# CAMELS-DE: Benchmark dataset for hydrology – significance, current status and outlook

**Alexander Dolich**[1], Pia Ebeling[2], Michael Stölzle[3], Jens Kiesel[4], Jonas Götte[5], Björn Guse[4,6], Sibylle Haßler[1], Mirko Mälicke[1], Larisa Tarasova[2], Ingo Heidbüchel[2], Corina Hauffe[7], Hannes Müller-Thomy[8], and Ralf Loritz[1]

[1]Karlsruhe Institute of Technology, Institute of Water and River Basin Management, Chair of Hydrology, Germany (alexander.dolich@kit.edu)

[2]Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

[3]University of Freiburg, Faculty of Environment and Natural Resources, Freiburg, Germany

[4]Christian-Albrechts-University Kiel, Department of Hydrology and Water Resources Management, Kiel, Germany

[5]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

[6]GFZ German Research Centre for Geosciences, Potsdam, Germany

[7]Technical University Dresden, Chair of Hydrology, Dresden, Germany

[8]Technical University Braunschweig, Department of Hydrology and River Basin Management, Braunschweig, Germany

CAMELS datasets are recognized in the hydrological community as consistent and comprehensive benchmark datasets for hydrological and meteorological analyses. CAMELS stands for "Catchment Attributes and MEteorology for Large-sample Studies". CAMELS datasets link landscape and catchment attributes (e.g. land use, geology, soil properties), hydrological time series (e.g. water level, discharge) and meteorological time series (e.g. precipitation, air temperature) in a large number of catchment areas. They clearly indicate the uncertainties and processing of individual variables and thus enable the comparison of models and data in different landscapes, but also contribute to the general understanding of hydrological processes across landscapes. This is crucial for assessing the consequences of the climate crisis and improves the basis for water resource management decisions. Although CAMELS datasets are intensively used in other countries, such a dataset is still lacking for Germany.

This contribution highlights the crucial importance of consistent and easily accessible benchmark datasets for hydrological research and education. We discuss both the challenges faced so far in compiling the dataset and the future ambitions of the project. In addition, an overview is given of the scope of the first version of the CAMELS-DE data set, which will include around 2,000 measuring stations with daily time series of discharge and water level with an average length of nearly 50 years in mainly small and medium-sized catchments. Also included are the landscape and catchment attributes as well as meteorological time series. A key focus is on the easy availability and straightforward import of data into programming environments. We discuss how such benchmark datasets not only increase efficiency in the use of environmental data, but also play a key role in ensuring the reproducibility of research results. Especially in the age of machine

learning learning, they form an indispensable basis for modern, data-driven hydrology. By integrating CAMELS-DE into the research landscape, we want to emphasize that data publications and benchmark datasets are much more than a by-product of a doctoral thesis, but rather the basis and key to modern environmental science.