

Predicting critical transitions in dynamical systems from time series using nonstationary probability density modeling

Frank Kwasniok*

College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom

(Received 29 July 2013; revised manuscript received 19 October 2013; published 25 November 2013)

A time series analysis method for predicting the probability density of a dynamical system is proposed. A nonstationary parametric model of the probability density is estimated from data within a maximum likelihood framework and then extrapolated to forecast the future probability density and explore the system for critical transitions or tipping points. A full systematic account of parameter uncertainty is taken. The technique is generic, independent of the underlying dynamics of the system. The method is verified on simulated data and then applied to prediction of Arctic sea-ice extent.

DOI: [10.1103/PhysRevE.88.052917](https://doi.org/10.1103/PhysRevE.88.052917)

PACS number(s): 05.45.Tp, 05.90.+m

I. INTRODUCTION

Many complex dynamical systems which are subject to slowly varying external conditions may exhibit critical transitions or tipping points; that is, they show a structural change in the observed macroscopic behavior when a critical threshold of a control parameter is crossed. Such tipping points are linked to bifurcations or phase transitions of the underlying dynamical system. At a critical value of a control parameter, an attractor becomes unstable and an alternative attractor emerges. Examples of real-world systems range from financial markets [1], over ecological systems [2], to the climate system [3]. Critical transitions in these systems may be of huge socioeconomic importance.

In recent years, there has been much interest in extracting early-warning signals of critical transitions from time series in order to detect, anticipate, or even predict impending tipping points [4]. Particular emphasis lies in generic indicators which do not require detailed knowledge of the underlying dynamics of the system. The techniques discussed so far are based on critical slowing down [5,6], detrended fluctuation analysis [7], increasing variance [8,9], multiple equilibria in the underlying potential or the quasistationary probability density [10,11], and (nonlinear) softening in the dynamics [12,13].

The present paper extends recent work on detection of critical transitions based on analysis of quasistationary probability densities and their modality [10,11]. In these earlier studies, the probability density is assumed to be stationary over a chosen data window. Here, a method is developed for estimating from time series data a nonstationary model of a slowly evolving probability density. This model is then extrapolated into the future, allowing for genuine prediction (rather than just detection or anticipation) of the nature and timing of future transitions in the probability density of the system. These transitions may be mere shifts (e.g., trends or cyclic variations in the mean or the variance) or may lead to bifurcations of the dynamical structure (e.g., changes in the number or stability of system states), including emergence of an alternative, previously unseen stable state.

The paper first outlines the general methodology. The technique is then verified on simulated data and applied to the prediction of Arctic sea-ice extent.

II. METHODOLOGY

A. Modeling the quasistationary probability density

We consider a high-dimensional, complex system which is subject to slow variations in its parameters. These variations may cause changes in the statistical characteristics of the emergent macroscopic behavior of the system. A scalar variable x of the system is observed. The quasistationary probability density $p(x;t)$ is introduced as the (marginal) stationary density or invariant measure the system adopts under constant parameters equal to those at time t . It is represented as

$$p(x;t) = Z^{-1}(t) \exp[-U(x;t)], \quad (1)$$

with a time-dependent normalization constant given by

$$Z(t) = \int_{-\infty}^{+\infty} \exp[-U(x;t)] dx. \quad (2)$$

The potential function $U(x;t)$ is approximated by a polynomial ansatz

$$U(x;t) = \sum_{i=1}^M a_i(t) x^i. \quad (3)$$

The degree of the polynomial M is an even integer in order to obtain a physically meaningful and normalizable stationary density. The low-order polynomial ansatz focuses the model on the large-scale features of the probability density. The time-dependent expansion coefficients $a_i(t)$ are themselves expanded in terms of prescribed time-dependent functions $f_{i,j}(t)$ as

$$a_i(t) = \sum_{j=0}^{J_i} \alpha_{i,j} f_{i,j}(t). \quad (4)$$

We always set $f_{i,0}(t) = 1$; thus the case $J_i = 0$ for all i corresponds to a time-invariant probability density [10,11,14]. For convenience of notation the constant part is not explicitly separated here from the time-dependent part. Natural choices

*F.Kwasniok@exeter.ac.uk

for the time-dependent functions $f_{i,j}(t)$ for $j > 0$ are trends (polynomial or other), cycles (represented by cosine and sine functions), or other external covariates whose time series is known.

B. Parameter estimation

Given a learning data set of length N , $\{x_1, \dots, x_N\}$, with $x_n = x(t_n)$ for $n = 1, \dots, N$, the model parameters are estimated according to the maximum likelihood principle. The time series may be evenly or unevenly sampled. We introduce the m th moment of the quasistationary probability density at time t_n given by

$$\mu_{m,n} = \frac{Z_{m,n}}{Z_{0,n}}, \quad (5)$$

with the auxiliary quantity

$$Z_{m,n} = \int_{-\infty}^{+\infty} x^m \exp[-U(x; t_n)] dx \quad (6)$$

for $m = 0, \dots, 2M$ and $n = 1, \dots, N$. We have $Z_{0,n} = Z(t_n)$. The likelihood function of the data is

$$L = \prod_{n=1}^N p(x_n; t_n); \quad (7)$$

the log-likelihood function is

$$\log L = \sum_{n=1}^N [-\log Z_{0,n} - U(x_n; t_n)]. \quad (8)$$

The natural logarithm is used throughout the paper. The log likelihood is maximized with respect to the parameters $\alpha_{i,j}$ by finding a stationary point of the gradient using the Newton-Raphson method. The gradient of the log likelihood is

$$\frac{\partial \log L}{\partial \alpha_{i,j}} = \sum_{n=1}^N f_{i,j,n} (\mu_{i,n} - x_n^i), \quad (9)$$

with the abbreviation $f_{i,j,n} = f_{i,j}(t_n)$; the elements of the Hessian matrix \mathbf{H} are

$$\frac{\partial^2 \log L}{\partial \alpha_{i,j} \partial \alpha_{k,l}} = \sum_{n=1}^N f_{i,j,n} f_{k,l,n} (\mu_{i,n} \mu_{k,n} - \mu_{i+k,n}). \quad (10)$$

The integrals $Z_{m,n}$ can be evaluated efficiently using standard numerical integration techniques on a large enough finite- x interval with relatively few mesh points as the integrand is rather smooth and decays rapidly for $|x| \rightarrow \infty$. We here use the (composite) Simpson rule.

Occasionally, the Newton-Raphson method is found to diverge if the chosen initial guess for the parameters is too far away from the solution. The following simple extension (see [15] for a similar algorithm) helped overcome divergence in the examples tested while keeping the rapid convergence of the Newton-Raphson method close to the solution: The step taken by the Newton-Raphson method at each iteration can be shown to be a descent direction of $\|\frac{\partial \log L}{\partial \alpha}\|^2$. If the proposed step in α is found not to decrease $\|\frac{\partial \log L}{\partial \alpha}\|^2$, the step length is halved until first encountering a decrease in $\|\frac{\partial \log L}{\partial \alpha}\|^2$ and that point taken as the next iterate.

For all examples reported in the present paper, the Hessian matrix \mathbf{H} of $\log L$ was found to be negative definite at all points in parameter space occurring as iterates during the maximization procedure. This hints at a possible strict convexity of the optimization problem, which would imply the existence of a unique and global maximum of the likelihood function. Convexity appears conceivable in the present context; unlike methods for estimating parameters in ordinary or stochastic differential equations, the present method does not construct any trajectories and thus does not involve the variation of the numerical integration of a nonlinear function with respect to the parameters. A general proof, though, appears not to be straightforward.

The method guarantees a positive leading-order polynomial coefficient at all data points [$a_M(t_n) > 0$], which is necessary for a normalizable probability density. No constraints need to be applied to achieve this as the likelihood function already acts as a barrier. For a given finite data range, a crossing below zero of the leading-order coefficient for some n would “leak” probability mass to infinity and thus decrease the likelihood on that data point to arbitrarily low values.

The present method, both in the time-independent case ($J_i = 0$ for all i) and the time-dependent case, is independent of the underlying dynamics of the system. For any fixed t , the quasistationary probability density of Eq. (1) is the stationary density of a one-dimensional Langevin equation describing noise-driven motion in the potential landscape $U(x; t)$ with a certain noise level. However, this one-dimensional dynamical model is not assumed to be valid here. The present method just empirically models the probability density of x and does not refer to any transition probabilities. It also does not assume a one-dimensional Fokker-Planck equation for the evolution of the probability density of x . The underlying dynamics may actually be high-dimensional and nongradient. As such, the technique is able to capture critical transition mechanisms beyond the one-dimensional saddle-node bifurcation scenario which is usually considered.

The method is clearly based on two time-scale assumptions. First, the parameter variation needs to be slow enough for an ensemble of system states to equilibrate instantaneously to the new quasistationary density to a good approximation. That means the solution of the time-dependent Fokker-Planck equation of the system initialized with its stationary solution at some point needs to always stay close to its stationary solution, at least for the marginal density of x . Second, the dynamics of the system needs to be such that a single trajectory explores state space fast enough; that is, the Fokker-Planck equation initialized with a point distribution representing a particular initial condition reaches its stationary solution quickly, at least for the marginal density of x . Only then can the data set be regarded as a representative sample of the quasistationary density at any time. Moreover, the likelihood function of Eq. (7) assumes independent data points. However, in most applications there are temporal correlations, depending on the dynamics of the system and the sampling interval of the data. Violation of any of these assumptions may result in biases in the parameter estimates.

In principle, the method is still applicable in the presence of observational noise. It would then model the observed probability density of x given as a convolution of the genuine

probability density of the system and that of the observational noise without explicitly disentangling the two components.

C. Model selection

The parameters M and J_i are hyperparameters of the method, controlling the complexity of the probability density model. Moreover, different choices for the functions $f_{i,j}(t)$ may be conceivable in a particular application. Natural methods for model comparison and selection are minimization of an information criterion or a likelihood ratio test when considering nested models. Here, the Bayesian information criterion is chosen as being equal to $p \log N - 2 \log L$, where p is the number of parameters in the model, N is the length of the time series, and L is the likelihood for the optimal parameter set. The Bayesian information criterion quite strongly penalizes additional parameters. This appears to be appropriate in the present context; in particular, when dealing with short and noisy time series, one would like to go for the most parsimonious model.

D. Parameter uncertainty

The maximum likelihood framework naturally provides information on parameter uncertainty and allows us to take full systematic account of it. For large enough N , the estimation errors of the parameters are Gaussian, and their covariance matrix is the inverse of the observed Fisher information matrix $\mathbf{I} = -\mathbf{H}(\boldsymbol{\alpha}^*)$, where $\boldsymbol{\alpha}^*$ denotes the parameter vector maximizing the likelihood. Uncertainty information for the probability density itself or diagnostics derived from it can then be obtained by Monte Carlo simulation drawing a large ensemble from the Gaussian parameter error distribution.

The Fisher information matrix is based on the assumption of independent data points. In the case of temporal correlation the uncertainty estimate derived from it may be expected to be a lower bound on the true uncertainty. Ideally, N should be replaced by an effective sample size which is yet hard to obtain in the present context.

E. Predicting the future probability density

Having estimated the probability density model from the learning data set, it can be extrapolated beyond the learning data window to predict the future evolution of the quasistationary probability density of the system. Given the functions $f_{i,j}(t)$ at some future time t , the probability density $p(x; t)$ is evaluated and uncertainty information is obtained as described above using the error covariance matrix of the parameters.

A positive leading-order polynomial coefficient, and thus a normalizable probability density, is no longer guaranteed when extrapolating beyond the learning data window. The leading-order coefficient (or the decay of the probability density for large $|x|$) should be monitored. The occurrence of a non-normalizable probability density clearly serves as an indicator that the model is inappropriately specified and/or the range of validity of the extrapolation is exceeded.

F. The special case of a time-invariant probability density

As a by-product, the time-independent case ($J_i = 0$ for all i) provides a method for obtaining a polynomial fit to the logarithm of a constant probability density in a given data window. It is an improvement on an earlier approach [10,14] which suggests a polynomial weighted least-squares fit to $-\log p(x)$, where $p(x)$ is obtained by a kernel density estimator. That method combines nonparametric and parametric elements in a slightly ad hoc manner and requires the specification of a smoothing bandwidth. The present likelihood-based method is more systematic, working directly on the given data points, and provides a natural framework for model selection (determining the degree M of the polynomial) and dealing with parameter uncertainty. Moreover, the likelihood approach guarantees a positive leading-order coefficient $\alpha_{M,0}$, which is not the case with the earlier method.

III. RESULTS

Various aspects of the method are illustrated and verified on a couple of simulated data sets. The technique is then applied to a real data set from the area of climate science.

A. Ornstein-Uhlenbeck process with time-dependent parameters

The first example system is an Ornstein-Uhlenbeck process with drifting parameters. The governing equation is

$$\dot{x} = -\gamma(t)x + c(t) + \sigma \eta. \tag{11}$$

η denotes a white Gaussian noise process with zero mean and unit variance. The standard deviation of the stochastic forcing is set as $\sigma = 2$. The system is considered on the time interval $[0, 200]$. The damping coefficient and the deterministic forcing have linear trends given by $\gamma(t) = 6 - \frac{t}{50}$ and $c(t) = \frac{t}{50}$. The exact quasistationary probability density of the system at time t is a Gaussian with mean $\frac{c(t)}{\gamma(t)}$ and variance $\frac{\sigma^2}{2\gamma(t)}$.

Figure 1(a) displays a sample trajectory of the system initialized with $x(0) = 0$; the upward drift of the mean and the increase in variance due to the decreasing damping are clearly visible. Here and in the following, the Euler-Maruyama scheme [16] with step size 10^{-4} was used to simulate stochastic differential equations. The learning data window extends from $t = 0$ to $t = 150$. The sampling interval is $\delta t = 0.05$, resulting in a learning data set of size $N = 3001$.

For the Ornstein-Uhlenbeck process the time-dependent Fokker-Planck equation can be easily solved, and it is thus possible to illustrate and check the validity of the time-scale assumptions of the method. For Gaussian initial densities the system can be fully described in terms of the mean $M = \langle x \rangle$ and the variance $V = \langle x^2 \rangle - \langle x \rangle^2$, where $\langle \cdot \rangle$ here denotes the average over an ensemble of system states. The evolution equations for the mean and the variance are $dM/dt = -\gamma(t)M + c(t)$ and $dV/dt = -2\gamma(t)V + \sigma^2$ [17]. They are decoupled, and we have linear relaxation towards the quasistationary values at a relaxation time scale γ^{-1} for the mean and $(2\gamma)^{-1}$ for the variance. As initial densities we consider an ensemble representing the quasistationary density at $t = 0$, that is, $M(0) = 0$ and $V(0) = 1/3$, as well as point distributions representing particular initial conditions, that is,

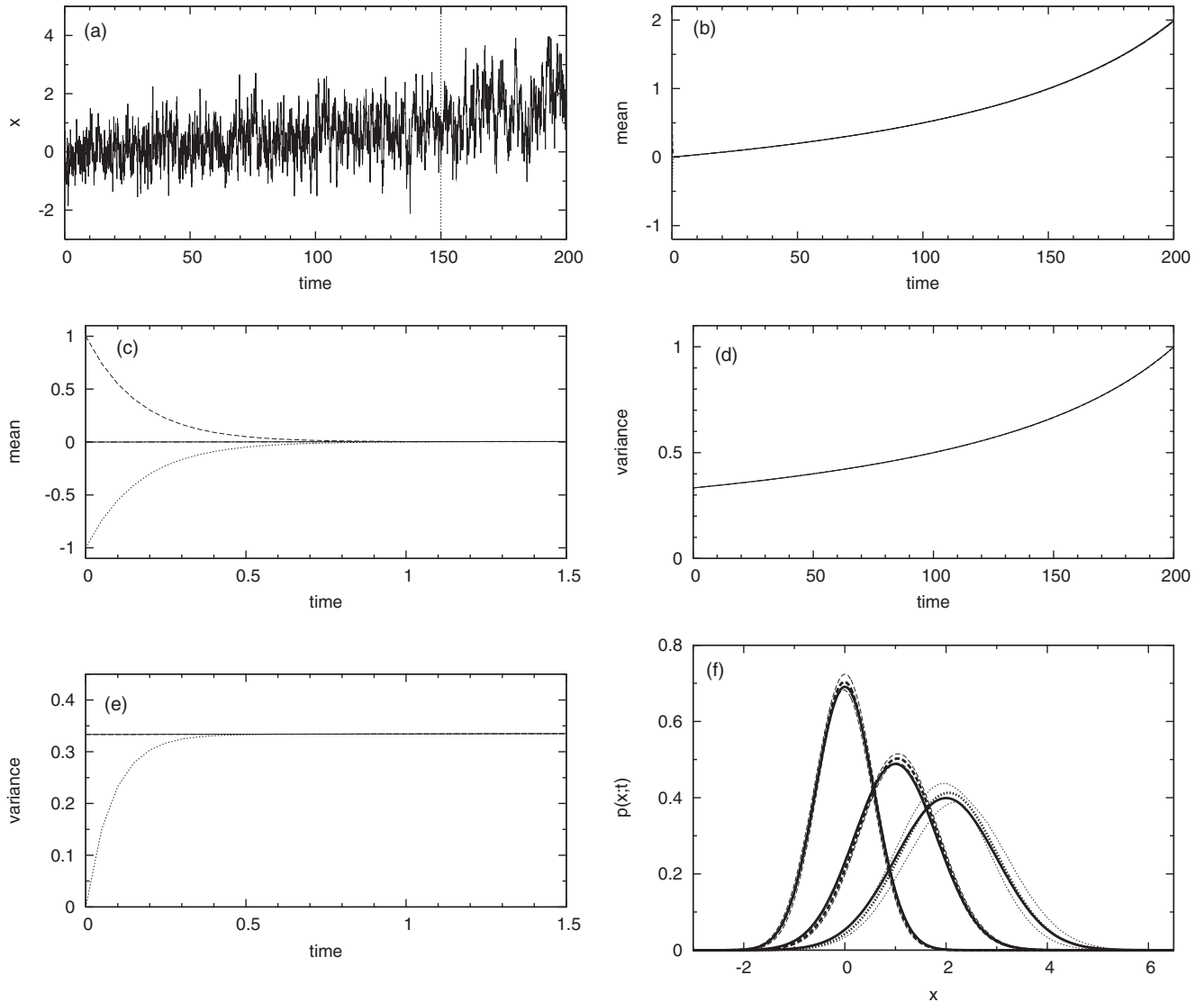


FIG. 1. Ornstein-Uhlenbeck process with time-dependent parameters. (a) Time series of the system. The dotted vertical line indicates the end of the learning data window. (b) Quasistationary mean (solid line) and time evolution of the mean for initial conditions $M(0) = 0$ (dot-dashed line), $M(0) = 1$ (dashed line), and $M(0) = -1$ (dotted line). (c) Close-up of (b). (d) Quasistationary variance (solid line) and time evolution of the variance for initial conditions $V(0) = 1/3$ (dashed line) and $V(0) = 0$ (dotted line). (e) Close-up of (d). (f) Evolution of the probability density: exact probability density at $t = 0$, $t = 150$, and $t = 200$ (solid lines, from left to right), reconstructed probability density at $t = 0$ and $t = 150$ (dashed lines), and predicted probability density at $t = 200$ (dotted line). The thin dashed and dotted lines indicate a 90% confidence interval.

$M(0) = x(0)$ and $V(0) = 0$, for the three example values, $x(0) = 0$, $x(0) = 1$, and $x(0) = -1$. Figures 1(b)–1(e) show the results of numerical integrations of the evolution equations of the mean and the variance for these cases. All initial conditions quickly asymptote to the stationary density, and the system almost instantaneously equilibrates to the changing quasistationary density. There is virtually no difference between the solution of the time-dependent Fokker-Planck equation and the quasistationary solution. This is due to the very short relaxation time scale of the system. Thus both time-scale assumptions of the method are extremely well satisfied here.

The potential function is chosen as a quadratic with linear trends in the coefficients: $U(x;t) = (\alpha_{1,0} + \alpha_{1,1}t)x + (\alpha_{2,0} + \alpha_{2,1}t)x^2$. An ensemble of 200 members is used to sample the uncertainty. Figure 1(f) shows the reconstructed probability density at the start and the end of the learning

data window as well as the predicted probability density at $t = 200$. They are calculated as the mean of the ensemble. The uncertainty is indicated by a central 90% confidence interval; that is, the 5% and 95% quantiles of the ensemble are given. Both the reconstructed probability densities and the forecast future probability density are quite accurate. Expectedly, the uncertainty is larger for out-of-sample prediction than for in-sample reconstruction.

B. Bifurcation in the underlying dynamics

We consider noise-driven motion in a nonstationary potential landscape which exhibits a bifurcation. The governing equation is

$$\dot{x} = -V'(x;t) + \sigma\eta. \quad (12)$$

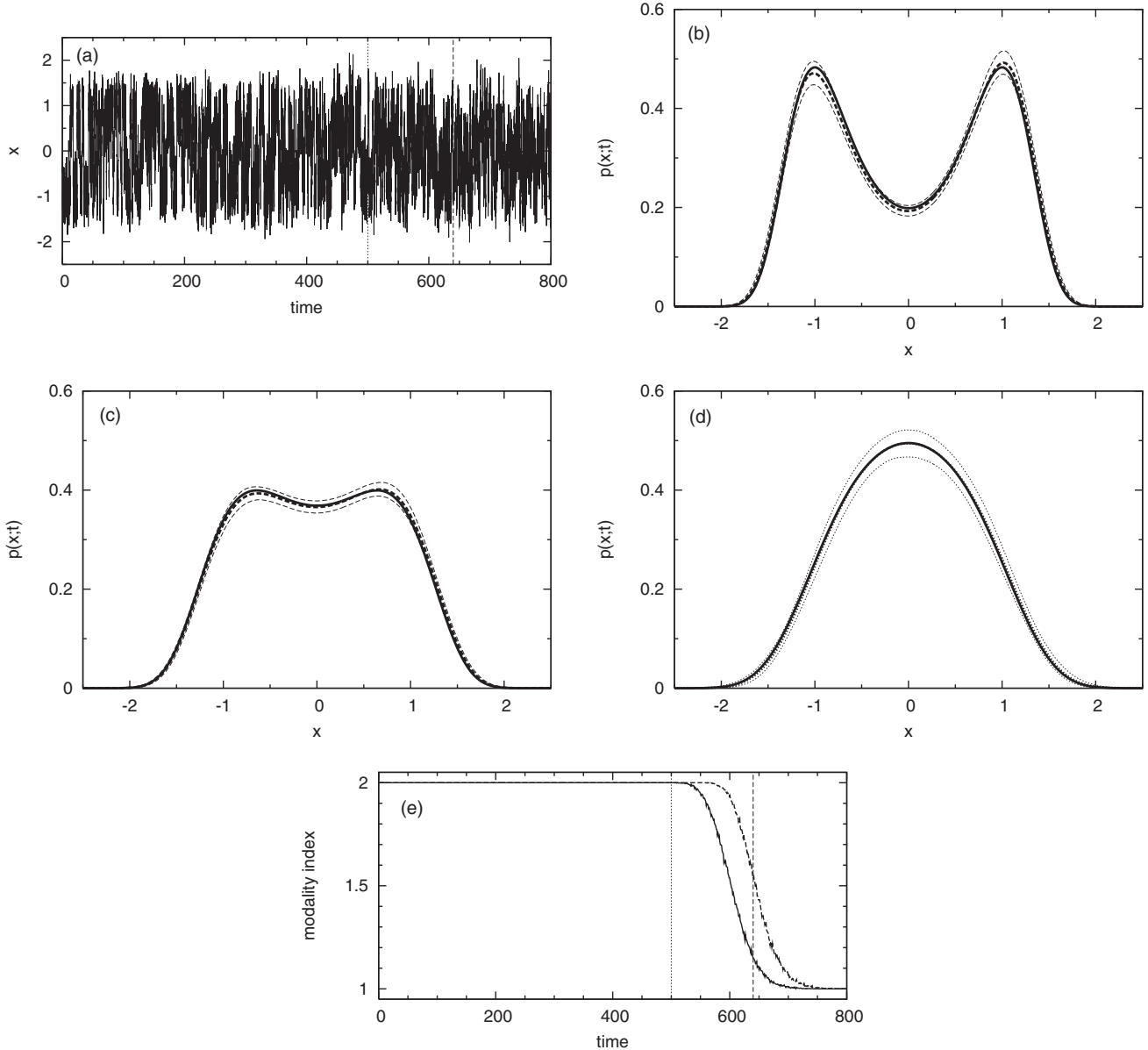


FIG. 2. Noise-driven motion in a potential landscape bifurcating from two wells to one well. (a) Time series of the system. The dotted vertical line indicates the end of the learning data window, and the dashed vertical line shows the bifurcation point. (b) Exact (solid line) and reconstructed (dashed line) probability density at $t = 0$. Thin dashed lines indicate a 90% confidence interval. (c) Exact (solid line) and reconstructed (dashed line) probability density at $t = 500$. Thin dashed lines indicate a 90% confidence interval. (d) Exact (solid line) and predicted (dotted line) probability density at $t = 800$. Thin dotted lines indicate a 90% confidence interval. (e) Modality index of the reconstructed or predicted quasistationary probability density as defined by local maxima (solid line) and pairs of inflection points (dashed line) as a function of time. The dotted vertical line indicates the end of the learning data window, and the dashed vertical line shows the bifurcation point.

η is a white Gaussian noise with zero mean and unit variance. The potential $V(x; t)$ slowly evolves from the double-well potential $V_1(x) = x^4 - 2x^2$ at initial time $t = 0$ to the single-well potential $V_2(x) = \frac{1}{4}x^4 + \frac{1}{2}x^2$ at final time $t = 800$ according to $V(x; t) = (1 - \frac{t}{800})V_1(x) + \frac{t}{800}V_2(x)$. The bifurcation from two stable states to one stable state occurs at $t = 640$. The noise standard deviation is $\sigma = 1.5$. The exact quasistationary probability density of the system at time t is known to be $p(x; t) \sim \exp[-2V(x; t)/\sigma^2]$.

Figure 2(a) displays a sample trajectory of the system initialized with $x(0) = -1$. The time series up to $t = 500$

forms the learning data set. The sampling interval is $\delta t = 0.05$; the length of the learning data set is $N = 10001$. The potential function is a fourth-order polynomial with linear trends in the coefficients: $U(x; t) = \sum_{i=1}^4 (\alpha_{i,0} + \alpha_{i,1}t)x^i$. An ensemble of 1000 realizations drawn from the parameter error distribution is used. Figures 2(b), 2(c), and 2(d) show the reconstructed probability density at the start and the end of the learning data window as well as the probability density predicted beyond the bifurcation point at $t = 800$, together with their uncertainties. Both the reconstructed and predicted probability densities are very accurate.

In order to diagnose the evolution of the modality of the probability density we introduce two different notions of the number of modes (or states) in the quasistationary probability density $p(x;t)$ at some time t . A straightforward definition of the number of modes is given by the number of local maxima in the probability density (or local minima in the potential function). We denote this integer number by S_1 . A slightly wider definition is based on inflection points in the probability density. The curvature of $p(x;t)$ is $p''(x;t) = Z^{-1}(t)[U'^2(x;t) - U''(x;t)] \exp[-U(x;t)]$. It is positive at $\pm\infty$; inflection points occur in pairs. We count one mode (or state) for each pair of inflection points and denote this integer number by S_2 . Put differently, S_2 is the number of disjoint intervals on which $p(x;t)$ has negative curvature, but not necessarily a local maximum. This definition of the number of states is different from that in [10,11], which is based on inflection points of the potential function rather than the probability density. We then define a modality index as the fractional number given by $m_1 = \langle S_1 \rangle$ or $m_2 = \langle S_2 \rangle$, where $\langle \cdot \rangle$ denotes the mean over the ensemble of realizations drawn from the parameter error distribution. We always have $S_1 \leq S_2$ and thus $m_1 \leq m_2$.

Figure 2(e) displays the modality index as a function of time. The exact probability density has $S_1 = S_2 = 2$ for $0 \leq t < 640$ and $S_1 = S_2 = 1$ for $640 \leq t \leq 800$. The system is correctly reconstructed as bimodal in the learning data window. Then both modality indices indicate the impending bifurcation and eventually drop to one after the bifurcation point. The timing of the bifurcation is somewhat better captured with the inflection point definition of states.

C. Transition to a previously unseen state

Again, stochastically driven motion in a nonstationary potential is considered, governed by the equation

$$\dot{x} = -V'(x;t) + \sigma\eta, \quad (13)$$

where η is a Gaussian white noise process with mean zero and unit variance. The system is considered on the time interval $[0,1000]$. The potential is given by $V(x;t) = \frac{1}{4}x^4 - \frac{1}{2}x^2 - \lambda(t)x$, with $\lambda(t) = \frac{t}{900}\lambda_0$ and $\lambda_0 = 2\sqrt{3}/9$. At $t = 0$, this is a symmetric two-well potential with local minima at $x = -1$ and $x = 1$. Then one well gradually shallows and the other deepens until at $\lambda = \lambda_0$; that is, at $t = 900$ a bifurcation from double well to single well occurs. The noise level is chosen as small as $\sigma = 0.1$. The system is then virtually confined for a long time to the well it initially started in and only has a considerable probability of switching to the other state when approaching the bifurcation point.

Figure 3(a) shows a sample trajectory of the system initialized with $x(0) = -1$; here, we observe a noise-induced early escape to the other state at about $t = 833$. The time interval $[0,700]$ is used as the learning data window. The sampling interval is $\delta t = 0.05$; the length of the learning data set is $N = 14001$. The potential function is a quadratic with linear trends in the coefficients: $U(x;t) = (\alpha_{1,0} + \alpha_{1,1}t)x + (\alpha_{2,0} + \alpha_{2,1}t)x^2$. An ensemble of size 200 is drawn from the parameter error distribution to assess the uncertainty.

Figure 3(b) displays the reconstructed empirical probability density of the time series at the start and at the end of the

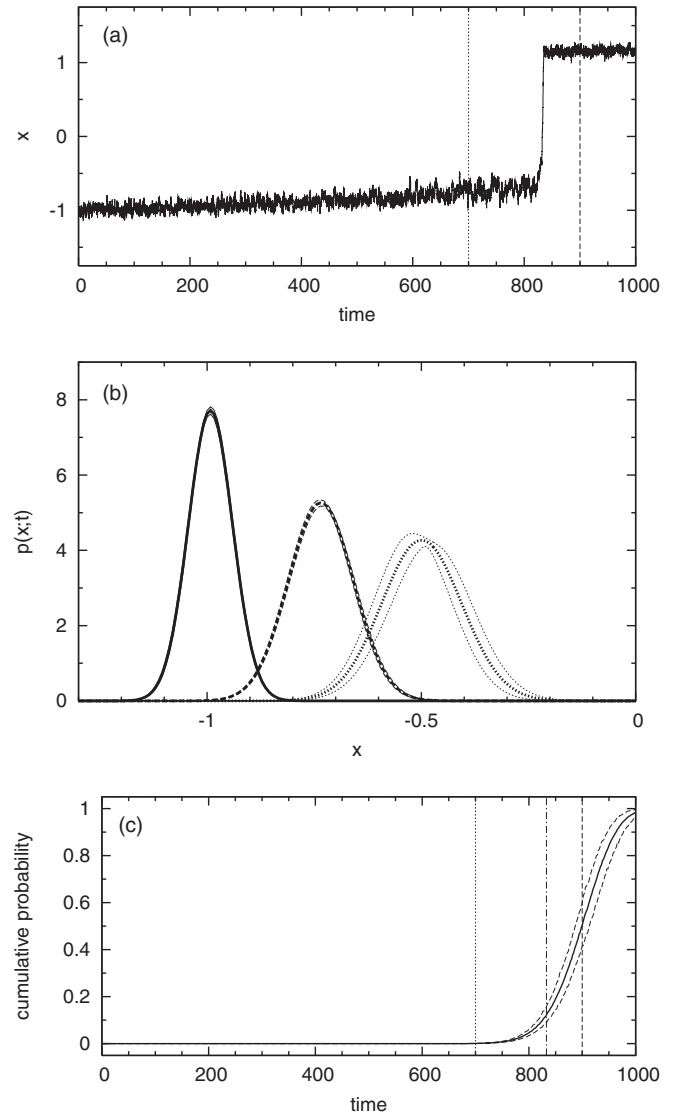


FIG. 3. Noise-driven motion in a potential landscape with transition to a previously unseen state. (a) Time series of the system. The dotted vertical line indicates the end of the learning data window, and the dashed vertical line shows the bifurcation point. (b) Reconstructed empirical probability density of the time series at $t = 0$ (solid line) and $t = 700$ (dashed line) as well as the extrapolated probability density at $t = 900$ (dotted line). The corresponding thin lines indicate a 90% confidence interval. (c) Cumulative probability $\Pr(x > -0.5)$ calculated from the reconstructed or predicted probability density as a function of time (solid line) with a 90% confidence interval (dashed lines). The dotted vertical line indicates the end of the learning data window, the dot-dashed vertical line shows the time of noise-induced early escape, and the dashed vertical line shows the bifurcation point.

learning data set as well as an extrapolation to the bifurcation point. It should be noted that these probability densities are filtered by the system trajectory and are confined to the part of state space the time series actually has explored. The second time-scale assumption discussed above is violated here. The true quasistationary probability density of the system (as sampled by an infinitely long trajectory) at $t = 0$ and $t = 700$ is bimodal with increasing probability mass around $x = 1$. At the bifurcation point, the exact probability density of the system

has a single maximum with most of the probability mass around $x = 1$. The method cannot be expected to reconstruct or predict these as it cannot learn about the unseen state. What the evolution of the empirical probability density of the time series does show is the shift in the mean and the increasing variance due to the softening of the dynamics when approaching the bifurcation point. The latter is a generic feature of a critical transition [8,9] and is picked up by the method well ahead of the bifurcation point, and a further softening is extrapolated towards the bifurcation point.

Figure 3(c) displays the reconstructed or predicted cumulative probability for $x > -0.5$. It is virtually zero in the learning data window and is projected to increase rapidly towards the bifurcation point, indicating an increasing probability of the system to move away from its current state. The cumulative probability $\Pr(x > -0.5)$ is a diagnostic of the probability density, indicating its changing shape; it cannot be easily interpreted as or related to an actual escape probability of the trajectory as the dynamics of the system is not modeled here.

The analysis can be refined in several ways. Higher-order (e.g., quadratic) trends may be important, particularly close to the bifurcation point. A fourth-order polynomial potential function allows for simultaneous analysis of also skewness [18] and kurtosis and their time evolution. In particular, the sign and time evolution of the skewness would give an indication on which side of the current state the new state will emerge.

D. Gaussian process with cyclic variation and drift

The system here is a Gaussian process

$$x(t) \sim \mathcal{N}(\mu(t), \sigma^2), \tag{14}$$

with time-dependent mean $\mu(t)$. The mean exhibits a linear trend and a cyclic variation with period 1, whose amplitude again has a linear trend: $\mu(t) = \frac{t}{40} + (1 + \frac{t}{40}) \sin(2\pi t + \frac{\pi}{4})$. The process is considered on the time interval $[0,40]$. The standard deviation is $\sigma = 0.2$.

Figure 4(a) shows a sample time series of the system. The first half of the time series up to $t = 20$ is used as a learning

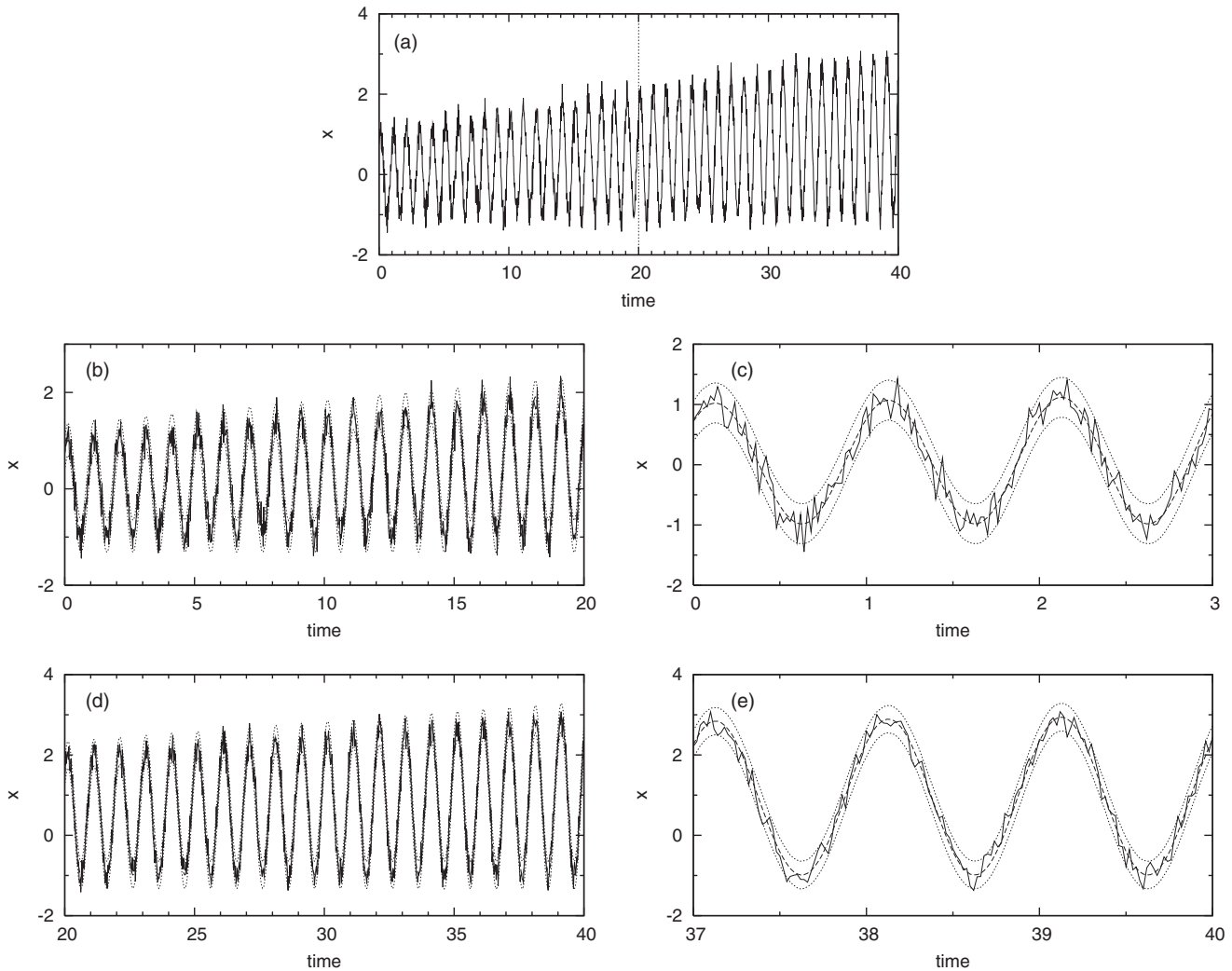


FIG. 4. Gaussian process with cyclic variation and drift. (a) Time series of the system. The dotted vertical line indicates the end of the learning data window. (b) In-sample prediction: observed time series (solid line), estimate of the mean state (dashed line), and 90% confidence interval for the data points (dotted lines). (c) Close-up of (b). (d) Out-of-sample prediction: observed time series (solid line), estimate of the mean state (dashed line), and 90% confidence interval for the data points (dotted lines). (e) Close-up of (d).

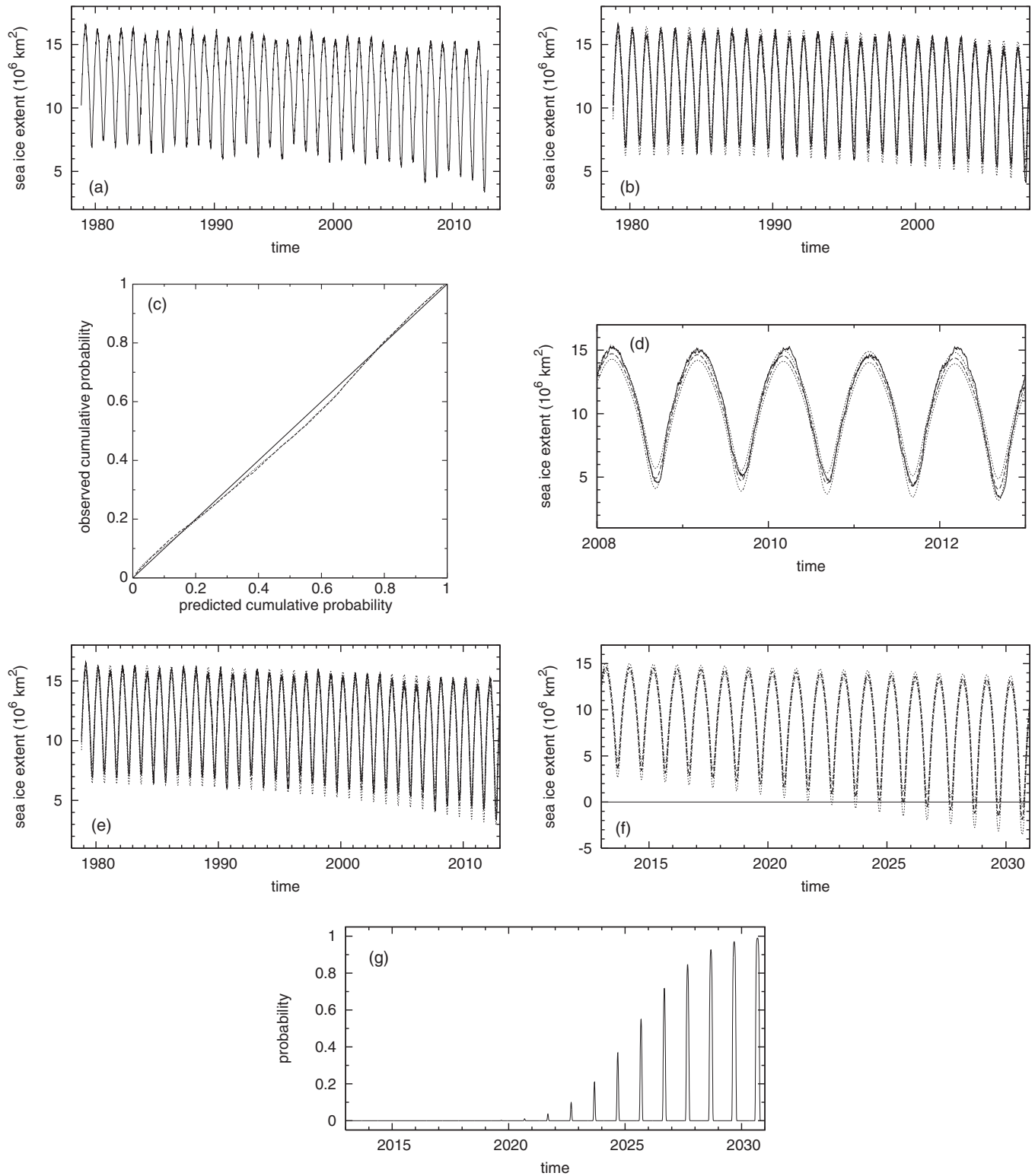


FIG. 5. Prediction of Arctic sea-ice extent. (a) Time series of Arctic sea-ice extent from 26 October 1978 to 31 December 2012. (b) In-sample prediction in the learning data set from 1978 to 2007: observed time series (solid line), estimate of the mean state (dashed line), and 90% confidence interval for the data points (dotted lines). (c) Probability plot for in-sample prediction in the learning data set from 1978 to 2007 (dashed line) and from 1978 to 2012 (dotted line). (d) Out-of-sample prediction for the period 2008–2012: observed time series (solid line), estimate of the mean state (dashed line), and 90% confidence interval for the data points (dotted lines). (e) In-sample prediction for the learning data set from 1978 to 2012: observed time series (solid line), estimate of the mean state (dashed line), and 90% confidence interval for the data points (dotted lines). (f) Prediction of future sea-ice extent: estimate of the mean state (dashed line) and 90% confidence interval for the data points (dotted lines). (g) Probability of future sea-ice loss.

data set. The sampling interval is $\delta t = 0.02$, resulting in a learning data set of length $N = 1001$. An ensemble of size 200 drawn from the parameter error distribution is used to describe the uncertainty. The potential function to capture all the features of the system is $U(x; t) = [\alpha_{1,0} + \alpha_{1,1}t + (\alpha_{1,2} + \alpha_{1,3}t) \cos 2\pi t + (\alpha_{1,4} + \alpha_{1,5}t) \sin 2\pi t]x + \alpha_{2,0}x^2$.

In-sample prediction within the learning data set [Figs. 4(b) and 4(c)] and out-of-sample prediction in the unseen second half of the data set [Figs. 4(d) and 4(e)] are performed. The mean state and a central 90% confidence interval for the individual data points (given by the 5% and 95% quantiles) are forecast. The uncertainty intervals comprise the inherent stochasticity of the process as well as the estimation uncertainty of all the parameters. Both in-sample and out-of-sample predictions are very accurate; the mean state and the uncertainty are faithfully captured.

E. Predicting Arctic sea-ice extent

As an application to a real-world data set, the method is used to predict Arctic sea-ice extent. The Arctic summer sea ice and its possible loss in the future have been identified as one of several potential tipping elements in the Earth's climate system [3].

Figure 5(a) displays the time series of Arctic sea-ice extent from 26 October 1978 to 31 December 2012. It is provided by the National Snow and Ice Data Centre (NSIDC), derived from satellite data sources [19]. Observations were recorded every second day until 9 July 1987 and daily thereafter. The time series has $N = 10\,856$ data points.

The data are characterized by an annual cycle with increasing amplitude as well as an overall downward trend. The annual cycle is pronouncedly nonsinusoidal; this is due to the strongly nonlinear threshold processes involved in the freezing and melting of sea ice. The fluctuations around the annual cycle are non-Gaussian; they are actually strongly leptokurtic. These features are taken into account by a fourth-order potential function, a quadratic overall trend, and a representation of the annual cycle using higher harmonics with a linear trend in some of them: $U(x; t) = [\alpha_{1,0} + \alpha_{1,1}t + \alpha_{1,2}t^2 + (\alpha_{1,3} + \alpha_{1,4}t) \cos 2\pi t + (\alpha_{1,5} + \alpha_{1,6}t) \sin 2\pi t + (\alpha_{1,7} + \alpha_{1,8}t) \cos 4\pi t + (\alpha_{1,9} + \alpha_{1,10}t) \sin 4\pi t + \alpha_{1,11} \cos 6\pi t + \alpha_{1,12} \sin 6\pi t]x + \alpha_{2,0}x^2 + \alpha_{3,0}x^3 + \alpha_{4,0}x^4$. An ensemble of size 200 is used to sample the uncertainty.

In order to assess the performance of the method within the time period covered by the data set we first withhold the last five years of data. Only the period from 26 October 1978 to 31 December 2007 is used as a learning data set, resulting

in $N = 9029$ data points. Figure 5(b) illustrates in-sample prediction. The predicted mean and central 90% confidence interval for the data points are given together with the actual time series. Both the mean prediction and the uncertainty are quite accurate. The uncertainty is larger in summer than in winter. In order to assess the accuracy of the reconstructed probability density in more detail a probability plot is shown in Figure 5(c). The predicted cumulative probability is plotted against the relative frequency of observing a data point below the predicted quantile. The points should ideally lie on the unit diagonal. The probability density turns out to be very reliable across the whole range of quantiles. Figure 5(d) displays out-of-sample prediction for the withheld time period 2008–2012. The accelerated downward trend and increase in the amplitude of the annual cycle are well captured. The actual data points lie well within the predicted uncertainty interval, except for some anomalies.

For prediction of future sea-ice extent the model is learned anew from the whole available data set. Figure 5(e) shows the in-sample performance of the model. Again, the model describes the learning data set very well, along with the newly included five years with the accelerated trends. The probability density is again reliable over the whole range of quantiles [Fig. 5(c)]. Figure 5(f) displays the results for prediction of future sea-ice extent. Figure 5(g) gives the cumulative probability mass put on values below zero, which can be naturally interpreted as the probability of complete sea-ice loss. The model projects the complete loss of the Arctic summer sea ice during the 2020s. The predicted mean is below zero for the first time in 2025. There is a substantial probability of complete summer sea ice loss from the early 2020s; it approaches 1 towards the end of the 2020s. The uncertainty of the prediction increases further into the future.

IV. CONCLUSIONS

A method for predicting the probability density of a dynamical system from time series data has been derived. It is based on estimating a nonstationary probability density model on a learning data window and then extrapolating it into the future. The predicted future density can be used to forecast critical transition phenomena of interest. The technique has been verified on simulated data with different shift, bifurcation, and tipping scenarios. The method was then applied to prediction of Arctic sea ice extent, projecting the complete loss of summer sea ice during the 2020s. The discussed approach should be useful in various scientific areas where critical transitions play a role.

[1] D. Sornette and A. Johansen, *Phys. A* **245**, 411 (1997).
 [2] M. Scheffer, S. R. Carpenter, J. A. Foley, C. Folke, and B. Walker, *Nature (London)* **413**, 591 (2001).
 [3] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber, *Proc. Natl. Acad. Sci. USA* **105**, 1786 (2008).
 [4] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara, *Nature (London)* **461**, 53 (2009).

[5] H. Held and T. Kleinen, *Geophys. Res. Lett.* **31**, L23207 (2004).
 [6] V. Dakos, M. Scheffer, E. H. van Nes, V. Brovkin, V. Petoukhov, and H. Held, *Proc. Natl. Acad. Sci. USA* **105**, 14308 (2008).
 [7] V. N. Livina and T. M. Lenton, *Geophys. Res. Lett.* **34**, L03712 (2007).
 [8] S. R. Carpenter and W. A. Brock, *Ecol. Lett.* **9**, 311 (2006).
 [9] P. D. Ditlevsen and S. J. Johnsen, *Geophys. Res. Lett.* **37**, L19703 (2010).

- [10] V. N. Livina, F. Kwasniok, and T. M. Lenton, *Clim. Past* **6**, 77 (2010).
- [11] V. N. Livina, F. Kwasniok, G. Lohmann, J. W. Kantelhardt, and T. M. Lenton, *Clim. Dyn.* **37**, 2437 (2011).
- [12] J. M. T. Thompson and J. Sieber, *Int. J. Bifurcation Chaos* **21**, 399 (2011).
- [13] J. Sieber and J. M. T. Thompson, *Philos. Trans. R. Soc. A* **370**, 1166 (2012).
- [14] F. Kwasniok and G. Lohmann, *Phys. Rev. E* **80**, 066104 (2009).
- [15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, 3rd ed. (Cambridge University Press, Cambridge, 2007).
- [16] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin, 1999).
- [17] C. Gardiner, *Stochastic Methods*, 4th ed. (Springer, Berlin, 2010).
- [18] V. Guttal and C. Jayaprakash, *Ecol. Lett.* **11**, 450 (2008).
- [19] The data are available at <ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/north/daily/data/>