

EGU21-6312

<https://doi.org/10.5194/egusphere-egu21-6312>

EGU General Assembly 2021

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Learning from mistakes - Assessing the performance and uncertainty in process-based models

Benjamin Roesky^{1,2}, Moritz Feigl^{2,3}, Mathew Herrnegger³, Karsten Schulz³, and Masaki Hayashi²

¹BGC Engineering Inc., Toronto, Canada

²Department of Geoscience, University of Calgary, Calgary, Canada

³Institute for Hydrology and Water Management, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

Typical applications of process- or physically-based models aim to gain a better process understanding of certain natural phenomena or to estimate the impact of changes in the examined system caused by anthropogenic influences, such as land-use or climate change. To adequately represent the physical system, it is necessary to include all (essential) processes in the applied model and to observe relevant inputs in the field. However, model errors, i.e. deviations between observed and simulated values, can still occur. Other than large systematic observation errors, simplified, misrepresented or missing processes are potential sources of errors. This study presents a set of methods and a proposed workflow for analyzing errors of process-based models as a basis for relating them to process representations.

The evaluated approach consists of three steps: (1) prediction of model errors with a machine learning (ML) model using data that might be associated with model errors (e.g., model input data), (2) derivation of variable importance (i.e. contribution of each input variable to prediction) for each predicted model error using SHapley Additive exPlanations (SHAP), (3) clustering of SHAP values of all predicted errors to derive groups with similar error generation characteristics. By analyzing these groups of different error/variable association, hypotheses on error generation and corresponding processes can be formulated. This analysis framework can ultimately lead to improving hydrologic understanding and prediction.

The framework is applied to the physically-based stream water temperature model HFLUX in a case study for modelling an alpine stream in the Canadian Rocky Mountains. Initial statistical tests show a significant association of model errors with available meteorological and hydrological variables. By using these variables as input features, the applied ML model is able to predict model residuals. Clustering of SHAP values results in four distinct error groups that can be related to tree shading, sensible and latent heat flux and longwave radiation emitted by trees.

Model errors are rarely random and often contain valuable information. Assessing model error associations is ultimately a way of enhancing trust in implemented processes and of providing information on potential areas of improvement to the model.