

Supplementary Information: Bayesian parameter estimation for biochemical reaction networks using region-based adaptive parallel tempering

Benjamin Ballnus^{*†}, Steffen Schaper[‡], Fabian J. Theis ^{*†}
and Jan Hasenauer^{*†}

^{*}Helmholtz Zentrum München – German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg 85764, Germany

[†]Technische Universität München, Center for Mathematics, Chair of Mathematical Modeling of Biological Systems, Garching 85748, Germany

[‡]Bayer AG, Engineering & Technologies, Applied Mathematics, Leverkusen 51369, Germany

1 RAmPART Pseudocode

In the main manuscript, we presented the essential idea behind the Region-based Adaptive PARallel Tempering algorithm (RAmPART) and its structure. In this section, we provide the pseudocode of (RAmPART) **and default values for its inputs**. Warm-up and sampling phase are presented individually. In the warm-up phase, an adaptive parallel tempering algorithm is used to gather a sample to estimate the parameters of a Gaussian Mixture Model (GMM) (Figure 1). In the sampling phase, the Gaussian mixture model is used to initialize a region-based parallel tempering algorithm which adapts the regional proposal covariances (Figure 2). In the pseudocode, we denote the number sample iterations N , the chain initialization point $\theta^{[0]}$, the initial proposal covariance matrix \mathbf{C} , the initial **inverse** temperature latter β (with $\beta_1 = 1$ and $\beta_L = 1/\tau_{max}$), the number of parallel tempered chains L , the maximum number of regions R_{max} , the number of cross validations when training the GMM N_{rep} , the adaptation parameters α , ν_τ and η_τ and the fraction of global proposals compared to local proposals p_g . **Please note, we are using a heuristic for initializing β , which is based on the discussion in [5] and personal experience. An overview of all input parameters is shown in Table 1.** The algorithm returns the posterior parameter chain $\theta_1^{[0]}, \dots, \theta_1^{[N-1]}$ as well as the corresponding posterior values $p^{[0]}, \dots, p_1^{[N-1]}$. **An overview of all output parameters is provided in Table 2.**

2 Running the warm-up phase in practice

In order to sample efficiently during the sampling phase, RAmPART first gathers a training sample in its warm-up phase. The training sample is used to train a GMM capturing the features of the underlying posterior landscape. To ensure robustness, (C1) the training sample has to be sufficiently representative for the posterior topology (e.g. multiple modes or pronounced tails have to be roughly covered) and (C2) the EM algorithm has to converge to a reasonable optimum.

To ensure that C1 holds, sufficiently many parallel chains, sufficiently many warm-up iterations and an appropriate maximum temperature τ_{max} are required. C2 can be ensured by initializing the EM algorithm for the fitting of the GMM at sufficiently many starting points. In practice we found that it is more difficult to ensure C1 than C2. For the presented benchmark examples and biological applications, $N_{warmUp} = 10^5$ samples and $N_{rep} = 5$ led to good results in all runs.

Should RAmPART fail to obtain a reasonable GMM in its warm-up phase, the sampling performance during the run is comparable to that of an adaptive PT algorithm. There is a computational overhead caused by training the GMM and the assessment of the region during the sampling phase. However, this computational overhead is negligible if the evaluation of the point-wise evaluation of the posterior distribution is computationally de-

Algorithm 1: Warm-up Phase

```

input :  $N, \theta^{[0]}, \mathbf{C}, \beta, L, R_{max}, N_{rep}, \alpha, \nu_\tau, \eta_\tau$ 
output:  $R, \theta^{[N-1]}, w, \mu, \Sigma, \beta, m, C$ 

// Initialize
for  $\ell \leftarrow 1$  to  $L$  do
     $\eta_\ell \leftarrow 1, m_\ell \leftarrow \theta^{[0]}$ 
     $p_\ell = \pi(\theta^{[0]} | \mathcal{D})$ 
for  $i \leftarrow 0$  to  $N - 1$  do
    // Random walk step
    for  $\ell \leftarrow 1$  to  $L$  do
        Propose parameter  $\theta' \sim \mathcal{N}(\theta_\ell^{[i]} | m_\ell, \eta_\ell^2 C_\ell)$ 
        Evaluate posterior  $p' \leftarrow \pi(\theta' | \mathcal{D})$ 
        Acceptance chance  $p \leftarrow (p' / p_\ell)^{\beta_\ell}$ 
        Draw uniformly  $v \sim U([0, 1])$ 
        if  $v < p$  then
             $\theta_\ell^{[i+1]} \leftarrow \theta'$ 
             $p_\ell = p'$ 
    // Random walk proposal density adaptation
    for  $\ell \leftarrow 1$  to  $L$  do
         $\gamma^{[i]} = 1 / (1 + i)^\alpha$ 
         $m_\ell \leftarrow (1 - \gamma^{[i]}) m_\ell + \gamma^{[i]} \theta_\ell^{[i]}$ 
         $C_\ell \leftarrow (1 - \gamma^{[i]}) C_\ell + \gamma^{[i]} (\theta_\ell^{[i]} - m_\ell)(\theta_\ell^{[i]} - m_\ell)^T$ 
         $\eta_\ell \leftarrow \eta_\ell \exp(\gamma^{[i]}(p - 0.234))$ 
    // Chain swapping
    for  $\ell \leftarrow L$  to  $2$  do
         $\Delta\beta_{\ell-1} \leftarrow \beta_{\ell-1} - \beta_\ell$ 
        Swap probability  $p_{swap, \ell-1} \leftarrow (p_\ell / p_{\ell-1})^{\Delta\beta_{\ell-1}}$ 
        Draw uniformly  $w \sim U([0, 1])$ 
         $A_{\ell-1} \leftarrow (w < p_{swap})$ 
        if  $A_{\ell-1}$  then
             $\theta_\ell \leftrightarrow \theta_{\ell-1}$ 
             $p_\ell \leftrightarrow p_{\ell-1}$ 
    // Temperature adaptation
     $\kappa^{[i]} \leftarrow \nu_\tau / (\eta_\tau(i + 1 + \nu_\tau))$ 
    for  $\ell \leftarrow 1$  to  $L - 2$  do
         $\Delta S_\ell \leftarrow \kappa^{[i]}(A_\ell - A_{\ell+1})$ 
         $\Delta\tau_\ell \leftarrow 1/\beta_{\ell+1} - 1/\beta_\ell$ 
    for  $\ell \leftarrow 2$  to  $L - 1$  do
         $\beta_\ell \leftarrow 1 / \sum_{m=2}^\ell \Delta\tau_{m-1} \exp(\Delta S_{m-1})$ 
// Train GMM
for  $k \leftarrow 1$  to  $N_{replicates}$  do
    for  $n \leftarrow 1$  to  $R_{max}$  do
        Fit GMM  $w_{n,k}, \mu_{n,k}, \Sigma_{n,k} \leftarrow \text{EM-Algorithm}(\theta_1^{[0]}, \dots, \theta_1^{[N]}; n, \text{random seed } k)$ 
         $BIC_{n,k} \leftarrow -2 \log \left( \sum_{m=0}^N \text{GMM}(\theta_\ell^{[m]} | w_{n,k}, \mu_{n,k}, \Sigma_{n,k}) \right) +$ 
         $n \log(N) ((n-1)/n + 2 \dim_\theta + (\dim_\theta - 1) \dim_\theta / 2)$ 
    Select best GMM  $(R, K) \leftarrow \underset{n,k}{\text{argmin}}(BIC_{n,k})$ 
    Define regions  $w, \mu, \Sigma \leftarrow w_{R,K}, \mu_{R,K}, \Sigma_{R,K}$ 

```

Figure 1: Pseudocode for warm-up phase of RAmPART.

Algorithm 2: Sampling Phase

```

input :  $N, N_{warmUp}, \theta^{[0]}, C, \beta, L, R_{max}, N_{rep}, \alpha, \nu_\tau, \eta_\tau, p_g$ 
output: non-tempered parameter chain  $\theta_1^{[0]}, \dots, \theta_1^{[N-1]}$ ,
          non-tempered posterior values  $p^{[0]}, \dots, p_1^{[N-1]}$ 

// Run warm-up phase
 $(R, \theta^{[0]}, w, \mu, \Sigma, \beta, m, C) \leftarrow \text{WarmUpPhase}(N_{warmUp}, \theta^{[0]}, C^{[0]}, \beta^{[0]}, L, R_{max}, N_{rep}, \alpha, \nu_\tau, \eta_\tau)$ 
// Initialize
for  $\ell \leftarrow 1$  to  $L$  do
  New region label  $r_{prop, \ell} \leftarrow \underset{r}{\operatorname{argmax}} \left( \mathcal{N}(\theta_\ell^{[0]} | \mu_r, w_r \Sigma_r) \right)$ 
   $p_\ell = \pi(\theta^{[0]} | \mathcal{D})$ 
  for  $r \leftarrow 1$  to  $R$  do
    Local proposal parameters  $\eta_{\ell, r} \leftarrow 1, m_{\ell, r} \leftarrow m, C_{\ell, r} \leftarrow C$ 
    Adaptation times  $j_{\ell, r} \leftarrow 0$ 
for  $i \leftarrow 0$  to  $N - 1$  do
  // Random Walk Step
  for  $\ell \leftarrow 1$  to  $L$  do
    Set old region label  $r_\ell \leftarrow r_{prop, \ell}$ 
    Adaptation fading  $j_{\ell, r_\ell} \leftarrow j_{\ell, r_\ell} + 1$ 
    Draw uniformly  $u \sim U([0, 1])$ 
    if  $u < 0.5$  then
       $\theta' \sim \mathcal{N}(\theta_\ell^{[i]} | m_{\ell, r_\ell}, \eta_{\ell, r_\ell}^2 C_{\ell, r_\ell})$ 
    else
       $\theta' \sim \mathcal{N}(\theta_\ell^{[i]} | m_\ell, \eta_\ell^2 C_\ell)$ 
    Get new region label  $r_{prop, \ell} \leftarrow \underset{r}{\operatorname{argmax}} (\mathcal{N}(\theta' | \mu_r, w_r \Sigma_r))$ 
    Evaluate posterior  $p' \leftarrow \pi(\theta' | \mathcal{D})$ 
    Forward probability  $T_{for} \leftarrow (1 - p_g) \mathcal{N}(\theta' | \theta_\ell, C_{\ell, r_\ell}) + p_g \mathcal{N}(\theta' | \theta_\ell, C_\ell)$ 
    Backward probability  $T_{back} \leftarrow (1 - p_g) \mathcal{N}(\theta_\ell | \theta', C_{\ell, r_{prop, \ell}}) + p_g \mathcal{N}(\theta_\ell | \theta', C_\ell)$ 
    Acceptance chance  $p \leftarrow (p' / p_\ell)^{\beta_\ell} (T_{back} / T_{for})$ 
    Draw uniformly  $v \sim U([0, 1])$ 
    if  $v < p$  then
       $\theta_\ell^{[i+1]} \leftarrow \theta'$ 
       $p_\ell \leftarrow p'$ 
  // Random walk proposal density adaptation
  for  $\ell \leftarrow 1$  to  $L$  do
     $\gamma = 1 / j_{\ell, r_\ell}^\alpha$ 
     $m_{\ell, r_\ell} \leftarrow (1 - \gamma) m_{\ell, r_\ell} + \gamma \theta_\ell^{[i]}$ 
     $C_{\ell, r_\ell} \leftarrow (1 - \gamma) C_{\ell, r_\ell} + \gamma (\theta_\ell^{[i]} - m_{\ell, r_\ell})(\theta_\ell^{[i]} - m_{\ell, r_\ell})^T$ 
     $\eta_{\ell, r_\ell} \leftarrow \eta_{\ell, r_\ell} \exp(\gamma(p - 0.234))$ 
  // Chain swapping
  for  $\ell \leftarrow L$  to  $2$  do
     $\Delta \beta_{\ell-1} \leftarrow \beta_{\ell-1} - \beta_\ell$ 
    Swap probability  $p_{swap, \ell-1} \leftarrow (p_\ell / p_{\ell-1})^{\Delta \beta_{\ell-1}}$ 
    Draw uniformly  $w \sim U([0, 1])$ 
     $A_{\ell-1} \leftarrow (w < p_{swap})$ 
    if  $A_{\ell-1}$  then
       $\theta_\ell \leftrightarrow \theta_{\ell-1}$ 
       $p_\ell \leftrightarrow p_{\ell-1}$ 
       $r_{prop, \ell} \leftrightarrow r_{prop, \ell-1}$ 
  // Temperature adaptation
   $\kappa^{[i]} \leftarrow \nu_\tau / (\eta_\tau (i + 1 + \nu_\tau))$ 
  for  $\ell \leftarrow 1$  to  $L - 2$  do
     $\Delta S_\ell \leftarrow \kappa^{[i]} (A_\ell - A_{\ell+1})$ 
     $\Delta \tau_\ell \leftarrow 1 / \beta_{\ell+1} - 1 / \beta_\ell$ 
  for  $\ell \leftarrow 2$  to  $L - 1$  do
     $\beta_\ell \leftarrow 1 / \sum_{m=2}^\ell \Delta \tau_{m-1} \exp(\Delta S_{m-1})$ 

```

Figure 2: Pseudocode for sampling phase of RAMPART.

manding.

3 Benchmark Problems

3.1 20-dimensional Gaussian mixture distribution

As mentioned in the main manuscript, the 20-dimensional Gaussian mixture distribution is given by

$$\pi_{\text{gm}}(\theta|\mathcal{D}) \propto \left(\sum_{i=1}^2 \mathcal{N} \left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \middle| \begin{pmatrix} \mu_{i,1} \\ \mu_{i,2} \end{pmatrix}, \Sigma \right) \right) \prod_{j=3}^{20} \mathcal{N}(\theta_j | 25, \sigma^2), \quad (1)$$

with $\mu_1 = (-50, -50)^t$, $\mu_2 = (50, 50)^t$, $\Sigma = 250 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ and $\sigma = 1$. The box constraints are summarized in Table 3.

3.2 20-dimensional blurred ring distribution

The 20-dimensional blurred ring distribution is defined by

$$\pi_{\text{ring}}(\theta|\mathcal{D}) \propto \mathcal{N}(r(\theta) | r_0, \sigma_r^2) \prod_{j=3}^{20} \mathcal{N}(\theta_j | 0, \sigma^2) \quad (2)$$

with $r(\theta) = \sqrt{\theta_1^2 + \theta_2^2}$, $\sigma = 1$, $r_0 = 50$ and $\sigma_r = 5$. The box constraints are summarized in Table 4.

4 Application Problems

4.1 mRNA Transfection

We consider the ODE model for mRNA transfection introduced by [2]

$$[\text{GFP}] = k_{TL}[\text{mRNA}] - \beta[\text{GFP}], \quad [\text{GFP}(t_0)] = 0, \quad (3)$$

$$[\text{mRNA}] = -\delta[\text{mRNA}], \quad [\text{mRNA}(t_0)] = m_0, \quad (4)$$

in which $[\text{GFP}]$ denotes the concentration of green fluorescent protein and $[\text{mRNA}]$ denotes the concentration of mRNA of green fluorescent protein. The mRNA is released in the cell at t_0 and the initial concentration is m_0 . The mRNA is translated with rate k_{TL} . Degradation rates for mRNA and protein are δ and β , respectively.

The analytical solution for the ODE model is given by

$$[\text{mRNA}(t)] = \begin{cases} 0, & t < t_0, \\ m_0 \exp(-\delta(t - t_0)), & \text{otherwise,} \end{cases} \quad (5)$$

and

$$[\text{GFP}(t)] = \begin{cases} 0, & t < t_0 \\ k_{TL}m_0 (\exp(-\beta(t-t_0)) - \exp(-\delta(t-t_0))) / (\delta - \beta), & t \geq t_0 \wedge \delta \neq \beta \\ k_{TL}m_0(t-t_0) \exp(-\delta(t-t_0)), & t \geq t_0 \wedge \delta = \beta \end{cases} \quad (6)$$

The experimental data were collected for $[\text{GFP}(t)]$ at 150 time points in the time interval $t \in [2, 27]$ hours [2]. For such data, the parameters k_{TL} and m_0 are structurally non-identifiable, as $[\text{GFP}(t)]$ merely depends on the product $\kappa = k_{TL}m_0$. We address this problem by estimating merely κ . In addition, β and δ are only locally structurally identifiable. The values of the two parameters can be interchanged without altering the observable $[\text{GFP}(t)]$. In previous studies, it was assumed that the mRNA half-life is smaller than the protein half-life, implying $\beta < \delta$. As this is not necessarily correct, we do not apply this constraint.

Simulation and data are compared using an error model assuming the measurement noise is normally distributed with standard deviation σ . As σ is unknown, it is taken into account as an additional parameter for parameter estimation. The parameter constraints are reported in Table 5.

4.2 JAK2/STAT5 Signaling

We consider the ODE model for JAK2/STAT5 signaling introduced by [3], which is based on the original publication by [4]. The ODE model is defined by

$$\frac{\partial}{\partial t}[\text{STAT}] = (\Omega_{nuc} \cdot p_4 \cdot [\text{nSTAT}_5] - \Omega_{cyt} \cdot [\text{STAT}] \cdot p_1 \cdot u(t)) / \Omega_{cyt} \quad (7)$$

$$\frac{\partial}{\partial t}[\text{pSTAT}] = [\text{STAT}] \cdot p_1 \cdot u(t) - 2p_2 \cdot [\text{pSTAT}]^2 \quad (8)$$

$$\frac{\partial}{\partial t}[\text{pSTAT2}] = p_2 \cdot [\text{pSTAT}]^2 - p_3 \cdot [\text{pSTAT2}] \quad (9)$$

$$\frac{\partial}{\partial t}[\text{npSTAT2}] = -(\Omega_{nuc} \cdot p_4 \cdot [\text{npSTATmpSTAT}] - \Omega_{cyt} \cdot p_3[\text{pSTAT2}]) / \Omega_{nuc} \quad (10)$$

$$\frac{\partial}{\partial t}[\text{nSTAT}_1] = -p_4([\text{nSTAT}_1] - 2[\text{npSTAT2}]) \quad (11)$$

$$\frac{\partial}{\partial t}[\text{nSTAT}_2] = p_4([\text{nSTAT}_1] - [\text{nSTAT}_2]) \quad (12)$$

$$\frac{\partial}{\partial t}[\text{nSTAT}_3] = p_4([\text{nSTAT}_2] - [\text{nSTAT}_3]) \quad (13)$$

$$\frac{\partial}{\partial t}[\text{nSTAT}_4] = p_4([\text{nSTAT}_3] - [\text{nSTAT}_4]) \quad (14)$$

$$\frac{\partial}{\partial t}[\text{nSTAT}_5] = p_4([\text{nSTAT}_4] - [\text{nSTAT}_5]) \quad (15)$$

in which u is the time dependent level of phosphorylated Epo receptor and the initial conditions are defined by $\mathbf{x}(0) = \mathbf{0}$ for all states except for $[\text{STAT}](0) = [\text{STAT}]_{tot}$. The phosphorylated Epo receptor initiates JAK2/STAT5 signalling and we model it using a cubic spline function. This function has the values $u(0) = sp_1$, $u(5) = sp_2$, $u(10) = sp_3$, $u(20) = sp_4$, $u(60) = sp_5$.

For the process, we have measurement data for the amount of phosphorylated STAT and total STAT in the cytosol,

$$y_{[\text{pSTAT}]} = O_{[\text{pSTAT}]} + s_{[\text{pSTAT}]} / [\text{STAT}]_{tot} ([\text{pSTAT}] + 2[\text{pSTAT}2]) \quad (16)$$

$$y_{[\text{tSTAT}]} = O_{[\text{tSTAT}]} + s_{[\text{tSTAT}]} / [\text{STAT}]_{tot} ([\text{STAT}] + [\text{pSTAT}] + 2[\text{pSTAT}2]) \quad (17)$$

and the concentration of phosphorylated Epo receptor

$$y_{[\text{pEpoR}]} = u(t). \quad (18)$$

The unknown parameters of the models are $\boldsymbol{\theta} = (p_1, p_2, p_3, p_4, [\text{STAT}]_{tot}, sp_1, sp_2, sp_3, sp_4, sp_5, O_{[\text{tSTAT}]}, O_{[\text{pSTAT}]}, s_{[\text{tSTAT}]}, s_{[\text{pSTAT}]}, \sigma_{[\text{pSTAT}]}, \sigma_{[\text{tSTAT}]}, \sigma_{[\text{pEpoR}]})^t$, the volumes of cytosol and nucleus (Ω_{cyt} and Ω_{nuc}) are treated as constants. All parameters are estimated in log-space. The box constraints are defined in Table 6.

5 Additional Summary Statistics

In addition to the summary statistics in the main manuscript, we assessed the overall running times (Table 7) and the number of runs which provide a representative sample (Table 8). For the calculation of the computation times, all runs were considered regardless of whether they provided a representative sample or not. We define a run as converged, if it has $ESS > 0$ after applying the analysis pipeline introduced in [1]. Since PT and RAmPART were executed with multiple parallel chains (40, 40, 30, 60 for Gaussian mixture, blurred ring, mRNA transfection and JAK2/STAT5 signaling), the computation times differ by orders of magnitude.

Remark: The summary statistics reported in the main manuscript, we calculated either

$$ESS'_{run} = ESS_{run} \cdot EQ_{\text{all runs}} \quad (19)$$

or

$$(ESS_{run}/t_{run})' = ESS_{run}/t_{run} \cdot EQ_{\text{all runs}} \quad (20)$$

for each run, which was identified to explore the posterior. Here, $EQ_{\text{all runs}}$ denotes the fraction of runs of a certain algorithm which provided a representative sample (please refer to [1] for an in-depths explanation).

In comparison to PT, RAmPART requires a warm-up phase which increases the computational cost. In the warm-up phase, a "small" sample is generated and used to train a GMM for defining the regions for the sampling phase. Figure 3 shows the absolute running times of all algorithms

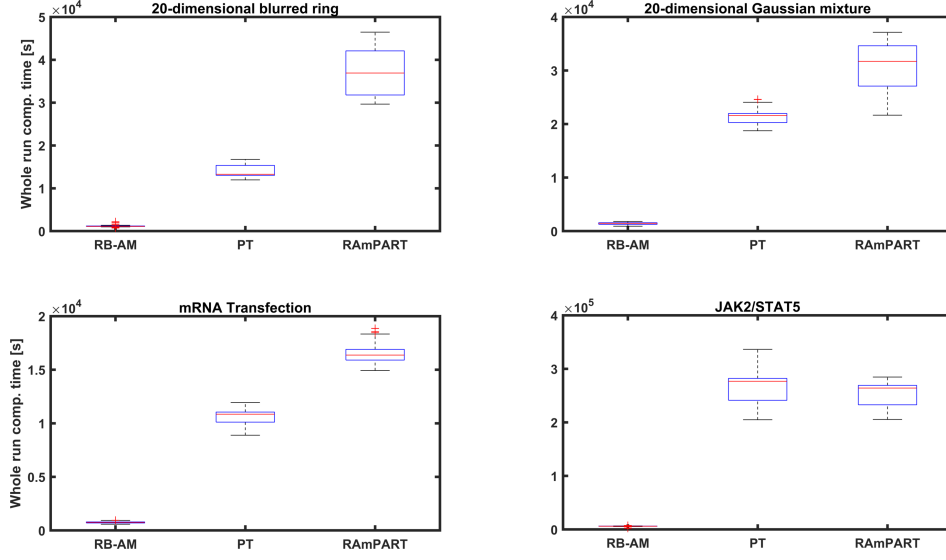


Figure 3: Running time comparison between RB-AM, PT and RAmPART.

for the benchmark problems and the biological applications. The 20-dimensional blurred ring, the 20-dimensional Gaussian mixture and the mRNA-Transfection example have a swift objective function evaluation, as no ODE has to be solved numerically. For these examples, the computation time increases due to the overhead generated by RAmPART. For the JAK2/STAT5 model, the additional computational effort is small compared to the objective evaluation. This is probably the most realistic scenario for parameter estimation problems in systems biology. However, as shown in the main manuscript, for all presented benchmark problems and biological applications, this computational overhead is worth the effort, as the overall ESS/s increased.

References

- [1] B. Ballnus, S. Hug, K. Hatz, L. Görlitz, J. Hasenauer, and F. J. Theis. Comprehensive benchmarking of Markov chain Monte Carlo methods for dynamical systems. *BMC Syst. Biol.*, 11(63), June 2017.
- [2] C. Leonhardt, G. Schwake, T. R. Stögbauer, S. Rappl, J. T. Kuhr, T. S. Ligon, and J. O. Rädler. Single-cell mRNA transfection studies: Delivery, kinetics and statistics by numbers. *Nanomedicine: Nanotechnology, Biology, and Medicine*, 10(4):679–688, May 2014.

- [3] C. Maier, C. Loos, and J. Hasenauer. Robust parameter estimation for dynamical systems from outlier-corrupted data. *Bioinformatics*, 33(5):718–725, Mar. 2017.
- [4] I. Swameye, T. G. Müller, J. Timmer, O. Sandra, and U. Klingmüller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc. Natl. Acad. Sci. USA*, 100(3):1028–1033, Feb 2003.
- [5] V. Vyshemirsky and M. A. Girolami. Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6):833–839, 2008.

Table 1:
Default input parameters used in RAmPART.

Description	Symbol	Default Value
Number of sampling iterations	N	10^6
Number of warm-up iterations	N_{warmUp}	10^5
Initial chain positions	$\theta_\ell^{[0]}$	Initial chain positions
Initial covariance matrix	\mathbf{C}	$10^6 \cdot \mathbf{I}$, \mathbf{I} Identity matrix
Number of tempered chains	L	20
Maximum temperature	τ_{max}	2000
Initial inverse temperatures	β	$1 / \left(\frac{L-1-i}{L-1} + \frac{i}{L-1} \cdot \tau_{max}^{(1/1000)} \right)^{1000}$ for $i = 0, \dots, L-1$
Maximum number of regions allowed	R_{max}	10
Number of EM runs on the training sample	N_{rep}	5
Covariance adaptation velocity factor	α	0.51
Temperature adaptation velocity factor 1	ν_τ	10^3
Temperature adaptation velocity factor 2	η_τ	10
Global proposal density contribution factor	p_g	0.5

Table 2:
Output parameters used in RAmPART.

Phase	Description	Symbol
Warm-up	Selected number of GMM modes	R
Warm-up	Last chain position in warm-up phase	$\theta^{[N-1]}$
Warm-up	Selected GMM mode weights	\mathbf{w}
Warm-up	Selected GMM mode centers	μ
Warm-up	Selected GMM mode covariances	Σ
Warm-up	Training sample mean	\mathbf{m}
Warm-up	Training sample covariance	\mathbf{C}
Sampling	Non-tempered parameter chain	$\theta_1^{[0]}, \dots, \theta_1^{[N-1]}$
Sampling	Non-tempered posterior values	$p_1^{[0]}, \dots, p_1^{[N-1]}$

Table 3: The parameter constraints for the 20-dimensional Gaussian mixture distribution.

Parameter Name	θ_{min}	θ_{max}
$\theta_{Gauss,1-20}$	-100	100

Table 4: The parameter constraints for the 20-dimensional blurred ring distribution.

Parameter Name	θ_{min}	θ_{max}
$\theta_{Ring,1-2}$	-200	200
$\theta_{Ring,3-20}$	-20	20

Table 5: The parameter constraints for the model of mRNA transfection.

Parameter Name	θ_{min}	θ_{max}
$\log_{10}(t_0)$	-2	1
$\log_{10}(k_{TL}m_0)$	-5	5
$\log_{10}(\beta)$	-5	5
$\log_{10}(\delta)$	-5	5
$\log_{10}(\sigma)$	-2	2

Table 6: The parameter constraints for the JAK2/STAT5 signaling application.

Parameter Name	θ_{min}	θ_{max}
p_1	-5	5
p_2	-3	6
p_3	-5	5
p_4	-3	6
$[\text{STAT}]_{tot}$	-5	5
sp_1	-5	5
sp_2	-5	5
sp_3	-5	5
sp_4	-5	5
sp_5	-6	5
$O_{[\text{tSTAT}]}$	-5	5
$O_{[\text{pSTAT}]}$	-5	5
$s_{[\text{tSTAT}]}$	-5	5
$s_{[\text{pSTAT}]}$	-5	5
$\sigma_{[\text{pSTAT}]}$	-5	5
$\sigma_{[\text{tSTAT}]}$	-5	5
$\sigma_{[\text{pEpoR}]}$	-5	5

Table 7: Summarized running times per iteration. This includes runs which were not able to converge as well.

	Gaussian Mixture	Blurred Ring	mRNA Transfection	JAK2/STAT5 Signaling
RB-AM	$1.4 \cdot 10^{-1}$	$1.2 \cdot 10^{-1}$	$7.5 \cdot 10^{-2}$	$2.6 \cdot 10^{-1}$
PT	1.4	2.1	1.1	$2.7 \cdot 10^1$
RAmPART	3.0	3.7	1.6	$2.5 \cdot 10^1$

Table 8: Summarized the number of converged runs in each of the examples. Any chain which has a non-vanishing ESS counts as converged. In total, 100 runs were started.

	Gaussian Mixture	Blurred Ring	mRNA Transfection	JAK2/STAT5 Signaling
RB-AM	0	0	0	0
PT	90	12	100	86
RAmPART	95	25	100	99