

<sup>1</sup> **Using Curvature to Select the Time Lag for Delay Reconstruction**

<sup>2</sup> Varad Deshmukh,<sup>1</sup> Elizabeth Bradley,<sup>1,2</sup> Joshua Garland,<sup>2</sup> and James D. Meiss<sup>3</sup>

<sup>3</sup>)*Department of Computer Science, University of Colorado, Boulder, CO, USA*

<sup>5</sup>)*Santa Fe Institute, Santa Fe, NM, USA*

<sup>6</sup>)*Department of Applied Mathematics, University of Colorado, Boulder, CO, USA*

<sup>8</sup> (Dated: 11 April 2020)

We propose a curvature-based approach for choosing good values for the time-delay parameter  $\tau$  in delay reconstructions. The idea is based on the effects of the delay on the geometry of the reconstructions. If the delay is chosen too small, the reconstructed dynamics are flattened along the main diagonal of the embedding space; too-large delays, on the other hand, can overfold the dynamics. Calculating the curvature of a two-dimensional delay reconstruction is an effective way to identify these extremes and to find a middle ground between them: both the sharp reversals at the ends of an insufficiently unfolded reconstruction and the folds in an overfolded one create spikes in the curvature. We operationalize this observation by computing the mean over the Menger curvature of 2D reconstructions for different time delays. We show that the mean of these values gives an effective heuristic for choosing the time delay. In addition, we show that this curvature-based heuristic is useful even in cases where the customary approach, which uses average mutual information, fails—e.g., noisy or filtered data.

9 **Delay-coordinate reconstruction, the foundation of nonlinear time-series analysis, involves**  
 10 **two free parameters: the embedding dimension  $m$  and the delay  $\tau$ . A number of heuristic**  
 11 **methods are available for choosing good values for these parameters: notably, the false near**  
 12 **neighbor method of Kennel et al. for  $m$  and the average mutual information (AMI) of Fraser**  
 13 **& Swinney for  $\tau$ . The AMI approach selects a  $\tau$  that attempts to produce independent co-**  
 14 **ordinates in the reconstructed trajectories. Taking a geometric view of this problem, we**  
 15 **develop a curvature-based method for this task. By computing statistics on the curvature**  
 16 **of a 2D reconstruction, we can identify a delay that unfolds the dynamics without intro-**  
 17 **ducing overfolding: i.e., between the extremes that can cause the embedding to not be a**  
 18 **faithful representation of the full state-space dynamics. As in AMI, this involves identifying**  
 19 **the first minimum in a plot of the statistic versus  $\tau$ —something that is sometimes difficult**  
 20 **with AMI because the minima can be shallow or even nonexistent. Using a suite of exam-**  
 21 **ples, we demonstrate that the minima in curvature-based statistics are effective in producing**  
 22 **embeddings whose correlation dimensions match those of the true dynamics. The curvature**  
 23 **heuristic is quite robust in the face of data issues like noise, smoothing, and shorter samples**  
 24 **of the dynamics, and its minima are generally more distinct, which makes the choice easier.**

---

## 25 I. INTRODUCTION

26 Delay-coordinate embedding, or the method of delays, is a well-established technique for dy-  
 27 namical reconstruction of time-series data.<sup>1-3</sup> This method, which generates a reconstruction by  
 28 plotting scalar time-series data against delayed versions of itself on orthogonal axes, involves two  
 29 free parameters—a time delay,  $\tau$ , and an embedding dimension,  $m$ . Examples of reconstructions  
 30 for the classic Lorenz attractor are shown in Fig. 1. Though the embedding theorems offer some  
 31 theoretical guidance regarding the selection of  $\tau$  and  $m$ , one must fall back upon heuristics for  
 32 choosing their values when faced with finite-precision data from an unknown system.

33 Over the past decades, the nonlinear dynamics community has devoted significant effort to  
 34 developing effective methods for estimating  $m$  and  $\tau$ . This paper offers a contribution to that  
 35 arsenal: a new, geometry-based method for determining  $\tau$ . The challenge is this: if the delay  
 36 is too small, the reconstructed dynamics are flattened along the main diagonal of the embedding  
 37 space; see, e.g., Fig. 1(b). Delays that are too large, on the other hand, can overfold the dynamics,  
 38 as in Fig. 1(c). A common approach to solving this problem, due to Fraser & Swinney,<sup>4</sup> seeks a  
 39  $\tau$  that maximizes independence between  $\tau$ -separated points in the time series, thereby separating  
 40 the trajectories. As an alternative, we propose to use the *curvature* of trajectories to find a  $\tau$   
 41 that effectively unfolds the dynamics while avoiding overfolding. The idea is based upon the  
 42 observation that both sharp reversals like those in Fig. 1(b) and overfolds, as in in Fig. 1(c), can  
 43 create regions of large curvature in a 2D projection of the reconstructed dynamics.

45 We operationalize this observation by computing various statistics over the local curvature  
 46 of 2D reconstructions using a discrete curvature due to Menger: the curvature,  $c$ , of three non-  
 47 collinear points—say  $x$ ,  $y$ , and  $z$ —is the inverse of the radius of the unique circumcircle through  
 48 these points.<sup>5</sup> It can be seen that

$$c(x, y, z) = \frac{1}{r} = \frac{4A}{|x - y||y - z||z - x|}, \quad (1)$$

49 where  $A$  is the area of the triangle  $xyz$ . In our case, as illustrated in Fig. 2,  $x$ ,  $y$ , and  $z$  are three

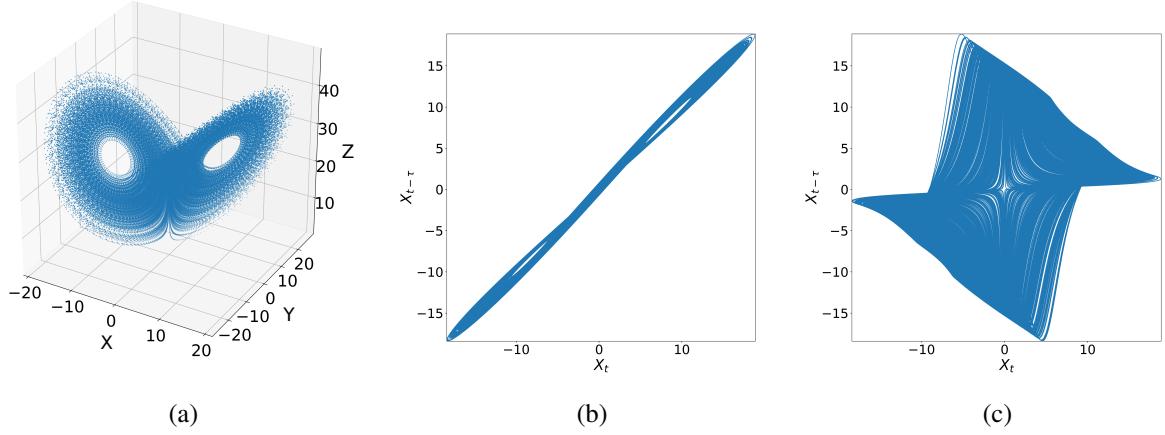


FIG. 1: The Lorenz attractor: (a) full state-space trajectory; 2D delay reconstructions from the  $x(t)$  time series with (b)  $\tau = 1$  and (c)  $\tau = 30$ .

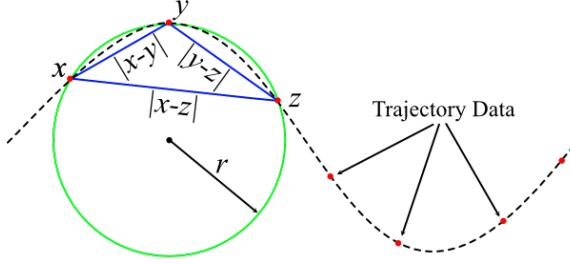


FIG. 2: The local curvature of a triplet of successive points selected from a trajectory of a dynamical system.

successive points along a reconstructed trajectory of a dynamical system, and we define

$$M_j(\tau) = c(\vec{x}_{j-1}, \vec{x}_j, \vec{x}_{j+1}). \quad (2)$$

Of course, when the trajectory is relatively straight, the local curvature will be small; if it has a sharp turn, the local curvature will be large.

The variations of the curvature along a trajectory can be visualized by computing a local average. For example, 2D reconstructions of data from the Lorenz attractor are shown in Fig. 3 for four different values of the time delay  $\tau$ . In each case, the color represents the local average of (1) over trajectory points that fall in a grid cell, using a uniform  $500 \times 500$  grid. These heat maps bring out the effects of  $\tau$  upon the local curvature. When  $\tau$  is small, the reconstructed trajectory has sharp turns at its local maxima and minima, which manifest as dark blue regions in Fig. 3(a). Spikes in the curvature can also arise from overfolding, when the time delay is large, as in Fig. 3(c). We seek a middle ground between these two extremes by choosing a time delay that minimizes the average curvature,  $\bar{M}$ , along a 2D reconstructed trajectory; as we will see, this corresponds to Fig. 3(b). Specifically, the proposed heuristic is that  $\tau$  be chosen to give the first minimum of  $\bar{M}(\tau)$ .

In §IV, we evaluate the effectiveness of this curvature heuristic for a suite of examples (quasiperiodic, Lorenz, and driven damped pendulum dynamics). We compare our heuristic to the method of Fraser & Swinney<sup>4</sup> that chooses the delay giving the first minimum on a plot of the average mutual information (AMI) versus  $\tau$ ; see §II. Since our aim is to obtain a dynamical reconstruction that is diffeomorphic to the full state-space dynamics of the system, we use a dynamical

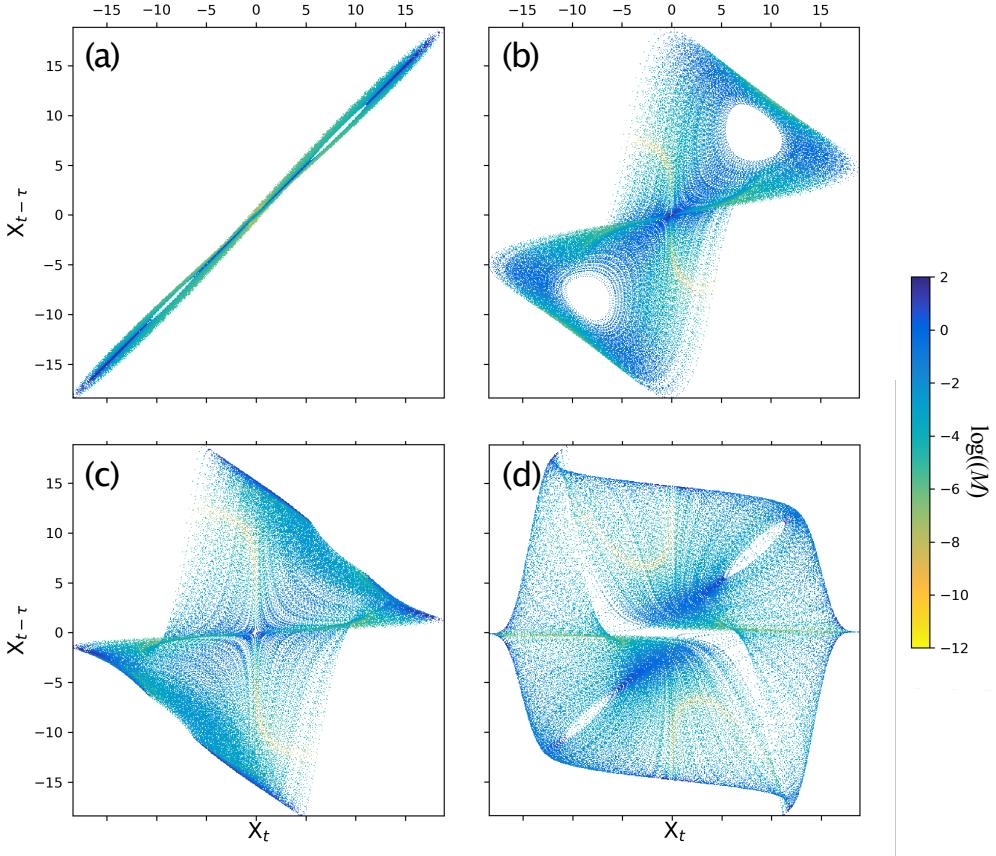


FIG. 3: Curvature heat maps of 2D delay reconstructions of the Lorenz attractor of Fig. 1: (a)  $\tau = 1$ , (b)  $\tau = 18$ , (c)  $\tau = 30$ , and (d)  $\tau = 60$ . The color represents the local value of the log of the curvature, (2). Panels (a) and (c) correspond to panels (b) and (c) of Fig. 1, respectively.

invariant—the correlation dimension,  $d_2$ —to compare the full and reconstructed dynamics. As described in more detail in §III, this involves computing  $d_2$  for embeddings constructed using the  $\tau$  values suggested by the curvature- and AMI-based heuristics and comparing these to the correlation dimension of the full state-space trajectory. The curvature-based heuristic matches the performance of AMI in many of our examples and outperforms it in others: e.g., when the data suffers from defects such as noise, limited trajectory length, or smoothing. Moreover, the first minima of  $\bar{M}(\tau)$  are generally deeper than those for AMI, making the choice of an effective  $\tau$  more clear.

## II. PARAMETER SELECTION IN DELAY RECONSTRUCTION

As mentioned above, success in delay reconstruction involves selection of values for two free parameters: the time delay,  $\tau$ , and the embedding dimension,  $m$ . A large number of useful heuristics have been proposed in the literature for estimating these, both separately and together.<sup>4,6–20</sup> In this paper, we focus on selecting  $\tau$ , even though the two parameters have interacting effects.<sup>81</sup> (Indeed, these interactions support an elegant method for selecting them at the same time.<sup>21</sup>)

Given a scalar time series  $\{x_t, t \in \mathbb{N}\}$ , an  $m$ -dimensional, time-delay reconstruction corresponds

<sup>83</sup> to the sequence of delay vectors

$$\vec{x}_j = [x_j, x_{j-\tau}, \dots, x_{j-(m-1)\tau}]^T \in \mathbb{R}^m \quad (3)$$

<sup>84</sup> for a given delay  $\tau$ . The time delay or “lag”  $\tau$  defines the number of steps separating each coordinate. The theoretical constraints on the time delay are far from stringent, requiring only  $\tau > 0$ .<sup>1,2</sup>  
<sup>85</sup> This only applies in the case of infinitely long, noise-free time series and infinite precision arithmetic, however—idealizations that are never realized in practice. As a result, the selection of  $\tau$   
<sup>86</sup> plays an important role in the practical application of this methodology.<sup>4,8–14</sup>

<sup>89</sup> The fact that the time delay does not play into the underlying mathematical framework is a  
<sup>90</sup> double-edged sword. Because the theoretical constraints are so loose, there is no practical way  
<sup>91</sup> to derive an “optimal” lag, or even know what criteria such a lag would satisfy.<sup>8</sup> Casdagli et al.<sup>22</sup>  
<sup>92</sup> provide a discussion of this theory and the impacts of  $\tau$  on reconstructing an attractor for a noisy  
<sup>93</sup> observation function. Unfortunately, their discussion gives no practical methods for estimating  
<sup>94</sup>  $\tau$ , even though it does nicely delineate a range of  $\tau$  between *redundancy* and *irrelevance*. For  
<sup>95</sup> very small  $\tau$ ,  $x_j$  and  $x_{j-\tau}$  are effectively indistinguishable. This is especially a problem in the  
<sup>96</sup> presence of noise and finite precision. In this situation, the reconstruction coordinates are highly  
<sup>97</sup> *redundant*: i.e., they contain nearly the same information about the system.<sup>6,22</sup> The implication  
<sup>98</sup> is that a *very* small  $\tau$  is not a good choice because the additional coordinates in (3) add almost  
<sup>99</sup> nothing new to the model. Choosing an arbitrarily *large*  $\tau$  is undesirable as well, because it makes  
<sup>100</sup> the coordinates of the reconstruction become causally unrelated. In such a case, the measurement  
<sup>101</sup> of  $x_{j-\tau}$  is *irrelevant* in predicting  $x_j$ .<sup>22</sup> Useful  $\tau$  values lie somewhere between these two extrema.

<sup>102</sup> In practice, finding  $\tau$  values in this middle ground can be quite challenging. The most com-  
<sup>103</sup> monly used method for this involves computing the time-delayed average mutual information or  
<sup>104</sup> AMI,  $I(x_j, x_{j-\tau})$ , for a range of  $\tau$ . Fraser & Swinney argue that selecting  $\tau$  to give the first min-  
<sup>105</sup> imum of AMI will minimize the redundancy of the embedding coordinates, thereby maximizing  
<sup>106</sup> the information content of the overall delay vector.<sup>4</sup> This standard method is not without prob-  
<sup>107</sup> lems. For some time series—e.g., processes with memory longer than  $\tau$ — $I(x_j, x_{j-\tau})$  does not  
<sup>108</sup> have a minimum. This occurs in any autoregressive process, for instance, and in real-world data as  
<sup>109</sup> well.<sup>23–25</sup> Even if a minimum exists, it can be shallow, requiring a subjective choice on the part of  
<sup>110</sup> the practitioner as to its location, or even its existence. Noise and other data issues, such as coarse  
<sup>111</sup> temporal resolution of the time series, can also affect the performance of any  $\tau$ -selection method,  
<sup>112</sup> including AMI.

### <sup>113</sup> III. METHODS

<sup>114</sup> In the absence of formal guidelines for selecting optimal parameters, standard practice in the  
<sup>115</sup> delay-reconstruction literature is to illustrate the usefulness of a parameter-selection method in  
<sup>116</sup> some specific context: often, its accuracy in estimating dynamical invariants,<sup>7,11,12,14</sup> in or maxi-  
<sup>117</sup> mizing forecast accuracy from a reconstruction.<sup>10,17</sup>

<sup>118</sup> In this paper, we show that our time-delay selection criterion allows for an accurate calculation  
<sup>119</sup> of the correlation dimension  $d_2$  of an attractor. We perform the calculations using the Grassberger-  
<sup>120</sup> Procaccia algorithm,<sup>26</sup> which approximates  $d_2$  by looking for the power law

$$C(\varepsilon) \sim \varepsilon^{d_2}, \quad (4)$$

<sup>121</sup> as the scale parameter  $\varepsilon \rightarrow 0$ . Here  $C(\varepsilon)$  is the correlation sum

$$C(\varepsilon) = \frac{1}{N(N-T)} \sum_{i=1}^N \sum_{j=1}^{i-T} \Theta[\varepsilon - ||\vec{x}_i - \vec{x}_j||], \quad (5)$$

<sup>122</sup> where  $N$  is the number of points in the trajectory,  $\Theta(x)$  is the Heaviside step function, and  $T$  is  
<sup>123</sup> the “Theiler window,” chosen to ensure that the temporal spacing is large enough to represent  
<sup>124</sup> an independently identically distributed sample. If (4) holds over some sufficiently large scaling  
<sup>125</sup> region on a log-log plot, then its slope estimates the correlation dimension. A fast algorithm to  
<sup>126</sup> compute the correlation sum is available in the TISEAN package.<sup>8,27</sup>

<sup>127</sup> There are several practical challenges that can affect the computation of  $d_2$ . These include  
<sup>128</sup> selection of the parameters  $N$  and  $T$ , as well as the range for  $\varepsilon$ . A persistence-based approach—i.e.,  
<sup>129</sup> finding a large range for each parameter that gives consistent values of the correlation dimension—  
<sup>130</sup> is perhaps the best way to make these choices.<sup>8,20</sup> To determine an appropriate scaling range for  
<sup>131</sup> (4), the standard practice is to require that the nearly linear relationship between  $\log(C(\varepsilon))$  and  
<sup>132</sup>  $\log(\varepsilon)$  exist over a *considerable* range of scales. This is necessarily subjective and makes the  
<sup>133</sup> process challenging, if not impossible, to automate. We discuss the specifics of our approach to  
<sup>134</sup> this problem in the Appendix.

## <sup>135</sup> IV. RESULTS

<sup>136</sup> In this section, we demonstrate the performance of our curvature-based heuristic using three  
<sup>137</sup> example systems: quasiperiodic motion on a two-torus, the classic Lorenz system, and a driven  
<sup>138</sup> damped pendulum. In each case, we compute a representative trajectory and choose one of the  
<sup>139</sup> state-space variables as the measurement function for the time-delay reconstruction. We then  
<sup>140</sup> compute the mean Menger curvature  $\bar{M}(\tau)$  of a 2D delay reconstruction of those data for a range  
<sup>141</sup> of values of  $\tau$ , using a C++ implementation of (1). We choose the time-delay,  $\tau_C$ , as the first  
<sup>142</sup> minimum of that curve. For the purposes of comparison, we also compute the average mutual  
<sup>143</sup> information profile  $I(x_t; x_{t-\tau})$  using the `mutual` command in the TISEAN package,<sup>8,27</sup> choosing  
<sup>144</sup>  $\tau_I$  at the first minimum of that curve. For the Lorenz and pendulum examples, we use TISEAN’s  
<sup>145</sup> `d2` to calculate the correlation dimension of the full state-space trajectory and compare it to the  
<sup>146</sup> correlation dimension of embeddings constructed with  $\tau_C$  and  $\tau_I$ . Details of these computations  
<sup>147</sup> are given in the Appendix, including the correlation sum plots and a discussion of the nuances  
<sup>148</sup> of choosing the scaling region. For the two-torus, we compare the calculated  $d_2$  value of the  
<sup>149</sup> embedding to the known dimension of the system. We also explore and discuss the effects of  
<sup>150</sup> noise, low-pass filtering, and data length using the Lorenz and pendulum systems.

### <sup>151</sup> A. Quasiperiodic Dynamics on a Two Torus

We begin our discussion with a simple quasiperiodic system—incommensurate rotation on a two-torus with the flow:

$$\begin{aligned} x &= (R + r \cos(2\pi t)) \sin(2\pi\phi t), \\ y &= (R + r \cos(2\pi t)) \cos(2\pi\phi t), \\ z &= r \sin(2\pi t). \end{aligned} \quad (6)$$

<sup>152</sup> Here, we use the radii  $R = 1$  and  $r = 0.2$  and frequency ratio  $\phi = \frac{1}{2}(1 + \sqrt{5})$ , the golden ratio.  
<sup>153</sup> For this example, we generate a time series of 200,000 points with a time spacing of 0.001. The  
<sup>154</sup> experimental time series,  $x(t)$ , is the first variable of (6) as a function of time.

<sup>155</sup> The mean curvature and AMI plots for 2D reconstructions of these data are shown in Fig. 4(a).  
<sup>156</sup> The first minima of the two profiles fall at the same value:  $\tau_I = \tau_C = 142$ . Fig. 4(b) shows

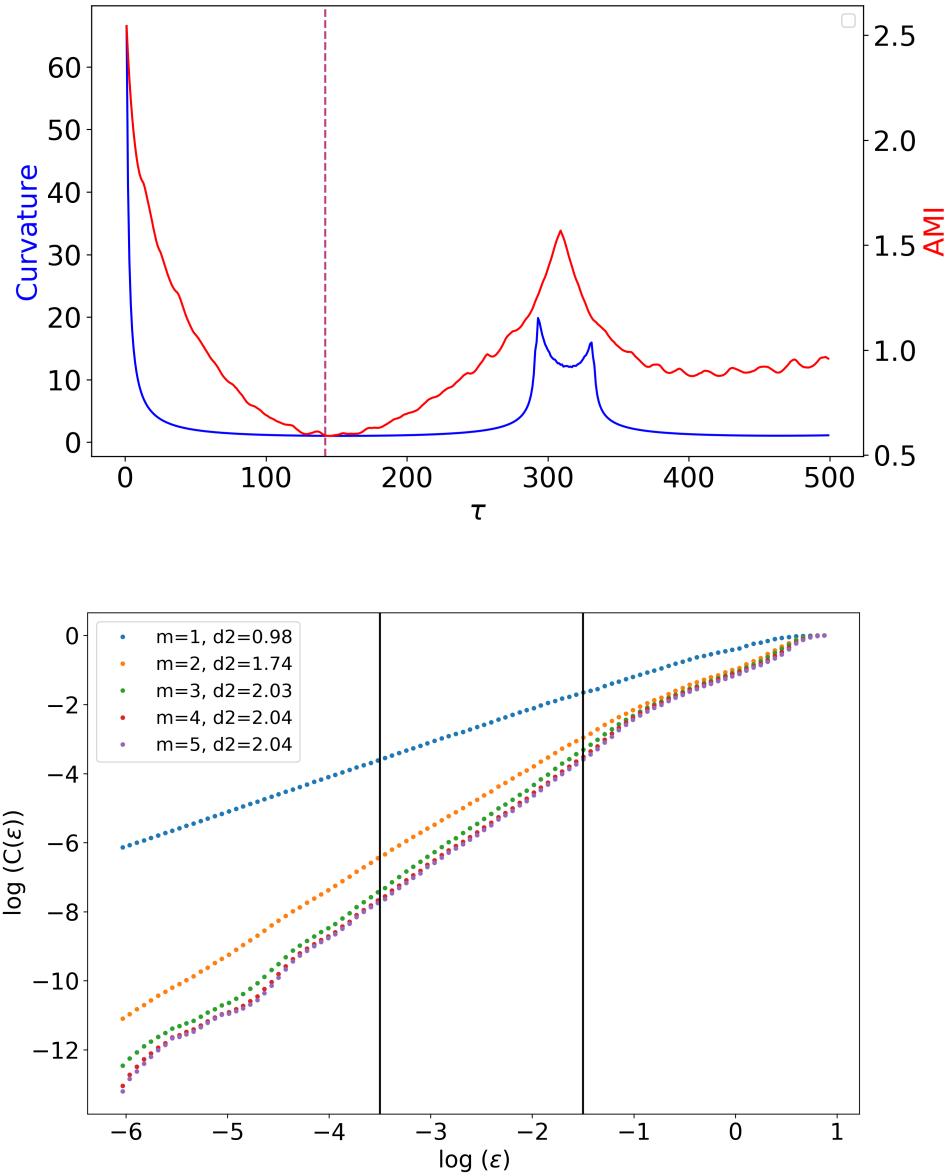


FIG. 4: Top: Mean curvature (blue) and average mutual information (red) profiles as a function of time delay, for 2D reconstructions of the two-torus dynamics from  $x(t)$ . Bottom: Correlation sums (5) for  $m$ -dimensional delay reconstructions with  $\tau = 142$ . The scaling region between the two vertical lines is used to compute  $d_2$ .

<sup>157</sup>

<sup>158</sup>

<sup>159</sup> the correlation sum plots for reconstructions constructed with this  $\tau$  value and  $m$  ranging up to  
<sup>160</sup> five, the value generically sufficient for embedding a two-dimensional invariant set, according to

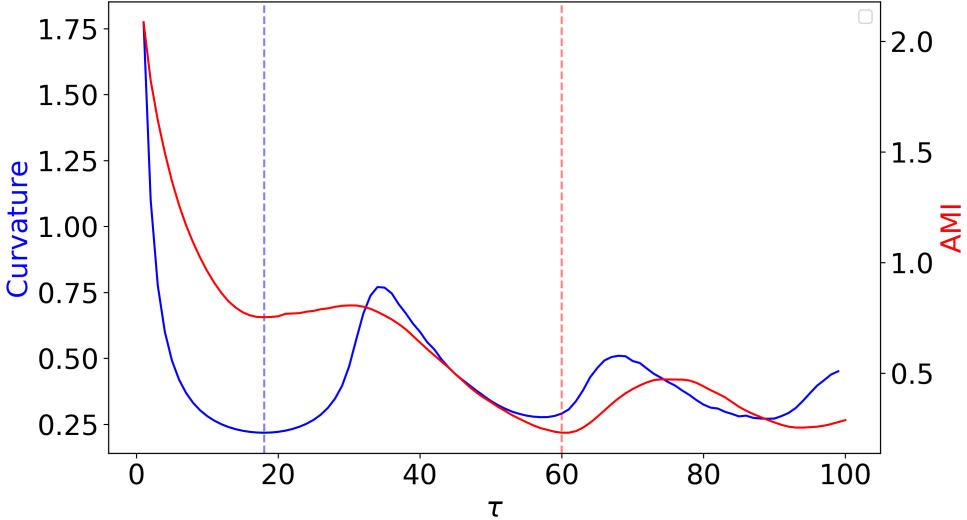


FIG. 5: Mean curvature and AMI profiles for the Lorenz data: the  $x$  coordinate of Fig. 1(a).

<sup>161</sup> Takens.<sup>1</sup> The correlation dimension converges to 2.04 for  $m \geq 4$ , close to the true dimension of  
<sup>162</sup> the two-torus. Interestingly, the profile of the mean curvature has a wide flat region around its first  
<sup>163</sup> minimum, suggesting that any value  $\tau_C \in [100, 200]$  could be appropriate. Indeed, the correlation  
<sup>164</sup> dimension is unchanged for time delays across this interval. It is often advantageous, for a variety  
<sup>165</sup> of reasons,<sup>7,10,17,22</sup> to choose the time delay as small as possible to maintain the accuracy of the  
<sup>166</sup> reconstructed dynamics. The fact that the curvature suggests a wide range for  $\tau$  would allow a  
<sup>167</sup> practitioner to exploit this advantage by selecting a smaller  $\tau$ ; by contrast, the AMI curve would  
<sup>168</sup> constrain that choice to a single, larger value.

## <sup>169</sup> B. Classic Lorenz Attractor

As a second example, we consider the canonical Lorenz system,<sup>28</sup>

$$\begin{aligned} \dot{x} &= \sigma(y - x), \\ \dot{y} &= x(\rho - z) - y, \\ \dot{z} &= xy - \beta z, \end{aligned} \tag{7}$$

<sup>170</sup> for the standard parameters  $\sigma = 10$ ,  $\beta = \frac{8}{3}$ , and  $\rho = 28$ . We use a fourth-order Runge-Kutta  
<sup>171</sup> method with a fixed time step of 0.01 and integrate the system to  $t = 1000$ . Discarding the initial  
<sup>172</sup> transient  $t < 100$ , we obtain a time series of length 90,000. As before, we take  $x(t)$  to be the  
<sup>173</sup> measurement function. Fig. 5 shows the mean curvature and average mutual information plots for  
<sup>174</sup> these data. The first minima of both curves fall near  $\tau = 18$ . Note, though, that the minimum in  
<sup>175</sup> the curvature profile appears significantly deeper. Indeed, based on this AMI profile, it is quite  
<sup>176</sup> likely that a practitioner would actually choose  $\tau_l = 60$ .

<sup>177</sup> To formalize the notion of depth of a minimum, we define

$$\Delta = (H_{\max} - H_{\min})/H_{\max}, \tag{8}$$

where  $H_{min}$  is the height of the curve at that minimum and  $H_{max}$  the height at the subsequent (here, higher  $\tau$ ) maximum. For the first minimum in the AMI profile,  $\Delta = 0.074$ ; for  $\bar{M}$ ,  $\Delta = 0.72$ . The deeper minimum of the  $\bar{M}$  curve is a significant advantage in this regard, as it makes the choice far more clear. Nevertheless, both values of  $\tau$  yield correlation dimensions that are good matches to the correlation dimension of the full 3D trajectory ( $d_2 \approx 2.05$ ): for  $\tau_C$ , we find  $d_2 = 2.04$ , and for  $\tau_I$ ,  $d_2 = 2.07$ .<sup>a</sup>

Even though, in this particular case, the two embeddings give nearly the same dimensions, a comparison of panels (b) and (d) of Fig. 3 suggests other issues that are potential problems. In particular, the smaller  $\tau$  produces a far less geometrically complicated reconstruction. The folds and kinks in the  $\tau = 60$  embedding—Fig. 3(d)—could be problematic if noise were present. This is precisely where we turn our attention next.

### C. Noisy Lorenz Dynamics

To explore the effects of noise on the choice of time delay, we use the same Lorenz trajectory as in §IV B but add *iid* noise to each point in the trajectory time series with a uniform distribution of amplitude  $0.1x_{max} = 1.922$ , i.e., a noise-to-signal ratio of 10%. For a fair comparison of the reconstruction to the full, 3D trajectory, it is important to add noise to each component. The correlation dimension calculated from the full, noisy trajectory is  $d_2 \approx 2.30$ ,<sup>b</sup> slightly larger than for the noise-free case, as one would expect.

The AMI and curvature profiles for 2D reconstructions of these noisy data are shown in Fig. 6. As in the noise-free case, the AMI profile has a comparatively shallow first minimum ( $\Delta = 0.066$ ), now at  $\tau = 20$ , and a more well-defined minimum ( $\Delta = 0.59$ ) at  $\tau = 60$ . This suggests the choice  $\tau_I = 60$ . Note that the first minimum of  $\bar{M}$ , at  $\tau_C = 21$ , is still well-defined ( $\Delta = 0.41$ ). The corresponding correlation dimensions<sup>c</sup> are  $d_2 = 2.28$  for  $\tau_C$  and  $3.83$  for  $\tau_I$ —the latter far higher than the correct value. These results suggest that the curvature-based heuristic works well in the face of noise, perhaps because the smaller  $\tau$  that it finds produces less overfitting of the reconstructed trajectory. To explore this further, we varied the noise-to-signal ratio from 0.001 to 0.1, and found that  $\tau_C$  remains relatively steady in the range [18, 21], producing a dimension consistent with the full trajectory. By contrast, the value of  $d_2$  for the AMI reconstruction steadily diverges from the correct value as the noise grows. This adds to our confidence about the robustness of the curvature-based heuristic with respect to noise.

### D. Driven Damped Pendulum

In this section, we consider a pendulum with natural frequency  $\nu_0$  that is subject to linear damping and a time-periodic force:

$$\begin{aligned}\dot{\theta} &= \omega, \\ \dot{\omega} &= -\beta\omega - \nu_0^2 \sin(\theta) + A \cos(\alpha t),\end{aligned}\tag{9}$$

The coordinates of the 3D, extended phase space  $\mathbb{S} \times \mathbb{R} \times \mathbb{S}$  are angle  $\theta$ , angular momentum  $\omega$ , and time  $t$ . To fix parameters, we choose  $\nu_0^2 = 98$ , damping  $\beta = 2.5$ , and a driving force with

<sup>a</sup> More details are in the Appendix and Fig. 10.

<sup>b</sup> The scaling regions for the correlation dimension calculations need to be chosen carefully in such situations to factor in the noise levels; see the Appendix for more discussion.

<sup>c</sup> See Fig. 11 in the Appendix.

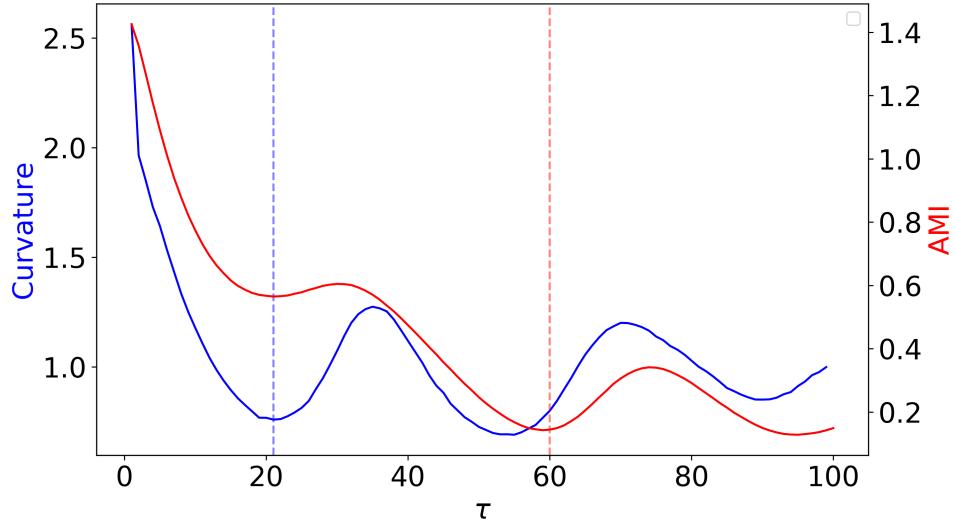


FIG. 6: Mean curvature and AMI profiles for the noisy Lorenz data. Compare to the curves without noise in Fig. 5.

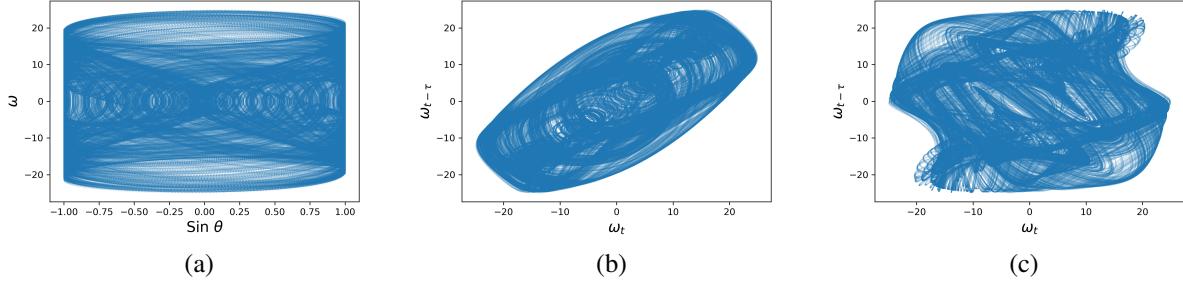


FIG. 7: The driven damped pendulum: (a) a projection of the full state space onto  $(\sin \theta, \omega)$ ; delay reconstructions of  $\omega$  for (b)  $\tau = 120$ ; (c)  $\tau = 250$ .

amplitude  $A = 91$  and frequency  $\alpha = 0.75v_0$ . This system has a chaotic attractor. As in the Lorenz example, we solve the system (9) using fourth-order Runge-Kutta, now with time step of 0.001. Discarding the first  $10^5$  points to eliminate transient behavior, we keep a time series of  $10^6$  points. To avoid issues with periodicity in  $\theta$  and  $t$ , we project the time series onto the three variables  $\{\sin(\theta(t)), \omega(t), \sin(\alpha t)\}$ , as seen in Fig. 7(a). For the time-delay reconstruction experiments, we take  $\omega$  as the measurement function. Profiles of the mean curvature and AMI of this signal are shown in Fig. 8. The former has a broad plateau in the range  $50 \lesssim \tau \lesssim 120$ , which again provides the flexibility to choose the lowest possible  $\tau$  that successfully reconstructs the dynamics. The AMI has a first minimum at  $\tau_l = 250$ . For this  $\tau$  value, as well as for values across the range  $50 \lesssim \tau \lesssim 120$ , the calculated correlation dimension  $d_2 = 2.22$ , which matches the correlation dimension of the full state-space dynamics. In this case, the curvature heuristic appears to match the performance of AMI, perhaps because  $d_2$  is relatively insensitive to the choice of  $\tau$ , as in the Lorenz example. However, there are significant geometrical differences between the resulting 2D reconstructions; see Fig. 7(b)-(c). The larger  $\tau_l$  produces an overfolded reconstruction, increasing

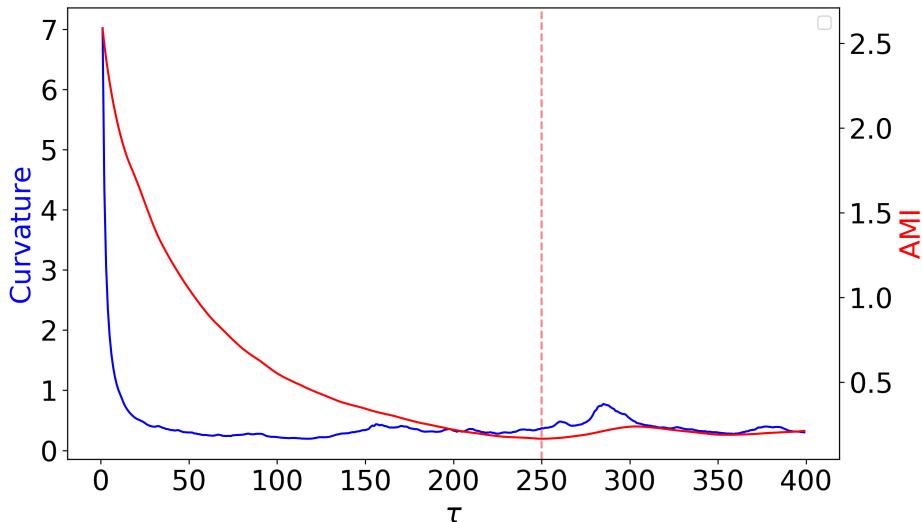


FIG. 8: Mean curvature and AMI profiles for the driven damped pendulum using  $\omega$  as the measurement function.

the curvature along the trajectory. This can, as demonstrated in the previous example, increase noise sensitivity.

### E. Other Data Effects

Data limits—shorter traces or coarser temporal sampling—are another practical issue in delay reconstruction. To explore the effects of data length on the curvature-based heuristic, we repeat the pendulum experiment with a shorter trajectory, keeping only the first 200,000 points, one fifth of the previous time series. The resulting AMI and mean-curvature profiles (not shown) are essentially identical to those in Fig. 8, and the correlation dimension for the  $\tau_C$  reconstruction ( $d_2 = 2.20$ ) is close to that of the full dynamics ( $d_2 = 2.18$ ); however, for a reconstruction using  $\tau_I$ ,  $d_2$  increases to 2.35.<sup>d</sup> This suggests that the effects of the larger  $\tau$  can be more significant for smaller data sets. Indeed, if there are fewer points, those that are artificially close, due to overfitting, will have a larger effect on the correlation function.

Another issue that arises in the practice of nonlinear time-series analysis is data smoothing. This often occurs during the measurement process<sup>25</sup> or in data-processing pipelines,<sup>23,29,30</sup> but can also occur naturally through, for example, diffusive processes.<sup>31</sup> To explore the potential effects of this upon the different  $\tau$ -selection heuristics, we ran the Lorenz time series (§IV B) through a moving average filter on a data window  $[j - 60, j + 59]$  for the  $j$ th point:

$$\tilde{x}_j = \frac{1}{120} \sum_{i=-60}^{i=59} x_{j+i}.$$

The mean curvature and AMI profiles for the smoothed time series  $\tilde{x}$  are shown in Fig. 9. The first clear minimum for the curvature moves slightly, to  $\tau_C = 16$ , with a depth  $\Delta = 0.65$ . The AMI

<sup>d</sup> See Figs. 13 and 14 in the Appendix for the calculations.

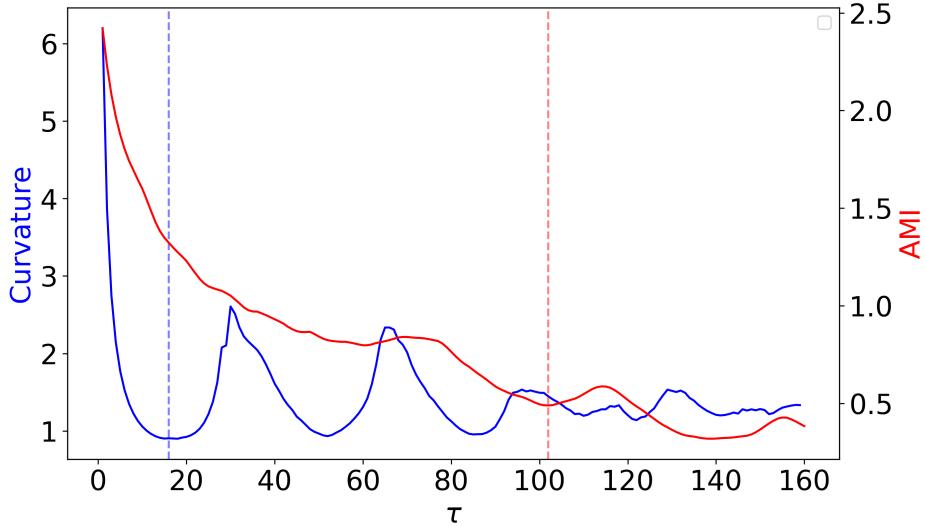


FIG. 9: Mean curvature and AMI profiles for the filtered Lorenz data. Compare to Fig. 5.

profile, though, is significantly distorted from that seen in the non-filtered data of Fig. 5: there is now no minimum at  $\tau = 18$  and only a weak one at  $\tau = 60$  ( $\Delta = 0.05$ ). This leaves the minimum at  $\tau = 102$  ( $\Delta = 0.17$ ) as the reasonable choice for  $\tau_I$ .

This deformation makes sense: a moving-average filter decreases the independence of neighboring points, thereby increasing the AMI and obscuring any minima that occur at lower  $\tau$  values. As in the previous case of limited data, the larger  $\tau_I$  creates problems with the dimension calculation:  $d_2$  for the  $\tau_I$  embedding is 1.89—a significant under-estimate. In contrast,  $d_2 = 2.11$  for the  $\tau_C$  embedding, which is still close to the value  $d_2 = 2.05$  of the full, filtered dynamics.<sup>e</sup> This is an encouraging result; obtaining a reconstruction that matches the underlying dynamics, even though the data are smoothed, is quite useful.

Over- and under-sampling—*i.e.*, temporal spacing between data points that is far smaller or far larger than the time scales of the dynamics—are also issues in the practice of nonlinear time-series analysis. Under-sampled data are, of course, a challenge to any method and there is very little recourse in that situation, as one should not make up data. Over-sampling causes different problems. For the curvature heuristic, points spaced too closely along the trajectory will be nearly co-linear, which can cause numerical issues in (1). AMI, in contrast, is relatively immune to this problem; oversampling simply moves its first minimum to a higher  $\tau$  value. One can determine whether over-sampling is an issue using standard best practices: *viz.*, repeating the analysis with every  $n^{th}$  point and observing whether the results change.<sup>8,20</sup>

## 269 V. CONCLUSIONS AND FUTURE WORK

The curvature heuristic that we have proposed is the following: choose a time-delay  $\tau$  to be the value giving the first local minimum of the average of the local Menger curvature, (2), along a segment of a 2D delay reconstruction of a scalar time series  $\{x_j\}$ . The set of experiments presented in the previous section, which are summarized in Table I, suggest that this heuristic is useful for

<sup>e</sup> Details presented in Fig. 12 of the Appendix.

<sup>274</sup> selecting an appropriate embedding delay. In particular, in every case, the  $\tau$  value suggested by  
<sup>275</sup> that heuristic gives a correlation dimension from a time-delay embedding (*i.e.*, a reconstruction  
<sup>276</sup> with a sufficiently large dimension  $m$ ) that agrees with the dimension of the full attractor—within  
<sup>277</sup> reasonable error bounds.

		Full Dynamics	Curvature	AMI
Two-Torus	True dynamics	2.0	2.04	2.04
	$\tau$	-	142	142
Lorenz	$d_2$	2.05	2.04	2.07
	$\tau$	-	18	60
Lorenz + Noise	$d_2$	2.30	2.28	3.83
	$\tau$	-	21	60
Pendulum (long)	$d_2$	2.22	2.22	2.21
	$\tau$	-	120	250
Pendulum (short)	$d_2$	2.18	2.20	2.35
	$\tau$	-	120	250
Lorenz + Filtering	$d_2$	2.05	2.11	1.89
	$\tau$	-	16	102

TABLE I: The dimensions for various systems and their 2D delay reconstructions using the time delay chosen by mean curvature and AMI heuristics.

<sup>278</sup> Just like the conventional AMI heuristic, which is based on averaging the mutual information,  
<sup>279</sup> this minimum can be computed before one knows the correct embedding dimension. The cur-  
<sup>280</sup> vature of a trajectory in a 2D reconstruction is a geometrical signal that can be used to diagnose  
<sup>281</sup> problems with the choice of time delay  $\tau$ . For example, when  $\tau$  is too small, the reconstructed  
<sup>282</sup> trajectory lies near the diagonal—it is not properly unfolded. This leads to sharp reversals in the  
<sup>283</sup> 2D reconstruction near local maxima of the time series, which locally increases their curvature.  
<sup>284</sup> Similarly, when  $\tau$  is too large, a 2D delay reconstruction tends to overfold: such folds again result  
<sup>285</sup> in larger curvature on parts of that reconstruction.

<sup>286</sup> The curvature heuristic appears, from our examples, to provide some advantages regarding the  
<sup>287</sup> choice of minimum: sometimes, a more-distinct first minimum than that exhibited by the AMI;  
<sup>288</sup> in other cases, a wider first minimum that allows useful flexibility in the choice of  $\tau$ . This is  
<sup>289</sup> important, as ambiguities in identifying the first minimum of AMI can lead to overfolded recon-  
<sup>290</sup> structions. We have also shown that the curvature heuristic is robust to noise, while the AMI  
<sup>291</sup> heuristic can be less so. The curvature heuristic also appears to be less influenced by shorter data  
<sup>292</sup> lengths and low-pass filtering. In practice, limited data, as well as data that has been aggregated in  
<sup>293</sup> some way, are quite common; in these cases, our heuristic may prove to be useful.

<sup>294</sup> In the future, we plan to study in more detail the full distribution of the Menger curvature along  
<sup>295</sup> trajectory sequences. From the point of view of Frenet, a smooth trajectory defines a curve in phase  
<sup>296</sup> space that generically has an associated orthonormal frame and a set of generalized curvatures.<sup>f</sup>  
<sup>297</sup> It is possible that additional information can be gleaned from the full statistics of one or more  
<sup>298</sup> of these curvatures—information that would lead to a better criterion than the average. It would  
<sup>299</sup> also be useful to evaluate the effectiveness of the curvature in selecting a time delay for other  
<sup>300</sup> purposes, such as in nonlinear forecasting. Akin to Pecora *et al.*,<sup>21</sup> one could perhaps leverage  
<sup>301</sup> curvature-based metrics to select  $m$  and  $\tau$  simultaneously.

<sup>f</sup> For example, in 3D the second curvature is the torsion.

<sup>302</sup> **ACKNOWLEDGMENTS**

<sup>303</sup> The authors acknowledge support from National Science Foundation Grant CMMI-1537460.  
<sup>304</sup> In addition, JDM was partially supported by NSF grant DMS-1812481. JG was supported by an  
<sup>305</sup> Omidyar and Applied Complexity Fellowship at the Santa Fe Institute. The authors would like to  
<sup>306</sup> thank Holger Kantz and Erik Boltt for numerous useful discussions.

<sup>307</sup> **DATA AVAILABILITY**

<sup>308</sup> The data that supports the findings of this study are available within the article.

<sup>309</sup> **REFERENCES**

- <sup>310</sup> <sup>1</sup>F. Takens, “Detecting strange attractors in fluid turbulence,” in *Dynamical systems and turbulence* (Springer, Berlin, 1981) pp. 366–381.
- <sup>312</sup> <sup>2</sup>T. Sauer, J. Yorke, and M. Casdagli, “Embedology,” *J. Statistical Physics* **65**, 579–616 (1991), <https://doi.org/10.1007/BF01053745>.
- <sup>314</sup> <sup>3</sup>N. Packard, J. P. Crutchfield, J. Farmer, and R. Shaw, “Geometry from a time series,” *Phys. Rev. Lett.* **45**, 712–716 (1980), <https://doi.org/10.1103/PhysRevLett.45.712>.
- <sup>316</sup> <sup>4</sup>A. Fraser and H. Swinney, “Independent coordinates for strange attractors from mutual information,” *Phys. Rev. A* **33**, 1134–1140 (1986), <https://doi.org/10.1103/PhysRevA.33.1134>.
- <sup>318</sup> <sup>5</sup>L. Blumenthal and K. Menger, *Studies in Geometry* (W. H. Freeman and Company, San Francisco, 1970) see p. 319.
- <sup>320</sup> <sup>6</sup>J. Gibson, J. Farmer, M. Casdagli, and S. Eubank, “An analytic approach to practical state space reconstruction,” *Physica D* **57**, 1–30 (1992), [https://doi.org/10.1016/0167-2789\(92\)90085-2](https://doi.org/10.1016/0167-2789(92)90085-2).
- <sup>323</sup> <sup>7</sup>E. Olbrich and H. Kantz, “Inferring chaotic dynamics from time-series: On which length scale determinism becomes visible,” *Phys. Lett. A* **232**, 63–69 (1997), [https://doi.org/10.1016/S0375-9601\(97\)00351-4](https://doi.org/10.1016/S0375-9601(97)00351-4).
- <sup>326</sup> <sup>8</sup>H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997) <https://doi.org/10.1017/CBO9780511755798>.
- <sup>328</sup> <sup>9</sup>T. Buzug and G. Pfister, “Comparison of algorithms calculating optimal embedding parameters for delay time coordinates,” *Physica D* **58**, 127–137 (1992), [https://doi.org/10.1016/0167-2789\(92\)90104-U](https://doi.org/10.1016/0167-2789(92)90104-U).
- <sup>331</sup> <sup>10</sup>J. Garland, R. G. James, and E. Bradley, “Leveraging information storage to select forecast-optimal parameters for delay-coordinate reconstructions,” *Phys. Rev. E* **93**, 022221 (2016), <https://doi.org/10.1103/PhysRevE.93.022221>.
- <sup>334</sup> <sup>11</sup>W. Liebert, K. Pawelzik, and H. Schuster, “Optimal embeddings of chaotic attractors from topological considerations.” *Europhysics Letters* **14**, 521–526 (1991), <https://doi.org/10.1209/0295-5075/14/6/004>.
- <sup>337</sup> <sup>12</sup>T. Buzug and G. Pfister, “Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behavior of strange attractors,” *Phys. Rev. A* **45**, 7073–7084 (1992), <https://doi.org/10.1103/PhysRevA.45.7073>.
- <sup>340</sup> <sup>13</sup>W. Liebert and H. Schuster, “Proper choice of the time delay for the analysis of chaotic time series,” *Physics Letters A* **142**, 107–111 (1989), [https://doi.org/10.1016/0375-9601\(89\)90169-2](https://doi.org/10.1016/0375-9601(89)90169-2).

- <sup>343</sup><sup>14</sup>M. Rosenstein, J. Collins, and C. De Luca, “Reconstruction expansion as a geometry-based  
344 framework for choosing proper delay times,” *Physica D* **73**, 82–98 (1994), [https://doi.org/10.1016/0167-2789\(94\)90226-7](https://doi.org/10.1016/0167-2789(94)90226-7).
- <sup>346</sup><sup>15</sup>L. Cao, “Practical method for determining the minimum embedding dimension of a scalar  
347 time series,” *Physica D* **110**, 43–50 (1997), [https://doi.org/10.1016/S0167-2789\(97\)00118-8](https://doi.org/10.1016/S0167-2789(97)00118-8).
- <sup>349</sup><sup>16</sup>D. Kugiumtzis, “State space reconstruction parameters in the analysis of chaotic time series—the  
350 role of the time window length,” *Physica D* **95**, 13–28 (1996), [https://doi.org/10.1016/0167-2789\(96\)00054-1](https://doi.org/10.1016/0167-2789(96)00054-1).
- <sup>352</sup><sup>17</sup>J. Garland and E. Bradley, “Prediction in projection,” *Chaos* **25**, 123108 (2015), <https://doi.org/10.1063/1.4936242>.
- <sup>354</sup><sup>18</sup>M. Kennel, R. Brown, and H. Abarbanel, “Determining minimum embedding dimension using a  
355 geometrical construction,” *Phys. Rev. A* **45**, 3403–3411 (1992), <https://doi.org/10.1103/PhysRevA.45.3403>.
- <sup>357</sup><sup>19</sup>R. Hegger, H. Kantz, and T. Schreiber, “Practical implementation of nonlinear time series  
358 methods: The TISEAN package.” *Chaos* **9**, 413–435 (1999), <https://doi.org/10.1063/1.166424>.
- <sup>360</sup><sup>20</sup>E. Bradley and H. Kantz, “Nonlinear time-series analysis revisited,” *Chaos* **25**, 097610 (2015),  
<https://doi.org/10.1063/1.4917289>.
- <sup>362</sup><sup>21</sup>L. Pecora, L. Moniz, J. Nichols, and T. Carroll, “A unified approach to attractor reconstruction,”  
*Chaos* **17**, 013110 (2007), <https://doi.org/10.1063/1.2430294>.
- <sup>364</sup><sup>22</sup>M. Casdagli, S. Eubank, J. Farmer, and J. Gibson, “State space reconstruction in the presence of  
365 noise,” *Physica D* **51**, 52–98 (1991), [https://doi.org/10.1016/0167-2789\(91\)90222-U](https://doi.org/10.1016/0167-2789(91)90222-U).
- <sup>366</sup><sup>23</sup>T. Myklowicz, A. Diwan, and E. Bradley, “Computers are dynamical systems,” *Chaos* **19**,  
367 033124 (2009), <https://doi.org/10.1063/1.3187791>.
- <sup>368</sup><sup>24</sup>J. Garland and E. Bradley, “On the importance of nonlinear modeling in computer performance  
369 prediction,” in *Advances in Intelligent Data Analysis XII* (Springer Lecture Notes in Computer  
370 Science, 2013) [https://doi.org/10.1007/978-3-642-41398-8\\_19](https://doi.org/10.1007/978-3-642-41398-8_19).
- <sup>371</sup><sup>25</sup>T. R. Jones, W. H. Roberts, E. J. Steig, K. Cuffey, B. Markle, and J. White, “Southern hemisphere  
372 climate variability forced by northern hemisphere ice-sheet topography,” *Nature* **554**,  
373 351–355 (2018), <https://doi.org/10.1038/nature24669>.
- <sup>374</sup><sup>26</sup>P. Grassberger and I. Procaccia, “Measuring the strangeness of strange attractors,” *Physica D* **9**,  
375 189–208 (1983), [https://doi.org/10.1016/0167-2789\(83\)90298-1](https://doi.org/10.1016/0167-2789(83)90298-1).
- <sup>376</sup><sup>27</sup>R. Hegger, H. Kantz, and T. Schreiber, “Tisean 3.0.1 nonlinear time series analysis,” (2007),  
[https://www.pks.mpg.de/~tisean/Tisean\\_3.0.1/index.html](https://www.pks.mpg.de/~tisean/Tisean_3.0.1/index.html).
- <sup>378</sup><sup>28</sup>E. Lorenz, “Deterministic nonperiodic flow,” *J. the Atmospheric Sciences* **20**, 130–141 (1963),  
379 [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- <sup>380</sup><sup>29</sup>F. Pennekamp, A. C. Iles, J. Garland, G. Brennan, U. Brose, U. Gaedke, U. Jacob, P. Kratina,  
381 B. Matthews, S. Munch, M. Novak, G. M. Palamara, B. C. Rall, B. Rosenbaum, A. Tabi,  
382 C. Ward, R. Williams, H. Ye, and O. L. Petchey, “The intrinsic predictability of ecological  
383 time series and its potential to guide forecasting,” *Ecological Monographs* **89**, e01359 (2019),  
384 <https://doi.org/10.1002/ecm.1359>.
- <sup>385</sup><sup>30</sup>J. Garland and E. Bradley, “Predicting computer performance dynamics,” in *Advances in In-*  
*386 telligent Data Analysis X*, Vol. 7014 (Springer Lecture Notes in Computer Science, 2011)  
387 [https://doi.org/10.1007/978-3-642-24800-9\\_18](https://doi.org/10.1007/978-3-642-24800-9_18).
- <sup>388</sup><sup>31</sup>T. Jones, K. Cuffey, J. White, E. Steig, C. Buizert, B. Markle, J. McConnell, and M. Sigl,  
389 “Water isotope diffusion in the wais divide ice core during the holocene and last glacial,” *Journal*

<sup>390</sup> of Geophysical Research: Earth Surface **122**, 290–309 (2017), <https://doi.org/10.1002/2016JF003938>.

<sup>392</sup><sup>32</sup>D. Viswanath, “The fractal property of the Lorenz attractor,” Physica D **190**, 115–128 (2004),  
<sup>393</sup> <https://doi.org/10.1016/j.physd.2003.10.006>.

<sup>394</sup><sup>33</sup>V. Martínez, R. Domínguez-Tenreiro, and L. Roy, “Hausdorff dimension from the minimal  
<sup>395</sup> spanning tree,” Phys. Rev. E **4**, 735–738 (1993), <https://doi.org/10.1103/PhysRevE.47.735>.  
<sup>396</sup>

<sup>397</sup> APPENDIX: CORRELATION DIMENSION CALCULATIONS

<sup>398</sup> Determination of the correlation dimension from the correlation sum plots produced by  
<sup>399</sup> TISEAN’s d2 tool requires care in choosing the scaling region, a significant straight segment  
<sup>400</sup> on the log-log plot. We make an initial choice by hand, then use the python polyfit tool to fit a  
<sup>401</sup> line to that segment, and iteratively expand or contract its width so as to minimize the fit error. In  
<sup>402</sup> the plots in this Appendix, the chosen regions are delineated with vertical lines.

<sup>403</sup> For validation purposes, we compare the correlation dimensions of the true and reconstructed  
<sup>404</sup> dynamics. This requires that one choose the dimension  $m$  of the reconstruction, and a too-small  
<sup>405</sup> choice for that value will cause the d2 results to be incorrect. Faced with data from an unknown  
<sup>406</sup> system, standard practice entails using a heuristic like the false near neighbor method<sup>18</sup> to esti-  
<sup>407</sup> mate  $m$ , then validating that choice by repeating the  $d_2$  calculation over a range of  $m$  values and  
<sup>408</sup> looking for convergence. For the examples in this paper, where we know the equations, that is  
<sup>409</sup> unnecessary; we can perform these comparisons with the full trajectories and correct embeddings:  
<sup>410</sup> i.e., reconstructions with  $m = 2d + 1$ , where  $d$  is the state-space dimension of the full system (this  
<sup>411</sup> is generically a sufficient condition for an embedding<sup>1,2</sup>). In the plots in this Appendix, we re-  
<sup>412</sup> peat the d2 calculation for reconstructions over the range  $1 \leq m \leq 2d + 1$ , where  $d$  is the (known)  
<sup>413</sup> state-space dimension of the system, in order to understand how the d2 results change, but we only  
<sup>414</sup> report the value for  $m = 2d + 1$  as the correct correlation dimension.

<sup>415</sup> Error estimates for  $d_2$  are notoriously problematic, since these complex algorithms have many  
<sup>416</sup> free parameters and subjective choices about interpretation of the plots.<sup>8</sup> In our examples, the  
<sup>417</sup> computed values of  $d_2$  are relatively insensitive to the choice of the scaling interval, provided the  
<sup>418</sup> initial hand-selected choice falls within the linear region. Upon varying the interval by 20%, we  
<sup>419</sup> typically find that the changes in dimension are about 0.2%

<sup>420</sup> The Theiler window,  $T$  in (5), is another important free parameter in d2 computations. Recall  
<sup>421</sup> that the correlation sum estimates the number of neighboring points present in an  $\varepsilon$ -ball around  
<sup>422</sup> sample points on the attractor. By default, these also include the points located along the imme-  
<sup>423</sup> diate trajectory of the attractor at the sample point: that is, its immediate temporal neighbors, in  
<sup>424</sup> forward and backwards time. The  $d_2$  estimate will be biased if the points in the  $\varepsilon$ -ball consist  
<sup>425</sup> primarily of these neighbors; indeed, if all of the points in the ball are immediate temporal neigh-  
<sup>426</sup> bors, the correlation dimension algorithm will return  $d_2 = 1$ . The Theiler window addresses this  
<sup>427</sup> issue by defining a set of temporal neighbors around each sample point that will be ignored in the  
<sup>428</sup> correlation sum.

<sup>429</sup> We choose the Theiler window so as to exclude points along the immediate trajectory segment  
<sup>430</sup> with a span given by the maximum of the scaling region used in the calculation. This is a circular  
<sup>431</sup> problem, since the choice of the Theiler window affects the scaling region and its limits. We solve  
<sup>432</sup> this iteratively, choosing a small Theiler window as a starting estimate to get a first approximation  
<sup>433</sup> of the correlation sum plots with a visible scaling region, setting the scaling region limits, updat-  
<sup>434</sup> ing the Theiler window, as stated, to the upper limit, recomputing the correlation sum plots, and  
<sup>435</sup> repeating until the results stabilize.

<sup>436</sup> The correlation plots for the Lorenz data of §IV B are shown in Fig. 10. The dimension cal-  
<sup>437</sup> culation yields an estimate of  $d_2 = 2.05$  over the scaling region  $\log(\varepsilon) \in [-1.5, 1]$  with a Theiler  
<sup>438</sup> window of 271 data points. It is accepted that the Hausdorff dimension of the standard Lorenz  
<sup>439</sup> attractor is between  $d_H = 2.06$  and  $2.16$ .<sup>32,33</sup> On the other hand, it is also known that  $d_2 \leq d_H$ , and  
<sup>440</sup> this dimension has been estimated to be  $d_2 = 2.05 \pm 0.01$ .<sup>26</sup> Our estimates are in line with these  
<sup>442</sup> results.

<sup>443</sup> For the noisy Lorenz data of §IV C, the correlation sum plots for the delay reconstructions—

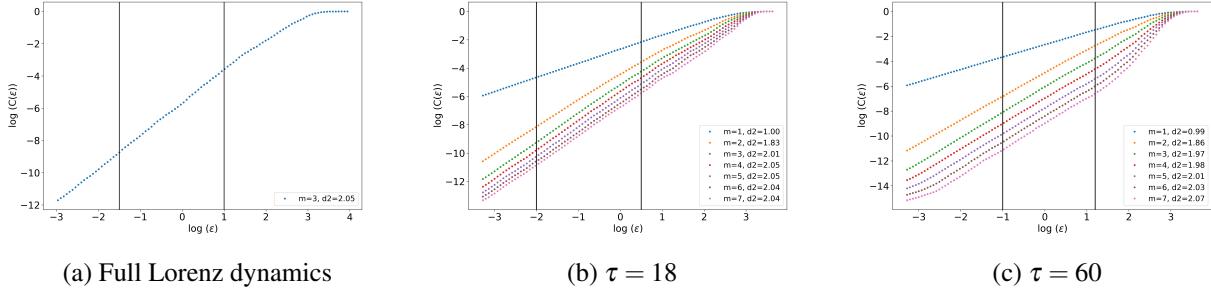


FIG. 10: Log-log plots of  $C(\epsilon)$  and computed correlation dimension of the Lorenz data set for the full dynamics, giving  $d_2 = 2.05$ , and using  $\tau = 18$ —the value suggested by the mean curvature heuristic—and  $\tau = 60$ , the first clear minimum of AMI for that data set.

Fig. 11(b) and (c)—have a more-interesting shape: there is a second scaling region for small  $\epsilon$  that arises from the added noise.<sup>g</sup> Indeed, when there is uniformly distributed *iid* noise with a maximum size  $\epsilon_{max}$ , then points within a radius  $\epsilon_{max}$  will be dominated by the noise and thus tend to fill out a ball of full dimension in the embedding space. Thus for reconstructed data we expect a computed  $d_2 \approx m$  for any scaling domain with  $\epsilon < \epsilon_{max}$ . For Fig. 11,  $\epsilon_{max} = 1.912$ , so  $\log(\epsilon_{max}) = 0.65$ . In the figures, this threshold corresponds to a knee in the curves below which their slopes are roughly  $m$ . This confirms that the lower scaling region is due to the noise. The correlation sum plot for  $\tau_C = 21$  has a relatively broad scaling region with a slope of  $d_2 = 2.28$  for  $m = 7$ , but for  $\tau_I = 60$ , the scaling region above the noise threshold is narrower. This region gives  $d_2 = 3.83$ , a significant over-estimate of the correlation dimension of this system.

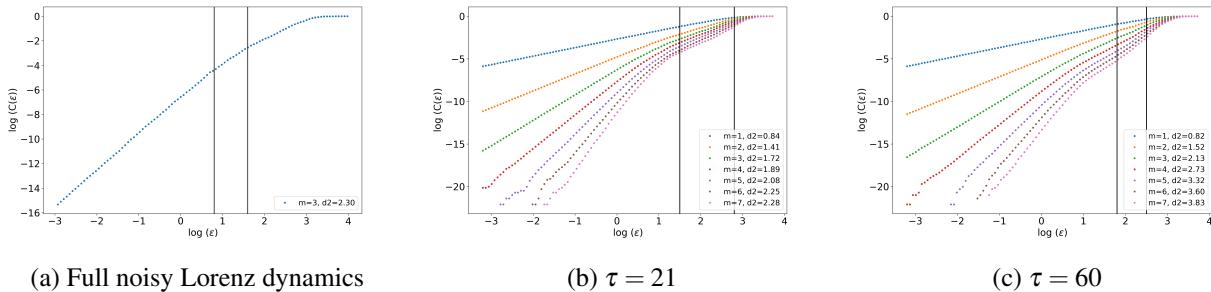


FIG. 11: Log-log plots of  $C(\epsilon)$  and calculation of the correlation dimension of the noisy Lorenz data set for the full dynamics, and using  $\tau = 21$ —the value suggested by the curvature—and  $\tau = 60$ , the second, clearer minimum of AMI.

In the third variant of the Lorenz experiments, discussed in § IV E, the goal is to determine whether the curvature-based heuristic produces a reconstruction that matches the original dynamics even if the data are distorted by smoothing. To explore this, we compared the full dynamics, from Fig. 10(a), to those for  $\tau_C$  and  $\tau_I$  reconstructions of the *filtered* time-series data in Fig. 12. The correlation sum plots for the two reconstructions are significantly different. The  $\tau_I$  embed-

<sup>g</sup> The Theiler windows for these calculation, and for the rest of the examples in this section, were chosen as described above.

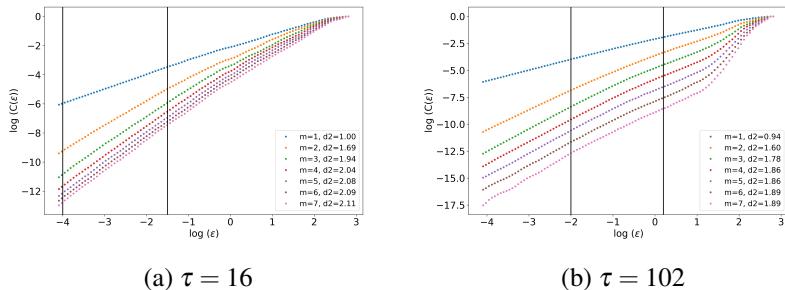


FIG. 12: Log-log plots of  $C(\varepsilon)$  and calculations of the correlation dimension of the filtered Lorenz data set using  $\tau = 16$ —the value suggested by the curvature heuristic—and  $\tau = 102$ , the first clear minimum of AMI. For the full state space, from Fig. 10,  $d_2 = 2.05$ .

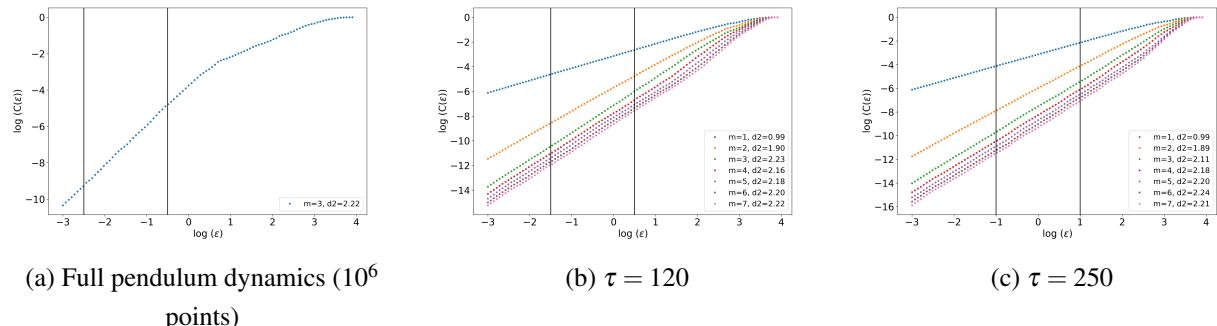


FIG. 13: Log-log plots of  $C(\epsilon)$  and calculations of the correlation dimension of the  $10^6$  point pendulum data set for the full dynamics giving  $d_2 = 2.22$ , and using  $\tau = 120$ —a representative value in the range suggested by the curvature heuristic and  $\tau = 250$ —the value suggested by AMI.

<sup>459</sup> ding in Fig. 12(a) has a narrow<sup>h</sup> scaling region,  $\log(\varepsilon) \in [-2, 0.2]$ . A fit in this region gives  
<sup>460</sup> the underestimate  $d_2 = 1.89$ . The  $\tau_C$  embedding, on the other hand, has a broad scaling region,  
<sup>461</sup>  $\log(\varepsilon) \in [-4, -1.5]$ , yielding  $d_2 = 2.11$ , closer to the dimension 2.05 of the unfiltered, full dy-  
<sup>462</sup> namics.

463 Figs. 13 and 14 show the correlation plots for the driven damped pendulum example discussed  
 464 in §IV D and §IVE, respectively. For the  $10^6$ -point trajectory, the correlation plot for the full  
 465 dynamics, Fig. 13(a), has a broad scaling region  $\log(\varepsilon) \in [-2.5, 0.5]$ ; a fit for this region yields  
 466  $d_2 = 2.22$ . The  $\tau_C$  and  $\tau_I$  reconstructions of these data in panels (b) and (c) of the figure also have  
 467 broad scaling regions ( $\log(\varepsilon) \in [-1, 1.5]$ ), with  $d_2 = 2.22$  and  $d_2 = 2.21$ , respectively—both good  
 468 approximations to the true value. The scaling regions for the shorter pendulum data set, shown  
 469 in Fig. 14, are identical to those for the longer trajectory, but their slopes are different. For the  
 470  $\tau_I = 120$  reconstruction,  $d_2 = 2.20$ , closely matching that for the full dynamics,  $d_2 = 2.18$ . The  
 471  $\tau_I = 250$  reconstruction, on the other hand, overestimates the dimension, giving  $d_2 = 2.35$ .

<sup>h</sup> A broader scaling region, e.g., starting at  $\log(\varepsilon) = -4$ , is not appropriate because of the slight bend in the  $m = 7$  profile at  $\log(\varepsilon) = -2$ .

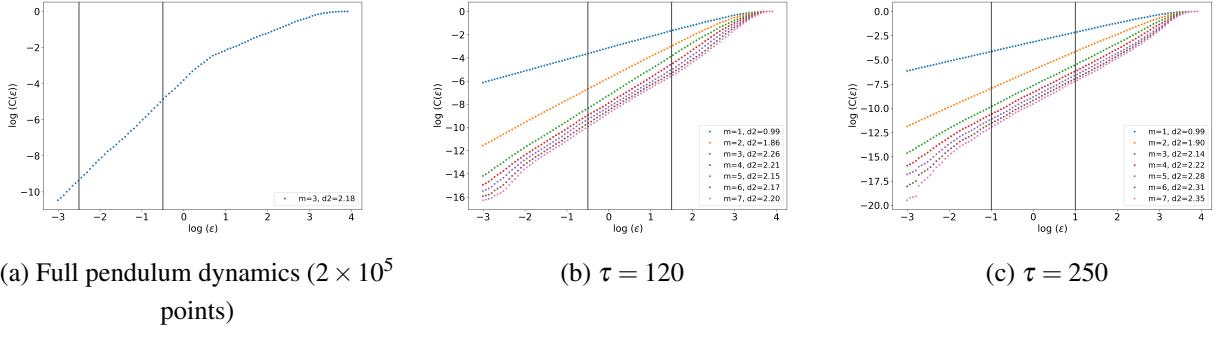


FIG. 14: Calculations of the correlation dimension of the shorter pendulum data set for the full dynamics giving  $d_2 = 2.18$ , and using  $\tau = 120$ —a representative value in the range suggested by the curvature heuristic—and  $\tau = 250$ , the value suggested by AMI.