.**

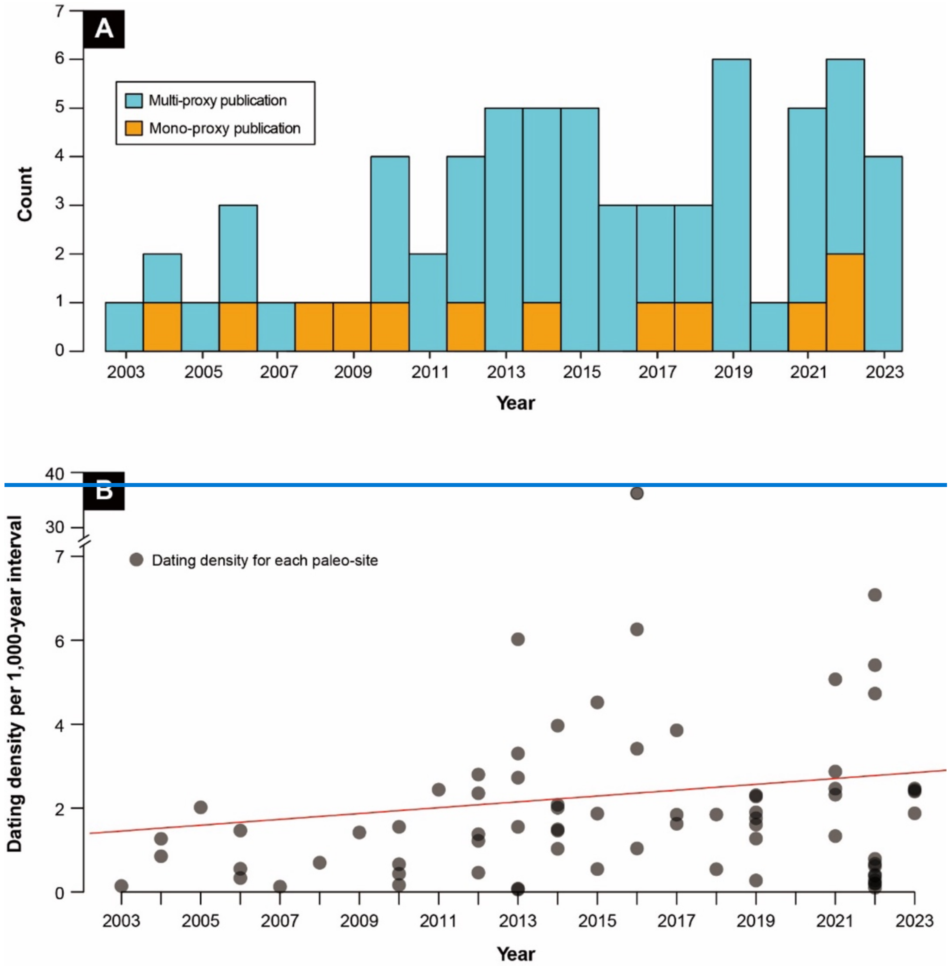
3.3. 3.3. Dataset III: Chron-Depth CollectionGeochronology Data

The chronological dataset corresponding to 72 paleo-sites areis documented separately. This dataset consists of nine categories: SiteID, dating method (Non-absolute dating: Core top or estimated; Absolute dating: C-14 AMS, OSL, or U-Th dating), sample ID, dating material, mean and 1σ-error of age (years before present), depth (cm), and mean and 1σ-error of $\delta^{13}C$ (‰). For the sample ID, AMS Lab IDs are prioritized particularly for (‰), and Chronological age type. In these fields, most chronological data from the original publications are documented without calibration, except in two standardized cases: core top ages are converted from AD to calendar years BP using 1950 CE as the reference point, and post-bomb radiocarbon dates~~; but when unavailable, identification indices~~ are expressed with "pMC." The field of Chronological age type indicates timescale in which the date is reported, including uncalibrated ^{14}C years before present (1950 CE). Further details on other data fields are provided in the ~~source articles are recorded. The total number of age controls per paleo-site ranges from 2 to 51 with a median of 8 dates (Fig. 11).~~ Supplementary Document as well as the Column description worksheet within the dataset file (see Section 4.).

Most absolute dates ~~are concentrated~~fall within the Holocene ~~epoch (≤11,700 years BP), accounting for 90 % of 14C dates, 68% of~~ (Figure 11). OSL dates~~, and 50% of~~ often constrain the ages of sediment sequences older than ~55,000 years BP, the upper limit of ^{14}C dating (Figure 11B). Although U-Th ~~dates.~~dating has been applied to sediments exceeding the maximum range of OSL ages, its usage for terrestrial sediments in South Korea remains limited (Figure 11C).



~~Figure 11. Histograms of dates over ages. (A) Radiocarbon dates (range of pre-1950 dates: >53,671–27 ^{14}C years BP; median: 4790 ^{14}C years BP). (B) OSL dates (range: 124,400–140 years BP; median: 7,114 years BP). (C) U-Th dates (range: 208,000–6900 years BP; median: 10,000 years BP).~~



~~Figure 12. Temporal trends in (A) the proportion of multi-proxy to mono-proxy publications and (B) dating density per 1,000-year interval. In (A), 64 publications are shown as two articles studying surface pollen samples are excluded. In (B), the dating density, defined as the total number of dates 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 the same site, the earlier one was selected. The red line represents a linear regression (y = 0.07x – 138).~~

4. ~~4 Data~~Dataset Availability

Our three datasets are accessible on two platforms: Figshare and the GeoEcoKorea website. Figshare hosts tabular datasets in .xlsx format at https://doi.org/10.6084/m9.figshare.28236596 ~~(Kim and Byun, 2025).~~. The first dataset file, **Publication Metadata.xlsx**, is composed of ~~a single worksheet~~two worksheets: a) Publication dataset and b) Column description. The second file, **Site Inventory.xlsx** contains three worksheets: a) Inventory dataset, b) Column description (identical to Table 1), and c) OtherProxy_Abbreviation. The third file, ~~Chron-Depth Collection~~**Geochronology Data**.xlsx has two worksheets, a) ~~Chron-Depth dataset~~Geochronological data and b) Column description.~~ ~~

Concurrently, the GeoEcoKorea webpage presents site locations on an interactive map at https://geoecokorea.org/2025_Site_Inventory. The bibliographic and site inventory datasets are integrated, converted to GeoJSON format, and visualized on OpenStreetMap-based interactive maps customized with the "Leaflet" JavaScript library. Each site is ~~georeferenced, featuring~~represented by an interactive ~~markers with~~ marker that opens a pop-up window displaying site-specific metadata~~ and a~~, along with direct ~~download link~~links to the corresponding ~~Chron-~~

bibliographic information and a downloadable **Geochronology** **Data** ~~Collection dataset (two~~(three worksheets: Readme, Geochronological data, and ~~Chron-Depth~~Column description) in .xlsx format.

5. ~~5~~ Summary and Dataset Reuse Potential

Our datasets provide a comprehensive overview of South Korean Quaternary paleoecological data published in peer-reviewed journals from 2003 to 2023. The datasets highlight the ~~relatively~~ high spatial density of paleorecords available from ~~the country, which can be especially useful for future applied studies aimed at reconstructing paleoclimate changes at high resolution or testing new hypotheses regarding local ecosystem turnover~~various depositional environments, including wetlands, coastal deposits, lagoons, estuaries, rivers, and volcanic crater wetlands in ~~Northeast Asia where long-term environmental data are not widely accessible.~~Jeju Island. While ~~most~~~67% of records ~~(~70 %)~~ are covered only the Holocene, ~~~30 % records~~the remaining 33% extend ~~to~~into the ~~late~~ Pleistocene. The quality of these records has improved over the past decades, particularly with the increasing dominance of multiproxy approaches and improved chronological constraints (~~Fig.~~Figure 12), suggesting continued progress in the future. ~~The compiled sites reflect the regional diversity of depositional environments, including wetlands, coastal deposits, lagoons, estuaries, rivers, and volcanic crater wetlands in Jeju Island. The age dating results provide a useful~~

Our site inventory offers practical reference ~~data~~ for future ~~study site selections.~~ site selection. Korea's compact geography and clearly differentiated depositional settings (Figure 7A) allow for temporal variability of sediment accumulation rates—defined as "deposition time" in similar studies (e.g., Crann et al., 2015; Goring et al., 2012)—across the settings without the profound effects of large-scale continental gradients. In addition, when the Korean paleorecords in the inventory are integrated, they can provide foundational datasets for testing new hypotheses regarding the drivers of local ecosystem turnover in Northeast Asia. The records span both human-cultivated lowlands and relatively undisturbed upland settings (Figure 6A), enabling comparisons that help separate climate- from human-induced impacts on past landscape changes during the Holocene. The interpretability of past vegetation dynamics is enhanced by extensive surface pollen samples collected across elevation gradients, which serve as modern analogues of climate–vegetation relationships (Figure 6B). These features can contribute to understanding regional-scale ecosystem dynamics in response to climate and anthropogenic forcings.

Although the site-specific findings can be explored through the main text of related publications in international peer-reviewed journals, the '~~interoperability~~interoperability' and 'reusability' of original data (and thus citation of the paper) has been relatively rare in international collaborative research. ~~For example, the most~~A recent ~~release of the EMPD (European Modern Pollen Database) did not include any records from Korea (Davis et al., 2020), so the next version may consider inviting the authors of the surface~~global pollen ~~sites compiled in this dataset. Similarly, a recent~~ data compilation study for ~~pollen-based~~the Northern Hemisphere ~~paleoclimate reconstruction~~, for example, included only a single dataset from Korea (Herzschuh et al., 2023~~), despite the potential for additional data to be explored through our dataset. Additionally, the~~). Our dataset framework ~~developed for our datasets served~~can serve as ~~a useful tool for systematically characterizing site information and sample metadata~~not only to characterize the inventory of paleo-studies in South Korea, and following the descriptions in this paper, future efforts can readily adapt the framework for but also a template for organizing similar inventories in other underrepresented regions. Moving forward, the relational database, integrated with the open-access web platform, will enhance the 'findability' and 'accessibility' of the original data, contributing to the FAIR data sharing and the advance of this research field, regionally and globally.

Author contribution

SK: Data curation, Formal analysis – data synthesis, Methodology, and Visualization. EB: Project administration. SK & EB: Conceptualization, Writing – original draft preparation, review & editing.

Competing interests

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

ACKNOWLEDGEMENTS

The authors have no conflicts of interest to declare.

Acknowledgements

The research presented in this paper was funded by the Yonsei University Research Fund (2024-22-0529). Additionally, this research was supported by two grants from the National Research Foundation of Korea (NRF): NRF-2018R1A5A7025409, funded by the South Korean Ministry of Science & ICT, and RS-2023-00301702, awarded through the Global Learning & Academic Research Institution for Master's · PhD Students and Postdocs (LAMP) Program, funded by the Ministry of Education. The authors are grateful to Absur Khan Siam, Rifakat Alim Rashkee, and Yerim Han, undergraduate students at the Korea Advanced Institute of Science and Technology, for their assistance in the initial collection and organization of metadata. We also are deeply grateful to our two reviewers, John (Jack) Williams and Jessica Blois, for their thoughtful, detailed, and constructive feedback, which substantially improved the clarity and quality of the manuscript and datasets.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

Birks, H. H. and Birks, H. J. **References**

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

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

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

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

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

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

Finsinger, W., Mauri, A., Phelps, L. N., Zanon, M., Abegglen, R., Åkesson, C. M., Alba-Sánchez, F., Anderson, R. S., Antipina, T. G., Atanassova, J. R., Beer, R., Belyanina, N. I., Blyakharchuk, T. A., Borisova, O. K., Bozilova, E., Bukreeva, G., Bunting, M. J., Clò, E., Colombaroli, D., Combourieu-Nebout, N., Desprat, S., Di Rita, F., Djamali, M., Edwards, K. J., Fall, P. L., Feurdean, A., Fletcher, W., Florenzano, A., Furlanetto, G., Gaceur, E., Galimov, A. T., Gałka, M., García-Moreiras, I., Giesecke, T., Grindean, R., Guido, M. A., Gvozdeva, I. G., Herzschuh, U., Hjelle, K. L., Ivanov, S., Jahns, S., Jankovska, V., Jiménez-Moreno, G., Karpińska-Kołaczek, M., Kitaba, I., Kołaczek, P., Lapteva, E. G., Latalowa, M., Lebreton, V., Leroy, S., Leydet, M., Lopatina, D. A., López-Sáez, J. A., Lotter, A. F., Magri, D., Marinova, E., Matthias, I., Mavridou, A., Mercuri, A. M., Mesa-Fernández, J. M., Mikishin, Y. A., Milecka, K., Montanari, C., Morales-Molino, C., Mrotzek, A., Muñoz Sobrino, C., Naidina, O. D., Nakagawa, T., Nielsen, A. B., Novenko, E. Y., Panajiotidis, S., Panova, N. K., Papadopoulou, M., Pardoe, H. S., Pędziszewska, A., Petrenko, T. I., Ramos-Román, M. J., Ravazzi, C., Rösch, M., Ryabogina, N., Sabariego Ruiz, S., Salonen, J. S., Sapelko, T. V., Schofield, J. E., Seppä, H., Shumilovskikh, L., Stivrins, N., Stojakowits, P., Svobodova Svitavska, H., Święta-Musznicka, J., Tantau, I., Tinner, W., Tobolski, K., Tonkov, S., Tsakiridou, M., et al.: The Eurasian Modern Pollen Database (EMPD), version 2, Earth Syst. Sci. Data, 12, 2423–2445, https://doi.org/10.5194/essd-12-2423-2020, 2020.

Farley, S. S., Dawson, A., Goring, S. J., and Williams, J. W.: Situating Ecologyecology as a Big-Data Sciencebig-data science: Current Advances, Challenges, and Solutions, Bioscienceadvances, challenges, and solutions, BioScience, 68, 563–576, https://doi.org/10.1093/biosci/biy068,https://doi.org/10.1093/biosci/biy068, 2018.

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.

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.: A guide to the processing and standardization of global palaeoecological data for large--scale syntheses using fossil pollen, Glob. Ecol. Biogeogr.,Global Ecology and Biogeography, 32, 1377–1394, https://doi.org/10.1111/geb.13693,https://doi.org/10.1111/geb.13693, 2023.

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, Review of Palaeobotany and Palynology, 223, 104–115, https://doi.org/10.1016/j.revpalbo.2015.09.008, 2015.

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

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

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

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 Indo-Pacific Pollen Database-a Neotoma constituent database, Climate of the Past, 20, 2473–2485, https://doi.org/10.5194/cp-20-2473-2024, 2024.

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

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

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., Yang, D.-Y., Hong, S.-S., and Yi, S.: Quaternary environmental studies in South Korea, Episodes, 47, 511–535, https://doi.org/10.18814/epiiugs/2024/02403s09,https://doi.org/10.18814/epiiugs/2024/02403s09, 2024.

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

~~Kim, S. H. and Byun, E.: A geospatial inventory dataset of study sites in a Korean.~~ Quaternary ~~paleoecology database. figshare. Dataset. https://doi.org/10.6084/m9.figshare.28236596.v2, 2025~~Research 96: 28-38, https://doi.org/10.1017/qua.2020.47, 2020.

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

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

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

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

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

Park, J.: The Applicability of Stable Isotope Analyses on Sediments to Reconstruct Korean Paleoclimate. Journal of the 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,~~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.~~

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,~~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,~~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,~~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_Calendar BP, and U-Th_Calendar_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

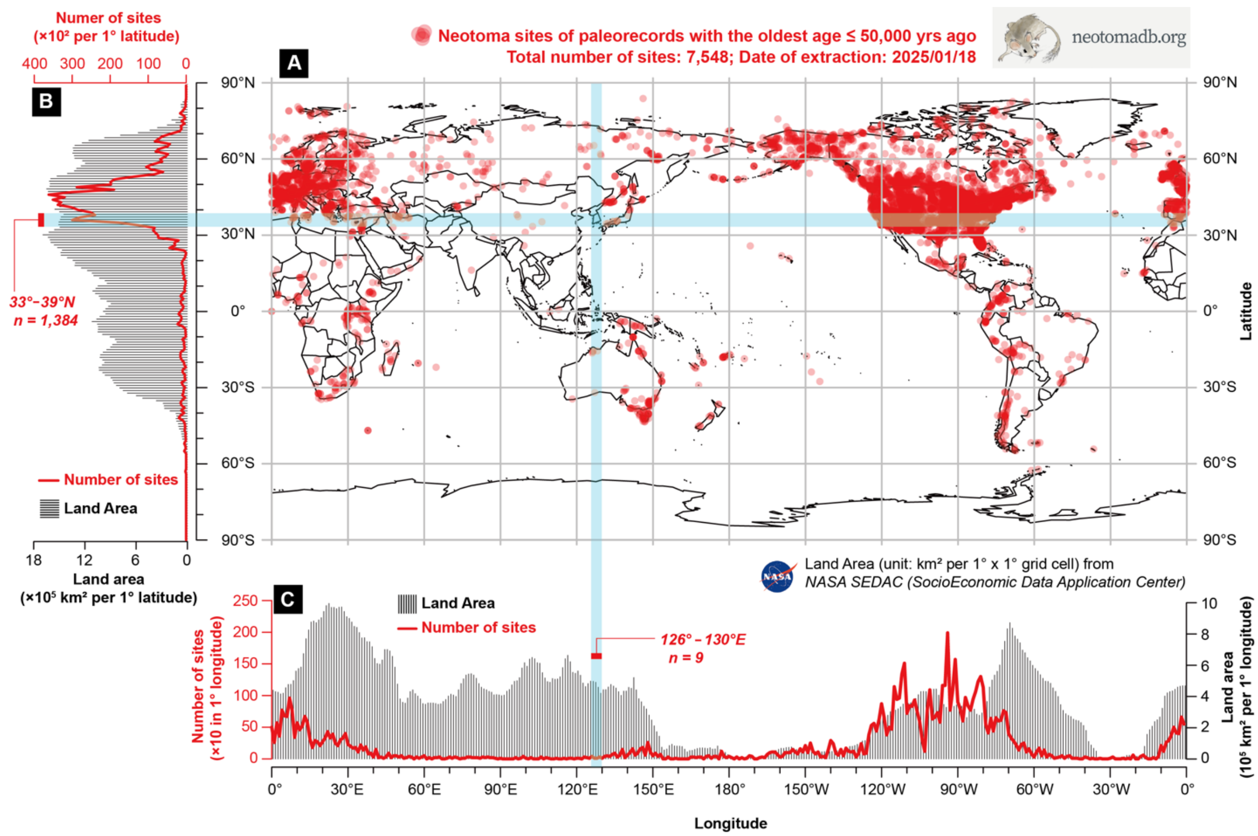


Figure 1. (A) Global distribution of sites with paleorecords (age range: ≤ 50,000 years ago) uploaded to the Neotoma Paleoecology Database (Total sites: 7,548; Extraction date: 2025/01/18). These records were selected based on the availability of data within the last 50,000 years, excluding sites having 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).

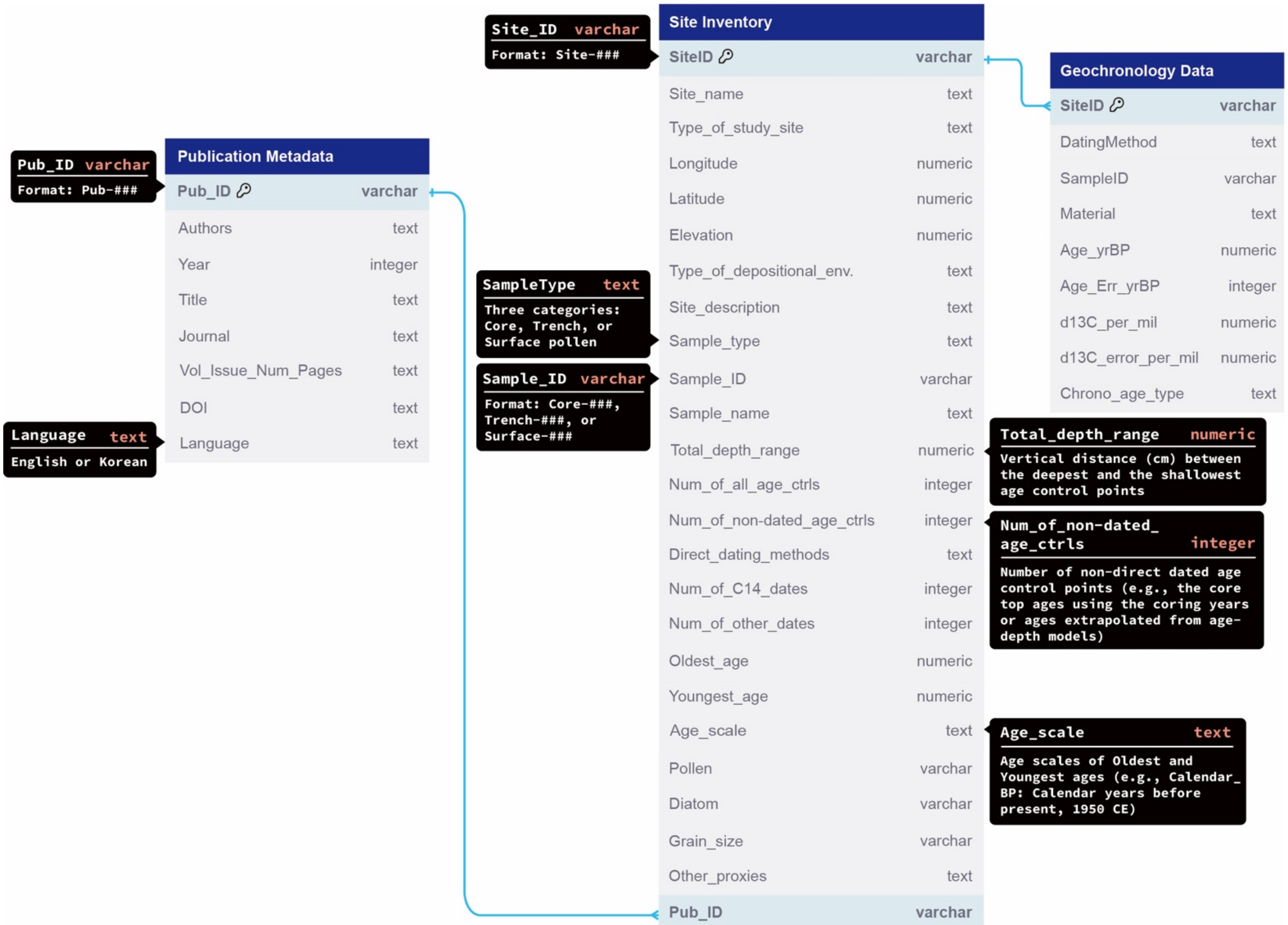


Figure 2. Entity-Relationship Diagram (ERD) illustrating the three datasets built in this study. This ERD was created using dbdiagram.io. Data types, following standard conventions of SQL (Standard Query Language), include varchar (variable-length character strings), text (long-form text), numeric (decimal values), and integer (whole number). Table 1 provides details of data fields.

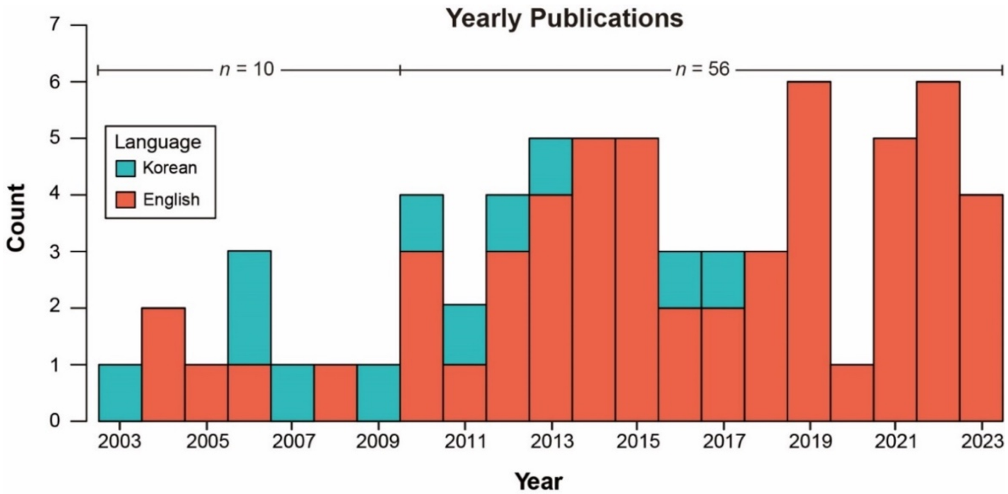


Figure 3. Barplot of annual publications from 2003 to 2023.

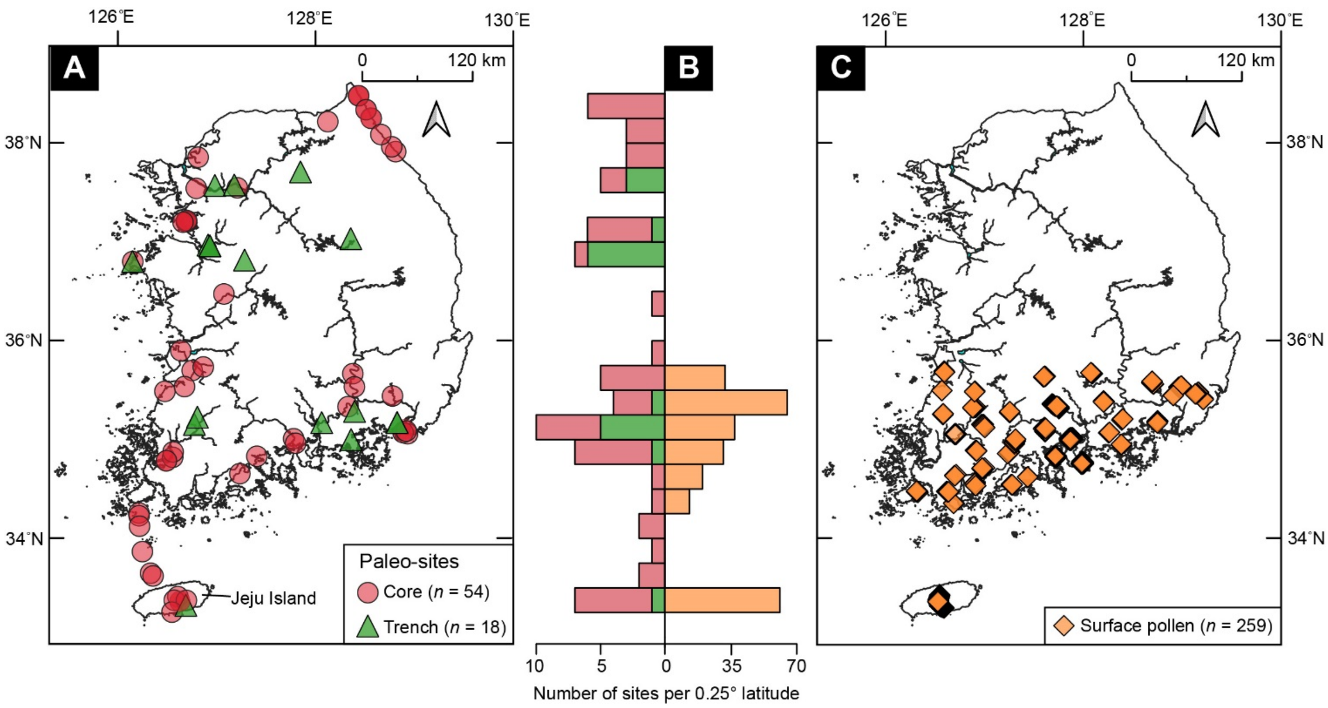


Figure 4. Geographic distributions of sites. (A) Paleo-sites. (B) Surface pollen sites. Site markers may appear overlapped in areas with high site density. Elevational distributions of the sites are shown in Figure 6.

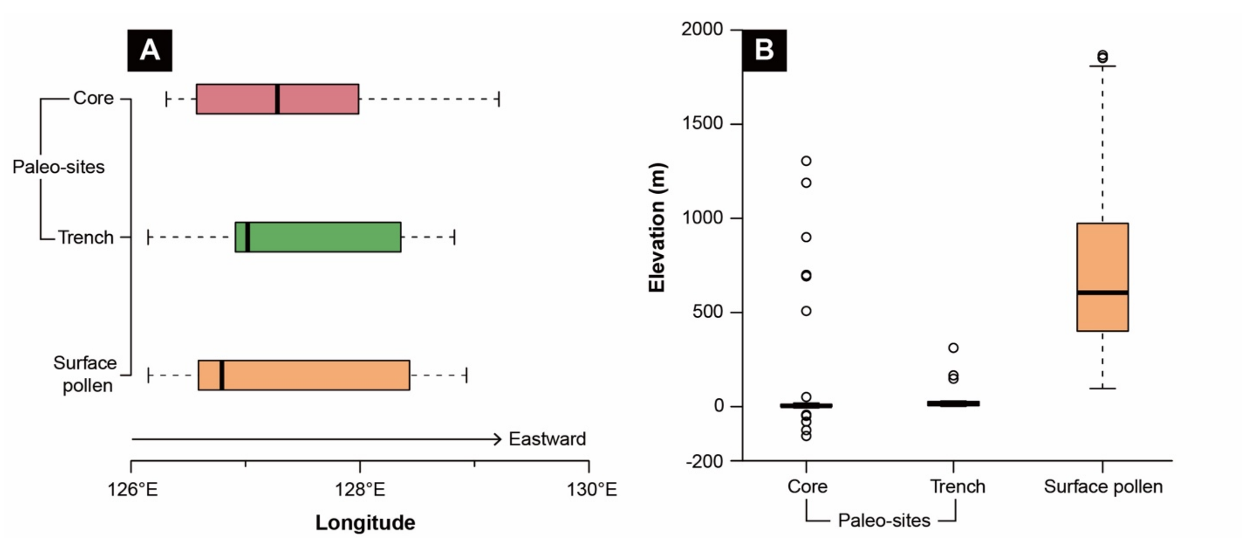


Figure 5. Boxplots of site locations and elevations. (A) Longitudinal distribution of sites. (B) Elevational distribution of sites.

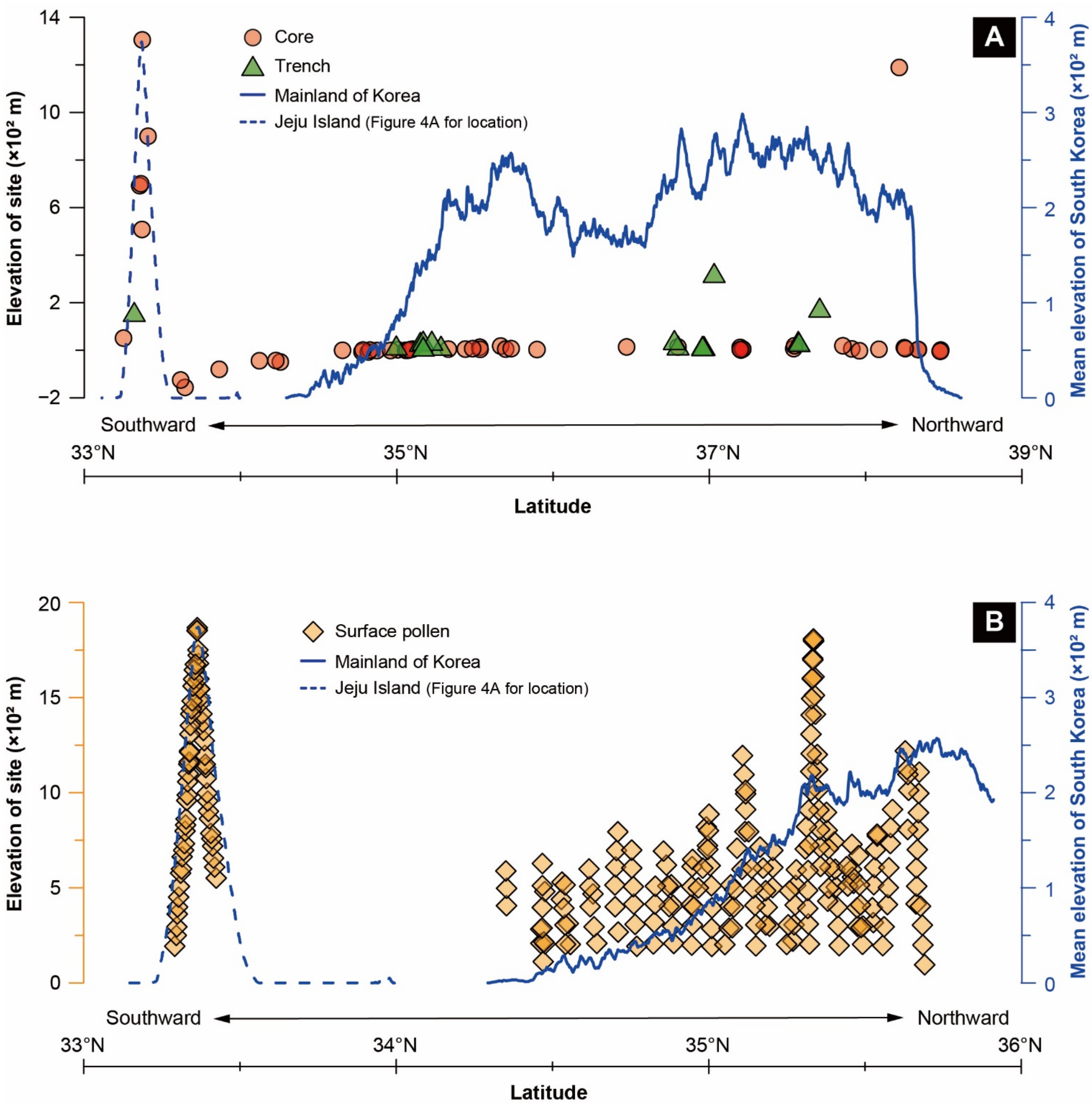


Figure 6. Elevational distribution of sites. (A) Paleo-sites. (B) Surface pollen sites. These plots have different latitudinal ranges (A: 33°N to 39°N, B: 33°N to 36°N). The blue lines display mean elevations along a latitudinal gradient, averaged elevations across all longitudes within the mainland (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.

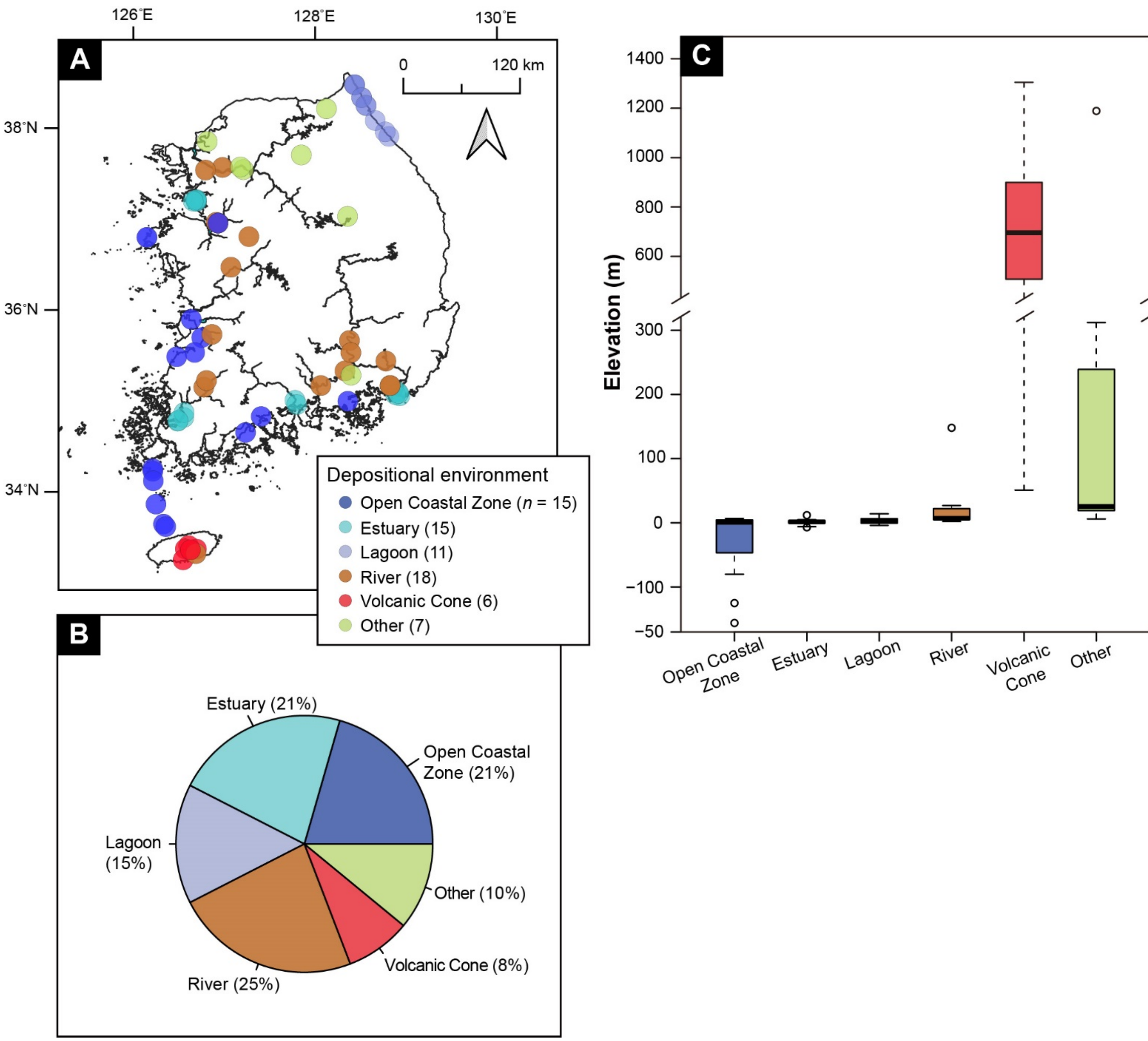


Figure 7. Paleo-sites (*n* = 72) categorized by six types of depositional environments. (A) Geographic locations of sites. (B) Proportion of sites by depositional environment. (C) Boxplots of elevations. The median elevations for each environmental type are as follows: Open Coastal Zone (1 m), Estuary (1.5 m), Lagoon (3 m), River (10 m), Volcanic Cone (696 m), and Other (21 m).

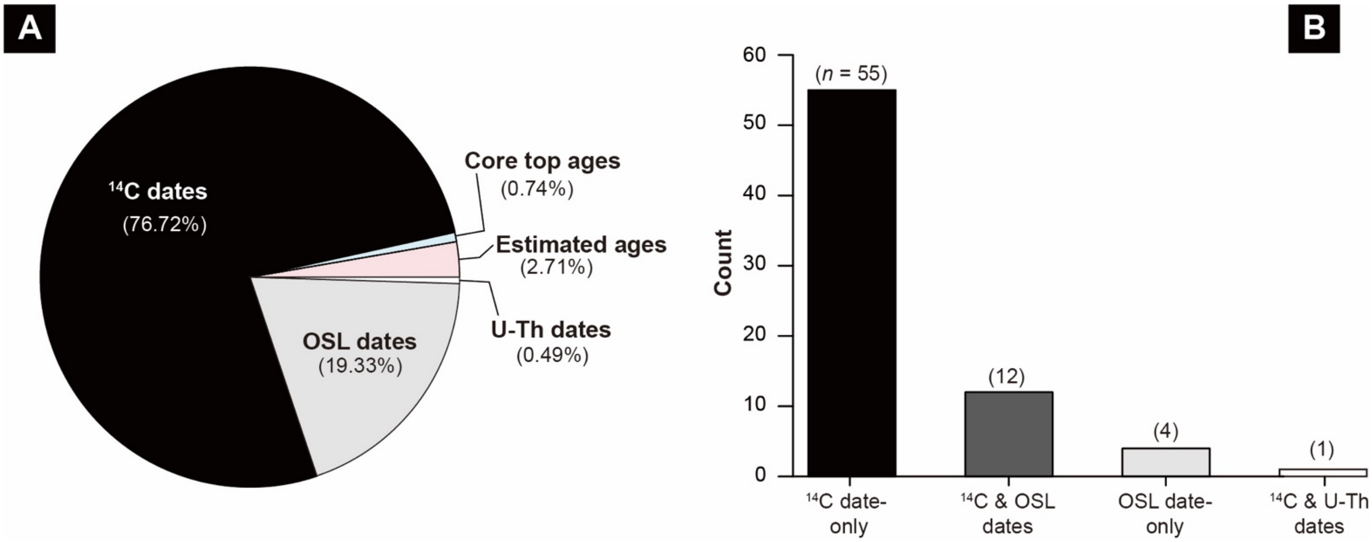


Figure 8. (A) Pie chart showing the proportional distribution of age control types, including three direct dating methods, ^{14}C, OSL, and U-Th, and two indirect controls, Core top ages and Estimated ages derived from age-depth modeling. Percentages may not total 100% due to rounding. (B) Bar plot showing 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.

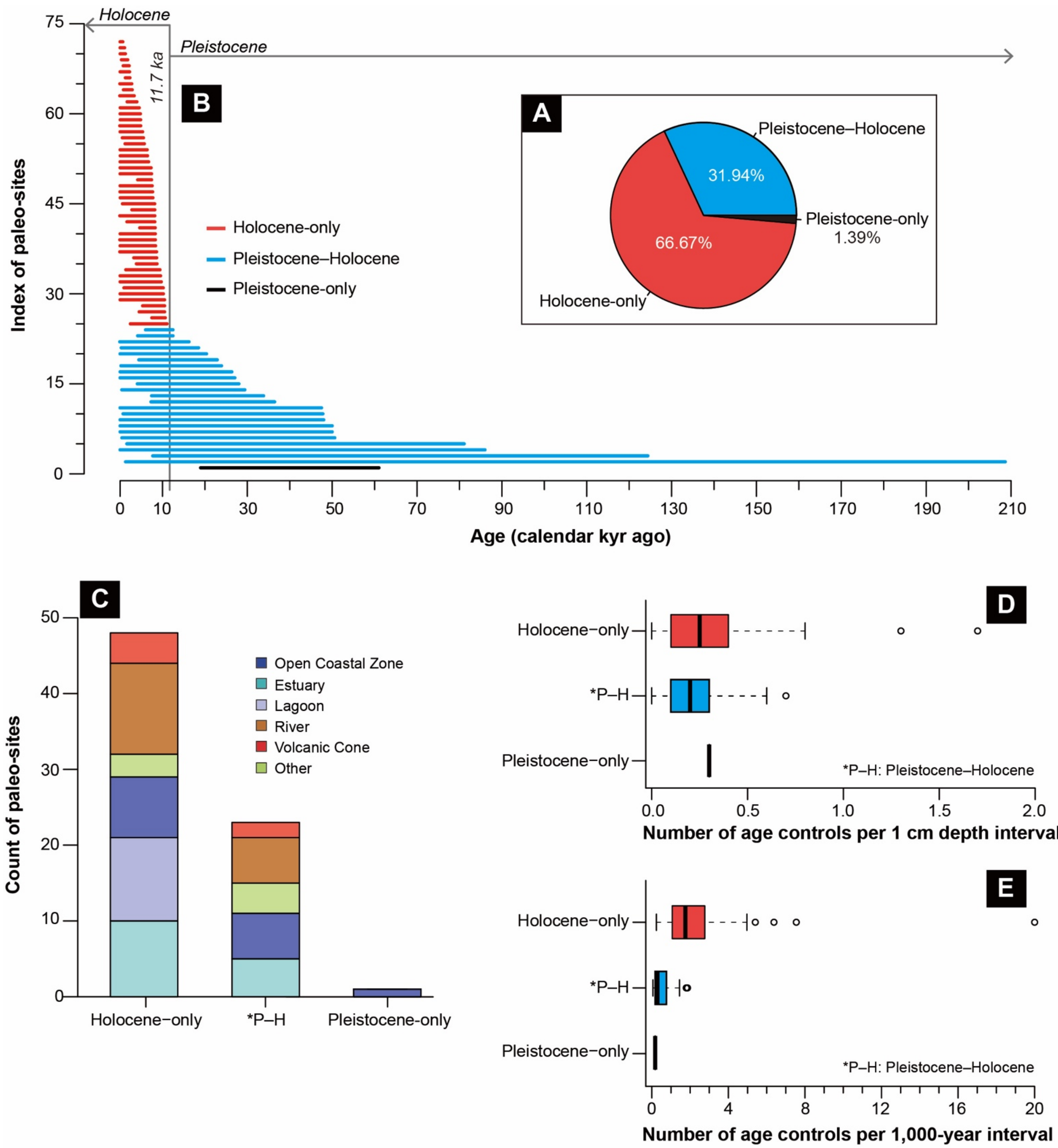


Figure 9. (A) Distribution of paleo-sites by three groups of age-coverage (kyr = 1,000 yea): Holocene-only (*n* = 48), Pleistocene–Holocene (23), and Pleistocene-only (1). (B) Chronological extent of records from paleo-sites, ordered by descending oldest ages within each age-coverage group. (C) Counts 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.

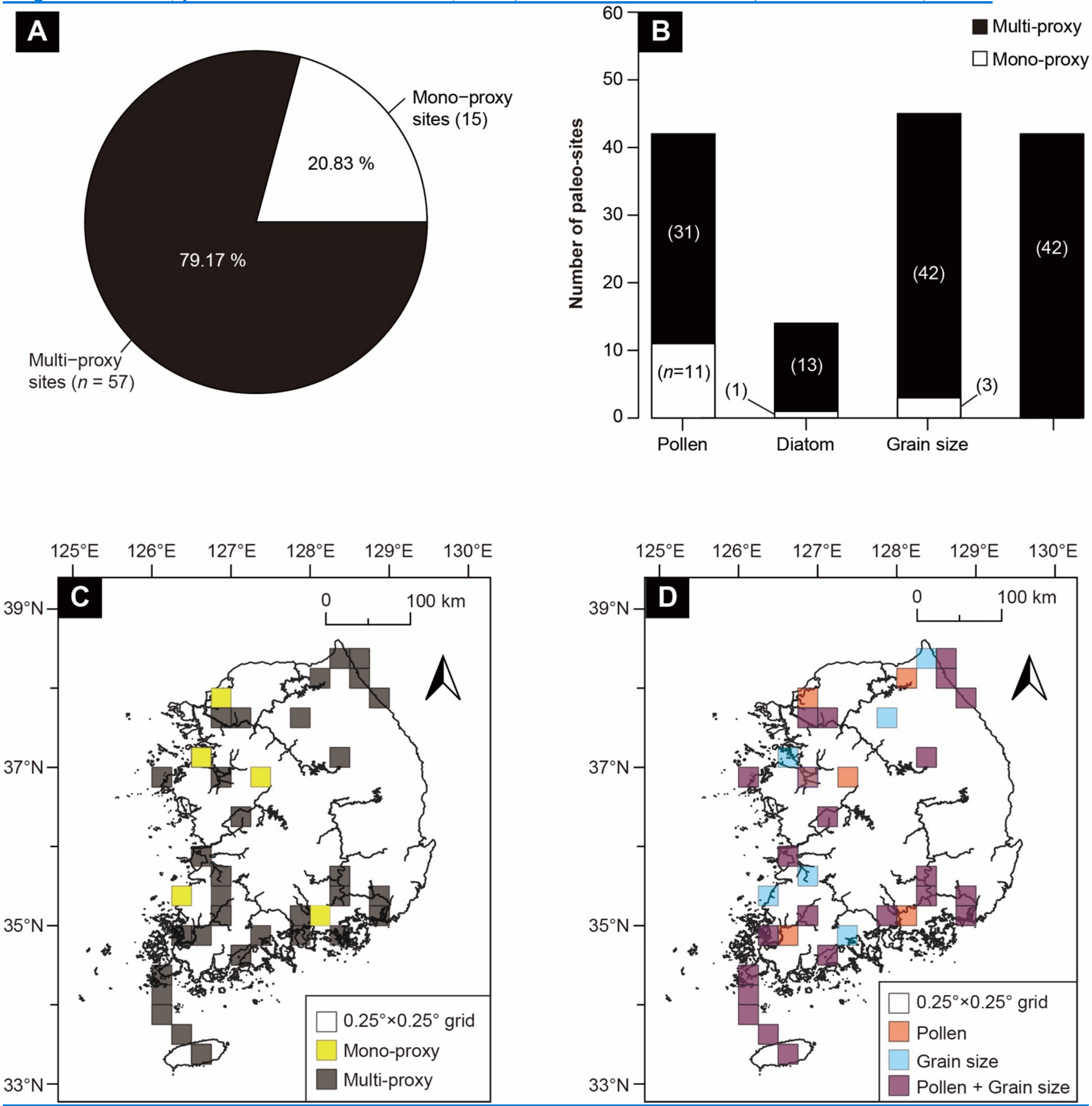


Figure 10. (A) Proportion of mono-proxy and multi-proxy sites. (B) Number of paleo-sites by four proxy types. (C & D) Geospatial distribution of proxy-based records for a 0.25° × 0.25° grid. These maps display grids containing proxy-based records from paleo-sites in South Korea. In (C), grids are classified 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.

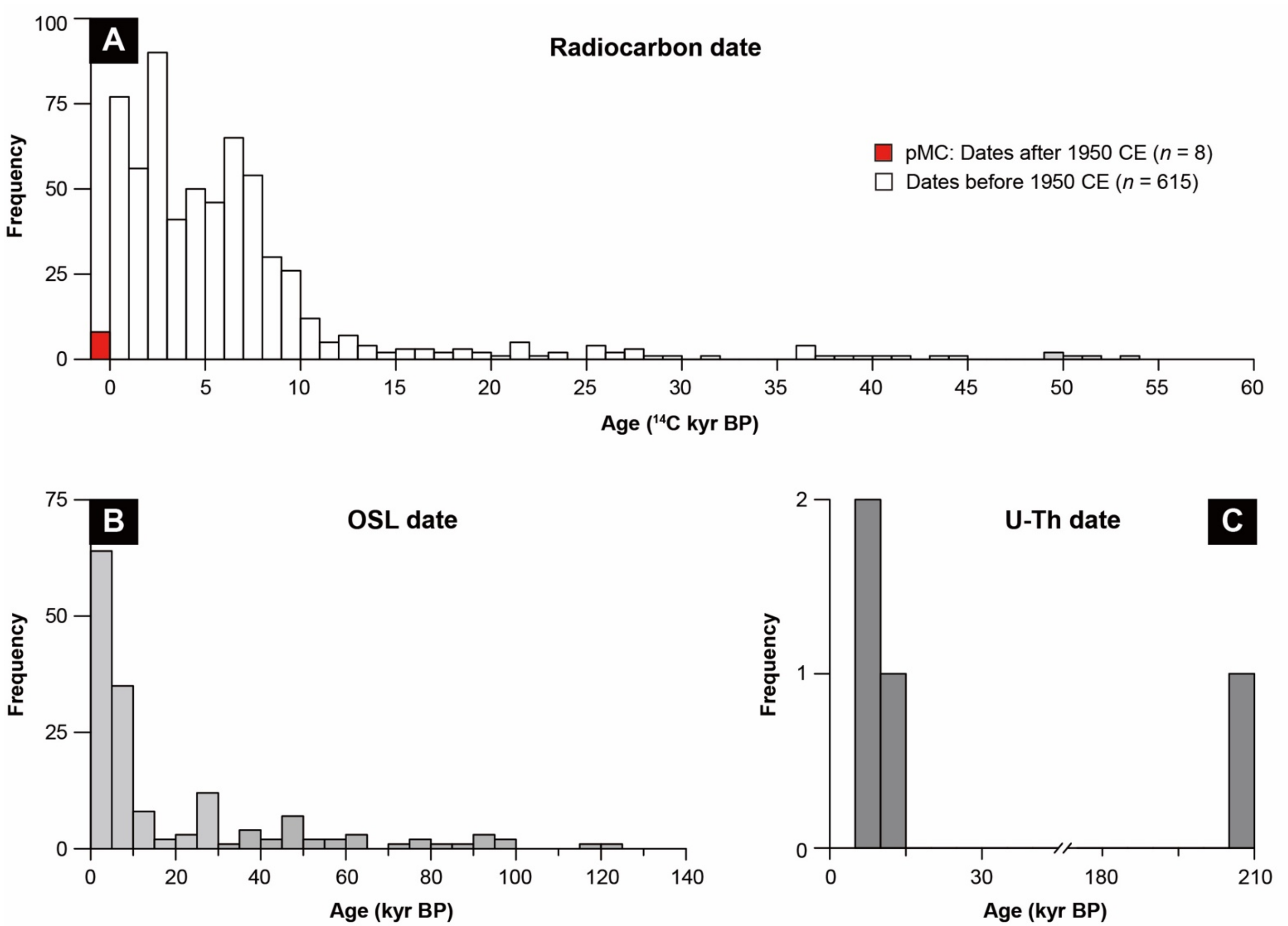


Figure 11. Histograms of dates over ages (unit: kyr BP, where means 1,000 years before present). (A) Radiocarbon dates (range of pre-1950 dates: >53,671 to 27 ^{14}C years BP; median: 4,790 ^{14}C years BP). (B) OSL dates (range: 124,400 to 140 years BP; median: 7,114 years BP). (C) U-Th dates (range: 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.

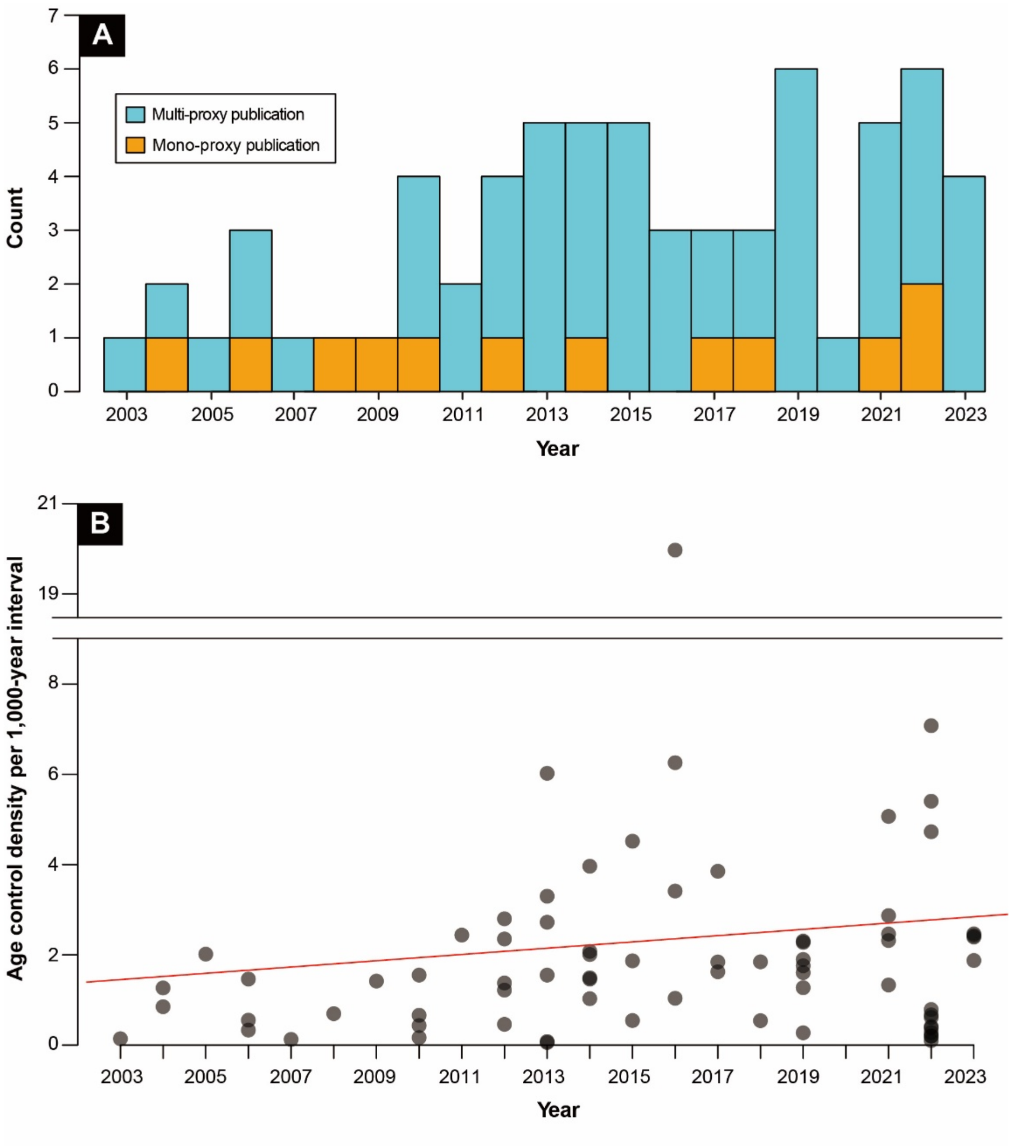


Figure 12. Temporal trends in (A) the number of multi-proxy and mono-proxy publications and (B) age control density per 1,000-year interval. In (A), 64 publications are shown because two articles studying surface pollen samples are excluded. In (B), the age control density, defined as the total number 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 ($y = 0.06x – 112.27$; $R^2 = 0.02$; $p\text{-}value = 0.30$).