

EGU21-2192

<https://doi.org/10.5194/egusphere-egu21-2192>

EGU General Assembly 2021

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



## Better Than Just Average: The Many Faces of Bayesian Model Weighting Methods and What They Tell Us about Multi-Model Use

Marvin Höge<sup>1</sup>, Anneli Guthke<sup>2</sup>, and Wolfgang Nowak<sup>2</sup>

<sup>1</sup>Eawag (SIAM), Dübendorf, Switzerland (marvin.hoege@eawag.ch)

<sup>2</sup>University of Stuttgart, Stuttgart, Germany

In environmental modelling it is usually the case that multiple models are plausible, e.g. for predicting a certain quantity of interest. Using model rating methods, we typically want to elicit a single best one or the optimal average of these models. However, often, such methods are not properly applied which can lead to false conclusions.

At the examples of three different Bayesian approaches to model selection or averaging (namely 1. Bayesian Model Selection and Averaging (BMS/BMA), 2. Pseudo-BMS/BMA and 3. Bayesian Stacking), we show how very similarly looking methods pursue vastly different goals and lead to deviating results for model selection or averaging.

All three yield a weighted average of predictive distributions. Yet, only Bayesian Stacking has the goal of averaging for improved predictions in the sense of an actual (optimal) model combination. The other approaches pursue the quest of finding a single best model as the ultimate goal - yet, on different premises - and use model averaging only as a preliminary stage to prevent rash model choice.

We want to foster their proper use by, first, clarifying their theoretical background and, second, contrasting their behaviors in an applied groundwater modelling task. Third, we show how the insights gained from these Bayesian methods are transferrable to other (also non-Bayesian) model rating methods and we pose general conclusions about multi-model usage based on model weighting.