Data quality control - an essential prerequisite of science

Thursday 4 April 2024 15:40 (15 minutes)

Any model can only be as good as the input and validation data. Models of various scales ranging from single soil columns to the complete earth system need time series of input data. Large scale models often rely on global products derived from satellite data or weather model re-analysis. However, classical on-site measured time series still serve as input of such compiled products and are also used to validate the model output. Furthermore, local effects often cannot be adequately represented in products of limited spatial resolution. In the Arctic, the required local reference data are scarce and raw data series often include corrupted or missing values. Our long-term observatories were installed more than 25 years ago and therefore provide very valuable data on soil and climate characteristics at Bayelva (Svalbard) and Samoylov (Lena River Delta, Siberia). In addition to the tremendous effort to maintain the sites and calibrate the instruments, it is essential to perform a thorough quality control of each measured variable. Our routine does not only cover removing values outside the physical limits, but also includes manual checks for plausibility both within a single time series and in comparison to similar variables. We use soil temperature data of the Samoylov observatory, Lena River Delta, Siberia, to highlight the effect of using raw data, quality-controlled data, and gap filling on estimated soil warming trends. In particular, we consider the timing of missing values in our analysis. We show that removing implausible values and a proper handling of data gaps are essential prerequisites of any data analysis including both physical and statistical modelling efforts.

Primary author: GRÜNBERG, Inge (AWI)

Co-authors: GROENKE, Brian (Alfred-Wegener-Institute); MIESNER, Frederieke (Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research); Prof. BOIKE, Julia (Alfred-Wegener-Institute)

Presenter: GRÜNBERG, Inge (AWI)

Session Classification: Session 2: Data Quality and HPC for Data Science

Track Classification: Data Quality and HPC for Data Science