Supporting Information for "Extending GLUE with Multilevel Methods to Accelerate Statistical Inversion of Hydrological Models"

M. G. Rudolph[1], T. Wöhling[2,3], T. Wagener[4], A. Hartmann[1]

[1]Institute of Groundwater Management, Technische Universität Dresden, Dresden, Germany

[2]Chair of Hydrology, Institute of Hydrology and Meteorology, Technische Universität Dresden, Dresden, Germany

[3]Lincoln Agritech, Lincoln, New Zealand

[4]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

Contents of this file

Introduction

This supporting information provides additional text and figures describing the results shown and discussed in the main article "Extending GLUE with Multilevel Methods to Accelerate Statistical Inversion of Hydrological Models". Texts S1 and S2 provide additional details on the derivation of MLGLUE. Texts S3 and S4 provide more detailed descriptions of results for the example inverse problems with a rainfall-runoff model and a groundwater flow model, respectively. Figures S1 and S2 illustrate results for the rainfall-runoff modelling example. Figures S3 and S4 illustrate results for the groundwater flow modelling example.

Text S1 - Derivation of MLGLUE, the Wrong Combination of MLMC and GLUE

Assuming that likelihood thresholds are given on each level prior to sampling, a straightforward approach to combining MLMC and GLUE would be to use an MLMC algorithm (e.g., Giles, 2015) directly. Then, only model simulations would be considered that correspond to likelihoods that are above the level-dependent likelihood threshold. With that, as most MLMC samples come from lower levels, posterior parameter samples would mainly be comprised of samples from the posterior distribution corresponding to the coarser-level models. We aim, however, at generating samples that come from the posterior distribution on the finest level. This combination is therefore not purposeful. Otherwise we could directly use the model on level $\ell = 0$ to perform statistical inversion on a single level, which contradicts the actual aim of the methodology.

Text S2 - Derivation of MLGLUE, Level-Dependent Likelihood Thresholds

Using level-dependent likelihood thresholds instead of the highest-level threshold for all levels is motivated by the construction of the MLDA algorithm (Lykkegaard et al., 2023) as well as by the original delayed acceptance MCMC algorithm (Christen & Fox, 2005). In MLDA, different target densities are considered on each level because the likelihood function - seen as a (hyper-) surface in the parameter space - depends on the model used on a corresponding level. In the sense of Bayes' theorem, those densities can be considered to be Bayesian posterior densities. This is an intuitive construction; conside evaluations of the quantity of interest on different levels, made with the same parameter samples, $\{\mathbf{Q}_\ell(\boldsymbol{\theta}^{(i)}), \mathbf{Q}_{\ell+1}(\boldsymbol{\theta}^{(i)})\}_{i=1}^N$, as well as corresponding likelihods $\{\widetilde{\mathcal{L}}_\ell(\boldsymbol{\theta}^{(i)}|\widetilde{\mathbf{Y}}), \widetilde{\mathcal{L}}_{\ell+1}(\boldsymbol{\theta}^{(i)}|\widetilde{\mathbf{Y}})\}_{i=1}^N$:

$$\mathbf{Q}_\ell(\boldsymbol{\theta}_i) \neq \mathbf{Q}_{\ell+1}(\boldsymbol{\theta}_i) \tag{1}$$

$$\Rightarrow \widetilde{\mathcal{L}}_\ell(\boldsymbol{\theta}^{(j)}|\widetilde{\mathbf{Y}}) \neq \widetilde{\mathcal{L}}_{\ell+1}(\boldsymbol{\theta}^{(j)}|\widetilde{\mathbf{Y}}) \tag{2}$$

$$\Rightarrow \widetilde{\mathcal{L}}_{T,\ell} \neq \widetilde{\mathcal{L}}_{T,\ell+1} \tag{3}$$

Therefore, level-dependent likelihood thresholds instead of a single highest-level threshold used on all levels need to be considered to accurately reflect the variations within the hierarchy of models.

Text S3 - Additional Description of Results, Rainfall-Runoff Modelling

With MLDA, a total number of $N_{\ell=L} = 2,000$ samples were computed on the highest level using $n_{chains} = 20$ and a subsampling rate of 5, resulting in a total of $N_{MLDA} = 2,000 \cdot 5^2 \cdot 20 = 1,000,000$ samples from the prior distribution. No samples were burnt from the 20 MLDA chains on the highest level, resulting in $\widehat{R} = 1.0$ for all 5 parameters. Out of the $40,020$ remaining samples (including randomly initialized samples on the highest level), only $8,204$ effective samples could be used (mean effective sample size estimate for the bulk of the posterior). Therefore, a thinning of 5 was applied, resulting in $8,020$ effective samples. With MCMC, a total number of $N_{\ell=L} = 50,000$ samples were computed on the highest level using $n_{chains} = 20$, resulting in a total of $N_{MCMC} = 50,000 \cdot 20 = 1,000,000$ samples from the prior distribution. No samples are burnt from the 20 MCMC chains, resulting in $\widehat{R} = 1.00$ for all parameters. Out of the $1,000,020$ remaining samples (including randomly initialized samples), only $16,353$ effective samples could be used (mean effective sample size estimate for the bulk of the posterior). Therefore, a thinning of 62 was applied, resulting in $16,140$ effective samples.

Text S4 - Additional Description of Results, Groundwater Flow

With MLDA, a total number of $N_{\ell=L} = 250$ samples were computed on the highest level using $n_{chains} = 32$, resulting in a total of $N_{MLDA} = 250 \cdot 5^3 \cdot 32 = 1,000,000$ samples from the prior distribution. No sample was burnt from the 32 MLDA chains, resulting in a mean Gelman-Rubin statistic of $\overline{\widehat{R}} = 1.02$ ($\widehat{R}_{min} = 1.01$, $\widehat{R}_{max} = 1.03$), averaged over all 51 parameters. Out of the $8,032$ remaining samples (including randomly initialized samples on the highest level), only $1,982$ effective samples could be used (mean effective sample size estimate for the bulk of the posterior). Therefore, a thinning of 4 was applied, resulting in $2,008$ effective samples. With MCMC, a total number of $N_{\ell=L} = 31,250$ samples were computed on the highest level using $n_{chains} = 32$ and a subsampling rate of 5, resulting in a total of $N_{MCMC} = 31,250 \cdot 32 = 1,000,000$ samples from the prior distribution. The initial sample was burnt from the 32 MCMC chains, resulting in a mean Gelman-Rubin statistic of $\overline{\widehat{R}} = 1.02$ ($\widehat{R}_{min} = 1.01$, $\widehat{R}_{max} = 1.03$), averaged over all 51 parameters. Out of the $1,000,000$ remaining sample, only $2,080$ effective samples can be used (mean effective sample size estimate for the bulk of the posterior). Therefore, a thinning of 481 was applied, resulting in $2,080$ effective samples.

References

Christen, J. A., & Fox, C. (2005). Markov Chain Monte Carlo Using an Approximation. *Journal of Computational and Graphical Statistics*, *14*(4). doi: 10.1198/106186005X76983

Giles, M. B. (2015, May). Multilevel Monte Carlo methods. *Acta Numerica*, *24*, 259–328. Retrieved 2022-11-24, from https://www.cambridge.org/core/product/identifier/S096249291500001X/type/journal_article doi: 10.1017/S096249291500001X

Lykkegaard, M. B., Dodwell, T. J., Fox, C., Mingas, G., & Scheichl, R. (2023, March). Multilevel Delayed Acceptance MCMC. *SIAM/ASA Journal on Uncertainty Quantification*, *11*(1), 1–30. Retrieved 2023-03-16, from https://epubs.siam.org/doi/10.1137/22M1476770 doi: 10.1137/22M1476770

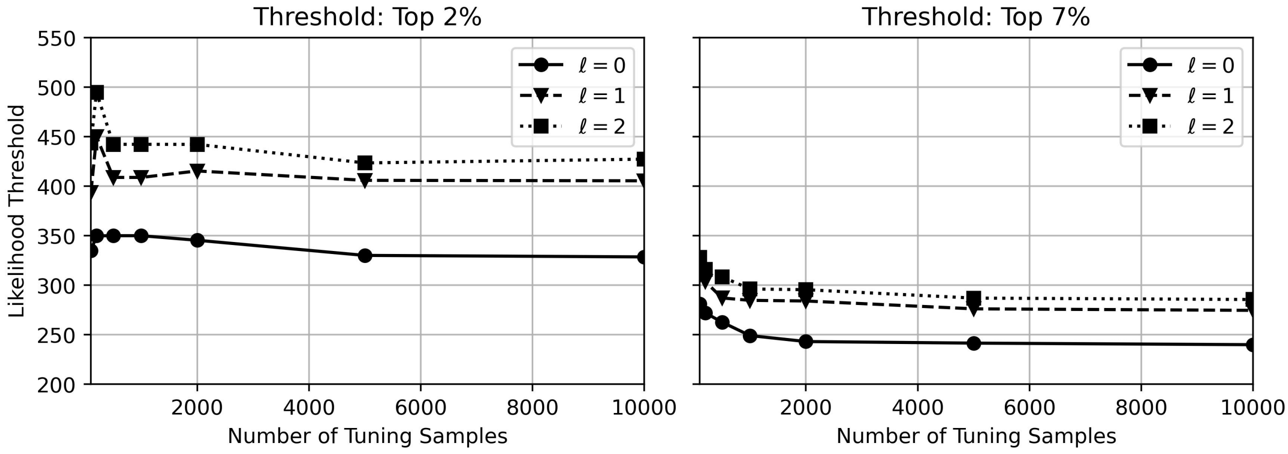

Figure S1. Level-dependent likelihood thresholds for the rainfall-runoff modelling example, estimated with different numbers of tuning samples for threshold settings corresponding to the top 2 % (left) and the top 7 % (right)

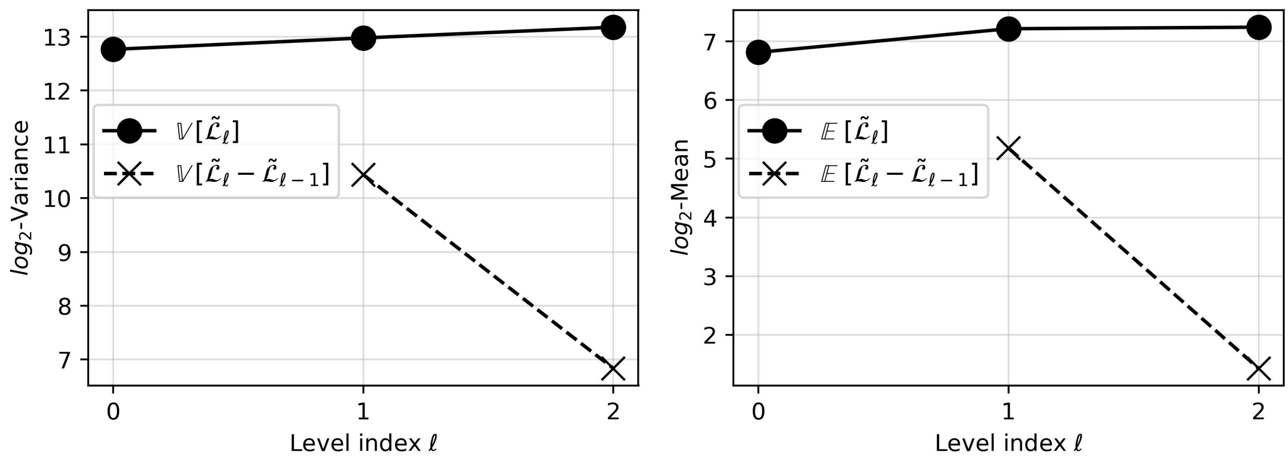

Figure S2. Relations between levels for the linear regression example, using an informal likelihood

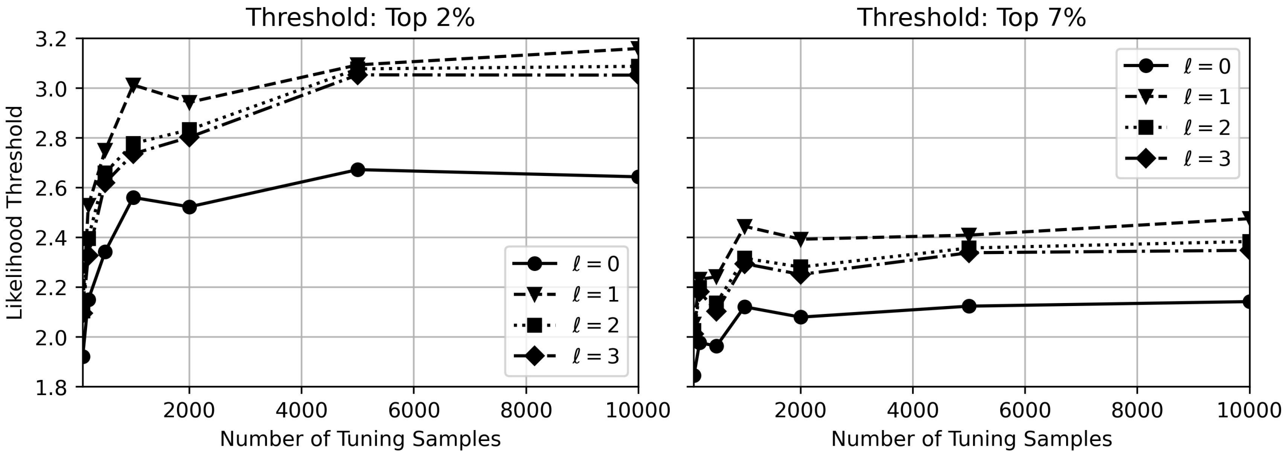

Figure S3. Level-dependent likelihood thresholds for the groundwater flow example, estimated with different numbers of tuning samples for threshold settings corresponding to the top 2 % (left) and the top 7 % (right)

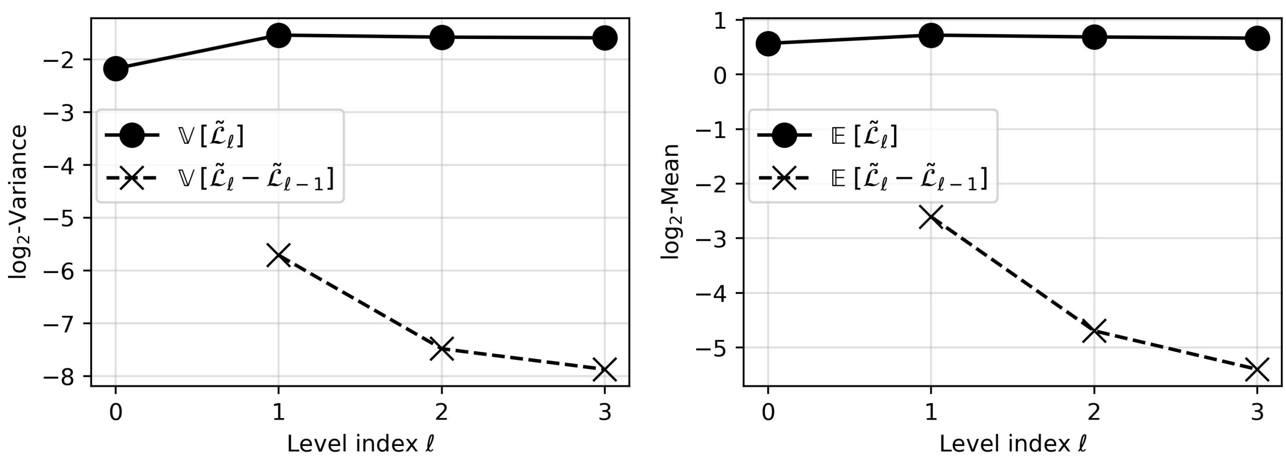

Figure S4. Relations between levels for the groundwater flow example, using an informal definition of the likelihood