

UNIVERSITÉ DE LAUSANNE

DOCTORAL THESIS

Software and Numerical Tools for Paleoclimate Analysis

Author:

Philipp S. SOMMER

Supervisor:

Dr. Basil A. S. Davis



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

Sommer, Philipp S. (2020). "Frontmatter". In: *Software and Numerical Tools for Paleoclimate Analysis*. Ed. by Philipp S. Sommer. University of Lausanne. Chap. 0, pp. 1–18. DOI: [10.5281/zenodo.3757356](https://doi.org/10.5281/zenodo.3757356).

IMPRIMATUR

Vu le rapport présenté par le jury d'examen, composé de

Président de la séance publique :	M. le Professeur Christian Kull
Président du colloque :	M. le Professeur Christian Kull
Directeur de thèse :	M. le Professeur Grégoire Mariéthoz
Co-directeur de thèse :	M. le Docteur Basil Davis
Expert externe :	M. le Professeur John Robert Tipton
Expert externe :	M. le Professeur Christoph Cornelius Raible

Le Doyen de la Faculté des géosciences et de l'environnement autorise l'impression de la thèse de

Monsieur Philipp SOMMER

Titulaire d'un
Master in Integrated Climate System Sciences
De l'Université de Hambourg

intitulée

Software and Numerical Tools for Paleoclimate Analysis

Lausanne, 28 février 2020

Pour le Doyen de la Faculté des géosciences et de
l'environnement


Professeur Christian Kull

“The purpose of computing is insight, not numbers.”

Richard Wesley Hamming

UNIVERSITÉ DE LAUSANNE

Summary

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

Doctor of Science

Software and Numerical Tools for Paleoclimate Analysis

by Philipp S. SOMMER

Data-model comparisons of Holocene (11,700 years ago to present) climate provide an ideal basis for evaluating climate model performance outside the range of modern climate variability. The Holocene is recent enough so that boundary conditions of the underlying physics and forcing are well known, while paleoenvironmental archives are abundant and dated with enough precision to comprehensively reconstruct climate. To date, efforts to reconstruct the spatial patterns of Holocene climate change have been mainly focused on the mid-Holocene (about 6'000 years ago), but significant discrepancies have already been identified in data-model comparisons.

These data-model discrepancies can be investigated using instrumental datasets covering continental or hemispheric scales which allow us to reconstruct large-scale climatic features, such as atmospheric dynamics or latitudinal temperature gradients. The generation of these datasets for times prior to the 19th century however faces considerable challenge because there are very few direct measurements of climatic variables. We rely on climate proxies as indirect measurements of the paleo climate. The most abundant one is fossil pollen data, i.e. pollen that are produced by vegetation and can be preserved over thousands of years in terrestrial (or coastal) archives (e.g. lake sediments). This proxy is available from all non-glaciated continents over the world in many different climate regimes, and the primary data is becoming increasingly accessible through large publicly available and community-driven relational databases. Our ability to use this proxy for continental-scale climate reconstructions, however, depends on our ability to analyze, explore and find patterns in these rich and heterogeneous databases. In particular, this requires a proper understanding of the uncertainties that are related to the indirect measurement of climate.

In the first part of this thesis, I present three new software tools that tackle the challenge to make this large amount of data accessible, and to build and develop a continental-scale pollen database. These tools cover a wide range of possible applications to leverage our work with site-based proxy data to a continental scale. The first tool I present is a web framework that is built around a map-based interactive database viewer, developed primarily for the Eurasian Modern Pollen Database, EMPD. This new tool makes the database accessible to other researchers and to the general public, and it allows a continuous and stable development of the community-driven database. In addition to the EMPD, I present an extension of this viewer that makes a large northern-hemispheric fossil pollen database accessible and allows its visual exploration.

The second tool tackles the challenge to fill the gaps in certain geographic areas in the pollen database. *straditize* is a digitization software for stratigraphic diagrams, and pollen diagrams in particular. It can be used to generate new data for the pollen

database from publications of the pre-digital era, i.e. from publications where the primary pollen data is not accessible anymore but through the visualization in form of a pollen diagram in a peer-reviewed publication.

Finally, I present the generic python visualization framework *psyplot*, that bridges the gap between visualization, computation and publication in the day-to-day work of scientists, and that has been used in multiple parts of the thesis. This flexible software can be integrated and enhanced by a variety of applications and already contains multiple convenient visualization methods useful for climate science, particularly the visualization of geo-referenced data and it handles data that is too large to fit into memory or lives on different structured or unstructured grids.

The second part of my thesis contains two new statistical methods to estimate large-scale paleo climatic environments based on modern day relationships. The first one, *pyleogrid*, uses a large pollen database and turns it into a gridded climate reconstruction that can cover continental, hemispheric or even global scales. This software focuses a lot on the integration of the intrinsic uncertainties in the proxy data. The outcome of this gridding procedure allows a comparison of computational climate models with an independent observational database that comes with reliable estimates of uncertainty.

The last chapter of this thesis applies a converse strategy and uses modern statistical relations within climate variables to inform a computational model. The global weather generator (*GWGEN*) has been parameterized with thousands of global weather stations and provides a statistical tool that downscales monthly to daily climatology on a global scale. This tool can be embedded in a global paleo vegetation model where it efficiently simulates the necessary daily meteorology.

UNIVERSITÉ DE LAUSANNE

Résumé

Faculté des géosciences et de l'environnement

Institut des dynamiques de la surface terrestre

Docteur ès Sciences

Software and Numerical Tools for Paleoclimate Analysis

by Philipp S. SOMMER

Les comparaisons données-modèles du climat de l'Holocène (d'il y a 11 700 ans à aujourd'hui) fournissent une base idéale pour évaluer la performance des modèles climatiques en dehors de la plage moderne de variabilité climatique. L'Holocène est assez récent pour que les conditions aux limites et les différents forçages soient bien connus, tandis que les archives paléoenvironnementales sont abondantes et datées avec suffisamment de précision pour reconstruire complètement le climat. Jusqu'à présent, les efforts pour reconstruire les changements climatiques de l'Holocène spatialement se sont principalement concentrés sur l'Holocène moyen (il y a environ 6 000 ans), mais des divergences significatives ont déjà été identifiées lors des comparaisons données-modèles.

Ces écarts entre les modèles et les données peuvent être étudiés à l'aide de collection de données d'observation couvrant des échelles continentales ou hémisphériques qui nous permettent de reconstruire des caractéristiques climatiques à grande échelle, comme la dynamique atmosphérique ou les gradients de température latitudinaux. La génération de ces ensembles de données pour des périodes antérieures au XIXe siècle est cependant confrontée à un défi considérable, car il existe très peu de mesures directes de ces variables climatiques. Nous nous appuyons sur des mesures indirectes du paléoclimat à l'aide d'indicateurs climatiques. Le plus abondant est le pollen fossile, c'est-à-dire le pollen produit par la végétation et qui peut être conservé pendant des milliers d'années dans des archives terrestres ou côtières (e.g. les sédiments lacustres). Ce « proxy » est disponible sur tous les continents non couverts par les glaces à travers de nombreux régimes climatiques différents à travers le monde, et les données primaires sont de plus en plus accessibles par le biais de grandes bases de données relationnelles accessibles au public et gérées par la communauté. Notre capacité à utiliser ce proxy pour les reconstructions climatiques à l'échelle continentale, cependant, dépend de notre capacité d'analyser, d'explorer et de trouver des modèles dans ces bases de données riches et hétérogènes. En particulier, cela exige une bonne compréhension des incertitudes liées à la mesure indirecte du climat.

Dans la première partie de cette thèse, je présente trois nouveaux outils logiciels qui relèvent le défi de rendre accessible cette grande quantité de données et de construire et développer une base de données de pollen à l'échelle continentale. Ces outils couvrent un large éventail d'applications possibles pour tirer parti de notre travail avec des données proxy basées sur des sites à l'échelle continentale.

Le premier outil que je présente est une application Web qui s'articule autour d'un visualiseur de base de données interactif basé sur des cartes, développé principalement pour la base de données eurasienne moderne sur le pollen (EMPD, Eurasian Modern Pollen Database). Ce nouvel outil rend la base de données accessible à d'autres chercheurs et au grand public, et il permet un développement continu et

stable. En plus des données de l'EMPD, je présente une extension de ce visualiseur qui rend accessible une vaste base de données sur les pollens fossiles de l'hémisphère Nord et permet son exploration visuelle.

Le deuxième outil s'attaque au défi de combler les lacunes dans certaines zones géographiques de la base de données de pollen. *straditize* est un logiciel de numérisation pour les diagrammes stratigraphiques, et les diagrammes de pollen en particulier. Il peut être utilisé pour générer de nouvelles données pour la base de données sur le pollen à partir de publications de l'ère pré-numérique, c'est-à-dire de publications dont les données primaires sur le pollen ne sont plus accessibles, mais dont la visualisation sous forme de diagramme de pollen est possible dans une publication.

Enfin, je présente le package python de visualisation générique *psyplot*, qui comble l'écart entre la visualisation, le calcul et la publication dans le travail quotidien des scientifiques, et qui a été utilisé dans plusieurs parties de la thèse. Ce logiciel flexible peut être intégré et amélioré par une variété d'applications et contient déjà de multiples méthodes de visualisation pratiques utiles pour la climatologie, en particulier la visualisation de données géoréférencées et il traite des données qui sont trop grandes pour tenir en mémoire ou qui vivent sur différentes grilles, structurées ou non.

La deuxième partie de ma thèse contient deux nouvelles méthodes statistiques pour estimer les environnements paléoclimatiques à grande échelle basées sur les relations modernes. La première, *pyleogrid*, utilise une grande base de données de pollen et la transforme en une reconstruction climatique maillée qui peut couvrir des échelles continentales, hémisphériques ou même globales. Ce logiciel se concentre sur l'intégration des incertitudes intrinsèques des données proxy. Le résultat de cette méthode de maillage permet de comparer des modèles climatiques computationnels avec une base de données d'observation indépendante qui fournit des estimations fiables de l'incertitude.

Le dernier chapitre de cette thèse applique une stratégie inverse et utilise les relations statistiques des variables climatiques modernes pour informer un modèle. Le générateur météorologique global (*GWGEN*, global weather generator) a été paramétré avec des données provenant de milliers de stations météorologiques mondiales et fournit un outil qui permet de passer de l'échelle mensuelle à l'échelle quotidienne (« *downscaling* ») pour le monde entier. Cet outil peut être intégré dans un modèle global de paléovégétation où il simule efficacement la météorologie quotidienne nécessaire.

Acknowledgements

This work and my entire PhD would not have been possible without the lot of support I received from numerous people. Most of all I have to thank my supervisor Basil Davis for becoming my mentor after the first year of my PhD and introducing me into the great world of paleo-climatic research. I especially want to thank him for his support and understanding during my critical third year, which would have been so much worse without it. I am deeply grateful for his advice, ideas and experiences that constantly improved my understanding of the big climate picture, I am thankful for his support during the rush in the last weeks of my PhD. I further want to thank (my unofficial second supervisor) Manuel Chevalier for his great experience, his open mind and his impressive skill to quickly familiarize with new topics and discuss them. Thank you both that I could always come to you and discuss any issue with my research. All the software packages that I present in this thesis would not have been useful at all without your opinions and your input.

I also want thank Jed O. Kaplan for the collaboration that we had during the development of the weather generator and for being my supervisor during the first year of my PhD. And I want to thank John Tipton for his inspiring methodological experience and his great advice during our collaboration, particularly for his input in the *pyleogrid* chapter. Special thanks goes to Grégoire Mariéthoz for his methodological advice with several statistical issues that I had during the last four years.

It has been a great pleasure to work in the Institute of Earth Surface Dynamics at the University of Lausanne and I want to thank the Swiss National Science Foundation, as well as the Faculty of Geosciences and Environment at UNIL for the support of our HORNET (200021_169598) and the ACACIA (CR10I2_146314) projects. I was very lucky to do my research in such a wonderful environment, particularly with Mathieu Gravey, Inigo Irarrazaval, Luiz Gustavo Rasera, Harsh Beria, Anthony Michelin (for all the little teasing), and especially with Dilan Rech, Lucien Goldenschue, Leanne Phelps, Ryan Hughes, John Shekeine and Andrea Kay. I thank you all for the nice discussions that we had, for your advice in the preparation of multiple talks, and for being good friends during the past four years.

Finally, this entire work would not have been possible without the skills I developed during my Master at the University of Hamburg and the Max-Planck-Institute for Meteorology. I am very thankful for the possibilities I had that allowed me to learn programming during this time, and I am thankful for the ongoing friendship with my former colleagues, Stefan Hagemann, Tobias Stacke and Philipp de Vrese.

To conclude this long list here, which is still not finished, I want to thank my family, particularly my parents and brothers for the ongoing wonderful relationship that we have, and I want to thank my new family, Sylvia Weiß-Fiebig and Gernot Fiebig for their support of my new little family that I got during this PhD. This little family is the best thing in my life and I am extremely thankful to the most important part of it, Bianca, for staying with me after all the difficulties we experienced, being such a good friend, a wonderful partner and a loving mom.

Dedicated to Bianca and Leo, my personal key to happiness...

List of Abbreviations

APD African Pollen Database. 13, 30, 139

API Application programming interface. 17, 55, 57, 60, 88

CDO Climate Data Operator. 54, 55

CI Continuous Integration. 16, 17, 30, 32, 35

EMPD Eurasian Modern Pollen Database. 18, 29, 30, 31, 32, 33, 35, 36, 73, 139

EPD European Pollen Database. 73

ESM Earth System Model. 12, 54

FOSS Free and Open-Source Software. 15, 16, 17

GAM Generalized Additive Model. 77, 78, 79

GUI graphical user interface. 17, 18, 53, 54, 58, 59, 60, 61, 62, 140

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

LAPD Latin American Pollen Database. 13, 30, 139

LGM Last Glacial Maximum. 12

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

MCMC Markov chain Monte Carlo. 82, 83

NAPD North American Pollen Database. 13

NCDC National Climatic Data Center. 14

NOAA National Oceanic and Atmospheric Administration. 14

PMIP Paleoclimate Modelling Intercomparison Project. 12