28 August 2026

LinkedEarth: a research ecosystem for open and reproducible paleoclimatology

Deborah Khider[1], Julien Emile-Geay[2], Varun Ratnakar[3], Nicholas P. McKay[4], Jordan P. Landers[5], Yolanda Gil[6], Daniel Garijo[7], David Edge[8], Cody C. Routson[9], Tanaya Gondhalekar[10], Feng Zhu[11]

1. University of Southern California Information Sciences Institute
2. University of Southern California
3. University of Southern California
4. Northern Arizona University
5. University of Southern California
6. University of Southern California
7. Universidad Politécnica de Madrid
8. Northern Arizona University
9. Northern Arizona University
10. University of Southern California
11. National Center for Atmospheric Research

Abstract

Paleoclimate research increasingly relies on the synthesis of heterogeneous proxy records to investigate climate variability across spatial and temporal scales. However, the reuse and integration of paleoclimate data remain limited by inconsistent formats, fragmented analytical workflows, divergent nomenclature and practices, and implicit methodological assumptions. Improved open data access, while helpful, is insufficient to ensure reproducibility and enable large-scale scientific discovery on its own. Here, we present an integrated framework for advancing open and reproducible paleoclimatology through three complementary developments. First, we build on community standards to harmonize the representation of paleoclimate data and metadata, connecting structured data formats with controlled vocabularies via an ontology. Second, we implement these semantic relationships within a knowledge graph, which integrates data and metadata into a unified, queryable structure that supports consistent access across heterogeneous datasets. Third, we introduce a curated gallery of reproducible and executable workflows that document complete analytical pipelines from data discovery to interpretation. Representative examples illustrate how these three pillars can be combined to enable transparent, efficient, and scalable paleoclimate analyses. This integrated approach provides the infrastructure necessary to support large-scale synthesis and reproducible research, laying the foundation for both human- and machine-assisted scientific discovery in paleoclimatology.

Keywords

Open and Reproducible Science, Paleoclimatology, FAIR principles, Open Data, Open Software

LinkedEarth: a research ecosystem for open and reproducible paleoclimatology

Deborah Khider[1], Julien Emile-Geay[2], Varun Ratnakar[1], Nicholas McKay[3], Jordan Landers[2], Yolanda Gil[1], Daniel Garijo[4,1], David Edge[3], Cody Routson[3], Tanaya Gondhalekar[2], Feng Zhu[5]

[1]Information Sciences Institute, University of Southern California
[2]Department of Earth Sciences, University of Southern California
[3]School of Earth and Sustainability, Northern Arizona University
[4]Universidad Politécnica de Madrid
[5]NSF National Center for Atmospheric Research

Key Points:

- We present a research ecosystem to enable open and reproducible paleoclimatology, based on data standards, open software and workflows
- Knowledge graphs integrate data, metadata, and semantic structure to support complex queries and consistent data access.
- Open software and reproducible workflows can accelerate discovery and transparently document methods and assumptions.

Corresponding author: Deborah Khider, `khider@usc.edu`

Abstract

Paleoclimate research increasingly relies on the synthesis of heterogeneous proxy records to investigate climate variability across spatial and temporal scales. However, the reuse and integration of paleoclimate data remain limited by inconsistent formats, fragmented analytical workflows, divergent nomenclature and practices, and implicit methodological assumptions. Improved open data access, while helpful, is insufficient to ensure reproducibility and enable large-scale scientific discovery on its own. Here, we present an integrated framework for advancing open and reproducible paleoclimatology through three complementary developments. First, we build on community standards to harmonize the representation of paleoclimate data and metadata, connecting structured data formats with controlled vocabularies via an ontology. Second, we implement these semantic relationships within a knowledge graph, which integrates data and metadata into a unified, queryable structure that supports consistent access across heterogeneous datasets. Third, we introduce a curated gallery of reproducible and executable workflows that document complete analytical pipelines from data discovery to interpretation. Representative examples illustrate how these three pillars can be combined to enable transparent, efficient, and scalable paleoclimate analyses. This integrated approach provides the infrastructure necessary to support large-scale synthesis and reproducible research, laying the foundation for both human- and machine-assisted scientific discovery in paleoclimatology.

Plain Language Summary

Scientists study past climates using natural archives such as tree rings, ice cores, and ocean sediments. Combining these records into a unified picture is difficult because each dataset is stored, described, and analyzed differently, often without transparent or repeatable methods. We present a framework that standardizes paleoclimate data organization, integrates them into a knowledge graph for enhanced searchability, and provides open computational notebooks documenting each analysis step. We demonstrate how this approach enables climate reconstruction across time periods, temperature pattern analysis, and model comparison. By improving data discoverability and method transparency, we facilitate more efficient scientific collaboration and knowledge reuse. These advances may enable computational assistance for paleoclimate data analysis, accelerating discovery.

1 Introduction

Modern paleoclimatology increasingly depends on synthesizing and comparing records across space and time, and with climate model output. Advances in proxy development, analytical techniques, and data archiving have produced a rich and expanding body of paleoclimate observations spanning diverse environmental settings and temporal resolutions (Lisiecki & Raymo, 2005; Cook et al., 2009; Shakun et al., 2012; Lin et al., 2014a; Snyder, 2016; Emile-Geay et al., 2017; Neukom et al., 2019; Routson et al., 2019; Konecky et al., 2020; Jonkers et al., 2020; Kaufman et al., 2020; Tierney, Zhu, et al., 2020; Osman et al., 2021; Walter et al., 2023; Clark et al., 2024; Evans et al., 2025). However, individual proxy records primarily reflect local environmental conditions and archive-specific processes, making it difficult to infer large-scale climate behavior from a single record alone (Cook, 1987; von Storch et al., 2004; Evans et al., 2013; Dolman et al., 2021; McPartland et al., 2025). Large-scale synthesis of proxy records across multiple proxy systems, regions, and timescales is therefore essential for addressing central paleoclimate questions: quantifying the magnitude and spatiotemporal patterns of natural climate variability, understanding the response of the climate system to internal and external forcings, and determining what aspects of modern climate change are unprecedented. By integrating independent lines of evidence, multiproxy synthesis enables the identification of coherent climate signals (Abram et al., 2016; Emile-Geay et al., 2017; Kaufman et al., 2020), improves the robustness of inferred variability (Neukom et al., 2019), helps constrain climate sensitivity (Hargreaves et al., 2012; Sherwood et al., 2020; Cropper et al., 2023; Cooper et al., 2026), and provides the spatial context necessary for investigating climate dynamics (Singh et al., 2018; Tierney, Poulsen, et al., 2020; Feng et al., 2021; Steinman et al., 2022; Luo et al., 2023) and evaluating climate model simulations (e.g. Braconnot et al., 2012; Harrison et al., 2015).

Realizing the full potential of this synthesis depends on the availability of open data, software, and workflows (defined here as the integrated sequences of computational and analytical steps used to process, analyze, and interpret paleoclimate data (Gil et al., 2016)). Addressing this challenge requires coordinated efforts across publishers, funding agencies, and scientific communities. Publishers (Stall et al., 2017) and funding agencies (National Academies of Sciences, Engineering, and Medicine., 2022; Nugroho et al., 2015) do increasingly mandate open science practices, recognizing that reproducibility extends beyond data sharing to encompass transparent documentation of scientific methods, assumptions, and provenance.

Translating these broad principles into paleoclimate research requires addressing several specialized methodological challenges. Paleoclimate time series are often irregularly sampled in time, strongly autocorrelated (raising the bar for statistical significance assessment), and characterized by low signal-to-noise ratios with large uncertainties in both age and measured values. Consequently, they require specialized analytical approaches for tasks such as age modeling (Blaauw & Christen, 2011; Parnell et al., 2008; Lee et al., 2023; Bronk Ramsey, 2009; Lisiecki & Lisiecki, 2002; Lin et al., 2014b; Hohmann et al., 2024; Parrenin et al., 2024), uncertainty propagation (Khider et al., 2014, 2015; Tierney & Tingley, 2014, 2018; Carson et al., 2019), frequency-domain analysis (Foster, 1996; Ghil et al., 2002; Thomson, 2009; Khider et al., 2022; N. McKay et al., 2021; Schulz & Mudelsee, 2002; Torrence & Compo, 1998), correlation and significance testing (Khider et al., 2022; Hu et al., 2017), proxy system modeling (Dee et al., 2015; Dolman & Laepple, 2018; Weitzel et al., 2025), and climate field reconstruction (Tingley et al., 2012; Smerdon et al., 2023; Zhu et al., 2024; Tierney et al., 2025). Although numerous methods have been developed to address these challenges, relatively few are implemented in accessible, well-documented software packages supported by tutorials (Blaauw & Christen, 2011; Parnell et al., 2008; Khider et al., 2022; N. McKay et al., 2021; Lisiecki & Lisiecki, 2002; Bronk Ramsey, 2009; Greene et al., 2019). As a result, researchers frequently reimplement similar analytical procedures, leading to fragmentation that increases the time required to conduct paleoclimate studies and makes it difficult to reproduce or extend previous analyses (Peng, 2011; Stodden et al., 2016; Sandve et al., 2013; Wilson et al., 2017; Gil et al., 2016).

Analyzing paleoclimate records is more than the sum of data and software: it requires *workflows* – structured, reproducible computational graphs encoding the full provenance chain from raw proxy observations to scientific conclusions. Unlike a collection of scripts, a workflow makes data dependencies explicit, specifying not only what computations are performed, but in what order, with what inputs, and under what parameter configurations. This is critical in the paleosciences where heterogeneous proxy archives (e.g., tree rings, corals, ice cores, sediment cores) must be ingested, harmonized, and passed through multi-stage processing chains before any environmental inference can be made. A well-designed workflow also serves as executable documentation: it captures methodological choices in a form that is both human-readable and machine-runnable, enabling

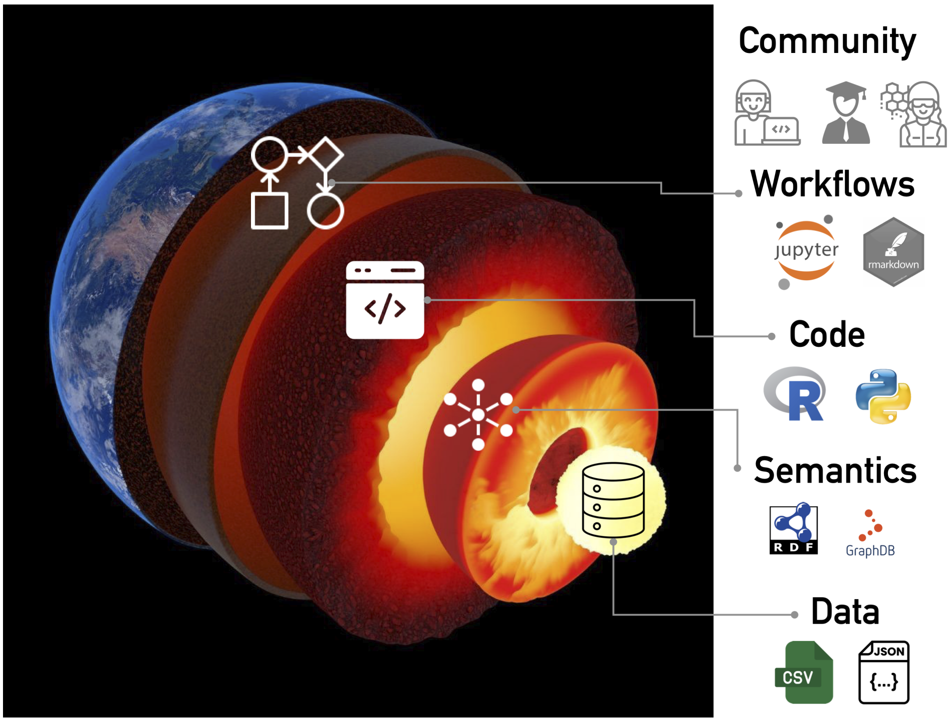

Figure 1. The LinkedEarth ecosystem: at the core sit standardized data, augmented by appropriate semantics (sections 2 and 3). This enables code (section 4) to leverage this structure and metadata, in turn supporting transparent replicable, and (ideally) reproducible workflows (section 5). This ecosystem supports the work of a scientific community, who uses (and contributes to) open datasets, standards, software and workflows.

other researchers to reproduce, audit, or extend analyses without reverse-engineering the original author's intent. Crucially, workflows support systematic parameter exploration, transforming what would otherwise be a one-off analysis into a reusable scientific instrument.

In this paper, we describe an ecosystem (Figure 1) to advance open and reproducible paleoclimatology through coordinated development of three fundamental pillars: (1) *Data* – standardized representations and semantic frameworks (Sections 2 and 3); (2) *Software* – tools for data manipulation, analysis, and interpretation (Section 4); and (3) *Workflows* – curated galleries of transparent, reproducible analyses (Section 5). These components link standardized datasets with analytical tools and their underlying assumptions, enabling researchers to focus on scientific questions rather than data preparation and methodological reimplementation. The ultimate goal is to build capacity in the analysis and synthesis of paleoclimate and paleoceanographic data, to accelerate and strengthen the inferences made from them. The present article summarizes work done in this area over the past decade, and provides an outlook to future challenges and opportunities.

2 Data Standards as a Foundation for Open and Reproducible Paleoclimatology

Despite the scientific importance of proxy synthesis, assembling analysis-ready datasets remains a major challenge that is primarily limited by the time and expertise required for manual data curation. Paleoclimate records are typically produced by small research teams and archived using conventions that reflect the practices of individual investigators, laboratories or synthesis efforts (e.g. PAGES 2k). As a result, key aspects of the interpretation (e.g., the climate variable represented, the seasonality of the signal, calibration assumptions, and the treatment of chronological uncertainty) are often documented inconsistently across datasets and are generally embedded within narrative descriptions in peer-reviewed publications, technical reports (e.g. IODP) and academic theses. Integrating records therefore requires substantial domain expertise to interpret these assumptions and harmonize them across studies, a process that must often be repeated for each new investigation.

Addressing this bottleneck requires community standards that preserve the scientific meaning of paleoclimate observations in a manner that can be consistently interpreted and reused. Following the EarthCube Leadership Council (2020), we use the term "standard" to mean *a public specification documenting some practice or technology that is adopted and used by a community.* Such standards are sometimes specified in one or more documents describing the technical characteristics of these artifacts or practices.

Such standardization rests on three complementary components: a **common structure** for organizing paleoclimate datasets, a **controlled terminology** to describe key metadata elements, and **shared guidelines** for reporting the information required to interpret and reproduce analyses. Together, these elements provide the foundation for stewardship of paleoclimate data, ensuring that datasets are not only accessible but also interpretable, interoperable, and reusable within analysis workflows.

Over the past decade, these three components have been addressed through complementary community initiatives. We briefly review these efforts here before introducing a semantic framework that connects them and enables their use in interoperable and computable paleoclimate workflows.

2.1 Community Standards for Paleoclimate Data and Metadata

The discussion below is nonexhaustive and only focuses on the efforts re-used as part of the work presented here.

2.1.1 The Linked PaleoData Format: a Standard Structure for Paleoclimate Data

Synthesis efforts as part of the PAGES 2k compilation (PAGES 2k Consortium, 2013) motivated the creation of a universal, machine-readable file format that could accommodate the flexibility and complexity of paleoclimate data: the Linked PaleoData (LiPD) format (N. McKay & Emile-Geay, 2016), which organizes paleoclimate data and metadata in a standard and machine-readable way. In LiPD, metadata are stored in a JSON-LD file while the data themselves are organized in one or more tables saved as `.csv` files. LiPD defines six distinct components: root metadata (e.g., dataset name, version, and unique identifier); geographic metadata (e.g., coordinates); publication metadata; funding metadata; PaleoData, which includes all measured variables (e.g., tree-ring width) and inferred environmental variables (e.g., temperature); and ChronData, which mirrors PaleoData for information pertaining to time and chronology. Together, these components provide the structural rigidity necessary to develop robust software tools while remaining sufficiently extensible to capture the diversity of paleoclimate datasets, as exemplified by the PAGES 2k compilation, version 2 (PAGES 2k Consortium, 2017). LiPD was intentionally designed to accommodate additional properties as needed to describe new datasets or analytical approaches (Gil et al., 2017). In practice, many such extensions have arisen during synthesis projects and are typically adopted through consensus among the researchers involved in these efforts (e.g. Konecky et al., 2020; Walter et al., 2023).

2.1.2 The Paleoenvironmental Standard Terms Thesaurus: A Standard Terminology for Paleoclimate Data

Developed by the U.S. National Oceanic and Atmospheric Administration (NOAA) World Data Service for Paleoclimatology (WDS-Paleo), the Paleoenvironmental Standard Terms (PaST) Thesaurus aimed to define common terminology for describing paleoclimate datasets, including variable names, archive types, error descriptions, units, and analytical methods (Morrill et al., 2021). PaST is a hierarchical vocabulary of standardized variable names and definitions, represented using the Simple Knowledge Organization System (SKOS) data model, and designed to evolve as community needs arise (Morrill et al., 2021). The standard is used operationally by WDS-Paleo, which also accepts datasets in the LiPD format provided that their metadata conform to the PaST vocabulary.

While PaST provides a controlled vocabulary for describing paleoclimate observations, its integration within a broader semantic framework required some extensions and adaptations to ensure compatibility with the data structures defined in LiPD. These alignments allow controlled terminology to be consistently associated with structured dataset elements and are described in more details in section 2.2.

2.1.3 The Paleoclimate Community reporTing Standard: a Reporting Standard for Paleoclimate Data

The Paleoclimate Community reporTing Standard (PaCTS) (Khider et al., 2019) represents the first crowdsourced effort to establish reporting guidelines for paleoclimate data, following similar initiatives in the climate modeling community such as the Climate and Forecast (CF) conventions (Gregory et al., 2020). PaCTS defines three categories of metadata: *essential* (without which the data cannot be reused), *recommended* (to enable reproducibility of the original study), and *desired* (additional metadata collected during the investigation that may support further analyses but are not strictly required for reuse). Agreement on the classification of metadata fields was achieved through a survey distributed to the paleoclimate community. In practice, however, the community identified a large number of properties as essential, reflecting aspirational goals for comprehensive reporting. As a result, the standard proved challenging to implement at scale. The work presented here builds on PaCTS by identifying the subset of metadata most commonly required in paleoclimate analysis workflows, helping to distinguish information necessary for reuse and interoperability from additional contextual metadata.

2.2 A Semantic Framework for Paleoclimate Data Standards

The standards described above address complementary aspects of paleoclimate data stewardship: LiPD provides a structured representation of datasets, the PaST Thesaurus defines controlled terminology for describing key metadata elements, and PaCTS establishes reporting guidelines for documenting paleoclimate observations. However, these efforts were developed largely independently and therefore require a framework capable of connecting their concepts in a consistent and machine-interpretable manner. The LinkedEarth Ontology (LEO, Section 2.2.2) provides this semantic layer by formally representing the relationships among paleoclimate data structures, controlled vocabularies, and reporting conventions. In the following subsections, we describe how PaST terminology was adapted and integrated within the LiPD data model, and how these concepts are represented within LEO.

2.2.1 A Controlled Vocabulary for Paleoclimate Datasets

Standardizing paleoclimate datasets has historically been challenging because of their diversity and often individualized characteristics. To accommodate this diversity, the LiPD format was designed to be flexible and extensible. However, for some key metadata fields this flexibility led to a proliferation of variable names specific to individual projects, which hindered effective querying and integration across synthesis efforts.

To address this issue, we standardized the terminology for six key metadata fields: variable name, units, seasonality (which portion of the seasonal cycle the environmental sensor records; for example, trees in the Northern Hemisphere typically grow during spring and summer), interpretation variable (e.g., temperature), proxy (e.g., ring width), and archive type (e.g., wood). For these categories, there was substantial overlap between the terminology used in existing LiPD datasets and the concepts defined in the PaST Thesaurus, although coverage varied across categories.

For some terms, alignment with PaST was straightforward. For example, the PaST nomenclature for seasonality was adopted almost directly. For other categories, the correspondence was more complex. In some cases, LiPD datasets used terminology that was more granular than the concepts currently defined in PaST (e.g., organic geochemical ratios). In other cases, differences between the LiPD framework and the WDS-Paleo data model meant that key LiPD concepts were absent from PaST. For instance, LiPD uses the concept of archive types first introduced by Evans et al. (2013), whereas the concept of `Lake Sediment` as an archive category does not exist in the PaST hierarchy.

To support interoperability between LiPD datasets and the PaST vocabulary, we developed an extended vocabulary that maps LiPD terminology to PaST concepts where possible and introduces additional terms where necessary. This extended vocabulary is available online (`https://lipdverse.org/vocabulary/`), where all terms and their relationships to PaST are documented and provided with persistent Universal Resource Identifiers (URIs).

2.2.2 The LinkedEarth Ontology

LiPD and its aligned vocabulary have given rise to the LinkedEarth Ontology (LEO, Ratnakar et al. (2026); see TextS1 for design philosophy and considerations), a community-led ontology designed to represent paleoclimate datasets and the scientific concepts necessary to interpret them. While controlled vocabularies standardize terminology, they do not capture the relationships among the entities involved in paleoclimate data, such as the links between archives, proxies, measurements, environmental interpretations, and chronological information. LEO addresses this gap by providing a formal representation of these relationships, enabling paleoclimate datasets to be described in a consistent and machine-interpretable manner.

Ontologies describe objects within a domain by defining classes of entities, their properties, and the relationships among them. For example, in the LEO, a `Dataset` is linked to its geographic `Location` through the property `hasLocation`, while a `DataTable` is related to the variables (i.e. columns) it contains through the property `hasVariable`. By making these relationships explicit, the ontology encodes the structure and meaning of paleoclimate data in a form that can be consistently interpreted, queried, and reused across studies.

Ontologies have had a profound impact in other scientific domains, particularly in biomedical research, where they underpin widely used resources ranging from genomics (Ashburner et al., 2000) to disease classification (Osborne et al., 2009). Ontologies have

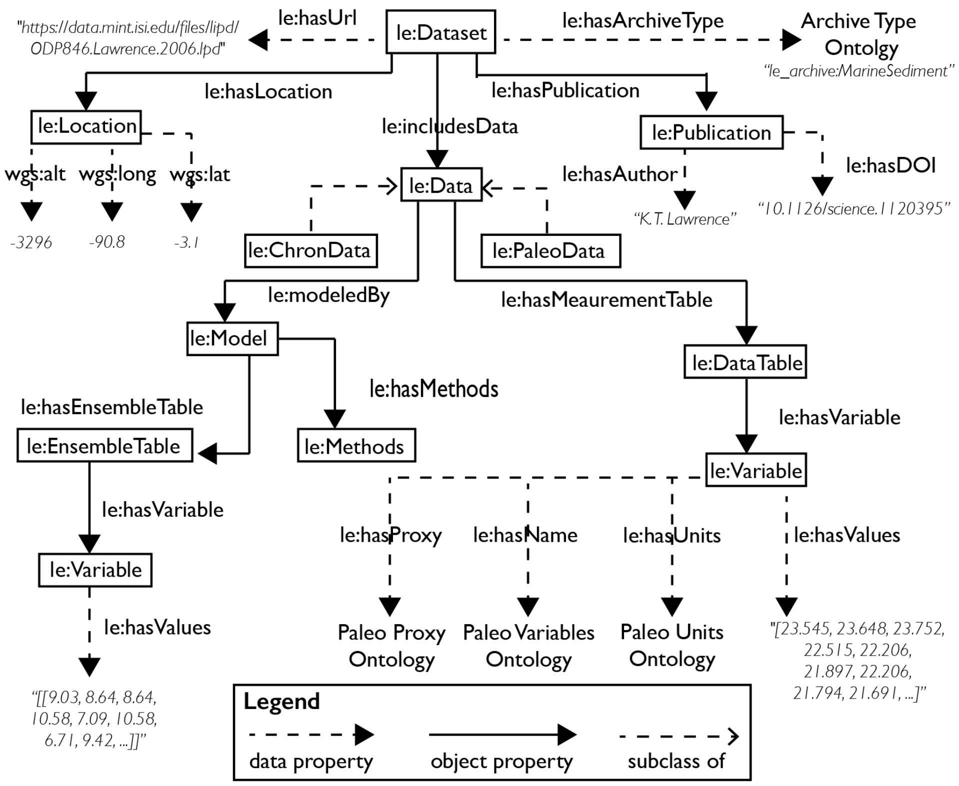

Figure 2. Visual depiction of a paleoclimate dataset (Lawrence et al., 2006) described using the LinkedEarth Ontology, illustrating the interaction between LiPD-based structural components and controlled vocabulary modules.

also been developed in the environmental sciences to support the integration of heterogeneous observations (Buttigieg et al., 2013). Standard ontologies support open data practices by facilitating accessibility, interoperability, and reuse: they provide a shared vocabulary for consistent discovery, an interlingua for mapping across heterogeneous data sources, and a formal structure that precisely defines dataset characteristics and their relationships. To be practically useful, however, ontologies must strike a balance between stability and flexibility, remaining sufficiently structured for reliable use in software applications while accommodating the evolution of scientific knowledge.

The LEO is organized into seven modules, each addressing a key aspect of paleoclimate datasets. The core module, the LinkedEarth Core Ontology, is based on the LiPD data model and represents the structure and provenance of paleoclimate datasets, including links to publications and associated metadata. Additional modules—Archive Type, Paleo Variables, Paleo Proxy, Paleo Units, and Interpretation—encode the controlled vocabulary described in Section 2.2.1, including their associated synonyms. This modular design enables automatic standardization of datasets by mapping non-standard terms to controlled vocabulary concepts, while also supporting consistent querying across datasets that use heterogeneous terminology. For example, a dataset containing the variable name `G. ruber` δ^{18}`O` can be mapped to the standardized term `d18O` in the controlled vocabulary. The original variable name is preserved in the dataset, but queries can now reliably retrieve all records associated with `d18O`, regardless of how the variable was originally labeled.

LEO also introduces an Instrument Ontology, which is not explicitly represented in LiPD. This module captures information about the devices used to produce observations (e.g., mass spectrometers), providing additional context for interpreting measurements.

By explicitly representing both the structure of paleoclimate datasets and the relationships among their components, LEO provides the semantic foundation required to transform collections of LiPD datasets into a knowledge graph that supports complex queries across paleoclimate data. Figure 2 illustrates how these modules interact to represent a paleoclimate dataset within a unified semantic framework.

3 Structuring Paleoclimate Knowledge for Reuse and Interoperability

To make LEO's formal representations practically useful for scientific analysis, they must be instantiated within a data structure that supports efficient search, access, integration, and reuse. While an ontology can be viewed as analogous to a schema in a relational database, ontologies offer distinct advantages: they support knowledge graphs in which data are represented as interconnected entities, allowing relationships to be expressed flexibly and queried across multiple dimensions without requiring predefined table structures (Hogan et al., 2021).

In our framework, paleoclimate datasets encoded in the `LiPD` format are instantiated as a knowledge graph called `LiPDGraph`, using LEO to define their structure. Datasets, variables, proxies, and associated metadata are modeled as nodes connected by explicitly defined relationships. By integrating data and metadata within a unified structure, `LiPDGraph` enables consistent access to heterogeneous datasets and supports complex queries across collections of paleoclimate records.

3.1 Representing LiPD-serialized Data as a Knowledge Graph

To transform `LiPD` files into their graph representation, we map the metadata keys defined in `LiPD` to properties in LEO by defining a schema map along with custom transformation functions (e.g., inferring controlled vocabulary terms). The same schema map is also used to reverse the process, allowing graph representations to be converted back into `LiPD` files and ensuring interoperability for offline use. Each individual dataset is stored as its own graph, preserving its internal structure while enabling integration across datasets—a requirement for querying across `LiPDGraph`.

In addition to metadata, the values associated with each data column are extracted from the original `.csv` files and stored directly within the graph. To our knowledge, `LiPDGraph` is unique among graph-based data systems in that both data and metadata are stored within the same structure, rather than linking to external files. This integrated approach enhances interoperability by eliminating the need to track, locate, and synchronize separate data and metadata files across computational environments. Scientists can seamlessly access and combine data and metadata within a single query or programmatic interface, such as through a `Pandas` DataFrame (Pandas development team, 2020), enabling consistent workflows across tools, programming languages, and research contexts. Concrete examples of this improved accessibility are provided in Section 5.

The use of serialized representations (strings) for storing numerical arrays improves the performance of large queries but introduces some limitations, as direct querying over individual values is not supported. To address this, summary properties (e.g., minimum and maximum values) are extracted during LiPD file creation and stored as metadata within the graph. The selection of these derived properties is guided by common scientific examples identified through synthesis efforts and community engagement. As new examples emerge, additional metadata properties can be incorporated, leveraging the flexibility of the `LiPD` format and LEO.

3.2 Datasets in the `LiPDGraph`

As of March 2026, `LiPDGraph` contains 5,316 datasets across 16 synthesis compilation efforts (Emile-Geay et al., 2017; Kaufman et al., 2020; Konecky et al., 2020; Walter et al., 2023; Arzey et al., 2024; N. P. McKay et al., 2024; Chadwick et al., 2022; Hancock et al., 2023; Routson et al., 2021; Comas-Bru et al., 2019, 2020; Hancock et al., 2024). During the curation and assembly of these compilations, disciplinary working groups iteratively refine the representation of their data to optimize discoverability and reusability. This process often necessitates the definition of new metadata properties when existing LEO properties cannot fully capture domain-specific information. Upon integration into `LiPDGraph`, these new properties are formally incorporated into the LEO, enabling more expressive queries and facilitating their adoption by other researchers.

This collaborative, community-driven approach to ontology curation (Gil et al., 2017) accelerates the evolution of the schema while ensuring that it remains responsive to the needs of the paleoclimate research community. The underlying knowledge graph architecture provides the flexibility to accommodate schema extensions while maintaining the formal structure required for querying and downstream analysis. This design allows researchers to focus on scientifically meaningful data representation without requiring deep expertise in ontology engineering.

3.3 Accessing `LiPDGraph` Programmatically

`LiPDGraph` can be accessed through a SPARQL endpoint[1], enabling expressive semantic queries across datasets. To support reproducibility and accommodate evolving data standards, we maintain multiple versioned snapshots of the knowledge graph, each accessible through a dedicated endpoint. This approach ensures that analyses conducted at a given time can be reproduced using the same version of `LiPDGraph` that was available when the original work was performed.

Because SPARQL requires familiarity with query languages that are not commonly used in the paleoclimate community, we provide a Python interface that exposes commonly used queries through user-friendly APIs (Ratnakar & Khider, 2025). This pack-

[1] `https://linkedearth.graphdb.mint.isi.edu/repositories/LiPDVerse-dynamic`

age allows researchers to interact with the knowledge graph using familiar tools, lowering the barrier to adoption and facilitating integration with existing analysis workflows.

Finally, we have developed a Visual Studio Code extension[2] to browse existing datasets saved locally or in the `LiPDGraph`, edit existing files (only registered users can push changes to the `LiPDGraph`), or create new ones.

4 Software Infrastructure for Paleoclimate Data Analysis

Open data and their structured representation are only a third of the story. To support open and reproducible science, researchers also require reliable, field-tested and well-documented software implementing the analytical methods upon which their workflows rest. The LinkedEarth ecosystem is thus predicated on a close integration between data and software[3]. While most of the relevant packages have been described in detail elsewhere, this section briefly recounts some of the essential packages supporting the science applications of Section 5.

`PyLiPD` (Ratnakar & Khider, 2025) provides a Python interface for reading, manipulating, and querying `LiPD`-formatted datasets. By enabling seamless conversion between `LiPD` files and common data structures (e.g., `Pandas` DataFrames), `PyLiPD` lowers the technical barrier to paleoclimate data analysis and allows researchers to apply the full breadth of Python scientific computing tools to paleoclimate records. The package also provides programmatic access to `LiPDGraph`, allowing researchers to query across large synthesis compilations without requiring familiarity with semantic query languages.

`Pyleoclim` (Khider et al., 2022) is a Python package for paleoclimate timeseries analysis and visualization, implementing a range of specialized methods for paleoclimate data including outlier detection, spectral analysis, wavelet analysis and correlation assessment. All methods are "paleo-aware" in the sense that they can natively deal with the challenges of paleoclimate timeseries (irregular sampling, noisiness, high autocorrelation). By consolidating these methods within a single, well-documented package, `Pyleoclim` addresses the fragmentation problem described in Section 1: researchers no longer need to reimplement common analytical procedures or assemble incompatible code from mul-

[2] `https://marketplace.visualstudio.com/items?itemName=KaryaLimited.lipd-vscode`

[3] `https://linked.earth/tools/`

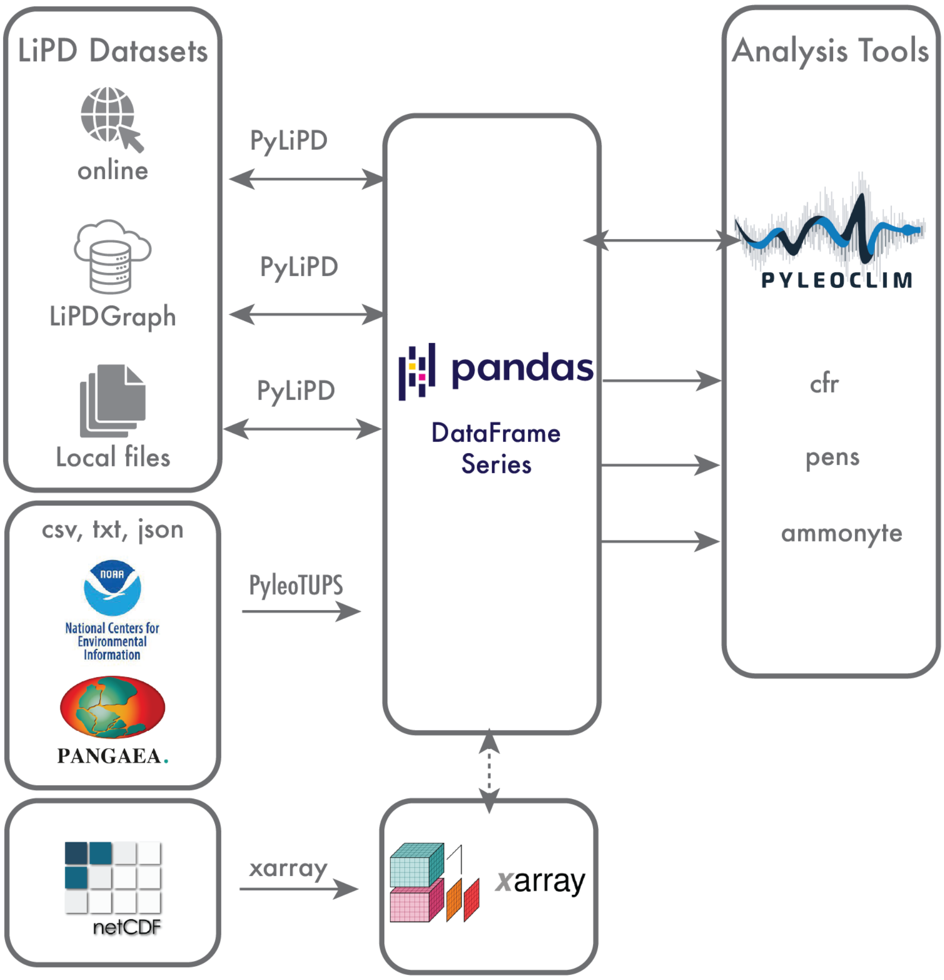

Figure 3. Python ecosystem of software tools for paleoclimate data analysis. Tools such as `PyLiPD` (Ratnakar & Khider, 2025) and `PyleoTUPS` (Oswal et al., 2025) read datasets and output to `Pandas` DataFrames, which can then be consumed by analysis tools such as `Pyleoclim` (Khider et al., 2022), `cfr` (Zhu et al., 2024), `Ammonyte` (James, 2023) or `pens` (Emile-Geay et al., 2025), thus enabling reproducible, interoperable, and efficient scientific workflows.

tiple sources. The package is designed to operate directly on `Pandas` DataFrames populated by `PyLiPD`, enabling seamless integration within analysis pipelines.

`lipdR` (Heiser et al., 2016) provides the R-language foundation for the LinkedEarth software ecosystem: a package for reading, writing, and manipulating `LiPD`-formatted datasets that underlies all other R tools described here. Analogous in function to `PyLiPD` in the Python stack, `lipdR` enables seamless conversion between `LiPD` files and native R data structures, allowing researchers to leverage R's statistical and visualization libraries on standardized paleoclimate records without bespoke format parsing. By establishing a common data access layer, `lipdR` ensures that improvements to the `LiPD` format propagate consistently through the entire R package family.

`geoChronR` (N. McKay et al., 2021) is an R package that provides complementary functionality for paleoclimate chronology modeling and age-uncertain data analysis. As many paleoclimate researchers work in R, `geoChronR` ensures that the LinkedEarth infrastructure is accessible across the dominant statistical programming environments in paleoclimatology. The package implements advanced methods for age modeling and uncertainty propagation, critical components of any paleoclimate analysis.

`actR` (N. McKay et al., 2024), or the Abrupt Change Toolkit in R, extends the R ecosystem with specialized methods for detecting abrupt transitions in paleoclimate time series. The package implements statistical tests for shift detection and change-point analysis, designed to account for the irregular sampling and strong autocorrelation that characterize paleoclimate records. Building on the ensemble and visualization infrastructure shared across the R package family, `actR` propagates detection results through age-uncertain ensembles, providing robust assessments of the timing and magnitude of abrupt events across multi-proxy datasets.

`compositeR` (N. McKay & Hancock, 2023) addresses the construction of composite records from collections of paleoclimate time series. This is a widely used approach for identifying coherent climate signals across heterogeneous datasets that is complicated by irregular sampling, age uncertainty, and differing proxy resolutions. The package implements both time-normalized and spatially gridded compositing methods that natively accommodate these challenges through the age-uncertain ensemble framework shared by the R package family. By enabling systematic synthesis of multi-proxy datasets within reproducible workflows, `compositeR` supports the large-scale event detection and synthesis analyses that are increasingly central to paleoclimate science.

Beyond these core packages, the ecosystem includes domain-specific tools such as `cfr` (Zhu et al., 2024), a Python package for climate field reconstruction that enables researchers to reconstruct large-scale climate fields from sparse paleoclimate observations, and other specialized analysis tools referenced throughout the workflows presented in the following section. Together, these software packages provide the computational infrastructure necessary to implement open and reproducible principles in paleoclimate research.

5 Reproducible Workflows as Accelerators of Scientific Insight

Building upon the foundations of data standards, semantic data resources and free and open software tools described above, we now describe reproducible analytical pipelines (workflows) that operate seamlessly across the LinkedEarth ecosystem. We first describe the infrastructure supporting these workflows through the `PaleoBooks` gallery, a centralized platform for transparent and reusable computational narratives. We then illustrate the scientific potential of this approach with three representative examples that demonstrate how the LinkedEarth ecosystem can be leveraged to address common challenges in paleoclimate research. These examples showcase the practical benefits of coordinated stewardship across the three pillars: they operate directly on standardized, well-structured datasets; they leverage established, well-documented software tools; and they capture scientific reasoning and procedural knowledge in transparent, reproducible, and reusable computational narratives.

5.1 PaleoBooks: A Community Gallery for Transparent and Reusable Paleoclimate Workflows

Each of the examples described in this section are drawn from a growing gallery of reproducible notebooks, referred to as `PaleoBooks`, which document complete analysis workflows from data discovery to interpretation. Each PaleoBook is a transparent, executable and reproducible computational narrative (aka "notebook", whether Jupyter (Kluyver et al., 2016) or Rmarkdown (Allaire et al., 2026)) that showcases how to use open science resources in the paleosciences, with the intent of spreading open science practices through easily repurposed examples.

The `PaleoBooks` gallery[4] serves as a centralized, browsable, and searchable index of such examples, bridging paleoclimate observations, analytical methods, and climate model output. While some `PaleoBooks` serve purely as tutorials, a growing fraction were designed to accompany peer-reviewed scientific investigations as fully executable forms of supplementary material, allowing readers to follow the progression from raw data to published figures and to interrogate each step of the analysis. This approach shifts the

[4] `https://linked.earth/PaleoBooks/`

role of computational workflows from isolated methodological descriptions to shareable research objects.

PaleoBooks fulfill complementary roles within the scholarly landscape of paleoclimate and paleoceanography. As scientific companions to publications, they provide detailed, executable accounts of analytical pipelines, including robustness checks, sensitivity analyses, and methodological choices that are often omitted from traditional articles due to space limitations. By making these elements accessible, PaleoBooks enhance transparency and facilitate critical evaluation and reuse of published results. In parallel, they support compliance with emerging journal and funding requirements (e.g., Fox et al. (2021); Erdmann et al. (2021)) for data and code availability by providing structured, citable, and reusable documentation of computational workflows. This approach is inspired by similar community galleries in other domains, such as Project Pythia (Camron et al., 2025), though PaleoBooks offer a more flexible contribution model. Indeed, while existing PaleoBooks (as of July 2026) have largely been contributed by the present authors or their close collaborators, the model is designed to encourage anyone to submit notebooks, as long as they follow the same principles.

Beyond their role in research dissemination, PaleoBooks also serve as educational and community resources. Didactic notebooks provide step-by-step implementations of common but nontrivial analytical tasks, enabling students and researchers to learn by interacting directly with real-world examples. These workflows capture domain-specific expertise (e.g. data preprocessing strategies, parameter selection, and diagnostic evaluation) that is typically transmitted informally within research groups, transforming it into reusable and extensible knowledge. In addition, PaleoBooks can highlight the integration of multiple software packages within coherent pipelines, offering practical guidance on how tools such as PyLiPD (Ratnakar & Khider, 2025), Pyleoclim (Khider et al., 2022), as well as mainstream libraries (e.g. xarray (Hoyer & Hamman, 2017), pandas (Pandas development team, 2020), NumPy (Harris et al., 2020), and matplotlib (Hunter, 2007)) may be leveraged to make paleoclimate analyses easier, more rigorous, and more reproducible.

The gallery is designed to support both contributors and users (Figure 4). Researchers can develop their own PaleoBooks by adapting a standardized template, publishing them as Jupyter Books, and linking them to persistent identifiers (e.g., DOIs) to ensure citabil-

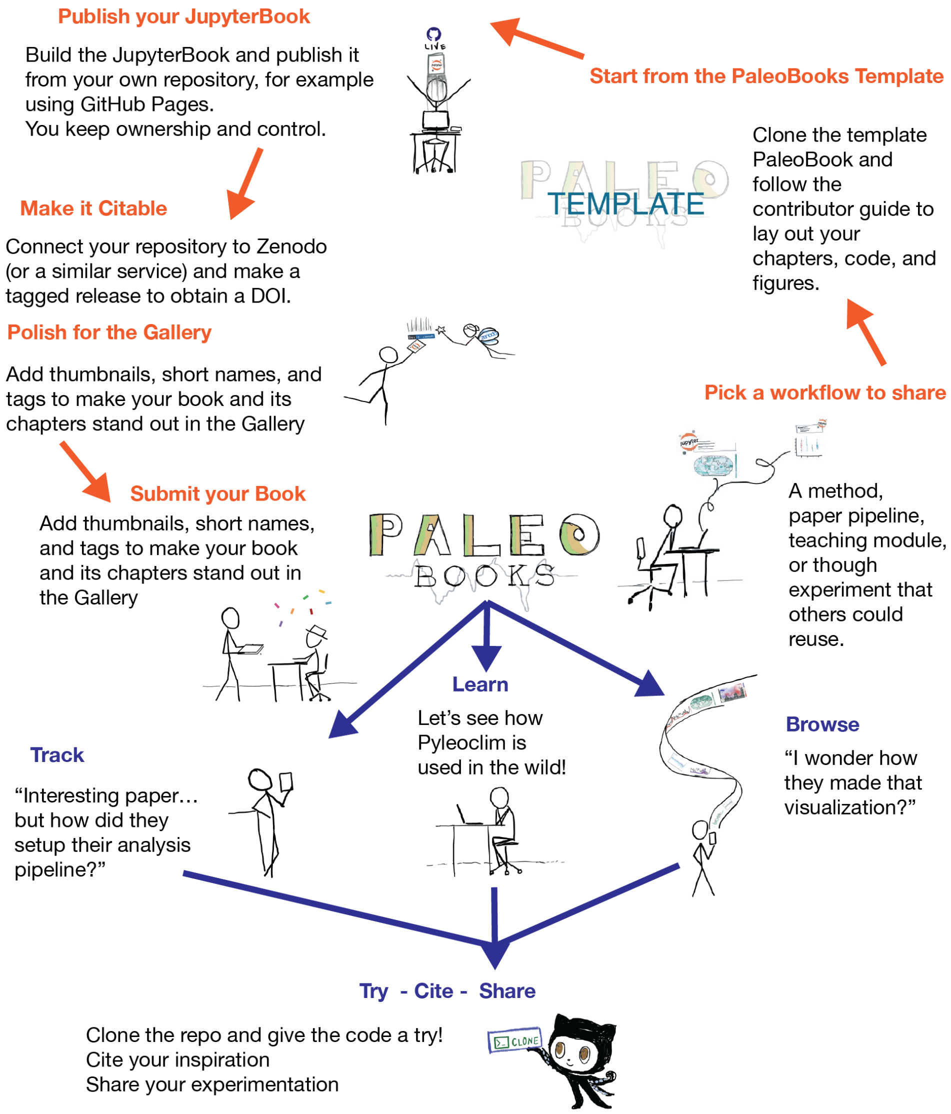

Figure 4. Researcher interaction with the `PaleoBooks` gallery as a contributor (orange) and as a user (blue). Contributions to the gallery promote principles of open and reproducible science while enabling researchers to retain full ownership of their content. Users are encouraged to cite the corresponding book to give credit to contributors.

ity and long-term accessibility. Once submitted, these contributions are indexed within the gallery with metadata and tags that facilitate discovery across topics, methods, and datasets. For users, the gallery provides an entry point to explore existing workflows, reproduce published analyses, and adapt them to new problems without reinventing the wheel, thereby accelerating the pace of scientific discovery.

By treating workflows as first-class, shareable research products, `PaleoBooks` extend open and reproducible principles beyond data and software to encompass the full analytical process. In doing so, they provide a scalable mechanism for capturing, disseminating, and reusing the computational knowledge that underpins paleoclimate research, complementing the standardized data structures and semantic frameworks described in previous sections. We now showcase three such `PaleoBooks`:

- Frequency-Domain Analysis of Paleoclimate Variability Across Timescales (Section 5.2)
- Interhemispheric Precipitation Changes over the Past Millennium (Section 5.3)
- Global Temperature Reconstruction over the Common Era (Section 5.4)

5.2 Frequency-Domain Analysis of Paleoclimate Variability Across Timescales

A common approach to investigating climate variability is the analysis of paleoclimate records in the frequency domain, including spectral, wavelet, and coherence methods. These analyses are used to identify periodic signals within individual records and to evaluate relationships between proxy records and potential forcings. When applied across datasets and timescales, they can provide insights into the drivers of climate variability and the robustness of inferred signals. A collection of reproducible workflows on frequency-based investigations of paleoclimate records (Khider, Emile-Geay, et al., 2026) showcases these frequency-domain techniques across multiple temporal scales and data types. These workflows leverage the following parts of the LinkedEarth ecosystem: they use `LiPDGraph` and `PyLipd` (Ratnakar & Khider, 2025) to semantically query and programmatically assemble standardized paleoclimate datasets (filtered by metadata criteria such as proxy type, temporal coverage, and archive type), and then apply frequency-domain methods (spectral analysis, Hilbert envelope extraction, and bandpass filtering) implemented in `Pyleoclim` (Khider et al., 2022) and the wider scientific Python stack (e.g., Virtanen et al. (2020)).

Frequency-domain analysis of paleoclimate records involves several interpretive considerations and methodological approaches:

Characterizing scaling behavior across timescales: Paleoclimate records often exhibit power-law scaling in their spectra, reflecting how variability is distributed across timescales. Estimating scaling exponents provides a complementary diagnostic of climate variability, enabling comparison across datasets and evaluation of how energy is transferred between temporal scales. This example demonstrates the workflow in practice: the Holocene temperature proxy compilation (Kaufman et al., 2020) was queried through `LiPDGraph`, which enabled filtering by standardized metadata (e.g., archive type: ice core, speleothem, marine sediment) to assemble consistent datasets. Spectral scaling exponents (β) were then estimated using `Pyleoclim`'s spectral analysis functionality, extending earlier work by Huybers & Curry (2006) and Zhu et al. (2019). At centennial–millennial scales, there are hints of steeper scaling at higher latitudes and in ice cores, though archive-level processes like firn diffusion (Johnsen et al., 2000) may partly explain this pattern. At millennial scales, the much larger sample reveals no consistent geographic or archive-dependent trend, with most records exhibiting $\beta \gtrsim 1$ (Figure 5a-b) – consistent with fractal scaling over the Holocene as reported by Zhu et al. (2019). These results underscore the need for longer records extending to the Last Glacial Maximum to more fully characterize the continuum of temperature variability. Detailed analysis workflows are available at `https://linked.earth/FreqFun/b-spectral-analysis-temp12k/`.

Interpreting coherence and phase relationships: Coherence analysis is commonly used to assess relationships between climate and external forcings. Two contrasting examples illustrate both the pitfalls and value of this approach. First, we used `Pyleoclim` to assess the wavelet coherence (Grinsted et al., 2004) between solar irradiance (Vieira et al., 2011) and global mean temperature over the Common Era (Neukom et al., 2019). The analysis finds limited coherence, with isolated peaks exceeding the red noise benchmark accompanied by highly variable phase angles, indicating no consistent phase relationship at centennial-to-millennial scales. This is consistent with attribution work by Barboza et al. (2019), who found that solar forcing had a negligible impact on reconstructed temperature variability.

By contrast, the same method finds significant coherence in orbital bands between JJA insolation at 35°N and the planktic δ^{18}O record at IODP Site U1385 (Hodell et al., 2023). Furthermore, we isolated millennial-scale climate variability (MCV) from the benthic δ^{18}O record at the same site using `Pyleoclim`'s Hilbert envelope analysis and extracted the obliquity band via bandpass filtering (Figure 5c). In doing so, we reproduced a central finding of Hodell et al. (2023): that MCV amplitude is inversely related to obliquity, indicating that orbital forcing modulates the intensity of millennial variability prior to the Mid-Pleistocene Transition. The complete analysis workflow is available at `https://linked.earth/FreqFun/iodp339-u1385/`. These examples underscore that high coherence does not necessarily imply a causal relationship, and phase relationships must be interpreted cautiously in light of uncertainties and potential confounding factors.

Assessing statistical significance: Spectral peaks and coherence relationships can arise from stochastic variability in autocorrelated time series. The use of appropriate null models (e.g., red-noise benchmarks) is therefore essential to distinguish meaningful variability from background noise. This principle is illustrated through `Pyleoclim`'s spectral analysis of temperature-sensitive records over the Holocene (Kaufman et al., 2020), which were assembled by querying `LiPDGraph` and reveal numerous spectral peaks in the 800–3000 year band without any obvious link to location or proxy archive (Figure 5d). Detailed analysis workflows are available at `https://linked.earth/FreqFun/c-spectral-analysis-peaks-temp12k/`. These widespread peaks highlight a critical interpretive challenge: the presence of spectral features at periods similar to potential forcings does not necessarily indicate a causal relationship. For example, apparent alignment between spectral peaks in climate records and those of solar forcing cannot alone establish a solar driver without supporting evidence from mechanistic or causal methods (e.g. Landers et al., 2025).

Non-stationary and intermittent signals: Periodic components in paleoclimate records may vary in strength through time and are not always persistently expressed. Time–frequency methods (e.g., wavelet analysis) are therefore required to identify when and over which intervals specific periodicities dominate, i.e. to assess their temporal stability. For instance, `Pyleoclim`'s wavelet analysis of the EPICA Dome C δD record (Jouzel et al., 2007) over the past 800,000 years reveals strong 100 kyr variability throughout the record, with precession and obliquity signals concentrated primarily in the most recent part (for details, see `https://linked.earth/FreqFun/epicadomec-explore/`).

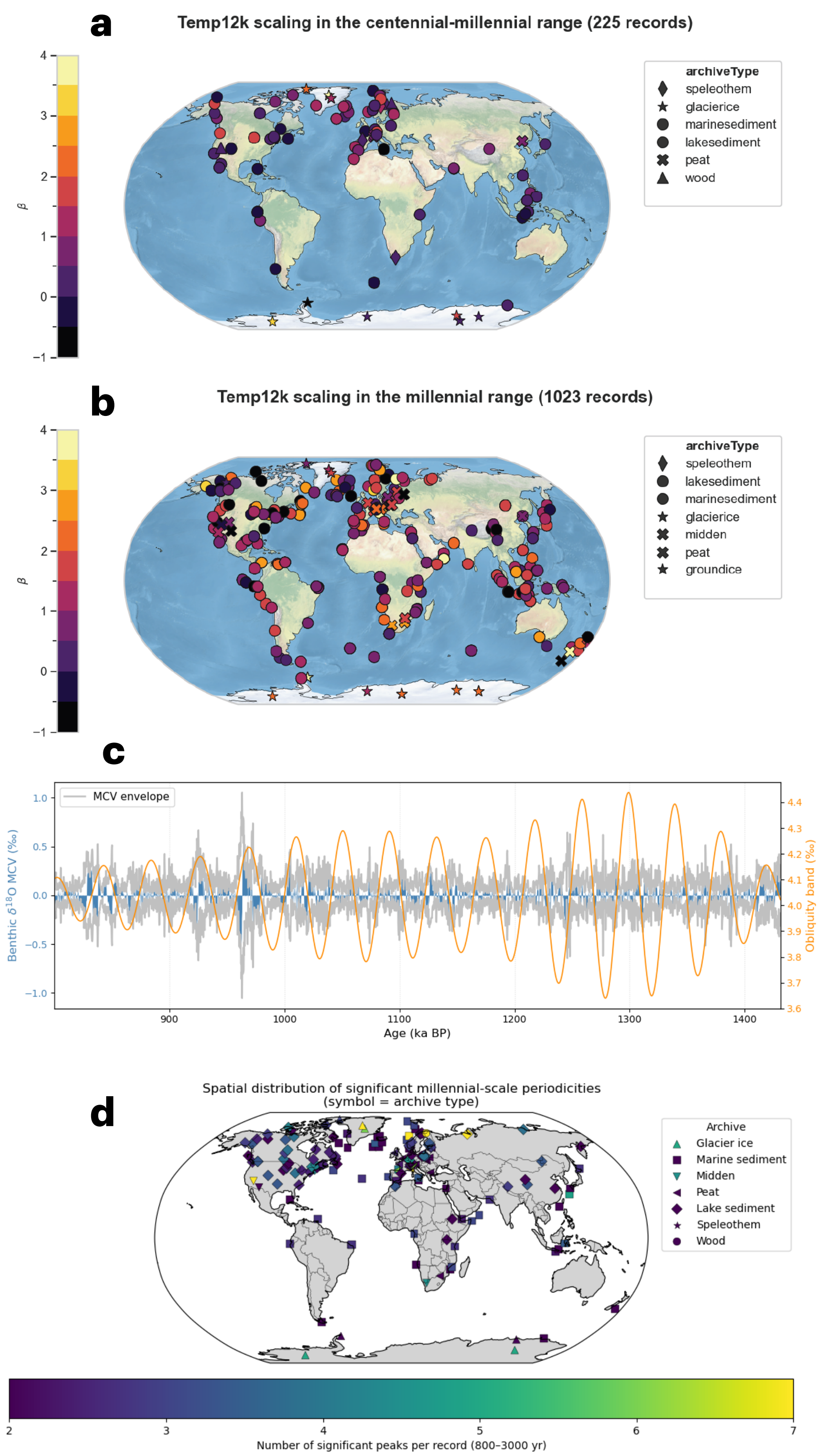

Figure 5. Applications of the LinkedEarth ecosystem to frequency-domain analysis. (a–b) Spectral scaling behavior of Holocene temperature records; (c) millennial-scale climate variability and orbital forcing from IODP Site U1385; (d) spatial distribution of millennial-scale spectral peaks over the Holocene. See Section 5.2 for details.

Together, these four examples illustrate how the LinkedEarth ecosystem enables frequency-domain analyses to be applied to a wide array of paleoclimate datasets and timescales within transparent and reproducible workflows.

5.3 Interhemispheric Precipitation Changes over the Past Millennium

In two days at a Project Pythia workshop (Camron et al., 2025), one of the authors (D. Khider) partially reproduced the results of Steinman et al. (2022) using the LinkedEarth ecosystem, a task that previously required weeks of manual data assembly and analysis.

Steinman et al. (2022) synthesized paleoclimate records and model simulations from the Neotropics spanning the past millennium and revealed an interhemispheric antiphasing pattern in precipitation, where northern and southern tropical regions experienced opposite wetting and drying trends, particularly during the Little Ice Age (LIA). The antiphasing was driven by a southward shift of the Intertropical Convergence Zone (ITCZ), likely triggered by volcanic forcing that cooled the Northern Hemisphere more than its Southern counterpart. Methodologically, the study highlights both the potential and the challenges of integrating heterogeneous proxy observations with model output, including differences in observed variables, temporal uncertainty, and the mixed environmental signals recorded by proxies.

Khider, Sundar, & Ratnakar (2026) reproduced key elements of this analysis using the LinkedEarth ecosystem in combination with complementary scientific Python libraries. The workflow leverages LinkedEarth tools for data discovery, loading, and analysis: standardized proxy records were accessed and semantically queried through `LiPDGraph`, then loaded into analysis-ready datasets using `PyLiPD` (Ratnakar & Khider, 2025), which handled the transformation of LiPD-formatted data into convenient data structures. Hydroclimate-sensitive proxy records were filtered by standardized metadata (temporal coverage, sampling resolution, and paleoclimate interpretation) and combined with climate model output from the CESM Last Millennium Ensemble (LME, Otto-Bliesner et al. (2016)). The assembled datasets were then analyzed using complementary Python tools: `Pyleoclim` (Khider et al., 2022) for paleoclimate-specific analysis, `xarray` (Hoyer & Hamman, 2017) for handling gridded model output, and `eofs` (Dawson, 2016) for principal component analysis (PCA). The latter was applied across these records and model fields to identify dominant modes of variability shared across proxy and model domains. The number of proxy records in the LiPDGraph is not sufficient to derive meaningful scientific conclusions and a direct comparison to the study by Steinman et al. (2022). However, this workflow demonstrates the importance of structuring both proxy and model data within a common analytical framework to enable rapid assembly and analysis of heterogeneous paleoclimate datasets—a task that previously required weeks of manual curation. The reproducible workflow is documented in the `paleoPCA-cookbook` PaleoBook (`https://projectpythia.org/paleoPCA-cookbook/`), which provides a complete, executable implementation accessible to other researchers wishing to adapt this approach to new datasets or scientific questions.

5.4 Global Temperature Reconstruction over the Common Era

Our third example illustrates how the LinkedEarth ecosystem enables the construction of large-scale synthesis datasets and their use in climate field reconstructions (CFRs), focusing on global temperature variability over the past two millennia (Gondhalekar et al., 2025). This `PaleoBook` documents the end-to-end process of assembling proxy records, screening for duplicates, generating a reconstruction, and evaluating the results against independent benchmarks.

A central challenge in global temperature reconstruction lies in assembling heterogeneous proxy records into analysis-ready datasets. Proxy observations differ widely in archive type, temporal resolution, spatial coverage, and interpretation, and are often distributed across individual studies with inconsistent metadata conventions. This workflow demonstrates how the LinkedEarth ecosystem addresses this challenge through programmatic access to standardized datasets via the `LiPDGraph` and standardized data loading tools. The workflow illustrates two complementary approaches: (1) direct ingestion of LiPD files using `PyLiPD` (Ratnakar & Khider, 2025), which parses standardized metadata structures and loads data into analysis-ready formats, and (2) semantic querying of the `LiPDGraph`, which enables filtering by controlled vocabulary terms directly into `pandas` data structures. In both cases, records are systematically filtered based on standardized metadata criteria, including temporal coverage, sampling resolution, proxy type, and paleoclimate interpretation. They are then harmonized into consistent data structures suitable for downstream analysis. By leveraging standardized representations and controlled vocabularies embedded in the LiPD format, this process dramatically reduces the need for manual data curation and ensures that filtering assumptions are explicitly encoded and reproducible.

Once assembled, these datasets are used to generate a global temperature reconstruction using `cfr` (Zhu et al., 2024), a LinkedEarth-developed framework for climate field reconstruction. The workflow demonstrates how LinkedEarth metadata standards enable methodological flexibility: proxy-climate relationships are specified by determining the seasonal sensitivity of each proxy record, estimated using two complementary approaches. The first uses "class-based seasonality," where seasonality is assigned based on archive type and proxy interpretation (e.g., tree rings are assumed summer-sensitive, ice core water isotopes are assumed winter-sensitive). The second approach leverages the rich metadata found in this compilation through "meta-seasonality", where expert annotations documenting the likely seasonal response (explicitly encoded in the LiPD metadata) guide the assignment. The workflow then proceeds with calibration against instrumental data and quantifies uncertainties in both proxy observations and reconstruction parameters. By integrating these steps within a documented and executable notebook, the workflow makes explicit the assumptions underlying the reconstruction and demonstrates how standardized metadata enables transparent, reproducible, and methodologically flexible paleoclimate synthesis.

The final stages of the workflow focus on validation and comparison with existing reconstructions (Figure 6). The reconstructed temperature fields are evaluated against independent datasets and compared to a published global temperature reconstruction (LMRv2.1, Tardif et al. (2019)), providing a benchmark for assessing their robustness and consistency. These comparisons highlight both agreements and discrepancies across reconstructions, reflecting differences in proxy selection, proxy seasonality assignment, and treatment of uncertainties. By embedding validation within the same reproducible framework, the workflow facilitates transparent evaluation of reconstruction skill and supports iterative refinement of both data selection and methodological choices.

The reconstructions reveal notable discrepancies in specific periods. LMRv2.1 and its update (here named "PReSto2k"), which employ class-based seasonality, show more prominent cooling around 550 CE, whereas the reconstructions using metadata-based seasonality exhibit greater cooling around 1800 CE. The latter feature corresponds to the 1815 Tambora eruption (responsible for the Year Without a Summer, (Stommel &

Stommel, 1979)) and its documented climatic signal, demonstrating how methodological choices can influence the detection of culturally significant climate events.

Figure 6. Global Mean Surface Temperature anomaly (GMSTa) reconstructions over the Common Era using different methodological approaches. Four reconstruction variants are shown: LMRv2.1 with offline data assimilation and class-based seasonality assignment; PReSto2k reconstructions employing class-based seasonality, metadata-based seasonality (meta-seasonality), and meta-seasonality combined with spatial interpolation. Each reconstruction displays the median estimate (solid red line) with uncertainty bands representing the 25.0% to 75.0% credible interval (darker red) and the 2.5% to 97.5% credible interval (lighter red). The figure illustrates how different proxy seasonality assignments influence reconstruction patterns, particularly evident in the period around 550 CE and 1800 CE, where class-based and metadata-based approaches diverge. From Gondhalekar et al. (2025).

Together, these `PaleoBooks` demonstrate synergies within and beyond the LinkedEarth ecosystem: standardized data structures and controlled vocabularies (LiPD format and PAST thesaurus) enable systematic discovery and filtering of proxy records; specialized software tools (`PyLiPD` for data loading, `LiPDGraph` for semantic querying, and `cfr` for reconstruction or `Pyleoclim` for timeseries analysis) provide integrated computational workflows; complementary external packages (e.g. `pandas`, `xarray` or `eofs`, `matplotlib`) are used for data acccess, analysis and visualization.

By making each step of the workflow transparent and executable, this approach significantly lowers the barrier to entry for large-scale paleoclimate synthesis. The time savings are substantial: the paleoPCA workflow (Section 5.3) was completed in two days using LinkedEarth tools, whereas comparable synthesis previously required weeks of manual data assembly. This efficiency, combined with reproducibility and methodological transparency, enables more rapid and robust investigation of climate variability over the Common Era and beyond.

6 Discussion

6.1 Open Science and the FAIR Principles

The advances presented in this work are grounded in a fundamental shift in how the scientific community conceives of reproducibility and data stewardship. Reproducibility – understood here as the ability of independent researchers to obtain consistent results using the same data, code, and analytical methods – has long been a cornerstone of the scientific method. However, as paleoclimate datasets and analytical tools grow increasingly complex, reproducibility alone is insufficient. The FAIR principles (Findable, Accessible, Interoperable, and Reusable) provide a complementary framework that operationalizes reproducibility across the full lifecycle of data, software, and workflows (M. D. Wilkinson et al., 2016; Lamprecht et al., 2020; Barker et al., 2022; Goble et al., 2020; S. R. Wilkinson et al., 2025).

In this view, open and reproducible research and FAIR are two sides of the same coin. Open science emphasizes the ethical and practical imperative to make research artifacts available to the scientific community; FAIR principles specify technical and semantic requirements that make those artifacts maximally useful. A dataset may be open but difficult to discover; software may be publicly available but incompatible with existing tools; workflows may be transparent but documented in ways that make them difficult for others to understand or adapt. The FAIR framework addresses these challenges by establishing concrete standards for data and metadata discovery (Findable), open access without unnecessary barriers (Accessible), seamless integration with other resources (Interoperable), and structured documentation that enables repurposing (Reusable).

The LinkedEarth ecosystem advances FAIR principles at each level:

- **Data.** LiPD and LEO make paleoclimate datasets *Findable* through persistent identifiers and enforced descriptive metadata that support semantic search across heterogeneous collections. Datasets are *Accessible* via a public endpoint using open, standardized protocols. The explicit representation of domain knowledge through controlled vocabularies and a formal ontology enhances *Interoperability*, enabling paleoclimate data to be integrated with tools and workflows from across the scientific ecosystem. Standardization promotes *Reusability* by documenting units, chronology, interpretation, and provenance within structured metadata, allowing data to be confidently applied to new research questions.
- **Software.** Pyleoclim, PyLiPD, and PyleoTUPS are *Findable* through persistent identifiers assigned to each package and to every release, with machine-readable citation metadata. They are *Accessible* from open archives and package repositories using standardized protocols, with archival copies preserved independently of the development infrastructure. *Interoperability* follows from reading and writing community formats and exchanging data through the standard structures of the scientific Python ecosystem, allowing the packages to compose with existing analysis workflows. Documentation, tutorials, tests, explicit licensing, and open development practice support *Reusability.*
- **Workflows.** PaleoBooks are *Findable* through persistent identifiers assigned to each book and release, and indexed both in a public archive and in a curated community gallery. They are *Accessible* as executable notebooks retrievable through open protocols, with archived copies preserved independently of the rendered site. *Interoperability* follows from building on community data formats and standard analysis packages, so that the studies draw on the same infrastructure as the wider ecosystem. Narrative documentation of scientific questions, method choices, and interpretation, alongside executable code and its outputs, supports *Reusability*, allowing readers to evaluate and extend published analyses.

A detailed assessment of the ecosystem's alignment with FAIR principles across data, software, and workflows is provided in the supplementary material (TextS2).

6.2 Machine-Actionable Knowledge and Emerging Paradigms

Despite the advances presented in this work, achieving truly open and reproducible paleoclimatology remains an ongoing challenge. Standardized data structures, controlled vocabularies, and reproducible workflows substantially reduce the barriers to data reuse and synthesis, but they do not eliminate the need for domain expertise in interpreting proxy records, selecting appropriate analytical methods, or more fundamentally, asking relevant questions. Many aspects of paleoclimate research, including the interpretation of proxy–climate relationships, the treatment of uncertainties, and the choice of analytical approaches, remain implicitly encoded in expert knowledge and must often be reinvented for each new study.

These limitations highlight an important transition in the evolution of open and reproducible science. To date, most efforts have focused on making data, software, and workflows accessible and interpretable to human users. However, as paleoclimate datasets and analytical tools continue to grow in scale and complexity, there is an increasing need to make scientific knowledge not only human-readable but also machine-actionable. This shift would enable computational systems to assist in tasks such as dataset discovery, metadata interpretation, method selection, and workflow construction, thereby reducing the cognitive and technical burden on researchers and enabling more scalable and systematic analyses.

The ecosystem developed in this work provides a foundation for this transition. By standardizing data representation (LiPD), formalizing domain knowledge (LEO), structuring data as a knowledge graph (`LiPDGraph`), and capturing analytical workflows as executable research objects (`PaleoBooks`), we establish the prerequisites for machine-assisted reasoning over paleoclimate data and workflows. Building on this foundation, emerging approaches such as artificial intelligence (AI)-assisted research systems (e.g., `autoTS` (Khider et al., 2021) and `PaleoPAL` (Ratnakar & Khider, 2026), which was used to create some of the Notebooks presented here) aim to integrate these components within interactive environments that guide users through the process of scientific analysis. Such systems can leverage structured metadata and semantic relationships to interpret datasets, recommend appropriate analytical methods for a particular research question, and assemble reproducible workflows while preserving transparency in the underlying reasoning.

Looking forward, the integration of structured scientific knowledge with AI-assisted workflows would transform paleoclimate research from a collection of manually assembled analyses into a more scalable, interoperable, and collaborative enterprise. By enabling both humans and machines to access, interpret, and reuse paleoclimate data and workflows, this paradigm supports more efficient synthesis across datasets, more transparent evaluation of scientific results, and ultimately a deeper understanding of the climate system across timescales.

7 Conclusion

This paper has presented an integrated approach to advancing open and reproducible paleoclimatology through coordinated stewardship of three fundamental pillars: data, software, and workflows. The developments presented here are motivated by fundamental scientific challenges in paleoclimatology, including the need to reconstruct large-scale climate variability from heterogeneous proxy records, to understand the response of the climate system to external forcing, and to assess the uniqueness of recent climate change.

Data. The foundation of this ecosystem rests on standardizing paleoclimate data and metadata representation through the development of community standards (e.g., LiPD) and LEO semantic framework.

Structuring paleoclimate knowledge as a knowledge graph (`LiPDGraph`) transforms standardized data into an accessible and queryable resource. The `LiPDGraph` instantiates the semantic relationships defined by LEO.

Software. Beyond standardization and structure, paleoclimate science requires a robust ecosystem of software tools that implement the analytical methods needed to extract scientific insight from paleoclimate records. The LinkedEarth ecosystem includes specialized packages such as `PyLiPD` for data access and manipulation, `Pyleoclim` for timeseries analysis and visualization, `geoChronR` for age modeling, and `cfr` for climate field reconstruction, among others.

Workflows. The third pillar extends open principles to the analytical processes that transform data and software into scientific insight through a curated gallery of transparent, reproducible workflows (`PaleoBooks`).

Collectively, these three pillars form a coherent ecosystem that operationalizes open and reproducible principles for paleoclimate science and facilitates the types of large-scale, integrative analyses that are increasingly central to the field. Standardized data representations provide the common language that enables software tools to operate reliably and interoperably; software tools implement the analytical methods that transform data into insight; and reproducible workflows demonstrate how to combine data and software to answer scientific questions. By addressing data standardization, software infrastructure, and workflow design in concert, we provide paleoclimate researchers with the infrastructure needed to move beyond isolated datasets and bespoke analyses toward more interoperable science.

This infrastructure does not prescribe particular scientific questions or analytical methods; rather, it provides the scaffolding upon which diverse research communities can build. As paleoclimate science continues to grapple with increasingly complex datasets, finer temporal and spatial scales, and more sophisticated questions about climate variability and change, the ability to discover, understand, combine, and reuse data, knowledge, and workflows becomes increasingly critical. This framework accelerates the transition toward open and reproducible paleoclimatology and positions the community to conduct the large-scale syntheses that this rich and growing body of observations now affords.

Open Research Section

All datasets used in this study are accessible through `LiPDGraph` and LiPDverse for LiPD-formatted datasets, and through NSF JetStream2 and Figshare (Khider, 2026) for NetCDF files. Each PaleoBook is being developed in a public repository on Github. Each Book includes a citation file for proper accreditations, Zenodo DOIs for unique identifiers, and a computational environment file (`environment.yml`) for reproducibility. The Books presented in this study can be accessed through their invidual GitHub repository while rendedred in HTML: FreqFun v2.0.0 (`https://github.com/LinkedEarth/FreqFun`; `http://linked.earth/FreqFun/`); PaleoPCA v2026.3.30 (`https://github.com/khider/paleoPCA`; `https://khider.github.io/paleoPCA/`); presto2k_cfr_pb v1.0.1 (`https://github.com/LinkedEarth/presto2k_cfr_pb`; `https://linked.earth/presto2k_cfr_pb/`). All Books are accessible through the `PaleoBooks` gallery (`https://linked.earth/PaleoBooks/`).

AI Disclosure

The authors used the GRAIL platform to prepare the manuscript, which includes an integrated AI agent. GRAIL assisted with the following tasks: conducting a literature review (with all articles subsequently vetted for inclusion in the introduction), generating text from bullet points (manually edited and refined), and improving prose clarity. Khider, Ratnakar, and Emile-Geay are currently developing an AI assistant called PaleoPAL. Khider tested PaleoPAL on select notebooks presented here; all generated code underwent author review. Notebooks produced by PaleoPAL are marked in the Preamble, with prompts retained in the markdown for full transparency. Gondhalekar used Claude and Claude Code to assist with the results of section 5.4.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

D. Khider was supported through the following grants from the US National Science Foundation (NSF): #1541029, #2126510, #2425885, and #2402618. J. Emile-Geay was supported by NSF grants #1541029, #1948822, #2002518, #2126510, #1347213, #2402476 and #2425885. J. Landers was supported by NSF Grant #2126510. T. Gondhalekar was supported by NSF Grant #2402476. N. McKay was supported by NSF grant #1347221, #1540996, #1740667, #1903548, #1929460, #1948746, #2126510, #2324733, and #2411268. D. Edge was supported by NSF Grant #2324733, #2411268, #2126510, and #1948746. C. Routson was supported by NSF grant #2002328. V. Ratnakar was supported by NSF grants #1541029 and #2126510. F. Zhu was supported by NSF grants #2303530 and #2202777. D. Garijo and Y. Gil were supported by NSF grant #1541029.

References

Abram, N. J., McGregor, H. V., Tierney, J. E., Evans, M. N., McKay, N. P., & Kaufman, D. S. (2016, AUG). Early onset of industrial-era warming across the oceans and continents. *Nature*, *536*(7617), 411-418. Retrieved from `http://dx.doi.org/10.1038/nature19082` doi: 10.1038/nature19082

Allaire, J., Xie, Y., & ... (2026). rmarkdown: Dynamic documents for r [Computer software manual]. Retrieved from https://github.com/rstudio/rmarkdown (R package version 2.31)

Arzey, A. K., McGregor, H. V., Clark, T. R., Webster, J. M., Lewis, S. E., Mallela, J., ... Fischer, M. J. (2024, OCT). Coral skeletal proxy records database for the great barrier reef, australia. *Earth System Science Data*, *16*(10), 4869-4930. Retrieved from http://dx.doi.org/10.5194/essd-16-4869-2024 doi: 10.5194/essd-16-4869-2024

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... Sherlock, G. (2000, MAY). Gene ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25-29. Retrieved from http://dx.doi.org/10.1038/75556 doi: 10.1038/75556

Barboza, L. A., Emile-Geay, J., Li, B., & He, W. (2019). Efficient Reconstructions of Common Era Climate via Integrated Nested Laplace Approximations. *Journal of Agricultural, Biological and Environmental Statistics*. Retrieved from https://rdcu.be/bLojo doi: 10.1007/s13253-019-00372-4

Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., ... Honeyman, T. (2022, October). Introducing the fair principles for research software. *Scientific Data*, *9*(1). Retrieved from http://dx.doi.org/10.1038/s41597-022-01710-x doi: 10.1038/s41597-022-01710-x

Blaauw, M., & Christen, J. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis*, *6*(3), 457-474. doi: doi:10.1214/11-BA618

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., ... Zhao, Y. (2012). Evaluation of climate models using palaeoclimate data. *Nature Climate Change*, *2*(6), 417–424. doi: 10.1038/nclimate1456

Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon*, *51*(1), 337-360.

Buttigieg, P., Morrison, N., Smith, B., Mungall, C. J., & Lewis, S. E. (2013). The environment ontology: contextualising biological and biomedical entities. *Journal of Biomedical Semantics*, *4*(1), 43. Retrieved from http://dx.doi.org/10.1186/2041-1480-4-43 doi: 10.1186/2041-1480-4-43

Camron, M. D., Khider, D., Rose, B. E. J., Grover, M. A., & May, R. M. (2025, January). The project pythia hackathon: Developing scientists' skills and community in open source development and education. In *105th american meteorological society (ams) annual meeting.* Retrieved from https://ui.adsabs.harvard.edu/abs/2025AMS...10557426C/abstract

Carson, J., Crucifix, M., Preston, S. P., & Wilkinson, R. D. (2019, APR). Quantifying age and model uncertainties in palaeoclimate data and dynamical climate models with a joint inferential analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *475*(2224), 20180854. Retrieved from http://dx.doi.org/10.1098/rspa.2018.0854 doi: 10.1098/rspa.2018.0854

Chadwick, M., Crosta, X., Esper, O., Thöle, L., & Kohfeld, K. E. (2022, AUG). Compilation of southern ocean sea-ice records covering the last glacial-interglacial cycle (12-130 ka). *Climate of the Past*, *18*(8), 1815-1829. Retrieved from http://dx.doi.org/10.5194/cp-18-1815-2022 doi: 10.5194/cp-18-1815-2022

Clark, P. U., Shakun, J. D., Rosenthal, Y., Köhler, P., & Bartlein, P. J. (2024, FEB). Global and regional temperature change over the past 4.5 million years. *Science*, *383*(6685), 884-890. Retrieved from http://dx.doi.org/10.1126/science.adi1908 doi: 10.1126/science.adi1908

Comas-Bru, L., Harrison, S. P., Werner, M., Rehfeld, K., Scroxton, N., & Veiga-Pires, C. (2019, AUG). Evaluating model outputs using integrated global speleothem records of climate change since the last glacial. *Climate of the Past*, *15*(4), 1557-1579. Retrieved from http://dx.doi.org/10.5194/cp-15-1557-2019 doi: 10.5194/cp-15-1557-2019

Comas-Bru, L., Rehfeld, K., Roesch, C., Amirnezhad-Mozhdehi, S., Harrison, S. P., Atsawawaranunt, K., ... Zhang, H. (2020, OCT). Sisalv2: a comprehensive speleothem isotope database with multiple age-depth models. *Earth System Science Data*, *12*(4), 2579-2606. Retrieved from http://dx.doi.org/10.5194/essd-12-2579-2020 doi: 10.5194/essd-12-2579-2020

Cook, E. R. (1987). The decomposition of tree-ring series for environmental studies. *Tree-Ring Bulletin*, *47*, 37–59.

Cook, E. R., Seager, R., Heim, R. R., Vose, R. S., Herweijer, C., & Woodhouse, C. (2009, DEC). Megadroughts in north america: placing ipcc projections of hydroclimatic change in a long-term palaeoclimate context. *Journal of Quaternary Science*, *25*(1), 48-61. Retrieved from http://dx.doi.org/10.1002/jqs.1303 doi: 10.1002/jqs.1303

Cooper, V. T., Armour, K. C., Hakim, G. J., Tierney, J. E., Burls, N. J., Proistosescu, C., ... Dong, Y. (2026, 2026/05/21). Paleoclimate pattern effects help constrain climate sensitivity and 21st-century warming [doi: 10.1073/pnas.2511370123]. *Proceedings of the National Academy of Sciences*, *123*(4), e2511370123. Retrieved from `https://doi.org/10.1073/pnas.2511370123` doi: 10.1073/pnas.2511370123

Cropper, S., Thackeray, C. W., & Emile-Geay, J. (2023, 2023/10/26). Revisiting a Constraint on Equilibrium Climate Sensitivity From a Last Millennium Perspective. *Geophysical Research Letters*, *50*(20), e2023GL104126. Retrieved from `https://doi.org/10.1029/2023GL104126` doi: 10.1029/2023GL104126

Dawson, A. (2016, APR). eofs: A library for eof analysis of meteorological, oceanographic, and climate data. *Journal of Open Research Software*, *4*(1), 14. Retrieved from `http://dx.doi.org/10.5334/jors.122` doi: 10.5334/jors.122

Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E., & Thompson, D. M. (2015). Prysm: An open-source framework for proxy system modeling, with applications to oxygen-isotope systems. *Journal of Advances in Modeling Earth Systems*, *7*, 1220-1247. doi: 10.1002/2015MS000447

Dolman, A. M., Kunz, T., Groeneveld, J., & Laepple, T. (2021, APR). A spectral approach to estimating the timescale-dependent uncertainty of paleoclimate records - part 2: Application and interpretation. *Climate of the Past*, *17*(2), 825-841. Retrieved from `http://dx.doi.org/10.5194/cp-17-825-2021` doi: 10.5194/cp-17-825-2021

Dolman, A. M., & Laepple, T. (2018, NOV). Sedproxy: a forward model for sediment-archived climate proxies. *Climate of the Past*, *14*(12), 1851-1868. Retrieved from `http://dx.doi.org/10.5194/cp-14-1851-2018` doi: 10.5194/cp-14-1851-2018

EarthCube Leadership Council. (2020, May 7). *Recommended standards and specifications for EarthCube projects* (Tech. Rep.). EarthCube. Retrieved from `https://doi.org/10.6075/J0QR4VMG` doi: 10.6075/J0QR4VMG

Emile-Geay, J., Hakim, G. J., Viens, F., Zhu, F., & Amrhein, D. E. (2025). Temporal comparisons involving paleoclimate data assimilation: Challenges and remedies. *Journal of Climate*, *38*(5), 1365–1385. Retrieved from `https://journals.ametsoc.org/view/journals/clim/38/5/JCLI-D-24-0101.1.xml` doi: 10.1175/JCLI-D-24-0101.1

Emile-Geay, J., McKay, N. P., Kaufman, D. S., von Gunten, L., Wang, J., Anchukaitis, K. J., . . . Zinke, J. (2017, Jul). A global multiproxy database for temperature reconstructions of the common era. *Scientific Data*, *4*(1). Retrieved from http://dx.doi.org/10.1038/sdata.2017.88 doi: 10.1038/sdata.2017.88

Erdmann, C., Stall, S., Gentemann, C., Holdgraf, C., Fernandes, F. P. A., Gehlen, K. P.-v., & Corvellec, M. (2021). *Guidance for AGU Authors - Jupyter Notebooks.* Zenodo. Retrieved from https://zenodo.org/record/4774440 doi: 10.5281/ZENODO.4774440

Evans, M. N., Lücke, L. J., Fan, K. J., & Zhu, F. (2025). A database of databases for common era paleoclimate applications. *Earth System Science Data Discussions*, *2025*, 1–25. Retrieved from https://essd.copernicus.org/preprints/essd-2025-364/ doi: 10.5194/essd-2025-364

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., & Anchukaitis, K. J. (2013). Applications of proxy system modeling in high resolution paleoclimatology. *Quaternary Science Reviews*, *76*, 16-28. doi: 10.1016/j.quascirev.2013.05.024

Feng, X., Ding, Q., Wu, L., Jones, C., Baxter, I., Tardif, R., . . . Steig, E. J. (2021). A multidecadal-scale tropically driven global teleconnection over the past millennium and its recent strengthening. *Journal of Climate*, *34*(7), 2549–2565. doi: 10.1175/JCLI-D-20-0216.1

Foster, G. (1996). Wavelets for period analysis of unevenly sampled time series. *The Astronomical Journal*, *112*(4), 1709-1729.

Fox, P., Erdmann, C., Stall, S., Griffies, S. M., Beal, L. M., Pinardi, N., . . . Legg, S. (2021). *Data and Software Sharing Guidance for Authors Submitting to AGU Journals.* Zenodo. Retrieved from https://zenodo.org/record/5124741 doi: 10.5281/ZENODO.5124741

Ghil, M., Allen, M., Dettinger, M., Ide, K., Kondrashov, D., Mann, M., . . . Yiou, P. (2002). Advanced spectral methods for climatic time series. *Reviews of Geophysics*, *40*(1), 1-41. doi: doi:10.1029/2001RG000092

Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., . . . Yu, X. (2016). Toward the geoscience paper of the future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science*, *3*(10), 388-415. doi: 10.1002/2015ea000136

Gil, Y., Garijo, D., Ratnakar, V., Khider, D., Emile-Geay, J., & McKay, N. (2017). A controlled crowdsourcing approach for practical ontology extensions and metadata annotations. In *The semantic web - iswc 2017* (p. 231-246). Springer International Publishing. Retrieved from http://dx.doi.org/10.1007/978-3-319-68204-4_24 doi: 10.1007/978-3-319-68204-4_24

Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M. R., … Schober, D. (2020, JAN). FAIR computational workflows. *Data Intelligence*, *2*(1-2), 108–121. Retrieved from https://doi.org/10.1162/dint_a_00033 doi: 10.1162/dint_a_00033

Gondhalekar, T., Khider, D., & Emile-Geay, J. (2025, OCT). *PReSto2k: updating LMRv2.1 with LinkedEarth Tools.* Retrieved from https://linked.earth/presto2k_cfr_pb/README.html

Greene, C. A., Thirumalai, K., Kearney, K. A., Delgado, J. M., Schwanghart, W., Wolfenbarger, N. S., … Blankenship, D. D. (2019, JUL). The climate data toolbox for matlab. *Geochemistry, Geophysics, Geosystems*, *20*(7), 3774-3781. Retrieved from http://dx.doi.org/10.1029/2019GC008392 doi: 10.1029/2019gc008392

Gregory, J., Eaton, B., Drach, B., Taylor, K., Hankin, S., Blower, J., … Lee, D. (2020). *Netcdf climate and forecast (cf) metadata conventions v1.8* (Report). Retrieved from {https://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.html}

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004, NOV). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, *11*, 561–566. doi: 10.5194/npg-11-561-2004

Hancock, C. L., Erb, M. P., McKay, N. P., Dee, S. G., & Ivanovic, R. F. (2024, DEC). A global data assimilation of moisture patterns from 21 000-0 bp (damp-21ka) using lake level proxy records. *Climate of the Past*, *20*(12), 2663-2684. Retrieved from http://dx.doi.org/10.5194/cp-20-2663-2024 doi: 10.5194/cp-20-2663-2024

Hancock, C. L., McKay, N. P., Erb, M. P., Kaufman, D. S., Routson, C. R., Ivanovic, R. F., … Valdes, P. (2023, June). Global synthesis of regional holocene hydroclimate variability using proxy and model data. *Paleoceanography and Paleoclimatology*, *38*(6). Retrieved from http://dx.doi.org/10.1029/2022PA004597 doi: 10.1029/2022pa004597

Hargreaves, J. C., Annan, J. D., Yoshimori, M., & Abe-Ouchi, A. (2012). Can the last glacial maximum constrain climate sensitivity? *Geophysical Research Letters*, *39*(24), L24702. Retrieved from http://dx.doi.org/10.1029/2012GL053872 doi: 10.1029/2012GL053872

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., . . . Oliphant, T. E. (2020). *Array programming with NumPy* (Vol. 585). Retrieved from https://doi.org/10.1038/s41586-020-2649-2 doi: 10.1038/s41586-020-2649-2

Harrison, S. P., Bartlein, P. J., Izumi, K., Li, G., Annan, J., Hargreaves, J., . . . Kageyama, M. (2015, JUL). Evaluation of cmip5 palaeo-simulations to improve climate projections. *Nature Climate Change*, *5*(8), 735-743. Retrieved from http://dx.doi.org/10.1038/nclimate2649 doi: 10.1038/nclimate2649

Heiser, McKay, Emile-Geay, & Khider. (2016). *lipdr.* Zenodo. Retrieved from https://zenodo.org/record/60665 doi: 10.5281/ZENODO.60665

Hodell, D. A., Crowhurst, S. J., Lourens, L., Margari, V., Nicolson, J., Rolfe, J. E., . . . Wolff, E. W. (2023, MAR). A 1.5-million-year record of orbital and millennial climate variability in the north atlantic. *Climate of the Past*, *19*(3), 607-636. Retrieved from http://dx.doi.org/10.5194/cp-19-607-2023 doi: 10.5194/cp-19-607-2023

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., . . . Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, *54*(4). doi: 10.1145/3447772

Hohmann, N., De Vleeschouwer, D., Batenburg, S., & Jarochowska, E. (2024, September). Nonparametric estimation of age-depth models from sedimentological and stratigraphic information. Retrieved from http://dx.doi.org/10.5194/egusphere-2024-2857 doi: 10.5194/egusphere-2024-2857

Hoyer, S., & Hamman, J. (2017). xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, *5*(1), 10. Retrieved from http://dx.doi.org/10.5334/jors.148 doi: 10.5334/jors.148

Hu, J., Emile-Geay, J., & Partin, J. (2017). Correlation-based interpretations of paleoclimate data - where statistics meet past climates. *Earth and Planetary Science Letters*, *459*, 362-371. doi: 10.1016/j.epsl.2016.11.048

Hunter, J. D. (2007). *Matplotlib: A 2d graphics environment* (Vol. 9) (No. 3). IEEE Computer Society. Retrieved from https://doi.org/10.1109/MCSE.2007.55 doi: 10.1109/MCSE.2007.55

Huybers, P., & Curry, W. (2006, MAY). Links between annual, milankovitch and continuum temperature variability. *Nature*, *441*(7091), 329-332. Retrieved from `http://dx.doi.org/10.1038/nature04745` doi: 10.1038/nature04745

James, A. (2023). *Ammonyte: A python package designed for conducting non-linear time series analysis of paleoclimate data.* Zenodo. Retrieved from `https://zenodo.org/record/7843881` doi: 10.5281/ZENODO.7843881

Johnsen, S. J., Clausen, H. B., Cuffey, K. M., Hoffmann, G., Schwander, J., & Creyts, T. (2000). Diffusion of stable isotopes in polar firn and ice: the isotope effect in firn diffusion. In T. Hondoh (Ed.), *Physics of ice core records* (pp. 121–140). Sapporo, Japan: Hokkaido University Press. doi: 10.7916/D8KW5D4X

Jonkers, L., Cartapanis, O., Langner, M., McKay, N., Mulitza, S., Strack, A., & Kucera, M. (2020, MAY). Integrating palaeoclimate time series with rich metadata for uncertainty modelling: strategy and documentation of the palmod 130k marine palaeoclimate data synthesis. *Earth System Science Data*, *12*(2), 1053-1081. Retrieved from `http://dx.doi.org/10.5194/essd-12-1053-2020` doi: 10.5194/essd-12-1053-2020

Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., ... Wolff, E. W. (2007, AUG). Orbital and millennial antarctic climate variability over the past 800, 000 years. *Science*, *317*(5839), 793-796. Retrieved from `http://dx.doi.org/10.1126/science.1141038` doi: 10.1126/science.1141038

Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., ... Zhilich, S. (2020). A global database of holocene paleotemperature records. *Scientific Data*, *7*(1), 115. Retrieved from `https://doi.org/10.1038/s41597-020-0445-3` doi: 10.1038/s41597-020-0445-3

Khider, D. (2026, 4). Precipitation Water Isotopes from the CESM Last Millennium Ensemble. Retrieved from `https://figshare.com/articles/dataset/Precipitation_Water_Isotopes_from_the_CESM_Last_Millennium_Ensemble/32030349` doi: 10.6084/m9.figshare.32030349.v2

Khider, D., Athreya, P., Ratnakar, V., Gil, Y., Rey, C. M. D., Zhu, F., ... Emile-Geay, J. (2021). Towards automating time series analysis for the paleogeosciences.. Retrieved from `https://api.semanticscholar.org/CorpusID:231646876`

Khider, D., Emile-Geay, J., McKay, N. P., Gil, Y., Garijo, D., Ratnakar, V., ... Zhou, Y. (2019, Oct). Pacts 1.0: A crowdsourced reporting standard for paleoclimate data. *Paleoceanography and Paleoclimatology*, *34*(10), 1570-1596. Retrieved from http://dx.doi.org/10.1029/2019PA003632 doi: 10.1029/2019pa003632

Khider, D., Emile-Geay, J., Zhu, F., James, A., & Landers, J. (2026). *Freqfun: Fun with paleoclimate analysis in the frequency domain.* Zenodo. Retrieved from https://zenodo.org/doi/10.5281/zenodo.19071441 doi: 10.5281/ZENODO.19071441

Khider, D., Emile-Geay, J., Zhu, F., James, A., Landers, J., Ratnakar, V., & Gil, Y. (2022, October). Pyleoclim: Paleoclimate timeseries analysis and visualization with python. *Paleoceanography and Paleoclimatology*, *37*(10). Retrieved from http://dx.doi.org/10.1029/2022PA004509 doi: 10.1029/2022pa004509

Khider, D., Huerta, G., Jackson, C., Stott, L., & Emile-Geay, J. (2015). A bayesian, multivariate calibration for globigerinoides ruber mg/ca. *Geochemistry Geophysics Geosystems*, *16*(9), 2916-2932. doi: 10.1002/2015GC005844

Khider, D., Jackson, C. S., & Stott, L. D. (2014). Assessing millennial-scale variability during the holocene: A perspective from the western tropical pacific. *Paleoceanography*, *29*(3), 143-159. doi: 10.1002/2013pa002534

Khider, D., Sundar, S., & Ratnakar, V. (2026). *Investigating interhemispheric precipitation changes over the past millennium.* Zenodo. Retrieved from https://zenodo.org/doi/10.5281/zenodo.19337843 doi: 10.5281/ZENODO.19337843

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., … Willing, C. (2016). Jupyter notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (p. 87 - 90).

Konecky, B. L., McKay, N. P., Churakova (Sidorova), O. V., Comas-Bru, L., Dassié, E. P., DeLong, K. L., … Members, I. P. (2020). The iso2k database: a global compilation of paleo-$\delta^{18}o$ and $\delta^{2}h$ records to aid understanding of common era climate. *Earth System Science Data*, *12*(3), 2261–2288. Retrieved from https://essd.copernicus.org/articles/12/2261/2020/ doi: 10.5194/essd-12-2261-2020

Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., … Capella-Gutierrez, S. (2020). Towards fair principles for research software. *Data Science*, *3*(1), 37-59. Retrieved from https://content.iospress.com/articles/data-science/ds190026 doi: 10.3233/DS-190026

Landers, J. P., Emile-Geay, J., James, A. K., Munch, S. B., Khider, D., & Bard, E. (2025). A causal examination of the solar influence on Holocene climate. *ESS Open Archive*. doi: 10.22541/essoar.176598901.17168843/v1

Lawrence, K. T., Liu, Z., & Herbert, T. D. (2006, APR). Evolution of the eastern tropical pacific through plio-pleistocene glaciation. *Science*, *312*(5770), 79-83. Retrieved from `http://dx.doi.org/10.1126/science.1120395` doi: 10.1126/science.1120395

Lee, T., Rand, D., Lisiecki, L. E., Gebbie, G., & Lawrence, C. (2023, OCT). Bayesian age models and stacks: combining age inferences from radiocarbon and benthic δ^{18}o stratigraphic alignment. *Climate of the Past*, *19*(10), 1993-2012. Retrieved from `http://dx.doi.org/10.5194/cp-19-1993-2023` doi: 10.5194/cp-19-1993-2023

Lin, L., Khider, D., Lisiecki, L., & Lawrence, C. (2014a). Probabilistic sequence alignment of stratigraphic records. *Paleoceanography*, *29*(976-989), 976-989. doi: 10.1002/2014PA002713

Lin, L., Khider, D., Lisiecki, L., & Lawrence, C. (2014b). Probabilistic sequence alignment of stratigraphic records. *Paleoceanography*, *29*(976-989), 976-989. doi: 10.1002/2014PA002713

Lisiecki, L. E., & Lisiecki, P. A. (2002). Application of dynamic programming to the correlation of paleoclimate records. *Paleoceanography*, *17*(4), 1-1-1-12. doi: 10.1029/2001pa000733

Lisiecki, L. E., & Raymo, M. E. (2005). A pliocene-pleistocene stack of 57 globally distributed benthic δ^{18}o records. *Paleoceanography*, *20*(PA1003). doi: 1010.1029/2004PA001071

Luo, X., Dee, S., Lavenhouse, T., Muñoz, S., & Steiger, N. (2023, 2023/08/17). Tropical pacific and north atlantic sea surface temperature patterns modulate mississippi basin hydroclimate extremes over the last millennium. *Geophysical Research Letters*, *50*(2), e2022GL100715. Retrieved from `https://doi.org/10.1029/2022GL100715` doi: https://doi.org/10.1029/2022GL100715

McKay, N., Edge, D., Hancock, C., Emile-Geay, J., & Erb, M. (2024). *The abrupt change toolkit in R.* Zenodo. Retrieved from `https://zenodo.org/records/11284536` doi: 10.5281/zenodo.11284536

McKay, N., & Emile-Geay, J. (2016). Technical note: The linked paleo data framework - a common tongue for paleoclimatology. *Climate of the Past*, *12*, 1093-1100. doi: 10.5194/cp-12-1093-2016

McKay, N., Emile-Geay, J., & Khider, D. (2021, MAR). geochronr - an r package to model, analyze, and visualize age-uncertain data. *Geochronology*, *3*(1), 149-169. Retrieved from `http://dx.doi.org/10.5194/gchron-3-149-2021` doi: 10.5194/gchron-3-149-2021

McKay, N., & Hancock, C. (2023). *nickmckay/compositer.* Zenodo. Retrieved from `https://zenodo.org/records/7951514` doi: 10.5281/zenodo.7951514

McKay, N. P., Kaufman, D. S., Arcusa, S. H., Kolus, H. R., Edge, D. C., Erb, M. P., ... Telles, F. (2024, August). The 4.2 ka event is not remarkable in the context of holocene climate variability. *Nature Communications*, *15*(1). Retrieved from `http://dx.doi.org/10.1038/s41467-024-50886-w` doi: 10.1038/s41467-024-50886-w

McPartland, M. Y., Münch, T., Dolman, A. M., Hébert, R., & Laepple, T. (2025, NOV). The colors of proxy noise. *Climate of the Past*, *21*(11), 1917-1931. Retrieved from `http://dx.doi.org/10.5194/cp-21-1917-2025` doi: 10.5194/cp-21-1917-2025

Morrill, C., Thrasher, B., Lockshin, S. N., Gille, E. P., McNeill, S., Shepherd, E., ... Bauer, B. A. (2021, Jun). The paleoenvironmental standard terms (past) thesaurus: Standardizing heterogeneous variables in paleoscience. *Paleoceanography and Paleoclimatology*, *36*(6). Retrieved from `http://dx.doi.org/10.1029/2020PA004193` doi: 10.1029/2020pa004193

National Academies of Sciences, Engineering, and Medicine. (2022). *Next generation earth systems science at the national science foundation.* National Academies Press. Retrieved from `https://doi.org/10.17226/26042` doi: 10.17226/26042

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., ... von Gunten, L. (2019, JUL). Consistent multidecadal variability in global temperature reconstructions and simulations over the common era. *Nature Geoscience*, *12*(8), 643-649. Retrieved from `http://dx.doi.org/10.1038/s41561-019-0400-0` doi: 10.1038/s41561-019-0400-0

Nugroho, R. P., Zuiderwijk, A., Janssen, M., & de Jong, M. (2015, AUG). A comparison of national open data policies: Lessons learned. *Transforming Government: People, Process and Policy*, *9*(3), 286–308. Retrieved from `https://doi.org/10`

.1108/tg-03-2014-0008 doi: 10.1108/tg-03-2014-0008

Osborne, J. D., Flatow, J., Holko, M., Lin, S. M., Kibbe, W. A., Zhu, L., ... Chisholm, R. L. (2009, Jul). Annotating the human genome with disease ontology. *BMC Genomics*, *10*(S1). Retrieved from http://dx.doi.org/10.1186/1471-2164-10-S1-S6 doi: 10.1186/1471-2164-10-s1-s6

Osman, M. B., Tierney, J. E., Zhu, J., Tardif, R., Hakim, G. J., King, J., & Poulsen, C. J. (2021, NOV). Globally resolved surface temperatures since the last glacial maximum. *Nature*, *599*(7884), 239-244. Retrieved from http://dx.doi.org/10.1038/s41586-021-03984-4 doi: 10.1038/s41586-021-03984-4

Oswal, D., Khider, D., & Pujara, J. (2025). *Pyleotups: Automated paleoclimate data extraction and processing.* Zenodo. Retrieved from https://zenodo.org/doi/10.5281/zenodo.16009165 doi: 10.5281/ZENODO.16009165

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., ... Strand, G. (2016). Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model. *Bulletin of the American Meteorological Society*, *97*(5), 735–754. doi: 10.1175/BAMS-D-14-00233.1

PAGES 2k Consortium. (2013, 05). Continental-scale temperature variability during the past two millennia. *Nature Geosci*, *6*(5), 339–346. Retrieved from http://dx.doi.org/10.1038/ngeo1797 doi: 10.1038/ngeo1797

PAGES 2k Consortium. (2017, 07). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*, 170088 EP. doi: 10.1038/sdata.2017.88

Pandas development team, T. (2020, February). *pandas-dev/pandas: Pandas.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.3509134 doi: 10.5281/zenodo.3509134

Parnell, A., Haslett, J., Allen, J., Buck, C. E., & Huntley, B. (2008). A flexible approach to assessing synchroneity of past events using bayesian reconstructions of sedimentation history. *Quaternary Science Reviews*, *27*(19-20), 1872-1885.

Parrenin, F., Bouchet, M., Buizert, C., Capron, E., Corrick, E., Drysdale, R., ... Rasmussen, S. O. (2024, DEC). The paleochrono-1.1 probabilistic model to derive a common age model for several paleoclimatic sites using absolute and relative dating constraints. *Geoscientific Model Development*, *17*(23), 8735-8750. Retrieved from http://dx.doi.org/10.5194/gmd-17-8735-2024 doi:

10.5194/gmd-17-8735-2024

Peng, R. D. (2011, DEC). Reproducible research in computational science. *Science*, *334*(6060), 1226-1227. Retrieved from `http://dx.doi.org/10.1126/science.1213847` doi: 10.1126/science.1213847

Ratnakar, V., & Khider, D. (2025, NOV). Pylipd: A python package for the manipulation of paleoclimate datasets. *Journal of Open Source Software*, *10*(115), 8861. Retrieved from `http://dx.doi.org/10.21105/joss.08861` doi: 10.21105/joss.08861

Ratnakar, V., & Khider, D. (2026). *Paleopal: An ai assistant for paleoclimatology.* Zenodo. Retrieved from `https://zenodo.org/doi/10.5281/zenodo.19341101` doi: 10.5281/ZENODO.19341101

Ratnakar, V., Khider, D., Gajiro, D., Emile-Geay, J., McKay, N., Edge, D., … Bradley, E. (2026). *The linked earth ontology: A modular, extensible representation of open paleoclimate data.* Zenodo. Retrieved from `https://zenodo.org/doi/10.5281/zenodo.19056116` doi: 10.5281/ZENODO.19056116

Routson, C. C., Kaufman, D. S., McKay, N. P., Erb, M. P., Arcusa, S. H., Brown, K. J., … Cumming, B. F. (2021, APR). A multiproxy database of western north american holocene paleoclimate records. *Earth System Science Data*, *13*(4), 1613-1632. Retrieved from `http://dx.doi.org/10.5194/essd-13-1613-2021` doi: 10.5194/essd-13-1613-2021

Routson, C. C., McKay, N. P., Kaufman, D. S., Erb, M. P., Goosse, H., Shuman, B. N., … Ault, T. (2019). Mid-latitude net precipitation decreased with arctic warming during the holocene. *Nature*, *568*(7750), 83–87.

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013, OCT). Ten simple rules for reproducible computational research. *PLoS Computational Biology*, *9*(10), e1003285. Retrieved from `http://dx.doi.org/10.1371/journal.pcbi.1003285` doi: 10.1371/journal.pcbi.1003285

Schulz, M., & Mudelsee, M. (2002). Redfit: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. *Computers and Geosciences*, *28*, 421-426.

Shakun, J. D., Clark, P. U., He, F., Marcott, S. A., Mix, A. C., Liu, Z., … Bard, E. (2012, APR). Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature*, *484*(7392), 49-54. Retrieved from `http://dx.doi.org/10.1038/nature10915` doi: 10.1038/nature10915

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., ... Zelinka, M. D. (2020, 2021/02/28). An assessment of earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), e2019RG000678. Retrieved from `https://doi.org/10.1029/2019RG000678` doi: 10.1029/2019RG000678

Singh, H. K. A., Hakim, G. J., Tardif, R., Emile-Geay, J., & Noone, D. C. (2018). Insights into atlantic multidecadal variability using the last millennium reanalysis framework. *Climate of the Past*, *14*(2), 157–174. Retrieved from `https://www.clim-past.net/14/157/2018/` doi: 10.5194/cp-14-157-2018

Smerdon, J. E., Cook, E. R., & Steiger, N. J. (2023). The historical development of large-scale paleoclimate field reconstructions over the common era. *Reviews of Geophysics*, *61*(4). doi: 10.1029/2022rg000782

Snyder, C. W. (2016, SEP). Evolution of global temperature over the past two million years. *Nature*, *538*(7624), 226-228. Retrieved from `http://dx.doi.org/10.1038/nature19798` doi: 10.1038/nature19798

Stall, S., Robinson, E., Wyborn, L., Yarmey, L. R., Parsons, M. A., Lehnert, K., ... B., H. (2017). Enabling FAIR data across the Earth and space sciences. *Eos*, *98*. doi: 10.1029/2017EO08842

Steinman, B. A., Stansell, N. D., Mann, M. E., Cooke, C. A., Abbott, M. B., Vuille, M., ... Fernandez, A. (2022, April). Interhemispheric antiphasing of neotropical precipitation during the past millennium. *Proceedings of the National Academy of Sciences*, *119*(17). Retrieved from `http://dx.doi.org/10.1073/pnas.2120015119` doi: 10.1073/pnas.2120015119

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., ... Taufer, M. (2016, DEC). Enhancing reproducibility for computational methods. *Science*, *354*(6317), 1240-1241. Retrieved from `http://dx.doi.org/10.1126/science.aah6168` doi: 10.1126/science.aah6168

Stommel, H., & Stommel, E. (1979, June). The year without a summer. *Scientific American*, *240*(6), 176–187. Retrieved from `https://www.jstor.org/stable/24965226`

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., ... Noone, D. (2019, JUL). Last millennium reanalysis with an expanded proxy database and seasonal proxy modeling. *Climate of the Past*, *15*(4), 1251-1273. Retrieved from http://dx.doi.org/10.5194/cp-15-1251-2019 doi: 10.5194/cp-15-1251-2019

Thomson, D. J. (2009). Time-series analysis of paleoclimate data. In *Encyclopedia of paleoclimatology and ancient environments* (p. 949-959). Springer Netherlands. Retrieved from http://dx.doi.org/10.1007/978-1-4020-4411-3_222 doi: 10.1007/978-1-4020-4411-3_222

Tierney, J. E., Judd, E. J., Osman, M. B., King, J. M., Truax, O. J., Steiger, N. J., … Anchukaitis, K. J. (2025, MAY). Advances in paleoclimate data assimilation. *Annual Review of Earth and Planetary Sciences*, *53*(1), 625-650. Retrieved from http://dx.doi.org/10.1146/annurev-earth-032320-064209 doi: 10.1146/annurev-earth-032320-064209

Tierney, J. E., Poulsen, C. J., Montañez, I. P., Bhattacharya, T., Feng, R., Ford, H. L., … others (2020). Past climates inform our future. *Science*, *370*(6517). doi: 10.1126/science.aay3701

Tierney, J. E., & Tingley, M. P. (2014). A bayesian, spatially-varying calibration model for the tex86 proxy. *Geochimica et Cosmochimica Acta*, *127*, 83-106. doi: 10.1016/j.gca.2013.11.026

Tierney, J. E., & Tingley, M. P. (2018, MAR). Bayspline: A new calibration for the alkenone paleothermometer. *Paleoceanography and Paleoclimatology*, *33*(3), 281-301. Retrieved from http://dx.doi.org/10.1002/2017PA003201 doi: 10.1002/2017pa003201

Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., & Poulsen, C. J. (2020, AUG). Glacial cooling and climate sensitivity revisited. *Nature*, *584*(7822), 569–573. doi: 10.1038/s41586-020-2617-x

Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., & Rajaratnam, B. (2012). Piecing together the past: statistical insights into paleoclimatic reconstructions. *Quaternary Science Reviews*, *35*, 1-22. doi: 10.1016/j.quascirev.2012.01.012

Torrence, C., & Compo, G. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*, 61-78.

Vieira, L. E. A., Solanki, S. K., Krivova, N. A., & Usoskin, I. (2011, MAY). Evolution of the solar irradiance during the holocene. *Astronomy & Astrophysics*, *531*, A6. Retrieved from http://dx.doi.org/10.1051/0004-6361/201015843 doi: 10.1051/0004-6361/201015843

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … van Mulbregt, P. (2020). *SciPy 1.0: Fundamental algorithms for scientific computing in python* (Vol. 17). Retrieved from https://www.scipy.org doi: 10.1038/s41592-019-0686-2

von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., González-Rouco, F., & Tett, S. F. B. (2004, OCT). Reconstructing past climate from noisy data. *Science*, *306*(5696), 679-682. Retrieved from http://dx.doi.org/10.1126/science.1096109 doi: 10.1126/science.1096109

Walter, R. M., Sayani, H. R., Felis, T., Cobb, K. M., Abram, N. J., Arzey, A. K., … the PAGES CoralHydro2k Project Members (2023). The CoralHydro2k database: a global, actively curated compilation of coral δ^{18}O and Sr / Ca proxy records of tropical ocean hydrology and temperature for the Common Era. *Earth System Science Data*, *15*(5), 2081–2116. Retrieved from https://essd.copernicus.org/articles/15/2081/2023/ doi: 10.5194/essd-15-2081-2023

Weitzel, N., Racky, M., & Braschoss, L. (2025). *cupsm v1.0.0: A python package for proxy system modeling with data cubes.* Zenodo. Retrieved from https://zenodo.org/doi/10.5281/zenodo.14624116 doi: 10.5281/ZENODO.14624116

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016, Mar). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1). Retrieved from http://dx.doi.org/10.1038/sdata.2016.18 doi: 10.1038/sdata.2016.18

Wilkinson, S. R., Aloqalaa, M., Belhajjame, K., Crusoe, M. R., de Paula Kinoshita, B., Gadelha, L., … Goble, C. (2025, February). Applying the fair principles to computational workflows. *Scientific Data*, *12*(1). Retrieved from http://dx.doi.org/10.1038/s41597-025-04451-9 doi: 10.1038/s41597-025-04451-9

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017, JUN). Good enough practices in scientific computing. *PLOS Computational Biology*, *13*(6), e1005510. Retrieved from http://dx.doi.org/10.1371/journal.pcbi.1005510 doi: 10.1371/journal.pcbi.1005510

Zhu, F., Emile-Geay, J., Hakim, G. J., Guillot, D., Khider, D., Tardif, R., & Perkins, W. A. (2024, APR). cfr (v2024.1.26): a python package for climate field reconstruction. *Geoscientific Model Development*, *17*(8), 3409-3431. Retrieved from `http://dx.doi.org/10.5194/gmd-17-3409-2024` doi: 10.5194/gmd-17-3409-2024

Zhu, F., Emile-Geay, J., McKay, N. P., Hakim, G. J., Khider, D., Ault, T. R., … Kirchner, J. W. (2019, APR). Climate models can correctly simulate the continuum of global-average temperature variability. *Proceedings of the National Academy of Sciences*, *116*(18), 8728-8733. Retrieved from `http://dx.doi.org/10.1073/pnas.1809959116` doi: 10.1073/pnas.1809959116