

**Dynamic models with p parameters are identified by $2p + 1$
random features**

Michael Wieck-Sosa¹ and Cosma Rohilla Shalizi^{1,2}

¹*Department of Statistics & Data Science,
Carnegie Mellon University, Pittsburgh, PA 15213 USA*

²*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501 USA**

(Dated: July 17, 2026)

Abstract

A foundational principle in nonlinear dynamics is that the structure of a dynamical system can be recovered from a small number of generic measurements or coordinates. We develop an analogous principle for the identification of dynamic models for time series *with noise*, which builds on previous identification results for noiseless dynamical systems. The noise is allowed to be non-iid, non-Gaussian, and dependent on the state. Our results cover noisily observed differential equations and discrete-time dynamical systems, as well as stochastic models with process noise. We illustrate the utility of this identification principle using a Lorenz-63 model and a Hénon map model, both with observational noise.

8 I. INTRODUCTION

9 Since Packard *et al.* [1], one aim of nonlinear dynamics has been to recover the attractors
10 of deterministic dynamical systems from time series of a single observable function of the
11 state. Long before that, time series analysts had long practiced predicting the next obser-
12 vation from a vector of delayed values of the observable, and it was intuitively plausible
13 that adding more delayed values would give more information about the state, and so about
14 future observations. What made “geometry from a time series” or “state space reconstruc-
15 tion” into both a research program and a practical tool was the proof by Takens [2] that,
16 when the state had a finite dimension, it could be reconstructed exactly, up to a smooth
17 change of coordinates, by a *finite* number of delayed observations, at most $2d + 1$ when the
18 state has d dimensions. Or, rather, all of this is true for “typical” or “generic” choices of the
19 observable and delay.

20 “Geometry from a time series” only allows one to reconstruct the unobserved state up to a
21 smooth change of coordinates. As such, it cannot support inferences about specific functional
22 forms for the underlying dynamics, or for numerical parameters describing those dynamics.
23 On the other hand, scientists increasingly express their dynamical ideas as models with
24 very specific functional forms, with a variety of physically-meaningful but *a priori* unknown
25 parameters. For the models to be useful, those parameters must be estimated from data[3].

26 A logically prior question to estimation, however, is whether the parameters of the model

* Contact author: cshalizi@cmu.edu

can be inferred from the data available, i.e., whether the observed data contain any information about the parameters at all. This is what statisticians call “parameter identification”, or just “identification”[4]. Speaking loosely (we are more precise below), the parameter vector θ is identified by the random vector X when differences in θ imply differences in the distribution of the random variable: $\theta_1 \neq \theta_2$ only if $p(x; \theta_1) \neq p(x; \theta_2)$ for at least some x . When a model is *un*-identified, there are at least two parameter vectors which lead to exactly the same distribution, hence no conceivable data analysis can distinguish those parameters. Most computational models advanced in scientific practice would be identified, if we could observe the full state of the system[5]; we should, of course, be so lucky.

Our main result here is two-fold. First, for a large class of dynamical systems observed through stochastic noise, a p -dimensional vector of parameters can be identified through the distribution of $2p + 1$ generic functions of observables, or “features”[6]. In fact, there is generically a diffeomorphism between the parameter space and the *expectation values* of those features. We may actually choose the features *randomly*, from a fixed, data-independent distribution over a suitably-rich class of basis functions, e.g., Fourier basis functions for vector-valued processes. We show all this by adapting results from the state-space reconstruction literature, especially the fractal Whitney embedding prevalence theorem from Sauer *et al.* [7]. Second, because of that diffeomorphism, and because the full distribution of those observables is not required for identification, we can construct practical estimation procedures for inferring the parameters from time series data, either in the long-time (“ergodic”) limit for stationary or asymptotically-mean-stationary processes, or in the dense-sampling (“in-fill”) limit for nonstationary processes.

A. Related work

As mentioned, the “geometry from a time series” program traces back to Packard *et al.* [1]. This initiated the procedure of forming vectors from finitely many delayed values of a scalar signal, producing what is now called a *time-delay embedding*, and analyzing the resulting time-delay embedding space. Takens [2] established that using $2d + 1$ delays yields a time-delay embedding space that is topologically equivalent to the original d -dimensional state space. The number of required delays can be smaller in many cases, with sharper bounds depending on the box-counting dimension of the attractor [7]. Time-delay embeddings

became a major source of inspiration for later works in nonlinear dynamics; see Kantz and Schreiber [8] for a comprehensive overview.

The Takens [2] embedding theorem has been extended in several directions. Sauer *et al.* [7] show that almost every time-delay map is an embedding. Similarly, Barański *et al.* [9] show that the time-delay map formed by a polynomial perturbation of a given locally Lipschitz observation function is almost surely an embedding. Stark *et al.* [10, 11] extend the Takens [2] embedding theorem to forced systems and stochastic systems. Casdagli *et al.* [12] use statistical methods to account for the effect of iid observational noise on the state space reconstruction. Recently, Botvinick-Greenhouse *et al.* [13] established a measure-theoretic generalization using optimal transport theory. These results all trace back to the classical Whitney [14] embedding theorem from differential topology.

Related results also appear in parameter identification and observability. Aeyels [15, 16] proved that the observability of a d -dimensional dynamical system can be achieved by taking $2d + 1$ exact measurements at random times. For analytically parametrized dynamical systems, Sontag [17] proved that $2p + 1$ generic experiments are sufficient to distinguish between distinct p -dimensional parameter values. However, these results, unlike ours, are restricted to dynamical systems *without noise*.

II. RESULTS

A. Genericity

To begin, we state the five *genericity axioms* introduced by Yorke and Ott [18]. For a topological vector space V , we desire that a definition of “generic” satisfies the following:

1. A generic subset of V is dense in V .
2. If $F \supset E$ and E is generic, then F is generic.
3. A countable intersection of generic sets is generic.
4. Every translate of a generic set is generic.
5. A subset E of finite-dimensional Euclidean space is generic if and only if E is a set of full Lebesgue measure.

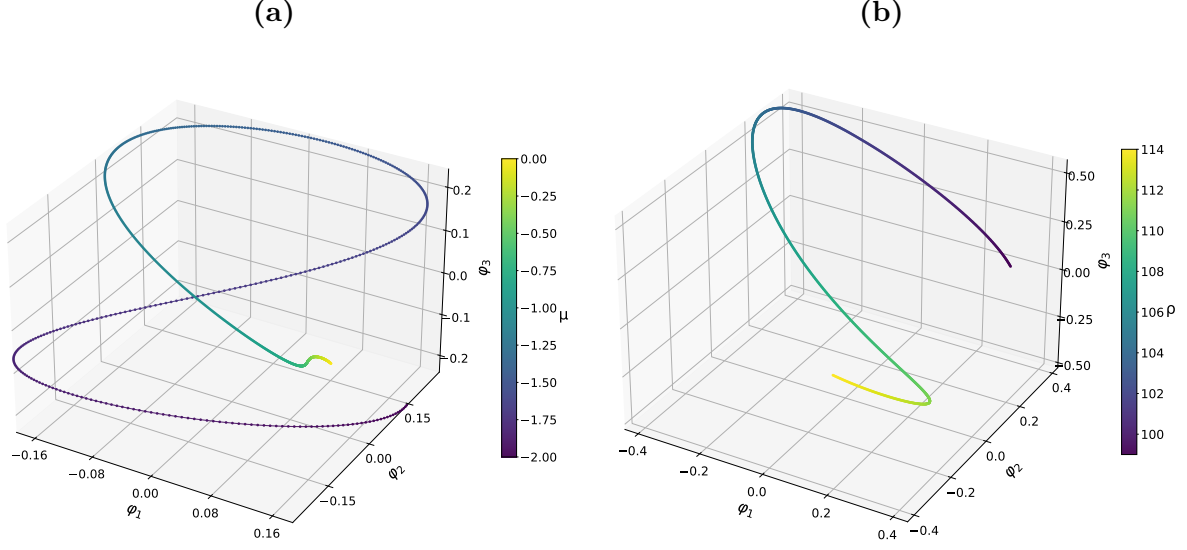


FIG. 1. The same $k = 2p + 1 = 3$ random features $\varphi = (\varphi_1, \varphi_2, \varphi_3)$ as in Eq. (1) identify every parameter value in both the Gaussian distribution model and the Lorenz-63 model.

(a) Let $(X_t^{(r)}(\mu))_{r=1}^s$, $s = 10$, be $X_t^{(r)}(\mu) \stackrel{\text{iid}}{\sim} N((\mu, \mu^2, \mu^3)^\top, (1, 1, 1)^\top)$, $r = 1, \dots, s$. The plot shows $s^{-1} \sum_{r=1}^s \varphi(X_t^{(r)}(\mu))$ as a function of the mean parameter μ . Color denotes $\mu \in [-2, 0]$.

(b) Let $(X_t^{(r)}(\rho))_{r=1}^s$, $s = 10$, be the noisily observed Lorenz-63 model as in Eq. (17) at time $T = 1$ with fixed parameter values $\sigma = 10$, $\beta = 8/3$, and $\lambda^{(1)} = \lambda^{(2)} = \lambda^{(3)} = 1$. The plot shows $s^{-1} \sum_{r=1}^s \varphi(X_t^{(r)}(\rho))$ as a function of the Rayleigh number ρ . Color denotes $\rho \in [99, 114]$.

84 The topological notion of genericity uses *residual sets*, which are countable intersections of
85 open, dense sets. By the Baire category theorem, residual subsets of Baire spaces are dense.
86 However, the topological notion of genericity does not satisfy Axiom 5 because the Lebesgue
87 measure of these residual sets can be zero. Thus, we would like to use a measure-theoretic
88 notion of genericity.

89 Unfortunately, the notion of “probability one” in finite-dimensional spaces does not ex-
90 tend in a canonical way to infinite-dimensional spaces. Thus, since the space of all smooth
91 maps is an infinite-dimensional space, it is difficult to assert that “almost every” smooth
92 map has a given property; see Sauer *et al.* [7] and Hunt *et al.* [19] for more discussion. To
93 formalize the notion that “almost every” map has a property, we use the idea of *prevalence*
94 from Yorke and Ott [18].

95 **Definition 1** (Prevalence [18]). Let V be a completely metrizable topological vector space.
96 A Borel set $E \subset V$ is said to be *prevalent* if there exists a Borel measure μ on V such that:

1. $0 < \mu(C) < \infty$ for some compact subset C of V .
 2. The set $E+v$ has full μ -measure (complement of $E+v$ has measure zero) for all $v \in V$.
- If $E \subset V$ contains a prevalent Borel set, then it is common to say that “almost every” element of V lies in E .

Next, we state a classical embedding result for smooth maps from Sauer *et al.* [7], which involves the notion of an *immersion*.

Definition 2 (Immersion [20]). A smooth map $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^k$ is an *immersion* at $\theta \in \mathbb{R}^p$ if its differential map $D_\theta \Phi$ at θ is injective from $\mathbb{R}^p$ to $\mathbb{R}^k$.

The following result establishes that “almost every” smooth map (in the sense of prevalence from Definition 1) from $\mathbb{R}^p$ to $\mathbb{R}^k$, $k \geq 2p + 1$, has the following properties.

Theorem 1 (Fractal Whitney Embedding Prevalence Theorem [7]). *Let Θ be a compact subset of $\mathbb{R}^p$ of box-counting dimension p' , and let k be an integer greater than $2p'$. For almost every smooth map $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^k$:*

1. Φ is one-to-one on Θ .
2. Φ is an immersion on each compact subset of a smooth manifold contained in Θ .

B. Setting and notation

For a vector $x \in \mathbb{R}^d$, denote the ℓ^q norm by $\|x\|_q$ and the Euclidean norm by $\|x\| = \|x\|_2$. For a random vector X with distribution depending on $\theta \in \Theta$, $\|X\|_{\mathcal{L}^q(\theta)} = (\mathbb{E}_\theta \|X\|^q)^{1/q}$, where $\mathbb{E}_\theta(\cdot)$ denotes expectation with respect to this θ -dependent distribution. For any natural number $j \in \mathbb{N}$, denote $[j] = \{1, 2, \dots, j\}$. For a sequence of random vectors X_t , $t = 1, \dots, n$, each taking values in $\mathbb{R}^d$, we denote a subsequence as $X_{t_1:t_2} = (X_{t_1}, \dots, X_{t_2})^\top$, which takes values in $\mathbb{R}^{(t_2-t_1+1) \times d}$ for some $t_1, t_2 \in [n]$. For real numbers $x, y \in \mathbb{R}$, denote $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$.

We observe X_t^{obs} , $t = 1, \dots, n$, taking values in $\mathbb{R}^d$ for some dimension $d \in \mathbb{N}$. The observed sequence is assumed to come from a known parametric model with an unknown p -dimensional parameter vector $\theta_0 \in \Theta \subset \mathbb{R}^p$, $p \in \mathbb{N}$. At any value of θ , we can simulate $s \in \mathbb{N}$ realizations of a time series $(X_{1:n}^{(r)}(\theta))_{r \in [s]}$ of length n . To refer to the i -th dimension of

the r -th realization, we use the notation $X_{1:n}^{(r)(i)}$. When we do not need to refer to a particular realization or dimension, we drop the corresponding superscript.

C. Random Fourier features

We aim to estimate the p -dimensional parameter vector θ_0 by matching the values of $2p + 1$ random features of the simulated and observed data. We randomly draw $k = 2p + 1$ functions $\varphi_1, \dots, \varphi_k$ from a distribution D over a suitably nice function class $\mathcal{F}$, where each function $\varphi_i : \mathbb{R}^{(m+1) \times d} \rightarrow \mathbb{R}$ takes a length $m + 1$ subsequence of the d -dimensional time series as input. The number of lags m is chosen based on the dynamics of the model, e.g., set $m = m^*$ for an order m^* Markov process.

Let $x_1, \dots, x_{m+1}$ be vectors in $\mathbb{R}^d$ and denote $x = (x_1, \dots, x_{m+1})^\top \in \mathbb{R}^{(m+1) \times d}$. Define

$$\varphi_i(x) = \cos \left(\sum_{j=1}^{m+1} \Omega_{i,j} \cdot x_j + \alpha_i \right), \quad (1)$$

where $\Omega_{i,j} \stackrel{\text{iid}}{\sim} N(0, I_d)$ and $\alpha_i \stackrel{\text{iid}}{\sim} U(-\pi, \pi)$ for $i = 1, \dots, k$ and $j = 1, \dots, m + 1$. Denote

$$\varphi = (\varphi_1, \dots, \varphi_k), \quad (2)$$

so that $\varphi : \mathbb{R}^{(m+1) \times d} \rightarrow \mathbb{R}^k$. Define the i -th random feature at time $t = m + 1, \dots, n$ as

$$f_{t,i}^{\text{obs}} = \varphi_i(X_{t-m:t}^{\text{obs}}), \quad f_{t,i}^{(r)}(\theta) = \varphi_i \left(X_{t-m:t}^{(r)}(\theta) \right),$$

and write all random features at time t as

$$f_t^{\text{obs}} = (f_{t,1}^{\text{obs}}, \dots, f_{t,k}^{\text{obs}}), \quad f_t^{(r)}(\theta) = (f_{t,1}^{(r)}(\theta), \dots, f_{t,k}^{(r)}(\theta)). \quad (3)$$

D. Identification of stationary dynamic models

For processes with meaningful stationary distributions, we consider the average of the random features over time. The following *time-average* estimator was studied in our companion paper [21], and was shown to be consistent under additional assumptions, e.g., decaying temporal dependence. In general, different estimators and random features may be used.

In this setting, as n grows, we observe the process further into the future. Define

$$F^{\text{obs}} = \frac{1}{n - m} \sum_{t=m+1}^n f_t^{\text{obs}}, \quad \bar{F}^{\text{sim}}(\theta) = \frac{1}{n - m} \sum_{t=m+1}^n \bar{f}_t^{\text{sim}}(\theta), \quad (4)$$

142 where $\bar{f}_t^{\text{sim}}(\theta) = \frac{1}{s} \sum_{r=1}^s f_t^{(r)}(\theta)$. The time-average estimator is defined as

$$\hat{\theta}^{\text{TA}} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\| \hat{Q}_n^{\text{TA}}(\theta) \right\|, \quad (5)$$

143 where the empirical discrepancy function is given by

$$\hat{Q}_n^{\text{TA}}(\theta) = F^{\text{obs}} - \bar{F}^{\text{sim}}(\theta). \quad (6)$$

144 Any suitable optimization procedure may be used to find the minimizer of $\left\| \hat{Q}_n^{\text{TA}}(\cdot) \right\|$.

145 Denote by $P_{\theta,t}^{(r)}$ the distribution of $\left[X_{t-m}^{(r)}(\theta), \dots, X_t^{(r)}(\theta) \right]^\top$, where m is from Eq. (2). We
 146 require the average of the distributions $(P_{\theta,t}^{(r)})_{t \in \mathbb{N}}$ over time to converge weakly to a limiting
 147 distribution P_θ , *uniformly* over $\theta \in \Theta$. This allows us to randomly initialize the model, so
 148 that the method can handle dynamical systems with unknown initial values.

Assumption 1. *For all $\theta \in \Theta$, there exists a limiting distribution P_θ such that, for all $r = 0, 1, \dots, s$, we have, as $n \rightarrow \infty$,*

$$\sup_{\theta \in \Theta} \left| \int_{\mathbb{R}^{(m+1) \times d}} \phi(x) d\bar{P}_{\theta,n}^{(r)}(x) - \int_{\mathbb{R}^{(m+1) \times d}} \phi(x) dP_\theta(x) \right| \rightarrow 0,$$

149 *for all continuous, bounded functions $\phi : \mathbb{R}^{(m+1) \times d} \rightarrow \mathbb{R}$, where $\bar{P}_{\theta,n}^{(r)} = \frac{1}{n-m} \sum_{t=m+1}^n P_{\theta,t}^{(r)}$.*

150 We introduce the mapping $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^k$, whose restriction to Θ is given by

$$\Phi|_\Theta(\theta) = \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) dP_\theta(x), \quad \theta \in \Theta. \quad (7)$$

151 The limiting expectation values of the random features coincide with Φ restricted to Θ .

152 Next, we impose conditions so Theorem 1 can be applied, and that, for “generic” choices
 153 of the random features, Φ is C^1 , Φ^{-1} exists, and Φ^{-1} is also C^1 . Let $\mathcal{P}(\mathbb{R}^{(m+1) \times d})$ be the
 154 space of probability measures on $\mathbb{R}^{(m+1) \times d}$.

155 **Assumption 2.** *The following conditions hold:*

- 156 1. Θ is a compact subset of $\mathbb{R}^p$ with non-empty interior.
- 157 2. The map $\theta \mapsto P_\theta$ is a bijection.
- 158 3. For some open set $O \supset \Theta$, the map $\theta \mapsto P_\theta$ admits an extension $\mathbf{p} : O \rightarrow \mathcal{P}(\mathbb{R}^{(m+1) \times d})$
 159 that is Fréchet C^1 on O with respect to the total variation norm.

160 In other words, there exists an open set $O \subset \mathbb{R}^p$ with $\Theta \subset O$ and a map $\mathbf{p} : O \rightarrow$
161 $\mathcal{P}(\mathbb{R}^{(m+1) \times d})$, such that $\mathbf{p}(\theta) = P_\theta$ for all $\theta \in \Theta$. Furthermore, $\mathbf{p}$ is Fréchet C^1 on O
162 with respect to the total variation norm. We link $\mathbf{p}$ from Assumption 2 to Φ through the
163 definition of the function class

$$\mathcal{C}_\varphi = \left\{ \Phi : \mathbb{R}^p \rightarrow \mathbb{R}^k \mid \Phi|_O(\theta) = \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) d\mathbf{p}(\theta)(x), \Phi \text{ is } C^1 \text{ on } \mathbb{R}^p \setminus \Theta \right\}.$$

164 Let $\mathcal{C} = \bigcup_{\varphi \in \mathcal{F}^k} \mathcal{C}_\varphi$, so that $\mathcal{C}$ consists of functions $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^k$ whose restrictions to $O \subset \mathbb{R}^p$
165 from Assumption 2 coincide with the expectations of the random features (i.e. with respect
166 to $\mathbf{p}(\theta), \theta \in O$, for some $\varphi \in \mathcal{F}^k$), and which admit a C^1 extension to the rest of $\mathbb{R}^p$.

167 The following result establishes that $2p + 1$ random features are sufficient to identify a
168 dynamic model with a p -dimensional parameter.

169 **Theorem 2.** *Suppose Assumptions 1-2 hold. Then the limiting expectation values of the k*
170 *random features are C^1 -smooth in the parameter values, so $\mathcal{C} \subset C^1(\mathbb{R}^p, \mathbb{R}^k)$. Furthermore,*
171 *once $k \geq 2p + 1$, almost every C^1 smooth map from $\mathbb{R}^p$ to $\mathbb{R}^k$ is one-to-one on Θ , and is an*
172 *immersion on each compact subset of a smooth manifold contained in Θ .*

Suppose the randomly chosen functions from Eq. (2) yield an embedding. Indeed, by
Theorem 2, they will, almost surely. Then we can distinguish parameters $\theta_1 \neq \theta_2$ by look-
ing at the expectation values of the random features $\Phi(\theta_1)$ and $\Phi(\theta_2)$. In particular, the
population discrepancy function

$$Q^{\text{TA}}(\theta) = \Phi(\theta_0) - \Phi(\theta),$$

173 corresponding to the empirical discrepancy function from Eq. (6), will be nonzero for $\theta \neq \theta_0$.

174 E. Identification of nonstationary dynamic models

175 For many nonstationary processes, e.g., ODEs with observational noise, the time-average
176 of the random features may not converge to a limiting value that uniquely identifies the
177 parameters. Hence, a local approach is required. The following *rolling-window* estimator
178 was studied in our companion paper [21], and was shown to be consistent.

179 In this setting, as n grows, we gain an increasing number of observations related to each
180 local structure of the nonstationary process. That is,

$$X_t^{\text{obs}} = \tilde{X}_t^{\text{obs}}(t/n), \quad X_t^{(r)}(\theta) = \tilde{X}_t^{(r)}(t/n, \theta), \quad (8)$$

for some random vectors $\tilde{X}_t^{\text{obs}}(u)$, $\tilde{X}_t^{(r)}(u, \theta)$ that are well-defined at every *rescaled* time $u \in [0, 1]$. Note that X_t^{obs} and $X_t^{(r)}(\theta)$ both depend on n , although we suppress this in the notation. For the technical details about the data generating mechanism from Eq. (8), see the companion paper [21]. Denote the rolling averages of the random features up to time t with window size $w \in \mathbb{N}$ by

$$F_t^{\text{obs}} = \frac{1}{w \wedge t} \sum_{j=(t-w) \vee (m+1)}^t f_j^{\text{obs}}, \quad \bar{F}_t^{\text{sim}}(\theta) = \frac{1}{w \wedge t} \sum_{j=(t-w) \vee (m+1)}^t \bar{f}_j^{\text{sim}}(\theta), \quad (9)$$

where $\bar{f}_t^{\text{sim}}(\theta) = \frac{1}{s} \sum_{r=1}^s f_t^{(r)}(\theta)$. The rolling-window estimator is defined as

$$\hat{\theta}^{\text{RW}} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left| \hat{Q}_n^{\text{RW}}(\theta) \right|, \quad (10)$$

where the empirical discrepancy function is given by

$$\begin{aligned} \hat{Q}_n^{\text{RW}}(\theta) &= \frac{1}{n-m} \sum_{t=m+\tau+L}^n [\|\Delta_{t-L}\|^2 + K_t(\theta)], \\ K_t(\theta) &= 2 (\Delta_{t-L})^\top ([f_t^{\text{obs}} - \bar{f}_t^{\text{sim}}(\theta)] - [\Delta_{t-L}]), \\ \Delta_{t-L} &= F_{t-L}^{\text{obs}} - \bar{F}_{t-L}^{\text{sim}}(\theta), \end{aligned} \quad (11)$$

and where $\tau \in \mathbb{N}$ is an initial time-offset and $L \in \mathbb{N}$ is a lag. We use an optimization procedure to find the minimizer of $|\hat{Q}_n^{\text{RW}}(\cdot)|$. We select $w = w_n$, $\tau = \tau_n$, and $L = L_n$ as follows: the lag L_n is selected as $\lceil \frac{1}{10} \log(n)^2 \rceil$, the window size w_n is selected as the integer within $[n^{\frac{1}{2}}, n^{\frac{3}{4}}]$ that minimizes the sum of squared distances $\sum_{t=m+1+L}^n \|F_{t-L}^{\text{obs}} - f_t^{\text{obs}}\|^2$, and the offset is selected as $\tau_n = w_n$.

For all rescaled times $u \in [0, 1]$, let $P_{\theta,u}$ be the distribution of $[\tilde{X}_{t-m}(u, \theta), \dots, \tilde{X}_t(u, \theta)]^\top$, which does not depend on the simulation index r . We introduce mappings $\Phi_u : \mathbb{R}^p \rightarrow \mathbb{R}^k$, $u \in [0, 1]$, whose restrictions to Θ satisfy

$$\Phi_u|_{\Theta}(\theta) = \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) dP_{\theta,u}(x). \quad (12)$$

The expectation values of the random features at the rescaled time $u \in [0, 1]$ coincide with Φ_u restricted to the parameter space Θ .

Typically, it is not possible to embed Θ into $\mathbb{R}^k$ using Φ_u at a particular rescaled time $u \in [0, 1]$. This is because we may have $P_{\theta_1,u} = P_{\theta_2,u}$ for two different parameters $\theta_1 \neq \theta_2$. The following reparametrization assumption is needed.

Assumption 3. For all $u \in [0, 1]$, there exists a Lipschitz mapping $g_u : \Theta \rightarrow \tilde{\Theta}_u$ defined by $g_u(\theta) = \tilde{\theta}_u$ where $\tilde{\Theta}_u = \{g_u(\theta) : \theta \in \Theta\}$ is a subset of $\mathbb{R}^{\tilde{p}}$ for some $\tilde{p} \in \mathbb{N}$, and there exists a reparametrization such that

$$\tilde{X}'_t(u, \tilde{\theta}_u) = \tilde{X}_t(u, \theta).$$

Let $\tilde{P}_{\tilde{\theta}_u, u}$ be the distribution of $\left[\tilde{X}'_{t-m}(u, \tilde{\theta}_u), \dots, \tilde{X}'_t(u, \tilde{\theta}_u)\right]^\top$. We introduce mappings $\tilde{\Phi}_u : \mathbb{R}^{\tilde{p}} \rightarrow \mathbb{R}^k$ whose restrictions to $\tilde{\Theta}_u$ satisfy

$$\tilde{\Phi}_u|_{\tilde{\Theta}_u}(\tilde{\theta}_u) = \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) d\tilde{P}_{\tilde{\theta}_u, u}(x). \quad (13)$$

The *local* expectations of the random features are $\tilde{\Phi}_u(\tilde{\theta}_u)$, $\tilde{\theta}_u \in \tilde{\Theta}_u$. Next, we impose conditions so that Theorem 1 may be applied at each rescaled time $u \in [0, 1]$. Let $\mathcal{P}(\mathbb{R}^{(m+1) \times d})$ be the space of probability measures on $\mathbb{R}^{(m+1) \times d}$.

Assumption 4. The following conditions hold for all rescaled times $u \in [0, 1]$:

1. $\tilde{\Theta}_u$ is a compact subset of $\mathbb{R}^{\tilde{p}}$, and Θ is a compact subset of $\mathbb{R}^p$ with non-empty interior.
2. The map $\tilde{\theta}_u \mapsto \tilde{P}_{\tilde{\theta}_u, u}$ is a bijection.
3. For some open set $\tilde{O}_u \supset \tilde{\Theta}_u$, the map $\tilde{\theta}_u \mapsto \tilde{P}_{\tilde{\theta}_u, u}$ admits an extension $\tilde{\mathbf{p}}_u : \tilde{O}_u \rightarrow \mathcal{P}(\mathbb{R}^{(m+1) \times d})$ that is Fréchet C^1 on $\tilde{O}_u$ with respect to the total variation norm.

In other words, for all $u \in [0, 1]$, there exists an open set $\tilde{O}_u \subset \mathbb{R}^{\tilde{p}}$ with $\tilde{\Theta}_u \subset \tilde{O}_u$ and a map $\tilde{\mathbf{p}}_u : \tilde{O}_u \rightarrow \mathcal{P}(\mathbb{R}^{(m+1) \times d})$ such that $\tilde{\mathbf{p}}_u(\tilde{\theta}_u) = \tilde{P}_{\tilde{\theta}_u, u}$. Also, $\tilde{\mathbf{p}}_u$ is Fréchet C^1 on $\tilde{O}_u$ with respect to the total variation norm. We link $\tilde{\mathbf{p}}_u$, $u \in [0, 1]$, to $\tilde{\Phi}_u$, $u \in [0, 1]$, through the definition of

$$\tilde{\mathcal{C}}_{u, \varphi} = \left\{ \tilde{\Phi}_u : \mathbb{R}^{\tilde{p}} \rightarrow \mathbb{R}^k \mid \tilde{\Phi}_u|_{\tilde{O}_u}(\tilde{\theta}_u) = \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) d\tilde{\mathbf{p}}_u(\tilde{\theta}_u)(x), \tilde{\Phi}_u \text{ is } C^1 \text{ on } \mathbb{R}^{\tilde{p}} \setminus \tilde{\Theta}_u \right\},$$

for $u \in [0, 1]$. Let $\tilde{\mathcal{C}}_u = \bigcup_{\varphi \in \mathcal{F}^k} \tilde{\mathcal{C}}_{u, \varphi}$ be the functions $\tilde{\Phi}_u : \mathbb{R}^{\tilde{p}} \rightarrow \mathbb{R}^k$ whose restrictions to $\tilde{O}_u \subset \mathbb{R}^{\tilde{p}}$ from Assumption 4 coincide with the local expectations of the random features with respect to $\tilde{\mathbf{p}}_u(\tilde{\theta}_u)$, $\tilde{\theta}_u \in \tilde{O}_u$, for some $\varphi \in \mathcal{F}^k$, and which admit a C^1 extension to the rest of $\mathbb{R}^{\tilde{p}}$.

Theorem 3. Suppose Assumptions 3-4 hold. Then, for all $u \in [0, 1]$, the local expectation values of the k random features are C^1 -smooth in the parameter values, so $\tilde{\mathcal{C}}_u \subset C^1(\mathbb{R}^{\tilde{p}}, \mathbb{R}^k)$.

221 Also, for all $u \in [0, 1]$, once $k \geq 2p + 1$, almost every C^1 smooth map from $\mathbb{R}^{\tilde{p}}$ to $\mathbb{R}^k$ is one-
 222 to-one on $\tilde{\Theta}_u$, and is an immersion on each compact subset of a smooth manifold contained
 223 in $\tilde{\Theta}_u$.

224 We require that $(P_{\theta_1, u})_{u \in [0, 1]}$ and $(P_{\theta_2, u})_{u \in [0, 1]}$ are sufficiently distinguishable for $\theta_1 \neq \theta_2$.

225 **Assumption 5.** For all $\theta_1, \theta_2 \in \Theta$, if $\theta_1 \neq \theta_2$, then there exists a subset $\mathcal{U}_{\theta_1, \theta_2} \subseteq [0, 1]$ with
 226 positive Lebesgue measure such that $P_{\theta_1, u} \neq P_{\theta_2, u}$ for all rescaled times $u \in \mathcal{U}_{\theta_1, \theta_2}$.

If the random features from Eq. (2) yield an embedding (by Theorem 3, they almost surely will), then we can distinguish parameters $\theta_1 \neq \theta_2$ by comparing the local expectation values of the random features $\Phi_u(\theta_1)$, $\Phi_u(\theta_2)$ at $u \in \mathcal{U}_{\theta_1, \theta_2}$ from Assumption 5, and the population discrepancy function

$$Q^{\text{RW}}(\theta) = \int_0^1 \|\Phi_u(\theta_0) - \Phi_u(\theta)\|^2 du,$$

227 corresponding to the empirical discrepancy function from Eq. (11), will be nonzero for $\theta \neq \theta_0$.

228 III. APPLICATIONS TO ATMOSPHERIC CONVECTION MODELS

229 *Hénon map model.* Hénon [22] introduced a discrete-time model that exhibits chaotic
 230 behavior, which is a simplified model for the Poincaré section of the Lorenz-63 system [23].
 231 For times $t = 1, \dots, n$, the dynamical system is defined by

$$\begin{aligned} Z_t^{(1)} &= 1 - a(Z_{t-1}^{(1)})^2 + Z_{t-1}^{(2)}, \\ Z_t^{(2)} &= bZ_{t-1}^{(1)}, \end{aligned} \tag{14}$$

232 where $Z_t^{(1)}$, $Z_t^{(2)}$ are the state variables. The unknown initial values $Z_0^{(1)}$ and $Z_0^{(2)}$ are
 233 sampled from $U[-0.7, 0.7]$ and $U[-0.2, 0.2]$ distributions, respectively. The components of
 234 the parameter vector are: a , which relates to the forcing or stress, and b , which relates to
 235 the damping or dissipation.

236 For times $t = 1, \dots, n$, we observe a noisy version of the first coordinate

$$X_t = Z_t + \sigma \varepsilon_t, \tag{15}$$

237 where $\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, 1)$ and σ controls the noise level. We use iid Gaussian noise to simplify the
 238 exposition, but in general we allow the noise to be non-iid and non-Gaussian. We aim to

$$a_0 = 1.4, \quad b_0 = 0.3, \quad \sigma_0 = 0.1.$$

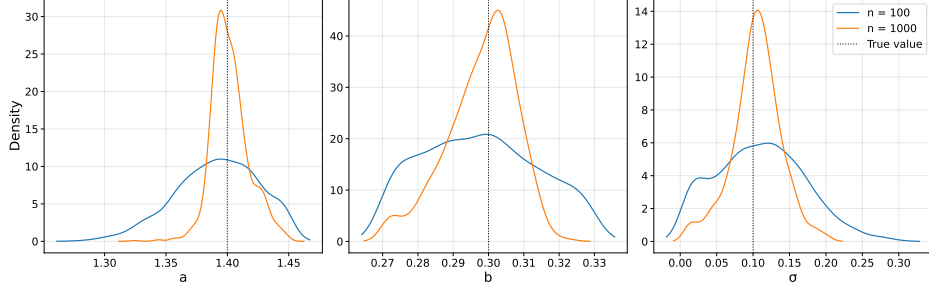


FIG. 2. **Hénon map model.** Density of parameter estimates using $2p + 1 = 7$ random features.

240 *Lorenz-63 model.* Lorenz [23] introduced a model for atmospheric convection that
 241 has become a canonical example of chaotic dynamics. For time horizon $T = 1$, let
 242 $(Z^{(1)}(v), Z^{(2)}(v), Z^{(3)}(v))_{v \in [0, T]}$ be defined by the system of ODEs

$$\begin{aligned} \frac{dZ^{(1)}}{dv}(v) &= \sigma(Z^{(2)}(v) - Z^{(1)}(v)), \\ \frac{dZ^{(2)}}{dv}(v) &= Z^{(1)}(v)(\rho - Z^{(3)}(v)) - Z^{(2)}(v), \\ \frac{dZ^{(3)}}{dv}(v) &= Z^{(1)}(v)Z^{(2)}(v) - \beta Z^{(3)}(v), \end{aligned} \tag{16}$$

243 where $Z^{(1)}(v)$, $Z^{(2)}(v)$, $Z^{(3)}(v)$ are the state variables at time $v \in [0, T]$. $Z^{(1)}(v)$ is pro-
 244 portional to the intensity of the convection, $Z^{(2)}(v)$ relates to the horizontal temperature
 245 variation, and $Z^{(3)}(v)$ relates to the vertical temperature variation. The known initial values
 246 are $Z^{(1)}(0) = 1$, $Z^{(2)}(0) = 1$, $Z^{(3)}(0) = 1$. The components of the parameter vector are: σ ,
 247 which relates to the Prandtl number, ρ , which relates to the Rayleigh number, and β , which
 248 relates to the fluid layer's physical dimensions.

249 We assume equally spaced observations, but this is not necessary. For $t = 1, \dots, n$, we
 250 observe

$$\begin{aligned} X_t^{(1)} &= Z^{(1)}(Tt/n) + \lambda^{(1)}\varepsilon_t^{(1)}, \\ X_t^{(2)} &= Z^{(2)}(Tt/n) + \lambda^{(2)}\varepsilon_t^{(2)}, \\ X_t^{(3)} &= Z^{(3)}(Tt/n) + \lambda^{(3)}\varepsilon_t^{(3)}, \end{aligned} \tag{17}$$

251 where $\varepsilon_t^{(1)}, \varepsilon_t^{(2)}, \varepsilon_t^{(3)} \stackrel{\text{iid}}{\sim} N(0, 1)$ and $\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}$ control the noise level. We aim to estimate

$$\begin{aligned} \sigma_0 &= 10, & \rho_0 &= 100.5, & \beta_0 &= 8/3, \\ \lambda_0^{(1)} &= 1, & \lambda_0^{(2)} &= 1, & \lambda_0^{(3)} &= 1. \end{aligned}$$

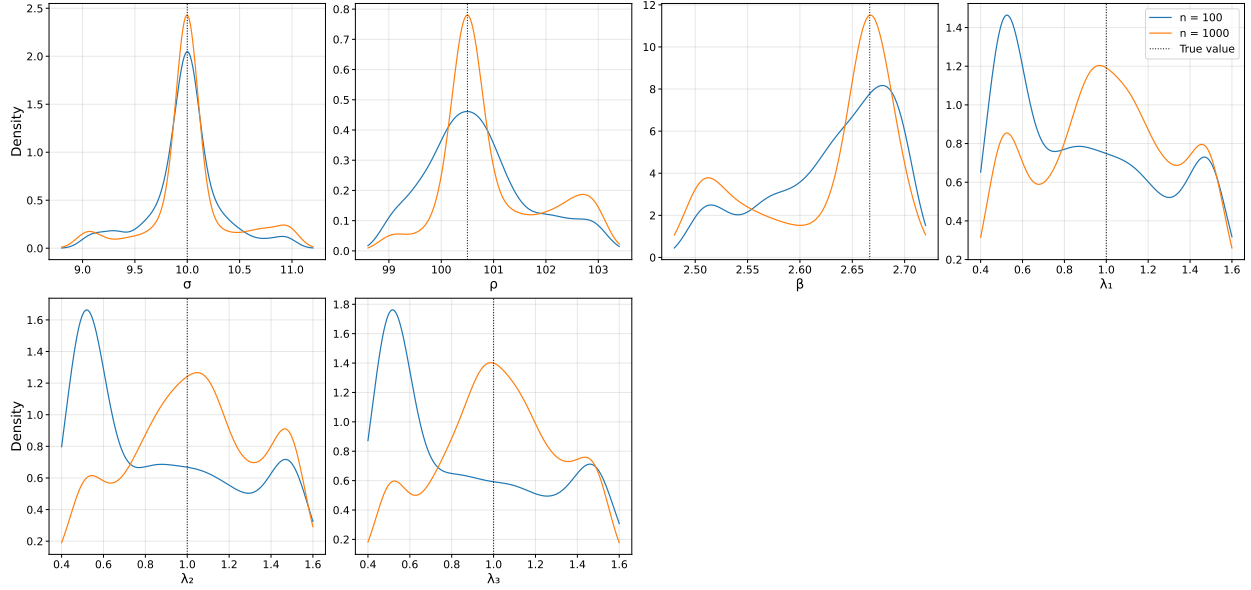


FIG. 3. **Lorenz-63 model.** Density of parameter estimates using $2p + 1 = 13$ random features.

IV. DISCUSSION

We present identification results for dynamic models *with noise*, which are inspired by previous results in nonlinear dynamics. It should be possible to extend our identification principle beyond dynamic models for time series. For example, it would be of interest to extend our ideas to parametric models for spatiotemporal data and network data. In general, the essential component is a class of random features for the dependence structure at hand.

Our identification results lead to broadly applicable estimation strategies based on matching $2p + 1$ random features of the simulated and observed data. Existing simulation-based estimation approaches are costly, because they either require user-chosen summary statistics or computationally intensive neural network training for learning summary statistics. In contrast, our approach requires very little input from the user. To get standard errors, a model-based bootstrap procedure can be used. In a separate manuscript, we show that the estimators are asymptotically normal, and we explain how to obtain valid confidence sets.

Appendix A: Proofs

1. Proof of Theorem 1

See the proof of Theorem 2.3 in [7].

□

2. Proof of Theorem 2

The proof is a streamlined version of the proof of Theorem 2 in the companion paper [21].

Step 1: Limiting expectations. We claim that, for all $\theta \in \Theta$, the limiting expectation values of the k random features are given by

$$\Phi(\theta) = \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) dP_\theta(x).$$

Observe that, for all $\theta \in \Theta$ and all $r \in \{0, 1, \dots, s\}$, as $n \rightarrow \infty$, we have

$$\left\| \frac{1}{n-m} \sum_{t=m+1}^n \mathbb{E}_\theta \left[\varphi \left(\left[X_{t-m}^{(r)}(\theta), \dots, X_t^{(r)}(\theta) \right]^\top \right) \right] - \Phi(\theta) \right\| = o(1), \quad (\text{A1})$$

by upper bounding the ℓ_2 norm $\|\cdot\|$ by the ℓ_1 norm (i.e. over the k random features) using the monotonicity of norms, linearity, and applying the weak convergence from Assumption 1.

Step 2: Smoothness. We show $\mathcal{C}$ from Theorem 2 consists of C^1 -smooth functions Φ . We establish that $\theta \mapsto \Phi(\theta)$ is differentiable on O from Assumption 2. For any $\varphi \in \mathcal{F}^k$, define

$$T_\varphi(\mu) = \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) d\mu(x),$$

which maps probability measures on $\mathbb{R}^{(m+1) \times d}$ to $\mathbb{R}^k$. For $\theta \in O$, denote the derivative

$$D\Phi(\theta)[h] = D \left(\int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) d\mathbf{p}(\theta)(x) \right) [h] = \int_{\mathbb{R}^{(m+1) \times d}} \varphi d(D\mathbf{p}(\theta)[h]),$$

where $h \in \mathbb{R}^p$. For all $\theta \in O$, we have, as $\|h\| \rightarrow 0$,

$$\begin{aligned}
& \frac{\|\Phi(\theta + h) - \Phi(\theta) - D\Phi(\theta)[h]\|}{\|h\|} \\
& \stackrel{(1)}{=} \frac{\left\| \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) d\mathbf{p}(\theta + h)(x) - \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) d\mathbf{p}(\theta)(x) - \int_{\mathbb{R}^{(m+1) \times d}} \varphi d(D\mathbf{p}(\theta)[h]) \right\|}{\|h\|} \\
& \stackrel{(2)}{=} \frac{\|T_\varphi(\mathbf{p}(\theta + h)) - T_\varphi(\mathbf{p}(\theta)) - T_\varphi(D\mathbf{p}(\theta)[h])\|}{\|h\|} \\
& \stackrel{(3)}{=} \frac{\|T_\varphi(\mathbf{p}(\theta + h) - \mathbf{p}(\theta) - D\mathbf{p}(\theta)[h])\|}{\|h\|} \\
& \stackrel{(4)}{\leq} \|\varphi\|_\infty \frac{\|\mathbf{p}(\theta + h) - \mathbf{p}(\theta) - D\mathbf{p}(\theta)[h]\|_{\text{TV}}}{\|h\|} \\
& \stackrel{(5)}{\leq} \frac{\|\mathbf{p}(\theta + h) - \mathbf{p}(\theta) - D\mathbf{p}(\theta)[h]\|_{\text{TV}}}{\|h\|} \\
& \stackrel{(6)}{\rightarrow} 0,
\end{aligned}$$

by (1) the definition of Φ , (2) the definition of T_φ , (3) the linearity of T_φ , (4) upper bounding by the sup norm multiplied by the total variation norm, (5) the boundedness of the random Fourier features so that $\|\varphi\|_\infty \leq 1$, and (6) because $\mathbf{p}$ is Fréchet C^1 on O by Assumption 2.

Next, we establish the continuity of $\theta \mapsto D\Phi(\theta)$ on O . By Assumption 2, we know that the map $\theta \mapsto D\mathbf{p}(\theta)$ is continuous on O . That is, for all $\theta \in O$ and all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $\theta' \in O$, if $\|\theta - \theta'\| < \delta$, then

$$\|D\mathbf{p}(\theta) - D\mathbf{p}(\theta')\|_{\text{op}} = \sup_{\|h\| \leq 1} \|D\mathbf{p}(\theta)[h] - D\mathbf{p}(\theta')[h]\|_{\text{TV}} < \epsilon.$$

Observe that, for all $\theta \in O$ and all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $\theta' \in O$, if $\|\theta - \theta'\| < \delta$, then for $\|D\Phi(\theta) - D\Phi(\theta')\|_{\text{op}} = \sup_{\|h\| \leq 1} \|D\Phi(\theta)[h] - D\Phi(\theta')[h]\|$ we have

$$\begin{aligned}
& \|D\Phi(\theta) - D\Phi(\theta')\|_{\text{op}} \\
& \stackrel{(1)}{=} \left\| \int_{\mathbb{R}^{(m+1) \times d}} \varphi d(D\mathbf{p}(\theta)) - \int_{\mathbb{R}^{(m+1) \times d}} \varphi d(D\mathbf{p}(\theta')) \right\|_{\text{op}} \\
& \stackrel{(2)}{=} \|T_\varphi(D\mathbf{p}(\theta)) - T_\varphi(D\mathbf{p}(\theta'))\|_{\text{op}} \\
& \stackrel{(3)}{=} \|T_\varphi(D\mathbf{p}(\theta) - D\mathbf{p}(\theta'))\|_{\text{op}} \\
& \stackrel{(4)}{\leq} \|\varphi\|_\infty \|D\mathbf{p}(\theta) - D\mathbf{p}(\theta')\|_{\text{op}} \\
& \stackrel{(5)}{\leq} \|D\mathbf{p}(\theta) - D\mathbf{p}(\theta')\|_{\text{op}} \\
& \stackrel{(6)}{\leq} \epsilon,
\end{aligned}$$

by (1) the definition of $D\Phi$, (2) the definition of T_φ , (3) the linearity of T_φ , (4) upper bounding by the sup norm multiplied by the operator norm, (5) the boundedness of the random Fourier features so that $\|\varphi\|_\infty \leq 1$, and (6) because $\mathbf{p}$ is Fréchet C^1 on O by Assumption 2. Hence, the map $\theta \mapsto \Phi(\theta)$ is C^1 on O . Note that the Whitney [24] extension theorem ensures the existence of a C^1 extension to the rest of $\mathbb{R}^p$. Therefore, $\mathcal{C}$ consists of C^1 -smooth functions, i.e. $\mathcal{C} \subset C^1(\mathbb{R}^p, \mathbb{R}^k)$.

Step 3: Embeddings. By Assumption 2, Θ is a compact subset of $\mathbb{R}^p$ with non-empty interior, so the box-counting dimension is p . Theorem 1 implies that, once $k \geq 2p + 1$, embeddings are prevalent in $C^1(\Theta, \mathbb{R}^k)$, i.e. “almost every” smooth map from $\mathbb{R}^p$ to $\mathbb{R}^k$ is one-to-one on Θ and is an immersion on each compact subset of a smooth manifold contained in Θ .

□

3. Proof of Theorem 3

The proof is a streamlined version of the proof of Theorem 5 in the companion paper [21].

Step 1: Local expectations. We claim that $\forall \tilde{\theta}_u \in \tilde{\Theta}_u$ the local expectation values of the random features are given by

$$\tilde{\Phi}_u(\tilde{\theta}_u) = \int_{\mathbb{R}^{(m+1) \times d}} \varphi(x) d\tilde{P}_{\tilde{\theta}_u, u}(x), \quad u \in [0, 1].$$

Observe that, for all $\theta \in \Theta$, all $u \in [0, 1]$, and all $r \in \{0, 1, \dots, s\}$, we have

$$\tilde{\Phi}_u(\tilde{\theta}_u) = \mathbb{E}_{\tilde{\theta}_u} \left[\varphi \left(\left[\tilde{X}'_{t-m}(u, \tilde{\theta}_u), \dots, \tilde{X}'_t(u, \tilde{\theta}_u) \right]^\top \right) \right] = \mathbb{E}_\theta \left[\varphi \left(\left[\tilde{X}'_{t-m}(u, \theta), \dots, \tilde{X}'_t(u, \theta) \right]^\top \right) \right],$$

by using the equivalent representation from Assumption 3.

Step 2: Smoothness. To show that $\tilde{\Phi}_u, u \in [0, 1]$, are C^1 -smooth in the parameter values, analogous arguments to those used in Step 2 of the proof of Theorem 2 can be used. By Assumption 4, $\tilde{\mathbf{p}}_u$ is Fréchet C^1 on $\tilde{O}_u$ (cf. Assumption 2), so the same arguments hold when replacing $\theta, \Phi, P_\theta, \mathcal{C}, O, \mathbf{p}$ with $\tilde{\theta}_u, \tilde{\Phi}_u, \tilde{P}_{\tilde{\theta}_u, u}, \tilde{\mathcal{C}}_u, \tilde{O}_u, \tilde{\mathbf{p}}_u$ at some $u \in [0, 1]$. We omit the details due to the similarity.

Step 3: Embeddings. The box-counting dimension of $\tilde{\Theta}_u$ is less than or equal to p , for all $u \in [0, 1]$, because: (1) Θ has box-counting dimension p because it is a compact subset of $\mathbb{R}^p$ with non-empty interior by Assumption 4, and (2) $\tilde{\Theta}_u = g_u(\Theta)$ and g_u is Lipschitz by

Assumption 3. Thus, we have $\dim(\tilde{\Theta}_u) = \dim(g_u(\Theta)) \leq \dim(\Theta) = p$. Theorem 1 implies that for each rescaled time $u \in [0, 1]$ (i.e. applying the theorem at each u), once $k \geq 2p + 1$, embeddings are prevalent in $C^1(\tilde{\Theta}_u, \mathbb{R}^k)$. That is, “almost every” smooth map from $\mathbb{R}^{\tilde{p}}$ to $\mathbb{R}^k$ is one-to-one on $\tilde{\Theta}_u$ and is an immersion on each compact subset of a smooth manifold contained in $\tilde{\Theta}_u$.

□

-
- [1] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Geometry from a time series, *Physical Review Letters* **45**, 712 (1980).
- [2] F. Takens, Detecting strange attractors in turbulence, *Symposium on Dynamical Systems and Turbulence, Warwick 1980*, 366 (1981).
- [3] Terminology differs confusingly across related fields. Among statisticians, inferring values of parameters from data is called “estimation”, while, e.g., climatologists and some economists are more apt to speak of “calibration”. We will generally follow conventions of statistics, but will try to define terms clearly.
- [4] By contrast, what control engineering and signal processing calls “system identification” is closer to statistics’ “estimation”.
- [5] There are rare exceptions with genuine redundancies or symmetries in the model parameterization, but these are usually accidental, and/or can be eliminated by, e.g., picking a canonical value from each congruence class.
- [6] In statistical theory, any (measurable) function of the data is called a “statistic”. We find this term leads to confusion among non-statisticians, so we will speak of “features”, as is common in data mining and machine learning.
- [7] T. Sauer, J. A. Yorke, and M. Casdagli, Embedology, *Journal of Statistical Physics* **65**, 579 (1991).
- [8] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, England, 2003).
- [9] K. Barański, Y. Gutman, and A. Śpiewak, A probabilistic takens theorem, *Nonlinearity* **33**, 4940 (2020).
- [10] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, Takens embedding theorems for forced

and stochastic systems, *Nonlinear Analysis* **30**, 5303–5314 (1997).

[11] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, Delay embeddings for forced systems. ii. stochastic forcing, *Journal of Nonlinear Science* **13**, 519 (2003).

[12] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, State space reconstruction in the presence of noise, *Physica D: Nonlinear Phenomena* **51**, 52 (1991).

[13] J. Botvinick-Greenhouse, M. Oprea, R. Maulik, and Y. Yang, Measure-theoretic time-delay embedding, *Journal of Statistical Physics* **192** (2025).

[14] H. Whitney, Differentiable manifolds, *Annals of Mathematics* **37**, 645 (1936).

[15] D. Aeyels, Generic observability of differentiable systems, *SIAM Journal on Control and Optimization* **19**, 595 (1981).

[16] D. Aeyels, On the number of samples necessary to achieve observability, *Systems and Control Letters* **1**, 92 (1981).

[17] E. D. Sontag, For differential equations with r parameters, $2r + 1$ experiments are enough for identification, *Journal of Nonlinear Science* **12**, 553 (2003).

[18] J. Yorke and W. Ott, Prevalence, *Bulletin of the American Mathematical Society* **42**, 263 (2005).

[19] B. R. Hunt, T. Sauer, and J. A. Yorke, Prevalence: a translation-invariant almost every on infinite-dimensional spaces, *Bulletin of the American Mathematical Society* **27**, 217 (1992).

[20] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*, Vol. 2 (Springer-Verlag, Berlin, 1990).

[21] M. Wieck-Sosa and C. R. Shalizi, Estimating dynamic models by matching random features (2026), manuscript submitted for publication.

[22] M. Hénon, A two-dimensional mapping with a strange attractor, *Communications in Mathematical Physics* **50**, 69 (1976).

[23] E. N. Lorenz, Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences* **20**, 130–141 (1963).

[24] H. Whitney, Analytic extensions of differentiable functions defined in closed sets, *Transactions of the American Mathematical Society* **36**, 63 (1934).