Documentation on calibration routines

1 Scoring function

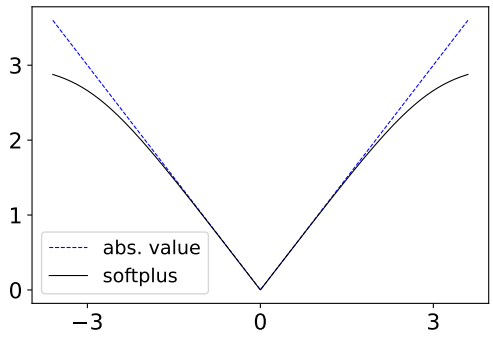

Figure 1: Softplus correction applied to residuals for ADM1

For AM2, the residuals are computed in the log space. To prevent the error from exploding between two predictions with negligible values, but with one some order of magnitude smaller than the other, an offset $\eta = 10^{-8}$ was introduced before computing the log. All in all, the score is computed as

$$S^2(\theta) = \frac{1}{N} \sum_{k,t} \log \left(\frac{\mathrm{AD}_{k,t}(\theta) + \eta}{\mathrm{Obs}_{k,t} + \eta} \right)^2$$

For ADM1, a softplus saturation function is applied to each log residual to ensure the score remains bounded by $S_{\max} = 3$. The corrective function is

$$\mathrm{softplus}(r) = \frac{1 + \exp(-2S_{\max})}{2} * \log \left(\frac{1 + \exp(2S_{\max})}{1 + \exp(2(S_{\max} - |r|))} \right).$$

A plot is given in Figure 1. The score is then defined as

$$S(\theta) = \sqrt{\frac{1}{N} \sum_{k,t} \mathrm{softplus} \left(\log \left(\frac{\mathrm{AD}_{k,t}(\theta) + \eta}{\mathrm{Obs}_{k,t} + \eta} \right) \right)^2}.$$

For both AD models, sets of parameters for which the ODE solver fails are attributed a score of $S_{\max}$.

2 Standard calibration

2.1 Sensitivity analysis

A global sensitivity analysis was conducted in the case of ADM1 datasets to select parameters which were thereafter calibrated. The routine used was based on Morris method [Morris, 1991], and adjusted to account for multidimensional responses. Consistent with the calibration, the analysis was conducted on the logarithm of the parameters. The plausible ranges were inferred from the prior, using the intervals $[\log(\theta_i) - 2\sigma_i, \log(\theta_i) + 2\sigma_i]$. These intervals were further discretized into 8 levels. From this grid of parameter values, chains of parameters (θ^d) are constructed as described in Morris method, such that (θ^d) and (θ^{d+1}) differ only in their d th component, and their d th components are neighbors in the grid. The impact of parameter component d is then assessed as

$$\text{sensitivity}_d = \sqrt{\frac{1}{N}\sum_{k,t} \text{softplus}\left(\log\left(\frac{\text{AD}_{k,t}(\theta^d)+\eta}{\text{AD}_{k,t}(\theta^d)+\eta}\right)\right)^2},$$

matching the formula used to compute the score.

A total of 96 chains are constructed, and the sensitivity values for each component averaged over these draws. Parameter whose sensitivity values was above 0.025 were selected for calibration.

The global sensitivity analysis performed selected from 9 to 14 parameters depending on the datasets. While all parameters were considered, only 14 parameters were at least selected once (K_{S_c4+}, k_{m_c4+}, K_{S_ac}, k_{m_ac}, K_{S_pro}, k_{m_pro}, k_{m_su}, k_{m_aa}, k_{dec}, $K_{I,NH3}$, $pH_{UL:LL\ aa}$, $pH_{LL\ aa}$, $pH_{UL:LL\ ac}$, $pH_{LL\ ac}$). The parameters selected were

- k_{m_c4+}, k_{m_pro}, k_{m_ac}, k_{dec}, K_{S_c4+}, K_{S_pro}, K_{S_ac}, $pH_{LL\ aa}$, $pH_{UL:LL\ aa}$ for LN
- K_{S_c4+}, k_{m_c4+}, K_{S_ac}, k_{m_ac}, K_{S_pro}, k_{m_pro}, k_{m_su}, k_{m_aa}, k_{dec}, $K_{I,NH3}$, $pH_{UL:LL\ aa}$, $pH_{LL\ aa}$, $pH_{UL:LL\ ac}$, $pH_{LL\ ac}$ for HN
- k_{m_c4+}, k_{m_pro}, k_{m_ac}, k_{dec}, K_{S_c4+}, K_{S_pro}, K_{S_ac}, $pH_{LL\ aa}$, $pH_{UL:LL\ aa}$ for LF
- K_{S_c4+}, k_{m_c4+}, K_{S_ac}, k_{m_ac}, K_{S_pro}, k_{m_pro}, k_{m_su}, k_{m_aa}, k_{dec}, $K_{I,NH3}$, $pH_{UL:LL\ aa}$, $pH_{LL\ aa}$, $pH_{UL:LL\ ac}$for HF

Code for the sensitivity analysis routine can be found in 'aduq' library ('global_sensitivity' in 'aduq.pyadm1.SA' module).

2.2 Optimisation algorithm

Calibration is achieved by minimising the score. Since computing gradients for ADM1 is expensive, a gradient-free minimisation technique based on CMA-ES algorithm [Hansen, 2016] is used to perform the minimisation procedure for both models. The initial parameter values are the default ones for mesophilic digester specified by Rosén and

Jeppsson [2006] for ADM1 and Bernard et al. [2001] for AM2 (see appendix A), while the initial covariance matrix matches the covariance of the prior distributions used for VarBUQ.

For the main calibration procedure, the solver was deemed to have converged when the mean evolution of the score over the thirty last optimisation step was below 10^{-8}, or when the covariance diagonal elements all satisfy $\sqrt{\mathrm{Cov}_{i,i}} < 10^{-8}$. The maximum number of step was 250.

For the bootstrap procedure, the criteria for the score evolution was diminished to $2 * 10^{-4}$.

Code for the standard calibration algorithm can be found in 'aduq' library ('optim_cma' function from 'aduq.optim' for general implementation, 'optim_cma_adm1' and respectively 'optim_cma_am2' from respectively 'aduq.pyadm1.optim' and 'aduq.pyam2.optim' for ADM1 (resp. AM2) specific implementations).

3 VarBUQ algorithm

3.1 Variational class

The class of distributions on which Catoni's bound is optimized consists of subsets of Gaussian distributions whose covariance are block diagonal, the blocks being previously known. Thirteen groups are considered for ADM1:

- $k_{\mathrm{hyd\_CH}}$, $k_{\mathrm{hyd\_PR}}$, $k_{\mathrm{hyd\_LI}}$, k_{dis}
- $k_{\mathrm{m\_aa}}$, $K_{\mathrm{S\_aa}}$
- $k_{\mathrm{m\_fa}}$, $K_{\mathrm{S\_fa}}$, $K_{\mathrm{I\ H2\_fa}}$
- $k_{\mathrm{m\_c4+}}$, $K_{\mathrm{S\_c4+}}$, $K_{\mathrm{I,\ H2\_c4+}}$
- $k_{\mathrm{m\_pro}}$, $K_{\mathrm{S\_pro}}$, $K_{\mathrm{I\ H2\_pro}}$
- $k_{\mathrm{m\_ac}}$, $K_{\mathrm{S\_ac}}$
- $k_{\mathrm{m\_h2}}$, $K_{\mathrm{S\_h2}}$
- $k_{\mathrm{m\_su}}$, $K_{\mathrm{S\_su}}$
- $\mathrm{pH}_{\mathrm{LL\ ac}}$, $\mathrm{pH}_{\mathrm{UL:LL\ ac}}$, $K_{I,nh3}$
- $\mathrm{pH}_{\mathrm{LL\ aa}}$, $\mathrm{pH}_{\mathrm{UL:LL\ aa}}$
- $\mathrm{pH}_{\mathrm{LL\ h2}}$, $\mathrm{pH}_{\mathrm{UL:LL\ h2}}$
- $K_{\mathrm{S\_IN}}$
- k_{dec}

Two groups are considered for AM2:

- $\mu_{1\max}$, K_{S1}
- $\mu_{2\max}$, K_{S2}, K_{I2}

3.2 Gradient estimation

Starting from Catoni's bound, the estimation of the posterior is a minimisation problem, where the objective function is

$$\tilde{S}(\gamma) = \pi(\gamma)[S] + \lambda \mathrm{KL}\left(\pi(\gamma), \pi_0\right).$$

Writing this in integral form and noting $\ell(\gamma, \theta) := \log\left(\frac{\mathrm{d}\pi(\gamma)}{\mathrm{d}\pi_0}(\theta)\right)$ and $\tilde{KL}(\gamma) = \mathrm{KL}\left(\pi(\gamma), \pi_0\right)$,

$$\tilde{S}(\gamma) = \int S(\theta) \exp(\ell(\gamma, \theta)) d\pi_0(\theta) + \lambda \mathrm{KL}\left(\pi(\gamma), \pi_0\right)$$

so that

$$\begin{aligned}
\nabla \tilde{S}(\gamma) &= \int S(\theta) \exp(\ell(\gamma, \theta)) \nabla \ell(\gamma, \theta) \mathrm{d}\pi_0(\theta) + \lambda \nabla \tilde{\mathrm{KL}}(\gamma) \\
&= \int S(\theta) \nabla \ell(\gamma, \theta) \mathrm{d}\pi(\gamma)(\theta) + \lambda \nabla \tilde{\mathrm{KL}}(\gamma) \\
&= \mathbb{E}_{\theta \sim \pi(\gamma)}[S(\theta) \nabla \ell(\gamma, \theta)] + \lambda \nabla \tilde{\mathrm{KL}}(\gamma).
\end{aligned}$$

For Gaussian distributions, the Kullback-Leibler divergence has an expression which can be directly differentiated. The derivative of the mean score term is expressed as an expectation, which can be approximated using an i.i.d. sample of θ from $\pi(\alpha)$. This gives an unbiased estimator, with variance decaying as $\frac{1}{\sqrt{n}}$. This slow decay of the variance implies that the sample size should be somewhat large to obtain good approximations. Moreover, since the sample $(\theta)_i$ has to be sampled from a specific distribution, the naive unbiased estimator needs to reevaluate a high number of scores at each step. If the gradient steps are small, the samples from the previous distribution $\pi(\alpha_{t-k})$ should be similar to samples from the current distribution $\pi(\alpha_t)$ - implying that the information from the previous samples is meaningful. For this reason, the estimator is constructed in the following manner. For any $k > 0$,

$$\mathbb{E}_{\theta \sim \pi(\gamma_t)}[S(\theta) \nabla \ell(\gamma_t, \theta)] = \mathbb{E}_{\theta \sim \pi(\gamma_{t-k})}[S(\theta) \exp\left(\ell(\gamma_t, \theta) - \ell(\gamma_{t-k}, \theta)\right) \nabla \ell(\gamma_t, \theta)].$$

This implies that using the sample $(\theta_{i,t-k})_{i \in [1,K]}$ drawn i.i.d. from $\pi(\gamma_{t-k})$,

$$\tilde{D}_{t,k} := K^{-1} \sum_{i=1}^{K} S(\theta_{i,t-k}) \exp\left(\ell(\gamma_t, \theta_{i,t-k}) - \ell(\gamma_{t-k}, \theta_{i,t-k})\right) \nabla \ell(\gamma_t, \theta_{i,t-k})$$

is an unbiased estimator of $\nabla(\pi(\cdot)[S])(\gamma_t)$. As such, the final estimator of the gradient is constructed as

$$\tilde{D}_t = \frac{1-\beta}{1-\beta^{\tilde{k}_{\max}+1}} \sum_{k=0}^{\tilde{k}_{\max}} \beta^k \tilde{D}_{t,k}$$

where $0 \leq \beta \leq 1$ and $k_{\max} = \min(k_{\max}, t)$

The gradient descent is accelerated, resulting in step $S_t = gS_{t-1} + (1-g)\eta\tilde{D}_t$.

3.3 Step removal procedure

The sample $(\theta_{i,t})_{i\in[1,K]}$ can be used to estimate the mean score for distribution γ_t. In the small step ($\eta \ll 1$) and small sample size regime, the standard deviation of the mean score can be significantly higher than the step difference of the objective $|S(\gamma_t) - S(\gamma_{t-1})|$, so that refusing all steps resulting in $\hat{S}(\gamma_t) > \hat{S}(\gamma_{t-1})$ can harm the efficiency of the algorithm.

Still, it is possible to remove steps for which it is highly probable that the objective is increased, by checking whether $\hat{S}(\gamma_t) > \hat{S}(\gamma_{t-1}) + \text{Tol} * \sigma_t$ where σ_t is the standard deviation of $(S(\theta_{i,t}))_{i\in[1,K]}$ and Tol is the 90% quantile of the normal distribution.

If a step t is removed, the current distribution is set to γ_{t-1}, the speed is reset to 0 and the sample $(\theta_{i,t})_{i\in[1,K]}$ is erased.

Code for VarBUQ routine can be found in 'aduq' library ('optim_VI_wc' function from 'aduq.bayes' module for the general algorithm, while ADM1 (resp. AM2) specific code can be found in 'aduq.pyadm1.var_buq' module for ADM1 specific code (resp 'aduq.pyam2.var_buq' module).

4 Results of calibration routines

The calibration algorithms were performed using the hyperparameters specified in Table 1.

The calibrated model obtained by standard and Bayesian calibration are specified in Table 2. For the Bayesian calibration, only the mean parameter in log-space is tabulated.

The minimized scores achieved by the calibration procedure are given in Section 4.

Table 1: Hyperparameters for calibration routines.

(a) Bayesian calibration

	step_size	PAC-Bayes. temp.	corr_eta	gen_decay	momentum	per_step	k
ADM1	0.025	0.001	0.7	0.20	0.985	160	8000
AM2	0.040	0.002	0.7	0.15	0.960	256	20000

(b) Standard calibration

	per_step	radius_ini	radius_factor	no_change_max	keep_frac	cov_updt_speed
ADM1	96	0.2	0.7	96	0.3	0.06
AM2	480	0.1	0.7	1024	0.3	0.08

Table 2: Calibrated parameters

(a) ADM1

	θ_{N}	$\hat{\theta}_{\mathrm{LN}}$	$\hat{\pi}_{\mathrm{LN}}$	$\hat{\theta}_{\mathrm{HN}}$	$\hat{\pi}_{\mathrm{HN}}$	θ_F	$\hat{\theta}_{\mathrm{LF}}$	$\hat{\pi}_{\mathrm{LF}}$	$\hat{\theta}_{\mathrm{HF}}$	$\hat{\pi}_{\mathrm{HF}}$
k_{dis}	0.355	0.500	0.575	0.500	0.601	0.233	0.500	0.436	0.500	0.493
k_{hyd_CH}	10.0	10.0	10.5	10.0	9.90	5.19	10.0	9.74	10.0	9.34
k_{hyd_PR}	8.65	10.0	9.13	10.0	9.99	11.9	10.0	9.91	10.0	8.83
k_{hyd_LI}	23.4	10.0	9.54	10.0	10.3	2.40	10.0	10.4	10.0	9.41
k_{m_su}	37.7	30.0	32.0	2.42	30.4	47.7	30.0	32.0	107	27.5
k_{m_aa}	44.6	50.0	47.7	49.5	55.6	48.7	50.0	50.6	2.23	51.3
k_{m_fa}	5.23	6.00	5.39	6.00	6.53	4.47	6.00	7.10	6.00	5.45
k_{m_c4+}	21.3	89.6	18.1	123	18.8	17.6	36.8	17.1	44.0	16.3
k_{m_pro}	12.7	92.9	11.9	26.2	12.6	9.91	49.4	14.3	9.35	14.9
k_{m_ac}	10.7	80.0	10.3	20.7	10.6	10.2	53.1	8.39	9.29	9.75
k_{m_h2}	27.9	35.0	33.2	35.0	32.7	28.7	35.0	36.9	35.0	35.0
k_{dec}	1.57e-02	1.99e-02	1.84e-02	6.16e-03	1.66e-02	1.45e-02	1.24e-02	1.65e-02	2.77e-02	2.08e-02
K_{S_IN}	9.31e-05	1.00e-04	9.96e-05	1.00e-04	9.93e-05	9.46e-05	1.00e-04	1.01e-04	1.00e-04	9.87e-05
K_{S_su}	0.427	0.500	0.499	0.500	0.543	0.293	0.500	0.469	0.500	0.467
K_{S_aa}	0.309	0.300	0.314	0.300	0.307	0.287	0.300	0.304	0.300	0.303
K_{S_fa}	0.389	0.400	0.441	0.400	0.402	0.579	0.400	0.419	0.400	0.427
K_{S_c4+}	0.300	1.18	0.236	2.00	0.256	0.399	0.898	0.372	0.939	0.350
K_{S_pro}	0.115	0.857	0.100	0.303	0.115	6.48e-02	0.380	9.39e-02	5.07e-02	0.103
K_{S_ac}	0.147	1.21	0.129	0.210	0.133	0.243	1.50	0.153	0.227	0.194
K_{S_h2}	6.65e-06	7.00e-06	7.64e-06	7.00e-06	6.11e-06	6.12e-06	7.00e-06	7.49e-06	7.00e-06	7.33e-06
$K_{I,H2\_fa}$	5.01e-06	5.00e-06	4.93e-06	5.00e-06	5.05e-06	5.60e-06	5.00e-06	5.15e-06	5.00e-06	4.99e-06
$K_{I,H2\_c4+}$	1.06e-05	1.00e-05	1.00e-05	1.00e-05	9.98e-06	1.11e-05	1.00e-05	1.04e-05	1.00e-05	9.85e-06
$K_{I,H2\_pro}$	3.15e-06	3.50e-06	3.49e-06	3.50e-06	3.55e-06	3.37e-06	3.50e-06	3.66e-06	3.50e-06	3.55e-06
$K_{I,NH3}$	1.88e-03	1.80e-03	1.91e-03	7.88e-04	1.75e-03	2.05e-03	1.80e-03	1.75e-03	2.91e-03	2.09e-03
$pH_{UL:LL\ aa}$	1.16	1.08	1.55	1.26	1.33	1.36	0.311	1.45	1.37	1.47
$pH_{LL\ aa}$	5.11	5.18	3.61	6.00	3.48	4.43	2.63	3.80	5.21	3.64
$pH_{UL:LL\ ac}$	1.02	1.00	1.01	0.826	0.986	1.00	1.00	0.988	1.00	1.01
$pH_{LL\ ac}$	4.53	5.00	4.98	5.33	4.90	4.76	5.00	4.97	3.80	4.93
$pH_{UL:LL\ h2}$	0.524	1.00	1.03	1.00	1.05	1.61	1.00	0.941	1.00	0.966
$pH_{LL\ h2}$	5.50	5.00	4.99	5.00	5.02	6.05	5.00	4.94	5.00	4.80

(b) AM2

	θ^*_{N}	$\hat{\theta}_{\mathrm{LN}}$	$\hat{\pi}_{\mathrm{LN}}$	$\hat{\theta}_{\mathrm{HN}}$	$\hat{\pi}_{\mathrm{HN}}$	θ^*_{F}	$\hat{\theta}_{\mathrm{LF}}$	$\hat{\pi}_{\mathrm{LF}}$	$\hat{\theta}_{\mathrm{HF}}$	$\hat{\pi}_{\mathrm{HF}}$
$\mu_{1\mathrm{max}}$	1.23	0.55	1.17	0.91	1.18	0.99	0.86	1.02	0.91	0.86
$\mu_{2\mathrm{max}}$	1.05	0.99	0.74	0.90	0.78	1.32	2.00	0.95	1.21	1.16
K_{S1}	7.51	5.76	7.34	6.25	7.20	15.98	15.38	15.98	15.13	14.40
K_{S2}	11.0	11.0	9.84	10.1	9.09	4.97	5.36	4.5	4.85	4.68
K_{I2}.	196.0	135	256	259	259	345	328	252	290	260

θ^* stands for the true sets of parameters used to generate the data. $\hat{\theta}$ stands for the result of the standard calibration. $\hat{\pi}[\theta]$ stands for the mean parameter found by VarBUQ using the chosen a priori PAC Bayesian temperature in the log space, after application of an exponential transform (i.e. $\exp(\hat{\pi}[\log(\theta)])$)

Table 3: Summary of standard and Bayesian calibration results

	$S(\theta^*)$	$S(\hat{\theta})$	$\hat{\pi}_\lambda[S(\theta)]$	$\mathrm{KL}(\hat{\pi}_\lambda, \pi)$	$\hat{\pi}_{2\lambda}[S(\theta)]$	$\mathrm{KL}(\hat{\pi}_{2\lambda}, \pi)$	$\hat{\pi}_{8\lambda}[S(\theta)]$	$\mathrm{KL}(\hat{\pi}_{8\lambda}, \pi)$
AM2 LN	0.09045	0.08922	0.09401	4.423	n.a.	n.a.	n.a.	n.a.
AM2 HN	0.09297	0.09239	0.09657	4.459	n.a.	n.a.	n.a.	n.a.
AM2 LF	0.08971	0.08957	0.09296	9.152	n.a.	n.a.	n.a.	n.a.
AM2 HF	0.08640	0.08583	0.08942	9.936	n.a.	n.a.	n.a.	n.a.
ADM1 LN	0.09902	0.09711	0.10089	13.18	0.10289	11.02	0.11589	6.02
ADM1 HN	0.10106	0.09496	0.10712	11.90	0.10833	10.35	0.12426	5.78
ADM1 LF	0.09313	0.09241	0.09687	13.99	0.10112	10.47	0.11368	6.24
ADM1 HF	0.10449	0.09682	0.11031	16.50	0.11362	13.67	0.13391	6.57

$S(\theta^*)$ is the score obtained by the true set of parameters, $S(\hat{\theta})$ the score obtained by the set of parameters obtained by the standard calibration. $\hat{\pi}_{k\lambda}[S(\theta)]$ is the average score of the posterior obtained by the Bayesian calibration with PAC-Bayesian temperature $k\lambda$, where λ is the PAC-Bayesian temperature chosen a priori. $\mathrm{KL}(\hat{\pi}_{k\lambda}, \pi)$ is the Kullback–Leibler divergence between the posterior obtained by the Bayesian calibration with PAC-Bayesian temperature $k\lambda$ and the prior.

References

O. Bernard, Z. Hadj-Sadok, D. Dochain, A. Genovesi, J.-P. Steyer, Dynamical model development and parameter identification for an anaerobic wastewater treatment process, Biotechnology and Bioengineering 75 (2001) 424–438. URL: https://doi.org/10.1002/bit.10036.

N. Hansen, The CMA evolution strategy: A tutorial, arXiv preprint arXiv:1604.00772 (2016). URL: https://doi.org/10.48550/arXiv.1604.00772.

M. D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (1991) 161–174. URL: https://doi.org/10.1080/00401706.1991.10484804.

C. Rosén, U. Jeppsson, Aspects on ADM1 implementation within the BSM2 framework, Department of Industrial Electrical Engineering and Automation, Lund University, Lund, Sweden (2006) 1–35.