Judging Model Reduction of Complex Systems

Jie Sun,* Erik M. Bollt,† and Takashi Nishikawa‡
Department of Mathematics & Computer Science, Clarkson University, Potsdam, NY 13699-5815

Model reduction is a common goal in the study of complex systems. The quality of such reduction, however, may not be reflected correctly in the stepwise prediction error in the model, since it ignores the global geometry of the dynamics. Here we introduce a general two-step framework, consisting of dimensionality reduction of the time series followed by the modeling of the resulting time series, and propose to use the shadowing distance to measure the quality of the second step. Using coupled oscillator networks as a prototypical example, we demonstrate that our approach can outperform those based on stepwise error and suggest that it sheds new light on the problem of identifying and modeling low-dimensional dynamics in large-scale complex systems.

PACS numbers: 89.75.-k, 05.45.Tp, 05.45.Xt,

Model reduction is an important concept found across various fields of science and engineering. Finding a low-dimensional model that captures the gross features of a high-dimensional system is a fundamental problem in the physics of complex systems, which touches upon such disparate fields as PDE [1], dynamical systems [2], and network theory [3]. Consider, for example, a complex system modeled by a network of coupled oscillators [4, 5], which may support processes such as disease spreading [6], the evolution of a food web [7], or the dynamics of a power grid [8]. It is often useful and desirable in such a system to “average” across parts, so that a system with only a few dynamical units is sufficient to model the large-scale dynamical behavior of the original system. In much the same way as community detection methods partition the nodes of a complex network into groups based on topological similarity [9, 10], dynamical units in a complex system may be grouped based on dynamical similarity.

For a given high-dimensional system, there are often many different ways to obtain a low-dimensional reduced model, and this leads to a natural and fundamental question: how do we choose the best one? For example, is it best to simply average the equations for individual units to obtain a reduced model? Would it be better to use a weighted average of the units reflecting their various roles within the system or to introduce an extra component into the model to compensate for the loss of information due to dimensionality reduction of the time series? Addressing these questions requires quantifying the quality of model reduction.

For a given time series, a best model of the same dimension as the time series is commonly selected based on the least-squares (LS) criterion, which minimizes the sum of squared stepwise errors. However, for chaotic systems this criterion can be inappropriate, or even misleading [11–13], due to the nature of the cost function that it is based on local, rather than global and geometric, features of the system. If reduced models of lower dimension are considered, the LS criterion can still be formulated, but its appropriateness is even more questionable.

To the best of our knowledge, there has been no previous work to address this problem of quantifying the quality of model reduction beyond the LS criterion. In this Letter, we introduce a new framework (illustrated in Fig. 1), which decomposes the problem into two steps: 1) find a low dimensional representation of the measured time series (bottom horizontal arrow), and then model the resulting time series (up arrow). Here we propose to use the shadowing distance (defined in the text) to measure the quality of the second step.

![Fig. 1: (Color online) Two-step approach for judging model reduction of large complex systems. Since direct comparison is impossible due to the difference in dimensionality (top horizontal arrow), we first find a low dimensional representation of the measured time series (bottom horizontal arrow), and then model the resulting time series (up arrow). Here we propose to use the shadowing distance (defined in the text) to measure the quality of the second step.](image-url)

\[ J(f_a) = (1 - \mu)\eta + \mu\epsilon \]  

where \(\mu\in [0,1]\) can be chosen (by the modeler) to emphasize either part of the model reduction process.

The first step is essentially a problem of finding a low-dimensional manifold that best fits the original time series as a set of points in high-dimensional state space. Once a best-fit manifold is found, the loss of information,
model trajectory can follow the time series. This can be regarded as a measure of the level of confidence in saying that the noisy time series came from the model, and thus as a measure of how good the model is for the time series \( \{x_t\}_{t=1}^T \). Symbolically, given a time series \( \{x_t\}_{t=1}^T \), we define the shadowing distance for a given model \( f \) as:

\[
\epsilon_{sd}(f) \equiv \inf_{y_t \in D} \max_{1 \leq t \leq T} ||x_t - y_t||, 
\]

where \( D \) is the state space, \( y_t \) for \( t \geq 2 \) is given recursively by \( y_{t+1} = f(y_t) \), and \( ||\cdot|| \) denotes the Euclidean distance. This definition can be naturally extended to continuous-time systems.

To estimate \( \epsilon_{sd} \), we find a numerical shadowing trajectory using a procedure developed for systems of arbitrarily high dimension, both in discrete time [19] and in continuous time [20]. Starting with the given time series, we iteratively generate a less and less noisy trajectory that still stays close to the time series. This leads to a trajectory which has stepwise error within the machine precision and whose distance to the time series is likely to be near its minimum possible value, \( \epsilon_{sd} \). Since we do not expect a long shadowing trajectory for a very noisy time series, such as a reduced time series resulting from (nonlinear) dimensionality reduction, we take segments of the time series of length \( T_n \leq T \), and average \( \epsilon_{sd} \) computed for these segments. An appropriate choice of \( T_n \) is \( \approx \log(\delta^{-1})/\lambda \), where \( \delta \) is the machine precision and \( \lambda \) is the maximum Lyapunov exponent of the system, since a longer trajectory is likely to suffer from the accumulation of computational error.

A critical feature of \( \epsilon_{sd} \) is that it correctly reflects the model’s capability in producing a trajectory that closely follows the entire time series. In contrast, the stepwise error, which can be measured by \( \epsilon_{se}(f) \equiv \sqrt{\sum_{t=1}^{T} ||f(x_t) - x_{t+1}||^2/T} \) for a given time series, may be small even when large error accumulation along the time series is unavoidable. This is clearly illustrated in Fig. 2 for non-chaotic time series from the standard circle map. The contrast would be even more dramatic for chaotic systems. Based on this observation, we propose to use the optimal shadowing (OS) criterion, in which we select a model with the smallest shadowing distance, instead of one with smallest \( \epsilon_{se} \) (LS criterion).

To demonstrate that our approach is well-suited for addressing fundamental questions on model reduction, consider a general network of coupled discrete-time dynamical systems described by

\[
x^{(i)}_{t+1} = g(x^{(i)}_t, a^{(i)}) - \sigma \sum_{j=1}^{n} \ell_{ij} g(x^{(j)}_t, a^{(j)}),
\]

where \( \{x^{(i)}\}_{i=1}^{n} \) with \( x^{(i)}_t \in \mathbb{R}^d \) represents the states of the \( n \) oscillators at time \( t \), the function \( g(x, a) \) describes the dynamics driving individual oscillators (which can

FIG. 2: (Color online) Shadowing distance \( \epsilon_{sd} \) vs stepwise error \( \epsilon_{se} \) for judging model quality. A periodic time series \( \{x_t\}_{t=1}^T \) was generated by the standard circle map [15], \( x_{t+1} = x_t + \Omega - 0.12 \sin(2\pi x_t) \) with \( \Omega = 0.35 \) and a random choice of \( x_1 \). Using the same map with \( \Omega = 0.35 + \Delta \Omega \) as a model for this time series, we have \( \epsilon_{se} = \|\Delta \Omega\| \) [black solid curve in panel (a)]. Despite the small \( \epsilon_{se} \), error can accumulate over time and reach 0.5 (the maximum possible error on the unit circle) for \( \Delta \Omega \approx 0.002 \), due to the difference in the rotation number for the two maps [panel (b)]. In contrast, \( \epsilon_{sd} \) [blue solid and red dashed curve in panel (a)], computed with \( T_n = T \) correctly reflects this effect, as well as the dependence on \( T \) and asymmetry with respect to \( \Delta \Omega \), none of which is captured at all by \( \epsilon_{se} \).

\( \eta \), can be quantified by measuring the amount of residual, i.e., deviation from the manifold. There are some relatively well-developed methods for this problem, such as the principal component analysis [2] for finding linear manifolds and ISOMAP [14] for finding nonlinear ones.

The second step, on the other hand, is a subtle problem that appears largely unexplored in the literature. To simplify our discussion, consider a scalar time series produced by an unknown one-dimensional map, where the goal is to measure the quality of a given map as a model for the unknown map. On the one hand, small stepwise error (or equivalently, small difference in the two maps as functions) may not imply that the model is good, since the error can accumulate over time, leading to poor long-term prediction. This is particularly problematic for a chaotic system due to the sensitivity to initial conditions. On the other hand, for a chaotic system large accumulated error does not necessarily mean that the model is poor, since this can occur even when the two maps match perfectly, due to sensitivity to perturbations.

To avoid this fundamental difficulty, we propose a measure of model quality based on the concept of shadowing [16–18], which was originally designed to test whether a simulated time series suffering from numerical inaccuracy could have come from a given model system. The shadowing distance \( \epsilon_{sd} \) for a given finite-length time series is defined as the smallest distance within which a
be different and are parameterized here by \( a \), and \( \sigma \) is the global coupling strength. The effective diffusive coupling among the oscillators is represented by the discrete Laplacian matrix, obtained from the adjacency matrix \( B = [b_{ij}] \) of the network by \( \ell_{ij} = - (1 - \delta_{ij}) b_{ij} + \delta_{ij} \sum_k b_{ik} \), where \( \delta_{ij} \) is the Kronecker delta. The coupling function has been chosen to have the same form as the individual dynamics \( g(x, a) \), which corresponds to the situation where each oscillator receives a direct signal from the output of its neighbors.

As a concrete example, consider a complex network of logistic maps: \( g(x, a) = ax(1 - x) \). The network structure is generated by the Krapivsky-Redner (KR) model [21]: start with two nodes and one directed link from node 2 to 1 (the “ancestor” of node 2); at each step, a new node \( k \geq 3 \) is added to the network, making a directed link to either a randomly chosen node from \( \{1, \ldots, k - 1\} \) (with probability \( 1 - r \)) or its ancestor (with probability \( r \)). In our example we take \( r = 0.5 \) and \( n = 1000 \) (total number of nodes). Each directed link is then assigned weight 1, and the associated backward link is created with weight 0.01. We set the node parameters as \( a^{(i)} = 3.96 + (i/n) \cdot 0.04 \), filling the interval [3.96, 4] uniformly. The coupling strength is chosen as \( \sigma = 0.75 \). With this set of parameters, the oscillators are nearly synchronized [22], i.e., the trajectory of any one oscillator stays close to that of any other oscillator.

Is there a simplified system that can model this 1000-dimensional coupled chaotic system? Since the oscillators are nearly synchronized, a reasonable guess might be a one-dimensional model \( f(x) = \bar{a}x(1 - x) \) where \( \bar{a} = \frac{1}{n} \sum_i a^{(i)} \) is the average parameter. Such uniform averaging, however, does not respect the non-homogeneity of the network, resulting from the fact that the oscillators introduced at early stage of network generation have much larger influence than the others. Since the parameters for these nodes are closer to 3.96, we expect that an optimal value for the parameter \( a \) in the family of candidate models \( f_a(x) = ax(1 - x) \) would be smaller than the average \( \bar{a} \). But what is this optimal value?

In Fig. 3(a) we compare the candidate models using both stepwise error (\( \epsilon_{sd} \)) and shadowing distance (\( \epsilon_{sd} \)). For the reduced time series, we take the average trajectory, \( x_t = \frac{1}{n} \sum_i x_i^{(i)} \) (but we will discuss other possibilities below). The two criteria give clearly distinct values of \( a \) as the best choice. The estimate of the Lyapunov exponent of the original system computed from the best model selected by the LS criterion is more than 10% off the actual value, while the optimal model via the OS criterion gives an estimate within 1%. In Fig. 3(b) we compare the predictive power of the optimal models selected by the two criteria. For the prediction of more than three time steps away from present, the LS based optimal model suffered from prediction error more than 20% larger than the OS based optimal model, another evidence that model quality is better judged by the proposed OS criterion.

We now address the question of whether the average trajectory is the best choice as a reduced time series. It is indeed the best if we minimize only the dimensionality reduction error \( \eta = 0 \) in Eq. (1), but allowing a slightly larger reduction error might reduce the shadowing error \( \epsilon_{sd} \) significantly. To explore this interplay between the two components of the model reduction process and how their relative emphasis influence the choice of the best model, consider the same networked logistic map system used for Fig. 3. We now consider weighted average of the individual trajectories: \( x_t = \frac{1}{n} \sum_i w^{(i)} x_i^{(i)} \) where \( w^{(i)} = k_i^2 / \sum_j k_j^2 \) for given \( \beta \). Here \( k_i \) is the out-degree of node \( i \). The dimensionality reduction error is then computed as \( \eta(\beta) = \sqrt{\frac{1}{n} \sum_i |x_t - x_i^{(i)}|^2 / (nT)} \). We define the shadowing distance \( \epsilon_{sd}(\beta) \) for a given \( \beta \) to be the minimum value of \( \epsilon_{sd} \) over the parameter \( a \) of the reduced model \( f(x, a) = ax(1 - x) \). In Fig. 4 we show the \( \beta \)-dependency for the model reduction error \( \zeta(\beta) = (1 - \mu) \eta(\beta) + \mu \epsilon_{sd}(\beta) \), for different values of \( \mu \), which controls the relative emphasis on the two parts. If \( \mu = 0 \) (upper left panel of Fig. 4), the best choice is \( \beta = 0 \).
The three other panels in Fig. 4), giving more weight to nodes with larger out-degree yields a smaller model reduction error. It is interesting that for a wide range of \( \mu \), the value \( \beta = 1 \) leads to the best weight distribution.

To show that our OS criterion can be effective even when individual trajectories are related nonlinearly, consider the simplest case of a two-unit system

\[
\begin{align*}
    x_{t+1} &= f(x_t) + \sigma(\Phi^{-1}(g(y_t)) - f(x_t)), \\
    y_{t+1} &= g(y_t) + \sigma(\Phi[f(x_t)] - g(y_t)).
\end{align*}
\]

Here we choose \( f(x) = 3.96x(1-x) \), \( g(y) = \sqrt{4y(1-y)} \), \( \sigma = 0.3 \), and \( \Phi(x) = \sqrt{x} \). Although naive comparison of \( x_t \) and \( y_t \) might lead to the conclusion that no model reduction is possible, nonlinear reduction techniques can be used to find intrinsic coherence between \( x_t \) and \( y_t \) [Fig. 5(a)]. A one-dimensional model for the single trajectory \( \{x_t\} \) is thus sufficient to model the system. Considering again the family of models of the form \( f_a(x) = ax(1-x) \), Fig. 5(b) compares the predictive power of the optimal models selected by the LS and OS criteria. It clearly shows that the error in making 7- to 10-step predictions is as much as 45% more for the LS-selected model than for the OS-selected model.

To summarize, we have proposed a general approach for judging the quality of low-dimensional reduced models for high-dimensional complex systems. A key to our approach is the decomposition of the problem into the spatial and temporal domain where the modeling error can be assessed separately. Another prominent feature is the use of a shadowing-based criterion in the temporal domain, which can be combined with any nonlinear dimensionality reduction technique in the spatial domain.

We have shown that the models selected based on our criterion are better than those selected by the traditional LS criterion, both in terms of the Lyapunov exponent estimates and of short-term prediction error.

We have provided evidence that this approach is useful in addressing fundamental questions about model reduction for a large network of nearly synchronized oscillators. As an example, we discovered an interesting phenomenon for the asymmetric KR networks of logistic maps, that the best model reduction is often obtained when the maps are averaged with weights proportional to the out-degrees. We emphasize that our framework applies readily to systems in which individual oscillators have a multi-dimensional state space and connected through arbitrary network structure. We thus suggest that it lays a promising foundation for tackling more general model reduction problems in the physics of complex systems.

**Acknowledgements** — E.M.B. and J.S. were supported by the Army Research Office under Grant 51950-MA.

---

* Electronic address: sunj@clarkson.edu
† Electronic address: bolltem@clarkson.edu
‡ Electronic address: tnishika@clarkson.edu


Another possibility is to use the distribution of shadowing times. See [13] for details.