Chaos, Solitons and Fractals 94 (2017) 8–15

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Regularized forecasting of chaotic dynamical systems

Erik M. Bollt*

Department of Mathematics, Clarkson University, Potsdam, New York 13699, USA

ARTICLE INFO

Article history:
Received 27 January 2016
Revised 9 October 2016
Accepted 11 October 2016

Keywords:
Chaos
Forecasting
Inverse problems
Tikhonov regularization
Time delay embedding

ABSTRACT

While local models of dynamical systems have been highly successful in terms of using extensive data sets observing even a chaotic dynamical system to produce useful forecasts, there is a typical problem as follows. Specifically, with *k*-near neighbors, kNN method, local observations occur due to recurrences in a chaotic system, and this allows for local models to be built by regression to low dimensional polynomial approximations of the underlying system estimating a Taylor series. This has been a popular approach, particularly in context of scalar data observations which have been represented by time-delay embedding methods. However such local models can generally allow for spatial discontinuities of forecasts when considered globally, meaning jumps in predictions because the collected near neighbors vary from point to point. The source of these discontinuities is generally that the set of near neighbors varies discontinuously with respect to the position of the sample point, and so therefore does the model built from the near neighbors. It is possible to utilize local information inferred from near neighbors as usual but at the same time to impose a degree of regularity on a global scale. We present here a new global perspective extending the general local modeling concept. In so doing, then we proceed to show how this perspective allows us to impose prior presumed regularity into the model, by involving the Tikhonov regularity theory, since this classic perspective of optimization in ill-posed problems naturally balances fitting an objective with some prior assumed form of the result, such as continuity or derivative regularity for example. This all reduces to matrix manipulations which we demonstrate on a simple data set, with the implication that it may find much broader context.

1. Introduction – local models, local forecasting

Forecasting chaotic dynamical systems, from measured data, is a topic that has seen a great deal of activity, at least for the last thirty years, facilitated by the time-delay embedding methods. In the time-delay embedding literature, forecasting from observed states, embedding the states and then fitting local models based on regression to the behavior of *k*-near neighbors (kNN) was put forward and somewhat matured by the mid1990s, [4–10]. So, in [13] we discuss the role of local models in model selection as it relates to spatial scale, and some of that is reviewed here. In particular we have been interested in how local modeling [11], and see also [12], relates to local polynomial models.

Specifically note that local models are built for the transformation based on observing the orbits of near neighbors, and hoping that there are a lot of sampled orbit segments due to recurrence and a long orbit sample, then in any small neighborhood there would be many samples. The "k" in kNN means to collect those *k* points from the data set that are closest to the forecast point. If there is noise, or otherwise, model error, then a degree of smoothing is implied by a least squares estimation of the local transformation. While a higher ordered model will tend to well fit more terms in a Taylor polynomial estimation of the local model, there are inherently many more parameters to be fitted when using a high degree polynomial, and so a much larger *k* would be required, and hence correspondingly the neighborhood would be larger. So as discussed in [13], there is a balancing between fine scale, data density, and smoothing when performing local modeling by least squares regression alone; here we add to this discussion that regularity can be emphasized directly by using Tikhonov regularization concepts derived from convex optimization theory, [18,19] and also found in advanced matrix analysis, [16], to find locally useful forecasts, which also have good global regularity properties, and with less data than perhaps a kNN method on its own.

When a global model is not be available to forecast evolution of a given point, a useful forecast may still be possible by observing the evolution of *k* nearby points using regression to appropriately "average" between them. In terms of a polynomial basis, this approach develops a least squares regression of the first few terms of a Taylor's series for the unknown global model, in the neighborhood of the point to be forecast, using the *k* neighbors as data.

* Corresponding author.
E-mail address: bolltem@clarkson.edu

http://dx.doi.org/10.1016/j.chaos.2016.10.007
0960-0779/

Enough near neighbors must be chosen to allow for the minimal fitting of the polynomial model, and furthermore, somewhat more than the minimal number of points should be chosen to allow for some degree of smoothing. This is akin to the familiar statistical issues faced when fitting a line to noisy data; more than 2 points should be used to confidently specify the line. Scale of the model is a major issue: there are competing demands between local truncation error that push toward small neighborhoods, but smoothing and confidence push toward larger k, leading to larger neighborhoods when using finite data sets. This trade-off was the topic of [13]. In this paper, we furthermore address an important issue overlooked in all previous studies on kNN local modeling, which is that if a model is developed for each neighborhood based on near neighbors, then since two nearby points may have a different set of near neighbors, this leads to lack of smoothness (regularity) of the forecasts. We address this problem here, by expanding our previous work to include regularization by utilizing concepts of Tikhonov regularization theory.

Consider a dynamical system,

$$T : M \to M,$$
$$y_{n+1} = T(y_n). \tag{1.1}$$

Let $y \in M \subset \mathbb{R}^d$. Assume that from the nonlinear dynamical system, we have a large collection of observed iterations as an orbit of Eq. (1.1), $\{y_i\}_{i=0}^N$, such that $y_{i+1} = T(y_i)$. Here, T will stand for a discrete time mapping as the transformation, throughout this paper, and note that if we have a continuous time process, then the discrete time mapping may come either by Poincare' section, or by time delay embedding, of a flow. For the modeling discussion below, these actually need not be a single orbit, but for the regularity discussion to follow, it is best if we include that assumption now. Furthermore assume that there is uniformly enough regularity of T so that there exists a Taylor's expansion through order-K, which we will exploit in the next Section 2. The standard discussion of local modeling is to put forward that these local polynomials can be estimated by nearby sampled points and their images, generally by a regression method. However, we also generally expect that these local models will vary continuously, or vary continuously with respect to higher ordered derivatives of T, with respect to variations in the sample point. We will show here that this desirable and physically expected property can be emphasized with Tikhonov regularity theory.

Now the idea is that for any point w as an initial condition that we may wish to forecast but that may not be amongst the observed orbit values, $w \notin \{y_i\}_{i=0}^N$, we proceed with local models built from first collecting near neighbors to w, amongst the data. A standard way to forecast a dynamical system, when presented with many previous states, is to collect "k-near" neighbors (kNN) in the phase space, and in some manner, average, regress, or otherwise associate the current forecast to those previous forecasts. The simplest version of these associations would be the method of analogues [9] from classical weather forecasting, namely forecasts are identified with the most same measured state. From [11–13], we review local forecasting in terms of local polynomial models.

Note that perhaps we may either estimate a discrete time map $T(y)$ from many observations as just stated, with the hope that there is low dimensionality, or a popularly common scenario is that we will only observe a single scalar time series, measured from a vector valued, y and the time delay embedding representation will be used. That is, a time-series from a "chaotic" dynamical system allows a data-only analysis by embedding attractor reconstruction, [1,2,4,6,7,10]. Recall that if an autonomous dynamical system,

$$\dot{x} = F(x), x(t) \in \mathbb{R}^n, \text{ and } x(t_0) = x_0, \tag{1.2}$$

has an invariant attractor A then an experimentalist who does not know the underlying global model Eq. (1.2) may not even know which are the correct variables to measure. Generally, a single-data channel can be considered to be a scalar measurement function $h[x(t)] : \mathbb{R}^n \to \mathbb{R}$. Given a set of measurements $\{h[x(t_i)]\}_{i=0}^N$, with uniformly spaced time samples t_i, the time-delay embedding is a vector,

$$\boldsymbol{y}(t) = < h[x(t)], h[x(t-\tau)], h[x(t-2\tau)], \ldots, h[x(t-d\tau)] >, \tag{1.3}$$

and one generally chooses τ to be some multiple of the sampling rate $\Delta t = t_{i+1} - t_i$. Takens proved [3] that, for topologically generic measurement function h, if the attractor A is a smooth m-dimensional manifold, then if one chooses the delay dimension to be $d \geq 2m+1$, then Eq. (1.3) is an embedding, meaning there exists a one-to-one function $G : A \to \mathbb{R}^d$, and G is a diffeomorphism. Sauer, *et. al* [8] proved an extension to allow for nonsmooth A, and even fractal A. To reconstruct the attractor, both of these results assume that the data is clean, and the data set is arbitrarily long. Neither assumption is physically realizable, but nonetheless, time-delay reconstruction has found many applications to nonlinear modeling and to prediction. See [1,2,4,6,9,10].

Local linear regression of the observed evolution of k-nearest neighbors $\{\boldsymbol{y}_j(t)\}_{j=1}^k$, to their images $\{\boldsymbol{y}_j(t+\tau)\}_{j=1}^k$, has emerged as the most popular method to predict "the next $y(t)$." The idea is that a Taylor's series of the (unknown) function f_τ, which evolves (flows) initial conditions $y(t)$, according to the differential equation, Eq. (1.2), is well approximated by the linear truncation, *if the near neighbors are "near enough."* Error analysis, such as that found in [10], is based on this local-truncation error, and therefore considers the Luyapunov exponents. There is naturally a conflict of demands since on the one hand, a) small local truncation error demands that neighborhoods be small, and therefore k must not be chosen too large, using a fixed (linear) model, but on the other hand, b) statistical fluctuations demand that k be chosen large enough to infer a degree of smoothing. The problem we study here is that it is well know that those points which are the near neighbors to any given sample point may not vary continuously with position in space. So the predictions likewise may vary discontinuously. Therefore we have developed a perspective here to emphasize that regularity is a desirable property. In many ways, this work should be considered as analogous to the standard local forecasting, but simply an enhanced alternative version. The emphasized regularity of forecasts therefore improves plausibility of forecasts in that there will be fewer jumps between forecasts of nearby initial conditions due simply to the artificial reason that the near neighbors set may differ.

2. Basis for local polynomial regression

Assuming that the transformation T has enough regularity to justify a Taylor polynomial at each point w, to the degree sought. For example, a local affine model of T at w, $T|_w$ is,

$$y = T_0 + DT \cdot h, \tag{2.1}$$

regressed over k-nearest neighbors of w, $\{y_{k_j}\}_{j=1}^k \subset \{y_i\}_{i=0}^N$, where DT is related to the Jacobian derivative in a neighborhood of w and $h = w - y$. This may be thought of as a local truncation of a Taylor's series. We index the k-nearest points to w by k_j, ordered $k_1 < k_2 < .. < k_k$ monotonically with respect to distance from w, assuming an underlying metric space. For a "good fit," just as realized by any Taylor polynomial, fit is better if h is small. So we would demand that $\{y_i\}_{i=0}^N$ "fills" the space adequately so that for any w we are likely to select that the k-nearest data points will be "close-enough" for a good estimate. A sufficient condition for a long orbit

to $\{y_i\}_{i=0}^N$ likely "fill" the space, is to assume that T is ergodic. Expanding the precision suggests a quadratic model of $T|_w$ as,

$$y = T_0 + DT \cdot h + \frac{1}{2} h^t \cdot H \cdot h, \tag{2.2}$$

where H is related to the Hessian matrix of second derivatives, and higher ordered models follow similarly. However, we will generally work here with local affine models and assume that there is enough data density to justify this; note that higher order models require more parameters to be fitted and therefore more data for good statistics of the fitting to balance against the concept that higher ordered models have generally good local truncation error. Thus it can be a losing proposition to increase the order of the model to reduce error, considering limited data.

The d parameters of T_0 and the d^2 parameters of DT of Eq. (2.1) may be estimated by least squares according,

$$\boldsymbol{Y} = \boldsymbol{X} \cdot \boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{2.3}$$

by corresponding normal equations [14,15], $\boldsymbol{X}^t \cdot \boldsymbol{X} \cdot \boldsymbol{\alpha} = \boldsymbol{X}^t \cdot \boldsymbol{Y}$, which is the convenient matrix form of linear regression, and this maintains the same vector form regardless of the degree of the fitted polynomial. For an unbiased model, expectation of the random variable is $E(\boldsymbol{\epsilon}) = \mathbf{0}$. The word "linear" refers to the linearity of coefficients which combine multiple linearly independent terms. For the affine model, one chooses,

$$\boldsymbol{Y} = \begin{bmatrix} y_{k_1+1}^t \\ y_{k_2+1}^t \\ \cdots \\ y_{k_k+1}^t \end{bmatrix}, \text{ and } \boldsymbol{X} = \boldsymbol{X}_1 = \begin{bmatrix} 1 & y_{k_1}^t \\ 1 & y_{k_2}^t \\ \cdots & \\ 1 & y_{k_k}^t \end{bmatrix}, \tag{2.4}$$

while for the quadratic model, Eq. (2.2),

$$\boldsymbol{X} = [\boldsymbol{X}_1 : \boldsymbol{X}_2], \tag{2.5}$$

and,

$$\boldsymbol{X}_2 = \begin{bmatrix} y_{k_1} y_{k_1}^t & y_{k_2} y_{k_1}^t & \cdots & y_{k_d} y_{k_1}^t \\ y_{k_1} y_{k_2}^t & y_{k_2} y_{k_2}^t & \cdots & y_{k_d} y_{k_2}^t . \\ \cdots & & & \\ y_{k_1} y_{k_k}^t & y_{k_2} y_{k_k}^t & \cdots & y_{k_d} y_{k_k}^t \end{bmatrix}, \tag{2.6}$$

is a convenient way to write quadratic terms. Formally, whatever the order of the local polynomial estimation, in this matrix notation such as Eq. (2.5), the normal equations are,

$$\mathbf{X}^t \cdot \mathbf{X} \cdot \mathbf{a} = \mathbf{X}^t \cdot \mathbf{Y}, \tag{2.7}$$

and the fitted parameters are,

$$\boldsymbol{a} = (\boldsymbol{X}^t \cdot \boldsymbol{X})^{-1} \cdot \boldsymbol{X}^t \cdot \boldsymbol{Y}, \tag{2.8}$$

if this inverse exists, and usefully if it is not an ill-conditioned problem. The more numerically stable solution Eq. (2.3) is by Penrose-Pseudo Inverse $\boldsymbol{X}^+$ of $\boldsymbol{X}$, as found by the Singular Value Decomposition, (SVD) [16],

$$\boldsymbol{a} = \boldsymbol{X}^+ \cdot \boldsymbol{Y}. \tag{2.9}$$

Note that if $\mathbf{Y}$ has just one column (in the case that the phase space, $M \subset \mathbb{R}^d$ is $d = 1$-dimensional), then this pseudo-inverse solution, $\boldsymbol{X}^+ \cdot \boldsymbol{Y}$ solves the Euclidean norm of the least squares solution $\| \cdot \|_2$ of Eqs. (2.3) and (2.4), (ordinary least squares, OLS), but in the case that it has multiple columns, $d > 1$, this is the so-called "multiple right hand side" (MRHS) regression problem and the optimization is in terms of the Frobenius-norm, $\| \cdot \|_F$. This is summarized,

$$\boldsymbol{a} = \boldsymbol{X}^+ \cdot \boldsymbol{Y} = \text{argmin} \|\mathbf{X} \cdot \mathbf{a} - \mathbf{Y}\|_*, \text{ where } * : \\ = 2 \text{ if } d = 1, \text{ and } * := F \text{ if } d > 1. \tag{2.10}$$

This standard fact is reviewed in the Appendix B, Section 9, as it is important for extension of these concepts to a Tikhonov regularized version which we will develop in subsequent sections.

3. Global models

In practice, the methods of the previous section can only be interpreted as a global modeling strategy in the following sense. Since for each point w to be forecast, a model $T|_w$ can be built from the data using the k-near neighbors following the kNN and modeling prescription Eqs. (2.4)–(2.9), then each of the matrice of fitted parameters, pointwise, are a function of w, $\mathbf{a}(w)$. Note that the step of collecting the k-near neighbors for each chosen w is a nonlinear thresholding function.

In this section we will rewrite the method of Eqs. (2.4)–(2.9), as a single equation so as to emphasize global aspects, but more importantly, as a precursor step to develop the regularity method in the following section. That is, given a finite set of w values to be forecast, formally rewriting the multiple instances of Eqs. (2.4)–(2.9) is possible in the form of a single equation. This is more than a formality, since it will allow us to expand to a single Tikhonov regularized form. Note that the rewriting of the collection of local models as a single global equation will allow us to move toward adding Tikhonov forms to emphasize regularity.

We reiterate the assumption that there are L states w where we will repeat Eqs. (2.4)–(2.9), to fit parameters $\mathbf{a}(w)$. Denote these, $\{w_j\}_{j=1}^L$, and assume that no w_j is amongst the fitting data to prevent triviality, $w_j \notin \{y_i\}_{i=0}^N$, $j = 1, .., L$. Building and collecting a local model of the transformation T, one for each w_j, reminds of building a manifold as an atlas of charts, [17].

For each w_j we write an equation of the form, Eq. (2.4), indexed to emphasize that the k-near neighbors involved have been collected to be near a specific data point w_j,

$$\boldsymbol{Y}(w_j) = \begin{bmatrix} y_{k_1^j+1}^t \\ y_{k_2^j+1}^t \\ \cdots \\ y_{k_k^j+1}^t \end{bmatrix} = \begin{bmatrix} | \\ y_{k^j+1}^t \\ | \end{bmatrix}, \text{ and } \boldsymbol{X}(w_j) \\ = \begin{bmatrix} 1 & y_{k_1^j}^t \\ 1 & y_{k_2^j}^t \\ \cdots & \\ 1 & y_{k_k^j}^t \end{bmatrix} = \begin{bmatrix} 1 & | \\ & y_{k^j}^t \\ & | \end{bmatrix}, \tag{3.1}$$

We have rewritten in this equation also in a second form for $\boldsymbol{Y}(w_j)$ and $\boldsymbol{X}(w_j)$ in terms of $y_{k^j}^t$ denoting running through the index values of the near neighbors of w_j, $0 \le k_i^j \le N$ for $i = 1, .., k$ by the vertical lines | as space holders, that they are a different notation for the matrices on the left but with less indexing. Finally, with these we write,

$$\text{Form 1: } \begin{bmatrix} | \\ y_{k^j+1}^t \\ | \end{bmatrix} = \begin{bmatrix} 1 & | \\ & y_{k^j}^t \\ & | \end{bmatrix} \cdot \mathbf{a}^j \tag{3.2}$$

which states the fitted $\mathbf{a}^j$ of $\boldsymbol{Y}(w_j) = \boldsymbol{X}(w_j) \cdot \mathbf{a}^j$, and superscripts j throughout, are just used as indexing rather than exponents. Summarizing sizes, Eq. (3.6) describes that a $k \times d$ matrix equals a $k \times d + 1$ matrix times a $d + 1 \times d$ matrix.

Now for clarity of presentation, consider the simplest scenario, that $d = 1$. As such, we are fitting a one-dimensional system, from observed data, (e.g. such as the logistic map). In this case $\mathbf{a} = [a_{1,1}, a_{2,1}]^t$ is a 2×1 matrix, denoting a line $y = a_{1,1} + a_{2,1} x$ fitting the mapping T of kNN model, one line for each set of $\{k_i^j\}_{i=1}^k$ for each w_j. We have called Eq. (3.2) to be "Form 1," since now we will write other equivalent forms. By "equivalent form", we mean that when expanded, they denote equations defining exactly the same set of points (line). That is, generally their solution sets are identi-

cal. We define another equivalent form to Eq. (3.6) as follows,

$$\text{Form 2: } \begin{bmatrix} 1 & 0 \\ a^j_{1,1} & a^j_{2,1} \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & .. & 1 \\ y^t_{k^j_1+1} & y^t_{k^j_2+1} & .. & y^t_{k^j_k+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & .. & 1 \\ y^t_{k^j_1} & y^t_{k^j_2} & .. & y^t_{k^j_k} \end{bmatrix} \tag{3.3}$$

It is immediate to see for this simple $d = 1$ case that this Form 2 is equivalent to Form 1. We further may rewrite Form 2 to become,

$$\text{Form 3: } A^j \cdot \mathbf{\Phi}^j = \mathbf{\Phi}^{j'}, \tag{3.4}$$

where, $\mathbf{\Phi}^j$ and $\mathbf{\Phi}^{j'}$, are the $2 \times k$ matrices of data in Eq. (3.3); the $'$ denotes the T image, so $y_{i+1} := y'_i = T(y_i)$. The important distinction between Form 1 and Form 2 is that while they denote statements of equivalent solution sets, with equivalent parameters encoded in a 2×1 matrix $\mathbf{a}$, or alternatively stated in 2×2 matrix form of Eq. (3.3), A^j, the "dummy" 1's included in the state,

$$\phi = [1\ y]^t, \text{ so, } \mathbf{\Phi}^j = [\phi_{k^j_1}, \phi_{k^j_2}, .., \phi_{k^j_k}], \tag{3.5}$$

and likewise for $\mathbf{\Phi}^{j'}$, will allow for a better interpretation of the global fit in an operator setting; the resulting global operator actually represents a transfer operator which we will explore in future works.

Now we proceed to build a global model, and then following that, we will regularize. We label with a "G" for global version of Form 1, across all the data $\{w_j\}_{j=1}^L$ to be fitted,

$$\text{Form 1G: } \begin{bmatrix} | \\ y^t_{k^1+1} \\ | \\ - \\ | \\ y^t_{k^2+1} \\ | \\ - \\ : \\ - \\ | \\ y^t_{k^L+1} \\ | \end{bmatrix}_{kL \times d} = \begin{bmatrix} | & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 1\ y^t_{k^1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ | & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ - & - & - & - \\ \mathbf{0} & | & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1\ y^t_{k^2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & | & \mathbf{0} & \mathbf{0} \\ - & - & - & - \\ & : & & \\ - & - & - & - \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & | \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 1\ y^t_{k^L} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & | \end{bmatrix}_{kL \times L(d+1)} \cdot \begin{bmatrix} \mathbf{a}^1 \\ \mathbf{a}^2 \\ : \\ \mathbf{a}^L \end{bmatrix}_{L(d+1) \times d}, \tag{3.6}$$

where we have included the statement of sizes of the matrices in the equation. Thus we see that Form 1G is uncoupled between the individual collections k^j. So this is simply L statements of uncoupled versions of Form 1, Eq. (3.6), one for each $j = 1, .., L$, stacked and offset. The Tikhonov regularity, to be imposed in the next section, will be the concept that will end up coupling the blocks of the equations. For now, labeling these three matrices, in order, we solve,

$$\mathcal{Y} = \mathcal{X} \cdot \mathbb{A}, \text{ with either least squares or Frobenius} - \text{norm solution,}$$
$$\mathbb{A} = \mathcal{X}^+ \cdot \mathcal{Y}, \tag{3.7}$$

is again equivalent to Eq. (2.9) L-times, by lack of coupling between any groups of the k-neighbor pairings. Again as in Eq. (2.10), and reviewed in Appendix B 9, Eq. (3.7) yields either the OLS solution minimizing 2-norm objective if $d = 1$, but the Frobenius-norm objective if $d > 1$ corresponding to MRHS.

So what is the advantage of writing a global Form 1G like this when it reduces to smaller and more computationally efficient Form 1, L-times? There are two advantages. Form 1G allows us perspective to describe global transfer operator through the following Form 2G, to be explored in future works. The second advantage which we will emphasize here is the possibility of regularization discussed in the next section.

Let,

$$\text{Form 2G: } \mathcal{A} \cdot \mathbf{\Phi} = \mathbf{\Phi}' \text{ globally extends Eq. (3.3) across, } j = 1,..,L, \text{ by,} \tag{3.8}$$

$$\mathcal{A}_{L(d+1) \times L(d+1)} = diag(A^1, A^2, .., A^L), \mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}^1 \\ - \\ \mathbf{\Phi}^2 \\ - \\ : \\ - \\ \mathbf{\Phi}^L \end{bmatrix}_{L(d+1) \times k}, \mathbf{\Phi}' = \begin{bmatrix} \mathbf{\Phi}^{1'} \\ - \\ \mathbf{\Phi}^{2'} \\ - \\ : \\ - \\ \mathbf{\Phi}^{L'} \end{bmatrix}_{L(d+1) \times k}, \tag{3.9}$$

Note that A^j parameter matrices, Φ^j and Φ^j data matrices and the state variable ϕ are defined in Eqs. (3.3)-(3.5). The *diag* statement used is as usual the $L(d+1) \times L(d+1)$ matrix with smaller $d+1 \times d+1$ matrices A^j on the diagonal. Thus $\mathbf{\Phi}' = \mathcal{A} \cdot \mathbf{\Phi}$ has an interesting interpretation as a transfer operator.

As already mentioned in the introduction, the lack of regularity is clearer upon inspection of Eq. (3.6). There is nothing explicit to connect the estimated parameter values to each other besides the implicit concept that the near neighbors distances should varying continuously. In [11] we showed that this connection suggests continuity in fitted parameters but not necessarily any greater regularity. One would expect at least continuity in forecasts of $T(w)$, and more likely higher derivatives are continuous reflecting the true regularity of the underlying T, if is assumed that $T \in C^r(M)$ for some $r \geq 0$. Stated generally, if we have a large data set, $\{y_i\}_{i=0}^N$ sampled from an ergodic T, then we would expect that the kNN method will converge to a continuous solution descriptive of T, with respect to increasing N. However, there will tend to be jumps in practice, since as w varies only slightly, then the specific k-near neighbors, $\{y_{k_j}\}_{j=1}^k$ will not change, and so the local model will vary only slightly. On the other hand, two nearby w values may have different sets of k neighbors and so we see there can be jumps of fitted parameters, and correspondingly the model will jump, as will forecasts. So in the following we discuss how to reward regularity by coupling nearby forecasts within the Tikhonov regularity formalism.

4. Tikhonov regularization

The goal of conferring continuity or higher ordered regularity of the resultant predictions can be stated within the context of Tikhonov regularity theory. Given $\mathcal{X} \in \mathbf{R}^{kL \times L(d+1)}$ from Eq. (3.6), then suppose a regularization matrix, $\mathcal{B} \in \mathbf{R}^{L(d+1) \times L(d+1)}$. For initial presentation, let $d = 1$ to describe the standard Tikhonov formulation of an optimization problem. Proofs and background of this part of convex optimization theory leading to closed formula solutions of a Tikhonov regularized problem can be found in variational calculus terms in [18], and also in [19–21], but in this work

we follow the finite vector space representation in terms of matrix theory computations as written and proved in [16], in particular on pp. 309, Theorem 6.1.1, that we now review. Let,

$$\min_{\mathbb{A}} \|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_2^2+\lambda\|\mathcal{B}\cdot\mathbb{A}\|_2^2. \tag{4.1}$$

whose solution can be designed to emphasize the desirable regularity properties of $\mathbb{A}$, depending on how we define $\mathcal{B}$, and depending on the specific choice of λ. We reiterate that since temporarily we have assumed $d=1$, then in fact $\mathbb{A}$ is of size, $L(d+1)\times d = 2L\times 1$; see Eqs. (3.6) and (3.7). It has been shown, which we draw from [16] specialized to this setting, that the normal equations follow as,

$$(\mathcal{X}^t\cdot\mathcal{X}+\lambda\mathcal{B}^t\cdot\mathcal{B})\cdot\mathbb{A}=\mathcal{X}^t\cdot\mathcal{Y}, \tag{4.2}$$

are nonsingular when $nul(\mathcal{X})\cap nul(\mathcal{B})=\{0\}$. Solution of this problem can be described by inverting the left hand side matrix, but it is numerically practical by means that Van Loan [16] describes as Generalized Singular Value Decomposition (GSVD) for such problems with two matrices. On the other hand, rewrite Eq. (4.1) with,

$$C=\mathcal{X}, x=\mathbb{A}, b=\mathcal{Y}, \text{ and } L=\mathcal{B}, \tag{4.3}$$

as,

$$\min_x \|C\cdot x-b\|_2^2+\lambda\|L\cdot x\|_2^2=\min_x\left\|\begin{bmatrix}C\\ \sqrt{\lambda}L\end{bmatrix}\cdot x-\begin{bmatrix}b\\0\end{bmatrix}\right\|_2^2, \tag{4.4}$$

(which we see is equivalent by expanding terms $[C^t|\sqrt{\lambda}L^t]\cdot\begin{bmatrix}C\\ \sqrt{\lambda}L\end{bmatrix}\cdot x$ and $[C^t|\sqrt{\lambda}L^t]\cdot\begin{bmatrix}b\\0\end{bmatrix}$). Then we get the second version of the standard Tikohonov problem which becomes an OLS, to estimate

$$\min\|D\cdot x-e\|_2^2, \text{ by, } x^{LS}=D^+\cdot e, \text{ with,}$$
$$D=\begin{bmatrix}C\\ \sqrt{\lambda}L\end{bmatrix}, \text{ and } e=\begin{bmatrix}b\\0\end{bmatrix}. \tag{4.5}$$

Therefore, for our general problem of MRHS, $d>1$, it becomes clear how to generalize the Tikohonov problem, but in terms of the Frobenius norm. Problem Eqs. (3.6) and (3.7) becomes,

$$\min_{\mathbb{A}} \|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_F^2+\lambda\|\mathcal{B}\cdot\mathbb{A}\|_F^2. \tag{4.6}$$

We have just shown that Tikohonov-Frobenius form of the problem has a general solution by pseudo-inverse of an appropriately stated matrix, considering the discussion of the relationship between 2-norm and F-norm solutions in Appendix B 9 depending on the number of columns of the RHS, which we summarize by the following:

The general Tikhonov regularization problem,

$$\mathbb{A}_{TR}=argmin_{\mathbb{A}}\|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_*^2+\lambda\|\mathcal{B}\cdot\mathbb{A}\|_*^2. \tag{4.7}$$

has a solution as a Penrose pseudo-inverse as follows,

$$\mathbb{A}_{TR}=\begin{bmatrix}\mathcal{X}\\ \sqrt{\lambda}\mathcal{B}\end{bmatrix}^+\cdot\begin{bmatrix}\mathcal{Y}\\ \mathbf{0}\end{bmatrix} \tag{4.8}$$

and the minimization is in terms of $*=2$-norm if $d=1$, and $*=F$-norm if $d>1$.

Note that, $\begin{bmatrix}\mathcal{X}\\ \sqrt{\lambda}\mathcal{B}\end{bmatrix}$ is $(k+d+1)L\times L(d+1)$ as it is a stacking of $kL\times L(d+1)$ and $L(d+1)\times L(d+1)$ matrices, and likewise, $\begin{bmatrix}\mathcal{Y}\\ \mathbf{0}\end{bmatrix}$ is $(k+d+1)L\times d$ and the zeros matrix must be of appropriate size, $L(d+1)\times L(d+1)$.

5. Defining a Laplacian for regularity

It remains to discuss how to design the regularity matrix $\mathcal{B}$ to confer the level of continuity regularity expected in the inferred system T that we are modeling. There are several different common ways to develop L within Tikohonov theory. For example, choosing the identity matrix, $\mathcal{B}=I$ penalizes "large" x in addition to the data fidelity which emphasizes the fitting $C\cdot x\approx b$, and this yields what is also called "ridge regression". On the other hand, choosing,

$$\mathcal{B}=\begin{bmatrix}2&-1&0&0&\dots&0&-1\\ -1&2&-1&0&\dots&0&0\\ 0&-1&2&-1&\dots&0&0\\ \dots&&&&&&\\ 0&0&\dots&&0-1&2&-1\\ -1&0&0&\dots&0&-1&2\end{bmatrix} \tag{5.1}$$

is a commonly used Laplacian that emphasizes regularity of the approximate second derivative, and we will adjust this concept to meet our needs. This matrix $\mathcal{B}$ rewards adjacent indexed values to have small differences between them. The key word here is "adjacent", and this requires clarification to emphasize that those points to be regularized must respect the underlying topology of the space, as sampled by the data orbit that may not have a specific structure. This particular $\mathcal{B}$ is not likely the correct matrix to describe the regularity that is appropriate for sampled data w_j, and so in what follows we describe how to build a comparably defined Laplacian matrix $\mathcal{B}$ that respects the actual proximity between data points to reflect the underlying topology in the resultant estimation. Albeit, this concept is simple to describe, but the technical indexing details are not as simple as perhaps one would expect, due to necessity of keeping track of all the near neighbors' indices of the unstructured data $\{y_i\}_{i=0}^N$. So we have included these details in the Appendix A 8 as to how to form $\mathcal{B}$ reflecting s-nearest neighbors $\{k_i^j\}$ from the set $\{w_j\}_{i=1}^L$ for each w_j, and then associating this to a repeated Laplacian regularity matrix, $\mathcal{B}$.

6. Illustrative results

There are several general schemes of choosing best λ level of regularization in Eq. (4.7) to balance the effects of over regularization of a large λ against over fitting when λ is too small which puts too much emphasis on data fit with not enough attention to regularity. There are many schemes within the practice of regularity theory, most notably the U-curve method and the L-curve method, but we feel that in the dynamical systems setting of prediction, a simple scheme as follows can be described, considering Fig. 1 as an example. Here we see the balance between $\|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_*^2+\lambda\|\mathcal{B}\cdot\mathbb{A}\|_*^2$ as λ is swept through a logarithmic scale. The best least squares fit $\|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|$ is expected to occur when no other competing concern is balanced into the full functional, which is at $\lambda=0$. So clearly we see that the blue curve is monotonically increasing with λ. But regularity is the competing term in the Tikhonov form, and so $\|\mathcal{B}\cdot\mathbb{A}\|_*^2$ decreases by design as it is included increasingly more importantly as λ increases. Nonetheless when multiplied by λ, the full regularity term $\lambda\|\mathcal{B}\cdot\mathbb{A}\|_*^2$ initially increases with λ, for a resultant sum the green curve $\|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_*^2+\lambda\|\mathcal{B}\cdot\mathbb{A}\|_*^2$, which while it is at a minimum at smallest λ, one must design how important to consider the second factor, the regularity is for their problem.

Our interpretation of this balance of choosing λ is generally that quality prediction is most important, but with a small value of λ, this can greatly improve the regularity costing only a minor loss of data fidelity (prediction). Thus we interpret that $\lambda=10^{-1}$ describes such a sweet spot. In broader terms, Fig. 1 can be described as revealing both the strengths and weaknesses of bring-

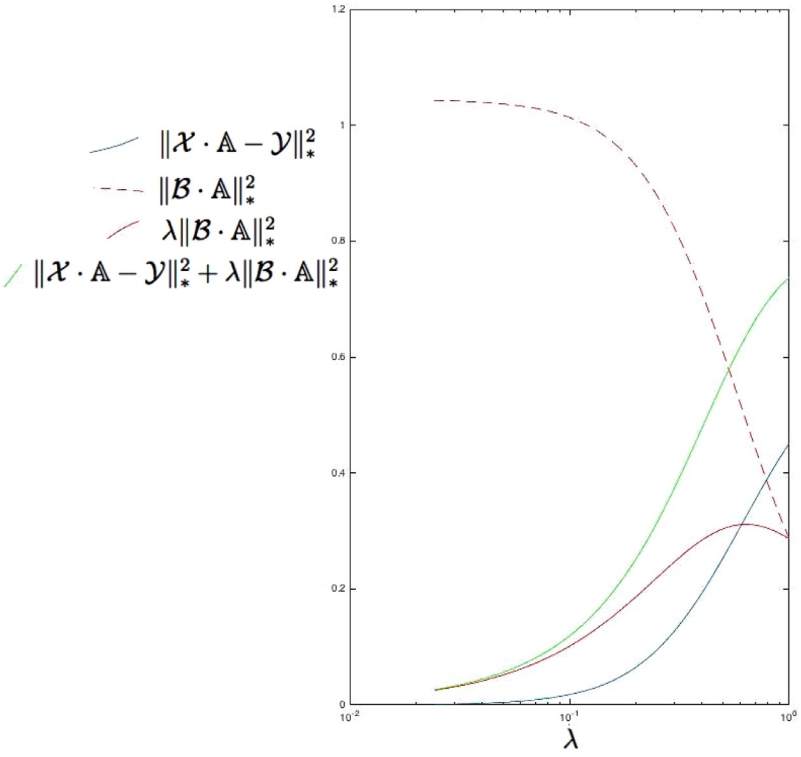

Fig. 1. Tikhonov regularized predictions curves as a function of the regularity parameter λ are varied through a logarithmic scale. Considering the Tikhonov regularity expression, Eq. (4.7), (Blue) Data fidelity $\|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_*^2$ illustrating the quality of fit emphasizes least squares for small λ but sacrifices this part of the full objective function as λ increases. (Red dashed) Regularity $\|\mathcal{B}\cdot\mathbb{A}\|_*^2$ as λ increases becomes more important in the optimization problem and thus decreases, but (Red) $\lambda\|\mathcal{B}\cdot\mathbb{A}\|_*^2$ is nonetheless increasing with λ starting at small λ, resulting in (Green) a total cost of the objective function balancing the two competing terms in, $\|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_*^2+\lambda\|\mathcal{B}\cdot\mathbb{A}\|_*^2$. In this instance, considering the blue data fidelity to be of primary important (model fit) to be of high quality then inferring a small degree of regularity red dashed curve, suggests a good balance of these at roughly $\lambda=0.1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ing a global perspective to prediction, explicitly regularizing, and as contrasted to traditional local modeling. In this term, consider that when $\lambda=0$ then regularity is not emphasized and we have the special case which is traditional local forecasting as measured on the global scale by averaging across the sample space, as represented by the norm bars. Notice that $\lambda=0$ leads to the smallest forecast error on average, $\|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_*^2$ (also called "data-fidelity" in the language of Tikhonov regularization), but also low regularity, $\|\mathcal{B}\cdot\mathbb{A}\|_*^2$ meaning that forecasts can vary more significantly between nearby points.

On the other hand, we can over regularize as we see that for larger λ, say $\lambda=1$, then regularity is large which may be a desirable property in general, but at both the cost of good forecasts since we see errors in terms of large data-fidelity (globally averaged error) $\|\mathcal{X}\cdot\mathbb{A}-\mathcal{Y}\|_*^2$, are bad and perhaps primary. Also, even a little too much regularity may not be desirable if it is beyond what is natural to the system, but how much is too much would be unknown. Therefore we recommend a relatively conservative small $\lambda=0.1$ since this tends to improve regularity without a great global cost of fidelity. While the standard principled "*U*-curve" method, [19] may not be appropriate in this setting since good forecasts are so important we recommend the more conservative smaller λ values.

7. Conclusions

Prediction of chaotic and complex time series is clearly a topic of intrinsic interest and so it is not a surprise that it has been a long running research thread within the community to extend methods to interpret and predict from observed data. While local methods based on regression of local polynomial models have been a leading method for decades, especially when the data is represented in time delay embedding coordinates. We have discussed how such predictions are lacking in global perspective and as such they can lack in global regularity one may expect of the underlying physical system. Not only have we presented a method to cope with global modeling, but since it is based on the well founded Tikhonov regularity theory, our approach suggests naturally that suppressing spurious fluctuations should be approached within aspects that incorporate both local and global information within a natural framework. While we expect our method applies equally well to predicting data from discrete time maps, as well as flow maps from differential equations, even if data is presented in time delay embedding coordinates, we have not included the later theory here since it is so well covered in many places. This leaves the focus of this paper with the aspect which is new. The new element here is how to link an atlas of local models and then to impose regularity upon these models within the Tikhonov framework. As the imposed regularity λ is adjustable then an apriori assumption of the degree of regularity can be imposed as a principle of least alteration to the local predictions while positively improving the global regularization.

Acknowledgments

This work was supported by the Office of Naval Research under grant contracts N00014-11-1-0715, and N00014-15-1-2093.

Appendix A. Regularity Laplacian for unstructured data

From Section 4, we require a regularity matrix, $\mathcal{B}$ that agrees with the association structure of the data $\{w_j\}_{i=1}^L$. All notation is the same as in that section. We must define our *Q*, Eq. (5.1), to emphasize the actual proximity of the points w_j in the underlying

phase space, reflected in each jth row of Q. We will build a graph Laplacian that rewards regularity when underlying points w_i and w_j are close in the phase space M. Recall that each $d+1$ successive rows of $\mathbb{A}$, from Eqs. (3.7) and (4.1), $1 \le j \le L$, reflect the reference points successively $\{w_j\}_{j=1}^{L}$, so that each row corresponds to a model developed for a w_j. Consistent with the notation of Eq. (3.1) for the k-near neighbors, kNN, let $\{k_i^j\}_{i=1}^{s}$ be the indices selecting the points s-closest to w_j from $\{w_j\}_{j=1}^{L}$ (not including w_j itself) in order of distance, which are $\{w_{k_i^j}\}_{i=1}^{s} \subset \{w_j\}_{j=1}^{L}$. We can catalogue this adjacency of points by an adjacency matrix the generates a directed graph. Let G be an $L \times L$ matrix that consists of all zeros, except for certain indices that we define to be ones as follows,

$$G_{j,k_i^j} = 1 \text{ since } y_{k_i^j} \text{ is the } i^{\text{th}} \text{ closest point to } w_j, \quad 1 \le i \le s, 1 \le j \le L, \tag{8.1}$$

reflecting a connection when two points are close. It follows that the degree of the jth vertex of this generated graph is the jth row sum of G, but by construction, these are all s,

$$deg(j) = \sum_{i=1}^{L} G_{j,i} = s, \text{ for all } 1 \le j \le L. \tag{8.2}$$

Finally the graph Laplacian is constructed from the adjacency matrix by the standard computation,

$$Q = diag(deg(j)) - G, \tag{8.3}$$

that is the degrees of each vertex on the diagonal and negative ones at each position of G that is an adjacency one. And since,

$$deg(j) = s, \tag{8.4}$$

then,

$$Q = diag(deg(j)) - G = sI - G, \tag{8.5}$$

where I is the identity matrix. Of course one does not need to use $s = k$, where recall that we used k-near neighbors for the regression steps leading to Eq. (3.1). Notice for sake of example that if we choose, $s = 2$ and in the very special case that all of the data is linearly ordered on a circle, then Q constructed by Eqs. (8.1)–(8.3) is similar to that standard graph Laplacian of a linear lattice shown in Eq. (5.1).

The general scenario of unstructured data in a space of several dimensions, respecting the relative closeness of the sample points in the underlying phase space yields a graph Laplacian Q with perhaps not readily apparent structure. To allow for the structure of each one row in the graph of G, Eq. (8.1) encoded in Laplacian Q, to emphasize continuity between each of the successive corresponding $d+1$ rows of $\mathbb{A}$ corresponding to the same point in phase space w_j. Since $\mathbf{A}$ is $L(d+1) \times L(d+1)$, then so $\mathcal{B}$ must be the same size, but we wish $\mathcal{B}$ to describe the regularity encoded by Q just defined in Eq. (8.3).

Appendix B. On OLS, and the Frobenius norm.

Here we briefly review a well known fact, that Eq. (2.10) describes the relationship between ordinary least squares (OLS) when the right side data is a vector, where the minimizer is in terms of the two-norm of the residual. The general problem that arises in this paper occurs when the right hand side data has multiple columns, (called multiple right hand side, MRHS). The residual of the analogously formed minimizer is in terms of the Frobenius norm. Note that we have changed notation in this appendix to reflect standard use of least squares to solve problems $Ax = b$.

Let $\mathbf{A}$, be $m \times n$, and $\mathbf{b}_i$ be $n \times 1$ be given data, and $m > n$ as the overdetermined case of more rows than columns, (stated more carefully, the column space of $\mathbf{A}$ is full) and we wish to solve,

$$\mathbf{A} \cdot \mathbf{x}_i \approx \mathbf{b}_i, \tag{9.1}$$

for any one of $\mathbf{A}$ fixed data matrix, but several data vectors, $\mathbf{b}_i$, $i = 1, .., p$. Then for each i, one may solve the least squares problem,

$$\mathbf{x}_i^{LS} = \text{argmin} \|\mathbf{A} \cdot \mathbf{x} - \mathbf{b}_i\|_2^2. \tag{9.2}$$

Note that this may be solved by either of several popular methods, theoretically equivalently but algorithmically differently, with sometimes important different computational complexity as well as stability issues each. These include QR decomposition, SVD[16] to Penrose-pseudo inverse $\mathbf{A}^+$, or directly by normal equations, $(\mathbf{A}^t\mathbf{A}) \cdot \mathbf{x} = \mathbf{A}^t \cdot \mathbf{b}_i$. In any case, assume we have a suitable minimizer for optimization, for $\mathbf{x}_i^{LS}$, for each $i = 1, .., p$. Then the following is straightforward to describe the "least squares-like" solution of the corresponding matrix problem,

$$\mathbf{A} \cdot \mathbf{X} \approx \mathbf{B}, \tag{9.3}$$

called the MRHS (multiple right hand sided) problem:

Theorem: If $\mathbf{A}$ is $m \times n$, and $\mathbf{B} = [\mathbf{b}_1 | \dots | \mathbf{b}_p] \mathbf{x}_i^{LS} = \text{argmin} \|\mathbf{A} \cdot \mathbf{x}_i - \mathbf{b}_i\|_2$, for each $i = 1, .., p$, and $p > 1$, then if $\mathbf{X}^{LS} = [\mathbf{x}_1^{LS} | \dots | \mathbf{x}_1^{LS}]$ is a minimizer in the sense of the Frobenius-norm, rather than the 2-norm, meaning,

$$\mathbf{X}^{LS} = \text{argmin} \|\mathbf{A} \cdot \mathbf{X} - \mathbf{B}\|_F = \mathbf{A}^+ \cdot \mathbf{B}. \tag{9.4}$$

Recall that the Frobenius-norm $\|\mathbf{C}\|_F$ of a $m \times n$ matrix $\mathbf{C}$ is defined,

$$\|\mathbf{C}\|_F^2 = \sum_{i=1, j=1}^{m,n} |C_{i,j}|^2, \tag{9.5}$$

and the 2-norm $\|\mathbf{v}\|_2$ of a $m \times 1$, vector, is defined, $\|\mathbf{v}\|_2^2 = \sum_{i=1}^{m} v_i^2$.

So the proof of this fact follows the partitioning of the matrices. If the residuals squared of each, $i = 1, .., p$,

$$\|\mathbf{r}_i\|_2^2 = \|\mathbf{A} \cdot \mathbf{x}_i - \mathbf{b}_i\|_2^2. \tag{9.6}$$

Then,

$$\begin{aligned} \|\mathbf{A} \cdot \mathbf{X} - \mathbf{B}\|_F^2 &= \|\mathbf{A} \cdot [\mathbf{x}_1 | \dots | \mathbf{x}_p] - [\mathbf{b}_1 | \dots | \mathbf{b}_p]\|_F^2 \\ &= [\mathbf{A} \cdot \mathbf{x}_1 - \mathbf{b}_1 | \dots | \mathbf{A} \cdot \mathbf{x}_p - \mathbf{b}_p\|_F^2 = \\ &= \|\mathbf{r}_1 | \dots | \mathbf{r}_p]\|_F^2 = \sum_{j=1}^{m} \sum_{i=1}^{p} \|[\mathbf{r}_i]_j\|_2^2 = \sum_{i=1}^{p} \|\mathbf{r}_i\|_2^2, \end{aligned} \tag{9.7}$$

where $\mathbf{r}_i$ denotes the ith residual vector, and $[\mathbf{r}_i]_j$ denotes the jth (scalar) indexed value of that vector.

References

[1] Abarbenel H, Brown R, Sidorowich J, Tsimring L. The analysis of observed chaotic data in physical systems. Rev Mod Phys 1993;65(4).
[2] Farmer J, Sidorowich J. Predicting chaotic time series. Phys Rev Lett 1987;59:845.
[3] Takens F, Rand D, Young L-S. Dynamical systems and turbulence. In: Lecture notes in mathematics, 898. Springer, Berlin; 1980. p. 366.
[4] Kantz H, Schreiber T. Nonlinear time series analysis. New York: Cambridge U. Press; 1997.
[5] Farmer J, Sidorowich J. Exploiting chaos to predict the future and reduce noise. In: Lee YC, editor. Evolution, learning, and cognition. World Scientific Press; 1988b. p. 277.
[6] Abarbanel H. Analysis of observed chaotic data. New York: Springer; 1996.
[7] Weigenbend AS, Gershenfeld NA. Times series prediction: forecasting the future and understanding the past. Reading, MA: Addison-Wesley; 1993.
[8] Sauer T, Yorke J, Casdagli M. Embedology. J Stat Phys 1991b;65(3/4.).
[9] Eckmann J, Ruelle D. Ergodic theory of chaos and strange attractors. Rev Mod Phys 1985b;57:617.
[10] Farmer J, Sidorowich J. Exploiting chaos to predict the future and reduce noise. In: Lee YC, editor. Evolution, learning, and cognition. World Scientific Press; 1988c. p. 277.
[11] Yao C, Bollt E. Modeling and nonlinear parameter estimation with kronecker product representation for coupled oscillators and spatiotemporal systems. Physica D 2007;227(1):78–99.

[12] Fan J, Gijbels I. Local polynomial modelling and its applications. London: Chapman & Hall; 1996.
[13] Bollt E. Model selection, confidence, and scaling in predicting chaotic time-series. Int J Bifurcation Chaos (IJBC) Appl Sci Eng 2000;10(6):1407–22.
[14] Draper NR, Smith H. Applied regression analysis. New York: John Wiley & Sons; 1981.
[15] Neter J, Kutne MH, Nachtsheim CJ, Wasserman W. Applied linear regression models, third edition. Chicago: IRWIN; 1996.
[16] Golub GH, Loan FV. Matrix computations. Baltimore: Johns Hopkins University Press; 2013.
[17] Munkres J. 2000. Topology (2nd Edition) 2nd Edition, Pearson; 2 edition.
[18] Vogel CR. Computational methods for inverse problems. SIAM, Frontiers in mathematics; 2002.
[19] Basnayake R, Bollt EM. A multi-time step method to compute optical flow with scientific priors for observations of a fluidic system. In: Froyland Gary, Bahsoun Wael, Bose Chris, editors. BIRS book chapter, in ergodic theory, open dynamics, and coherent structures, 70. Springer proceedings in mathematics statistics; 2014. p. 59–79.
[20] Luttman A., Bollt E.M., Basnayake R., Kramer S., Tullaro N.B. A stream function framework for estimating fluid flow from digital imagery, CHAOS. 2013. 23, 033134.
[21] Luttman A, Bollt E, Basnayake R, Kramer S. A stream function approach to optical flow with applications to fluid transport dynamics. In: Applied mathematics and mechanics, vol. 11; 2012. p. 855–6.