

EGU21-8584, updated on 03 Nov 2021

<https://doi.org/10.5194/egusphere-egu21-8584>

EGU General Assembly 2021

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



A Benchmark for Bivariate Causal Discovery Methods

Christoph Käding and Jakob Runge

German Aerospace Center (DLR), Institute of Data Science, Climate Informatics Group, Germany (christoph.kaeding@dlr.de)

The Earth's climate is a highly complex and dynamical system. To better understand and robustly predict it, knowledge about its underlying dynamics and causal dependency structure is required. Since controlled experiments are infeasible in the climate system, observational data-driven approaches are needed. Observational causal inference is a very active research topic and a plethora of methods have been proposed. Each of these approaches comes with inherent strengths, weaknesses, and assumptions about the data generating process as well as further constraints.

In this work, we focus on the fundamental case of bivariate causal discovery, i.e., given two data samples X and Y the task is to detect whether X causes Y or Y causes X . We present a large-scale benchmark that represents combinations of various characteristics of data-generating processes and sample sizes. By comparing most of the current state-of-the-art methods, we aim to shed light onto the real-world performance of evaluated methods. Since we employ synthetic data, we are able to precisely control the data characteristics and can unveil the behavior of methods when their underlying assumptions are met or violated. Further, we give a comparison on a set of real-world data with known causal relations to complete our evaluation.