A BOOTSTRAP DETECTION FOR OPERATIONAL DETERMINISM

Qiwei Yao and Howell Tong

Institute of Mathematics and Statistics, University of Kent at Canterbury

Canterbury, Kent CT2 7NF, U.K.

Abstract

We propose a bootstrap detection for *operationally deterministic* versus stochastic nonlinear modelling and illustrate the method with both simulated and real data sets.

Keywords: Bandwidth; Kernel regression; Opertional determinism; Stochastic noise.

Corresponding author: Howell Tong, Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury, Kent CT2 7NF, U.K. Tel: ++44 1227 823658, Fax: ++44 1227 827932, Email: ht1@ukc.ac.uk.

1 Introduction

It is well known that the pseudo-randomness associated with a deterministic chaotic system has low dimensional attractors. By contrast, randomness is typically associated with a stochastic system. It is always an interesting and challenging problem to distinguish between deterministic chaos and stochastic randomness. Besides the many obvious philosophical implications, into which we shall not enter, there are important practical consequences. For example, if we accept that the data at hand are essentially generated by a deterministic dynamical system (DDS), then we can call on a substantial collection of modern tools, e.g. the Takens embedding, fractal dimension, Lyapunov exponent, map reconstruction and others, for their analysis and prediction. On the other hand, if the data are stochastic, a stochastic dynamical system (SDS) will provide a more appropriate model. In fact, in this case, many of the standard tools in DDS are not directly applicable without modification. For example, even fundamental notions such as embedding dimension and initial value sensitivity would have to be carefully re-defined and routine applications of tools designed for a DDS may lead to misleading conclusions.

What distinguishes a DDS is the fact that it is completely free of (stochastic) dynamic noise (also called system noise). Let us agree to ignore observation noise in our discussion. In a real practical situation, we might be willing to relax the 'purity' of a DDS by allowing an 'insignificant' amount of dynamic noise because pragmatism might dictate this. We are then faced with the inevitable problem of detecting *operational determinism*. In practice, this requires us to develop sufficiently powerful statistical tools which have a high detection rate of any underlying dynamic noise.

At varying degrees of sophistication, several ingeneous tools have been developed recently for the above purpose [1–3]. A common feature of these tools is the use of locally valid models over the reconstructed state space for good short-term prediction. Normally, a DDS would allow the locally valid models to be really very local (e.g. covering only a very small number of observations) and this observation forms the basis of the above methods, which differ mainly in the implementation of the strategy. Either directly or indirectly the above methods have utilised the so-called nearest neighbours. For example, the most recent nearest-neighbour technique for detecting operational determinism seems to be due to Casdagli [3], who constructed an ingenious forecasting algorithm using the k nearest neighbours. He has suggested that 'a small value of k corresponds to a deterministic approach to modelling. The largest value of k corresponds to fitting a stochastic linear autoregressive model. Intermediate values of k correspond to fitting non-linear stochastic models'. It is well known in the statistical literature that k-nearest neighbours are statistics with complex properties. This is perhaps one of the reasons why to-date the methods mentioned above have not been given a firm theoretical foundation. Further, there has been no objective way to assess how small is *small* for the value of k, which should be properly addressed since the number of neighbours used in estimation should depend on the sample size. The calibration effect due to sampling fluctuation should not be overlooked either.

In this paper we propose a new statistical method for detecting operational determinism based on the use of the bandwidth statistic in the kernel smoothing, and demonstrate its practical utility via real and simulated data. Taking advantage of the simpler sampling properties of the kernel smoothing, we have arrived at a conclusion which is almost the same as Casdagli's mentioned above, the only difference being the role of the number of nearest neighbours, which is now replaced by the bandwidth. Furthermore, we have introduced a bootstrap method to assess how small is small for a bandwidth to claim that the system is *operationally* deterministic, which also reduce the calibration effect due to sampling fluctuation. For the sake of brevity, we shall omit some of the more esoteric mathematical details but make them available to the interested readers upon request.

Let $\{Y_t,\ -d+1 \le t \le n\}$ be a sample from a strictly stationary discrete-time time series generated by an unknown model

$$Y_t = f(Y_{t-1}, \dots, Y_{t-d}) + \epsilon_t \equiv f(X_t) + \epsilon_t, \tag{1}$$

where $X_t = (Y_{t-1}, \dots, Y_{t-d})^\tau$, $\epsilon_t = Y_t - f(X_t) = Y_t - \mathrm{E}\{Y_t|X_t\}$. Of interest is to decide whether ϵ_t is small enough to be negligible from a data analytical point of view, based on the observed data $\{Y_t,\ -d+1 \le t \le n\}$. We call the system operationally deterministic if ϵ_t is negligible. Obviously, this includes a purely deterministic system ($\epsilon_t \equiv 0$) as a special case.

2 Kernel regression estimation

We start with the estimation of the regressive function $f(x) = \mathrm{E}\{Y_1|X_1 = x\}$. Given the observations $\{Y_t;\ -d+1 \le t \le n\}$, one of the conventional nonparametric estimators of $f(x)$ is the Nadaraya-Watson kernel regression estimator, which can be viewed as the solution of the following *local* least squares problem

$$\hat{f}(x) = \arg\min_a \sum_{t=1}^{n} \{Y_t - a\}^2 K\left(\frac{X_t - x}{h}\right),$$

where $K(.)$ is a kernel function, and $h > 0$ is the bandwidth. Typically, $K(x) \searrow 0$ fairly quickly when $||x|| \nearrow \infty$, and h is small. For example, if $K(x) \propto 1$ for $||x|| \le 1$ and $K(x) = 0$ otherwise, $\hat{f}(x)$ is the average of all the Y_t's for which $||X_t - x|| \le h$. Therefore, similar to the nearest-neighbour method, the kernel regression is a local average, or more precisely, a locally weighted average in general. It is clear that the bandwidth h controls the amount of data which is effectively used in the estimation, and thus plays a role similar to the number of nearest neighbours in the nearest neighbour estimation. For further discussion of kernel regression, we refer to Härdle [4].

Note that for z around x, we have the approximation

$$f(z) \approx f(x) + \dot{f}(x)(z - x).$$

This suggests the locally linear regression estimator: $\hat{f}_{n,h}(x) = \hat{a}$, where $(\hat{a}, \hat{b})$ minimize

$$\sum_{t=1}^{n} \{Y_t - a - b^\tau (X_t - x)\}^2 K\left(\frac{X_t - x}{h}\right).$$

It has been pointed out that the locally linear regression method has various advantages over the conventional Nadaraya-Watson method [5,6]. Intuitively it is easy to understand that the locally linear fit can accommodate more local variation of the curve $f(.)$ than the locally constant fit.

In kernel regression, the quality of the estimation depends critically on the bandwidth h. For the purpose of prediction, we would ideally choose h which gives the best prediction for the future observations $\{Y_{n+1}, ..., Y_{n+n_2}\}$, *i.e.* h minimizes the following mean squared errors:

$$M_n(h) = \frac{1}{n_2} \sum_{1}^{n_2} \{\hat{f}_{n,h}(X_{n+t}) - f(X_{n+t})\}^2 w(X_t),$$

where $w(.)$ is a weight function. We have shown that if model (1) satisfies some mild regularity conditions and $\text{Var}(\epsilon_t) > 0$,

$$M_n(h) = \frac{h^4 \sigma_0^4}{4} \int [\text{tr}\{\ddot{f}(x)\}]^2 p(x) w(x) \text{d}x + \frac{1}{nh^d} \int \sigma^2(x) w(x) \text{d}x \int K^2(u) \text{d}u + o\left(h^4 + \frac{1}{nh^d}\right), \quad (2)$$

as both n and m converge to ∞, where $\ddot{f}(x) = \frac{\partial^2}{\partial x \partial x^\tau} f(x)$, $\sigma^2(x) = \text{Var}(Y_1|X_1 = x)$, $p(x)$ is the marginal probability density function of X_1, $\int uu^\tau K(u)\text{d}u = \sigma_0^2 I_d$ and I_d denotes the $d \times d$ identity matrix. In fact, it can be proved that for a (purely) deterministic and ergodic model, (2) still holds with $\sigma^2(x) \equiv 0$. On ignoring the higher order term on the LHS of the above, the minimizer of $M_n(h)$ is

$$h_n = n^{-\frac{1}{d+4}} \left\{ \frac{\int K^2(u)\text{d}u \int \sigma^2(x) w(x) \text{d}x}{\sigma_0^4 \int [\text{tr}\{\ddot{f}(x)\}]^2 p(x) w(x) \text{d}x} \right\}^{\frac{1}{d+4}}. \quad (3)$$

We have three options: (i) use $h = h_n \approx 0$ when the noise is small enough (i.e. $\sigma^2(x)$ is small enough) such that the second term of the RHS of (2) can be ignored; (ii) use $h = h_n = \infty$ when the model is linear (i.e. $\text{tr}[\ddot{f}(x)] \equiv 0$); (iii) use $h = h_n \in (0, \infty)$ when the model is nonlinear and stochastic. These are exactly in the same spirit as Casdagli's suggestion mentioned before.

Note that h_n given in (3) involves several unknown quantities and cannot be directly used in practice. Various data-driven methods to determine h have been developed among which the cross-validation approach is the most frequently used method and offers an estimate which is equivalent to h_n asymptotically, see, for example, [8].

To speed up the computation for the cross-validation method, we propose a modified version as follows: we first split the sample into two subsets $\{(X_t, Y_t) : 1 \le t \le m\}$ and $\{(X_t, Y_t) : m+1 \le t \le n\}$. We estimate $f(.)$ using the first m observations and let it be denoted by $\hat{f}_{m,h}(.)$. We choose h such that $\hat{f}_{m,h}(.)$ gives the best prediction for Y_t for $m+1 \le t \le n$ in the sense that $h = \tilde{h}_m$ minimizes

$$\frac{1}{n-m} \sum_{t=m+1}^{n} \{Y_t - \hat{f}_{m,h}(X_t)\}^2 w(X_t).$$

According to (3), the bandwidth with the whole sample should be

$$\hat{h}_n = \left(\frac{m}{n}\right)^{\frac{1}{d+4}} \tilde{h}_m. \tag{4}$$

We have proved that $\hat{h}_n$ and h_n are asymptotically equivalent.

3 Detection for operationally deterministic systems

We now propose a method to detect that system (1) is operationally deterministic. It will be based on the statistic $\hat{h}_n$ defined as in (4). In practice, the selected bandwidth is always positive. Thus, the event that the selected bandwidth is close to zero would indicate that model (1) is operationally deterministic. Of course, it remains to decide how close is *close* in this context. Furthermore, there exists a potential danger that the small value of the selected bandwidth is due to sampling fluctuations. To overcome the problems mentioned above, we propose using the bootstrap method, which is a computational device to obtain tail-probabilities when the latter do not admit analytical expressions. The basic idea is to calibrate the $\hat{h}_n$ at hand by reference to the distribution of $\hat{h}_n$ obtained by repeated sampling. This is another aspect where our method differs from the existing ones [1–3] in that we have taken particular care to reduce the chance of fortuitous calibration. Note that the bootstrap method is different from and predates the method of surrogate data. For an elementary account of the former, see, e.g., Efron and Tibshirani [9]. The following constitutes our bootstrap detector for operational determinism. We represent the observations $\{Y_t, -d+1 \le t \le n\}$ in the form $\{(X_t, Y_t), 1 \le t \le n\}$ with $X_t = (Y_{t-1}, \dots, Y_{t-d})^\tau$.

1. For the given data $\{(X_t, Y_t), 1 \le t \le n\}$, obtain the estimate $\hat{h}_n$ as given in (4).

2. Obtain the locally linear regression estimator $\hat{f}_{n,h}(.)$ using $h = \hat{h}_n$, and calculate the residuals $\hat{\epsilon}_t = Y_t - \hat{f}_{n,\hat{h}_n}(X_t)$ for $t = 1, \dots, n$.

3. Bootstrap: draw n independent random numbers $i_1, \dots, i_n$ from the uniform distribution with the sample space $\{1, \dots, n\}$, and define $\epsilon_t^* = \hat{\epsilon}_{i_t}$ for $t = 1, \dots, n$. Form the bootstrap sample

$\{(X_t, Y_t^*),\ 1 \le t \le n\}$ with

$$Y_t^* = \hat{f}_{n,\hat{h}_n}(X_t) + \epsilon_t^*.$$

4. Obtain an estimate $\hat{h}_n^*$ from the sample $\{(X_t, Y_t^*),\ 1 \le t \le n\}$ as in Step 1. Especially, the search for $\hat{h}_n^*$ around h_n is conducted on finer grids than those used in Step 1.

5. Repeat Steps 3 and 4 N times, and count the frequency of occurrence of the event that $\hat{h}_n^* \le \hat{h}_n$. Then the relative frequency α (= frequency/N) is taken as a measure of how plausible the data are generated by an operationally deterministic model.

Remark. In the above, values of α near to 1 provide evidence of the system being operationally deterministic; values of α around 0.5 provide evidence of the system being nonlinear and stochastic (i.e. $\ddot{f}(x) \not\equiv 0$). Small values of α may be taken to indicate that $f(.)$ is linear, or Y_t is simply the 'white noise' (i.e. $f(x) \equiv 0$).

The key idea of the proposed method can be explained as follows. For a noise-free, or an almost noise-free system, (3) indicates that $\hat{h}_n$ would take as small a value as computations permit. (Note that too small an h will cause overflow in computation.) Since the search for $\hat{h}_n^*$ is conducted on finer grids around $\hat{h}_n$, $\hat{h}_n^*$ tends to take a slightly smaller value than $\hat{h}_n$.

Finally, we note that the above bootstrap detector can be defined in terms of any reasonable data-driven bandwidth selector.

4 Examples

To illustrate the above method, we report simulation studies for five examples, which include some purely deterministic models. The application of the detector to four real data sets is reported in Example 6.

In all the examples below, we search for $\hat{h}_n$ among 100 possible values. We always set $K(.)$ equal to the Gaussian kernel, and $N = 100$ for the number of bootstrap replications. For each simulated model, we replicate the Monte Carlo experiments 100 times. For the first three examples, we always set $n = 300$ for the sample size, $m = 200$ for the estimation of $\hat{h}_n$, and $w(.)$ equal to the indicator function of the 90% inner samples. For Examples 3.4 and 3.5, we set $n = 500$, $m = 300$ and $w(.) \equiv 1$.

Example 1. Let

$$Y_t = 0.246 Y_{t-1}(16 - Y_{t-1}) + \sigma\epsilon_t, \tag{5}$$

Table 1. The bootstrap detection for model (5)

σ	Mean(α)	Variance(α)	Mean($\hat{h}_n$)	Variance($\hat{h}_n$)	Mean($\hat{h}_n^*$)	Variance($\hat{h}_n^*$)
0.07	0.5134	0.0750	0.0425	0.0001	0.0436	0.0001
0.04	0.6185	0.1105	0.0354	0.0001	0.0354	0.0001
0.01	0.6618	0.0390	0.0207	0.0000	0.0196	0.0001
0.008	0.6831	0.0653	0.0209	0.0000	0.0190	0.0000
0.005	0.9558	0.0451	0.0201	0.0000	0.0162	0.0000

Table 2. The bootstrap detection for the regression of $Y_t = X_{t+m}$ on X_t, where X_t is determined by (6)

m	Mean(α)	Variance(α)	Mean($\hat{h}_n$)	Variance($\hat{h}_n$)	Mean($\hat{h}_n^*$)	Variance($\hat{h}_n^*$)
1	1.0000	0.0000	0.0201	0.0000	0.0124	0.0000
3	0.9834	0.0041	0.0201	0.0000	0.0131	0.0000
5	0.9192	0.0083	0.0694	0.0109	0.0586	0.0098
7	0.3882	0.0183	1.6044	1.1220	1.6521	1.1474
9	0.3005	0.0273	3.6475	3.1301	3.9109	3.2202
11	0.2985	0.0279	3.6187	3.1285	3.8787	3.2227

Table 3. The bootstrap detection for the tent map (7) and its time reversal

	Mean(α)	Variance(α)	Mean($\hat{h}_n$)	Variance($\hat{h}_n$)	Mean($\hat{h}_n^*$)	Variance($\hat{h}_n^*$)
Model (7)	0.9596	0.0549	0.0115	0.0000	0.0115	0.0000
Time reversal	0.3260	0.0248	0.8936	0.4257	1.0378	0.4407

where $\sigma > 0$ is a constant, and $\{\epsilon_t,\ t \geq 1\}$, are independent random variables with the same distribution as the random variable 0.5η, and η is equal to the sum of 48 independent random variables each uniformly distributed on $[-0.5, 0.5]$. According to the central limit theorem, we can treat ϵ_t as almost standard normal. However, it has a bounded support $[-12,\ 12]$. The simulation has been carried out for the cases with σ equal to five different values between 0.07 and 0.005. The average α-values in 100 replications of the Monte Carlo experiments are reported in Table 1, which show that the bootstrap detector has no difficulties in identifying the model being nonlinear and stochastic when $\sigma \geq 0.01$. But when $\sigma = 0.005$, the bootstrap detector shows that we could operationally treat (5) as a deterministic model. Note that the noise-to-signal ratio ($\equiv \frac{\sigma}{\{\mathrm{Var}(Y_t)\}^{1/2} - \sigma + |\mathrm{E}Y_t|}$) is about 0.05% when $\sigma = 0.005$. The means and variances of $\hat{h}_n$ in the 100 replications , together with their bootstrap counterparts (in 10000 (= 100×100) replications), are also included in the table.

Example 2. For the transformed standard logistic model (with coefficient 4)

$$X_{t+1} = 0.25X_t(16 - X_t), \tag{6}$$

we consider the cases $Y_t = X_{t+m}$, $m = 1, 3, 5, 7, 9$, and 11. The results are reported in Table 2. We can see that the bootstrap detector has no difficulties in confirming that we can model X_{t+k} as a deterministic function of X_t for $m \leq 5$. However, for $m \geq 7$ the α-values are considerably smaller, which shows that now it will be difficult to model X_{t+m} as a deterministic function of X_t with the given data.

Example 3. For the tent map

$$X_{t+1} = \begin{cases} aX_t & 0 \leq X_t < 0.5 \\ a(1 - X_t) & 0.5 \leq X_t < 1 \end{cases} \tag{7}$$

with $a = 2$, time reversal gives the stochastic model

$$X_t = \frac{1}{2} + \frac{1}{2}\epsilon_t(1 - X_{t+1}),$$

where $\{\epsilon_t\}$ is a sequence of independent random variables and ϵ_t equals 1 or -1 with equal probability [10,11]. Therefore, although the original time series is generated from a purely deterministic model, the reverse time series can be viewed as a sequence generated by a stochastic model. We apply the bootstrap detector to both original and reversed time series. In our simulation, we use $a = 1.9999999$ as a surrogate for $a = 2$ which is unrealisable[11]. We set $(a_l, a_u) = (0.05, 5)$. The results are reported in Table 3. The detector confirms that the time series from (7) is operationally deterministic. However,

Table 4. The bootstrap detection for model (8)

σ	Mean(α)	Variance(α)	Mean($\hat{h}_n$)	Variance($\hat{h}_n$)	Mean($\hat{h}_n^*$)	Variance($\hat{h}_n^*$)
0.09	0.6091	0.2168	0.1391	0.0.0035	0.1407	0.0029
0.07	0.5540	0.2262	0.1358	0.0038	0.1431	0.0046
0.05	0.8287	0.0764	0.1421	0.0046	0.1054	0.0031
0.03	0.9361	0.0512	0.1289	0.0028	0.0910	0.0018
0.01	0.9675	0.0190	0.1303	0.0029	0.0851	0.0019

Table 5. The bootstrap detection for the modelling of $Y_t = X_{t+m}$ on X_t,

where X_t is determined by (9)

m	Mean(α)	Variance(α)	Mean($\hat{h}_n$)	Variance($\hat{h}_n$)	Mean($\hat{h}_n^*$)	Variance($\hat{h}_n^*$)
1	0.9425	0.0378	0.1312	0.0026	0.0862	0.0013
3	0.9246	0.0353	0.1095	0.0015	0.0773	0.0013
5	0.7087	0.1273	0.1240	0.0044	0.1140	0.0068
7	0.2884	0.1750	0.3565	0.1704	0.4572	0.0951
9	0.3091	0.1867	0.4391	0.2035	0.4907	0.1029
11	0.2699	0.1015	0.6999	0.3217	0.8394	0.1572

Table 6. The bootstrap detection for the four real data sets

data set	Laser data	GSL	Sunspot numbers	Measles data
regressors	Y_{t-1}, Y_{t-2}	Y_{t-12}, Y_{t-24}, Y_{t-36}, Y_{t-48}	Y_{t-1}, Y_{t-2}, Y_{t-4}	Y_{t-1}, Y_{t-4}, Y_{t-7}
n	998	3379	289	151
m	700	2500	200	101
α-value	1.00	1.00	0.43	0.27
$\hat{h}_n$	0.1090	0.0713	0.4123	0.3833
Mean($\hat{h}_n^*$)	0.0610	0.0357	0.4162	0.4762
Variance($\hat{h}_n^*$)	0.0000	0.0000	0.0057	0.0065

for the reversed time series the estimated α-value is much smaller than 0.5 (cf. Remark in last section).

Example 4. Let us consider the model

$$Y_t = 20 - 0.0645Y_{t-1}^2 + 0.3Y_{t-2} + 4\sin(0.05Y_{t-3}) + \sigma\epsilon_t, \tag{8}$$

where $\sigma > 0$ is a constant, and $\{\epsilon_t\}$ is the same as in (5). The simulation has been carried out for five different values of σ between 0.01 and 0.09. The results are reported in Table 4. The detector confirms that when $\sigma \geq 0.07$, the model is nonlinear and stochastic. However, when $\sigma \leq 0.03$, the detector suggests that we could operationally treat (8) as a deterministic model. Note that when $\sigma = 0.03$, the ratio of noise to signal is about 0.11%.

Example 5. For the purely deterministic model

$$X_{t+1} = 1 - 1.3Y_{t-1}^2 + 0.3Y_{t-2} + 0.2\sin(Y_{t-3}), \tag{9}$$

we apply the detector to the cases that $Y_t = X_{t+m}$ for $m = 1, 3, 5, 7, 9$, and 11. The results are reported in Table 5. For $m \leq 3$, it has been identified as an operationally deterministic model. For $m \geq 7$, the detector seems to suggest that it would not be prudent to model X_{t+m} as a deterministic function of X_{t-1}, X_{t-2} and X_{t-3} with given data. The detector fails to make a clear suggestion for the case when $m = 5$.

Example 6. We have applied the bootstrap to the folowing four real data sets:

(i) NH_3-FIR Laser data set. This is the first data set used in the *Santa Fe Time Series Prediction and Analysis Competition.* The 1000 data were generated by a physics laboratory experiment (NH_3 laser), which is believed to generate Lorenz-like chaos [12]. We have fitted the data with a nonlinear autoregressive model of order 2, as determined by the cross-validation method [13].

(ii) The Great Salt Lake (GSL) data from Utah [14]. We have fitted the biweekly volume data of the GSL with a nonlinear autoregressive model with sampling time 12 and order 4. The order was determined by the cross-validation method.

(ii) Wolf's annual sunspot numbers (1700-1992). We have fitted the data with the optimal subset regression model determined by the the cross-validation method, and the optimal regression subset [15] consists of the lagged variables at lags 1, 2, and 4.

(iii) The monthly New York measles data. In order to avoid possible outliers, we use only the first 158 points. We have fitted the data, on the natural log base, with the optimal subset regression model determined by the the cross-validation method, and the optimal regression subset consists of the lagged variables at lags 1, 4, and 7.

We have standardized the data first in each case. The results of the detection are summarized in Table 6, which show that the bootstrap detector identifies that both the Laser data the Great Salt Lake data are operationally deterministic. This conclusion is not surprising as far as the Laser data are concerned. For the GSL data, Sangoyomi [14] have given an explanation for operational determinism. For the other two data sets, nonlinear stochastic models are suggested.

References

[1] Farmer, J.D. and Sidorowich, J.J., Predicting chaotic time series. *Phys. Rev. Lett.*, **59**, 845-848 (1987).

[2] Sugihara, G. and May, R.M., Nonlinear forecasting as a way of distinguishing chaos from measurement errors in time series. *Nature*, **344**, 734-741 (1990).

[3] Casdagli, M., Chaos and deterministic versus stochastic non-linear modelling. *J. Roy. Statist. Soc.* **B**, **54**, 303- 328 (1992).

[4] Härdle, W., *Applied Nonparametric Regression* (Cambridge University Press, Cambridge, 1990).

[5] Fan, J., Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998-1004 (1992).

[6] Hastie, T. and Loader, C., Local refression: automatic kernel carpentry. *Statist Science*, **8**, 120-143 (1993).

[7] Wand, M.P. and Jones, M.C., *Kernel Smothing* (Chapman & Hall, London, 1995).

[8] Härdle, W. and Vieu, P., Kernel regression smothing of time series. *J. Time Series Anal.*, **13**, 209-232 (1992).

[9] Efron, B. and Tibshirani, R.J., *An introduction to the bootstrap*, (Chapman & Hall, London, 1993).

[10] Tong, H. and Cheng, B., A Note on one-dimensional chaotic maps under time reversal. *Adv. Appl. Prob.*, **24**, 219-220 (1992).

[11] Lawrance, A.J. and Spencer, N.M., Statistical aspects of curved chaotic map models and their stochastic reversals. In *Complex Stochastic Systems and Engineering* ed. D. M. Titterington, 55-58, (Clarendon Press, Oxford, 1995).

[12] Weigend, A.S. and Gershenfeld, N.A., *Time Series Prediction.* (Addison-Wesley, Reading, 1994).

[13] Cheng, B. and Tong, H., On consistent non-parametric order determination and chaos. *J. Roy. Statist. Soc.* **B**, **54**, 427-449 (1992).

[14] Sangoyomi, T.B., Lall, U. and Abarbanel, H.D.I., Nonlinear dynamics of the Great Salt Lake: dimension estimation. *Water resources Research*, **32**, 149-159 (1996).

[15] Yao, Q. and Tong, H., On subset selection in non-parametric stochastic regression. *Statistica Sinica*, **4**, 51-70 (1994).