Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Physica D 229 (2007) 1–8
PHYSICA D
www.elsevier.com/locate/physd

# A note on the Gamma test analysis of noisy input/output data and noisy time series

Antonia J. Jones[a], D. Evans[a], S.E. Kemp[b,*]

[a] *School of Computer Science, Cardiff University, PO Box 916, Cardiff CF24 3XF, UK*
[b] *Department of Computing and Mathematical Sciences, Faculty of Advanced Technology, University of Glamorgan, Pontypridd, Wales, CF37 1DL, UK*

Received 9 November 2005; received in revised form 27 September 2006; accepted 18 December 2006
Available online 28 February 2007
Communicated by C.K.R.T. Jones

**Abstract**

In a smooth input/output process $y = f(\boldsymbol{x})$, if the input data $\boldsymbol{x} \in \mathbb{R}^d$ is noise free and only the output data $y$ is corrupted by noise, then a near optimal smooth model $\hat{g}$ will be a close approximation to $f$. However, as previously observed, for example in [H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, 2nd ed., Cambridge Univ. Press, 2004], if the input data is also corrupted by noise then this is no longer the case. With noise on the inputs, the best predictive smooth model based on noisy data need not be an approximation to the actual underlying process; rather, the best predictive model depends on both the underlying process *and* the noise. A corollary of this observation is that one cannot readily infer the nature of a process from noisy data. Since almost all data has associated noise this conclusion has some unsettling implications. In this note we show how these effects can be *quantified* using the Gamma test.

In particular we examine the Gamma test analysis of noisy time series data. We show that the noise level on the best predictive smooth model (based on the noisy data) can be much higher than the noise level on individual time series measurements, and we give an upper bound for the first in terms of the second.


*Keywords:* The Gamma test; Non-linear analysis; Process identification; Noise estimation; Time series

## 1. Introduction

In the analysis of time series, we often hypothesize that the variable of interest is just one of a number of variables of a complex dynamic system, described by a system of differential equations. Following the work of Takens [10] we seek to predict the next value (output) based on a number $d$ of previous values (input). In this context, the input is called a *delay vector* and $d$ is called the *embedding dimension*.

For a time series $(z_t)$, Takens' theorem [10] and its subsequent extensions ensure, under a broad range of circumstances, that there exists a *smooth* function $f$ with bounded partial derivatives such that

$$z_t = f(z_{t-1}, z_{t-2}, \dots z_{t-d}) \tag{1}$$

which, provided $d$ is sufficiently large to unfold the dynamics, can be used as the basis for a recursive one-step prediction.

By *smooth function*, we mean throughout that $f$ and its partial derivatives of first (and possibly higher) orders exist, are continuous over a compact region, and are therefore bounded. To be explicit we suppose $|\nabla f|^2 \leq B$ over the region in question.

### 1.1. Stochastic time series

We draw a distinction between the subject of our paper – what we have called *noisy* time series – and that of *stochastic* time series. For a univariate stochastic time series with additive noise (often assumed to be Gaussian), the process is defined according a recursive rule of the form

$$z_t = f(z_{t-1}, z_{t-2}, \dots z_{t-d}) + e_t \tag{2}$$

where $f$ is a smooth function and $e_t$ is a realization of some random variable (if $f$ is linear, these are called *linear* stochastic

* Corresponding address: The Hollies, Curtis Lane, Headley, GU35 8PH Bordon, Hampshire, UK. Fax: +44 0 7880740220.
*E-mail address:* kemp.samuel@googlemail.com (S.E. Kemp).

0167-2789/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2006.12.013

time series). The significant fact is that the noise $e_{t-1}$ associated with the *previous* value $z_{t-1}$ of the time series feeds through to affect the next value $z_t$, so this noise plays a role in determining the evolution of the time series.[1]

### 1.2. Noisy time series

We consider the case where a noise-free time series $(z_t)$ is observed under additive noise, i.e.

$$y_t = z_t + r_t \qquad (t = 1, 2, 3, \ldots, M) \tag{3}$$

where the *true* values $z_t$ are subject to independent and identically distributed random perturbations $r_t$ having expectation zero.

We assume that the noise-free value $z_t$ is determined by a smooth function $f : \mathbb{R}^d \to \mathbb{R}$ of some number $d$ of the previous noise-free values $z_{t-1}, \ldots, z_{t-d}$,

$$z_t = f(z_{t-1}, \ldots, z_{t-d}). \tag{4}$$

Thus we imagine that in reality, the variable $z_t$ (part of a high-dimensional non-linear dynamical process) is evolving according to an unknown but smooth rule such as (4), but that what we *actually* observe, typically as a consequence of measurement error, are the corrupted values $y_t = z_t + r_t$. Importantly, and in contrast with stochastic time series, the noise associated with previous values such as $y_{t-1}$ does *not* feed through to affect the value $y_t$.

Of course, in many real world situations a time series may be *both* stochastic and noisy. However, here we seek to examine just those features specifically relating to noisy time series.

In the context of non-linear dynamic systems time series state space reconstruction, the noise that we have considered is termed 'observational noise'. The question of optimal prediction for time series under observational noise is also considered as a special case in Casdagli et al. [1], which studies in considerable detail the more general issues surrounding state space reconstruction under noise.

### 1.3. Effective noise

For $d \in \mathbb{N}$, let $\boldsymbol{x}_{d+1}, \ldots, \boldsymbol{x}_M$ denote the noisy delay vectors:

$$\boldsymbol{x}_t = (y_{t-1}, \ldots, y_{t-d}) \in \mathbb{R}^d. \tag{5}$$

Using only the noisy time series data $(\boldsymbol{x}_t, y_t)$, we seek to identify a smooth function $g : \mathbb{R}^d \to \mathbb{R}$ that 'best explains' the observed behaviour of the time series. We first clarify what is meant by 'best explains', i.e. what is an optimal smooth model in this context.

Let $S = \{h : \mathbb{R}^d \to \mathbb{R} \mid h \text{ smooth}, |\nabla h|^2 \leq B\}$, i.e. $S$ is the class of smooth functions in the sense described earlier. For each $h$ consider the mean squared error

$$\mathrm{MSE}(h) = \mathcal{E}((y - h(\boldsymbol{x}))^2) \tag{6}$$

where the expectation is taken over all realizations of the input/output pair $(\boldsymbol{x}, y)$. The set of *optimal* predictive smooth data models is defined to be

$$S_{\mathrm{opt}} = \{g \in S : \mathrm{MSE}(g) \leq \mathrm{MSE}(h) \text{ for all } h \in S\}. \tag{7}$$

Let $g \in S_{\mathrm{opt}}$. We write

$$y = g(\boldsymbol{x}) + R \tag{8}$$

where $R$ is a zero-mean random variable, called the *effective noise* on the output, which accounts for all variation in the output that cannot be accounted for by *any* smooth transformation of the input.

Note that

$$\mathcal{E}(R^2) = \mathcal{E}((y - g(\boldsymbol{x}))^2) = \mathrm{MSE}(g) \tag{9}$$

so the variance of the effective noise coincides with the minimum achievable mean squared error by a smooth data model based on the given selection of inputs.

To best model the time series data, we need to identify a function $\hat{g} \in S$ which is as close as possible to an optimal data model $g \in S_{\mathrm{opt}}$. Such a model will have close to minimal $\mathcal{E}\left((y - \hat{g}(\boldsymbol{x}))^2\right)$ and will not change significantly as more and more data is used in the model construction, i.e. as $M \to \infty$. We describe such a model as 'asymptotically stable'.

By (6) and (8),

$$\mathrm{MSE}(\hat{g}) = \mathcal{E}(R^2) + \mathcal{E}((g(\boldsymbol{x}) - \hat{g}(\boldsymbol{x}))^2). \tag{10}$$

Once the model selection process has been completed, it is tempting to assume that $g = f$, and hence that $\hat{g}$ is an approximation to the original function $f$ that generated the noise-free data $z_t$. However, as observed in Kantz and Schreiber [6] and as we illustrate here, this is not necessarily the case. The main contribution of this note is to illustrate how these differences can be quantified using the Gamma test.

### 1.4. Model construction

In practice, given a noisy time series $(y_t)$, we seek to construct an asymptotically stable model $\hat{g}$ for which the *empirical* mean squared error, defined by

$$\mathrm{MSE}_{\mathrm{emp}}(\hat{g}) = \frac{1}{M-d} \sum_{t=d+1}^{M} (y_t - \hat{g}(\boldsymbol{x}_t))^2 \tag{11}$$

is as close as possible to $\mathcal{E}(R^2)$, the variance of the effective noise. By (10), this ensures that $\hat{g}$ is as close as possible to an optimal predictive data model $g \in S_{\mathrm{opt}}$ (in a mean squared sense).

Our tool for estimating $\mathrm{var}(R)$ is the *Gamma test* [5], and we show how the results of a Gamma test analysis should be interpreted first when applied to input/output data with noisy inputs, and second when applied to noisy time series data.

[1] This type of noise is often called 'dynamic noise' in the literature on dynamic systems, see for example Casdagli et al. [1].