Estimation of autocorrelation timescales with Approximate Bayesian Computations

Roxana Zeraati [1,2], Tatiana A. Engel [3,*,†], Anna Levina [4,2,5,*,‡]

[1]Graduate School of Neural Information Processing, University of Tübingen, Germany

[2]Max Planck Institute for Biological Cybernetics, Tübingen, Germany

[3] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

[4] Department of Computer Science, University of Tübingen, Germany

[5] Bernstein Center for Computational Neuroscience Tübingen, Tübingen, Germany

[*] These authors contributed equally to this work

August 11, 2020

Abstract

Timescales characterize the pace of change for many dynamic processes in nature: radioactive decay, metabolization of substances, memory decay in neural systems, and epidemic spreads. Measuring timescales from experimental data can reveal underlying mechanisms and constrain theoretical models. Timescales are usually estimated by fitting the autocorrelation of sample time-series with exponential decay functions. We show that this standard procedure often fails to recover the correct timescales, exhibiting large estimation errors due to a statistical bias in autocorrelations of finite data samples. To overcome this bias, we develop a method using adaptive Approximate Bayesian Computations. Our method estimates the timescales by fitting the autocorrelation of sample data with a generative model based on a mixture of Ornstein-Uhlenbeck processes. The method accounts for finite sample size and noise in data and returns a posterior distribution of timescales quantifying the estimation uncertainty. We demonstrate how the posterior distribution can be used for model selection to compare alternative hypotheses about the dynamics of the underlying process. Our method accurately recovers the correct timescales on synthetic data from various processes with known ground truth dynamics. We illustrate its application to electrophysiological recordings from the primate cortex.

1 Introduction

Dynamic changes in many stochastic processes occur over typical periods known as timescales. The timescales of different processes in nature range broadly from milliseconds in protein folding [1, 2] or neural signalling [3–5], to days in spreading of infectious diseases [6–8], to years in demographic changes of populations [9, 10], and up to millennia in the climate change [11, 12].

[†]engel@cshl.edu

[‡]anna.levina@uni-tuebingen.de

Accurate measurements of timescales from experimental data are necessary to uncover mechanisms controlling the dynamics of underlying processes and to determine the operating regime of these dynamics [13]. Inferring the operating regime, e.g., how close the system is to a critical point, can reveal general principles that explain the collective behavior of complex systems [14–16] and predict their future evolution [17–19]. Thus experimentally measured timescales constrain theoretical models and enable accurate predictions in practical applications.

The timescales of a stochastic process are defined by the exponential decay rates of its autocorrelation function. Accordingly, timescales are usually estimated by fitting the autocorrelation of a sample time-series with exponential decay functions [3, 13, 20–23]. However, the shape of autocorrelations can be altered by many factors unrelated to the dynamics of the processes under study. For example, autocorrelations of *in vivo* neural activity can contain components arising from a specific trial structure of a behavioral task or slow drifts in the average activity. To correct for these irrelevant factors, several techniques were developed based on data resampling, in particular, trial shuffling and spike jittering methods [24–27]. These methods remove from the autocorrelation the average autocorrelation of surrogate data, which are designed to match the irrelevant factors in the real data but are otherwise random. The success of these methods critically depends on the ability to construct the appropriate surrogate data that exactly reproduce the irrelevant factors.

The shape of empirical autocorrelations is also affected by the fact that sample autocorrelation is a biased estimator. The values of sample autocorrelation computed from a finite time series systematically deviate from the true autocorrelation [28–33]. This statistical bias deforms the shape of empirical autocorrelations and hence can affect the estimates of timescales obtained by direct fitting of exponential functions. The magnitude of the bias mainly depends on the length of the sample time-series, but also on the value of the true autocorrelation at each time-lag. The expected value and variance of the autocorrelation bias can be derived analytically in some simple cases, such as a Markov process with a single timescale [28, 31]. However, the analytical derivation is not tractable for more general processes that involve multiple timescales or have additional temporal structure. Moreover, since the bias depends on the true autocorrelation itself, which is unknown, it cannot be corrected by constructing appropriate surrogate data as in shuffling or jittering methods. How autocorrelation bias affects the estimation of timescales was not studied systematically.

We show that the autocorrelation bias due to a finite sample size significantly affects the timescales estimated by the direct fitting of exponential functions, resulting in large systematic errors. To correct for this bias, we develop a method based on adaptive Approximate Bayesian Computations (aABC) that estimates timescales by fitting the autocorrelation with a generative model. ABC is a family of likelihood-free inference algorithms for estimating model parameters when the likelihood function cannot be calculated analytically [34]. The aABC algorithm approximates the multivariate posterior distribution of parameters of a generative model using population Monte-Carlo sampling [35]. Our generative model is based on a mixture of Ornstein-Uhlenbeck processes—one for each estimated timescale—which have exponentially decaying autocorrelations. The generative model can be further augmented with the desired noise model (e.g., a spike generation process) and additional temporal structure to match the statistics of the data. Our method accurately recovers the correct timescales from finite data samples for various synthetic processes with known ground truth dynamics. We demonstrate how the inferred posterior distributions can be used for model selection to compare alternative hypotheses about the dynamics of the underlying process. To illustrate an application of our method, we estimate timescales of ongoing spiking activity in the primate visual cortex during

a behavioral task. The general method presented here can be adapted to various types of data and can find broad applications in neuroscience, epidemiology, physics, and other fields.

2 Results

2.1 Bias in timescales estimated by exponential fitting

Timescales of a stochastic process $A(t')$ are defined by the exponential decay rates of its autocorrelation function. The autocorrelation is the correlation between the values of the process at two time points separated by a time lag t. For stationary processes, the autocorrelation function only depends on the time lag:

$$AC(t) = \frac{\mathrm{E}\left[(A(t') - \mu)\,(A(t'+t) - \mu)\right]_{t'}}{\sigma^2}. \tag{1}$$

Here μ and σ^2 are, respectively, the mean and variance of the process, which are constant in time, and $\mathrm{E}[\cdot]_{t'}$ is the expectation over t'. Different normalizations of autocorrelation are also used in literature, but our results do not depend on a specific choice of normalization.

In experiments or simulations, the autocorrelation needs to be estimated from a finite sample of empirical data. A data sample from the process $A(t')$ constitutes a finite time-series measured at discrete times t'_i ($i = 1 \ldots N$, where N is the length of the time-series). For example, the sample time-series can be spike-counts of a neuron in discrete time bins, or a continuous voltage signal measured at a specific sampling rate. Accordingly, the sample autocorrelation is defined for a discrete set of time lags t_j. For empirical data, the true values of μ and σ are unknown. Hence, several estimators of the sample autocorrelation were proposed, which use different estimators for the sample mean $\hat{\mu}$ and sample variance $\hat{\sigma}^2$ [28, 29]. One possible choice is:

$$\hat{AC}(t_j) = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{N-j} \left(A(t'_i) - \hat{\mu}_1(j)\right)\left(A(t'_{i+j}) - \hat{\mu}_2(j)\right), \tag{2}$$

with the sample variance $\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(A(t'_i)^2 - \frac{1}{N^2}(\sum_{i=1}^{N} A(t'_i))^2)$ and two different sample means $\hat{\mu}_1(j) = \frac{1}{N-j}\sum_{i=1}^{N-j} A(t'_i)$, $\hat{\mu}_2(j) = \frac{1}{N-j}\sum_{i=j+1}^{N} A(t'_i)$. However, for any of these choices, the sample autocorrelation is a biased estimator: for a finite length time-series the values of the sample autocorrelation systematically deviate from the true autocorrelation [28–33]. This statistical bias deforms the shape of the sample autocorrelation and therefore may affect the estimation of timescales by direct fitting of exponential decay functions.

To investigate how the autocorrelation bias affects the timescales estimated by exponential fitting, we tested how accurately this procedure recovers the correct timescales on synthetic data with a known ground truth. We generated synthetic data from several stochastic processes for which the autocorrelation function can be computed analytically. The exponential decay rates of the analytical autocorrelation provide the ground-truth timescales. Each synthetic dataset consisted of 500 independent realizations of the process (i.e. trials) with fixed duration. Such trial-based data are typical in neuroscience but usually with a smaller number of trials. We computed the sample autocorrelation for each trial using Eq. 2, averaged them to reduce the noise, and then fitted the average autocorrelation with the correct analytical functional form to estimate the timescale parameters. We repeated the entire procedure 500 times to obtain a distribution of estimated timescales for multiple independent realizations of the data.

We considered three ground truth processes which differed in the number of timescales, additional temporal structure and noise. First, we used an Ornstein–Uhlenbeck (OU) process (Fig. 1A):

$$\dot{A}_{\mathrm{OU}}(t') = -\frac{A_{\mathrm{OU}}(t')}{\tau} + \sqrt{2D}\xi(t'), \tag{3}$$

where $\xi(t')$ is a Gaussian white noise with zero mean, and the diffusion parameter D sets the variance $\mathrm{Var}[A_{\mathrm{OU}}(t')] = D\tau$ [36, 37]. The autocorrelation of the OU process is an exponential decay function [38]

$$AC_{\mathrm{OU}}(t) = e^{-t/\tau}. \tag{4}$$

Accordingly, the parameter τ provides the ground truth timescale.

Second, we used a linear mixture of an OU process and an additional oscillatory component (Fig. 1B):

$$A(t') = \sqrt{(1-c_1)}A_{\mathrm{OU}}(t') + \sqrt{2c_1}\sin(\phi + 2\pi f t'). \tag{5}$$

Here the coefficient $c_1 \in [0, 1]$ sets the relative weights of two components without changing the total variance of the process, and the phase ϕ is drawn on each trial from a uniform distribution on $[0, 2\pi]$. This process resembles the firing rate of a neuron modulated by a slow oscillation [39, 40]. The autocorrelation of this process is given by

$$AC(t) = (1-c_1)e^{-t/\tau} + c_1\cos(2\pi f t), \tag{6}$$

hence the ground truth timescale is defined by the OU parameter τ. When fitting this functional form to the sample autocorrelation, we assumed that the frequency f is known and only estimated the timescale τ and the coefficient c_1.

Finally, we used an inhomogeneous Poisson process with the instantaneous rate modeled as a linear mixture of two OU processes with different timescales (Eq. 11, Fig. 1C). The OU mixture was shifted, scaled and rectified to produce the Poisson rate with the desired mean and variance (Eq. 15). Similar doubly-stochastic processes are often used to model spiking activity of neurons [41]. The total variance of this process consists of two contributions: the rate variance and the Poisson process variance (Eq. 17) [42]. We simulated this process in discrete time bins t'_i by sampling event-counts from a Poisson distribution with the given instantaneous rate (see Methods 4.1.3). The autocorrelation at all lags $t_j (j > 0)$ arises only from the autocorrelation of the rate, since the Poisson process has no temporal correlations. The drop of autocorrelation between t_0 and t_1 mainly reflects the Poisson-process variance, and for all non-zero lags $t_j (j > 0)$ the autocorrelation is given by

$$AC(t_j) = \frac{\sigma^2 - \mu}{\sigma^2}\left(c_1 e^{-t_j/\tau_1} + (1-c_1)e^{-t_j/\tau_2}\right). \tag{7}$$

Here τ_1 and τ_2 are the ground-truth timescales defined by the parameters of two OU processes, c_1 is the mixing coefficient, and μ and σ^2 are the mean and variance of the event-counts, respectively. The variance of event-counts is related to the variance of the rate via Eq. 18. We fitted this analytical formula to the sample autocorrelation starting from the lag t_1. We assumed that the mean and variance of the rate are known and only fitted the timescales τ_1 and τ_2 and the coefficient c_1.

For all three processes, the sample autocorrelation exhibits a negative bias: the values of the sample autocorrelation are systematically below the ground-truth autocorrelation function (Fig. 1, left). This bias is clearly visible in the logarithmic-linear scale, where the ground-

truth exponential decay turns into a straight line. The sample autocorrelation deviates from the straight line and even becomes systematically negative at intermediate lags (hence not visible in logarithmic scale) for processes with a strictly positive ground-truth autocorrelation (Fig. 1A,C). The deviations from the ground truth are larger when the timescales are longer or when multiple timescales are involved. The negative bias decreases for longer trial durations (Fig. 1, left inset), but it is still substantial for realistic trial durations such as in neuroscience data.

Due to the negative bias, a direct fit of the sample autocorrelation with the correct analytical function cannot recover the ground-truth timescales systematically underestimating them (Fig. 1, middle, right). When increasing the duration of each trial, the timescales obtained from the direct fits become closer to the ground-truth values (Supplementary Fig. 1). This observation indicates that timescales estimated from datasets with different trial durations cannot be directly compared, as differences in the estimation bias may result in misleading interpretation. Thus direct fitting of a sample autocorrelation, even when a correct analytical form of autocorrelation is known, is not a reliable method for measuring timescales in experimental data.

2.2 Estimating timescales by fitting generative models with Approximate Bayesian Computations

Since direct fitting of the sample autocorrelation cannot estimate timescales reliably, we developed an alternative method based on fitting the sample autocorrelation with a generative model. Using a generative model with known ground-truth timescales, we can generate synthetic data matching the essential statistics of the observed data, i.e. with the same duration and number of trials, mean and variance. Hence the sample autocorrelations of the synthetic and observed data will be affected by a similar statistical bias when their shapes match. As a generative model, we chose a linear mixture of OU processes—one for each estimated timescale—which if necessary can be augmented with additional temporal structure (e.g., oscillations) and noise. The advantage of using a mixture of OU processes for modelling the observed data is that the analytical autocorrelation function of this mixture explicitly defines the timescales. We set the number of components in the generative model in accordance with our hypothesis about the shape of the autocorrelation in the data, e.g., the number of timescales, additional temporal structure and noise; see Methods 4.1. Then, we optimize parameters of the generative model to match the shape of the autocorrelation between the synthetic and observed data. The timescales of the optimized generative model provide an approximation for the timescales in the observed data, without the bias corruption.

For complex generative models, calculating the likelihood function can be computationally expensive or even intractable. Therefore, we optimize the generative model parameters using adaptive Approximate Bayesian Computations (aABC) [35] (Fig. 2). aABC is an iterative algorithm that minimizes the distance between summary statistics of synthetic and observed data. We choose the autocorrelation as the summary statistics and define the distance d between the autocorrelations of synthetic and observed data as

$$d(t_m) = \frac{1}{m} \sum_{j=0}^{m} \left(AC_{\text{observed}}(t_j) - AC_{\text{synthetic}}(t_j) \right)^2 , \tag{8}$$

where t_m is the maximum time-lag considered in computing the distance. Due to stochasticity

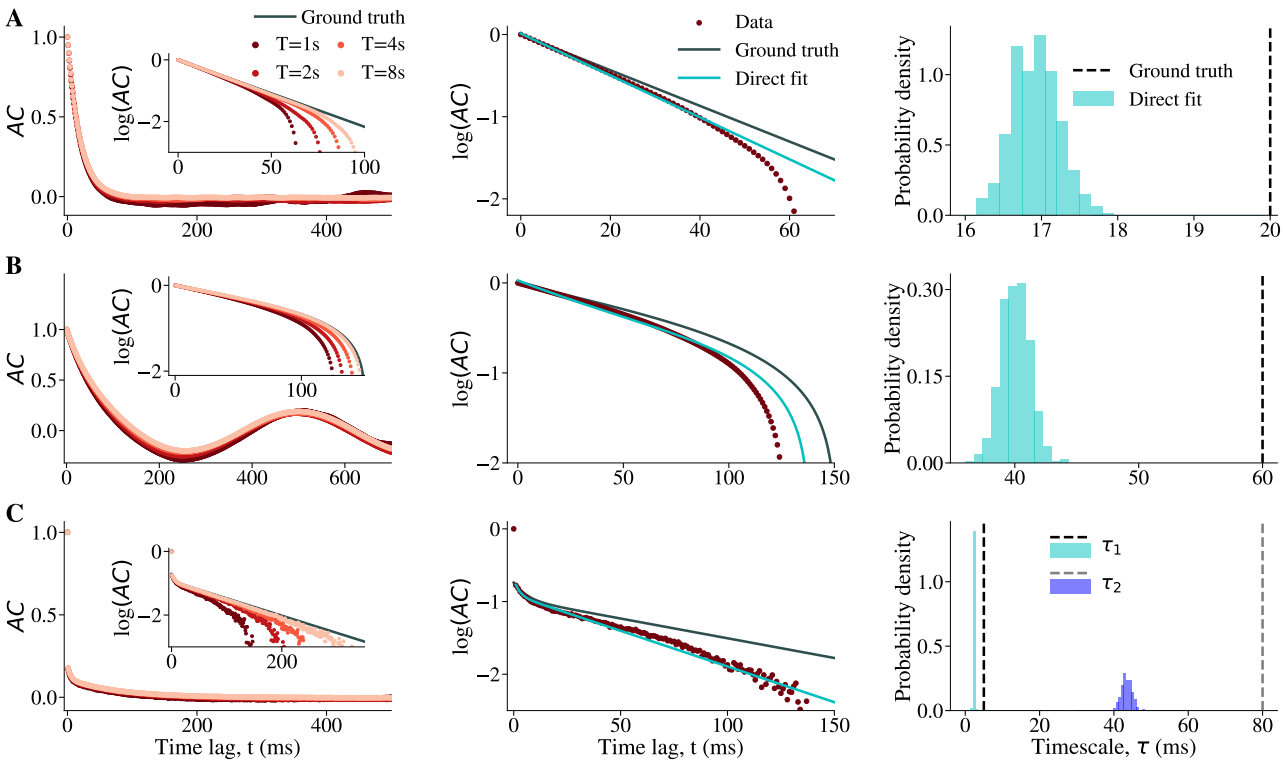


Fig. 1: **Bias in sample autocorrelations and in timescales estimated by direct exponential fitting.** **(A)** Data are generated from an OU process with the timescale $\tau = 20$ ms. **Left**: Sample autocorrelation (colored dots) systematically deviates from the ground-truth autocorrelation (grey line). The shape of sample autocorrelation depends on the time-series duration (T) approaching the ground truth with increasing T (inset: close up in logarithmic-linear scale). **Middle:** Direct fitting of the analytical autocorrelation function (cyan) to sample autocorrelation (brown, $T = 1$ s) does not recover the ground-truth autocorrelation shape (grey). **Right:** The ground-truth timescales largely deviate from the distribution of timescales estimated by direct exponential fitting across 500 independent realizations of the same process. **(B)** Same as A for data from a linear mixture of an OU process with the timescale $\tau = 60$ ms and an oscillation with frequency $f = 2$ Hz (Eq. 5). **(C)** Same as A for data from an inhomogeneous Poisson process with the instantaneous rate generated from a linear mixture of two OU processes with timescales $\tau_1 = 5$ ms and $\tau_2 = 80$ ms. All simulation parameters are provided in Methods 4.4.

of the observed data, point estimates (i.e. assigning a single value to a timescale) are not reliable as different realizations of the same stochastic process lead to slightly different autocorrelation shapes (Supplementary Fig. 2). aABC overcomes this problem by estimating the joint posterior distribution of the generative model parameters, which quantifies the estimation uncertainty. The uncertainty depends on the maximum time-lag t_m. Smaller t_m excludes noisy values in the tail of autocorrelation, which can produce narrower posterior distributions in fewer iterations. On the other hand, larger t_m may be necessary for estimating long timescales.

The aABC algorithm operates iteratively (Fig. 2). First, we choose a multivariate prior distribution over the parameters of the generative model and set an initial error threshold ε at a rather large value. On each iteration, we draw a sample of the parameter-vector θ. On the first iteration, θ is drawn directly from the prior distribution. On subsequent iterations, θ is drawn from a proposal distribution defined based on the prior distribution and the parameter samples accepted on the previous iteration. We use the generative model with the sample parameters θ to generate synthetic data. If the distance d between the autocorrelations of the observed

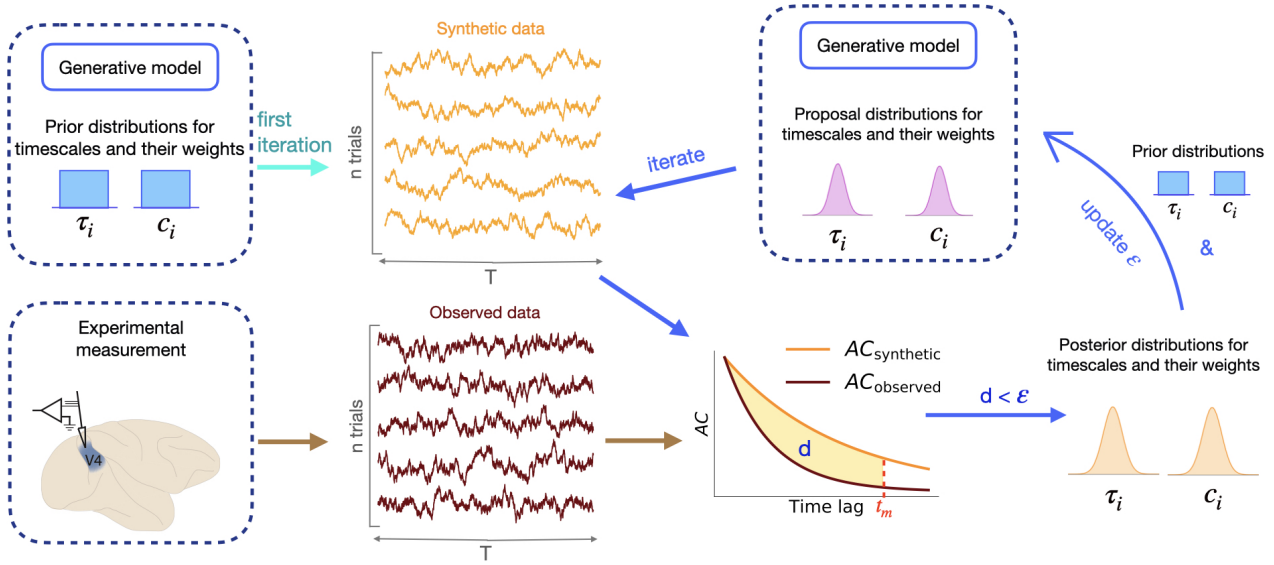


Fig. 2: **Estimation of timescales with adaptive Approximate Bayesian Computations.** aABC estimates timescales by fitting the sample autocorrelation of observed data with a generative model. At the first iteration of the algorithm, parameters of the generative model are drawn from the multivariate prior distribution, e.g., a uniform distribution (upper left). Synthetic data are generated from the generative model with these parameters. If the distance d between the autocorrelations of synthetic and observed data is smaller than a specific error threshold ε, these parameters are added to the multivariate posterior distribution (lower right). In the subsequent iterations, new parameters are drawn from a proposal distribution which is computed based on the posterior distribution of the previous iteration and the initial prior distribution (upper right, see Methods 4.2).

and synthetic data (Eq. 8) is smaller than the error threshold, the sample θ is accepted. We then repeatedly draw new parameter samples and evaluate d until a fixed number of parameter samples are accepted. For each iteration, the fraction of accepted samples out of all drawn parameter samples is recorded as the acceptance rate $accR$. Next, the proposal distribution is updated using the samples accepted on the current iteration, and the new error threshold is set at the first quartile of the distance d for the accepted samples. The iterations continue until the acceptance rate reaches a specified value. The last set of accepted parameter samples is treated as an approximation for the posterior distribution (see Methods 4.2). To visualize the posterior distribution of a parameter (e.g., a timescale), we marginalize the obtained multivariate posterior distribution over all other parameters of the generative model.

We tested our method on the same synthetic data from the processes with known ground-truth timescales that we used to demonstrate the bias of direct exponential fitting (Fig. 3 cf. Fig. 1). For all three processes, the shape of the sample autocorrelation of the observed data is accurately reproduced by the autocorrelation of synthetic data generated using the maximum *a posteriori* (MAP) estimate of parameters from the multivariate posterior distribution (Fig. 3, left). The posterior distributions inferred by the aABC algorithm include the ground-truth timescales (Fig. 3, middle). The variance of posterior distributions quantifies the uncertainty of estimates. In our simulations, the number of trials controls the signal to noise ratio in sample autocorrelation, and consequently the width of the posterior distributions (Supplementary Fig. 3). The width of the posterior distributions also depends on the final acceptance rate (Fig. 3, right). A smaller final acceptance rate results in a narrower posterior distribution but requires larger number of iterations.

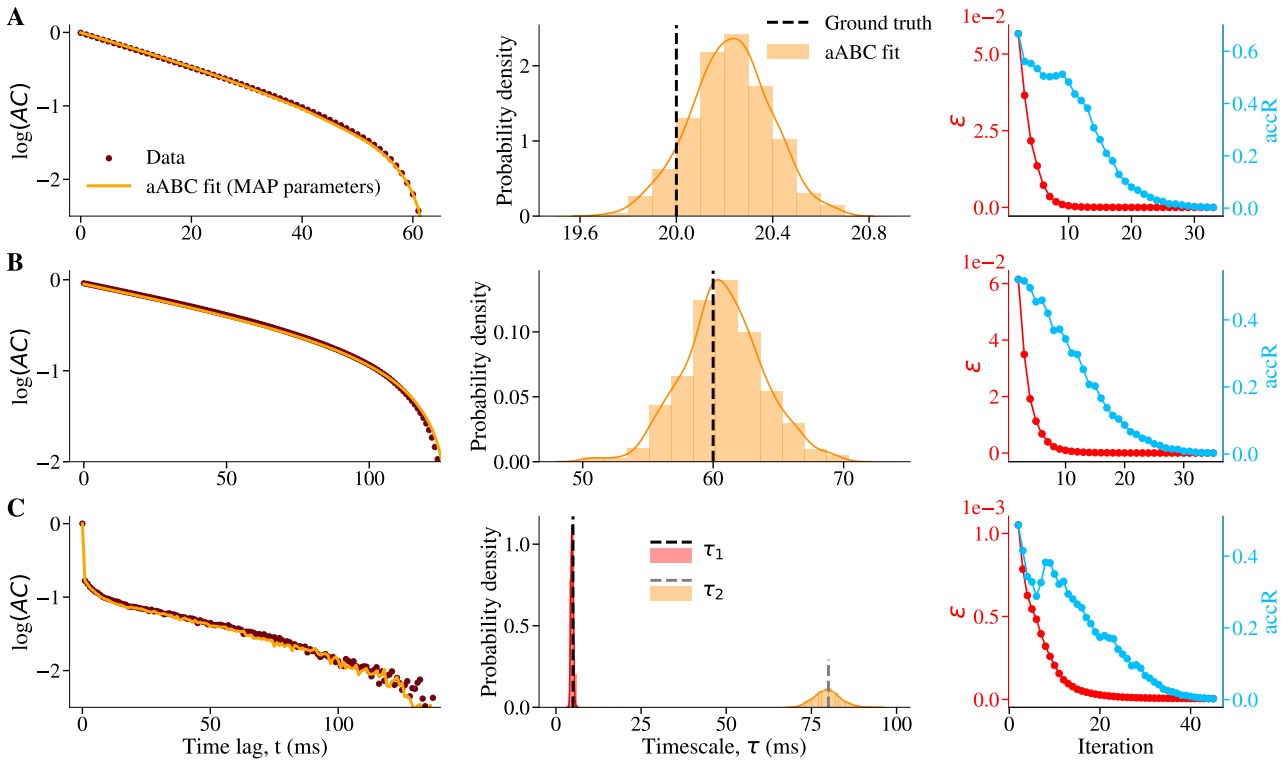


Fig. 3: **The aABC algorithm accurately recovers the ground-truth timescales and quantifies the estimation uncertainty.** The same synthetic data as in Fig. 1 for $T = 1$ s. (**A**) Data are generated from an OU process with the timescale $\tau = 20$ ms. **Left:** The shape of the data autocorrelation (brown) is accurately reproduced by the autocorrelation of synthetic data from the generative model with the MAP estimate parameters (orange). $\tau_{\mathrm{MAP}} = 20.2$ ms. **Middle:** The marginal posterior distributions (histograms overlaid with Gaussian kernel smoothing) include the ground-truth timescales. The width of the posteriors indicates the estimation uncertainty. **Right:** The convergence of the aABC algorithm is defined based on the acceptance rate $accR$ (blue), which decreases together with the error threshold ε (red) over the iterations of the algorithm. Data are plotted from the second iteration. Initial error threshold for all fits was set to 1. **(B)** Same as A for data from a linear mixture of an OU process with the timescale $\tau = 20$ ms and an oscillation with frequency $f = 2$ Hz (Eq. 5). $\tau_{\mathrm{MAP}} = 60.5$ ms, **(C)** Same as A for data from an inhomogeneous Poisson process with two timescales $\tau_1 = 5$ ms and $\tau_2 = 80$ ms. $\tau_{1,\mathrm{MAP}} = 4.7$ ms, $\tau_{2,\mathrm{MAP}} = 80$ ms. See Methods 4.4 for other parameters.

2.3 Estimating the timescale of activity in a branching network model

So far, we demonstrated that our aABC method accurately recovers the ground-truth timescales within the same model class, i.e. when the generative model and the process that produced the observed data are the same. However, the timescale inference based on OU processes is broadly applicable to data outside this model class, i.e. when the mechanism that generated the exponential decay of autocorrelation in the data is different from an OU process.

We tested our inference method on data from a different model class with known ground-truth autocorrelation function. Specifically, we applied our method to estimate the timescale of the global activity in a branching network model [43–45]. A branching network consists of k interconnected binary neurons, each described by the state variable $x_i \in \{0, 1\}$, where $x_i = 1$ indicates that neuron i is active, and 0 that it is silent (Fig. 4A). We considered a fully-connected network. Each active neuron can activate other neurons with the probability

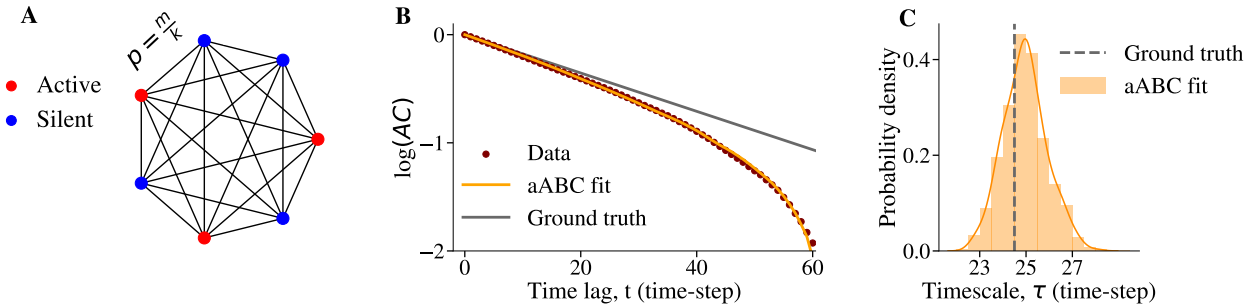


Fig. 4: **The aABC inference with a generative model based on the OU process recovers the ground-truth timescale of activity in a branching network.** (**A**) A schematic of a fully connected branching network with binary neurons. Branching parameter ($m = 0.96$) and number of neurons ($k = 10^4$) define the probability of activity propagation. Each neuron receives an external input with probability $h = 10^{-3}$. (**B**) The shape of the sample autocorrelation of simulated activity in the branching network (brown) deviates from the ground truth (grey), but is accurately reproduced by the autocorrelation of synthetic data generated from a one-timescale OU process with the MAP estimate timescale from the aABC method. $\tau_{\text{MAP}} = 24.9$. The data contained 100 trials of 500 time-steps. (**C**) The aABC posterior distribution includes the ground-truth timescale. $\tau_{\text{ground truth}} = 24.5$, $\tau_{\text{aABC}} = 24.9 \pm 0.9$ (mean $\pm$ std). Fitting parameters are provided in Methods 4.4.

$p = m/k$ and then, if not activated by other neurons, it becomes inactive again in the next time-step. Additionally, at every time-step, each neuron can be activated with a probability h by an external input. For a small input strength h, the state of the network's dynamics is governed by a branching parameter m ($m = 1$ corresponds to the critical state). The autocorrelation function of the global activity $A(t') = \sum_i x_i(t')$ in this network is known analytically $AC(t_j) = \exp(t_j \ln(m))$ [13]. Thus the ground-truth timescale of this activity is given by $\tau = -1/\ln(m)$.

We simulated the binary network model to generate the time-series data of the global activity. We then used aABC to fit the sample autocorrelation of these data with a one-timescale OU process as the generative model. The inferred posterior distribution is centered on the theoretically predicted timescale of the branching network, and the MAP estimate parameters accurately reproduce the shape of the sample autocorrelation of the network activity (Fig. 4B,C). These results show that our method can be used to estimate timescales in diverse types of data with different underlying mechanisms.

2.4 Model selection with Approximate Bayesian Computations

For experimental data, the correct generative model is usually unknown. For example, it may not be obvious *a priori* how many timescales should the generative model contain. Several alternative hypotheses may be plausible, and we need a procedure to select which one is more likely to explain the observed data. Assuming that autocorrelation function is a sufficient summary statistics for estimating timescales, we can use aABC to approximate the Bayes factor for selecting between alternative models [46–49].

We developed a procedure for selecting between two alternative models M_1 and M_2 based on their aABC fits. Models are compared using a goodness of fit measure that describes how well each model fits the data. The goodness of fit can be measured by the distance d between the autocorrelations of synthetic and observed data (Eq. 8). However, d is a noisy measure because of the finite sample-size of synthetic data and the uncertainty in the model parameters. Therefore, we compare the distributions of distances generated by two models with parameters

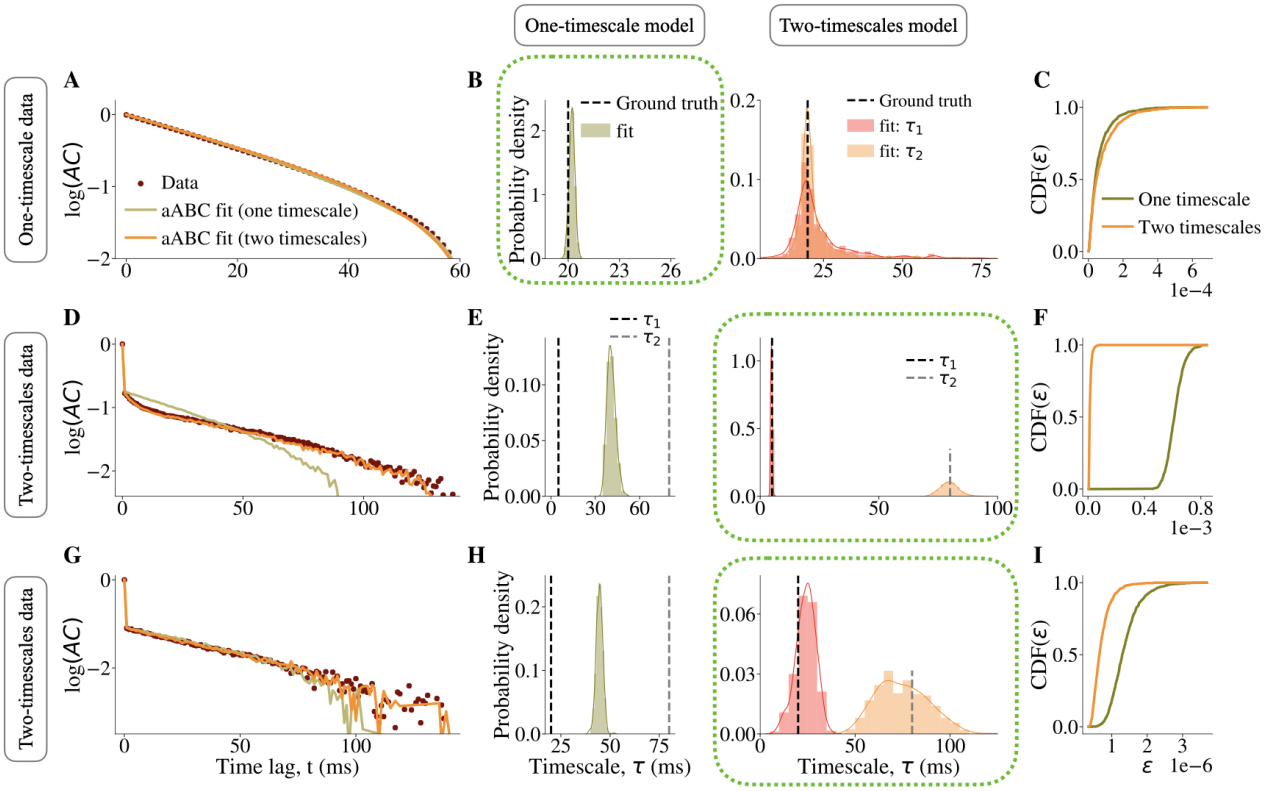


Fig. 5: **Model selection with ABC.** (**A-C**) Data are generated from a one-timescale OU process with $\tau = 20$ ms. (**A**) The data autocorrelation (brown) is fitted with a one-timescale (olive, M_1) and two-timescale (orange, M_2) generative models (sample autocorrelations for MAP estimate parameters are shown). (**B**) Marginal posterior distribution of the timescale estimated by the one-timescale model includes the ground truth and has small variance (**left**). Marginal posterior distributions of two timescales estimated by the two-timescale model overlap and have large variance (**right**). (**C**) Cumulative distributions of distances d for two models, $\mathrm{CDF}_{M_i}(\varepsilon)$. Since $\mathrm{CDF}_{M_2}(\varepsilon) < \mathrm{CDF}_{M_1}(\varepsilon)$ for all ε, the one-timescale model is selected (green box in B). (**D-F**) Data are generated from an inhomogeneous Poisson process with two timescales: $\tau_1 = 5$ ms, $\tau_2 = 80$ ms. (**D**) Same format as A. (**E**) Marginal posterior distributions for two timescales estimated by the two-timescale model include the ground-truth (**right**). Marginal posterior distributions for the timescale estimated by the one-timescale model falls in between the ground-truth timescales (**left**). (**F**) Cumulative distributions of distances d for two models. Since $\mathrm{CDF}_{M_2}(\varepsilon) > \mathrm{CDF}_{M_1}(\varepsilon)$ for all ε, the two-timescale model is selected (green box in E). (**G-I**) Same as D-F, but for $\tau_1 = 20$ ms, $\tau_2 = 80$ ms. For all fits, $\mathrm{CDF}_{M_i}(\varepsilon)$ are computed from 1000 samples. Simulation and fitting parameters are provided in 4.4.

uncertainty defined by their posterior distributions. To approximate the distributions of distances, we generate multiple samples of synthetic data from each model with parameters drawn from its posterior distribution and compute the distance d for each sample. If the distributions of distances are significantly different (e.g., can be tested using Wilcoxon rank-sum test), then we continue with the model selection, otherwise the summary statistics does not allow us to distinguish these models.

For selecting between M_1 and M_2, we use the distributions of distances to estimate the Bayes factor, which is the ratio of marginal likelihoods of the two models [50, 51]. Assuming both models are *a priori* equally probable ($p(M_1) = p(M_2)$), the Bayes factor can be approximated

using the models acceptance rates for a specific error threshold ε [46, 48, 52]

$$B_{21}(\varepsilon) = \frac{accR_{M_2}(\varepsilon)}{accR_{M_1}(\varepsilon)}. \tag{9}$$

$B_{21}(\varepsilon) > 1$ indicates that the model M_2 is more likely to explain the observed data and vice versa. To eliminate the dependence on a specific error threshold, we compute the acceptance rates and the Bayes factor with a varying error threshold. For each error threshold ε, the acceptance rate is given by the cumulative distribution function of the distances $\mathrm{CDF}_{M_i}(\varepsilon) = p_{M_i}(d < \varepsilon) = accR_{M_i}(\varepsilon)$ $(i = 1, 2)$. Hence, the ratio between CDFs of two models gives the value of the Bayes factor for every error threshold $B_{21d}(\varepsilon) = \mathrm{CDF}_{M_2}(\varepsilon)/\mathrm{CDF}_{M_1}(\varepsilon)$. We select the model M_2 if $B_{21d}(\varepsilon) > 1$, i.e. if $\mathrm{CDF}_{M_2}(\varepsilon) > \mathrm{CDF}_{M_1}(\varepsilon)$ for all ε, and vice versa.

We evaluated our model selection method using synthetic data from three example processes with known ground truth, so that the correct number of timescales is known. Specifically, we used an OU process with a single timescale (Fig. 5A) and two different examples of an inhomogeneous Poisson process with two timescales. In the first example of an inhomogeneous Poisson process, the ground-truth timescales were well separated, so that the shape of the data autocorrelation suggested that the underlying process had multiple timescales (Fig. 5D). In the second example of an inhomogeneous Poisson process, we chose the ground-truth timescales to be more similar, so that a simple visual inspection of the data autocorrelation could not clearly suggest the number of timescales in the underlying process (Fig. 5G). For all three example processes, we fitted the data with one-timescale (M_1) and two-timescale (M_2) generative models using aABC and selected between these models using Bayes factors. The one- and two-timescale models were based on a single OU process or a linear mixture of two OU processes, respectively. For the data from inhomogeneous Poisson processes, the generative model also incorporated an inhomogeneous Poisson noise.

For the example OU process with a single timescale, the one- and two-timescale models fitted the shape of the data autocorrelation almost equally well (Fig. 5A). The marginal posterior distributions of two timescales estimated by the two-timescale model are not significantly different from each other (Fig. 5B), which indicates that the one-timescale model possibly better describes the data. To select between two models, we compare the cumulative distributions of distances $\mathrm{CDF}_{M_i}(\varepsilon)$ (Fig. 5C). Although the two-timescales model has more parameters, it has significantly larger distances than the one-timescale model (Wilcoxon rank-sum test, $P = 0.002$, mean $d_{M_1} = 6 \times 10^{-5}$, mean $d_{M_2} = 8 \times 10^{-5}$). The two-timescales model has a larger average distance because its posterior distribution has larger variance (larger uncertainty), which leads to a higher possibility to sample a combination of parameters with a larger distance. Since $\mathrm{CDF}_{M_2}(\varepsilon) < \mathrm{CDF}_{M_1}(\varepsilon)$ (i.e. $B_{21}(\varepsilon) < 1$), the one-timescale model is preferred over the two-timescale model, in agreement with the ground-truth generative process.

For both example inhomogeneous Poisson processes with two timescales, the shape of the data autocorrelation is better matched by the two-timescale than by the one-timescale model (the difference is subtle for the second example, Fig. 5D,G). The marginal posterior distributions of two timescales estimated by the two-timescale model are clearly separated and include the ground-truth values, whereas the timescale estimated by the one-timescale model is in between the two ground-truth values (Fig. 5E,H). The two-timescale model has significantly smaller distances (Wilcoxon rank-sum test, Fig. 5F: $P < 10^{-10}$, mean $d_{M_1} = 6 \times 10^{-4}$, mean $d_{M_2} = 1.5 \times 10^{-5}$; Fig. 5I: $P < 10^{-10}$, mean $d_{M_1} = 10^{-6}$, mean $d_{M_2} = 7 \times 10^{-7}$). Since $\mathrm{CDF}_{M_2}(\varepsilon) > \mathrm{CDF}_{M_1}(\varepsilon)$ (i.e. $B_{21d}(\varepsilon) > 1$) for all error thresholds, the two-timescale model provides a better

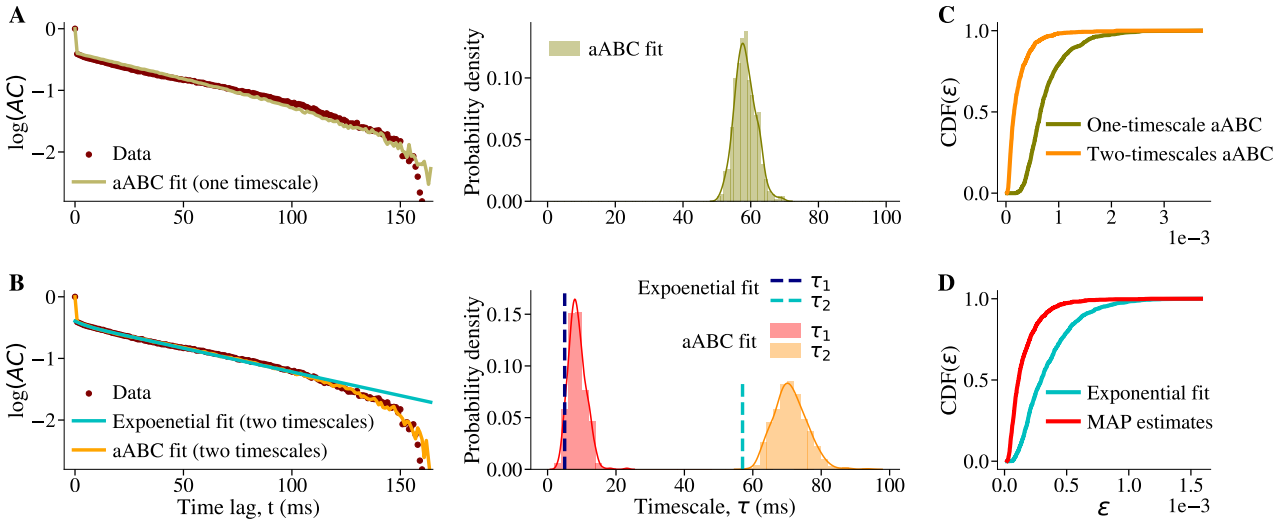


Fig. 6: **Estimating timescales of ongoing neural activity and comparing hypotheses about their number with aABC.** (**A**) **Left:** The autocorrelation of neural spike-counts (brown) is fitted with a one-timescale doubly-stochastic model using aABC (olive, sample autocorrelation for MAP estimate parameters). **Right:** Posterior distribution of the timescale. $\tau_{\text{MAP}} = 58$ ms. (**B**) **Left:** Same data as in A is fitted directly with a double exponential function (cyan) and with a two-timescale doubly-stochastic model using aABC (orange, sample autocorrelation for MAP estimate parameters). **Right:** Timescales estimated by the direct exponential fit (τ_1 - blue, τ_2 -cyan) and marginal posterior distributions of timescales inferred with aABC (τ_1 - red, τ_2 - orange). $\tau_{1,\text{exp}} = 5$ ms, $\tau_{2,\text{exp}} = 57$ ms, $\tau_{1,\text{MAP}} = 8$ ms, $\tau_{2,\text{MAP}} = 70$ ms. (**C**) Cumulative distribution of distances d for one-timescale (M_1) and two-timescale (M_2) models. Since $\text{CDF}_{M_2}(\varepsilon) > \text{CDF}_{M_1}(\varepsilon)$ for all ε, the two-timescale model is selected. (**D**) Cumulative distributions of distances d between the autocorrelation of neural data and synthetic data from the two-timescale doubly-stochastic model with parameters either from the direct fit (cyan) or MAP estimate with aABC (red). The MAP parameters have smaller distances, i.e. describe the autocorrelation of neural data more accurately than the direct fit. Fitting parameters are provided in Methods 4.4.

description of the data for both examples, in agreement with the ground truth. Thus, our method selects the correct generative model even for a challenging case where the shape of the data autocorrelation does not suggest the existence of multiple timescales.

2.5 Estimating timescales of ongoing neural activity

To illustrate an application of our method to experimental data, we inferred the timescales of ongoing spiking activity in the primate visual cortex. The spiking activity was recorded from the visual area V4 with a 16-channels micro-electrode array [53]. During recordings, a monkey was fixating a central dot on a blank screen for 3 s on each trial. To estimate the timescales, we pooled the spiking activity recorded across all channels and computed the spike-count autocorrelation of the pooled activity with a bin-size of 1 ms.

Previously, the autocorrelation of neural activity in several brain regions was modeled as an exponential decay with a single timescale [3]. To determine whether a single timescale is sufficient to describe the temporal dynamics of neural activity in our data, we compared the one-timescale (M_1) and two-timescale (M_2) models and selected the model that better describes the data. As a generative model for neural spike-counts, we used a doubly-stochastic process [41, 42], where spike-counts are generated from an instantaneous firing rate modelled as one OU process

(M_1) or a mixture of two OU processes (M_2). To account for the non-Poisson statistics of the spike-generation process, we sampled spike-counts from a gamma distribution [54] (see Methods 4.1.3). We fitted both models with the aABC algorithm and selected between the models using Bayes factors (Fig. 6A-C). The two-timescale model provided a better description for the data, since it had smaller distances and $\mathrm{CDF}_{M_2}(\varepsilon) > \mathrm{CDF}_{M_1}(\varepsilon)$ for all the error thresholds (Fig. 6C, Wilcoxon rank-sum test, $P < 10^{-10}$, mean $d_{M_1} = 8 \times 10^{-4}$, mean $d_{M_2} = 2 \times 10^{-4}$).

We further compared our method with a direct exponential fit of the sample autocorrelation which is usually employed to infer the timescales of neural activity [3, 13, 20–23]. We fitted the sample autocorrelation with a double exponential function

$$AC(t) = c_1 e^{-t/\tau_1} + (1 - c_1) e^{-t/\tau_2} \tag{10}$$

and compared the result with the two-timescale aABC fit (Fig. 6B). Similar to what we observed with synthetic data, the direct exponential fit produced timescales that were systematically smaller than the MAP estimates with the aABC fit. Since the ground-truth timescales are not available for biological data, we used a sampling procedure to evaluate whether the data is better described by the timescales from the direct fit or from the MAP estimates with the aABC fit. We generated multiple samples of synthetic data using the two-timescale doubly-stochastic generative model with the parameters from either the direct fit or MAP estimates from aABC. For each sample, we measured the distance d between the autocorrelation of synthetic and neural data to obtain the distribution of distances for both types of fits (Fig. 6D). The distances were significantly smaller for synthetic data generated with the parameters from the MAP estimate with aABC than from the direct exponential fit (Wilcoxon rank-sum test, $P < 10^{-10}$, mean distance of MAP parameters 10^{-4}, mean distance of exponential fit parameters 3×10^{-4}). Our method performs better because it accounts for the bias in the autocorrelation shape which is ignored by the direct fit. These results suggest that our method estimates the timescales of neural activity more accurately than a direct exponential fit and, moreover, allows for comparing alternative hypotheses about the underlying dynamics.

3 Discussion

We demonstrated that direct exponential fitting often fails to recover the correct timescales due to a statistical bias in the shape of sample autocorrelation. To overcome this problem, we developed a method based on adaptive Approximate Bayesian Computations. Our method fits the sample autocorrelation of the data with a generative model which is based on a mixture of Ornstein-Uhlenbeck processes (one for each estimated timescale) and can incorporate additional temporal structure and noise. Our method infers a posterior distribution of timescales consistent with the observed data. The width of the posterior distribution depends on the noise in the data and quantifies the estimation uncertainty, e.g., it can be used as a confidence interval. The posterior distributions can be used for model selection to compare alternative hypotheses about the dynamics of the underlying process. Our method is suitable for comparing timescales between datasets with different durations and between processes that exhibit different temporal dynamics (e.g., oscillations).

The statistical bias in sample autocorrelation arises primarily due to the deviation of the sample mean from the ground-truth mean. If the ground-truth mean of the process is known, then using the true mean instead of the sample mean for computing the autocorrelation largely eliminates the bias. In this case, direct fitting of exponential decay functions can produce

satisfactory estimates, but our method additionally provides confidence intervals. When the true mean is unknown but it is known that it is the same across all trials, it is beneficial to estimate a single sample mean from the whole dataset instead of estimating the mean for trials individually. However, this assumption does not always hold. For example, the mean of spiking neural activity can change across trials because of changes in animal's arousal state. If the assumption of a constant mean is violated in the data, estimating the sample mean from the whole dataset leads to strong distortions of the autocorrelation shape introducing additional slow timescales [55].

One of the design choices in our method is the definition of distances between the autocorrelation of the synthetic and observed data. Here we used linear differences, but it is also possible to compute distances in the logarithmic-linear scale. This latter choice allows us to better resolve the behavior in the tail of the autocorrelation at large time-lags. However, the tail is also subject to the largest noise, and emphasis on the tail slows down the convergence of the aABC algorithm.

We validated our method using data from synthetic processes, where the exponential decay rate of autocorrelation is known analytically by design. We further illustrated our method using spiking neural activity data, where the underlying ground-truth process is unknown. In this case, the posterior distributions estimated by our method can be used to approximate the Bayes factor for selecting between different generative models that represent alternative hypotheses about neural dynamics. Since the Bayes factor takes the model complexity into account, it can be used for comparing models with different number of parameters, e.g., to infer the number of timescales in the data.

Our method reliably estimates correct timescales when the standard direct exponential fitting fails, but this advantage comes with a price of higher computational costs for our method. Thus, when long time-series data are available and statistical bias does not corrupt the results, the direct exponential fit of sample autocorrelation may be preferred. To test whether this is the case, a generative model (e.g., based on a mixture of OU processes) with parameters obtained from the direct exponential fit can be used to generate multiple samples of synthetic data each with the same amount of data as in the experiment. For each sample of synthetic data, timescales can be estimated again by direct exponential fitting and the obtained distribution of timescales can be compared to the original timescales estimated from the experimental data. If there is no significant difference, the standard direct exponential fit may be sufficiently accurate.

The general approach of inferring timescales with aABC based on OU processes can be adapted to various types of data and different generative models. Our method can find broad applications in different fields including neuroscience, epidemiology, and physics. This general approach is particularly favorable for data organized in short trials or trials of different durations, when standard exponential fitting is unreliable, and it allows for comparing timescales between datasets with different trial durations and data amount.

4 Methods

4.1 Generative models

We used several generative models based on a linear mixture of OU processes—one for each estimated timescale—sometimes augmented with additional temporal structure (e.g., oscillations) and noise.

4.1.1 Ornstein–Uhlenbeck process with multiple timescales

We define an OU process with multiple timescales $A(t')$ as a linear mixture of OU processes $A_k(t')$ with timescales τ_k, $k \in \{1, \ldots, n\}$, zero mean and unit variance:

$$A(t') = \sum_{k=1}^{n} \sqrt{c_k} A_k(t'), \quad \sum_{k=1}^{n} c_k = 1, \quad c_k \in [0, 1]. \tag{11}$$

Here n is the number of timescales in the mixture and c_k are the mixing coefficients. We simulate each OU process A_k by iterating its time-discrete version using the Euler scheme [38]

$$A_k(t'_{i+1}) = \left(1 - \frac{\Delta t'}{\tau_k}\right) A_k(t'_i) + \sqrt{2 D_k \Delta t'} \eta_k(t'_i), \tag{12}$$

where $\Delta t' = t'_{i+1} - t'_i$ is the discretization time-step and $\eta_k(t'_i)$ is a random number generated from a normal distribution. We set the unit variance for each OU process $\mathrm{Var}(A_k) = D_k \tau_k = 1$ by fixing $D_k = 1/\tau_k$. The parameter-vector θ for a linear mixture of n OU processes consists of $2n - 1$ values: n timescales τ_k and $n - 1$ coefficients c_k (see Eq. 11).

We match the mean and variance of the multi-timescale OU process to the sample mean $\hat{\mu}$ and sample variance $\hat{\sigma}^2$ of the observed data using a linear transformation:

$$A_{\text{trans}}(t') = \hat{\sigma} A(t') + \hat{\mu}. \tag{13}$$

We use the process $A_{\text{trans}}(t')$ as a generative model for data fitting and hypothesis testing (Methods 4.2).

4.1.2 Multi-timescale Ornstein–Uhlenbeck process with an oscillation

To obtain a generative model with an oscillation, we add to a multi-timescale OU process (Eq. 11) an oscillatory component with the weight c_{k+1}:

$$A(t') = \sum_{k=1}^{n} \sqrt{c_k} A_k(t') + \sqrt{2 c_{k+1}} \sin(\phi + 2\pi f t'), \quad \sum_{k=1}^{n+1} c_k = 1, \quad c_k \in [0, 1]. \tag{14}$$

Here we assume that the frequency f is known and only fit the weight c_{k+1}. In applications, the frequency f can be estimated from the power spectrum of data time-series, or it can also be fitted with aABC as an additional parameter. For each trial, we draw the phase ϕ independently from a uniform distribution on $[0, 2\pi]$. We use the linear transformation Eq. 13 to match the mean and variance of this generative process to the observed data.

4.1.3 Doubly stochastic process with multiple timescales

The doubly stochastic process with multiple timescales is generated in two steps: first generating time-varying rate and then generating event-counts from this rate. To generate the time-varying rate, we scale, shift and rectify a multi-timescale OU process (Eq. 11) using the transformation

$$A_{\text{trans}}(t') = \max\left(\sigma' A(t') + \mu', 0\right). \tag{15}$$

The resulting rate $A_{\text{trans}}(t')$ is non-negative and for $\mu' \gg \sigma$ it has the mean $\mathrm{E}[A_{\text{trans}}(t'_i)] \approx \mu'$ and variance $\mathrm{Var}[A_{\text{trans}}(t'_i)] \approx \sigma'^2$. We then draw event-counts s for each time-bin $[t'_i, t'_{i+1}]$ from

an event-count distribution $p_{\text{count}}(s|\lambda(t'_i))$, where $\lambda(t'_i) = A_{\text{trans}}(t'_i)\Delta t'$ is the mean event-count and $\Delta t' = t'_{i+1} - t'_i$ is the bin size (in our simulations $\Delta t' = 1$ ms). A frequent choice of $p_{\text{count}}(s|\lambda(t'_i))$ is a Poisson distribution

$$p_{\text{count}}(s|\lambda(t'_i)) = \frac{(\lambda(t'_i))^s}{s!} e^{-\lambda(t'_i)}, \quad \lambda(t'_i) = A_{\text{trans}}(t'_i)\Delta t', \tag{16}$$

which results in an inhomogeneous Poisson process.

To match the mean and variance of the doubly stochastic process to the observed data, we need to estimate the mean rate μ' and both the variance of the rate σ'^2 and the variance of the event-count distribution $\sigma^2_{s|\lambda(t'_i)}$. According to the law of total expectation, the mean rate $\mu' = \hat{\lambda}/\Delta t'$, where $\hat{\lambda}$ is the sample mean of the observed event-counts. According to the law of total variance [42], the total variance of event-counts σ^2 arises from two contributions: the variance of the rate and the variance of the event-count distribution:

$$\sigma^2 = \text{Var}[\lambda(t'_i)] + \text{E}[\sigma^2_{s|\lambda(t'_i)}] = (\Delta t')^2 \sigma'^2 + \text{E}[\sigma^2_{s|\lambda(t'_i)}]. \tag{17}$$

For the Poisson distribution, the variance of the event-count distribution is equal to its mean: $\sigma^2_{s|\lambda(t'_i)} = \lambda(t'_i)$. However, the condition of equal mean and variance does not always hold in experimental data [54]. Therefore, we also use other event-count distributions, in particular a Gaussian and gamma distribution. We define α as the variance over mean ratio of the event-count distribution $\alpha = \sigma^2_{s|\lambda(t'_i)}/\lambda(t'_i)$. For the Poisson distribution, $\alpha = 1$ always holds. For other distributions, we assume that α is constant (i.e. does not depend on the rate). With this assumption, the law of total variance Eq. 17 becomes

$$\sigma^2 = (\Delta t')^2 \sigma'^2 + \alpha \Delta t' \mu', \tag{18}$$

where $\mu' = \text{E}[A_{\text{trans}}(t'_i)]$ is the mean rate. From Eq. 18 we find the rate variance

$$\sigma'^2 = \frac{1}{(\Delta t')^2} \left(\hat{\sigma}^2 - \alpha \hat{\lambda} \right), \tag{19}$$

where $\hat{\sigma}^2$ is the sample variance of event-counts in the observed data. We find that with the correct values of μ', σ'^2 and α, both the Gaussian and gamma distributions of event-counts produce comparable estimates of timescales (Supplementary Fig. 4).

To calculate σ'^2 with Eq. 19, we first need to estimate α from the observed data. We estimate α from the drop of autocorrelation of event-counts between the time-lags t_0 and t_1. Since event-counts in different time-bins are drawn independently, this drop mainly reflects the difference between the total variance of event-counts and variance of the rate (Eq. 17, we neglect a small decrease of the rate autocorrelation between t_0 and t_1) and does not depend on timescales. Thus, we find α with a grid search that minimizes the distance between the autocorrelation at t_1 of the observed and synthetic data from the generative model with fixed timescales. Alternatively, α can be fitted together with all other parameters of the generative model using aABC. We find that since α is almost independent from other parameters, aABC finds the correct value of α first and then fits the rest of parameters. The MAP estimate of α converges to the same value as estimated by the grid search, but aABC requires more iterations to get posterior distributions for estimated timescales with a similar variance (Supplementary Fig. 5). Therefore, for practical purposes, it is preferred to first find α with the grid search and then

get narrower posteriors in fewer aABC iterations.

4.2 Optimizing generative model parameters with adaptive Approximate Bayesian Computations

We optimize parameters of generative models with adaptive Approximate Bayesian Computations (aABC) following the algorithm from Ref. [35, 56]. aABC is an iterative algorithm to approximate the multivariate posterior distribution of model parameters. It uses population Monte-Carlo sampling to minimize the distance between the summary statistics of the observed and synthetic data from the generative model. We use sample autocorrelation as the summary statistics and define the distance d as the mean squared deviation between the observed and synthetic autocorrelations up to time-lag t_m (Eq. 8).

On the first iteration of the algorithm, the parameters of the generative model $\theta_r^{(1)}$ are drawn from the prior distribution. We use a multidimensional uniform prior distribution $\pi(\theta)$ over fitted parameters (e.g., timescales and their weights). The domain of prior distribution for the timescales is chosen to include a broad range below and above the timescales estimated by the direct exponential fits of data autocorrelation. For weights of timescales, we use uniform prior distributions on $[0, 1]$. The model with parameters $\theta_r^{(1)}$ is used to generate synthetic time-series $A(t')$ with the same duration and number of trials as in the observed data. Next, we compute the distance d between the autocorrelations of the synthetic and observed data. If d is smaller than the error threshold ε (initially set to 1), the parameters are accepted and added to the multivariate posterior distribution. Each iteration of the algorithm is repeated until 500 parameters samples are accepted.

On subsequent iterations, the same steps are repeated but with parameters drawn from a proposal distribution and with an updated error threshold. On each iteration, the error threshold is set at the first quartile of the accepted sample distances from the previous iteration. The proposal distribution is computed for each iteration ξ based on the prior distribution and the accepted samples from the previous iteration:

$$\hat{\pi}_\xi(\theta^{(\xi)}) \propto \sum_{r=1}^{N} \omega_r^{(\xi-1)} K_\xi(\theta^{(\xi)}|\theta_r^{(\xi-1)}). \tag{20}$$

Here $\omega_r^{(\xi-1)}$ is the importance weight of the accepted sample $\theta_r^{(\xi-1)}$ from the previous iteration

$$\omega_r^{(\xi-1)} \propto \frac{\pi(\theta_r^{(\xi-1)})}{\hat{\pi}(\theta_r^{(\xi-1)})}. \tag{21}$$

K_ξ is the random walk kernel for the population Monte Carlo algorithm, which is a multivariate Gaussian with the mean $\theta_r^{(\xi-1)}$ and the covariance equal to twice the covariance of all accepted samples from the previous iteration $\Sigma = 2\mathrm{Cov}[\theta^{(\xi-1)}]$:

$$K_\xi(\theta^{(\xi)}|\theta_r^{(\xi-1)}) = \frac{1}{\sqrt{2\pi^\kappa|\Sigma|}} \exp\left(\frac{-1}{2}\left(\theta^{(\xi)} - \theta_r^{(\xi-1)}\right)^T \Sigma^{-1} \left(\theta^{(\xi)} - \theta_r^{(\xi-1)}\right)\right). \tag{22}$$

Here κ is the number of fitted parameters, and $|\Sigma|$ is the determinant of Σ.

The convergence of the algorithm is defined based on the acceptance rate $accR$, which is the

number of accepted samples divided by the total number of drawn samples on each iteration. The algorithm terminates when the acceptance rate reaches $accR_{\min}$, which is set to $accR_{\min} = 0.003$ in our simulations. Smaller $accR_{\min}$ leads to narrower posterior distributions and smaller final error threshold, but requires more iterations to converge, i.e. longer simulation time. To find the MAP estimates, we smooth the final joint posterior distribution with a multivariate Gaussian kernel and find its maximum with a grid search.

4.3 Description of neural recordings

Experimental procedures and data pre-processing were described previously [53]. In brief, a monkey was trained to fixate a central dot on a blank screen for 3 s on each trial. Spiking activity was recorded with a 16-channel micro-electrode array inserted perpendicularly to the cortical surface to record from all layers in the visual area V4. For fitting, we used a recording session with 81 trials. We pooled the activity across all channels and calculated the population spike-counts in bins of 1 ms. Then, we computed the autocorrelation of spike-counts using Eq. 2.

4.4 Parameters of simulations and aABC fits in figures

For all fits, the initial error threshold was set to $\varepsilon = 1$. The aABC iterations continued until $accR \leqslant 0.003$ was reached. All datasets (except for the branching network) consisted of 500 trials each of 1 s duration. The dataset for the branching network (Fig. 4) consisted of 100 trials with 500 time-steps.

Figures	Data parameters
Fig. 3A, Fig. 5A	$\tau = 20$, $\mu = 0$, $\sigma = 1$
Fig. 3B	$\tau = 60$, $f = 2$ Hz, $\mu = 0$, $\sigma = 1$
Fig. 3C, Fig. 5D	$\tau_1 = 5$, $\tau_2 = 80$, $c_1 = 0.4$, $\mu' = 1$, $\sigma' = 0.5$
Fig. 5G	$\tau_1 = 20$, $\tau_2 = 80$, $c_1 = 0.4$, $\mu' = 0.3$, $\sigma' = 0.18$

Table 1: Simulation parameters for data autocorrelations.

Figures	Priors	t_m
Fig. 3A, Fig. 5B, Fig. 4	$\tau : U[0, 60]$	50
Fig. 3B	$\tau : U[0, 120]$	100
Fig. 3C, Fig. 5E (Right)	$\tau_1 : U[0, 60]$, $\tau_2 : U[20, 140]$	110
Fig. 5E (Left)	$\tau : U[0, 140]$	110
Fig. 5H (Right)	$\tau_1 : U[0, 60]$, $\tau_2 : U[40, 150]$	105
Fig. 5H (Left)	$\tau : U[0, 150]$	105
Fig. 6B	$\tau_1 : U[0, 60]$, $\tau_2 : U[40, 150]$	150
Fig. 6A	$\tau : U[0, 150]$	150

Table 2: aABC fit parameters.

Acknowledgments

This work was supported by a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation, endowed by the Federal Ministry of Education and Research (RZ, AL),

SMART*START* 2 program provided by Bernstein Center for Computational Neuroscience and Volkswagen Foundation (RZ), NIH grant R01 EB026949 (TAE), and the Pershing Square Foundation (TAE). We acknowledge the support from the BMBF through the Tübingen AI Center (FKZ: 01IS18039B). We thank N.A. Steinmetz and T. Moore for sharing the electrophysiological data, which are presented in Ref. [53] and are archived at the Stanford Neuroscience Institute server at Stanford University.

Author contributions

RZ, TAE and AL designed the research, discussed results and wrote the paper. RZ wrote the computer code, performed simulations, and analyzed the data.

References

[1] Thomas Veitshans, Dmitri Klimov, and Devarajan Thirumalai. "Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties". In: *Folding and Design* 2.1 (1997), pp. 1–22.

[2] Jan Kubelka, James Hofrichter, and William A Eaton. "The protein folding 'speed limit'". In: *Current opinion in structural biology* 14.1 (2004), pp. 76–88.

[3] John D Murray et al. "A hierarchy of intrinsic timescales across primate cortex". In: *Nature neuroscience* 17.12 (2014), p. 1661.

[4] Yang Yang et al. "Millisecond-scale differences in neural activity in auditory cortex can drive decisions". In: *Nature Precedings* (2008), pp. 1–1.

[5] Kamran Diba et al. "Millisecond timescale synchrony among hippocampal neurons". In: *Journal of Neuroscience* 34.45 (2014), pp. 14984–14994.

[6] Denis Mollison and Mollison Denis. *Epidemic models: their structure and relation to data*. Vol. 5. Cambridge University Press, 1995.

[7] Denis Mollison, Valerie Isham, and Bryan Grenfell. "Epidemics: models and data". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 157.1 (1994), pp. 115–129.

[8] Sang Woo Park et al. "The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak". In: *medRxiv* (2020).

[9] Richard P Cincotta, Jennifer Wisnewski, and Robert Engelman. "Human population in the biodiversity hotspots". In: *Nature* 404.6781 (2000), pp. 990–992.

[10] Miikka Tallavaara et al. "Human population dynamics in Europe over the Last Glacial Maximum". In: *Proceedings of the National Academy of Sciences* 112.27 (2015), pp. 8232–8237.

[11] Ulrike Rippke et al. "Timescales of transformational climate change adaptation in sub-Saharan African agriculture". In: *Nature Climate Change* 6.6 (2016), pp. 605–609.

[12] Paul A Baker et al. "Tropical climate changes at millennial and orbital timescales on the Bolivian Altiplano". In: *nature* 409.6821 (2001), pp. 698–701.

[13] Jens Wilting and Viola Priesemann. "Inferring collective dynamical states from widely unobserved systems". In: *Nature communications* 9.1 (2018), pp. 1–7.

[14] William Bialek. *Biophysics: searching for principles*. Princeton University Press, 2012.

[15] William Bialek et al. "Statistical mechanics for natural flocks of birds". In: *Proceedings of the National Academy of Sciences* 109.13 (2012), pp. 4786–4791.

[16] Gašper Tkačik et al. "Thermodynamics and signatures of criticality in a network of neurons". In: *Proceedings of the National Academy of Sciences* 112.37 (2015), pp. 11508–11513.

[17] Anne Sornette and D Sornette. "Self-organized criticality and earthquakes". In: *EPL (Europhysics Letters)* 9.3 (1989), p. 197.

[18] Vadim M Uritsky, Alexander J Klimas, and Dimitris Vassiliadis. "Analysis and prediction of high-latitude geomagnetic disturbances based on a self-organized criticality framework". In: *Advances in Space Research* 37.3 (2006), pp. 539–546.

[19] O Ramos, Ernesto Altshuler, and KJ Måløy. "Avalanche prediction in a self-organized pile of beads". In: *Physical review letters* 102.7 (2009), p. 078701.

[20] Sean E Cavanagh et al. "Autocorrelation structure at rest predicts value correlates of single neurons during reward-guided choice". In: *Elife* 5 (2016), e18937.

[21] Rishidev Chaudhuri et al. "A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex". In: *Neuron* 88.2 (2015), pp. 419–431.

[22] Takamitsu Watanabe, Geraint Rees, and Naoki Masuda. "Atypical intrinsic neural timescale in autism". In: *Elife* 8 (2019), e42256.

[23] Carsen Stringer et al. "Inhibitory control of correlated intrinsic variability in cortical networks". In: *Elife* 5 (2016), e19695.

[24] Carlos D Brody. "Slow covariations in neuronal resting potentials can lead to artefactually fast cross-correlations in their spike trains". In: *Journal of Neurophysiology* 80.6 (1998), pp. 3345–3351.

[25] Valérie Ventura, Can Cai, and Robert E Kass. "Trial-to-trial variability and its effect on time-varying dependency between two neurons". In: *Journal of neurophysiology* 94.4 (2005), pp. 2928–2939.

[26] Asohan Amarasingham et al. "Conditional modeling and the jitter method of spike resampling". In: *Journal of Neurophysiology* 107.2 (2012), pp. 517–531.

[27] Marlene R Cohen and Adam Kohn. "Measuring and interpreting neuronal correlations". In: *Nature neuroscience* 14.7 (2011), p. 811.

[28] FHC Marriott and JA Pope. "Bias in the estimation of autocorrelations". In: *Biometrika* 41.3/4 (1954), pp. 390–402.

[29] A Sree Rama Sastry. "Bias in estimation of serial correlation coefficients". In: *Sankhyā: The Indian Journal of Statistics* (1951), pp. 281–296.

[30] Bradley E Huitema and Joseph W McKean. "Autocorrelation estimation and inference with small samples." In: *Psychological Bulletin* 110.2 (1991), p. 291.

[31] John S White. "Asymptotic expansions for the mean and variance of the serial correlation coefficient". In: *Biometrika* 48.1-2 (1961), pp. 85–94.

[32] ZA Lomnicki and SK Zaremba. "On the estimation of autocorrelation in time series". In: *The Annals of Mathematical Statistics* 28.1 (1957), pp. 140–158.

[33] Maurice G Kendall. "Note on bias in the estimation of autocorrelation". In: *Biometrika* 41.3-4 (1954), pp. 403–404.

[34] Mikael Sunnåker et al. "Approximate bayesian computation". In: *PLoS computational biology* 9.1 (2013).

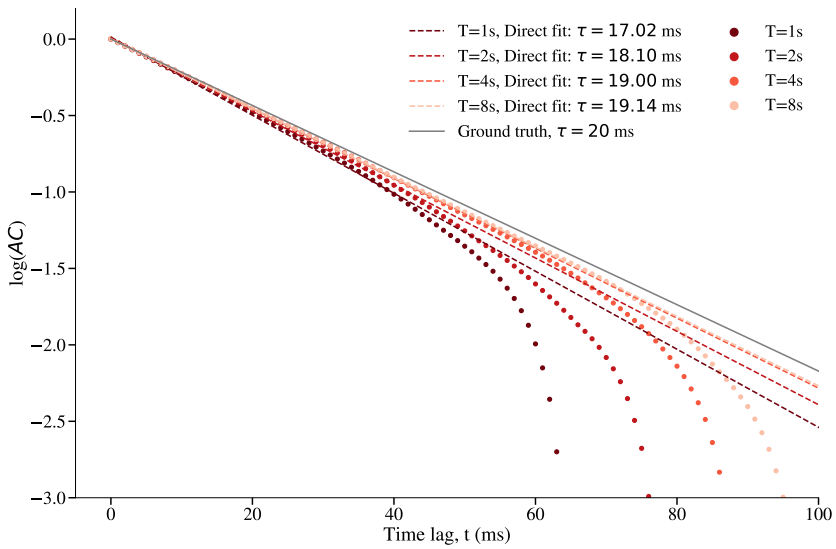
[35] Mark A Beaumont et al. "Adaptive approximate Bayesian computation". In: *Biometrika* 96.4 (2009), pp. 983–990.

[36] George E Uhlenbeck and Leonard S Ornstein. "On the theory of the Brownian motion". In: *Physical review* 36.5 (1930), p. 823.

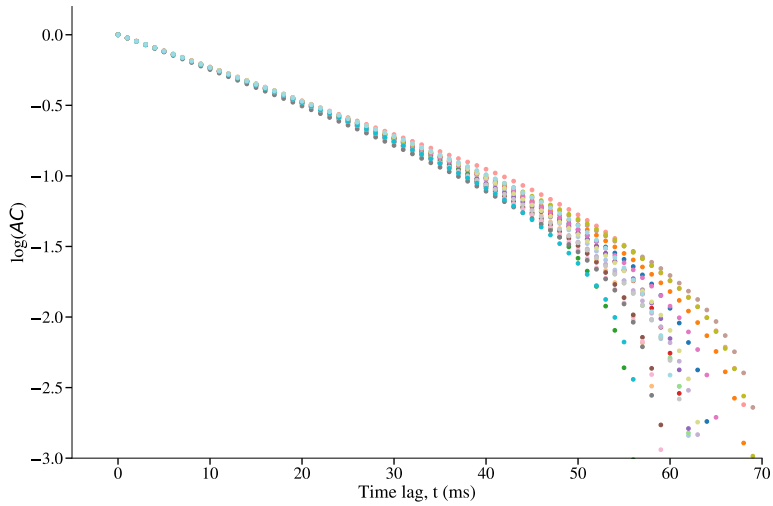
[37] Hannes Risken. "Fokker-planck equation". In: *The Fokker-Planck Equation*. Springer, 1996, pp. 63–95.

[38] Benjamin Lindner. "A brief introduction to some simple stochastic processes". In: *Stochastic Methods in Neuroscience* 1 (2009).
[39] Ryan C Kelly et al. "Local field potentials indicate network state and account for neuronal response variability". In: *Journal of computational neuroscience* 29.3 (2010), pp. 567–579.
[40] Alexander S Ecker et al. "State dependence of noise correlations in macaque primary visual cortex". In: *Neuron* 82.1 (2014), pp. 235–248.
[41] Mikhail Genkin and Tatiana A Engel. "Beyond generalization: Enhancing accurate interpretation of flexible models". In: *BioRxiv* (2019), p. 808261.
[42] Anne K Churchland et al. "Variance as a signature of neural computations during decision making". In: *Neuron* 69.4 (2011), pp. 818–831.
[43] Clayton Haldeman and John M Beggs. "Critical branching captures activity in living neural networks and maximizes the number of metastable states". In: *Physical review letters* 94.5 (2005), p. 058101.
[44] Johannes Zierenberg et al. "Description of spreading dynamics by microscopic network models and macroscopic branching processes can differ due to coalescence". In: *Physical Review E* 101.2 (2020), p. 022301.
[45] Johannes Zierenberg et al. "Tailored ensembles of neural networks optimize sensitivity to stimulus statistics". In: *Physical Review Research* 2.1 (2020), p. 013115.
[46] Aude Grelaud et al. "ABC likelihood-free methods for model choice in Gibbs random fields". In: *Bayesian Analysis* 4.2 (2009), pp. 317–335.
[47] Xavier Didelot et al. "Likelihood-free estimation of model evidence". In: *Bayesian analysis* 6.1 (2011), pp. 49–76.
[48] Christian P Robert et al. "Lack of confidence in approximate Bayesian computation model choice". In: *Proceedings of the National Academy of Sciences* 108.37 (2011), pp. 15112–15117.
[49] Jean-Michel Marin et al. "Relevant statistics for Bayesian model choice". In: *Journal of the Royal Statistical Society: Series B: Statistical Methodology* (2014), pp. 833–859.
[50] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
[51] Andrew Gelman et al. "Bayesian Data Analysis Second edition Corrected version (30 Jan 2008)". In: (1995).
[52] Tina Toni and Michael PH Stumpf. "Simulation-based model selection for dynamical systems in systems and population biology". In: *Bioinformatics* 26.1 (2010), pp. 104–110.
[53] Tatiana A Engel et al. "Selective modulation of cortical state during spatial attention". In: *Science* 354.6316 (2016), pp. 1140–1144.
[54] Robbe LT Goris, J Anthony Movshon, and Eero P Simoncelli. "Partitioning neuronal variability". In: *Nature neuroscience* 17.6 (2014), p. 858.
[55] Carlos D Brody. "Correlations without synchrony". In: *Neural computation* 11.7 (1999), pp. 1537–1551.
[56] Robert C. Morehead and Alex Hagen. *A Python package for Approximate Bayesian Computation*. https://github.com/rcmorehead/simpleabc. 2014.

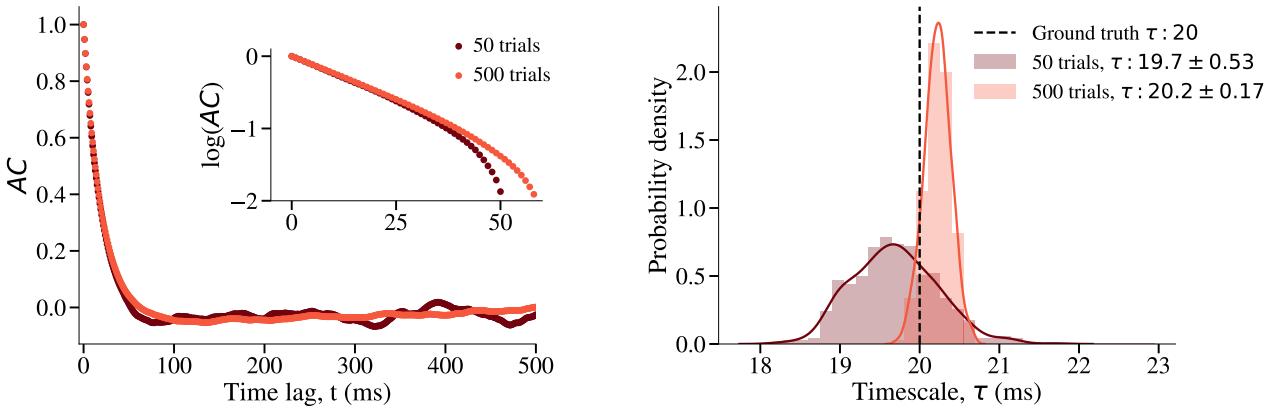
5 Supplementary figures



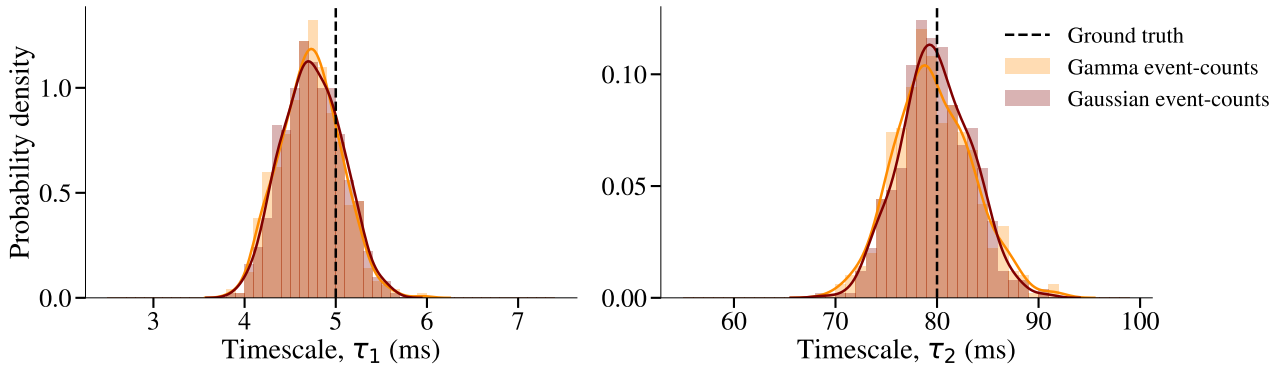
Supplementary Fig. 1: **Direct exponential fit produces different estimated timescale for datasets from the same generative process but with different trial duration.** Each dataset is generated from an OU process with $\tau = 20$ ms and 500 trials. Dots - sample autocorrelation, dashed lines - direct exponential fit, solid line - ground truth.



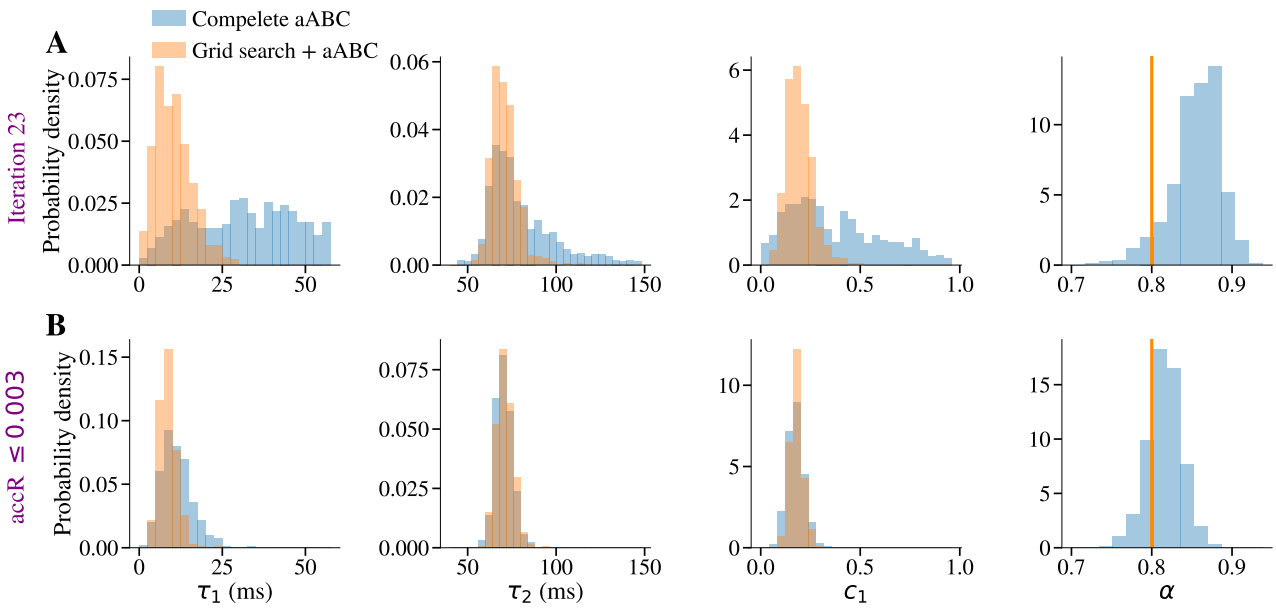
Supplementary Fig. 2: **The shape of sample autocorrelation is slightly different across realizations of the same generative process due to noise in the data.** Each trace is the average autocorrelation of 500 trials with $T = 1$ s generated from an OU process with $\tau = 20$ ms.



Supplementary Fig. 3: **Variance of posterior distributions estimated with aABC captures the estimation uncertainty and depends on the signal to noise ratio.** Increasing the number of trials in observed data reduces the noise in sample autocorrelation **(left)** and the estimation uncertainty (narrower posterior distribution) **(right)**. Estimated τ : mean $\pm$ std. The data was generated from an OU process with $\tau = 20$ ms and consisted of trials with $T = 1$ s.



Supplementary Fig. 4: **Doubly stochastic generative models with the Gaussian and gamma distributions of event-counts produce similar aABC fits.** We fitted the inhomogeneous Poisson process example from Fig. 1C using aABC with generative models which had a Gaussian or gamma event-count distributions with $\alpha = 1$. Both generative models recover the ground-truth timescales and produce posterior distributioins that are not significantly different (Wilcoxon rank-sum test: $P_{\tau_1} = 0.23$, $P_{\tau_2} = 0.37$).



Supplementary Fig. 5: **Estimating the parameter α of the doubly stochastic generative model with aABC and with the grid search.** We fitted the neural data autocorrelation in Fig. 6 considering α as an additional parameter of the generative model estimated with aABC (complete aABC). We compared the results of complete aABC with first estimating α with a grid search and then fitting the rest of parameters with aABC (grid search + aABC). **(A)** Starting from the same error threshold $\varepsilon = 1$, after the same number of iterations (23), the complete aABC has a larger variance of marginal posterior distributions than the grid search + aABC. This result indicates that the grid-search method reaches narrower posteriors in fewer iterations. Moreover, for the complete aABC, the posterior of α is relatively narrow, while posteriors of all other parameters are wider at iteration 23. This observation shows that the complete aABC fits α first and almost independently from the other parameters of the generative model. **(B)** After running both methods until they reach $accR \leq 0.003$, MAP estimates with complete aABC converge toward the estimations from the grid search + aABC.