



## Disinformative data in large-scale hydrological modelling

Anna Kauffeldt (1), Sven Halldin (1), Allan Rodhe (1), Chong-Yu Xu (2,1), Ida Westerberg (3,1)

(1) Dept. of Earth Sciences, Uppsala University, Uppsala, Sweden (anna.kauffeldt@geo.uu.se), (2) Dept. of Geosciences, University of Oslo, Oslo, Norway, (3) IVL Swedish Environmental Research Institute, Stockholm, Sweden

Large-scale hydrological modelling has become an important tool for the study of global and regional water resources, climate impacts, and water-resources management. However, modelling efforts over large spatial domains are fraught with problems of data scarcity, uncertainties and inconsistencies between forcing and evaluation data. Model-independent methods to screen and analyse data for such problems are needed. This study aims at identifying two types of data inconsistencies in global datasets using a pre-modelling analysis, inconsistencies that can be disinformative for subsequent modelling. Firstly, four hydrographic datasets were examined in terms of how well basin areas were represented in the flow networks. It was found that most basins could be well represented in both gridded basin delineations and polygon-based ones, but some basins exhibited large area discrepancies between hydrographic datasets and archived basin areas. Secondly, the consistency between climate data (precipitation and potential evaporation) and discharge data was examined for the possibility of water-balance closure. It was found that basins exhibiting too high runoff coefficients were abundant in areas where precipitation data were likely affected by snow undercatch, and that the occurrence of basins exhibiting losses exceeding the energy limit were strongly dependent on the potential-evaporation data, both in terms of numbers and geographical distribution. These results emphasise the need for pre-modelling data analysis to identify dataset inconsistencies as an important first step in any large-scale study. Applying data-screening methods before modelling increases our chances to draw robust conclusions from subsequent model simulations.