Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows

Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows (Appendix)

George Papamakarios
University of Edinburgh

David C. Sterratt
University of Edinburgh

Iain Murray
University of Edinburgh

A Detailed description of simulator models used in experiments

A.1 A toy problem with complex posterior

This toy model illustrates that even simple models can have non-trivial posteriors. The model has 5 parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_5)$ sampled from a uniform prior as follows:

$$\theta_i \sim \mathcal{U}(-3,\, 3) \quad \text{for} \quad i = 1, \dots, 5. \tag{1}$$

Given parameters $\boldsymbol{\theta}$, the data $\mathbf{x}$ is generated as follows:

$$\mathbf{m}_{\boldsymbol{\theta}} = (\theta_1, \theta_2) \tag{2}$$
$$s_1 = \theta_3^2 \tag{3}$$
$$s_2 = \theta_4^2 \tag{4}$$
$$\rho = \tanh(\theta_5) \tag{5}$$
$$\mathbf{S}_{\boldsymbol{\theta}} = \begin{pmatrix} s_1^2 & \rho s_1 s_2 \\ \rho s_1 s_2 & s_2^2 \end{pmatrix} \tag{6}$$
$$\mathbf{x}_j \sim \mathcal{N}(\mathbf{m}_{\boldsymbol{\theta}}, \mathbf{S}_{\boldsymbol{\theta}}) \quad \text{for} \quad j = 1, \dots, 4 \tag{7}$$
$$\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_4). \tag{8}$$

The data $\mathbf{x}$ is 8-dimensional. The likelihood is:

$$p(\mathbf{x} \,|\, \boldsymbol{\theta}) = \prod_{j=1}^{4} \mathcal{N}(\mathbf{x}_j \,|\, \mathbf{m}_{\boldsymbol{\theta}}, \mathbf{S}_{\boldsymbol{\theta}}). \tag{9}$$

In our experiments, we took the ground truth parameters to be:

$$\boldsymbol{\theta}^* = (0.7,\ -2.9,\ -1,\ -0.9,\ 0.6), \tag{10}$$

and simulated the model with parameters $\boldsymbol{\theta}^*$ to get observed data $\mathbf{x}_o$.

A.2 M/G/1 queue model

The M/G/1 queue model [8] describes how a server processes a queue of arriving customers. Our experimental setup follows Papamakarios and Murray [6]. There are 3 parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ sampled from a uniform prior as follows:

$$\theta_1 \sim \mathcal{U}(0, 10) \tag{11}$$
$$\theta_2 - \theta_1 \sim \mathcal{U}(0, 10) \tag{12}$$
$$\theta_3 \sim \mathcal{U}(0, 1/3). \tag{13}$$

Let I be the total number of customers, s_i be the time the server takes to serve customer i, a_i be the time customer i arrived, and d_i be the time customer i departed. Take $a_0 = d_0 = 0$. The M/G/1 queue model is described by:

$$s_i \sim \mathcal{U}(\theta_1, \theta_2) \tag{14}$$
$$a_i - a_{i-1} \sim \mathrm{Exp}(\theta_3) \tag{15}$$
$$d_i - d_{i-1} = s_i + \max(0, a_i - d_{i-1}). \tag{16}$$

In our experiments we used $I = 50$. The data $\mathbf{x}$ is 5-dimensional, and is obtained by (a) calculating the 0th, 25th, 50th, 75th and 100th quantiles of the set of inter-departure times $\{d_i - d_{i-1}\}_{1:I}$, and (b) linearly transforming the quantiles to have approximately zero mean and unit covariance matrix. The parameters of the linear transformation were determined by a pilot run. We took the ground truth parameters to be:

$$\boldsymbol{\theta}^* = (1,\ \ 5,\ \ 0.2), \tag{17}$$

and simulated the model with parameters $\boldsymbol{\theta}^*$ to get observed data $\mathbf{x}_o$.

A.3 Lotka–Volterra population model

The Lotka–Volterra model [10] is a Markov jump process describing the evolution of a population of predators interacting with a population of prey, and has four parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_4)$. Let X be the number of predators, and Y be the number of prey. According to the model, the following can take place:

- With rate $\exp(\theta_1) XY$ a predator may be born, increasing X by one.

- With rate $\exp(\theta_2)X$ a predator may die, decreasing X by one.
- With rate $\exp(\theta_3)Y$ a prey may be born, increasing Y by one.
- With rate $\exp(\theta_4)XY$ a prey may be eaten by a predator, decreasing Y by one.

Our experimental setup follows that of Papamakarios and Murray [6]. We used initial populations $X = 50$ and $Y = 100$. We simulated the model using the Gillespie algorithm [2] for a total of 30 time units. We recorded the two populations every 0.2 time units, which gives two timeseries of 151 values each. The data $\mathbf{x}$ is 9-dimensional, and corresponds to the following timeseries features:

- The mean of each timeseries.
- The log variance of each timeseries.
- The autocorrelation coefficient of each timeseries at lags 0.2 and 0.4 time units.
- The cross-correlation coefficient between the two timeseries.

Each feature was normalized to have approximately zero mean and unit variance based on a pilot run. The ground truth parameters were taken to be:

$$\boldsymbol{\theta}^* = (\log 0.01,\ \ \log 0.5,\ \ \log 1,\ \ \log 0.01), \tag{18}$$

and the observed data $\mathbf{x}_o$ were generated from a simulation of the model at $\boldsymbol{\theta}^*$. In our experiments we used two priors: (a) a broad prior defined by:

$$p_{\text{broad}}(\boldsymbol{\theta}) \propto \prod_{i=1}^{4} I(-5 \leq \theta_i \leq 2), \tag{19}$$

and (b) a prior corresponding to the oscillating regime, defined by:

$$p_{\text{osc}}(\boldsymbol{\theta}) \propto \mathcal{N}\big(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^*, 0.5^2\big)\, p_{\text{broad}}(\boldsymbol{\theta}). \tag{20}$$

A.4 Hodgkin–Huxley cortical pyramidal neuron model

In neuroscience, the formalism developed by Hodgkin–Huxley in their classic model of the squid giant axon [4] is used to model many different types of neuron. In our experiments, we used a slightly modified version of a regular-spiking cortical pyramidal cell [7], for which NEURON [1] simulation code is available in ModelDB.[1]

[1] `https://senselab.med.yale.edu/ModelDB/ShowModel.cshtml?model=123623`

The model is formulated as a set of five coupled ordinary differential equations (ODEs) and describes how the electrical potential $V(t)$ measured across the neuronal cell membrane varies over time as a function of current $I_{\text{e}}(t)$ injected through an electrode. In essence, the membrane is a capacitor punctuated by conductances formed by multiple types of ion channel through which currents flow. The currents charge and discharge the membrane capacitance, causing the membrane potential to change, as described by the first ODE, the membrane equation:

$$C_{\text{m}} \frac{\mathrm{d}V}{\mathrm{d}t} = -I_\ell - I_{\text{Na}} - I_{\text{K}} - I_{\text{M}} - I_{\text{e}}. \tag{21}$$

Here, $C_{\text{m}} = 1\,\mu\text{F}\,\text{cm}^{-2}$ is the specific membrane capacitance, and I_ℓ, I_{Na}, I_{K} and I_{M} are the ionic currents flowing through 'leak' channels, sodium channels, potassium delayed-rectifier channels and M-type potassium channels respectively. Each ionic current depends on a conductance that corresponds to how many channels are open, and on the difference between the membrane potential and an equilibrium potential. For example, for the leak current:

$$I_\ell = g_\ell\,(V - E_\ell), \tag{22}$$

where g_ℓ is the leak conductance, and E_ℓ is the leak equilibrium potential. Here the conductance g_ℓ is constant through time, but for the sodium, potassium and M-type channels, the conductances vary over time, as described by the product of a fixed conductance and time-varying state variables:

$$I_{\text{Na}} = \overline{g}_{\text{Na}}\, m^3 h\,(V - E_{\text{Na}}) \tag{23}$$

$$I_{\text{K}} = \overline{g}_{\text{K}}\, n^4\,(V - E_{\text{K}}) \tag{24}$$

$$I_{\text{M}} = \overline{g}_{\text{M}}\, p\,(V - E_{\text{K}}). \tag{25}$$

Here, $\overline{g}_{\text{Na}}$, $\overline{g}_{\text{K}}$ and $\overline{g}_{\text{M}}$ are the per-channel maximum conductances, m, h, n and p are state variables that range between 0 and 1, and E_{Na} and E_{K} are the sodium and potassium reversal potentials. The state variables evolve according to differential equations of the form first introduced by Hodgkin and Huxley [4]:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha_x(V)(1 - x) - \beta_x(V)\, x \quad \text{for } x \in \{m, h, n\} \tag{26}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{p - p_\infty(V)}{\tau_{\text{p}}(V)}, \tag{27}$$

where $\alpha_x(V)$, $\beta_x(V)$, $p_\infty(V)$ and $\tau_{\text{p}}(V)$ are nonlinear functions of the membrane potential. We use the published equations [7] for $\alpha_{\text{m}}(V)$, $\beta_{\text{m}}(V)$, $\alpha_{\text{h}}(V)$, $\beta_{\text{h}}(V)$ and $\alpha_{\text{n}}(V)$, which contain a parameter V_{T}, and the published equations for $p_\infty(V)$ and $\tau_{\text{p}}(V)$, the latter of

which contains a parameter $\tau_{\max}$. We use a generalized version of $\beta_{\mathrm{n}}(V)$:

$$\beta_{\mathrm{n}}(V) = k_{\beta\mathrm{n}1} \exp\left(-\frac{V - V_{\mathrm{T}} - 10}{k_{\beta\mathrm{n}2}}\right) \tag{28}$$

in which $k_{\beta\mathrm{n}1}$ and $k_{\beta\mathrm{n}2}$ are adjustable parameters (rather than $0.5\,\mathrm{ms}^{-1}$ and $40\,\mathrm{mV}$). In order to simulate the model, we use NEURON [1] to solve the ODEs numerically from initial conditions of:

$$m = h = n = p = 0 \quad \text{and} \quad V = -70\,\mathrm{mV}, \tag{29}$$

using the "CNexp" method and a time-step of $25\,\mu\mathrm{s}$. At each time step the injected current I_{e} is drawn from a normal distribution with mean $0.5\,\mathrm{nA}$ and variance σ^2. The duration of the simulation is $100\,\mathrm{ms}$ and the voltage is recorded, which generates a timeseries of 4001 voltage recordings.

Our inference setup follows Lueckmann et al. [5]. There are 12 parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{12})$ to infer, defined as:

$$\begin{aligned}
\theta_1 &= \log(g_\ell) & \theta_7 &= \log(-E_{\mathrm{K}}) \\
\theta_2 &= \log(\overline{g}_{\mathrm{Na}}) & \theta_8 &= \log(-V_{\mathrm{T}}) \\
\theta_3 &= \log(\overline{g}_{\mathrm{K}}) & \theta_9 &= \log(k_{\beta\mathrm{n}1}) \\
\theta_4 &= \log(\overline{g}_{\mathrm{M}}) & \theta_{10} &= \log(k_{\beta\mathrm{n}2}) \\
\theta_5 &= \log(-E_\ell) & \theta_{11} &= \log(\tau_{\max}) \\
\theta_6 &= \log(E_{\mathrm{Na}}) & \theta_{12} &= \log(\sigma).
\end{aligned} \tag{30}$$

The data $\mathbf{x}$ is taken to be 18 features of the voltage timeseries $V(t)$, in particular:

- The mean and log standard deviation of $V(t)$.
- The normalized 3rd, 5th and 7th moments of $V(t)$.
- The logs of the normalized 4th, 6th and 8th moments of $V(t)$.
- The autocorrelation coefficients of $V(t)$ at lags $k \times 2.5\,\mathrm{ms}$ for $k = 1, \ldots, 10$.

The features are linearly transformed to have approximately zero mean and unit covariance matrix; the parameters of the transformation are calculated based on a pilot run. The ground truth parameters $\boldsymbol{\theta}^*$ are taken to be:

$$\begin{aligned}
\theta_1^* &= \log\left(10^{-4}\right) & \theta_7^* &= \log(100) \\
\theta_2^* &= \log(0.2) & \theta_8^* &= \log(60) \\
\theta_3^* &= \log(0.05) & \theta_9^* &= \log(0.5) \\
\theta_4^* &= \log\left(7 \times 10^{-5}\right) & \theta_{10}^* &= \log(40) \\
\theta_5^* &= \log(70) & \theta_{11}^* &= \log(1000) \\
\theta_6^* &= \log(50) & \theta_{12}^* &= \log(1).
\end{aligned} \tag{31}$$

The prior over parameters is:

$$\theta_i \sim \mathcal{U}(\theta_i^* - \log 2,\ \theta_i^* + \log 1.5) \quad \text{for} \quad i = 1, \ldots, 12. \tag{32}$$

Finally, the observed data $\mathbf{x}_o$ were generated by simulating the model at $\boldsymbol{\theta}^*$.

B Full experimental results

In this section, we include the full set of experimental results. For each simulator model, we report:

- The approximate posterior computed by SNL.
- The trade-off between accuracy and simulation cost. This is reported for all methods.
- The full results of the simulation-based calibration test, consisting of one histogram per parameter.
- The distance-based convergence diagnostic, i.e. the distance between simulated and observed data vs the number of rounds. This is reported for SNL, SNPE-A and SNPE-B.
- The goodness-of-fit diagnostic, i.e. the Maximum Mean Discrepancy between simulated data and data generated from the likelihood model, for a given parameter value (we use the true parameters $\boldsymbol{\theta}^*$). We report this for SNL, NL and a baseline Gaussian fit.

B.1 A toy problem with complex posterior

Figure 1 shows the results. The exact posterior $p(\boldsymbol{\theta} \mid \mathbf{x}_o)$ is plotted in Figure 1a. Even though the prior is uniform and the likelihood is Gaussian, the posterior is complex and non-trivial: it has four symmetric modes due to the two squaring operations in Equations (3) and (4), and vertical cut-offs due to the hard constraints imposed by the prior. We can see that the SNL posterior (Figure 1b) approximates the exact posterior well.

B.2 M/G/1 queue model

Figure 2 shows the results. The SNL posterior is shown in Figure 2a. We can see that the posterior is concentrated around the true parameters. Parameter θ_1 is particularly well constrained. From the description of the model in Equations (11)–(13), it directly follows that:

$$\theta_1 \leq \min_i (d_i - d_{i-1}). \tag{33}$$

The data $\mathbf{x}$ is an invertible linear transformation of the quantiles of $\{d_i - d_{i-1}\}_{1:I}$ including the $0th$ quantile, which is precisely equal to $\min_i (d_i - d_{i-1})$. Hence, the data $\mathbf{x}$ imposes a hard constraint on the maximum

possible value of θ_1, as correctly captured by the SNL posterior. On the other hand, the data is less informative about θ_2 and θ_3, hence the parameters are less constrained.

B.3 Lotka–Volterra population model

Figure 3 shows the results. The SNL posterior is shown in Figure 3a; the extent of each plot corresponds to the broad prior $p_{\text{broad}}(\boldsymbol{\theta})$. We can see that the posterior is tightly concentrated around the true parameters, which suggests that the data $\mathbf{x}$ is highly informative about the parameters $\boldsymbol{\theta}$.

B.4 Hodgkin–Huxley cortical pyramidal neuron model

The results are shown in Figure 4. The SNL posterior is shown in Figure 5; the extent of each plot corresponds to the uniform prior. In the SNL posterior, it can be seen that all parameters are clustered around their true values (red dots and lines).

The sodium and equilibrium potentials are relatively tightly clustered, and the potassium equilibrium potential less so:

$$E_{\text{Na}} = 50 \pm 2 \text{ mV} \tag{34}$$

$$E_{\text{K}} = -99 \text{ mV (range } [-121\,\text{mV}, -90\,\text{mV}]) \tag{35}$$

$$E_{\ell} = -70 \pm 4 \text{ mV}. \tag{36}$$

The tight clustering reflects that concentrations, and hence equilibrium potentials, are maintained within a range by neuronal ion exchangers and and glial buffering [9]. Furthermore, when regulation of ion concentration fails, pathological brain states can arise [9]. The longer tail of the potassium equilibrium potential posterior might be due to it having relatively little influence on the mean of $V(t)$, since at lower potentials the potassium conductance $\overline{g}_{\text{K}} n^4$ will be relatively small, so, according to Equation (24), the potassium current will also be small. The quantity V_{T}, which adjusts the threshold of spike initiation, is also fairly tightly controlled, which will tend to keep the firing rate around a constant value.

In contrast to the equilibrium potentials, the conductances vary more, within a factor of 1.8 for g_{ℓ}, and a factor of 3 for $\overline{g}_{\text{Na}}$, $\overline{g}_{\text{K}}$ and $\overline{g}_{\text{M}}$. Moreover, $\overline{g}_{\text{Na}}$ and $\overline{g}_{\text{K}}$ are correlated, which is consistent with their opposing depolarizing and hyperpolarizing influences on the membrane potential. A higher sodium conductance could lead to the cell being hyper-excitable, but this should be counteracted by a greater potassium conductance. This allows for their wide range, and is consistent with the biological evidence for diverse but correlated sets of channel conductances underlying particular activity patterns [3]. In contrast, $\overline{g}_{\text{M}}$ appears to have relatively little influence over the output, and is not correlated with any other parameters.

The parameter $\tau_{\max}$, which also relates to the M-type potassium channel, also has little effect. Further simulations of other neuron types could be undertaken to see if these parameters are generally loosely constrained, which could then lead to experimentally testable predictions about the density and variability of M-type conductances.

The parameters $k_{\beta \text{n}1}$ and $k_{\beta \text{n}2}$ also have relatively wide ranges, and there is a weak correlation between the two. Increasing $k_{\beta \text{n}1}$ effectively increases the half-activation voltage of potassium conductances, and increasing $k_{\beta \text{n}2}$ makes the slope of transition less pronounced. The lack of posterior at high $k_{\beta \text{n}1}$ (high threshold) and low $k_{\beta \text{n}2}$ (sharper transition) might cause the neuron not to repolarize quickly enough, and hence be hyper-excitable.

Finally, we note that the posterior computed by SNL is qualitatively consistent with the posterior reported by Lueckmann et al. [5] in Figure G.2 of their article.

References

[1] T. Carnevale and M. Hines. *The NEURON Book*. Cambridge University Press, 2006.

[2] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

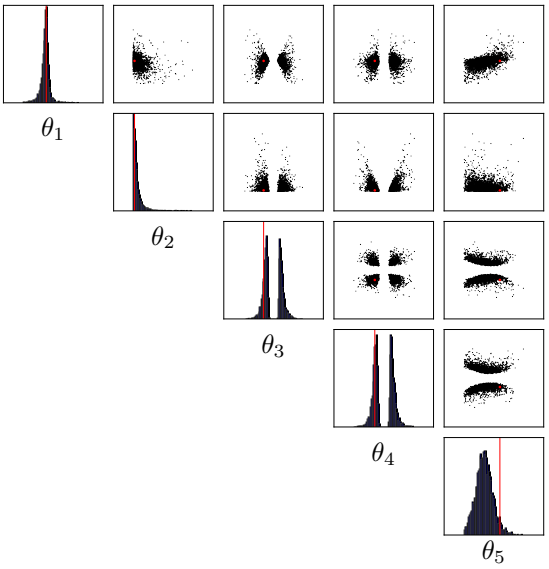
[3] M. S. Goldman, J. Golowasch, E. Marder, and L. F. Abbott. Global structure, robustness, and modulation of neuronal models. *The Journal of Neuroscience*, 21(14):5229–5238, 2001.

[4] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117:500–544, 1952.

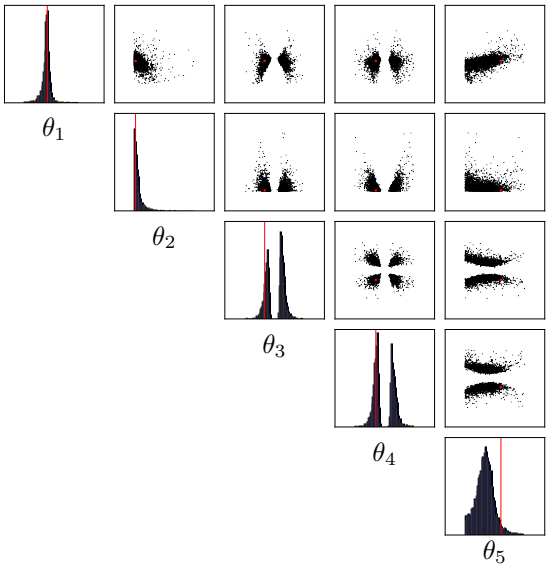
[5] J.-M. Lueckmann, P. J. Goncalves, G. Bassetto, K. Öcal, M. Nonnenmacher, and J. H. Macke. Flexible statistical inference for mechanistic models of neural dynamics. *Advances in Neural Information Processing Systems 30*, 2017.

[6] G. Papamakarios and I. Murray. Fast ϵ-free inference of simulation models with Bayesian conditional density estimation. *Advances in Neural Information Processing Systems 29*, 2016.

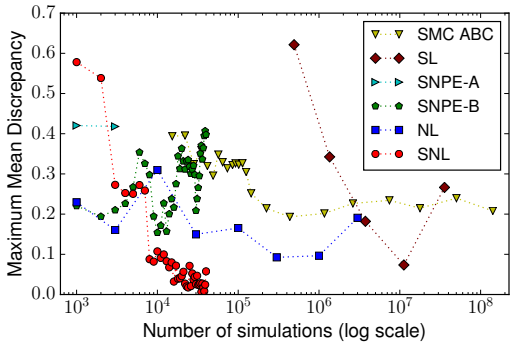
[7] M. Pospischil, M. Toledo-Rodriguez, C. Monier, Z. Piwkowska, T. Bal, Y. Frégnac, H. Markram, and A. Destexhe. Minimal Hodgkin–Huxley type models for different classes of cortical and thalamic neurons. *Biological Cybernetics*, 99(4):427–441, 2008.



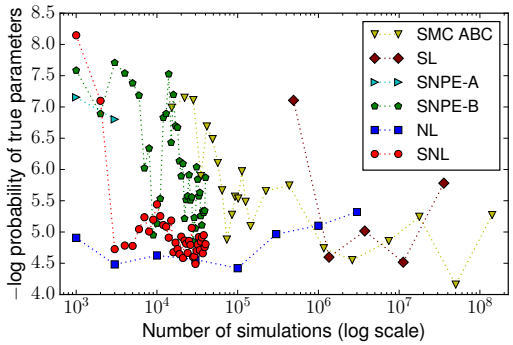
(a) MCMC samples from exact posterior. True parameters indicated in red.



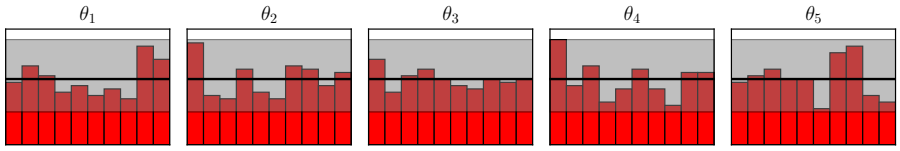
(b) MCMC samples from SNL posterior. True parameters indicated in red.



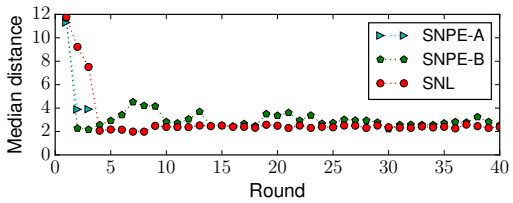
(c) Maximum Mean Discrepancy vs simulation cost.



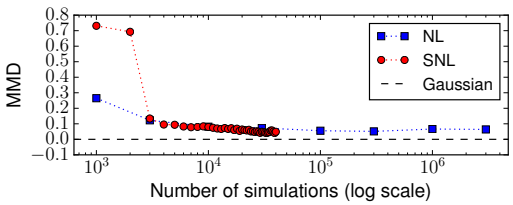
(d) Minus log probability of true parameters vs simulation cost.



(e) Calibration test for SNL. Histogram outside gray band indicates poor calibration.

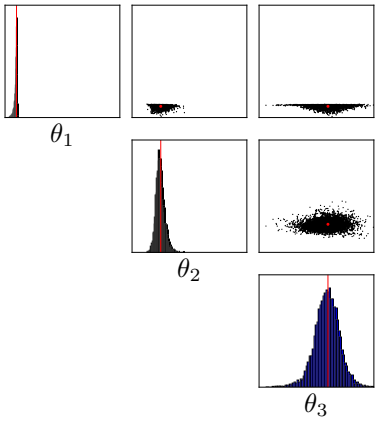


(f) Median distance from simulated to observed data.

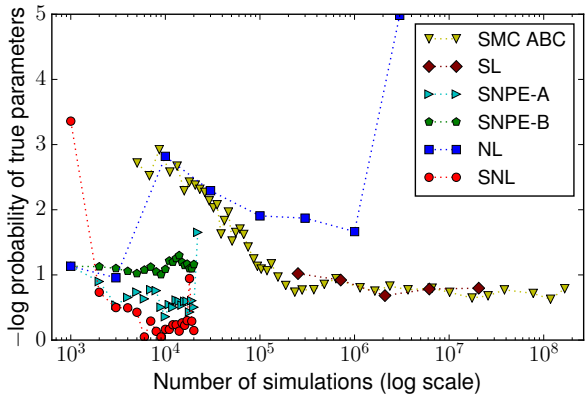


(g) Likelihood goodness-of-fit vs simulation cost, calculated at true parameters.

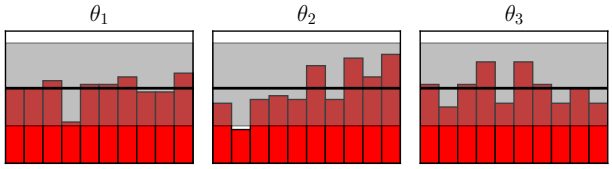
Figure 1: A toy model with complex posterior.



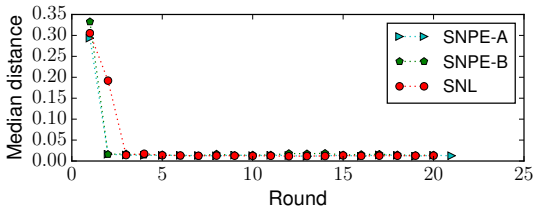
(a) MCMC samples from SNL posterior. True parameters indicated in red.



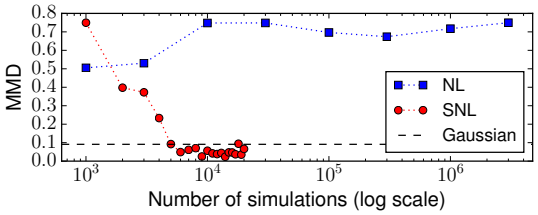
(b) Minus log probability of true parameters vs simulation cost.



(c) Calibration test for SNL. Histogram outside gray band indicates poor calibration.

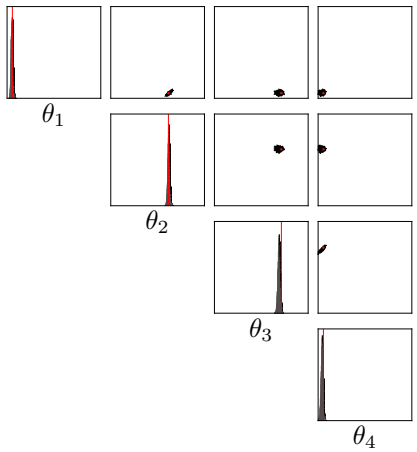


(d) Median distance from simulated to observed data.

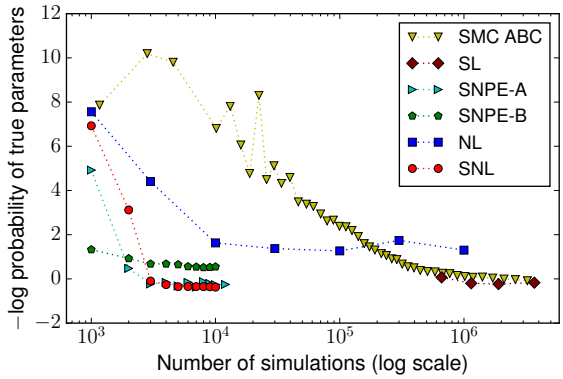


(e) Likelihood goodness-of-fit vs simulation cost, calculated at true parameters.

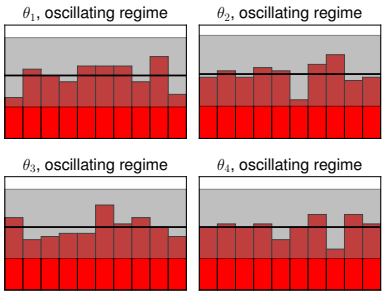
Figure 2: M/G/1 queue model.



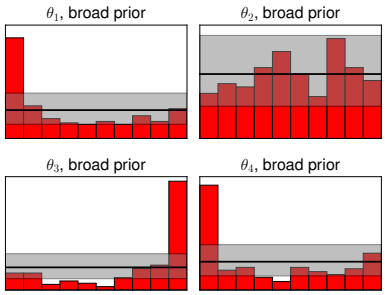
(a) MCMC samples from SNL posterior. True parameters indicated in red.



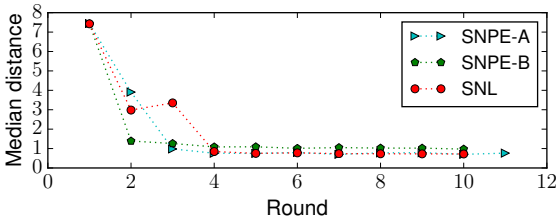
(b) Minus log probability of true parameters vs simulation cost.



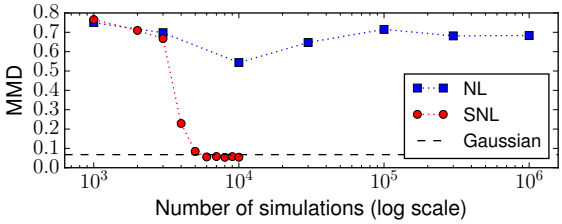
(c) Calibration test for SNL, oscillating regime.



(d) Calibration test for SNL, broad prior.

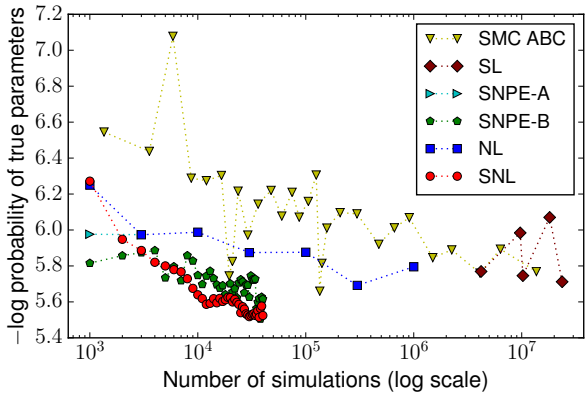


(e) Median distance from simulated to observed data.

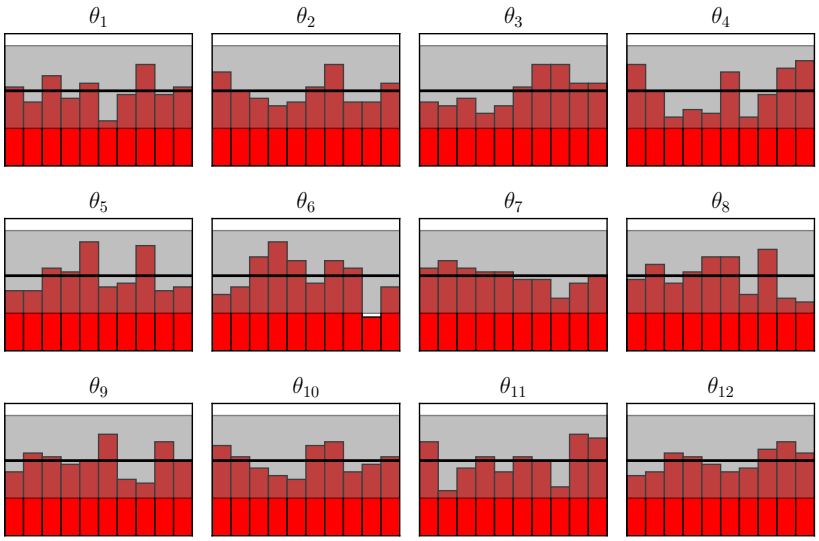


(f) Likelihood goodness-of-fit vs simulation cost, calculated at true parameters.

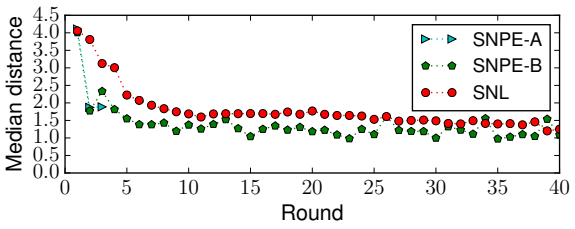
Figure 3: Lotka–Volterra population model.



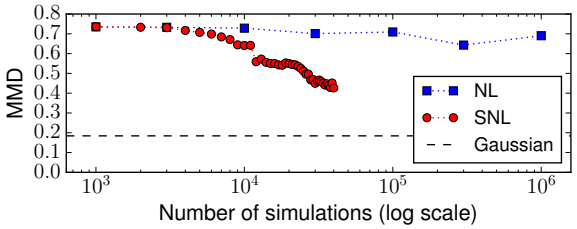
(a) Minus log probability of true parameters vs simulation cost.



(b) Calibration test for SNL. Histogram outside gray band indicates poor calibration.



(c) Median distance from simulated to observed data.



(d) Likelihood goodness-of-fit vs simulation cost, calculated at true parameters.

Figure 4: Hodgkin–Huxley neuron model.

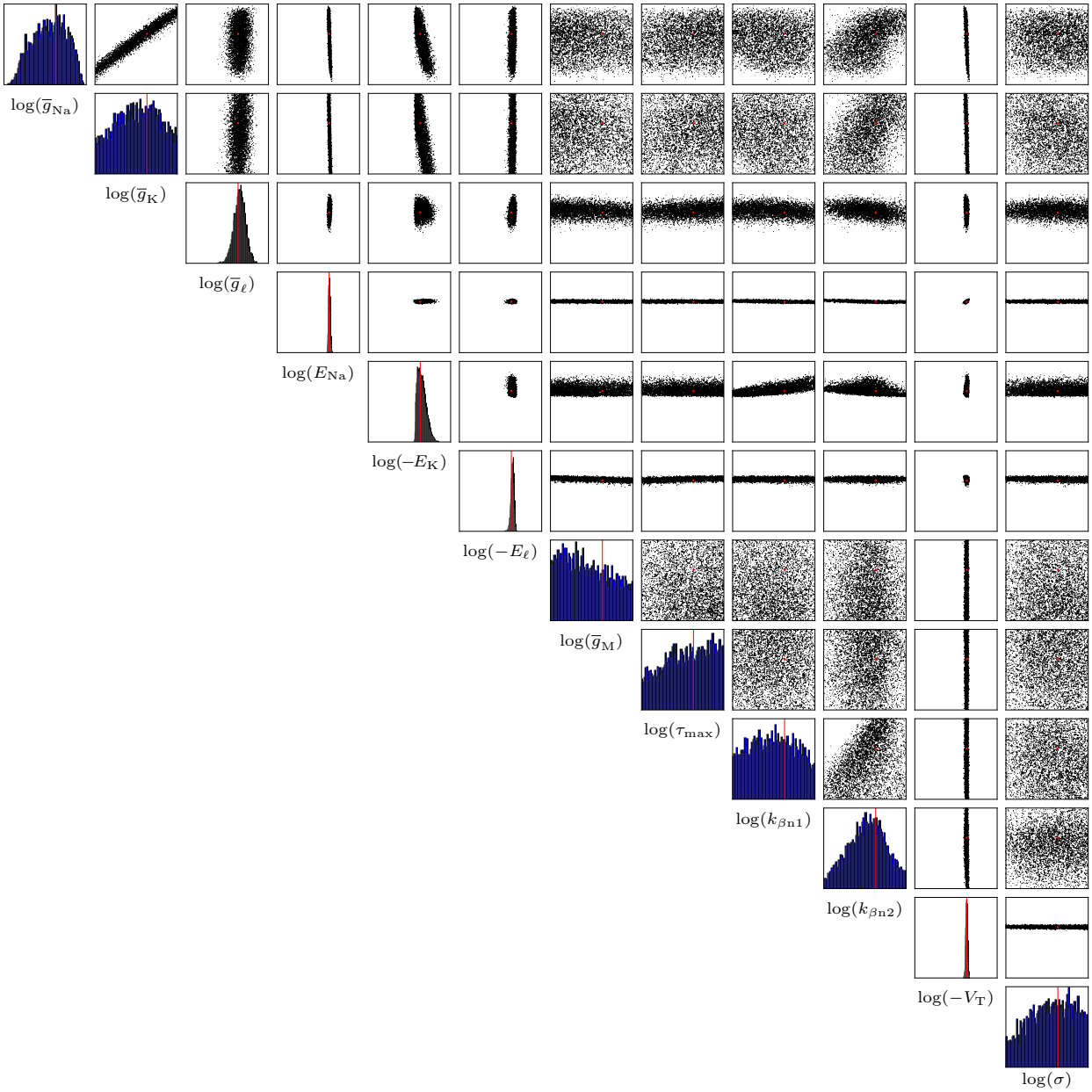


Figure 5: Hodgkin–Huxley model: MCMC samples from SNL posterior. True parameters $\boldsymbol{\theta}^*$ are indicated in red. The range of each histogram is $[\theta_i^* - \log 2,\ \theta_i^* + \log 1.5]$ for $i = 1, \ldots, 12$. Compare with Figure G.2 by Lueckmann et al. [5].

[8] A. Y. Shestopaloff and R. M. Neal. On Bayesian inference for the M/G/1 queue with efficient MCMC sampling. *arXiv:1401.5548*, 2014.

[9] G. G. Somjen. Ion regulation in the brain: implications for pathophysiology. *The Neuroscientist*, 8(3):254–267, 2002.

[10] D. J. Wilkinson. *Stochastic Modelling for Systems Biology, Second Edition*. Taylor & Francis, 2011.