

# Time series analysis in astronomy: Limits and potentialities

R. Vio<sup>1</sup>, N. R. Kristensen<sup>2</sup>, H. Madsen<sup>2</sup>, and W. Wamsteker<sup>3</sup>

<sup>1</sup> Chip Computers Consulting s.r.l., Viale Don L. Sturzo 82, S.Liberale di Marcon, 30020 Venice, Italy  
ESA-VILSPA, Apartado 50727, 28080 Madrid, Spain  
e-mail: robertovio@tin.it

<sup>2</sup> Department of Informatics and Mathematical Modelling, Technical University of Denmark, Richard Petersens Plads,  
2800 Kgs. Lyngby, Denmark  
e-mail: nrk@imm.dtu.dk; hm@imm.dtu.dk

<sup>3</sup> ESA-VILSPA, Apartado 50727, 28080 Madrid, Spain  
e-mail: willem.wamsteker@esa.int

Received 11 October 2004 / Accepted 19 January 2005

**Abstract.** In this paper we consider the problem of the limits concerning the physical information that can be extracted from the analysis of one or more time series (light curves) typical of astrophysical objects. On the basis of theoretical considerations and numerical simulations, we show that with no a priori physical model there are not many possibilities to obtain interpretable results. For this reason, the practice to develop more and more sophisticated statistical methods of time series analysis is not productive. Only techniques of data analysis developed in a specific physical context can be expected to provide useful results. The field of *stochastic dynamics* appears to be an interesting framework for such an approach. In particular, it is shown that modelling the experimental time series by means of the stochastic differential equations (SDE) represents a valuable tool of analysis. For example, besides a more direct connection between data analysis and theoretical models, in principle the use of SDE permits the analysis of a continuous signal independent of the characteristics (e.g., frequency, regularity, ...) of the sampling with which the experimental time series were obtained. In this respect, an efficient approach based on the extended Kalman filter technique is presented. Its performances and limits are discussed and tested through numerical experiments. Freely downloadable software is made available.

**Key words.** methods: data analysis – methods: statistical

## 1. Introduction

The study of the light curves of astrophysical objects has always been an important tool for astronomers. The reason is simple: an effective way to get insight on the structure of a given physical system is to study its evolution over time. Some examples are the reconstruction of the structure of binary star systems, the understanding of the nature of the pulsars and the determination of the sizes of the central regions of active galactic nuclei. However, in spite of these remarkable successes, in many other situations the analysis of light curves has not proved useful. This is because often the time evolution of a dynamical system, describable in terms of a set of generic physical quantities (*state-variables*)  $\mathbf{x}(t)$ <sup>1</sup>, is governed by an  $n$ -dimensional system of differential equations (*state-equation*) with the general form

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{s}(t), t], \quad (1)$$

<sup>1</sup> Hereafter, vector quantities will be denoted in boldface.

where  $t$  is the time coordinate, symbol “ $\dot{\cdot}$ ” means derivation with respect to  $t$ , and  $\mathbf{f}[\cdot]$  is a  $n$ -dimensional (possibly non-linear) function. The  $m$ -dimensional vector  $\mathbf{s}(t)$  represents independent processes whose time evolution does not depend on  $\mathbf{x}(t)$ , such as, for example, the processes that take place in regions external to the system of interest. In general, the quantities  $\mathbf{x}(t)$  are not directly observable and the experimental time series  $\{\mathbf{y}_{t_k}\}_{k=0}^N$  are obtained through the *measurement-equation*

$$\mathbf{y}_{t_k} = \mathbf{h}[\mathbf{x}(t_k), t_k] + \mathbf{e}_{t_k}. \quad (2)$$

Here,  $\mathbf{h}[\cdot]$  is a  $l$ -dimensional (possibly non-linear) function,  $\{t_k\}_{k=0}^N$  is the set of sampling time instants and  $\{\mathbf{e}_{t_k}\}_{k=0}^N$  represents the measurement errors. Usually, the number  $l$  of available time series is smaller than  $n$ , and often  $l = 1$ . This means that the observed signals provide information only on a projection of the dynamics of the system of interest. Curiously, instead of developing new methodologies for the analysis of the observed signals in a given physical context, in the past much effort has been

spent on attempts to devise more and more sophisticated techniques for the *statistical* characterization of  $\mathbf{y}(t_k)$  (e.g. AR and ARMA modelling, maximum entropy power spectra, wavelets analysis, ...). The expectation of these efforts was that such a characterization could be able to provide hints on the functional form of  $f[\cdot]$ . However, the results have very often been disappointing since the experimental time series do not contain all the information necessary for such a task.

This does not mean that the classic statistical analysis of the time series is unproductive. However, it has to represent only a starting point, otherwise there is the risk that the studies on the time evolution of the astrophysical objects could merely consist of a collection of data and the elaboration of some generic statistical measures with no direct physical meaning.

On the basis of this argument, Vio et al. (1992) stressed the necessity to start directly from model (1), as provided by a given theoretical model, and to use time series as a test for such a model. From the consideration that many astrophysical objects show evolutions that are unpredictable over time, they suggest modifying Eq. (1) to

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), \mathbf{w}(t), t, \boldsymbol{\theta}], \quad (3)$$

with  $\mathbf{u}(t)$  and  $\mathbf{w}(t)$  representing deterministic and random processes, respectively, and then to solve them via a numerical approach. In other words, the functional form of  $f[\cdot]$  is assumed known out of a set  $\boldsymbol{\theta}$  of parameters. In this way, once the values of  $\boldsymbol{\theta}$  are fixed, it is possible to obtain “*synthetic*” light curves that can be compared to the observed ones. The reason to add the random process  $\mathbf{w}(t)$ , typically a continuous Gaussian white noise<sup>2</sup>, is a consequence of the fact that this term represents the interaction of the physical system of interest with its surroundings and/or the action of complex processes that cannot be directly included in the model (e.g. gas turbulence). In general, such processes are characterized by a huge number of degrees of freedom and therefore they can be assumed to have a stochastic nature. In practice, this means study of the time evolution of a given physical system in the context of the so-called *stochastic dynamics*, i.e., through the modelling of the observed time series by means of stochastic differential equations (SDE).

Although *stochastic dynamics* is an approach widely used for the study and simulation of realistic scenarios in many fields of applied science and engineering such as fluid dynamics, structural and mechanical engineering, avionics, material properties, financial sciences... (van Kampen 1981; Ghanem & Spanos 1991; Kloeden et al. 1997; Garcia-Okjalvo & Sancho 1999; Gardiner 2004), in astronomy it is not well known. However, as we will show on the basis of some numerical experiments, even the action of quite weak noise sources on simple nonlinear dynamical systems can produce deep modifications of their behaviour over time. This means that in real scenarios the noise component must be considered as intrinsic to the physics of the systems and not only a secondary factor.

<sup>2</sup> NB. Hereafter, with the term noise we will indicate only the random processes perturbing the dynamics of a given physical system and not the contamination due to the measurement errors.

Consequently, in many situations, *stochastic dynamics* constitutes the most natural approach in the analysis of the experimental time series.

In Vio et al. (1992) it is suggested that the value of  $\boldsymbol{\theta}$  has to be derived from physical considerations. Here, we provide some tools that permit us to estimate  $\boldsymbol{\theta}$  directly from the data.

In Sect. 2 some arguments are presented that support the necessity of stochastic modelling in the study of the physical systems and in Sect. 3 an example in the astronomical context is provided. In Sect. 4 it is shown that modelling the time series through discrete models is unsuited to many physical systems. Hence, in Sect. 5 the use of SDE is suggested and some tools are provided to fit them to the experimental time series. An example application is given in Sect. 6. The conclusions and some final comments are presented in Sect. 7.

## 2. Why is stochastic modelling necessary?

In this section we will consider the time evolution of the simple dynamical system

$$\dot{x}(t) = -x^3(t) + 6x^2(t) - 11x(t) + 6 \quad (4)$$

suffering the influence of different kinds of noise. This is a nonlinear model with the particularity that the associated potential is characterized by two regions, both of them with their own stable point of equilibrium (see Fig. 1). We have deliberately chosen to represent an unsophisticated physical situation since our aim is to show that, even in this simplified scenario, the fact of not considering the noise as a fundamental component of the dynamics makes it essentially impossible to get insights into the characteristics of the system under study.

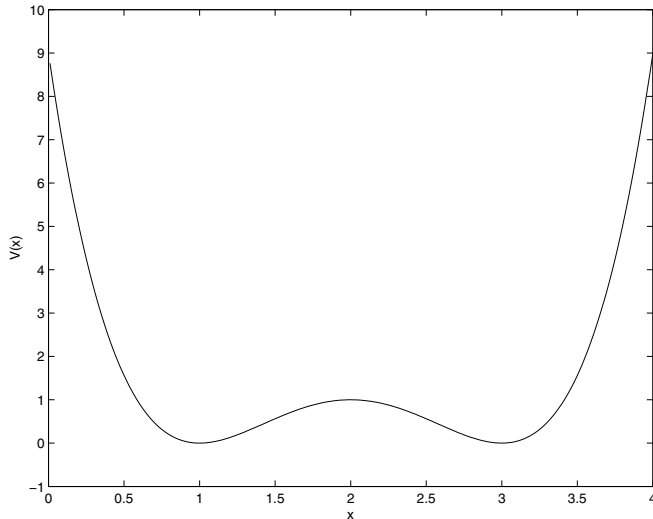
### 2.1. Perturbation by an additive noise process

Figures 2a,b show two realizations of the stochastic differential equation

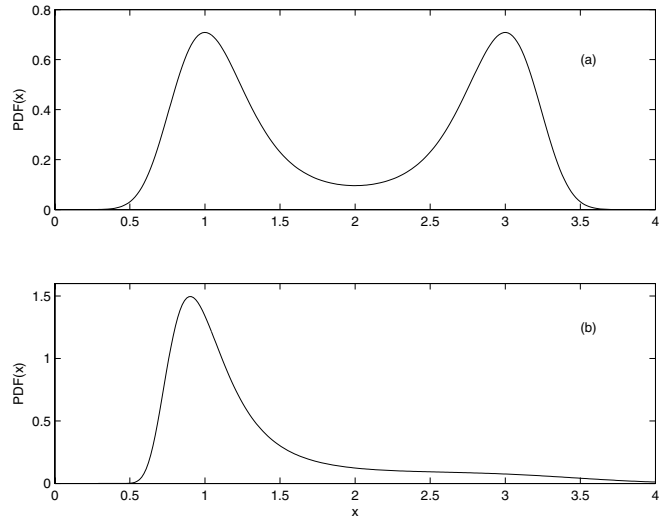
$$\dot{x}(t) = -x^3(t) + 6x^2(t) - 11x(t) + 6 + \sigma w(t), \quad (5)$$

where  $w(t)$  is a zero-mean, unit-variance and continuous white noise process. In both the simulations the same realization of  $w(t)$  is used but two different values, 0.1 and 0.5, have been adopted for the constant  $\sigma$ . Notice that in this example  $w(t)$  acts as a simple additive perturbing process.

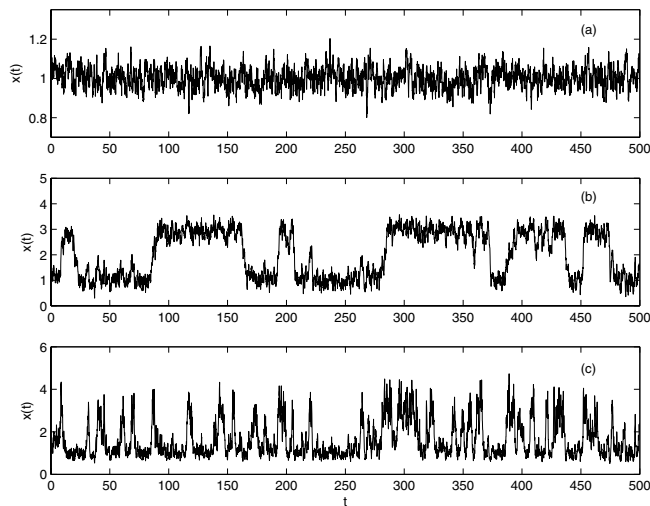
Clearly, the two signals show different time evolution. The reason lies in the two points of equilibrium that characterize the dynamics of model (4). Indeed, the only possibility for the system to jump from one region to the other is represented by the perturbation  $\sigma w(t)$ . If such perturbation is strong enough, then the jumps will be frequent and the bi-stable characteristic of the system will be revealed even by short observed signals. Conversely, if the perturbation is small, then it is possible that to observe a single jump requires a long time. From the statistical analysis of  $x(t)$  it is still possible to obtain some results in this unfavourable situation. For example, the Keenan test (Keenan 1985), a test devised to verify the nonlinearity of the time series, is able to detect the nonlinear nature of the data sequence shown in Fig. 2a at a confidence level of 95%. However,



**Fig. 1.** Potential function associated with the dynamical model (4).



**Fig. 3.** Probability density functions associated a) with the dynamical model (5); and b) with the dynamical model (6).



**Fig. 2.** a), b) Time series obtained from the dynamical model (5) with  $\sigma$  equal, respectively, to 0.1 and 0.5; c) time series obtained from the dynamical model (6) with  $\sigma$  equal to 0.5.

if no jump is detected, even the most sophisticated statistical analyses will be unable to provide more detailed information on the functional form of model (4).

### 2.2. Perturbation by a multiplicative noise process

The situation worsens when  $x(t)$  affects the intensity of the perturbation process. This is shown by Fig. 2c that presents a realization of the stochastic system

$$\dot{x}(t) = -x^3(t) + 6x^2(t) - 11x(t) + 6 + \sigma x(t)w(t). \quad (6)$$

Here, the realization of process  $w(t)$  is the same used as in the previous example, and the value of  $\sigma$  is equal to that used in Fig. 2b. With respect to model (5), now the standard deviation of the noise component at time  $t$ , given by  $\sigma x(t)$ , is not constant but depends on the value of the signal at the same time instant.

A comparison of Fig. 2b with Fig. 2c indicates that, although the only modification regards the noise component, the

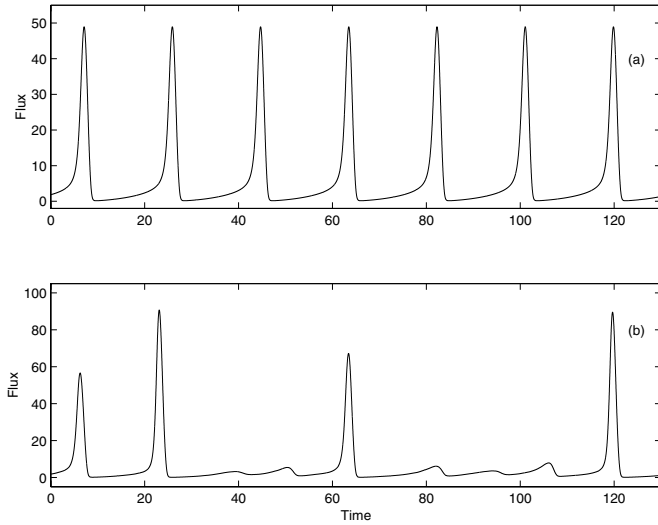
dynamical behaviour of  $x(t)$  has suffered deep changes. This is illustrated in Figs. 3a,b which show the probability density functions (PDF) associated with the two processes. From Fig. 3b it is evident that the bi-stable structure of model (4) is no longer detectable. The problem is that signal  $x(t)$ , because of the term  $x(t)w(t)$ , is no longer able to furnish direct information on the deterministic part of system (6). This illustrates that an approach based only on the statistical analysis of the experimental data can be quite non-informative, no matter how sophisticated the technique applied.

Unfortunately, it is highly probable that a situation like this one constitutes a typical situation for many astrophysical objects. For example, it is to be expected that the luminosity of a given object could depend on the quantity of gas accreting onto it, while it is conceivable that the accretion rate is influenced by the energy emitted by the object itself.

### 3. An astronomical example

In order to show that neglecting the stochastic component of a nonlinear dynamical system can be risky also in the case of astrophysical systems, here we give an example based on the *pair-production instability model* by Moskalik & Sikora (1986), that has been proposed to explain the strongly variable emission of high-energy radiation from the central regions of the active galactic nuclei (AGN). For simplicity, the details of this model will be not given.

According to this model, the strongly variable emission in the hard X-ray electromagnetic waveband of AGNs can be modelled via a scenario where the gas, accreting a black hole residing in the central regions of these objects, suffers a “pair production instability” (an instability due to the creation of electron-positron pairs). The time evolution of such a system can be formalized via the following set of differential equations describing, respectively, 1) pair creation and annihilation; 2) photon production, absorption and escape; 3) electron/positron heating and cooling; and 4) the proton density



**Fig. 4.** **a)** light curve  $n_\gamma(t)$  obtained by the numerical integration of the system of Eqs. (7)–(10); **b)** the same light curve when Eq. (10) is substituted with Eq. (11).

changes:

$$\frac{dn_+}{dt} = \dot{n}_+^{\text{cre}} - \dot{n}_+^{\text{ann}} \quad (7)$$

$$\frac{dn_\gamma}{dt} = \frac{\dot{u}_{\text{cbr}}}{kT_e} - \frac{3n_j}{1 + \tau_T} \frac{c}{R} - 2 \frac{dn_+}{dt} \quad (8)$$

$$\frac{du_e}{dt} = \dot{u}_{\text{ep}} - \dot{u}_{\text{cbr}} \quad (9)$$

$$\frac{dn_p}{dt} = \lambda_0 - \frac{n_p}{t_{\text{ep}}} \quad (10)$$

where  $u_e \simeq n_e kT_e$ ,  $T_e$  = electron temperature,  $n_e$ ,  $n_+$ ,  $n_\gamma$ , and  $n_p$  are, respectively, the densities of the electrons, positrons, photons, and protons,  $\tau_T = n_e R \sigma_T$ ,  $\sigma_T$  = Thomson cross-section,  $\dot{n}_+^{\text{cre}}$  = pair creation rate,  $\dot{n}_+^{\text{ann}}$  = pair annihilation rate,  $\dot{u}_{\text{ep}}$  = rate of energy transfer from protons to electrons,  $\dot{u}_{\text{cbr}}$  = electron cooling rate,  $\lambda_0$  = quantity proportional to the accretion rate of the gas surrounding the central regions of AGNs, and  $t_{\text{ep}}$  = electron energy transfer time. In the present context, the observed X-ray light curves can be assumed to be proportional to the quantity  $n_\gamma(t)$ .

For certain values of the parameters, this system gives rise to periodic hard X-ray flares. In the recent past, this model has enjoyed a certain fame because of this ability. However, a serious drawback of the above scenario is that the accretion rate of the gas is supposed to be strictly constant. Of course, this is a strong assumption and it raises some doubts about the reliability of the periodic behaviour under realistic conditions.

Figures 4a,b show what happens to a periodic light curve, obtained from the numerical solution of Eqs. (7)–(10), when the accretion rate  $\lambda_0$  is perturbed by a continuous, additive white noise process with standard deviation  $\sigma = 0.15 \lambda_0$ . Formally, this means substituting Eq. (10) with:

$$\frac{dn_p}{dt} = [\lambda_0 + \sigma w(t)] - \frac{n_p}{t_{\text{ep}}}, \quad (11)$$

a much more realistic assumption. Although the perturbation is not so strong, the periodic behaviour of the light curve has almost completely disappeared.

The indication provided by this experiment is that the modelling of astrophysical systems via deterministic equations can be misleading. Indeed, some features can be predicted but, probably, are very difficult to actually observe in experimental data.

#### 4. Is discrete modelling appropriate?

In the previous sections it has been shown that neglecting the noise component in the dynamics of the astrophysical systems can be risky with regard to their expected time evolution. Moreover, a simple statistical approach appears inadequate. Hence, some modelling is necessary. Since the experimental time series are discrete in nature, one could be tempted to adopt a discrete approach to model the observed signals. This is not appropriate because, if the continuous nature of the signals is not taken into account, the results provided by any method of analysis in general depend on the sampling time. Consider a simple stochastic model:

$$\dot{x} = -\theta x(t) + \sigma w(t); \quad \theta \geq 0, \quad (12)$$

where  $\theta$  and  $\sigma$  are constants. When this system is observed at a set of discrete, evenly spaced time instants, its dynamics can be described in terms of the discrete model

$$x_{k+1} = \alpha x_k + w_k, \quad (13)$$

where  $x_k = x(t_k)$ , and  $\{w_k\}_{k=0}^N$  is the realization of a discrete white noise process. From the point of view of classical time series analysis, Eq. (13) represents an AR(1) model. This fact could suggest the possibility of obtaining information on system (12) by means of the classical discrete AutoRegressive modelling. Unfortunately, this is not true because the relationship between  $\theta$  and  $\alpha$  (Vio et al. 1992) is given by

$$\alpha = \exp(-\theta \Delta t), \quad (14)$$

i.e., the results provided by the discrete model depend on  $\Delta t$ .

Equation (14) could suggest of adopting discrete models to fit the observed time series and then to use the results to recover the parameters of the continuous models. However, such an approach is not practical. In the case of nonlinear systems it can even not be possible to determine the functional form that links the continuous and the corresponding discrete models (for example, see Timmer 2000). Considering the following one-dimensional model

$$\dot{x}(t) = \mu[x(t), \theta] + \sigma[x(t), \theta]w(t). \quad (15)$$

If  $\Delta t$  is sufficiently small, i.e.  $\Delta t \equiv \delta t$ , Eq. (15) can be approximated by

$$x_{k+1} = x_k + \mu[x_k, \theta]\delta t + w_k \sigma[x_k, \theta](\delta t)^{1/2}, \quad (16)$$

i.e., with a discrete model,

$$x_{k+1} = h[x_k] + \epsilon_k, \quad (17)$$

where  $h[x_k]$  is a function that can be related back to the parameter vector  $\theta$  and  $\{\epsilon_k\}$  are independent, discrete Gaussian random quantities. However, things are more difficult if, as it usually happens in practical applications,  $\Delta t$  is not small enough for approximation (16) to hold. In fact, it is often possible to model a discrete sequence of data in the form (Priestley 1988)

$$x_{k+1} = k[x_k, x_{k-1}, x_{k-2}, \dots] + \eta_k, \tag{18}$$

with  $\{\eta_k\}$  discrete random quantities not necessarily Gaussian and/or independent, but the relationship between  $k[\cdot]$  and  $\mu[\cdot]$  can be not defined. This has important consequences. For example, according to the frequency sampling, a given time series can display different types of nonlinearities or even appear as a linear process. Moreover,  $\theta$  cannot be estimated from  $k[\cdot]$ .

These considerations indicate that discrete modelling suffers intrinsic limitations that make it unsuited to the description of physical systems. Consequently, an approach based on continuous models appears more appropriate (see also Maybeck 1979, 1982).

### 5. SDE for modelling time series

An effective use of continuous models for the analysis of experimental signals requires two distinct operations: a) fit of specific dynamical models to the time series; b) validation of the results through the generation of synthetic signals to compare with the experimental ones. In turn, these operations require the capability to estimate the parameters in SDE from discrete observations and the numerical solution of this kind of equations.

#### 5.1. Parameter estimation in SDE

In the past, various approaches have been suggested to estimate the parameters in SDE given discrete observations (e.g., see Bibby & Sorensen 1995; Timmer 2000; Brandt & Santa-Clara 2002; Durham & Gallant 2002, and reference therein). However, most of them have serious limits in computational burden, and/or the impossibility of dealing with measurement errors, and/or difficulty in numerical implementation.

Here, we present a general and flexible approach that is applicable to systems in the form

$$\dot{x}(t) = f[x(t), u(t), t, \theta] + \sigma[u(t), t, \theta] w(t); \tag{19}$$

$$y_{t_k} = h[x(t_k), u(t_k), t_k, \theta] + e_{t_k}, \tag{20}$$

where  $x(t) \in \mathbb{R}^n$  is the vector of the state variables,  $u(t) \in \mathbb{R}^m$  is the vector of known input deterministic variables,  $y_{t_k} \in \mathbb{R}^l$  is the vector of observable variables,  $\theta \in \mathbb{R}^p$  is the vector of parameters,  $f[\cdot] \in \mathbb{R}^n$ ,  $\sigma[\cdot] \in \mathbb{R}^{n \times n}$  and  $h[\cdot] \in \mathbb{R}^l$  are (possibly) nonlinear functions,  $w(t)$  is an n-dimensional, standard, continuous white noise process with covariance given by the identity matrix, and  $e_{t_k}$  is the vector of the measurements errors supposed to be zero mean, Gaussian quantities with covariance matrix  $S(u_{t_k}, t_k, \theta)$ . The quantities  $\{e_{t_k}\}$  and  $w(t)$  are supposed to be mutually independent for all  $t$  and  $t_k$ . This model is not as general as model (3). However, it is of interest in various practical applications. Moreover, as shown in Appendix A, through

an appropriate transformation it is often possible to transform the more general state model

$$\dot{x}(t) = f[x(t), u(t), t, \theta] + \sigma[x(t), u(t), t, \theta] w(t), \tag{21}$$

(i.e., with  $\sigma[\cdot]$  that depends also on  $x(t)$  to the form (19).

Given a particular model structure, the *maximum likelihood* (ML) estimation of the unknown parameters can be performed by finding the parameters  $\theta$  that maximize the likelihood function of a given sequence of measurements, say  $y_{t_0}, y_{t_1}, \dots, y_{t_k}, \dots, y_{t_N}$ . By introducing the notation

$$\mathcal{Y}_k = [y_{t_k}, y_{t_{k-1}}, \dots, y_{t_1}, y_{t_0}], \tag{22}$$

the likelihood function is the joint probability density

$$L(\theta; \mathcal{Y}_N) = p(\mathcal{Y}_N | \theta), \tag{23}$$

or equivalently:

$$L(\theta; \mathcal{Y}_N) = p(y_{t_0} | \theta) \prod_{k=1}^N p(y_{t_k} | \mathcal{Y}_{k-1}, \theta). \tag{24}$$

Here the rule  $p(A \cap B) = p(A|B)p(B)$  has been applied to form a product of conditional probability densities. In order to obtain an exact evaluation of the likelihood function, the initial probability density  $p(y_{t_0} | \theta)$  must be known and all the subsequent conditional densities must be determined by successively solving Kolmogorov's forward equation and applying Bayes' rule (Jazwinski 1970), but this approach is computationally unfeasible in practice. Given that in Eq. (19) the term containing  $\sigma(t)$  does not depend on  $x(t)$ , a more efficient alternative can be proposed. In particular, since the dynamics of model (19) is driven by Gaussian, white noise processes, it is reasonable to assume that, under some regularity conditions and for sampling time steps  $\Delta t_k$  not too large (see Appendix B), the conditional PDFs  $p(y_{t_k} | \mathcal{Y}_{k-1}, \theta)$  can be well approximated by Gaussians. Since the Gaussian density is completely characterized by its mean and covariance, by introducing the notation

$$\widehat{y}_{t_k|t_{k-1}} = E\{y_{t_k} | \mathcal{Y}_{k-1}, \theta\}, \tag{25}$$

$$R_{t_k|t_{k-1}} = V\{y_{t_k} | \mathcal{Y}_{k-1}, \theta\}, \tag{26}$$

$$\epsilon_{t_k} = y_{t_k} - \widehat{y}_{t_k|t_{k-1}}, \tag{27}$$

where  $E[\cdot]$  and  $V[\cdot]$  are, respectively, the mean and covariance operators, the likelihood function (24) can be written in the form

$$L(\theta; \mathcal{Y}_N) = p(y_{t_0} | \theta) \prod_{k=1}^N \frac{\exp\left(-\frac{1}{2} \epsilon_{t_k}^T R_{t_k|t_{k-1}}^{-1} \epsilon_{t_k}\right)}{(\det[R_{t_k|t_{k-1}}])^{1/2} (2\pi)^{l/2}}. \tag{28}$$

For a fixed  $\theta$ , the quantities  $\epsilon_{t_k}$  and  $R_{t_k|t_{k-1}}$  can be computed by means of an *extended Kalman filter* (Kristensen et al. 2004, see also Appendix B). Further, conditioning on  $y_{t_0}$  and taking the negative logarithm  $l(\theta) = -\ln[L(\theta; \mathcal{Y}_k | y_{t_0})]$  gives

$$l(\theta) \propto \sum_{k=1}^N \left( \ln(\det[R_{t_k|t_{k-1}}]) + \epsilon_{t_k}^T R_{t_k|t_{k-1}}^{-1} \epsilon_{t_k} \right). \tag{29}$$

The ML estimate of  $\theta$  (and optionally of  $y_{t_0}$ ) can now be determined by solving the nonlinear optimization problem

$$\widehat{\theta} = \arg \min_{\theta} [l(\theta)]. \quad (30)$$

An estimate of the uncertainty of  $\widehat{\theta}$  is obtained by the fact that the ML-estimator is asymptotically normal with mean  $\theta$  and covariance  $\Sigma$  given by the lower bound of the Cramer-Rao inequality, i.e.,

$$\Sigma = \mathbf{H}^{-1}. \quad (31)$$

Here, the Hessian matrix  $\mathbf{H} = \{h_{ij}\}$  is given by

$$h_{ij} = -\mathbb{E} \left\{ \frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j} \right\}, \quad (32)$$

that can be approximated with

$$h_{ij} \approx - \left\{ \frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j} \right\}_{\theta=\widehat{\theta}}. \quad (33)$$

### 5.2. Numerical solution of SDE

Although the theory behind SDE is complex, their numerical integration does not present more problems than the integration of the deterministic differential equations (Kloeden et al. 1997; Higham 2001). Indeed, for a general SDE,

$$\dot{x}(t) = a[t, x(t)]dt + b[t, x(t)]w(t), \quad (34)$$

the simplest integration scheme, i.e. the Euler method, is given by

$$x_{t_{k+1}} = x_{t_k} + \mu_{t_k} \delta t_k + \sigma_{t_k} w_{t_k} \sqrt{\delta t_k}. \quad (35)$$

Here,  $\mu_{t_k} = \mu[t_k, x_{t_k}]$ ,  $\sigma_{t_k} = \sigma[t_k, x_{t_k}]$ ,  $\delta t_k = t_{k+1} - t_k$ , and  $\{t_k\}_{k=0}^N$  represents a set of not necessarily equispaced time instants. A pitfall of this method is that its order of strong convergence  $\gamma$  is rather small, say 1.0 for additive noise and 0.5 for multiplicative noise. For this reason, sometimes it is preferable to use the Milstein scheme,

$$x_{t_{k+1}} = x_{t_k} + \bar{\mu}_{t_k} \delta t_k + \sigma_{t_k} w_{t_k} \sqrt{\delta t_k} + \frac{1}{2} \sigma_{t_k} \sigma'_{t_k} w_{t_k}^2 \delta t_k, \quad (36)$$

that has  $\gamma = 1$  for both kinds of noise (for additive noise the two schemes coincide). Here,  $\sigma_{t_k} = \sigma[t_k, x_{t_k}]$ ,  $\bar{\mu}_{t_k} = \mu[t_k, x_{t_k}] - \frac{1}{2} \sigma_{t_k} \sigma'_{t_k}$ , and the symbol “'” denotes differentiation with respect to  $x(t)$ . Integration schemes with higher values of  $\gamma$  are possible, however they are rather more cumbersome to implement (e.g., see Kloeden et al. 1997). The algorithms (35) and (36) can be extended to deal with systems of SDE (Kloeden et al. 1997).

### 6. A worked example

To illustrate the potential of stochastic modelling in the analysis of time series, we have considered a sequence of X-ray observations of the low mass X-ray binary Sco X-1 (van der Klis, priv. comm.) made with the *Proportional Counter Array* (PCA) on the Rossi-XTE spacecraft (Bradt et al. 1993). The time series used contains 1000 data points; the first 300 points are

shown in Fig. 5. Sampling is regular with a time step of 0.015 s. As shown in Fig. 6, this time series presents a power-spectrum typical of QPOs, i.e., a broad peak superimposed on a steeply decreasing continuum. If one interprets such a fact as due to a single driving mechanism, a possible model for reproducing the observed signal is

$$\dot{x}(t) = -ax(t) + \sigma_1 w(t); \quad (37)$$

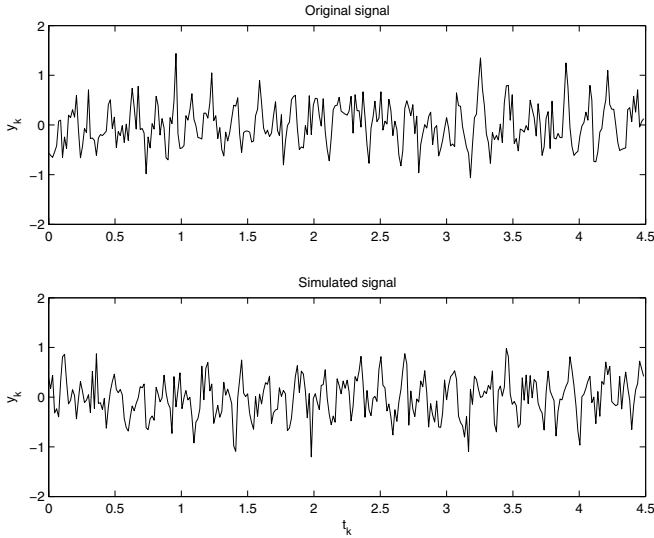
$$y_k = x(t_k) + A \cos[\omega_0 t_k + \sigma_2 x(t_k)] + e_k. \quad (38)$$

Here,  $a$ ,  $A$ ,  $\sigma_1$ , and  $\sigma_2$  are unknown constants that are to be estimated. The frequency  $\omega_0$  and the variance of the measurement errors  $e_k$  (assumed independent and Gaussian) are estimated through the central position of the broad peak and the high frequency level in the estimated power-spectrum, respectively. The process  $x(t)$  is assumed to be not observable.

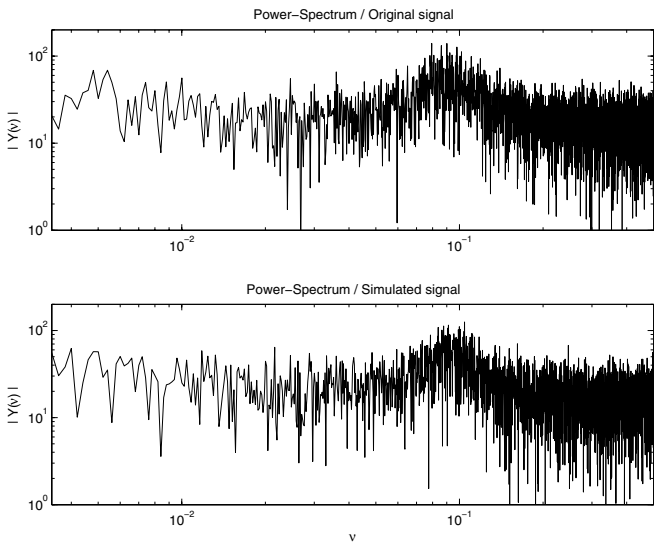
The idea behind this model is that the luminosity of the object is determined by a driving linear stochastic process  $x(t)$  (e.g. the accretion rate) superimposed on a periodic process that, however, is perturbed by  $x(t)$  itself. Figures 5 and 6 display the results obtainable with the methodology explained in the previous section. Figure 7 shows that the stochastic model (37)–(38) is able to satisfactorily reproduce the main spectral features of the experimental signal. Of course, this does not mean that such a simple model corresponds to a real scenario. The only thing that it is possible to claim is the compatibility of model (37)–(38) with observations (in a rigorous study, a more complete validation is necessary). However, this is the most that can be obtained from any technique of data analysis.

### 7. Conclusions

The time series usually available in astronomy are able to characterize only a subset of the system of equations that describe the dynamics of the physical system under study. For this reason, although in principle it is always possible to determine a statistical model able to reproduce the experimental data, without any a priori physical model there are not many possibilities to obtain a reliable reconstruction of the physical scenario investigated. In general, this means that an approach to the analysis of time series exclusively based on the experimental data will provide inconclusive results, and that the practice to search for more and more sophisticated statistical techniques is not productive. In many situations, the only possibility for physical insights is to carry out the analysis in a well-defined physical context. In this respect, *stochastic dynamics*, i.e. modelling the time series with stochastic differential equations, appears to be a promising tool. A benefit of working within such a framework is that one is forced to provide a mathematical/physical formalization of the starting hypotheses adopted in the analysis of the signals (e.g. linearity, non-linearity, type of non-linearity, ...). In this way, there is no risk of misunderstandings concerning the interpretation of the final results. Additionally, a more direct relation between complex physical models and the limited observational material is made possible permitting a more efficient interaction between data and theory.



**Fig. 5.** Upper panel: original (mean subtracted) time series of SCO-X1; lower panel: typical time series obtainable through the fit of model (37)–(38) (see text). Although the time series used in the analysis contains 1000 data points, here, for ease of comparison, only the first 300 are shown. Time is in seconds.



**Fig. 6.** Upper panel: power-spectrum  $|Y(v)|$  of the first 5000 data of the time series of SCO-X1; lower panel: power-spectrum  $|Y(v)|$  of a typical 5000 points realization of the process given by the fitted model (37)–(38). The frequency  $v$  is in Nyquist units.

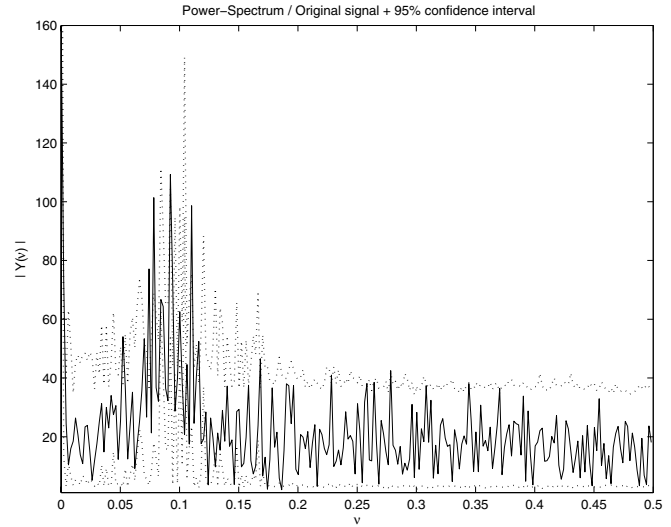
A code implementing some efficient numerical tools for modelling the experimental time series with stochastic differential equations is freely downloadable from <http://www.imm.dtu.dk/ctsm/>.

*Acknowledgements.* We thank Prof. M. van der Klis for the availability of the data on Sco X-1.

## Appendix A: A multivariate transformation

In this section, a bijective transformation

$$z(t) = \Psi[x(t), t] \quad (\text{A.1})$$



**Fig. 7.** Power-spectrum  $|Y(v)|$  of the first 5000 data of the time series of SCO-X1 and the 95% confidence interval for the power-spectrum of the simulated signal in Fig. 6. For each frequency, the interval has been estimated on the basis of the 2.5th and 97.5th percentiles obtained by 1000 different realizations of the noise process in the measurement-equation (38). For clearness, the power-spectra have been downsampled by a factor of 10.

is proposed to transform the state-equation

$$\dot{x}(t) = f[x(t), u(t), t, \theta] + \sigma[x(t), u(t), t, \theta]w(t), \quad (\text{A.2})$$

to

$$\dot{z}(t) = \tilde{f}[z(t), u(t), t, \theta] + \tilde{\sigma}[u(t), t, \theta]w(t). \quad (\text{A.3})$$

Here,  $\Psi[\cdot]$  is assumed continuously differentiable with respect to  $t$  and twice continuously differentiable with respect to  $x(t)$ . The covariance matrix of  $w(t)$  is assumed to be equal to the identity matrix. Further assumptions are:

1. All the elements in  $\sigma[x(t), u(t), t, \theta]$  are strictly nonzero, i.e.

$$\sigma_{ij}[x(t), u(t), t, \theta] \neq 0, \quad i, j = 1, 2, \dots, n. \quad (\text{A.4})$$

2. For each  $i$  there exists only one  $\sigma_{ij}$  as a function of one and only one state-variable  $x_{v(i)}(t)$ , where  $v(i)$  is different for each  $i$ , i.e.,

$$\sigma_{ij}[x(t), u(t), t, \theta] = \sigma_{ij}[x_{v(i)}(t), u(t), t, \theta]. \quad (\text{A.5})$$

3. The functions  $\sigma_{ij}[x_{v(i)}(t), u(t), t, \theta]$  are bijective and  $\sigma_{ij}^{-1}[x, u(t), t, \theta]$  are integrable with respect to  $x$ .

Given these assumptions, it can be shown (Nielsen et al. 2001) that the transformation

$$\Psi_l[x_{v(i)}(t)] = \int \frac{dx}{\sigma_{ij}[x(t), u(t), \theta]} \Big|_{x=x_{v(i)}(t)}, \quad (\text{A.6})$$

$l, i, j = 1, 2, \dots, n$ , fullfills Eq. (A.3).

After the transformation (A.1), the state-equation (A.3) contains the same parameters  $\theta$  as the original state-equation (A.2). Moreover, the measurement-equation (20),

$$y_{t_k} = h[x(t_k), u(t_k), t_k, \theta] + e_{t_k}, \quad (\text{A.7})$$

can be used in its original form since the state-equation  $x(t)$  is obtainable from the inverse transformation  $x(t) = \Psi^{-1}[z(t), t]$ .

## Appendix B: Kalman filter for estimating parameters in SDE

In Sect. 5.1 it is shown that the parameters  $\theta$  of a system of SDEs can be estimated through the maximization of the likelihood function (30). This requires the quantities  $\epsilon_{t_k}$  and  $\mathbf{R}_{t_k|t_{k-1}}$  that, however, are unknown and have to be estimated. Here, we propose an approach based on the *continuous-discrete extended Kalman filter*. In particular, if at a given time instant  $t_k$  the quantities  $\theta$ ,  $\widehat{\mathbf{x}}_{t_k|t_{k-1}} = E[\mathbf{x}_{t_k}|\mathbf{x}_{t_{k-1}}]$ , and  $\mathbf{P}_{t_k|t_{k-1}} = E[\mathbf{x}_{t_k}\mathbf{x}_{t_k}^T|\mathbf{x}_{t_{k-1}}]$  are fixed, then the procedure is based on the iterated solution of the following sequence of equations

1. The *output prediction equations*

$$\widehat{\mathbf{y}}_{t_k|t_{k-1}} = \mathbf{h}[\widehat{\mathbf{x}}_{t_k|t_{k-1}}, \mathbf{u}_{t_k}, t_k, \theta]; \quad (\text{B.1})$$

$$\mathbf{R}_{t_k|t_{k-1}} = \mathbf{C}\mathbf{P}_{t_k|t_{k-1}}\mathbf{C}^T + \mathbf{S}. \quad (\text{B.2})$$

2. The *innovation equation*

$$\epsilon_{t_k} = \mathbf{y}_{t_k} - \widehat{\mathbf{y}}_{t_k|t_{k-1}}. \quad (\text{B.3})$$

3. The *Kalman gain equation*

$$\mathbf{K}_{t_k} = \mathbf{P}_{t_k|t_{k-1}}\mathbf{C}^T\mathbf{R}_{t_k|t_{k-1}}^{-1}. \quad (\text{B.4})$$

4. The *updating equations*

$$\widehat{\mathbf{x}}_{t_k|t_k} = \widehat{\mathbf{x}}_{t_k|t_{k-1}} + \mathbf{K}_{t_k}\epsilon_{t_k}; \quad (\text{B.5})$$

$$\mathbf{P}_{t_k|t_k} = \mathbf{P}_{t_k|t_{k-1}} - \mathbf{K}_{t_k}\mathbf{R}_{t_k|t_{k-1}}\mathbf{K}_{t_k}^T. \quad (\text{B.6})$$

5. The *prediction equations*

$$\frac{d\widehat{\mathbf{x}}_{t|t_k}}{dt} = \mathbf{f}[\widehat{\mathbf{x}}_{t|t_k}, \mathbf{u}_t, t, \theta], \quad t \in [t_k, t_{k+1}); \quad (\text{B.7})$$

$$\frac{d\mathbf{P}_{t|t_k}}{dt} = \mathbf{A}\mathbf{P}_{t|t_k} + \mathbf{P}_{t|t_k}\mathbf{A}^T + \sigma\sigma^T, \quad t \in [t_k, t_{k+1}). \quad (\text{B.8})$$

Equations (B.7) and (B.8) provide the quantities  $\widehat{\mathbf{x}}_{t_{k+1}|t_k}$  and  $\mathbf{P}_{t_{k+1}|t_k}$  that can be used to start a new iteration of the sequence. Here,  $\sigma = \sigma(\mathbf{u}_{t_k}, t_k, \theta)$ ,  $\mathbf{S} = \mathbf{S}(\mathbf{u}_{t_k}, t_k, \theta)$ , and

$$\mathbf{A} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}_t} \right|_{\mathbf{x}=\widehat{\mathbf{x}}_{t_k|t_{k-1}}, \mathbf{u}_{t_k}, t=t_k, \theta}, \quad (\text{B.9})$$

$$\mathbf{C} = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}_t} \right|_{\mathbf{x}=\widehat{\mathbf{x}}_{t_k|t_{k-1}}, \mathbf{u}_{t_k}, t=t_k, \theta}. \quad (\text{B.10})$$

Initial conditions for the iteration are  $\widehat{\mathbf{x}}_{t|t_0} = \mathbf{x}_{t_0}$  and  $\mathbf{P}_{t|t_0} = \mathbf{P}_{t_0}$ , which may either be pre-specified or considered as additional parameters to estimate.

In this procedure, the most problematic point is represented by the integration of Eqs. (B.7)–(B.8). Here, however, two considerations are necessary. The first one is that only Eq. (B.8) makes use of the linearization of function  $\mathbf{f}[\cdot]$ , whereas Eq. (B.7) is exact (i.e., no approximation is used). The second one is that both equations are deterministic. These points are important since they imply that a) Eqs. (B.7)–(B.8) can be

solved through higher order schemes for ordinary differential equations (ODE); b) the conditional expectation  $\widehat{\mathbf{x}}_{t|t_k}$  is propagated with no error, i.e. the linearization of function  $\mathbf{f}[\cdot]$  can be considered to have only second order effects on the overall procedure, since only the propagation of the conditional covariance  $\mathbf{P}_{t|t_k}$  is affected. The consequence is that the procedure can be expected to provide satisfactory results also in the case of sampling time steps that are comparatively large.

Concerning the ODE solver, various possibilities are available. One of the most effective is Adam's method in the implementation of Hindmarsh (1983). This is an algorithm equipped with an automatic determination of the integration time step that permits one to solve Eqs. (B.7)–(B.8) simultaneously. Moreover, at each integration step, the matrices  $\mathbf{A}$  and  $\sigma$  are re-evaluated. This provides a better approximation to the true state propagation solution because it more accurately reflects the true time-varying nature of the two matrices.

## References

- Akaike, H. 1971, *Ann. Inst. Stat. Math.*, 23, 163
- Bibby, B. M., & Sorensen, M. 1995, *Bernoulli*, 1, 17
- Bradt, H. V., Rotschild, R. E., & Swank, J. H. 1993, *A&AS*, 97, 355
- Brandt, M. W., & Santa-Clara, P. 2002, *J. Fin. Econ.*, 63, 161
- Durham, G. B., & Gallant, R. 2002, *J. Busin. Econ. Stat.*, 20, 297
- Garcia-Ojalvo, J., & Sancho, J. M. 1999, *Noise in Spatially Extended Systems* (Berlin: Springer-Verlag)
- Gardiner, C. W. 2004, *Handbook of Stochastic Methods*, 3rd ed. (London: Springer)
- Ghanem, R. G., & Spanos, P. 1991, *Stochastic Finite Elements. A Spectral Approach* (Berlin: Springer-Verlag)
- Higham, D. J. 2001, *SIAM Rev.*, 43, 525
- Hindmarsh, A. C. 1983, *ODEPACK, A Systematized Collection of ODE Solvers*, in *Scientific Computing*, ed. R. S. Stepleman, IMACS Transactions on Scientific Computation (Amsterdam: North-Holland), 1, 55
- Honerkamp, J. 1993, *Stochastic Dynamical Systems* (New York: VCH)
- Jazwinski, A. H. 1970, *Stochastic Processes and Filtering Theory* (New York: Academic Press)
- Keenan, D. M. R. 1985, *Biometrika*, 72, 39
- Kloeden, P. E., Platen, E., & Schurz, H. 1997, *Numerical Solution of SDE Through Computer Experiments* (Berlin: Springer-Verlag)
- Kristensen, N. R., Madsen, H., & Jorgensen, S. B. 2004, *Automatica*, 40, 225
- Maybeck, P. S. 1979, *Stochastic models, estimation, and control* (London: Academic Press), 1
- Maybeck, P. S. 1982, *Stochastic models, estimation, and control* (London: Academic Press), 2
- Moskalik, P., & Sikora, M. 1986, *Nature*, 319, 649
- Nielsen, J. N., & Madsen, H. 2001, *Automatica*, 37, 107
- Priestley, M. B. 1988, *Non-linear and Non-stationary Time Series Analysis* (London: academic Press)
- Timmer, J. 2000, *Chaos, Solitons & Fractals*, 11, 2571
- van Kampen, N. G. 1981, *Stochastic Processes in Physics and Chemistry* (Amsterdam: North-Holland)
- Vio, R., Cristiani, S., Lessi, O., & Provenzale, A. 1992, *ApJ*, 391, 518