

EGU21-15768

<https://doi.org/10.5194/egusphere-egu21-15768>

EGU General Assembly 2021

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



## An empirical study on the GEOTop hydrological model optimal estimation and uncertainty reduction using supercomputers

Giacomo Bertoldi<sup>1</sup>, **Stefano Campanella**<sup>1,2</sup>, Emanuele Cordano<sup>3</sup>, and Alberto Sartori<sup>4</sup>

<sup>1</sup>Eurac research, Institute for the Alpine Environment, Bolzano, Italy

<sup>2</sup>National Institute of Oceanography and Applied Geophysics – OGS, Trieste, Italy

<sup>3</sup>Rendena 100 s.r.l., Italy

<sup>4</sup>SISSA, Trieste, Italy

Proper characterization of uncertainty remains a major research and operational challenge in Earth and Environmental Systems Models (EESMs). In fact, model calibration is often more an art than a science: one must make several discretionary choices, guided more by his own experience and intuition than by the scientific method. In practice, this means that the result of calibration (CA) could be suboptimal. One of the challenges of CA is the large number of parameters involved in EESM, which hence are usually selected with the help of a preliminary sensitivity analysis (SA). Finally, the computational burden of EESMs models and the large volume of the search space make SA and CA very time-consuming processes.

This work applies a modern HPC approach to optimize a complex, over parameterized hydrological model, improving the computational efficiency of SA/CA. We apply the derivative-free optimization algorithms implemented in the Facebook Nevergrad Python library (Rapin and Teytaud, 2018) on a HPC cluster, thanks to the Dask framework (Dask Development Team, 2016).

The approach has been applied to the GEOTop hydrological model (Rigon et al., 2006; Endrizzi et al., 2014) to predict the time evolution of variables as soil water content and evapotranspiration for several mountain agricultural sites in South Tyrol with different elevation, land cover (pasture, meadow, orchard), soil types.

We performed simulations on one-dimensional domains, where the model solves the energy and water budget equations in a column of soil and neglects the lateral water fluxes. Even neglecting the distribution of parameters across layers of soil, considering a homogeneous column, one has tens of parameters, controlling soil and vegetation properties, where only a few of them are experimentally available.

Because the interpretation of global SA could be difficult or misleading and the number of model evaluations needed by SA is comparable with CA, we employed the following strategy. We performed CA using a full set of continuous parameters and SA after CA, using the samples collected during CA, to interpret the results. However, given the above-mentioned computational challenges, this strategy is possible only using HPC resources. For this reason, we focused on the

computational aspects of calibration from an HPC perspective and examined the scaling of these algorithms and their implementation up to 1024 cores on a cluster. Other issues that we had to address were the complex shape of the search space and robustness of CA and SA against model convergence failure.

HPC techniques allow to calibrate models with a high number of parameters within a reasonable computing time and exploring the parameters space properly. This is particularly important with noisy, multimodal objective functions. In our case, HPC was essential to determine the parameters controlling the water retention curve, which is highly not linear. The developed framework, which is published and freely available on GitHub, shows also how libraries and tools used within the machine learning community could be useful and easily adapted to EESMs CA.