Available online at www.sciencedirect.com
SCIENCE DIRECT®
PHYSICS LETTERS A
Physics Letters A 318 (2003) 373–379
www.elsevier.com/locate/pla

Towards coherent estimation of correlation dimension

A. Guerrero [a,1], L.A. Smith [a,b,*]

[a] *Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK*
[b] *Centre for the Analysis of Time Series, London School of Economics, Houghton St., London WC2A 2AE, UK*

Received 9 March 2002; received in revised form 28 August 2003; accepted 12 September 2003

Communicated by A.P. Fordy

Abstract

A new method for coherent estimation of scaling exponents is presented and demonstrated in the context of the correlation dimension. The method is based on contrasting the distribution of Takens' estimators at a given length scale (which is known to be Gaussian) with the distribution of those estimators at smaller length scales (which is again Gaussian, but typically has larger variance). Requiring consistency with all smaller length scales allows a coherent (that is, internally consistent) estimate of the correlation dimension. It is not possible, of course, to place (non-trivial) bounds on the true dimension with any finite sample. The technique is developed and illustrated on sets where the dimension is known a priori. Macroscopic structure of more typical fractal sets is shown to limit the accuracy with which the correlation dimension is known, even for well studied sets like the Hénon attractor.

PACS: 05.45.Tp; 02.60.Pn

Keywords: Correlation dimension; Takens' estimator; Coherent estimation; Scaling exponents; Gaussian distribution; Grassberger–Procaccia algorithm; Fractal dimension

1. Introduction

The correlation dimension is one of the set of Rényi dimensions which characterise the scaling properties of a distribution of points on an M-dimensional space [8]. While it is, perhaps, the most frequently estimated dimension, it is not possible to put absolute bounds on the error of estimation from a finite sample (at least, not beyond the trivial bounds of zero and M), simply because we can neither take the limit of vanishing length scales nor consider infinite number of points. This Letter presents a new approach that allows one to make coherent estimates of the correlation dimension; such estimates are consistent with all the available information. This is accomplished by exploiting the (distributional) properties of the Takens' estimator, which are well defined at each length scale, and then imposing a consistency constraint between all length scales observed.

We first illustrate the technique on sets where the dimension is known a priori and then consider the

* Corresponding author.
E-mail addresses: martine1@maths.ox.ac.uk (A. Guerrero), l.smith@lse.ac.uk (L.A. Smith).
[1] Present address: Canadian Centre for Climate Modelling and Analysis, University of Victoria, PO Box 1700, Victoria, BC V8W 2Y2, Canada.

0375-9601/$ – see front matter
doi:10.1016/j.physleta.2003.09.023

results for the set corresponding to the original Hénon attractor. Additional options applicable only to time series data will be presented elsewhere.

2. The correlation dimension and Takens' estimator

Consider a set in $\mathbb{R}^M$ with a probability measure μ. Given a set of points x_i $(i = 1, 2, \ldots, n)$ drawn at random from this set, the correlation integral is defined as [3],

$$C_2(r) = \lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} \Theta\big(r - \|x_i - x_j\|\big), \tag{1}$$

where $\Theta(x)$ is the Heaviside function which is zero for a negative argument and one otherwise. The sum gives the number of pairs of points (i, j) whose distance is less than r and $C_2(r)$ reflects the probability that two randomly chosen points are closer than r. As r approaches zero, we expect the correlation integral to behave like

$$C_2(r) = \Phi(r) \cdot r^{\nu}, \tag{2}$$

where ν is the correlation exponent and is equal to the generalised Rényi dimension d_2 [8]. $\Phi(r)$ is a function that reflects the lacunarity of the set [1,10,15].

The Takens' estimator, $T_2(r)$, is based on the distances between randomly selected points. Assume that these distances are independent and randomly distributed according to the probability

$$P(r_p < r) = C_2(r) = \Phi(r) \cdot r^{\nu}. \tag{3}$$

The likelihood of observing a pair of points separated by the distance r is thus expressed as a function of the parameters ν and Φ. Assuming Φ is constant, the value of ν that maximises the probability of finding the observed distances r_p, is given by the Takens' estimator [14],

$$T_2(r) = \left[\frac{-1}{N_p - 1} \sum_{p=1}^{N_p} \log\left(\frac{r_p}{r}\right)\right]^{-1}, \tag{4}$$

where r_p are the distances between randomly chosen points which are smaller than r. In the limit $r \to 0$ and $N_p \to \infty$ then $T_2(r) \to d_2$, assuming this limit exists. In this case Φ is constant (or approaches a constant as $r \to 0$) and the Takens' estimator is in fact optimal [16]. For a discussion of cases where lacunarity cannot be ignored, see [15]. In cases where $\lim_{r\to 0} \Phi(r)$ is not a constant, T_2 usually fails to converge in the same limit. As noted by Borovkova et al. [2], the original Takens' estimator is in fact biased but the slight correction (included in Eq. (4)) of replacing N_p by $N_p - 1$ in the denominator, yields an unbiased minimum variance estimator.[2] In the next section we give additional details on the estimator itself.

3. Coherent estimation

For finite values of N_p (the number of distances), the distribution of the Takens' estimator at a given length scale is Gaussian with mean equal to $\overline{T}_2(r)$, and a standard deviation that increases due to sampling uncertainty as r decreases [12]. Our uncertainty in $T_2(r)$ will be smallest at the largest values of r, yet the quality of $T_2(r)$ as an estimator of d_2 is greatest at the smallest values of r. We thus seek a *coherent* estimate of d_2: the estimate of $T_2(r^\star)$ and its standard error where $r^\star$ is chosen such that $T_2(r^\star)$ is internally consistent with all $T_2(r)$, $r < r^\star$.

To illustrate our main result, we consider a set of points randomly chosen with uniform density on a line segment. Fig. 1 shows 64 different sample estimators $\widehat{T}_2(r)$ with $N_p = 2^{12}$ distances. Since we know the dimension a priori in this case, we use it to illustrate the characteristic behaviors of our approach. The distribution of $\widehat{T}_2$ is tightest (small standard deviation) at large length scales, hence the uncertainty in the estimator is smallest at the least relevant length scales. In Fig. 1, for example, at $r = 2^{-1}$ we have a very precise value of $\overline{\widehat{T}}_2(2^{-1})$ of 0.860 ± 0.004. Examination of the graph shows that this value cannot be taken as coherent in the limit $r \to 0$, as it is not even consistent with $r = 2^{-3}$ ($\overline{\widehat{T}}_2(2^{-3}) = 0.969 \pm 0.007$).

At the smallest scales where there are, by construction, the fewest pairs of points; the distribution of T_2 is often observed not to be Gaussian. This is seen in Fig. 1 (for $r = 2^{-10}$). Intuitively, as $\widehat{T}_2(r)$ is positive

[2] If the difference between $1/N_p$ and $1/(N_p - 1)$ is significant, then the data set is unlikely to be of interest in terms of dimension estimation.

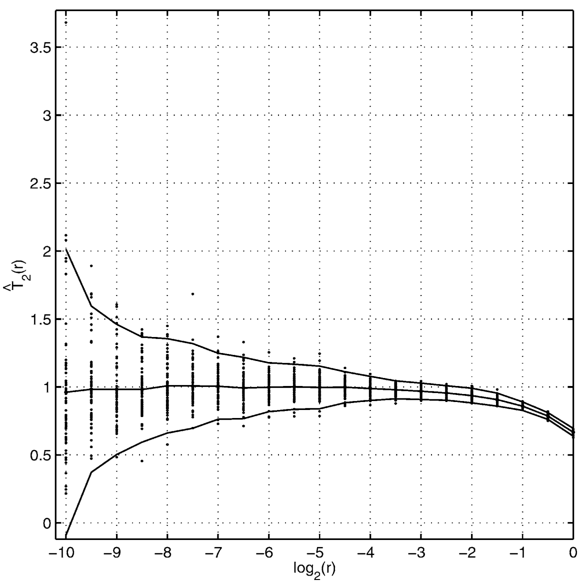

Fig. 1. $\widehat{T}_2$ for a line segment of 64 independent realisations each with 2^{12} distances. The dots show the estimates, and the lines are $\overline{\widehat{T}}_2(r)$ and $\overline{\widehat{T}}_2(r) \pm 1.96\,\mathrm{std}(\widehat{T}_2(r))$.

definite, if the standard deviation of $\widehat{T}_2(r)$ approaches its mean $\overline{\widehat{T}}_2(r)$, the distribution cannot be symmetric about $\overline{\widehat{T}}_2(r)$ (as negative values are not possible) hence it will be skewed and thus not Gaussian. We will restrict our attention to those length scales large enough that the distribution of $T_2(r)$ is arguably Gaussian. In general, the variance will increase with decreasing r, as shown in the figure.

The departure from Gaussianity observed at small length scales indicates that the Takens' estimator is not, in fact, BLUE (Best Linear Unbiased Estimator). This has been noted and corrected for by Judd [6]. Our approach generalises immediately to the Judd estimator which is BLUE to within the limits stated in [6]. Our approach can also be generalised to make estimation of other methods coherent, the point of this Letter is to illustrate a coherent Takens' estimator.

In short, our approach aims to find the largest length scale (i.e., the one with smallest variance) where the sample is arguably *internally consistent* with the results at all smaller length scales. We then report the length scale (where the diameter of the set is normalised to one) and the standard error of $\overline{\widehat{T}}_2$ at this length scale. It is important to note that this does not guarantee the true dimension lies within our reported standard error, only that our estimate is coherent given all the available data.

4. Internally consistent estimators

Our goal is to find the largest value of r, $r^\star$, at which the estimator $T_2(r)$ is *internally consistent* with all estimates at smaller r_i. We call this property coherence. There are many ways to define internally *consistent*. Here we will say two estimates are consistent if they are *statistically not distinguishable*. This can be assessed by comparing the sample mean of their estimates. In this case, the 2 samples to be compared are constituted by independent estimators (each one corresponding to an independent realization) at different length scales. Thus, we test the null hypothesis,

$$H_0: \quad T_2(r_i) = T_2(r_j) \tag{5}$$

against the two-sided alternative hypothesis. This can be done with a t-student test. If one cannot reject H_0 at a certain significance level, we will say that $T_2(r_i)$ and $T_2(r_j)$ are consistent. If the estimators at the two smallest length scales available are consistent, one can repeat the same process making all pairwise tests for all length scales until one rejects the hypothesis for any two length scales. In this manner one can find the maximum length scale $r^\star$ for which all the estimators at smaller ones are consistent.

It is important to note that all pairwise tests must be computed between all length scales and not just the corresponding to consecutive length scales. This is due to the fact that, if $T_2(r_i)$ is consistent with $T_2(r_{i+1})$ and $T_2(r_{i+1})$ is consistent with $T_2(r_{i+2})$ (with $r_i < r_{i+1} < r_{i+2}$); this does not imply that $T_2(r_i)$ is consistent with $T_2(r_{i+2})$. For consecutive pairwise tests, failure to reject the null hypothesis (Eq. (5)) will become more likely as $r_{i+1}/r_i \to 1$. In practice, simply decreasing the step size in $\log_2(r)$ (i.e., r_{i+1}/r_i) should not be able to affect a decision of coherence.

5. Examples

5.1. Line segment

Fig. 2 shows the results for a line segment in greater detail. The arrow indicates the coherent estimator,

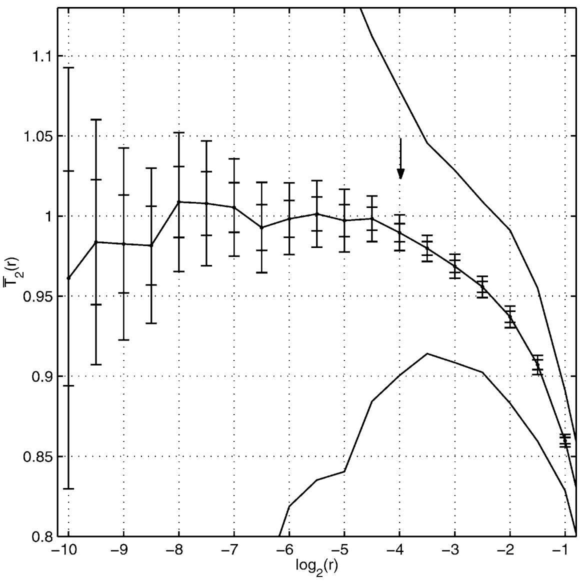

Fig. 2. Detail of Fig. 1 showing $\overline{\overline{\widehat{T}}}_2(r)$ for a line segment with error bars $\pm\,\mathrm{std}(\overline{\overline{\widehat{T}}}_2(r))$ and $\pm 1.96\,\mathrm{std}(\overline{\overline{\widehat{T}}}_2(r))$.

$\overline{\overline{\widehat{T}}}_2(2^{-4}) = 0.990 \pm 0.011$ found using t-tests at a significance level of $\alpha = 0.05$. The error bars we quote are the $1.96\times$ standard errors of $\overline{\overline{\widehat{T}}}_2(r)$ which correspond to a 95% confidence level. Note that at this significance level and with this sample size these do cover the true value of 1.00 (for $r^\star$), but the 68% error bars fail to cover it. Also note that visual inspection of Fig. 2 yields immediate identification of where the result tends to drift away from d_2. Ad hoc tests (such as the overlap of the 95% level pooled estimator) may be developed in preference to the t-test.

5.2. Cat map

Next we provide an example where the estimation of d_2 is hampered by macroscopic structure which cuts off at a (known) finite length scale. Following [11], we consider the set of points corresponding to delay reconstruction of the cat map in 3D. This set consists of a series of 11 sheets. When viewed within the Takens' estimator, this yields variations in $\widehat{T}_2$ which reach values of 2.4 at $r = 2^{-2.25}$ (see Fig. 3). At length scales smaller than the separation between the closest sheets, the estimator converges towards an estimate of

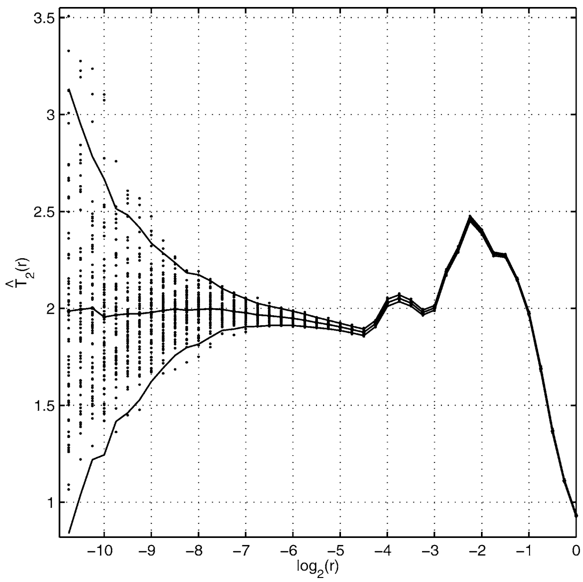

Fig. 3. $\widehat{T}_2$ for the 3D cat map of 64 independent realisations each with 2^{24} distances. The dots show the estimates, and the lines are $\overline{\overline{\widehat{T}}}_2(r)$ and $\overline{\overline{\widehat{T}}}_2(r) \pm 1.96\,\mathrm{std}(\widehat{T}_2(r))$.

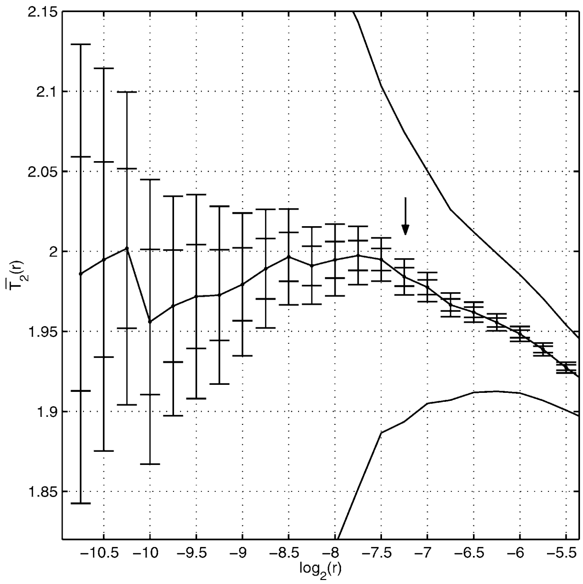

Fig. 4. Detail of Fig. 3 showing $\overline{\overline{\widehat{T}}}_2(r)$ for the cat map with error bars $\pm\,\mathrm{std}(\overline{\overline{\widehat{T}}}_2(r))$ and $\pm 1.96\,\mathrm{std}(\overline{\overline{\widehat{T}}}_2(r))$.

2.0, suffering only from the usual edge effects. Fig. 4 shows the estimates using 2^{24} inter-point distances. The t-tests yield $r^\star = -7.25$ and $\overline{\overline{\widehat{T}}}_2(r^\star) = 1.984 \pm 0.011$.

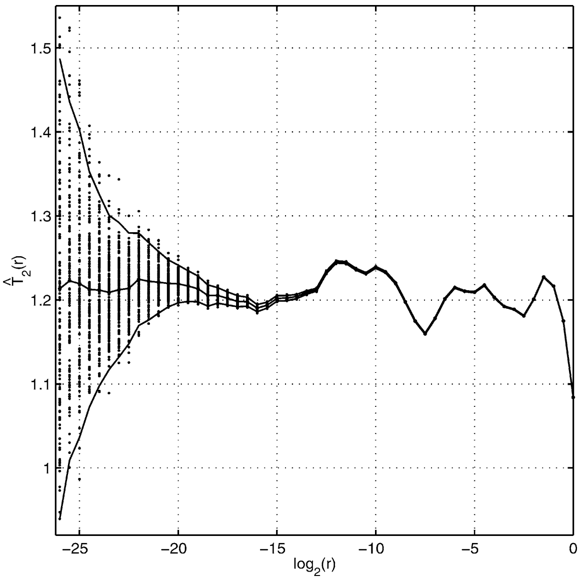

Fig. 5. $\widehat{T}_2$ for the 2D Hénon attractor of 64 independent realisations each with 2^{38} distances. The dots show the estimates, and the lines are $\overline{\widehat{T}}_2(r)$ and $\overline{\widehat{T}}_2(r) \pm 1.96\,\mathrm{std}(\widehat{T}_2(r))$.

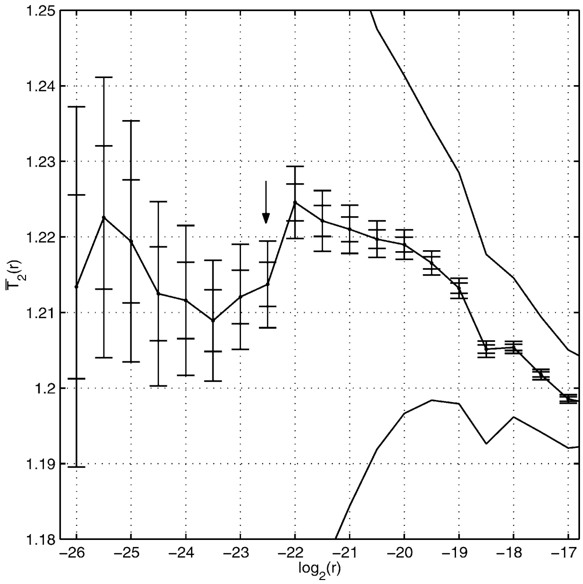

Fig. 6. Detail of Fig. 5 showing $\overline{\widehat{T}}_2(r)$ for the 2D Hénon attractor with error bars $\pm\,\mathrm{std}(\overline{\widehat{T}}_2(r))$ and $\pm 1.96\,\mathrm{std}(\overline{\widehat{T}}_2(r))$.

5.3. Hénon attractor

Figs. 5 and 6 illustrate our method applied to points taken from the Hénon attractor. Fig. 5 reveals the expected large scale structure, the precise estimates of $T_2(r)$ for $r > 2^{-10}$ are clearly *not* coherent. Applying the algorithm yields an estimate of 1.214 ± 0.006 which is coherent for $r \leqslant -22.5$. This is best seen in the zoom (Fig. 6). Note that while our result at $r^\star = -22.5$ is coherent, the rapid growth in the standard error at, say, $r = 2^{-22}$ does not rule out lacunarity oscillations of the same magnitude as observed at $r = 2^{-12}$. We stress that coherence is but a necessary condition for meaningful dimension estimates; it is not a sufficient condition. While this estimate differs from that quoted by Grassberger et al. [3], it is, in fact, consistent with the uncertainty in the estimate determined by constructing an ensemble of estimates using the method in [4]. Details of both calculations will be presented elsewhere.

Unlike traditional methods, the Takens' estimator requires all interpoint distances to be independent. In order to choose independent points in the attractor we used probabilistic sampling of a data stream. Thus, instead of a fixed set of points, a 'stream' of points was sampled (pseudo) randomly; pairs were then drawn from this sample, their interpoint distance measured and then the points were discarded (for further discussion see [12]). This stream of points is obtained by following a trajectory that has reached a stationary state. While the number of points needed using this approach is greater than those required by a traditional approach, where all interpoint distances between a fixed set of points are used, the information content is also greater.

In order to make statistical comparisons between the estimates at different length scales, the method of coherent estimation requires a sample of estimates at each length scale. In the examples above, we use a sample with a least 32 members, thus requiring at least $32N_p$ interpoint distances. On the other hand, traditional methods such as the Grassberger–Procaccia algorithm [3] only require one sample of interpoint distances to estimate d_2. This drawback is compensated by the fact that coherent estimation provides statistical error bars which allow us to have a meaningful measure of the uncertainty of the estimate of d_2.

Fitting a slope to a $\log C_2(r)$ vs. $\log(r)$ plot, as in the Grassberger–Procaccia algorithm [3], requires defining, either implicitly or explicitly, a 'scaling range' [7,11]; the Takens' estimator avoids this by first considering each length scale on its own merits. It is important, however, to first verify that the data follows a scaling law and that $\overline{\overline{T}}_2(r)$ converges as $r \to 0$. Internally consistent estimation allows us to determine whether $\overline{\overline{T}}_2(r)$ converges by assessing the equality of estimates at different length scales.

6. Restrictions and necessary conditions

The crucial issue in a test of coherence is the consistency of the reported estimate with all smaller length scales. For simplicity, we have adopted a t-test, but other alternatives exist (such as the ANOVA). Without doubt, custom tests can be devised for estimating different statistics. As in most statistical tests, there is a balance between theoretical exactness and ease of use. The key argument of this Letter is to test the consistency of the reported result with that at all smaller scales, not the particular test used to do so.

Some important assumptions have been made when using the t-test; namely independent and Gaussianly distributed observations and equality in variances. The Gaussianity assumption is met in this context (see Section 3). The observations within a sample are indeed independent (by construction), but this is not the case for observations between samples, as the estimators come (in this case) from the same realisations. Moreover, the assumption of equality of variances is not valid and it is, in fact, an important issue that the distributions at different length scales may have different variances. It is often argued [5] that, relatively large differences in the population variances have relatively small consequences for the conclusions derived from a t-test. In addition, in the case of similar sized length scales, their variances will not be in general very different. Therefore, we can put more trust in the validity of the test in these cases; which is often where the null hypothesis Eq. (5) is rejected.

Making multiple comparisons using the significance level α in all pairwise comparisons does not guarantee that the significance level of the overall test is also α. One possibility is to use the Bonferrioni method [5] which sets the level of each test at α/J (J is the number of samples), or a modification of this method given by Simes [9].

An alternative to carrying out all the $\binom{J}{2}$ different t-tests is to use an Analysis of Variance test (ANOVA). This test compares the means of multiple samples. Like the t-test, it also assumes that the sample variances are equal. In this context the variance increases as $r \to 0$ so, when using an ANOVA test, distributions with a whole range of variances will be considered. On the other hand, for pairwise t-tests just two different variances are used. Therefore, in this context it may be more adequate to use pairwise comparisons. Nevertheless, we strongly recommend that both analysis are made and a decision is made based on both results. In fact, for the examples shown here the ANOVA and t-test are in rough agreement.

When randomly sampling directly from a distribution, the distances r_p are indeed independent. This is not the case for a time series; when sampling points from a trajectory there are several things that can cause the points x_i, x_j to be correlated. Points should be sampled so as they are not close in time. Otherwise, their spatial separation could only reflect their closeness in time. It is also advisable to use probabilistic uneven sampling as, even if x_i and x_j are long separated in time, the distance $|x_i - x_j|$ is often correlated to $|x_{i+1} - x_{j+1}|$ (see [12]).

The Takens' estimator relies on $T_2(r)$ at small length scales to estimate the correlation dimension. This can be problematic in the case where noise is present since $T_2(r)$ is strongly influenced at length scales smaller than the noise level. One possible way to overcome this problem is by rescaling the interpoint distances using l_n, the maximum possible noise magnitude [13]. In nearly all practical situations, however, this quantity is unknown or not known to sufficient accuracy. In principle, it is possible to develop a maximum likelihood estimators of d_2 and l_n. It has been found by Schouten et al., however, that the numerical calculation of these estimators is not straightforward. See [13] and references within for different methods that deal with noise corrupted data.

The problem of the influence of noise when estimating d_2 using the Takens' estimator can be overcome, to a certain extent, given the distributional properties of $T_2(r)$. In the case where the interpoint distances are corrupted by noise, the Takens' estimator will effectively overestimate the true correlation dimension of the uncorrupted attractor. This effect will be stronger at the smallest length scales. Unfortunately, these are the length scales we are most interested on, as they better approximate the true dimension of the attractor. As mentioned in previous sections, in the absence of noise, the Takens' estimator is known to have a Gaussian distribution. When noise is present, due to the fact that $T_2(r)$ in overestimated, the distribution will no longer be Gaussian, as it becomes skewed by large positive values and negative values are not possible ($T_2(r)$ is positive definite). Thus, the lack of Gaussianity at these length scales, can be used as an indicator of the presence of noise and these estimates should not be taken on account when obtaining a coherent estimate of d_2.

The departure from Gaussianity was already encountered (Section 3) at the smallest length scales, due to the small number of interpoint distances. As in this case, it is advised that coherent estimation is restricted to those length scales large enough that the distribution of $T_2(r)$ is Gaussian. Finally, even though the strong influence of noise is an important drawback of the Takens' estimator, coherent estimation does not allow us to have false confidence in the estimate of d_2, by providing broad error bands that reflect the presence of noise.

7. Conclusions

We have introduced a method for dimension estimation that yields coherent statistics. This approach is of interest whenever the underlying statistic is defined only in the limit, as it provides an estimate which is consistent with measures at all smaller length scales. This is a significant improvement in understanding estimates of correlation dimension, as it provides constraints on how accurate a given estimate is likely to be. As with all cases where the limit cannot be taken, these estimates are lower bounds on the remaining uncertainty. It is the fact that they are sometimes viewed as 'surprisingly large' that is their strength.

Finally, we note that much of the interest in dimension estimation arises in the context of time series observations. In this Letter, we have restricted attention to well-defined sets of points; in the time series context, one is often interested in a variety of different sets which may be closely related (e.g., the original attractor, and delay reconstructions in various embedding dimensions). It should be clear that vast data sets may be required to extract insight from any dimension algorithm; an advantage of the coherence approach is that it is easily applied to data streams (rather than fixed size data sets). By shifting the analysis to a data stream and continuing the analysis until coherent results are achieved with small uncertainties, we allow the dynamical system to reveal the data requirements, thereby avoiding ad hoc generalisations about scaling.

Acknowledgements

The authors would like to thank the anonymous reviewers for their suggestions. This work was supported by a CONACYT grant and ONR DRI grant N00014-99-1-0056.

References

[1] A. Arneodo, G. Grasseau, E.J. Kostelich, Phys. Lett. A 124 (1987) 426.
[2] S. Borovkova, R. Burton, H. Dehling, Ann. Appl. Probab. 9 (1999) 376.
[3] P. Grassberger, I. Procaccia, Phys. Rev. Lett. 50 (1983) 346.
[4] P. Grassberger, Phys. Lett. A 128 (1988) 369.
[5] W.L. Hays, Statistics, Holt, Rinehart & Winston, New York, 1988.
[6] K. Judd, Physica D 56 (1992) 216.
[7] K. Judd, Physica D 71 (1994) 421.
[8] A. Rényi, Acta Math. Hungarica 6 (1953) 285.
[9] J.R. Simes, Biometrika 73 (1986) 751.
[10] L.A. Smith, J.D. Fournier, E.A. Spiegel, Phys. Lett. A 114 (1986) 465.
[11] L.A. Smith, Phys. Lett. A 133 (1988) 283.
[12] L.A. Smith, in: Proc. Int. School of Physics "Enrico Fermi" CXXXIII, Societá Italiana di Fisica, 1997, p. 177.
[13] J.C. Schouten, F. Takens, C.M. van den Bleek, Phys. Rev. E 50 (1994) 1851.
[14] F. Takens, in: Dynamical Systems and Bifurcations, in: Lecture Notes in Mathematics, Vol. 1125, Springer, Berlin, 1985, p. 99.
[15] J. Theiler, Phys. Lett. A 133 (1988) 195.
[16] J. Theiler, Int. J. Bifur. Chaos 3 (1993) 765.