Detection of local mixing in time-series data using permutation entropy

Michael Neuder and Elizabeth Bradley[*]
Department of Computer Science, University of Colorado, Boulder, Colorado 80309, USA

Edward Dlugokencky
National Oceanic and Atmospheric Administration, Boulder, Colorado 80305, USA

James W. C. White
Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado 80309, USA

Joshua Garland[†]
Santa Fe Institute, Santa Fe, New Mexico 87501, USA

(Received 28 October 2020; revised 5 January 2021; accepted 22 January 2021; published 26 February 2021)

Mixing of neighboring data points in a sequence is a common, but understudied, effect in physical experiments. This can occur in the measurement apparatus (if material from multiple time points is pulled into a measurement chamber simultaneously, for instance) or the system itself, e.g., via diffusion of isotopes in an ice sheet. We propose a model-free technique to detect this kind of *local mixing* in time-series data using an information-theoretic technique called permutation entropy. By varying the temporal resolution of the calculation and analyzing the patterns in the results, we can determine whether the data are mixed locally, and on what scale. This can be used by practitioners to choose appropriate lower bounds on scales at which to measure or report data. After validating this technique on several synthetic examples, we demonstrate its effectiveness on data from a chemistry experiment, methane records from Mauna Loa, and an Antarctic ice core.

DOI: 10.1103/PhysRevE.103.022217

I. INTRODUCTION

In many fields of science, data are measured in gas or fluid states where mixing can occur: in measurement chambers and laboratory piping, for instance. If that process has the effect of intermingling samples from different time points, it can obfuscate the results. A continuous flow analysis system in an ice-core laboratory, for example, melts ice from multiple years into a single bolus, mixing the isotopes from each data point (e.g., 6000 BCE) with those from surrounding ice (e.g., 5990–6010 BCE). Similar effects can also be at work in the system under study, e.g., diffusion of isotopes in the ice sheet, which mixes climate data from successive years. In the face of these challenges, choosing an appropriate interval at which to measure, analyze, or report the data is often an imperfect balance between time, money, laboratory capabilities, and scientific need. Frequently, educated guesses are the only way to make these critical choices.

In this paper, we describe a technique to detect these effects in a data set; not only their presence, but also the scale on which they operate, which we term the *local mixing scale*. This gives practitioners a way to know precisely where to draw the line, in terms of sampling the system and reporting the data. Critically, our approach is model-free, providing results without any need for domain-specific knowledge. The technique is based on a method called permutation entropy, which provides an estimate of the rate at which new information is produced in a data sequence; in effect, a measure of predictability. By varying the "stride" of this calculation, we can detect whether or not the data bear the scars of mixing.

The underlying idea is as follows: for most time series, predictability tends to decrease as one extends the horizon, so permutation entropy will generally increase with the stride of the calculation because the data points involved span wider and wider temporal ranges. If successive data points are measured on a smaller scale than the mixing scales that are inherent in the data, though, each measurement is essentially a single draw from a local distribution composed of the data points in its local neighborhood. By definition, this added randomness will raise the entropy rate. Reversal of the normal pattern of the relationship between the stride of the calculation and the permutation entropy values, then, is an effective way to detect local mixing. Used in tandem with bin averaging, this method also allows one to detect the scale of those mixing effects and adjust one's procedures accordingly.

In Sec. II, we describe the techniques involved in our approach, beginning with some background on permutation entropy and a description of the metric that we have developed to capture the effects described in the previous paragraph. We validate our technique using two synthetic examples (Secs. III A and III B) and explore its utility in the context

[*]Also at Santa Fe Institute, Santa Fe, New Mexico 87501, USA.
[†]Author to whom correspondence should be addressed: joshua@santafe.edu

of three real-world time-series data sets: one from a chemistry experiment involving gas mixtures (Sec. III C), a second from an Antarctic ice core (Sec. III D), and a third from the Mauna Loa methane records (Sec. III E). We discuss these results and their implications in Sec. IV and conclude, with some thoughts about broader applications and future directions, in Sec. V.

II. MATERIALS AND METHODS

A. Permutation entropy

Permutation entropy [1] is a method for estimating the Shannon entropy rate [2] of an arbitrary time-series data set. The calculation focuses on the local relationships among a sequence of points by mapping their values to an ordering of the same length. The three-point sequence [7,3,11], for instance, would map to the ordering [1,0,2] because $3 < 7 < 11$. The statistics of these ordinal sequences or *permutations* are calculated as follows. Let $|\cdot|$ denote set cardinality and Ord define a function that calculates the ordering of a given sequence of length ℓ, as in the $\ell = 3$ example above. Additionally, define a time series as the sequence $(x_i)_{i\in I_N}$, where the index set is $I_N = \{k \in \mathbb{N} \setminus \{0\} \mid k \leqslant N\}$. Then the probability of the appearance in the time series of ℓ sequences of data points whose values map to a specific ordering π can be estimated by

$$\Pr[\pi] = \frac{|\{n \in I_{N-\ell+1} \mid \mathrm{Ord}(x_n, x_{n+1}, \dots, x_{n+\ell-1}) = \pi\}|}{N - \ell + 1}. \tag{1}$$

By calculating the probability of each of the $\ell!$ orderings across the whole time series, one constructs the permutation entropy or PE:

$$\mathrm{PE} = -\frac{1}{\log_2 \ell!} \sum_{\pi} \Pr[\pi] \log_2 \Pr[\pi]. \tag{2}$$

This is an estimate of the average rate at which new information appears in a time series per observation [1]. With the $1/\log_2 \ell!$ normalization, its values range from 0 to 1. If PE is low, the observations, on average, contain a significant amount of information about the past.[1] This is the case for a simple sine wave. For a time series created by sampling $\sin(x)$ at a spacing of $\Delta x = 0.0001$, for instance, the PE is approximately 0.024. As PE increases, more of the information in each observation is new. The PE of the time series in the top panel of Fig. 1 (the canonical Lorenz system, described at more length later in this paper) is ≈ 0.28, reflecting the ease of predicting a noiseless deterministic signal.[2] If the signal were *wholly* random, the PE would be 1. The value of the PE is a function of the subsequence length ℓ (please see [4,6,7] for more discussion of this parameter and its implications).

As defined above, PE captures the permutation entropy of an entire data set in the form of a single value. One can also perform this calculation in a sliding window across a time series to increase the temporal resolution of the results. This can be a good idea, for example, if the dynamics of the system at hand are nonstationary. In this variant of the calculation, each PE value captures the statistics of the orderings in a window around the associated time point. As shown in the bottom panel of Fig. 1, this windowed analysis brings out subtle changes in the predictability of the Lorenz time series in the top panel as the system moves through different parts of its state space. However, it does introduce another free parameter into the calculation: the window width w. See [8,9] for discussion of the associated choices and issues.

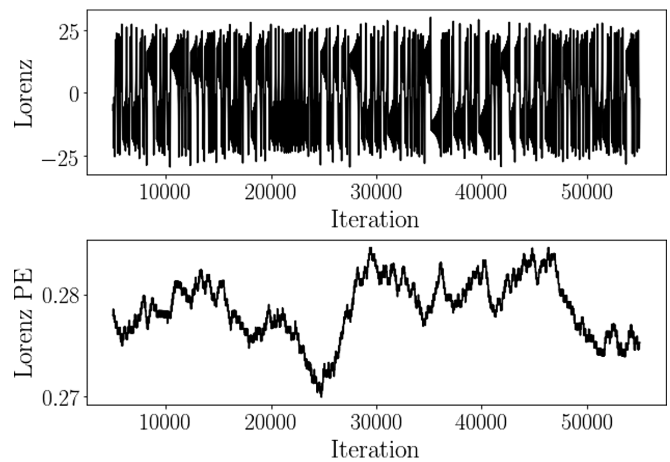

FIG. 1. Permutation entropy example. Top: a short segment of a chaotic signal from the Lorenz system. Bottom: permutation entropy of the Lorenz time series in the top panel.

All of the calculations reported in this paper use $\ell = 4$ and $w = 5000$, which are values chosen following best practices in the permutation-entropy literature [3,7,8,10]. In a window of this length, each permutation of length four will have approximately 200 opportunities to appear at least once, which should be enough to rule out any forbidden ordinals [11].

B. Permutation entropy at varying temporal resolutions

The examples and equations in the previous section assume that one always works with ordinal permutations that are constructed from the values of *successive* points in a time series. It can also be useful to coarsen the grain of this procedure using τ-separated points instead, thereby changing the temporal resolution of the calculation. This variant of the PE technique, which was introduced in Ref. [8], requires a few modifications to the first step in the calculations. Specifically, Eq. (1) becomes

$$\Pr[\pi] = \frac{|\{n \in I_{N-(\ell+1)\tau} \mid \mathrm{Ord}(x_n, x_{n+\tau}, \dots, x_{n+(\ell-1)\tau}) = \pi\}|}{N - (\ell + 1)\tau}, \tag{3}$$

where τ defines the spacing between the points in the time series that are used to construct the ℓ-element ordinal permutations. With $\tau = 2$, for instance, the first permutation constructed from the sequence [1,4,6,2,5,3] would be [0,2,1] because $1 < 5 < 6$.

The spacing parameter τ is the focus of the work reported here, and the central point of leverage for our technique.

[1] As such, PE has been shown to correlate with predictability (or lack thereof) [3–5].

[2] The integration step size used to generate the time series also plays a role here; if it is very small, successive points are highly correlated and the PE will be lower.

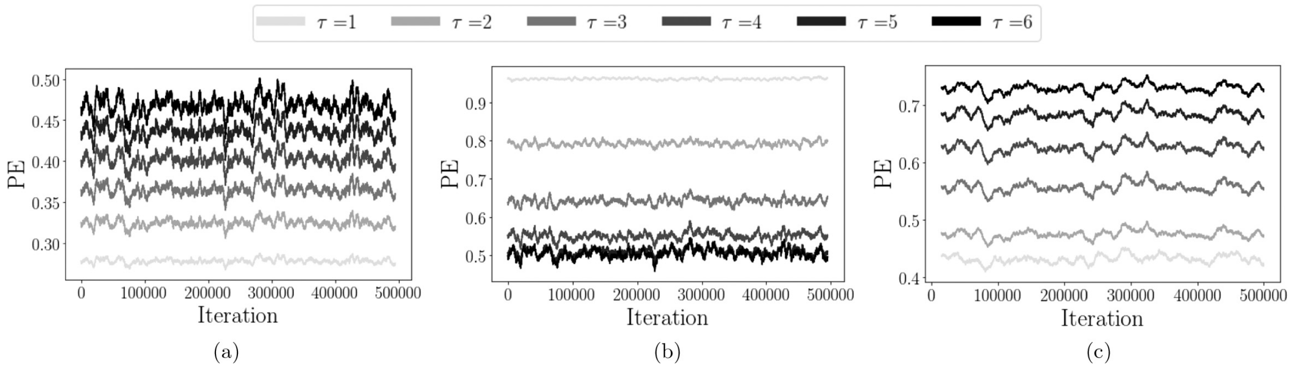

FIG. 2. Lorenz permutation entropy. (a) Permutation entropy (PE) of the full version of the Lorenz signal from the top panel of Fig. 1, calculated with a range of values of τ, the spacing between data points used to construct the permutations. (The trace in the bottom panel of Fig. 1 is a short segment of the $\tau = 1$ trace in this image.) (b) If local mixing effects are artificially added to the same signal, the ordering of the PE traces is reversed. (c) A simple bin-averaging operation removes the local mixing effects, restoring the normal order of the traces.

Because this parameter controls the "stride" of the calculation, changing its value allows one to understand the information dynamics of the system at different resolutions. This can reveal some interesting patterns: persistence of features in PE results across a range of τ values, for instance, indicates an effect in the underlying signal that spans multiple timescales. In a wide range of data sets, we have observed that *increasing the τ value generally raises the PE curves*. These data sets include a number of classic chaotic systems, two of which are used as examples later in this paper, as well as transient deterministic dynamics (e.g., the examples used in Ref. [8]), various paleoclimate records (e.g., [9,12]), various 1-s and 1-min financial price data sets, computer performance data from the experiments reported in Refs. [13,14], and experimental data from a driven damped pendulum, which are available at [15]. Figure 2(a) demonstrates this in the context of the Lorenz signal from Fig. 1.

A monotonic increase in the PE values with increasing τ simply reflects decreasing predictability over the longer time span sampled by each permutation. When there is local mixing in a time series, though, that reasoning no longer applies. Rather, because the data points used to construct individual permutations are measured on a smaller scale than the mixing scales that are inherent in the data, each element of those permutations is essentially a single draw from a local distribution composed of the data points in the neighborhood of each individual measurement. By definition, this added randomness will raise the PE. As τ increases, though, the points used to generate the orderings will spread out across those local distributions. When that point spacing exceeds the local distribution width (i.e., the largest mixing scale at work in the data), the extrinsic PE increase [4] caused by that mixing will be reduced and eventually eliminated. The schematic in Fig. 3 illustrates the mechanism underlying this effect.

The takeaway here is that, when local mixing is at work in a data set, its PE will not increase monotonically with τ. Rather, the vertical ordering of the PE curves in a plot like the ones in Fig. 2 will be reversed, with the low-τ traces at the top and the high-τ traces at the bottom, the opposite of what we expect for a deterministic system. This reversal is not only a clear indicator of a specific issue in the data. With some additional experimentation, as described in Sec. II D, this effect can actually bring out the scale of the mixing. First, though, we need a metric for assessing the relationships between permutation entropy curves calculated with different values of τ.

C. Reversal metric

The purely monotone nature of the vertical orderings of the different PE traces in Fig. 2 persists across the entire span of the time series, but that is not always the case; for some

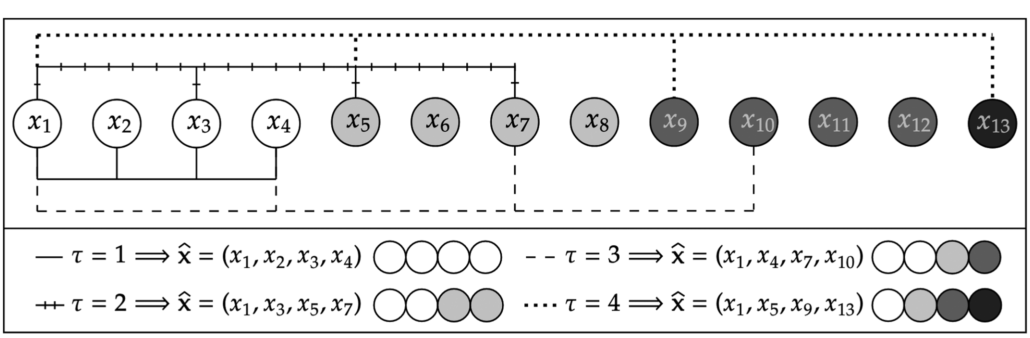

FIG. 3. Mixing and permutation entropy. If the values of neighboring points are not independent and the τ value is small, then the points used to construct each permutation are effectively drawn from a single local distribution, schematized here with different shades. In this case, even though the underlying data may be deterministic, the relationships of these draws will be random and the permutation entropy will be artificially high. Higher τ values widen the span of the calculation beyond the spread of the local distributions, thereby mitigating this effect.

data sets, the traces will touch, or even cross, at different points in the time series. In order to assess the overall form of these patterns, what we want is a measure of how close this vertical ordering is to a purely monotone-increasing pattern at different points in the time series. To quantify this formally, we develop the following metric. Let $\tau_{\min}$ and $\tau_{\max}$ denote the smallest and largest values of τ used in the PE calculation, with $\tau_{\min} < \tau_{\max}$. Additionally, let $\mathbf{v}_i$ denote the vector of monotone-increasing values of τ:

$$\mathbf{v}_i = [\tau_{\min} \quad \tau_{\min+1} \quad \ldots \quad \tau_{\max}]^T.$$

If $\tau_{\min} = 1$ and $\tau_{\max} = 6$, for instance, this vector would be

$$\mathbf{v}_i = [1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6]^T.$$

We construct a *focal τ-sequence vector*, $\mathbf{v}$, that captures the ordering of the values of PE at a given time n. (For the example in Fig. 2, $\mathbf{v} = [1\ 2\ 3\ 4\ 5\ 6]^T\ \forall\ n$.) We define the reversal metric $\mathcal{R}$ as the distance between this focal τ-sequence vector and the purely monotone-increasing vector $\mathbf{v}_i$:

$$\mathcal{R}(\mathbf{v}) = \|\mathbf{v} - \mathbf{v}_i\|/\lambda, \tag{4}$$

where $\mathbf{v} \in \mathcal{S}_{\tau_{\max}}$, the set of permutations of order $\tau_{\max}$, and $\|\cdot\|$ denotes the $L1$ norm. The scaling term λ, which is defined as

$$\lambda = \max_{\mathbf{y}} \|\mathbf{y} - \mathbf{v_i}\|$$

with $\mathbf{y} \in \mathcal{S}_{\tau_{\max}}$, normalizes the value of $\mathcal{R}$ to run from 0 to 1, where 0 indicates perfect monotone-increasing order of the different PE curves with τ. This is the case for every time point in Fig. 2(a). $\mathcal{R} = 1$, on the other hand, indicates perfect monotone-decreasing order, as in Fig. 2(b), where local-mixing effects have been artificially added to the Lorenz signal via the procedure described in the following section. One can average these pointwise values of $\mathcal{R}$ across the time series, or over subsequences of it, in order to assess the overall reversal pattern in the corresponding span. In the examples in Figs. 2(a) and 2(b), that average value $\overline{\mathcal{R}}$ is identical for all subsequences of the data. If the reversal patterns vary across the time series, though, that will not be the case, as will be shown in Sec. III E.

The way of measuring distance that is encoded in Eq. (4), which is a normalized version of Spearman's footrule [16,17], is more appropriate in our application than the Kendall-τ distance [17,18] (which is essentially the number of pairs on which the relative order of two permutations disagrees) because a reversal of, say, the $\tau = 1$ and $\tau = 6$ traces is more meaningful than a reversal of the $\tau = 3$ and $\tau = 4$ traces because of the timescales involved. Other choices are of course possible for the form of this metric, as discussed in Sec. IV.

D. Data manipulation techniques

To validate our conjectures about the effect of local mixing on the τ orderings of PE traces, we perform a number of tests using synthetic time-series data sets $\mathbf{x} = (x_n)_{n\in I_N}$ from well-known dynamical systems, manipulating them in order to create *Ansätze* that replicate the effects of that process on those data; specifically, how it causes the interchange of information within some "mixing window" that includes k points before and after each measurement. To do this, we create a local normal distribution at each point x_n whose mean and variance are calculated from the points in that mixing window around that point. To produce the corresponding point of the *Ansatz*, $\hat{x}_n$, we make a random draw from that distribution:

$$\hat{x}_n \sim \text{Normal}\big(\mu_n, \sigma_n^2\big),$$

where μ_n and σ_n are the mean and standard deviation of the points of $\mathbf{x}$ in the window:

$$[x_{n-k}, \ldots, x_n, \ldots, x_{n+k}].$$

This is a comparatively simple model of the effects of local mixing on a series of data points; it does not capture long-range effects, for instance. But for a wide range of experimental situations, this is a useful approximation.

In the following section, we also make use of binned averaging, where one uses groups of j consecutive points of a time series $\mathbf{y}$ to create the elements of a new series $\bar{\mathbf{y}}$:

$$\bar{y}_n = \left(\frac{y_{(n-1)j+1} + y_{(n-1)j+2} + \cdots + y_{nj}}{j}\right)_{n\in I_{N/j}}. \tag{5}$$

This is not a *rolling* bin average; rather, the first point of $\bar{\mathbf{y}}$ is the average of the first j points of $\mathbf{y}$, the second is the average of $[y_{j+1} \ldots y_{2j}]$, and so on. This is intended to mimic what happens in experimental practice when j successive points in a raw data set are averaged together—in the system, in the post processing, or in the laboratory apparatus. In the analyses that follow, this technique serves two purposes: as an extra validation step in the synthetic examples and as a diagnostic of mixing scales in the real-world data sets.

III. RESULTS

As validation and demonstration cases for our mixing-detection technique, we use two synthetic examples: classic systems from the field of nonlinear dynamics and three real-world time-series data sets. The synthetic-data examples allow us to manipulate the time series, as described in Sec. II D, in order to validate our reasoning about mixing and τ reversal. The experimental data sets—chemical sensor measurements from a gas mixing experiment, water-isotope data from an Antarctic ice core, measured by a spectrometer, and atmospheric methane from the Global Monitoring Laboratory observatory on Mauna Loa, Hawaii—provide a view into the utility of this technique in real-world settings. Figure 4 shows data from the three real-world data sets.

A. Lorenz system

The Lorenz system [19] is defined by the following ordinary differential equations:

$$\dot{x} = a(y - x),$$
$$\dot{y} = x(r - z) - y,$$
$$\dot{z} = xy - bz.$$

To create a synthetic data set from this system, we choose $a = 16,\ b = 4$, and $r = 45$, parameter values that are known to produce chaotic behavior, and integrate the ordinary differential equations using a fourth-order Runge-Kutta method for 500 000 steps with a step size of $h = 0.005$, starting from the

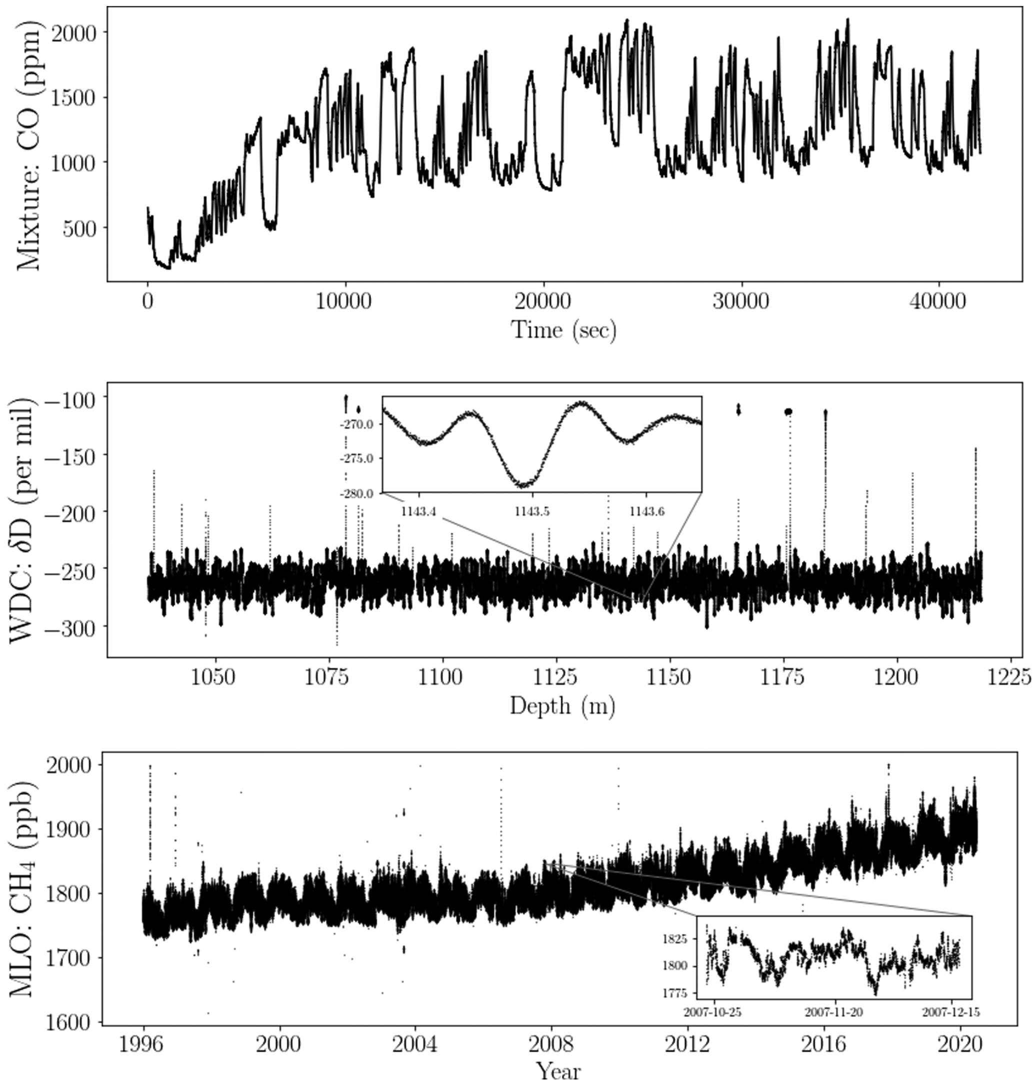

FIG. 4. Data. The three experimental time-series data sets used as examples in this paper: chemical sensor data from a gas-mixing experiment, water-isotope data from an Antarctic ice core (WDC), and atmospheric methane records from Mauna Loa (MLO). Insets are provided for WDC and MLO to see finer details of each time series, e.g., seasonal temperature variation in the WDC signal.

initial condition $[x_0, y_0, z_0] = [-13, -12, 52]$. The x coordinate of this trajectory is the baseline time series $\mathbf{x}$ for the set of tests reported in this section.

Using the process described in Sec. II D, we manipulate this high-resolution trajectory in order to replicate the effects of local mixing in real experiments. The details are as follows. At each point x_n in the original time series, we calculate the mean and variance of the points in a window of width seven, centered on that point (i.e., $[x_{n-3} \dots x_{n+3}]$) and then make a random draw from a normal distribution with those parameters to obtain the corresponding point of a new time series $\hat{x}_n$. We choose a mixing window of width seven for this first experiment because $k = 3$ gives us points from many distributions for each τ.

As discussed previously, this $\hat{\mathbf{x}}$ signal, whose PE is shown Fig. 2(b), is an *Ansatz* that is specifically designed to let us explore the effects of local mixing in a controlled experiment. As expected, the monotone-increasing vertical order that is apparent in the PE traces calculated from the original time series is reversed by the local-mixing operation. The corresponding value of the reversal metric $\overline{\mathcal{R}}$, computed across the span of the PE traces in Fig. 2(b), is 1, reflecting the perfect monotone-decreasing order of those traces with τ. (This would also be the case for $\overline{\mathcal{R}}$ computed across any *subsequence* of those traces, since the ordering is reversed everywhere: i.e., $\mathcal{R} = 1 \ \forall\ n$.) The PE values are higher than in Fig. 2(a). This simply reflects the additional randomness introduced by the local mixing operation. Careful examination shows that that effect decreases with τ: the $\tau = 6$ trace in Fig. 2(b), for instance, has roughly the same mean value as in Fig. 2(a), but the movement in the smaller-τ traces is comparatively larger. This is due to the mitigation of the mixing effects that occurs with larger τ values.

Our original conjecture was that one could effectively remove the added local mixing effects in $\hat{\mathbf{x}}$ by taking a binned average of that signal using the technique described in Sec. II D, with a bin size j roughly equal to the size of the mixing window used in the creation of that *Ansatz*. This conjecture can be easily tested using the *Ansatz* signal, as binning at that scale should "undo" the τ reversal produced by that

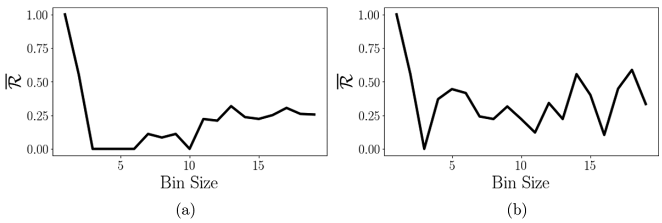

FIG. 5. Bin size and mixing scales. Plots of $\overline{\mathcal{R}}$ versus bin size for the (a) Lorenz and (b) Mackey-Glass examples.

synthetic mixing manipulation, thereby restoring the normal ordering.[3] We propose the following *heuristic* for choosing the value of j to remove the local-mixing effects: bin average the time series using varying bin sizes j, calculating $\overline{\mathcal{R}}$ across the span of the time series at each one, and choose the first bin size where $\overline{\mathcal{R}}$ goes to zero; if $\overline{\mathcal{R}}$ never reaches zero, choose the first minimum of this curve. Figure 5(a) shows such a plot for the Lorenz *Ansatz*.

As the bin size increases, the curves start reordering themselves, with $\overline{\mathcal{R}}$ reaching zero at $j = 3$. Beyond that, the ordering shifts because of the complicated relationships between the different timescales of the analysis (τ, ℓ, the mixing window, and the bin size), as well as other crucial dynamical timescales, such as the irrelevance and redundance timescales [20]. The PE traces in Fig. 2(c), computed from the Lorenz *Ansatz* $\hat{\mathbf{x}}$ binned with $j = 3$, are in clear monotone-increasing order with τ, consistent with the $\overline{\mathcal{R}}$ value of 0.

These results confirm that the binning process—a data-manipulation step used to eliminate the simulated effects of local mixing in the *Ansatz*—does indeed restore the normal ordering. The bin size indicated by our heuristic is smaller than the local mixing window that was used to create the *Ansatz*, so our original conjecture was off by a scale factor. This is discussed further below. Note, too, that the binning operation does not restore *the exact form* of the original traces in Fig. 2(a), just their ordering. The PE values in Fig. 2(c) are higher because of the residual effects of the added mixing. The difference in the smoothness of the traces in these two images is partially due to the altered window size—from the reduction in the number of points in the $\overline{\mathbf{x}}$ signal—as well as the difference in the scales of the plots.

This series of experiments validates the conjecture that, when local mixing occurs in a data set, high sampling rates can lead to reversal of the normal ordering of permutation entropy with changing τ. *This observation is a potentially useful way to detect the presence of local mixing in experimental data.* The bin size that restores the correct ordering is an indicator of the scale of the effect, and that size can be chosen from a curve like the one in Fig. 5(a). Our experiments show that this claim holds for different mixing windows, as well as for different w and ℓ values. In all cases, that bin size is smaller than, but on the same order as, the mixing window used in the creation of the *Ansatz*. The implications of this, for experimental practice, are that this operation *effectively identifies the order of magnitude of the mixing scale in the data, though not its specific value*.

These results are encouraging, but this is only one system, and a low-dimensional one at that. In the following section, we replicate these steps with a more complicated dynamical system.

B. Mackey-Glass system

The Mackey-Glass system [21] is described by a delay-differential equation of the following form:

$$\dot{x} = \beta \frac{x(t - t_0)}{1 + x(t - t_0)^q} - \gamma x, \quad \gamma, \beta, q > 0. \tag{6}$$

The delay here makes this system effectively infinite dimensional, even though there is only a single state variable. Like the Lorenz system, Mackey-Glass is known to exhibit deterministic chaos for some values of the parameters γ, β, t_0, and q. Using $\beta = 0.2$, $\gamma = 0.1$, $t_0 = 17$, and $q = 10$, we integrate Eq. (6) with the same Runge-Kutta solver for 1.5×10^6 steps from an initial state of $x = 1.2$ using a step size of 0.1 to obtain our test data $\mathbf{x}$.

The procedure for exploring τ reversal in this example is exactly the same as in the Lorenz system in the previous section. Figure 6 shows PE traces of the $\mathbf{x}$, $\hat{\mathbf{x}}$, and $\overline{\mathbf{x}}$ signals.

Figure 6(a) demonstrates normal, monotone-increasing order, with $\mathcal{R} = 0$ at all time points and a corresponding value of $\overline{\mathcal{R}} = 0$ computed across the span of the time series. For the locally mixed *Ansatz* $\hat{\mathbf{x}}$ in Fig. 6(b), which was created with a mixing window of nine, the order is reversed and $\overline{\mathcal{R}} = 1$. As before, bin averaging with a bin size $j = 3$ chosen at the first minimum of the $\overline{\mathcal{R}}$ versus j curve for this signal [Fig. 5(b)] completely restores the monotone-increasing order of the PE traces, as shown in Fig. 6(c). Again, the form of these results persists for different mixing windows, as well as for different values of the parameters of the PE calculations. As in the Lorenz examples, the bin size needed to fully restore the normal ordering is always smaller than, but on the same order as, the mixing window used in the creation of the *Ansatz*. In other words, this value provides a useful litmus test for the scale of that effect, even if it does not indicate the exact value.

These results reaffirm the overall relationship between local mixing effects and τ reversal. Next, we move to data from physical experiments.

C. Gas-mixture data

As a first real-world demonstration case for these techniques, we use a data set from the UCI Machine Learning Repository [22] that was produced during a chemistry experiment involving gas mixing. The values in this time series were measured by a sensor measuring carbon monoxide and ethylene mixtures in air under changing concentration conditions [23]. We focus on the ninth sensor in the array, a Firago TGS-2600 instrument that reports particle concentrations in parts per million. Calculations on the raw data from this sensor yielded PE $\approx 1\ \forall\ \tau$ across the full range of the time series: values that suggest that the signal is either all noise or completely obscured by noise. To distinguish between these

[3]For systems with delays, such as those modeled with delay-differential equations, a wider binning window may be required because of the associated temporal propagation of information due to the delay term.

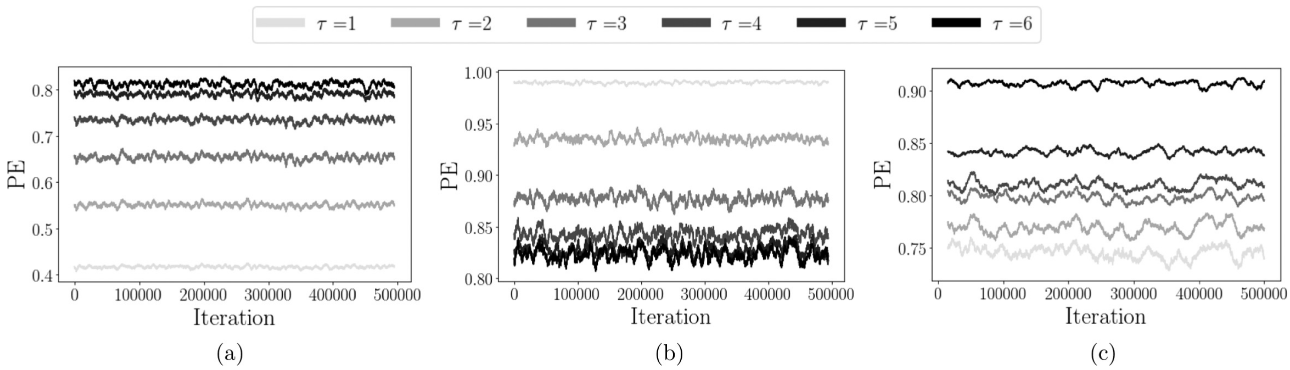

FIG. 6. Reversal results in the Mackey-Glass system. PE traces from the Mackey-Glass system with $1 \leqslant \tau \leqslant 6$ for (a) the original time series (b) with added local-mixing effects (c) after a bin-averaging operation was applied to remove those effects.

two cases, we performed a low-pass filtering step before proceeding with further analysis. The first plot in Fig. 4 shows the resulting time series, which is spaced at 0.25 s. PE traces calculated from this filtered time series demonstrate reversed τ ordering across the entire span of the data set with $\overline{\mathcal{R}} = 1$ [see Fig. 7(a)]. This reversed ordering, which was hidden by the observational noise in the raw signal, suggests that local mixing was at work during the collection of these data: in the chamber, perhaps, or the compartment in the sensor. To ascertain whether this was indeed the case, we perform a second bin-averaging step on the data, as in the synthetic examples in the previous sections, and observe how that operation affects the ordering of the PE curves. As in those previous examples, we can repeat this analysis using varying bin sizes in order to estimate the scale of the mixing effects.

In this example, the results of that procedure are quite clear, though the shape of the $\overline{\mathcal{R}}$ versus bin size curve is somewhat different than in our synthetic examples [see Fig. 8(a)]. As before, the PE traces move towards normal ordering as the bin size increases, but $\overline{\mathcal{R}}$ then reaches a broad minimum that is zero at $j = 15$ and extends to $j = 54$ (not shown).[4] Figure 7(b) shows that the PE traces for data bin averaged with $j = 15$ (i.e., 3.75 s) are in fully normal order. *These results not only confirm that local mixing is at work in the raw data, but also suggest a scale for those effects.* This, in turn, has implications for the resolution at which these data should be reported.

D. Antarctic ice-core data

Water isotopes in ice cores, which are viewed as climate proxies [24], can be measured at very high resolution in modern continuous flow analysis (CFA) equipment [25]. During this process, the ice is continuously melted and then piped through tubing into an optical spectrometer that measures the ratios of heavy to normal variants of oxygen and hydrogen isotopes in the ice. The CFA apparatus in the Stable Isotope Laboratory at the Institute for Arctic and Alpine Research at the University of Colorado can perform these measurements at submillimeter resolution. Basic fluid dynamics makes mixing inevitable in the different stages of this analysis pipeline (see [25] for discussion of the associated issues). Diffusion causes these isotopes to mix in the ice sheet as well, before the samples reach the laboratory, so there are at least two potential sources of mixing in these data, perhaps with different scales.

The West Antarctic Ice Sheet Divide Core (WDC) is 3300 m long, so a data set gathered from this core with submillimeter spacing contains many million points. Here, we focus on a short segment of that record: the ice from 1035.4 to 1368.2 m, which captures climate information from roughly 4500–6500 years ago [10,26]. The second trace in Fig. 4 shows the hydrogen isotope (δD) data from this core. These data are reported in delta (δ) notation relative to the baseline Vienna standard mean ocean water (VSMOW) and normalized to the standard light Antarctic water ($\delta\mathrm{D} = -428.0$) scale. The δ value is determined by $\delta = 1000(\rho_{\mathrm{sample}}/\rho_{\mathrm{VSMOW}} - 1)$, where ρ is the isotopic ratio D/H (i.e., $^2\mathrm{H}/^1\mathrm{H}$). Please see [9] for more details about these data.

As is customary in this field, the raw, high-resolution data (0.3 mm spacing, in this data set) undergo a number of postprocessing steps: removal of outliers, interpolation to fill gaps where data are missing, etc. The final step in the Stable Isotope Laboratory's postprocessing procedure involves a binned average of the data, with the bin scale explicitly chosen to reduce any mixing effects. In this case, the ice-core scientists chose a bin size of 0.5 cm (i.e., 17 points at a 0.3 mm spacing) to construct the published version of this data, which is available at [26].

This provides a unique opportunity for our work: using permutation entropy, we can not only assess whether local mixing was indeed at work in the raw data, but also potentially identify the associated scales, and even validate our results against expert judgment about those scales. To that end, we calculate PE and $\overline{\mathcal{R}}$ for the raw 0.3-mm spaced data. Similarly to the gas mixture data in Sec. III C, the higher-τ PE traces are below the lower-τ traces across the full span of the data set [see Fig. 7(c)]. The corresponding $\overline{\mathcal{R}} = 1$ value reflects this consistent reversal of the normal monotone-increasing order across the entire time series. As in the gas-mixture example, this suggests that local mixing effects are present in those data. The τ reversal does not identify the *mechanics* of the effect, or its source. In this example, there are many potential culprits: the laboratory apparatus and the physics of the ice sheet, among other things. See Sec. IV for more discussion.

[4]We do not have enough data to extend the calculation beyond $j = 54$.

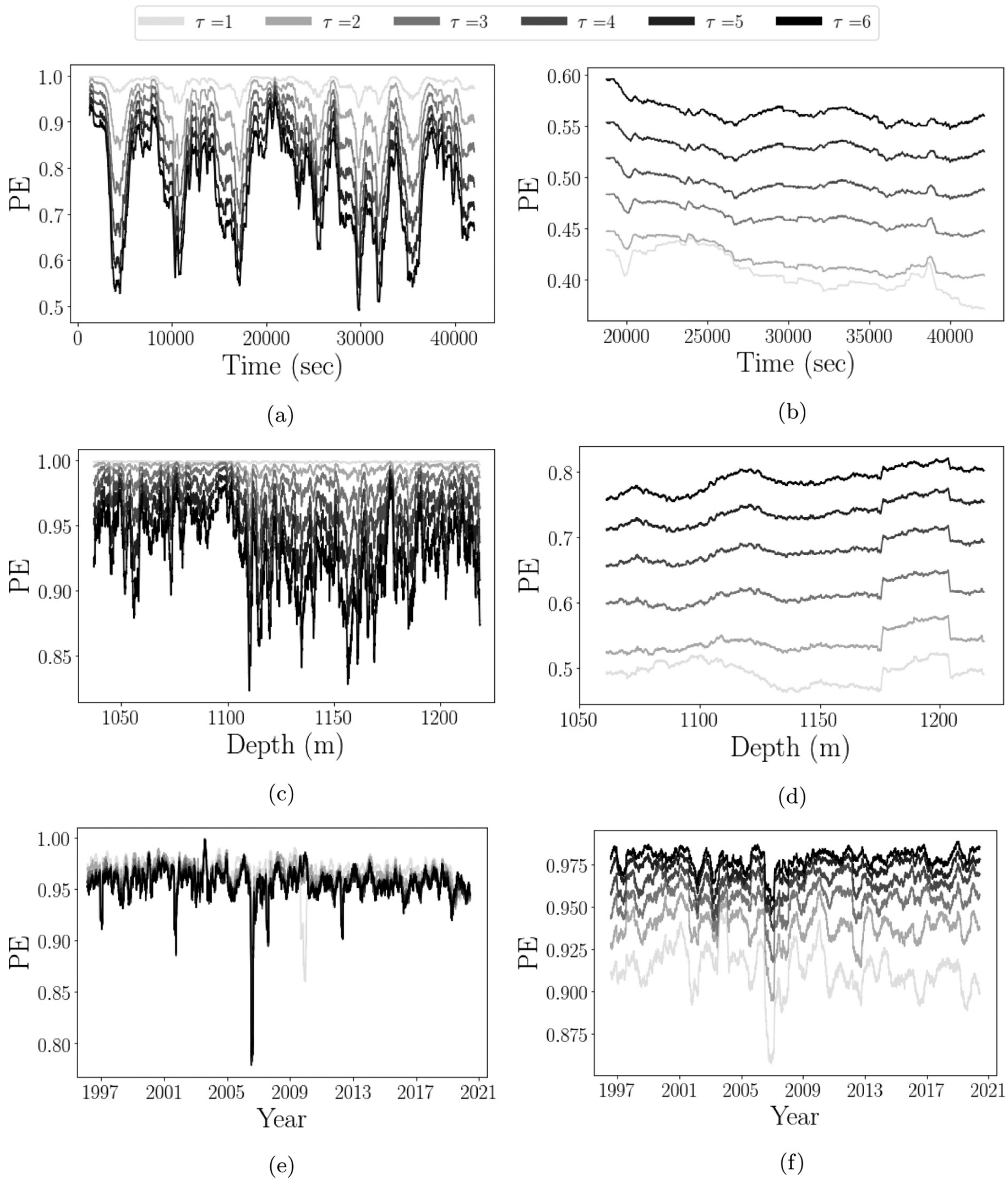

FIG. 7. Reversal results in experimental data sets. From top to bottom: PE traces for the gas mixture, WAIS Divide, and Mauna Loa data. In all cases, the images in the left and right columns show the calculations for the raw and bin-averaged data, respectively. The traces in the latter are temporally offset by the width of the binning window times the width of the PE calculation window. (This offset is essentially invisible in the Mauna Loa data because of the number of data points in the time series.) (a) Gas mixture: raw. (b) Gas mixture: binned with $j = 15$. (c) WDC: raw. (d) WDC: binned with $j = 15$. (e) Mauna Loa: raw. (f) Mauna Loa: binned with $j = 4$.

Note that we have moved to speaking of spatial scales and resolutions here (depths in the core) rather than temporal scales. Over this span of the WDC, there is a linear relationship between depth and age, with 0.5 cm of ice translating to 1/34th of a year of climate record, so these depthwise data can be viewed as time-series data. This would not be the case if we were using a longer span of the core; particularly one that extended to its deeper levels, where the nonlinearity of the age-depth relationship more strongly affects the temporal spacing of the data points.

To explore the mixing scales in this ice-core record, we again use bin averaging, computing $\overline{\mathcal{R}}$ from the raw 0.3-mm data, binned at a range of bin sizes. The shape of this curve, which is shown in Fig. 8(b), is similar to the one for the gas-mixture data, reaching a broad minimum at $j = 15$ with $\overline{\mathcal{R}} = 0$. That is, bin averaging the raw data using $j = 15$ restores the normal order of the PE traces; this is clearly visible in Fig. 7(d). As before, this is an indication of the scale of the mixing effects that are present in the data. In this case, where we have expert judgment as a point of comparison, that scale can be corroborated. Recall that the laboratory used a bin size of 17 to construct the data to be reported on the US Antarctic Program Data Center website, basing that choice on expert knowledge about the CFA apparatus. *In other words, our technique not only matches expert reasoning about the data, but does so in a manner that is completely model-free,*

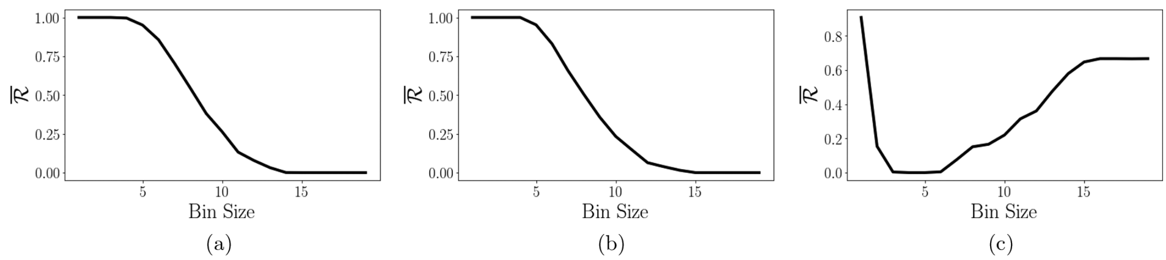

FIG. 8. $\overline{\mathcal{R}}$ as a function of bin size. (a) Gas mixture, (b) WAIS Divide, and (c) Mauna Loa methane.

requiring no knowledge about the system that produced the data or the apparatus that was used to measure it.

This analysis brings out the effects of *any* local mixing process, of course—not just laboratory effects. Since the isotopes mix via diffusion in the ice sheet, there are also potential scientific implications of these results, as discussed further in Sec. IV.

E. Mauna Loa methane data

NOAA's Global Monitoring Laboratory has been measuring atmospheric composition from Mauna Loa Observatory in Hawaii (altitude = 3397 m) at high temporal resolution for decades. Air is pumped continuously at $\approx$6 l/min from two inlets 40 m above the surface to a manifold inside the laboratory where $\approx$100 ml/min of the flows is fed to a calibrated analyzer that measures the abundance of atmospheric CH_4 at fixed intervals [27]. Over the period on which we focus in this paper, two well-calibrated analysis methods were used. In both cases, the analyzer alternates between the two inlets. From 1996 to 9 April 2019, measurements were by gas chromatography with flame ionization detection (GC/FID) at 15-min intervals, reported in parts per billion (ppb).[5] For each measurement, a 10-ml (at standard temperature and pressure) slug of air is injected into the GC/FID system every 15 min. The response of the detector is compared with bracketing measurements of standard to quantify CH_4. From 11 April 2019 to present, a cavity ring-down spectrometer (CRDS) was used. This instrument records nine 10-s averages of CH_4 every 5 min and averages them into a single "5-min average" value, also reported in ppb. It is calibrated with a suite of standards every 4 wk relative to a reference cylinder of natural air; this reference is measured once per hour during normal monitoring to track drift in the analyzer.

Because PE calculations require data with a constant temporal spacing, we first downsampled the 5-min segment of the data to produce an 826 633 point time series at an even 15-min spacing, beginning at 0007 on 1 January 1996 and ending at 2350 on 7 June 2020. We then scanned the data for missing or damaged values, replacing each of them with the last good value. The third plot in Fig. 4 shows this record beginning in March of 1996.

The ordering of the PE traces of this important data set, which are shown in Fig. 7(e), is more complex than in the previous two examples. The overall pattern is reversed, but not perfectly, with $\overline{\mathcal{R}} = 0.907$, rather than 1.0, as in all of the previous examples. In other words, local mixing effects are almost certainly in play here, but the indication is not as clear as in the gas-mixture and ice-core examples, where $\mathcal{R} = 1$ at every point in time (and $\overline{\mathcal{R}} = 1$ for the whole span). Moreover, nonstationarity is obviously an issue, viz., the downward spikes in PE values, especially at $\tau = 1$ (e.g., in 2004), for instance, and the short regions where the τ ordering of the traces changes (e.g., in 2010 and from 2019 on). Some of these issues can be traced back to the data-cleaning and quality-control processes; many of the dips, for instance, correspond to regions where a large fraction of the data are missing or damaged. Replacing those points with the last good value effectively creates a constant signal which, by definition, obeys normal τ ordering. This is a well-known challenge in the permutation entropy literature; one that causes problems for all data-cleaning strategies, including ours [4,28,29].

As before, we can repeat the $\overline{\mathcal{R}}$ calculation for a range of bin sizes in order to confirm the presence of local mixing and estimate the associated scale. In this case, as Fig. 8(c) shows, the first minimum of $\overline{\mathcal{R}}$ falls at $j = 4$. The $\overline{\mathcal{R}}$ value at this point is 0, indicating that a binning operation at a scale of 1 h eliminates the local mixing effects. This is reflected in the ordering of the resulting traces, shown in Fig. 7(f), which are in monotone-increasing order with τ. *Again, our results are in good alignment with standard practice in this field, which prescribes a 1-h reporting cadence.* As in the gas-mixture and ice-core examples in the previous section, this has implications for both analysis and reporting. Unlike those examples, though, the curve in Fig. 7(f) rises strongly for $j > 5$. This behavior, which more closely resembles the two synthetic examples in Fig. 5, is likely due to the timescales involved, in the systems and in the calculations, as well as to data limitations, which precluded a broader span on the horizontal axes of Figs. 7(d) and 7(e).

When one suspects nonstationarities, it can be informative to perform a more focused analysis; specifically, to compute $\overline{\mathcal{R}}$ over *windows* of the PE traces, rather than over their entire temporal extent. Figure 9 shows the results of this calculation for the raw and binned MLO methane data that uses a window size of 5000 points. This plot brings out the nonstationarities of the data very clearly. From 1996 through early 2019, $\overline{\mathcal{R}}$ oscillates between 1.0 and 0.7, with occasional drops to $\approx$0.6 and one even lower, in 2010. These results confirm that the mixing effects in this segment of the time series vary across the time series: the ordering of the curves is largely reversed across most of the span of the data, but the magnitude of that reversal varies with time. Many of these nonstationarities are due to the quality-control and data-cleaning processes, as described above. Between June 2009 and May 2010, for

[5] 10^{-9} mol CH_4 per mol dry air.

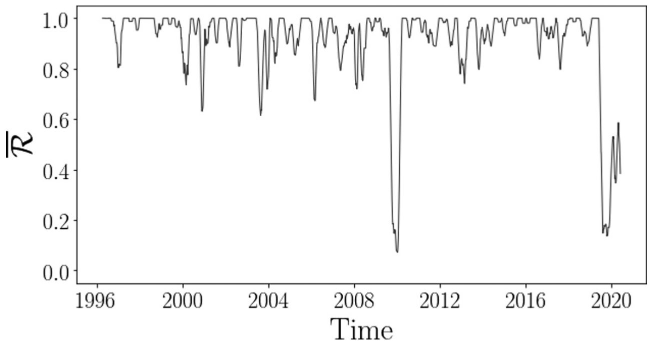

FIG. 9. Windowed $\overline{\mathcal{R}}$ analysis for the Mauna Loa data. Each point corresponds to the value of $\overline{\mathcal{R}}$ over the previous 5000 points.

instance, at the time of the obvious spike in the middle of the curve, 5066 of the 31 000 points were missing or damaged. Standard interpolation strategies for filling these gaps—a necessary step because PE calculations require data that are evenly spaced in time—employ smooth functions: lines, curves, etc. Data spaced in this manner are inherently predictable, so the PE values in the repaired spans will be low and the τ ordering will be monotone increasing, hence the drop in $\overline{\mathcal{R}}$. These effects are visible in PE traces themselves [Fig. 7(e) and the magnified views that appear in the Supplemental Material [30], which zoom in on the time spans around 2000 and 2009], but they really stand out in the windowed $\overline{\mathcal{R}}$ plot in Fig. 9. The same effects are also at work in many of the smaller dropouts in $\overline{\mathcal{R}}$, e.g., a temporary ten times increase in analyzer noise that occurred in June 2003, causing many of the points in this time span to be flagged during the quality-control process as suspicious, and therefore removed in the data-cleaning process.

One of the most interesting features of Fig. 9 is the precipitous drop in $\overline{\mathcal{R}}$ in the spring of 2019. Recall that the Observatory switched from a gas chromatograph to a cavity ring-down spectrometer at 0155 on 11 April 2019, at the same time as the shift in measurement cadence. This newer instrument has a factor of 5–10 less short-term noise than the gas chromatograph that it replaced. *Our metric detects this quite effectively from the raw data, not only flagging the timing of the change, but indicating clearly that the introduction of this new instrument significantly reduced the mixing effects in the data.* This provides additional corroboration of the efficacy of our technique, and also has implications for the scientific analysis, as described in the following section.

IV. DISCUSSION

The synthetic-data experiments in Secs. III A and III B, which use *Ansätze* that are expressly designed to mimic the effects of the process that we term local mixing, confirm our conjectures about τ reversal of PE curves as an indicator of local mixing in a data set. In both cases, the $\overline{\mathcal{R}}$ metric shows that the monotone-increasing order of the PE traces for the original data is reversed when information is artificially interchanged between neighboring data points. In both examples, a bin-averaging operation on these artificially mixed time series restores the normal order, as expected, since that operation compresses the distribution that is effectively created by that mixing down to its mean. The bin sizes required to accomplish this are on the same scale as, but roughly half the size of, the width of the mixing window that we use to create the *Ansätze*.[6] One can leverage this observation to estimate the mixing scale: specifically, when the $\overline{\mathcal{R}}$ value of the PE curve reaches zero.

Applied in tandem with bin averaging, this reversal metric has an even more powerful role in the examples in Secs. III C–III E, which involve data from real-world experiments. In all of these cases except a few spans of the Mauna Loa data, the τ-reversal metric suggests that the resolution of these data sets was actually high enough that the raw data are essentially repeated draws from a distribution composed from the data in a local window around each point. Permutation entropy brings this out quite effectively, if one varies the τ parameter in the calculation and evaluates the ordering of those results. Even though this operation may not reveal the exact distribution of the mixing effects, it has important practical implications, as it gives an indication of the scales of those effects, not just their presence or absence. This can be used, together with expert knowledge and other data-analytic techniques, to select an appropriate lower bound on the interval at which one should analyze, or report, the data.

Departures from pure monotonicity in the PE traces are a complex and interesting question. If $\overline{\mathcal{R}}$ does not reach zero for any bin size, for instance, one can choose the first minimum of a curve like the ones in Fig. 5 or 8 to perform the binning, but that is not guaranteed to remove all of the mixing effects, especially if nonstationarities are involved, causing the mixing scales to vary across the data set. Deconvolving multiple nonstationary mixing processes would call for strategies far more sophisticated than simple binning at a fixed resolution across the whole time series. A formal solution to this would be a major challenge, given the complex nature of the nonlinear statistics that underpin permutation-entropy analyses. Nonetheless, a windowed $\overline{\mathcal{R}}$ analysis like the one shown in Fig. 9 can help ascertain whether there are different regimes in the data, with different noise and mixing properties in each, as in the different segments of the Mauna Loa methane data, before and after the instrument change.

A major advantage of our technique is its model-free nature; it identifies the mixing scale regardless of the cause. This has disadvantages as well, though: if two or more mixing processes are at work, our analysis will bring out the one with the largest scale. If the scales are widely separated and one has very fine-grained data, it might be possible to extend this analysis to detect the different scales. This is a potential issue in the WDC data, where there are two known sources of mixing: diffusion of the isotopes in the ice sheet and fluid mixing in the laboratory analysis pipeline. Diffusion is a function of density, and density in an ice core is a function of depth, so the effects are not simple. Standard procedure in this laboratory is to use estimates of the accumulation rate and

[6] Calculations of the ratio of the standard deviation of the points in the mixing window to the absolute value of the difference between two adjacent points explain this to some extent, yielding values between 1.2 and 2.3 for the Lorenz and Mackey-Glass experiments with a range of integrator time steps.

temperature to make an educated guess at the diffusion scale at each depth. In constructing the published data set from the raw data, the analysts chose a sample rate that exceeded that resolution—in effect, intentionally oversampling the data—and then selected a bin size that spaced the reported data points safely above their estimates of the mixing scales. If those estimates are good, this procedure will obviously make the two mixing scales (diffusion in the ice sheet and fluid dynamics in the laboratory apparatus) very similar, so we do not expect to be able to deconvolve these two effects from this data set.

Similarly, there are a number of different potential sources of the mixing effects in the Mauna Loa data. While the sample lines are unlikely to be at issue because the flow through them can be approximated as plug flow, there is noise in the analyzer. There are also known sources of organized natural variability in the study system itself, which span the timescales from days to decades: (1) diurnal cycle related to local wind regime, (2) synoptic scale variability occurring on timescales of weather fronts (a few days), (3) seasonal cycle related to changing atmospheric loss process rate, and (4) long-term trend related to imbalance between emissions and losses. We did not find any correlation between $\overline{\mathcal{R}}$ and meteorological parameters, although Dlugokencky *et al.* [27] showed that, on average, the diurnal cycle in atmospheric CH_4 correlated with dew point, because both are impacted by daily cycling between up-slope and down-slope wind regimes. Of course, there could be convolved effects from multiple causes, but those would be difficult to study in isolation with a permutation-entropy analysis for the same reasons mentioned at the end of the previous paragraph.

The alignment between our results and scientific knowledge about the experimental situation is encouraging. The $\overline{\mathcal{R}}$ calculations also clearly bring out the timing, as well as the salutary effects, of the 2019 instrument change on Mauna Loa, for instance. In the ice-core data, the $\overline{\mathcal{R}}$ metric revealed that a 15-point averaging window restored the normal order of the PE traces, when the experts chose a 17-point window. Similarly, our method suggested a 1-h mixing window for the Mauna Loa data, which is exactly the cadence that is used in practice. In the data gathered by the newer instrument, the same strategy indicated a significant reduction in the mixing scales, corresponding to a 25-min cadence.

While corroboration of the known properties of a new instrument is not that exciting, there can be unknown mechanisms at work in one's data, leaving traces that are all but invisible to normal analysis techniques. Indeed, this was the driving force behind the creation of the WDC data set used for this work. After a permutation entropy analysis revealed noise introduced by the CRDS instrument that had been used to analyze this section of the ice core [10], the analysis was repeated using a newer instrument, removing the noise effects. Finally, the results of the analysis described in the previous sections not only corroborate the reasoning behind the approaches taken by the experts, but could also provide a more systematic way to figure out exactly how to squeeze as much information out of a study system as nature will allow. Of course, this does not mean that data measured below the mixing scale are useless; indeed, one must sample below a scale in order to identify its lower boundary.

There are many avenues for extension of this work. The metric used in the pointwise τ-reversal metric ($\mathcal{R}$) and the temporally aggregated nature of its average, $\overline{\mathcal{R}}$, have pros and cons. $\mathcal{R}$ is based solely on the $L1$ norm of the distance between the focal τ-sequence vector $\mathbf{v}$ and the purely monotone vector $\mathbf{v}_i$. While this is a sensible measure of an overall pattern of monotonicity, it does incorporate some bias: an interchange of the $\tau = 6$ and $\tau = 1$ PE traces, for instance, will affect the $\mathcal{R}$ value more strongly than an interchange of the $\tau = 3$ and $\tau = 4$ traces. Because $\overline{\mathcal{R}}$ is an average of $\mathcal{R}$ across a segment of the data set, it cannot capture temporal detail that occurs below the scale of the segment, such as regimes of reversed and normal ordering at different points in the data. Since many real-world data sets are nonstationary, it might be useful to develop a formalized strategy for a windowed $\overline{\mathcal{R}}$ analysis with varying window sizes. (The window size in Fig. 9 is tuned by hand to bring out the operative effects without excessive detail.) One could compute some statistics on the lengths of reversed and normal regimes from these results; even just the maxima and minima in $\overline{\mathcal{R}}$ might be informative. This could be particularly useful in ice cores, where large diffusion events are theorized to occur [31]. In cases like this, these τ-reversal metrics could provide insight not only into the instruments and sampling rates that are appropriate for different timescales, but also into the scientific questions of a field. Applications of the techniques described in this paper are by no means limited to purely temporal data, either, as evidenced by the WDC example. Since mixing is an important issue in spatial and spatiotemporal data sets, that is an additional advantage of the technique, and an avenue for further investigation.

V. CONCLUSION

In experimental practice, a delicate balance can arise between properly sampling the dynamics of a study system and oversampling them and thereby obfuscating the underlying signal with measurement-related effects. The Nyquist rate provides some guidance for this when the system is linear and its highest frequency is known, but those conditions are not always met in real-world situations. In this paper, we have described an information-theoretic technique for detecting one such oversampling effect, which arises when local mixing is present. Critically, this method requires no domain-specific knowledge about the data, or about the system that produced it. One simply calculates the permutation entropy across the time series at different temporal resolutions and examines the relationship between the resulting traces. For most time series, permutation entropy will increase with the stride of the calculation (the τ parameter in the previous sections). This simply reflects decreasing predictability over the longer time span sampled by each permutation. Reversal of that ordering suggests that local mixing has occurred.[7] The basic intuition here is straightforward: if successive data points are measured on a smaller scale than the mixing scales that are inherent in the data, each measurement is essentially a single draw from a local distribution composed of the data points in its

[7]In data from purely random process, of course, where temporal ordering is meaningless, τ order cannot bring out local mixing.

local neighborhood. By definition, this added randomness will raise the entropy rate. A reversal of the normal ordering of the traces, then, is an indication that mixing may be at work in the data. One can extend this reasoning to estimate the scale of the mixing effects by performing bin averaging of the signal for a range of bin sizes. (However, this does not extend to discontinuous data, such as iterates of maps, where a binned average does not have the appropriate smoothing effects.) Mixing is not the only data issue that $\overline{\mathcal{R}}$ can flag; we are currently investigating rounding effects, for instance, which can also change PE values and reshuffle the τ ordering.

Our method can not only be used by practitioners to detect mixing in their data; it can also help them choose appropriate lower bounds on the scales at which to perform the analysis, allowing them to squeeze as much information as possible out of their study system while avoiding spurious effects. Sampling frequency is often an imperfect balance between time, money, scientific need, and what the system of interest allows. Educated guesses are frequently our only means of making decisions about how often to sample. Considerable time and money could be saved if we had some meaningful feedback about where to draw the line in terms of sampling. Accuracy is also an issue here, since local mixing obfuscates data; in a situation like this, knowing the scales at which one should perform an analysis is of obvious value. Moreover, a method that identifies oversampling scales can be used to establish solid guidelines for data reporting. Our method does not only work on *time* series data, as is clear from our ice-core example, where the τ-reversal metric revealed mixing on spatial scales in the core that aligned with expert reasoning about diffusion in the ice sheet.

Permutation entropy and its variants (e.g., [32]) have become a staple in time-series analysis, especially in anomaly detection and forecasting. Some of that work has leveraged the τ parameter to focus on different temporal scales. To the best of our knowledge, though, no one has used the ordering of PE traces calculated across different τ values to understand the properties of the data.

The gas-mixture data is located in the UCI Machine Learning Repository [33]. The ice-core data is located in the U.S. Antarctic Program Data Center [34]. The hourly averages of the Mauna Loa data can be found at [35].

ACKNOWLEDGMENTS

This research was funded by US National Science Foundation (NSF) Grant No. 1807478. J.G. was also partially supported by an Applied Complexity Fellowship at the Santa Fe Institute.

M.N., E.B., and J.G. crafted the research questions. M.N., E.B., and J.G. designed the analyses. M.N. and J.G. conducted the analyses. M.N., E.B., J.W., E.D., and J.G. wrote the manuscript and contributed to discussion and idea development.

The authors declare that they have no competing financial interests.

[1] C. Bandt and B. Pompe, Permutation Entropy: A Natural Complexity Measure for Time Series, Phys. Rev. Lett. **88**, 174102 (2002).

[2] J. Amigó, *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That* (Springer Science & Business Media, New York, 2010).

[3] J. Garland, R. James, and E. Bradley, Model-free quantification of time-series predictability, Phys. Rev. E **90**, 052910 (2014).

[4] F. Pennekamp, A. C. Iles, J. Garland, G. Brennan, U. Brose, U. Gaedke, U. Jacob, P. Kratina, B. Matthews, S. Munch *et al.*, The intrinsic predictability of ecological time series and its potential to guide forecasting, Ecol. Monogr. **89**, e01359 (2019).

[5] S. V. Scarpino and G. Petri, On the predictability of infectious disease outbreaks, Nat. Commun. **10**, 1 (2019).

[6] A. Myers and F. A. Khasawneh, On the automatic parameter selection for permutation entropy, Chaos **30**, 033130 (2020).

[7] M. Riedl, A. Müller, and N. Wessel, Practical considerations of permutation entropy, Eur. Phys. J.: Spec. Top. **222**, 249 (2013).

[8] Y. Cao, W.-W. Tung, J. B. Gao, V. A. Protopopescu, and L. M. Hively, Detecting dynamical changes in time series using the permutation entropy, Phys. Rev. E **70**, 046217 (2004).

[9] J. Garland, T. R. Jones, M. Neuder, J. W. White, and E. Bradley, An information-theoretic approach to extracting climate signals from deep polar ice cores, Chaos **29**, 101105 (2019).

[10] J. Garland, T. R. Jones, M. Neuder, V. Morris, J. W. White, and E. Bradley, Anomaly detection in paleoclimate records using permutation entropy, Entropy **20**, 931 (2018).

[11] J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, Combinatorial detection of determinism in noisy time series, Europhys. Lett. **83**, 60005 (2008).

[12] J. Garland, T. R. Jones, E. Bradley, R. James, and J. W. White, A first step toward quantifying the climate's information production over the last 68,000 years, in *International Symposium on Intelligent Data Analysis* (Springer, New York, 2016), pp. 343–355.

[13] J. Garland and E. Bradley, Predicting computer performance dynamics, in *International Symposium on Intelligent Data Analysis* (Springer, New York, 2011), pp. 173–184.

[14] T. Myktowicz, A. Diwan, and E. Bradley, Computers are dynamical systems, Chaos **19**, 033124 (2009).

[15] E. Bradley, Damped-Driven Pendulum Dataset, https://bit.ly/2UAuADp.

[16] P. Diaconis and R. L. Graham, Spearman's footrule as a measure of disarray, J. R. Stat. Soc.: Ser. B (Method.) **39**, 262 (1977).

[17] M. G. Kendall, *Rank Correlation Methods* (Griffin, Spokane, 1948).

[18] H. Abdi, The Kendall rank correlation coefficient, *Encyclopedia of Measurement and Statistics* (Sage, Thousand Oaks, CA, 2007), p. 508.

[19] E. N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. **20**, 130 (1963).

[20] L. De Micco, J. G. Fernández, H. A. Larrondo, A. Plastino, and O. A. Rosso, Sampling period, statistical complexity, and chaotic attractors, Physica A **391**, 2564 (2012).

[21] M. Mackey and L. Glass, Oscillation and chaos in physiological control systems, Science **197**, 287 (1977).
[22] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml.
[23] J. Fonollosa, S. Sheik, R. Huerta, and S. Marco, Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring, Sens. Actuators, B **215**, 618 (2015).
[24] W. Dansgaard, Stable isotopes in precipitation, Tellus **16**, 436 (1964).
[25] T. R. Jones, J. W. White, E. J. Steig, B. H. Vaughn, V. Morris, V. Gkinis, B. R. Markle, and S. W. Schoenemann, Improved methodologies for continuous-flow analysis of stable water isotopes in ice cores, Atmos. Meas. Tech. **10**, 617 (2017).
[26] T. R. Jones, J. Garland, V. Morris, M. Neuder, E. Bradley, and J. W. C. White, Resampling of Deep Polar Ice Cores Using Information Theory, U.S. Antarctic Program (USAP) Data Center, doi: 10.15784/601274.
[27] E. J. Dlugokencky, L. P. Steele, P. M. Lang, and K. A. Masarie, Atmospheric methane at Mauna Loa and Barrow observatories: Presentation and analysis of in situ measurements, J. Geophys. Res.: Atmos. **100**, 23103 (1995).
[28] M. McCullough, K. Sakellariou, T. Stemler, and M. Small, Counting forbidden patterns in irregularly sampled time series. I. The effects of under-sampling, random depletion, and timing jitter, Chaos **26**, 123103 (2016).
[29] K. Sakellariou, M. McCullough, T. Stemler, and M. Small, Counting forbidden patterns in irregularly sampled time series. II. Reliability in the presence of highly irregular sampling, Chaos **26**, 123104 (2016).
[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.103.022217 for Fig. 1 which shows the raw and bin-averaged data for three experimental data sets that we analyze. Figure 2 shows two zoomed in views of the PE traces of the raw Mauna Loa data.
[31] T. R. Jones, K. M. Cuffey, J. W. C. White, E. J. Steig, C. Buizert, B. R. Markle, J. R. McConnell, and M. Sigl, Water isotope diffusion in the WAIS Divide ice core during the Holocene and last glacial, J. Geophys. Res.: Earth Surf. **122**, 290 (2017).
[32] B. Fadlallah, B. Chen, A. Keil, and J. Príncipe, Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information, Phys. Rev. E **87**, 022911 (2013).
[33] archive.ics.uci.edu/ml/machine-learning-databases/00322/.
[34] J. White *et al.*, “Stable Isotopes of Ice in the Transition and Glacial Sections of the WAIS Divide Deep Ice Core” U.S. Antarctic Program (USAP) Data Center, 2019, https://doi.org/10.15784/601274.
[35] ftp://aftp.cmdl.noaa.gov/data/trace_gases/ch4/in-situ/surface/.