We sincerely thank both reviewers, Dr. John Williams and Dr. Jessica Blois, for their thoughtful, detailed, and constructive feedback. Their comments have greatly improved the clarity, usability and quality of our datasets, manuscript, and web-based platform.

To response in a structured manner, we have organized the comments into five categories:
1) Dataset, 2) Manuscript, 3) Table, 4) Figure, and 5) Webpage.
Each comment is indexed using a consistent format—for example, [JB-Dataset-01], where "JB" refers to Dr. Jessica Blois, "Dataset" indicates the category, and "01" denotes the comment number within that category. In this framework, *Manuscript* refers to the main text, while *Table* and *Figure* correspond to tables and figures included in the manuscript.

Reviewer comments are followed by our point-by-point responses, which are highlighted in blue for clarity.

RC1: 'Comment on essd-2025-130', John Williams

This is a useful summary of a major data compilation of paleoecological data in Korea. To my knowledge, this is the first such compilation of these data. The data compilation is somewhat small in volume and spatial scope, but it represents an important step forward in building a culture of data sharing and FAIR data in Korea. Papers like this are really helpful for showing the data now available and as a starting point for future community-building efforts, building open databases, and conducting large-scale scientific research.

1. **[JW-Dataset-01]** Of the three datasets built (see Fig. 2), only one of them has a detailed description of variable names, in Table 1. The other two datasets need a similarly detailed table. Ideally these two new tables would be put in the main text, but if there's not enough space, the two new tables could go into supplementary information. Specific comments about the data schema shown in Figure 2 and Table 1.

   Reply: We appreciate the reviewer's suggestion about the inclusion of descriptions for data fields of the other two datasets, Dataset I (Publication Metadata) and Dataset III (Geochronology Data), as tables in either the main text or supplementary documents. In fact, each of the two dataset files uploaded to Figshare already contains additional worksheets titled "Column description," which provide comprehensive details on all data fields, similar to what is presented in Table 1. We understand that this was not sufficiently stated in the original manuscript, which may have led to the impression that only Dataset II: Site Inventory (the core dataset of our study) included data schema and field definitions.

   To avoid redundancy between our manuscript and dataset files, we have decided not to reproduce the tables for data structures in Dataset I & III. However, we fully agree that the visibility of the worksheets must be improved.

   Thus, in our revised manuscript, we explicitly highlight the presence of the "Column description" worksheets and give clear guidance for ESSD readers to access them via Figshare. Last, we have also added a supplementary document that describes all data fields in each dataset.

1) **[JW-Dataset-02]** 'Chron-Depth Collection' is an awkward name… recommend naming to 'Age-Depth Data' or 'Geochronology Data'.

   Reply: As suggested, we have revised its name to "Geochronology Data" throughout the manuscript and datasets.

2) **[JW-Dataset-03]** The Chron-Depth Collection table lacks a field for the depth of each age control, which is essential data/metadata; for age-depth models to be built or rebuilt, each age control must have a depth and time coordinate.

   Reply: We appreciate this comment and would like to clarify that the table includes a field for the depth of each age control. This field is provided in the 7th column, labeled "Depth_cm."

3) **[JW-Dataset-04]** The Chron-Depth Collection table also should include a field indicating whether a date is in radiocarbon years or calendar/calibrated years. Ideally should also indicate whether the age datum is 1950 CE (which is the standard for radiocarbon years) or something else.

   Reply: As suggested, we have added a new field, labeled "Chronological_age_type," to the table in Figshare and each site's geochronology dataset downloadable through the interactive map (link) on GeoEcoKoera (GEK). This field specifies the timescale and age format in which the age is reported, using one of three values: Radiocarbon_BP (*The datum of radiocarbon years for present is specified to 1950 CE), OSL_Calendar_BP, or U-Th_Calendar_BP. The description of this field is provided in the 2nd worksheet of the Figshare table and the 3rd worksheet of each geochronology dataset on the GEK map.

4) **[JW-Dataset-05]** In Site Inventory, what does 'Year' store and is this information redundant with the 'Year' field in the Publication table? A standard principle of good database design is that each piece of information should be stored in only one location.

   Reply: As suggested, we have removed the Year field from the Site Inventory dataset, and another redundant field, Language. These modifications are reflected in the Entity-Relationship Diagram (ERD) in Figure 2 and Table 1.

5) **[JW-Dataset-06]** In the Site Inventory, it is a little strange, or at least not internally consistent to store age as Oldest, Youngest endpoints, while depth is stored as an interval. Ideally Depth would also be stored as Top, Bottom endpoints.

   Reply: We understand the reviewer's concerns about the inconsistency of endpoints between age and depth. However, age reversals are commonly found in geochronological data in settings such as South Korean lagoons, rivers, and estuaries. Depth endpoints do not always correspond to age endpoints. Thus, we have decided to report Oldest and Youngest ages based on age-depth models presented in the original publications.

For clarity for ESSD's board readership, we have added the following explanatory notes to Section 3.2.4 of our revised manuscript.

"*For paleo-sites, these two end values represent ages of the youngest and oldest age controls used in age-depth models in the original studies. The ages at the deepest and the shallowest depth points are not selected due to the frequent occurrence of age reversals within sediment sequences.*"

6) **[JW-Dataset-07]** Elevation (m): Elevation above sea level?

Reply: Yes. And as suggested, we have specified that elevation values are references to mean sea level throughout our manuscript's text, Table 1, and the worksheet of Column description in the Site Inventory dataset.

7) **[JW-Dataset-08]** Recommend renaming 'Specifications on environment' to 'Site Description'

Reply: As suggested, we have renamed the field to "Site Description" in Figure 2, Table 1, and the worksheet of Column description in the Site Inventory file.

8) **[JW-Dataset-09]** For Sample type, recommend renaming 'Surface pollen' to 'Surface sample' so that other kinds of surface data can be stored in the database.

9)

Reply: We agree that the "Sample type" field should have a certain flexibility to accommodate new types of modern samples. Rather than renaming "Surface Pollen," we revised our datasets to implement a two-tier classification of study sites. The field of "Type of study site" is moved from Dataset I: Publication Metadata to Dataset II: Site Inventory. The relocated field distinguishes paleo-sites from surface sites. Then, the "Sample type" field allows further categorization, Core and Trench under paleo-site and Surface Pollen for surface site. This new framework retains "Surface Pollen," while allowing flexible entries of new sample types, for example Speleothem under paleo-sites and Surface Diatom under surface site.

10) **[JW-Dataset-10]** For Sample type, note that what this database calls 'Trench', Neotoma calls 'Section'. 'Section' is a little more general, because 'Trench' implies a dug trench, while 'Section' can be any stratigraphic section or outcrop, e.g. including trenches, riverbanks, or rock outcrops. See also comment on L179 below.

Reply: We appreciate the reviewer's suggestion and the reference to Neotoma's terminology. We agree that "Section" is a more general term, and we will consider adopting it when contributing our curating data into global platforms such as Neotoma. However, we have decided to retain the term "Trench" in the datasets in this study and the GEK (GeoEcoKorea) database to reflect its common usage in the Korean Quaternary

research community. Many paleo-sites in South Korea have been investigated through trenching, either as part of geoarcheological research or as stratigraphic exposures uncovered during land development projects. To maintain consistency with local research practices, we have added a clarification in the revised manuscript.

* I have copied this as the reply to the comments on L179.

11) **[JW-Manuscript Table-01]** Could probably rename 'Written language' as 'Language'. Also, there seems to be a confusion here, because 'Language' in Fig. 1 is part of the Publication table, but Table 1 is listed as describing the variables in the Site table.

Reply: As noted in the reviewer's comment, **[JW-Dataset-05]**, it was recommended to avoid the redundancy of fields in multiple datasets. In response, we have omitted Year and Language fields from the Site Inventory dataset. This resolves the confusion regarding the name of 'Written language'. These changes have been reflected in Table 1, Site Inventory dataset, and Figure 2.

12) **[JW-Table-02]** Table 1 lists a category called 'Proxy' and fields including Pollen, Diatom, etc. However, these fields aren't shown in Figure 2's entity relationship diagram. In general, Table 1 and Figure 2 should be checked to ensure that there is a 1:1 match between all variable names shown in each location.

Reply: As suggested, we have updated that the three 'Proxy' fields (Pollen, Diatom, and Other proxies) are included in the table of Site Inventory in Figure 2.

2. **[JW-Manuscript-01]** The paper needs to be very careful in how it handles and reports ages, because it is using at least three different time scales, creating major potential for confusion. First, it uses the standard AD/BC (or CE/BCE) timescale for information such as publication year. Second, it uses the radiocarbon timescale when reporting radiocarbon dates, which are expressed in years before present, uses 1950 CE as 'present' and has a non-linear relationship to the calendar-year timescale. Third, the paper uses the calendar-year timescale for OSL and U/Th dates, which also are expressed in years before present and can use various years as 'present'.

    1) For all variables that store year, the paper should clearly establish what time scale is being used.
    2) For all geochronological datasets, the dataset should formally store the datum used for 'present' or the paper should establish that in all cases, 'present' is treated as 1950 CE.
    3) For all radiocarbon dates, they should be reported in their original radiocarbon-year dates with 1sigma error. The use of the radiocarbon timescale should be explicitly noted as a metadata field. The paper optionally could also include calendar-year conversions
    4) for each radiocarbon date as additional fields, along with information about the calibration curve and program used to convert from radiocarbon years to calendar years.

5) **[JW-Dataset-11]** For all other geochronological dates, the use of the calendar-year timescale should be explicitly noted as a metadata field.

Reply: We appreciate the reviewer's constructive feedback about the importance of clarifying time scales of our datasets and manuscript. In response, we have been updated both for consistency and transparency in the presentation of time scales.

The summary of updates in Dataset II: Site Inventory and Dataset III: Geochronological Data follows as:

Dataset III is intended to contain chronological data without calibration and conversion. A new field, "Chrono_age_type," has been added to specify different time scales and formats of core top ages and those used for ^{14}C, OSL, and U-Th dates. This field indicated whether the reported age is expressed:

- Calendar_BP: calibrated calendar years before present, where "present" is defined as 1950 CE

- Radiocarbon_BP: uncalibrated radiocarbon years before present, , where "present" is defined as 1950 CE

- OSL_yrs_BP: years before the measurement year, based on Optically Stimulated Luminescence dating

- U-Th_yrs_BP: years before present, where "present" is defined as 1950 CE

In Dataset II, we have standardized two age endpoints (Youngest and Oldest ages) as calendar years before present, using 1950 CE as the reference point. For radiocarbon dates, the upper and lower bounds of the 95% calibrated age range using the IntCal20 and Marin20 calibration curves are used for the two ends. For OSL and U-Th dates, the upper (Youngest) and lower (Oldest) bounds of the 1σ-error range are selected. OSL-based ages are converted to years before present (=1950 CE), while U-Th dating ages are not changed as their reference year is defined as 1950 CE (Dutton et al., 2017).

[Reference for U-Th dates] Dutton et al. (2017) 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.

In addition, a new field, "Age scale," has been added to specify the origin and method of driving these calendar-based ages. This field is formatted like:

"Oldest: yrs_BP (OSL converted); Youngest: yrs_BP (C-14 calibrated with the Marine20)"

The above chronological data processing is described in our revised manuscript. The relevant passage in section 3.2.5 is:

"*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."

3. **[JW-Manuscript-02]** In the Summary, this paper does a good job of describing how this site inventory could be used for future work. That noted, I think the paper would be strengthened if its closing provided more of a forward-looking vision and potential future directions, in at least three ways. Papers like this can set the tone for a community and inspire others.

    1) **Communicate more of a scientific vision** (this could be done in both Intro and Conclusions/Summary). What kinds of research questions can be enabled when large open paleoecological databases are available? Cite more of the recent papers that have used large paleoecological datasets and discuss how these approaches could be used to advance questions in Korean paleoecology

    2) I know that the politics of data sharing and openness can be sensitive, but I think it would still be good to add **a gentle call to authors to contribute their data and thereby further enhance the database by obtaining and adding the actual proxy datasets.** This call follows naturally from the scientific vision… what scientific research questions are enabled by the inventory in its current form? What new questions would be possible if all datasets were also available?

    3) It would also be good to talk about **the longer-term vision / plan for the GEK**. One option is to keep it as a standalone database; another would be for it to join Neotoma as a Constituent Database. There are pros and cons to each option. A standalone database allows for more autonomy and local control but raises questions of sustainability; small databases find it hard to persist over time. You could point to other regional / national efforts such as the LAPD, APD, and Chinese Pollen Database as various options. The first two have been merged into Neotoma and the latter has remained a standalone database. The Japanese Pollen Database is somewhat in between, with some but not all of these data in Neotoma. You could discuss options without reaching a firm decision on which to pursue, so as to leave options open. Could also here talk about the importance of data governance and making sure that people feel like they have a voice in how their data are curated… this could e.g. lead to a working group, South Korean Pollen

Council, or other group… describe options with invitations for others to join this effort. Williams et al. (2018) provides a good description of Neotoma data governance. Note of course that the authors have a much better sense of academic culture and norms in South Korea, so are best positioned to figure out how best to gently advance community-oriented data governance, while not pushing too hard.

Reply: We appreciate these thoughtful and forward-looking comments. In response, we have elaborated and expanded our introduction (Section 1) and conclusion (Section 5).

1) Scientific vision: We have clarified the types of scientific questions that can be addressed using the compiled dataset, such as investigating ecosystem responses to Holocene climate variability and human activity across diverse depositional settings. In both the Introduction and Section 5, we have highlighted how Korean paleoecological data—when integrated across proxies and sites—can support regionally relevant syntheses and contribute to broader understanding of past environmental change in Northeast Asia. In addition, we have also noted its potential to help fill spatial gaps in global-scale synthesis efforts.

However, as our manuscript is a data paper, we have focused on reusing our current datasets, but slightly mention prospective regional- and global-scale synthesis questions.

2 & 3) Data contribution call & Long-term vision and governance: In Section 5, we have emphasized GEK's role in fostering a collaborative data-sharing culture and outlined both our long-term vision (supporting a community-driven Korean paleoecology database) and short-term plan (applying to become a Neotoma's constituent database). We also included a gentle—though perhaps subtle—call for Korean researchers to contribute their datasets to enhance collaboration and improve the visibility of Korean paleoecology in global syntheses.

See line-by-line comments below and attached PDF for more detailed comments. My edits to the PDF are fairly extensive but are mainly focused on readability and brevity. I've tried to keep substantive changes out of the PDF edits, but have flagged below in line-by-line comments any wording edits that might create a substantive change in tone or emphasis.

LINE-BY-LINE COMMENTS

[JW-Manuscript-03] L24-25: This closing sentence in the abstract about data sharing initiatives is good, but it would also be good to talk about scientific vision and future questions / opportunities that are being enabled.

Reply: We have added forward-looking perspectives that articulates GEK's scientific vision. The updated sentence highlights the potential of the database to support regional synthesis, enable cross-site comparisons of Holocene ecosystem dynamics, and contribute Korean paleoecological data to global-scale reconstructions. It also emphasizes the promotion of a collaborative and inclusive data-sharing culture to facilitate future research opportunities.

[JW-Manuscript-04] L50: Should also add mention of open-source statistical packages, e.g. those developed by Mottl et al for rate-of-change analyses or Simpson for hierarchical GAMs.

Reply: As suggested, three R packages—neotoma2 for acquiring data directly from Neotoma and R-Fossilpol for pressing pollen datasets into standardized and reproducible format, and R-Ratepol for estimating the rate of compositional change within palynological assemblages—have been added into the introduction of our revised manuscript.

[JW-Manuscript-05] L53 and elsewhere: note that 'assumptions' is often being used when 'analyses' or other wording would be better. Usually a scientist makes assumptions in the absence of data… the assumptions are precursors to formulation of hypotheses and the analysis of data. Much of the paper is focusing on this later stage of hypothesis testing and analysis.

Reply: As suggested, we have edited "different interpretive assumptions" to "different interpretive approaches" in the sentence of the revised manuscript.

[JW-Manuscript-06] L62: See edits to PDF, have revised text to clarify how regional databases can affiliate with Neotoma and advantages.

Reply: As suggested, we have edited the text according to the edits provided in the PDF.

[JW-Text-07] L64: See proposed rewording

Reply: As suggested, we have reworded the text according to the proposed revision.

[JW-Figure-01] L85 and Fig 1:

*Clarify what the blue bands represent, i.e. they are intended to highlight Korea and the absence of data

*Possibly could truncate the map and associated panel B at 65S and 85N given the absence of sites at these latitudes

Reply: As suggested, we have added a note in the caption of Figure 1 clarifying the meaning of the blue bands, which states: "*Blue band: The blue band indicates the Korean Peninsula and spatial gaps in site coverage.*" However, we decided not to truncate the map and associated panel B at 65°S and 85°N, as the land area per 1° longitudinal band shown in Figure 1C is calculated as an average across the full -90° to 90° latitude range.

[JW-Figure-02] L115 and Fig 2:

*typo in the second table: 'depositionoanal'

Reply: As suggested, we have corrected "depositionoanal" to "depositional" in the second table of Figure 2.

[JW-Manuscript-08 & Dataset-12] L134-139: This paragraph needs to be written for a broad ESSD audience, who won't e.g. know what the space-for-time substitution method is. Could cite Chevalier et

al. (2020). Also ‘paleo-site’ here and elsewhere reads as jargony, suggest instead ‘fossil site’ or ‘paleoecological record’. And recommend here and elsewhere ‘surface sample’ instead of ‘surface site’ or ‘surface pollen’.

Reply: As suggested, we have revised all instances of ‘Others’ to ‘Other’ in Figure 7, Table 1 and Site Inventory dataset, and throughout the manuscript text.

[JW-Manuscript-09] L150: What is meant by ‘distinct geographic coordinates’?

Reply: The relevant sentence has been revised to: “*Each site is georeferenced using geographic coordinates and elevation.*”

[JW-Manuscript-10] L164 and elsewhere: Avoid using a hyphen to indicate numeric ranges, because it can be easily confused for a negative sign. Rewrite here to ’51 to 1,305 m’

Reply: As suggested, we have revised all relevant instances of numeric ranges in the manuscript text.

[JW-Manuscript-11 & Dataset-13] Fig 6 / L215 and elsewhere: Change ‘Others’ to ‘Other’ when using it as a categorical.

Reply: As suggested, we have revised all instances of ‘Others’ to ‘Other’ in Figure 7, Table 1 and Site Inventory dataset, and throughout the manuscript text.

[JW-Dataset-15] L172: Note that what the paper calls a sample (surface pollen, core, trench), Neotoma calls a collection unit. In Neotoma’s naming system, a site (such as a lake) can have multiple collection units (e.g. multiple cores per lake) and multiple samples per core, distributed by depth. Understood that this paper’s data schema doesn’t have to closely align with Neotoma… still, aligning language now will avoid much confusion later.

Reply: We have noted this distinction and will take it into account when revising our terminology or integrating with Neotoma-compatible frameworks in future updates.

[JW-Dataset-14] L179: Given these (useful) definitions, I’d argue that ‘Section’ or ‘Stratigraphic Section’ should be used instead of ‘Trench’. ‘Trench’ usually implies a ditch dug by human action. The manuscript describes instead naturally occurring outcrops, which aren’t trenches.

Reply: We appreciate the reviewer’s suggestion and the reference to Neotoma’s terminology. We agree that “Section” is a more general term, and we will consider adopting it when contributing our curating data into global platforms such as Neotoma. However, we have decided to retain the term “Trench” in the datasets in this study and the GEK (GeoEcoKorea) database to reflect its common usage in the Korean Quaternary research community. Many paleo-sites in South Korea have been investigated through trenching, either as part of geoarcheological research or as stratigraphic exposures uncovered during land development projects. To maintain consistency with local research practices, we have added a clarification in the revised manuscript.

* I have copied the reply to the comments of [JW-Dataset-10].

[JW-Manuscript-12] L191 and elsewhere: edit 'Open-coastal zone' to 'Open Coastal Zone' and edit 'Volcanic cone' to 'Volcanic Cone', capitalizing these names because you are proposing them as formally named categories.

Reply: As suggested, we have revised the text to capitalize the depositional environment categories as forma name through our manuscript and datasets.

[JW-Manuscript-13] L191-197: Can remove all quotation marks, because the capitalized names will be enough to establish these as defined entities.

Reply: As suggested, we have removed the quotation marks of the capitalized category in the manuscript.

[JW-Dataset-15] L200: Given that so many samples are soil samples, recommend adding 'Soil' as a category and reducing the scope of 'Other'.

Reply: We appreciate the reviewer's comment. Soil samples are relatively overrepresented in the current dataset of Site Inventory, but this may change in future updates. To maintain flexible classification scheme, we have decided to retain soil samples within "Other" category.

[JW-Manuscript-14] L213/Fig. 7: Typo 'coatal'

Reply: We have corrected the typo to "coastal" in Figure 7.

[JW-Dataset-16] L222: Entering ages and depths as zero for the surface samples is not a good idea, because no measurements actually exist. Setting ages to zero is a particular problem, because in the radiocarbon timescale, 0 = 1950 CE, which would be an incorrect age for these samples. For depth, recommend leaving blank, and for age, recommend using the sampling year.

Reply: This comment is addressed in our response to [JW-Dataset-11].

[JW-Dataset-17 & Manuscript-15] L222-223: Somewhere in this paper, state explicitly a) that you are using the radiocarbon timescale for all ages and b) remind ESSD reader that on this timescale, 1950 CE = 0. Also, because some time-related variables use the AD/BC calendar year (e.g. for publications) while others us the radiocarbon timescale (e.g. for ages), be very explicit about which systems are being used for which variables.

Reply: This comment is addressed in our response to [JW-Dataset-11].

[JW-Figure-03] L226: The statement that 96% of sites use only 14C dating doesn't appear to match Fig. 8B, where the proportion looks more like 85%.

Reply: We thank the reviewer for the constructive comment. In response, the format of Figure 8B has been revised: instead of a stacked bar chart, we now present a histogram that categorizes paleo-sites by dating methods—^{14}C only, ^{14}C–OSL, OSL only, and ^{14}C–U-Th. The purpose of this revised figure is to emphasize the predominance of sites dated exclusively using ^{14}C. Accordingly, the associated text has been revised to highlight the dominance of ^{14}C-only chronologies, rather than stating a specific percentage, which states: "*Among these absolute age controls, 14C dating has been exclusively used to construct the geochronology of sediments at 55 out of 72 paleo-sites (Figure 8B).*"

[JW-Figure-04] L230 / Fig 8: I don't really understand the histogram plot shown in Fig. 8B. How do the three histograms differ from each other? They appear to contain overlapping information.

Reply: This comment is addressed in our response to [JW-Figure-03].

[JW-Manuscript-16] L236-237: I don't understand what this parenthetical about -73 years is trying to convey.

Reply: This comment is addressed in our response to [JW-Dataset-11].

[JW-Figure-05] L238 & Fig. 9: Recommend renaming 'last glacial-interglacial cycle' to 'Pleistocene and Holocene'. The last glacial-interglacial cycle is usually reserved for sites that span the last 100,000 years, i.e. a full glacial-interglacial cycle. This renaming would also better align the other two categories 'Pleistocene-only' and 'Holocene-only'

Reply: We have revised commented parts with errors in Figure 9A to E. First, we have renamed the last glacial-interglacial cycle to "Pleistocene–Holocene," which is abbreviated as "P–H." In Figure 9B, we have updated the horizonal bar plot colors, x-axis title (calendar kyr ago, in the caption, adding a note that kyr = 1,000 years), and adding a vertical line of the beginning of the Holocene. Figure 9D and E, the revised boxplots have no scale breaks and proper titles (Number of age controls per 1 cm depth interval and 1,000-year interval) of x-axes.

[JW-Figure-06] L275 & Fig 9:

*typo: 'Interglcacial'

*x-axis title in Panel B is incorrect, it should indicate thousands of years ago, not years ago

*Panel D: Check the math used to calculate the number of dates per 1 cm depth interval. The reported values for the Pleistocene sites (anywhere from 1 date per cm to 2.5 dates per cm) is unimaginably high.

*The use of multiple scale breaks in the x-axis is non-traditional. Recommend either just one scale break or perhaps experimenting with a log scale.

Reply: This comment is addressed in our response to [JW-Figure-05].

[JW-Figure-07] L280 & Fig 10:

*Panel D legend: delete word 'record'

*Panel D: the colors in the legend don't appear to match the colors used in the map

Reply: As suggested, we have revised Panel D's legend by removing 'record' of Pollen record and correcting the colors of blocks corresponding to the grid blocks in the maps.

[JW-Manuscript-17] L292-293: Per comments above, because radiocarbon dates and U/Th use different timescales,

Reply: This comment is addressed in our response to [JW-Figure-08].

[JW- Figure-08] L300 and Fig. 11:

*What does 'pMC' stand for?

*1950 what? CE?

*Note that this figure does a good job of clearly differentiating calendar years vs. 14C years by indicating 14C kyr BP for some plots and kyr BP for others. This model should be adopted throughout ms. Be sure to define kyr abbreviation somewhere.

Reply: As suggested, we have updated Figure 11 and its caption. In Panel A, pMC is explained as dates after 1950 CE. In the caption, we have added the definition of kyr as 1,000 years before preset, and a note: "*Note: The reference year for "present" is 1950 CE for radiocarbon and U–Th dates, and the measurement year for OSL dates."*

[JW-Figure-09] *L311: Please report the Pearsons correlation coefficient and significance test for this regression.

Reply: As suggested. we have now added the coefficient, p-value, R2 and regression equation in Figure 12's caption.

RC2: 'Comment on essd-2025-130', Jessica L. Blois

This paper and dataset compiles and examines records of Korean Quaternary paleoenvironmental proxy records and associated modern datasets. The authors performed a literature search to locate potential records, imposed some light quality controls to filter out some records, then assembled a database of metadata about the datasets. They then examined the dataset for different measures of comprehensiveness and bias, related to different paleoecological and geographic attributes. Overall, the authors have compiled a very nice dataset that highlights, and then starts to solve, a gap in data accessibility for Korean Quaternary data. As noted by the authors, there are currently no dataset from Korea in the Neotoma Paleoecology Database and the region is underrepresented in global syntheses, despite there being a lot of good work on Korean Quaternary paleoenvironments in the literature. This paper represents a strong starting point for making Quaternary data from Korea more accessible.

I have four primary comments, and then some minor comments.

[JB-Manuscript-01] 1. I would suggest adding references to the first two paragraphs of the paper, to bolster the literature basis of this introduction and help point readers from outside the subdiscipline to relevant literature. As the introduction stands, the authors make a number of assertions about the benefits of making data FAIR, but it would be very helpful to ground these assertions in specific examples or cases where compiling data across sites, or across proxy types, led to better or different inferences than a single site study alone.

Reply: We thank the reviewer for the helpful suggestion. In response, we have added key references to the first three paragraphs to strengthen the literature basis and guide readers unfamiliar with the subdiscipline. We also included specific examples from recent studies demonstrating how multi-proxy and multi-site syntheses, supported by FAIR data practices, have led to broader ecological insights.

The list of literatures newly cited is:

For multi-proxy approaches:

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

- Mann, M. E.: The Value of Multiple Proxies, Science, 297, 1481–1482, https://doi.org/10.1126/science.1074318, 2002.

For multi-site syntheses:

- Gaillard et al., 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.

- Kaufman et al., A global database of Holocene paleotemperature records, Scientific Data, 7, 1–34, https://doi.org/10.1038/s41597-020-0445-3, 2020.

- Mottl et al., 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.

[JB-Website-01] 2. Data accessibility:

The abstract says "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." The relationship between the database and the viewer seems clear on the surface, but when digging more into the viewer (see comment #3) and other parts of the text, there appear to be some points of disconnect.

Specifically, the GeoEcoKorea viewer is stated to provide access to archived raw data in some cases. But I don't see these links in either the database description or in the viewer. Database description: Is there somewhere in the three linked tables that provides links to the raw data or a field/variable storing the link? The ERD in Figure 2 does not appear to include any fields that are linked to an original dataset (maybe through DOI in the Publication Metadata table) or linked to the raw data. GeoEcoKorea viewer: the description of GeoEcoKorea in section '4, Data Availability', does not mention links to raw data.

Perhaps this is a place where "raw data" needs to be defined? I took this phrase to mean the base proxy data, like pollen or diatom counts, but perhaps this simply refers to the chron-depth dataset specifically, which is provided in some cases? Regardless, some additional clarity here would be helpful.

Reply: We understand the ambiguity round the meaning of "raw data" and of how access to both "raw geochronological data" and "bibliographic information" is implemented in the GeoEcoKorea map viewer. To clarify this, we have revised the abstract to define "raw data" as uncalibrated geochronological datasets. The revised sentence in the abstract is:

"*To enhance accessibility, we have developed GeoEcoKorea, an open-access platform featuring interactive maps where each site marker displays site-level metadata and links to bibliographic information and uncalibrated geochronological datasets, when available.*"

We have also added clearer explanations in both the manuscript and the GeoEcoKorea webpage regarding how each site marker in the GeoEcoKorea viewer to the corresponding datasets. In Section 4 of the revised manuscript, the explanation is:

"*Each site is represented by an interactive marker that opens a pop-up window displaying site-specific metadata, along with direct links to the corresponding bibliographic information and a downloadable Geochronology Data (three worksheets: Readme, Geochronological data, and Column description) in .xlsx format*".

In addition, for GEK's viewer (Figure RC1), we have updated the text of the left side of the viewer to clarify how to link to publication and geochronological data. The revised text is:

"*This interactive map displays the locations of the study sites introduced in our manuscript (currently under review; [Link to Preprint]). The map integrates bibliographic information, site information, proxies, and geochronological datasets. Each site is marked with a colored circle, which reveals site-specific metadata.* ***Clicking a site marker*** *opens a pop-up window displaying site-level metadata. A downloadable tabular* ***geochronological dataset is available via the "Download File" link in each pop-up****. Each entry also includes* ***a PublicationID (e.g., Pub-052); clicking the ID reveals the full citation information***".

3. GeoEcoKorea viewer

a. First, building the viewer in addition to the database is a huge effort. The authors have developed a nice, simple display that is easy to navigate and understand.

[JB-Website-02] b. The authors might consider coloring the markers differently between the surface sample and sedimentary records, for ease of location and navigation.

Reply: As suggested, we have assigned different colored markers for paleo-sites (blue) and surface pollen sites (orange) on the map (Link to the map) (Figure RC1). Additionally, a new category has been created for the 9 paleo-sites having datasets archived in Neotoma as a part of GEK's early efforts to join as a constituent database of Neotoma (the upload was completed after the submission of this manuscript). Each category is now clearly defined and can be viewed on the map, both separately and collectively.

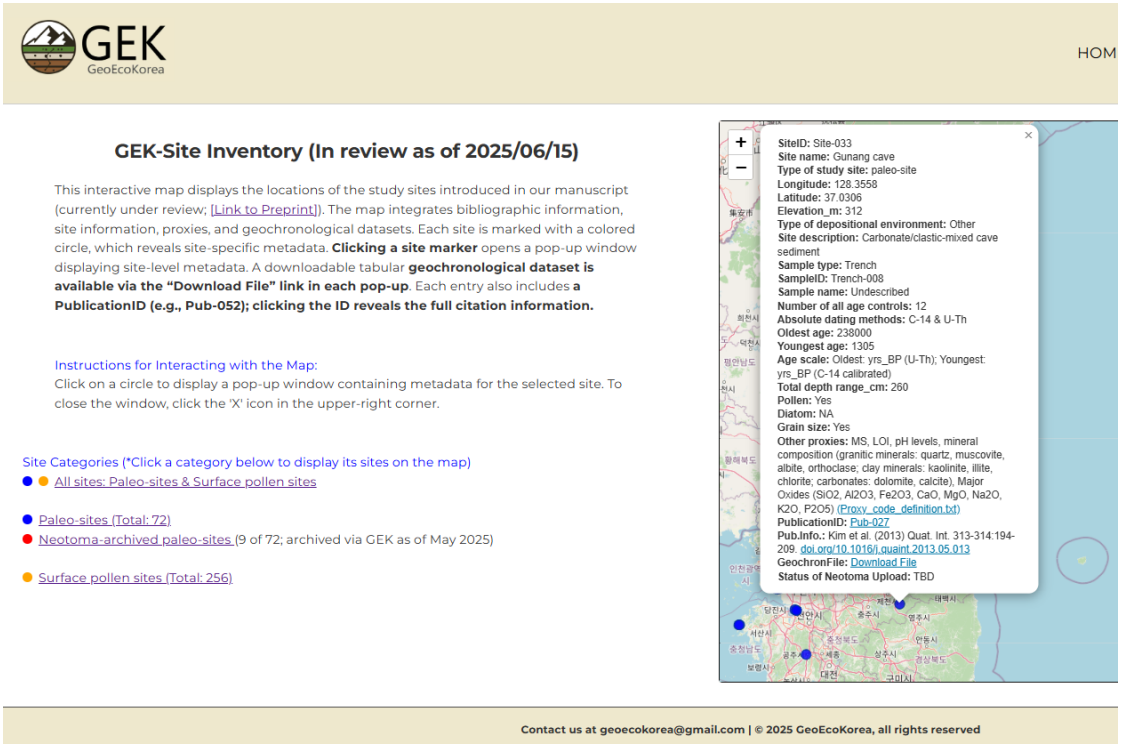

Figure RC1. GEK's Updated Map Viewer (https://geoecokorea.org/)

[JB-Website-03] c. Related to my comment about data accessibility (comment #2), in GeoEcoKorea on page https://geoecokorea.org/itnl_pubs, there is a statement that "International data must first be uploaded to the Neotoma DB, after which GEK will establish a link to it." Where would this link be stored in the database? I don't see any specific fields. Are there any links currently established? If so, you could perhaps highlight one of them.

Reply: This comment is mostly addressed in our response to [JB-Website-01 & 02].

Our site-level metadata are used as a foundation for us to contact lead authors and encourage submissions of their proxy datasets (e.g., pollen, charcoal, and geochronological datasets) to Neotoma via GEK. Currently, datasets from 9 sites listed in the inventory have been successfully uploaded in Neotoma.

To reflect these updates, we have added a field titled "Status of Neotoma Upload" to the map interface. When a user clicks on a site marker, a pop-up window appears displaying site-level metadata, including a hyperlinked DOI to the corresponding Neotoma dataset (when available). In addition, GEK's relevant webpages, including the page https://geoecokorea.org/itnl_pubs of the comment, have displayed links to isolate and view sites with datasets currently archived in Neotoma.

4. Links to Neotoma Paleoecology Database.

- For transparency, I want to note that I am the Associate Chair of the Neotoma Paleoecology Database (and am very supportive of this effort to make Korean data more visible!).
- **[JB-Manuscript-02]** In the Introduction, the authors mention the Neotoma database and it is also mentioned in GeoEcoKorea (see #3c). For example, the Introduction says "The Neotoma community also helps regional databases independently manage their own servers and affiliate with Neotoma for global networking." However, then in section '5 Summary and Dataset Reuse Potential', the future plans for the Korean database do not mention any affiliation with Neotoma, though it's clear from browsing the GeoEcoKorea that they are planned (see item c in my comments about the viewer). It may be helpful to elaborate in Section 5 or elsewhere on how the authors envision (or actually are) linking to Neotoma.

  Reply: As suggested, our revised manuscript has mentioned GEK's ongoing efforts (uploading proxy datasets from 9 Korean sites to Neotoma) and near future plan, becoming a Neotoma's constituent database. The relevant passage is:

  "*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 essential 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 fuel our regional initiatives to encourage and support Korean Quaternary researchers to contribute their data, thereby fostering a more collaborative and inclusive data-sharing culture.*"

- **[JB-Dataset-01, Website-04]** While this is absolutely not necessary for publication, if the plan is to continue enhancing links with Neotoma, consider adding a table that provides a cross-walk between the variables stored in your database and the variable names in Neotoma and providing that on the GeoEcoKorea website and/or in the FigShare supporting files. This could help enhance inclusion of Korean data into syntheses that use Neotoma for the majority of their data searches.

  Reply: This comment is mostly addressed in our response to [JB-Website-01 to 03 & JB-Manuscript-02].

Minor comments:

- **[JB-Figure-01]** Figure 6. There is a note here that "The location of Jeju Island is shown in Figure 3A." But I think the figure number needs to be updated – I think Jenu Island is shown first in Figure 4.

  Reply: As suggested, we have updated the note, *"Jeju Island (Figure 4A for location)"*, in the legend of Figure 6.

- **[JB-Figure-02]** Figure 9:
  - In panel B, the color of the horizontal lines seems to vary slightly – different shades of red, different shades of blue. This may be purely an individual color screen issue! But if there is indeed color shading, then perhaps figure out how to more clearly differentiate some of the blue shades from the black/gray 'Pleistocene-only' sites.

    Reply: As suggested, we have updated the horizonal bar plot colors in Figure 9B. If the slight variation in shades persists, it may be due to screen or resolution differences. However, these distinctions become clearer when the figure is viewed at a larger scale.

  - I had a difficult time interpreting the LIC acronym in panel D. I originally thought it meant "last interglacial", but then realized it essentially means "encompassing portions of both the Pleistocene and the Holocene". And also, in examining panel B, it's clear that at least some sites extend into the penultimate glacial-interglacial transition as well. So perhaps change this to "late Quaternary"?

    Reply: We have revised commented parts with errors in Figure 9. The last glacial-interglacial cycle has been renamed as "Pleistocene–Holocene," which is abbreviated as "P–H.".

- **[JB-Figure-03]** Figure 11: I would define in the caption what 'pMC' means, noted in red in panel A.

  Reply: As suggested, we have updated Figure 11 and its caption. In Panel A, pMC is explained as dates after 1950 CE.

- **[JB-Dataset-02]** In the FigShare files, the "Publication Metadata" file does not contain a tab with the Column descriptions, whereas the other two files do. I would add this to the Publication Metadata file.

  Reply: As suggested, we have added the Column description worksheet to the Publication Metadata file.