

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

Moritz Feig¹, Benjamin Roesky², Mathew Herrnegger¹, Karsten Schulz¹, and Masaki Hayashi³

¹University of Natural Resources and Life Sciences Vienna

²BGC Engineering Inc

³University of Calgary

September 24, 2021

Abstract

Typical applications of process- or physically-based models aim to gain a better process understanding or provide the basis for a decision-making process. To adequately represent the physical system, models should include all essential processes. However, model errors can still occur. Other than large systematic observation errors, simplified, misrepresented, inadequately parametrized 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: (i) training a machine learning (ml) error-model using the input data of the process-based model and other available variables, (ii) estimation of local explanations (i.e., contributions of each variable to a individual prediction) for each predicted model error using SHapley Additive exPlanations (SHAP) in combination with principal component analysis, (iii) 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. That can ultimately lead to improvements in process understanding and prediction. The approach is applied to a process-based stream water temperature model HFLUX in a case study for modelling an alpine stream in the Canadian Rocky Mountains. By using available meteorological and hydrological variables as inputs, the applied ml model is able to predict model residuals. Clustering of SHAP values results in three distinct error groups that are mainly related to shading and vegetation emitted longwave radiation. 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.

Hosted file

`Feigl_et_al_manuscript.docx` available at <https://authorea.com/users/435642/articles/538521-learning-from-mistakes-assessing-the-performance-and-uncertainty-in-process-based-models>

Hosted file

`Feigl_et_al_table01.docx` available at <https://authorea.com/users/435642/articles/538521-learning-from-mistakes-assessing-the-performance-and-uncertainty-in-process-based-models>











