https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

Open Access Earth System Science Data Discussions

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

Soo Hyun Kim[1], 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 (eb@yonsei.ac.kr)

Abstract. Ecological insights beyond human-observable time scales are derived from geologically preserved records in lake and wetland 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 Korean Quaternary paleoecology data remain underrepresented in these global efforts, despite a growing body of the relevant research. Here, we organize an inventory of 328 paleoecological study sites (72 paleo-sites for sediment records and 256 surface sites for surface pollen samples) in South Korea, compiled from 66 research articles published between 2003 and 2023. We have structured three datasets related to this inventory: (1) Publication Metadata, which provides citation details of the 66 articles; (2) Site Inventory, which contains geospatial, depositional environments, chronological ranges, proxies, and indexed publications; and (3) Chron-Depth Collection, which includes chronological details (dating methods, age, 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, Estuary, Lagoon, River, Volcanic cone, and Others. A total of 784 chronological controls (14C, OSL, and U-Th) were analyzed from 72 sediment records, and the majority based on radiocarbon dating. Pollen, diatoms, grain-size analysis, and geochemical markers have been extensively used as paleoenvironmental proxies, with multiproxy analyses becoming 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.

1 Introduction

Paleoecologists and paleoclimatologists in Quaternary (2.58 Ma to present) research area investigate the past ecosystems, paleoclimate, and paleoenvironmental changes during the current geologic time period, making use of abundant well-preserved sediments in certain depositional environments. While direct observations of past environments are typically limited to the past few decades, analysis of environmental proxies preserved in the Quaternary sediments allows

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

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 to reconstruct past vegetation changes require different assumptions 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.

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. For this purpose, the FAIR (findable, accessible, interoperable, and reusable) guiding principles have been proposed (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/) is one of the most remarkable contributions to the field (Williams et al., 2018). Research community-based support is available throughout the database use, from securing servers and curating data uploads to providing a user-friendly interface for global data search and downloads. The Neotoma community also helps regional databases independently manage their own servers and affiliate with Neotoma for global networking. With this possibility in mind, this work is an initiative effort and aims to develop a Korea-based research community database for

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

domestic Quaternary paleoenvironmental data. The goal is to establish a detailed regional context 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 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, which is consistent with the low representation of data from Korea in the global paleoecology database (Fig. 1) and in a recent data synthesis study (e.g., Herzschuh et al., 2023). Thus, this paper presents the first inventory dataset of the research and site metadata from a total of 328 study sites in South Korea (33˚-38˚N, 124˚-132˚E), compiled from 66 peer-reviewed publications over the past 20 years (2003-2023). The dataset provides metadata on study sites and sediment samples, including latitude, longitude, elevation, depositional setting, 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, 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.

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

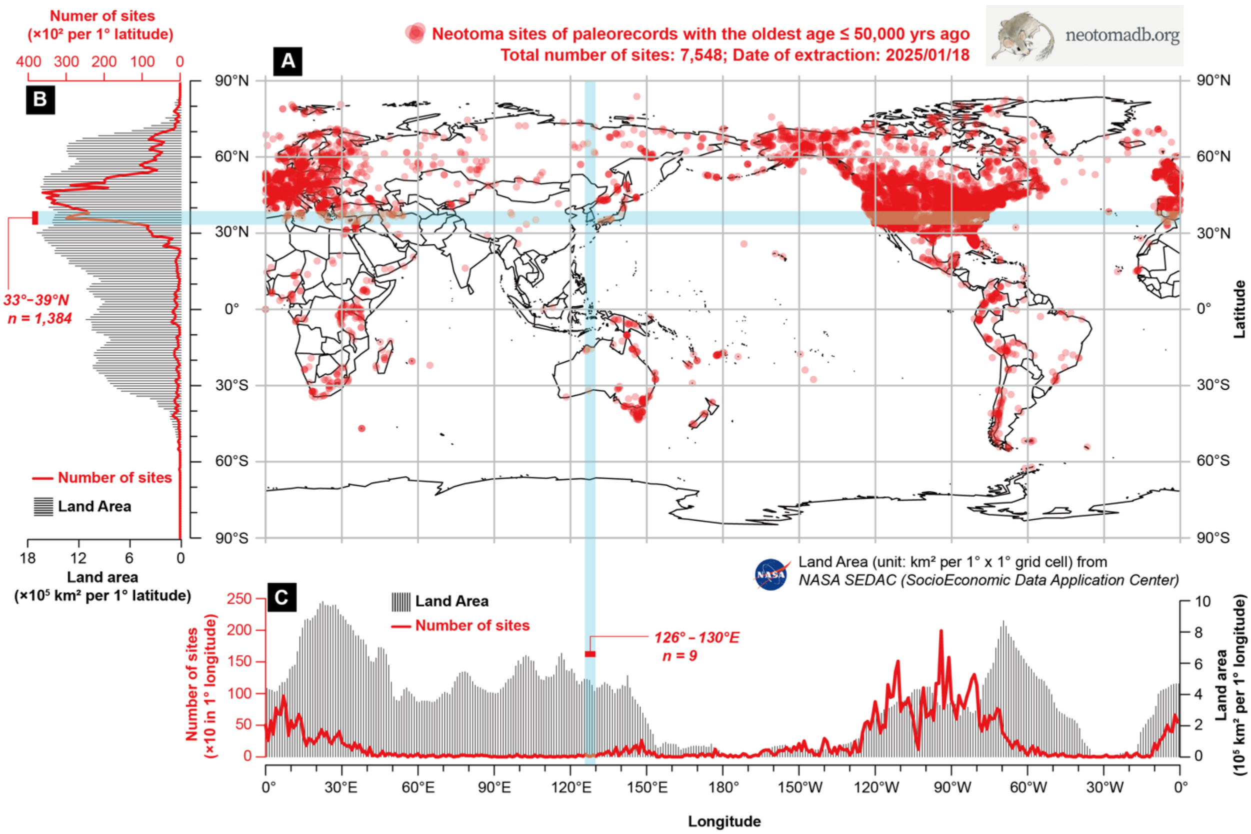


Figure 1. (A) Global distribution of sites having paleorecords (age range: ≤ 50,000 years ago) uploaded to the Neotoma database (Total sites: 7,548; Extraction date: 2025/01/18). These paleo-sites were selected based on the availability of data within the last 50,000 years, excluding sites having only modern data–surface samples, pollen trap, and datasets with the oldest age below zero. (B) Number of sites and total land area (10^5 km^2) per 1° latitude. The total 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 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 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).

2 Data Collection

We compiled 66 research papers published between 2003 and 2023, which presented proxy data from sediments and surface pollen samples in South Korean territory. The literature search was conducted in the following 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 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/), we examined additional journal articles by the authors identified in the second step.

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

Open Access Earth System Science Data Discussions

Step 4. Compilation of research articles: We organized articles related to Korean Quaternary proxy records published up to 2023 in the journals listed 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 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**. The first two datasets are linked through Publication ID, and the last two datasets through Site ID (Fig. 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 comprises chronological information used for sediment records of the sites.

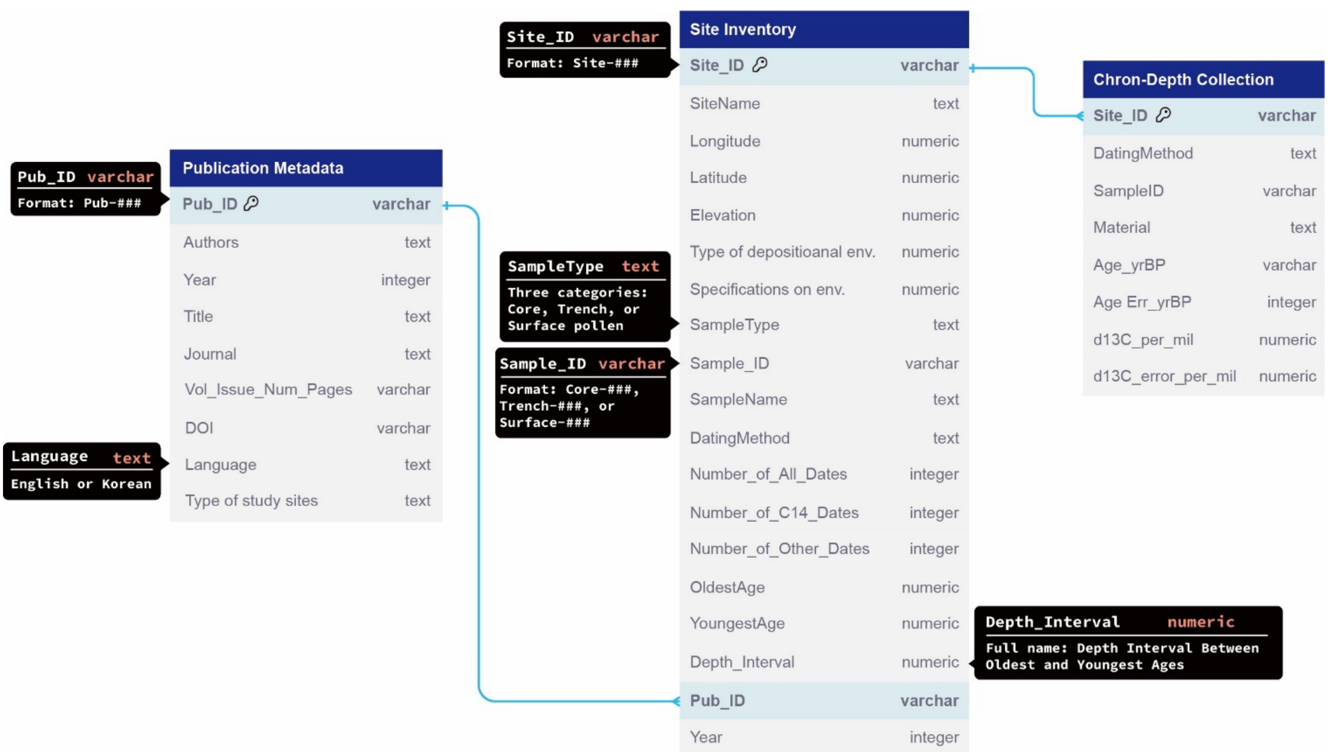


Figure 2. Entity-Relationship Diagram (ERD) illustrating the linking of three datasets in this study. The ERD was created using dbdiagram.io. The detailed description of data fields in the Site Inventory dataset is in Table 1.

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

Open Access Earth System Science Data Discussions

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.
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.

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

3 Data Processing and Description

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 include 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 trend since 2010, driven by the growing number of English-written papers in international journals (Fig. 3).

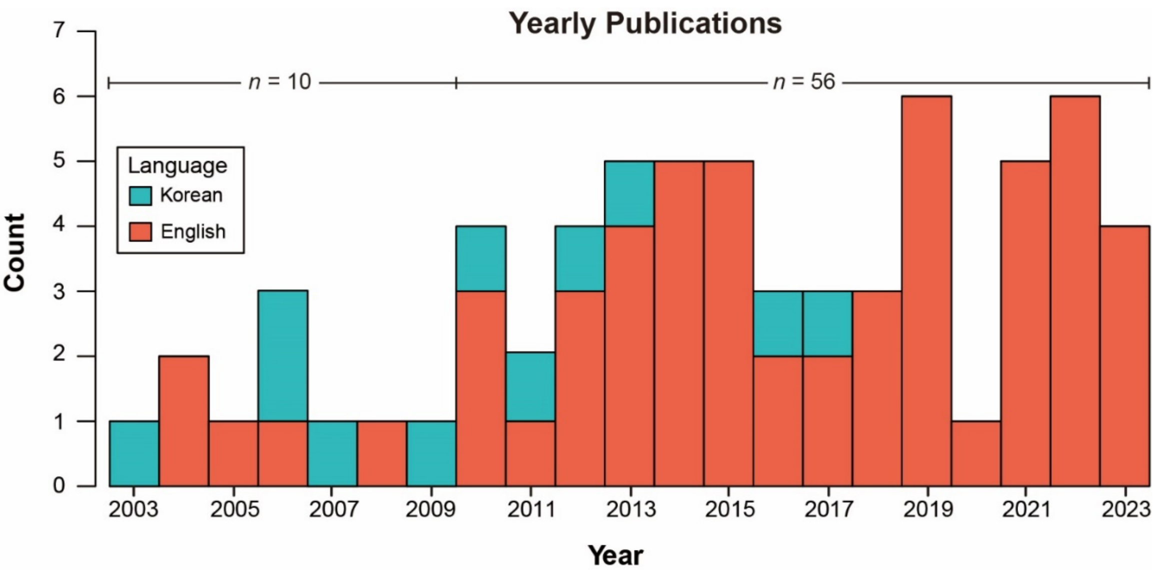


Figure 3. Barplot of annual publications between 2003 and 2023.

Regarding type of study 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 paleo-sites where stratigraphic sediments were collected 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 articles covered surface sites 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).

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).

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

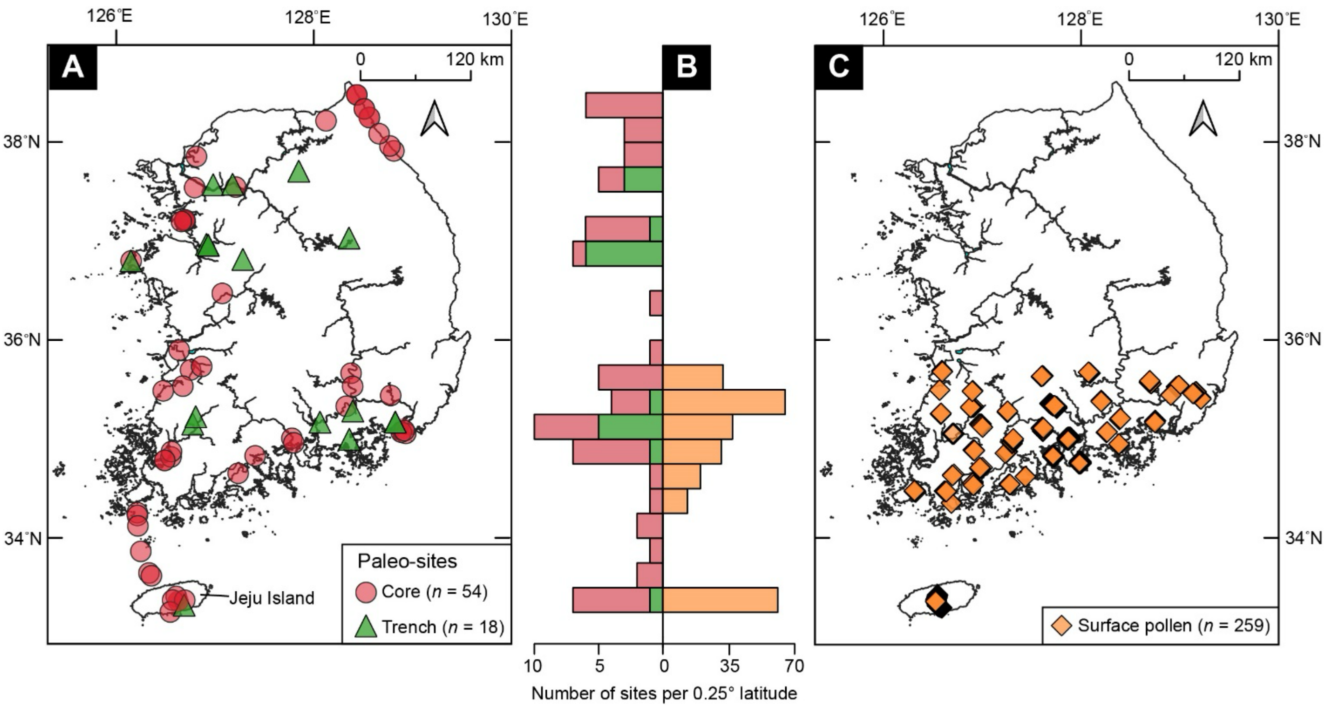


Figure 4. Geographic distributions of sites. (A) Paleo-sites. (B) Surface pollen sites.

3.2.1 Geospatial coverage

Sites with distinct geographic coordinates were treated as separate entries. We prioritized coordinates and elevations reported in the papers when available. If not, we estimated them using site maps from the articles, cross-referencing with Google Earth. Site locations were recorded in four-digit decimal degree coordinates with elevations in meters. The distribution of paleo-sites and surface sites shows no significant variation in longitude but distinct patterns in elevation (Fig. 5) and in latitude (Fig. 6). In the geographic extent (33.2508˚ – 38.4808˚N, 126.1486˚ – 128.9719˚E) of paleo-sites, their elevations (range: -156 – 1,305 m; median: 5 m) are consistent with the tendency of depositional environments for stratigraphic sediments typically found at low elevations (Fig. 5A), although their latitudinal distribution exhibits a broad coverage across South Korea except for the central eastern region (Fig. 6A). In contrast, the surface pollen sites (33.2916˚ – 35.6913˚N, 126.3090˚ – 129.2132˚E) span a wider range of elevation (range: 95 – 1,867 m; median: 605 m) compared to the paleo-sites (Fig. 5B) but are concentrated in the southern regions (Fig. 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

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

modern pollen-climate references (Fig. 6B). It is interesting to note the high elevations of Jeju's volcanic-cone wetland sites (Fig. 7), which locally shift the statistical distribution of the island paleo-sites (range: 51-1,305 m; median 692 m).

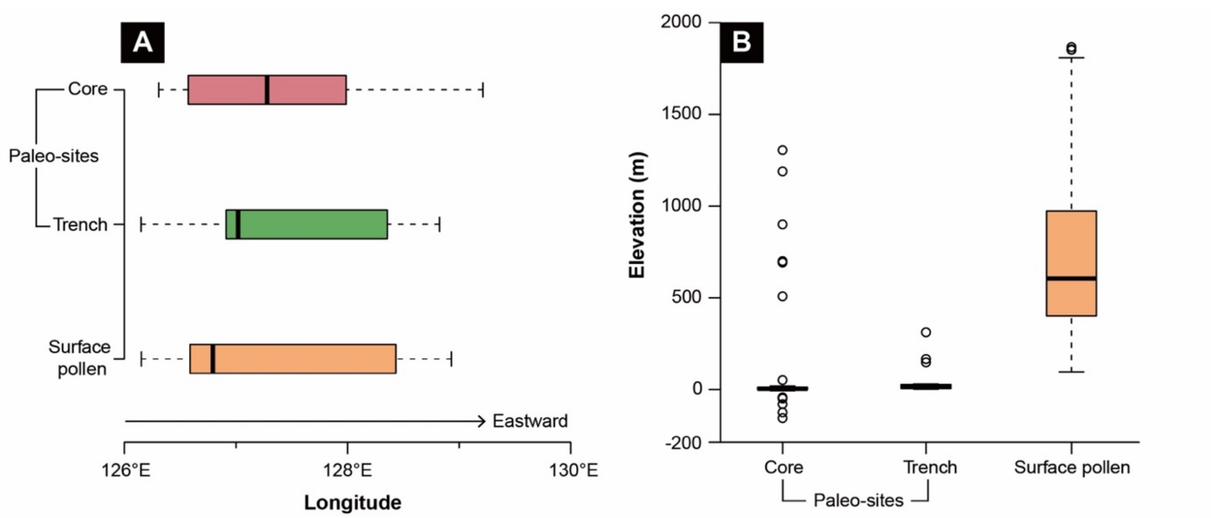
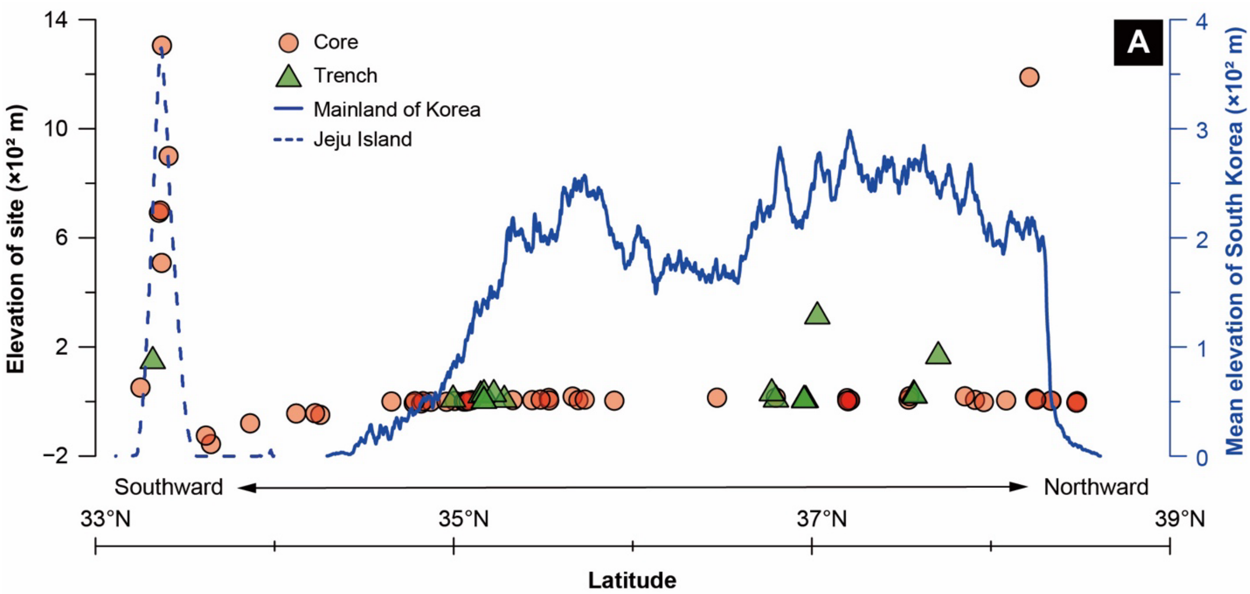
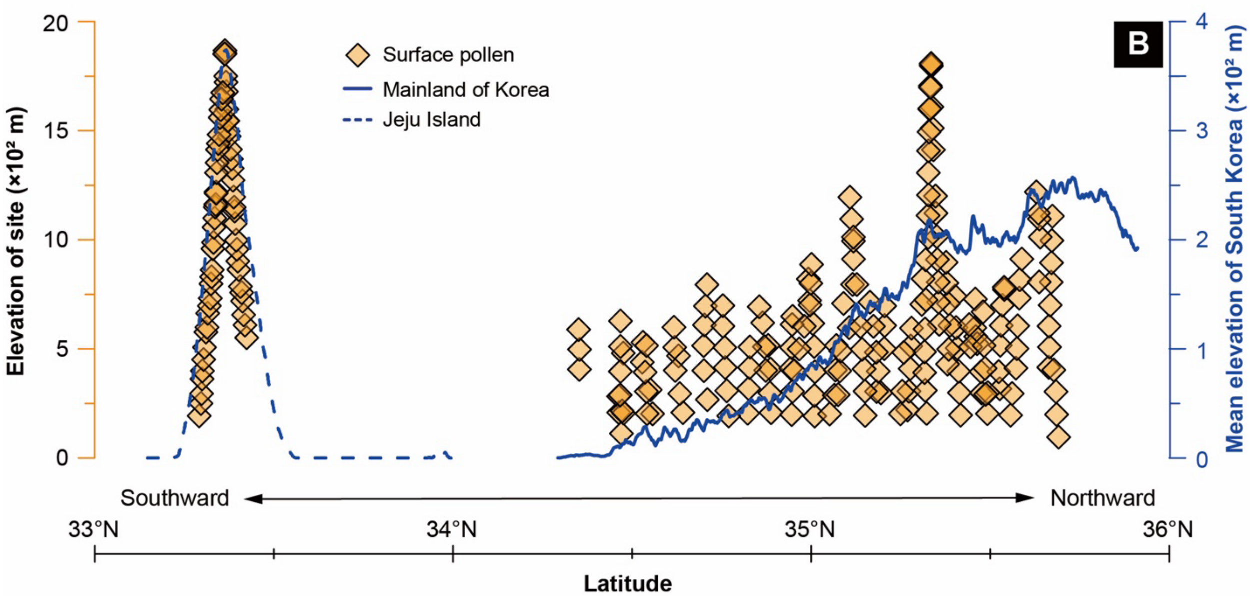


Figure 5. Boxplots of site locations and elevations. (A) Longitudinal distribution of sites. (B) Vertical distribution of sites.

3.2.2 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, 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 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. 4).





185 **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.**

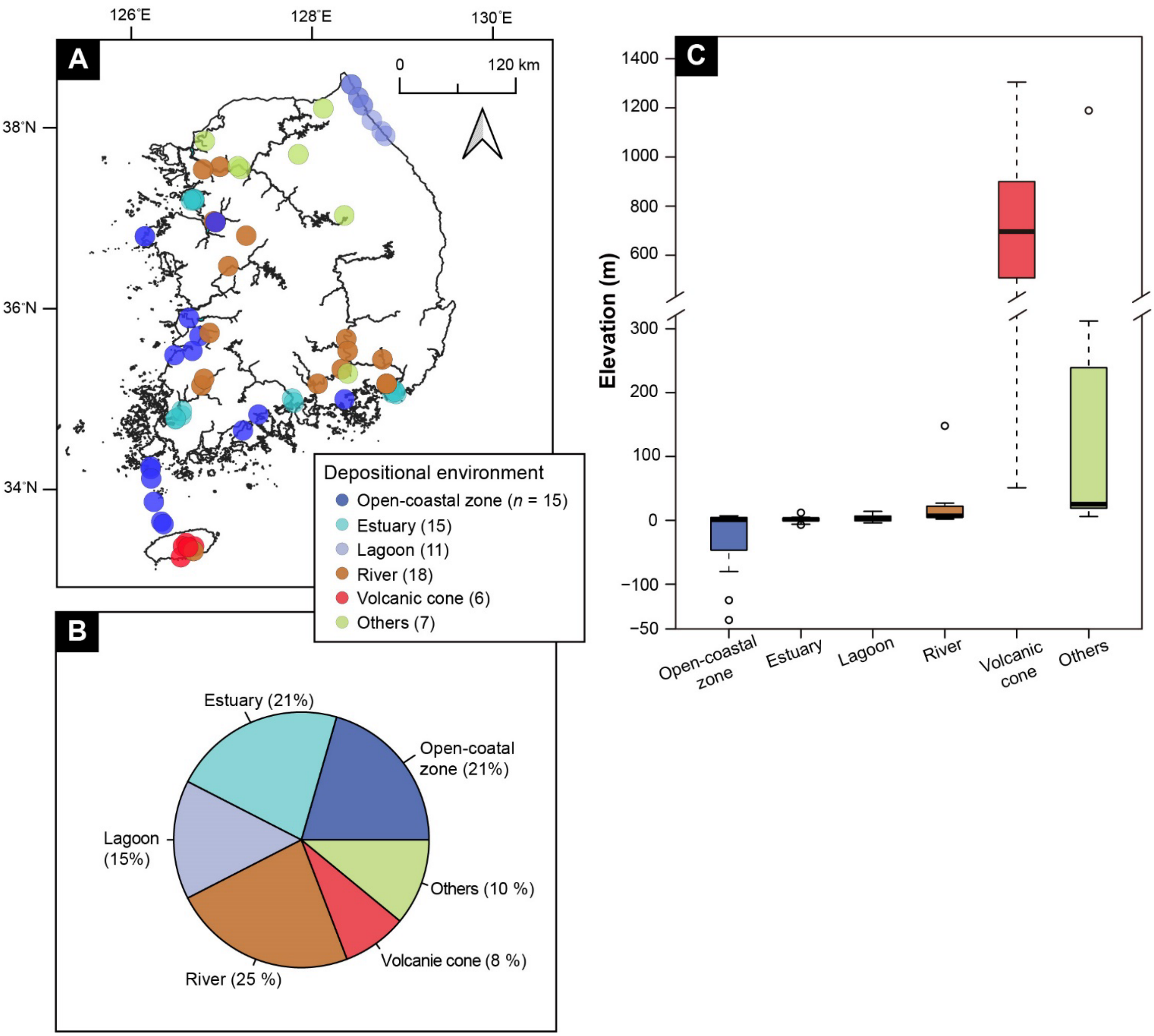
3.2.3 Depositional setting

The depositional environments of paleo-sites are categorized into six groups: Open-coastal zone, Estuary, Lagoon, River, Volcanic cone, and Others (Table 2). The "Open-coastal zone" encompasses a set of environments dominantly influenced by oceanic depositional processes, such as beach, 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 associated with rivers and streams. The "Volcanic 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" 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" category. The paleo-sites can be broadly categorized into two groups, coastal and terrestrial sites. The coastal sites (n = 41), including Open-coastal zone, Estuary, and Lagoon, outnumber non-coastal (terrestrial) sites (n = 31), including River, Volcanic cone, and Others, with River being the popular sampling sites (Fig. 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

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025



215 **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).**

3.2.4 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).

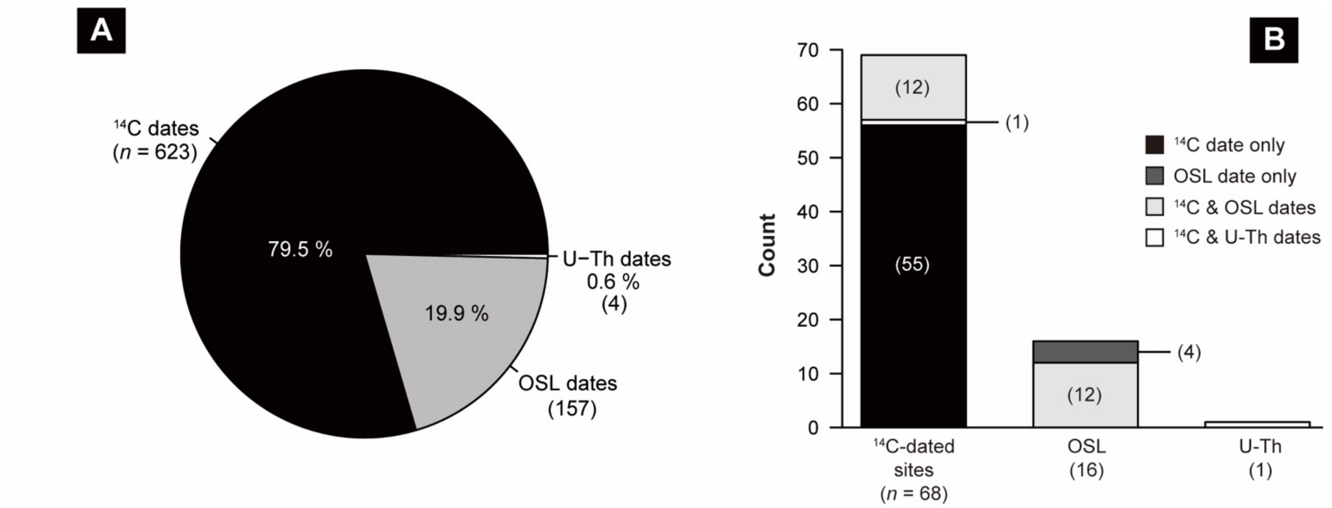


Figure 8. (A) Pie chart of absolute dates by dating methods, and (B) Bar plot of paleo-sites by dating methods.

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).

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

Open Access Earth System Science Data Discussions

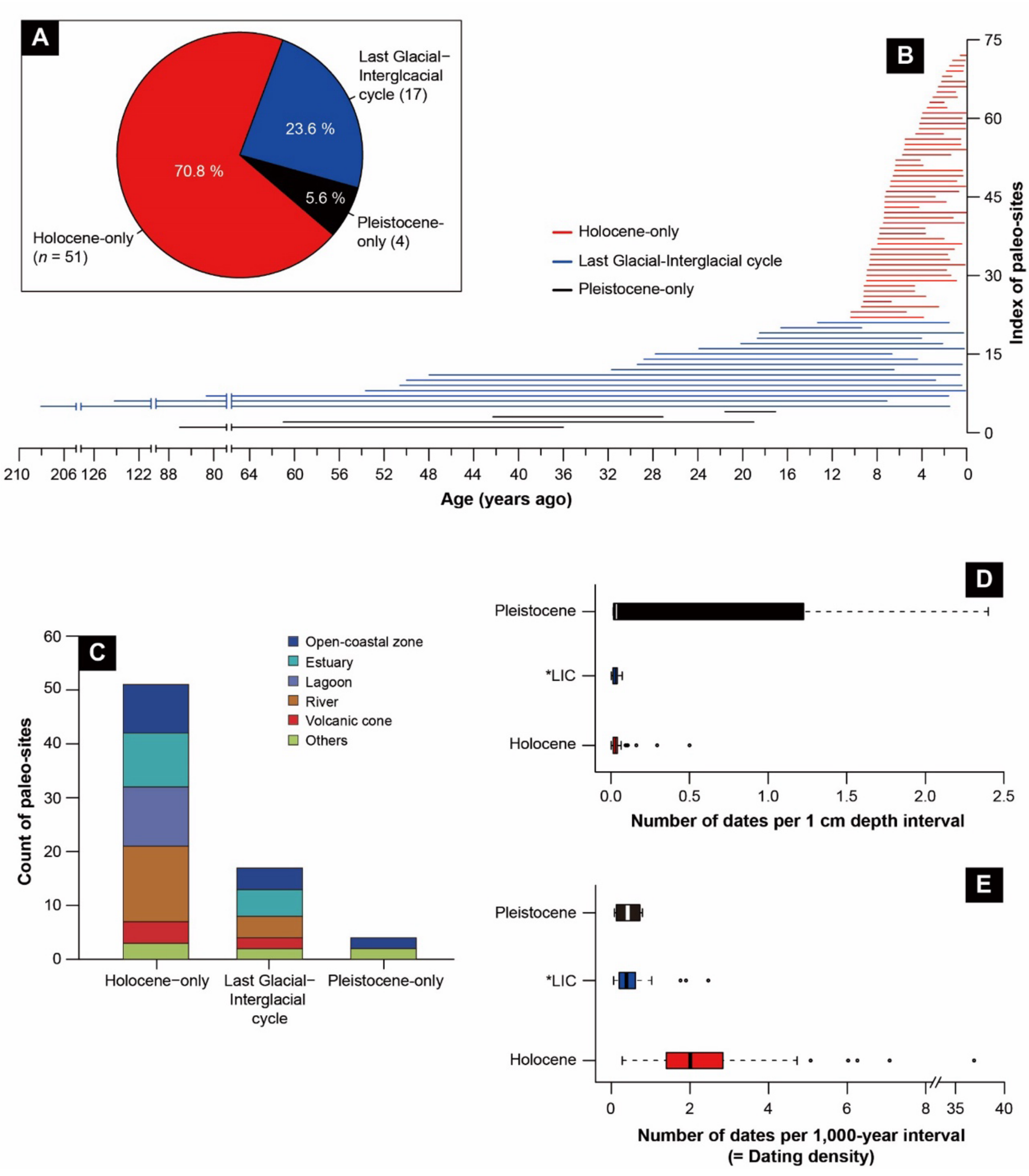
In addition, our chronological data are useful for deriving the dating density, which indicates the number of dates per time interval (Lacourse and Gajewski, 2020). The dating density (represented as the inverse of the age control density suggested in Flantua et al., 2015) is calculated by dividing the total number of dates by the gap between oldest and youngest ages. This metric is used to assess the accuracy 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 dating density than Pleistocene-only and Last glacial-interglacial cycle sites (Fig. 9E).

3.2.5 Proxy

The proxy-related data are 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 the most frequently investigated proxies, with some sites even showing mono-proxy reconstructions that rely on only one of them (Fig. 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 δ13C) are more commonly used than inorganic proxies such as mineralogical compositions, trace element content, and magnetic susceptibility.

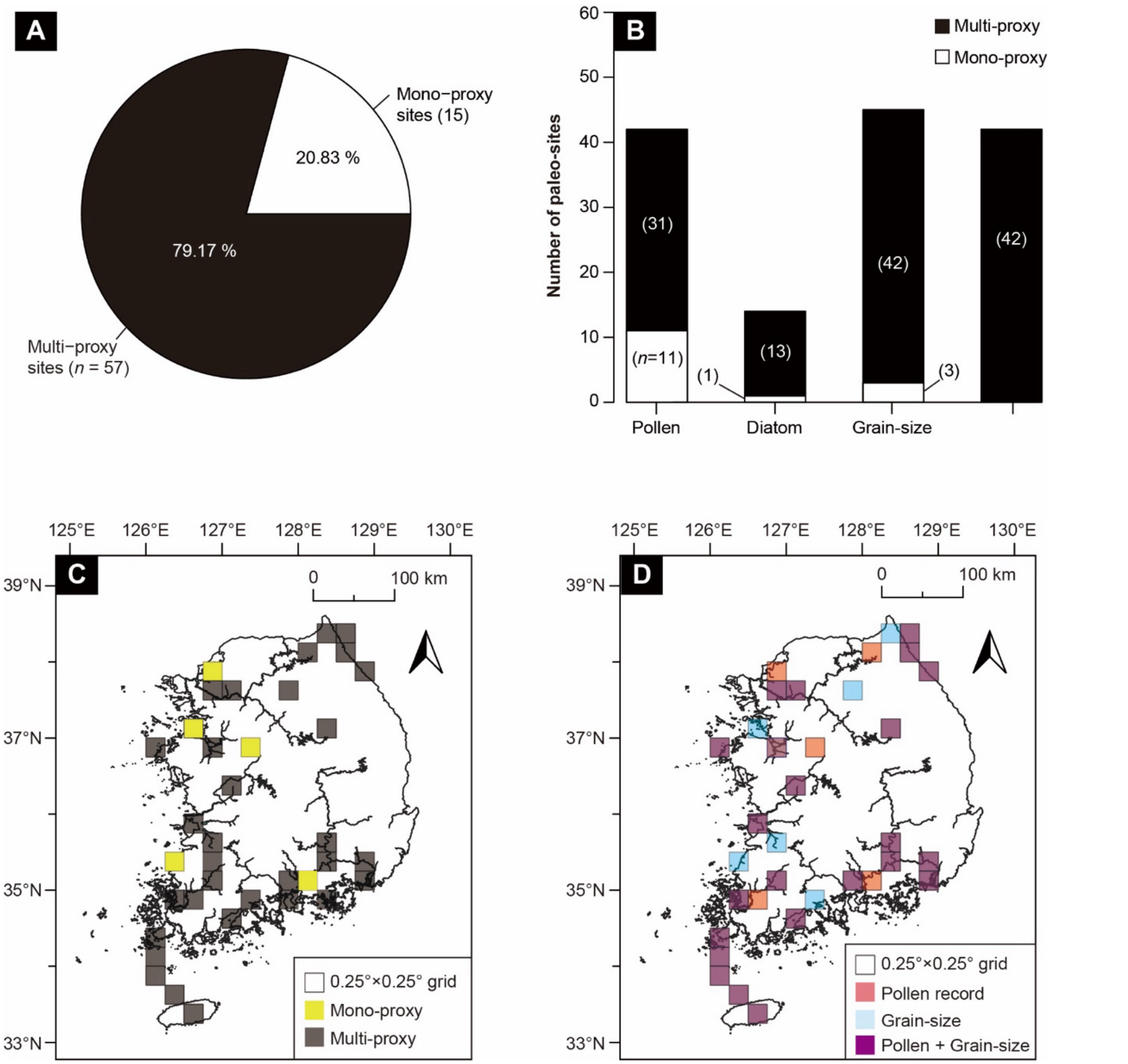
Overall, for surface samples, our dataset includes only pollen as the primary indicator used to construct taxon-based transfer functions for paleotemperatures estimates (Lee et al., 2022; Park and Park 2015). While non-palynological organisms, such as ostracods from brine and freshwater swamps and subterranean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

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025



275 **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.**

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025



285 **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.**

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

Open Access Earth System Science Data Discussions

3.3. Dataset III: Chron-Depth Collection

The chronological dataset corresponding to 72 paleo-sites are documented separately. This dataset consists of nine categories: SiteID, dating method (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 δ13C (‰). For the sample ID, AMS Lab IDs are prioritized particularly for radiocarbon dates; but when unavailable, identification indices 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). Most dates are concentrated within the Holocene epoch (≤11,700 years BP), accounting for 90 % of 14C dates, 68% of OSL dates, and 50% of U-Th dates.

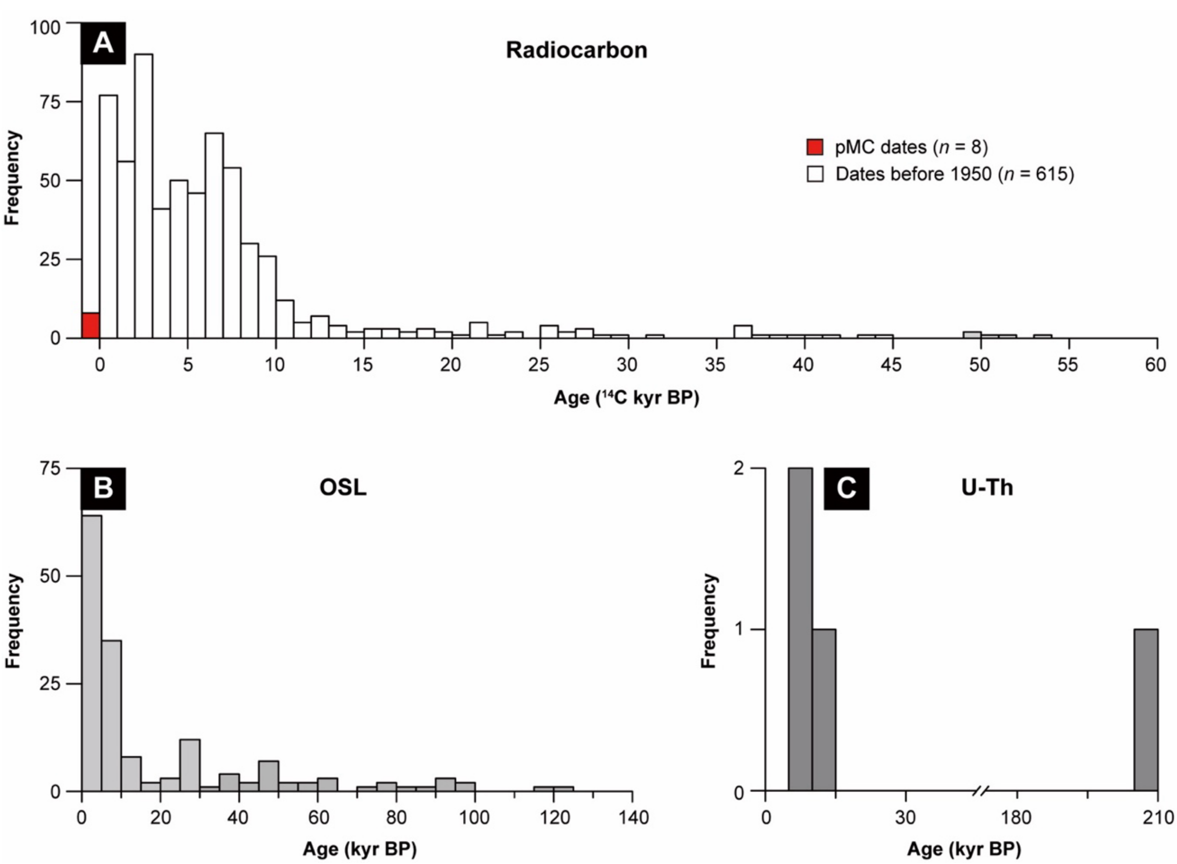


Figure 11. Histograms of dates over ages. (A) Radiocarbon dates (range of pre-1950 dates: >53,671–27 ^{14}C years BP; median: 4790 ^{14}C years BP). (B) OSL dates (range: 124,400–140 years BP; median: 7,114 years BP). (C) U-Th dates (range: 208,000–6900 years BP; median: 10,000 years BP).

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

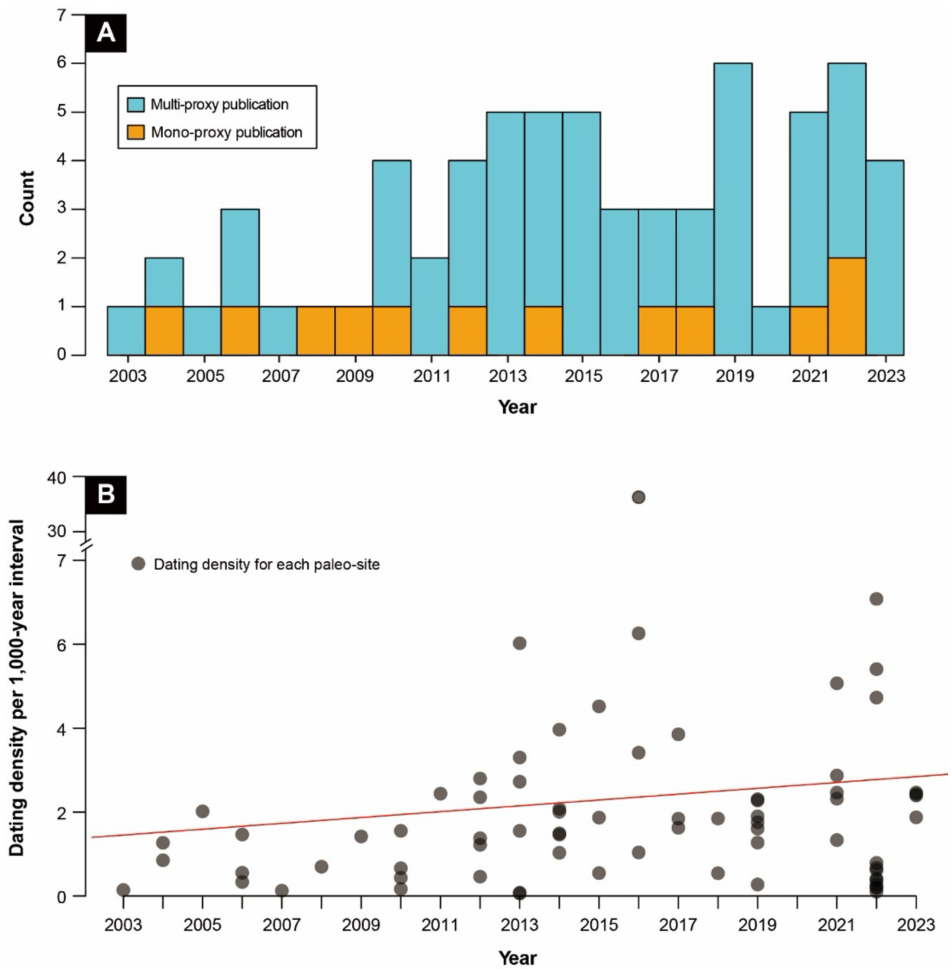


Figure 12. Temporal trends in (A) the proportion of multi-proxy to mono-proxy publications and (B) dating density per 1,000-year interval. In (A), 64 publications are shown as two articles studying surface pollen samples are excluded. In (B), the dating density, defined as the total number of dates by the difference between the oldest and youngest ages (unit: 1,000 years), is calculated for sediments from 72 paleo-sits. 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 Data 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. 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.xlsx has two worksheets, a) Chron-Depth dataset and b) Column description. Concurrently, the GeoEcoKorea webpage presents site locations on an interactive map at https://geoecokorea.org/2025_Site_Inventory. The

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

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 interactive markers with a pop-up displaying site-specific metadata and a direct download link to the corresponding Chron-Data Collection dataset (two worksheets: Readme and Chron-Depth) in .xlsx format.

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 in Northeast Asia where long-term environmental data are not widely accessible. While most records (~70 %) are covered only the Holocene, ~30 % records extend to 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. 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 reference data for future study site selections. Although the site-specific findings can be explored through the main text of related publications in international peer-reviewed journals, the 'interoperability and 'reusability' of original data (and thus citation of the paper) has been relatively rare in international collaborative research. For example, the most 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 pollen sites compiled in this dataset. Similarly, a recent data compilation study for pollen-based Northern Hemisphere paleoclimate reconstruction 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 framework developed for our datasets served as a useful tool for systematically characterizing site information and sample metadata of paleo-studies in South Korea, and following the descriptions in this paper, future efforts can readily adapt the framework for 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.

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

Competing interests

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.

References

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, 2018.

Davis, B. A. S., Chevalier, M., Sommer, P., Carter, V. A., 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., Latałowa, 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 Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions, Bioscience, 68, 563–576, 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., 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., 32, 1377–1394, https://doi.org/10.1111/geb.13693, 2023.

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, 2023.

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, 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.

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.

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.

Nahm WH.: 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, 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.

https://doi.org/10.5194/essd-2025-130
Preprint. Discussion started: 11 April 2025

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.

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.

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.