UNIVERSITÉ DE LAUSANNE

DOCTORAL THESIS

Software and Numerical Tools for Paleoclimate Analysis

Author:
Philipp S. SOMMER

Supervisor:
Dr. Basil A. S. Davis

Unil

UNIL | Université de Lausanne

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

in the

Davis Group
Institute of Earth Surface Dynamics (IDYST)
Faculty of Geosciences and Environment (GSE)

Jury members:
Prof. Dr. Christian Kull (*president*), Prof. Dr. Grégoire Mariéthoz (*director*),
Basil A. S. Davis (*co-director*),Prof. Dr. John Robert Tipton (*examiner*),
Prof. Dr. Christoph Cornelius Raible (*examiner*)

March 9, 2020

Please cite this chapter as Sommer, 2020

Chapter 7

Conclusions

Paleoclimatological large-scale reconstructions allow an independent evaluation of global climate models skill to simulate climates outside the range of modern climate variability. In the previous chapters of this thesis I described several new open-source tools that can be used to leverage the single site-based proxy-climate reconstruction onto continental, or even global scale, by using a combination of thousands of different records.

For the Eurasian Modern Pollen Database (EMPD) (Davis et al., in prep) I developed a web-framework to communicate and manage the community-driven database in a transparent and sustainable way (chapter 2). The framework consists of an interactive web-based interface into the data and an automated administration webapp. The entire framework is based on the free webservices provided by the version control platform Github and as such allow to trace back every change to the database and provides a variety of tools to manage new contributions and/or changes to the database. This methodology assures stable and intuitive access to the database, independent of the available funding, and contributors or maintainers. One can think of many further potential applications for this framework that can be applied to any regional pollen (or in general, proxy) database. The EMPD is only one example, other potential use cases are the Latin American Pollen Database (LAPD) (Flantua et al., 2015), or the African Pollen Database (APD) (Vincens et al., 2007). The method can also be applied to communicate a study-specific collection of proxy sites and use already implemented analysis and visualization tools for the data, or add new methods specific to the scope of the study. The future plans with this project therefore include a further generalization of the methods, particularly the visualization methods of the EMPD- and POLNET-viewer (section 2.3), to make it widely applicable. The integration with Github allows an easy way to share the source, and to host the interactive interface on the same platform without any costs.

The next tool I presented is the stratigraphic digitization software straditize in chapter 3 (Sommer et al., 2019). This package transforms stratigraphic diagrams, i.e. diagrams where the analysis of samples are plotted against a common y-axis, usually representing age or depth. The potential applications for this software are numerous because of the existence of hundreds of pollen datasets (and more) that are only available as pollen diagrams in the publications. This software provides the unique possibility to make this data from the pre-digital era accessible in a reasonable amount of time. Further extensions to this package will involve the support of new diagram types (e.g. multiple lines in a single diagram column). A strong focus will lie on the documentation of the software in order to make it easier and accessible. This will involve video tutorials, more tutorials for the various diagrams directly embedded into the software, and there are still some parts of the software that are not yet sufficiently document.

In chapter 4, I further described the interactive visualization framework psyplot (Sommer, 2017), a cross-platform open source python project that combines plotting and data management into a single framework that can be used conveniently via command-line and a graphical user interface (GUI). The software differs from most of the visual analytic software such that it focuses on extensibility and flexibility and can therefore be used in a variety of research questions. It particularly provides the basis for the straditize software (chapter 3) and for multiple of the analysis in the chapters 5 and 6. psyplot currently provide visualization methods that range from simple line plots, to density plots, regression analysis and geo-referenced visualization in two dimensions. In the future, this will be enhanced with 3D visualization methods to provide the first visualization tool in climate research that can be used conveniently from the command line (Sommer, 2019).

In the second part of my thesis I described two new computational models that leverage site-based information into models for large-scale prediction of paleo environments. The first one is the pyleogrid package, a new method for a spatio-temporal gridded climate estimate from a database of proxy-climate reconstruction. It is a probabilistic extension of the method by Mauri et al., 2015 and Davis et al., 2003 that provides reliable uncertainties by incorporating the intrinsic dating and proxy-climate reconstruction uncertainties of the individual sites, in order to generate a product that can conveniently be used for data-model intercomparison. In addition, this framework contains two novel methods, (1) a model to predict dating uncertainties based on the age of the pollen sample and the time-difference to the closest chronological control point, and (2) a new probabilistic version of the modern analogue technique (MAT), based on constrained Gibbs sampling algorithms for the age of the samples and the climate reconstructions. The ensemble method is very scalable, both in terms of the size of the spatio-temporal domain, and the computational resources that are used for the prediction. This method will be used in the near future to generate a climate reconstruction for the entire northern hemisphere that is based on the POLNET database as described in Davis and Kaplan, 2017. Further developments will also concentrate on a revision of the age uncertainty estimate using the recent database by Wang et al., 2019 which contains standardized chronologies for more than 500 datasets.

Finally, the second model in chapter 6 describes the new global weather generature GWGEN (Sommer and Kaplan, 2017), a statistical model for a temporal downscaling of monthly to daily climatology. This model can be applied on the entire globe whilst being parameterized based on a large dataset of weather stations with more than 50 million records. The aim is to provide a tool that can be implemented in a global vegetational model for paleo environments. These models require daily meteorology as input which poses a considerable challenge considering the long simulation period they have.

I additionally developed another model in collaboration, that is not related to paleo and therefore not included in the main part of this thesis. This model, the Integrated Urban Complexity Model (IUCM) (Cremades and Sommer, 2019), presents a new method that simulates the transformation of urban areas while focusing on a low energy consumption from urban mobility. I mention this model here because it also incorporates the infrastructural methodologies that I developed for psyplot and GWGEN. This additional use-case highlights one of the important aspects of open-source software development, that is a flexible, extensible and sustainable modular framework, where the packages related to a specific product can be used in multiple other products. Other examples for it are the *docrep* and *sphinx-nbexamples* packages (Sommer, 2018a,b) that I primarily developed for psyplot but are used in a variety

of recently developed packages now (e.g. Abernathey et al., 2017; Banihirwe et al., 2019; Uchida, 2018).

References

Abernathey, Ryan, Takaya Uchida, Julius Busecke, Dhruv Balwada, Ci Zhang, and Anirban Sinha (July 2017). *xgcm/xgcm: v0.1.0*. Version v0.1.0. DOI: 10.5281/zenodo.826926. URL: https://doi.org/10.5281/zenodo.826926.

Banihirwe, Anderson, Matthew Long, Alper Altuntas, Riley Brady, Michael Levy, and Sudharsana K J L (Apr. 2019). *NCAR/esmlab: v2019.04.27*. Version v2019.04.27. DOI: 10.5281/zenodo.2652704. URL: https://doi.org/10.5281/zenodo.2652704.

Cremades, R. and Philipp S. Sommer (2019). "Computing climate-smart urban land use with the Integrated Urban Complexity model (IUCm 1.0)". In: *Geoscientific Model Development* 12.1, pp. 525–539. DOI: 10.5194/gmd-12-525-2019. URL: https://www.geosci-model-dev.net/12/525/2019/.

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, Basil A. S., Manuel Chevalier, Philipp S. Sommer, et al. (in prep). "The Eurasian Modern Pollen Database (EMPD), Version 2". In: *Earth System Science Data ESSD*.

Davis, Basil A. S. and Jed O. Kaplan (Feb. 2017). *HORNET Holocene Climate Reconstruction for the Northern Hemisphere Extra-tropics*. SNF-Research-Plan. last accessed Jan, 30th, 2018. URL: http://p3.snf.ch/project-169598#.

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 (Dec. 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.

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 multiproxy 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.

Sommer, Philipp S. (Aug. 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.

— (2018a). *docrep: A Python Module for intelligent reuse of docstrings*. Last accessed: 2018-02-03. URL: https://github.com/Chilipp/docrep (visited on 02/03/2018).

— (2018b). *sphinx-nbexamples: Create an examples gallery with sphinx from Jupyter Notebooks*. Last accessed: 2018-02-03. URL: https://github.com/Chilipp/sphinx-nbexamples (visited on 02/03/2018).

— (2019). *psy-vtk: A VTK plugin for psyplot*. Last accessed: 2019-05-27. URL: https://github.com/Chilipp/psy-vtk (visited on 05/27/2019).

Sommer, Philipp S. and Jed O. Kaplan (Oct. 2017). "A globally calibrated scheme for generating daily meteorology from monthly statistics: Global-WGEN (GW-GEN) 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 A. S. Davis (Feb. 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.

Uchida, Takaya (Aug. 2018). *xgcm/xrft: First Official Release*. Version v0.1-beta. DOI:
    10.5281/zenodo.1402636. URL: https://doi.org/10.5281/zenodo.1402636.

Vincens, Annie, Anne-Marie Lézine, Guillaume Buchet, Dorothée Lewden, and An-
    nick 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.

Wang, Yue, Simon J. Goring, and Jenny L. McGuire (2019). "Bayesian ages for pollen
    records since the last glaciation in North America". In: *Scientific Data* 6.1, p. 176.
    ISSN: 2052-4463. DOI: 10.1038/s41597-019-0182-7. URL: https://doi.org/10.
    1038/s41597-019-0182-7.

List of Abbreviations

APD African Pollen Database. 15, 32, 139

API Application programming interface. 19, 55, 57, 60, 88

CDO Climate Data Operator. 54, 55

CI Continuous Integration. 18, 19, 32, 34, 37

EMPD Eurasian Modern Pollen Database. 20, 31, 32, 33, 34, 35, 37, 38, 73, 139

EPD European Pollen Database. 73

ESM Earth System Model. 14, 54

FOSS Free and Open-Source Software. 17, 18, 19

GAM Generalized Additive Model. 77, 78, 79

GUI graphical user interface. 19, 20, 53, 54, 58, 59, 60, 61, 62, 140

JJA summer (June, July and August). 72, 75, 84

LAPD Latin American Pollen Database. 15, 32, 139

LGM Last Glacial Maximum. 14

MAT modern analogue technique. 75, 84, 90, 95, 97, 140

MCMC Markov chain Monte Carlo. 82, 83

NAPD North American Pollen Database. 15

NCDC National Climatic Data Center. 16

NOAA National Oceanic and Atmospheric Administration. 16

PMIP Paleoclimate Modelling Intercomparison Project. 14