

Efficient, nonparametric removal of noise and recovery of probability distributions from time series using nonlinear-correlation functions: Additive noise

Mainak Dhar, Joseph A. Dickinson and Mark A. Berg^{a)}

Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina 29208, USA

(Accepted, J. Chem. Phys., JCP23-AR-01693R)

^{a)} Author to whom correspondence should be addressed: berg@sc.edu

ABSTRACT

Single-molecule and related experiments yield time series of an observable as it fluctuates due to thermal motion. In such data, it can be difficult to distinguish fluctuating signal from fluctuating noise. We present a method of separating signal from noise using nonlinear-correlation functions. The method is fully nonparametric: no *a priori* model for the system is required, no knowledge of whether the system is continuous or discrete is needed, the number of states is not fixed, and the system can be Markovian or not. The noise-corrected, nonlinear-correlation functions can be converted to the system's Green's function; the noise-corrected moments yield the system's equilibrium-probability distribution. As a demonstration, we analyze synthetic data from a three-state system. The correlation method is compared to another fully nonparametric approach—time-binning to remove noise, and histogramming to obtain the distribution. The correlation method has substantially better resolution in time and state space. We develop formulas for the limits on data quality needed for signal recovery from time series and test them on data sets of varying size and signal-to-noise ratio. The formulas show that the signal-to-noise ratio needs to be on the order of or greater than one-half before convergence scales at a practical rate. With experimental benchmark data, the positions and populations of the states and their exchange rates are recovered with an accuracy similar to parametric methods. The methods demonstrated here are essential components in building a complete analysis of time series using only high-order correlation functions.

I. INTRODUCTION

Separating experimental data into signal and noise is an old problem. For kinetics, the most common type of data is the response of an observable as it relaxes after a perturbation. In these experiments, averaging multiple experimental runs is the obvious way to reduce noise. This method works because, at each delay, the noise fluctuates, but the signal is constant. An alternative type of kinetic data is a time series of thermal fluctuations of the same observable in an unperturbed system. In these experiments, both signal and noise fluctuate. Averaging multiple experimental runs is not possible; it eliminates the signal as much as it reduces the noise. Most existing methods for analyzing noisy time series require some degree of prior knowledge of (or assumptions about) the system. Nonparametric methods avoid these assumptions. Time binning, histogramming and linear-correlation functions are well-established nonparametric methods, but by themselves, they are inefficient and incomplete. This paper will demonstrate efficient nonparametric methods for analyzing time series with additive noise based on nonlinear-correlation functions. The increased efficiency translates into improved resolution both in state space and in time.

Time series are generated in many different experiments with different detectors and different types of noise, for example, single ion-channel recordings,¹ single-emitter measurements,²⁻⁷ fluorescence-,^{8,9} photon-,¹⁰ or x-ray-correlation¹¹ spectroscopies. Many of these experiments are inherently low signal and high noise. Even when the dynamics are relatively simple, characterizing them can be difficult. On the other hand, modern electronics can collect large quantities of data. The problems are to define the limits to compensating for a low signal-to-noise ratio with a large data set and to find practical methods that approach these limits.

This paper will introduce basic concepts using additive noise, which is the simplest type of noise to treat and is a reasonable model for many detectors. Single-molecule and fluorescence-correlation spectroscopies are particularly important experiments that are widely used in biology and material science. They also include photon noise, which is more complex than additive noise. At least under certain circumstances, these complexities can be ignored, and single-molecule data can be treated with the proposed methods. Moreover, the current treatment of additive noise lays the foundation for a full treatment of photon noise.¹²

Many new types of time-series analysis have been introduced in recent years.¹³⁻¹⁵ They can provide good time resolution in the presence of high noise, but most require prior information about the system. Among these methods, fully parametric ones require the most information. A complete model of the system is specified, except for the values of certain parameters in the model. The parameters are fit to the data by maximum-likelihood or Bayesian¹⁶⁻²² methods. If the state space is known to be continuous, the model might be based on diffusion equations.^{23,24} More often, methods are aimed at discrete-state systems with Markovian dynamics. (Markovian dynamics imply exponential kinetics between the system's states and multiexponential observed

decays. In Markovian models, nonexponential, stretched or heterogeneous dynamics are explained by inferring the existence of multiple states that cannot be resolved by the experimental observable.) A model based on a master equation can be used, resulting in a hidden-Markov model.²⁵⁻³⁷ Introducing Bayesian-nonparametric ideas allows the number of states to vary, within the same discrete-state-Markov framework.³⁸⁻⁴¹ Noise is not removed *per se* in these methods. Instead, a specific description of the noise is incorporated into the overall model.

Unfortunately, detailed prior knowledge about the system is often not available. Instead, multiple models are proposed and fit. This introduces a model-selection problem. In addition, the models considered are typically restricted to a purely discrete state-space, to a purely continuous state-space, and/or to Markovian dynamics. The proposed methods seek to circumvent these restrictions.

A flexible and popular method of nonparametric noise removal is time binning.⁴²⁻⁴⁴ The only assumption is that the noise fluctuations are rapid compared to the signal. However, time binning creates a competition between noise reduction and time resolution. Moreover, the trade-off is inefficient: the time resolution is reduced by the square of the noise reduction. Paradoxically, the effective time resolution of the experiment can be orders-of-magnitude lower than the instrument's time resolution when the noise is high.

Time binning is often combined with histogramming to estimate the equilibrium distribution.^{35,45-47} However, the histogram can be broadened by measurement noise and blurred by kinetics within the time bins. By introducing models of these processes, one can attempt to disentangle their effects.⁴⁸⁻⁵²

Improvement over simple time binning is possible, but again at the cost of introducing assumptions. Most often, one assumes that the system consists of a small number of discrete states undergoing sudden transitions. Various methods, from simple thresholding^{53,54} to more sophisticated change-point methods,⁵⁵⁻⁶¹ identify these transitions. By their nature, these methods do not work when the state space is continuous, for example, for molecular reorientation³ or intrinsically disordered proteins.⁵

Both time binning and change-point methods aim to remove the noise from the time series itself to estimate an ideal, noise-free trajectory. The ideal trajectory can then be analyzed with a parametric model or by model-free methods, such as dwell-time histograms or transition-correlation functions.⁶²⁻⁶⁵ However, the ideal trajectory is problematic as a statistical quantity. The accuracy of state identification at a given time or the time of a specific transition cannot be increased arbitrarily by increasing the size of the data set.

Correlation methods avoid many of the issues introduced by the above approaches. Consider the linear-correlation function,

$$M_{11}[D](\tau) = \langle D(\tau)D(0) \rangle, \quad (1)$$

which is a functional of the time-series data $D(t)$. It is nonparametric: no model of the system is used, no restrictive assumptions are needed, and no model-selection problem is introduced. It can be used with any type of state space, either continuous or discrete, and with any type of dynamics, either Markovian or not. No ideal trajectory is involved. As a result, its accuracy can be increased arbitrarily by collecting more data, either through longer time series or by measuring multiple time series. The time resolution is equivalent to the instrument's time resolution, even when the noise level is high. Unfortunately, much of the important information about the system is not captured by this single function.

Beyond the linear-correlation function, there is a large set of high-order correlation functions that do contain complete information about the data. One can take a nonlinear function of the time series before the calculation, creating a nonlinear-correlation function, and/or one can use more than two time points, creating a multidimensional-correlation function (a correlation function with multiple time delays). Using powers as a basis set for all nonlinear functions leads us to the moment-correlation functions,

$$M_{k...lm}[D](\tau_N, \dots, \tau_1) = \langle D(\tau_N + \dots + \tau_1)^k \dots D(\tau_1)^l D(0)^m \rangle. \quad (2)$$

Most work has focused on the linear-correlation function $M_{11}(\tau)$,⁶⁶⁻⁷¹ which is just the lowest, nontrivial element of this set. (Although, see Refs. 72,73 for early work with high-order functions). One objection to the linear-correlation function is that it loses information on heterogeneity in the sample. However, it has been recognized that this information is in the multidimensional members of the set. We have shown that $M_{111}(\tau_2, \tau_1)$ or $M_{121}(\tau_2, \tau_1)$ contain information on the heterogeneity of rate constants and that $M_{1111}(\tau_3, \tau_2, \tau_1)$ measures the time evolution of those rate constants.⁷⁴⁻⁷⁹ Marcus, van Hippel and coworkers have used $M_{111}(\tau_2, \tau_1)$ to evaluate discrete-state models containing an unobserved coordinate.^{80,81} These multidimensional-correlation methods should be compared to other time-series methods that use multiple coordinates, such as wavelets^{56,82,83} or polyspectra.⁸⁴⁻⁸⁶

Here, we postpone issues of heterogeneity and focus on one-dimensional (1D), nonlinear-correlation functions,

$$M_{kl}[D](\tau) = \langle D(\tau)^k D(0)^l \rangle, \quad n = k + l. \quad (3)$$

(These time-domain functions can be equivalently represented by nonlinear spectral densities.⁸⁷) In a previous paper, we showed that a set of these functions can be converted to the Green's function of the system,⁸⁷ which has complete information about the dynamics along the observed coordinate. However, that procedure requires a noise-free time series and a known equilibrium-

probability distribution. It has been successful with computer simulations,^{79,88} but it cannot be applied to unmodified experimental data. This paper presents a method for analyzing noisy time-series analysis that solves these problems while retaining the advantages of a correlation approach.

The major goals of this paper are threefold. First, we seek a practical method to remove noise from nonlinear-correlation functions while only assuming that the noise-correlation time is faster than the signal-correlation time (Sec. III.A). The noise-removal process is shown to be effective and accurate using a set of synthetic data (Sec. II).

More attention is given to the second goal (Sec. III.B)—finding the equilibrium-probability distribution—which is a less well-studied problem. A typical nonparametric approach is making a histogram of the time series after time binning. In the correlation approach, the nonlinear correlation functions are first reduced to noise-corrected moments of the distribution. The distribution is recovered from these moments. This latter step is known to be unstable and error-sensitive.⁸⁹⁻¹⁰⁰ We demonstrate a numerically stable route for our version of the problem and show that the noise-corrected moments can be sufficiently accurate for this route to work. The demonstration is made on synthetic data and is compared to time-binned histograms. There are substantial improvements in the resolution both in time and in state space.

With the combination of noise-corrected correlation functions and an accurate equilibrium distribution, the system's Green's function can be found.⁸⁷ From these results, the dynamical equation governing the system is reconstructed. Although the details of this step have been covered elsewhere, a brief demonstration is given in Sec. III.C for completeness.

The third major goal of the paper is to establish the limitation on data quality needed for this reconstruction, in particular, the minimum series length, signal-to-noise ratio, and ensemble size required. This goal is achieved in part by looking at smaller synthetic-data sets, in part with a general theory, and in part by treating a set of experimental, benchmark data (Sec. IV). Equations that estimate the amount of data needed and the time resolution versus the signal-to-noise ratio and the complexity of the equilibrium distribution are derived (Sec. IV.B). These formulas can guide experimental design.

In addition to these practical issues, we address two fundamental questions: Can collecting more data always compensate for a high noise level and can the time resolution of the experiment approach the time resolution of the instrument? For the first question, we find that an arbitrarily small signal-to-noise ratio can always be compensated by an unlimited amount of data, as required by statistical consistency. However, below a signal-to-noise ratio of one-half, the required amount increases at an impractical rate. For the second question, the time resolution of the results can approach the time resolution of the instrument, but only if the signal-to-noise ratio is above one-half.

An overarching goal is to build a comprehensive, nonparametric time-series analysis based on correlation methods. Here, two important pieces of that program—noise removal and extracting the probability distribution—are shown to be viable. However, this demonstration does not exhaust the needs of real experiments. Photon-counting noise, multiple observables, and slow noise (e.g., baseline drift)¹⁰¹ are important examples of topics that still need to be addressed. However, this paper lays the foundations needed to treat these problems in the future.

II. DATA WITH ADDITIVE NOISE

A. Additive noise in equilibrium–fluctuation measurements

A dynamic system has an observable property $X(t)$ that varies with time t . We assume that the system has reached an equilibrium or stable steady state at the start of the measurement. The system's static properties are described by its equilibrium distribution $P_{\text{eq}}(X)$. Its dynamics are described by a Green's function $G(X_1|X_0; \tau)$, the conditional probability of observing X_1 at time $t + \tau$, if it has intensity X_0 at time t . If the system is small, thermal motion will cause $X(t)$ to fluctuate in time, even though the system remains at equilibrium during the measurement. Such noise-free measurements of the system property itself are found in computer experiments.^{29,79,88} Without noise, the equilibrium distribution is simply the histogram of $X(t)$, and the Green's function can be extracted from the full set of moment-correlation functions of the time series $M_{kl}[X](\tau)$ [Eq. (3)].^{79,87,88}

In a physical experiment, an ideal detector would yield a noise-free signal $S(t)$ that is linearly related to the system property,

$$S(t) = \alpha X(t) + B. \quad (4)$$

The constant α represents the overall detection efficiency and yields the signal as an intensity. (Intensities include units of inverse time, for example, a current or counts per second.) The detector may also create a background signal B , which will not be removed by the proposed methods. For additive noise, the effects of the detection efficiency and background are relatively simple, but when we move to nonadditive cases, their effects will be more complex.¹² For simplicity, we take $\alpha = 1$ throughout this paper.

A real detector generates data $D(t_j)$ at time points t_j with measurement noise that varies randomly from point to point with conditional probability $P_\epsilon(D|S)$. In practice, there is a finite, noise-correlation time T_ϵ required between measurements to obtain independent values of the noise. This time defines the measurement time resolution. More precisely, at T_ϵ , the noise correlation has decayed to a level small compared to the signal correlation. We assume that the raw data is collected in time steps equal to the measurement time resolution: $t_{j+1} - t_j = T_\epsilon$.

Time binning is the primary, existing, nonparametric method for correcting measurement noise. To create binned data $D(t_i, N_b)$, the raw data is averaged over N_b time steps,

$$D(t_j, N_b) = \frac{1}{N_b} \sum_{i=j}^{j+N_b} D(t_i), \quad (5)$$

reducing the time resolution of the experiment to $T_r = N_b T_\varepsilon$. The magnitude of the noise is expected to decrease as $N_b^{-1/2}$. However, the magnitudes of intensities, such as $S(t)$ or $D(t)$, do not change size as a result of binning.

The simplest type of noise is additive. It is often caused by electronic amplifiers. Additive noise is defined by

$$D(t) = S(t) + \varepsilon(t), \quad (6)$$

where $S(t)$ is noise free, and $\varepsilon(t)$ is a stochastic process with a noise distribution $P_\varepsilon(\varepsilon)$. The noise distribution has a zero mean, $E(\varepsilon) = 0$, so a simple average of $D(t)$ is unbiased. [An expectation value is indicated by $E(\dots)$.] The absolute size of the noise is given by σ_ε , the standard deviation of $P_\varepsilon(\varepsilon)$, and its relative size is expressed by the signal-to-noise ratio,

$$\text{SNR} = \frac{\langle S \rangle}{\sigma_\varepsilon}. \quad (7)$$

The key distinction between signal and noise lies in their correlation properties. This is true for noise removal by either time binning or correlation methods. The noise-correlation time is short,

$$E\left(\varepsilon(t + T_\varepsilon)^k \varepsilon(t)^l\right) = E\left(\varepsilon^k\right) E\left(\varepsilon^l\right), \quad (8)$$

whereas the signal-correlation time is long,

$$\langle S(nT_\varepsilon) S(0) \rangle = \langle S^2 \rangle \quad (9)$$

for n not too large. Because the noise distribution is independent of the signal size S , the noise is uncorrelated with the signal,

$$\left\langle E\left(\varepsilon(\tau)^k S(0)^l\right) \right\rangle = E(\varepsilon^k) \langle S^l \rangle \quad (10)$$

for all τ , including $\tau = 0$.

As a result of Eq. (10), the distribution of the measured data $P_D(D)$ is a convolution of the signal distribution $P_S(S)$ and the noise distribution,

$$P_D(D) = \int_{-\infty}^{\infty} P_S(S) P_\varepsilon(D - S) dS. \quad (11)$$

Noise-removal is essentially a blind deconvolution problem requiring that $P_D(D)$ be separated into $P_S(D)$ and $P_\varepsilon(\varepsilon)$ without knowing either of them.

The lack of correlation between signal and noise [Eqs. (10) and (11)] no longer holds for photon noise, and the lack of bias no longer holds for photon counting.¹² However, the separation of correlation times [Eqs. (8) and (9)] persists in these cases. Section IV.D will show that the methods developed here can continue to work in these cases, despite the lack of full, formal support at this time.

B. Generating synthetic data

Synthetic data were generated from the three-state system shown in Fig. 1. The system properties X_i for the states i form a vector $\mathbf{X} = \{0, 0.5, 1.0\}$. The equilibrium populations $P_{\text{eq},i}$ are in the ratio of 5:3:4, forming a vector $\mathbf{P}_{\text{eq}} = \{5/12, 1/4, 1/3\}$. All of the transitions are Markovian with rates from state i to j of k_{ij} (supplementary material, Sec. SI). We used a background of $B = 0.125$ to avoid having a state with the special value of zero signal.

Although this model has a discrete state space, the proposed method is equally applicable to continuous state spaces. In fact, the recovered distribution will always be broadened into a quasi-continuous function because the resolution of the recovery will not be perfect. The state properties X_i convert between the discrete-state probabilities $P_{\text{eq},i}$ and the continuous-state probability density, $P_{\text{eq}}(X) = \sum_i P_{\text{eq},i} \delta(X - X_i)$. Cumulative-distribution functions, for example

$$F_X(X) = \int_0^X P_X(X') dX' , \quad (12)$$

make it easier to compare peak areas between continuous and discrete descriptions. Figure 1 shows an example.

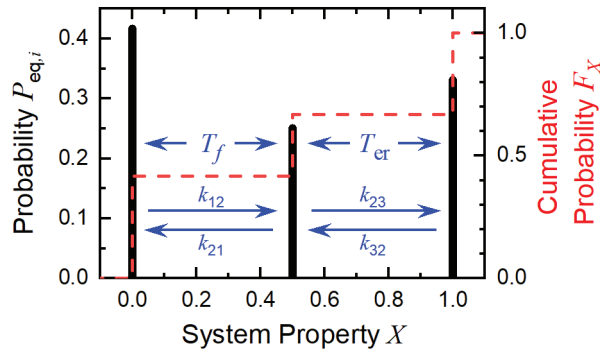


FIG. 1. The model used to generate the synthetic data. The equilibrium probabilities of the states versus the value of the system property are shown as black bars (left scale) along with their state-to-state rates and equilibration times (blue arrows). The cumulative-distribution function is shown as a dashed, red curve (right scale).

The dynamics are conveniently characterized by the fast-equilibration time constant, $T_f = (k_{12} + k_{21})^{-1}$, between states 1 and 2, and the slow-equilibration time constant, $T_{er} = (k_{23} + k_{32})^{-1}$, between states 2 and 3. Because T_{er} is the slowest relaxation in the model, it has been identified with the ergodic time, the time needed to fully explore the system's configurations. To make the results more concrete, we will quote times for a time step of $T_\epsilon = 1 \mu\text{s}$. However, all results in this paper are independent of the absolute timescale, and times can be uniformly rescaled to other values.

The fast-equilibration time is set to $T_f = 100T_\epsilon = 100 \mu\text{s}$. In the absence of noise, the time resolution is more than adequate to measure this time, but at high noise levels, the equilibration will be obscured by time binning. The slow-equilibration time is set to $T_{er} = 10^4 T_\epsilon = 10 \text{ ms}$, which can be resolved under all reasonable levels of noise. However, its measurement can be limited by the finite length of the time series T_L .¹⁰²⁻¹⁰⁴ Initially, we wish to focus on the problem of measurement noise, so we first look at a long series of $T_L = 10^4 T_{er} = 100 \text{ s}$. (Shorter series are considered in Sec. IV.C.) Solving the kinetic equations with these parameters (supplementary material, Sec. SI) gave the noise-free time series $S(t)$.

The form of the noise distribution is not important in our method. For simplicity, we used Gaussian noise,

$$P_\epsilon(\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left(-\frac{\epsilon^2}{2\sigma_\epsilon^2}\right). \quad (13)$$

The initial calculations were done with $\sigma_\epsilon = 1$ or $\text{SNR} = 0.583$. (The effects of changing the signal-to-noise ratio will be examined in Sec. IV.A.) Equation (6) was then used to generate the data $D(t)$.

The results can be seen as the black points in Fig. 2(a). The noise-free signal $S(t)$ is shown as red points for comparison. A transition from state 2 to state 3 is marked by a blue vertical line. A histogram of the unbinned data $P_D(D)$ is shown in Fig. 3 ($N_b = 1$). The states and transitions are difficult to identify without some type of noise correction.

Examples of noise correction by time binning are shown in Figs. 2(b)–2(d). Corresponding histograms of the binned time series as a function of the extent of binning $P_D(D, N_b)$ are shown in Fig. 3. The goals are for the binned data (black points) to approximate the noise-free time series (red points) in Fig. 2 and for the histogram (Fig. 3) to approximate the true distribution (Fig. 1).

Achieving these goals is frustrated because binning puts time resolution and noise removal in conflict. Too little binning does not remove enough noise: transitions are difficult to identify, and the peaks in the histogram are poorly resolved. Too much binning reduces the time resolution: the fast-exchanging states blur, and their transition rates cannot be measured. Many analysts would

select $N_b = 100$ (Fig. 3, blue, filled curve) as the best compromise, but the choice is subjective. In this data set, no genuinely good choice is possible, despite the large amount of data available.

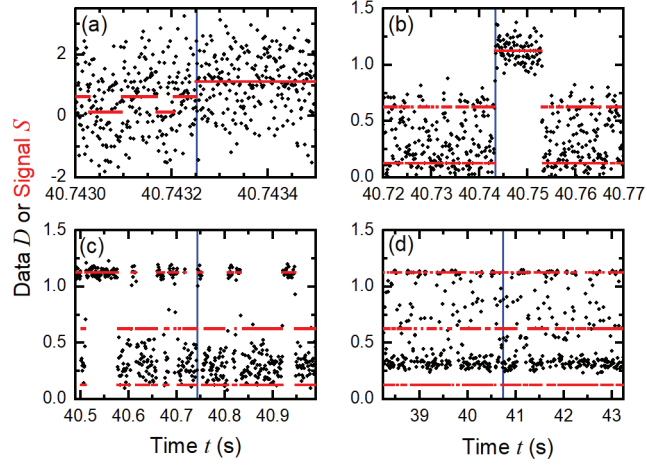


FIG. 2. Effect of time binning on the time series. Segments of the noise-free signal (red squares) and the noisy data (black circles) with different amounts of binning: (a) $N_b = 1$ (no binning); (b) $N_b = 100$; (c) $N_b = 10^3$; (d) $N_b = 10^4$. Only points from non-overlapping bins are plotted. The time scale is changed in each subpanel to show the same number of time bins. A fixed time, which is centered on a transition from state 2 to state 3, is marked by a vertical, blue line. Note the different vertical scale in (a).

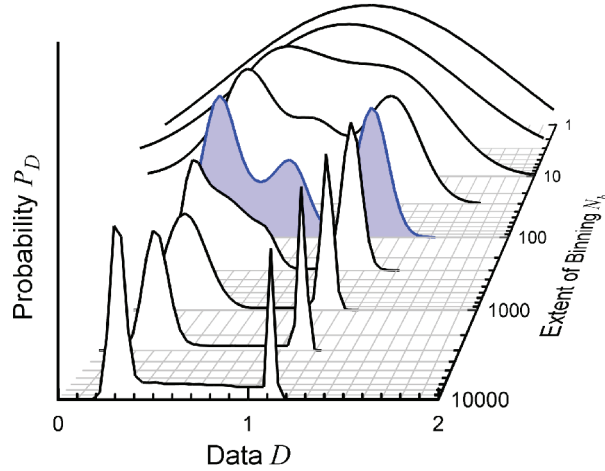


FIG. 3. Effect of time binning on the histogram. Histograms of the time-binned data versus the extent of binning. Increasing the binning reduces the broadening due to measurement noise, but increases the blurring due to poor time resolution. The filled, blue curve at $N_b = 100$ is taken as the best compromise.

This data set has problems with time resolution for the fast dynamics and with adequate sampling of the slow dynamics. The proposed correlation methods will ultimately address both problems.

III. RECOVERING DYNAMICS FROM NOISY DATA

A. Recovering nonlinear-correlation functions

Equation (3) defined the moment-correlation functions as functionals of a time series. The current goal is to relate the moment-correlation functions of data $M_{kl}[D](\tau)$ to those of the system property $M_{kl}[X](\tau)$. Conventional analysis focuses on the linear-correlation function $M_{11}[D](\tau)$, where the effects of noise and background are simple. Using Eqs. (4), (6), (8), and (10) in Eq. (3) gives

$$\begin{aligned} M_{11}[D](\tau) &= \sigma_\varepsilon^2 \delta(\tau) + \langle S(\tau)S(0) \rangle \\ &= \sigma_\varepsilon^2 \delta(\tau) + \langle X(\tau)X(0) \rangle + 2B \langle X \rangle + B^2. \end{aligned} \quad (14)$$

In addition to the property-correlation function, $M_{11}[X](\tau) = \langle X(\tau)X(0) \rangle$, there is a delta-function (first) term due to noise and a constant (last two terms) due to the background. The constant is removed by defining Δ as the operation of subtracting the $\tau = \infty$ value: $\Delta M_{kl}[D](\tau) = M_{kl}[D](\tau) - M_{kl}[D](\infty)$. The noise is removed by dropping the first delay point:

$$\Delta M_{11}[X](\tau) = \Delta M_{11}[S](\tau) \approx \Delta M_{11}[D](\tau + T_\varepsilon). \quad (15)$$

An example is shown in Fig. 4(a). The solid, black curve is calculated from the noise-free signal $S(t)$, and the dashed, green curve is calculated from the noisy data $D(t)$. The delta-function component is off-scale, and its size is shown by the labeled point. Aside from the $\tau = 0$ point, the two curves overlap perfectly. (Correlation functions are calculated with quasi-logarithmic time bins. Although restricted to monotonic decays, this method is faster and less noisy at long times than linear-time or Fourier-transform methods.¹⁰⁵)

This simple procedure does not extend to higher orders. Starting with $(k, l) = (3, 1)$, there are cross terms between lower-order functions and both the noise and the background:

$$\begin{aligned} \Delta M_{31}[D](\tau \neq 0) &= \Delta M_{31}[S](\tau) + 3 \langle \varepsilon^2 \rangle \Delta M_{11}[S](\tau) \\ &= \Delta M_{31}[X](\tau) + 3 \left(\langle \varepsilon^2 \rangle + B^2 \right) \Delta M_{11}[X](\tau) \\ &\quad + 3B \Delta M_{21}[X](\tau). \end{aligned} \quad (16)$$

Figure 4(b) compares $\Delta M_{31}[D](\tau)$ from noisy data to $\Delta M_{31}[S](\tau)$ from noise-free signal. The difference is large: the noisy result has been divided by a factor of 2.5 to keep it on the same scale

as the noise-free result. Once the two are on the same scale, one can also see that the decay shape is distorted. Even when the noise is delta correlated, its effects are not confined to a single delay.

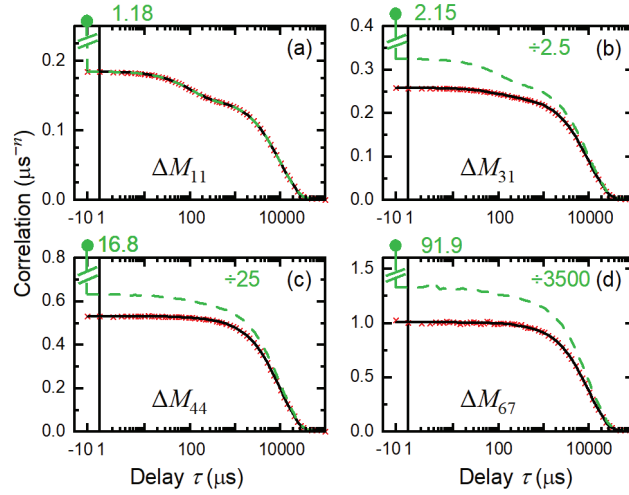


FIG. 4. Effect of noise and noise-correction on nonlinear-correlation functions. Moment-correlation functions with the infinite-time value subtracted for (a) $(k, l) = (1, 1)$, (b) $(k, l) = (3, 1)$, (c) $(k, l) = (4, 4)$, and (d) $(k, l) = (6, 7)$. The solid, black curves are calculated from the noise-free signal $\Delta M_{kl}[S](\tau)$, the dashed, green curves are from the noisy data $\Delta M_{kl}[D](\tau)$, and the red crosses are noise-corrected data $\Delta M'_{kl}[D](\tau)$. The noisy results have been divided by the factors shown. The horizontal scale is linear below 1 to show the point at $\tau = 0$.

A general formula for the noise-induced distortions is

$$\begin{aligned}
 M_{kl}[D](\tau) = & M_{kl}[S](\tau) \\
 & + \sum_{i=2}^{k-1} \binom{k}{i} \langle \varepsilon^i \rangle M_{k-i,l}[S](\tau) \\
 & + \sum_{j=2}^{l-1} \binom{l}{j} \langle \varepsilon^j \rangle M_{l,k-j}[S](\tau) \\
 & + \sum_{i=2}^{k-1} \sum_{j=2}^{l-1} \binom{k}{i} \binom{l}{j} \langle \varepsilon^i \rangle \langle \varepsilon^j \rangle M_{k-i,l-j}[S](\tau); \quad \tau \neq 0.
 \end{aligned} \tag{17}$$

The first term contains the desired signal-correlation function. The next term was in the previous example [Eq. (16), Fig. 4(b)]; the third term is its symmetry companion. Above $(k, l) = (3, 3)$, the last term in Eq. (17) also becomes time dependent, further increasing the effect of noise. Examples for $\Delta M_{44}(\tau)$ and $\Delta M_{67}(\tau)$ are shown in Figs. 4(c) and 4(d). As a result of the combinatorial factors in Eq. (17), the desired first term becomes overwhelmed by noise-related terms as the order

increases. Thus, correcting noise in nonlinear-correlation functions is more difficult than for linear ones.

To treat high-order cases, a noise-corrected, moment-correlation function is defined as

$$M'_{kl}[D](\tau) = \left\langle D(\tau + nT_\varepsilon)^{k'} D(0)^{l'} \right\rangle; \quad n = k + l, \quad (18)$$

where the primed powers are defined by

$$D(t)^{k'} = D((k-1)T_\varepsilon + t) \dots D(T_\varepsilon + t) D(t). \quad (19)$$

In the primed power, the data points are separated by the correlation time for the noise T_ε . As a result, the factors of $\langle \varepsilon^i \rangle$ in Eq. (17) become $\langle \varepsilon^{i'} \rangle = E(\varepsilon)^i = 0$ [Eq. (8)]. Noise effects are eliminated. On the other hand, if the noise-correlation time is substantially faster than the fastest signal-decay time, $T_\varepsilon \ll T_\beta$, the signal-correlation functions will be unaffected. Thus, Eq. (17) becomes

$$M'_{kl}[D](\tau) = M_{kl}[S](\tau). \quad (20)$$

The accuracy of this result is shown in Fig. 4. The noise-corrected, moment-correlation functions are shown as red crosses. They agree very well with the results calculated from noise-free signal (solid, black), despite correcting several orders-of-magnitude of distortion.

B. Recovering the equilibrium distribution

1. Recovering distributions from their moments

In our correlation-based approach, the distribution is recovered from its moments μ_n . These are defined as functionals of a probability distribution, for example,

$$\mu_n[P_D] = \int_{-\infty}^{\infty} dD D^n P_D(D). \quad (21)$$

They are also the zero-time value of the moment-correlation function calculated from the corresponding time series,

$$\mu_n[D] = \left\langle D^n \right\rangle = M_{kl}[D](0); \quad n = k + l. \quad (22)$$

Our problem is to invert Eq. (21), that is, to find the distribution $P_{\text{eq}}(X)$ from a sequence of moments derived from a noisy time series.

Simply applying Eq. (22) to the noisy data is an inadequate starting point. The zero-time values of the uncorrected correlation functions are strongly affected by noise (Fig. 4). Inverting those values would give $P_D(D)$, the distribution of the noisy data (Fig. 3, $N_b = 1$). The success of the noise-corrected correlation functions suggests that we should start with their zero-time values, which we define as the noise-corrected moments,

$$\mu'_n[D] = M'_{lk}[D](0); \quad n = k + l. \quad (23)$$

Their values can also be calculated directly, without generating the entire correlation function:

$$\mu'_n[D] = \langle D(nT_\varepsilon) \dots D(T_\varepsilon) D(0) \rangle = \langle D^{n'} \rangle. \quad (24)$$

If the signal does not decay significantly over nT_ε , then

$$\mu_n[P_S] \approx \mu'_n[D], \quad (25)$$

and inverting the noise-corrected moments will give the distribution of the signal.

The problem of finding a function from its moments is a classic problem in pure mathematics.^{89,90} Data analysis is always restricted to a finite domain, $X_{\min} \leq X \leq X_{\max}$, in which case, any function can be recovered uniquely and exactly from an infinite number of error-free moments. Equation (24) removes measurement noise, but unfortunately, the noise-corrected moments still contain sampling error due to the finite length of the time series. That error increases with order, so we only have a finite moment sequence, $\boldsymbol{\mu}' = \{\mu'_n\}$ for $n = 0$ to N_μ , will be accurate enough to be useful.

To estimate the minimum number of moments needed for the inversion $N_{\mu,\min}$, first consider discrete-state systems, which always have a highest state at $X_{\max}$ with a probability $P_{X,\max}$. In the limit of high orders, $n \rightarrow \infty$, the behavior of the moments $\mu_n^\infty[P_X]$ is dominated by this state, with

$$\ln \mu_n^\infty[P_X] = \ln P_{X,\max} + n \ln X_{\max}. \quad (26)$$

As an example, the noise-corrected moments from our data are shown in Fig. 5 (black circles). Moments in the asymptotic region are redundant.

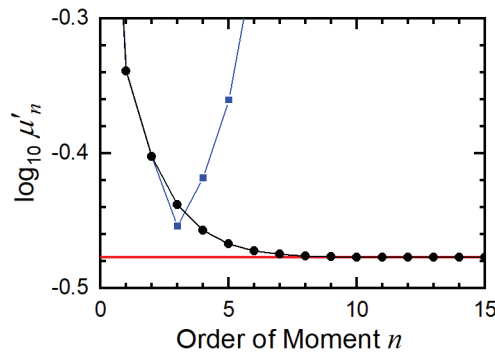


FIG. 5. The noise-corrected moments of our simulated, discrete-state data (black curve and circles). The red line is the asymptote. The moments of a continuous-state Gaussian (blue curve and squares) are shown for comparison.

In a continuous-state system, the asymptote will not be a straight line; the Gaussian moments in Fig. 5 (blue squares) are an example. However, we are only interested in problems where the probability density decays quickly at high intensities. An asymptotic region will still exist that is dominated by the behavior of the high-intensity, low-probability tail of the distribution, but the shape of this tail is usually not important for interpreting the data. Thus, moments in the asymptotic region of a continuous-state system are also redundant for practical purposes.

The effective start of the asymptotic region n_{asy} can be defined by the point where the deviation from the asymptote becomes less than the maximum allowable error level in the measured moment $\omega_{\varepsilon, \text{max}}$. From our experience [described below, see Eq. (42)],

$$\frac{\mu_n - \mu_n^\infty}{\mu_n^\infty} = \omega_{\varepsilon, \text{max}} = 5 \times 10^{-3} \quad \text{for } n = n_{\text{asy}}. \quad (27)$$

As illustrated by Fig. 6, this number increases with the complexity of the distribution. Moments have been calculated for three models that have the same first moments and the same asymptotic moments, but that have two (red circles), three (green squares) or four (blue triangles) states ($\mathbf{X} = \{1/2, 2\}$, $\mathbf{P}_{\text{eq}} = \{2/3, 1/3\}$; $\mathbf{X} = \{0, 1, 2\}$, $\mathbf{P}_{\text{eq}} = \{1/3, 1/3, 1/3\}$; and $\mathbf{X} = \{0, 1/2, 3/2, 2\}$, $\mathbf{P}_{\text{eq}} = \{1/3, 1/6, 1/6, 1/3\}$; respectively). The deviation from the asymptote is plotted on a linear scale in the main panel of Fig. 6. The more states in the distribution, the slower the approach to the asymptote. From the inset with a log-linear scale, one can use Eq. (27) to estimate n_{asy} : if there are two states, $n_{\text{asy}} = 4$ –5; if there are three, $n_{\text{asy}} = 7$ –8; if there are four, $n_{\text{asy}} \approx 16$.

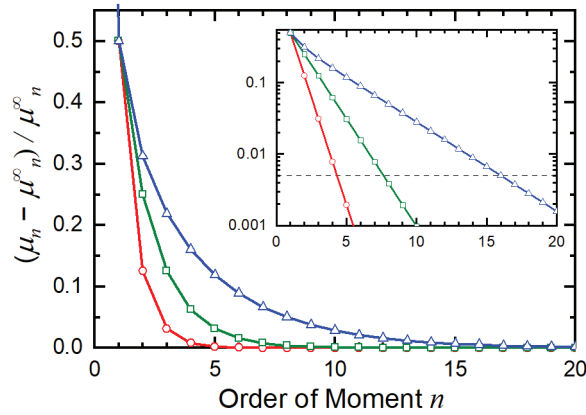


FIG. 6. The deviation of the moments from their asymptotic values [Eq. (26)] versus their order for three models of increasing complexity: 2-state (red circles), 3-state (green squares), and 4-state (blue triangles). The inset are the same data on a log-linear scale. The dashed reference line indicates the start of the asymptotic region [Eq. (27)].

This argument can be reversed when inverting a finite sequence of moments. One must first decide on the maximum complexity of the distribution to be considered and thus on a maximum value of n_{asy} . Adding moments to the sequence beyond this number will not be helpful (although not necessarily harmful either). Conversely, moments up to this number help to distinguish between different distributions. Thus, the minimum number of moments needed for a successful inversion $N_{\mu,\text{min}}$ is taken to be n_{asy} : $N_{\mu,\text{min}} = 4-5$, if two states are considered; $N_{\mu,\text{min}} = 7-8$, if three are considered; $N_{\mu,\text{min}} \approx 16$, if four states are considered.

For models and data of practical concern, one needs to consider sequences of 10 or more moments. In established treatments of inverting moment sequences, only a few (3–5) moments are considered. In the well-known method of moments, one selects a distribution function from a standard library, and its parameters are matched to the moments.¹⁰⁶ The generalized method of moments allows more flexibility in the choosing the function.¹⁰⁷ Alternatively, a series expansion around a base distribution can be matched to the moments.^{95-98,106} These parametric methods are generally used with continuous distributions. The primary nonparametric approaches have been based on the maximum-entropy method.^{91,93,94,100} It explicitly favors broad, continuous distributions. Our problem differs from this previous work: we seek a nonparametric method that works well on long moment sequences and that is agnostic about whether the distribution is discrete, continuous or a mixture of the two.

2. Finding an initial solution

We first discretize Eq. (21). A vector of N_P points is taken along the signal axis, $\mathbf{S} = \{S_i\}$ with $S_1 = S_{\text{min}}$, and $S_{N_P} = S_{\text{max}}$. The recovered distribution is given by the vector of values at these points, $\mathbf{P}_r(\mathbf{S}) = \{P_r(S_i)\}$. It should satisfy the equation

$$\boldsymbol{\mu}'[D] = \mathbf{V}(\mathbf{S}) \cdot \mathbf{P}_r, \quad (28)$$

where the Vandermonde matrix $\mathbf{V}(\mathbf{S})$ is defined by¹⁰⁸

$$\mathbf{V}(\mathbf{S}) = \begin{pmatrix} \mathbf{S}^0 \\ \mathbf{S}^1 \\ \vdots \\ \mathbf{S}^{N_\mu} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ S_1 & S_2 & \dots & S_{N_P} \\ \vdots & \vdots & \ddots & \vdots \\ S_1^{N_\mu} & S_2^{N_\mu} & \dots & S_{N_P}^{N_\mu} \end{pmatrix}, \quad (29)$$

and N_μ is the number of moments used. In our calculations, the points are equally spaced by a distance δS , although evenly spaced point are not required. We also use $S_{\text{min}} = 0.0125$, $S_{\text{max}} = 2.0125$, and $N_P = 81$ ($\delta S = 0.025$). These values were chosen so the points in the calculation do not accidentally lie at the exact positions of the model states. Calculations with a greater point density ($N_P = 161$, $\delta S = 0.0125$) show no significant changes (supplementary material, Fig. S2).

Even with an infinite number of moments, inverting $\mathbf{V}(\mathbf{S})$ is ill-conditioned: the solution is extremely sensitive to errors in the moments.¹⁰⁸ With a finite number of moments, the problem is also underdetermined: the number of moments is much smaller than the number of points in the solution, $N_P \gg N_\mu$. Both problems are addressed by looking for a solution that nearly solves Eq. (28), but that also satisfies certain “regularizing” conditions.¹⁰⁹ As the measure of the distance from an exact solution, we define a fitting error based the fractional error in the measured moments,

$$\chi(\mathbf{P}_r)^2 = \frac{1}{N_\mu} \sum_{n=1}^{N_\mu} \left[\frac{(\mathbf{V} \cdot \mathbf{P}_r)_n - \mu'_n[D]}{\mu'_n[D]} \right]^2. \quad (30)$$

The correct solution will have a error of χ^* , which is equal to the (unknown) error in the input moments. A simple minimization of χ will drive it to a value below this value, a process known as “fitting the error.”

One set of regularizing constraints has been implicitly imposed when the solution was confined to be between $S_{\min}$ and $S_{\max}$. We assume that our experiment is measuring light intensities, so we take $S_{\min}$ near zero with confidence. Visual inspection of Fig. 5 gives $S_{\max} \approx 1$ from the zero slope of the asymptote and $\log_{10} P_{S,\max} \approx -0.47$ from its intercept. We take $S_{\max}$ near two to be conservative. This assumption will be verified if the recovered distribution approaches zero before reaching $S_{\max}$. If not, the data themselves will have indicated that $S_{\max}$ needs to be increased. Thus, we have strong confidence in this boundary as well.

The most important regularization conditions in our problem prove to be additional boundary constraints. Specifically, we add the requirement that all the elements of $\mathbf{P}_r$ be nonnegative, consistent with their role as probabilities. With the same justification, we require that the zeroth moment be exactly one, equivalent to $\mathbf{P}_r$ being a normalized probability distribution. (Probabilities must also be less than one, but this result is already implied by these two constraints.) General treatments of regularization often emphasize regularization functions. However, using regularizing functions without these boundaries can result in wildly incorrect solutions and unrealistically small fitting errors (supplementary material, Sec. SII).

To implement these boundary conditions, the first element of $\boldsymbol{\mu}'$ and the first row of $\mathbf{V}$ are dropped in Eq. (30) and nonnegativity and normalization constraints are added to the minimization of χ :

$$\begin{aligned} \min_{\mathbf{P}_r} \chi(\mathbf{P}_r)^2 \quad \text{subject to} \quad & \sum_{i=1}^{N_P} P_{r,i} = 1 \\ & \text{and} \quad P_{r,i} \geq 0; \quad i = 1, \dots, N_P. \end{aligned} \quad (31)$$

This problem, as well as its extensions below, is a quadratic minimization with linear constraints and constant boundaries. Fast and efficient algorithms are available for this problem.¹¹⁰ See the supplementary material, Sec. SIII for details.

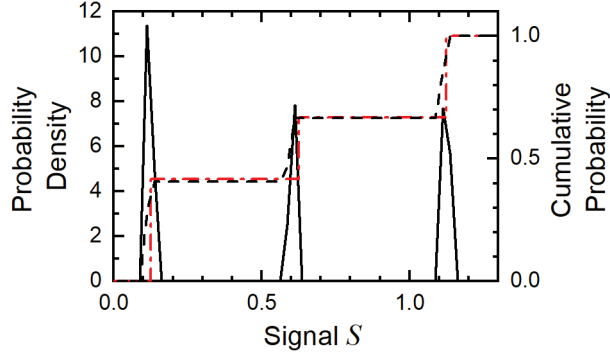


FIG. 7. The recovered probability distribution P_r (solid, black peaks) and cumulative distribution F_r (black, dashed-dotted line) using only nonnegativity and unit-area constraints [Eq. (31) or Eq. (32) with $\alpha = \beta = 0$] with $N_\mu = 13$ moments. The error in fitting the measured moments is $\chi_0 = 0.8 \times 10^{-4}$. The cumulative distribution of the correct solution F_S is the red, dashed line.

Figure 7 shows the solution to Eq. (31) using $N_\mu = 13$ noise-corrected moments from our data (Fig. 5). The error in matching the measured moments, $\chi_0 = 0.8 \times 10^{-4}$, is small, but much larger than for the unconstrained minimum ($\chi = 8.8 \times 10^{-14}$, supplementary material Fig. S1). (We use χ_0 to denote the fitting error for the solution regularized only with boundary conditions.) Comparing the recovered cumulative distribution (dash-dotted, black line) with the known solution (dashed, red line), we see that the solution is accurate both qualitatively and quantitatively: it correctly identifies the model as having three discrete states, the peak areas are correct, and the correct positions are within the recovered linewidths.

3. Exploring the error range

If we were working with real experimental data, we would not know whether this solution and its value of χ_0 were correct. For example, by the arguments given by Tibshirani in connection with LASSO regularization,¹¹¹ a cusp in the solution domain favors sparse solutions (solutions with many zero-probability points). The boundary conditions that we have imposed create such cusps at $P_r(S_{\min}) = 0$ and $P_r(S_{\max}) = 0$. Does this effect suppress interpeak intensity or artificially narrow the peaks in Fig. 7? Or should the peaks be even narrower? Would a solution more like the histogram in Fig. 3 be equally consistent with the data? We need to explore the range of other solutions that fit the data nearly as well as the solution to Eq. (31).

To do this, we combine regularizing functions $R_i(\mathbf{P}_r)$ with the boundary constraints in Eq. (31) :

$$\min_{\mathbf{P}_r} \left[\chi(\mathbf{P}_r)^2 + \beta R_1(\mathbf{P}_r) + \alpha R_2(\mathbf{P}_r) \right] \quad \text{subject to} \quad \sum_{i=1}^{N_p} P_{r,i} = 1$$

$$\text{and} \quad P_{r,i} \geq 0; i = 1, \dots, N_p. \quad (32)$$

Regularizing functions encode prior knowledge or assumptions about the correct solution. Increasing values of β and α reflect increasing confidence in the prior assumptions and will push χ to higher values. The values of β and α themselves are not meaningful, but the ratio χ/χ_0 reflects the reduction in the fit to the data. Thus, the range of solutions with χ/χ_0 modestly higher than one defines the uncertainty range of the recovered distribution.

The coefficients' relationships to the fitting error, $\beta(\chi)$ and $\alpha(\chi)$, are initially unknown. Values of β and α are assumed, and the implied value of χ is calculated from the minimized solution. An example of $\beta(\chi)$ is given in the supplementary material [Fig. S4(a)].

Regularization is used in a broad range of inverse and machine-learning problems, and many regularization functions are available.¹¹² In general, there are two competing ideas about what makes a better distribution, and so, we have included two regularizing functions. The idea that a smoother, broader distribution is better is expressed by the smoothing regularizer $R_1(\mathbf{P}_r)$. We choose it to be

$$R_1(\mathbf{P}_r) = \frac{1}{N_p - 1} \|\mathbf{D} \cdot \mathbf{P}_r\|^2, \quad (33)$$

where

$$\mathbf{D} = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix}. \quad (34)$$

The matrix $\mathbf{D}$ is a discrete approximation to the first derivative of the solution. The function $R_1(\mathbf{P}_r)$ measures the mean-squared magnitude of the solution's derivative and is a standard form in Tikhonov regularization.¹⁰⁹

Figure 8 looks at solutions with varying smoothness and breadth. By increasing β in Eq. (32) (with $\alpha = 0$), χ/χ_0 increases from 1 to 2.5. The acceptable fitting error is not perfectly defined, but this range covers the likely possibilities. Although the peaks broaden, they remain distinct and well-separated by near-zero-intensity regions. The peak areas change very little. These are robust features of the solution that cannot be eliminated without losing the fit to the data.

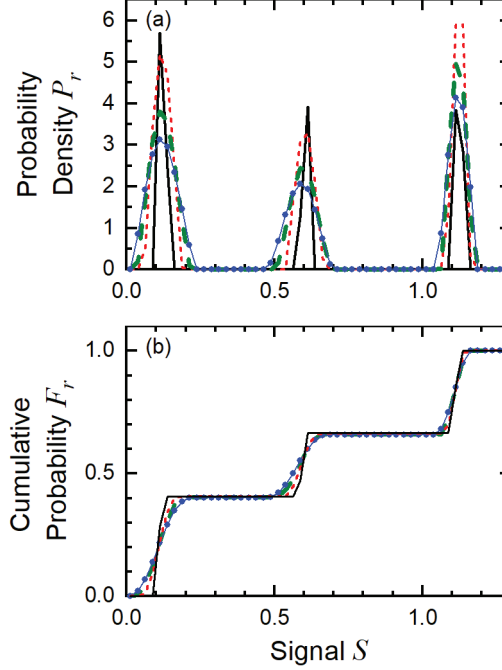


FIG. 8. Effect of the smoothing regularizer $R_1(\mathbf{P}_r)$ on the recovered distribution. The fitting error increases with increasing β [Eq. (32), $\alpha = 0$, $N_\mu = 13$]: $\chi/\chi_0 = 1$ (thin, black curve; from Fig. 7), 1.22 (short-dashed, red curve), 1.8 (dashed, green curve), 2.5 (blue points and line). The $\chi/\chi_0 = 1$ probability density [thin, black curve; panel (a)] has been divided by 2.0. (a) Probability densities and (b) cumulative probabilities.

Alternatively, the simplest and best solution can be taken to be the one with the smallest number of variables, parameters, or states.¹¹³ In our case, we favor solutions with large regions of zero or near-zero probability (sparse solutions) with $R_2(\mathbf{P}_r)$. This sparsity regularizer can narrow large peaks, eliminate small ones, or suppress baseline noise. Although the particular solutions in Fig. 8 are already quite sparse, we will demonstrate its effect for use in other problems.

A popular sparsity regularizer is the LASSO function,¹¹¹

$$R_3(\mathbf{P}_r) = \sum_{i=1}^{N_p} |P_{r,i}|. \quad (35)$$

This function cannot be used in the current problem because its value is fixed by the nonnegativity and normalization constraints in Eq. (32). As an alternative, we sum over only the points that are already low:

$$R_2(\mathbf{P}_r, P_{\text{sp}}) = \sum_{P_{r,i} < P_{\text{sp}}} P_{r,i}. \quad (36)$$

This function will try to take probability away from points with probability below P_{sp} and redistribute it to points with values larger than P_{sp} .

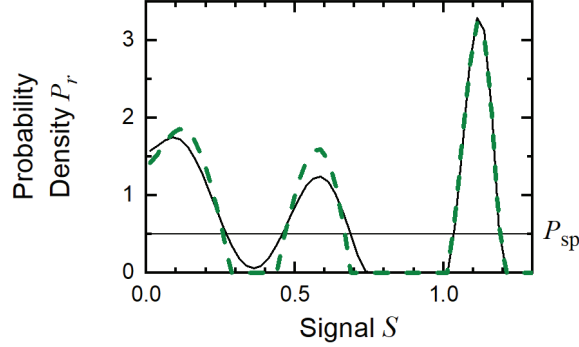


FIG. 9. Effect of the sparsity regularizer $R_2(\mathbf{P}_r; P_{\text{sp}} = 0.53)$ on the recovered distribution. The $N_\mu = 14$ probability density without the sparsity regularizer [thin, black curve; $\alpha = 0$; $\chi/\chi_0 = 1.21$; from Fig. 10(c)] and with the sparsity regularizer (dashed, green curve; $\chi/\chi_0 = 1.21$). The horizontal, grey line marks the level of P_{sp} .

As an example, Fig. 9 shows a solution from a moment sequence with an additional, high-error moment included [$N_\mu = 14$, see Fig. 10(c) below]. This error leads to a solution with broad peaks. In principle, there are two parameters to be adjusted, α and P_{sp} . In this particular problem, increasing α has only a small effect on χ . Thus, we take the large α limit, which forces all the points below P_{sp} to zero. As Fig. 9 shows the two peaks can be cleanly separated with very little increase in χ/χ_0 . Thus, there is no evidence in the data that the interpeak probability is real.

The final question is how many moments N_μ should be used in Eq. (32) [via Eq. (30)]. Above some order, the moments will have more error than new information, resulting in a poorer solution for a longer moment sequence. On the other hand, setting N_μ too small will throw away moments that do contain useful information, also reducing the quality of the result.

These effects are seen in Fig. 10, which shows results from our data for different values of N_μ . Using only $N_\mu = 6$ gives a noisy result [Figs. 10(a) and 10(e)]. Adding another one or two moments [$N_\mu = 7-8$, Figs. 10(a) and 10(e)] makes the distribution narrower and better resolved. We conclude that $\mu'_7[D]$ contains critical new information and that $\mu'_8[D]$ has some additional information. The minimum sequence length, $N_{\mu,\text{min}} = 7-8$, is consistent with our earlier estimate for a three-state system (see Fig. 6).

At the other extreme, the solutions with $N_\mu = 17-20$ [Figs. 10(d) and 10(h)] change erratically as moments are added. The moments above $n = 17$, have so much error that they degrade the solution. The moments from $n = 14-16$ [Figs. 10(c) and 10(g)] have less error, but enough to be detrimental. Thus, there is also a maximum number of usable moments, $N_{\mu,\text{max}} = 13$ in this case.

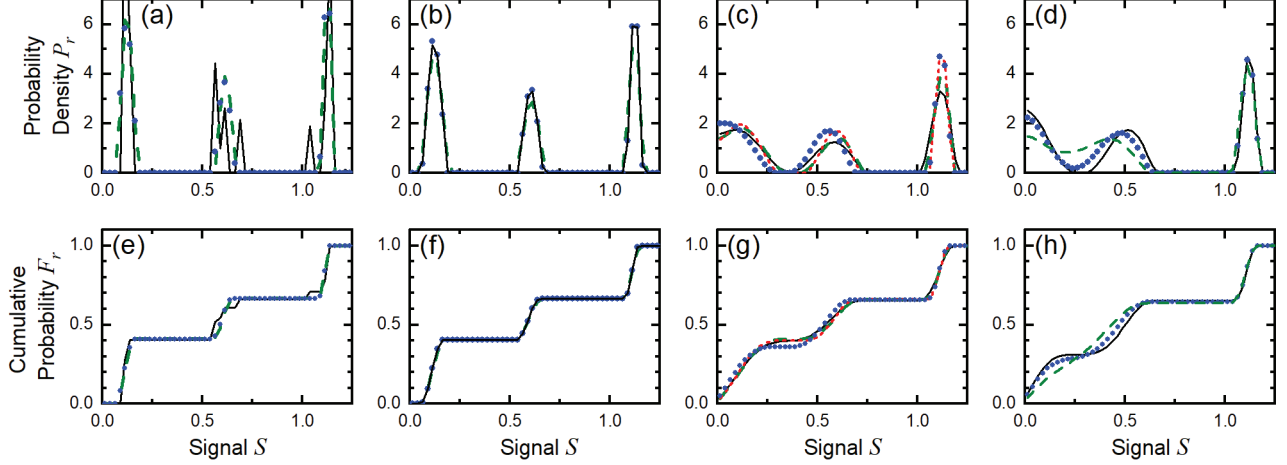


FIG. 10. Effect of the number of moments used N_μ on the recovered distribution. (a) and (e): $N_\mu = 6$ (thin, black curve), 7 (dashed, green curve), 8 (blue points). (b) and (f): $N_\mu = 10$ (dashed, green curve), 12 (blue points), 13 (thin, black curve). (c) and (g): $N_\mu = 14$ (thin, black curve), 15 (dashed, green curve), 16 (short dashed, red curve), 17 (blue points). (d) and (h): $N_\mu = 18$ (thin, black curve), 19 (blue points), 20 (dashed, green curve). Top row [(a)–(d)]: probability densities. Bottom row [(e)–(h)]: cumulative probabilities. The value of β has been adjusted to $\chi/\chi_0 \approx 1.22$ to smooth out minor differences between the solutions. Sparsity is not enforced ($\alpha = 0$).

Between $N_{\mu,\min}$ and $N_{\mu,\max}$ [$N_\mu = 8$ – 13 , Figs. 10(b) and 10(f)], the solutions are almost independent of the number of moments used. The moments $\mu'_9[D]$ – $\mu'_{13}[D]$ are accurate, but they are in the asymptotic region and do not better define a three-state system. Although, these moments do not change the solution, they remain useful. The fact that a fourth peak fails to appear up to $N_\mu = 13$ increases our confidence that it does not exist. We conclude that for this data set, $N_\mu = 13$ is optimal. (By chance, the break at this point is particularly striking in this realization of the time series. See Fig. S3 in the supplementary material.)

C. Final results

1. Equilibrium distribution

We have explored the space of solutions consistent with the data and spanned by the parameters N_μ , β , and P_{sp} . Throughout this space, the solutions have the same major physical features: three narrow peaks with unchanging positions and areas. The solution is not mathematically unique, but it is physically unique. To conclude, a single solution is chosen to represent the acceptable solutions. Statistical approaches to optimizing regularization parameters are not helpful (supplemental material, Sec. SIV), so we make a more subjective judgement.

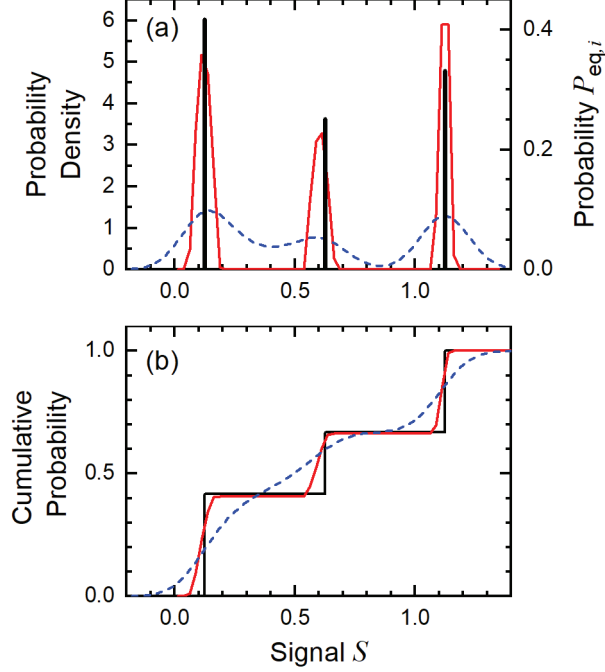


FIG. 11. Comparison of the distribution from a correlation-function approach $P_r(S)$ (solid, red curves) to one from time-binning and histogramming $P_D(S, N_b)$ (short-dashed, blue curves; from Fig. 3, $N_b = 100$) and to the known solution [(a) black bars, (b) solid, black lines; from Fig. 1]. The correlation-function solution is the $\chi/\chi_0 = 1.22$ result from Fig. 8 ($N_\mu = 13$, $\chi_0 = 0.8 \times 10^{-4}$, $\beta = 0.0014$, $\chi/\chi_0 = 1.22$, $\alpha = 0$). (a) Probability densities and probabilities and (b) cumulative probabilities.

Our choice is shown in Fig. 11 (red, solid curves), where it is compared to the original model from Fig. 1 (black bars and solid curve). The recovered distribution is not only qualitatively correct, but it is also quantitatively reliable: the mean positions of the states and the state populations are accurate. In contrast to a hidden-Markov fit, the peaks have finite widths. With real data, the widths would have indicated the extent of substructure within the states that would be consistent with the data.

This result comes from a data set that is noisy, $\text{SNR} = 0.58$, but otherwise favorable: the data set is large, $N_P = 10^8$; the instrument time-resolution is high compared to the fastest dynamics, $T_f/T_e = 100$; and the slow-dynamics are well sampled, $T_L/T_{\text{er}} = 10^4$. Despite the favorable data, time binning is not able to provide similar results. Figure 11 also shows the histogram with optimal time binning (blue, short dashed curve, see Fig. 3). The time resolution of this histogram is 100 times the instrumental time resolution, $T_r/T_e = 100$. The peaks are poorly resolved due to a combination of residual noise and imperfect time resolution. Figure 3 shows that the binning time would have to be increased by more than 30-fold to narrow the long-lived, high-intensity peak as

much as the correlation method does. Binning that drastic would reduce the time resolution of the experiment to 3000 times the instrument's time resolution.

The time resolution of the correlation method is not precisely defined. However, if even 10% of the decay was not resolved, 10% of the population of the two low-intensity peak would be averaged to values between the two peaks. The lack of interpeak intensity shows that $T_r / T_e < 10$. Thus, noise-removal by correlation methods offers major improvements over time binning in both time and state-space resolution, without introducing *a priori* restrictions on the solution.

2. System dynamics

With the equilibrium distribution known, the conversion of the noise-corrected moment-correlation functions can be converted to the system's Green's function. As the full details are covered in Ref. 87, we present only a brief, simplified conversion for purposes of illustration. The linewidths in the equilibrium distribution are ignored: three delta-function states are taken at the positions of the peaks in Fig. 11(a), with their probabilities given by the plateaus in Fig. 11(b). This distribution defines the orthogonal polynomials $Q_k(S)$ or "modes" for the system. Without linewidths, only three modes and only correlation functions only up to order $n = 4$ are required. The mode-correlation functions, $C_{kl}(\tau) = \langle Q_k(\tau)Q_l(0) \rangle$, were then calculated from the noise-corrected, moment-correlation functions (Fig. 4). In conjunction with the equilibrium distribution, the mode-correlation functions were converted to the Green's function of the signal $G(S_1|S_0; \tau)$.

The results are displayed in the top row in Fig. 12 as the more symmetrical, joint-probability distribution $P(S_1, S_0; \tau) = G(S_1|S_0; \tau)P_{\text{eq}}(S_0)$. Delta functions have been plotted as bars with their heights indicating their probabilities. At $\tau = 0$, the joint distribution should be diagonal, $P(S_1, S_0; 0) = \delta(S_1 - S_0)P_{\text{eq}}(S_0)$. The small, off-diagonal elements in Fig. 12(a) could have been eliminated by refitting our simplified model to the data. The utility of the joint distribution comes at $\tau = 100 \mu\text{s}$ and 1 ms [Figs. 12(b) and 12(c)], the mid- and end-points of the first decay in the autocorrelation function [Fig. 4(a)]. During this period, the cross-peaks between the low- and mid-intensity peaks grow, showing that the fast kinetics are between these two states. Cross-peaks with the high-intensity peak only grow in near $\tau = 10 \text{ ms}$ [Fig. 12(d)], the time of the second decay of the autocorrelation function. This decay can be assigned to the equilibration of this state. Thus, higher correlation functions allow the assignment of times to particular state-to-state transitions, a feature lacking in the autocorrelation function alone. At $\tau = 110 \text{ ms}$ [Fig. 12(e)], the distribution matches the equilibrium distribution, $P(S_1, S_0; 110 \text{ ms}) = P_{\text{eq}}(S_1)P_{\text{eq}}(S_0)$, confirming that the system equilibrates within this time.

An alternative analysis is shown in the bottom row of Fig. 12. The mode-correlation matrix can always be diagonalized at one delay, here chosen to be $\tau = 1 \text{ ms}$. Applying the same rotation to all times gives the eigenstate-correlation function $E_{kl}(\tau)$. If this matrix is diagonal at all times,

the system has eigenstates, and the diagonal elements $E_{kk}(\tau)$ are the eigendecays. As expected, this condition holds for our model, and the eigendecays are shown in Fig. 12(i). The corresponding right eigenstates of the Green's function $R_k(S_0)$ are derived from the diagonalizing matrix⁸⁷ and are shown in the bottom row of Fig. 12. As expected, the first eigenstate R_0 is the assumed equilibrium distribution [Fig. 12(f)], and its eigendecay E_{00} is constant [Fig. 12(i), black]. The second eigenstate R_1 is associated with transitions between the low- and mid-intensity states [Fig. 12(g)]. Its eigendecay E_{11} fits a single exponential with a time constant $T_f = (k_{12} + k_{21})^{-1} = 99 \mu\text{s}$, a good match to the true value of $100 \mu\text{s}$. The last eigenstate R_2 is associated with decay of the high-intensity state [Fig. 12(h)]. Its eigendecay is also exponential with a time constant $T_{\text{er}} = (k_{3(1/2)} + k_{(1/2)3})^{-1} = 10.0 \text{ ms}$, which again matches the true value. From these time constants and the state populations from Fig. 11, the original master equation of the system can be reconstructed.

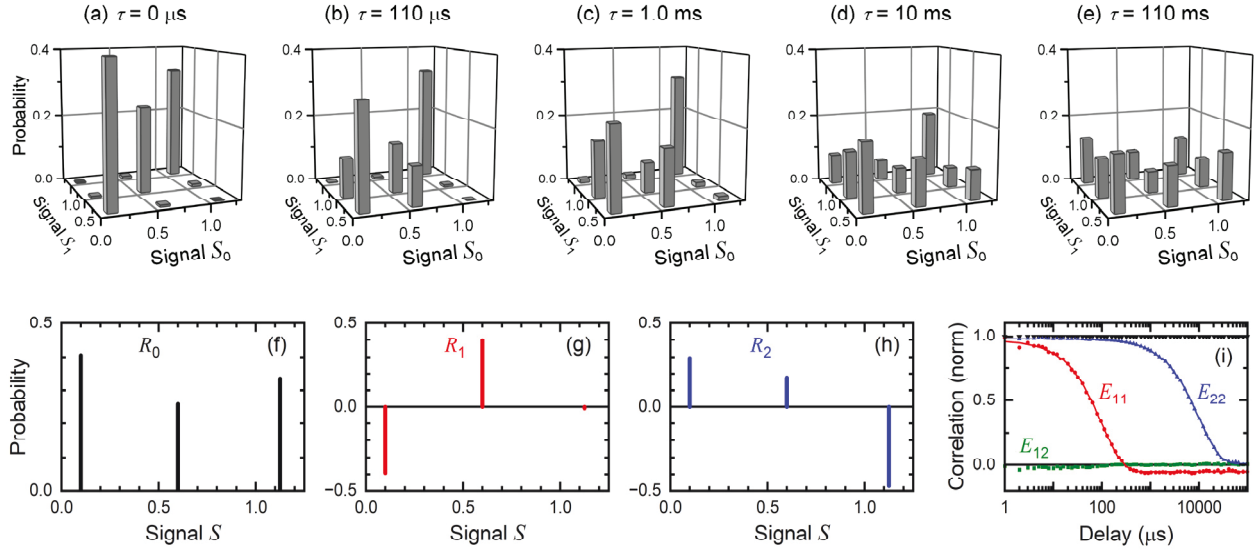


FIG. 12. Dynamic quantities extracted from the noise-corrected, moment-correlation functions (Fig. 4). Top row, (a)–(e): The joint-probability distribution $P(S_1, S_0; \tau)$. Bottom row: The eigenstates $R_k(S)$ [(f)–(h)] and their corresponding eigendecays $E_{kl}(\tau)$ [(i), points]. The eigendecays E_{11} (red circles) and E_{22} (green, up triangles) are fit with single exponentials (curves). The eigendecays E_{00} (black, down triangles) and E_{12} (green, squares) are constant.

[There is an ambiguity about which state the high-intensity state couples to. This ambiguity is inherent in kinetics when there is a wide timescale separation between decay times. If the decay times were close, the cross-peaks between the lower and highest intensity states in Fig. 12(d) would not rise at the same rate, and the ambiguity would be resolved.]

The fact that the eigendecays are single exponentials shows that the system is Markovian in the identified states and that no hidden states remain.⁸⁷ Further analysis with multidimensional-

correlation functions is not needed as it would only duplicate the information from the one-dimensional analysis. We can be confident that all the useful information in the data has been extracted.

IV. ROBUSTNESS OF CORRELATION METHODS

A. A floor in the usable SNR

The last section demonstrated that correlation methods work well for one noisy, but large, data set. This section explores the limits of this approach as the data quality decreases, in this subsection as the noise increases and in the next subsection as the data-set size decreases. To begin, we must distinguish between measurement noise and sampling error. Measurement noise is caused by the experimental apparatus and disappears when σ_ϵ goes to zero; sampling error is caused by the finite size of the data set and disappears when T_L goes to infinity. Sampling error manifests as residual fluctuations along the correlation function after noise correction [see Fig. 4(d)]. The issues are similar at all delays, so we focus on the error in the $\tau = 0$ moments and the resulting error in the recovered distribution.

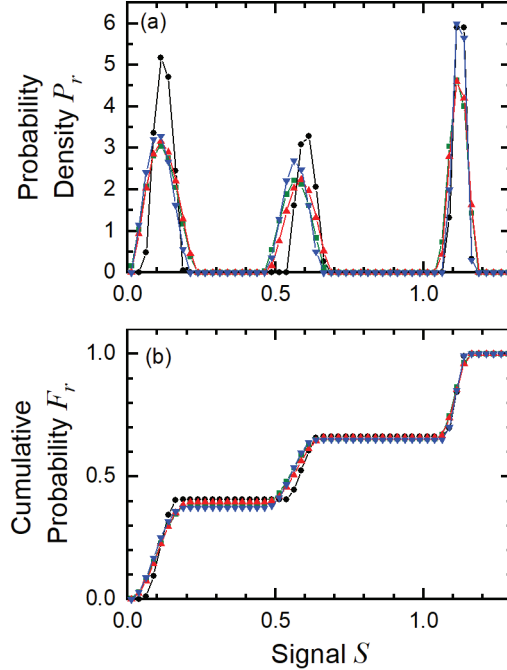


FIG. 13. Distributions recovered from four realizations of a long, medium-noise time series ($\text{SNR} = 0.58$, $N_t = 10^8$, $N_{\text{en}} = 1$). All cases were smoothed until $\chi/\chi_0 = 1.22$, sparsity was not enforced ($\alpha = 0$), and $N_\mu = 13$. (a) Probability densities and (b) cumulative probabilities. The result from Fig. 11 has black circles.

We first test the method's precision by looking at repeated measurements. Three additional realizations of the data were generated with the same parameters used in Sec. III.B, but with different seeds for the random-number generators for both signal and noise. Distributions recovered from each realization are shown in Fig. 13, along from the one presented in the previous section (Sec. 0). The differences are small and are the result of sampling error, not residual measurement noise.

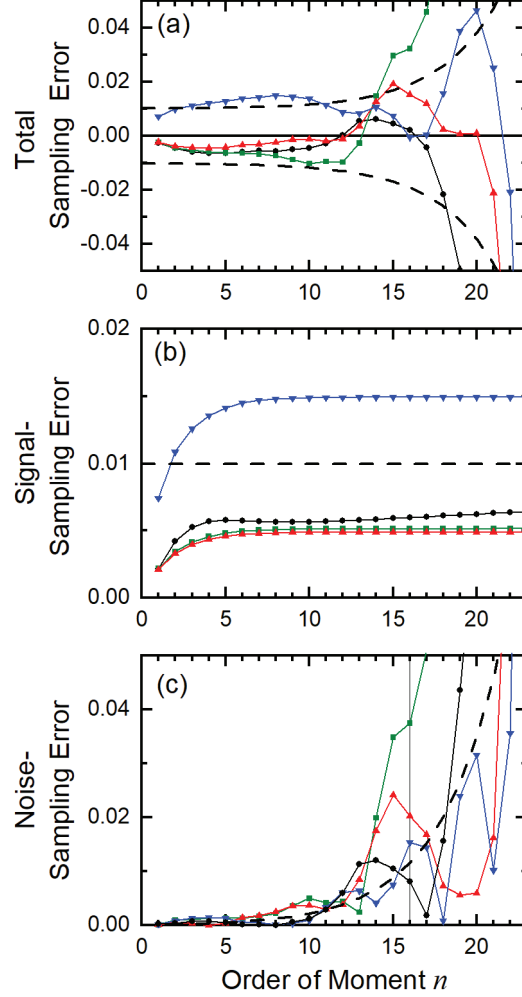


FIG. 14. Deviation of the measured moments from the true moments for the realizations used in Fig. 13 (same color and symbol code). (a) The total sampling error $\zeta_{n,i}$, (b) the absolute value of the signal-sampling error $|\zeta_{S,n,i}|$, and (c) the absolute value of the noise-sampling error $|\zeta_{\varepsilon,n,i}|$. The dashed, black curves are the predictions of the theory for (a) $\pm\omega_n$ [Eq. (38)], (b) ω_S [Eq. (39)], and (c) $\omega_{\varepsilon,n}$ [Eq. (40)].

We next look at the error contributed by individual moments. The sampling error $\xi_{n,i}$ in a specific moment is defined as the fractional deviation of a finite-length, noise-corrected moment from the value calculated analytically from the true distribution,

$$\xi_{n,i} = \frac{\mu'_n[D_i] - \mu_n[P_S]}{\mu_n[P_S]}. \quad (37)$$

The subscript i labels a specific realization. Figure 14(a) shows results from the realizations used in Fig. 13. As anticipated, the error increases for higher moments, but the pattern of increase is peculiar. For moments $n \lesssim 14$, the magnitude of the errors is small and nearly independent of the order. For $n \gtrsim 18$, the error is large and increasing rapidly with order. Thus, there is a crossover in the character of the sampling error. This crossover results in the maximum number of usable moments, $N_{\mu,\max} = 13$, being relatively well defined, as seen in Fig. 10.

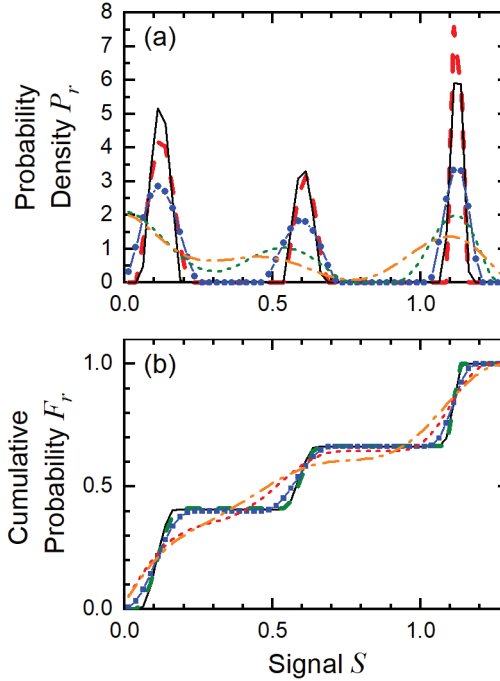


FIG. 15. Distributions recovered from long time series ($N_t = 10^8$, $N_{\text{en}} = 1$) with different signal-to-noise ratios: SNR = 1.16 (dashed, red curve), 0.58 (solid, black curve), 0.39 (blue points and line), 0.33 (short-dashed, green curve), and 0.29 (dash-dotted, orange curve) ($\sigma_\varepsilon = 0.5, 1.0, 1.5, 1.75$ and 2.0 , respectively). The numbers of moments used were $N_\mu = 15, 13, 8, 7$, and 7 , respectively. All cases were smoothed until $\chi/\chi_0 \approx 1.22$, and sparsity was not enforced ($\alpha = 0$). (a) Probability densities and (b) cumulative probabilities. The result from Fig. 11 has a black curve.

To see the effect of changing the signal-to-noise ratio, we constructed data sets with different noise levels. In addition to the case analyzed above with $\text{SNR} = 0.58$, three cases with higher noise; $\text{SNR} = 0.39$, 0.33 , and 0.29 ; and one with lower noise, $\text{SNR} = 1.16$, were examined. To eliminate variations in the sampling noise, we used a single realization of the signal and one of the noise and combined them in different ratios. The optimal number of moments was determined as in Fig. 10. The resulting distributions are shown in Fig. 15. There is a slow loss of resolution as the signal-to-noise level drops to $\text{SNR} = 0.39$, and then the resolution deteriorates rapidly for lower values. Thus, there is a floor to the acceptable signal-to-noise ratio for a given data-set size.

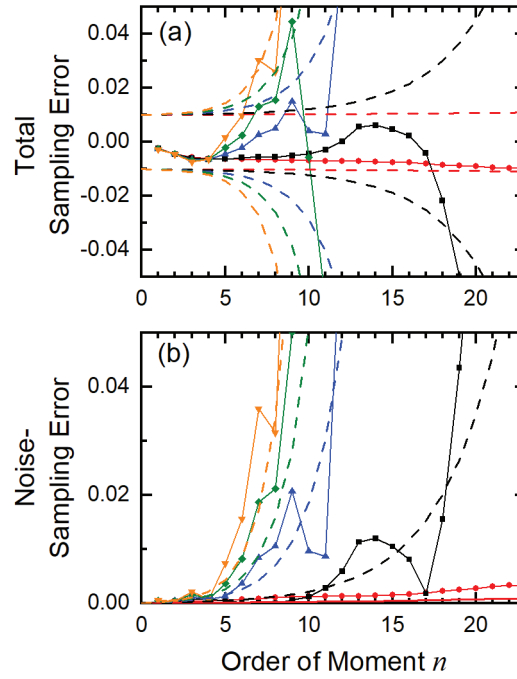


FIG. 16. Deviation of the measured moments from the true moments for the time series used in Fig. 15: $\text{SNR} = 1.16$ (red circles), 0.58 (black squares), 0.39 (blue, up triangles), 0.33 (green diamonds), and 0.29 (orange, down triangles). (a) The total sampling error $\zeta_{n,i}$ and (b) the absolute value of the noise-sampling error $|\zeta_{\varepsilon,n,i}|$. [See Fig. 14(b) (black circles) for the signal-sampling error.] The dashed curves are the predictions of the theory for (a) $\pm\omega_n$ [Eq. (38)] and (b) $\omega_{\varepsilon,n}$ [Eq. (40)].

The same crossover in sampling error with order seen in Fig. 14(a) is also seen for different signal-to-noise ratios [Fig. 16(a)]. As the signal-to-noise ratio drops, the crossover occurs at smaller orders. As a result, the maximum number of usable moments drops from $N_{\mu,\text{max}} = 15$ (or more) with $\text{SNR} = 1.16$ to $N_{\mu,\text{max}} = 6$ with $\text{SNR} = 0.27$. Figure 6 estimated that a minimum of $N_{\mu,\text{min}} = 7-8$ moments are needed to accurately recover a three-state system. When the signal-to-

noise ratio limits the number of accurate moments to below this minimum, the distribution recovery deteriorates. Thus, the floor in the acceptable signal-to-noise ratio is directly linked to the crossover from constant to rapidly rising sampling error.

B. Limits to data-set size and time resolution from sampling error

An explanation of the crossover and its dependence on the signal-to-noise ratio requires a more quantitative analysis of sampling error and how it varies with the quality of the data set. Data-set quality depends on its signal-to-noise ratio, its dynamic range in time, $N_t = T_L / T_\varepsilon$, and its total size N_{er} . Although our examples above consisted of a single time series, more generally, one collects data from an ensemble of N_{en} series from different members of the sample, so the total data-set size is $N_{\text{er}} = N_t N_{\text{en}}$. (See Sec. IV.C for more discussion of ensemble averaging.)

In addition to time and ensemble averages over one realization of the data, we also need to consider statistics over many hypothetical realizations of the data. The supplemental material (Sec. SV) shows that noise removal is unbiased. It also calculates the expected variance of the n th-order sampling error, $\omega_n^2 = \text{Var}(\xi_n)$. (Previous work has looked at the variance of uncorrected, linear, auto- and cross-correlation functions.^{105,114}) The result can be divided into two terms:

$$\omega_n^2 = \omega_S^2 + \omega_{\varepsilon,n}^2, \quad (38)$$

the signal-sampling error ω_S and the noise-sampling error $\omega_{\varepsilon,n}$. The signal-sampling error results from incomplete sampling of the signal distribution. It occurs even in the absence of measurement noise and does not depend on the signal-to-noise ratio. In contrast, the noise-sampling error goes to zero as the signal-to-noise ratio becomes large. It results from the need to effectively sample the noise distribution. Although we are not interested in the noise distribution, Eq. (11) makes it clear that noise correction is really a blind separation of the data into two components, signal and noise. Consequently, information about the noise distribution is just as important as information about the signal distribution. This point is further emphasized later in Sec. V, which explicitly determines the noise distribution from the same moments we have been using.

Asymptotic approximations for the two sources of sampling noise (supplementary material, Sec. SV) are sufficient to illustrate the important trends:

$$\omega_S^2 = \frac{2a-1}{N_{\text{en}} \left(1 + \frac{\langle T_L \rangle}{T_{\text{er}}} \right)} \quad (39)$$

and

$$\omega_{\varepsilon,n}^2 = \frac{2a}{N_{\text{er}}} \left[\left(1 + \frac{1}{(2b\text{SNR})^2} \right)^n - 1 \right]. \quad (40)$$

The constants a and b depend on the details of the distribution, but are of order unity. Both types of error go to zero as the size of the data set goes to infinity, as expected. More importantly, the signal-sampling error is independent of the order of moment, whereas the noise-sampling error is exponential in the error.

To test these predictions against our data, the total error for each realization was decomposed by extending Eq. (38) to individual time series,

$$\xi_{n,i}^2 = \xi_{S,n,i}^2 + \xi_{\varepsilon,n,i}^2. \quad (41)$$

The signal-sampling errors $\xi_{S,n,i}$ were calculated from the noise-free time series and are plotted in Fig. 14(b). The noise-sampling errors $\xi_{\varepsilon,n,i}$ were then calculated from Eq. (41) using the values from Figs. 14(a), 16(a), and 14(b). The results are plotted in Fig. 14(c) and 16(b) as colored lines with symbols. The predictions of Eqs. (38)–(40) are plotted as dashed curves. Because a and b would not be known with real data, we have used generic values of $a = b = 1$. The predictions are of the right magnitude and show the correct trends with order and signal-to-noise ratio.

The combination of these effects explains the crossover in the total-sampling error. In Figs. 14(a) and 16(a), the constant, signal-sampling error dominates at low n . Above some order, the exponential dependence of the noise-sampling error takes over, and the total rises rapidly. The position of this crossover as a function of the signal-to-noise ratio is correctly predicted [Fig. 16(a)]. Thus, this phenomenon is not specific to our examples; it is a general effect.

In real experiments, the errors of individual moments are not known; only the maximum number of useful moments $N_{\mu,\text{max}}$ is discernable. Before estimating its value, we note that the signal-sampling error has mild consequences. Signal-sampling error creates a moment sequence that corresponds to a physically realistic distribution with the correct number of states and the correct state positions. In contrast, noise-sampling error can distort a moment sequence in any way possible. As a result, false peaks can appear, state widths can be broadened, or peak positions and areas can be incorrect. Indeed, noise-sampling error may create a moment sequence that does not correspond to any real distribution.

Thus, we focus on the stronger effects of noise-sampling error. The maximum error that a moment can have without degrading the recovery is $\omega_{\varepsilon,\text{max}}$. As the order increases and the noise-sampling error rises, there will be a maximum order that does not have excessive error n_{max} , which is defined as the order where the noise-sampling error is equal to $\omega_{\varepsilon,\text{max}}$. It is the counterpart to n_{asy} , defined in Eq. (27): n_{asy} defines the largest moment that provides new information about the

distribution greater than the error in the moment; $n_{\max}$ defines the largest moment that will not disrupt the distribution recovery because of its inaccuracy. In between, the moments are accurate, but redundant. These values then define the minimum and maximum sequence lengths, $N_{\mu,\min} = n_{\text{asy}}$ and $N_{\mu,\max} = n_{\max}$, that produce accurate distributions. So long as $N_{\mu,\max} > N_{\mu,\min}$, it is still possible find an N_{μ} that gives a good recovery.

These definitions can be put into Eq. (40). Dropping the right-hand one and taking $a = b = 1$, the value of $N_{\mu,\max}$ is estimated to be

$$N_{\mu,\max} = \frac{\ln\left(\frac{1}{2} N_{\text{er}} \omega_{\varepsilon,\max}^2\right)}{\ln\left[1 + (2\text{SNR})^{-2}\right]}. \quad (42)$$

A value of $\omega_{\varepsilon,\max} = 5 \times 10^{-3}$ matches our results well. This value gives $N_{\mu,\max} = 13$ [Eq. (42)] and $N_{\mu,\min} = 7-8$ [Fig. 6, Eq. (27)]. Figure 10 shows accurate, stable distributions are recovered in this range, and poor solutions are found outside it. As the signal-to-noise ratio drops from 1.16 to 0.29 in Fig. 15, $N_{\mu,\max}$ drops: $N_{\mu,\max} = 42, 13, 7.3, 6.0, 5.2$. Good distributions are found for the three high signal-to-noise cases, where $N_{\mu,\max} \geq N_{\mu,\min}$, and poor distributions are recovered for the two low signal-to-noise cases, where $N_{\mu,\max} < N_{\mu,\min}$. This formula is accurate enough to estimate the position of the signal-to-noise floor.

With these formulas, we can address the fundamental questions posed in the introduction. First, can a larger data set compensate for an arbitrarily low signal-to-noise ratio? The restriction on data-set size and signal-to-noise ratio is given by the requirement that $\omega_{\varepsilon,n} \leq \omega_{\varepsilon,\max}$ for $n = N_{\mu,\min}$. Putting this requirement into Eq. (40), gives different behavior for high and low signal-to-noise ratios. In the low-noise limit, $\text{SNR} \gg 1/2$,

$$\left(\frac{N_{\text{er}}}{N_{\mu,\min}}\right)^{1/2} 2\text{SNR} \gtrsim \sqrt{2} \omega_{\varepsilon,\max}^{-1}, \quad (43)$$

whereas in the high-noise limit, $\text{SNR} \ll 1/2$,

$$N_{\text{er}}^{1/2} (2\text{SNR})^{N_{\mu,\min}} \gtrsim 2 \omega_{\varepsilon,\max}^{-1}. \quad (44)$$

In either case, an arbitrarily low SNR can be compensated by an arbitrarily large data-set size. In the high signal-to-noise ratio regime [Eq. (43)], the data-set size must be increased by the square of the decrease in signal-to-noise ratio. This behavior is typical. However, compensation is impractical in the low signal-to-noise regime. In Eq. (44), SNR is raised to a high power. In the case we have been looking at with $N_{\mu,\min} = 7$, a 2-fold decrease in signal-to-noise ratio requires a 1.6×10^4 -fold increase in the data-set size. Thus, there is a minimum $\text{SNR} \approx 1/2$, below which it is impractical to collect enough data to make up for a high noise level.

The other fundamental question was: How is the time resolution of the experiment T_r related to the time resolution of the instrument T_ε ? If the signal-to-noise ratio is above one-half, the main restriction on the time resolution is the time averaging in the primed powers [Eq. (19)]. A simple average over the time range covered by a moment sequence gives an estimate of $T_r/T_\varepsilon = \frac{1}{2}(N_{\mu,\min} + 1)$. Equations (43) and (44) imply that if the experimental signal-to-noise ratio is below one-half, then the data must be binned to reach this value. The extent of binning needed is $N_b = (2\text{SNR})^{-2}$. Thus, the ratio of the experimental time resolution to the measurement time resolution is approximately

$$\frac{T_r}{T_\varepsilon} \approx \frac{N_{\mu,\min} + 1}{2} \begin{cases} 1; & \text{SNR} \gtrsim \frac{1}{2} \\ 1/(2\text{SNR})^2; & \text{SNR} \lesssim \frac{1}{2} \end{cases}. \quad (45)$$

Time binning shows a conflict between noise removal and time resolution and creates the paradox of the experimental time resolution being orders-of-magnitude lower than the instrumental resolution. These problems are eliminated. The time resolution of a nonparametric method can be similar to that of a more restricted, parametric analysis.³¹

C. Using smaller data sets and shorter series

It remains to test our methods on smaller data sets, closer to the limits implied by the theory presented above and closer to common experimental conditions. The effects of time range N_t , data-set size N_{er} , ensemble size N_{en} , and signal-to-noise ratio SNR are all intertwined. Rather than a full exploration of these parameters, we look at two, smaller data sets that illustrate the most salient points: a medium-sized data set with 100 times fewer data points ($N_{\text{er}} = 10^6$) than in the earlier, large sets ($N_{\text{er}} = 10^8$), and a small data set with 10^4 times fewer points ($N_{\text{er}} = 10^4$). The new data sets also contain $N_{\text{en}} = 100$ series, in contrast to the single series ($N_{\text{en}} = 1$) in the earlier sets. Thus, we will see the consequences of ensemble analysis.

The new data sets have much shorter series as well. The average series length in the medium set is 10^4 times shorter ($\langle N_t \rangle = 10^4$) than in the previous examples ($N_t = 10^8$). This length is only as long as the slowest relaxation time of the system, $\langle T_L \rangle = T_{\text{er}} = 10$ ms. The series lengths in the small data set are 10^6 shorter ($N_t = 10^2$) than in the earlier sets. This length is only as long as the fastest relaxation time in the system, $\langle T_L \rangle = T_f = 100$ μs .

Finally, the medium data set has nonuniform series lengths, a common feature in experimental data. The lengths of individual series $T_{L,i}$ were chosen from an exponential with a time constant of T_{er} . Series shorter than $\frac{1}{3}T_{\text{er}}$ or longer than $3T_{\text{er}}$ were rejected. This set mimics the effects of photobleaching in single-molecule experiments.

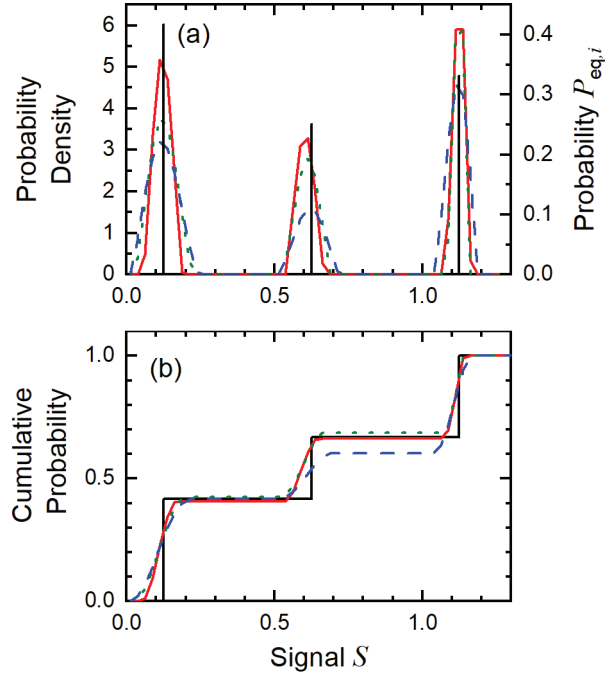


FIG. 17. Distributions recovered from data sets of various size [$N_{er} = 10^8$ (red, solid curve), 10^6 (green, dashed curve), 10^4 (blue, dotted curve)] compared to the known solution [(a) black bars, (b) solid, black lines; from Fig. 1]. The number of data points was decreased through a combination of reducing the time range of the series [$N_t = 10^8$ (red), 10^4 (green), 10^4 (blue)], increasing the number of series measured [$N_{en} = 1$ (red), 100 (green), 100 (blue)] and increasing the signal-to-noise ratio [SNR = 0.58 (red), 1.16 (green), 1.75 (blue)]. (a) Probability densities and (b) cumulative probabilities. Recovery was done with the $N_\mu = 13$ (red), 14 (green), 13 (blue), $\chi/\chi_0 = 1.22$, $\alpha = 0$.

Distributions recovered from the large, medium and small data sets are compared to the original model in Fig. 17. To get acceptable results, the signal-to-noise ratio was raised as the size decreased: compared to the large sets (SNR = 0.58 and $N_{er} = 10^8$), the signal in the medium set was doubled when the size was cut 100-fold (SNR = 1.16 and $N_{er} = 10^6$), and the signal in the small set was trebled when the size was cut 10^4 -fold (SNR = 1.75 and $N_{er} = 10^4$). The sizes and signal-to-noise ratios satisfy the inequalities required by Eq. (43): for the medium set, $880 > 280$, and for the small set, $130 \approx 280$ (compared to $4400 \gg 280$ for the large set). The number of states and the peak positions are always accurate [Fig. 17(a)]. Because the small data set is crossing this limit, there is some error in the peak areas [Fig. 17(b)], and the peaks are slightly broader. Overall, a much smaller data set can be compensated by a modest increase in the signal-to-noise ratio.

Turning to the correlation functions, the total number of data points is not a full description of data quality: how the size is divided between series length and number of series is also important.

The role of ensemble analysis is not always clear in the literature. Many parametric methods do a simultaneous fit over an ensemble of time series.³⁷ Correlation methods naturally use an ergodic average, one mixing time and ensemble averages. In this paper, Eqs. (23) and (24) are based on time and ensemble averages, respectively, and are taken to be equivalent. This ergodic assumption should hold for systems at equilibrium or in a stable, steady state.

However, ensemble averaging is sometimes portrayed as antithetical to single-molecule measurements.^{82,115-119} One objection is that time series themselves cannot be ensemble averaged without losing essential information. However, by not analyzing series directly, correlation methods avoid this problem. Another objection is that the linear-correlation function—as well as the 1D, nonlinear-correlation functions treated here—average over any heterogeneity in the ensemble. However, it has now been shown that the information on heterogeneity is contained in (ergodically-averaged) two- and three-dimensional correlation functions.^{74-78,80,81}

Skepticism about ensemble averaging has led to the use of “single-molecule” correlation functions $\tilde{M}'_{kl}[D_i](\tau)$, which are restricted to time averaging over a single series,^{3,102-104,120-125}

$$\tilde{M}'_{kl}[D_i](\tau) = \frac{\langle \delta D_i(\tau)^{k'} \delta D_i(0)^{l'} \rangle}{\langle \delta D_i^{(k+l)'} \rangle}, \quad (46)$$

where different series are identified by the subscript i . Following conventional practice, the data have also been centered by subtracting the mean calculated from the same time series,

$$\delta D_i(t) = D_i(t) - \langle D_i \rangle. \quad (47)$$

Single-molecule averages work well, if the series are sufficiently long. For example, in our large data set, the long length ($T_L = 10^4 T_{\text{er}}$) reduces the signal-sampling error to only 2.6% [Eq. (39)]. As a result, both linear- and nonlinear-correlation functions are quite good (Fig. 4). (However, data centering [Eq. (47)] does not give a long-time limit of zero for nonlinear-correlation functions, as it does for linear ones.)

On the other hand, single-molecule correlation functions become susceptible to signal-sampling error when the series are short. Vanden Bout and Kaufman have detailed this problem for the linear case.¹⁰²⁻¹⁰⁴ In our medium-sized data set, where the slow relaxation is poorly sampled ($\langle T_L \rangle = T_{\text{er}}$), the estimated signal-sampling error for a single series is 180% [Eq. (39)]. Figures 18(a) and 18(b) show selected single-molecule correlation functions, linear and nonlinear respectively, from this data set. The linear-correlation functions fluctuate widely and are strongly biased toward short relaxation times. The nonlinear-correlation functions deteriorate very rapidly with increasing order; the 4th-order case shown in Fig. 18(b) is the highest that could be meaningfully plotted. Moreover, when series are short, neither single-molecule correlation

functions themselves nor statistics derived from them can be subsequently ensemble averaged because of the bias introduced by the nonlinear manipulations in Eq. (46).

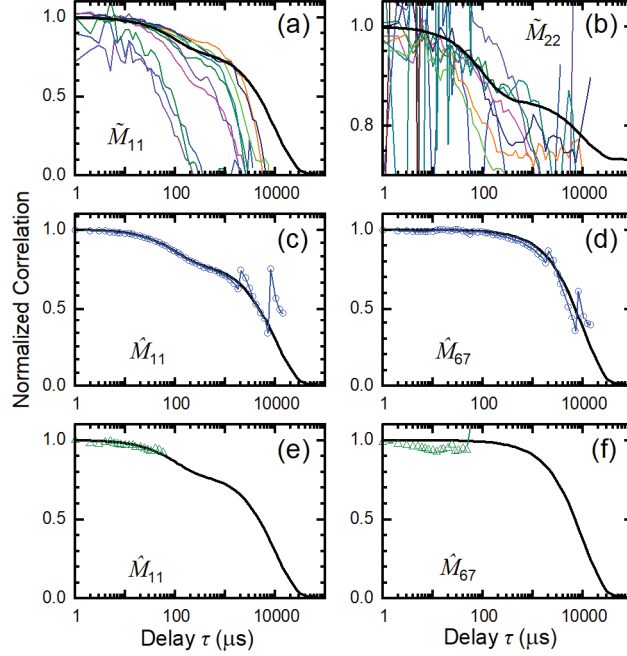


FIG. 18. Normalized correlation functions from shorter time series. In each subpanel, the full, black curve is the correct function. In (a) and (b), the single-molecule correlation functions for selected short time series ($\langle T_L \rangle = 10^4 \mu\text{s}$, medium data set) are shown as thin, colored curves. Ensemble-averaged moment-correlation functions are shown for short series ($\langle T_L \rangle = 10^4 \mu\text{s}$, medium data set) in (c) and (d) (blue circles) and for very short series ($T_L = 10^2 \mu\text{s}$, small data set) in (e) and (f) (green triangles).

To allow the use of short time series, uncentered, unnormalized moment-correlation functions [Eq. (3)] have been used so far. These can be ensemble averaged without bias. Because we have already looked at the distributions derived from the zero-time points of these functions, we can now focus on the kinetics alone, which are represented by normalized moment-correlation functions,

$$\hat{M}_{kl}[D](\tau) = \frac{M'_{kl}[D](\tau) - \mu'_k[D]\mu'_l[D]}{\mu'_{k+l}[D] - \mu'_k[D]\mu'_l[D]}. \quad (48)$$

To avoid bias, the normalization at $\tau = 0$ and zeroing at $\tau = \infty$ are only done with ensemble-averaged quantities.

Figures 18(c) and 18(d) compare these functions from the medium-sized data set with the full correlation function. Unavoidably, the slow relaxation is incomplete because few of the individual

series span the relaxation. However, the dynamics within the series length are correctly represented, including the whole of the fast exchange between the two, low intensity states. Even the fact that slow relaxation is incomplete is evident because the correlation functions do not decay to zero. Because the simulated photobleaching reduces the number of series in the ensemble after $\tau = 3$ ms, the error increases noticeably after this time. However, the variation in series lengths introduces no fundamental problem in calculating the ergodic average or observing the early dynamics.

The small data set has series too short to yield any information on the kinetics [Figs. 18(e) and 18(f)]. Nonetheless, the distribution can be correctly recovered (Fig. 19). Intuitively, the static distribution should be measurable, even when the dynamics are not. It is satisfying to see the correlation methods match this expectation.

Overall, short time series do not create a problem for correlation methods. As in all methods, dynamics beyond the series length cannot be measured. Otherwise, short series can be compensated for by increasing the ensemble size. The demand on the total number of data points is moderate for moderately complex dynamics and is within the reach of many experiments.

D. Test on experimental data

Synthetic data has the advantage that the truth behind the data is known, and the accuracy of the results can be judged. However, real experiments can have issues that are not included in our synthetic examples: nonadditive noise, drifting background, or other, unknown complications. To test for sensitivity to these issues, we analyzed FRET data that have been published by Schuler as a benchmark for time-series analysis.^{37,126} We compare our nonparametric results to the results of 14 parametric methods that have been tested on this data.³⁷

The data set results from the interaction of the nuclear-coactivator binding domain of CBP/p300 with the intrinsically disordered activation domain of the steroid receptor coactivator 3 and was measured by confocal, single-photon detection.¹²⁷ It consists of $N_{\text{en}} = 19$ FRET series with a time bin of $T_e = 10$ ms and an average length of $\langle T_L \rangle = 119$ s, for a total of $N_{\text{er}} = 226,100$ points. The signal-to-noise ratio is $\text{SNR} = 4$. Evaluating Eq. (43) with $N_{\mu, \text{min}} = 8$ gives $1300 > 200$, meaning that the data should be sufficient to characterize three or fewer states. For two states ($N_{\mu, \text{min}} = 5$), the estimated time resolution [Eq. (45)] is $T_r = 20$ ms.

Our recovered distribution [Fig. 19] shows two, distinct states. The results (colored curves) are stable over a broad range of the number of moments used, consistent with the high quality of the data. The correlation-based resolution is clearly better than the histogram (short-dashed, grey curves).

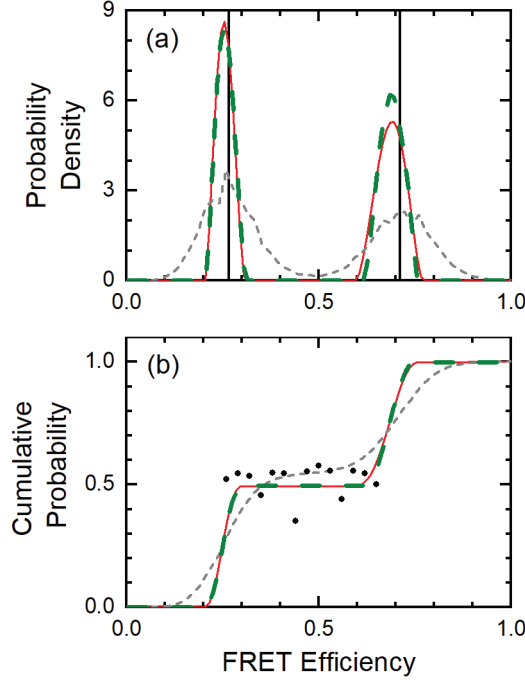


FIG. 19. Distributions recovered from the experimental data set ($\text{SNR} = 4$, $\langle N_t \rangle = 1.19 \times 10^4$, $N_{\text{en}} = 19$): histogram $P_D(D)$ (grey, short dashed curve), correlation method $P_r(S)$ with $N_\mu = 6$ (dashed, green curve) or 15 (red, thin curve). (a) Probability distributions. The mean positions from 14 parametric analyses are shown as black, vertical lines. (b) Cumulative distributions. The population of the low intensity state from 14 individual parametric analyses are shown as points. The points are in the numerical order given in Ref. 37, left to right.

The nonlinear-correlation functions from this data are simple: they all fit single exponentials with the same time constant of $T = 175$ ms (Fig. 20 and the supplementary material, Fig. S5). This result is sufficient to indicate that the system is a two-state Markovian process;⁸⁷ the complexity of generating a full Green's function is not necessary. Furthermore, all the information for such a process is contained in the 1D correlation functions; a multidimensional analysis is not needed. These results are reasonable and self-consistent. Correlation methods prove to be robust enough to be used with real, experimental data.

Although the correct solution is unknown, we can compare to the consensus of parametric analyses.³⁷ The averaged state positions from the parametric analyses are shown as vertical, black lines in Fig. 19(a). The populations of the low-intensity state as found by the individual parametric analyses (black points) are compared to the plateau in our cumulative distribution function in Fig. 19(b). Our value for the equilibration time $T = 175$ ms is consistent with the parametric results, which have a mean of $T = 171$ ms and a range of $T = 149$ –200 ms.

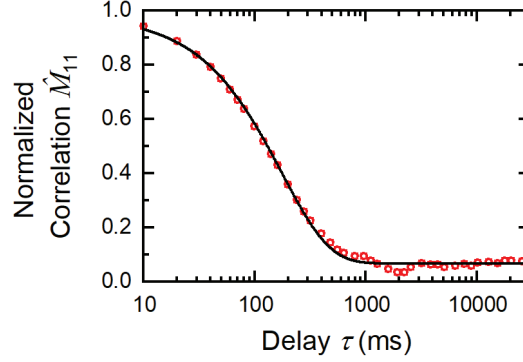


FIG. 20. The normalized moment-correlation function from the experimental data set (red circles) and a single exponential fit (black curve).

Our average FRET efficiencies and our population for the low-intensity state are slightly lower than the parametric consensus (Fig. 19), and our correlation functions do not relax to zero (Fig. 20). Both effects can be assigned to the imperfect separation of noise and signal time scales, which only differ by a factor of 10. The time averaging in the noise-corrected moments reduces all their sizes slightly. Smaller moments will both shift the distribution toward smaller values and under correct the long-time values of the correlation functions [Eq. (48)].

On the other hand, the parametric analyses were restricted *a priori* to two states and to states with zero width. The correlation analysis allows a broader range of possibilities. Based on the constancy of the distribution using up to $N_\mu = 15$ moments, a 3-state, a 4-state, or a similarly complex, continuous distribution were within the available solution space, but they are not consistent with the data. The correlation analysis also allows the states to have a finite width. Widths as broad as those in Fig. 19(a) would be consistent with the data, a result not available in the parametric analyses. Thus, parametric and nonparametric approaches are consistent with each other, but each has its expected strengths. Nonparametric methods require less prior information and allow a broader range of solutions. Parametric methods find more precise parameter values, if their assumptions are correct.

V. RECOVERING THE NOISE DISTRIBUTION

Although we have phrased the problem as removing the noise from the data, in fact, one is implicitly separating the data into two components, signal and noise. Our synthetic data has Gaussian noise, but that fact has not been used, so the noise distribution is formally unknown. In real experiments, noise comes from multiple sources, and its distribution is often truly unknown. For completeness, we will show that the noise distribution can also be recovered with correlation methods.

In addition, there is the question of how well our methods for converting noise-corrected moments to a distribution will work for a continuous distribution. The moments of a continuous distribution are not asymptotically exponential, as the moments of a discrete distribution are [Fig. 5]. Also, as mentioned in the discussion of Fig. 7, the boundary conditions we use may favor sparse solutions. This exercise will test our regularization methods on a continuous distribution.

Finally, this exercise will allow us to introduce a second basis for separating signal from noise. Up to this point, signal and noise have been distinguished by the difference in their time scales. However, signal and additive noise can also be distinguished by their lack of correlation [Eq. (10)]. To exploit this property, we look at the n th-order cumulants $\kappa_n[P]$ of a distribution P . They are algebraic combinations of moments of the same or lower order.¹⁰⁶ For example,

$$\begin{aligned}\kappa_1[P] &= \mu_1[P] \\ \kappa_2[P] &= \mu_2[P] - \mu_1[P]^2 \\ \kappa_3[P] &= \mu_3[P] - 3\mu_2[P]\mu_1[P] - 6\mu_1[P]^3,\end{aligned}\tag{49}$$

and so on. These formulas can be inverted to give the moments in terms of the cumulants of lower order. The cumulants' important property is that they are additive for uncorrelated variables.¹⁰⁶ Thus, Eq. (10) implies that

$$\kappa_n[P_D] = \kappa_n[P_S] + \kappa_n[P_\varepsilon].\tag{50}$$

This result is used as follows: The moments without noise correction, $\mu_n[D]$ [Eq. (22)], are converted to the cumulants of the data, $\{\mu_n[D]\} \rightarrow \{\kappa_n[D]\}$. The noise-corrected moments, $\mu'_n[D]$ [Eqs. (23) and (24)], are converted to the cumulants of the signal, $\{\mu'_n[D]\} \rightarrow \{\kappa_n[S]\}$. Subtraction according to Eq. (50) gives the cumulants of the noise $\kappa_n[\varepsilon]$. These are converted back to moments of the noise, $\{\kappa_n[\varepsilon]\} \rightarrow \{\mu_n[\varepsilon]\}$. Equation (32) was then used to recover the noise distribution $P_{\varepsilon,r}(\varepsilon)$ from its moments. We used a numerical method proposed by Smith to convert between moments and cumulants.¹²⁸

Figure 21 shows results corresponding to the signal distribution shown in Fig. 11. Using only boundary constraints [Eq. (31)] yields an erroneous answer for the noise distribution (thin, green curve) and an unreasonably small fitting error, $\chi = 6.4 \times 10^{-14}$. However, simply adding smoothing until χ is equal to the signal-fitting error dramatically improves the result (red curve). The tendency to produce sparse results is easily corrected. Thus, correlation methods can be used with both continuous and discrete systems.

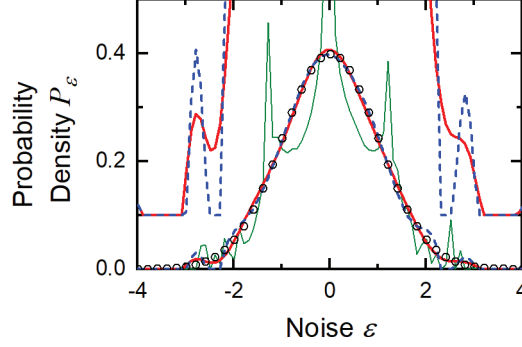


FIG. 21. The recovered distribution of the noise $P_{\epsilon,r}(\epsilon)$ with various regularization conditions ($N_\mu = 13$, $N_P = 81$, $\epsilon_{\min} = -4$, and $\epsilon_{\max} = 4$ in all cases): boundary constraints only (thin, green curve; $\beta = 0$; $\alpha = 0$; $\chi = 6.4 \times 10^{-14}$), smoothing added (solid, red curve; $\beta = 0.149$; $\alpha = 0$; $\chi = 7.7 \times 10^{-5}$), and both smoothing and sparsity added (dashed, blue line; $\beta = 0.149$; $P_{\text{sp}} = 0.025$; $R_2 = 0.016$; $\chi = 1.2 \times 10^{-4}$). The correct, Gaussian distribution $P_\epsilon(\epsilon)$ is shown as black circles. On the left and right, the tails of the curves have been vertically offset and magnified 10 times.

The tails of the recovered distribution have low-intensity satellite peaks at $\epsilon = \pm 2.8$. One can test whether these tails are real by trying to eliminate them with the sparsity regularizer. The blue curve shows an example where χ has been allowed to double in an attempt to eliminate intensity below $P_{\text{sp}} = 0.025$, and thus, the tails beyond $\epsilon = \pm 2.1$. Unlike with the signal distribution (Fig. 9), the low-intensity regions cannot be eliminated without seriously degrading the fit. The data require that the noise distribution have low-intensity tails, although it is not possible to recover the details of their shape.

VI. DISCUSSION

1. Parametric, correlation analysis

A nonparametric data analysis is normally a prelude to proposing a specific, interpretive model. In the example of Fig. 11, it would be natural to propose a three-discrete-state, Markov model. Alternatively, the structure of the system may be known from previous studies. In either case, one wants to make a parametric fit to the data. Rather than revert to a method based on analyzing the time series directly, it is possible to parametrically fit the noise-corrected moments and nonlinear-correlation functions, rather than directly fitting the time series. The implementation of a parametric, correlation analysis is specific to the model proposed, but it is relatively straightforward.

One advantage of a parametric, correlation analysis is the reduction in the size of the problem. For example, our large data set contained 10^8 data points, but only 13 moments are needed to

describe the distribution, and only about 10^3 points in the moment-correlation functions are needed to characterize the dynamics. As a result, even repeated analyses are computationally quick, in contrast to parametric fits to raw time series.³¹

2. Non-Markovian dynamics and multidimensional-correlation functions

In calculating the nonlinear-correlation functions, there is no restriction that the dynamics between observable states be either exponential or multiexponential. Various stretch, non-Markovian decays can also be found. States that are degenerate in the observable, but distinct along some other, “hidden” coordinate, are not distinguished by 1D correlation functions, resulting in non-Markovian dynamics of the observable states. Exploring dynamics along this hidden coordinate requires a second stage of analysis using multidimensional-correlation functions [Eq. (2)].⁷⁴⁻⁷⁸ In a parametric, hidden-Markov model, hidden system coordinates and the noise process are dealt with in a single stage of analysis, rather than in two.

The noise-correction methods used here should extend in a simple way to multidimensional functions. To see this point, consider the linear, multidimensional-correlation function of overall order n , $M_{1,\dots,1}[D](\tau_{n-1}, \dots, \tau_1)$ [see Eq. (2)]. Noise is confined to places where one or more of the delays τ_i is zero. Removing the noise requires shifting by one point along every time axes to eliminate these regions,

$$M'_{1,\dots,1}[D](\tau_{n-1}, \dots, \tau_1) = \left\langle D[\tau_{n-1} + \dots + \tau_1 + (n-1)T_\varepsilon] \dots \times D(\tau_1 + T_\varepsilon) D(0) \right\rangle. \quad (51)$$

The noise correction of the linear-correlation function [Eq. (15)] is just the lowest-order example of this idea. Setting the first $k-1$ and the last $l-1$ delays to zero,

$$M'_{kl}[D](\tau) = M'_{1,\dots,1}[D](0, \dots, 0, \tau, 0, \dots, 0), \quad (52)$$

recovers the noise-corrected, 1D, nonlinear-correlation functions [Eq. (18)]. The extension to nonlinear, multidimensional-correlation functions just requires setting a different set of delays to zero, for example

$$\begin{aligned} M'_{121}(\tau_2, \tau_1) &= \left\langle D(\tau_2 + \tau_1) D(\tau_1)^2 D(0) \right\rangle \\ &= M'_{1111}(\tau_2, 0, \tau_1) \end{aligned} \quad (53)$$

It is axiomatic that the full set of high-order correlation functions [Eq. (2)] contains complete information about a stationary time series. However, this set is so large that its undirected use is impractical; one needs to identify manageable subsets that contain specific pieces of information. Earlier work has shown that the distribution along hidden coordinates is contained in two-dimensional (2D) correlation functions,⁷⁴⁻⁷⁸ and that the dynamics along these coordinates are

contained in three-dimensional correlation functions.⁷⁵⁻⁷⁸ Most recently, we have shown that complete information about the dynamics along the observed coordinate is contained in the full set of 1D, nonlinear correlation functions.^{78,87,88} This paper contributes to this classification by showing that complete information about the distribution along the observed coordinate is contained in the “zero-dimensional” (0D) moments.

The discussion of $N_{\mu,\max}$, the number of measurable 0D moments, propagates to higher dimensional functions. The 0D moments are a point on the 1D functions of the same order, the 1D functions are slices of the 2D functions of the same order, and so on. If the 0D moments above $N_{\mu,\max}$ are inaccurate, all the higher dimensional functions with order above $N_{\mu,\max}$ will be inaccurate as well.

Similarly, the number of moments needed for a successful analysis $N_{\mu,\min}$ also propagates to higher dimensions. Just as the 0D moments reduce to a distribution $P_{\text{eq}}(X)$ along one X dimension, the 1D moments reduce to a Green’s function $G(X_1|X_0; \tau)$ along two X dimensions, and so on. If $N_{\mu,\min}$ moments are need to resolve the X -axis in $P_{\text{eq}}(X)$, the same order will be needed to resolve the X -axes in $G(X_1|X_0; \tau)$. Similar reasoning holds for even higher orders.

VII. SUMMARY AND CONCLUSIONS

Using a broad array of high-order correlation functions to yield comprehensive information about a time series is a new, and only partially proven, idea. This paper has taken two important steps to further this idea: it has shown that correlation-based methods can separate signal fluctuations from measurement noise, and it has shown that they can recover equilibrium-probability distributions.

The paper first showed how to remove measurement noise, only assuming that noise fluctuates more rapidly than signal does. With an example of synthetic data, we showed that accurate correlation-functions and moments can be extracted from noisy data. With these quantities, the dynamics along the observed coordinate were fully defined. Correlation results were compared to time-binning and histogramming, the other nonparametric methods in common use. The correlation results have substantially better resolution, both in time and in state space. The correlation methods were also tested on published, experimental data. Good results for both the distribution and the kinetics were found.

The paper then established the amount and quality of data needed for a correlation analysis. Although measurement noise can be removed, sampling error remains whenever a time series has a finite length. We showed that there are two components to this error with differing behavior: noise-sampling error and signal-sampling error. Approximate expressions for these errors were derived and used to predict the minimum data-set size and signal-to-noise ratio that are needed to recover a distribution of a given complexity. There is a floor to the usable signal-to-noise ratio of

approximately one-half, below which it is impractical to compensate for high noise with increased data collection. Above this floor, the time resolution of the experiment can be close to the time resolution of the detector. Below this floor, initial time binning is necessary, which limits the time resolution of the results.

Most of the work in the paper used the short correlation time of the noise to distinguish it from signal. In a final demonstration, we showed that the lack of correlation between signal and noise can also be used to discriminate between them. Cumulants were used to extract the noise distribution from our data. This work points toward methods for removing baseline drift and other types of noise that do not have a short correlation time.

Correlation methods are nonparametric: no model of the system or the noise is needed. The model-selection problem of parametric analysis is avoided. A parametric analysis will always give an apparently narrower error range because the range of solutions is limited *a priori*. However, a nonparametric approach has an important role. If a feature is identified in a parametric analysis, but it is not required in a nonparametric analysis, one can infer that the feature results from the restrictions applied, as much as from the data. Similarly, the requirements for data quality in a nonparametric analysis, as found here, inform a parametric analysis. If a parametric approach finds a solution from data that violates these requirements, one knows that the prior assumptions play a critical role. In either case, one can then assess one's confidence in the assumed prior knowledge.

Many types of experiment will require even further extensions of the ideas presented here. Some experiments involve multiple observables, for example, donor and acceptor intensities in FRET experiments or parallel and perpendicular polarizations in anisotropy measurements. In this case, both high-order auto- and cross-correlation functions need to be considered. Although more complex, there does not appear to be a fundamental barrier to this generalization.

Many experiments are dominated by photon noise, which is not additive. In addition, photon counting can introduce bias into the data. Many parametric methods have been extended to "photon-by-photon" analysis for this situation.^{24,25,27,28,33,41,66,72} A similar extension of nonparametric, correlation methods is possible and will be discussed in a forthcoming paper.¹²

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation – Chemical Measurement and Imaging program under Grant No. CHE-2003619 (with partial co-funding from the Chemical Structures, Dynamics, and Mechanisms–A program).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Mainak Dhar: Formal analysis (lead); Investigation (lead); Methodology (equal); Software (lead); Visualization (lead); Writing – review & editing (equal). **Joseph Dickinson:** Formal analysis (supporting); Investigation (supporting); Methodology (equal); Software (supporting); Writing – review & editing (supporting). **Mark Berg:** Conceptualization (lead); Formal analysis (supporting); Methodology (equal); Supervision (lead); Writing – original draft (lead); Writing – review & editing (equal).

SUPPLEMENTARY MATERIAL

See the supplementary material for more detail on generating synthetic data, regularization without boundary constraints, the minimization algorithm, optimizing the regularization parameters, deriving the sampling-error equations, and the experimental correlation functions.

DATA AVAILABILITY

The data that supports the findings of this study are available within the article and its supplementary material.

REFERENCES

- ¹ W. J. Bruno, J. Yang, and J. E. Pearson, "Using Independent Open-to-Closed Transitions to Simplify Aggregated Markov Models of Ion Channel Gating Kinetics," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6326-6331 (2005). <http://doi.org/10.1073/pnas.0409110102>.
- ² W. E. Moerner and D. P. Fromm, "Methods of Single-Molecule Fluorescence Spectroscopy and Microscopy," *Rev. Sci. Instrum.* **74**, 3597-3619 (2003). <http://doi.org/10.1063/1.1589587>.
- ³ K. Paeng and L. J. Kaufman, "Single Molecule Rotational Probing of Supercooled Liquids," *Chem. Soc. Rev.* **43**, 977-989 (2014). <http://doi.org/10.1039/c3cs60186b>.
- ⁴ T. Kondo, W. J. Chen, and G. S. Schlau-Cohen, "Single-Molecule Fluorescence Spectroscopy of Photosynthetic Systems," *Chem. Rev.* **117**, 860-898 (2017). <http://doi.org/10.1021/acs.chemrev.6b00195>.
- ⁵ B. Schuler, "Perspective: Chain Dynamics of Unfolded and Intrinsically Disordered Proteins from Nanosecond Fluorescence Correlation Spectroscopy Combined with Single-Molecule FRET," *J. Chem. Phys.* **149**, 010901 (2018). <http://doi.org/10.1063/1.5037683>.
- ⁶ S. Adhikari and M. Orrit, "Progress and Perspectives in Single-Molecule Optical Spectroscopy," *J. Chem. Phys.* **156**, 160903 (2022). <http://doi.org/10.1063/5.0087003>.
- ⁷ E. Lerner, *et al.*, "FRET-Based Dynamic Structural Biology: Challenges, Perspectives and an Appeal for Open-Science Practices," *eLife* **10**, e60416 (2021). <http://doi.org/10.7554/eLife.60416>.
- ⁸ E. L. Elson, "Brief Introduction to Fluorescence Correlation Spectroscopy," in *Methods in Enzymology*, edited by S. Y. Tetin (Academic Press, 2013) Vol. 518, pp. 11-41. <http://doi.org/https://doi.org/10.1016/B978-0-12-388422-0.00002-9>.
- ⁹ C. Eggeling and C. Hellriegel, "Special Issue on Developments in Fluorescence Correlation Spectroscopy and Related Techniques," *Methods* **140-141**, 1-222 (2018). <http://doi.org/10.1016/j.ymeth.2018.04.034>.
- ¹⁰ B. J. Berne and R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics* (John Wiley & Sons, New York, 1976).
- ¹¹ M. Sutton, "A Review of X-Ray Intensity Fluctuation Spectroscopy," *C. R. Physique* **9**, 657-667 (2008). <http://doi.org/10.1016/j.crhy.2007.04.008>.
- ¹² M. Dhar and M. A. Berg, "Efficient, Nonparametric Removal of Noise and Recovery of Probability Distributions from Time Series Using Nonlinear-Correlation Functions: Photon Noise" (in preparation).

- ¹³ *Single-Molecule Biophysics: Experiment and Theory*, edited by T. Komatsuzaki, M. Kawakami, S. Takahashi, H. Yang, and R. J. Silbey (Wiley, Hoboken, New Jersey, 2012) *Advances in Chemical Physics*, Vol. 146. <http://doi.org/10.1002/9781118131374>.
- ¹⁴ M. Tavakoli, J. N. Taylor, C.-B. Li, T. Komatsuzaki, and S. Pressé, “Single Molecule Data Analysis: An Introduction,” *Adv. Chem. Phys.* **162**, 205-305 (2017).
<http://doi.org/10.1002/9781119324560.ch4>.
- ¹⁵ C. Du and S. C. Kou, “Statistical Methodology in Single-Molecule Experiments,” *Statist. Sci.* **35**, 75-91 (2020). <http://doi.org/10.1214/19-STS752>.
- ¹⁶ G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (Addison-Wesley Pub. Co., Reading, Mass, 1973).
- ¹⁷ J. B. Witkoskie and J. Cao, “Single Molecule Kinetics. II. Numerical Bayesian Approach,” *J. Chem. Phys.* **121**, 6373-6379 (2004). <http://doi.org/10.1063/1.1785784>.
- ¹⁸ D. L. Ensign and V. S. Pande, “Bayesian Single-Exponential Kinetics in Single-Molecule Experiments and Simulations,” *J. Phys. Chem. B* **113**, 12410-12423 (2009).
<http://doi.org/10.1021/jp903107c>.
- ¹⁹ M. El Beheiry, S. Türkcan, M. U. Richly, A. Triller, A. Alexandrou, M. Dahan, and J.-B. Masson, “A Primer on the Bayesian Approach to High-Density Single-Molecule Trajectories Analysis,” *Biophys. J.* **110**, 1209-1215 (2016). <http://doi.org/10.1016/j.bpj.2016.01.018>.
- ²⁰ C. D. Kinz-Thompson, K. K. Ray, and R. L. Gonzalez Jr., “Bayesian Inference: The Comprehensive Approach to Analyzing Single-Molecule Experiments,” *Annu. Rev. Biophys.* **50**, 191-208 (2021). <http://doi.org/10.1146/annurev-biophys-082120-103921>.
- ²¹ N. James and M. Menzies, “Optimally Adaptive Bayesian Spectral Density Estimation for Stationary and Nonstationary Processes,” *Stat. Comput.* **32**, 45 (2022).
<http://doi.org/10.1007/s11222-022-10103-4>.
- ²² M. Bertolacci, O. Rosen, E. Cripps, and S. Cripps, “Adaptspec-X: Covariate-Dependent Spectral Modeling of Multiple Nonstationary Time Series,” *J. Comput. Graph. Stat.* **31**, 436-454 (2022). <http://doi.org/10.1080/10618600.2021.2000870>.
- ²³ J. B. Witkoskie and J. Cao, “Analysis of the Entire Sequence of a Single Photon Experiment on a Flavin Protein,” *J. Phys. Chem. B* **112**, 5988-5996 (2008).
<http://doi.org/10.1021/jp075980p>.
- ²⁴ K. R. Haas, H. Yang, and J.-W. Chu, “Expectation-Maximization of the Potential of Mean Force and Diffusion Coefficient in Langevin Dynamics from Single Molecule FRET Data Photon by Photon,” *J. Phys. Chem. B* **117**, 15591-15605 (2013).
<http://doi.org/10.1021/jp405983d>.

- ²⁵ M. Andrec, R. M. Levy, and D. S. Talaga, "Direct Determination of Kinetic Rates from Single-Molecule Photon Arrival Trajectories Using Hidden Markov Models," *J. Phys. Chem. A* **107**, 7454-7464 (2003). <http://doi.org/10.1021/jp035514+>.
- ²⁶ S. A. McKinney, C. Joo, and T. Ha, "Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling," *Biophys. J.* **91**, 1941-1951 (2006). <http://doi.org/10.1529/biophysj.106.082487>.
- ²⁷ M. Hajdziona and A. Molski, "Maximum Likelihood-Based Analysis of Single-Molecule Photon Arrival Trajectories," *J. Chem. Phys.* **134**, 054112 (2011). <http://doi.org/10.1063/1.3544494>.
- ²⁸ M. Pirchi, R. Tsukanov, R. Khamis, T. E. Tomov, Y. Berger, D. C. Khara, H. Volkov, G. Haran, and E. Nir, "Photon-by-Photon Hidden Markov Model Analysis for Microsecond Single-Molecule FRET Kinetics," *J. Phys. Chem. B* **120**, 13065-13075 (2016). <http://doi.org/10.1021/acs.jpcb.6b10726>.
- ²⁹ B. E. Husic and V. S. Pande, "Markov State Models: From an Art to a Science," *J. Am. Chem. Soc.* **140**, 2386-2396 (2018). <http://doi.org/10.1021/jacs.7b12191>.
- ³⁰ H. Li and H. Yang, "Statistical Learning of Discrete States in Time Series," *J. Phys. Chem. B* **123**, 689-701 (2019). <http://doi.org/10.1021/acs.jpcb.8b10561>.
- ³¹ J. Yoo, J.-Y. Kim, J. M. Louis, I. V. Gopich, and H. S. Chung, "Fast Three-Color Single-Molecule FRET Using Statistical Inference," *Nature Commun.* **11**, 3336 (2020). <http://doi.org/10.1038/s41467-020-17149-w>.
- ³² Z. Kilic, I. Sgouralis, W. Heo, K. Ishii, T. Tahara, and S. Pressé, "Extraction of Rapid Kinetics from smFRET Measurements Using Integrative Detectors," *Cell Rep. Phys. Sci.* **2**, 100409 (2021). <http://doi.org/10.1016/j.xcrp.2021.100409>.
- ³³ P. D. Harris, A. Narducci, C. Gebhardt, T. Cordes, S. Weiss, and E. Lerner, "Multi-Parameter Photon-by-Photon Hidden Markov Modeling," *Nature Commun.* **13**, 1000 (2022). <http://doi.org/10.1038/s41467-022-28632-x>.
- ³⁴ O. Opanasyuk, A. Barth, T.-O. Peulen, S. Felekyan, S. Kalinin, H. Sanabria, and C. A. M. Seidel, "Unraveling Multi-State Molecular Dynamics in Single-Molecule FRET Experiments. II. Quantitative Analysis of Multi-State Kinetic Networks," *J. Chem. Phys.* **157**, 031501 (2022). <http://doi.org/10.1063/5.0095754>.
- ³⁵ I. V. Gopich and H. S. Chung, "Theory and Analysis of Single-Molecule FRET Experiments," in *Protein Folding: Methods and Protocols*, edited by V. Muñoz (Springer US, New York, NY, 2022), pp. 247-282. http://doi.org/10.1007/978-1-0716-1716-8_14.

- ³⁶ A. Barth, O. Opanasyuk, T.-O. Peulen, S. Felekyan, S. Kalinin, H. Sanabria, and C. A. M. Seidel, “Unraveling Multi-State Molecular Dynamics in Single-Molecule FRET Experiments. I. Theory of FRET-Lines,” *J. Chem. Phys.* **156**, 141501 (2022).
<http://doi.org/10.1063/5.0089134>.
- ³⁷ M. Götz, *et al.*, “A Blind Benchmark of Analysis Tools to Infer Kinetic Rate Constants from Single-Molecule FRET Trajectories,” *Nature Commun.* **13**, 5402 (2022).
<http://doi.org/10.1038/s41467-022-33023-3>.
- ³⁸ Keegan E. Hines, John R. Bankston, and Richard W. Aldrich, “Analyzing Single-Molecule Time Series via Nonparametric Bayesian Inference,” *Biophys. J.* **108**, 540-556 (2015).
<http://doi.org/https://doi.org/10.1016/j.bpj.2014.12.016>.
- ³⁹ I. Sgouralis and S. Pressé, “An Introduction to Infinite HMMs for Single-Molecule Data Analysis,” *Biophys. J.* **112**, 2021-2029 (2017).
<http://doi.org/https://doi.org/10.1016/j.bpj.2017.04.027>.
- ⁴⁰ I. Sgouralis, S. Madaan, F. Djutanta, R. Kha, R. F. Hariadi, and S. Pressé, “A Bayesian Nonparametric Approach to Single Molecule Förster Resonance Energy Transfer,” *J. Phys. Chem. B* **123**, 675-688 (2019). <http://doi.org/10.1021/acs.jpcc.8b09752>.
- ⁴¹ M. Tavakoli, S. Jazani, I. Sgouralis, O. M. Shafraz, S. Sivasankar, B. Donaphon, M. Levitus, and S. Pressé, “Pitching Single-Focus Confocal Data Analysis One Photon at a Time with Bayesian Nonparametrics,” *Phys. Rev. X* **10**, 011021 (2020).
<http://doi.org/10.1103/PhysRevX.10.011021>.
- ⁴² L. P. Watkins and H. Yang, “Information Bounds and Optimal Analysis of Dynamic Single Molecule Measurements,” *Biophys. J.* **86**, 4015-4029 (2004).
<http://doi.org/10.1529/biophysj.103.037739>.
- ⁴³ D.-H. Lee, C.-T. Yuan, M. Tachiya, and J. Tang, “Influence of Bin Time and Excitation Intensity on Fluorescence Lifetime Distribution and Blinking Statistics of Single Quantum Dots,” *Appl. Phys. Lett.* **95**, 163101-163103 (2009). <http://doi.org/10.1063/1.3236772>.
- ⁴⁴ C. H. Crouch, O. Sauter, X. Wu, R. Purcell, C. Querner, M. Drndic, and M. Pelton, “Facts and Artifacts in the Blinking Statistics of Semiconductor Nanocrystals,” *Nano Lett.* **10**, 1692-1698 (2010). <http://doi.org/10.1021/nl100030e>.
- ⁴⁵ S. Kalinin, E. Sisamakias, S. W. Magennis, S. Felekyan, and C. A. M. Seidel, “On the Origin of Broadening of Single-Molecule FRET Efficiency Distributions Beyond Shot Noise Limits,” *J. Phys. Chem. B* **114**, 6197-6206 (2010). <http://doi.org/10.1021/jp100025v>.
- ⁴⁶ I. V. Gopich and A. Szabo, “Theory of Single-Molecule FRET Efficiency Histograms,” *Adv. Chem. Phys.* **146**, 245-297 (2011). <http://doi.org/10.1002/9781118131374.ch10>.

- ⁴⁷ Toma E. Tomov, R. Tsukanov, R. Masoud, M. Liber, N. Plavner, and E. Nir, "Disentangling Subpopulations in Single-Molecule FRET and ALEX Experiments with Photon Distribution Analysis," *Biophys. J.* **102**, 1163-1173 (2012). <http://doi.org/10.1016/j.bpj.2011.11.4025>.
- ⁴⁸ L. P. Watkins, H. Chang, and H. Yang, "Quantitative Single-Molecule Conformational Distributions: A Case Study with Poly-(L-Proline)," *J. Phys. Chem. A* **110**, 5191-5203 (2006). <http://doi.org/10.1021/jp055886d>.
- ⁴⁹ E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov, and S. Weiss, "Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments," *J. Phys. Chem. B* **110**, 22103-22124 (2006). <http://doi.org/10.1021/jp063483n>.
- ⁵⁰ D. A. Koster, C. H. Wiggins, and N. H. Dekker, "Multiple Events on Single Molecules: Unbiased Estimation in Single-Molecule Biophysics," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1750-1755 (2006). <http://doi.org/10.1073/pnas.0510509103>.
- ⁵¹ S. Kalinin, A. Valeri, M. Antonik, S. Felekyan, and C. A. M. Seidel, "Detection of Structural Dynamics by FRET: A Photon Distribution and Fluorescence Lifetime Analysis of Systems with Multiple States," *J. Phys. Chem. B* **114**, 7983-7995 (2010). <http://doi.org/10.1021/jp102156t>.
- ⁵² M. Backovic, E. S. Price, C. K. Johnson, and J. P. Ralston, "A Distribution-Based Method to Resolve Single-Molecule Forster Resonance Energy Transfer Observations," *J. Chem. Phys.* **134**, 145101 (2011). <http://doi.org/10.1063/1.3568946>.
- ⁵³ P. A. Frantsuzov, S. Volkán-Kacsó, and B. Jankó, "Model of Fluorescence Intermittency of Single Colloidal Semiconductor Quantum Dots Using Multiple Recombination Centers," *Phys. Rev. Lett.* **103**, 207402 (2009). <http://doi.org/10.1103/PhysRevLett.103.207402>.
- ⁵⁴ T. G. Terentyeva, H. Engelkamp, A. E. Rowan, T. Komatsuzaki, J. Hofkens, C.-B. Li, and K. Blank, "Dynamic Disorder in Single-Enzyme Experiments: Facts and Artifacts," *ACS Nano* **6**, 346-354 (2011). <http://doi.org/10.1021/nn203669r>.
- ⁵⁵ L. P. Watkins and H. Yang, "Detection of Intensity Change Points in Time-Resolved Single-Molecule Measurements," *J. Phys. Chem. B* **109**, 617-628 (2005). <http://doi.org/10.1021/jp0467548>.
- ⁵⁶ H. Yang, "Change-Point Localization and Wavelet Spectral Analysis of Single-Molecule Time Series," *Adv. Chem. Phys.* **146**, 217-243 (2012). <http://doi.org/10.1002/9781118131374.ch9>.
- ⁵⁷ D. L. Ensign and V. S. Pande, "Bayesian Detection of Intensity Changes in Single Molecule and Molecular Dynamics Trajectories," *J. Phys. Chem. B* **114**, 280-292 (2009). <http://doi.org/10.1021/jp906786b>.

- ⁵⁸ J. N. Taylor, D. E. Makarov, and C. F. Landes, "Denoising Single-Molecule FRET Trajectories with Wavelets and Bayesian Inference," *Biophys. J.* **98**, 164-173 (2010). <http://doi.org/10.1016/j.bpj.2009.09.047>.
- ⁵⁹ B. Shuang, D. Cooper, J. N. Taylor, L. Kisley, J. Chen, W. Wang, C. B. Li, T. Komatsuzaki, and C. F. Landes, "Fast Step Transition and State Identification (STaSI) for Discrete Single-Molecule Data Analysis," *J. Phys. Chem. Lett.* **5**, 3157-3161 (2014). <http://doi.org/10.1021/jz501435p>.
- ⁶⁰ J. N. Taylor, C.-B. Li, D. R. Cooper, C. F. Landes, and T. Komatsuzaki, "Error-Based Extraction of States and Energy Landscapes from Experimental Single-Molecule Time-Series," *Sci. Rep.* **5**, 9174 (2015). <http://doi.org/10.1038/srep09174>.
- ⁶¹ Y. J. Bae, N. A. Gibson, T. X. Ding, A. P. Alivisatos, and S. R. Leone, "Understanding the Bias Introduced in Quantum Dot Blinking Using Change Point Analysis," *J. Phys. Chem. C* **120**, 29484-29490 (2016). <http://doi.org/10.1021/acs.jpcc.6b09780>.
- ⁶² M. A. Ochoa, P. Chen, and R. F. Loring, "Single Turnover Measurements of Nanoparticle Catalysis Analyzed with Dwell Time Correlation Functions and Constrained Mean Dwell Times," *J. Phys. Chem. C* **117**, 19074-19081 (2013). <http://doi.org/10.1021/jp4065246>.
- ⁶³ W. Xu, J. S. Kong, and P. Chen, "Single-Molecule Kinetic Theory of Heterogeneous and Enzyme Catalysis," *J. Phys. Chem. C* **113**, 2393-2404 (2009). <http://doi.org/10.1021/jp808240c>.
- ⁶⁴ J. Cao and R. J. Silbey, "Generic Schemes for Single-Molecule Kinetics. 1: Self-Consistent Pathway Solutions for Renewal Processes," *J. Phys. Chem. B* **112**, 12867-12880 (2008). <http://doi.org/10.1021/jp803347m>.
- ⁶⁵ I. V. Gopich and A. Szabo, "Theory of the Statistics of Kinetic Transitions with Application to Single-Molecule Enzyme Catalysis," *J. Chem. Phys.* **124**, 154712 (2006). <http://doi.org/10.1063/1.2180770>.
- ⁶⁶ H. Yang and X. S. Xie, "Probing Single-Molecule Dynamics Photon by Photon," *J. Chem. Phys.* **117**, 10965-10979 (2002). <http://doi.org/10.1063/1.1521154>.
- ⁶⁷ R. Verberk and M. Orrit, "Photon Statistics in the Fluorescence of Single Molecules and Nanocrystals: Correlation Functions Versus Distributions of On- and Off-Times," *J. Chem. Phys.* **119**, 2214-2222 (2003). <http://doi.org/10.1063/1.1582848>.
- ⁶⁸ M. Pelton, G. Smith, N. F. Scherer, and R. A. Marcus, "Evidence for a Diffusion-Controlled Mechanism for Fluorescence Blinking of Colloidal Quantum Dots," *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14249-14254 (2007). <http://doi.org/10.1073/pnas.0706164104>.
- ⁶⁹ G. Hinze and T. Basché, "Statistical Analysis of Time Resolved Single Molecule Fluorescence Data without Time Binning," *J. Chem. Phys.* **132**, 044509 (2010). <http://doi.org/10.1063/1.3303634>.

- ⁷⁰ J. Cui, A. P. Beyler, T. S. Bischof, M. W. B. Wilson, and M. G. Bawendi, “Deconstructing the Photon Stream from Single Nanocrystals: From Binning to Correlation,” *Chem. Soc. Rev.* **43**, 1287-1310 (2014). <http://doi.org/10.1039/C3CS60330J>.
- ⁷¹ J. Houel, *et al.*, “Autocorrelation Analysis for the Unbiased Determination of Power-Law Exponents in Single-Quantum-Dot Blinking,” *ACS Nano* **9**, 886-893 (2015). <http://doi.org/10.1021/nn506598t>.
- ⁷² H. Yang and X. S. Xie, “Statistical Approaches for Probing Single-Molecule Dynamics Photon-by-Photon,” *Chem. Phys.* **284**, 423-437 (2002). [http://doi.org/10.1016/S0301-0104\(02\)00672-9](http://doi.org/10.1016/S0301-0104(02)00672-9).
- ⁷³ J. B. Witkoskie and J. Cao, “Single Molecule Kinetics. I. Theoretical Analysis of Indicators,” *J. Chem. Phys.* **121**, 6361-6372 (2004). <http://doi.org/10.1063/1.1785783>.
- ⁷⁴ S. D. Verma, D. A. Vanden Bout, and M. A. Berg, “When is a Single Molecule Homogeneous? A Multidimensional Answer and Its Application to Molecular Rotation near the Glass Transition,” *J. Chem. Phys.* **143**, 024110 (2015). <http://doi.org/10.1063/1.4926463>.
- ⁷⁵ S. D. Verma, S. A. Corcelli, and M. A. Berg, “Rate and Amplitude Heterogeneity in the Solvation Response of an Ionic Liquid,” *J. Phys. Chem. Lett.* **7**, 504–508 (2016). <http://doi.org/10.1021/acs.jpcllett.5b02835>.
- ⁷⁶ M. A. Berg and J. R. Darvin, “Measuring the Dynamics of a Hidden Coordinate: Rate Exchange from 3D Correlation Functions,” *J. Chem. Phys.* **145**, 054119 (2016). <http://doi.org/10.1063/1.4960186>.
- ⁷⁷ M. A. Berg and H. Kaur, “Non-Parametric Analysis of Nonexponential and Multidimensional Kinetics: I. Quantifying Rate Dispersion, Heterogeneity and Exchange,” *J. Chem. Phys.* **146**, 054104 (2017). <http://doi.org/10.1063/1.4974508>.
- ⁷⁸ H. Kaur, S. D. Verma, K. Paeng, L. J. Kaufman, and M. A. Berg, “Biphasic Rate Exchange in Supercooled *o*-Terphenyl from an Ensemble Analysis of Single-Molecule Data,” *Phys. Rev. E* **98**, 040603(R) (2018). <http://doi.org/10.1103/PhysRevE.98.040603>.
- ⁷⁹ H. Kaur and M. A. Berg, “Jump-Precursor State Emerges Below the Crossover Temperature in Supercooled *o*-Terphenyl,” *Phys. Rev. E* **103**, L050601 (2021). <http://doi.org/10.1103/PhysRevE.103.L050601>.
- ⁸⁰ C. Phelps, B. Israels, M. C. Marsh, P. H. von Hippel, and A. H. Marcus, “Using Multiorder Time-Correlation Functions (TCFs) to Elucidate Biomolecular Reaction Pathways from Microsecond Single-Molecule Fluorescence Experiments,” *J. Phys. Chem. B* **120**, 13003–13016 (2016). <http://doi.org/10.1021/acs.jpcb.6b08449>.

- ⁸¹ B. Israels, C. S. Albrecht, A. Dang, M. Barney, P. H. von Hippel, and A. H. Marcus, "Submillisecond Conformational Transitions of Short Single-Stranded DNA Lattices by Photon Correlation Single-Molecule Förster Resonance Energy Transfer," *J. Phys. Chem. B* **125**, 9426-9440 (2021). <http://doi.org/10.1021/acs.jpcb.1c04119>.
- ⁸² H. Yang, "Detection and Characterization of Dynamical Heterogeneity in an Event Series Using Wavelet Correlation," *J. Chem. Phys.* **129**, 074701 (2008). <http://doi.org/10.1063/1.2969074>.
- ⁸³ K. Chen, B. Wang, J. Guan, and S. Granick, "Diagnosing Heterogeneous Dynamics in Single-Molecule/Particle Trajectories with Multiscale Wavelets," *ACS Nano* **7**, 8634-8644 (2013). <http://doi.org/10.1021/nn402787a>.
- ⁸⁴ J. M. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications," *Proc. IEEE* **79**, 278-305 (1991). <http://doi.org/10.1109/5.75086>.
- ⁸⁵ C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis* (PTR Prentice Hall, Upper Saddle River, New Jersey, 1993).
- ⁸⁶ J. J. G. de la Rosa, A. Agüera-Pérez, J. C. Palomares-Salas, and A. Moreno-Muñoz, "Higher-Order Statistics: Discussion and Interpretation," *Measurement* **46**, 2816-2827 (2013). <http://doi.org/10.1016/j.measurement.2013.04.055>.
- ⁸⁷ S. R. Hodge and M. A. Berg, "Nonlinear Measurements of Kinetics and Generalized Dynamical Modes: I. Extracting the One-Dimensional Green's Function from a Time Series," *J. Chem. Phys.* **155**, 024122 (2021). <http://doi.org/10.1063/5.0053422>.
- ⁸⁸ S. R. Hodge, S. A. Corcelli, and M. A. Berg, "Nonlinear Measurements of Kinetics and Generalized Dynamical Modes: II. Application to a Simulation of Solvation Dynamics in an Ionic Liquid," *J. Chem. Phys.* **155**, 024123 (2021). <http://doi.org/10.1063/5.0053424>.
- ⁸⁹ J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (American Mathematical Society, Providence, Rhode Island, 1943).
- ⁹⁰ N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis* (Society for Industrial and Applied Mathematics, 2020). <http://doi.org/10.1137/1.9781611976397>.
- ⁹¹ L. R. Mead and N. Papanicolaou, "Maximum Entropy in the Problem of Moments," *J. Math. Phys.* **25**, 2404-2417 (1984). <http://doi.org/10.1063/1.526446>.
- ⁹² G. Talenti, "Recovering a Function from a Finite Number of Moments," *Inverse Probl.* **3**, 501-517 (1987). <http://doi.org/10.1088/0266-5611/3/3/016>.

- ⁹³ J. N. Kapur and H. K. Kesavan, "Entropy Optimization Principles and Their Applications," in *Entropy and Energy Dissipation in Water Resources*, edited by V. P. Singh and M. Fiorentino (Springer Netherlands, Dordrecht, 1992), pp. 3-20. http://doi.org/10.1007/978-94-011-2430-0_1.
- ⁹⁴ A. Tagliani, "On the Application of Maximum Entropy to the Moments Problem," *J. Math. Phys.* **34**, 326-337 (1993). <http://doi.org/10.1063/1.530385>.
- ⁹⁵ G. A. Athanassoulis and P. N. Gavriliadis, "The Truncated Hausdorff Moment Problem Solved by Using Kernel Density Functions," *Probabilist. Eng. Mech.* **17**, 273-291 (2002). [http://doi.org/10.1016/S0266-8920\(02\)00012-7](http://doi.org/10.1016/S0266-8920(02)00012-7).
- ⁹⁶ S. B. Provost, "Moment-Based Density Approximants," *Mathematica Journal* **9**, 727-756 (2005).
- ⁹⁷ R. M. Mnatsakanov, "Hausdorff Moment Problem: Reconstruction of Probability Density Functions," *Statistics & Probability Letters* **78**, 1869-1877 (2008). <http://doi.org/10.1016/j.spl.2008.01.054>.
- ⁹⁸ T. M. P. Ngoc, "A Statistical Minimax Approach to the Hausdorff Moment Problem," *Inverse Probl.* **24**, 045018 (2008). <http://doi.org/10.1088/0266-5611/24/4/045018>.
- ⁹⁹ R. M. Mnatsakanov and A. S. Hakobyan, "Recovery of Distributions via Moments," in *Optimality: The Third Erich L. Lehmann Symposium*, edited by J. Rojo (Institute of Mathematical Statistics, 2009), Institute of Mathematical Statistics Lecture Notes - Monograph Series Vol. 57, pp. 252-265. <http://doi.org/10.1214/09-LNMS5715>.
- ¹⁰⁰ G. L. Bretthorst, "The Maximum Entropy Method of Moments and Bayesian Probability Theory," *AIP Conf. Proc.* **1553**, 3-15 (2013). <http://doi.org/10.1063/1.4819977>.
- ¹⁰¹ D. T. Hoang, K. Paeng, H. Park, L. M. Leone, and L. J. Kaufman, "Extraction of Rotational Correlation Times from Noisy Single Molecule Fluorescence Trajectories," *Anal. Chem.* **86**, 9322-9329 (2014). <http://doi.org/10.1021/ac502575k>.
- ¹⁰² C. Y. Lu and D. A. Vanden Bout, "Effect of Finite Trajectory Length on the Correlation Function Analysis of Single Molecule Data," *J. Chem. Phys.* **125**, 124701 (2006). <http://doi.org/10.1063/1.2352748>.
- ¹⁰³ C.-Y. J. Wei, C.-Y. Lu, Y. H. Kim, and D. A. Vanden Bout, "Determining If a System is Heterogeneous: The Analysis of Single Molecule Rotational Correlation Functions and Their Limitations," *J. Fluoresc.* **17**, 797-804 (2007). <http://doi.org/10.1007/s10895-007-0234-9>.
- ¹⁰⁴ S. A. Mackowiak and L. J. Kaufman, "When the Heterogeneous Appears Homogeneous: Discrepant Measures of Heterogeneity in Single-Molecule Observables," *J. Phys. Chem. Lett.* **2**, 438-442 (2011). <http://doi.org/10.1021/jz1016872>.
- ¹⁰⁵ J. A. Hanson and H. Yang, "A General Statistical Test for Correlations in a Finite-Length Time Series," *J. Chem. Phys.* **128**, 214101 (2008). <http://doi.org/10.1063/1.2931943>.

- ¹⁰⁶ A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics* (Oxford University Press, New York, 1987).
- ¹⁰⁷ A. R. Hall, *Generalized Method of Moments* (Oxford University Press, New York, 2005).
- ¹⁰⁸ W. H. Press, S. A. Teukolsky, W. T. Vetterling, and P. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1992).
- ¹⁰⁹ A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems* (Kluwer Academic, Dordrecht, 1995).
<http://doi.org/10.1007/978-94-015-8480-7>.
- ¹¹⁰ Å. Björck, *Numerical Methods for Least Squares Problems* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996).
- ¹¹¹ R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Statist. Soc. B* **58**, 267-288 (1996). <http://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- ¹¹² P. C. Hansen, "Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems," *Numer. Algorithms* **6**, 1-35 (1994). <http://doi.org/10.1007/BF02149761>.
- ¹¹³ K. Sjöstrand, L. H. Clemmensen, R. Larsen, G. Einarsson, and B. Ersbøll, "SpaSM: A Matlab Toolbox for Sparse Statistical Modeling," *J. Stat. Softw.* **84**, 1-37 (2018).
<http://doi.org/10.18637/jss.v084.i10>.
- ¹¹⁴ J. A. Hanson and H. Yang, "Quantitative Evaluation of Cross Correlation between Two Finite-Length Time Series with Applications to Single-Molecule FRET," *J. Phys. Chem. B* **112**, 13962-13970 (2008). <http://doi.org/10.1021/jp804440y>.
- ¹¹⁵ X. S. Xie, "Single-Molecule Spectroscopy and Dynamics at Room Temperature," *Acc. Chem. Res.* **29**, 598-606 (1996). <http://doi.org/10.1021/ar950246m>.
- ¹¹⁶ W. E. Moerner and M. Orrit, "Illuminating Single Molecules in Condensed Matter," *Science* **283**, 1670-1676 (1999). <http://doi.org/10.1126/science.283.5408.1670>.
- ¹¹⁷ W. P. Ambrose, P. M. Goodwin, J. H. Jett, A. Van Orden, J. H. Werner, and R. A. Keller, "Single Molecule Fluorescence Spectroscopy at Ambient Temperature," *Chem. Rev.* **99**, 2929-2956 (1999). <http://doi.org/10.1021/cr980132z>.
- ¹¹⁸ E. Barkai, Y. Jung, and R. Silbey, "Theory of Single-Molecule Spectroscopy: Beyond the Ensemble Average," *Annu. Rev. Phys. Chem.* **55**, 457-507 (2004).
<http://doi.org/10.1146/annurev.physchem.55.111803.143246>.
- ¹¹⁹ F. Kulzer, T. Xia, and M. Orrit, "Single Molecules as Optical Nanoprobes for Soft and Complex Matter," *Angew. Chem. Int. Ed.* **49**, 854-866 (2010).
<http://doi.org/10.1002/anie.200904858>.
- ¹²⁰ A. Schob, F. Cichos, J. Schuster, and C. von Borczyskowski, "Reorientation and Translation of Individual Dye Molecules in a Polymer Matrix," *Eur. Poly. J.* **40**, 1019-1026 (2004).
<http://doi.org/10.1016/j.eurpolymj.2004.01.016>.

- ¹²¹ R. Zondervan, F. Kulzer, G. C. G. Berkhout, and M. Orrit, “Local Viscosity of Supercooled Glycerol near T_g Probed by Rotational Diffusion of Ensembles and Single Dye Molecules,” *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12628-12633 (2007).
<http://doi.org/10.1073/pnas.0610521104>.
- ¹²² E. Braeken, G. De Cremer, P. Marsal, G. Pèpe, K. Müllen, and R. A. L. Vallée, “Single Molecule Probing of the Local Segmental Relaxation Dynamics in Polymer above the Glass Transition Temperature,” *J. Am. Chem. Soc.* **131**, 12201-12210 (2009).
<http://doi.org/10.1021/ja901636v>.
- ¹²³ G. Hinze, T. Basche, and R. A. L. Vallee, “Single Molecule Probing of Dynamics in Supercooled Polymers,” *Phys. Chem. Chem. Phys.* **13**, 1813-1818 (2011).
<http://doi.org/10.1039/c0cp01654c>.
- ¹²⁴ S. Adhikari, M. Selmke, and F. Cichos, “Temperature Dependent Single Molecule Rotational Dynamics in PMA,” *Phys. Chem. Chem. Phys.* **13**, 1849-1856 (2011).
<http://doi.org/10.1039/c0cp01959c>.
- ¹²⁵ D. Bingemann, R. M. Allen, and S. W. Olesen, “Single Molecules Reveal the Dynamics of Heterogeneities in a Polymer at the Glass Transition,” *J. Chem. Phys.* **134**, 024513 (2011).
<http://doi.org/10.1063/1.3516516>.
- ¹²⁶ M. Götz and S. Schmid, “Datasets and Simulator of the Kinsoftchallenge & Related Article,” Zenodo, Version v0, (2021). <http://doi.org/10.5281/zenodo.5701310>
- ¹²⁷ F. Zosel, A. Soranno, K. J. Buholzer, D. Nettels, and B. Schuler, “Depletion Interactions Modulate the Binding between Disordered Proteins in Crowded Environments,” *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13480-13489 (2020). <http://doi.org/doi:10.1073/pnas.1921617117>.
- ¹²⁸ P. J. Smith, “A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa,” *Am. Stat.* **49**, 217-218 (1995). <http://doi.org/10.2307/2684642>.

Supplementary Material for Efficient, nonparametric removal of noise and recovery of probability distributions from time series using nonlinear-correlation functions: Additive noise

Mainak Dhar, Joseph A. Dickinson and Mark A. Berg^{a)}

Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina 29208, USA

SI. GENERATING SYNTHETIC DATA

The time series of states $s(t)$ with $s = 1, 2$, or 3 was generated from a simulation of a master equation for the time-dependent probability $\mathbf{P}(t)$ of the states,

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{k} \cdot \mathbf{P}(t), \quad (\text{S1})$$

with the rate matrix

$$\mathbf{k} = \begin{pmatrix} -k_{12} & k_{21} & 0 \\ k_{12} & -(k_{23} + k_{21}) & k_{32} \\ 0 & k_{23} & -k_{32} \end{pmatrix}. \quad (\text{S2})$$

To ensure that the system equilibrates with the probabilities $\mathbf{P}_{\text{eq}}$, the detailed-balance equations were enforced:

$$k_{12} = \frac{P_{\text{eq},2}}{P_{\text{eq},1}} k_{21}, \quad (\text{S3})$$

$$k_{23} = \frac{P_{\text{eq},3}}{P_{\text{eq},2}} k_{32}, \quad (\text{S4})$$

and

$$k_{31} = k_{13}. \quad (\text{S5})$$

The detailed-balance equations and the information in the main text determine the rate constants $k_{12} = 3.75 \text{ ms}^{-1}$, $k_{21} = 6.25 \text{ ms}^{-1}$, $k_{23} = 57.1 \text{ s}^{-1}$, and $k_{13} = 42.9 \text{ s}^{-1}$.

The signal simulation requires the Green's function for a small time step δt ,

$$\delta \mathbf{G} = \mathbf{1} + \mathbf{k} \delta t. \quad (\text{S6})$$

The time step in the simulation was the same as the time step in the final time series. A series of random numbers $\{r_i\}$ was generated from a uniform distribution from zero to one for $i = 0, \dots, L - 1$. The first value was used to start the state-time series $\{s_i\}$ from the equilibrium distribution:

$$s_1 = \begin{cases} 1; & 0 < r_0 < P_{\text{eq},1} \\ 2; & P_{\text{eq},1} < r_0 < P_{\text{eq},1} + P_{\text{eq},2} \\ 3; & P_{\text{eq},1} + P_{\text{eq},2} < r_0 < 1 \end{cases} \quad (\text{S7})$$

The state-time series was propagated using the Green's function:

$$s_{i+1} = \begin{cases} 1; & 0 < r_i < G_{1s_i} \\ 2; & G_{1s_i} < r_i < G_{1s_i} + G_{2s_i} \\ 3; & G_{1s_i} + G_{2s_i} < r_i < 1 \end{cases} \quad (\text{S8})$$

Finally, the state-time series was converted to the signal-time series using the discrete version of Eq. (4),

$$S(t_i) = \alpha (X_{s_i} + B). \quad (\text{S9})$$

SII. REGULARIZATION WITHOUT BOUNDARY CONSTRAINTS

Singular-value decomposition (SVD) is a popular means of solving ill-conditioned problems without using boundary conditions.^{1,2} It is very fast and does not require a search for the minimum. Simple truncation of the singular values simultaneously minimizes $\chi(\mathbf{P}_r)$ and

$$R_0(\mathbf{P}_r) = \frac{1}{N_P} \sum_{i=1}^{N_P} P_{r,i}^2. \quad (\text{S10})$$

This function penalizes solutions with high peaks and generally favors smoother functions. When applied to our noise-corrected moments, it yields the solution in Fig. S1. This solution is clearly unphysical, with strong oscillations and many regions of negative probability, but it reproduces the noise-corrected moments with extreme accuracy, $\chi = 8.8 \times 10^{-14}$.

A Weiner cutoff of the singular values at a value β solves Eq. (32) with $R_1(\mathbf{P}_r)$, but without the boundary constraints. This process is also known as ridge regularization.³ The solutions do not improve significantly from Fig. S1.

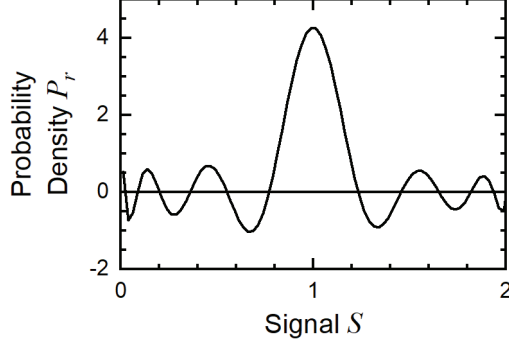


FIG. S1. The recovered probability distribution using only the regularizing function R_0 [Eq. (S10)] and no boundary conditions. The match to the measured moments is almost exactly ($\chi = 8.8 \times 10^{-14}$), but the result is far from the correct solution (Fig. 1). $N_\mu = 13$ moments.

SIII. MINIMIZATION ALGORITHM

The regularization conditions were chosen, in part, because minimization of a quadratic form with linear constraints and constant boundaries is simple and efficient.⁴ To set up the solution of Eq. (32) without a sparsity regularizer ($\alpha = 0$), we first look at a more general problem written in matrix form:

$$\begin{aligned} \min_{\mathbf{P}_r} & \left[\left\| \mathbf{W}(\tilde{\mathbf{V}} \cdot \mathbf{P}_r - \tilde{\boldsymbol{\mu}}') \right\|^2 + \beta^2 \left\| \mathbf{R}(\mathbf{P}_r - \mathbf{P}_0) \right\|^2 \right] \\ \text{subject to} & \quad 0 \leq P_{r,i}; \quad i = 1, \dots, N_P \\ \text{and} & \quad \mathbf{1} \cdot \mathbf{P}_r = 1 \end{aligned} \quad (\text{S11})$$

The smoothing regularizing function is $\mathbf{R}$. The matrix $\mathbf{W}$ is diagonal with elements giving the relative weights for the moments. The first row of the Vandermonde matrix $\mathbf{V}$ [Eq. (29)] is dropped to give $\tilde{\mathbf{V}}$ and the first (zeroth) element of $\boldsymbol{\mu}'$ is dropped to give $\tilde{\boldsymbol{\mu}}'$. The vector $\mathbf{P}_0$ is a preferred solution.

Although Bayesian statistics are usually parametric,⁵ the regularization methods used in this nonparametric problem mimic Bayesian ideas. Here, $\mathbf{P}_0$ mimics a prior distribution that will be improved by including the new data $\tilde{\boldsymbol{\mu}}'$. The regularizing parameter then controls the confidence given to the prior information relative to the new data. From this perspective, Eq. (32) assumes a “no-information” prior. These priors give zero for the regularization function and, therefore, do not appear explicitly. For example, the smoothing regularizer $R_1(P)$ [Eq. (33)] has a no-information prior that is flat, $P_0(S) = 1/(S_{\max} - S_{\min})$. The sparsity regularizer $R_2(P, P_{\text{sp}})$ [Eq. (36)] has any single-state distribution as a no-information prior, for example, $P_0(S) = \delta(S - \langle S \rangle)$.

Following Ref. 4, Eq. (S11) can be converted to

$$\begin{aligned} \min_{\mathbf{P}_r} & \left\| \mathbf{C} \cdot \mathbf{P}_r - \mathbf{d} \right\|^2 \quad \text{subject to} \quad 0 \leq P_{r,i}; \quad i = 1, \dots, N_P \\ \text{and} & \quad \mathbf{1} \cdot \mathbf{P}_r = 1 \end{aligned} \quad (\text{S12})$$

with the concatenated matrix

$$\mathbf{C} = \begin{pmatrix} \mathbf{W} \cdot \tilde{\mathbf{V}} \\ \beta \mathbf{R} \end{pmatrix} \quad (\text{S13})$$

and the concatenated vector

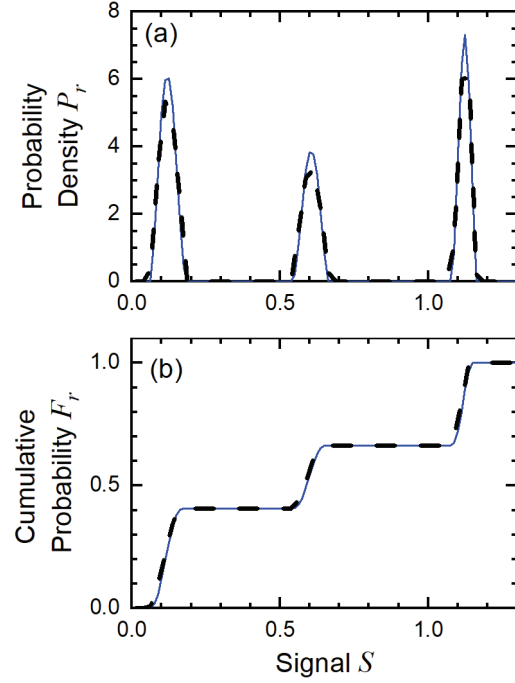


FIG. S2. Effect of the point density on the recovered distribution. The black, dashed lines are from Fig. 11 ($N_P = 81$, $\delta S = 0.025$). The blue, solid lines are from an identical calculation, except with twice the point density ($N_P = 161$, $\delta S = 0.0125$). (a) Probability densities and (b) cumulative probabilities.

$$\mathbf{d} = \begin{pmatrix} \mathbf{W} \cdot \tilde{\boldsymbol{\mu}}' \\ \beta \mathbf{R} \cdot \mathbf{P}_0 \end{pmatrix}. \quad (\text{S14})$$

To specialize to our problem, we use $\mathbf{R} = (N_P - 1)^{-1/2} \mathbf{D}$ for the regularization matrix [Eq. (34)]. With $\mathbf{W} = N_\mu^{-1/2} \text{diag}(\mu_n^{-1})$, the first term in Eq. (S11) matches the definition of χ^2 in Eq. (30). As the preferred solution, we take a constant, $\mathbf{P}_0 = 1/N_P$. This factor drops out, because $\mathbf{D}\mathbf{P}_0 = \mathbf{0}$.

The sparsity regularizer [Eq. (36)] is not quadratic and cannot be easily added to the minimization. However, it is linear and can be added as a constraint:

$$\begin{aligned} \min_{\mathbf{P}_r} & \left\| \mathbf{C} \cdot \mathbf{P}_r - \mathbf{d} \right\|^2 \quad \text{subject to} \quad 0 \leq P_{r,i}; \quad i = 1, \dots, N_P \\ \text{and} & \quad \mathbf{1} \cdot \mathbf{P}_r = 1 \\ \text{and} & \quad \sum_{P_i < P_{\text{sp}}} P_{r,i} = R_2 \end{aligned} \quad (\text{S15})$$

The sparsity is adjusted by choosing the value of $R_2(\alpha)$ instead of the value of α . Equations (S12) and (S15) were solved using the *lsqlin* function in MATLAB.

Figure S2 shows the effect of changing the point density on the minimization. The results are only slightly narrower with a higher point density. The peak widths are limited by the data, not by the calculation.

SIV.OPTIMIZING THE REGULARIZATION PARAMETERS

The most difficult issue in any regularization problem is choosing an optimal solution. That issue translates to specify optimal values for the regularization parameters; N_μ , β , α and P_{sp} ; and the correct level of fitting error χ^* that they imply. In the main text, we made a subjective judgement. However, more objective methods are also popular.⁶⁻¹² We consider some of them here, but we find that they do not work well on our problem.

The first quantity to consider in any fitting problem is the error in the fit to the data χ . Unfortunately, χ is not a direct indication of the accuracy of the result. As discussed in the main text, the signal-sampling error always creates a moment sequence that can be fit with no error. The noise-sampling error is more variable, but it can also create a moment sequence that is close to one that exactly fits an incorrect solution. The fitting error is only the component of the noise-sampling error that does not correspond to an allowed moment sequence. As a result, the optimum fitting error χ^* can be much less than the moment error ζ_n . Figure S3(a) shows the fitting errors corresponding to the moment errors in Fig. 16. The exponential rise of χ with order is similar to the rise in noise-sampling error in Fig. 16(b), but its magnitude is much less.

One can conjecture that the fitting error will be a fixed fraction c of the noise-sampling error, $\chi = c\zeta$. The expected fitting error X is then the average of the expected moment error [Eq. (40)]

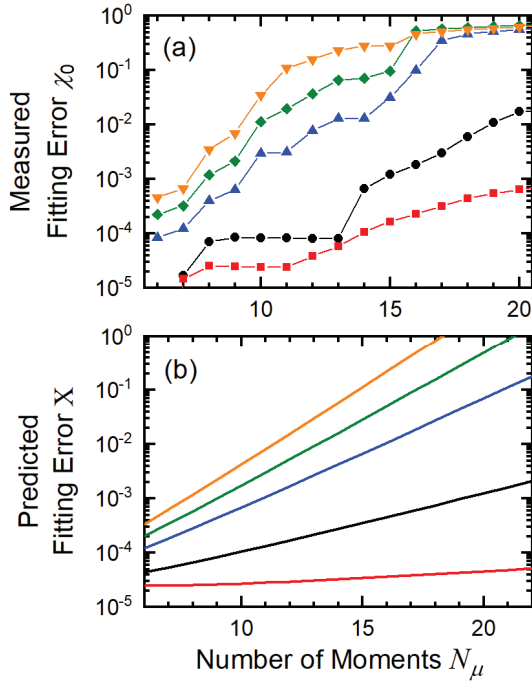


FIG. S3. Fitting error versus number of moments in the fit and the signal-to-noise ratio: SNR = 1.16 (red circles), 0.58 (black squares), 0.39 (blue, up triangles), 0.33 (green diamonds), and 0.29 (orange, down triangles). (a) Results corresponding to Fig. 16. (b) Predictions of Eq. (S17).

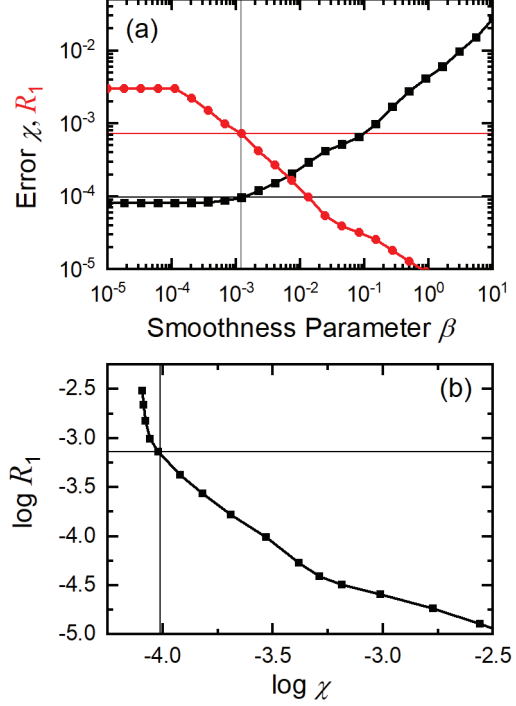


FIG. S4. (a) A plot of $\log \chi$ (red circles) and $\log R_1$ (black squares) vs $\log \beta$ (a “regularization pathway”) for the problem in Fig. 8. (b) A parametric plot of $\log R_1$ versus $\log \chi$ (the “L-curve”) for $\beta = 0-10$. The thin reference lines indicate our choice of solution (Fig. 11)

$$\begin{aligned} X^2 &= E(\chi^2) \\ &= \frac{1}{N_\mu} \int_0^{N_\mu} c^2 \omega_{\varepsilon,n}^2 dn. \end{aligned} \quad (\text{S16})$$

Integrating and keeping the leading term in N_μ gives

$$X^2 = \frac{2ac^2 A^{N_\mu}}{N_\mu N_{cr} \ln A}; \quad A = 1 + \frac{1}{(2b\text{SNR})^2}. \quad (\text{S17})$$

Figure S3(b) shows calculations with $2ac^2 = 2.25 \times 10^{-2}$. The predictions have a semiquantitative similarity to the results in Fig. S3(a), but do not match quantitatively. The correct size of the fitting error can be roughly anticipated. For example, the solution of Fig. S1 can be rejected as overfit. However, χ^* cannot be predicted with enough accuracy to select an optimum solution.

Some popular methods for choosing a regularized solution, for example, generalized cross-validation,⁷ rely on having uncorrelated errors.⁸ However, the errors in our problem are correlated. The signal-sampling error is highly correlated from moment to moment [Fig. 14(c)]. The noise-sampling error is more random, but each realization has its own discernable pattern [Fig. 14(b)].

It is generally recognized that the regularization parameters themselves, such as β or α , are not directly meaningful. However, one can look at patterns in their behavior. Figure S4(a) shows a regularization “pathway”, χ and R_1 as a function of the smoothing parameter β (same data as Fig. 8). Figure S4(b)

replots the data as an “L-curve”.^{8,9} Hanson and O’Leary argue that the corner of this curve, defined as the point of maximum curvature, is the optimum solution, even in the presence of correlated errors,⁸ although others have pointed out the limitations of this idea.^{11,12} However, for either of these plots to be useful, one needs to know the relative weight that should be given to matching the data versus having a smooth solution. Our choices, which are shown on the figure, weight the data more heavily.

SV. DERIVING THE SAMPLING-ERROR EQUATIONS

In deriving Eqs. (38)–(40), we start with the assumption that the noise is unbiased, that is,

$$E(D|S) = S. \quad (\text{S18})$$

Then, the primed power of the data [Eq. (19)] is also unbiased:

$$\begin{aligned} E(D'') &= E\left(E(D_n \dots D_1 | \{S_j\})\right) \\ &= E(E(D_n | S_1) \dots E(D_1 | S_1)) \\ &= E(S^n). \end{aligned} \quad (\text{S19})$$

In the first line, we have first taken the primed power conditioned on a specific signal trajectory $\{S_j\}$ and used the law of total expectation. Assuming fast noise and slow signal gives the second line. Equation (S18) then gives the final result. The noise-corrected moment is then unbiased:

$$E(\mu'_n[D]) = E\left(N_{\text{er}}^{-1} \sum_{i=1}^{N_{\text{er}}} D_i''\right) = \mu_n[P_S]. \quad (\text{S20})$$

Throughout this section, moments of a time series, e.g., $\mu'_n[D]$, indicate a value from a finite, sampled data set; moments on a probability distribution, e.g., $\mu_n[P_S]$, indicate an ideal, population statistic.

The variance of the noise-corrected moment is of more interest:

$$\begin{aligned} \text{Var}(\mu'_n[D]) &= \text{Var}\left(N_{\text{er}}^{-1} \sum_{i=1}^{N_{\text{er}}} D_i''\right) \\ &= E\left[\text{Var}\left(N_{\text{er}}^{-1} \sum_{i=1}^{N_{\text{er}}} D_i'' \middle| \{S_j\}\right)\right] \\ &\quad + \text{Var}\left(E\left(N_{\text{er}}^{-1} \sum_{i=1}^{N_{\text{er}}} D_i'' \middle| \{S_j\}\right)\right). \end{aligned} \quad (\text{S21})$$

The second line uses the law of total variance to again separate the result for a specific signal series. The assumptions of fast noise and slow signal yield

$$\begin{aligned} \text{Var}(\mu'_n[D]) &= E\left(N_{\text{er}}^{-2} \sum_{i=1}^{N_{\text{er}}} \text{Var}\left(D_i'' \middle| S_i\right)\right) \\ &\quad + \text{Var}\left(N_{\text{er}}^{-1} \sum_{i=1}^{N_{\text{er}}} E\left(D_i'' \middle| S_i\right)\right). \end{aligned} \quad (\text{S22})$$

Knowing that the processes are stationary and using Eq. (S19) gives

$$\begin{aligned} \text{Var}(\mu'_n[D]) &= N_{\text{er}}^{-1} E\left(\text{Var}\left(D'' \middle| S\right)\right) \\ &\quad + \text{Var}\left(N_{\text{er}}^{-1} \sum_{i=1}^{N_{\text{er}}} S_i^n\right). \end{aligned} \quad (\text{S23})$$

Dividing by $\mu_n[P_S]$ gives Eq. (38), which separates the total, fractional sampling error into two components. The noise-sampling error comes from the first term in Eq. (S23):

$$\omega_{\varepsilon,n}^2 = \frac{1}{\mu_n[P_S]^2 N_{\text{er}}} E\left(\text{Var}\left(D'' \middle| S\right)\right) \quad (\text{S24})$$

The second term in Eq. (S23) forms the signal-sampling error,

$$\omega_{S,n}^2 = \frac{\text{Var}(\mu_n[S])}{\mu_n[P_S]^2}. \quad (\text{S25})$$

The signal-sampling error is complicated by the long correlation time of the signal. We assume that the total data set can be separated into N_S effectively independent samples of the signal with

$$N_S = N_{\text{en}} \left(1 + \frac{T_L}{T_{\text{er}}}\right). \quad (\text{S26})$$

Each new series in the data set starts with an new sample and a new sample is generated every time a signal-correlation time passes within the series. The signal correlation has multiple time scales, so we just use the longest T_{er} . Equation (S25) is approximated by assuming the signal is constant within each of these samples:

$$\omega_{S,n}^2 = \frac{1}{\mu_n[P_S]^2} \text{Var}\left(N_S^{-1} \sum_{i=1}^{N_S} S_i^n\right). \quad (\text{S27})$$

By the independence of these samples,

$$\begin{aligned} \omega_{S,n}^2 &= \frac{1}{\mu_n[P_S]^2 N_S^2} \sum_{i=1}^{N_S} \text{Var}(S^n) \\ &= \frac{1}{N_S} \frac{\mu_{2n}[P_S] - \mu_n[P_S]^2}{\mu_n[P_S]^2}, \end{aligned} \quad (\text{S28})$$

The final formula depend on ratios of moments:

$$\omega_{S,n}^2 = \frac{1}{N_{\text{en}} \left(1 + \frac{T_L}{T_{\text{er}}}\right)} \left(\frac{\mu_{2n}[P_S]}{\mu_n[P_S]^2} - 1\right). \quad (\text{S29})$$

The important factor the noise-sampling error [Eq. (S24)] is

$$\begin{aligned}
\text{Var}(D^{n'} | S) &= \text{Var}\left(\prod_{i=1}^n D_i | S\right) \\
&= \prod_{i=1}^n \left(\text{Var}(D_i | S) + E(D_i | S)^2 \right) - \prod_{i=1}^n E(D_i | S)^2 \\
&= \left(\text{Var}(D | S) + S^2 \right)^n - S^{2n} . \quad (\text{S30})
\end{aligned}$$

The second line uses the variance of a product of uncorrelated variables. The last line come from Eq. (S18) and the stationarity of the noise. With the binomial formula, Eq. (S24) becomes

$$\omega_{\varepsilon,n}^2 = \frac{1}{\mu_n [P_S]^2 N_{\text{cr}}} \sum_{k=1}^n \binom{n}{k} E \left(\text{Var}(D | S)^k S^{2(n-k)} \right). \quad (\text{S31})$$

Note the absence of the $k = 0$ term.

At this point, we specialize to a particular type of noise additive noise. Equation (6) implies that

$$\text{Var}(D | S) = \sigma_{\varepsilon}^2. \quad (\text{S32})$$

Thus, the noise-sampling error is also reduced to ratios of errors:

$$\omega_{\varepsilon,n}^2 = \frac{1}{N_{\text{cr}}} \sum_{k=1}^n \binom{n}{k} \sigma_{\varepsilon}^{2k} \frac{\mu_{2(n-k)} [P_S]}{\mu_n [P_S]^2}, \quad (\text{S33})$$

So far, the approximations have been mild, but to simplify further, we need a stronger one. To evaluate the moment ratios, we approximate the moments with

$$\mu_n = \frac{(2b\mu_1)^n}{2a}, \quad (\text{S34})$$

where a and b are unitless constants of order unity. This form has the correct units and holds exactly in limit of high n [Eq. (26)]. In this asymptotic case,

$$a = \frac{1}{2P_{\text{max}}} \quad (\text{S35})$$

and

$$b = \frac{S_{\text{max}}}{2\mu_1}. \quad (\text{S36})$$

Because a successful distribution recovery requires working in or near the asymptotic region, this approach is reasonable. Using Eq. (S34) in Eqs. (S29) and (S33), give Eqs. (39) and (40), respectively.

SVI. EXPERIMENTAL CORRELATION FUNCTIONS

Figure S5 contains high-order correlation functions from the experimental data discussed in Sec. IV.D. The lowest order function has been shown in Fig. 20, along with a single-exponential fit with a time constant of $T = 175$ ms. The functions in Fig. S5 have also been fit with single exponentials constrained to have the same time constant.

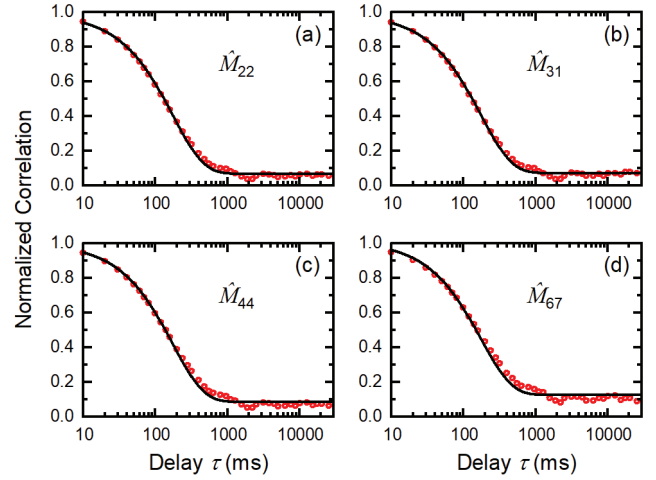


FIG. S5. High-order, normalized, moment-correlation functions (red circles) and single exponential fits (black curve) for the experimental data set. The time constants have been fixed to the value found for $\hat{M}_{11}$ (Fig. 20).

- ¹W. H. Press, S. A. Teukolsky, W. T. Vetterling, and P. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1992).
- ²R. Penrose, "On Best Approximate Solutions of Linear Matrix Equations," *Math. Proc. Cambridge Phil. Soc.* **52**, 17-19 (1956). <http://doi.org/10.1017/S0305004100030929>.
- ³M. E. S. Arashi, M. Arashi, and B. M. G. Kibria, *Theory of Ridge Regression Estimation with Applications* (John Wiley & Sons, 2019). <http://doi.org/10.1002/9781118644478>.
- ⁴Å. Björck, *Numerical Methods for Least Squares Problems* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996).
- ⁵G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (Addison-Wesley Pub. Co., Reading, Mass, 1973).
- ⁶M. E. Kilmer and D. P. O'Leary, "Choosing Regularization Parameters in Iterative Methods for Ill-Posed Problems," *SIAM J. Matrix Anal. Appl.* **22**, 1204-1221 (2001). <http://doi.org/10.1137/S0895479899345960>.
- ⁷G. H. Golub, M. Heath, and G. Wahba, "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter," *Technometrics* **21**, 215-223 (1979). <http://doi.org/10.2307/1268518>.
- ⁸P. C. Hansen and D. P. O'Leary, "The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems," *SIAM J. Sci. Comput.* **14**, 1487-1503 (1993). <http://doi.org/10.1137/0914086>.
- ⁹P. C. Hansen, "Analysis of Discrete Ill-Posed Problems by Means of the L-Curve," *SIAM Rev.* **34**, 561-580 (1992). <http://doi.org/10.1137/1034115>.
- ¹⁰C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Chapt. 25 (Prentice-Hall, Englewood Cliffs, NJ, 1974).
- ¹¹C. R. Vogel, "Non-Convergence of the L-Curve Regularization Parameter Selection Method," *Inverse Probl.* **12**, 535 (1996). <http://doi.org/10.1088/0266-5611/12/4/013>.
- ¹²H. W. Engl and W. Grever, "Using the L-Curve for Determining Optimal Regularization Parameters," *Numer. Math.* **69**, 25-31 (1994). <http://doi.org/10.1007/s002110050078>.