

Finding the Embedding Dimension and Variable Dependencies in Time Series

Hong Pi

Carsten Peterson

*Department of Theoretical Physics, University of Lund,
Sölvegatan 14A, S-223 62 Lund, Sweden*

We present a general method, the δ -test, which establishes functional dependencies given a sequence of measurements. The approach is based on calculating conditional probabilities from vector component distances. Imposing the requirement of continuity of the underlying function, the obtained values of the conditional probabilities carry information on the embedding dimension and variable dependencies. The power of the method is illustrated on synthetic time-series with different time-lag dependencies and noise levels and on the sunspot data. The virtue of the method for preprocessing data in the context of feedforward neural networks is demonstrated. Also, its applicability for tracking residual errors in output units is stressed.

1 Introduction

The behavior of a dynamic system is often modeled by analyzing a time series record of certain system variables. Using artificial neural networks (ANN) to model such systems has recently attracted much attention. The success of such models relies heavily upon identifying the underlying structure in the time series—it is advantageous to know in advance the embedding dimension, most relevant inputs, noise level, etc. In this paper we devise a simple and easy-to-use method based on continuity requirements on statistical measures for identifying such essential properties in a time series record. Even though the language is that of time series the approach applies to any continuous function mapping problem.

Time series can have a wide range of behavior ranging from being entirely random and uncorrelated to being completely deterministic. In reality one is often in between these two extremes. Existing approaches to determine dependencies are either based on entropy measures (Kolmogorov 1959; Farmer 1982), or on elaborate autocorrelation measures (Russell *et al.* 1980, Grassberger and Procaccia 1983; Brock *et al.* 1988; Savit and Green 1991). Our approach, which has its roots in the lat-

ter philosophy, aims at determining the embedding dimension, pinning down sensitivity on the various variables and establishing noise levels.

Brock *et al.* (1988) have devised a method (the BDS test) to test the null hypothesis of whether or not a sequence of numbers is IID (independently and identically distributed random numbers). This test was further developed by Savit and Green (1991) into a conditional probability approach in which the degree of variable dependence may be quantified. Although the latter method has its merit in brevity, the message it gives is not without ambiguities.

Inspired by the work of Savit and Green (1991) we propose a method (the δ -test) from a different viewpoint, which exploits the definition of function continuity. This definition, which can be easily connected to the behavior of the conditional probabilities, gives clear signatures (apart from ambiguities arising from insufficient statistics) with respect to variable dependencies and the embedding dimensionality. To our knowledge the proposed method does not exist in the literature despite its conceptual simplicity.

The existing approaches mentioned above in general aim at establishing some invariant fractal dimensional measures. The numerical implementations usually involve some box counting algorithms. Meaningful interpretations of the outcome from these algorithms as the box size is reduced rely heavily on a scaling assumption. The estimate of embedding dimension can be viewed as a byproduct from estimating fractal dimensions using these algorithms. In this paper we do not attempt to establish yet another invariant nonlinearity measure. Rather our aim is to pick out variable dependencies and identify the minimum embedding dimension directly from the data. By exploiting the properties of continuous functions we need no scaling assumption. Whereas the traditional line of approach leads naturally to the BDS statistic that tests against the null hypothesis of an IID sequence, the δ -test tests the hypotheses at both the extremes: IID or a deterministic map. In a large class of models, in particular the neural network models for time series prediction and system identification problems, the existence of a function mapping is inherently assumed. The δ -test provides good measures on the truthfulness of such assumptions and gives an estimate on how successful these models can be in reproducing the sequence of data. This is the strength the traditional approaches are lacking.

Successful explorations are made on different maps with and without noise and with a variety of time-lag dependencies. Also, the underlying dynamics of a sunspot series is studied.

The relevance of the method for feedforward network training is illustrated with the sunspot series, where it is shown that feeding the network with the established minimum embedding dimension vectors gives rise to state-of-the-art generalization performance. Also, the method can be used to track residual dependencies of the output errors in a Multilayer Perceptron (MLP).

2 General Formulation

Consider a discrete-time system, manifested as a time series x_t , $t = 1, 2, 3, \dots, N$, where we are interested in knowing if there exists a continuous map relating future values to the past ones (i.e., if it is possible to identify a state equation)¹

$$x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-d}) + r_t \quad (2.1)$$

The "noise"-term r_t represents an indeterminable part that originates either from insufficient dimension of the measurements or from real noise. In general r_t should decrease with d . If the system is completely deterministic, r_t should vanish entirely as d exceeds the minimum embedding dimension d_{\min} .

For a map given as a sequence of measurements, one wants to know (1) the minimum embedding dimension d_{\min} , (2) the sensitivity of x_t with respect to each of the dependent variables, and (3) an estimate of the size of the noise. By "variable dependence" we mean the primary dependence, not the induced ones. If $x_t = f(x_{t-1}) = f(f(x_{t-2}))$ we say that x_{t-1} is the primary dependent variable, the induced dependence on x_{t-2} is of no interest. Variables with no primary dependence are denoted "irrelevant."

We approach the problem by constructing conditional probabilities in embedding spaces of various dimensions d . The time series in equation 2.1 is represented as a series of N points $\mathbf{z}(i)$ in a $(d+1)$ -dimensional space ($d = 0, 1, 2, \dots$)

$$\mathbf{z}(i) = (z_0(i), z_1(i), \dots, z_k(i), \dots, z_d(i)) \quad (2.2)$$

where $z_k(t) = x_{t-k}$. The distances between the k th components of two vectors $\mathbf{z}(i)$ and $\mathbf{z}(j)$ are defined as

$$l_k(i, j) = |z_k(i) - z_k(j)|, \quad k = 0, 1, \dots, d \quad (2.3)$$

Given a set of positive numbers, ϵ and $\delta = (\delta_1, \dots, \delta_d)$, one can construct the following joint probabilities from the data

$$P(l_0 \leq \epsilon, \mathbf{l} \leq \delta) = \frac{1}{N_{\text{pair}}} n(l_0 \leq \epsilon, \mathbf{l} \leq \delta) \quad (2.4)$$

$$P(\mathbf{l} \leq \delta) = \frac{1}{N_{\text{pair}}} n(\mathbf{l} \leq \delta) \quad (2.5)$$

where N_{pair} is total number of vector pairs, and $n(l_0 \leq \epsilon, \mathbf{l} \leq \delta)$ and $n(\mathbf{l} \leq \delta)$ are the number of the pairs satisfying the corresponding distance constraints. Throughout this paper we freely use the notation $\mathbf{l} \leq \delta$ for $\{(l_1 \leq \delta_1), (l_2 \leq \delta_2), \dots, (l_d \leq \delta_d)\}$. Also we set $\delta_i = \delta$ for all i .

¹The series is assumed to be bounded and stationary and x_t can take either real or complex values.

Next we form the conditional probabilities

$$P_d(\epsilon | \delta) \equiv P(l_0 \leq \epsilon | 1 \leq \delta) = \frac{P(l_0 \leq \epsilon, 1 \leq \delta)}{P(1 \leq \delta)} \tag{2.6}$$

How is $P(l_0 \leq \epsilon | 1 \leq \delta)$ expected to vary under different conditions? The following important observations can be made:

1. For a completely random time series one has

$$P_0(\epsilon) = P_1(\epsilon | \delta) = \dots = P_d(\epsilon | \delta) = \dots \tag{2.7}$$

This identity, which should be understood in a statistical sense, holds for any choice of positive ϵ and δ .

2. If a continuous map exists as in equation 2.1 with no intrinsic noise, then for any $\epsilon > 0$ there exists a δ_ϵ such that

$$P_d(\epsilon | \delta) = 1 \text{ for } \delta \leq \delta_\epsilon \text{ and } d \geq d_0 \tag{2.8}$$

The smallest integer d_0 for which equation 2.8 holds is identified with $(d_{\min} - 1)$.

3. In the presence of noise r , $P_d(\epsilon | \delta)$ will no longer saturate to 1 as ϵ becomes smaller than the width Δr_{\max} of the noise.

Equation 2.8 is a direct consequence of the definition of function continuity, which states that if $z_0 = f(z_1, \dots, z_d)$, then for any $\epsilon > 0$ there exists a $\delta > 0$ such that the conditions $|z_1 - z'_1| < \delta, \dots, |z_d - z'_d| < \delta$ guarantee $|z_0 - z'_0| < \epsilon$. With the presence of noise, however, $|z_0 - z'_0| = |f(z_1, \dots) - f(z'_1, \dots) + r - r'|$, which as $\delta \rightarrow 0$ cannot be made smaller than $\Delta r = |r - r'|$. This justifies the statement number 3. If we assume a flat noise distribution extending from $-r$ to r with standard deviation $\sigma_r = r/\sqrt{3}$, we get $\Delta r_{\max} = 2r = 2\sqrt{3}\sigma_r$, which gives an upper limit estimate on σ_r knowing Δr_{\max} .

How does $P_d(\epsilon | \delta)$ vary as a function of δ for fixed ϵ ? For $\delta \rightarrow \infty$ the conditions have no effect. Hence one has $P_d(\epsilon | \delta)|_{\delta \rightarrow \infty} = P_0(\epsilon)$. As $\delta \rightarrow 0$, $P_d(\epsilon | \delta)$ should increase monotonically and saturate to 1 for $d \geq d_0$. This behavior is shown schematically in Figure 1a. The approach of Savit and Green (1991) was based on the identity $P_d(\epsilon | \delta)_{\delta=\epsilon} = P_{d-1}(\epsilon | \delta)_{\delta=\epsilon}$, and establishes variable dependence when this identity is violated. However, ambiguities arise whenever irrelevant variables induce sizable changes in $P_d - P_{d-1}$ at $\delta = \epsilon$. This effect often occurs due to nonuniform curvatures of trajectories. For this reason we examine the maxima

$$P_d(\epsilon) = \max_{\delta > 0} P_d(\epsilon | \delta) = P_d(\epsilon | \delta)|_{\delta \leq \delta_\epsilon} \tag{2.9}$$

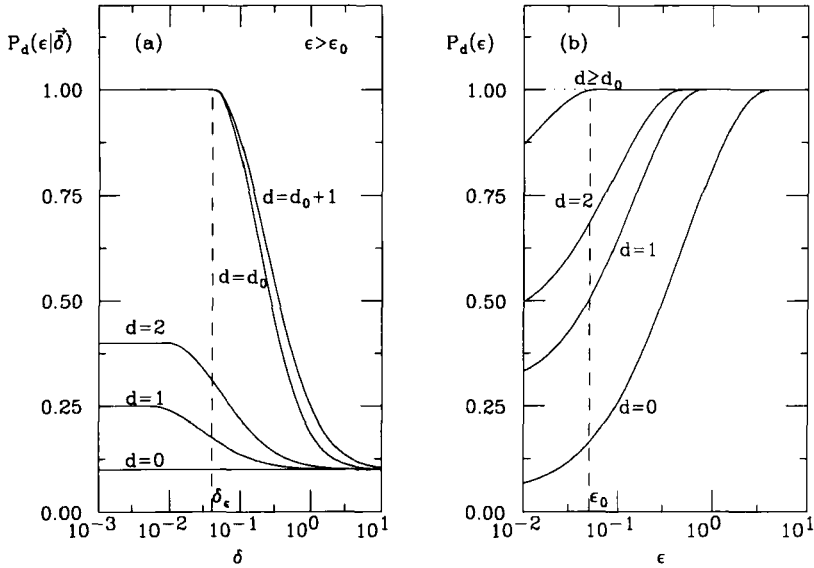


Figure 1: (a) $P_d(\epsilon | \delta)$ as a function of δ for fixed ϵ . (b) The maxima $P_d(\epsilon)$ as a function of ϵ . Saturation to 1 would be observed for $d \geq d_0$. In the presence of noise the saturation deviates from 1 around $\epsilon_0 \sim \Delta r_{\max}$.

Saturation of the maxima as d increases singles out the irrelevant variables. How the maxima $P_d(\epsilon)$ change with d and ϵ provides basically all the information we need (see Fig. 1b).

$P_d(\epsilon)$ measures how well the dynamics can be modeled in terms of the d variables. To quantify the dependence on each of the variables, it is convenient to define a *dependability index*²

$$\lambda_d(\epsilon) = \frac{P_d(\epsilon) - P_{d-1}(\epsilon)}{1 - P_0(\epsilon)}, \quad d = 1, 2, \dots \tag{2.10}$$

and its average over ϵ

$$\bar{\lambda}_d = \frac{\int_0^\infty d\epsilon \lambda_d(\epsilon)(1 - P_0(\epsilon))}{\int_0^\infty d\epsilon (1 - P_0(\epsilon))} = \frac{\int_0^\infty d\epsilon (P_d(\epsilon) - P_{d-1}(\epsilon))}{\int_0^\infty d\epsilon (1 - P_0(\epsilon))} \tag{2.11}$$

For a noise-free deterministic map, $P_d(\epsilon)$ saturates to 1 for $d \geq d_0$ and one has

$$\sum_{d=1}^{d_0} \bar{\lambda}_d = \sum_{d=1}^{d_0} \lambda_d(\epsilon) = 1 \tag{2.12}$$

²This is similar to the index defined by Savit and Green (1991) but with different definition of $P_d(\epsilon)$ and normalization.

A variable is considered irrelevant if its inclusion in the condition does not raise the conditional probability to a higher plateau ($\bar{\lambda}_d \approx 0$). In the approach of Savit and Green (1991) negative indices of statistical significance complicate the issue of identifying dependent variables. In our case it can be shown that $\lambda_{d=1} \geq 0$, $\lambda_{d_0} > 0$, and $\lambda_{d>d_0} = 0$. Thus negative indices for $d > d_0$ can arise only from statistical fluctuations. For $1 < d < d_0$ we expect that negative λ_d to be largely due to limited statistics in the maximization procedures. Anything beyond statistical problems can easily be clarified by inspecting whether the saturation to 1 is affected by treating the variable irrelevant.

So far the formalism has assumed an infinite amount of data. With limited statistics very low δ -values or large d s may give rise to a picture not as crisp as the one in Figure 1. To estimate the errors we use the standard estimator

$$\Delta P_d(\epsilon | \delta) = 2 \sqrt{\frac{P_d(1 - P_d)}{n(1 \leq \delta)}} \tag{2.13}$$

This error expression is not entirely adequate when correlations exist in time series. It nevertheless serves the purpose to signal whenever a statistically unreliable region is crossed into.

Theoretically $P_d(\epsilon | \delta)$ is expected to be a smooth and monotonically decreasing function of δ . It is generally flat near $\delta = 0$ and has a plateau that extends to a finite δ_ϵ . Finding its approximate maximum is not difficult, provided that there is a reasonable amount of statistics to probe the plateau region. In cases with very limited statistics it is advantageous to set an irrelevant variable k *inactive*, which means that the condition $l_k \leq \delta$ is omitted when computing $P_d(\epsilon | \delta)$ for $d > k$. By removing the unnecessary restrictive conditions in this way, improved statistics are obtained without affecting the evaluation of P_d s in higher dimensions.

3 Implementation Issues

Next we give a systematic procedure following the above on how to read off the key properties of $P_d(\epsilon | \delta)$ given a data set. We propose the following scheme:

0. Compute $P_0(\epsilon)$.
1. Starting with $d = 1$ compute $P_d(\epsilon | \delta)$ and find the maxima $P_d(\epsilon)$.
2. Evaluate the dependability index $\bar{\lambda}_d$.
3. (Optional) Variable elimination: if the d th variable is irrelevant ($\bar{\lambda}_d \sim O(0)$), set it inactive.
4. Increment d by 1 and repeat steps 1–3 until $P_d(\epsilon)$ saturates.

5. Identify d_0 for which $P_d(\epsilon)$ begins to saturate, and also the point ϵ_0 for which $P_{d_0}(\epsilon)$ begins to deviate from 1. The minimum embedding dimension is then $d_0 + 1$, and the noise width is estimated as ϵ_0 .

If option 3 is used, the dependability index in equation 2.10 is modified according to $\lambda_d(\epsilon) = (P_d(\epsilon) - P_{d'}(\epsilon))/(1 - P_0(\epsilon))$, where $d' \leq d - 1$ is the nearest *active* variable and the summations in equation 2.12 are restricted to active variables only.

In what follows ϵ and δ are expressed in units of σ , the standard deviation of the data set. To obtain a map of $P_d(l_0 \leq \epsilon, 1 \leq \delta)$ one can go through the data set and recount the statistics every time ϵ and δ are changed, or one can discretize the ϵ - δ plane, record the statistics in each bin, and then sum up contents in all bins progressively such that $P(l_0 \leq \epsilon, 1 \leq \delta)$ are obtained for a grid of ϵ - δ values by going through the data only once. We adopted the latter approach for the sake of computing speed and discretized the $\ln \epsilon$ - $\ln \delta$ plane by 30×30 bins with $\ln 10^{-2} \leq \ln \epsilon \leq \ln 4$ and $\ln 10^{-4} \leq \ln \delta \leq \ln 4$. A lower cut $\epsilon_{\min} = 0.1$ is used for the ϵ -integrations in equation 2.11. These parameters are not critical to the outcome of the method. It is possible to design algorithms without them.

4 Applications

We apply the method to two synthetic time series and the sunspot data problems.

For the **Logistic map** we generate 4000 time-steps according to

$$x_t = \eta x_{t-1}(1 - x_{t-1}) + r_t \quad (4.1)$$

with $\eta = 4$ giving $\sigma = 0.35$. The iterative noise r_t consists of random numbers uniformly distributed in $(-r, r)$. Two data sets are generated with $r = 0$ (1) and $r = 0.28\sigma$ (2), respectively. In order to keep the series bounded r_t is constrained such that $x_t \in (0, 1)$.

In Figure 2a we show $P_d(\epsilon|\delta)$ versus δ at $\epsilon = 0.085$ for various d . We observe a saturation of $P_d(\epsilon|\delta)$ to 1 for $d \geq 1$ in case (1), as expected for a one-dimensional noise-free map. The conditional probabilities still saturate in case (2) but only to ~ 0.55 , which indicates the presence of noise. In Figure 2b the maximized conditional probability $P_d(\epsilon)$ is shown as a function of ϵ for various dimensions d . For the noise-free data $P_d(\epsilon)$ saturates to 1 for $d \geq 1$ in the entire ϵ range. For the noisy data $P_d(\epsilon)$ still saturates approximately for $d \geq 1$ but the saturation value starts to fall off from 1 somewhere between 0.3 and 0.7, implying a noise level $r = |\Delta_{\max}|/2$ between 0.15σ and 0.35σ , which is consistent with the generated noise level. Calculating the dependability indices gives $\bar{\lambda}_1 = 1$ for the $r = 0$ data, $\bar{\lambda}_1 = 0.97$ for the noisy data and $\bar{\lambda}_{d \geq 2} \approx 0$.

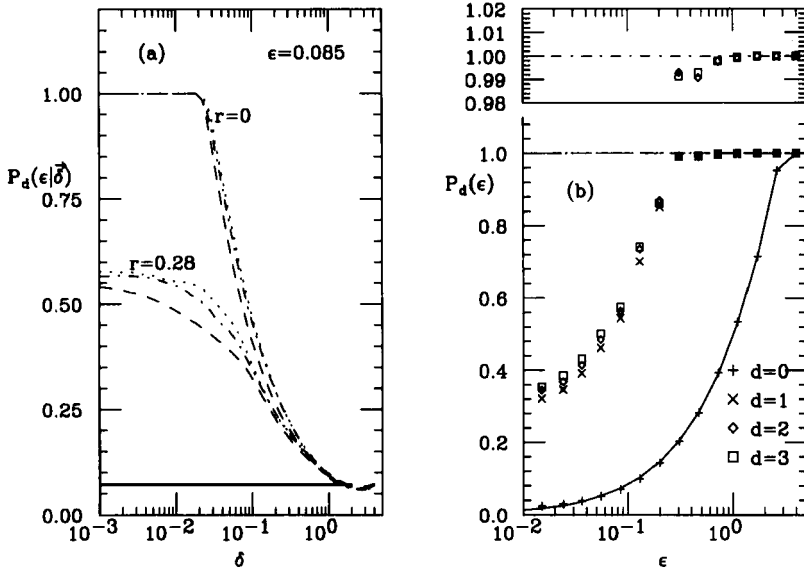


Figure 2: Logistic map: (a) $P_d(\epsilon | \delta)$ as a function of δ for noise-free ($r = 0$) and noisy data ($r = 0.28$). The curves represent $d = 0$ (solid), $d = 1$ (dashed), $d = 2$ (dash-dotted), and $d = 3$ (dotted), respectively. (b) $P_d(\epsilon)$ as a function of ϵ . The curves correspond to noise-free data (same notation as in a) and the points to the noisy data. The upper curve displays the fall-off of $P_d(\epsilon)$ from unity on an enlarged scale.

For the **Hénon map** we generate 4000 time-steps according to

$$x_t = 1 - a(x_{t-2} - r_{t-2})^2 + b(x_{t-4} - r_{t-4}) + r_t \tag{4.2}$$

with $a = 1.4$, and $b = 0.3$ giving $\sigma = 0.723$. This is the usual Hénon map with the dependencies stretched to larger lags and with the noise additively applied. Again we will use two data sets with $r = 0$ (1) and $r = 0.14\sigma$ (2). The variable elimination option is used here. We obtain $\bar{\lambda}_{1-4} = 0.002, 0.886, -0.023, 0.114$ for (1) and $0.052, 0.728, 0.004, 0.128$ for (2). The dependencies on x_{t-2} and x_{t-4} emerge as large values of $\bar{\lambda}_2$ and $\bar{\lambda}_4$. For (1) $\bar{\lambda}_2 + \bar{\lambda}_4 = 1$, indicating a noise-free map, whereas for (2) the value is 0.86 signaling the presence of noise.

The **sunspot data** (Priestley 1988) contain the annual averaged sunspot activities from 1700 to 1979 (280 points). This is a very limited statistics data set. The resolution limit, from which to extrapolate $\epsilon \rightarrow 0$ behavior, is set by $\epsilon \approx 0.95$. For that reason we explore two different input representations, linear and logarithmic—different sensitivities may give rise

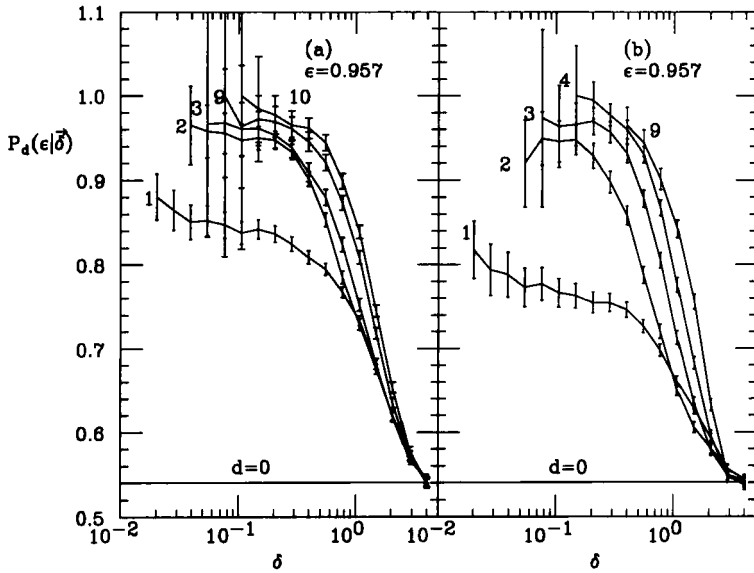


Figure 3: $P_d(\epsilon | \delta)$ for the sunspot data together with estimated errors shown as a function of δ for $\epsilon = 0.957$ and various dimensions d (marked on the curve); (a) is for the embedding (x_t, x_{t-1}, \dots) and (b) is for the embedding $(x_t, \ln(1+x_{t-1}), \dots)$. The δ s for which the plateau fail to rise are omitted.

to a more complete picture. Using the variable elimination option $P_d(\epsilon | \delta)$ are shown in Figure 3 for the two cases. From Figure 3a we see that the probability with conditional variables $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-9}$, and x_{t-10} approximately saturates to 1, whereas in the logarithmic representation (Fig. 3b) the variables $x_{t-1}, x_{t-2}, x_{t-3}$, and x_{t-4} show up to be relevant. Probing into smaller ϵ with more statistics would clarify the situation. Based on the results of Figure 3 we approximately determine the embedding dimension to be around 7 with $x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-9}$ and x_{t-10} as the most important variables, and we find the sunspot data are masked by large noises with amplitude on the order of 0.47σ .

5 Impact on Neural Network Learning

Feeding an MLP with redundant variables should be avoided since fitting to noise increases the difficulty of learning and may give rise to poor generalization. In Weigend *et al.* (1990) an MLP with a layer of 8 hidden nodes is trained with the sunspot series data. The authors experiment

with different number of time-lags as inputs, 6, 12, and 24, and conclude on the basis of generalization performance that 12 is optimal and hence that this number reflects the embedding dimension. We instead use the δ -test results above as a guide for the relevant input variables.

Following the procedures (without weight elimination) of Weigend *et al.* (1990) we have trained MLPs with 8 sigmoidal hidden units and 1 linear output unit and various number of input units. With x_{t-1} , x_{t-2} , x_{t-3} , x_{t-4} , x_{t-9} , x_{t-10} as inputs, we find for a 6-8-1 MLP the ratio of the mean square error to the variance³ (ARV) 0.073 on the test data from 1921 to 1955. This performance is as good as the one achieved by Nowlan and Hinton (1992) using more sophisticated algorithms. In Figure 4 we show the learning curves of the 6-input network together with the ones of a network using 12 lag variables as used by Weigend *et al.* (1990). It is also interesting to notice that after training the network with weight elimination, Weigend *et al.* (1990) find large connections to the hidden nodes from the inputs x_{t-1} , x_{t-2} , and x_{t-9} , consistent with the δ -test findings.

The δ -test is also very powerful when it comes to analyzing the residual error of output units in MLP learning. With perfect training the δ -test should identify the residual series as an independent random series. If this is not the case the test singles out the relevant input units from which information has not been fully extracted by the network.

6 Summary

We have devised a general method, the δ -test, for identifying dependencies in continuous functions. It is not limited to linear correlations and it determines the embedding dimensions, dependencies, and noise levels fairly accurately even in cases of low statistics. Automated procedures for setting bin sizes, cutoffs, etc. and error analysis are feasible.

Being based on conditional probabilities our approach at first sight appears very similar to that of Savit and Green (1991). However conceptually the two are rather distinct. The latter is based on Grassberger and Procaccia correlation integral using $\delta = \epsilon$. In contrast our method is based on the fundamental property of functional continuity ($\delta \rightarrow 0$). With the Savit and Green approach there are mainly two problems (Pi and Peterson 1993): (1) induced dependencies can show up in the index if nonuniform curvature is present in the function map; (2) there is not a saturation measure to indicate the critical embedding dimension and there is not an indication to quantify the noise level. We are able to extract information unambiguously from the region where $P_d(\epsilon|\delta)$ is maximized. There are also similarities between the Kolmogorov entropy method (Kolmogorov 1959) and ours in that both methods examine the behavior of a certain set of conditional probability distributions. However, evaluating

³The sunspot series has a variance $\sigma^2 = 1495$. We use $\sigma^2 = 1535$, a value quoted by Weigend *et al.* (1990), in order to compare our ARV with established results.

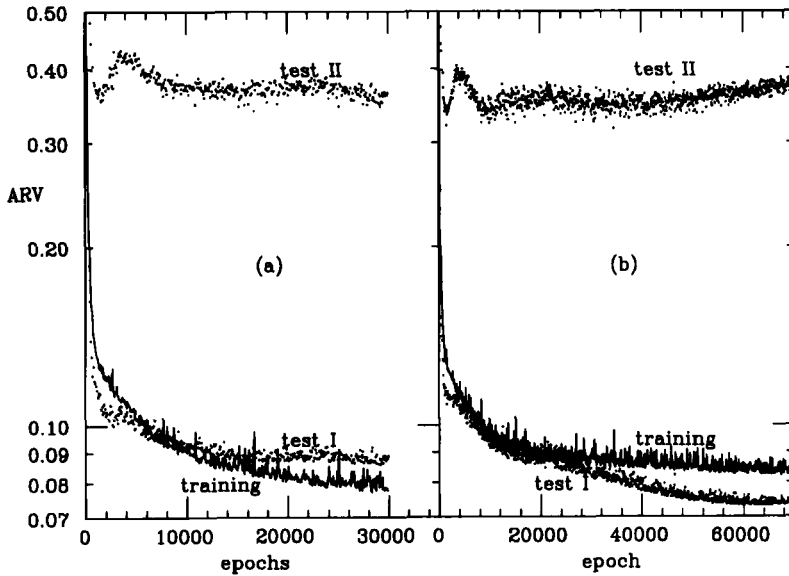


Figure 4: Learning curves for the sunspot data shown as ARV versus training epochs for a 12-8-1 network (a) and a 6-8-1 network (b). The solid lines are for the training set (1700–1920), the lower dotted lines are for the test set I (1921–1955) and the upper dotted lines are for the test set II (1956–1979). Large fluctuations on the curves have been filtered out.

entropy for a high-dimensional joint probability distribution requires a huge amount of statistics in contrast to the present method.

Acknowledgments

We are indebted to Richard Blankenbecler for bringing the work of Savit and Green (1991) to our attention.

References

- Brock, W. A., Dechert, W. D., Scheinkman, J. A., and LeBaron, B. 1988. A test for independence based on the correlation dimension. University of Wisconsin Preprint.
- Farmer, J. D. 1982. Information dimension and the probabilistic structure of chaos. *Z. Naturforschung* 37A, 1304–1325.

- Grassberger, P., and Procaccia, I. 1983. Measuring the strangeness of strange attractors. *Physica D* **9**, 189–208.
- Kolmogorov, A. N. 1959. *Dokl. Akad. Nauk SSSR* **124**, 754–755 [Entropy per unit time as a metric invariant of automorphisms. *Math. Rev.* **21**, 386 (1960)].
- Nowlan, S. J., and Hinton, G. 1992. Simplifying neural networks by soft weight-sharing. *Neural Comp.* **4**, 473–493.
- Pi, H., and Peterson, C. 1994. To be published.
- Priestley, M. B. 1988. *Non-linear and Non-stationary Time Series Analysis*, p. 223. Academic Press, New York.
- Russell, D. A., Hanson, J. D., and Ott, E. 1980. Dimension of strange attractors. *Phys. Rev. Lett.* **45**, 1175–1178.
- Savit, R., and Green, M. 1991. Time series and dependent variables. *Physica D* **50**, 95–116.
- Weigend, A., Huberman, B. A., and Rumelhart, D. 1990. Predicting the future: A connectionist approach. *Int. J. Neural Syst.* **1**, 193–209.

Received April 19, 1993; accepted September 13, 1993.

This article has been cited by:

1. Sayan Mukherjee, Sanjay Kumar Palit, Dilip Kumar Bhattacharya. 2014. Approximate discrete dynamics of EMG signal. *Applied Mathematics and Computation* **243**, 879-888. [[CrossRef](#)]
2. Sayan Mukherjee, Sanjay Kumar Palit, Santo Banerjee, M.R.K. Ariffin, D.K. Bhattacharya. 2014. Phase synchronization of instrumental music signals. *The European Physical Journal Special Topics* **223**, 1561-1577. [[CrossRef](#)]
3. Yanjun Gan, Qingyun Duan, Wei Gong, Charles Tong, Yunwei Sun, Wei Chu, Aizhong Ye, Chiyuan Miao, Zhenhua Di. 2014. A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling & Software* **51**, 269-285. [[CrossRef](#)]
4. Narges Salehnia, Mohammad Ali Falahi, Ahmad Seifi, Mohammad Hossein Mahdavi Adeli. 2013. Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis. *Journal of Natural Gas Science and Engineering* **14**, 238-249. [[CrossRef](#)]
5. Ricardo de A. Araújo, Tiago A.E. Ferreira. 2013. A Morphological-Rank-Linear evolutionary method for stock market prediction. *Information Sciences* **237**, 3-17. [[CrossRef](#)]
6. Ricardo A. Araújo. 2013. Evolutionary Learning Processes to Design the Dilation-Erosion Perceptron for Weather Forecasting. *Neural Processing Letters* **37:3**, 303-333. [[CrossRef](#)]
7. F. Liébana-Cabanillas, R. Nogueras, L.J. Herrera, A. Guillén. 2013. Analysing user trust in electronic banking using data mining methods. *Expert Systems with Applications* . [[CrossRef](#)]
8. Souhaib Ben Taieb, Gianluca Bontempi, Amir F. Atiya, Antti Sorjamaa. 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications* . [[CrossRef](#)]
9. Ricardo de A. Araújo. 2011. A morphological perceptron with gradient-based learning for Brazilian stock market forecasting. *Neural Networks* . [[CrossRef](#)]
10. Ricardo de A. Araújo. 2011. A robust automatic phase-adjustment method for financial forecasting. *Knowledge-Based Systems* . [[CrossRef](#)]
11. L. Uzal, G. Grinblat, P. Verdes. 2011. Optimal reconstruction of dynamical systems: A noise amplification approach. *Physical Review E* **84:1**. . [[CrossRef](#)]
12. Ricardo de A. Araújo. 2011. A class of hybrid morphological perceptrons with application in time series forecasting. *Knowledge-Based Systems* **24:4**, 513-529. [[CrossRef](#)]
13. Jan Viebig, Thorsten Poddig, Panagiotis Ballis-Papanastasiou. 2011. Regime-Dependent Nonlinear Analysis of Hedge Funds. *The Journal of Alternative Investments* **13:4**, 53-72. [[CrossRef](#)]

14. Ricardo de A. Araújo. 2011. Translation Invariant Morphological Time-lag Added Evolutionary Forecasting method for stock market prediction. *Expert Systems with Applications* **38**:3, 2835-2848. [[CrossRef](#)]
15. Ricardo de A. Araújo. 2010. Swarm-based translation-invariant morphological prediction method for financial time series forecasting. *Information Sciences* **180**:24, 4784-4805. [[CrossRef](#)]
16. Ricardo de A. Araújo. 2010. Hybrid intelligent methodology to design translation invariant morphological operators for Brazilian stock market prediction. *Neural Networks* **23**:10, 1238-1251. [[CrossRef](#)]
17. Zhihao Guo, Shaya Sheikh, Camelia Al-Najjar, Hyun Kim, Behnam Malakooti. 2010. Mobile ad hoc network proactive routing with delay prediction using neural network. *Wireless Networks* **16**:6, 1601-1620. [[CrossRef](#)]
18. Federico Montesino Pouzols, Amaury Lendasse, Angel Barriga Barros. 2010. Autoregressive time series prediction by means of fuzzy inference systems using nonparametric residual variance estimation. *Fuzzy Sets and Systems* **161**:4, 471-497. [[CrossRef](#)]
19. Ricardo de A. Araújo. 2010. A quantum-inspired evolutionary hybrid intelligent approach for stock market prediction. *International Journal of Intelligent Computing and Cybernetics* **3**:1, 24-54. [[CrossRef](#)]
20. Sunil Kukreja. 2009. Application of a least absolute shrinkage and selection operator to aeroelastic flight test data. *International Journal of Control* **82**:12, 2284-2292. [[CrossRef](#)]
21. M. Ciszak, F. Marino, A. Ortolan, T. Dal Canton. 2009. Identification of gravitational wave signals from chaotic astrophysical systems through phase space and attractor properties. *Physical Review D* **80**:4. . [[CrossRef](#)]
22. Manuel R. Arahal, Manuel Berenguel, Eduardo F. Camacho, Fernando Pavon. 2009. Selección de variables en la predicción de llamadas en un centro de atención telefónica. *Revista Iberoamericana de Automática e Informática Industrial RIAI* **6**:1, 94-104. [[CrossRef](#)]
23. Tiago A. E. Ferreira, Germano C. Vasconcelos, Paulo J. L. Adeodato. 2008. A New Intelligent System Methodology for Time Series Forecasting with Artificial Neural Networks. *Neural Processing Letters* **28**:2, 113-129. [[CrossRef](#)]
24. P. Verdes, P. Granitto, H. Ceccatto. 2006. Overembedding Method for Modeling Nonstationary Systems. *Physical Review Letters* **96**:11. . [[CrossRef](#)]
25. José Luis Rangel, Ursula Iturrarán-Viveros, A. Gustavo Ayala, Francisco Cervantes. 2005. Tunnel stability analysis during construction using a neuro-fuzzy system. *International Journal for Numerical and Analytical Methods in Geomechanics* **29**:15, 1433-1456. [[CrossRef](#)]
26. P. Verdes. 2005. Assessing causality from multivariate time series. *Physical Review E* **72**:2. . [[CrossRef](#)]

27. I LIND, L LJUNG. 2005. Regressor selection with the analysis of variance method#. *Automatica* **41**:4, 693-700. [[CrossRef](#)]
28. C. Goutte. 2000. Extraction of the relevant delays for temporal modeling. *IEEE Transactions on Signal Processing* **48**:6, 1787-1795. [[CrossRef](#)]
29. F Andersson. 2000. Algorithmic approaches for studies of variable influence, contribution and selection in neural networks. *Chemometrics and Intelligent Laboratory Systems* **51**:1, 61-72. [[CrossRef](#)]
30. Jiayu Lin, Zhiping Huang, Yueke Wang, Zhenken Shen. 2000. Selection of proper embedding dimension in phase space reconstruction of speech signals. *Journal of Electronics (China)* **17**:2, 161-169. [[CrossRef](#)]
31. Ya-Chen Hsu, Guanrong Chen Fuzzy Dynamical Modeling Techniques for Nonlinear Control Systems and Their Applications to Multiple-Input, Multiple-Output (Mimo) Systems 47-86. [[CrossRef](#)]
32. J Bomberger. 1998. Determination of model order for NARX models directly from input-output data. *Journal of Process Control* **8**:5-6, 459-468. [[CrossRef](#)]
33. S.A. Billings, S Chen The Determination of Multivariable Nonlinear Models for Dynamic Systems **7**, 231-278. [[CrossRef](#)]
34. H Pi. 1994. Delta 2.0 — A program for finding dependencies in tables of data. *Computer Physics Communications* **83**:2-3, 293-306. [[CrossRef](#)]