International Journal of Bifurcation and Chaos, Vol. 8, No. 11 (1998) 2191–2201

WAVELET RECONSTRUCTION OF NONLINEAR DYNAMICS

DAVID ALLINGHAM*, MATTHEW WEST† and ALISTAIR I. MEES‡
Centre for Applied Dynamics and Optimization, Department of Mathematics, University of Western Australia, Nedlands, WA 6907, Australia

Received September 26, 1997; Revised April 26, 1998

We investigate the reconstruction of embedded time-series from chaotic dynamical systems using wavelets. The standard wavelet transforms are not applicable because of the embedding, and we use a basis pursuit method which on its own does not perform very well. When this is combined with a continuous optimizer, however, we obtain very good models. We discuss the success of this method and apply it to some data from a vibrating string experiment.

1. Introduction

In recent years, the modeling of dynamical systems has been successfully carried out with methods such as radial basis functions and single-layer neural networks, both of which approximate the desired function with a sum of basis functions. Wavelet series use a similar summation of functions, and have been applied in fields such as the processing of one-dimensional signals and images. They have many nice theoretical properties as bases of function spaces, and this paper investigates whether or not these properties translate well to be used in modeling embedded time-series from dynamical systems. We outline two strategies for using wavelets for such modeling.

In Sec. 2 we discuss dictionary selection methods for building models and demonstrate that such techniques do not perform well. This is followed in Sec. 3 by the use of continuous optimization for model fitting, in a manner similar to that of [Cao *et al.*, 1995], and we demonstrate a hybrid method which successfully models dynamical systems using wavelets.

2. Dictionary Wavelet Models

Our approach to building models is founded on the idea of parsimony, in which a good model is one which can express the data very concisely. For compressing data via basis decomposition [Donoho, 1992] has shown that unconditional bases (also called frames) perform optimally. A frame is a spanning set of basis functions $\{v_i\}$ for which there exists a constant C such that

$$\sum_i |(f, v_i)|^2 \leq C\|f\|^2$$

for all functions f in the space. Wavelets form an unconditional basis (see e.g. [Daubechies, 1992]) and so are an appealing technique for modeling dynamical systems.

We used a dictionary model scheme. That is, a function f was represented by the superposition of a set of elements chosen from some large dictionary $\mathcal{D}$:

$$f = \sum_{i \in I} c_i d_i$$

*davida@maths.uwa.edu.au
†mwest@maths.uwa.edu.au
‡alistair@maths.uwa.edu.au

where c_i's are scalars, $d_i \in \mathcal{D}$, and the modeling problem is to choose the subset I. We approximate a function by using a finite number of dictionary elements, i.e.

$$f = \sum_{i=1}^{n} c_i d_i + R_n$$

where R_n is the residual. A dictionary such as ours for which any function has multiple representations in terms of the dictionary elements is called over-complete. This contrasts with a complete dictionary in which a given function has a unique representation.

For our modeling we used a dictionary composed of tensor product wavelet functions, but our conclusions also apply to more general classes. The elements of our wavelet dictionaries were of the form

$$d_{a,b}\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = \prod_{i=1}^{n} \psi\left(\frac{x_i - b_i}{a_i}\right),$$

where the a_i's and b_i's are constants specifying $d_{a,b}$ and $\psi : \mathbb{R} \to \mathbb{R}$ is fixed for the dictionary. Typically ψ was chosen so that

$$\int \psi(x)dx = 0\,.$$

For signal processing applications, a highly oscillatory ψ is often chosen because the signals themselves are of this nature when seen as functions of time. For the modeling of dynamical systems however, we are interested in smooth interpolation on higher-dimensional spaces and we have found that better results are obtained using a ψ with relatively low curvature. The systems we modeled were smooth functions, and so, smooth wavelets were also used. All the results presented use

$$\psi(x) = (2x^2 - 1)e^{-x^2}\,.$$

Also of significance is the coverage of the function domain in signal processing. The dimension of the function domain is the same as the dimension of the space (essentially meaning that the data is not sparse on the region of the space that it occupies), and the data is uniformly spread. For dynamical systems, the dimension of the attractor which we are trying to model is frequently less than the space in which it lies, and the points are scattered in a nonuniform way throughout that space. This sparseness means that the model is not very constrained, and so model trajectories can wander far from the true attractor.

For appropriately chosen sets of a's and b's, the dictionary $\mathcal{D} = \{d_{a,b}\}$ is orthogonal, and so

$$c_{a,b} = \frac{\langle f,\, d_{a,b}\rangle}{\langle d_{a,b},\, d_{a,b}\rangle^{\frac{1}{2}}}\,.$$

If the domain of f is discrete, then this decomposition is known as the discrete wavelet transform (DWT) [Daubechies, 1992]. For modeling dynamical systems, we use time-lag embedding to define in $\mathbb{R}^n$ a finite portion of a trajectory, and thus our function has a finite domain. Our data points are not, however, uniformly spaced and so standard wavelet dictionaries are no longer orthogonal, and the DWT cannot be performed. In fact, the decomposition of the data is no longer unique and the problem becomes one of selecting a suitable subset of the basis.

The work of Rissanen [1989] and Donoho [1992] suggests that the best models will be those containing relatively few basis elements. For this reason, we sought a decomposition which had few large coefficients. A well-known technique for doing this is the orthogonal matching pursuit (OMP) method, also called greedy selection. See, for example, [Mallat & Zhang, 1993] or [Pati *et al.*, 1993].

The orthogonal matching pursuit algorithm iteratively builds a model, by including at each step, the next best dictionary element in the model until the desired accuracy is reached. This technique produces better models when presented with over-complete dictionaries, and so we used such wavelet dictionaries.

There are a number of more elaborate techniques for basis selection, including the swapping method of [Judd & Mees, 1995] and the basis pursuit method of [Chen *et al.*, 1996]. We investigated the performance of the Judd and Mees algorithm and found that it did not substantially alter the conclusions we arrived at. We do not believe that other basis selection methods would have a significant impact either, and so, for reasons of computational efficiency, the results presented here are from models built using orthogonal matching pursuit.

We modeled the Hénon map

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 1 + y - 1.4x^2 \\ 0.3x \end{bmatrix}$$

by selecting a subset of our wavelet dictionary. For these models we were interested in both one-step

and free-run prediction. As a quantitative measure of the performance of the model, we used cross-validation. This involved calculating the prediction error of data from the system which was not used during model fitting. This was done by dividing the available data into a "learning" set (used to build the model) and a test set. For our models of the Hénon map built with 200 points of data (Secs. 3 and 4), the test set contained 10000 points; for our model of 200 points of real data (Sec. 5), the test set contained 2000 points. Cross-validation error is an approximation of the generalization error for the model, and, as such, we expect the model with the smallest cross-validation error to give the best short-term free-run behavior.

The results of fitting a wavelet model to 200 points of the Hénon attractor are shown in Figs. 1 and 2. This model selected vectors from an over-complete dictionary of size 6400. The cross-validation error calculated from 10000 data points is plotted against the model size (the number of wavelets chosen) in Fig. 3. From this plot it can be seen that the best model is of size 108. The difference between this model and the true Hénon function is plotted in Fig. 2. The position of the data points, to which the model was fitted, are shown in this figure by green circles. From this, we see that the model surface is extremely accurate on the data points but diverges quickly from the true function elsewhere, even in the region bounded by the attractor.

Despite this, the model still exhibits stable long-term free-run behavior with an attractor as shown in Fig. 1. From this we can see that the model attractor diverges prominently from the true attractor in several regions. Two such regions are enlarged in Figs. 1(b) and 1(c). These show that the model coincides very accurately with the true system at the points on which the model was fitted, but has inaccuracies elsewhere.

To quantify the short-term predictive capabilities of this model, we plotted the multi-step cross-validation errors in Fig. 4. This is done by taking a large number of points that were not used for

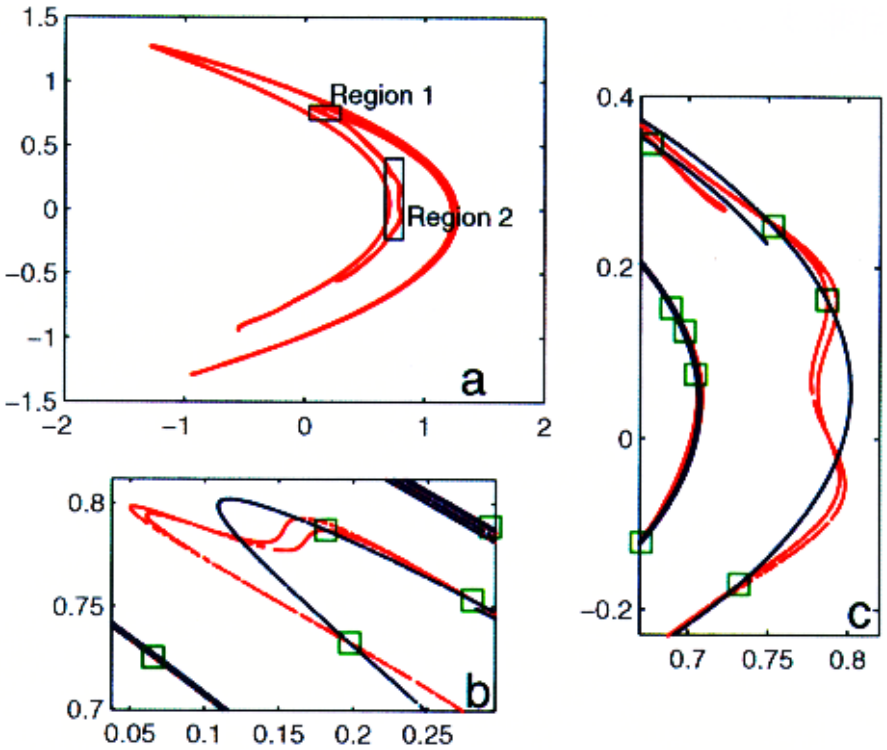

Fig. 1. (a) shows the attractor for the OMP-selected model of size 108 (shown in red); (b) and (c) show details of regions 1 and 2 respectively. The true attractor is in blue, and the data points used for fitting are the green boxes. Note how the model attractor passes directly through each data point but may not behave well elsewhere.

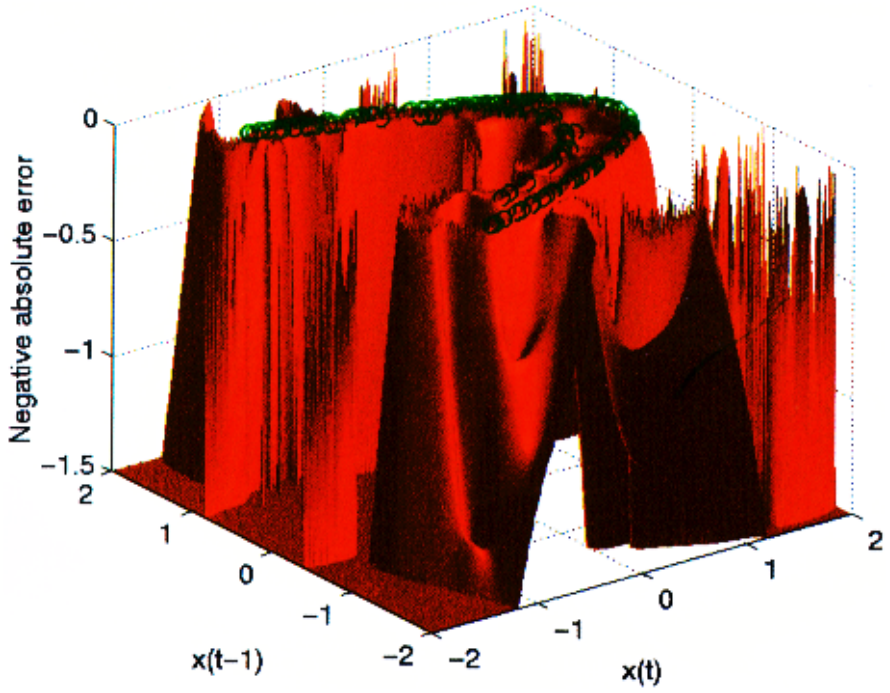

Fig. 2. The error surface for the OMP-selected model of size 108. The data used for fitting are the green circles. Note that the model has absolutely no generalization away from the attractor, not even in the region enclosed by the attractor.

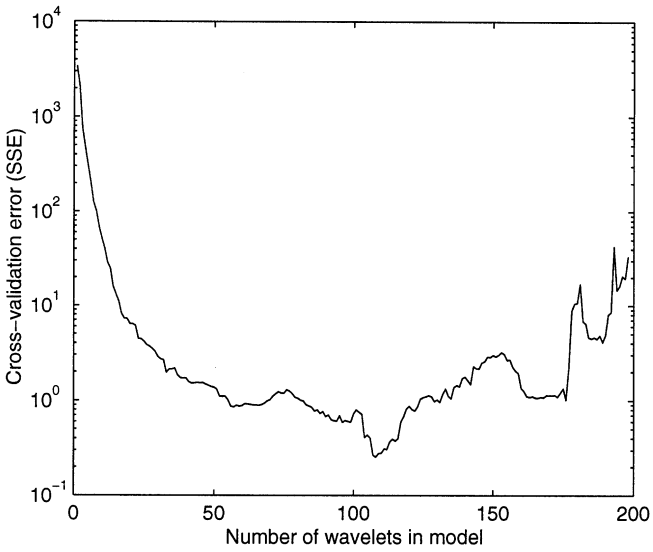

Fig. 3. Cross-validation versus model size for OMP-selected models.

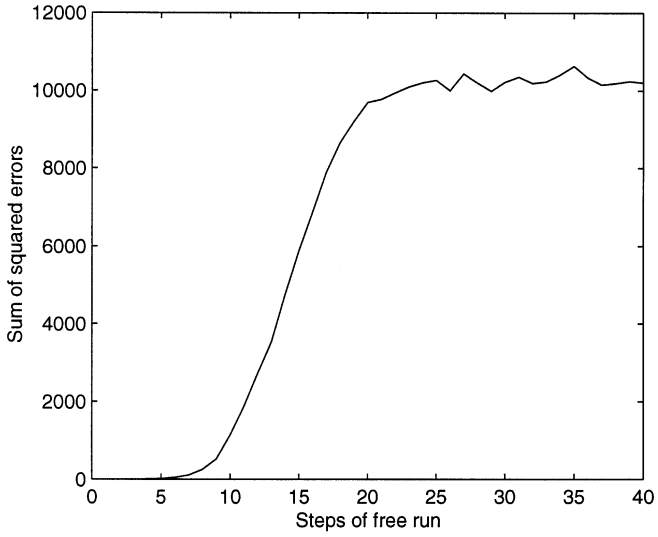

Fig. 4. Multi-step cross-validation for the OMP-selected model of size 108.

fitting the model and iterating the model a number of steps using these points. These iterated points were compared with the true values of the map, and the sum of squares error calculated. This plot shows that there is little useful predictability beyond 10 steps. It is possible to do much better, especially on such a simple system as Hénon.

3. Optimized Wavelet Models

The second approach to model building that we investigated, was continuous optimization of the parameters from some initial model. To perform least-squares optimization we used the Levenberg–Marquardt (LMO) [Marquardt, 1963] and Nelder

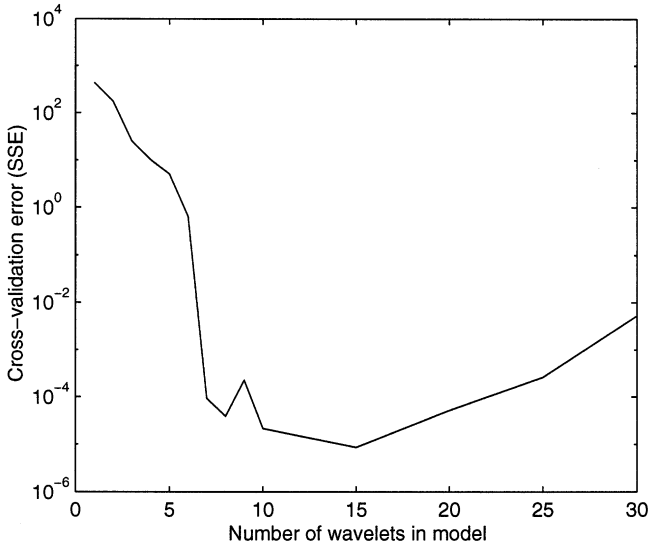

Fig. 5. Cross-validation error versus model size for LMO models. We choose the model of size 15 as it minimizes the cross-validation error.

and Mead's downhill simplex method [Press *et al.*, 1988], both standard optimizers with good performance in a wide range of problems. Occasionally, Levenberg–Marquardt optimization would be extremely slow in finding its way off flat sections of the error surface. In this event, we allowed the simplex algorithm to take over, later resuming the Levenberg–Marquardt technique because the simplex method was very slow to converge.

The use of these optimizers, however, was particularly sensitive to the choice of the initial model. When provided with random starting conditions, Levenberg–Marquardt optimization would usually become trapped in a poor local minimum of the error surface. This problem worsened as the number of wavelets in the model increased, making it virtually impossible to obtain a good model using more than 5 wavelets. Although a reasonable model of Hénon can be built using only 5 wavelets, this approach clearly would not generalize well to more complex systems. When successful models were built with this technique, they used very few wavelets and had lower cross-validation errors than the basis-selected models in Sec. 2.

To overcome the problem of sensitivity to initial conditions, we started the optimization with a model built by the basis-selection techniques outlined in Sec. 2. This hybrid method did in fact produce much better models of the dynamical systems. A plot of cross-validation errors versus model size

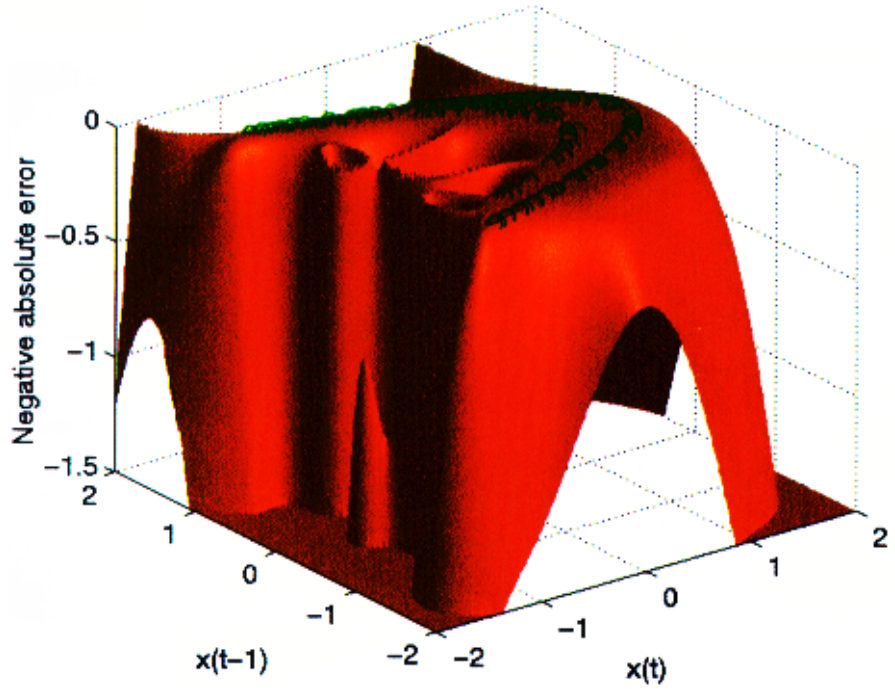

Fig. 6. Error surface for the LMO model of size 15. The data used for fitting are the green circles. Here, the improved generalization of this model can be observed, with the accurate part of the model extending away from the attractor.

is shown in Fig. 5. This indicates that the optimal model size is much smaller for this technique than if we use the greedy algorithm from Sec. 2 without LMO. From Fig. 5, we see that the optimal model contains 15 wavelets.

The error surface for the model containing 15 wavelets is plotted in Fig. 6. Comparing this to Fig. 2 we can clearly see the improved generalization capabilities of the optimized model. This was particularly reflected by the extended "plateau" surrounding the attractor. This improvement was also reflected in the free-run behavior, both in the short-term, as demonstrated by the multi-step cross-validation, and in the long-term attractor. These are plotted in Figs. 7 and 8, respectively. In these, it can be seen that this model has approximately twice as many steps of accurate short-term prediction as the previous technique, and that the model attractor is virtually identical to the true attractor, deviating by no more than 10^{-5}.

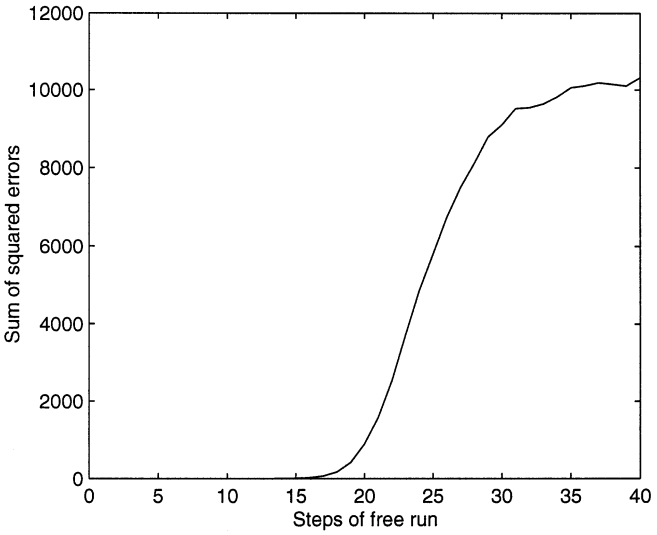

Fig. 7. Multi-step cross-validation for the LMO model of size 15.

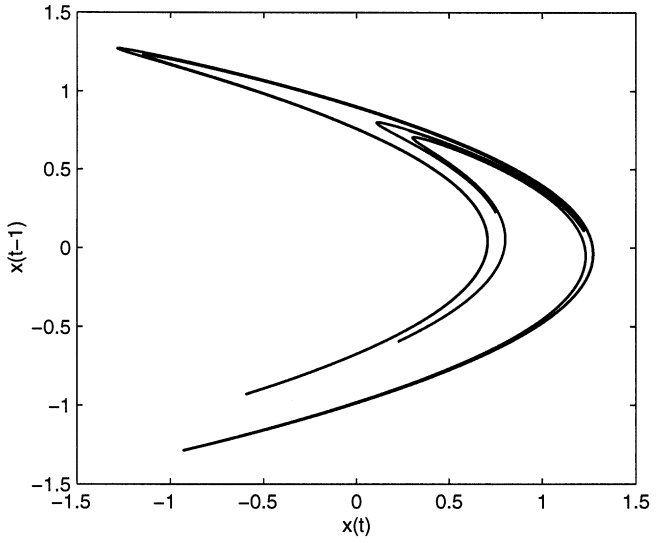

Fig. 8. The attractor for the LMO model of size 15.

This demonstrates the success of our hybrid basis-selection/Levenberg–Marquardt technique. Models were also built for the Ikeda system and these too performed extremely well.

4. Modeling With Additive Observational Noise

In Secs. 2 and 3 we described our two approaches to techniques for modeling dynamical systems. The data that we used to illustrate these methods was clean, but this is very unlike any real data which might be encountered. In this section we show the results of modeling the Hénon attractor with added observational noise. We expect that this will cause our selection methods to choose smaller models which do not perform as well as the models for the clean Hénon data.

Plots of cross-validation error against model size for orthogonal matching pursuit are shown in Fig. 9 for various levels of noise. It can easily be seen that the optimal number of wavelets decreased as more noise was added, but that the form of these plots remains the same. For Levenberg–Marquardt optimization with 3% noise, the cross-validation plot is shown in Fig. 10, from which it can be seen that the optimum model size occurs earlier and that the cross-validation error for that model is greater than for the noise-free model (as shown in Fig. 5). We chose the minimum model sizes for 3% noise to more fully compare the two techniques. For orthogonal matching pursuit, the minimum model size was 32; for Levenberg–Marquardt optimization, it was 6.

We compared these two models by considering multi-step cross-validation and the attractor geometry. For OMP, the model built for the noisy attractor is quite bad. It is shown in Fig. 11. This can be compared to the true attractor with noise as shown in Fig. 12. Although some of the same structure is present, it is very distorted. The multi-step cross-validation error for the model is plotted in Fig. 13. Comparing this to Fig. 4 we can see that the model of the noisy data performs very

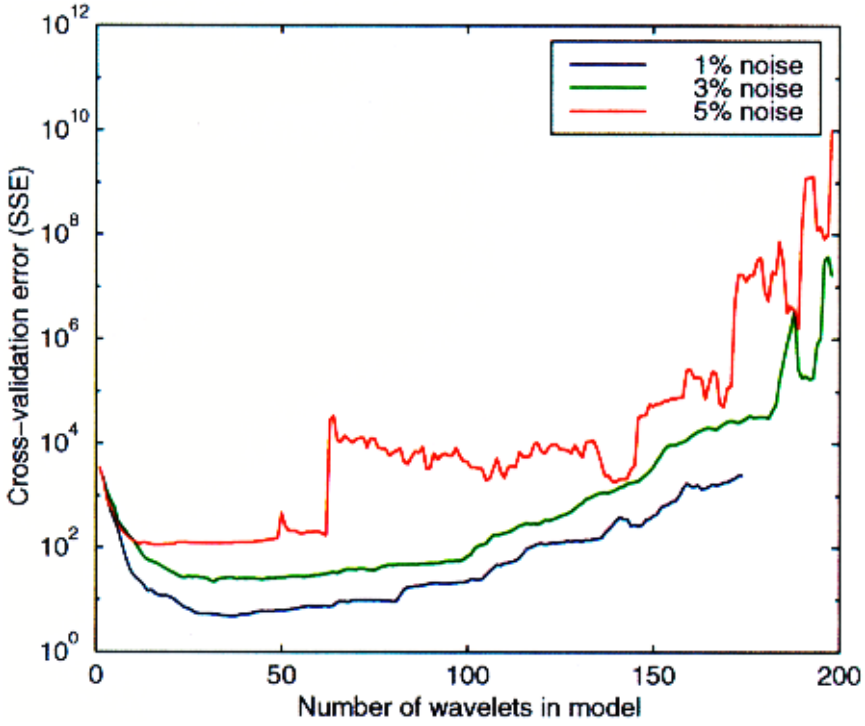

Fig. 9. Cross-validation error versus model size for OMP models with 1%, 3% and 5% additive observational noise.

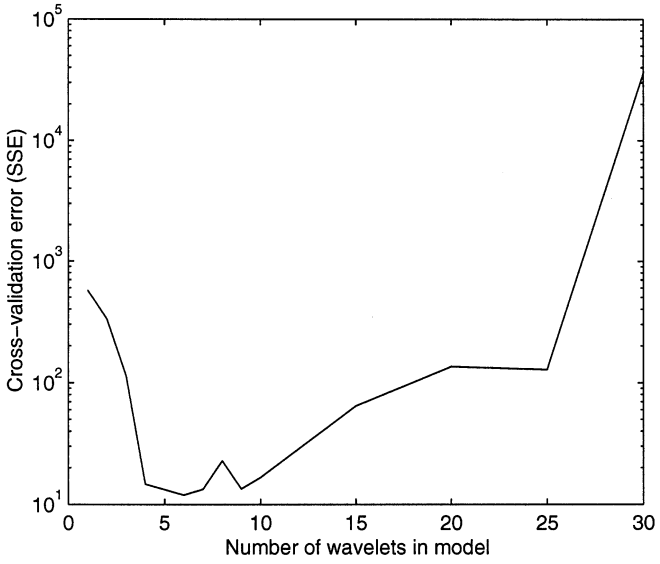

Fig. 10. Multi-step cross-validation for the LMO model of size 6 with 3% additive observational noise.

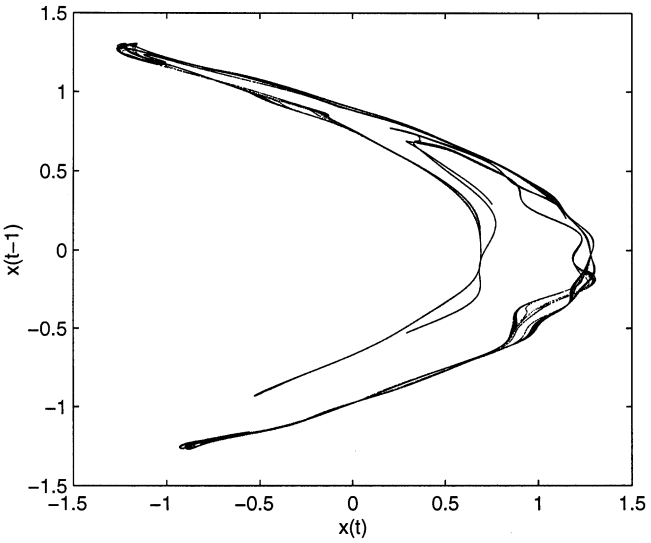

Fig. 11. The attractor for the OMP model of size 32 with 3% additive observational noise.

badly, with no real predictive ability beyond about 4 points ahead. Clearly, the performance of this modeling technique is much reduced in the presence of noise.

The model built using our hybrid method with 3% noise was far superior to the previous one. The multi-step cross-validation error is plotted in Fig. 14. The model's predictive capability extends to about 5 steps ahead, and the attractor closely resembles the true Hénon attractor for about 1000 iterations. After this, the model settled down to a periodic orbit. It is clear from the number of

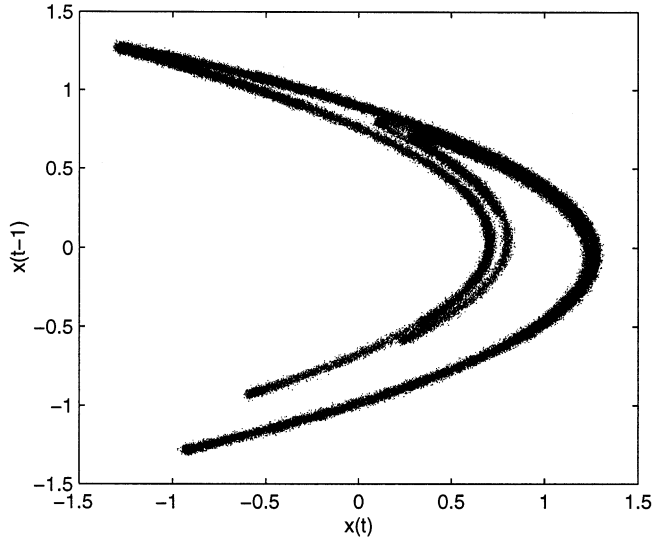

Fig. 12. The Hénon attractor with 3% additive observational noise.

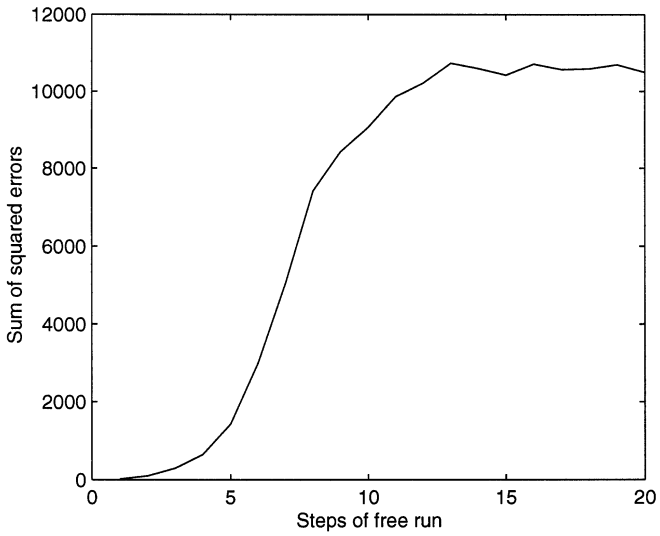

Fig. 13. Multi-step cross-validation for the OMP model of size 32 with 3% additive observational noise.

transient points for this periodic orbit that the model is only weakly stable, and so, a very small amount of dynamic noise would remove this orbit. To check this, 0.01% dynamic noise was added during the free-run calculations, resulting in the non-periodic attractor, shown in red in Fig. 15, which closely resembles the true attractor.

The models built for the noisy Hénon attractor demonstrate the importance of the continuous optimization step in our hybrid technique. Without this, the models that were built, performed very poorly, whereas the addition of this step allowed us

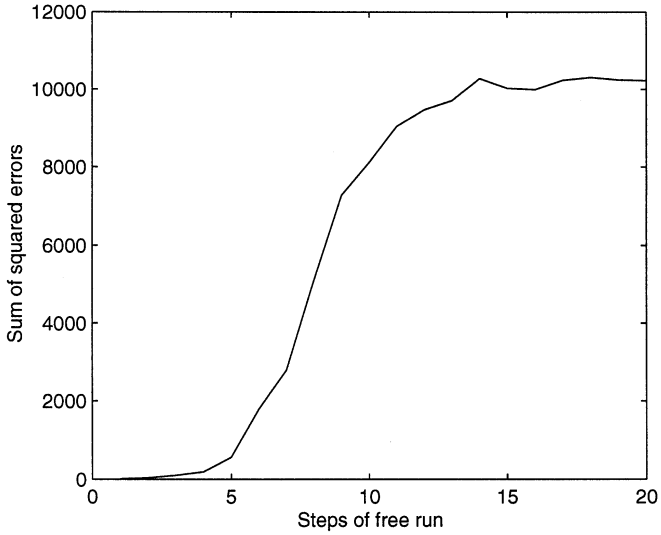

Fig. 14. Multi-step cross-validation for the LMO model of size 6 with 3% additive observational noise.

to build quite reasonable models of the noisy system, suggesting that our technique might be able to successfully model some real-world dynamical systems. In the next section, we apply our hybrid method to data from a vibrating guitar string.

5. Modeling of a Forced Oscillating String

Ultimately, the purpose of any modeling method is to model and study data from real systems. To illustrate such an application with our basis-selection/Levenberg–Marquardt method, we attempted to model the time-series from a forced oscillating string, using data collected by [Molteno & Tufillaro, 1990].

This data is sampled from 2 minutes of string vibration, and it starts out in a chaotic regime. As the string heats up it settles down to a periodic orbit. Judd and Mees [1996] have analyzed this data extensively and have demonstrated that it contains bifurcation structures consistent with a Shil'nikov mechanism. This was accomplished by modeling the experimental data and then analyzing the properties of the model. To test our modeling method, we also adopted this approach, so that we could compare our results with theirs.

We built a model of the initial chaotic region of the data using the technique described in Sec. 3, where we used basis-selection followed by continuous optimization, producing a model containing 6 wavelets. This model accurately captures the

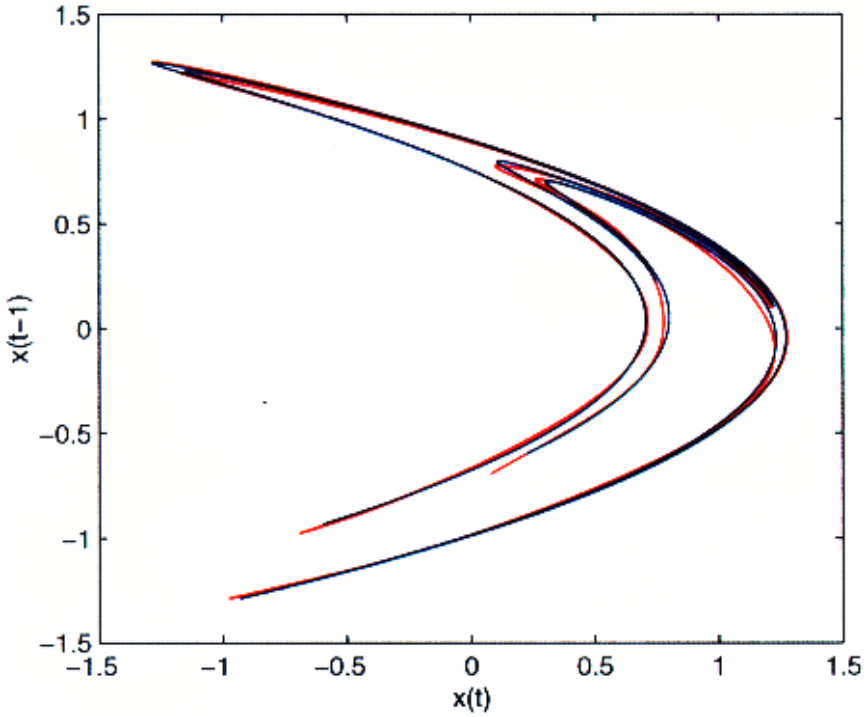

Fig. 15. The attractor for the LMO model of size 6 is shown in red, the true attractor in blue. The model was built from the noisy data in Fig. 12 (Hénon map with 3% additive observational noise).

periodic component of the behavior and for this reason, the multi-step cross-validation remains near 1 for many periods of free-run.

To demonstrate that a Shil'nikov-type bifurcation might be present in the data, it is necessary to calculate the fixed points of our model. This model has three fixed points. These are shown in two-dimensional projection in Fig. 16, along with the 200 points of data used to build the model. Two of these fixed points lie far from the data, and as such, we can make no comment on the reality of these fixed points. The third fixed point, located at $(-0.71, -0.71, -0.71)$, is in good agreement with fixed point A found in the paper of Judd and Mees. This is in approximately the location for a Shil'nikov mechanism.

Calculating the eigenvalues of the linearized system at this point, we obtained $\lambda_1 = 1.5486$ and $\lambda_2 = -0.0805 \pm 0.1148i$. This implies a stable two-dimensional manifold spiralling into the fixed point and a one-dimensional unstable manifold. This is the same type of fixed point as was found by Judd and Mees. Calculating the value $\delta = -\log|\mathrm{Re}(\lambda_2)|/\log\lambda_1$ for this fixed point, we obtain 4.492. This compares poorly with the value of 0.74 obtained by Judd and Mees, and in fact, does not support the existence of a Shil'nikov mechanism as that would require $\delta < 1$. The theorem of Shil'nikov, describes the behavior of the system close to the homoclinic orbit (which in this case approaches the fixed point). Loosely, for $\delta > 1$, there will be a stable periodic orbit (actually a fixed point here); for $\delta < 1$, chaotic motion will arise via bifurcation. For more detailed discussion of this process, see [Glendinning & Sparrow, 1983, 1984; Mees & Sparrow, 1987].

Building optimized wavelet models of other segments of the time-series gave extremely similar results. There was always a fixed point very close to $(-0.7, -0.7, -0.7)$ and the eigenvalues of the map at this point differed very little from those above.

This stability of the fixed point and its characterization, along with the accuracy of our models for free-run predictive purposes, suggest that the disagreement of our results with those of Judd and Mees is due to the type of model we are using rather than the quality of the fit. In regions of little or no data, the model is determined by the type of functions that the method being used utilizes. The radial basis models used by Judd and Mees contain only Gaussian functions with standard deviations on the order of one third of the attractor width. These are much less oscillatory than our wavelet functions and so the model surface will be

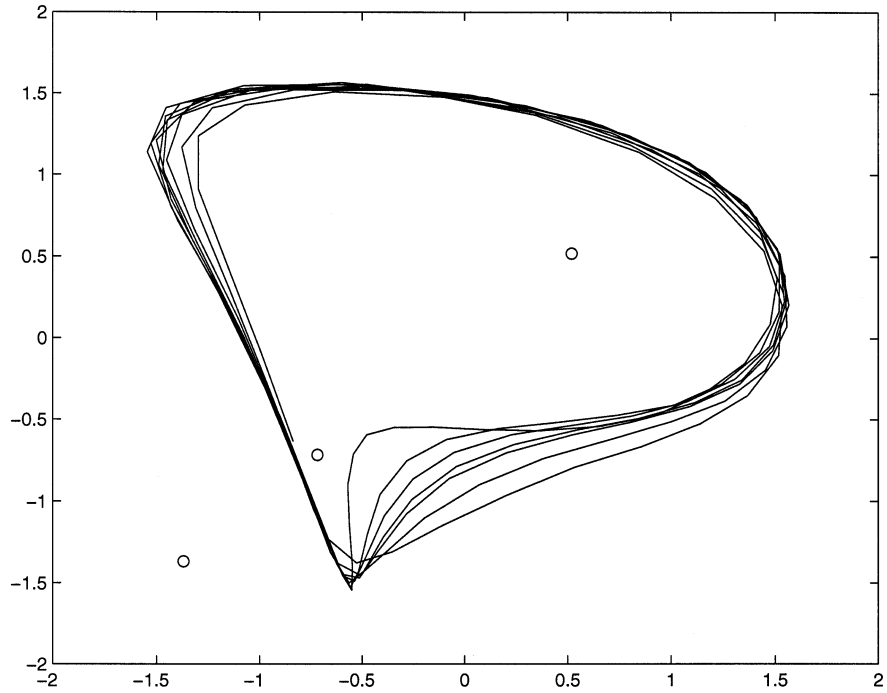

Fig. 16. The circles are the fixed points of the LMO model of the vibrating string data. This was built using 200 points of data, and these points form the orbits shown here.

relatively smoother compared to our wavelet model. If we suspect that the true system is quite smooth, then it is reasonable to assume that a very smooth model is required, and for this reason, the radial basis model might be expected to provide a more accurate approximation to the true map off the attractor, in such regions that we find the fixed point in question.

Nevertheless, the optimized wavelet model accurately captured many of the important characteristics of the system, justifying the use of this modeling approach for real-world data.

6. Conclusion

In this paper we applied a wavelet basis approach to modeling time-series from chaotic dynamical systems. We investigated the use of basis-selection and continuous optimization to build wavelet models, and found that a hybrid of the two, was far more successful than each, on their own. Our technique was applied successfully to some real-world data as well as to the noise-free data of the Hénon map.

The main property of wavelet bases that we hoped might confer some advantage, was that they form a frame (unconditional basis). Donoho [1992] has shown that such bases are optimal for statistical estimation. This did not carry over into our application of function approximation. The important properties for this type of modeling that we utilized, were the smoothness and simple nonoscillatory shape of the wavelets that were used. For this reason, we expect that the technique illustrated in this paper, would extend to building models of dynamical systems from a variety of other functions.

References

Cao, L., Hong, Y., Fang, H. & He, G. [1995] "Predicting chaotic time series with wavelet networks," *Phys.* **D85**, 225–238.

Chen, S. S., Donoho, D. L. & Saunders, M. A. [1996] "Atomic decomposition by basis pursuit," unpublished.

Daubechies, I. [1992] *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Series in Applied Mathematics, SIAM.

Donoho, D. L. [1992] "Unconditional bases are optimal bases for data compression and for statistical estimation," Technical Report, Department of Statistics, Stanford University.

Glendinning, P. & Sparrow, C. T. [1983] "Local and global bahaviour near homoclinic orbits," *J. Stat. Phys.* **35**, 645–697.

Glendinning, P. & Sparrow, C. T. [1983] "Bifurcations

near homoclinic orbits," in *XVIth Int. Congress of Theoretical and Applied Mechanics*.

Judd, K. & Mees, A. I. [1995] "On selecting models for nonlinear time series," *Phys.* **D82**, 426–444.

Judd, K. & Mees, A. I. [1996] "Modeling chaotic motions of a string from experimental data," *Phys.* **D92**, 221–236.

Mallat, S. & Zhang, Z. [1993] "Matching pursuit in a time-frequency dictionary," *IEEE Trans. Sig. Process.* **41**, 3397–3415.

Marquardt, D. W. [1963] "An algorithm for least-squares estimation of nonlinear parameters," *J. SIAM* **11**, 431–441.

Mees, A. I. & Sparrow, C. T. [1987] "Some tools for analyzing chaos," *Proc. IEEE* **75**, 1058–1070.

Molteno, T. C. A. & Tufillaro, N. B. [1990] "Torus doubling and chaotic string vibrations: Experimental results," *J. Sound Vib.* **137**, 327–330.

Pati, Y. C., Rezaiifar, R. & Krishnaprasad, P. S. [1993] "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Systems and Computers*, ed. Singh, A.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. [1988] *Numerical Recipes in C* Cambridge (Cambridge University Press).

Rissanen, J. [1989] *Stochastic Complexity in Statistical Inquiry*, Series in Computer Science **15** (World Scientific, Singapore).