

Title: Interactive and Reproducible Data Processing for Forest Science: Addressing the Efficiency Gap

Authors: Richard L. Peters¹, Alexander G. Hurley²

¹Tree Growth and Wood Physiology, TUM School of Life Sciences, Technical University of Munich, Hans-Carl-v.-Carlowitz-Platz 2, 85354 Freising, Germany

²Expertenrat für Klimafragen (ERK) | German Council of Experts on Climate Change
Seydelstr. 15 10117 Berlin

100 word Abstract:

Forest research is increasingly data-rich, requiring robust tools to address technical and logistical challenges in processing “big data.” Meeting best practices in data exploration, quality control, and reproducibility can be difficult when conventional methods, like diagnostic visualizations and tables, are impractical due to time constraints. To address these challenges, we developed datacleanr, an R package that provides an interactive interface for streamlined data exploration, quality control, and flexible processing of diverse tabular data types, including time series and georeferenced data. By enhancing workflows, datacleanr aims to increase confidence in research outputs and improve reproducibility across data-intensive forest research.