EGU2020-14019, updated on 28 Jan 2022
https://doi.org/10.5194/egusphere-egu2020-14019
EGU General Assembly 2020

Challenges in creating and exemplary applications of two cross-repository data compilations on sedimentary pollen and permafrost soil temperature

Mareike Wieczorek[1], Birgit Heim[1], Thomas Böhmer[1], Nadine Gebhardt[2], Annett Bartsch[3], and Ulrike Herzschuh[1,4,5]

[1]Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Polar Terrestrial Environmental Systems, Germany (mareike.wieczorek@awi.de)

[2]Faculty of Mathematics and Geography, Catholic University of Eichstätt-Ingolstadt, 85072 Eichstätt, Germany

[3]b.geos GmbH, 2100 Korneuburg, Austria

[4]Institute of Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany

[5]Institute of Environmental Sciences and Geography, University of Potsdam, 14476 Potsdam, Germany

Large scale analyses of climatic or ecological data are important to understand complex relationships. Often, such data are available in open repositories or national measurement programmes, others are only made available via the responsible researcher. However, merging data from various sources is often not straightforward, due to issues with the data itself or the metadata. Nevertheless, the application of such compilations offers various possibilities. In our working group, two large-scale compilations are currently constructed and applied. The Northern Hemispheric Pollen Compilation consists of data from NEOTOMA, European Pollen Database (EPD), PANGAEA and various authors. With the help of this compilation, we reconstruct climate and vegetation of large spatial and temporal scales. The circumpolar soil temperature dataset consist of data from the Global Terrestrial Network for Permafrost (GTN-P), Roshydromet, PANGAEA, Nordicana D and the National Science Foundation (NSF) Arctic Data Center. In its first version, the compilation has already been successfully applied to validate the ESA CCI Permafrost soil temperature map.

The various sources of errors and problems will be shown by the two compilations of (i) sedimentary pollen data and (ii) soil temperature data. The most general problem and error source are wrong or inaccurate coordinates. These errors arise out of coordinates provided with two decimals only, wrong conversion of DMS to decimal format, wrong coordinates etc. For most analyses, the most exact geographic position is a prerequisite, as e.g. lake size is an important parameter when reconstructing vegetation out of sedimentary pollen data. Sedimentary pollen records not located in a lake according to their given location thus need manual reposition according to the main researcher of a dataset or satellite maps. Further challenges concerning the pollen dataset pose various naming conventions or variable resolution in time. Furthermore, taxonomic resolution varies between datasets, making homogenization necessary.

But also for the soil temperature dataset, extensive checks were necessary, as even quality checked data comprise erroneous values. Furthermore, measured depths vary between datasets. For easy comparisons of soil temperature simulations against data, standardized depths were extracted. In a future step, interpolations between measured depths will help the end-users to extract the exactly needed depths and a compilation of available metadata on e.g. surrounding vegetation and borehole stratigraphy shall be provided.

All compilations will be made available on public repositories.