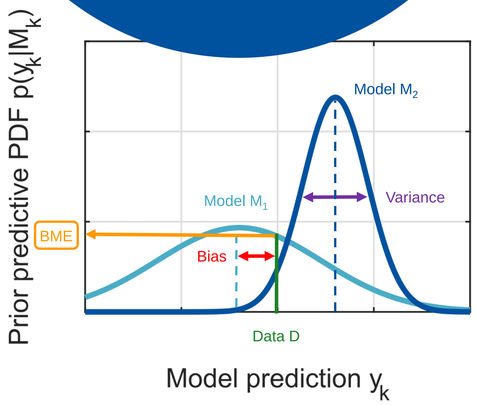
Bayesian Model Evidence as a Model Evaluation Metric

Anneli Guthke[1], Marvin Höge[2], Wolfgang Nowak[1]

[1] Institute for Modelling Hydraulic and Environmental Systems (LS³)/SimTech, University of Stuttgart, Germany. ✉ Anneli.guthke@iws.uni-stuttgart.de. [2] Center for Applied Geoscience, University of Tübingen, Germany.

We thank the DFG for financial support within the IRTG "Integrated Hydrosystem Modelling" (IRTG 1829, Tübingen), and within the Cluster of Excellence in Simulation Technology (EXC 310/1, Stuttgart).

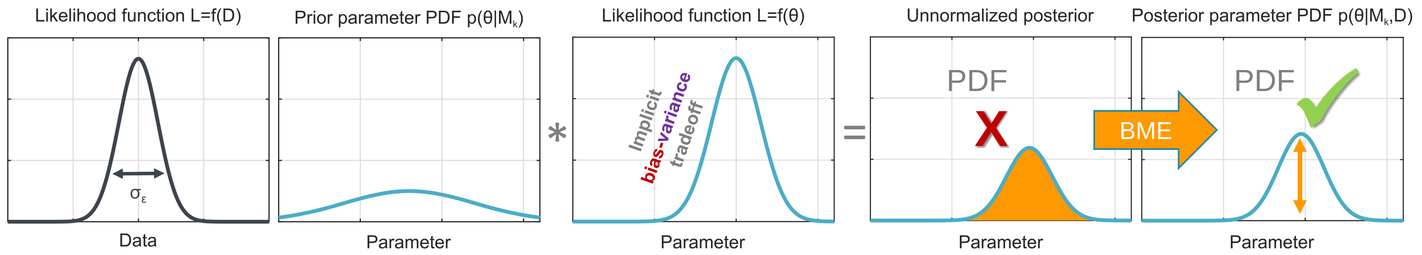
How to assess model quality in the face of uncertainty?



Uncertainty (input, parameters) and errors (model structure, measurements) trigger a **predictive distribution** instead of a deterministic forecast. How to rate model quality aspects **bias** and **variance**?

Bayesian model evidence

[Raftery, 1995]: $p(\boldsymbol{D}|M_k) = \int_{\Theta_k} p(\boldsymbol{D}|M_k,\boldsymbol{\theta}_k)\, p(\boldsymbol{\theta}_k|M_k) d\theta_k = \frac{p(\boldsymbol{D}|M_k,\boldsymbol{\theta}_k)\, p(\boldsymbol{\theta}_k|M_k)}{p(\boldsymbol{\theta}_k|M_k,\boldsymbol{D})}$



Bayesian model evidence…

Predictive log-score [Good, 1952] for independent predictions:

$$PLS = \log p(\boldsymbol{D}|M_k) = \sum_{j=1}^{N_d} \log p(D_j|M_k)$$

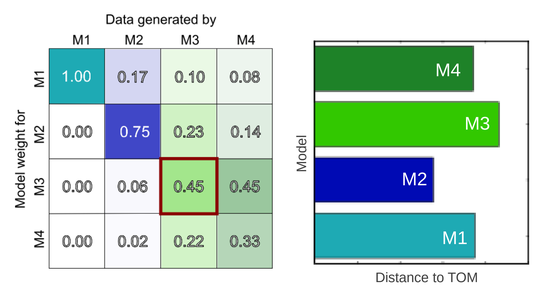
Information criteria (e.g., AIC [Akaike, 1973], BIC [Schwarz, 1978], KIC [Neuman, 2003]) for linear models and multi-Gaussian distributions [Schöniger et al., 2014]:

$$IC \approx -2 \log p(\boldsymbol{D}|M_k)$$

… related metrics…

- Investigating model quality for competing models with **Bayes factors** [Kass & Raftery, 1995]
- Weighting and combining models: **Bayesian model selection and averaging** [Hoeting et al., 1999]
- Ranking models under limited data: **Model justifiability analysis** [Schöniger et al., 2015a]
- Ranking models under noisy data: Comparison with **theoretically optimal model** TOM [Schöniger et al., 2015b]

$$BF = \frac{p(\boldsymbol{D}|M_1)\, P(M_1)}{p(\boldsymbol{D}|M_2)\, P(M_2)}$$



… and further analysis options to dig deeper!

References:

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, Int Symp on Info Theory. // Good, I. J. (1952), Rational decisions, J Roy Stat Soc B Met. // Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999), Bayesian model averaging: A tutorial, Stat Sci. // Kass, R., & Raftery, A. (1995). Bayes factors. J Am Stat Assoc. // Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, SERRA. // Raftery, A. E. (1995), Bayesian model selection in social research, Sociol Methodol. // Schöniger, A., Wöhling, T., Samaniego, L., & Nowak, W. (2014), Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence, WRR. // Schöniger, A., Illman, W.A., Wöhling, T., & Nowak, W. (2015a), Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection, JoH. // Schöniger, A., Wöhling, T., & Nowak, W. (2015b), A statistical concept to assess the uncertainty in Bayesian model weights and its impact on model ranking, WRR. // Schwarz, G. (1978), Estimating the dimension of a model, Ann Stat.