Improvements to local projective noise reduction through higher order and multiscale refinements

Jack Murdoch Moore, Michael Small, and Ali Karrech

Citation: [Chaos](#) **25**, 063114 (2015); doi: 10.1063/1.4922473

View online: <http://dx.doi.org/10.1063/1.4922473>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/chaos/25/6?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

[Smooth local subspace projection for nonlinear noise reduction](#)

[Chaos](#) **24**, 013121 (2014); 10.1063/1.4865754

[Noise reduction by recycling dynamically coupled time series](#)

[Chaos](#) **21**, 043110 (2011); 10.1063/1.3645969

[Noise-induced order in the chaos of the Belousov–Zhabotinsky reaction](#)

[J. Chem. Phys.](#) **129**, 014508 (2008); 10.1063/1.2946710

[Improvement of speech recognition by nonlinear noise reduction](#)

[Chaos](#) **17**, 023121 (2007); 10.1063/1.2743614

[Nonlinear noise reduction through Monte Carlo sampling](#)

[Chaos](#) **8**, 775 (1998); 10.1063/1.166363

Improvements to local projective noise reduction through higher order and multiscale refinements

Jack Murdoch Moore,^{1,2,a)} Michael Small,³ and Ali Karrech⁴

¹*School of Earth and Environment, The University of Western Australia, Crawley, WA 6009, Australia*

²*Mineral Resources Flagship, CSIRO, Australian Resources Research Centre, Kensington, WA 6152, Australia*

³*School of Mathematics and Statistics, The University of Western Australia, Crawley, WA 6009, Australia*

⁴*School of Civil, Mining and Environmental Engineering, The University of Western Australia, Crawley, WA 6009, Australia, The University of Western Australia, Crawley, WA 6009, Australia*

(Received 24 November 2014; accepted 25 May 2015; published online 15 June 2015)

The broad spectrum characteristic of signals from nonlinear systems obstructs noise reduction techniques developed for linear systems. Local projection was developed to reduce noise while preserving nonlinear deterministic structures, and a second order refinement to local projection which was proposed ten years ago does so particularly effectively. It involves adjusting the origin of the projection subspace to better accommodate the geometry of the attractor. This paper describes an analytic motivation for the enhancement from which follows further higher order and multiple scale refinements. However, the established enhancement is frequently as or more effective than the new filters arising from solely geometric considerations. Investigation of the way that measurement errors reinforce or cancel throughout the refined local projection procedure explains the special efficacy of the existing enhancement, and leads to a new second order refinement offering widespread gains. Different local projective filters are found to be best suited to different noise levels. At low noise levels, the optimal order increases as noise increases. At intermediate levels second order tends to be optimal, while at high noise levels prototypical local projection is most effective. The new higher order filters perform better relative to established filters for longer signals or signals corresponding to higher dimensional attractors. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4922473>]

Linear filters degrade the beautifully intricate structure of the (aptly named) attractors which characterise chaotic signals. Local projection was developed two decades ago as a method to reduce noise with minimal damage to the attractor structure. One decade ago, a second order refinement to local projection was proposed. It reduces distortion during noise reduction by accommodating attractor curvature. This paper uncovers an analytic motivation for the improvement, and extends it to accommodate higher order terms and to simultaneously consider the different information available at various resolutions. Statistical considerations are used to explicate the special efficacy of the second order enhancement, and to formulate a refinement of local projective noise reduction which is usually more effective than established methods. The different refinements identified and reviewed in this work are found to be appropriate to different types of data and noise levels.

I. INTRODUCTION

The broad spectrum characteristic of signals from nonlinear systems complicates the process of noise reduction which is often necessary in dealing with physical measurements. Linear filters obscure the structure underlying nonlinear signals (Theiler and Eubank, 1993). The filtering of

nonlinear signals requires techniques specifically developed for the purpose: nonlinear filters, reviewed by Kostelich and Schreiber (1993). A useful class of these is local projective filters, reviewed by Grassberger *et al.* (1993). Local projective filters correct the location of points in phase space informed by the locations of neighbouring points. Since their inception by Cawley and Hsu (1992); Sauer (1992), local projective filters have found continuous employment in spheres ranging from the supernally immediate one containing our neurons (Wan *et al.*, 2006) to the study of light curves from distant stars within the discipline which encompasses our entire universe (Jevtić *et al.*, 2012).

The general framework of local projection is frequently adapted to fit precisely the idiosyncrasies of particular systems. Wan *et al.* (2006) applied local projection to the wavelet transform of an electroencephalography (EEG) signal to remove interferential artifacts.

To mitigate the absence of pure determinism in electrocardiograms (ECGs), Kotas (2007) aligned individual heartbeats in time instead of space. This allowed her to separate maternal and foetal ECGs. Later, she extended the method by associating points in different heartbeats according to both temporal and spatial similarity (Kotas, 2008). Via this metric she cleaned a contaminated ECG (Kotas, 2008). More recently, Jevtić, Schweitzer, and Stine (2011) optimised parameters for local projective filtering of stellar light curves and applied their findings to noisily observed variations in light intensity as extrasolar planets pass between their star

^{a)}Electronic mail: jack.moore@research.uwa.edu.au

and the observer. They demonstrated that it is, in principle, possible to detect the presence of multiple extrasolar planets of disparate size (Jevtić *et al.*, 2011). Chelidze (2014) projected strands (sequences of temporally successive points in phase space) according to the relative position of neighbouring strands. By thus simultaneously exploiting temporal and spatial correlations, he attenuated noise in highly sampled flow-like data more effectively than classical local projection (Chelidze, 2014).

Kantz and Schreiber (2004) achieved more precise results by better accommodating attractor geometry. They chose as the origin of the projective subspace a weighted sum of the first and second order centres of mass of local phase space neighbourhoods (see Figure 1). Kantz and Schreiber (2004) credited Sauer (1992) with the inspiration for their improvement. He translated the adjustments made in each filtering step such that they had a mean of zero over a local neighbourhood (Sauer, 1992).

This paper describes an analytic motivation for the refinement of Kantz and Schreiber (2004) from which further higher order and multiple scale refinements naturally follow. In many cases, the extensions can reduce noise more effectively than established local projective filters. At low and intermediate levels of noise, a particular dual scale filter tends to be more effective than existing filters. At low noise levels, the new higher order filters become more advantageous relative to established filters as noise level increases. The new higher order filters also perform better relative to existing filters for longer signals or signals corresponding to higher dimensional attractors.

Section II describes established forms of local projection, including the second order refinement of Kantz and Schreiber (2004). Section III derives the higher order and multiscale filters proposed in this paper. It goes on to explore the way in which considerations of independently distributed measurement errors may be incorporated into previous multiscale and higher order results. Section IV introduces the classical nonlinear systems used to assess the filters. It also details the specific parameters of the local projective filtering process applied to each time series, which are those optimised by Grassberger *et al.* (1993) for prototypical local projective filters. Additionally, Sec. IV describes the application of the two most successful benchmarked filters to data from an NMR laser experiment. Section V summarises and suggests ways in which results could be extended.

FIG. 1. Following Figure 10.3 of Kantz and Schreiber (2004). The centres of mass of local neighbourhoods (○) do not, in general, lie exactly on the attractor. The second order centres of mass (●) tend to aberrate roughly twice as much.

II. BACKGROUND

A. Local projection

Consider a series of N scalar measurements $s(1), s(2), \dots, s(N) \in \mathbb{R}$ from a chaotic system. The $N - (m - 1)T$ vectors $\mathbf{x}_i \triangleq (s(i), s(i + T), \dots, s(i + (m - 1)T))^T \in \mathbb{R}^m$ are called m -histories and can be used to reconstruct phase space (Takens, 1981; Sauer *et al.*, 1991; and Noakes, 1991).

Local projection begins with the transformation of the m -histories $\mathbf{x}_i$ via an invertible matrix $\mathbf{R}$ according to

$$\mathbf{z}_i \triangleq \mathbf{R}\mathbf{x}_i. \quad (2.1)$$

Cawley and Hsu (1992); Sauer (1992) left $\mathbf{R}$ as the identity matrix, so that $\mathbf{z}_i = \mathbf{x}_i$. Local projection of an m -history $\mathbf{x}_i$ involves replacing $\mathbf{z}_i$ with $\mathbf{z}_i^o + \mathbf{P}_i^{(q)}(\mathbf{z}_i - \mathbf{z}_i^o)$ (or, equivalently, $\mathbf{x}_i$ with $\mathbf{x}_i^o + \mathbf{R}^{-1}\mathbf{P}_i^{(q)}(\mathbf{R}(\mathbf{x}_i - \mathbf{x}_i^o))$), where $\mathbf{P}_i^{(q)}$ is the matrix representing projection onto the subspace spanned by the eigenvectors corresponding to the $m - q$ largest eigenvalues of the covariance matrix

$$\mathbf{C}_i \triangleq \sum_{j=1}^{\nu_i} (\mathbf{z}_i^{(j)} - \mathbf{z}_i^o)(\mathbf{z}_i^{(j)} - \mathbf{z}_i^o)^T,$$

$\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}, \dots, \mathbf{x}_i^{(j)}$ denote the j nearest neighbours of $\mathbf{x}_i$, $\mathbf{z}_i^{(j)} \triangleq \mathbf{R}(\mathbf{x}_i^{(j)})$, ν_i is the number of nearest neighbours of $\mathbf{x}_i$ considered, $\mathbf{x}_i^o$ is the origin of the coordinate system in which the projection is performed, and $\mathbf{z}_i^o \triangleq \mathbf{R}(\mathbf{x}_i^o)$.

The validity of local projection as a method of reducing noise depends on the assumption that an embedding of the clean signal would lie near a submanifold of the embedding space, and the result that $\mathbf{Q}^{(q)} = \mathbf{P}_i^{(q)}$ maximises

$$\sum_{j=1}^{\nu_i} \|(\mathbf{Q}^{(q)}\mathbf{z}_i^{(j)} - \mathbf{z}_i^o)\|_2^2, \quad (2.2)$$

with respect to projections $\mathbf{Q}^{(q)}$ onto subspaces of $\mathbb{R}^m$ with codimension q . To see this, note that maximising (2.2) is equivalent to minimising

$$\sum_{j=1}^{\nu_i} \sum_{k=1}^q (\mathbf{v}_k)^T (\mathbf{z}_i^{(j)} - \mathbf{z}_i^o)(\mathbf{z}_i^{(j)} - \mathbf{z}_i^o)^T \mathbf{v}_k, \quad (2.3)$$

with respect to $\{\mathbf{v}_k\}_{k=1}^q$, subject to the orthonormality constraint that $(\mathbf{v}_k)^T \mathbf{v}_l = \delta_{kl}$. Kantz and Schreiber (2004) avoid algebraic complications by tackling a more general problem. Using Lagrange multipliers they minimise (2.3) subject to normality (but not orthogonality) constraints by considering

$$\sum_{j=1}^{\nu_i} \sum_{k=1}^q (\mathbf{v}_k)^T (\mathbf{z}_i^{(j)} - \mathbf{z}_i^o)(\mathbf{z}_i^{(j)} - \mathbf{z}_i^o)^T \mathbf{v}_k - \sum_{k=1}^q \lambda_k ((\mathbf{v}_k)^T \mathbf{v}_k - 1). \quad (2.4)$$

Requiring the derivative with respect to $\mathbf{v}_l$ of the above to be zero yields $\mathbf{C}_i \mathbf{v}_l = \lambda_l \mathbf{v}_l$. As the unit eigenvectors of a real symmetric matrix, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q$ comprise an orthonormal set.

Cawley and Hsu (1992); Sauer (1992) chose as the origin of the coordinate system $\mathbf{x}_i^o = \bar{\mathbf{x}}_i \triangleq \frac{1}{\nu_i} \sum_{j=1}^{\nu_i} \mathbf{x}_i^{(j)}$, the centre of mass of the nearest neighbours of $\mathbf{x}_i$. This minimises (2.3). However, other choices of $\mathbf{x}_i^o$ can yield superior results in local projection; these lead to the higher order and multiple scale filters discussed in this paper.

The projected m -histories will not, in general, correspond to a time delay embedding of a scalar time series. For this reason, projection is followed by averaging all coordinates which would correspond to the same element of a scalar time series. This intuitive approach can also be identified by solving a least squares error minimisation problem. (Cawley and Hsu, 1992)

Kantz and Schreiber (2004) enhanced their local projective filter by selecting a transformation $\mathbf{R}$ in (2.1) which is distinct from the identity. As Schreiber (1993) discusses, the exponential divergence characteristic of the forward evolution of a chaotic system contributes to reduced stability of the final coordinates of an m -history, but the corresponding exponential convergence during reverse evolution contributes to increased stability of earlier coordinates. Similarly, exponential convergence in reverse evolution destabilises the first coordinates but contributes to the stability of later coordinates. The net effect is to make the interior coordinates relatively stable. Kantz and Schreiber (2004) direct adjustments made during projection to these more robust values by choosing $\mathbf{R}$ to be a diagonal matrix for which R_{11} and R_{mm} are orders of magnitude larger than the other entries of the main diagonal.

Grassberger *et al.* (1993) indicated a tendency for the largest magnitude adjustments to be spurious, and improved results by rescaling these.

In addition, Grassberger *et al.* (1993) suggested increasing the projection codimension q with pass number as repeated iteration made it easier to distinguish signal and noise. They did not specify a precise rule for adjusting q , and in this work q is not varied between iterations. The codimensions used in the benchmarking of Sec. IV correspond to the final, maximal codimensions used by Grassberger *et al.* (1993).

B. A second order refinement

Kantz and Schreiber (2004) used a second order refinement to local projective noise reduction to demonstrate the fractal nature of experimental data from an NMR laser experiment in the spectacular fashion of their Figure 6.1. This is reproduced as Figure 5 of this paper using a novel second order filter to be introduced in Sec. III.

Note that the second order centre of mass $\bar{\bar{\mathbf{x}}}_i$ refers to the mean over the ν_i nearest neighbours of $\bar{\mathbf{x}}_i$ among the neighbourhood centres of mass $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_{N-(m-1)T}$. Kantz and Schreiber (2004) observed that the difference between $\mathbf{x}_i$ and the mean over nearest neighbours, $\mathbf{x}_i - \bar{\mathbf{x}}_i$, is roughly half the difference between $\mathbf{x}_i$ and the second order centre of mass, $\mathbf{x}_i - \bar{\bar{\mathbf{x}}}_i$ (see Figure 1). They deduced that $2\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}}_i$ is an estimate of $\mathbf{x}_i$ in the absence of noise, and hence can be a better point from which to apply local projection than is $\bar{\mathbf{x}}_i$. Figure 1 illustrates this observation. Section III uncovers conditions

under which it is not just approximately, but exactly, true; the section describes a precise sense in which one difference will “tend to be roughly twice” that of the other.

III. ANALYTIC MOTIVATION

A. Higher order refinements

For real numbers $\delta > 0$, define the continuous moving average operator I_δ by

$$(I_\delta f)(x) = \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} f(t) dt.$$

Kantz and Schreiber (2004) achieve a frequently more precise estimate of the m -history corresponding to a clean signal by considering a linear combination of centres of mass of first and second order. In analogy, require

$$f_n(0) = a_1(I_\delta f_n)(0) + a_2(I_\delta^2 f_n)(0), \quad (3.1)$$

to hold for all monomials $f_n(x) \triangleq \begin{cases} 1, & n = 0 \\ x^n, & n \neq 0 \end{cases}$ with $0 \leq n \leq n_{\max}$, and with $n_{\max}$ as large as possible. For (positive) odd n , both the left and right of (3.1) are zero. Setting $n = 0$ and $n = 2$ in (3.1) yields a system of two linear equations with the unique solution $a_1 = 2$ and $a_2 = -1$. Identifying the m -dimensional analogue of the operator I_δ with the centre of mass of points within a ball of radius δ in Euclidean m -space, this leads to the second order refinement of Kantz and Schreiber (2004).

Having (3.1) hold for all monomials $f_n(x) = x^n$, with $0 \leq n \leq 3$ is equivalent to having the approximation $p_3(x) = a_1(I_\delta p_3)(x) + a_2(I_\delta^2 p_3)(x)$ hold exactly for the third degree Taylor polynomial p_3 of any thrice differentiable function. Obvious analogues would hold exactly for multivariate Taylor polynomials of degree up to three in each variable.

It is natural to seek linear combinations of I_δ^i which exactly describe Taylor polynomials of higher degree. Induction establishes that for $j, n \in \mathbb{N}$

$$(I_\delta^j f_n)(x) = \frac{1}{\delta^{n+j}} \left(-\frac{1}{2}\right)^j \frac{n!}{(n+j)!} \times \sum_{r=0}^j \binom{j}{r} (-1)^r (x + (2r-j)\delta)^{n+j}. \quad (3.2)$$

This formula was used to confirm that for $k = 1, 2, \dots, 20$, the unique solution to the system of k linear equations

$$f_n(0) = \sum_{i=1}^k a_i (I_\delta^i f_n)(0) \quad n = 0, 2, \dots, 2(k-1) \quad (3.3)$$

is

$$a_i = (-1)^{i-1} \binom{k}{i}, \quad i = 1, 2, \dots, k. \quad (3.4)$$

Together, (3.4) and (3.3) are equivalent to $\forall n \in \{0, 2, \dots, 2(k-1)\}, ((1 - I_\delta)^k f_n)(0) = 0$. This may suggest

a route to proving that (3.4) is a solution to (3.3) for general positive integer k .

B. Multiscale refinements

It is possible to represent $f_n(0)$ using a linear combination of iterated moving averages of f_n evaluated at zero for integers n such that $0 \leq n \leq 5$ without using moving averages of order greater than two. This is achieved by using moving averages over multiple scales. Requiring that

$$f_n(0) = a_2(I_\delta^2 f_n)(0) + b_1(I_{\beta\delta} f_n)(0) \quad (3.5)$$

holds for integers n such that $0 \leq n \leq 5$ (with $\beta \geq 0$) gives the two solutions $a_2 = 0, b_1 = 1, \beta = 0$ (in the limit $\beta \rightarrow 0$, $I_{\beta\delta}$ is analogous to integration after multiplication by the dirac delta distribution) and $a_2 = 4, b_1 = -3, \beta = \sqrt{8/3}$.

Seeking formulae involving higher order centres of mass and/or more scales would be a matter of applying (3.2) and algebra.

As will be illustrated in Subsection IV B (with references to Figure 3 and Table I), the multiscale filter corresponding to (3.5) for $0 \leq n \leq 5$ often performs worse than the second order, single scale refinement of Kantz and Schreiber

TABLE I. Proportional increase in signal to noise ratio over eight iterations, ϵ_0/ϵ_8 . This metric is used in order to follow Table I of Grassberger *et al.* (1993) and corresponds to a logarithmic change in the signal to noise ratio of $20 \log_{10}(\epsilon_0/\epsilon_8)$ dB. Higher order filters have coefficients satisfying (3.3), and multiple scale filters have coefficients satisfying (3.5), for $0 \leq n \leq n_{\max}$. The final column corresponds to the filter of Subsection III C, with coefficients chosen to the minimise normalised expected error. The results corresponding to the most and second most effective filters are, respectively, bold and *italic*. In Figures 2(b), 3, 4, and 6, the symbols +, □, ▽, *, ×, ◇, and ○ correspond to the same filters as they do below.

System	N	Order Scales	$n_{\max}$	ϵ_0	1	2	3	4	5	2	2
					+	□	▽	*	×	◇	○
Henon	5000	100%	^a	1.72	1.53	1.35	1.22	1.30	1.53	<i>1.59</i>	
		30%		2.27	<i>2.31</i>	2.23	1.99	1.82	2.30	2.37	
		10%	^a	3.19	3.46	3.45	3.39	3.26	3.53	<i>3.50</i>	
		3%		3.23	3.73	3.70	3.69	3.68	3.53	3.74	
		1%	^a	2.92	3.43	3.42	3.39	3.27	2.97	<i>3.43</i>	
	20 000	10%		3.28	3.74	3.56	3.37	3.20	3.69	3.74	
		1%	^a	3.46	4.10	4.07	4.01	3.88	3.84	4.04	
	Ikeda	100%		1.95	1.89	1.63	1.38	1.43	<i>1.94</i>	1.91	
		30%		1.65	1.49	1.28	1.36	1.49	1.67	1.65	
		10%		1.84	2.03	2.10	2.03	1.98	1.87	2.05	
Lorenz (1963)	5 000	3%		1.52	1.73	1.75	1.69	1.61	1.49	<i>1.73</i>	
		1%		1.22	1.31	1.28	1.22	1.22	1.14	<i>1.31</i>	
	20 000	10%	^a	2.34	2.63	2.80	2.56	2.38	2.45	2.66	
		1%	^a	1.87	<i>2.31</i>	2.36	2.25	2.09	1.82	2.31	
	5 000	100%		1.69	1.37	1.24	1.39	<i>1.53</i>	1.40	1.46	
		30%		1.96	<i>2.11</i>	1.95	1.64	1.72	1.73	2.14	
		10%	^a	1.83	2.04	2.05	2.06	1.99	1.88	2.04	
		3%		1.64	1.80	1.85	1.82	1.75	1.66	1.81	
		1%		1.40	<i>1.57</i>	1.57	1.52	1.49	1.41	1.56	
	20 000	10%		1.98	2.17	2.20	2.16	2.07	1.98	<i>2.18</i>	
		1%		1.70	1.94	1.97	1.89	1.81	1.74	<i>1.95</i>	

^aConsidered by Grassberger *et al.* (1993) (see their Table I).

(2004), and rarely performs considerably better. This is surprising, since the multiscale filter should be better able to approximate the geometry of the underlying attractor. Section III C explains the result by showing that the second order, single scale filter of Kantz and Schreiber (2004) particularly effectively attenuates errors.

C. Attenuating error

The choice of coefficients ($a_2 = -1, b_1 = 2$) and scale ($\beta = 1$) corresponding to the refined filter of Kantz and Schreiber (2004) not only preserves the component of the geometry of the underlying attractor described by a multivariate Taylor polynomial of degree up to three in each variable but also attenuates errors particularly effectively.

For non-negative integer d define the discrete moving average operator J_d by

$$(J_d f)(j) = \frac{1}{2d+1} \sum_{i=-d}^d f(j+i).$$

Let $e(-2), e(-1), e(0), e(1), e(2), \dots \in \mathbb{R}$ be independent random variables with mean zero and unit variance. The $e(i)$ can be thought of as errors; in this section is sought a choice of a_2, b_1, β for which they are attenuated by the operator $a_2 J_d^2 + b_1 J_{\beta d}$. For $\beta \leq 2$,

$$\begin{aligned} & ((a_2 J_d^2 + b_1 J_{\beta d})e)(0) \\ &= \sum_{l=-\beta d}^{\beta d} \left(a_2 \frac{2d+1-|l|}{(2d+1)^2} + \frac{b_1}{2\beta d+1} \right) e(l) \\ &+ \sum_{r=\beta d+1}^{2d} a_2 \frac{2d+1-r}{(2d+1)^2} (e(r) + e(-r)). \end{aligned} \quad (3.6)$$

Of course, the result only holds for β, d such that $\beta d \in \mathbb{Z}$. Eventually, a limit $d \rightarrow \infty$ obviates this complication.

To preserve the geometry of the underlying attractor, require (3.5) to hold for $0 \leq n \leq 3$, yielding (for $\beta \neq \sqrt{2}$) $a_2 = A(\beta) \triangleq \frac{\beta^2}{-2+\beta^2}$ and $b_1 = B(\beta) \triangleq 1 - \frac{\beta^2}{-2+\beta^2}$. Consider the expectation of the square of the filtered error at $j=0$, $S_{a_2, b_1, \beta, d} \triangleq \langle (((a_2 J_d^2 + b_1 J_{\beta d})e)(0))^2 \rangle$, with these geometric constraints on a_2, b_1, β .

The desirable case of a dense embedding corresponds to a large window d . However, in the limit $d \rightarrow \infty$, $S_{a_2, b_1, \beta, d} \rightarrow 0$. Therefore, normalise the expectation $S_{a_2, b_1, \beta, d}$ by that corresponding to prototypical local projection. This gives

$$T_{a_2, b_1, \beta} \triangleq \lim_{d \rightarrow \infty} \frac{S_{a_2, b_1, \beta, d}}{S_{1, 0, \beta, d}}, \quad (3.7)$$

called the normalised expected error. Using (3.6) and the independence of the variables $e(i)$ it is straightforward to compute that $T_{a_2, b_1, \beta} = \frac{2a_2^2}{3} - \frac{1}{2}a_2 b_1 (-4 + \beta) + \frac{b_1^2}{\beta}$.

As Figure 2(b) shows, $T_{A(\beta), B(\beta), \beta}$ has a local minimum at $\beta = \beta^* \approx 0.96215$, for which $a_2 = a_2^* \triangleq A(\beta^*) \approx -0.86174$, $b_1 = b_1^* \triangleq B(\beta^*) \approx 1.86174$. These values are suggestively close to the coefficients of the refined local projective filter of

FIG. 2. Normalised expected error (3.7) as a function of scale factor β for local projective filters satisfying (3.5) for $0 \leq n \leq 3$. (a) The normalised expected error has vertical asymptotes $\beta = 0, \sqrt{2}$ and a horizontal asymptote at $2/3$. (b) The local minimum at $\beta \approx 0.96215$, marked $\circ$, is close to $\beta = 1$ (corresponding to Kantz and Schreiber (2004)), marked $\square$. The normalised expected error corresponding to $\beta = \sqrt{8/3}$ is marked $\diamond$.

Kantz and Schreiber (2004), namely, $a_2 = -1, b_1 = 2, \beta = 1$. Indeed, as Figure 2(b) shows, the ratio of normalised expected errors is close to unity, with $T_{2,-1,1}/T_{a_2^*, b_1^*, \beta^*} \approx 1.004$.

Similar analysis involving higher orders and/or more numerous scales could identify filters which effectively attenuate error while more precisely preserving geometry.

IV. RESULTS AND DISCUSSION

Following Kantz *et al.* (1993); Kantz and Schreiber (2004), the matrix $\mathbf{R}$ of (2.1) is diagonal with $R_{11} = R_{mm} = 10^3$ and other entries of the main diagonal set to 1. Again following Kantz *et al.* (1993); Kantz and Schreiber (2004), the highest magnitude corrections are assumed deleterious and rescaled to the mean magnitude. Specifically, the largest 5% are rescaled.

For a time delay embedding of a signal, the most obvious ways to calculate an analogue of I_δ at an m -history $\mathbf{x}_i$ are to take either the mean over neighbours within some fixed radius or the mean over some fixed number of neighbours. Fixing the number of neighbours in each pass seems to achieve consistently superior results, presumably because it is flexible with respect to the inhomogeneous density of the attractor. This paper presents the results of local projection with a fixed number of neighbours in each pass.

The appropriate scaling of population corresponding to a change in length scale β is estimated to be β^d , where d is a dimension of the underlying attractor. For $d = d_p$ the pointwise dimension, this estimate is exact in the limit as the number of points $N \rightarrow \infty$ and the neighbourhood radius $\delta \rightarrow 0$ (Barreira and Wolf, 2006). Populations are here scaled using $d = d_B$ the box counting dimensions of the Henon map ($d_B = 1.28$ (Grassberger, 1983)), the Ikeda map ($d_B = 1.7$ (Kantz, 1994)), and the Lorenz (1963) system ($d_B = 2.06$ (Grassberger and Procaccia, 1983)), and $d = d_C$ the correlation dimension of the experimental NMR laser system of Subsection IV C ($d_C = 1.5$ (Kantz, 1994)). Should it be necessary, the algorithms of Diks (1996) and Yu *et al.* (2000) can estimate the correlation dimension of an attractor underlying noisy data.

A. Benchmarks

Following Grassberger *et al.* (1993), the filters considered in this paper are benchmarked using time series from x -coordinates of the Lorenz (1963) system

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} \sigma(y - x) \\ -y + rx - xz \\ -bz + xy \end{pmatrix},$$

with $\sigma = 10, r = 28$, and $b = 8/3$ and a time step of $\Delta t = 0.2$ (the large value helps to avoid the advantages of temporal correlation); the Henon map

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 1 - ax_t^2 + y_t \\ bx_t \end{pmatrix},$$

with $a = 1.4$ and $b = 0.3$; and the Ikeda map

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 1 + 0.9(x_t \cos \phi_t - y_t \sin \phi_t) \\ 0.9(x_t \sin \phi_t + y_t \cos \phi_t) \end{pmatrix},$$

where $\phi_t = 0.4 - 6/(1 + x_t^2 + y_t^2)$. Each time series is normalised to unity standard deviation, whereafter Gaussian noise is added.

Again following Grassberger *et al.* (1993), the Lorenz (1963) time series are filtered in embedding dimension $m = 7$ using projection dimension $m - q = 3$, the Ikeda in $m = 7$ using $m - q = 2$, and the Henon in $m = 5$ using $m - q = 2$. Each time series is filtered with lag $T = 1$.

Grassberger *et al.* (1993) estimate the appropriate neighbourhood population ν for local projective filtering of N data corresponding to an attractor of dimension d by

$$\nu = (\epsilon N^{2/d})^{2d/(4+d)}, \quad (4.1)$$

where ϵ is the root mean square error. The algorithm to select neighbourhood size incorporates this formula.

Each iteration of the filter uses that neighbourhood population among a particular selection which most reduces noise, for which the resultant root mean square error ϵ would be least. Populations considered belong to $(1 : 1 : 20) \cup (25 : 5 : 200) \cup (250 : 50 : 2000) \cup (2500 : 500 : 20000)$, where $(a : b : c) \triangleq \{a, a + b, a + 2b, \dots, c\}$. To reduce computation, the first pass only considers populations which are less than or equal to the least upper bound in this set of the number of neighbours suggested by (4.1) with $d = d_B$ the box counting dimensions cited earlier in this section. Subsequent iterations consider only populations less than or equal to the optimal population of the previous pass. This greedy algorithm is not, in general, optimal with respect to multiple passes. Hence, it does not achieve the greatest noise

reduction possible through local projective filtering. It is merely a straightforward and consistent way of comparing the efficacy of different filters.

Kantz *et al.* (1993); Kantz and Schreiber (2004) found that eight passes tended to be the best trade-off between effective noise reduction and computational efficiency. Additionally, Hegger, Kantz, and Schreiber (1998) noticed that most noise reduction takes place in the first three iterations. Guided by these heuristics and following the presentation of Grassberger *et al.* (1993), Table I and Figure 4 present results after eight iterations, and Figure 3 shows the noise level after 1 – 8 iterations.

B. Results of benchmarking

Table I and Figure 3 present the results of filtering noisy nonlinear systems with local projective filters of a single scale and orders 1 (corresponding to the prototypical local projection of Cawley and Hsu (1992); Sauer (1992)), 2 (corresponding to the refinement of Kantz and Schreiber (2004)), up to 5, as well as dual scale local projective filters derived from purely geometric and from simultaneous geometric and statistical considerations. The four most frequently optimal filters are the single scale filters of orders 1, 2, and 3 and the filter derived via considerations of the normalised expected error. Figure 4 presents the results of eight iterations of each of these four filters to 20 000 points of data from the three benchmark systems with $\epsilon_0 = 10\%$ added Gaussian noise. The results of different filters are visually similar, but the three of order greater than two shrink the attractors less, and lead to smoother surfaces.

As presaged in Subsection III B, seen in Table I and exemplified by Figure 3, the multiscale filter corresponding to (3.5) for $0 \leq n \leq 5$ often performed worse than the second order, single scale refinement of Kantz and Schreiber (2004), and rarely performed considerably better. However, Table I also shows that the multiscale filter identified through simultaneous geometric and error attenuation considerations (with parameters $a_2 = a_2^*$, $b_1 = b_1^*$, $\beta = \beta^*$) almost always performed similar to, but usually marginally better than, the second order, single scale refinement of Kantz and Schreiber (2004) (with parameters $a_2 = -1$, $b_1 = 2$, $\beta = 1$).

This is consistent with the filters' equivalent geometric motivations and similar normalised expected errors (see Subsection III C).

It is perhaps surprising that the filter derived in Subsection III C is so effective since it is not clear that the analysis involving the discrete moving average operator J_d applies in dimension greater than one, as does that involving the continuous moving average operator I_δ . It would be interesting to perform similar analysis of a higher dimensional analogue of the normalised expected error.

Table I shows that as noise level increases from the lowest level considered, 1%, higher order filters become relatively more effective. This trend terminates at a higher noise level, from which the filter of Kantz and Schreiber (2004) and the filter derived in Subsection III C via considerations of error tend to be optimal. The highest noise level considered, 100%, favours prototypical local projective filters. The pattern can be rationalised by considering the relative importance of geometry preservation and error reduction at different noise levels. At low noise levels, geometry is critical and so filters of order two or greater are more effective. Higher order filters become relatively more effective as the relatively low level of noise increases. This may be because, for a given neighbourhood size, higher order centre of mass measurements incorporate more observations. At intermediate noise levels, both error attenuation and geometry are significant. Hence, the filters derived from geometric considerations which also correspond to a point near the minimum of the normalised expected error $T_{A(\beta),B(\beta),\beta}$ of Subsection III C are optimal. At high levels of noise when geometry is less important, prototypical local projection, which corresponds to the still lower value of normalised expected error $T_{0,1,\beta} = 1$ (compare with Figure 2), is most effective.

In fact, in noisy conditions the rigid process of benchmarking inhibits most filters of order two or greater. As Figure 3 shows, they plateau more slowly than the first order filter. This is more pronounced for higher noise levels, under which conditions most filters other than the prototypical local projective filter cannot near their plateau within eight iterations using neighbourhoods constrained by (4.1). The

FIG. 3. Relative error amplitude during local projective filtering of two time series. The markers +, □, ▽, *, ×, ◇, and ○ are as indicated in Table I. (a) 5000 points from the Henon map with $\epsilon_0 = 30\%$ added noise. After eight iterations, the dual scale filter corresponding to a local minimum of the normalised expected error is the most effective. The next most successful is the single scale, second order refined filter of Kantz and Schreiber (2004). (b) 20 000 points from the Lorenz (1963) system map with $\epsilon_0 = 1\%$ added noise. After eight iterations, the single scale, third order filter is the most effective. It is followed by the dual scale filter corresponding to a local minimum of the normalised expected error.

FIG. 4. Results of eight iterations of local projective filtering of 20 000 point time series from three different systems with $\epsilon_0 = 10\%$ added noise. Markers are as indicated in Table I.

dual scale filter identified purely geometrically employs a scale factor $\beta = \sqrt{8/3}$ and hence a larger neighbourhood size, and so escapes this trend.

Note that the box counting dimension of the Henon map, Ikeda map, and Lorenz (1963) system are, respectively, 1.28 ± 0.01 (Grassberger, 1983), approximately 1.7 (Kantz, 1994), and 2.06 ± 0.01 (Grassberger and Procaccia, 1983). Table I suggests that higher order filters become more advantageous relative to established ones as attractor dimension

increases. This may be because higher order polynomial fits better capture the more complicated geometric structure typical of higher dimensional chaotic systems.

Higher order polynomial fits require more data. Congruently, as Table I suggests, the optimal order tends to increase with the length of the time series considered. The instability of higher order polynomial fits might explain the relative ineffectiveness typical of filters of order greater than three.

Under the conditions of benchmarking, the order 3 filter was the most frequently optimal. The filter identified by minimising the normalised expected error is, uniquely, superior to both established filters in the majority of cases. Subsection IV C applies each to 39 883 measurements of an NMR laser estimated using the algorithm of Yu *et al.* (2000) to have a noise level $\epsilon_0 = 0.76\%$.

C. NMR laser data

This subsection presents the results of local projective filtering of 39 883 points of NMR laser data analysed and described by Kantz *et al.* (1993); Kantz and Schreiber (2004). The data are normalised to unity standard deviation. A pioneering version of the experiment is described by Bösigner, Brun, and Meier (1977).

The noise-free signal is not known and noise levels cannot be determined trivially or exactly. Instead, the formula (4.1) together with the result $d_C = 1.5 \pm 0.1$ (Kantz and Schreiber, 2004) and estimates of noise level obtained using the algorithm of Yu *et al.* (2000) (in dimension 10)

determine the neighbourhood population ν for each iteration.

Figure 5 showcases application to NMR laser data of the local projective filter identified by minimising the normalised expected error. The filter unveils a procession of twins cascading down length scale. Figure 6 demonstrates explicitly the increase in resolution possible through local projective filtering. Although Figures 6(b) and 6(c) pertain to different filters, visually they are very similar. This is not surprising since, as shown in Figure 6(d), they correspond to estimated noise levels less than $\epsilon_5 = 0.049\%$ and $\epsilon_5 = 0.023\%$, respectively.

The noise attenuation seen in Figure 6(d) is much greater than that recorded in Table I or seen in Figure 3. Of course, the algorithm of Yu *et al.* (2000) can only estimate the deviation of a signal from the strange attractor to which the data most closely correspond, which is a different quantity from the root mean square deviation of Figure 3. In particular, Figure 6 suggests that estimates of noise level via the method of Yu *et al.* (2000) together with (4.1) provide an appropriate neighbourhood size for local projective noise reduction.

FIG. 5. Counterpart of Figure 6.1 of Kantz and Schreiber (2004). A local projective filter identified through considerations of error attenuation makes clear the scale invariance of experimental chaotic data, in this case from an NMR laser experiment.

FIG. 6. Counterpart of Figure 10.4 of Kantz and Schreiber (2004). (a) NMR laser data prior to filtering. (b) After six iterations of the single scale, third order filter, denoted ∇ in (d). (c) After six iterations of the dual scale filter corresponding to a local minimum of the normalised expected error, denoted $\circ$ in (d). (d) Relative error amplitude during local projective filtering of NMR laser data. Five iterations of the filters reduce the error amplitude to 6% and 3%, respectively, of its initial value.

V. CONCLUSION

This paper examined the refined local projective filter of Kantz and Schreiber (2004), leading to series of filters involving different numbers of length scales and orders of centre of mass operators. The new filters were compared to established ones. Higher order filters were relatively more effective for data of greater length or corresponding to a higher underlying attractor dimension.

When reformulated for multiple scales, the geometric considerations which led to effective higher order filters did not consistently improve upon existing filters. However, the same geometric criteria, augmented by statistical analysis designed to minimise error, led to filters usually superior to established filters. In addition, the multiscale statistical and geometric considerations explicate the efficacy of the refinement of Kantz and Schreiber (2004).

As noise level increased from a low value (1%), higher order filters initially became relatively more effective. At intermediate noise levels ($\sim 30\%$), geometrically motivated filters corresponding to high error attenuation (including a new, dual scale filter) began to dominate until, at the highest noise level (100%), prototypical local projective filters prevailed.

As Sec. I mentions, Chelidze (2014) locally projected strands (sequences of temporally successive points in phase space) to attenuate noise in flow-like data more effectively than can classical local projection. It would be interesting to compare his method, which does not rely upon the refinement of Kantz and Schreiber (2004), with higher order and multiscale filters. Unfortunately, his 60 000 point time series from a Lorenz (1963) system, sampled at $\Delta t = 0.01$, cannot be compared confidently with the 5000 or 20 000 point time series, sampled at $\Delta t = 0.2$, considered in Sec. IV. It is not

even obvious which time series is most propitious; while the more frequent sampling of Chelidze (2014) offers stronger temporal correlation, those of Sec. IV correspond to longer trajectories and so to better coverage of the attractor. The method of Chelidze (2014) shows excellent performance for shorter but more highly sampled trajectories, while the novel higher order and multiscale refinements are conceived for longer trajectories with a lower sampling rate.

This work highlights several open problems. It would be satisfying to confirm that (3.4) is the unique solution to (3.3) for every positive integer k . It would be useful to find formulae for higher dimensional and/or more numerous length scale analogues of the normalised expected error. Together with geometric considerations, the latter formulae could lead to novel filters which precisely preserve geometry while minimising error.

It would be interesting to compare the results using the filter coefficients suggested by (3.3) together with monomial f_n with those obtained via other bases of function spaces, such as Hermitian polynomials. It could also be valuable to see how such analytically motivated formulae compare to coefficients numerically optimised for particular classes of problems.

The size of the neighbourhood used to identify projection directions ideally represents a trade-off between statistical stability (increasing with number of observations) and the validity of a linear approximation (decreasing with neighbourhood radius) (Grassberger *et al.*, 1993). The same is true of the size of the neighbourhood used to determine the origin of the coordinate system in prototypical local projection, but not the higher order local projection proposed in this paper. For computational convenience as well as to evince fair comparison, the neighbourhood sizes used to

determine projection directions and projection origins were set equal. There is no other obvious reason this need be the case; allowing them to differ would increase the flexibility of local projective filters and could improve their results.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful suggestions. J.M.M. extends heartfelt thanks to Thomas Poulet, Klaus Regenauer-Lieb, Manolis Veveakis, and Jürg Hauser for valuable feedback and for stimulating coffee room discussions. J.M.M. was supported by the Prescott Postgraduate Scholarship of the University of Western Australia and the OCE Ph.D. Scholarship of CSIRO. M.S. was supported by an Australian Research Council Future Fellowship (FT110100896) and Discovery Project (DP140100203).

- Barreira, L. and Wolf, C., "Pointwise dimension and ergodic decompositions," *Ergodic Theory Dyn. Syst.* **26**(3), 653–671 (2006).
- Bösiger, P., Brun, E., and Meier, D., "Solid-state nuclear spin-flip maser pumped by dynamic nuclear polarization," *Phys. Rev. Lett.* **38**(11), 602–605 (1977).
- Cawley, R. and Hsu, G. H., "Local-geometric-projection method for noise reduction in chaotic maps and flows," *Phys. Rev. A* **46**(6), 3057–3082 (1992).
- Chelidze, D., "Smooth local subspace projection for nonlinear noise reduction," *Chaos* **24**(1), 013121 (2014).
- Diks, C., "Estimating invariants of noisy attractors," *Phys. Rev. E* **53**(5), R4263–R4266 (1996).
- Grassberger, P., "On the fractal dimension of the Henon attractor," *Phys. Lett. A* **97**(6), 224–226 (1983).
- Grassberger, P. and Procaccia I., "Characterization of strange attractors," *Phys. Rev. Lett.* **50**(5), 346–349 (1983).
- Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C., and Schreiber, T., "On noise reduction methods for chaotic data," *Chaos* **3**(2), 127–141 (1993).
- Hegger, R., Kantz, H., and Schreiber, T., "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos* **9**(2), 413–435 (1999).
- Jevtić, N., Schweitzer, J., and Stine, P., "Optimizing nonlinear projective noise reduction for the detection of planets in mean-motion resonances in transit light curves," in *Topics in Chaotic Systems: Selected Papers from Chaos 2010 International Conference*, World Scientific Publishing (2011), pp. 191–198.
- Jevtić, N., Stine, P., and Schweitzer, J., "Nonlinear time series analysis of Kepler Space Telescope data: Mutually beneficial progress," *Astron. Nachr.* **333**(10), 983–986 (2012).
- Kantz, H., "A robust method to estimate the maximal Lyapunov exponent of a time series," *Phys. Lett. A* **185**(1), 77–87 (1994).
- Kantz, H. and Schreiber, T., *Nonlinear Time Series Analysis*, 2nd ed. (Cambridge University Press, New York, 2004).
- Kantz, H., Schreiber, T., Hoffmann, I., Buzug, T., Pfister, G., Flepp, L. G., Simonet, J., Badii, R., and Brun, E., "Nonlinear noise reduction: A case study on experimental data," *Phys. Rev. E* **48**(2), 1529–1538 (1993).
- Kostelich, E. J. and Schreiber, T., "Noise reduction in chaotic time-series data: A survey of common methods," *Phys. Rev. E* **48**(3), 1752–1763 (1993).
- Kotas, M., "Projective filtering of time-aligned beats for foetal ECG extraction," *Tech. Sci.* **55**(4), 331–339 (2007).
- Kotas, M., "Projective filtering of time warped ECG beats," *Comput. Biol. Med.* **38**(1), 121–137 (2008).
- Lorenz, E. N., "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**(2), 130–141 (1963).
- Noakes, L., "The Takens embedding theorem," *Int. J. Bifurcation Chaos* **1**(04), 862–872 (1991).
- Sauer, T., "A noise reduction method for signals from nonlinear systems," *Physica D* **58**(1), 193–201 (1992).
- Sauer, T., Yorke, J. A., and Casdagli, M., "Embedology," *J. Stat. Phys.* **65**(3–4), 579–616 (1991).
- Schreiber, T., "Extremely simple nonlinear noise-reduction method," *Phys. Rev. E* **47**(4), 2401–2404 (1993).
- Takens, F., "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980* (Springer-Verlag, 1981), pp. 366–381.
- Theiler, J. and Eubank, S., "Don't bleach chaotic data," *Chaos* **3**(4), 771–782 (1993).
- Wan, X., Iwata, K., Riera, J., Ozaki, T., Kitamura, M., and Kawashima, R., "Artifact reduction for EEG/fMRI recording: Nonlinear reduction of ballistocardiogram artifacts," *Clin. Neurophysiol.* **117**(3), 668–680 (2006).
- Yu, D., Small, M., Harrison, R. G., and Diks, C., "Efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy time series data," *Phys. Rev. E* **61**(4), 3750–3756 (2000).