Eur. Phys. J. Special Topics **165**, 35–44 (2008)

DOI: 10.1140/epjst/e2008-00847-y

THE EUROPEAN PHYSICAL JOURNAL
SPECIAL TOPICS

Regime prediction and predictability in nonlinear dynamical systems

F. Kwasniok[a]

School of Engineering, Computing and Mathematics, University of Exeter, North Park Road, Exeter EX4 4QF, UK

Abstract. Prediction and predictability properties of nonlinear dynamical systems are diagnosed and analysed empirically using nonlinear time series analysis techniques. The notion of predictability is relaxed from accurate prediction of individual trajectories to a coarse-grained view in which only probabilities of visiting certain regions of state space or regimes are forecast. The regimes and the transition probabilities between them are determined simultaneously by fitting a hidden Markov model to a time series of the system. Predictive information is then refined by building a nearest-neighbour model of the regime posterior distribution. The ideas are exemplified on the stochastically forced Lorenz system.

1 Introduction

Predicting the future time evolution of a dynamical system is an important task in many areas of science. Often the underlying dynamical equations are unknown and only a time series of the system is available. There are well-established methods for constructing statistical models from the observations [1–4]; possible approaches include local constant or analogue models, local polynomial models, radial basis functions and neural networks.

Sometimes a dynamical system is dominated on a coarse-grained scale by the switching between different areas of state space where the trajectory spends some time. Such regime behaviour may be already obvious by eye in a time series [5–7] or may be more subtle [8,9]. This leads to the notion of regime prediction where only probabilities of visiting certain regions of state space are forecast rather than individual trajectories. The perspective of regime prediction may be particularly adequate when determinism and thus predictability is weak, that is, for systems with a considerable stochastic component. Regime predictability then focuses on and condenses the little predictability available at all in the system. Moreover, it is presumably most interesting at longer prediction times. Regime predictability may be still present in a dynamical system at prediction timescales where predictability of individual trajectories measured by mean square errors or correlations is already lost.

The present approach combines the well-known methodologies of hidden Markov models [10] and nearest-neighbour models [4] to build statistical predictive models from data. Firstly, the most predictable coarse-grained structures or regimes are identified by fitting a hidden Markov model to a time series of the dynamical system under consideration. Secondly, a nearest-neighbour model of the regime posterior distribution is used to predict the regime posterior distribution at future time.

The paper is organised as follows: In section 2, the model system used as an example in the present study is introduced. Then the prediction methodology is outlined in detail in section 3.

[a] e-mail: F.Kwasniok@exeter.ac.uk

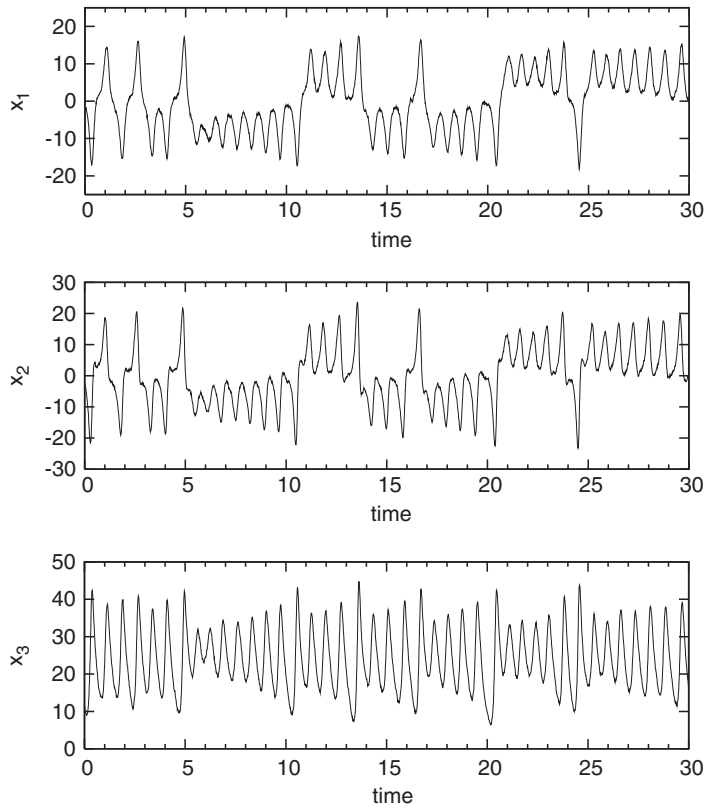

Fig. 1. Sample time series of the x_1-, x_2- and x_3-component of the Lorenz system with white Gaussian noise.

The results are presented in section 4; the paper closed with a brief discussion of the results and their implications in section 5.

2 The stochastically forced Lorenz system

The classical Lorenz model [5] augmented by white stochastic forcing [11] is used here as an example system:

$$\dot{x}_1 = -sx_1 + sx_2 + \sigma\eta_1 \tag{1}$$
$$\dot{x}_2 = -x_1x_3 + rx_1 - x_2 + \sigma\eta_2 \tag{2}$$
$$\dot{x}_3 = x_1x_2 - bx_3 + \sigma\eta_3 \tag{3}$$

where η_1, η_2 and η_3 denote pairwise independent white Gaussian noises with zero mean and unit variance: $\langle\eta_i\rangle = 0$ and $\langle\eta_i\eta_j\rangle = \delta_{ij}$. The standard parameter set $s = 10$, $r = 28$ and $b = 8/3$ is chosen. The noise standard deviation is $\sigma = 2$. The system is integrated in time numerically using the Euler forward scheme [12] with step size $h = 10^{-5}$. Reducing the step size by a factor of 10 does not change the statistical properties of the system. Figure 1 displays a sample trajectory of the model; the sampling interval in the graphs is 0.01. The noise level is moderate; shape and character of the trajectories are still close to those of the deterministic Lorenz system. The trajectory evolves around an attractor with two butterfly-wing-shaped lobes; it switches irregularly between the two lobes. On each of the wings, the state vector spirals around the unstable steady states at $(-6\sqrt{2}, -6\sqrt{2}, 27)$ and $(6\sqrt{2}, 6\sqrt{2}, 27)$, respectively.

3 Methodology

3.1 The hidden Markov model

Hidden Markov models (HMMs) [10] are a flexible and versatile tool for time series modelling, prediction and classification. In the following, the HMM is described in the form in which it is used in the present study.

3.1.1 Model setting

We consider an equally sampled data set of length N, $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, from the (deterministic or stochastic) dynamical system under consideration. Each data point is a vector $\mathbf{x} = (x_1, \ldots, x_m)^{\mathrm{T}}$ in m-dimensional state space, dropping the temporal index for convenience. It may be either measured data or data from a numerical simulation. The vector $\mathbf{x}$ may represent the original variables of the state space of the dynamical system (complete observation) or be obtained by a (linear or nonlinear) projection from a higher-dimensional state space (partial observation). In the latter case, it may be advisable to introduce additional variables using a time-delay embedding [13] in order to recover some of the information lost in the projection. The m-dimensional vector $\mathbf{x}$ is assumed to already include these. In the case of the Lorenz system, we have complete observation with $m = 3$ and $\mathbf{x} = (x_1, x_2, x_3)^{\mathrm{T}}$.

A discrete HMM with Gaussian output density is used to model the system in a coarse-grained sense. The HMM has a finite number of *internal* or *hidden states* $\{1, \ldots, K\}$ which are not observable directly. Each state is associated with a time-independent Gaussian output density into m-dimensional state space characterized by a mean $\boldsymbol{\mu}_i$ and a covariance matrix $\boldsymbol{\Gamma}_i$. The probability of observing the data vector $\mathbf{x}$ given the internal state $q = i$ is

$$g_i(\mathbf{x}) = p(\mathbf{x}|q = i) = \frac{1}{(2\pi)^{\frac{m}{2}} (\det \boldsymbol{\Gamma}_i)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \boldsymbol{\Gamma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)\right]. \tag{4}$$

This defines a probabilistic partitioning of state space; the internal states correspond to the *regimes* of the dynamical system and are referred to as such. Transitions among the regimes are governed by a time-independent matrix of transition probabilities

$$A_{ij} = p(q_{n+1} = j|q_n = i), \qquad i, j = 1, \ldots, K. \tag{5}$$

The initial probability distribution is

$$\theta_i = p(q_1 = i), \qquad i = 1, \ldots, K. \tag{6}$$

The parameters of the model are the means $\{\boldsymbol{\mu}_i\}_{i=1}^K$, the covariance matrices $\{\boldsymbol{\Gamma}_i\}_{i=1}^K$, the transition matrix $\mathbf{A}$ and the initial distribution $\boldsymbol{\theta}$. The number of regimes K is the only hyperparameter in the model which has to be fixed a priori. It reflects the degree of coarse-graining of state space.

It is not clear whether regime transitions in a nonlinear deterministic or stochastic system are actually Markovian. Coarse-graining and, probably more severely, projection onto a lower-dimensional space (in case of only partial observation of the system) are possible sources of non-Markovianity. Nevertheless, a first-order Markov model appears to be a reasonable working approximation. A short discussion on how to test this approximation is given in section 3.1.3.

3.1.2 Parameter estimation

Given an equally sampled time series of length N, $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, the parameters of the HMM are estimated according to the maximum likelihood principle using the *expectation-maximization* (EM) algorithm [14] which in the context of HMMs is the *Baum-Welch* or *forward-backward*

algorithm. Two auxiliary variables are introduced. The *forward variable* $\alpha_{i,n}$ is the probability of observing the partial sequence $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ together with the regime $q_n = i$:

$$\alpha_{i,n} = p(\mathbf{x}_1, \ldots, \mathbf{x}_n, q_n = i), \quad i = 1, \ldots, K; \quad n = 1, \ldots, N. \tag{7}$$

The forward variable satisfies the recursion relation

$$\alpha_{i,n+1} = g_i(\mathbf{x}_{n+1}) \sum_{j=1}^{K} \alpha_{j,n} A_{ji}, \quad i = 1, \ldots, K; \quad n = 1, \ldots, N-1 \tag{8}$$

with initialisation

$$\alpha_{i,1} = \theta_i g_i(\mathbf{x}_1), \quad i = 1, \ldots, K. \tag{9}$$

In a similar way the *backward variable* $\beta_{i,n}$ is introduced as the conditional probability of observing the partial sequence $\{\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N\}$ given the regime $q_n = i$:

$$\beta_{i,n} = p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | q_n = i), \qquad i = 1, \ldots, K; \quad n = 1, \ldots, N-1. \tag{10}$$

The backward variable obeys the recursion formula

$$\beta_{i,n} = \sum_{j=1}^{K} A_{ij} g_j(\mathbf{x}_{n+1}) \beta_{j,n+1}, \qquad i = 1, \ldots, K; \quad n = 1, \ldots, N-1 \tag{11}$$

together with the definition

$$\beta_{i,N} = 1, \qquad i = 1, \ldots, K. \tag{12}$$

Two more useful probabilities are introduced which can be expressed in terms of the forward and backward variables: the probability $\gamma_{ij,n}$ of having the successive pair of regimes $q_n = i$ and $q_{n+1} = j$ given all the data

$$\gamma_{ij,n} = p(q_n = i, q_{n+1} = j | \mathbf{x}_1, \ldots, \mathbf{x}_N) = \frac{\alpha_{i,n} A_{ij} g_j(\mathbf{x}_{n+1}) \beta_{j,n+1}}{\sum_{k=1}^{K} \sum_{l=1}^{K} \alpha_{k,n} A_{kl} g_l(\mathbf{x}_{n+1}) \beta_{l,n+1}},$$
$$i, j = 1, \ldots, K; \qquad n = 1, \ldots, N-1 \tag{13}$$

and the posterior probability of having the regime $q_n = i$ given all the data

$$\delta_{i,n} = p(q_n = i | \mathbf{x}_1, \ldots, \mathbf{x}_N) = \frac{\alpha_{i,n} \beta_{i,n}}{\sum_{j=1}^{K} \alpha_{j,n} \beta_{j,n}}, \qquad i = 1, \ldots, K; \qquad n = 1, \ldots, N. \tag{14}$$

The likelihood function of the data is

$$L = p(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \sum_{i=1}^{K} \alpha_{i,N} = \sum_{i=1}^{K} \alpha_{i,n} \beta_{i,n}, \qquad n = 1, \ldots, N. \tag{15}$$

The EM algorithm is iterative. The expectation step of the rth iteration is the calculation of $\alpha_{i,n}^{(r)}$ and $\beta_{i,n}^{(r)}$ from the current parameter estimates according to formulae (8), (9), (11) and (12) and then the calculation of $\gamma_{ij,n}^{(r)}$ and $\delta_{i,n}^{(r)}$ according to equations (13) and (14). The maximization step consists in updating the parameters according to the reestimation formulae

$$\boldsymbol{\mu}_i^{(r+1)} = \frac{\sum_{n=1}^{N} \mathbf{x}_n \delta_{i,n}^{(r)}}{\sum_{n=1}^{N} \delta_{i,n}^{(r)}}, \qquad i = 1, \ldots, K \tag{16}$$

$$\mathbf{\Gamma}_i^{(r+1)} = \frac{\sum_{n=1}^{N} \left(\mathbf{x}_n - \boldsymbol{\mu}_i^{(r+1)}\right)\left(\mathbf{x}_n - \boldsymbol{\mu}_i^{(r+1)}\right)^{\mathrm{T}} \delta_{i,n}^{(r)}}{\sum_{n=1}^{N} \delta_{i,n}^{(r)}}, \qquad i = 1, \ldots, K \tag{17}$$

$$A_{ij}^{(r+1)} = \frac{\sum_{n=1}^{N-1} \gamma_{ij,n}^{(r)}}{\sum_{n=1}^{N-1} \delta_{i,n}^{(r)}}, \qquad i, j = 1, \ldots, K \tag{18}$$

$$\theta_i^{(r+1)} = \delta_{i,1}^{(r)}, \qquad i = 1, \ldots, K. \tag{19}$$

Initial guesses for the parameters $\{\boldsymbol{\mu}_i^{(0)}\}_{i=1}^{K}$, $\{\mathbf{\Gamma}_i^{(0)}\}_{i=1}^{K}$, $\mathbf{A}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$ need to be provided to start the iteration. The algorithm is monotonically non-decreasing in likelihood and converges to a maximum of the likelihood function L.

3.1.3 Regime prediction and predictability

The mean posterior probability of observing the regime $q_n = i$ is introduced:

$$w_i = \frac{1}{N} \sum_{n=1}^{N} \delta_{i,n}, \qquad i = 1, \ldots, K. \tag{20}$$

The vector $\mathbf{w} = (w_1, \ldots, w_K)$ corresponds to the weights of the individual components in an ordinary Gaussian mixture model [15]. Given only a single data point $\mathbf{x}$, the posterior probability of the system being in regime $q = i$ is

$$P_i = p(q = i|\mathbf{x}) = \frac{w_i g_i(\mathbf{x})}{\sum_{j=1}^{K} w_j g_j(\mathbf{x})}, \qquad i = 1, \ldots, K. \tag{21}$$

For $N = 1$ and the choice $\boldsymbol{\theta} = \mathbf{w}$, one has $P_i = \delta_{i,1}$. $\mathbf{P} = (P_1, \ldots, P_K)$ is the posterior obtained from a static Gaussian mixture model but using the regimes obtained from the HMM. It appears to be necessary to resort to this static posterior in order to be able to assign a posterior to a single data point. The problem of regime prediction now consists in predicting the posterior distribution $\mathbf{P}^{\tau}$ a lead time τ ahead given an initial condition $\mathbf{x}^0$.

The matrix of transition probabilities $\mathbf{A}$ is a row-stochastic matrix, that is, $A_{ij} \geq 0$ for $i, j = 1, \ldots, K$ and $\sum_{j=1}^{K} A_{ij} = 1$ for $i = 1, \ldots, K$. According to the Perron-Frobenius theorem, the transition matrix (generically) possesses a real single leading eigenvalue $\lambda_1 = 1$; the corresponding real eigenvector $\mathbf{v}_1$ has only non-negative entries and represents the stationary distribution or invariant measure of the Markov chain when normalized with respect to the L_1-norm. In the following, we assume that all eigenvectors $\mathbf{v}_i$ are normalised with respect to the (complex) L_1-norm. The remaining eigenvalues have magnitude smaller than one and greater equal zero; they may come as real eigenvalues or in complex conjugate pairs.

Consider an initial probability distribution $\boldsymbol{\rho}^0$ expanded into the eigenvectors of the Markov model: $\boldsymbol{\rho}^0 = \mathbf{v}_1 + \sum_{i=2}^{K} c_i^0 \mathbf{v}_i$. After k steps, the initial distribution has evolved into the distribution $\boldsymbol{\rho}^k = \boldsymbol{\rho}^0 \mathbf{A}^k = \mathbf{v}_1 + \sum_{i=2}^{K} c_i^k \mathbf{v}_i$ with $c_i^k = c_i^0 \lambda_i^k$. Any initial distribution eventually converges to the stationary density of the Markov model: $\boldsymbol{\rho} = \lim_{k\to\infty} \boldsymbol{\rho}^k = \mathbf{v}_1$. A predictability timescale τ_i is associated with each eigenvector $\mathbf{v}_i$ defined as the time in which that mode decays to $1/e$ of its initial amplitude:

$$\tau_i = -\frac{\delta t}{\log|\lambda_i|}, \tag{22}$$

δt is the step size of the Markov model. The leading eigenvector $\mathbf{v}_1$ as the limit distribution is associated with an infinite timescale but does not really carry predictive information. The following eigenvalues and corresponding eigenvectors with moduli close to one are the slowly decaying probability modes which contribute most to predictability.

If the dynamics of regime transitions in the system under consideration were exactly first-order Markovian, the timescales τ_i would be independent of δt as the eigenvalues of transition matrices $\mathbf{A}$ and $\mathbf{A}'$ derived with step sizes δt and $\delta t' = \kappa \delta t$ with some positive real constant κ would then be related by $\lambda_i' = \lambda_i^\kappa$. This offers a possibility to check the consistency of HMMs with different time steps and thus the validity of the Markovian approximation.

The most obvious scheme for regime prediction is given by the Markov model itself. Given some initial condition $\mathbf{x}^0$ at time $t = 0$ with corresponding regime posterior $\mathbf{P}^0$, a forecast $\mathbf{f}^\tau = (f_1^\tau, \ldots, f_K^\tau)$ for the posterior $\mathbf{P}^\tau$ at lead time $\tau = k\delta t$, k being a positive integer, is obtained by iterating the Markov model k steps:

$$\mathbf{f}^\tau = \mathbf{P}^0 \mathbf{A}^k. \tag{23}$$

This model is referred to as the global Markov model in the following.

3.2 The nearest-neighbour prediction model

The global Markov model cannot be expected to provide overly good forecasts because due to the coarse-graining it does not make proper use of the information contained in a particular initial condition. Predictive information is refined by combining the HMM with a *nearest-neighbour model* [1–4]. Motivated by well-known methods from nonlinear time series analysis, a *locally constant* or *analogue model* of the regime posterior distribution based on nearest neighbours is built. Let $\mathcal{U}_M(\mathbf{x}^0)$ be the set of the M nearest neighbours of an initial condition $\mathbf{x}^0$ in a learning set of size N_l. Then a forecast is made by averaging over the posteriors of these nearest neighbours at lead time τ:

$$\mathbf{f}^\tau = \frac{1}{M} \sum_{\mathcal{U}_M(\mathbf{x}^0)} \mathbf{P}^\tau. \tag{24}$$

Nearest neighbours are taken with respect to the Euclidean norm. The number of nearest neighbours M and the size of the learning set N_l are parameters of the method controlling the locality of the model in state space.

3.3 Forecast characterization and verification

Two quantities based on information theory are introduced in order to characterize and evaluate the forecasts of the different models. The first one is prediction utility [16] defined by

$$R = \sum_{i=1}^{K} f_i \log \frac{f_i}{\rho_i}, \tag{25}$$

for a forecast at some prediction time τ, dropping the superscript for convenience now. Prediction utility is actually the relative entropy of the forecast distribution with respect to the stationary distribution of the Markov model and measures the information content in the forecast distribution over the stationary distribution. Utility is non-negative for any forecast distribution $\mathbf{f}$; it is zero if and only if $\mathbf{f} = \boldsymbol{\rho}$. Prediction utility is not a skill score; it just characterises the forecast.

To actually evaluate forecast quality the ignorance score [17] is used. It is here defined as

$$\mathrm{ign} = -\sum_{i=1}^{K} P_i \log f_i, \tag{26}$$

for some forecast $\mathbf{f}$ and corresponding true posterior distribution $\mathbf{P}$. This is actually a generalisation of the ignorance score in that here a probabilistic forecast is evaluated against a

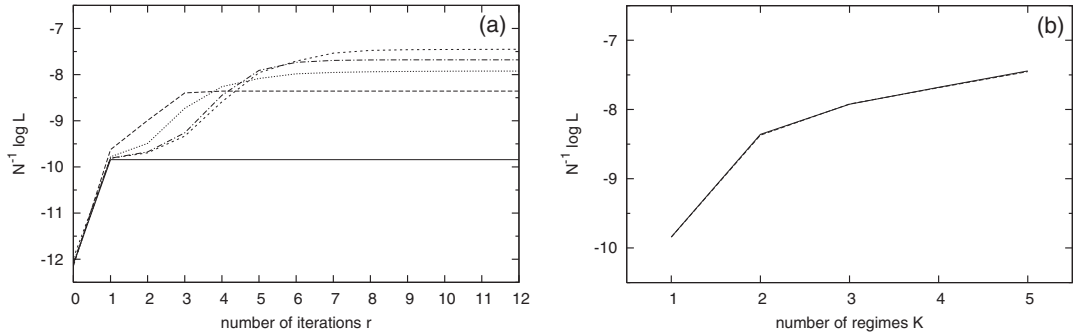

Fig. 2. (a) Convergence of the log-likelihood in the learning data set for $K = 1$ (solid), $K = 2$ (long-dashed), $K = 3$ (dotted), $K = 4$ (dot-dashed) and $K = 5$ (short-dashed). (b) In-sample (solid) and out-of-sample (dashed) log-likelihood as a function of the number of regimes K.

verification which is itself probabilistic whereas in [17] the verification is deterministic. Ignorance is non-negative and unbounded above; it measures the ignorance of someone who is in possession of the forecast $\mathbf{f}$ but not the true posterior $\mathbf{P}$. The mean ignorance score $\langle \text{ign} \rangle$ over an ensemble of initial conditions of any prediction scheme is bounded below by the mean entropy of the posterior distribution $\langle -\sum_{i=1}^{K} P_i \log P_i \rangle$, which is independent of the lead time τ.

4 Results

HMMs are determined for the stochastically forced Lorenz system with numbers of regimes ranging from $K = 2$ to $K = 5$. The estimation of the models is done in a learning time series of length $N = 50000$; the step length of the Markov chain is $\delta t = 0.1$, amounting to 5000 time units worth of data used in the estimation of the HMM. HMMs have also been estimated for step sizes $\delta t = 0.05$ and $\delta t = 0.2$. They are very close to the one at $\delta t = 0.1$ regarding the predictability timescales and the means and covariances of the regimes. Therefore only the case $\delta t = 0.1$ is presented here.

In Figure 2(a), the convergence of the log-likelihood with the number of iterations in the EM algorithm is illustrated. The algorithm converges quite fast; the likelihood function is virtually steady after 10 iterations for all K. Figure 2(b) gives the in-sample and out-of-sample log-likelihoods of the HMMs for various values of K. The out-of-sample values are calculated in a time series of length $N = 50000$ different from the learning set. In-sample and out-of-sample likelihood are virtually identical for all K; there is no sign of overfitting yet until $K = 5$. The likelihood function increases monotonically up to $K = 5$; it actually keeps increasing up to about $K = 10$ (not shown) exhibiting a very flat plateau.

Table 1 gives the magnitude of the eigenvalues and the corresponding predictability timescales of the HMMs with increasing numbers of regimes K. All models possess a second eigenvalue with a predictability timescale of about $\tau_2 \approx 1$. There is a clear gap between the second

Table 1. Moduli of eigenvalues and associated predictability timescales for hidden Markov models with different numbers of regimes K.

i	$K = 2$		$K = 3$		$K = 4$		$K = 5$	
	$\lvert\lambda_i\rvert$	τ_i	$\lvert\lambda_i\rvert$	τ_i	$\lvert\lambda_i\rvert$	τ_i	$\lvert\lambda_i\rvert$	τ_i
1	1.000	∞	1.000	∞	1.000	∞	1.000	∞
2	0.899	0.94	0.872	0.73	0.916	1.15	0.895	0.90
3			0.557	0.17	0.504	0.15	0.505	0.15
4					0.366	0.10	0.505	0.15
5							0.424	0.12

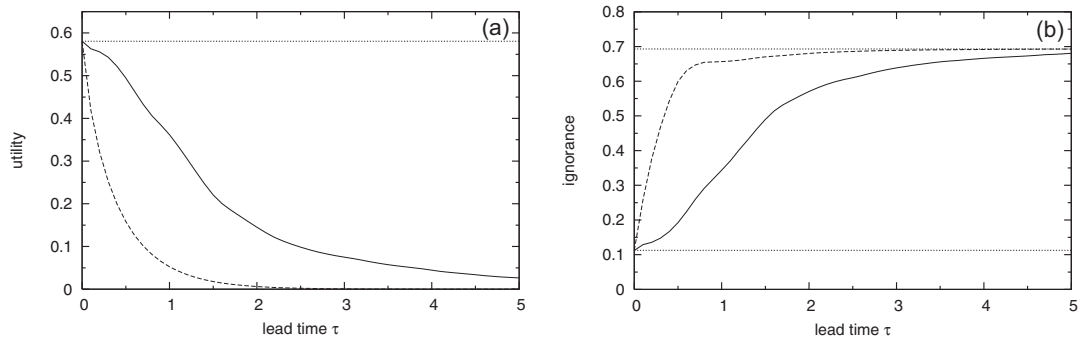

Fig. 3. (a) Mean utility of forecasts using the nearest-neighbour model (solid), the global Markov model (dashed) and the true posterior distribution (dotted). (b) Mean ignorance of forecasts using the nearest-neighbour model (solid), the global Markov model (dashed), the true posterior distribution (lower dotted) and the stationary distribution (upper dotted).

eigenvalue and the remainder of the spectrum whose timescales are smaller by a factor of about 4 to 7.5.

For a HMM with $K = 2$, the two regimes are located at positions $\boldsymbol{\mu}_1 = (-6.67, -6.50, 24.09)^{\mathrm{T}}$ and $\boldsymbol{\mu}_2 = (6.65, 6.47, 24.08)^{\mathrm{T}}$; they lie symmetrically on the wings of the attractor close to the unstable steady states. The matrix of transition probabilities is symmetric within estimation error and given by

$$\mathbf{A} = \begin{pmatrix} 0.950 & 0.050 \\ 0.051 & 0.949 \end{pmatrix}. \tag{27}$$

The stationary distribution is $\boldsymbol{\rho} = \mathbf{v}_1 = \mathbf{w} = (0.505, 0.495)$; the second eigenvector is $\mathbf{v}_2 = (-0.500, 0.500)$.

In the model with $K = 3$, the three regimes are centred at $\boldsymbol{\mu}_1 = (-8.08, -7.04, 27.64)^{\mathrm{T}}$, $\boldsymbol{\mu}_2 = (0.11, 0.17, 15.12)^{\mathrm{T}}$ and $\boldsymbol{\mu}_3 = (8.06, 6.93, 27.78)^{\mathrm{T}}$, respectively. The transition matrix is

$$\mathbf{A} = \begin{pmatrix} 0.876 & 0.124 & 0.000 \\ 0.155 & 0.685 & 0.160 \\ 0.000 & 0.132 & 0.868 \end{pmatrix}. \tag{28}$$

The stationary distribution of the Markov chain is $\boldsymbol{\rho} = \mathbf{v}_1 = \mathbf{w} = (0.362, 0.288, 0.350)$; the other eigenvectors are $\mathbf{v}_2 = (-0.500, 0.012, 0.488)$ and $\mathbf{v}_3 = (-0.243, 0.500, -0.257)$. The first and the third regime are very close to the regimes found with $K = 2$; a new regime occurs in the middle of the attractor in between the other two. The second eigenvector is very similar to the second eigenvector with $K = 2$. The third eigenvector involving the new regime versus the other two is only relevant on a very short timescale.

We conclude that the essential regime behaviour in this case is just the two wings of the attractor and the switching between them. It is already captured in the model with only two regimes. Therefore we will only consider that model further.

Figure 3 illustrates the mean utility and the mean ignorance score of the various regime forecasts. For the nearest-neighbour model, $M = 50$ nearest neighbours in a learning set of length $N_{\mathrm{l}} = 50000$ with sampling interval $\delta t = 0.1$ have been used. All statistics are calculated using an ensemble of $N_{\mathrm{v}} = 50000$ forecasts from initial conditions a time interval $\delta t = 0.1$ apart taken from a verification data set different from the learning set the model is estimated in and also different from the learning set the nearest neighbours are taken from. The mean utility of the nearest-neighbour model monotonically decreases to zero like the global Markov model but much more slowly. The ignorance score of both the Markov model and the nearest-neighbour model increase monotonically with lead time and then saturate at the value given by the stationary distribution. The nearest-neighbour model clearly outperforms the global

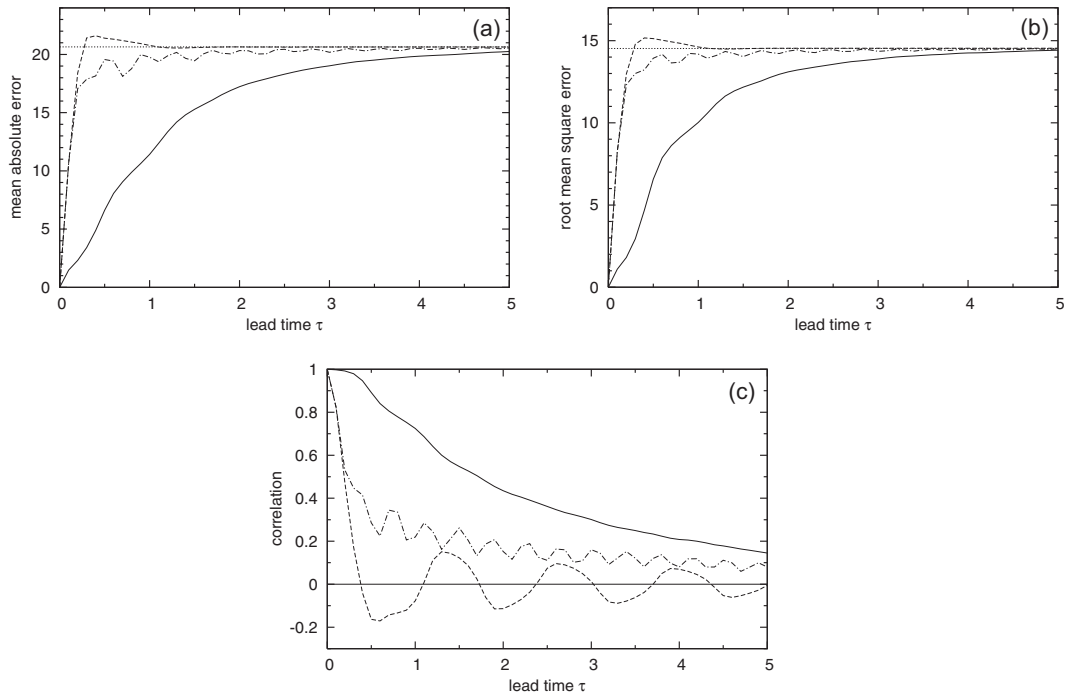

Fig. 4. (a) Mean absolute error of forecasts using the nearest-neighbour model (solid), the iterated AR(1) model fitted at lag $\delta t = 0.1$ (dashed) and the AR(1) model fitted at lag τ (dot-dashed); dotted horizontal line indicates the L_1-norm of the data. (b) Root mean square error of forecasts using the nearest-neighbour model (solid), the iterated AR(1) model fitted at lag $\delta t = 0.1$ (dashed) and the AR(1) model fitted at lag τ (dot-dashed); dotted horizontal line indicates the L_2-norm of the data. (c) Correlation skill of forecasts using the nearest-neighbour model (solid), the iterated AR(1) model fitted at lag $\delta t = 0.1$ (dashed) and the AR(1) model fitted at lag τ (dot-dashed).

Markov model. It still has substantial skill at $\tau \approx 2$ and there are traces of regime prediction skill up to about $\tau \approx 5$.

Figure 4 gives information on predictability in the system in the conventional sense of prediction of individual trajectories. The same nearest-neighbour model as for regime prediction is used but now as is usually done as a deterministic model for trajectory prediction [1–4]:

$$\hat{\mathbf{x}}^\tau = \frac{1}{M} \sum_{\mathcal{U}_M(\mathbf{x}^0)} \mathbf{x}^\tau. \tag{29}$$

The mean absolute error, the root mean square error and the correlation skill are shown. These are given by

$$\text{mae} = \left\langle \sum_{i=1}^{3} |x_i - \hat{x}_i| \right\rangle, \qquad \text{rmse} = \sqrt{\left\langle \sum_{i=1}^{3} (x_i - \hat{x}_i)^2 \right\rangle}.$$

and

$$\text{corr} = \frac{\left\langle \sum_{i=1}^{3} (x_i - \langle x_i \rangle)(\hat{x}_i - \langle \hat{x}_i \rangle) \right\rangle}{\sqrt{\left\langle \sum_{i=1}^{3} (x_i - \langle x_i \rangle)^2 \right\rangle} \sqrt{\left\langle \sum_{i=1}^{3} (\hat{x}_i - \langle \hat{x}_i \rangle)^2 \right\rangle}},$$

where $\hat{x}_i$ is a prediction at some lead time τ and x_i the corresponding verification. In the present system, the predictability timescale for trajectory prediction is about the same as for regime prediction. There is detectable skill in the predictions up to $\tau \approx 5$.

As a comparison to the global Markov model, prediction of the system with a first-order autoregressive process (AR(1) model) is examined. Firstly, an iterated one-step AR(1) model with step size $\delta t = 0.1$ as in the Markov model is considered. The predictions for lead time $\tau = k\delta t$, k being a positive integer, are given by $\hat{\mathbf{x}}^\tau = \langle \mathbf{x} \rangle + \mathbf{B}_1^k(\mathbf{x}^0 - \langle \mathbf{x} \rangle)$ with $\mathbf{B}_1 = \mathbf{B}_1\mathbf{B}_0^{-1}$ where $\mathbf{B}_0$ is the covariance matrix of the system and $\mathbf{C}_1$ the covariance matrix at lag 1 (corresponding to δt). Secondly, an AR(1) model specific to the prediction time τ is built, that is, $\hat{\mathbf{x}}^\tau = \langle \mathbf{x} \rangle + \mathbf{B}_k(\mathbf{x}^0 - \langle \mathbf{x} \rangle)$ with $\mathbf{B}_k = \mathbf{C}_k\mathbf{C}_0^{-1}$ where $\mathbf{C}_k$ is the covariance matrix at lag k (corresponding to τ). Both linear models rapidly lose any substantial skill. The model specific to the lead time is slightly better as it makes use of the predictive information in the oscillation of the x_3-component which is not visible in the covariance at lag 1. The matrix $\mathbf{B}_1$ has eigenvalues with $|\lambda_1| = |\lambda_2| = 0.758$ and $|\lambda_3| = 0.628$. These are associated with predictability timescales $\tau_1 = \tau_2 = 0.36$ and $\tau_3 = 0.22$. Predictability of trajectories based on the autocorrelations in the system decays faster than regime predictability in the global Markov model.

5 Discussion

A framework for regime prediction and predictability in nonlinear dynamical systems has been described and exemplified. In the stochastically forced Lorenz system, regime behaviour is dominated by just two regimes, basically the two wings of the attractor, which are symmetrical both statistically and dynamically. Things should be more interesting in systems with more than two non-symmetric regimes. An example is large-scale atmospheric circulation [8,9].

The present methodology holds the potential for improving prediction by combining dynamical and statistical models. Often a dynamical model for an observed system is available which is more or less good, yet imperfect, due to either parameter uncertainty or misspecification of the model class [6,7]. Such models may provide good short-term forecasts but miss regime residence times or the structure of regimes in the long term. A similar situation exists in numerical weather prediction: Current models have a good local propagator but model error is still there and sometimes they miss transitions between weather regimes. The statistical regime model built from data then really contains predictive information on top of the dynamical model and could be hybridized with the dynamical model, for example, in a Bayesian model averaging framework. Put differently, the coarse-grained regime model acts as a (hard or weak) constraint on the predictions of the dynamical model.

References

1. J.D. Farmer, J.J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987)
2. G. Sugihara, R.M. May, Nature **344**, 734 (1990)
3. H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, L.S. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993)
4. H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 2004)
5. E.N. Lorenz, J. Atmos. Sci. **20**, 130 (1963)
6. N. Rulkov, L. Tsimring, H. Abarbanel, Phys. Rev. E **50**, 314 (1994)
7. J. Timmer, H. Rust, W. Horbelt, H.U. Voss, Phys. Lett. A **274**, 123 (2000)
8. D.T. Crommelin, J. Atmos. Sci. **60**, 229 (2003)
9. D. Kondrashov, K. Ide, M. Ghil, J. Atmos. Sci. **61**, 568 (2004)
10. L.R. Rabiner, Proc. IEEE **77**, 257 (1989)
11. A. Sutera, J. Atmos. Sci. **37**, 246 (1980)
12. P.E. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, 1992), p. 632
13. T. Sauer, J.A. Yorke, M. Casdagli, J. Stat. Phys. **65**, 579 (1991)
14. A.P. Dempster, N.M. Laird, D.B. Rubin, J. Roy. Stat. Soc. **39**, 1 (1977)
15. B.W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, 1986), p. 175
16. R. Kleeman, J. Atmos. Sci. **59**, 2057 (2002)
17. M.S. Roulston, L.A. Smith, Mon. Wea. Rev. **130**, 1653 (2002)