UNIVERSITÉ DE LAUSANNE

DOCTORAL THESIS

Software and Numerical Tools for Palaeoclimate 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)

August 30, 2019

Declaration of Authorship

I, Philipp S. SOMMER, declare that this thesis titled, "Software and Numerical Tools for Palaeoclimate Analysis" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism."

Dave Barry

UNIVERSITÉ DE LAUSANNE

Abstract

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

Doctor of Science

Software and Numerical Tools for Palaeoclimate Analysis

by Philipp S. SOMMER

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too…

Acknowledgements

The acknowledgments and the people to thank go here, don’t forget to include your project advisor…

Contents

For/Dedicated to/To my…

Chapter 1

Introduction

1.1 Motivation

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 solved by high quality and flexible software packages.

Add reference.

Add reference.

The Neotoma Database, a global international database for palaeoenvironmental 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.

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

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. 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).

Add reference. ICON

Add reference. POLNET-gridding paper

In the following section 1.2 I will lay down the interest in the study of palaeoclimates, 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 palaeo-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 palaeo-climates

Add reference.
Add reference.
Add reference. cite World bank report?
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 additionally 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).

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 palaeo-climates, provide the only opportunity to evaluate an ESM under conditions very different than today. palaeo-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; 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.

Add reference. Check these

The end of this period can be used for data-model comparisons due to the availability of palaeo-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) .

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) .

Add reference. PMIP paper
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 bogsor 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.

Add reference.

Add reference.

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 palaeo-climate proxy (H. J. B. Birks and H. H. Birks, 1980) and has a long history in quantitative palaeo-climatologic reconstructions (e.g. Bradley, 1985; Nichols, 1967, 1969).

Add reference.

Add reference.

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 .

Add reference.

Add reference.

Add reference.

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

Add reference. Manus review paper

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

Add reference. cite some MAT, WAPLS, Bayesian, etc. papers

This paragraph should be rewritten based on Grimm, 2008, section 1.3.1! Another useful feature of this proxy for data-model intercomparisons is the existence of databases for fossil pollen assemblages. Freely available data comes from the European Pollen Database (EPD) (Vincens et al., 2007) or the North American Pollen Database (NAPD), that both started in the 80s, and more recent from Africa and Latin America (Flantua et al., 2015; Fyfe et al., 2009; Marchant et al., 2002)). Another recent attempt is the Neotoma database (Williams et al., 2018), a global multiproxy database that also incorporates many of the regional pollen databases.

The use of the above-mentioned proxies, particularly pollen, for palaeo-climate reconstruction has a long academic tradition in geology (Bradley, 1985) and provides the source of large-scale palaeo-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).

Add reference. add more..., Climate12K

Add paragraph about heterogeneity of data, distribution of data, accessibility of data – need for software tools

1.2.2 Model Simulations from the Holocene

add PMIP experments, highlight transient model runs

Add paragraph about size of data, distribution of data, accessibility of data – need for software tools

Add models for transient runs: LOVECLIM, TRACE, MPI-ESM-LR, HadCM3, FAMOUS

1.3 Software for Palaeoclimatology

1. History of Software development in Earth System Science
   - Development of Climate models
   - Statistics
   - Visualization
   - Distribution and Synthesis of Data

- Quantitative and large-scale reconstructions through proxies

2. An overview on open-source Software Development

   - Version control
   - Transparency
   - Automated tests through Continuous Integration
   - Accessible and extensive documentation
   - Distribution of Software through Package managers (conda, PyPi, Docker)

The usage of software is crucial for the quantitative reconstruction of Earth's Climate. Palaeoclimate 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 palaeoclimate 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 palaeoclimate 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).

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 palaeoclimatic 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 by M. Chevalier et al., 2014; Manuel Chevalier, 2019.

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 where the only possibility to publish the entire dataset (see also chapter 3). The generation of these diagrams

[1] https://products.office.com/en/excel

[2] http://polsystems.rits-palaeo.com

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, , see also 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 Simulations

Software and computational numerics play a crucial row for our understanding of climate since the beginning of the development of General Circulation Models (GCMs) after world war II (Edwards, 2010; Lewis, 1998; Phillips, 1956). 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.

Böttinger and Röber, 2019; Hoyer and Hamman, 2017; Nocke et al., 2008; Rautenhaus et al., 2018; Schneider, 2012; Treut et al., 2007, UV-CDAT

1.3.3 An Introduction into Open-Source Software Development

1.4 Challenges tackled in this thesis

- Visual analysis of large amounts of data (psyplot)
- Synthesizing and Distributing large amounts of proxy data (straditize, EMPD/POLNET viewer)
- Understanding and modelling past climates with statistical methods (Teleconnections, GWGEN)
- Thesis structure:

[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

1. Software tools: Chapters 2, 3, 4
2. Numerical tools: Chapters 5, 6
3. Appendix

References

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.

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, 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.

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.

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.

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.

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.

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.

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.

— (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.

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.

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.

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.

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

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.

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

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.

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.

Marcott, S. A., J. D. Shakun, P. U. Clark, and A. C. Mix (2013). “A Reconstruction of Regional and Global Temperature for the Past 11,300 Years”. In: *Science* 339.6124, pp. 1198–1201. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1228026. URL: http://www.ncbi.nlm.nih.gov/pubmed/23471405.

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.

Mauri, A., B. A. S. Davis, P. M. Collins, and J. O. Kaplan (2014). “The influence of atmospheric circulation on the mid-Holocene climate of Europe: a data-model comparison”. In: *Clim. Past* 10.5, pp. 1925–1938. ISSN: 1814-9324. DOI: 10.5194/cp-10-1925-2014. URL: http://www.clim-past.net/10/1925/2014/cp-10-1925-2014.pdf.

— (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.

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.

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/.

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.

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.

Schneider, Birgit (2012). "Climate model simulation visualization from a visual studies perspective". In: *Wiley Interdisciplinary Reviews: Climate Change* 3.2, pp. 185–193. DOI: 10.1002/wcc.162.

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.

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.

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.

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.

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.

Chapter 2

Psyplot

A flexible framework for interactive data analysis

Need to write chapter 2

- Summary (from JOSS Paper)
- Other visualization frameworks (short review)
- The psyplot framework design
- Integration into a graphical user interface

Chapter 3

Straditize

A digitization software for pollen diagrams

Need to write chapter 3

- quaternary paper
- straditize builds upon the psyplot GUI
- fill the gaps for large-scale climate reconstructions

Chapter 4

The EMPD- and POLNET web-interface to pollen data

Need to write chapter 4

- quaternary paper
- straditize builds upon the psyplot GUI
- fill the gaps for large-scale climate reconstructions

Chapter 5

GWGEN v1.0

A globally calibrated scheme for generating daily meteorology from monthly statistics

Abstract. While a wide range of earth system processes occur at daily and even sub-daily timescales, many global vegetation and other terrestrial dynamics models historically used monthly meteorological forcing, both to reduce computational demand and because global datasets were lacking. Recently, dynamic land surface modeling has moved towards resolving daily and subdaily processes, and global datasets containg daily or sub-daily meteorology have become available. These meteorological datasets, however, cover only the instrumental era of the last ca. 100 years at best, are subject to considerable uncertainty, and represent extremely large data files with associated computational costs of data input/output and file transfer. For periods before the recent past or into the future, global meteorological forcing can be provided by climate model output, but the quality of these data at high temporal resolution is low, particularly e.g., with daily precipitation frequency and amount. Here we present ARVE-WGEN, a globally applicable statistical weather generator for the temporal downscaling of monthly climatology to daily meteorology. Our weather generator is parameterized using a global meteorological database and simulates daily values of five common daily weather variables: minimum and maximum temperature, precipitation, cloud cover, and windspeed. ARVE-WGEN is lightweight, modular, and requires a minimal set of monthly mean variables as input. The weather generator may be used in a range of applications, for example, in global vegetation, crop, erosion, or hydrological models.

5.1 Introduction

The development of the first global vegetation models in the 1970's (e.g. Helmut Lieth, 1975) brought about the demand for meteorological forcing datasets with global extent and relatively high spatial resolution, e.g., 1x1°. While a global weather station-based monthly climate dataset was available at this time (Walter and H Lieth, 1967), limitations in computers and data storage allowed only the simplest treatment of these data. The first global simulations of the net primary productivity of the terrestrial biosphere (Helmut Lieth, 1975), thus used rasterized polygons of annual meteorological variables that had been crudely interpolated from the station-based climatology. A decade later saw the development of better computers and more sophisticated global vegetation models (I. C. Prentice et al., 1992; I. Prentice, 1989) that recognized the need for forcing at a sub-annual timestep and development of these models was done in parallel with the first global, gridded high resolution (0.5°)

monthly climatology (Leemans and Cramer, 1991). At the time, monthly meteorological data was the only feasible global data that could be produced, in terms of the raw station data available to feed the interpolation process, the processing time required to produce gridded maps, and the data storage and transfer capabilities of contemporary computer systems and networks. Global gridded monthly climate data became the standard for not only large-extent vegetation modeling (A. Haxeltine and I. C. Prentice, 1996; Alex Haxeltine et al., 1996; Kaplan et al., 2003; Kucharik et al., 2000; Woodward et al., 1995), but also for a wide range of studies on biodiversity and species distribution (e.g. Elith et al., 2006), vegetation trace gas emissions (e.g. Guenther et al., 1995), and even the geographic distribution of human diseases (e.g. Bhatt et al., 2013)

Over subsequent years, the global gridded monthly climate datasets were improved (New et al., 1999, 2002), developed with very high spatial resolution (Hijmans et al., 2005), and expanded beyond climatological mean climate to cover continuous timeseries over decades (Harris et al., 2014; Mitchell and Jones, 2005; New et al., 2000). The latter was an essential requirement for forcing dynamic global vegetation models (DGVMs) (e.g. Sitch et al., 2003). However despite increasing quality, spatial resolution, and temporal extent in these datasets, the basic time step remained monthly, partly for legacy reasons – models had been developed in an earlier era subject to computational limitations and therefore used a monthly timestep for efficiency even if this was no longer strictly a constraint – and partly because of the challenge in developing a global, high-resolution climate dataset with a daily or shorter timestep still presented a major data management challenge.

On the other hand, there was increasing awareness that accurate simulation of many earth surface processes required representation of processes at a shorter-than-monthly timestep. Global simulation of surface hydrology (Gerten et al., 2004), crop growth (Bondeau et al., 2007), or biogeophysical processes (Krinner et al., 2005) needed sub-monthly forcing to produce reliable results. To address this need for better forcing data, two main approaches were taken: either monthly climate data were downscaled online using a stochastic weather generator (Pfeiffer et al., 2013), or a sub-daily, high-resolution, gridded climate timeseries was generated directly by merging high-temporal-resolution reanalysis data (e.g., NCEP, 6h, 2.5°) with high-spatial-resolution monthly climate data (e.g., CRU, 0.5°). The latter process resulted in the CRUNCEP dataset (Viovy and Ciais, 2016; Wei et al., 2014), which, while global, is large even by modern standards (ca. 350 Gb), is not available at spatial resolution greater than 0.5°, and covers only the period 1901-2014.

Forcing data for global vegetation and other models with shorter than monthly resolution at higher spatial resolutions than 0.5°, or for any other period than the last ca. 120 years, e.g., for the future or the more distant past, may therefore only be available through downscaling techniques. One approach to overcome the limitations of currently available datasets could be to use GCM output directly, however, most GCM output currently available does not have greater than 0.5° spatial resolution, with the current generation of GCMs typically approaching ca. 1°x1° degree. Furthermore, there is a general observation that daily meteorology produced by GCMs is not realistic, particularly for precipitation (Dai, 2006; Stephens et al., 2010; Sun et al., 2006). An alternative approach is, therefore, to perform temporal downscaling on monthly meteorological data using a statistical weather generator.

Statistical weather generators were first developed primarily for crop and hydrological modeling at the field to catchment scale (Richardson, 1981; Woolhiser and Pegram, 1979; Woolhiser and Roldan, 1982). The weather generator was parameterized using daily meteorological observations at one or more weather stations close to the area of interest, although some attempts were made to generalize the parameterization over larger, sub-continental regions (e.g. D. S. Wilks, 1998, 1999b; Woolhiser and Roldán, 1986). Locally parameterized weather generators have been applied to a very wide range of studies (D. S. Wilks and Wilby, 1999; Daniel S. Wilks, 2010), and enhanced to include additional meteorological variables beyond the original precipitation, temperature, and solar radiation (e.g. Parlange and Katz, 2000). Applications of a weather generator at continental to global scales was still limited, however, because of the need to perform local parameterization.

The need to simulate daily meteorology in regions of the world with short, unreliable, or unavailable daily meteorological timeseries brought about the realization that certain features of weather generator parameterization might be generalized across a range of climates (S. Geng et al., 1986; Shu Geng and Auburn, 1987). This ultimately led to the development of globally applicable weather generators (Friend, 1998), and their incorporation in DGVMs (Bondeau et al., 2007; Gerten et al., 2004; Pfeiffer et al., 2013). The original global parameterization (S. Geng et al., 1986) of these weather generators was, however, limited to seven weather stations, mostly in the temperate latitudes. Friend, 1998 does not publish the parameters used in his global weather generator, but we assume these were the same as the original S. Geng et al., 1986; Shu Geng and Auburn, 1987 models. Given the availability of 1) large datasets of daily meteorology, and 2) computers powerful enough to process these data, we therefore decided that it would be valuable to revisit these parameterizations, perform a systematic and quantitative evaluation of the resulting downscaled meteorology, and potentially improve our ability to perform monthly-to-daily downscaling of common meteorological variables with a single, globally applicable parameterization.

In the following sections we describe GWGEN, a weather generator parameterized using more than 50 million daily weather observations, from all continents and latitudes. We demonstrate how updated schemes for simulating precipitation occurrence and amount, and for correcting bias in temperature and wind speed, further improve the quality of the model simulations. We perform an extensive model evaluation and parameter uncertainty analysis in order to settle on a parameter set that provides the most accurate, globally applicable results. We comment on the limitations of the model and priorities for future research. The GWGEN is an open-source, stand-alone model that may be incorporated into any number of models designed to work at global scale, including, e.g., vegetation, hydrology, climatology, and animal distribution models.

5.2 Model development

Our WGEN type weather generator follows the methodology described by Richardson, 1981, the simple method described by S. Geng et al., 1986. It uses a second-order Markov chain as described in D. S. Wilks, 1999a to calculate the precipitation occurrence and a hybrid Gamma-GP distribution (Neykov et al., 2014) to estimate the amount. Temperature, cloud cover and wind speed are calculated via cross correlation and depending on the wet/dry-state of the day. An additional quantile-based bias correction for wind speed and minimum temperature improves the simulation results significantly.

In this section, we describe the global weather station database and provide the underlying figures and equations that have been used for the parameterization of GWGEN. A full model description can be found in section 5.3.

5.2.1 Development of global weather station database

For the parametrization we used temperature, wind and cloud data from the Synoptic Cloud Reports (EECRA) databases (Hahn and Warren, 1999), as well as precipitation and temperature from the daily Global Historical Climatology Network (GHCN) (Menne et al., 2012a,b). The latter consists of roughly 100'000 different stations from which we selected the one with the longest record for each grid cell in a global grid $0.5 \times 0.5°$. Since several of the US stations measure precipitation in inches instead of millimeter per day, we furthermore selected the station that have all numbers from 0.1 to 1.0 millimeter per day in its record. This selection procedure resulted in 16'590 stations.

Our parameterization only uses months with complete records for each day of the month. The finally selected stations from the EECRA and GHCN data are displayed in figure 5.1.

5.2.2 Parameterization

Precipitation occurrence

To apply the Markov chain, we calculated the transition probabilities for a wet day being followed by a wet day (p_{11}), for a wet day being followed by a dry day being followed by a wet day (p_{101}) and for two dry days being followed by a wet day (p_{001}). We did this by first extracting all (complete) Januaries, Februaries, etc. for each of the parameterization stations separately. In a second step, all Januaries were merged together into one large multi-year series and the above mentioned transition probabilities were calculated and fitted towards the number of wet days in this series (see figure 5.2). Like-wise we did it for all Februaries, Marchs, etc..

The result of this procedure are the following relationships

$$p_{11} = 0.2549 + 0.7451 \cdot \frac{\text{wet days in month}}{\text{days in month}} \tag{5.1}$$

$$p_{101} = 0.8463 \cdot \frac{\text{wet days in month}}{\text{days in month}} \tag{5.2}$$

$$p_{001} = 0.7240 \cdot \frac{\text{wet days in month}}{\text{days in month}}. \tag{5.3}$$

Those equation are used in the weather generator to decide whether the current simulated day is a wet day or not based upon the total number of wet days in the month.

Precipitation amount

We apply a hybrid gamma-GP distribution consisting of the the gamma and the generalized pareto (GP) distribution. The gamma distribution function is defined as

$$f(x) = \begin{cases} \frac{x^{\alpha-1} \exp^{-\frac{x}{\theta}}}{\theta^{\alpha}\Gamma(\alpha)} & \text{for } x > 0 \\ 0 & \text{for } x = 0 \end{cases} \tag{5.4}$$

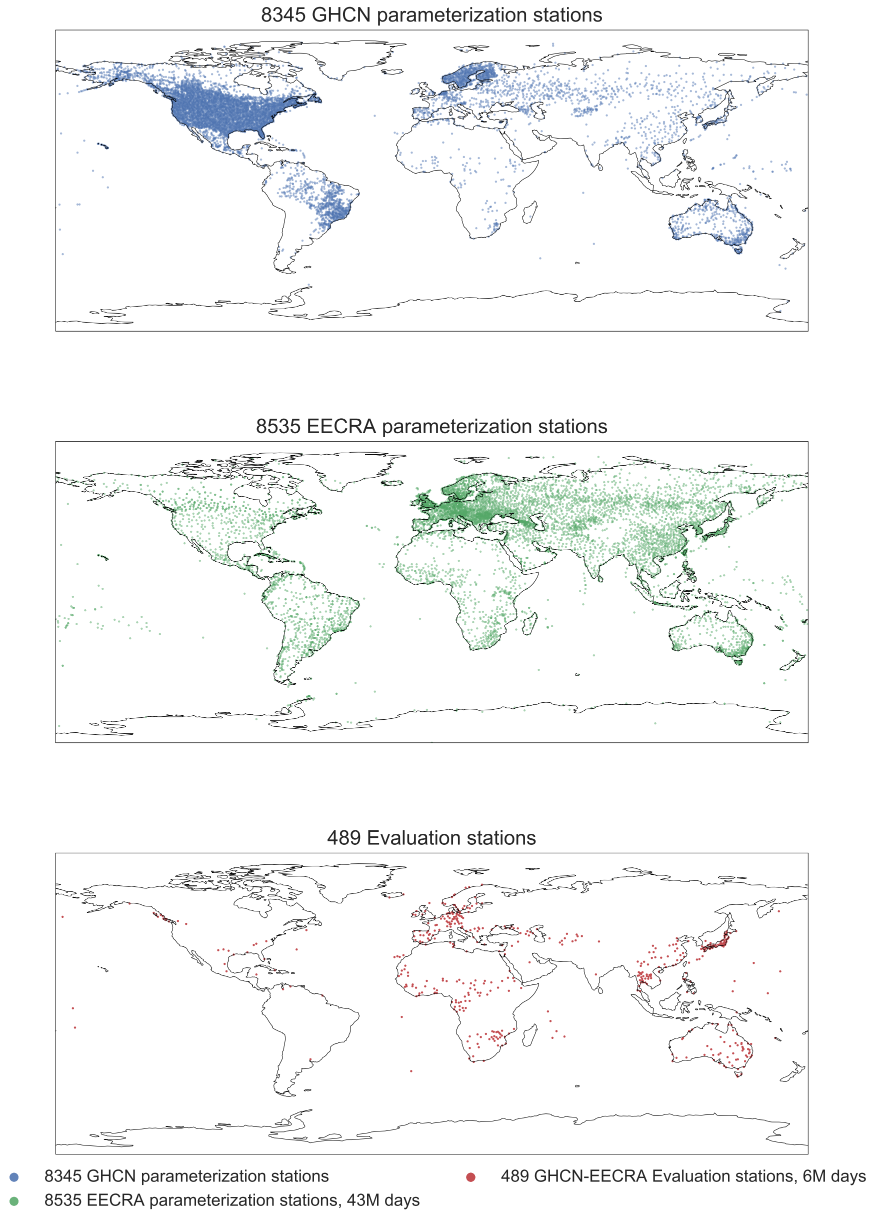

FIGURE 5.1: Stations that are used for parameterization and evaluation of the weather generator. The upper stations are used for parameterizing precipitation and temperature, the middle ones for cloud and wind parameterization, as well as the cross correlation of cloud, temperature and wind. The lower plot shows the evaluation stations.

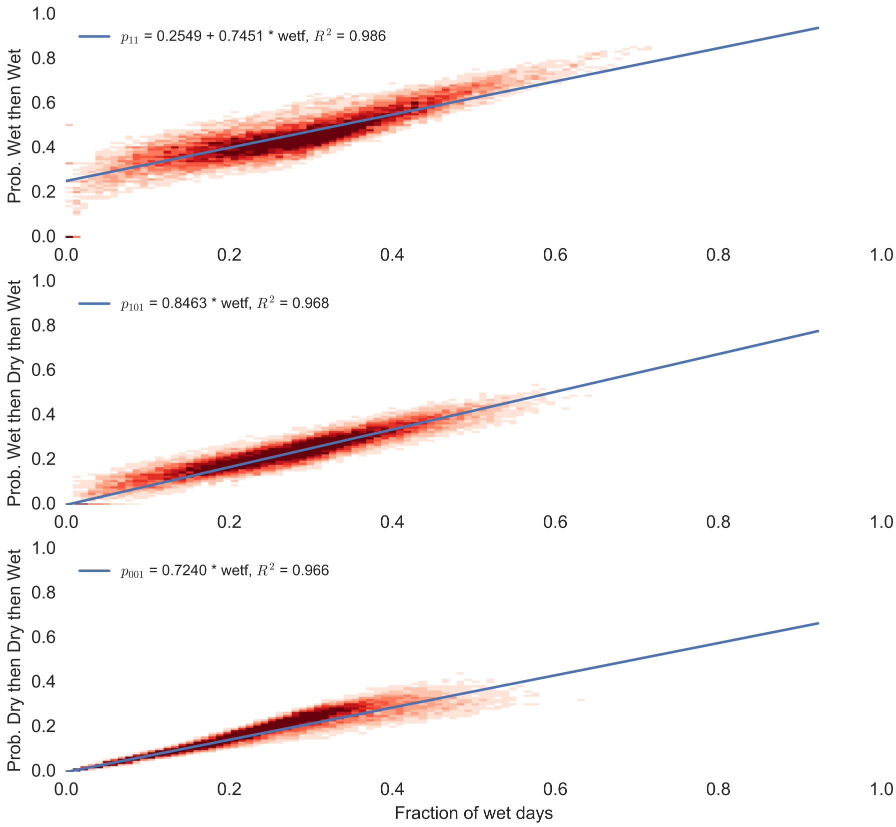

FIGURE 5.2: Transition probabilities vs. wet fraction. The red lines show the fit of the probability against the wet fraction. The fit for the p_{11} transition probability was forced to the point $(1, 1)$, the others were forced to $(0, 0)$.

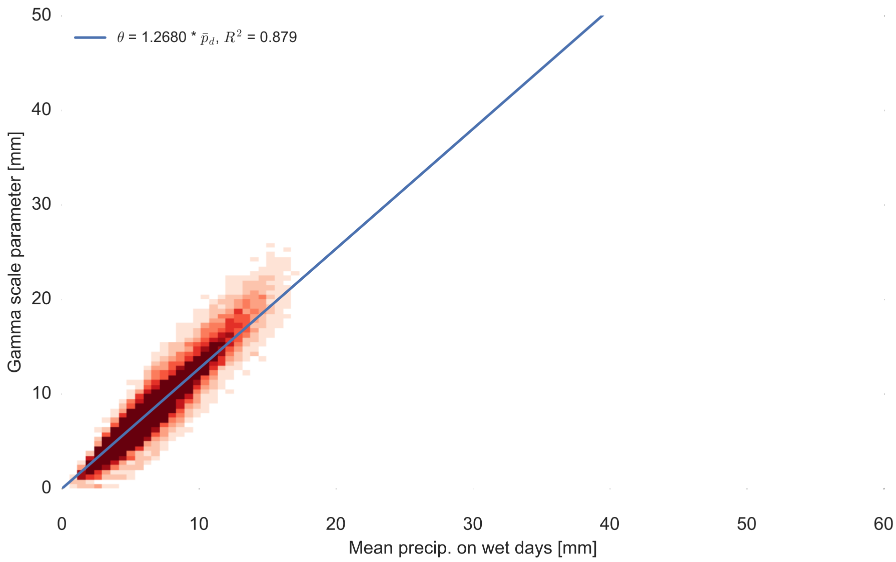

FIGURE 5.3: Mean precipitation - Gamma scale relationship. The blue line represents the best fit line of the mean precipitation on wet days to the estimated gamma scale parameter of the corresponding distribution. Each data point corresponds to one multi-year series of one month for one station.

where $\alpha > 0$ is the shape, and $\theta > 0$ the scale parameter. The generalized pareto (GP) distribution is defined via

$$g(x) = \frac{1}{\sigma}\left(1 + \frac{\xi\,(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}-1} \tag{5.5}$$

with $\sigma > 0$ being the scale parameter and $\xi \in \mathbb{R}$ the shape parameter. μ is the location parameter.

Following Furrer and Katz, 2008, we define the hybrid gamma-GP distribution as

$$h(x) = \begin{cases} f(x) & \text{for } x \leq \mu \\ (1 - F(\mu))\,g(x) & \text{for } x > \mu \end{cases}, \tag{5.6}$$

where $F(\mu)$ describes the cumulative gamma distribution function at the threshold μ. ibid. and Neykov et al., 2014 have shown, that this distribution better simulates the heavy tail of the precipitation distribution than the gamma distribution alone.

To determine the parameters of the hybrid distribution for precipitation, we started with the simple stategy by S. Geng et al., 1986. As we did above when calculating the markov chain parameters, we created multi-year series for each of the parameterization stations for each month and extracted the days with precipitation. If a series contained more than 100 entries, we fit a gamma distribution to it and estimated the α and θ parameters.

Following ibid., we then fit a regression line of the gamma scale parameter against the mean precipitation on wet days $\bar{p}_d$ (see figure 5.3) and found the relationship

$$\theta = 1.268\,\bar{p}_d. \tag{5.7}$$

As proposed by ibid., we use this relationship in our model to estimate the scale parameter of the distribution. The gamma shape parameter α is calculated dynamically in the weather generator via

$$\alpha = \frac{\bar{p}_d}{\theta}. \tag{5.8}$$

The GP scale parameter σ on the other hand is calculated during the simulation following Neykov et al., 2014 via

$$\sigma = \frac{1 - F(\mu)}{f(\mu)}. \tag{5.9}$$

The other parameters of the GP distribution are obtained through a sensitivity analysis described later.

Temperature

For each day we know from the Markov chain approach, whether the current simulated day is a wet or dry day. Based upon the simple linear relationships

$$\begin{aligned} x_{\text{wet}} &= a_{x,\text{wet}} + b_{x,\text{wet}} \cdot \bar{x} \\ x_{\text{dry}} &= a_{x,\text{dry}} + b_{x,\text{dry}} \cdot \bar{x} \end{aligned} \tag{5.10}$$

we adjust the monthly mean $\bar{x}$ of the variable $x \in \{T_{\min}, T_{\max}\}$. The intercept a and slope b of those linear relationships have been determined through a fit of the mean maximum and minimum temperature on wet or dry days in one month to

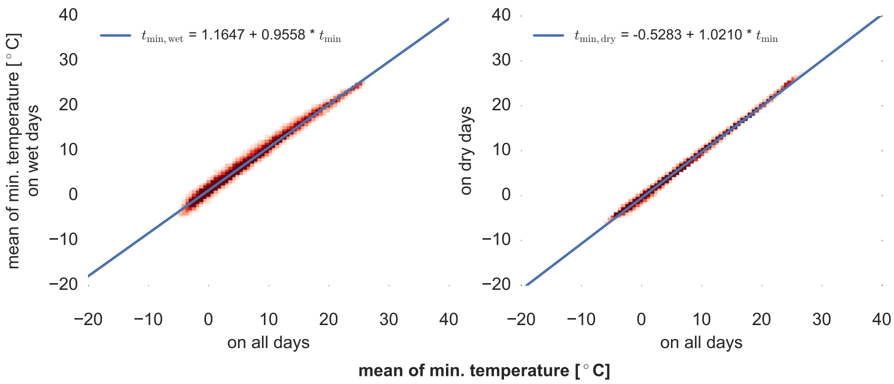

FIGURE 5.4: Correlation of minimum temperature on wet and dry days to the monthly mean. The y-axes show the mean minimum temperature on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 5.1.

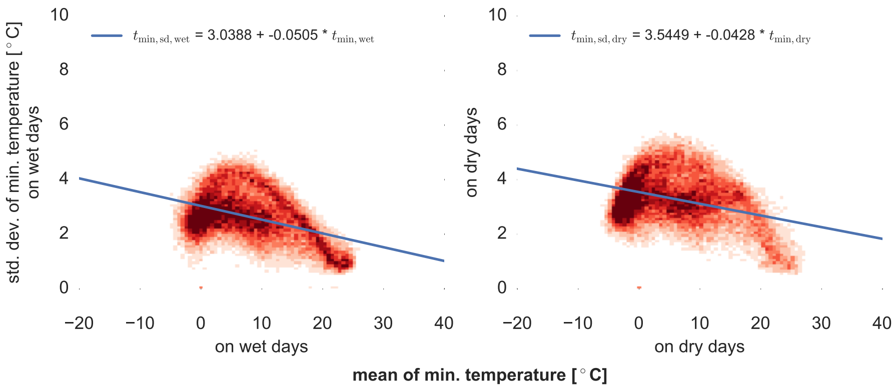

FIGURE 5.5: Correlation of standard deviation of the minimum temperature on wet and dry days to the monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 5.1.

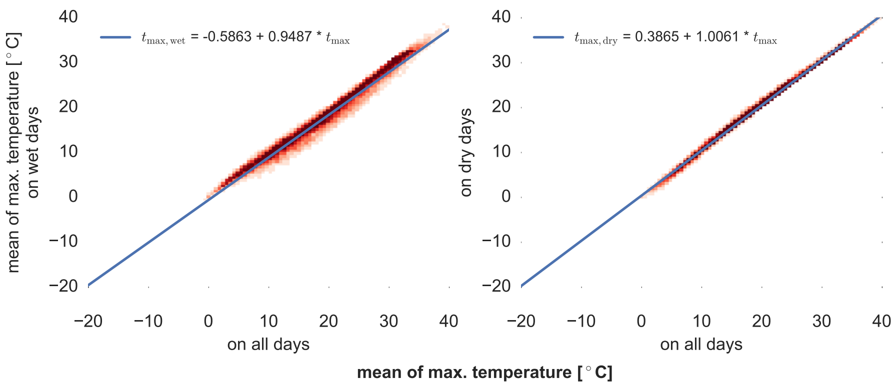

FIGURE 5.6: Correlation of maximum temperature on wet and dry days to the monthly mean. The y-axes show the mean maximum temperature on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 5.1.

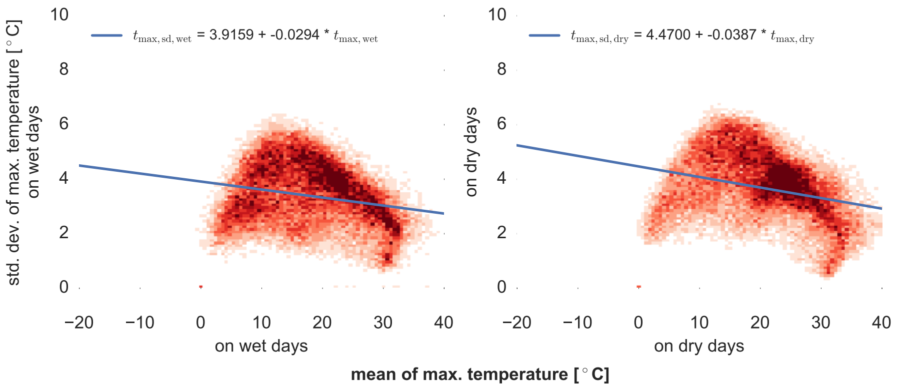

FIGURE 5.7: Correlation of standard deviation of the minimum temperature on wet and dry days to the monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 5.1.

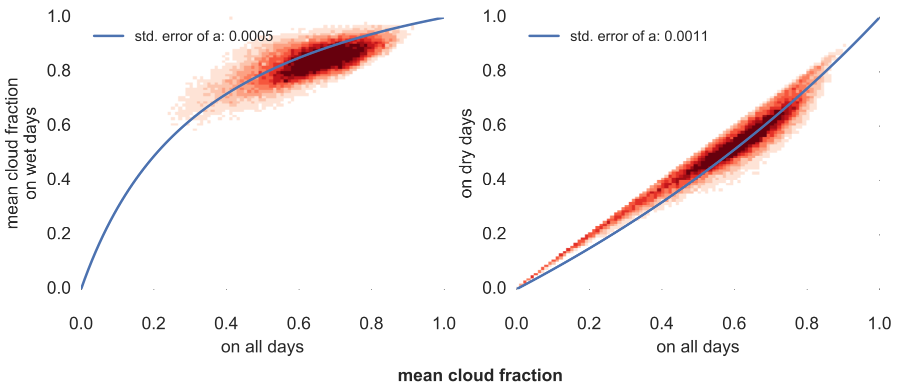

FIGURE 5.8: Correlation of cloud fraction on wet and dry days to the monthly mean.The y-axes show the mean cloud fraction on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 5.2.

the overall monthly mean (see figures 5.4, 5.6 and table 5.1) on our parameterization data.

The standard deviation σ_x of variable $x \in \{T_{\min}, T_{\max}\}$ is then calculated from the adjusted mean via

$$\begin{aligned}\sigma_{x,\mathrm{wet}} &= a_{\sigma,x,\mathrm{wet}} + b_{\sigma,x,\mathrm{wet}} \cdot x_{\mathrm{wet}} \\ \sigma_{x,\mathrm{dry}} &= a_{\sigma,x,\mathrm{dry}} + b_{\sigma,x,\mathrm{dry}} \cdot x_{\mathrm{dry}}.\end{aligned} \tag{5.11}$$

To estimate the values of the parameters a and b in equation (5.11), we used a fit to the GHCN data (figures 5.5, 5.7 and table 5.1). The linear model does not really satisfy the complex behaviour of the standard deviation, but since this value is only used for the random noise (see below), we think that the error is negligible.

TABLE 5.1: Fit results of temperature correlation for wet and dry days for figure 5.4 and 5.6. N is the total number of measurements used for the fit. The values of a and b correspond to the values in equation (5.10) and (5.11).

plot	variable	Intercept	R^2	Slope
5.6	$T_{\mathrm{max,dry}}$	0.3865	0.9907	1.0061
5.6	$T_{\mathrm{max,wet}}$	-0.5863	0.9539	0.9487
5.7	$T_{\mathrm{max,sd,dry}}$	4.4700	0.0722	-0.0387
5.7	$T_{\mathrm{max,sd,wet}}$	3.9159	0.0407	-0.0294
5.4	$T_{\mathrm{min,dry}}$	-0.5283	0.9940	1.0210
5.4	$T_{\mathrm{min,wet}}$	1.1647	0.9733	0.9558
5.5	$T_{\mathrm{min,sd,dry}}$	3.5449	0.1044	-0.0428
5.5	$T_{\mathrm{min,sd,wet}}$	3.0388	0.1386	-0.0505
5.11	$w_{\mathrm{sd,dry}}$	0	0.8959	0.5120
5.11	$w_{\mathrm{sd,wet}}$	0	0.8994	0.4404
5.10	w_{dry}	0	0.9933	0.9452
5.10	w_{wet}	0	0.9720	1.0929

TABLE 5.2: Fit results of cloud correlation for wet and dry days for figure 5.8

plot	variable	a	std. dev. of a
5.8	c_{dry}	0.4205	0.0011
5.8	c_{wet}	-0.7383	0.0005
5.9	$c_{\mathrm{sd,dry}}$	1.0417	0.0003
5.9	$c_{\mathrm{sd,wet}}$	0.9819	0.0005

Cloud fraction

To parameterize the cloud fraction, we first calculated the variable from the EECRA dataset. The original dataset contains eight measurements per day of the total cloud cover ranging from 0 (clear sky) to 8 (overcast). Hence, to calculate the daily cloud fraction, those values were averaged and divided by 8 to get the daily mean.

To adjust the monthly mean depending on the wet-dry state of the day, we could not use the simple linear relationship as we used for temperature because the cloud fraction bounded by the lower limit 0 and the upper limit 1. However, the cloud cover on wet days is usually greater or equal to the monthly mean cloud cover whereas the cloud cover on dry days is usually less or equal to the monthly mean cloud cover. This results in a concave curve for the wet case and a convex curve for dry days. Therefore we came up with the following formula for the fit:

$$\begin{aligned} \mathrm{c}_{\mathrm{wet}} &= \frac{-a_{\mathrm{c,wet}} - 1}{a_{\mathrm{c,wet}}^2 * c - a_{\mathrm{c,wet}}^2 - a_{\mathrm{c,wet}}} - \frac{1}{a_{\mathrm{c,wet}}} \\ \mathrm{c}_{\mathrm{dry}} &= \frac{-a_{\mathrm{c,dry}} - 1}{a_{\mathrm{c,dry}}^2 * c - a_{\mathrm{c,dry}}^2 - a_{\mathrm{c,dry}}} - \frac{1}{a_{\mathrm{c,dry}}} \end{aligned} \tag{5.12}$$

with $a_{\mathrm{c,wet}} < 0$ and $a_{\mathrm{c,dry}} > 0$.

The standard deviation on the other hand becomes 0 when the amount of mean

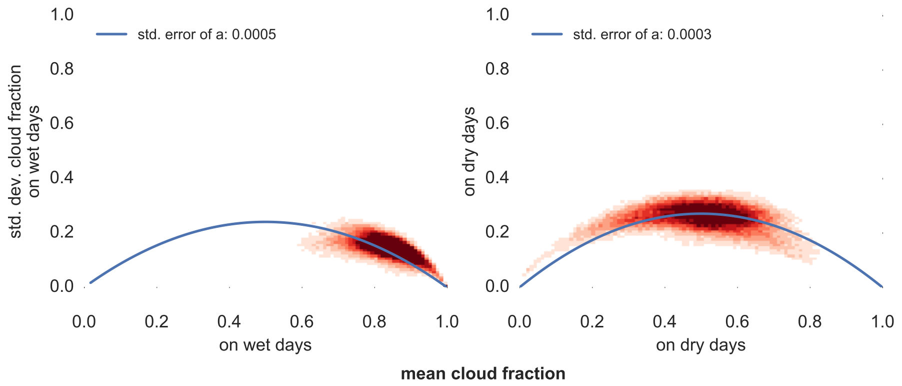

FIGURE 5.9: Correlation of standard deviation of the cloud fraction on wet and dry days to the corresponding monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 5.2.

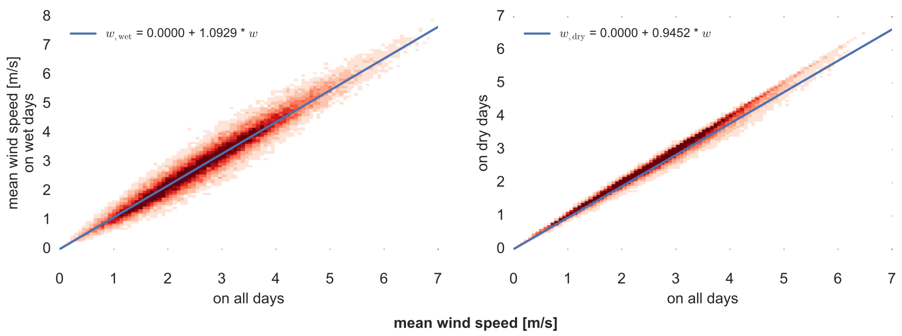

FIGURE 5.10: Correlation of wind speed on wet and dry days to the monthly mean.The y-axes show the mean cloud fraction on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 5.1.

monthly cloud fraction reaches the outer limits 0 and 1. Hence we have an concave parabola with the formula

$$\begin{aligned} \sigma_{c,\text{wet}} &= a_{c,\text{wet}}^2 \cdot c_{\text{wet}} \cdot (1 - c_{\text{wet}}) \\ \sigma_{c,\text{dry}} &= a_{c,\text{dry}}^2 \cdot c_{\text{dry}} \cdot (1 - c_{\text{dry}}) \end{aligned} \tag{5.13}$$

with $a_{c,\text{wet}}, a_{c,\text{dry}} \geq 0$. Results of the fits can be seen in figure 5.8, 5.9 and table 5.2.

Wind speed

The parameterization of wind speed is based upon the same equations (5.10) and (5.11) as temperature and shown in the figures 5.10 and 5.11. Following Parlange and Katz, 2000 we additionally use a square root transformation (see line 19 in the model algorithm 1).

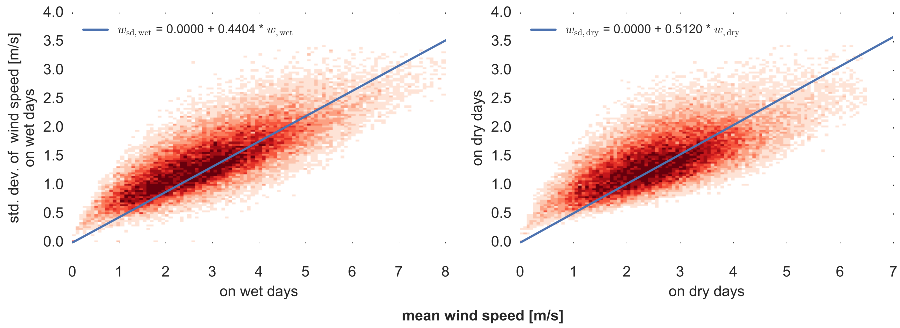

FIGURE 5.11: Correlation of standard deviation of the wind speed on wet and dry days to the corresponding monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 5.1.

Cross correlation

Following Richardson, 1981 we use cross correlation to calculate the residuals and, following Parlange and Katz, 2000, we use a square root transformation for the wind speed.

From the data in the EECRA database for the cloud parameterization (see figure 5.1) we get the

$$A = \begin{pmatrix} 0.913 & 0.033 & -0.021 & 0.001 \\ 0.489 & 0.137 & -0.073 & -0.046 \\ -0.002 & -0.046 & 0.592 & 0.026 \\ 0.011 & -0.044 & -0.019 & 0.667 \end{pmatrix} \qquad B = \begin{pmatrix} 0.362 & 0. & 0. & 0. \\ 0.114 & 0.803 & 0. & 0. \\ 0.145 & -0.061 & 0.783 & 0. \\ 0.081 & -0.016 & 0.066 & 0.737 \end{pmatrix} \tag{5.14}$$

where the columns and rows correspond to min. and max. temperature, cloud fraction and square root of wind speed.

The above matrices (5.14) where calculated via

$$A = M_1 M_0^{-1} \qquad BB^T = M_0 - M_1 M_0^{-1} M_1^T \tag{5.15}$$

from the lag-0 and lag-1 covariance matrices M_0 and M_1 with

$$M_0 = \begin{pmatrix} 1. & 0.572 & 0.025 & 0.032 \\ 0.572 & 1. & -0.101 & -0.045 \\ 0.025 & -0.101 & 1. & 0.127 \\ 0.032 & -0.045 & 0.127 & 1. \end{pmatrix} \qquad M_1 = \begin{pmatrix} 0.932 & 0.557 & -0.001 & 0.025 \\ 0.564 & 0.426 & -0.08 & -0.046 \\ -0.012 & -0.108 & 0.6 & 0.104 \\ 0.006 & -0.066 & 0.071 & 0.667 \end{pmatrix}. \tag{5.16}$$

5.2.3 Evaluation

To evaluate our model, we merged the EECRA dataset into the GHCN data. For each station in the EECRA database we identified the closest GHCN station within 1km. Doing so, resulted in about 1200 stations. We compared the simulated years against the observed years and hence took only the complete years per station. This resulted in about 500 evaluation stations, shown in figure 5.1.

We calculated the monthly input for our weather generator and compared the daily simulation from our model to the original daily data. Since we cannot expect

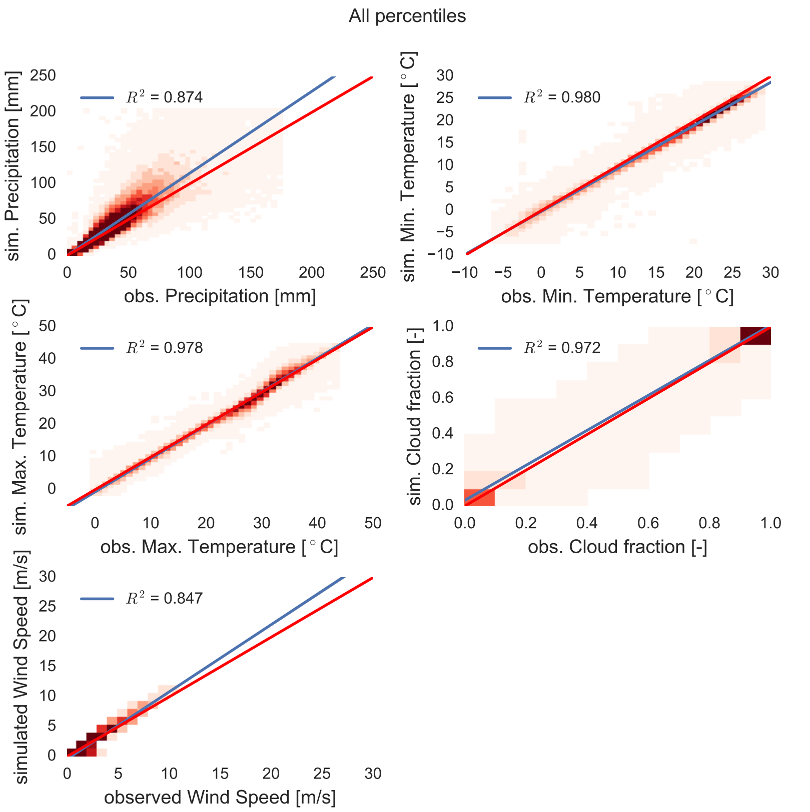

FIGURE 5.12: QQ-plots for all variables with all quantiles (1, 5, 10, 25, 50, 75, 90, 95 and 99) for $\mu = \ldots$ mm, $\xi = \ldots$. The blue lines are linear regression from simulation to observation. The red line shows the ideal fit (the identity line). Blue shaded areas represent the 95% confidence interval. Plots for wind speed and minimum temperature used the bias correction from subsection 5.2.4.

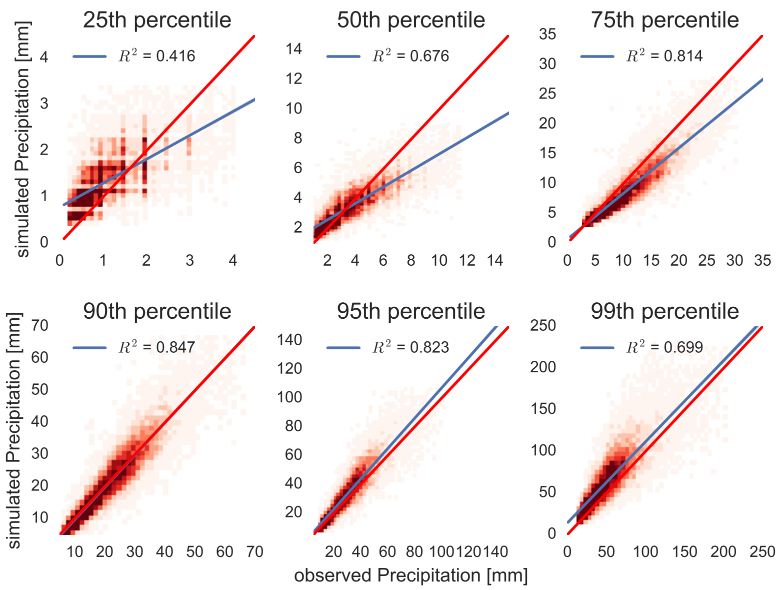

FIGURE 5.13: QQ-plot for different quantiles for precipitation for $\mu = 15\,\text{mm}$, $\xi = 0.08303$. The blue lines are linear regression from simulation to observation. The red line shows the ideal fit (the identity line). Blue shaded areas represent the 95% confidence interval.

from the weather generator to represent the exact timing during the month, we are restricted to compare the two distributions.

Figure 5.12 shows the comparison of the simulated versus the observed quantiles. For temperature, wind and cloud fraction, the model does an excellent job of downscaling monthly input to daily resolution. Also precipitation looks good when mangling all the quantiles, however, a closer look into figure 5.13 shows that the higher precipitation percentiles are well captured using the hybrid Gamma-GP distribution, the lower percentiles however show worse results. The same holds for the wind speed (not shown here). The lower values of the two variables, however, are very close to the precision of the observation (0.1 mm for precipitation and 0.1 m/s for wind speed). Secondly, they are ecologically not as important as the higher percentiles.[1]

In table 5.3 we also compare the simulated versus the observed frequencies. For very light rain (<=1mm), light rain (1-10mm), heavy rain (10-20mm) and very heavy rain (>20mm). As we can see, our model underestimates the occurence of very light rain events (21.6% instead of 27.0%) and overestimates the light rain events (55% instead of 47%) but performs much better than other climate models (Dai, 2006; Sun et al., 2006), especially when it comes to the heavy rain events.

5.2.4 Bias correction

After evaluating the results of GWGEN for wind speed and minimum temperature for the different quantiles (see previous subsection 5.2.3) we found a strong correlation between the deviation from the simulated and the observed quantile.

For the minimum temperature, we use an empirical distribution correction approach (quantile-mapping, Lafon et al., 2012) with the third order polynomial shown

[1] Note that the plots for wind speed and minimum temperature were bias corrected using the approach in subsection 5.2.4.

TABLE 5.3: Simulated and observed precipitation frequencies for certain ranges. The frequency is defined as the number of precipitation occurences in the specified range, divided by the total number of precipitation occurences.

Precipitation [mm]	Simulated	Observed
(0, 1]	0.216	0.270
(1, 10]	0.550	0.470
(10, 20]	0.115	0.133
(20, ∞)	0.120	0.127

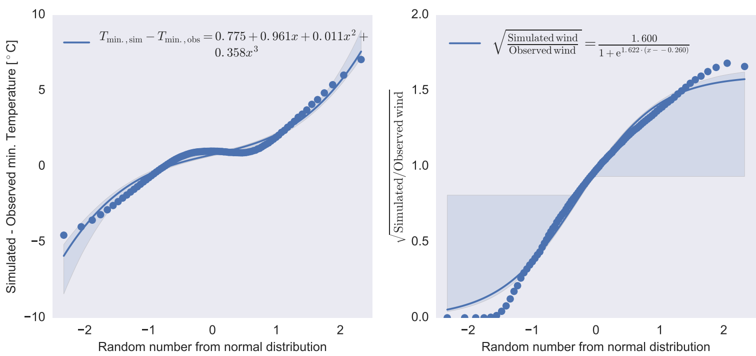

FIGURE 5.14: Basis for the minimum temperature (left) and wind bias correction (right). For the left plot (min. Temperature), each data point corresponds to the difference of a simulated percentile to the observed percentile. For the right plot (wind speed), each data point corresponds to the fraction of simulated to the observed square root of the wind speed for a given percentile. The random number on the x-axis represents the residual value from a normal distribution centered at 0 with standard deviation of unity, as it is used in the cross correlation approach (Richardson, 1981).

in figure 5.14 to transform the simulated minimum temperature data based upon the modeled quantile. The coefficients of this polynomial are calculated a posteriori based upon the bias from simulated to the observed quantile for all percentiles between 1 and 99. The resulting function $f_{T_{\min}}(u)$ is then used in the weather generator to correct the minimum temperature $T_{\min}$ via

$$T_{\min} = T_{\min,biased} - f_{T_{\min}}(u) \tag{5.17}$$

where $u \in \mathbb{R}$ corresponds to the number from the normal distribution drawn for the cross correlation (section 5.2.2 and Richardson, 1981).

Furthermore, we found that the fraction of simulated to observed wind speed for a given percentile follows strongly a logistic function or a third order polynomial (figure 5.14). Hence, as we did for minimum temperature, we use a quantile mapping approach and a posteriori correlate the square root of the fraction of simulated to observed wind speed with the percentile in the normal distribution. The resulting function $f(u)$, is then used inside the weather generator to correct the the wind speed w via

$$w = w_{\text{biased}}\, f(u). \tag{5.18}$$

.

5.2.5 Sensitivity analysis

Our hybrid Gamma-GP distribution has two parameters, the GP shape and the threshold parameter, which could not be related in a sufficiently sophisticated way to any of the simulated variables. Hence, we determined the parameters using a sensitivity analysis.

We chose two methods: the first one is the direct comparison of the quantiles (see previous section), the second one is a Kolmogorov-Smirnov (KS) test that evaluates whether two data samples come from significantly different distributions. Our criteria where

1. The R^2 correlation coefficient between simulated and observed quantiles
2. The fraction $\frac{\text{simulated precipitation}}{\text{observed precipitation}}$ from the slopes in 5.13 and it's deviation from unity
3. the fraction of simulated (station specific) years that are significantly different from the observation
4. The mean of the above values

We tried two different approaches for the threshold: firstly, a fixed crossover point, secondly a quantile based crossover point. For the latter, the model chooses to use the GP distribution if the quantile of the drawn random number is above a certain quantile. This introduces a flexible crossover point in our hybrid distribution which, however, did not improve the results significantly. Hence we decided to only show the results of the fixed crossover point.

The values of the crossover point for our sensitivity analysis were 2, 2.5, 3, 4 and from 5 to 100 in steps of 5. Furthermore we varied the GP shape parameter from 0 to 3 in steps of 0.1 (810 experiments in total). The results of this sensitivity analysis are shown in the supplementary material, figure 5.15.

In general we found that the three criteria 1, 2 and 3 could not be optimized all together at the same time. The R^2 is best for high thresholds and low GP shape parameters, the slope is best for intermediate threshold and a high GP shape and the KS statistic is best for low threshold and intermediate GP shape parameters.

However, R^2 did not vary that much (from 0.68 to 0.74) and from a visual evaluation of the corresponding quantile plots we saw that the higher quantiles (>90) were much better represented for a better KS result. Hence we chose to follow the KS test criteria, which is also the strictest of our evaluation methods but again compared the different quantile plots to get good results for the higher quantiles. Finally, we chose a threshold of 5 mm and a GP shape parameter of 1.5. For this setting, 83.6% of the simulated years do not show a significant difference compared to the observation, the mean R^2 of the plots in figure 5.13 is 0.70 and the mean deviation of the slope from unity is 0.18 and for the upper quantiles (90 to 100), 0.1.

Nevertheless, in total the results seem to be fairly independent of the two parameters since even the amount of years without significant differences vary from 70% to only 86%. It is however better than the gamma distribution alone with only 76.2% of station years not differing significantly and a slope deviation from unity for the upper quantiles of 0.26.

5.3 Model description

The parameterization described in section 5.2.2 is incorporated into a stand-alone model. It requires additional total monthly precipitation, the number of wet days, mean minimum and maximum temperature, mean cloud fraction and wind speed as input. The output are the same variables with daily resolution. This section summarizes the basic workflow in the model which is also shown schematically in algorithm 1.

The first approximation of the daily variables comes from smoothing the monthly time series using the algorithm described in Rymes and Myers, 2001.

For precipitation we then first use the markov chain approach from section 5.2.2 to decide the wet/dry state of the day. If the day is a wet day, we calculate the gamma parameters using the equations (5.7) and (5.8). The resulting distribution allows us to draw a random number, the precipitation amount of the currently simulated day. If we are above the threshold μ, we draw a second random number from the GP distribution parameterized via equation (5.9) and the chosen GP shape.

The next step modifies the means of temperature, wind speed and cloud fraction depending on the wet/dry state of the day (lines 10 and 14 in algorithm 1). After that, we use the cross-correlation approach described in Richardson, 1981 and lines 17 - 19 and calculate the other variables. Finally we use the quantile-based bias correction described in section 5.2.4 to correct the simulated wind speed and minimum temperature.

We restrict the weather generator to reproduce the exact number of wet days (± 1) as the input and to be within a 5% range of the total monthly precipitation (with a maximum allowed deviation of 0.5 mm). Hence, if the program cannot produce these results, the procedure described above is repeated (see line 3).

5.4 Applications and limitations

GWGEN is designed for downscaling monthly precipitation, minimum and maximum temperature, wind speed and cloud fraction to a daily resolution. Through the extensive amount of data used for the parameterization, GWGEN is applicable on the entire globe without the need for spatial information. However one should be

Algorithm 1 Basic workflow of GWGEN

Require: monthly precipitation P_{in} [mm], cloud cover fraction c_{in}, minimum ($T_{\text{min,in}}$ [°C]) and maximum ($T_{\text{max,in}}$ [°C]) temperature, wind speed w_{in} [m/s], number of wet days n_{in}

Output: daily P_i [mm/d], c_i, T_i [°C], w_i [m/s] and the wet/dry state $s_i \in \{0, 1\}$

for month m in *input* **do**
  smooth the monthly data using Rymes and Myers, 2001
  while $\left|\sum_{d_i \in m} P_i - P_{\text{in}}\right| > \min\left(5\% \cdot P_{\text{in}}, 0.5mm\right)$ or $|n_{\text{sim}} - n_{\text{in}}| > 2$ **do**
    for day d_i in m **do**
      Calculate p_{11}, p_{101}, p_{001} after equations (5.1) - (5.3) using n {Precipitation occurence after D. S. Wilks, 1999a}
      Use the Markov chain to determine whether d_i is wet ($s_i = 1$) or dry ($s_i = 0$)
      if $s_i = 1$ **then**
        Calculate θ, α and σ via eq. (5.7)-(5.9) {Precipitation amount after Neykov et al., 2014}
        Draw a random number P_i from the Gamma-GP distribution, eq. (5.6)
        Set $T_{\text{min},i} = T_{\text{min,wet}}, T_{\text{max},i} = T_{\text{max,wet}}, c_i = c_{\text{wet}}, w_i = w_{\text{wet}}$ from eq. (5.10) and (5.12) and tables 5.1, 5.2
        Set $\sigma_{T_{\text{min}},i} = \sigma_{T_{\text{min}},\text{wet}}, \sigma_{T_{\text{max}},i} = \sigma_{T_{\text{max}},\text{wet}}, \sigma_{w,i} = \sigma_{w,\text{wet}}, \sigma_{c,i} = \sigma_{c,\text{wet}}$ from eq. (5.11) and (5.13) and tables 5.1, 5.2
      else
        Set $P_i = 0\,\text{mm/d}$
        Set $T_{\text{min},i} = T_{\text{min,dry}}, T_{\text{max},i} = T_{\text{max,dry}}, c_i = c_{\text{dry}}, w_i = w_{\text{dry}}$ from eq. (5.10) and (5.12) and tables 5.1, 5.2
        Set $\sigma_{T_{\text{min}},i} = \sigma_{T_{\text{min}},\text{dry}}, \sigma_{T_{\text{max}},i} = \sigma_{T_{\text{max}},\text{dry}}, \sigma_{w,i} = \sigma_{w,\text{dry}}, \sigma_{c,i} = \sigma_{c,\text{dry}}$ from eq. (5.11) and (5.13) and tables 5.1, 5.2
      end if
      Draw 4 normally distributed random numbers $\epsilon \in \mathbb{R}^4$ {Cross correlation after Richardson, 1981}
      Set the residuals $\chi_i = \begin{pmatrix}\chi_{T_{\text{min}}} & \chi_{T_{\text{max}}} & \chi_c & \chi_w\end{pmatrix} = A\chi_{i-1} + B\epsilon \in \mathbb{R}^4$ with A and B from eq. (5.14)
      Calculate daily variables via
      $$T_{\text{min},i} = \chi_{T_{\text{min}}} \cdot \sigma_{T_{\text{min}},i} + T_{\text{min},i} \qquad c_i = \chi_c \cdot \sigma_{c,i} + c_i$$
      $$T_{\text{max},i} = \chi_{T_{\text{max}}} \cdot \sigma_{T_{\text{max}},i} + T_{\text{max},i} \qquad w_i = \left(\chi_w \cdot \sqrt{\sigma_{w,i}} + \sqrt{w_i}\right)^2$$
      Apply bias correction for T_{min} (eq. (5.17)) and w (eq. (5.18))
    end for
  end while
end for

careful when using it for small regional experiments (e.g. on the catchment scale) where other, more specialized and regional weather generators might be better.

5.5 Outlook and conclusions

Our weather generator model uses a global dataset of precipitation, temperature and cloudiness to downscale monthly to daily data. Our results show that some simple relationships are applicable on the whole globe and can be used to within a reasonable accuracy for the estimation of the weather distribution throughout a month.

Compared to the Gamma distribution the applied hybrid Gamma-GP distribution improves the model results due to its heavy tail. Further improvements might come through correlating the GP shape and location parameter to the region and seasonality (Maraun et al., 2009; Rust et al., 2009).

5.A Supplementary material

5.A.1 Sensitivity analysis

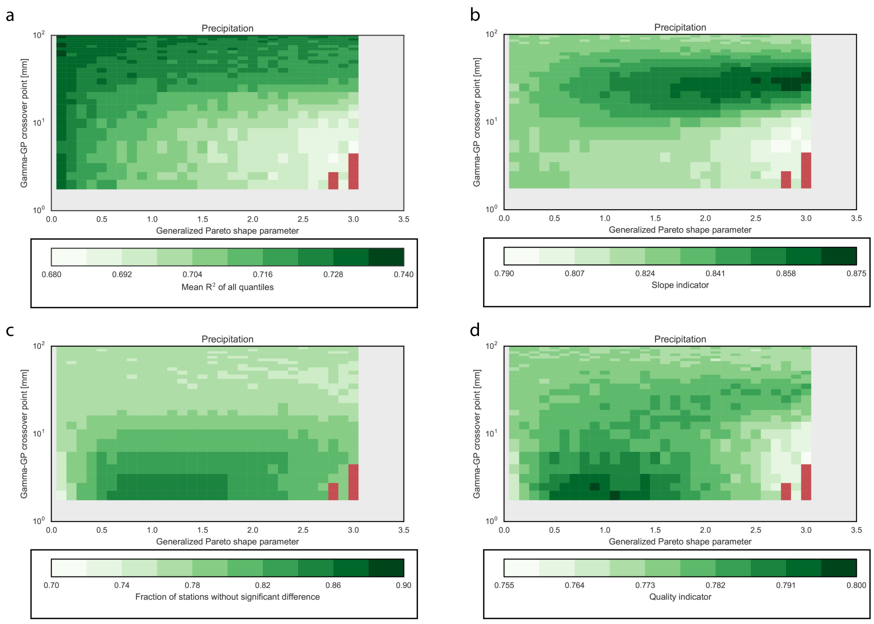

FIGURE 5.15: Results of the sensitivity analysis for the (a) correlation coefficient R^2, (b) deviation from a slope of unity, (c) the fraction of significant different station years, (d) the mean of (a) - (c). For the plots in (a) and (b) the mean of the 25th, 50th, 75th, 90th, 95th and 99th percentiles are used. In general, 1 (dark green) is best, 0 (white) is worst. The dark red fields indicate experiments that failed because of a too low threshold and too high GP shape parameter. Note also the logarithmic scale on the y-axis.

Author contributions. JOK conceived the model and analyses, wrote the prototype code and performed preliminary analyses, PS developed and documented the final version of the code (including parameterization), performed all of the final analyses, and okay, something else, created the graphical output. Both authors contributed to the writing of the manuscript

Acknowledgements. This work was supported by the European Research Council (COEVOLVE, 313797) and the Swiss National Science Foundation (ACACIA, CR10I2_146314). We thank Shawn Koppenhoefer for assistance compiling and querying the weather databases and Alexis Berne and Grégoire Mariéthoz for helpful suggestions on the analyses. We are grateful to NOAA NCDC and the University of Washington for providing free of charge the GHCN-Daily and EECRA databases, respectively.

References

Bhatt, Samir, Peter W. Gething, Oliver J. Brady, Jane P. Messina, Andrew W. Farlow, Catherine L. Moyes, John M. Drake, John S. Brownstein, Anne G. Hoen, Osman Sankoh, Monica F. Myers, Dylan B. George, Thomas Jaenisch, G. R. William Wint, Cameron P. Simmons, Thomas W. Scott, Jeremy J. Farrar, and Simon I. Hay (2013).

"The global distribution and burden of dengue". In: *Nature* 496.7446, pp. 504–507. DOI: 10.1038/nature12060. URL: http://dx.doi.org/10.1038/nature12060.

Bondeau, Alberte, Pascalle C. Smith, SÖNke Zaehle, Sibyll Schaphoff, Wolfgang Lucht, Wolfgang Cramer, Dieter Gerten, Hermann Lotze-Campen, Christoph MÜLler, Markus Reichstein, and Benjamin Smith (2007). "Modelling the role of agriculture for the 20th century global terrestrial carbon balance". In: *Global Change Biol.* 13.3, pp. 679–706. ISSN: 1354-1013 1365-2486. DOI: 10.1111/j.1365-2486.2006.01305.x.

Dai, Aiguo (2006). "Precipitation Characteristics in Eighteen Coupled Climate Models". In: *J. Climate* 19.18, pp. 4605–4630. DOI: 10.1175/jcli3884.1. URL: http://dx.doi.org/10.1175/JCLI3884.1.

Elith, Jane, Catherine H. Graham, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, Falk Huettmann, John R. Leathwick, Anthony Lehmann, Jin Li, Lucia G. Lohmann, Bette A. Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. M. Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Scachetti-Pereira, Robert E. Schapire, Jorge Soberón, Stephen Williams, Mary S. Wisz, and Niklaus E. Zimmermann (2006). "Novel methods improve prediction of species' distributions from occurrence data". In: *Ecography* 29.2, pp. 129–151. ISSN: 1600-0587. DOI: 10.1111/j.2006.0906-7590.04596.x. URL: http://dx.doi.org/10.1111/j.2006.0906-7590.04596.x.

Friend, A. D. (1998). "Parameterisation of a global daily weather generator for terrestrial ecosystem modelling". In: *Ecol. Modell.* 109.2, pp. 121–140. ISSN: 0304-3800. DOI: Doi10.1016/S0304-3800(98)00036-2.

Furrer, Eva M. and Richard W. Katz (2008). "Improving the simulation of extreme precipitation events by stochastic weather generators". In: *Water Resour. Res.* 44.12, n/a–n/a. ISSN: 00431397. DOI: 10.1029/2008wr007316.

Geng, S., F. W. T. P. Devries, and I. Supit (1986). "A Simple Method for Generating Daily Rainfall Data". In: *Agric. For. Meteorol.* 36.4, pp. 363–376. ISSN: 0168-1923. DOI: 10.1016/0168-1923(86)90014-6.

Geng, Shu and J. S. Auburn (1987). "Weather simulation models based on summaries of long-term data". In: *Weather and Rice: Proceedings of the international workshop on the Impact of Weather Parameters on Growth and Yield of Rice, 7-10 Apr 1986*. Ed. by International Rice Research Institute. Los Baños, Philippines: International Rice Research Institute, pp. 237–254.

Gerten, Dieter, Sibyll Schaphoff, Uwe Haberlandt, Wolfgang Lucht, and Stephen Sitch (2004). "Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model". In: *J. Hydrol.* 286.1-4, pp. 249–270. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2003.09.029.

Guenther, A., C. N. Hewitt, D. Erickson, R. Fall, C. Geron, T. Graedel, P. Harley, L. Klinger, M. Lerdau, W. A. Mckay, T. Pierce, B. Scholes, R. Steinbrecher, R. Tallamraju, J. Taylor, and P. Zimmerman (1995). "A Global-Model of Natural Volatile Organic-Compound Emissions". In: *Journal of Geophysical Research-Atmospheres* 100.D5, pp. 8873–8892. ISSN: 2169-897x. DOI: Doi10.1029/94jd02950.

Hahn, C.J. and S.G. Warren (1999). "Extended Edited Synoptic Cloud Reports from Ships and Land Stations Over the Globe, 1952-1996 (with Ship data updated through 2008)". In: DOI: 10.3334/CDIAC/cli.ndp026c. URL: http://dx.doi.org/10.3334/CDIAC/cli.ndp026c.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister (2014). "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset". In: *Int. J. Climatol.* 34.3, pp. 623–642. ISSN: 08998418. DOI: 10.1002/joc.3711.

Haxeltine, A. and I. C. Prentice (1996). "BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types". In: *Global Biogeochem. Cycles* 10.4, pp. 693–709. ISSN: 0886-6236. DOI: Doi10.1029/96gb02344.

Haxeltine, Alex, I. Colin Prentice, and Ian David Creswell (1996). "A coupled carbon and water flux model to predict vegetation structure". In: *J. Veg. Sci.* 7.5, pp. 651–666. ISSN: 1654-1103. DOI: 10.2307/3236377. URL: http://dx.doi.org/10.2307/3236377.

Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis (2005). "Very high resolution interpolated climate surfaces for global land areas". In: *Int. J. Climatol.* 25.15, pp. 1965–1978. ISSN: 0899-8418 1097-0088. DOI: 10.1002/joc.1276.

Kaplan, J. O., N. H. Bigelow, I. C. Prentice, S. P. Harrison, P. J. Bartlein, T. R. Christensen, W. Cramer, N. V. Matveyeva, A. D. McGuire, D. F. Murray, V. Y. Razzhivin, B. Smith, D. A. Walker, P. M. Anderson, A. A. Andreev, L. B. Brubaker, M. E. Edwards, and A. V. Lozhkin (2003). "Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections". In: *Journal of Geophysical Research-Atmospheres* 108.D19. ISSN: 2169-897x. DOI: Artn817110.1029/2002jd002559.

Krinner, G., Nicolas Viovy, Nathalie de Noblet-Ducoudré, Jérôme Ogée, Jan Polcher, Pierre Friedlingstein, Philippe Ciais, Stephen Sitch, and I. Colin Prentice (2005). "A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system". In: *Global Biogeochem. Cycles* 19.1, n/a–n/a. ISSN: 08866236. DOI: 10.1029/2003gb002199.

Kucharik, Christopher J., Jonathan A. Foley, Christine Delire, Veronica A. Fisher, Michael T. Coe, John D. Lenters, Christine Young-Molling, Navin Ramankutty, John M. Norman, and Stith T. Gower (2000). "Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance, and vegetation structure". In: *Global Biogeochem. Cycles* 14.3, pp. 795–825. ISSN: 1944-9224. DOI: 10.1029/1999GB001138. URL: http://dx.doi.org/10.1029/1999GB001138.

Lafon, Thomas, Simon Dadson, Gwen Buys, and Christel Prudhomme (2012). "Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods". In: *Int. J. Climatol.* 33.6, pp. 1367–1381. DOI: 10.1002/joc.3518. URL: http://dx.doi.org/10.1002/joc.3518.

Leemans, Rik and Wolfgang P Cramer (1991). "The IIASA database for mean monthly values of temperature, precipitation, and cloudiness on a global terrestrial grid". In: *International Institute for Applied Systems Analysis, Laxenburg, Austria.*

Lieth, Helmut (1975). "Modeling the Primary Productivity of the World". In: *Primary Productivity of the Biosphere*. Ed. by Helmut Lieth and Robert H. Whittaker. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 237–263. ISBN: 978-3-642-80913-2. DOI: 10.1007/978-3-642-80913-2_12. URL: http://dx.doi.org/10.1007/978-3-642-80913-2_12.

Maraun, D., H. W. Rust, and T. J. Osborn (2009). "The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics". In: *Int. J. Climatol.* 29.12, pp. 1731–1744. DOI: 10.1002/joc.1811. URL: http://dx.doi.org/10.1002/joc.1811.

Menne, Matthew J., Imke Durre, Bryant Korzeniewski, Shelley McNeill, Kristy Thomas, Xungang Yin, Steven Anthony, Ron Ray, Russell S. Vose, Byron E. Gleason, and Tamara G. Houston (2012a). "Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.22". In: DOI: 10.7289/V5D21VHZ. URL: http://dx.doi.org/10.7289/V5D21VHZ.

Menne, Matthew J., Imke Durre, Russell S. Vose, Byron E. Gleason, and Tamara G. Houston (2012b). "An Overview of the Global Historical Climatology Network-Daily Database". In: *J. Atmos. Oceanic Technol.* 29.7, pp. 897–910. DOI: 10.1175/jtech-d-11-00103.1. URL: http://dx.doi.org/10.1175/JTECH-D-11-00103.1.

Mitchell, Timothy D. and Philip D. Jones (2005). "An improved method of constructing a database of monthly climate observations and associated high-resolution grids". In: *Int. J. Climatol.* 25.6, pp. 693–712. ISSN: 0899-8418 1097-0088. DOI: 10.1002/joc.1181.

New, M., M. Hulme, and P. Jones (1999). "Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology". In: *J. Climate* 12.3, pp. 829–856. ISSN: 0894-8755. DOI: Doi10.1175/1520-0442(1999)012<0829:Rtcstc>2.0.Co;2.

— (2000). "Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate". In: *J. Climate* 13.13, pp. 2217–2238. ISSN: 0894-8755. DOI: Doi10.1175/1520-0442(2000)013<2217:Rtcstc>2.0.Co;2.

New, M., D. Lister, M. Hulme, and I. Makin (2002). "A high-resolution data set of surface climate over global land areas". In: *Climate Research* 21.1, pp. 1–25. ISSN: 0936-577x. DOI: DOI10.3354/cr021001.

Neykov, N. M., P. N. Neytchev, and W. Zucchini (2014). "Stochastic daily precipitation model with a heavy-tailed component". In: *Natural Hazards and Earth System Science* 14.9, pp. 2321–2335. ISSN: 1684-9981. DOI: 10.5194/nhess-14-2321-2014.

Parlange, Marc B. and Richard W. Katz (2000). "An Extended Version of the Richardson Model for Simulating Daily Weather Variables". In: *J. Appl. Meteorol.* 39.5, pp. 610–622. DOI: 10.1175/1520-0450-39.5.610. URL: http://dx.doi.org/10.1175/1520-0450-39.5.610.

Pfeiffer, M., A. Spessa, and J. O. Kaplan (2013). "A model for global biomass burning in preindustrial time: LPJ-LMfire (v1.0)". In: *Geosci. Model Dev.* 6.3, pp. 643–685. ISSN: 1991-959x. DOI: 10.5194/gmd-6-643-2013. URL: http://www.geosci-model-dev.net/6/643/2013/gmd-6-643-2013.pdf.

Prentice, I. C., W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud, and A. M. Solomon (1992). "A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate". In: *J. Biogeogr.* 19.2, pp. 117–134. ISSN: 0305-0270. DOI: Doi10.2307/2845499.

Prentice, I.C. (1989). *Developing a Global Vegetation Dynamics Model: Results of an IIASA Summer Workshop*. IIASA Research Report. IIASA, Laxenburg, Austria. URL: http://pure.iiasa.ac.at/3223/.

Richardson, C. W. (1981). "Stochastic simulation of daily precipitation, temperature, and solar radiation". In: *Water Resour. Res.* 17.1, pp. 182–190. ISSN: 00431397. DOI: 10.1029/WR017i001p00182.

Rust, H. W., D. Maraun, and T. J. Osborn (2009). "Modelling seasonality in extreme precipitation". In: *The European Physical Journal Special Topics* 174.1, pp. 99–111. DOI: 10.1140/epjst/e2009-01093-7. URL: http://dx.doi.org/10.1140/epjst/e2009-01093-7.

Rymes, M.D. and D.R. Myers (2001). "Mean preserving algorithm for smoothly interpolating averaged data". In: *Sol. Energy* 71.4, pp. 225–231. DOI: 10.1016/s0038-092x(01)00052-4. URL: http://dx.doi.org/10.1016/S0038-092X(01)00052-4.

Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. O. Kaplan, S. Levis, W. Lucht, M. T. Sykes, K. Thonicke, and S. Venevsky (2003). "Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model". In: *Global Change Biol.* 9.2, pp. 161–185.

ISSN: 1354-1013. DOI: 10.1046/j.1365-2486.2003.00569.x. URL: http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2486.2003.00569.x/abstract.

Stephens, Graeme L., Tristan L'Ecuyer, Richard Forbes, Andrew Gettelmen, Jean-Christophe Golaz, Alejandro Bodas-Salcedo, Kentaroh Suzuki, Philip Gabriel, and John Haynes (2010). "Dreary state of precipitation in global models". In: *Journal of Geophysical Research: Atmospheres* 115.D24, n/a–n/a. ISSN: 2156-2202. DOI: 10.1029/2010JD014532. URL: http://dx.doi.org/10.1029/2010JD014532.

Sun, Ying, Susan Solomon, Aiguo Dai, and Robert W. Portmann (2006). "How Often Does It Rain?" In: *J. Climate* 19.6, pp. 916–934. DOI: 10.1175/jcli3672.1. URL: http://dx.doi.org/10.1175/JCLI3672.1.

Viovy, N. and P. Ciais (2016). Online Database. URL: http://dods.extra.cea.fr/data/p529viov/cruncep.

Walter, H and H Lieth (1967). "Climate diagram world atlas". In: *VEB Gustav Fischer Verlag Jena, Jena*.

Wei, Y., S. Liu, D. N. Huntzinger, A. M. Michalak, N. Viovy, W. M. Post, C. R. Schwalm, K. Schaefer, A. R. Jacobson, C. Lu, H. Tian, D. M. Ricciuto, R. B. Cook, J. Mao, and X. Shi (2014). "The North American Carbon Program Multi-scale Synthesis and Terrestrial Model Intercomparison Project – Part 2: Environmental driver data". In: *Geosci. Model Dev.* 7.6, pp. 2875–2893. ISSN: 1991-9603. DOI: 10.5194/gmd-7-2875-2014. URL: http://www.geosci-model-dev.net/7/2875/2014/.

Wilks, D. S. (1998). "Multisite generalization of a daily stochastic precipitation generation model". In: *J. Hydrol.* 210.1-4, pp. 178–191. ISSN: 0022-1694. DOI: 10.1016/S0022-1694(98)00186-3.

— (1999a). "Interannual variability and extreme-value characteristics of several stochastic daily precipitation models". In: *Agric. For. Meteorol.* 93.3, pp. 153–169. ISSN: 0168-1923. DOI: 10.1016/S0168-1923(98)00125-7.

— (1999b). "Multisite downscaling of daily precipitation with a stochastic weather generator". In: *Climate Research* 11.2, pp. 125–136. ISSN: 0936-577x. DOI: 10.3354/cr011125.

Wilks, D. S. and R. L. Wilby (1999). "The weather generation game: a review of stochastic weather models". In: *Prog. Phys. Geog.* 23.3, pp. 329–357. ISSN: 0309-1333. DOI: 10.1177/030913339902300302.

Wilks, Daniel S. (2010). "Use of stochastic weathergenerators for precipitation downscaling". In: *Wiley Interdiscip. Rev. Clim. Change* 1.6, pp. 898–907. ISSN: 17577780. DOI: 10.1002/wcc.85.

Woodward, F. Ian, Thomas M. Smith, and William R. Emanuel (1995). "A global land primary productivity and phytogeography model". In: *Global Biogeochem. Cycles* 9.4, pp. 471–490. ISSN: 1944-9224. DOI: 10.1029/95GB02432. URL: http://dx.doi.org/10.1029/95GB02432.

Woolhiser, D. A. and G. G. S. Pegram (1979). "Maximum Likelihood Estimation of Fourier Coefficients to Describe Seasonal-Variations of Parameters in Stochastic Daily Precipitation Models". In: *J. Appl. Meteorol.* 18.1, pp. 34–42. ISSN: 0894-8763. DOI: Doi10.1175/1520-0450(1979)018<0034:Mleofc>2.0.Co;2.

Woolhiser, D. A. and J. Roldan (1982). "Stochastic Daily Precipitation Models: 2. A Comparison of Distributions of Amounts". In: *Water Resour. Res.* 18.5, pp. 1461–1468. ISSN: 0043-1397. DOI: DOI10.1029/WR018i005p01461.

Woolhiser, D. A. and José Roldán (1986). "Seasonal and Regional Variability of Parameters for Stochastic Daily Precipitation Models: South Dakota, U.S.A". In: *Water Resour. Res.* 22.6, pp. 965–978. ISSN: 00431397. DOI: 10.1029/WR022i006p00965.

Chapter 6

A model analysis on the stability of northern hemispheric teleconnections

Need to write chapter 6

- still too write…

Chapter 7

Conclusions

- New tools that have been developed
- Quality standards of the tools
- Further development and potential usage

Appendices

Todo list

Appendix A

Computing climate-smart urban land use with the Integrated Urban Complexity model (IUCm 1.0)

Need to write chapter A

Appendix B

Publications and Conference contributions

B.0.1 Peer-reviewed

Cremades, R. and P. 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/.

Sommer, Philipp, 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.

Weitzel, Nils, Sebastian Wagner, Jesper Sjolte, Marlene Klockmann, Oliver Bothe, Heather Andres, Lev Tarasov, Kira Rehfeld, Eduardo Zorita, Martin Widmann, Philipp S. Sommer, Gerd Schädler, Patrick Ludwig, Florian Kapp, Lukas Jonkers, Javier García-Pintado, Florian Fuhrmann, Andrew Dolman, Anne Dallmeyer, and Tim Brücher (2018). “Diving into the past – A paleo data-model comparison workshop on the Late Glacial and Holocene”. In: *Bulletin of the American Meteorological Society*. DOI: 10.1175/bams-d-18-0169.1.

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.

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.

B.0.2 Conference contributions

Sommer, P. S., B. A. S. Davis, and M. Chevalier (2019a). “Github and Open Research Data; an example using the Eurasian Modern Pollen Database”. In: *EGU General Assembly Conference Abstracts*. Vol. 21. EGU General Assembly Conference Abstracts, p. 5669. URL: https://meetingorganizer.copernicus.org/EGU2019/EGU2019-5669.pdf.

Sommer, Philipp S., Basil A. S. Davis, Manuel Chevalier, Jian Ni, and John Tipton (2019b). “The HORNET project: applying ’big data’ to reconstruct the climate of the Northern Hemisphere during the Holocene”. In: *20th Congress of the International Union for Quaternary Research (INQUA)*. International Union for Quaternary Research. URL: https://app.oxfordabstracts.com/events/574/program-app/submission/94623.

Sommer, P. S. (2018). “Psyplot: Interactive data analysis and visualization with Python”. In: *EGU General Assembly Conference Abstracts*. Vol. 20. EGU General Assembly Conference Abstracts. Provided by the SAO/NASA Astrophysics Data System, p. 4701. URL: http://adsabs.harvard.edu/abs/2018EGUGA..20.4701S.

Sommer, P. S., B. A. S. Davis, and M. Chevalier (2018a). “STRADITIZE: An open-source program for digitizing pollen diagrams and other types of stratigraphic data”. In: *EGU General Assembly Conference Abstracts*. Vol. 20. EGU General Assembly Conference Abstracts. Provided by the SAO/NASA Astrophysics Data System, p. 4433. URL: http://adsabs.harvard.edu/abs/2018EGUGA..20.4433S.

Sommer, Philipp S., Manuel Chevalier, and Basil A. S. Davis (2018b). “STRADITIZE: An open-source program for digitizing pollen diagrams and other types of stratigraphic data”. In: *AFQUA - The African Quaternary*. Nairobi (Kenya): AFQUA. URL: https://afquacongress.wixsite.com/afqua2018.

Sommer, P. and J. Kaplan (2017). “Quantitative Modeling of Human-Environment Interactions in Preindustrial Time”. In: *PAGES OSM 2017, Abstract Book*, pp. 129–129.

Sommer, P. (2016). “Psyplot: Visualizing rectangular and triangular Climate Model Data with Python”. In: *EGU General Assembly Conference Abstracts*. Vol. 18. EGU General Assembly Conference Abstracts. Provided by the SAO/NASA Astrophysics Data System, p. 18185. URL: http://adsabs.harvard.edu/abs/2016EGUGA..1818185S.

Sommer, P. and J. Kaplan (2016a). “Fundamental statistical relationships between monthly and daily meteorological variables: Temporal downscaling of weather based on a global observational dataset”. In: *EGU General Assembly Conference Abstracts*. Vol. 18. EGU General Assembly Conference Abstracts. Provided by the SAO/NASA Astrophysics Data System, EPSC2016–18183. URL: http://adsabs.harvard.edu/abs/2016EGUGA..1818183S.

— (2016b). “Fundamental statistical relationships between monthly and daily meteorological variables: Temporal downscaling of weather based on a global observational dataset”. In: *Workshop on Stochastic Weather Generators*. Vannes (France): University of Bretagne Sud. URL: https://www.lebesgue.fr/content/sem2016-climate-program.

Appendix C

New Software Tools - An Overview

This section mainly contains the latest version of the package, a short summary and an information table about where to find everything (Documentation, source code, etc.)

Need to write chapter C

C.1 Main packages

- psyplot
  - psy-simple
  - psy-maps
  - psy-reg
  - psyplot-gui
  - psy-strat
- straditize
- gwgen
- iucm
- EMPD
  - EMPD-admin
  - EMPD-viewer
  - EMPD-data
- POLNET
  - POLNET-viewer
  - POLNET-data

C.2 Other packages

- docrep
- sphinx-nbexamples
- model-organization
- funcargparse
- autodocsumm

Bibliography

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

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.

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.

Grimm, Eric C. (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.

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.

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.

Marcott, S. A., J. D. Shakun, P. U. Clark, and A. C. Mix (2013). "A Reconstruction of Regional and Global Temperature for the Past 11,300 Years". In: *Science* 339.6124, pp. 1198–1201. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1228026. URL: http://www.ncbi.nlm.nih.gov/pubmed/23471405.

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.

Mauri, A., B. A. S. Davis, P. M. Collins, and J. O. Kaplan (2014). "The influence of atmospheric circulation on the mid-Holocene climate of Europe: a data-model comparison". In: *Clim. Past* 10.5, pp. 1925–1938. ISSN: 1814-9324. DOI: 10.5194/

cp-10-1925-2014. URL: http://www.clim-past.net/10/1925/2014/cp-10-1925-2014.pdf.

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.

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.

Schneider, Birgit (2012). "Climate model simulation visualization from a visual studies perspective". In: *Wiley Interdisciplinary Reviews: Climate Change* 3.2, pp. 185–193. DOI: 10.1002/wcc.162.

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.

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.

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.