EGU23-11095, updated on 14 Apr 2024
https://doi.org/10.5194/egusphere-egu23-11095
EGU General Assembly 2023

Towards better identification of dominant controls in Earth system data

Robert Reinecke[1], Francesca Pianosi[2], and Thorsten Wagener[1]
[1]Institue of Environmental Science and Geography, University of Potsdam, Potsdam, Germany
[2]Department of Civil Engineering, University of Bristol, Bristol, UK

We are in a state of simultaneous exuberance and starvation of Earth system data. Model ensembles of increasing complexity provide petabytes of output, while remote sensing products offer terabytes of new data every day. On the other hand, we lack data on processes that are more challenging to observe, like groundwater recharge, or have data heavily impacted by un-quantified anthropogenic change. These problems leave us with highly imbalanced datasets.

Our ability to produce and collect mountains of data contrasts with our progress in improving scientific process understanding. How can we harness simulated and observed data alike to enhance our knowledge and test scientific hypotheses about process relationships given poorly known uncertainties? Our contribution discusses methods to approach this problem while being agnostic to the data source (model or observation). We introduce a new strategy that allows us to interrogate given datasets to identify correlational and possibly causal relationships between the variables included. We test the method on an ensemble of complex global hydrological models and observations to demonstrate its usefulness and limitations, i.e., from the ISIMIP experiments. We show that our approach can provide powerful insights into dominant process controls while scaling with large amounts of data.