UNIVERSITÉ DE LAUSANNE

DOCTORAL THESIS

Software and Numerical Tools for Paleoclimate Analysis

Author:
Philipp S. SOMMER

Supervisor:
Dr. Basil A. S. Davis

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Science

in the

The Davis Group
Institute of Earth Surface Dynamics (IDYST)

October 7, 2019

Contents

Chapter 1

Introduction

1.1 Motivation

rewrite the motivation – does not read well

Climate science and in particular the study of past climates face an increasing need for the analysis, standardization and sharing of data. Scientists made huge efforts to explore climate archives throughout the world to investigate the evolution of the Earths climate. In parallel, computational climate models grew in complexity and data ouput due to an increase of computational power and the availability of supercomputers. This generates new challenges for big data analysis that can only be solved by high quality and flexible software packages.

1: Add reference.

2: Add reference.

The Neotoma Database, a global international database for paleoenvironmental proxies (Williams et al., 2018) currently lists XXX datasets with in total XXX samples for past 12'000 years, the Holocene. Such data collections enable large-scale reconstructions of past climates that however face considerable challenges. They mainly arise from the heterogeneity of the data and the necessity of further quality control and standardization. One key problem is the accessibility of data. A lot of data is not available in standardized relational databases and either held private, or is stored in less standardized archives such as PANGAEA, or is not available in a digital format at all. The latter often results in the need of digitizing the associated data from a published diagram, a tedious and imprecise task (Sommer et al., 2019). Additionally handling such a big heterogeneous data resource and analyzing its contents is a key challenge and requires flexible visualization resources that efficiently allow the querying of spatial data with heterogeneous time and meta data information.

3: Add reference. `https://pangaea.de/`

4: Add reference. EMPD paper

An additional challenge arises from the combination with numerical models that usually operate on a structured (Edwards, 2010; Treut et al., 2007) or unstructured grid with a fixed timestep. The development and analysis of such models requires visualization techniques that are interoperable with the specific data structure of the model (e.g. S. A. Brown et al., 1993; Rew and G. Davis, 1990) while still being flexible enough for general purposes and computations (Hoyer and Hamman, 2017; Sommer, 2017). Additionally it requires techniques to process observational data to make it comparable with climate models (Mauri et al., 2015) or to feed a model with the data using data assimilation of statistical models (Sommer and Kaplan, 2017).

5: Add reference. ICON

6: Add reference. POLNET-gridding paper

In the following section 1.2 I will lay down the interest in the study of paleoclimates, both from the observational and the modellers perspective. This is continued by a section 1.3 which highlights the specific requirements and the historical development of software in paleo-science and concludes with section 1.3.3 that provides an overview on the contents of this thesis.

1.2 Learning from the Past – Why we study paleo-climates

7: Add reference.

8: Add reference.

9: Add reference. cite World bank report?

10: Add reference.

Mankind is facing large infrastructural challenges during this century, such as the loss of biodiversity, an exponentially growing world population and an acceleration of growth and globalization of markets. They all interact with a global climate change that may lead to a new environment none of us ever experienced. Any future global planning has to account highly diverse responses that range from regional to continental scales. The complex (climate) system will enter a state that is significantly different from everything we had since the beginning of the satellite era in the 19th century, the beginning of global meteorological data acquisition.

Our knowledge about this new climate is therefore mainly based on computational Earth System Models (ESMs). They face the challenge of simulating a new climate based on our present knowledge of the interactions between the different compartments Ocean, Land and Atmosphere. Running such a model for the entire Earth with a reasonable resolution is therefore very cost-intensive and requires large computational resources. The validation of it becomes technically difficult considering the large amount of data output, and conceptually difficult because of the aforementioned transition into a warmer world during the next century. We are entering a new state and it is questionable how well our models perform (Hargreaves et al., 2013; Karpechko, 2010; Ulden and Oldenborgh, 2006).

11: Add reference. check these references! taken from Achilles PhD thesis, there might be better ones

To evaluate their skill, we can only use our knowledge of the past climate from before the systematic measurement of temperature, precipitation, etc. These climates, also referred to as paleo-climates, provide the only opportunity to evaluate an ESM under conditions very different than today. paleo-climatic research has therefore been an integral part for climate sciences since the 80s (COHMAP Members, 1988; Joussaume and Taylor, 1995), particularly in the Paleoclimate Modelling Intercomparison Project (PMIP) (Braconnot et al., 2012, 2007a,b; Jungclaus et al., 2017; Kageyama et al., 2016; B. L. Otto-Bliesner et al., 2017).

The current geological period is the Quaternary. It is characterized by glacial-interglacial cycles mainly driven by orbital changes (Hays et al., 1976; Imbrie et al., 1993)that cause a varying insolation on the planet.

12: Add reference. Check these

The end of this period can be used for data-model comparisons due to the availability of paleo-climate archives. It started with the Last Interglacial (LIG) about 127’000 years ago and was followed by the Last Glacial Maximum (LGM) at 21’000 years ago. The warming of the atmosphere in the following interglacial has been interrupted by a rapid cooling, called the Younger Dryas, between 12’900 and 11’700 years ago, which then let to the onset of the current epoch, the Holocene (Walker et al., 2009) .

13: Add reference. check ibid.

Add some background on the Holocene. How did it change (global mean temperature estimate?), how was the insolation? CO_2 effects, impact of the ice sheets during the early holocene, changes in altitude, large-scale atmospheric circulation, human influences.

This epoch is of particular interest because the continental setup is comparable to nowadays while still having a climate that is significantly different from present day . Additionally we have a large set of proxies available to quantify the climate, independent from the model estimates, and for the entire globe (Wanner et al., 2008) .

14: Add reference. PMIP paper

15: Add reference. check ibid.

1.2.1 Proxy Data from the Holocene

Before 1850, there is almost no instrumental measurement of temperature. Instead we rely on archives such as lake sediments, glaciers, peat bogs, or speleothems that preserve climate proxies. The latter is a set of variables that are influenced by climate conditions and therefore allow an indirect measurement of climate parameters at ancient times, e.g. temperature, precipitation or sea-level. The most prominent proxies are isotopic compositions of δ^{18}O in glacial ice cores, marine sediments , peat bogs or speleothems ; bio-ecologic assemblages such as pollen , chironomids or diatoms in lake sediments; foraminifera and alkenone in marine sediments ; and the widths of tree rings.

16: Add reference.

17: Add reference.

18: Add reference.

19: Add reference.

20: Add reference.

21: Add reference.

22: Add reference.

23: Add reference.

The most abundant climate proxy, that I will also focus on in the next chapters, are pollen assemblages. It is the geographically most wide spread paleo-climate proxy (Harry John Betteley Birks and H. H. Birks, 1980) and has a long history in quantitative paleo-climatologic reconstructions (e.g. Bradley, 1985; Nichols, 1967, 1969).

The ability to serve as a proxy for the past arises from the chemically stable polymer sporopollenin, that allows it to be preserved over very long periods of time, in various environments such as lakes, wetlands or ocean sediments (Fægri et al., 1989; Havinga, 1967). Pollen are produced by seed-bearing plants (spermatophytes, Wodehouse, 1935) and as such have a high spatial continuity and prevalence. Their compositions (closely related to the surrounding vegetation) is highly dependent on the climate and allows the reconstruction of the latter through an inverse modelling approach (S. Brewer et al., 2007; Juggins, 2013; Juggins and H. John B. Birks, 2012).

24: Add reference. Don't know about ibid., took it from Manus review paper...

25: Add reference. Manus review paper

The usefulness for large-scale data-model intercomparisons additionally arises from the existence of regional databases for fossil pollen assemblages. The earliest examples are the European Pollen Database (EPD) and North American Pollen Database (NAPD) that both started around 1990 and developed a similar structure in order to be compatible (Fyfe et al., 2009; Grimm, 2008) . This let to the development of other regional pollen databases, such as the Latin American Pollen Database (LAPD) (LAPD, Flantua et al., 2015; Marchant et al., 2002) in 1994 or the African Pollen Database (APD) (Vincens et al., 2007) in 1996, and others (see Grimm, 2008). These attempts finally let to the development of the Neotoma database (Williams et al., 2018), a global multiproxy database that incorporates many of the regional pollen databases.

26: Add reference. Don't know about ibid., took it from Manus review paper...

27: Add reference. Other review papers?

The use of the above-mentioned proxies, particularly pollen, for paleo-climate reconstruction has a long academic tradition in geology (Bradley, 1985) and provides the source of large-scale paleo-climatic reconstructions in number of different studies (B. A. S. Davis et al., 2003; Marsicek et al., 2018; Mauri et al., 2015; Neukom et al., 2019a,b). They however have multiple uncertainties, often difficult to quantify and to consider (see chapter 6). A key challenge for a data-model comparison are dating uncertainties, or the influence of seasonality on the proxy (e.g. whether it represents summer, winter or annual temperature) and the quality of the record. Another challenge is the proper handling of uncertainties of the inverse modelling approach, spatial coverage of the proxy (see chapter 3), and, considering pollen assemblages, the various naming schemes for pollen taxa that need to be considered for large-scale reconstructions.

28: Add reference. add more..., Climate12K

29: Add reference. cite some MAT papers

30: Add reference. that North-US/South-US discrepancy...

1.2.2 Model Simulations of the Holocene

As mentioned in the earlier section 1.2, paleoclimate simulations of ESMs played in important role in previous intercomparisons. The Holocene analysis within past

Paleoclimate Modelling Intercomparison Project (PMIP) versions focused mainly on the mid-Holocene around 6000 years before present, a time period with a different latitudinal and seasonal distribution of incoming solar radiation (insolation) but greenhouse gas concentrations similar to the preindustrial period (B. L. Otto-Bliesner et al., 2017). The main findings from previous intercomparisons are an underestimation of polar amplification in PMIP2 and PMIP3 models due to sea ice and vegetational feedbacks, and an underestimation of the north-south temperature gradient over Europe (Simon Brewer et al., 2007; Basil A. S. Davis and Simon Brewer, 2009; Intergovernmental Panel on Climate Change, 2014; Masson-Delmotte et al., 2006; Zhang et al., 2010).

The focus on only a short time slice (mainly due to the high computational costs for running an ESM over a large simulation period) however has several complications. Climate changes, i.e. the shift into a different climatic state, cannot be simulated and high dating uncertainties hinder a credible comparison of models and proxies. Therefore multiple recent studies published and proposed increasing efforts for transient model simulations, i.e. simulations that cover multiple millenia during the last deglaciation (Ivanovic et al., 2016) and the Holocene (B. L. Otto-Bliesner et al., 2017). Several studies used Earth System Models of Intermediate Complexity (EMICs) for transient simulations that cover parts of the last 12'000 years (e.g. Gregoire et al., 2015; Menviel et al., 2011; Roche et al., 2011) and a few used a global coupled ESM (Bette L. Otto-Bliesner et al., 2014; Varma et al., 2012). As stated by Weitzel et al., 2019, those model results can clarify the role of internal climate variability for Holocene temperature trends and large-scale patterns.

31: Add reference.

Despite the computational costs for running these models, technical challenges arise from the size of the data that easily exceeds the size of multiple Gigabyte per climatic variable with a monthly resolution. It requires software that is able to deal with data too large to fit into memory (see section 1.3.2 and chapter 2) and automated techniques to identify patterns in the data.

1.3 Software for Paleoclimatology

The usage of software is crucial for the quantitative reconstruction of Earth's Climate. Paleoclimate research is facing an information overload problem and requires innovative methodologies in the realm of visual analytics, the interplay between automated analysis techniques and interactive visualization (Keim et al., 2008; Nocke, 2014). As such, a visual representation of the paleoclimate reconstruction has been essential for both, proxies (Bradley, 1985; Grimm, 1988; Nichols, 1967) and models (Böttinger and Röber, 2019; Nocke, 2014; Nocke et al., 2008; Phillips, 1956; Rautenhaus et al., 2018), although the visualization methods significantly differ due to the differences in data size and data heterogeneity.

The second important aspect for software and paleoclimate is the distribution of data to make it accessible to other researchers, the community and policy makers, which is commonly established through online accessible data archives and recently also through map-based web interfaces (Bolliet et al., 2016; Williams et al., 2018).

32: Add reference. cite some open-data publications

The following sections provide an overview on the different techniques used by modelers and palynologists to visualize and distribute their data and concludes with an introduction into Open-Source Software Development, which forms the basis of the software solutions that are presented later in this thesis (chapters 2, 3, and 4, and appendix C).

1.3.1 Sofware for Proxy Data Analysis, Visualization and Distribution

Due to the nature of stratigraphic data, proxies, especially pollen assemblages, are often treated as a collection of multiple time-series (one-dimensional arrays). The size of one dataset is generally small (in the range of kB) and can be treated as plain text files. Traditionally, numerical and statistical analysis are separated from the visualization.

In palynology, standard analytical tools are Microsoft Excel[1] and the R software for statistical computing (R Core Team, 2019). The latter also involves multiple packages for paleoclimatic reconstruction, such as `rioja` (Juggins, 2017) and `analogue` (Simpson, 2007; Simpson and Oksanen, 2019) or bayesian methods (Nolan et al., 2019; Tipton, 2017). Alternatively, desktop applications exist, such as Polygon[2] by Nakagawa et al., 2002 or the CREST software presented in M. Chevalier et al., 2014; Manuel Chevalier, 2019.

33: Add reference. add more?

It is a long-standing tradition to visualize stratigraphic data, and especially pollen data, in form of a stratigraphic (pollen) diagram (Bradley, 1985; Grimm, 1988). Especially during the 19th century, when it was not yet common to distribute data alongside a peer-reviewed publication, pollen diagrams have been the only possibility to publish the entire dataset (see also chapter 3). The generation of these diagrams is usually based on desktop applications such as C2 (Juggins, 2007), Tilia[3] (Grimm, 1988, 1991). A more recent implementation into the psyplot framework (Sommer, 2017, chapter 2) is also provided with the psy-strat plugin[4] (Sommer, 2019).

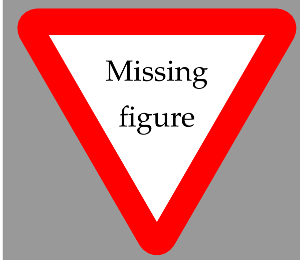
Raw pollen data is at present made available through web archives, such as PANGAEA[5] or the National Climatic Data Center (NCDC) by the National Oceanic and Atmospheric Administration (NOAA)[6] where researchers can create a DOI for their raw data. Collections of data, such as regional pollen databases or project specific collections (e.g. B. A. S. Davis et al., 2013; Whitmore et al., 2005) are usually published in one of the above-mentioned archives or associated with a publication. A different approach has been developed by Bolliet et al., 2016 to develop a small web application as an interface into the data collection, the *ClimateProxiesFinder* (Brockmann, 2016, chapter 4).

Outstanding compared to the previous data interfaces is the new infrastructure for the Neotoma database (Williams et al., 2018). It consists of the map-based web interface, the Neotoma Explorer[7], a RESTful api[8] that allows an interaction with other web services, the neotoma R package (Goring et al., 2015) and an interface into the Tilia software for stratigraphic and map-based visualizations (Williams et al., 2018). This rich functionality is, however, bound to the structure of Neotoma and as such, different from the Javascript-based approach developed in chapter 4 cannot easily be transferred to other projects.

1.3.2 The Development of Computational Climate Model Analysis

Software and computational numerics play a crucial role for our understanding of climate since the first General Circulation Models (GCMs) by Phillips, 1956 after

[1] https://products.office.com/en/excel
[2] http://polsystems.rits-paleo.com
[3] https://www.tiliait.com/
[4] https://psy-strat.readthedocs.io
[5] https://pangaea.de/
[6] https://www.ncdc.noaa.gov/data-access/paleoclimatology-data
[7] https://apps.neotomadb.org/Explorer
[8] https://api.neotomadb.org



Visualize multiple grids on the same map, e.g. by using the grid specifications from Treut et al., 2007.

FIGURE 1.1: A selection of grid sizes and formats since the first IPCC report

world war II (Edwards, 2010; Lewis, 1998). The first simulations and analysis of GCMs where limited by the available computational facilities, the model by Phillips, 1956 for example operated on a 17×16 grid simulating a surface with the size of roughly one tenth of the Earth. The possibilities of climate modeling increased rapidly, mainly due to a drastic increase in computational capacity and the availability of supercomputers. This let to an increase in speed by a factor of roughly one million between 1970 and 2007, permitting an increase in model complexity, length of the simulations, and spatial resolution (Treut et al., 2007). In the past decade, unstructured grids raised more and more attention (Skamarock et al., 2012; Zängl et al., 2014) especially with the focus on *seamless prediction* (Bauer et al., 2015; Hoskins, 2012) that allows a refined grid resolution in selected regions of the earth (Rautenhaus et al., 2018).

The varying grids, multi-dimensionality and volume of the data requires visual analytic methodologies much different from what is used with proxy data (section 1.3.1) and much more diverse. In general, scientists tend to disentangle the numeric post-processing of climate model output (such as computing aggregated statistics in time or space dimension) and then visualize these aggregated results (Böttinger and Röber, 2019; Schulz et al., 2013). Common post-processing software are, for example Climate Data Operators (CDOs) (Schulzweida, 2019), netCDF Operators (NCOs) (Zender, 2008, 2016; Zender and Mangalam, 2007) and R (R Core Team, 2019), or more recently also python packages such as `xarray` (Hoyer and Hamman, 2017). The choice of method thereby depends on the scientiests preference but also on the size of the data. Especially the analysis of transient model runs (section 1.2.2) requires software that is able to deal with data that is too big to fit into memory and requires parallel computation. Such an analysis can be executed with CDOs and `xarray` through it's interface with the parallel computing package dask (Dask Development Team, 2016; Rocklin, 2015) (see also chapter 6).

Rautenhaus et al., 2018 provide a detailed overview about how such a high amount of data is visualized. Methods range from 2D projection on a map to 3D interactive visualizations, depending on the background and knowledge of the researcher (Nocke, 2014). Nocke et al., 2008 recognize preference for script-based solutions such as Python, R or domain-specific languages such as NCL (D. Brown et al., 2012) that is still persistent today (Rautenhaus et al., 2018). Nocke, 2014, Schulz et al., 2013 and Rautenhaus et al., 2018 attribute this to the importance of comparability and reproducibility for research, and the usability in peer-reviewed publications. Therefore, 3D visualization, e.g. with ParaView (Ayachit, 2015), VAPOR

34: Add reference.

(Clyne et al., 2007) or Avizo, are mainly used by visualization experts rather than scientists (Nocke et al., 2008). However, Nocke, 2014 note that young researchers are more open for 3D visualization and especially with the newly emerging unstructured grids, they become more prominent. Besides psyplot (Sommer, 2017, chapter 2) and UV-CDAT, there exists no scripted method that easily visualizes climate model data on unstructured grids, without the need for interpolation to a rectilinear grid. These have however been implemented for ParaView (Röber et al., 2015) and Vapor (Jubair et al., 2015) and new software packages emerge e.g. Sullivan and Kaszynski, 2019; Sullivan and Trainor-Guitton, 2019.

Look into ibid.

35: Add reference.

1.3.3 Methods and Workflows in Open-Source Software Development

The importance and necessity of software for visualization and data analysis led to the development of the software packages presented in this thesis. Most of them are written in the programming language Python (Perez et al., 2011), on the one hand due to my personal preference, but mainly due to the recent developments in out-of-core computing with the establishment of xarray and dask (Dask Development Team, 2016; Hoyer and Hamman, 2017; Rocklin, 2015). Another important reason, especially for psyplot (chapter 2) and straditize (chapter 3) was the availability of a highly flexible and stable package for graphical user interfaces, PyQt, and the comparably simple possibility to implement an in-process python console into the PyQ5 application that allows to handle the software functionalities both, from the command line and from the GUI.

36: Add reference.

37: Add reference. jupyter qtconsole

The tools that I present in the following chapter are all available as open-source software packages. But modern Free and Open-Source Software (FOSS) development is not only about making the source code available, but rather about providing a sustainable and maintainable package that allows continuous and transparent development under the aspect of rapidly evolving environment. In the following sections, I will introduce the most important FOSS development concepts (e.g. Shaw, 2018; Stodden and Miguez, 2014) and the necessary vocabulary. These concepts are used by many of the well-established software packages, such as matplotlib (Hunter, 2007), numpy (T. E. Oliphant, 2006), and scipy (Jones et al., 2001).

Version Control

Version control systems record changes to a file and enables the user to roll-back to previous versions of it. The usage of a such a system is inevitable for sustainable FOSS packages. It enables contributions by other FOSS developers and the usage through external packages.

The packages I present in the following chapters are hosted on Github[9], a freely available web platform for hosting projects that are managed with git (Chacon et al., 2019).

Version control with git has a specific terminology (see also chapter 4). Central aspects are *repositories* (project folders), *commits* (change of the project files), *issues* (bug reports), *branches* and *forks* (copies of the (main) project), and *pull requests* (contributions to a project). The following list explains this vocabulary in a bit more detail because the terminology is used in several parts of this thesis, particular in chapter 2 and 4. A more complete list is provided in Github, Inc., 2019.

Repositories are the most basic elements of git and Github. It can be compared to a folder that contains all the necessary files associated with a project (e.g. the

[9] The packages are available at `https://github.com/Chilipp`. Other potential platforms for version control are sourceforge (`https://sourceforge.net`) and Bitbucket (`https://bitbucket.org`)

source code and documentation of a software package). It also contains all the different versions (revisions) of the project files.

Commits or revisions track the changes in the repository. Each commit is a change to a specific file (or a set of files) that is associated with a unique ID and a message of the author to describe the changes.

Issues are suggested improvements, bug reports or any other question to the repository. Every issue has an associated discussion page for the communication between repository owners and the users,

Branches are parallel versions of a repository. Often one incorporates new developments into a separate branch that does not affect the main version of the repository (the *master* branch) and merge the two versions when the new developments are fully implemented.

Forks are copies of repositories. When someone wants to contribute to a software package (repository) that does not belong to him, he can *fork* (copy) it, implement it's changes, and then create a *pull request* to contribute to the official version. Different from a branch, that is a (modified) copy of another branch, forks are copies of the entire repository, i.e. all existing branches.

Pull request are the proposed changes to a repository. One can create a fork of the repository from someone else, implement changes in this fork and then create a pull request to merge it into the original repository. Every pull request has an associated discussion page that allows the repository owner to moderate and discuss the suggested changes.

Webhooks are general methods for web development. Github can trigger a hook to inform a different web service (such as a Continuous Integration (CI) (section 1.3.3)) that a repository has changed or that someone contributed in a discussion. In chapter 4 we use Github webhooks for a automated administration of a repository.

Automated Tests, Test Coverage and Continuous Integration

The most important aspect for FOSS development, especially considering the rapid evolution of this area, is the existence of automated tests. One distinguishes unit tests (test of one single routine) and integration tests (tests of one or more routines within the framework) (Shaw, 2018). The boundary between the two tests is rather vague and the decision about what is used highly depends on the structure of the software that is supposed to be tested. For complex frameworks (such as psyplot or straditize), integration tests are needed to ensure the operability within the framework. Other more simple software packages, (such as docrep or model-organization, see appendix C.2) go well with unit tests only.

Another good standard for such an test suite is to use an automated test discovery tool (e.g. the Python unittest package (Python Software Foundation, 2019) or pytest (Krekel et al., 2004)) that also reports the test coverage (i.e. the fraction of the code that is tested by the test suite). These functionalities are then implemented on a CI service, such as Travis CI[10], Appveyor[11] or CircleCi[12] that are integrated into

[10] https://travis-ci.org/
[11] https://appveyor.com
[12] https://circleci.com/

the Github repository (section 1.3.3). Every commit to the Github repository, or any new pull requests then triggers the tests. This transparently allows to ensure the operability of the software, and the test coverage report ensures that the newly implemented functionality is properly tested. A software development concept that is build entirely on that is the test-driven development. Within this framework, new features are implemented by starting with the test that should be fulfilled by the new feature and then improving the software until this test pass (Beck, 2002).

Automated Documentation

Documentation is the key aspect of a sustainable software and much of the geoscientific code has a lack of proper documentation (based on personal experience). For the software in this thesis, four different levels of the documentation play an important role:

The Application programming interface (API) documentation is meant to document the major parts of the software code that is subject to be used by external scripts or packages. It is usually implemented in the code and documents the essential subroutines and methods of the software.

The graphical user interface (GUI) documentation provides help for the most high-level functionality for the software. The GUI is a user interface into the software through graphical elements (such as buttons, checkboxes, etc.). Unlike the API documentation, it should not require knowledge about programming.

The contributing and/or developers guide is targeting other software developers that might want to contribute to the software package. This document states how other software developers should contribute to the software and introduces the central structural aspects and frameworks of the software.

The manual (or also commonly referred to as *the* documentation) is the document that contains all necessary information for the software, such as installation instructions, tutorials, examples, etc.. It often includes some (or multiple) of the above parts.

The documentations for the software in this thesis have been automatically generated with Sphinx, a Python tool to generate documentations in various different formats (such as HTML, PDF, etc.) (Hasecke, 2019; Perez et al., 2011). It is also implemented as a webhook into the Github repository (see section 1.3.3) to automatically generate an up-to-date documentation of the software for each commit to the Github repository. This provides an additional automated test for the software, and especially it's high-level-interface, in addition to the automated test suite described above (section 1.3.3). Most of the manuals are hosted and build online with the free services offered by readthedocs.org.

Distribution through package managers and virtual environments

FOSS software is meant to be extensible and to build upon other FOSS packages. This requires an accurate and transparent handling of it's dependencies and requirements which is usually provided through the so-called packaging of the software (e.g. Torborg, 2016). There exists a variety of package managers and the choice most often depends on the framework of the software.

The software in this thesis is mainly distributed via two systems. The first one is python's own package manager *pip* which is based on the packages uploaded to pypi.org. The second one which got increasing importance during the recent past is the open-source Anaconda Distribution[13]. Both work on multiple operating systems (Windows, Linux and Mac OS), but the Anaconda Distribution contains also non-python packages (e.g. written in C or C++) for which the Python packages rely on, and it contains a rich suite of r-packages.

One step further, compared to package managers, are the distribution of virtual environments. These systems do not only provide the software, but also a full operating system and the installed dependencies. A popular platform (used also for the Eurasian Modern Pollen Database (EMPD) database) is provided through so-called Docker containers[14]. Compared to package managers, this system has the advantage of simplifying the installation procedure for the user because he only has to download the corresponding docker image. The docker image itself then runs independent of the local file system in a separate isolated mode.

1.4 Challenges tackled in this thesis

I present several new tools in this thesis that tackle the aspects described in the previous sections. It is divided into two parts: Part I are the software chapters 2, 3 and 4 that introduce new packages developed for paleoclimate analysis. Part II consists of the analysis chapters 5 and

6 that address two use-cases tackling the combination of observations and models

for an informed paleoclimate understanding.

The first part starts with the visualization framework psyplot in chapter 2, a suite of python packages that are designed for interactive visual analysis of data both from a GUI and the command line. The scope of this software is not limited to paleoclimate analysis and serves a more general purpose. As such, it serves as a base infrastructure for many of the topics described in the other chapters.

Straditize, described in the next chapter 3, addresses the problem of gathering paleo-climate information that has been collected during the pre-digital area. This software is a semi-automated digitization package for stratigraphic diagrams, pollen diagrams in particular. Straditize is built on top of psyplot and its GUI and as such provides rich interactive documentation and visualization methods.

Chapter 4 covers the last aspect of software usage for paleoclimate data: data distribution. In this chapter I describe new infrastructural tools for the sustainable management of a community-driven pollen database, the Eurasian Modern Pollen Database (EMPD). They consist of a flexible and lightweight map-based web interface, the EMPD-viewer, into the data and a webserver for an automated administration of the database. Within this section, I also present another use case for the EMPD-viewer that is adapted to a large northern-hemispheric database of fossil and modern pollen records.

The second part starts with the weather generator GWGEN in chapter 5, a statistical model that uses modern relationships in observational data to inform large-scale paleo climate models with temporally downscaled temperature, precipitation, cloud cover and wind speed records.

[13] `https://www.anaconda.com`

[14] `https://www.docker.com`

Finally, in chapter 6 I investigate the question to what extent large-scale atmospheric circulation features can be estimated from proxy data. In this analysis I analyze the long-term stability of spatial correlation patterns between surface temperature and northern hemispheric teleconnections based on three ESMs.

This thesis finishes with the conclusions in chapter 7 which summarizes the new tools and findings and provides an outlook for the further development of the methods presented. In the Appendix I present another work that has developed in a cooperation which is also based on the infrastructural tools from psyplot and GWGEN (appendix A), and I provide a list of the publications during my thesis (appendix B) and an overview about all the software packages that have been developed (appendix C).

not sure if I include this here

References

Ayachit, Utkarsh (2015). *The paraview guide: a parallel visualization application*. Kitware, Inc.

Bauer, Peter, Alan Thorpe, and Gilbert Brunet (2015). "The quiet revolution of numerical weather prediction". In: *Nature* 525.7567, pp. 47–55. DOI: 10.1038/nature14956.

Beck, Kent (2002). *Test Driven Development. By Example*. Addison Wesley. 192 pp. ISBN: 978-0-321-14653-3. URL: https://www.ebook.de/de/product/3253611/kent_beck_test_driven_development_by_example.html.

Birks, Harry John Betteley and Hilary H Birks (1980). *Quaternary palaeoecology*. Edward Arnold London.

Bolliet, Timothé, Patrick Brockmann, Valérie Masson-Delmotte, Franck Bassinot, Valérie Daux, Dominique Genty, Amaelle Landais, Marlène Lavrieux, Elisabeth Michel, Pablo Ortega, Camille Risi, Didier M. Roche, Françoise Vimeux, and Claire Waelbroeck (2016). "Water and carbon stable isotope records from natural archives: a new database and interactive online platform for data browsing, visualizing and downloading". In: *Climate of the Past* 12.8, pp. 1693–1719. DOI: 10.5194/cp-12-1693-2016.

Böttinger, Michael and Niklas Röber (2019). "Visualization in Climate Modelling". In: *International Climate Protection*. Ed. by Michael Palocz-Andresen, Dòra Szalay, Andràs Gosztom, Làszlò Sìpos, and Tìmea Taligàs. Cham: Springer International Publishing, pp. 313–321. ISBN: 978-3-030-03816-8. DOI: 10.1007/978-3-030-03816-8_39. URL: https://doi.org/10.1007/978-3-030-03816-8_39.

Braconnot, P., Sandy P Harrison, Masa Kageyama, Patrick J Bartlein, Valerie Masson-Delmotte, Ayako Abe-Ouchi, Bette Otto-Bliesner, and Yan Zhao (2012). "Evaluation of climate models using palaeoclimatic data". In: *Nature Climate Change* 2.6, p. 417. DOI: 10.1038/nclimate1456. URL: https://www.nature.com/articles/nclimate1456.

Braconnot, P., B. Otto-Bliesner, S. Harrison, S. Joussaume, J.-Y. Peterchmitt, A. Abe-Ouchi, M. Crucifix, E. Driesschaert, Th. Fichefet, C. D. Hewitt, M. Kageyama, A. Kitoh, M.-F. Loutre, O. Marti, U. Merkel, G. Ramstein, P. Valdes, L. Weber, Y. Yu, and Y. Zhao (2007a). "Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 2: feedbacks with emphasis on the location of the ITCZ and mid- and high latitudes heat budget". In: *Climate of the Past* 3.2, pp. 279–296. DOI: 10.5194/cp-3-279-2007. URL: https://www.clim-past.net/3/279/2007/.

Braconnot, P., Otto-Bliesner, S. P. Harrison, S. Joussaume, J.-Y. Peterchmitt, A. Abe-Ouchi, M. Crucifix, E. Driesschaert, Th. Fichefet, C. D. Hewitt, M. Kageyama, A. Kitoh, A. Laîné, M.-F. Loutre, O. Marti, U. Merkel, G. Ramstein, P. Valdes, S. L. Weber, Y. Yu, and Y. Zhao (2007b). "Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features". In: *Climate of the Past* 3.2, pp. 261–277. DOI: 10.5194/cp-3-261-2007. URL: https://www.clim-past.net/3/261/2007/.

Bradley, Raymond S (1985). *Quaternary paleoclimatology : methods of paleoclimatic reconstruction*. eng. Boston ; London [etc.]: Allen and Unwin. ISBN: 0045510679.

Brewer, S., Joel Guiot, and Doris Barboni (2007). "Pollen data as climate proxies". In: *Encyclopedia of Quaternary Science*. Elsevier, pp. 2497–2508. URL: https://hal.archives-ouvertes.fr/hal-00995404.

Brewer, Simon, J. Guiot, and F. Torre (2007). "Mid-Holocene climate change in Europe: a data-model comparison". In: *Climate of the Past* 3.3, pp. 499–512. DOI: 10.5194/cp-3-499-2007. URL: https://www.clim-past.net/3/499/2007/.

Brockmann, Patrick (2016). *ClimateProxiesFinder: dc.js + leaflet application to discover climate proxies*. Last accessed: 2019-08-30. URL: https://github.com/PBrockmann/ClimateProxiesFinder (visited on 10/06/2016).

Brown, Dave, Richard Brownrigg, Mary Haley, and Wei Huang (2012). "NCAR Command Language (NCL)". eng. In: DOI: 10.5065/d6wd3xh5.

Brown, Stewart A., Mike Folk, Gregory Goucher, Russ Rew, and Paul F. Dubois (1993). "Software for Portable Scientific Data Management". In: *Computers in Physics* 7.3, p. 304. DOI: 10.1063/1.4823180.

Chacon, Scott, Ben Straub, and Pro Git Contributors (2019). *Pro Git*. 2nd ed. Last accessed: 2019-08-31. URL: https://github.com/progit/progit2 (visited on 08/31/2019).

Chevalier, M., R. Cheddadi, and B. M. Chase (2014). "CREST (Climate REconstruction SofTware): a probability density function (PDF)-based quantitative climate reconstruction method". In: *Climate of the Past* 10.6, pp. 2081–2098. DOI: 10.5194/cp-10-2081-2014.

Chevalier, Manuel (2019). "Enabling possibilities to quantify past climate from fossil assemblages at a global scale". In: *Global and Planetary Change* 175, pp. 27–35. DOI: 10.1016/j.gloplacha.2019.01.016.

Clyne, John, Pablo Mininni, Alan Norton, and Mark Rast (2007). "Interactive desktop analysis of high resolution simulations: application to turbulent plume dynamics and current sheet formation". In: *New Journal of Physics* 9.8, pp. 301–301. DOI: 10.1088/1367-2630/9/8/301.

COHMAP Members (1988). "Climatic Changes of the Last 18,000 Years: Observations and Model Simulations". In: *Science* 241.4869, pp. 1043–1052. ISSN: 00368075, 10959203. URL: http://www.jstor.org/stable/1702404.

Dasgupta, Aritra, Jorge Poco, Enrico Bertini, and Claudio T. Silva (2016). "Reducing the Analytical Bottleneck for Domain Scientists: Lessons from a Climate Data Visualization Case Study". In: *Computing in Science & Engineering* 18.1, pp. 92–100. DOI: 10.1109/mcse.2016.7.

Dask Development Team (2016). *Dask: Library for dynamic task scheduling*. URL: https://dask.org.

Davis, B. A. S., S. Brewer, A. C. Stevenson, and J. Guiot (2003). "The temperature of Europe during the Holocene reconstructed from pollen data". In: *Quat. Sci. Rev.* 22.15-17, pp. 1701–1716. ISSN: 02773791. DOI: 10.1016/s0277-3791(03)00173-2.

Davis, B. A. S., M. Zanon, P. Collins, A. Mauri, J. Bakker, D. Barboni, A. Barthelmes, C. Beaudouin, A. E. Bjune, E. Bozilova, R. H. W. Bradshaw, B. A. Brayshay, S.

Brewer, E. Brugiapaglia, J. Bunting, S. E. Connor, J. L. de Beaulieu, K. Edwards, A. Ejarque, P. Fall, A. Florenzano, R. Fyfe, D. Galop, M. Giardini, T. Giesecke, M. J. Grant, J. Guiot, S. Jahns, V. Jankovska, S. Juggins, M. Kahrmann, M. Karpinska-Kolaczek, P. Kolaczek, N. Kuhl, P. Kunes, E. G. Lapteva, S. A. G. Leroy, M. Leydet, J. Guiot, S. Jahns, V. Jankovska, S. Juggins, M. Kahrmann, M. Karpinska-Kolaczek, P. Kolaczek, N. Kuehl, P. Kunes, E. G. Lapteva, S. A. G. Leroy, M. Leydet, J. A. L. Saez, A. Masi, I. Matthias, F. Mazier, V. Meltsov, A. M. Mercuri, Y. Miras, F. J. G. Mitchell, J. L. Morris, F. Naughton, A. B. Nielsen, E. Novenko, B. Odgaard, E. Ortu, M. V. Overballe-Petersen, H. S. Pardoe, S. M. Peglar, I. A. Pidek, L. Sadori, H. Seppa, E. Severova, H. Shaw, J. Swieta-Musznicka, M. Theuerkauf, S. Tonkov, S. Veski, W. O. van der Knaap, J. F. N. van Leeuwen, J. Woodbridge, M. Zimny, and J. O. Kaplan (2013). "The European Modern Pollen Database (EMPD) project". In: *Vegetation History and Archaeobotany* 22.6, pp. 521–530. ISSN: 0939-6314. DOI: 10.1007/s00334-012-0388-5. URL: http://link.springer.com/article/10.1007/s00334-012-0388-5.

Davis, Basil A. S. and Simon Brewer (2009). "Orbital forcing and role of the latitudinal insolation/temperature gradient". In: *Climate Dynamics* 32.2, pp. 143–165. ISSN: 1432-0894. DOI: 10.1007/s00382-008-0480-9. URL: https://doi.org/10.1007/s00382-008-0480-9.

Edwards, Paul N. (2010). "History of climate modeling". In: *Wiley Interdisciplinary Reviews: Climate Change* 2.1, pp. 128–139. DOI: 10.1002/wcc.95.

Fægri, K., P. E. Kaland, and K. Krzywinski (1989). *Textbook of pollen analysis.* Ed. 4. Chichester, UK: John Wiley & Sons Ltd.

Flantua, Suzette G.A., Henry Hooghiemstra, Eric C. Grimm, Hermann Behling, Mark B. Bush, Catalina González-Arango, William D. Gosling, Marie-Pierre Ledru, Socorro Lozano-Garcìa, Antonio Maldonado, Aldo R. Prieto, Valentì Rull, and John H. Van Boxel (2015). "Updated site compilation of the Latin American Pollen Database". In: *Review of Palaeobotany and Palynology* 223, pp. 104–115. DOI: 10.1016/j.revpalbo.2015.09.008.

Fyfe, Ralph M., Jacques-Louis de Beaulieu, Heather Binney, Richard H. W. Bradshaw, Simon Brewer, Anne Le Flao, Walter Finsinger, Marie-Josè Gaillard, Thomas Giesecke, Graciela Gil-Romera, Eric C. Grimm, Brian Huntley, Petr Kunes, Norbert Kühl, Michelle Leydet, Andrè F. Lotter, Pavel E. Tarasov, and Spassimir Tonkov (2009). "The European Pollen Database: past efforts and current activities". In: *Vegetation History and Archaeobotany* 18.5, pp. 417–424. DOI: 10.1007/s00334-009-0215-9.

Github, Inc. (2019). "GitHub glossary". In: Last accessed: 2019-08-31. URL: https://help.github.com/en/articles/github-glossary (visited on 08/31/2019).

Goring, Simon, Andria Dawson, Gavin L Simpson, Karthik Ram, Russell W Graham, Eric C Grimm, and Jack W. Williams (2015). "neotoma: A Programmatic Interface to the Neotoma Paleoecological Database". In: *Open Quaternary* 1.1, p. 2. URL: http://doi.org/10.5334/oq.ab.

Gregoire, Lauren J., Paul J. Valdes, and Antony J. Payne (2015). "The relative contribution of orbital forcing and greenhouse gases to the North American deglaciation". In: *Geophysical Research Letters* 42.22, pp. 9970–9979. DOI: 10.1002/2015gl066005. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015GL066005. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL066005.

Grimm, Eric C. (1988). "Data analysis and display". In: *Vegetation history*. Ed. by B. Huntley, T. Webb, B. Huntley, and T. Webb. Dordrecht: Springer Netherlands, pp. 43–76. ISBN: 978-94-009-3081-0. DOI: 10.1007/978-94-009-3081-0_3. URL: https://doi.org/10.1007/978-94-009-3081-0_3.

Grimm, Eric C. (1991). “Tilia and Tiliagraph”. In: *Illinois State Museum, Springfield* 101.

— (2008). “Neotoma: an ecosystem database for the Pliocene, Pleistocene, and Holocene”. In: *Illinois State Museum Scientific Papers E Series* 1. URL: https://www.neotomadb.org/uploads/NeotomaManual.pdf.

Hargreaves, J. C., J. D. Annan, R. Ohgaito, A. Paul, and A. Abe-Ouchi (2013). “Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene”. In: *Clim. Past* 9.2, pp. 811–823. ISSN: 1814-9332. DOI: 10.5194/cp-9-811-2013.

Hasecke, Jan Ulrich (2019). *Software-Dokumentation mit Sphinx: Zweite überarbeitete Auflage (Sphinx 2.0) (German Edition)*. Independently published. ISBN: 1793008779. URL: https://www.amazon.com/Software-Dokumentation-mit-Sphinx-%C3%83%C2%BCberarbeitete-Auflage/dp/1793008779?SubscriptionId=AKIAIOBINVZYXZQZ2U3A%5C&tag=chimbori05-20%5C&linkCode=xm2%5C&camp=2025%5C&creative=165953%5C&creativeASIN=1793008779.

Havinga, A.J. (1967). “Palynology and pollen preservation”. In: *Review of Palaeobotany and Palynology* 2.1-4, pp. 81–98. DOI: 10.1016/0034-6667(67)90138-8.

Hays, J. D., J. Imbrie, and N. J. Shackleton (1976). “Variations in the Earth’s Orbit: Pacemaker of the Ice Ages”. In: *Science* 194.4270, pp. 1121–32. ISSN: 0036-8075 (Print) 0036-8075 (Linking). DOI: 10.1126/science.194.4270.1121. URL: http://www.ncbi.nlm.nih.gov/pubmed/17790893.

Hoskins, Brian (2012). “The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science”. In: *Quarterly Journal of the Royal Meteorological Society* 139.672, pp. 573–584. DOI: 10.1002/qj.1991.

Hoyer, S. and J. Hamman (2017). “xarray: N-D labeled arrays and datasets in Python”. In: *Journal of Open Research Software* 5.1. DOI: 10.5334/jors.148. URL: http://doi.org/10.5334/jors.148.

Hunter, J. D. (2007). “Matplotlib: A 2D Graphics Environment”. In: *Computing in Science Engineering* 9.3, pp. 90–95. ISSN: 1521-9615. DOI: 10.1109/MCSE.2007.55.

Imbrie, J., A. Berger, E. A. Boyle, S. C. Clemens, A. Duffy, W. R. Howard, G. Kukla, J. Kutzbach, D. G. Martinson, A. McIntyre, A. C. Mix, B. Molfino, J. J. Morley, L. C. Peterson, N. G. Pisias, W. L. Prell, M. E. Raymo, N. J. Shackleton, and J. R. Toggweiler (1993). “On the structure and origin of major glaciation cycles 2. The 100,000-year cycle”. In: *Paleoceanography* 8.6, pp. 699–735. DOI: 10.1029/93pa02751. URL: https://ui.adsabs.harvard.edu/abs/1993PalOc...8..699I.

Intergovernmental Panel on Climate Change (2014). “Evaluation of Climate Models”. In: *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, pp. 741–866. DOI: 10.1017/CBO9781107415324.020.

Ivanovic, R. F., L. J. Gregoire, M. Kageyama, D. M. Roche, P. J. Valdes, A. Burke, R. Drummond, W. R. Peltier, and L. Tarasov (2016). “Transient climate simulations of the deglaciation 21–9 thousand years before present (version 1) – PMIP4 Core experiment design and boundary conditions”. In: *Geosci. Model Dev.* 9.7, pp. 2563–2587. DOI: 10.5194/gmd-9-2563-2016. URL: https://www.geosci-model-dev.net/9/2563/2016/.

Jones, Eric, Travis Oliphant, Pearu Peterson, et al. (2001). *SciPy: Open source scientific tools for Python*. [Online; accessed 2017-02-18]. URL: http://www.scipy.org/.

Joussaume, S and KE Taylor (1995). “Status of the paleoclimate modeling intercomparison project (PMIP)”. In: *World Meteorological Organization-Publications-WMO TD*, pp. 425–430.

Jubair, Mohammad Imrul, Usman Alim, Niklas Roeber, John Clyne, Ali Mahdavi-Amiri, and Faramarz Samavati (2015). "Multiresolution visualization of digital earth data via hexagonal box-spline wavelets". In: *2015 IEEE Scientific Visualization Conference (SciVis)*. IEEE. DOI: 10.1109/scivis.2015.7429508.

Juggins, Steve (2007). "C2: Software for ecological and palaeoecological data analysis and visualisation (user guide version 1.5)". In: *Newcastle upon Tyne: Newcastle University* 77. URL: https://www.staff.ncl.ac.uk/stephen.juggins/software/C2Home.htm.

— (2013). "Quantitative reconstructions in palaeolimnology: new paradigm or sick science?" In: *Quaternary Science Reviews* 64, pp. 20–32. ISSN: 0277-3791. DOI: 10.1016/j.quascirev.2012.12.014. URL: http://www.sciencedirect.com/science/article/pii/S0277379112005422.

— (2017). *rioja: Analysis of Quaternary Science Data*. R package version 0.9-21. URL: http://www.staff.ncl.ac.uk/stephen.juggins/.

Juggins, Steve and H. John B. Birks (2012). "Quantitative Environmental Reconstructions from Biological Data". In: *Tracking Environmental Change Using Lake Sediments: Data Handling and Numerical Techniques*. Ed. by H. John B. Birks, André F. Lotter, Steve Juggins, and John P. Smol. Dordrecht: Springer Netherlands, pp. 431–494. ISBN: 978-94-007-2745-8. DOI: 10.1007/978-94-007-2745-8_14. URL: https://doi.org/10.1007/978-94-007-2745-8_14.

Jungclaus, J. H., E. Bard, M. Baroni, P. Braconnot, J. Cao, L. P. Chini, T. Egorova, M. Evans, J. F. Gonzàlez-Rouco, H. Goosse, G. C. Hurtt, F. Joos, J. O. Kaplan, M. Khodri, K. Klein Goldewijk, N. Krivova, A. N. LeGrande, S. J. Lorenz, J. Luterbacher, W. Man, A. C. Maycock, M. Meinshausen, A. Moberg, R. Muscheler, C. Nehrbass-Ahles, B. I. Otto-Bliesner, S. J. Phipps, J. Pongratz, E. Rozanov, G. A. Schmidt, H. Schmidt, W. Schmutz, A. Schurer, A. I. Shapiro, M. Sigl, J. E. Smerdon, S. K. Solanki, C. Timmreck, M. Toohey, I. G. Usoskin, S. Wagner, C.-J. Wu, K. L. Yeo, D. Zanchettin, Q. Zhang, and E. Zorita (2017). "The PMIP4 contribution to CMIP6 – Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 *past1000* simulations". In: *Geosci. Model Dev.* 10.11, pp. 4005–4033. DOI: 10.5194/gmd-10-4005-2017. URL: https://www.geosci-model-dev.net/10/4005/2017/.

Kageyama, M., P. Braconnot, S. P. Harrison, A. M. Haywood, J. Jungclaus, B. L. Otto-Bliesner, J.-Y. Peterschmitt, A. Abe-Ouchi, S. Albani, P. J. Bartlein, C. Brierley, M. Crucifix, A. Dolan, L. Fernandez-Donado, H. Fischer, P. O. Hopcroft, R. F. Ivanovic, F. Lambert, D. J. Lunt, N. M. Mahowald, W. R. Peltier, S. J. Phipps, D. M. Roche, G. A. Schmidt, L. Tarasov, P. J. Valdes, Q. Zhang, and T. Zhou (2016). "PMIP4-CMIP6: the contribution of the Paleoclimate Modelling Intercomparison Project to CMIP6". In: *Geosci. Model Dev. Discuss.* 2016, pp. 1–46. DOI: 10.5194/gmd-2016-106. URL: https://www.geosci-model-dev.net/11/1033/2018/gmd-11-1033-2018.html.

Karpechko, A.Y. (2010). "Uncertainties in future climate attributable to uncertainties in future Northern Annular Mode trend. NAM AND FUTURE CLIMATE UNCERTAINTIES". In: *Geophysical Research Letters* 37. ISSN: 0094-8276. DOI: 10.1029/2010gl044717.

Keim, Daniel, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon (2008). "Visual Analytics: Definition, Process, and Challenges". In: *Information Visualization: Human-Centered Issues and Perspectives*. Ed. by Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, Chris North, Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North. Berlin, Heidelberg:

Springer Berlin Heidelberg, pp. 154–175. ISBN: 978-3-540-70956-5. DOI: 10.1007/978-3-540-70956-5_7. URL: https://doi.org/10.1007/978-3-540-70956-5_7.

Krekel, Holger, Bruno Oliveira, Ronny Pfannschmidt, Floris Bruynooghe, Brianna Laugher, and Florian Bruhin (2004). *pytest 5.1*. URL: https://github.com/pytest-dev/pytest.

Lewis, J. M. (1998). "Clarifying the Dynamics of the General Circulation: Phillips's 1956 Experiment". In: *Bulletin of the American Meteorological Society* 79.1, pp. 39–60. DOI: 10.1175/1520-0477(1998)079<0039:ctdotg>2.0.co;2. URL: https://ui.adsabs.harvard.edu/abs/1998BAMS...79...39L.

Marchant, Robert, Lucia Almeida, Hermann Behling, Juan Carlos Berrio, Mark Bush, Antoine Cleef, Joost Duivenvoorden, Maarten Kappelle, Paulo De Oliveira, Ary Teixeira de Oliveira-Filho, Socorro Lozano-Garìia, Henry Hooghiemstra, Marie-Pierre Ledru, Beatriz Ludlow-Wiechers, Vera Markgraf, Virginia Mancini, Marta Paez, Aldo Prieto, Olando Rangel, and Maria Lea Salgado-Labouriau (2002). "Distribution and ecology of parent taxa of pollen lodged within the Latin American Pollen Database". In: *Review of Palaeobotany and Palynology* 121.1, pp. 1–75. DOI: 10.1016/s0034-6667(02)00082-9.

Marsicek, Jeremiah, Bryan N. Shuman, Patrick J. Bartlein, Sarah L. Shafer, and Simon Brewer (2018). "Reconciling divergent trends and millennial variations in Holocene temperatures". In: *Nature* 554.7690, pp. 92–96. DOI: 10.1038/nature25464.

Masson-Delmotte, V., M. Kageyama, P. Braconnot, S. Charbit, G. Krinner, C. Ritz, E. Guilyardi, J. Jouzel, A. Abe-Ouchi, M. Crucifix, R. M. Gladstone, C. D. Hewitt, A. Kitoh, A. N. LeGrande, O. Marti, U. Merkel, T. Motoi, R. Ohgaito, B. Otto-Bliesner, W. R. Peltier, I. Ross, P. J. Valdes, G. Vettoretti, S. L. Weber, F. Wolk, and Y. YU (2006). "Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints". In: *Climate Dynamics* 26.5, pp. 513–529. ISSN: 1432-0894. DOI: 10.1007/s00382-005-0081-9. URL: https://doi.org/10.1007/s00382-005-0081-9.

Mauri, A., B. A. S. Davis, P. M. Collins, and J. O. Kaplan (2015). "The climate of Europe during the Holocene: a gridded pollen-based reconstruction and its multi-proxy evaluation". In: *Quat. Sci. Rev.* 112, pp. 109–127. ISSN: 0277-3791. DOI: 10.1016/j.quascirev.2015.01.013. URL: http://www.sciencedirect.com/science/article/pii/S0277379115000372.

Menviel, L., A. Timmermann, O. Elison Timm, and A. Mouchet (2011). "Deconstructing the Last Glacial termination: the role of millennial and orbital-scale forcings". In: *Quaternary Science Reviews* 30.9, pp. 1155–1172. ISSN: 0277-3791. DOI: 10.1016/j.quascirev.2011.02.005. URL: http://www.sciencedirect.com/science/article/pii/S0277379111000539.

Nakagawa, Takeshi, Pavel E. Tarasov, Kotoba Nishida, Katsuya Gotanda, and Yoshinori Yasuda (2002). "Quantitative pollen-based climate reconstruction in central Japan: application to surface and Late Quaternary spectra". In: *Quaternary Science Reviews* 21.18-19, pp. 2099–2113. DOI: 10.1016/s0277-3791(02)00014-8.

Neukom, Raphael, Luis A. Barboza, Michael P. Erb, Feng Shi, Julien Emile-Geay, Michael N. Evans, Jörg Franke, Darrell S. Kaufman, Lucie Lücke, Kira Rehfeld, Andrew Schurer, Feng Zhu, Stefan Brönnimann, Gregory J. Hakim, Benjamin J. Henley, Fredrik Charpentier Ljungqvist, Nicholas McKay, Veronika Valler, Lucien von Gunten, and P. A. G. E. S. 2k Consortium (2019a). "Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era". In: *Nature Geoscience* 12.8, pp. 643–649. ISSN: 1752-0908. URL: https://doi.org/10.1038/s41561-019-0400-0.

Neukom, Raphael, Nathan Steiger, Juan Josè Gòmez-Navarro, Jianghao Wang, and Johannes P. Werner (2019b). "No evidence for globally coherent warm and cold periods over the preindustrial Common Era". In: *Nature* 571.7766, pp. 550–554. DOI: 10.1038/s41586-019-1401-2.

Nichols, Harvey (1967). "The Post-glacial history of vegetation and climate at Ennadai Lake, Keewatin, and Lynn Lake, Manitoba (Canada)". In: E&G – Quaternary Science Journal 18.1. DOI: 10.23689/fidgeo-1124.

— (1969). "The Late Quaternary History of Vegetation and Climate at Porcupine Mountain and Clearwater Bog, Manitoba". In: *Arctic and Alpine Research* 1.3, p. 155. ISSN: 00040851. DOI: 10.2307/1550287. URL: http://www.jstor.org/stable/1550287.

Nocke, Thomas (2014). "Images for Data Analysis: The Role of Visualization in Climate Research Processes". In: *IMAGE POLITICS OF CLIMATE CHANGE: VISUALIZATIONS, IMAGINATIONS, DOCUMENTATIONS*. Ed. by Schneider, B and Nocke, T. Vol. 55. Image-Series, 55–77. ISBN: 978-3-8394-2610-4; 978-3-8376-2610-0.

Nocke, Thomas, Till Sterzel, Michael Böttinger, Markus Wrobel, et al. (2008). "Visualization of climate and climate change data: An overview". In: *Digital earth summit on geoinformatics*, pp. 226–232.

Nolan, Connor, John Tipton, Robert K Booth, Mevin B Hooten, and Stephen T Jackson (2019). "Comparing and improving methods for reconstructing peatland water-table depth from testate amoebae". In: *The Holocene* 29.8, pp. 1350–1361. DOI: 10.1177/0959683619846969.

Oliphant, Travis E (2006). *A guide to NumPy*. Vol. 1. Trelgol Publishing USA. URL: http://www.numpy.org/.

Otto-Bliesner, B. L., P. Braconnot, S. P. Harrison, D. J. Lunt, A. Abe-Ouchi, S. Albani, P. J. Bartlein, E. Capron, A. E. Carlson, A. Dutton, H. Fischer, H. Goelzer, A. Govin, A. Haywood, F. Joos, A. N. LeGrande, W. H. Lipscomb, G. Lohmann, N. Mahowald, C. Nehrbass-Ahles, F. S. R. Pausata, J.-Y. Peterschmitt, S. J. Phipps, H. Renssen, and Q. Zhang (2017). "The PMIP4 contribution to CMIP6 – Part 2: Two interglacials, scientific objective and experimental design for Holocene and Last Interglacial simulations". In: *Geosci. Model Dev.* 10.11, pp. 3979–4003. DOI: 10.5194/gmd-10-3979-2017. URL: https://www.geosci-model-dev.net/10/3979/2017/.

Otto-Bliesner, Bette L., James M. Russell, Peter U. Clark, Zhengyu Liu, Jonathan T. Overpeck, Bronwen Konecky, Peter deMenocal, Sharon E. Nicholson, Feng He, and Zhengyao Lu (2014). "Coherent changes of southeastern equatorial and northern African rainfall during the last deglaciation". In: *Science* 346.6214, pp. 1223–1227. ISSN: 0036-8075. DOI: 10.1126/science.1259531. eprint: https://science.sciencemag.org/content/346/6214/1223.full.pdf. URL: https://science.sciencemag.org/content/346/6214/1223.

Perez, Fernando, Brian E. Granger, and John D. Hunter (2011). "Python: An Ecosystem for Scientific Computing". In: *Computing in Science & Engineering* 13.2, pp. 13–21. DOI: 10.1109/mcse.2010.119.

Phillips, Norman A. (1956). "The general circulation of the atmosphere: a numerical experiment". In: *Quarterly Journal of the Royal Meteorological Society* 82.352, pp. 123–164.

Python Software Foundation (2019). *unittest – Unit testing framework*. URL: https://docs.python.org/3.7/library/unittest.html (visited on 09/02/2019).

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rautenhaus, Marc, Michael Böttinger, Stephan Siemen, Robert Hoffman, Robert M. Kirby, Mahsa Mirzargar, Niklas Rober, and Rudiger Westermann (2018). "Visualization in Meteorology—A Survey of Techniques and Tools for Data Analysis Tasks". In: *IEEE Transactions on Visualization and Computer Graphics* 24.12, pp. 3268–3296. DOI: 10.1109/tvcg.2017.2779501.

Rew, R. and G. Davis (1990). "NetCDF: an interface for scientific data access". In: *IEEE Computer Graphics and Applications* 10.4, pp. 76–82. DOI: 10.1109/38.56302.

Röber, Niklas, P. Adamidis, and Jörn Behrens (2015). "Visualization and Analysis of Climate Simulation Performance Data". In: *EGU General Assembly Conference Abstracts*. Vol. 17. EGU General Assembly Conference Abstracts, p. 15318.

Roche, D. M., H. Renssen, D. Paillard, and G. Levavasseur (2011). "Deciphering the spatio-temporal complexity of climate change of the last deglaciation: a model analysis". In: *Climate of the Past* 7.2, pp. 591–602. DOI: 10.5194/cp-7-591-2011. URL: https://www.clim-past.net/7/591/2011/.

Rocklin, Matthew (2015). "Dask: Parallel Computation with Blocked algorithms and Task Scheduling". In: *Proceedings of the 14th Python in Science Conference*. Ed. by Kathryn Huff and James Bergstra, pp. 130–136.

Schulz, Hans-Jorg, Thomas Nocke, Magnus Heitzler, and Heidrun Schumann (2013). "A Design Space of Visualization Tasks". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12, pp. 2366–2375. DOI: 10.1109/tvcg.2013.120.

Schulzweida, Uwe (2019). *CDO User Guide*. DOI: 10.5281/zenodo.2558193. URL: https://doi.org/10.5281/zenodo.2558193.

Shaw, Anthony (2018). *Getting Started With Testing in Python*. Last accessed: 2019-09-02. URL: https://realpython.com/python-testing/ (visited on 10/22/2018).

Simpson, G. L. (2007). "Analogue Methods in Palaeoecology: Using the analogue Package". In: *Journal of Statistical Software* 22.2, pp. 1–29.

Simpson, G. L. and J. Oksanen (2019). *analogue: Analogue and weighted averaging methods for palaeoecology*. R package version 0.17-3. URL: https://cran.r-project.org/package=analogue.

Skamarock, William C., Joseph B. Klemp, Michael G. Duda, Laura D. Fowler, Sang-Hun Park, and Todd D. Ringler (2012). "A Multiscale Nonhydrostatic Atmospheric Model Using Centroidal Voronoi Tesselations and C-Grid Staggering". In: *Monthly Weather Review* 140.9, pp. 3090–3105. DOI: 10.1175/mwr-d-11-00215.1.

Sommer, Philipp S. (2017). "The psyplot interactive visualization framework". In: *The Journal of Open Source Software* 2.16. DOI: 10.21105/joss.00363. URL: https://doi.org/10.21105/joss.00363.

— (2019). *psy-strat v0.1.0: A Python package for creating stratigraphic diagrams*. DOI: 10.5281/zenodo.3381753. URL: https://doi.org/10.5281/zenodo.3381753.

Sommer, Philipp S. and Jed O. Kaplan (2017). "A globally calibrated scheme for generating daily meteorology from monthly statistics: Global-WGEN (GWGEN) v1.0". In: *Geosci. Model Dev.* 10.10, pp. 3771–3791. DOI: 10.5194/gmd-10-3771-2017.

Sommer, Philipp S., Dilan Rech, Manuel Chevalier, and Basil Davis (2019). "straditize: Digitizing stratigraphic diagrams". In: *Journal of Open Source Software* 4.34, p. 1216. DOI: 10.21105/joss.01216. URL: https://doi.org/10.21105/joss.01216.

Stodden, Victoria and Sheila Miguez (2014). "Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research". In: *Journal of Open Research Software* 2.1. DOI: 10.5334/jors.ay.

Sullivan, C. and Alexander Kaszynski (2019). "PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK)". In: *Journal of Open Source Software* 4.37, p. 1450. DOI: 10.21105/joss.01450.

Sullivan, C. and Whitney Trainor-Guitton (2019). "PVGeo: an open-source Python package for geoscientific visualization in VTK and ParaView". In: *Journal of Open Source Software* 4.38, p. 1451. DOI: 10.21105/joss.01451.

Tipton, John (2017). *BayesComposition: Fit forward and inverse prediction Bayesian functional models for compositional data*. R package version 1.0. URL: https://github.com/jtipton25/BayesComposition.

Torborg, Scott (2016). *python-packaging: Tutorial on how to structure Python packages*. Revision 35daf993. URL: https://python-packaging.readthedocs.io (visited on 09/02/2019).

Treut, Hervé Le, Richard Somerville, Ulrich Cubasch, Yihui Ding, Cecilie Mauritzen, Abdalah Mokssit, Thomas Peterson, and Michael Prather (2007). "Historical Overview of Climate Change Science". In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, H.L. Miller, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Chap. 1, pp. 93–123. URL: https://www.ipcc.ch/site/assets/uploads/2018/03/ar4-wg1-chapter1.pdf.

Ulden, A.P. van and G.J. van Oldenborgh (2006). "Large-scale atmospheric circulation biases and changes in global climate model simulations and their importance for climate change in Central Europe". In: *Atmos Chem Phys* 6, pp. 863–881. ISSN: 1680-7324. DOI: 10.5194/acp-6-863-2006.

Varma, V., M. Prange, U. Merkel, T. Kleinen, G. Lohmann, M. Pfeiffer, H. Renssen, A. Wagner, S. Wagner, and M. Schulz (2012). "Holocene evolution of the Southern Hemisphere westerly winds in transient simulations with global climate models". In: *Climate of the Past* 8.2, pp. 391–402. DOI: 10.5194/cp-8-391-2012. URL: https://www.clim-past.net/8/391/2012/.

Vincens, Annie, Anne-Marie Lézine, Guillaume Buchet, Dorothée Lewden, and Annick Le Thomas (2007). "African pollen database inventory of tree and shrub pollen types". In: *Rev. Palaeobot. Palynol.* 145.1-2, pp. 135–141. ISSN: 00346667. DOI: 10.1016/j.revpalbo.2006.09.004.

Walker, Mike, Sigfus Johnsen, Sune Olander Rasmussen, Trevor Popp, Jørgen-Peder Steffensen, Phil Gibbard, Wim Hoek, John Lowe, John Andrews, Svante Björck, Les C. Cwynar, Konrad Hughen, Peter Kershaw, Bernd Kromer, Thomas Litt, David J. Lowe, Takeshi Nakagawa, Rewi Newnham, and Jakob Schwander (2009). "Formal definition and dating of the GSSP (Global Stratotype Section and Point) for the base of the Holocene using the Greenland NGRIP ice core, and selected auxiliary records". In: *J. Quat. Sci.* 24.1, pp. 3–17. ISSN: 02678179 10991417. DOI: 10.1002/jqs.1227.

Wanner, Heinz, Jürg Beer, Jonathan Bütikofer, Thomas J. Crowley, Ulrich Cubasch, Jacqueline Flückiger, Hugues Goosse, Martin Grosjean, Fortunat Joos, Jed O. Kaplan, Marcel Küttel, Simon A. Müller, I. Colin Prentice, Olga Solomina, Thomas F. Stocker, Pavel Tarasov, Mayke Wagner, and Martin Widmann (2008). "Mid- to Late Holocene climate change: an overview". In: *Quaternary Science Reviews*

27.19-20, pp. 1791–1828. ISSN: 0277-3791. DOI: 10.1016/j.quascirev.2008.06.013. URL: http://www.sciencedirect.com/science/article/pii/S0277379108001479.

Weitzel, Nils, Sebastian Wagner, Jesper Sjolte, Marlene Klockmann, Oliver Bothe, Heather Andres, Lev Tarasov, Kira Rehfeld, Eduardo Zorita, Martin Widmann, Philipp Sommer, Gerd Schädler, Patrick Ludwig, Florian Kapp, Lukas Jonkers, Javier Garcìa-Pintado, Florian Fuhrmann, Andrew Dolman, Anne Dallmeyer, and Tim Brücher (2019). "Diving into the Past: A Paleo Data–Model Comparison Workshop on the Late Glacial and Holocene". In: *Bulletin of the American Meteorological Society* 100.1, ES1–ES4. DOI: 10.1175/bams-d-18-0169.1.

Whitmore, J., K. Gajewski, M. Sawada, J.W. Williams, B. Shuman, P.J. Bartlein, T. Minckley, A.E. Viau, T. Webb, S. Shafer, P. Anderson, and L. Brubaker (2005). "Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications". In: *Quaternary Science Reviews* 24.16-17, pp. 1828–1848. DOI: 10.1016/j.quascirev.2005.03.005.

Williams, John W., Eric C. Grimm, Jessica L. Blois, Donald F. Charles, Edward B. Davis, Simon J. Goring, Russell W. Graham, Alison J. Smith, Michael Anderson, Joaquin Arroyo-Cabrales, Allan C. Ashworth, Julio L. Betancourt, Brian W. Bills, Robert K. Booth, Philip I. Buckland, B. Brandon Curry, Thomas Giesecke, Stephen T. Jackson, Claudio Latorre, Jonathan Nichols, Timshel Purdum, Robert E. Roth, Michael Stryker, and Hikaru Takahara (2018). "The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource". In: *Quaternary Research* 89.1, pp. 156–177. DOI: 10.1017/qua.2017.105.

Wodehouse, Roger Philip (1935). *Pollen grains: Their structure, identification and significance in science and medicine*. McGraw-Hill Book Co.

Zängl, Günther, Daniel Reinert, Pilar Rípodas, and Michael Baldauf (2014). "The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core". In: *Quarterly Journal of the Royal Meteorological Society* 141.687, pp. 563–579. DOI: 10.1002/qj.2378.

Zender, Charles S. (2008). "Analysis of self-describing gridded geoscience data with netCDF Operators (NCO)". In: *Environmental Modelling & Software* 23.10-11, pp. 1338–1342. DOI: 10.1016/j.envsoft.2008.03.004.

— (2016). "Bit Grooming: statistically accurate precision-preserving quantization with compression, evaluated in the netCDF Operators (NCO, v4.4.8+)". In: *Geoscientific Model Development* 9.9, pp. 3199–3211. DOI: 10.5194/gmd-9-3199-2016.

Zender, Charles S. and Harry Mangalam (2007). "Scaling Properties of Common Statistical Operators for Gridded Datasets". In: *The International Journal of High Performance Computing Applications* 21.4, pp. 485–498. DOI: 10.1177/1094342007083802.

Zhang, Q., H. S. Sundqvist, A. Moberg, H. Körnich, J. Nilsson, and K. Holmgren (2010). "Climate change between the mid and late Holocene in northern high latitudes – Part 2: Model-data comparisons". In: *Climate of the Past* 6.5, pp. 609–626. DOI: 10.5194/cp-6-609-2010.

Part I

New Software Tools for Paleoclimate Analysis

Part II

Numerical Analysis of Paleoclimate Data

Appendices

Todo list

Acronyms

APD African Pollen Database. 1, 4

API Application programming interface. 1, 2

CDO Climate Data Operator. 1, 2

CI Continuous Integration. 1, 4

EMIC Earth System Model of Intermediate Complexity. 1

EMPD Eurasian Modern Pollen Database. 1, 4, 6

EPD European Pollen Database. 1, 6

ESM Earth System Model. 1, 2,

FOSS Free and Open-Source Software. 1

GAM Generalized Additive Model. 6

GCM General Circulation Model. 1

GUI graphical user interface. 1, 2

JJA summer (June, July and August). 6

LAPD Latin American Pollen Database. 1, 4

LGM Last Glacial Maximum. 1

LIG Last Interglacial. 1

MAT modern analogue technique. 6

MCMC Markov chain Monte Carlo. 6

NAPD North American Pollen Database. 1

NCDC National Climatic Data Center. 1

NCO netCDF Operator. 1

NOAA National Oceanic and Atmospheric Administration. 1

PMIP Paleoclimate Modelling Intercomparison Project. 1