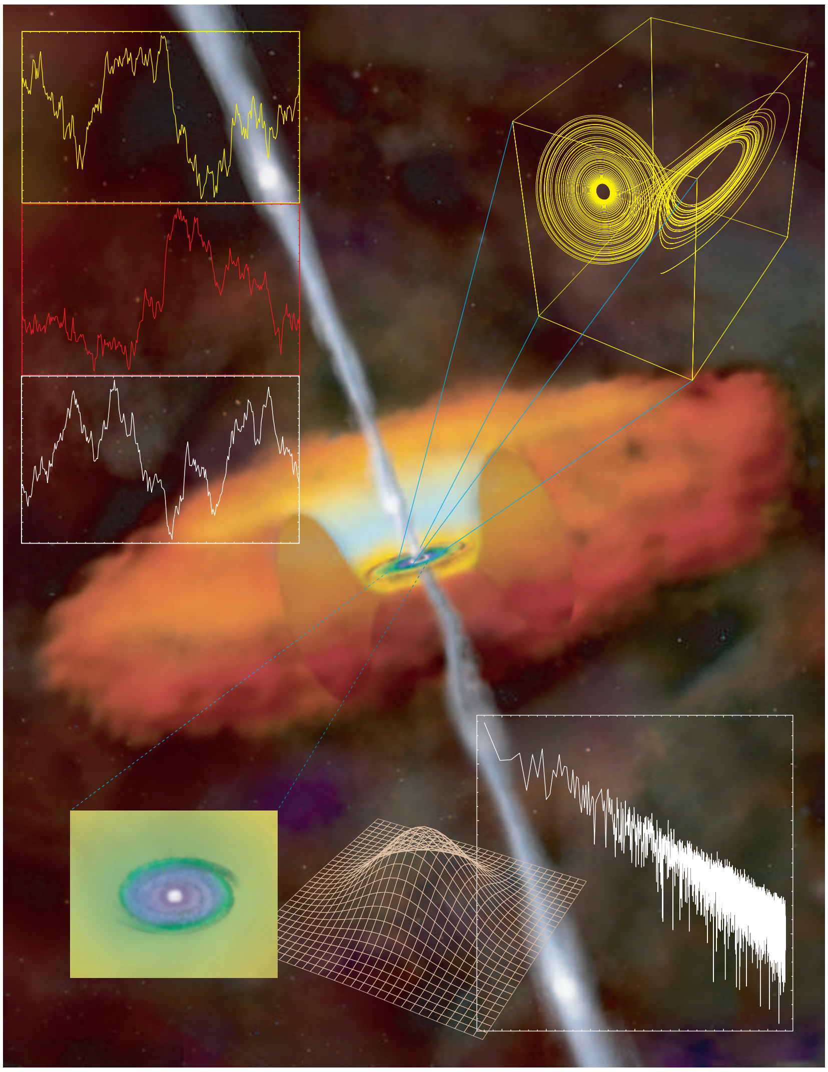
Dissertation

submitted to the
Combined Faculties for the Natural Sciences
and for Mathematics
of the Ruperto-Carola University of
Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Dipl.-Phys. Dimitrios Emmanoulopoulos
born in Athens, Greece

Oral examination: 25th July 2007

Nonlinear Time Series Analysis of BL Lac Light Curves



Referees

Prof. Dr. Stefan Wagner
Prof. Dr. John Kirk

The picture in the previous page (the AGN and the accretion disk in the inset) was taken from the Chandra Resources web site: `http://chandra.harvard.edu/resources/illustrations/agn/`. It illustrates in a vivid way the main effort which is made in this work to disclose any deterministic signature behind the apparent stochastic variability behavior of AGN through nonlinear time series analysis.

To my family

Στην οικογένειά μου

Zusammenfassung: In der vorliegenden Arbeit werden mit Hilfe der Methoden linearer und nichtlinearer Zeitreihenanalysen die zeitlichen Eigenschaften von BL Lac-Objekten in ausführlicher Art und Weise untersucht. Trotz in großer Menge gesammelter Daten haben die Variabilitätsuntersuchungen der letzten zwanzig Jahre keine wesentlichen Ansatzpunkte für das Verständnis der Verhaltensweise dieser Objekte geliefert. Üblicherweise wird sich auf unklare Zeitreihenanalysemethoden ohne jegliche mathematische Grundlage berufen, welche zu fehlerhaften Einschätzungen der Zeiteigenschaften und schließlich zu Missinterpretationen in der Modellierung der betrachteten Systemdynamik führen.

Anhand der bekannten Quellen Mrk 421 und Mrk 501 sollen zunächst die Mängel einiger aktueller Zeitreihenanalysemethoden aufgezeigt und die Notwendigkeit des Einsatzes der Zeitreihenanalyse höherer Ordnung demonstriert werden. Präsentiert wird dann eine eingehende Betrachtung moderner nichtlinearer Analysemethoden zusammen mit Beispielen, die es erlauben, auch für astronomische Zeitreihen Anwendung zu finden. Anschließend werden diese Methoden auf den Datensatz aus der durch RXTE über einen Zeitraum von neun Jahren gewonnenen Röntgenlichtuntersuchung der Quelle Mrk 421, angewendet. Dies soll die Beantwortung der folgenden Fragen ermöglichen: Ist eine Erklärung des Variabilitätsverhaltens dieser Quellen auf der Basis weniger physikalischer Parameter (deterministisches System) möglich oder ist dieses das Ergebnis zahlreicher Komponenten eines stochastischen Systems?

Schließlich werden die Ergebnisse der längsten Multiwellenlängen-Kampagne für BL Lac-Objekte, durchgeführt zwischen August und September 2004, vorgestellt. Die Quelle wurde hierbei sowohl von H.E.S.S. mit hochenergetischen γ-Strahlen (>100 GeV), als auch durch Röntgenstrahlen (2–10 keV) von RXTE und dem optischen R-Band dreier erdgebundener Observatorien beobachtet.

Abstract: In this work the time properties of the BL Lac objects are elaborated in a detailed manner through linear and nonlinear time series analysis methods. In spite of the large amount of available data in the last 20 years, the variability studies have not provided major progress for understanding the behavior of these objects. Vague time series analysis methods, lacking any mathematical foundation, are usually invoked revealing erroneous time properties in the data sets which then act misleadingly for modeling the dynamics of the system under study.

The flaws of some of the current time series analysis methods are reviewed thoroughly throughout this work for specific sources (Mrk 421, Mrk 501) and the need of employing higher order time series analysis methods is demonstrated. An extensive description of the modern nonlinear analysis methods is presented together with examples being implemented in a way to be applicable to astronomical time series. Then, these methods are applied to the X-ray data set of Mrk 421, obtained by RXTE, covering a time period of 9 years, giving some hints to answer the question: Is it possible to explain the variability behavior of these sources based on few physical parameters (deterministic system), or is it the result of numerous components yielding from a stochastic system?

Finally the results from the longest multiwavelength campaign, conducted during August–September 2004, for the BL Lac object PKS 2155-304 are presented. The source was observed in the very high energy γ-rays (>100 GeV) by H.E.S.S., in the X-rays (2–10 keV) by RXTE and in the optical (R-band) by three terrestrial observatories.

Contents

Abbreviations–Acronyms

ACF autocorrelation function

ADC analog digital converter

AGN active galactic nucleus

AMSL above mean sea level

ANU Australian national university

AR autoregressive model

ARCH autoregressive conditional heteroscedasticity model

ARF ancillary response file

ASM all-sky monitor

BeppoSAX Beppo satellite per astronomia X

BL Lac BL Lacertae

BLR broad-line region

CALDB calibration database

Cas A Cassiopeia A

CASS/UCSD center for astrophysics & space sciences at the university of California, San Diego

CCD charge-coupled device

CGRO Compton gamma ray observatory

CSA charge sensitive amplifier

CSR/MIT center for space research at the Massachusetts institute of technology

DFT discrete Fourier transform

DOF degrees of freedom

EA event analyzer

EAR exponential autoregression (or ExpAR)

EDS experimental data system

EGRET energetic gamma ray experiment telescope

EW equivalent width

EXOSAT European X-ray observatory satellite

ExpAR exponential autoregression (or EAR)

FITS flexible image transport system

FOV field of view

FR I Fanaroff-Riley type I galaxy

FR II Fanaroff-Riley type II galaxy

FSRQ flat spectrum radio quasars

FWHM full width at half maximum

FWZI full width at zero intensity

GRB γ-ray burst

GSFC/NASA Goddard space flight center of the national aeronautics and space administration

GTI good time interval file

HBL high-energy peaked BL Lacertae

HEAO high energy astrophysical observatory

HEASARC high energy astrophysics science archive research center

H.E.S.S. High Energy Stereoscopic System

HEXTE high energy X-ray timing experiment

HVPS high voltage power supply

IC inverse Compton

IDFT inverse discrete Fourier transform

IDV intraday variable

IID independent and identically distributed

KVA Kungliga Vetenskapsakademien (Royal Swedish Academy of Sciences)

LBL low-energy peaked BL Lacertae

LVPS low voltage power supply

MET mission elapsed time

MIT Massachusetts institute of technology

MJD modified Julian day

MOD method of delays

NASA national aeronautics and space administration

NHP null hypothesis probability

NLR narrow-line region

NSF normalized structure function

PCA proportional counter array

PCU proportional counter unit

Prin.Comp.Anal. principal component analysis

PSD power spectral density

PSPC position sensitive proportional counter

QSO quasi-stellar object

RIF remote interface

RMF redistribution matrix function

ROTSE robotic optical transient search experiment

RSP response matrix

RXTE Rossi X-ray timing explorer

SA shaping amplifier

SAA south Atlantic anomaly

SED spectral energy distribution

SETAR self-exciting threshold autoregression

SF structure function

SSC synchrotron self-Compton

SShC scanning shadow camera

SSO Siding Spring observatory

SSRQ steep spectrum radio quasars

Std2 Standard-2

SVD singular value decomposition

TPG test pulse generator

VHE very high-energy

VLBA Very Long Baseline Array

VLBI Very Long Baseline Interferometry

Chapter 1

INTRODUCTION

Time series analysis in astronomical research played a crucial role in the historical development of the field. Until the twentieth century, the heavens could be studied only with the human eye or with very simple telescopes used as “vision enhancers”. From this perspective the sky contains only the Sun, the Moon and the planets whose position exhibits diurnal, monthly and annual periodicities. Together with these objects stars and nebulae are also present but they seem to be completely unchanged with respect to all their observed characteristics (motion and brightness).

Simple methods of registering the consecutive positions of the Sun and the moon for several years created the first calendars. Astronomers in the Middle East had discerned *saros*, the 18.6 year cycle of eclipses. The pre-Columbian Mayans developed a detailed calendar based on eclipses and the periodic motion of Venus. All these methods comprised the first systematic study of astrophysical objects in terms of time series analysis. One could say that Tycho Brahe who accurately measured and tabulated planetary positions over decades together with his assistant Johanes Keppler who derived empirical relations from the given time series data sets regarding the planetary motions, were the first methodical time series analysts of modern astronomy.

During the last century sophisticated technological extensions of the human eye permit a deeper view of the universe. The charge-coupled devices (CCDs) mounted in the focus of the telescope permit more sensitive and more accurate measurements of the brightness and the color of an object than that achieved with the human eye. Moreover, modern telescopes and detectors overcome the eye’s restriction to a narrow range of wavelengths of light, spanning only a factor of 2, and permit a study of the sky over an incredible range of 10^{20} in wavelength.

The key to understand the rising interest in astronomical time series is that many objects in the sky which are innocuous or even invisible at optical wavelengths are spectacularly variable in brightness or other properties at other wavelengths. Astrophysical objects that were considered to be constant luminosity sources appear nowadays to be variable in various time scales. Variable stars, accretion binaries, pulsars and active galactic nuclei consist only a small fraction of the “zoo” of temporally variable astronomical beasts. This thesis deals only with the latter category of these objects but the time series methods used to analyze the behavior of these sources are applicable to any astrophysical source.

1.1 The Astronomical Time Series

While there is an enormous range of specific methodologies concerning the time series analysis of a given data set coming from a particular system, astronomical time series exhibit some special peculiarities. The already existing methods that have been developed from statisticians and mathematicians in order to be applicable in natural, social and health sciences as well as in engineering or even psychology can not be used directly in astronomical data sets. The latter differ in several aspects from the classical data sets with respect to their structure and the nature of their elements.

Firstly, the times of observations are often discontinuous and irregular. Observations in the optical regime can be interrupted by either the rise of the Sun or by the weather conditions. Concerning the satellite experiments the Earth occultations as well as the charged particles cause an inhomogeneous acquisition rate of the data. Therefore appropriate methods which are not affected from the sampling should be applied to the given data sets. In the case of interpolated data sets application of any analysis method should be thoroughly checked through simulations.

Secondly, the data are often acquired with noise, yielding measurement uncertainties which may also differ from observation to observation. These uncertainties can be taken into account by observing background regions and calibration sources. Thus, heteroscedastically weighted time series methods with known measurement errors/uncertainties errors are often needed.

Thirdly, it is usually assumed that a time series data set is linear and Gaussian. The simple and widely used method of linear regression based on the method of least squares, assumes Gaussianity concerning the parent distribution of the measurements. Linearity is also another "silent" assumption which is hidden behind the majority of the analyses methods. Autocorrelation and structure functions (SFs) make use only of the first two moments of the data set (i.e. mean value and variance) appropriate to fully describe Gaussian distributions of data sets coming from linear systems. Hence higher order time series analyses must be employed in order to use all the available information existing in the data sets.

Fourthly, there are few cases only which the scientific goal of an astronomical time series study is the prediction or the forecast. Rather, the astrophysicist seeks to characterize the time variations in terms of a physical model involving parameters that yield from a given analysis method. The main drawback of this approach especially in the field of active galactic nuclei (AGN) is that the origin of the variations is most of the times ignored. Therefore, whatever "time feature", is emerged from an analysis method, it is directly used to fulfill the needs of a model as an "appropriate" and "convenient" time scale element, irrespective of which part of the AGN is modeled (accretion disk or emitting regions in the jet). It is often a great challenge to link the results of a statistical analysis with specific physical source parameters in a way of having "one-to-one" correspondence.

Nevertheless, even flawless astronomical data sets which have the appropriate form and they are ready to accept any conventional analysis method, are sometimes treated with "vague" statistical methodologies. The latter are usually formulated based on more intuitional rather than mathematical arguments. An examination of the astronomical literature on variable objects reveals a surprising disregard for well consolidate time series analysis methods developed for other applications. Standard texts in time series and statistics are rarely consulted. The result of this insularity is that unnecessary methodological development occurs and methods may be accepted without careful evaluation of their mathematical merits or operational deficiencies. The reason behind this tendency is probably reflecting an effort of filling roughly the major gap that exists between the time properties of the deterministic and the random physical processes. A "fake" characteristic time scale can be used as a handle to pass from a completely random data set to something with fixed time properties. This inexistent time scales which are supposed to be embedded in the data sets most of the times create confusion and misunderstandings.

Concerning the AGN, in spite the fact of the large amount of available data in the last 30 years the variability studies have not provided major progress. Phenomenological methods are usually applied and peculiar time scales are derived with no mathematical support of their physical meaning. In this thesis examples concerning the aforementioned problems are going to be discussed in a thorough and detailed manner. At the same time numerous firmly established time series analysis methods are going to be presented and applied in astronomical data sets originating mainly from a special class of the AGN objects known as BL Lacertae (sect.1.2.1).

1.2 The Active Galactic Nuclei

1.2.1 General properties-Taxonomy

The term active galactic nuclei (AGN) denotes the ensemble of extragalactic objects sharing one common characteristic, which distinguishes them from all the other galaxies in the Universe, being that the emission from the nucleus largely outshines that of the whole galaxy. Various subclasses forming the AGN population, e.g. radio galaxies, quasars or Seyfert galaxies, have been known for quite a long time but only in relatively recent years attempts have been made to unify them in a coherent picture. According to this unified view all AGN subclasses share a common mechanism of energy production, the accretion of matter onto a supermassive black hole in their center. Their different observational properties arise, to a large degree, from their intrinsically anisotropic geometry and radiation pattern, from absorption as well as from relativistic effects.

The first property of this objects is that they emit radiation over the entire electromagnetic spectrum having a bolometric luminosity $L_{\mathrm{bol}} \geq 10^{44}$ erg sec^{-1} sometimes reaching up to 10^{48} erg sec^{-1}. In comparison, normal galaxies have $L_{\mathrm{bol}} \leq 10^{42}$ erg sec^{-1} and the bulk of their luminosity is emitted in the visible band, essentially produced by stars. Another property of AGN is that their spectra over the whole electromagnetic band are essentially of non-thermal origin contrary to normal galaxies where the spectrum is given by the integration of the stars, thermal spectra. Some AGN also reveal strong radio and X-ray emission in the form of spectacular jets. Finally the basic characteristic which is the major concern of this thesis is the variability. The observed emission of these objects is highly variable in all energy bands on time scales raging from years down to minutes. All these properties indicate that powerful physical mechanisms with very high efficiency (higher from the nuclear processes) originate inside these extragalactic objects. Despite the fact that AGN consist $\sim 1\%$ of the total galaxy population they consist an excellent site of studying unexplored physical processes.

Concerning the taxonomy of AGN there is a general but not universally accepted scheme. They are divided into two main categories according to their radio emission (fig.1.1). In radio-quiet galaxies, the ratio of the galaxy's 5 GHz luminosity to its optical luminosity in the B-band (4400 Å) is generally less than 10. In the opposite case the galaxy is considered radio-loud. Roughly 85% of AGN are hosted in radio-quiet galaxies.

The radio-quiet galaxies which are not known to be γ-ray emitters can be further subdivided in two classes. The first class consists of the quasi-stellar objects (QSOs) where the host galaxy is not resolved, leaving only the AGN visible. The second class includes the Seyfert galaxies where, in contrast to the previous case, the host galaxy is visible. The Seyfert galaxies come into two different species type I and type II. The former exhibit both narrow and broad emission lines as well as a strong non-stellar continuum and the latter exhibit only narrow lines and a weaker non-stellar continuum.

The radio-loud galaxies (also called quasars) come in two flavors, Fanaroff-Riley (FR) type I and II. The FR I galaxies are generally less luminous than their type II counterparts and they tend to have a weaker optical emission as well. The radio emission of the former galaxies is classified as *core dominated* as opposed to the *lobe dominated* emission region of the latter.

Of primary interest to very high-energy (VHE) γ-ray astronomy (>100GeV) is a subclass of FR I galaxies called BL Lacertae (BL Lac), named after the name of their prototype that was originally mistaken for a star. BL Lac have flat or inverted radio spectra with a smooth and featureless continuum and also emission lines are entirely absent or very weak unlike all other kind of AGN. Among all the AGN they exhibit the most variable emission, their flux is beamed and additionally they exhibit high polarization reaching values up to 30%. The current general radiation model for the case of BL Lac objects consists of the synchrotron emission from accelerated electrons at low energies and the inverse Compton (IC) emission, via photon scattering, at high energies (γ-rays). These processes form a broad band two component shape spectrum, covering 20 orders of magnitude, with two well defined peaks, one in the IR/optical or UV/X-ray band (due to synchrotron) and the other in the γ-ray regime (due to IC), separated by ~ 9 decades in frequency. The position of the synchrotron peak distinguishes the BL Lac into high-energy peaked (HBL) and low-energy peaked (LBL). It turns out that most of the HBL are TeV γ-ray sources and most of the LBL are EGRET sources (i.e. GeV sources).

Finally the FR II galaxies are divided into the flat spectrum radio quasars (FSRQ) flat and steep spectrum radio quasar (SSRQ). The first category is the analog of the BL Lac and both of them form separate category called blazars. It must be remarked that the optical spectra of BL Lac and FSRQ differ greatly. In fact, whereas FSRQ show strong broad emission lines, BL Lac have either weak emission lines of typical equivalent width (EW) of less than 5 Å or no emission lines at all in their optical spectra. The reason of putting them in a common category is because they share the same continuum properties such as strong variability and high polarization.

1.2.2 Orientation scenario

All the different types of objects presented in sect.1.2.1 are usually explained in terms of an orientation scenario of the AGN with respect to the viewing angle (Urry & Padovani 1995). In general type I objects are viewed close to the jet axis and type II are view perpendicular to it (fig.1.2).

Concerning the radio-quiet galaxies, Seyfert I generally exhibit broad and narrow lines meaning that the AGN is viewed from the observer at a small angle from the jet axis therefore both the broad-line region (BLR) and the narrow-line region (NLR) are visible. In the opposite case of Seyfert II the AGN is viewed along the plane of the torus therefore the BLR is obscured and only the NRL is visible.

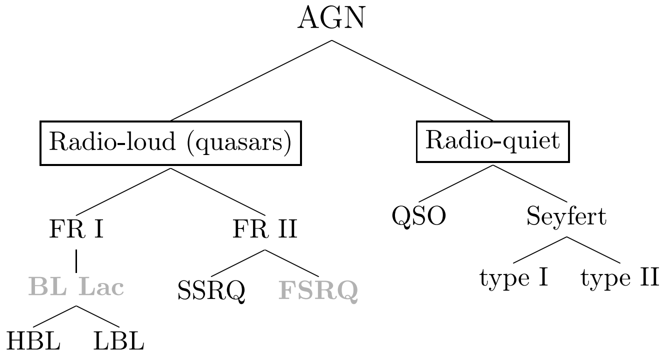


Figure 1.1: The taxonomy of AGN. The BL Lac and the FSRQ can be joined into one category, named blazars, depicting the ultra-variable character of these objects among all the AGN.

In the radio-loud regime, blazars are thought to be oriented such as their jets are aligned very close to the line of sight resulting "exotic" relativistic phenomena, such as the relativistic beaming and the superluminal motion.

If a source of radiation moves with relativistic velocities towards the observer the intensity of the radiation will be amplified by Doppler boosting in the direction of motion. The time intervals measured in the observer's frame will be shorter than in the rest frame of the source and consequently the measured frequencies will be higher. These effects are all direct consequences of the Lorentz transformations of special relativity and they are known as relativistic beaming. The parameter which quantifies the relativistic beaming is the Doppler factor:

$$\delta = \frac{1}{\Gamma(1-\beta\cos\theta)} \tag{1.1}$$

where Γ is the bulk Lorentz factor and θ is the angle between the line of sight of the observer and the direction of motion of the source.

The specific intensity of the source $I(\nu)$ in the observer's frame, is boosted with respect to that of the source rest frame $I'(\nu')$

$$I(\nu) = \delta^3 I'(\nu') \quad \left(\mathrm{erg\,sec^{-1}cm^{-2}Hz^{-1}sr^{-1}}\right) \tag{1.2}$$

where primed quantities refer to the rest frame of the source.

In many blazars single radio components (i.e. blobs) have been seen moving at apparent velocities $u_{\rm app}$ greater than the speed of light c. This has been interpreted as an effect of relativistic beaming in the following way. When a source is moving at a velocity u close to c along a direction forming a small angle with the observer's line of sight it "runs after" its emitting photons. This reduces the time intervals between the emission of two photons as measured in the observer's frame and the source appears to move faster than it actually does. The apparent velocity $\mathrm{u_{app}}$, is given by

$$u_{\rm app} = \frac{u\sin\theta}{c-u\cos\theta}c \tag{1.3}$$

Therefore for small angles of sight $0 < \cos\theta < 2uc/(c^2+u^2)$, the observed velocities exceed the speed of light: $u_{\rm app} > c$.

1.2.3 The SED of blazars

One defining characteristic of the AGN is their variable emission in almost all energy bands (sect.1.2.1). Among them the blazars are the most variable specie and as such they exhibit a very variable behavior ranging from minutes (Racine 1970; Miller et al. 1989; Papadakis et al. 2004) up to decades (Teräsranta et al. 2005). The usual way to illustrate the multiwavelength energy output of blazars (and general all the AGN) is through the so called *spectral energy distribution* (SED). This is a broad band spectrum covering the whole range of frequencies, from radio to γ-rays, and it is generally represented in a $\log(\nu f_\nu)$ versus ν plot. Such a plot has the advantage of showing approximately the emitted energy per unit logarithmic frequency

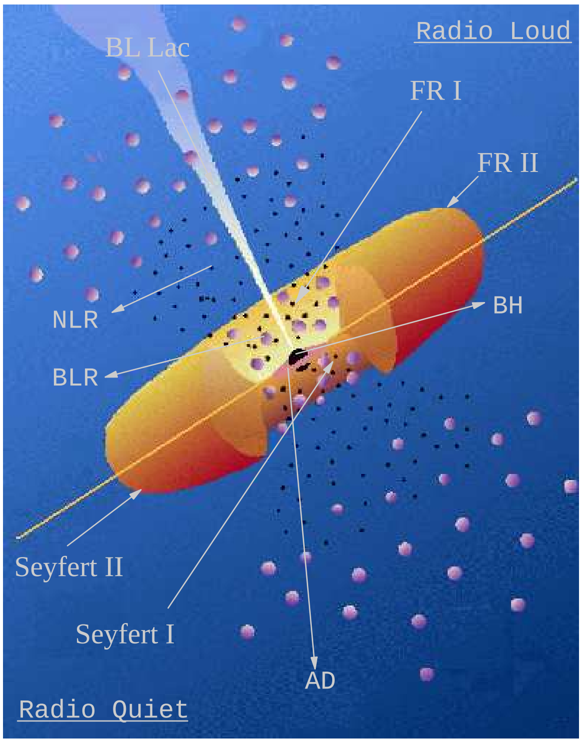


Figure 1.2: Schematic representation of the AGN orientation scenario (after Urry & Padovani 1995).

interval (or per decade of frequency), indicating immediately in which band most of the energy is released. The SED of blazars is characterized by two distinct components: one from radio to X-rays peaking at UV/soft X-rays and another one peaking at MeV-to-TeV γ-rays (Ulrich et al. 1997). The lower frequency component is attributed to synchrotron radiation of relativistic particles in magnetic fields. The higher frequency component is usually interpreted as IC radiation of "seed" photons coming either internally from the jet's synchrotron radiation, synchrotron self-Compton (SSC) or externally from various radiating sources (e.g. UV photons from the accretion disk, emission line clouds). The X-ray band comprises the link between these two regions therefore information about the temporal behavior of the source can greatly constrain the jet physics by providing information about the time properties of the radiation mechanism.

The time evolution of the SED (only in the X-ray regime) of the two most well studied BL Lac type blazars Mrk 421 (z=0.031) and Mrk 501 (z=0.033) during flaring events is shown in fig.1.3 (Xue et al. 2006). Several physical parameters for these two sources can be derived by fitting their SED with an SSC model. The general idea of the SSC model is that the observed X-ray emission comes from upscattered synchrotron photons (via the IC process) by the same population of electrons, assumed to follow a power-law, that has created them in the first place (via the synchrotron process). The derived source parameters are the power-law index of the emitting electrons, their maximum Lorenz factor, their total energy density and the magnetic field of the emitting source. The fixed model parameters are the Doppler factor δ, the minimum Lorenz factor of the electrons and the source size. The latter is the only parameter derived directly from the light curves and specifically from their time properties which are claimed to be directly connected with the linear size of the source (sect.8.1.1). This example shows an approach of how the time properties of a given data set can be used in order to probe into the physical parameters of the source.

1.2.4 The BL Lac object Mrk 421

Modeling the SED of the BL Lac objects is one way of deriving source parameters in a completely model depended way. In this work a try is made to extract as many information as possible directly from the light curves of these objects without making any assumption about the underlying radiation processes. The central point that one has to elucidate concerning the variability characteristics of the BL Lac objects is to check thoroughly for fixed time signatures in their flux variations which might be conserved for extended time periods. Thus sources providing with huge data sets must be employed in order to analyze their variability behavior homogeneously throughout the years. Probably the most extensively observed and well studied BL Lac object is Mrk 421 due to its interesting ultra-variable behavior. Among all

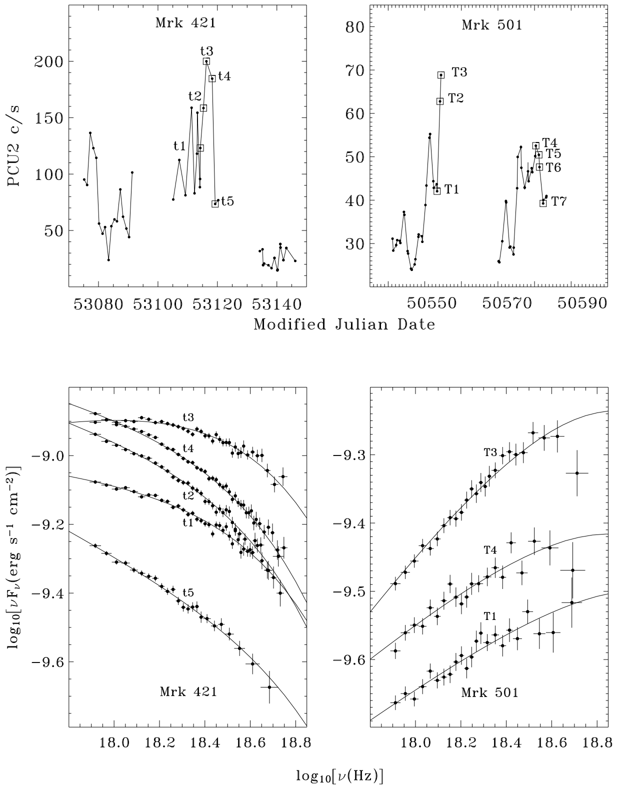


Figure 1.3: [Top panel] The light curves of the two BL Lac objects Mrk 421 and Mrk 501 during a flaring episode as registered by the PCU2 on board RXTE in the 2–60 keV energy band (taken from Xue et al. 2006). Detailed information about PCU2 and RXTE can be found in sect.2.1. [Bottom panel] The time evolution of the SED, focused in the X-ray regime, of the two BL Lac objects during the flaring events. The solid line represents the best fit models for the SSC scenario (taken from Xue et al. 2006).

the BL Lac, Mrk 421 (z=0.031) is the most bright source in the X-rays and it is the first extragalactic source detected in VHE (Punch et al. 1992). From the early X-ray observations with the *European X-ray observatory satellite* (EXOSAT) (George et al. 1988), Mrk 421 exhibited a very active behavior on ksec timescales together with spectral variations on daily time scales. Then observations with better X-ray time coverage of the source, conducted by ASCA and *Rossi X-ray Timing Explorer* (RXTE) started to reveal flares on time scales of about a day (Takahashi et al. 1996), fact that was firmly established in the late nineties with extended X-rays observations lasting around 10 days (Takahashi et al. 2000; Tanihata et al. 2001). Through simultaneous multifrequency campaigns it has been seen that the VHE flux behavior of the source shows correlations with the X-rays on daily time scales (Macomb et al. 1995) down to hourly time scales (Maraschi et al. 1999; Aharonian et al. 2003). In sub-hour time scales this correlation tends to loose (Błażejowski et al. 2005; Rebillot et al. 2006) declaring the fact that possibly some of the source parameters such as the volume may be changed with respect to time. Since for the last 10 years the RXTE satellite continues to observe Mrk 421 on a regular basis, these observations are going to be used for the analysis purposes of this thesis.

The chapter, following this introduction, contains a description of the basic instrumentation properties of the RXTE satellite together with the data analysis reduction procedures. In chap.3 some of the most powerful time series analysis methods are presented in a compact and comprehensible way adjusted to the needs of the astronomical data sets. The basic theories behind these methods together with simple examples are shown in order to familiarize the reader fully with concepts that are rarely used in the field of BL Lac astronomy. Then in chap.4 the main data sets which are used in this work concerning the BL Lac object Mrk 421 are presented. These consist of the X-ray observations obtained by the PCA and the ASM (instruments onboard RXTE) covering a time period of roughly 9 years. Up to now these are the biggest data sets, in this energy band, homogeneously analyzed for the given source and due to their length they give an additional opportunity to check about the validity of the debatable ASM data. In chap.5 the flaws of one of the most extensively used time series analysis method, the SF, are presented thoroughly through a series of simulations concerning some past ASCA X-ray observations of Mrk 501. The method is proven to be inappropriate for studying the time properties of BL Lac objects. A robust study of the latter necessitates the need of taking into account the higher order statistical moments of the data sets by making use of higher order time series analysis. Exactly this kind of analysis is performed in Chap.6 for the case of Mrk 421, aiming to reveal any possible evidence of deterministic behavior embedded in the short and/or in the long-term variability behavior of the source. The VHE variability behavior of another BL Lac object PKS 2155-304, during the period during August–September 2004, is presented in chap.7 as it was registered in the VHE by the *High Energy Stereoscopic System* (H.E.S.S.), in the X-ray band (2–10 keV) by RXTE and in the optical regime (R-band) by three terrestrial observatories. Finally chap.8 contains a discussion concerning the main results of this work, dealing mainly with BL Lac light curves, and their impact on the derivation of physical source parameters.

Chapter 2

INSTRUMENTATION AND DATA ANALYSIS

In this chapter the four instruments onboard the Rossi X-ray timing explorer (RXTE) satellite together with the data reduction methods are described in a concisely way. The *proportional counter array* (PCA) together with the *high energy X-ray timing experiment* (HEXTE) and the *all sky monitor* (ASM) consist the scientific instruments of the mission whereas the *experimental data system* (EDS) is the processing unit. The products coming from the three instruments are reduced based on the software, the methods and the criteria provided by GSFC/NASA.

2.1 The Rossi X-Ray Timing Explorer Satellite

The RXTE satellite, launched on 30/12/1995, is a NASA mission maintained and controlled by GSFC/NASA. The main characteristic of RXTE, as it is stated from its name, is the unprecedented time resolution since it was launched dedicated for observing variability processes of X-ray sources. Time scales from microseconds up to years can be covered by an instantaneous spectral range of 2 to 250 keV. Therefore historical archives containing the X-ray activity of the sources can be created and time series analysis on a long look basis can be conducted. The main characteristics of the observatory are listed in tab.2.1.

Table 2.1: Characteristics of the RXTE observatory

Launch date	30 December 1995
Launch vehicle	Delta-7920-10 rocket
Height	565–585 km
Total mass	3035 kg
Orbit type	low-earth circular
Period	$\sim$ 90 min
Inclination	22.99°

2.1.1 An overview of the onboard instruments

The mission carries in total three observing experiments and one microprocessor-based electronics package that controls the data acquisition. A set of five large area xenon proportional counter units (PCUs) forms the PCA sensitive in the energy range of 2–60 keV. Then, two large area sodium iodide scintillators sensitive in the energy range of 15–200 keV compose the HEXTE. Both of them are equipped with collimators yielding

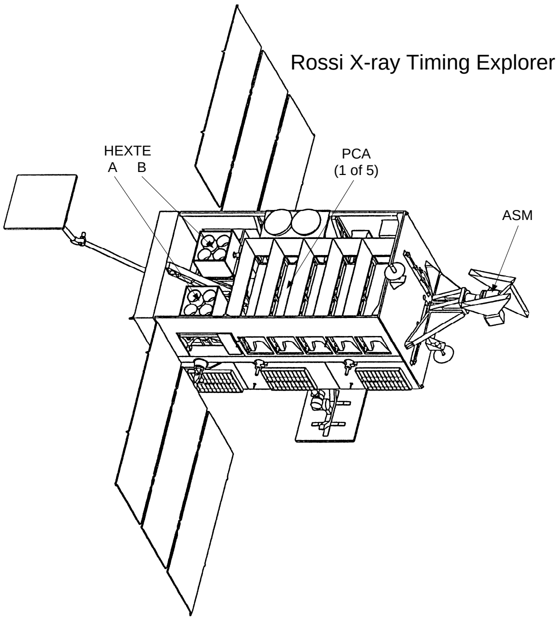


Figure 2.1: Arrangement of RXTE's instrumentation (taken from the web: `http://heasarc.gsfc.nasa.gov/docs/xte/xte_images.html`).

Table 2.2: General properties of the instruments onboard RXTE

	PCA	HEXTE	ASM
Energy range	2–60 keV	15–200 keV	2–12 keV
Energy resolution	<18% at 6 keV	15% at 60 keV	—
Time sampling	1 msec	8 μsec	90 minutes
FOV	1° FWHM	1° FWHM	6° × 90° FWHM/camera
Detectors	5 PCUs	2 clust. of 4 scintill.count.	3 scanning cameras
Collecting area	6500 cm^2	2×800 cm^2	90 cm^2

a FWHM of one degree. Finally, the ASM consists of three wide FOV scanning detectors, sensitive in the 2–10 keV energy range, monitors constantly 80% of the sky in every orbit. The events from both the PCA and the ASM are processed on board by the EDS in contradiction to HEXTE having its own built in data processor system. The EDS consists of 8 parallel processing systems known as *event analyzers* (EAs) which can carry out different analyses in parallel on the incoming data streams. The fig.2.1 shows the arrangement of the experiments upon the satellite and tab.2.2 lists some of their general properties.

The first experiment was developed by GSFC/NASA the second by CASS/UCSD and the other two by MIT.

2.1.2 Instrumentation properties of the PCA

The PCA experiment (Jahoda et al. 1996) consists of 5 sealed and collimated (1° FWHM), xenon-methane (Xe-CH_4) multi-anode PCUs sensitive to the energy range of 2–60 keV. The total effective area (fig.2.2) is 3000 cm^2 at 3 keV, 6000 cm^2 at 10 keV and 800 cm^2 at 50 keV. Since the 5 PCUs (numbered as PCU0, PCU1, PCU2, PCU3, PCU4) were designed to be essentially identical and operate independently all of the following description will refer to only a single detector module (fig.2.3).

Each PCU is covered with a thermal shield of aluminized Kapton polyimide 8.46 μm thick. This shield is part of the passive thermal design of PCA and each detector is thermally connected to the RXTE spacecraft. The thermal shield is not illuminated by direct sunlight except for observations which are conducted less than 45° away from the Sun. The collimating system of each unit consists of 5 *collimator modules*. Each module is formed from 76.2 μm beryllium-copper (Be-Cu) sheets which are tin-coated, stamped into half-hexagonal form, stacked and then heated causing the tin coating to solder the sheets

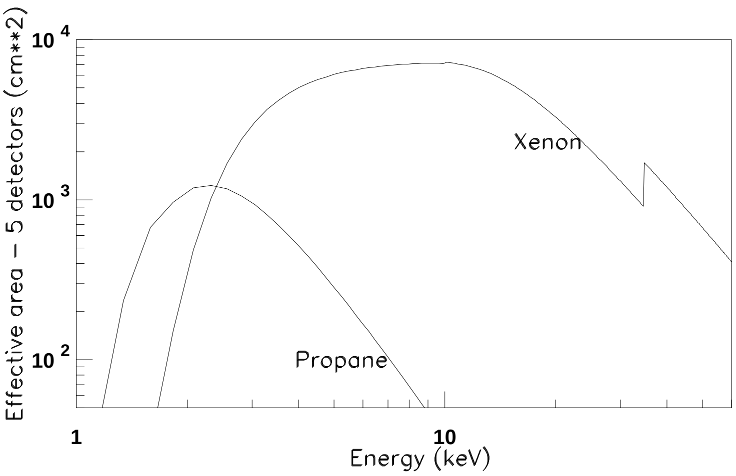


Figure 2.2: The effective area of the PCA as a function of the energy response of the complete PCA experiment for the propane and the xenon sections (taken from Schlegel 2006).

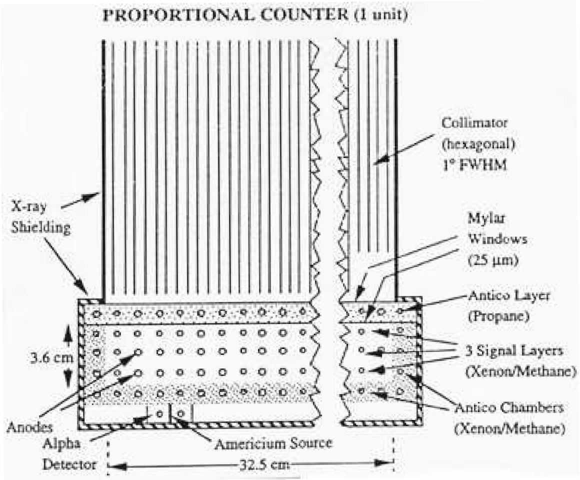


Figure 2.3: Transversal view of one PCU (taken from the web:
`http://heasarc.gsfc.nasa.gov/docs/xte/xte_images.html`).

Table 2.3: PCU signal chains

VP	All propane anodes connected together
XL1 and XR1	1st xenon layer of two interleaved sets of anodes
XL2 and XR2	2nd xenon layer of two interleaved sets of anodes
XL3 and XR3	3rd xenon layer of two interleaved sets of anodes
VX	Xenon veto layer
ALPHA	Events from the calibration source detector

together. Each module is a cube of 20.32 cm on a side and each hexagonal cell is 3.175 mm across the flats. The bottom surface is polished and coated with a small amount of uralane to protect the *Mylar window*. Finally, each of the five modules has a 1 cm^2 mirror attached to the front surface which were used in order to align them within their surrounding frame before they were fixed in position with epoxy.

Immediately behind the collimator is a 0.508 mm beryllium-copper shoe which duplicates the footprint of the first interior grid. A second window is held between the first and the second grid. These two windows together with the first grid form the boundaries a 1.3 cm deep volume filled with propane (VP). The propane ($CH_3CH_2CH_3$) volume serves as an electron veto region and charged particle anticoincidence shield, *antico propane layer*. The propane detector has twenty successive *anodes* across its width, spaced 1.3 cm apart. Each anode is separated from the other by thin solid aluminum walls that support the second window. The nominal pressure of propane is 1.05 atm. The double gas volume design requires that this pressure is less than or equal to the pressure behind the second window at all times during the filling and the operation of the detector.

Below the second window is the *xenon volume*. This consists of four wire grids, each with twenty anodes at the same spatial arrangement as for the propane volume. These anodes are separated by wire wall cathodes (5 across cell side) and this volume is normally filled with xenon plus 10% methane at a total pressure of 1.10 atm. All anode and cathode wires are made from gold coated stainless steel wire 50.8 μm in diameter. The wires are installed under tension sufficient to stay taut at the lowest survival temperature of -25° C. The upper three of the 4 xenon grids are used for the X-ray detections.

The bottom of the four xenon anode layers together with the nearest ones on the remaining two sides of the detector form a veto chamber for charged particles. The lowest surface ground plane of the bottom layer is defined by a beryllium-cooper back plate. Additionally this provides some shielding from events created at the rear of the counter. An Americium-241 (^{241}Am) source is mounted on this plate, providing a continuous low energy calibration signal of the tagged events. Ultimately each PCU contains a stack of electronics producing signals from nine independent anode chains summarized in tab.2.3.

In order to estimate the sensitivity levels of the three xenon layers for all the PCUs, it is assumed that a 1 mCrab source has a spectrum $N(E) = 0.005E^{-1.7}$ photons sec^{-1}cm^{-2} absorbed by an equivalent neutral hydrogen (H I) column of $N_{\rm H\,I}$=3×10^{20} atoms cm^{-2}. The estimated count rates summed over the five detectors are shown in tab.2.4. The signal to noise ratio S/N is defined as

$$S/N = \frac{st}{\sqrt{(s + cb + ib)t}} \tag{2.1}$$

where t is the time duration of the observation, s is the source count rate and cb and ib are the cosmic and instrumental background count rates respectively. Based on tab.2.4 the S/N as a function of time for the three layers for the PCA detector is shown in fig.2.4.

The energy resolution over the whole area of each PCU is exceedingly uniform and has a typical value for the xenon layers of 18% at 6 keV and 9% at 22 keV. The propane layer has a resolution of 18% at 5.9 keV. Externally each PCU is covered by a graded shield of tantalium (Ta), 1.524 mm thick over tin (Sn) to reduce the X-ray background flux can absorb the hard X-ray and γ-ray events generated in the spacecraft by cosmic ray impacts. The tin thickness (0.508 mm) is chosen to absorb escaped photons generated from interactions in the outer tantalium layer.

The electronics which are mounted on the back side of the detector consist of the following parts (for more details see Schlegel 2006, sect.4.2.3, p.28), charge sensitive amplifiers (CSAs), shaping amplifier (SA), test pulse generator (TPG), analog digital converter (ADC), remote interface (RIF), low voltage power supply (LVPS) and high voltage power supply (HVPS).

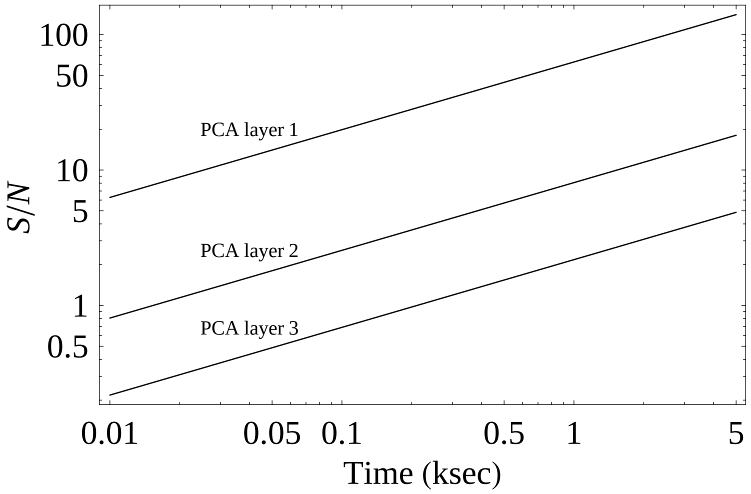


Figure 2.4: The S/N of the PCA layers as a function of time after summing the signals coming from all five PCUs for a source of 1 mCrab with a spectrum $N(E) = 0.005E^{-1.7}$ photons $\sec^{-1}\mathrm{cm}^{-2}$ absorbed by $N_{\mathrm{H\,I}}$=3×10^{20}atoms cm^{-2}.

Table 2.4: Sensitivity of PCA xenon layers in the energy band of 2–10 keV

PCA xenon layer	Source	Cosmic background	Instrumental background
1^{st}	10.20	8.54	7.66
2^{nd}	0.52	0.47	3.16
3^{rd}	0.13	0.07	3.36

2.1.3 The HEXTE

The HEXTE (Hink et al. 1992) (fig.2.5) consists of two clusters each one containing four phoswich ("phosphor sandwich") (NaI,CsI) scintillation detectors which are sensitive in the energy range of 15–200 keV. The total collecting area of the detector is 1600 cm^2 and the energy resolution of it is 15% at 60 keV. Each cluster can "rock" (beamswitch) along mutually orthogonal directions to provide background measurements 1.5° or 3.0° away from the source every 16 to 128 sec. Automatic gain control is provided by using a radioactive source (^{241}Am) mounted in each detector's FOV. In contrast to PCA and ASM both the time tagging of the events (with a time resolution of 8 μsec) and the telemetry stream, having an average data rate of 5 kb $\sec^{-1}$, are performed within HEXTE by its own processing unit. Data products include event mode, binned spectra and light curves, and a burst-triggered event buffer.

2.1.4 Instrumentation Properties of the ASM

The ASM (Levine et al. 1996) performs a nearly continuous monitoring of the X-ray sky. As such, it provides a record of the history of the X-ray sky during the lifetime of RXTE. The ASM consists of three basically *scanning shadow cameras* (SShCs) mounted on a rotating drive assembly (fig.2.6). Each camera has a FOV of 6°×90° i.e. FWHM (and a FWZI of $(\phi\times\theta)$=$(12^\circ\times110^\circ)$) and all of them cover about 80% of the sky upon the completion of one full rotation lasting 90 min. Two of the cameras share the same look direction but are canted by ±12° from each other, while the third camera looks in a direction parallel to the ASM drive axis. When the cameras rotated through 360°, the combined SShC exposure areas cover the entire sky, expect for a 45° half-angle cone centered toward the base of the instrument. The assembly dwells at a fixed position for $\sim$ 90 sec, followed by a rotation of 6°.

Each SShC has a position sensitive proportional counter, and the data is analyzed to give the total source intensity in the 2–12 keV energy band. The SShC consists mainly of a *position sensitive proportional counter* (PSPC), a mask plate, a collimator, a thermal shield and electronics. The PSPC is actually a multiwire proportional counter used for the detection and measurement of the position of each X-ray event in the coordinate perpendicular to the slits in the mask plate. The latter is a thin aluminum sheet, penetrated by a number of parallel slits and its role is to cast X-ray events of every X-ray source in the SShC FOV upon the PSPC. The collimator holds the slit mask in the desired position with respect to the PSPC and

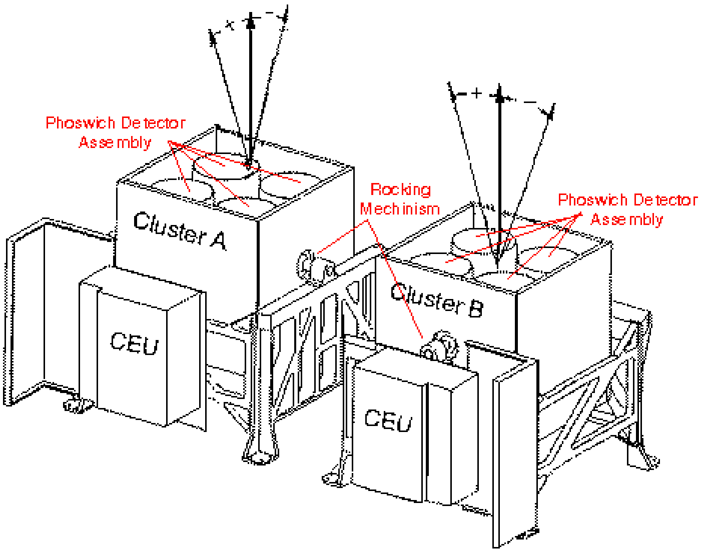


Figure 2.5: The HEXTE experiment and the spatial arrangement of the eight phoswich scintillators (taken from the web: `http://mamacass.ucsd.edu/hexte/pictures/`).

partially defines the overall FOV. Finally the thermal shield (layer like form) is situated in the slit mask and helps moderate temperature swings of the SShCs.

Sensitivity

In order to estimate the sensitivity of the ASM detector the SShC position histogram data have been simulated for a variety of X-ray source intensities, assuming that the source is in the center of the FOV, the SShC dwell time is 100 sec (the interval of the data accumulation lasts typically ∼ 90 sec), the position histograms include events from the full instrumental energy 2–10 keV and the total background is 49 counts $\mathrm{sec}^{-1}\mathrm{SShC}^{-1}$. The fig.2.7 shows the probability of obtaining a detection with a significance above 3σ as a function of X-ray intensity, the latter expressed in mCrab (1 mCrab$\simeq$1.06 μJy at 5.2 keV and a 10 mCrab source is estimated to produce a net count rate of 0.91 counts $\mathrm{sec}^{-1}\mathrm{SShC}^{-1}$). These simulations indicate that X-ray sources with intensities above 20 mCrab are detected at or above the 3σ significance level more than 90% of the time in a series of 100 sec dwells.

In practice, X-ray sources are not always centered in the FOV, and also for fields in the Galactic bulge there are likely to be a number of sources in the FOV at one time. On the other hand, these factors will be compensated, to some degree, by the fact that a source will appear in several dwells per celestial scan, given the geometry of the 3-SShC system and the nominal plan of 6° rotations between the dwells.

Energy Resolution

Detailed spectral studies of the registered events are not possible to be conducted with the ASM due to the lack of energy resolution. The whole experiment is able to categorize events in only three energy channels (1.3–3 keV, 3–5 keV and 5–12 keV) resulting a really poor energy resolution of the overall system.

2.1.5 The EDS interface

The EDS allows the observer to recover the most desired information by providing multiple-data processing options (from observations taken by PCA and ASM) on board the RXTE spacecraft. The EDS consists of three major types of subsystems: 6 PCA EAs, 2 ASM EAs, 2 system managers and 1 power distribution board. The data system can tag the relative arrival time of each event with an accuracy of 1μsec.

The large effective area and excellent time resolution capability of the PCA creates a great strain on telemetry resources. If the date were to be telemetered simply as 48 bits (i.e. time, detector and energy information) per individual PCA event, the nominal sustained PCA telemetry rate limit of ∼ 20 kb sec^{-1} would be exceeded for average total count rates above ∼ 420 counts sec^{-1}. This limit is equivalent to an X-ray source with intensity of only ∼ 50 mCrab and makes no allowance for non source background. Thus

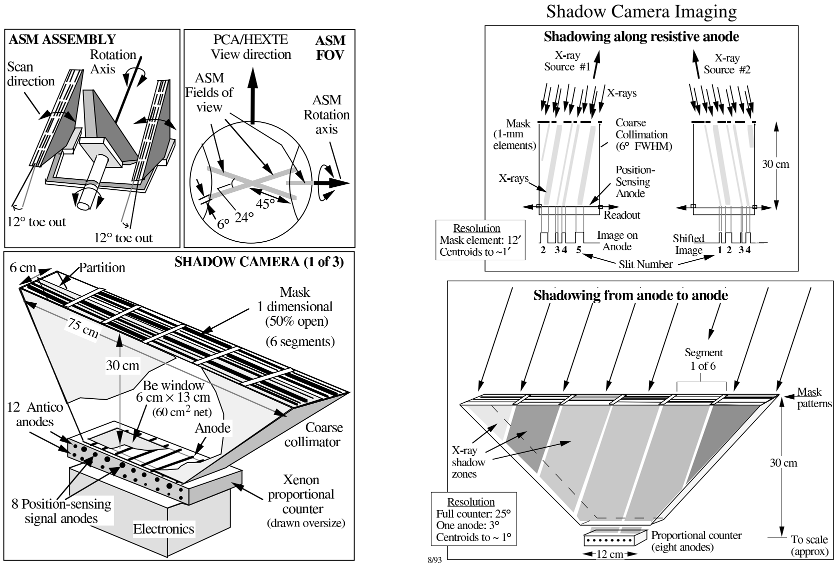


Figure 2.6: Schematic representation of the ASM and one SShC (taken from Schlegel 2006).

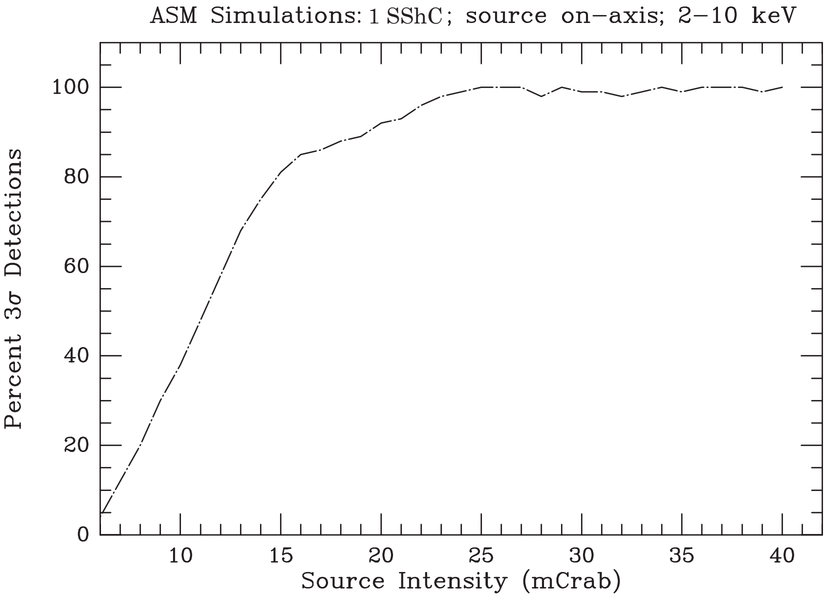


Figure 2.7: First order estimate (for fields without bright sources) of the ASM's detection threshold per celestial scan which nominally corresponds to a timescale of $\sim$ 90 min, i.e. 1 satellite orbit (taken from Schlegel 2006).

the EDS serves to compress the PCA data by means of time and energy binning. All the events detected from the PCA are transferred to the EDS without any selection criteria. There each PCA EA receives data from the five PCUs where the time tagging (down to 1μsec) is added as well as other reduction procedures (e.g. pulse profile averaging, Fourier power spectra) are carried out.

Similarly for the ASM the amount of raw information on an event by event basis is enormous to be telemetered directly to the ground. At 28 bits per event (i.e. not considering the time bits) the diffuse X-ray background emission ($\sim$ 40 counts SShC^{-1}) generates a raw data of 3.360 kb sec^{-1} which is above the nominal ASM telemetry rate of $\sim$ 3 kb sec^{-1}. Therefore in a similar way as before EDS compress the ASM data in two ways. The first ASM EA accumulates the data in the form of position histograms and the second one produces time series, background and pulse height spectra data products.

2.2 Data Reduction Procedures for the Proportional Counter Array

2.2.1 Data reduction software

The software used for reducing the PCA data is the HEAsoft v.6.0 provided by GSFC/NASA[1]. It is explicitly designed for research and advanced studies of scientific astronomical observations in *flexible image transport system* (FITS) format. The software consists of a collection of commands, tools and scripts coming from three major programs. The first program is the FTOOLS (Blackburn 1995) being appropriate for manipulating FITS files. This program contains also a file browser/editor/plotter with a graphical user interface, fv (Pence et al. 1997) as well as a library of C and Fortran subroutines for reading and writing data files in FITS data format, CFITSIO (Pence 1999). The second program is the XANADU which is appropriate for processing the data and producing scientific products. It consists of a spectral analysis program xspec (Arnaud 1996), a timing analysis program xronos (Stella & Angelini 1992) and an X-ray imaging program ximage (Giommi et al. 1992). Finally the last program comprising the HEAsoft collection is XSTAR (Bautista & Kallman 1999) appropriate for calculating the physical conditions and emission spectra of photoionized gases.

2.2.2 Data quality selection criteria

The PCA data are reduced based on the recommended criteria provided by GSFC/NASA[2]. Initially all the observations obtained with the EDS configuration *Standard-2* (Std2) mode are selected using the FTOOL `xdf`. The Std2 configuration is a type of binned mode running for all observations. Files containing Std2 data are in science array format (i.e. data binned at regular intervals by the spacecraft electronics). The science data are comprised by thirty columns of pulse-height histograms (i.e. energy spectra), one for each anode of each PCU, accumulated every 16 seconds.

For each Std2 observation a filter file is created using the FTOOL Perl script xtefilt containing information about the housekeeping data and the status of the satellite as well as its instruments within each observation. These files are then combined into a total one using the FTOOL fmerge listing as a function of time the values of various parameters such as: observing activity of all PCUs, Earth occultations, pointing offsets, high electron contamination and satellite passages from the *south Atlantic anomaly* (SAA).

The final step in order to select the optimum observations (i.e. observations taken under optimum conditions) is to choose those Std2 observations for which the all the aforementioned parameters are fulfilling some optimum observing criteria at the same time. The *good time interval* file (GTI), produced by the FTOOL maketime, contains exactly a list of times when all the parameters are within their allowed range at the same time. The range for each parameter is specified based on the following criteria:

- The elevation angle between the Earth's limb and the source should be grater than 10°.
- The offset between the source position and the satellite pointing should be grater than 0.02°.
- The time since SAA should be grater than 30 minutes. The SAA is caused due to the displacement of the center of the Earth's magnetic field with respect to its geographical center (by 280 miles) as well as the displacement between the magnetic and geographic poles of the Earth[3]. For satellite orbits

---

[1] `http://heasarc.gsfc.nasa.gov/ftools/`

[2] `http://heasarc.gsfc.nasa.gov/docs/xte/xhp_proc_analysis.html`

[3] The location of the north magnetic pole in 2005 was 82° 07′ North, 114° 04′ West, near Ellesmere Island, in Canada. Since 1970 its rate of motion has been accelerated from 9 km/year to 41 km/year and if it maintains its present speed and direction it will reach Siberia around 2055.

tilted between 20°–45° against Earth's equator and orbiting altitudes of $\sim$ 550 km (such as RXTE see tab.2.1), SAA becomes important because the spacecraft in those orbits periodically pass through a zone of a very high particle flux compared to the usual one outside this zone.

- Since the PCA is subject to contamination from electrons trapped in the Earth's magnetosphere or from solar flare activity, events based on this nature should be rejected in order not to be misinterpreted as source events. Such electrons, measured by the coincidence of events between the PCUs' antico propane layer and either of the two anodes in the first layer, increase the background at low energies. The parameter characterizing this activity is called ELECTRON2 and it is set to be smaller than 0.1 (see also sect.2.2.5).

Thus, only the Std2 observations which have been obtained within these "good times" are selected and these observations are those ones from which the scientific products are going to be extracted.

2.2.3 Background estimation

Previous proportional counter experiments in Earth orbit had successfully modeled the instrument background as a function of one parameter e.g. the A2 experiment on board the *first high energy astrophysical observatory* (HEAO-1) (Boldt & Garmire 1975) or several parameters e.g. Ginga (Hayashida et al. 1989). Some experiments such as the (EXOSAT) had also employed offset detectors (Taylor et al. 1981). While some early concepts for the PCAs included offset or movable detectors, the flight detectors are all fixed on the RXTE spacecraft and they are all pointing along the science axis (see fig.2.1). While the sky background is assumed to be constant at any one pointing position, the internal background may vary as the detectors move through different ambient conditions. Throughout the entire mission the background model is continuously evolving around a basic scheme. In the current implementation it consists of three parameters:

1. particle-induced background, arising from interactions between radiation or particles in the orbital environment with both the detector and the spacecraft.
2. the diffuse X-ray background emission entering through the collimator as X-rays.
3. the activation in the SAA.

The particle induced components are modeled by analyzing Earth data from the entire mission. The main particle component is found to be well correlated with many of the PCA anticoincidence rates, which are used to flag and veto non X-ray events. Many such rates are available and the current model uses the so-called Q6 rate, which measures coincidences that activate six of the eight PCU anode chains. In this way GSFC/NASA provides two parametrized files of the background one for the "faint" (<40 counts sec^{-1}PCU^{-1}) and one for the "bright" sources (>40 counts sec^{-1}PCU^{-1}). In addition to this primary particle component the satellite's passage through SAA induces radioactive decay terms into background. These can be accounted for by correlating the positional history of the satellite with the PCA count rate during occultations. The sky background is determined by subtracting the predicted particle components from "blank sky" pointings, averaging over several positions.

The final step for the data reduction is to create the background files corresponding to the selected "good time" Std2 observations. For this purpose the FTOOL `pcabackest` is used having as a primary function the estimation of the variable internal background component of every observation. In addition this FTOOL is capable of adding a constant term to approximate the sky background (coming from the aforementioned averaging procedure). It is assumed that the internal background varies on a relatively long time scale with respect to the 16 sec time resolution of the Std2 PCA observations. The basic idea is that the model can be evaluated once every 16 sec, and from that an equivalent of a Std2 data file can be created for each anode chain of each PCU. At the end for every Std2 data file a background file is produced with the same time duration as defined by the GTI files.

2.2.4 Scientific products

Light Curves

This is the last step for the production of the light curves and spectra. By making use of the FTOOL `saextrct`, the "source plus background" the light curves of the source are produced from the selected Std2 files together with the "background" light curves coming from the background files (sect.2.2.3).

All PCUs have three xenon layers (X1, X2, X3) each one consisting of two anode chains: left (L) and right (R) (tab.2.3). The 1st layer (XL1–XR1) detects roughly 90 percent of the cosmic photons and 50 percent of the internal instrumental background (tab.2.4). Therefore in order to improve the S/N (fig.2.4) events from the second and the third layer are excluded and only those from the first one are used in order to produce the scientific products. Finally the source light curves are produce after subtracting the two aforementioned light curves with the xronos command lcmath. In order to plot the light curves and save them in an ascii form file, the xronos routine lcurve and the xronos Perl script flc2ascii are used respectively.

Spectra

The production of the source spectrum is done in a similar way firstly by producing separately the "source plus background" spectrum and the "background spectrum" , and then disentangling the two. Note here that the latter is created based on the chosen parametrized background model file. In order to disentangle these two spectra and get the source spectrum alone the detector's response is necessary. The PCA response matrices are notionally divided in two parts: the *ancillary response file* (ARF), which accounts for the detector windows and collimator response and the *redistribution matrix function* (RMF), which accounts for the redistribution of photon energy amongst detector channels by the detecting medium. The product of the two is known as the response matrix (RSP) and this is used by the xspec program in order to produce the source spectrum. The PCA response depends on primarily three parameters[4]:

- Gain settings: The overall gain settings of the PCA have been changed five times since launch for operational reasons. These changes are significant and therefore data from more than one gain epoch should not be analyzed spectrally with the same RSP. It should be stressed here the fact that it is preferable to create the RSPs for all the observations, even if they are falling in the same epoch, due to the fact that there are small differences of the order of 2.2% due to small mission malfunctions (e.g. voltage fluctuations[5]).
- PCU ID: Individual PCUs have slightly different gains and moreover not all PCUs may be on at the same time during the data acquisition. Therefore, it is important to match the detector IDs of the response with those of the data.
- Anode ID: To boost the S/N, the spectra are extracted from just the top layer (anodes X1L, X1R) of the PCUs (see tab.2.4). The anodes included in the response should match those of the data.

For these reasons the Perl script pcarsp is employed in order to create from the "source plus background" spectrum file the appropriate ARF and RMF products and combine them to the corresponding RSP. This script uses also the *calibration database* (CALDB) provided by HEASARC[6] in order to consider for the estimation of the RSP the correct gain settings of the satellite during the corresponding observing period.

In order to fit the resulted source spectrum one can choose between additive or multiplicative spectral models of xspec. The selected model is then combined with the response matrix and the results of the fitted parameters are given together with the goodness of the fit (i.e. χ^2 and NHP). In AGN astronomy among the most common models in use are the *power-law* and the *broken power-law.* These models should be usually combined with an interstellar absorption photoelectric model in order to reflect the internal properties of the source which are independent from the intermediate material. The xspec uses for the effective absorption cross section per hydrogen atom as a function of energy (in 0.03–10 keV) the model of Morrison & McCammon (1983). The $N_{\rm H\,I}$ for a given source can be left as a free fit parameter or it can be specified by using the FTOOL nh. The latter program is using the H I map of Dickey & Lockman (1990) and requires as inputs the source's coordinates i.e. right ascension and declination.

2.2.5 Unfiltered background events

The background flaring activity, mainly caused by electrons, in sub-hour timescales can create some times artefacts which can be misinterpreted as source events. Low-energy electrons may be registered as X-ray events when they fail to trigger any of the vetoing logics of the detector (sect.2.1.2). The course of these

---

[4] A fourth parameter could be considered the channel binning. Although the response does not in fact depend on how the spectrum is binned, xspec requires that the binning of the response and that of the spectral files ("source plus background" and "background") should be identical.

[5] The complete list with the mission events can be found in: `http://heasarc.nasa.gov/docs/xte/whatsnew/big.html`

[6] `http://heasarc.gsfc.nasa.gov/docs/heasarc/caldb/`

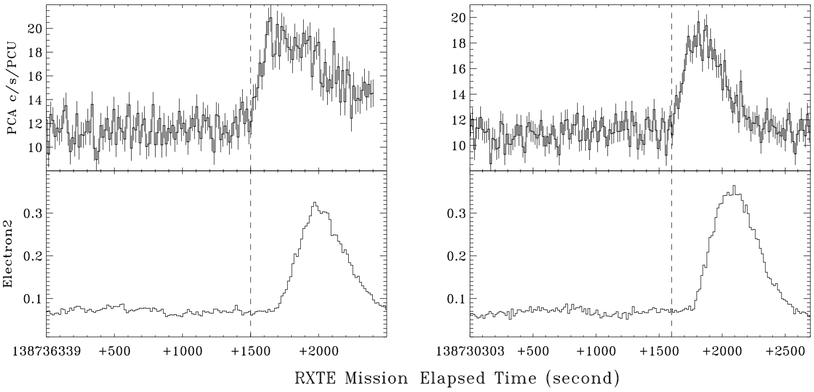


Figure 2.8: [Left panels] The upper plot shows the sub-hour flaring event of Mrk 501 during 1998 May 15–29 as reported by Catanese & Sambruna (2000) and the bottom plot shows the simultaneous behavior of the background electron events (denoted by Electron2). The vertical dashed line indicates the 200 sec lead of the X-ray flare with respect to the electron flare (taken from Xue & Cui 2005).
[Right panels] A similar X-ray event, occurring ∼ 1.5 h earlier, induced also by an electron flare (taken from Xue & Cui 2005).
Both light curves give the count rate in the energy range of 2–10 keV and the RXTE *mission elapsed time* (MET) is defined as the number of seconds since 1994 January 1 00:00:00 (UT).

events throughout an observation can be checked simply by plotting the parameter ELECTRON2, from the filter file (sect.2.2.2), as a function of time. Source flaring events occurring together with electron events is an indication of artefact flares not attributed to the source. For the case of the BL Lac object Mrk 501 a sub-hour flare was reported by Catanese & Sambruna (2000) which has been proven to be induced by an electron event by Xue & Cui (2005). The upper panels of fig.2.8 show two such events occurred during these observations of Mrk 501 between May 15–29, 1998 and the bottom panels show the corresponding behavior of the electron contamination parameter during these events.

In order to take into account these kind of unfiltered background events together with the X-ray light curves the ELECTRON2 parameter is plotted as a function of time and a thorough check is conducted for event correspondence. The X-ray events which appear to lead an electron flare by 100-500 sec are excluded from the overall analysis.

2.3 Data Reduction Procedures for the All-Sky Monitor

The ASM observations are becoming publicly available[7] by the ASM team at CSR/MIT in two forms. The first form is the *dwell observations* in which all the data points are given with a time resolution of ∼ 90 sec. Each point represents the fitted flux originating from each ASM dwell (see sect.2.1.4) and it is given together with a flux error and the observation time (in MJD). Together with each point a set of additional information is also available in order to check about the dwell quality. This consists of the number of sources in the FOV, the Earth's angle with respect to the ASM axes, the exposure time, number of SShCs in operational mode, the reduced χ^2 of the flux fit, the long-axis angle θ and the short-axis angle ϕ.

The second form is the one-day average observations in which each data point represents the one-day average of the fitted source fluxes coming from selected ASM dwells and it is given together with an a flux error and the mean observation time (in MJD). The selection criteria applied in order to choose the optimum dwells are the following: The number of sources in the FOV should be less than 16, the earth angle should be greater than 75°, the exposure time should be greater than 30 sec, the reduced χ^2 of the fit should be less than 1.5, the long-axis angle should have an angle θ between $-41.5°$ and $46°$ and at the same time the short-axis should be lying in an angle ϕ between $-5°$ and $5°$.

---

[7] `http://xte.mit.edu/`

Chapter 3

TIME SERIES ANALYSIS IN ASTRONOMY

Most of the time series analysis methods applied in astronomy are appropriate only for linear systems. However there is no reason why multi-component complex astrophysical systems such as AGN should generate emission through linear processes. The assumption of linearity is often made more for mathematical and computational convenience rather than it is believed to be true. In this chapter both the methods of linear and nonlinear time series are described in a compact and comprehensible way. Concerning the former there is a rich and extremely well written literature (e.g. Priestley 1981; Bendat & Piersol 1986; Brockwell & Davis 2002; Chatfield 2003) that focuses on various levels of both theory and applications allowing the reader to easily understand through examples the applicability of a given method. Moreover most of the times *ready-to-use* computer routines are also provided which can be then applied directly on a given data set. Concerning the nonlinear methods the literature is much more dispersed and untidy. The main sources of information, concerning a given method, are spread out in a extremely wide field of disciplines: physics, topology, economics, medicine, biology etc. In this chapter a try is made to present in a simple and comprehensible way the various nonlinear analysis tools giving a *step-by-step* methodology on how they can be applied directly on a data set. Additionally the physical interpretation of the results is discussed in depth helping to understand further when and why a given method should be applied and what are the capabilities/limitations of it. Despite the fact that all the nonlinear time series analysis methods make use of numerous topological theorems, based on a strict mathematical nomenclature, this chapter aims more to present the physical connection of these theorems with the given analysis methods rather than to go into the mathematical implications of them.

3.1 Basic Categories of Physical Systems

The physical systems can be broadly classified as being either deterministic or nondeterministic. The former are those which can be described by a small number of explicit mathematic relations (usually ≤ 15), low-dimensional systems (see sect.3.4.1) and their course in time can be readily predicted. In contrast the latter can not be usually described by exact mathematic relationships or the number of the equations is prohibitive in order to perform any analytical computation concerning their time evolution. The data sets yielding from both systems are named after them (i.e. deterministic data set, nondeterministic data set) and can be used in order to classify them further in smaller categories.

The general classification scheme of deterministic systems, is given in fig.3.1 after Bendat & Piersol (1986). The basic discriminating tool used to categorize the various types of deterministic systems is the power spectral density (PSD) (sect.3.3.1) of their data sets. Data representing deterministic phenomena can be categorized as being either *periodic* or *nonperiodic*. Periodic can be further categorized as being either sinusoidal or complex periodic. Nonperiodic data can be further categorized as being almost periodic or transient. In *sinusoidal* data, coming from *sinusoidal* systems e.g. the position of an unbalanced rotating weight as function of time, the PSD consists only of a single line at the cyclical frequency of the system. *Complex periodic* data e.g. a vibrating string, present various spectral lines in their PSD in frequencies

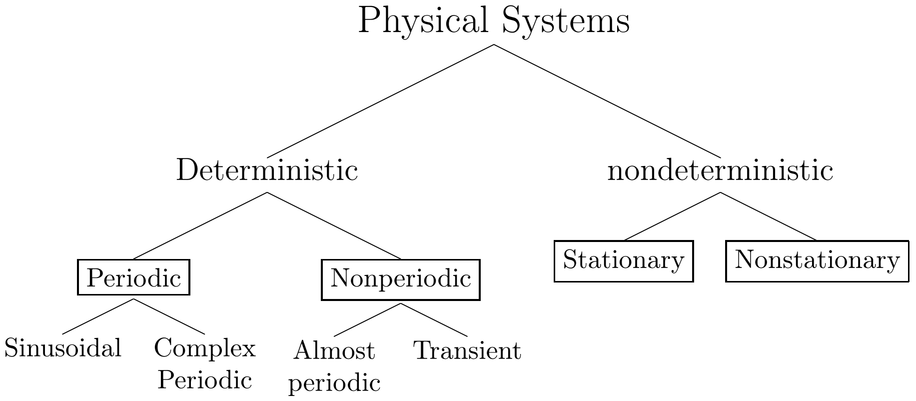


Figure 3.1: The classification scheme of deterministic and random systems (after Bendat & Piersol 1986).

that are all integral multiples of a fundamental frequency (harmonics). Conversely, *almost periodic* data e.g. systems mixing two or more unrelated periodic phenomena, exhibit a PSD also with lines but with no fundamental frequency (i.e. the fundamental period is infinitely long)[1]. Finally the *transient nonperiodic* data can be still explained by means of an explicit relation but this time the PSD is a continuous function. This means that the spectral power is not produced by some discrete frequencies but rather by a continuous range of frequencies. A data set exhibiting this kind of behavior can represent the free vibration of a damped mechanical system after removing the excitation form.

Usually the outcome of a nondeterministic physical systems is a random (or stochastic) data set in the sense that each observation of the phenomenon is unique and can not be described by few explicit independent variables (< 15), high-dimensional systems (see sect.3.4.1). Roughly speaking the number of independent variables is the same with the number of equations needed in order to fully describe the system. Random processes in nature are extremely common and they are met usually in systems comprising of multiple elements (e.g. particles). The output voltage of a resistor as a function of time and the microscopic pressure exerted in the walls of a container containing a gas are some of the systems producing random (or noisy) processes. Of course all the systems in nature can be described by a finite number of equations and in this sense all the systems can be seen as deterministic but this representation is computationally not feasible and practically not interesting for stochastic systems. Consider that in order to describe classically the behavior of 1 mole of He, having $\sim 6{\cdot}10^{23}$ atoms, one would need a set of $3 \times 6 \cdot 10^{23} + 3 \times 6 \cdot 10^{23}$ equations for the position and the velocity of the atoms respectively in order be able to fully describe the system. Practically speaking the behavior of a deterministic system described by more than 15 independent variables converges to that of a stochastic system. Stochastic data sets yielding are random in nature and must be described in terms of probabilistic statements and statistical averages. The following classification (fig.3.1), after Bendat & Piersol (1986), aims to categorize the various random data sets based on their statistical properties. In the statistical literature concerning the time series the terms random, stochastic, nondeterministic and noise processes/data sets are used exactly in the same context.

Consider a collection of light curves $\{x_n(t)\}$, produced by a random process $x(t)$ with n denoting the $\mathrm{n^{th}}$, output of the process (fig.3.2). The mean value of the realization of the random process at some time t_1 can be computed by taking the instantaneous value of each time series of the ensemble at time t_1.

$$\langle x(t_1)\rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n(t_1) \tag{3.1}$$

In a similar way, the autocovariance function between the values of the random process at two different times can be computed by taking the ensemble average of the product of instantaneous values at two times t_1 and $t_1 + \tau$.

$$V_{x,x}(t_1, t_1+\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left[x_n(t_1) - \langle x(t_1)\rangle\right] \left[x_n(t_1+\tau) - \langle x(t_1+\tau)\rangle\right] \tag{3.2}$$

For the general case where $\langle x(t_1)\rangle$ and $V_{x,x}(t_1, t_1+\tau)$ vary as time t_1 varies then the stochastic process $x(t)$ is said to be nonstationary. For the special case where these quantities remain constant as time t_1 varies,

---

[1] Specifically, the sum of two or more sinusoids will be periodic only when the ratio of all possible pairs of frequencies form rational numbers.

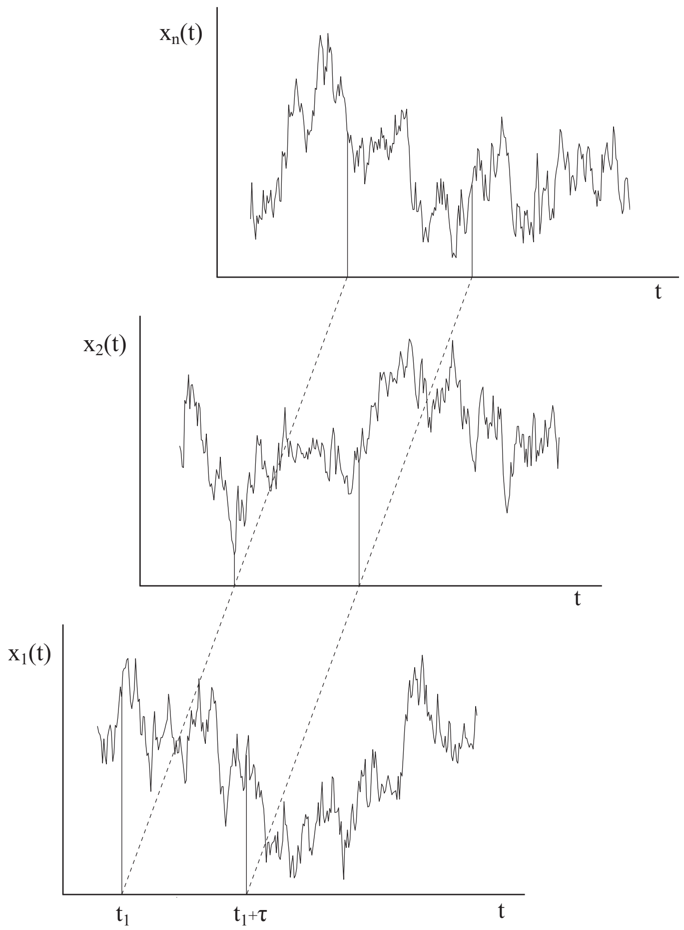


Figure 3.2: Collection of light curves (time series) defining a random process (after Bendat & Piersol 1986).

$\langle x(t_1)\rangle = \langle x(t)\rangle$ and $V_{x,x}(t_1, t_1+\tau) = V_{x,x}(\tau)$, the process is said to be *weakly stationary* or *stationary* in the wide sense. Finally when all possible moments and joint moments are time invariant the stochastic process is said to be *strongly stationary* in the strict sense.

Therefore the statistical properties of a random process can be determined by computing ensemble averages at specific time instants. Nevertheless in most of cases it is possible to describe the properties of a stationary random process by computing time averages over a specific time series (i.e. a light curve obtained in a specific time period T). For example consider for the $\mathrm{k^{th}}$ output of a random process, the mean value and the autocovariance function of this sample are given respectively by the following relations

$$\langle x_k(t)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T x_k(t)dt \tag{3.3}$$

and

$$V_{x_k,x_k}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \left[x_k(t) - \langle x_k(t)\rangle\right]\left[(x_k(t+\tau) - \langle x_k(t+\tau)\rangle\right] dt \tag{3.4}$$

If the random process $x(t)$ is stationary and both $\langle x_k(t)\rangle$ and $V_{x_k,x_k}(\tau)$ do not differ when computed over different sample functions (e.g. light curves) the random process is said to be *ergodic*. For ergodic random processes the time-averaged mean value and autocovariance function (as well as all the other time-averaged properties) are equal to the corresponding ensemble averaged values, that is $\langle x_k(t)\rangle = \langle x(t)\rangle$ and $V_{x_k,x_k}(\tau) = V_{x,x}(\tau)$. Note here that only stationary random processes can be ergodic and as such is the most important class of random processes since all the statistical properties can be determined by performing time averages over a single time series produced by the observed system.

3.1.1 Noise processes

The PSD (sect.3.3.1) of a random system is well represented by a power-law $\mathcal{P}(f) \propto f^{-a}$ with $0 \leq a < 3$ (see eq.3.10). Usually a special name for the noise process is given after the value of a.

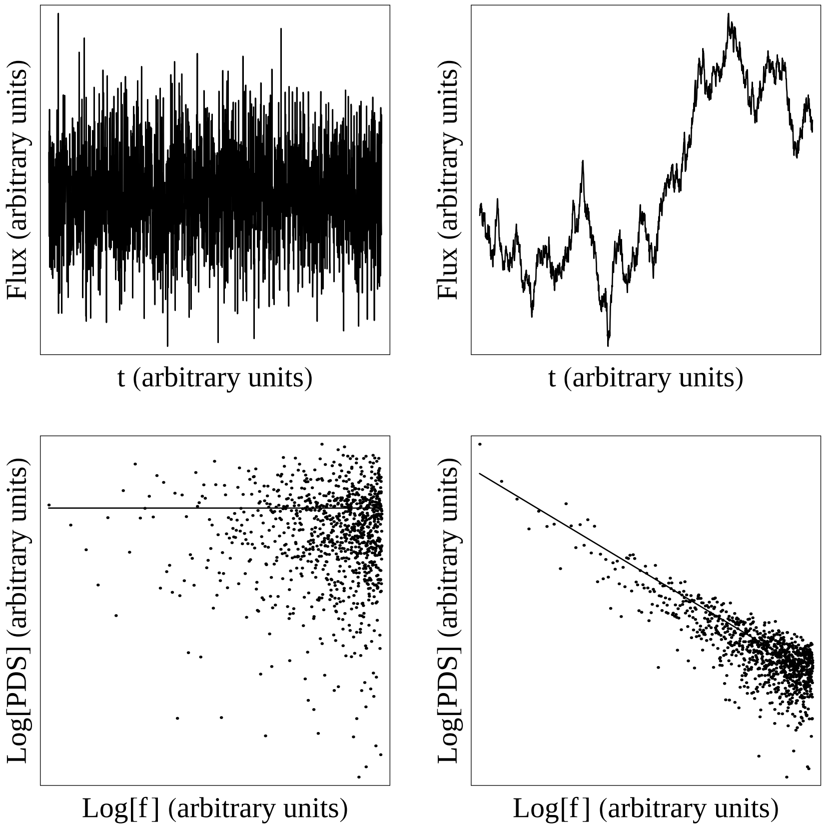


Figure 3.3: [Top panels] The left light curve represents a white noise process produced by a sequence of IID variables (2000 points). The right light curve represents a red noise process produced by summing cumulatively the outcome of the white noise process.
[Bottom panels] The periodogram estimates (eq.3.11), represented by the points, of the white and the red noise processes respectively (left-right plots). The solid line represent the underlying PSD (eq.3.10) going as f^0, in the case of the white noise, and f^{-2} in the case of the red noise.

- *white noise* for a=0 (fig.3.3). If $\{x_t\}$ is a sequence of uncorrelated random variables each with zero mean and variance S^2 then $\{x(t)\}$ is stationary having an autocovariance function $V_{x,x}(t, t+\tau) = S^2$ for $\tau = 0$ and $V_{x,x}(t, t+\tau) = 0$ for $\tau \neq 0$ independent of t. Such a sequence is referred to as white noise: $\{x_t\} \sim \mathrm{WN}(0, S^2)$. Data sets in which the observations are independent and identically distributed (IID) random variables are white noise processes but the opposite is not true. The first order autoregressive conditional heteroscedasticity model (ARCH) (Engle 1982) is a strictly stationary white noise process but it is not an IID sequence since it is not Gaussian. Only a strictly Gaussian white noise process consists of a sequence of IID variables.

- *red noise* (or random walk) for a=2 (fig.3.3). The red noise process $\{s_t\}, t = 0, 1, 2, \ldots$ with $s_0 = 0$ is obtained by cumulatively integrating IID random variables $\{x_t, t = 0, 1, 2, \ldots\}$: $s_t = x_1 + x_2 + \ldots x_t$ for $t = 1, 2, \ldots$ and it has a mean value of zero and variance tS^2. Since the autocovariance function: $V_{x,x}(t, t+\tau) = tS^2$ depends on time the process is nonstationary. A random walk process, consisting of a sequence of discrete steps of fixed length in one dimension (Brownian motion), is known to have an average distance between N successive steps equal to $\sqrt{2N/\pi} \approx \sqrt{N}$ (e.g. Harwit 2006).

There is a lot of perhaps unnecessary nomenclature concerning the “colors of noise” (pink, orange, green etc.). In practice, in AGN astronomy PSD with $1 < a < 3$ usually are mentioned as red noise spectra (Green et al. 1993; Lawrence & Papadakis 1993; Vaughan et al. 2003). These kind of spectra appear in all the BL Lac objects depicting exactly the fact that all the time scales are mixed in such away resulting an aperiodic behavior (no frequency is favored) with decreasing variability amplitude as the time scales become shorter (larger frequencies).

3.1.2 Noise in the time series

It is desirable for every time series analysis method to have the biggest possible data set one can get. The bigger the data set is the more information about the underlying mechanism can be extracted since more accurate estimations of its statistical properties can be derived. Nevertheless, even if the problem of few data points is solved there is the issue of the noise.

There are two types of noise in time series depending on whether or not they affect the feature outcome of the process it self.

- Observational noise: It originates from the instruments (e.g. CCDs, scintillators) that are used in order to register the signal. All the instruments have a given precision specifying a corresponding measurement uncertainty. The observational noise also includes the uncertainties due to photon statistics. Since the majority of astronomical detectors measure number of N events in a specific time interval δt there is an uncertainty associated with these measurements specified by the Poissonian statistics, $1/\sqrt{N}$. This kind of noise does not affect the feature evolution of the time series. Especially in the astrophysical systems the latter is affected only by the underlying dynamical process which is unrelated to the observational method.

- Dynamical noise: It originates from the dynamical process itself affecting the state of the system and hence altering its future course in time. By the time it is embodied into the realization of the process it comprises an inextricable part of the data set and has to be taken into account for the future evolution of the system. In order to see how dramatically the variable behavior of a dynamical system can change due dynamical noise check the example in sect.3.2.1.

These two kind of noises $n(t)$ are annexed to the time series in an additive or in a multiplicative way. Usually the parent distribution of the noise component is unknown and a Gaussian distribution is then considered with a given mean value and variance specified from the data set. One should always test the sensitivity of the results with respect to the chosen distribution if the latter is not known *a priori*.

3.1.3 The influence of the observational noise

The errors of the data points in the PCA/ASM light curves (sect.2.2.4, sect.2.3), are symmetric around the actual estimates, meaning that the distribution of the parent distribution is considered to be Gaussian. The question now is: How one can take into account these errors during the application of a statistical method in order to check how much they affect the result of the given method? By taking into account these errors one can consolidate the significance of a statistical result based on the quality of the data.

A neat and simple method to incorporate the measurements errors into a statistical method is through Monte Carlo simulations (Press et al. 1992). During each simulation every measurement is replaced by a randomly selected value coming from a Gaussian distribution having as mean the data point's actual value and as standard deviation its error. The same procedure is then repeated 1000 times and every time the same statistical method, intended to be used for analyzing the original data set, is applied to the surrogated light curves yielding an ensemble of results. Based on the distribution of the latter one can check how the measurement errors affect the method and up to which significance level the results are robust.

3.2 Linear, Nonlinear and Chaotic Dynamical Systems

A dynamical system provides a mathematical representation of a physical system based on a set of equations. The form of these equations describes the evolution of the physical system as a function of a single or multiple variables (e.g. time t, position (x, y, z)) characterizing at the same time its behavior. If the behavior of a system (no matter how much complicated and noisy is) can be analyzed as a sum of responses of simpler components then the system in principle can be characterized as linear. Strictly speaking a dynamical system is linear if the response characteristics are *additive* (superposition) and *homogeneous* (scaling). The term *additive* means that the output to a sum of inputs is equal to the sum of the outputs produced by each input individually. The term *homogeneous* means that the output to a sum of inputs is equal to the sum of the outputs produced by the input alone. Consider a system whose properties are described by an operator $\mathcal{L}$. Then the system's outcome in two valid inputs $x_1(t)$ and $x_2(t)$ is $y_1(t)$ and $y_2(t)$ respectively. Then the system is linear (the equation and the operator is called linear) if and only if

$$\left.\begin{array}{l}\mathcal{L}x_1(t) = y_1(t)\\ \mathcal{L}x_2(t) = y_2(t)\end{array}\right\} \Rightarrow \mathcal{L}\left[\lambda_1 x_1(t) + \lambda_2 x_2(t)\right] = \lambda_1 \mathcal{L}x_1(t) + \lambda_2 \mathcal{L}x_2(t) = \lambda_1 y_1(t) + \lambda_2 y_2(t) \tag{3.5}$$

Nonlinear systems include all the other dynamical systems that do not meet the requirements of the linear systems. Crudely speaking, a nonlinear system is one whose behavior is not simply the sum of its parts or their multiples and as such it is often difficult (or impossible) to model it. Moreover its evolution with respect to a given variable e.g. time, is sometimes difficult to predict or even impossible after a given time duration. Typical nonlinear systems are the simple pendulum, the Van der Pol oscillator, the Navier-Stokes equations etc.

A really interesting category of nonlinear systems are the chaotic systems. Chaotic systems have the unique ability to appear, through their realization, as random systems but in fact they comprise a special category of nonlinear systems characterized by a small number of independent variables (≤ 15) i.e. deterministic. One could give a loose definition by stating that a chaotic system is a pure deterministic nonlinear physical system with an apparent stochastic behavior.

3.2.1 Intermittency

A special category of nonlinear physical systems can exhibit segments of relative constant values (laminar phase) interspersed by erratic bursts. These sudden transitions from a low quiescent state to a higher more variable one are known as "intermittent" events and the phenomenon as such is called *intermittency* (Pomeau & Manneville 1980). Intermittent behavior can be triggered by either stochastic noise or nonlinearities and threshold exceedance (Platt et al. 1993). The "passage" to an apparent stochastic behavior is chaotic in nature and occurs aperiodically with no fixed duration.

In order to get a more vivid idea about how such a behavior can occur, consider the nonlinear dynamical system of an asymmetric "double-well" potential $V(x_t)$ (fig.3.4), with x_t being its displacement from 0

$$V(x_t) = ((x_t - 6)^2 - 1.5)((x_t - 8.5)^2 + 0.1) \tag{3.6}$$

having equilibrium points, specified by the equation $dV(x_t)/dx_t = 0$, g_1=5.561 (stable) g_2=7.746 (unstable) g_3=8.444 (stable). Consider also that the system suffers the influence of a white noise process n_t coming from a Gaussian parent distribution with zero mean and unit variance, then the realization of the process can be described in a form of time series after integrating the evolution equation

$$\frac{dV(x_t)}{dx_t} = -\frac{dx_t}{dt} + kn_t \tag{3.7}$$

where kn_t is the noise component of the dynamical process with k being a factor tuning firstly the nature of the noise process i.e. additive or multiplicative (sect.3.1.2) and secondly its power. In the case of having an additive noise process, k =const., and its value determines the energy region (A or B in fig.3.4) where the system is going to be evolved.

For small values of k the system experience very rarely large deviations from g_1 (fig.3.5,(a)) that can result a transition from A to B and therefore it evolves almost always in region A. For bigger values of k the system is able to "jump" to the higher energy state region B but it can only stay there for relatively short time periods (fig.3.5,(b)). Exactly this sudden and short duration transitions of the nonlinear system from A to B are characterized as intermittent behavior and they are triggered entirely by the stochastic noise component. The physical reason for this short-time duration stay within region B is due to the fact that this energy state is not surrounded by high "energy walls", as region A, therefore under the influence of even small perturbations, coming from the noise process, the system can easily escape from there and return back to A.

In the case of multiplicative noise $k \propto x_t$ (fig.3.5,(c)) the situation is slightly different in the sense that this time the intermittent transitions from A to B occur more often, since as the displacement of the system increases towards the higher energy state B (on the right side of g_1), the power of the noise component gradually augments as well (this of course applies also for the left "wing" of g_1). Therefore the increasing course of the system is perturbated successively more and more from the noise component, and once the system exceeds g_2 it enters the energy valley B where there another stable equilibrium point exists, g_3. Thus the system can also be evolved around it for longer time duration, in comparison to the k =const. case, since now the noise component is much more active. This burst-like activity is another expression of the intermittent behavior of the source.

The histogram in fig.3.5,(d) shows exactly the number of x_t events that occur in every case for 10000 events of the processes. It can be seen that for the cases (b) and (d) the main difference is that for the latter there are much more intermittent events occurring with higher displacement due to the multiplicative nature of the noise process. All the aforementioned sudden transitions from A to B emerge from the combination of the nonlinear nature of the process (i.e. shape of the $V(x_t)$ eq.3.6, fig.3.4) and the noise component.

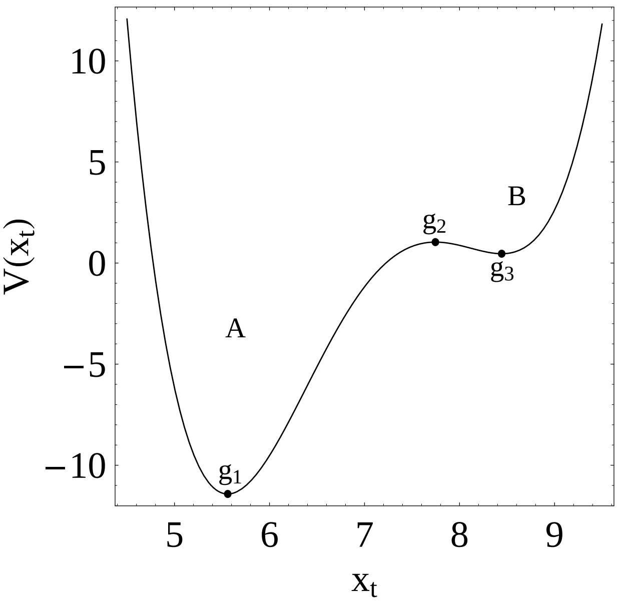


Figure 3.4: The asymmetric "double-well" potential $V(x_t) = ((x_t - 6)^2 - 1.5)((x_t - 8.5)^2 + 0.1)$ with x_t being its displacement from 0 and equilibrium points g_1=5.561 (stable) g_2=7.746 (unstable) g_3=8.444 (stable). The two valleys A and B define the two energetic levels of the system.

3.3 Linear Methods

3.3.1 The PSD

In general a continuous physical process can be described either in the time domain by the values of some equation X as a function of time t, e.g. $X(t)$, or else in the frequency domain where the process is specified by giving its amplitude Y as a function of frequency f, that is $Y(f)$. The functions $X(t)$ ($X : \mathbb{R} \to \mathbb{R}$) and $Y(f)$ ($Y : \mathbb{C} \to \mathbb{R}$, indicating phase also) are two different representations of the same process completely equivalent. These two representations are connected by means of the Fourier transform equations

$$Y(f) = \int_{-\infty}^{+\infty} X(t) e^{2\pi f t} dt \Leftrightarrow X(t) = \int_{-\infty}^{+\infty} Y(f) e^{-2\pi f t} df \tag{3.8}$$

In physics the primary way of describing a physical process is simply by representing a characteristic quantity of it (e.g. velocity, distance, electric current, temperature etc.) as a function of time, $X(t)$. Then, having this function it is possible to estimate $Y(f)$ and find whether or not some sinusoidal components are dominant.

The power contained in the frequency interval between f and $f + df$ is the sum of the modulus-squared of the sinusoidal amplitudes $Y(f)$ and $Y(-f)$.

$$\mathcal{P}(f) = |Y(f)|^2 + |Y(-f)|^2, \;\; 0 \leq f < \infty \tag{3.9}$$

The $\mathcal{P}(f)$ consists the power spectral density (PSD) of $X(t)$ and since $X(t) \in \mathbb{R}$

$$\mathcal{P}(f) = 2|Y(f)|^2 \tag{3.10}$$

The aforementioned method is appropriate to study a continuous real-valued process described by a continuous function $X(t)$ in the frequency domain. The next thing that one should query is the following: In case that $X(t)$ is given in a discretized form, $x(t)$, is it possible to estimate the PSD of the process described by $X(t)$? This is actually the usual case in astrophysics where a quantity (e.g. flux) coming from a continuous process is discretized due to observational sampling. Consider an evenly sampled light curve comprising a series of fluxes x_i with measurements errors $\sigma_{\mathrm{err},i}$ measured at discrete times t_i ($i = 1, 2, \ldots, N$) with a sampling time Δt. These measurements comprise a discretization of the continuous function $X(t)$ describing the emission process of the observed source.

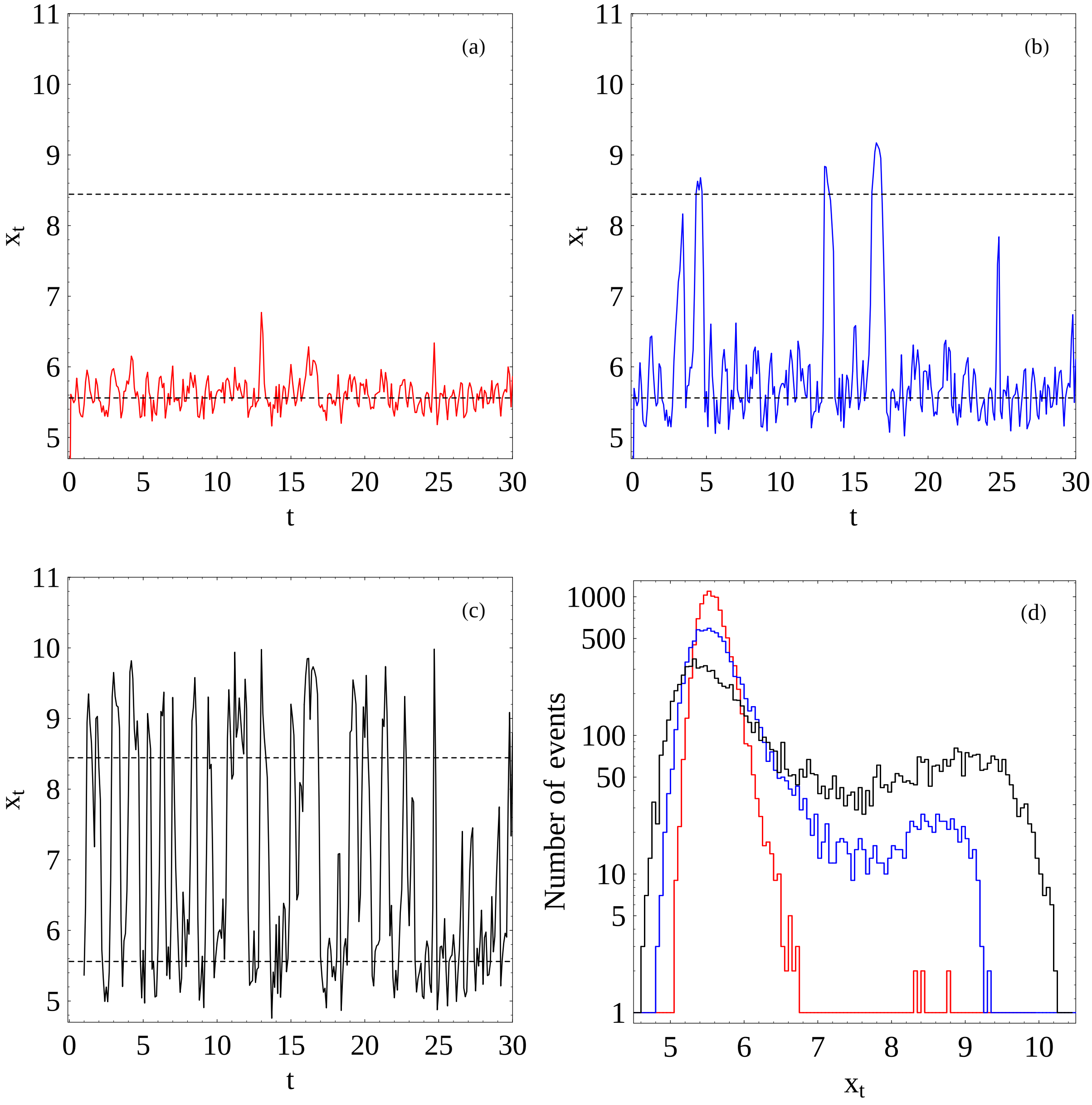


Figure 3.5: Time series composed from 10000 points after integrating eq.3.7 using a $\Delta t = 0.1$, for three different noise components kn_t. Here only the first 300 realizations are shown. The horizontal dashed lines indicate the two stable equilibrium points, g_1=5.561 and g_3=8.444.
[Top panels] (a) $k = 6$ (b) $k = 10$.
[Bottom panels] (c) $k = 3x_t$ (d) The histograms of the data sets (corresponding colors) for the complete data sets 10000 points long with a bin size of 0.05.

In this case the natural estimate of $\mathcal{P}(f)$ in a given frequency f is given by the periodogram, $P(f)$, being the modulus-squared of the discrete Fourier transform (DFT) of the light curve at $N/2$ evenly spaced frequencies $f_j = j/N\Delta t$ with j running from 1 to $N/2$ with $f_{N/2} \equiv f_{\mathrm{Nyq}}$ where f_{Nyq} is the Nyquist critical frequency.

$$P(f_j) = \underbrace{\frac{2\Delta t}{N}}_{C}\left|\sum_{i=1}^{N} x_i e^{2\pi f_j t_i}\right|^2 = C\left[\left|\sum_{i=1}^{N} x_i \cos(2\pi f_j t_i)\right|^2 + \left|\sum_{i=1}^{N} x_i \sin(2\pi f_j t_i)\right|^2\right] \tag{3.11}$$

Note here that the periodogram as an even function has the same shape between negative and positive frequencies. Therefore the parameter 2 in the normalization constant C is used in order to take into account the existing power in the negative frequencies, since the summation is dome only over the positive f_j, resulting the “one-sided” periodogram.

3.3.2 Binning the periodogram of a random process

In general the periodogram of a random process is a very fluctuating function and shows a great scatter around the underlying PSD (see fig.3.3). For a nondeterministic (see sect.3.1) Gaussian process the real and the imaginary parts of eq.3.11 follow a Gaussian distribution therefore the periodogram in a specific frequency is distributed as χ^2 distribution with two DOF, χ^2_2, around the PSD (Vaughan et al. 2003).

$$P(f_j) = \mathcal{P}(f_j)\frac{\chi^2_2}{2} \tag{3.12}$$

Neither by sampling a longer stretch of data at a given sample rate Δt nor by resampling the same stretch of data with a faster sampling rate ameliorates the accuracy of the periodogram estimates at a given frequency bin f_j. For the first case the reason is that more data points N simply produce estimates at a greater number of discrete frequencies f_j since $f_j = j/N\Delta t$. The Nyquist critical frequency remains the same but there is finer frequency resolution. For the second case with a finer sampling interval the Nyquist critical frequency extends to higher frequencies but the frequency resolution remains the same. This means that for real data the periodogram is an inconsistent estimator of the underlying PSD.

In order to reduce this scatter the periodogram must be smoothed in some fashion by averaging over the frequency bins. An average over M consecutive discrete frequencies reduces the scatter by a factor of $1/\sqrt{M}$. The main disadvantage of this method is that in order to have the distributions of the smoothed estimates close to Gaussian (appropriate for applying linear regression) one needs to average around 50 periodogram points. This leads both to severe bias in the lowest frequency points and to a substantial loss in frequency resolution. The optimum way to face this problem is to work with the logarithm of the periodogram rather than the periodogram itself (Papadakis & Lawrence 1993). The probability distribution function of the logarithm of the periodogram is more symmetric (i.e. Gaussian) than the distribution of the periodogram. Also the variances of the logarithm of the periodogram estimates are known, since they do not depend on the unknown value of the PSD. Finally especially for power-law like spectra $\mathcal{P}(f) \propto f^{-a}$ the estimates of the logarithm of the power density based on grouping of the logarithmic periodogram estimates are not biased due to the linear relationship between $\log(\mathcal{P}(f))$ and $\log f$. Briefly the method consists of the following steps.

- The periodogram $P(f_j)$ is estimated and the logarithm of it is computed, $\log(P(f_j))$. Then the logarithmic periodogram ordinates are grouped in n sets of size M and their average value is estimated

$$\langle \log(P(f_{\mathrm{avg},i})) \rangle = \frac{1}{M}\sum_j \log(P(f_j)) \tag{3.13}$$

  where the index j varies over M consecutive periodogram ordinates and the index i varies over the n subsets into which the logarithmic periodogram estimates are grouped. For the frequency $f_{\mathrm{avg},i}$ the geometric mean frequency of each group is considered

$$f_{\mathrm{avg},i} = \left[\prod_j f_j\right]^{1/M} \tag{3.14}$$

  By taking the mean value of the logarithm of eq.3.12 it is obvious that the logarithmic periodogram estimates are biased in the sense that $\langle \log(P(f_j)) \rangle = \langle \log(\mathcal{P}(f_j)) \rangle + \langle \log(\chi^2_2/2) \rangle$ although $\langle P(f_j) \rangle =$

$\mathcal{P}(f_j)$. However this is a constant bias $\langle\log(\chi_2^2/2)\rangle = -0.253$ (Abramowitz & Stegun 1970) therefore by adding 0.253 to each one of the logarithms of the periodogram estimates the resulted values consist unbiased estimates of the logarithm of the power spectral density

$$\log(\mathcal{P}(f_{\mathrm{avg},i})) = \langle\log(P(f_{\mathrm{avg},i}))\rangle + 0.253 \tag{3.15}$$

- The variance of the logarithm of the periodogram is independent of the frequency and known, in contrast with the variance of the periodogram estimates. By taking the variance of the logarithm of eq.3.12 this reads $\mathrm{var}[\log(\chi_2^2/2)] \approx 0.310$ (Abramowitz & Stegun 1970). In this sense the variance of the binned logarithmic periodogram

$$\mathrm{var}[\log(\mathcal{P}(f_{\mathrm{avg},i}))] = \frac{0.310}{M} \tag{3.16}$$

3.3.3 The variance and the PSD

The PSD of a continuous process in a given frequency f, $\mathcal{P}(f)$ reveals the distribution of the variability power corresponding to a given frequency (Vaughan et al. 2003). The integration of this function between two frequencies f_1 and f_2 yields the contribution to the expectation value of the intrinsic, "true, inherent", variance, of the under study system, due to variations between the corresponding time scales $(\frac{1}{f_1}, \frac{1}{f_2})$. As it is dictated from Parserval's theorem

$$\langle S^2\rangle = \int_{f_1}^{f_2} \mathcal{P}(f)df \tag{3.17}$$

Since the "natural" estimate of the PSD in the case of discrete data is the periodogram $P(f)$, the integrated periodogram yields the observed variance of the given data set consisting a realization of the process. Consequently, the total variance of a data set is equal to the value of the integrated periodogram over the frequency range (f_1, f_{Nyq}) in the theoretical limit of $N \to \infty$ (practically $N > 100$)

$$\sum_{i=1}^{N/2} P(f_j)\Delta f = S^2 \tag{3.18}$$

where S^2 is the *unbiased* estimator[2] of the variance of a data set x_i, with an arithmetic mean $\overline{x}$, and it is defined as

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2 \tag{3.19}$$

Also due to the fact that the periodogram in a specific frequency is distributed as x_2^2 around the PSD (eq.3.12), the variance is a sum of x_2^2 distributions weighted by the PSD

$$S^2 = \frac{1}{N\Delta t}\sum_{i=1}^{N/2} \mathcal{P}(f_i)\chi_2^2 \tag{3.20}$$

Thus while the expectation value of the variance is equal to the integrated PSD (eq.3.17) each realization of the process, given in a form of time series, might have a different variance even if the parent variability process is stationary (see sect.6.1). These fluctuations can be understood in a statistical sense since they are governed by the Gaussian statistical rules of noise processes. In the field of AGN this means that the statistical properties between two or more light curves obtained in different periods might be different but the physical parameters of the system responsible for the variability may remain the same. The method presented in sect.6.1 aims exactly to distinguish between random fluctuations resulted from a stationary and thus unaltered physical system and changes coming from genuine differences concerning the parameters of the physical process.

---

[2] Concerning the terms *unbiased* and *biased* estimator of the variance there is a long discussion going on in various statistical books (e.g. Bendat & Piersol 1986) which is absolutely necessary in order to consolidate the correct mathematical definition of the quantity but inessential for the statistical analysis of almost any real-life data set. The sample variance, having in the denominator N instead of $N-1$, is considered to be a "biased" version of the sample's variance since it requires an *a priori* knowledge of the mean, $\overline{x}$, rather than a direct estimation of it from the data set. Of course the impact of this matter concerning the pragmatic needs of any statistical method with adequate number of points is simply unimportant e.g. for 200 points the results differ by 0.002%.

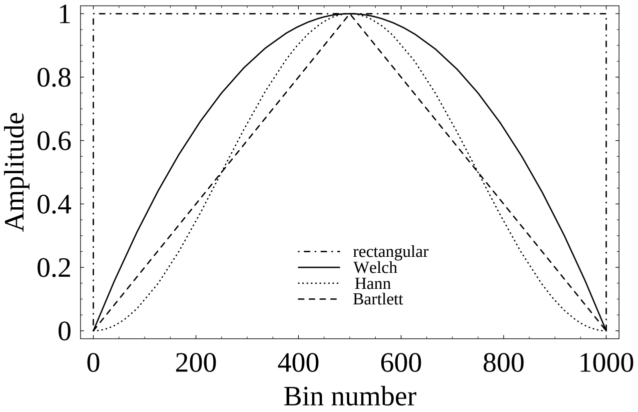


Figure 3.6: The most commonly used window functions in the PSD estimation (after Press et al. 1992).

3.3.4 Aliasing and windowing effects

The periodogram should be calculated up to Nyquist critical frequency $f_{\rm Nyq}$ since for higher frequency values f_a the estimated power is a combination of true power at f_a and its aliases $2f_a \pm f_{\rm Nyq}, 4f_a \pm f_{\rm Nyq}, \ldots$. This effect "folds back" in a way power from high frequencies to lower ones. Nevertheless in the case of a power-law-like PSD the spectral density decreases logarithmically as frequency increases so the *fold-back* power from low to high frequencies is small compared to the observed value. Moreover this aliasing effect affects only the high frequency part of the periodogram by flattening it which is dominated anyhow from the experimental noise. Finally the astronomical data are not simply sampled but they are averaged over specific time intervals Δt fact that further reduces the amount of the power which id aliased back.

Apart from the aliasing effects windowing effects are also present. Obtaining a sample of N observations x_i from a continuous process $X(t)$ and calculating the periodogram is effectively equivalent with multiplying a continuous function by a rectangular window function of time (fig.3.6), being 1 for t_i $(i = 1, \ldots, N)$ and 0 elsewhere. By the convolution theorem this "multiplication" is equivalent with the convolution of the data's Fourier transform with the window's Fourier transform. Since the latter turns "on" and "off" suddenly (i.e. in zero time) substantial components at high frequencies appear with an oscillatory lobe-form transferring power from low to high frequencies. This effect which is known as *spectral leakage* and it is affected mainly by the shape of the window function. Several types of window functions exist in general (Welch, Parzen, Barlett, Hamm, Blackman-Harris, Tukey, Kaiser etc.) but for the purpose of AGN time series analysis there is effectively no difference between them. A useful general rule concerning the window functions is that the smoother the window function goes from 0 to 1 and then back to 0, the less power "leaks" to adjacent frequency bins. An approximate formula giving the number of bins where the leakage is extended towards the one direction $n_{\rm offset}$ around a given frequency is the following (Press et al. 1992)

$$n_{\rm offset} \approx \frac{N}{\text{rise and fall time}} \tag{3.21}$$

Throughout this work, for the computation of the periodogram the Welch window function is used (fig.3.6) given by

$$w_j = 1 - \left(\frac{j - \frac{N}{2}}{\frac{N}{2}} \right)^2 \tag{3.22}$$

3.3.5 The SF

The use of the structure function (SF) in time series analysis was first introduced by Provenzale et al. (1992). The first order SF for a time series data set $x(t)$, represents a measurement of the mean squared difference between all the values, N, which are separated in time by τ (time delay), $(x(t), x(t+\tau))$. In simple words, it consists the mean estimate of the variance (eq.3.19) of the data points which are separated in time by τ.

$$SF(\tau) = \frac{1}{N} \sum_{i=1}^{N} (x_i - x_{i+\tau})^2 = \left\langle (x(t) - x(t+\tau))^2 \right\rangle \tag{3.23}$$

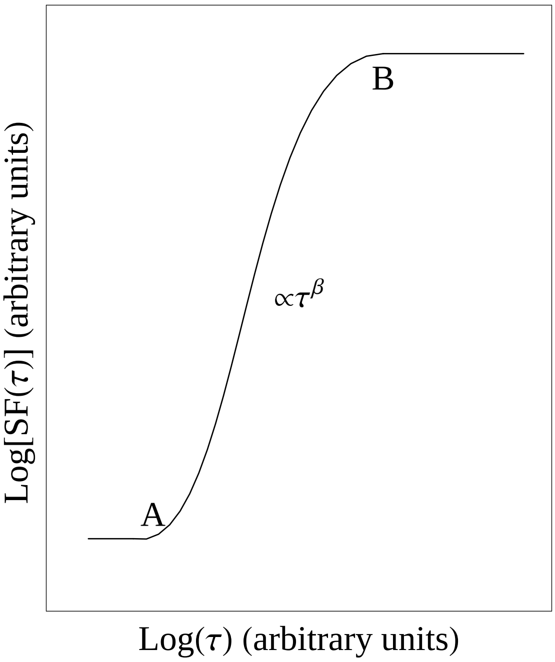


Figure 3.7: Schematic representation of the shape of the SF for a red noise process characterized by two plateaus and one slope, β, situated between the letters A and B (the SF shape after Hughes et al. 1992).

The shape of the SF for a red noise process (sect.3.1.1) is ideally described in a *log-log* representation by three components (Hughes et al. 1992): a slope between two plateaus (fig.3.7). The first "break" A is indicative of the variance of the measurements errors and the second "break" B gives an estimation concerning the sample variance.

The SF as a general statistical method was introduced by Kolmogorov (1941a,b) in order to discern scaling patterns in turbulent flows. In the field of astrophysics it was introduced in the middle 1980s by Simonetti et al. (1985) as a simple and capable function to describe the variability properties of extra-galactic radio sources. Thenceforth it has been extensively used in the time series analysis of AGN light curves with a completely wrong and misleading way, concerning especially the physical interpretation of the abscissa of the second "break" B, which is supposed to indicate a characteristic time property of the observed system. As it is shown in chap.5, completely random red noise data sets, representing artificial light curves, show exactly the same "breaks" as the observed ones, indicating that this quantity has nothing to do with any "characteristic time scale" embedded in the data set.

Concerning the slope of the SF, β, in the case of a white noise process (sect.3.1.1) it is equal to zero depicting exactly the fact that the mean variance between all the possible values is the same. For the case of a red noise process it can be shown that $\beta = 1$ (see following sections).

SF and ACF

For a stationary real process $x(t)$, $i = 1, 2, \ldots, T$, there is a direct relation between the SF and the ACF, $R_{1,1}(\tau)$ (eq.3.44) of the process. The autocovariance function of a time process $x(t)$ is defined in a discrete form by eq.3.4

$$V_{x,x}(\tau) = \left\langle \left(x(t) - \overline{x(t)} \right) \left(x(t+\tau) - \overline{x(t)} \right) \right\rangle \tag{3.24}$$

By expanding the terms inside the mean and considering that for a stationary process $\overline{x(t)} = \overline{x(t+\tau)} = \left\langle \overline{x(t)} \right\rangle =$ const. this reads

$$V_{x,x}(\tau) = \langle x(t)x(t+\tau) \rangle - \overline{x(t)}^2 \tag{3.25}$$

The variance S^2 of a process $x(t)$ is defined based on 3.19 having in the denominator N instead of $N - 1$ (i.e. the biased estimator) due to the fact that theoretically $\overline{x(t)}$ is estimated directly from the parent distribution and not from the data set (see footnote on sect.3.3.3).

$$S^2 = \left\langle \left(x(t) - \overline{x(t)} \right)^2 \right\rangle \tag{3.26}$$

Expansion of the terms inside the mean yields

$$S^2 = \overline{x(t)^2} - \overline{x(t)}^2 = V_{x,x}(0) \tag{3.27}$$

From eq.3.23

$$SF(\tau) = \overline{x(t)^2} - 2\left\langle x(t)x(t+\tau)\right\rangle + \overline{x(t+\tau)^2} \tag{3.28}$$

By means of eq.3.27

$$SF(\tau) = 2\left(S^2 + \overline{x(t)^2} - \left\langle x(t)x(t+\tau)\right\rangle \right) \tag{3.29}$$

Finally from eq.3.25 and eq.3.44 the SF for a stationary process is given by

$$\begin{aligned} SF(\tau) &= 2\left(S^2 - V_{x,x}(\tau)\right) \\ &= 2S^2\left(1 - \frac{V_{x,x}(\tau)}{S^2}\right) \\ &= 2S^2\left(1 - R_{1,1}(\tau)\right) \end{aligned} \tag{3.30}$$

Based on eq.3.30 the normalized SF (NSF) is defined as:

$$\begin{aligned} NSF(\tau) &= \frac{SF(\tau)}{S^2} \\ &= 2\left(1 - R_{1,1}(\tau)\right) \end{aligned} \tag{3.31}$$

As $t \rightarrow \infty$, $R_{1,1}(\tau) \rightarrow 0$ (i.e. there is no linear correlation between the various measurements) meaning that $SF(\tau) \rightarrow 2$ and $NSF(\tau) \rightarrow 2S^2$. In real life time series as the time delay τ increases the linear correlation between the measurements is gradually diminishing till the value τ_{uncor} where the measurements with separation time $\tau \geq \tau_{\mathrm{uncor}}$ have zero linear correlation. This means that as τ increases the SF is increasing (eq.3.30) till the value $\tau_{\mathrm{break}} = \tau_{\mathrm{uncor}}$ and from then on it continues its course around a fixed level $2S^2$. Since in real time series there is no infinite number of observations some times the observed segments of the parent distribution (data sets) might no be representative of the latter (heteroscedasticity, see sect.3.6.2). A direct consequence of that is either the occurrence of SF "break" at $\tau_{\mathrm{break}} \neq \tau_{\mathrm{uncor}}$ or the absence of a clear plateau where the distribution of points after τ_{uncor} is continuously increasing or decreasing without having a fixed mean value around $2S^2$.

SF and PSD

There is also a direct relation between the SF and the PSD (sect.3.3.1) being valid under some certain conditions which are rarely taken into account. Consider a zero mean stationary real time process $x(t)$ with an autocovariance function (eq.3.25)

$$V_{x,x}(\tau) = \left\langle x(t)x(t+\tau)\right\rangle = \left\langle x(t)x^*(t+\tau)\right\rangle \tag{3.32}$$

where the asterisk denotes complex conjugation. By rewriting the terms inside the mean as a function of their Fourier transform, $Y(f)$, this relation reads

$$\begin{aligned} V_{x,x}(\tau) &= \left\langle \int_{-\infty}^{+\infty} Y(f)e^{-2\pi i f t} df \int_{-\infty}^{+\infty} Y^*(f)e^{2\pi i f (t+\tau)} df \right\rangle \\ &= \left\langle \int_{-\infty}^{+\infty} Y(f)Y^*(f)e^{2\pi i f \tau} df \right\rangle \\ &= \left\langle \int_{-\infty}^{\infty} \left|Y(f)\right|^2 e^{2\pi i f \tau} df \right\rangle \end{aligned} \tag{3.33}$$

and since none of the terms is varying as a function of τ (stationarity)

$$V_{x,x}(\tau) = \int_{-\infty}^{\infty} \left|Y(f)\right|^2 e^{2\pi i f\tau} df \tag{3.34}$$

According to eq.3.30

$$\begin{aligned} SF(\tau) &= 2\left[V_{x,x}(0) - V_{x,x}(\tau)\right] \\ &= 2\left[\int_{-\infty}^{\infty} \left|Y(f)\right|^2 df - \int_{-\infty}^{\infty} \left|Y(f)\right|^2 e^{2\pi i f\tau} df\right] \\ &= 2\int_{-\infty}^{\infty} (1 - e^{2\pi i f\tau})\left|Y(f)\right|^2 df \end{aligned} \tag{3.35}$$

Based on the Euler's formula, $e^{i\kappa} = \cos(\kappa) + i\sin(\kappa)$, and ignoring the imaginary part which expresses only the phases

$$SF(\tau) = \int_{-\infty}^{\infty} (1 - \cos(2\pi f\tau))\left|Y(f)\right|^2 df \tag{3.36}$$

Considering eq.3.10 and the fact that the PSD is defined for $0 \leq f < \infty$ (eq.3.9)

$$SF(\tau) = 2\int_{0}^{\infty} (1 - \cos(2\pi f\tau))\mathcal{P}(f) df \tag{3.37}$$

For the majority of AGN the PSD can be well represented by a power-law $\mathcal{P}(f) = \kappa f^{-a}$, with $1 < a < 3$ (sect.3.1.1) and κ a positive constant. For this range of parameters the integration in eq.3.37 has an analytical solution

$$SF(\tau) = -2^a \pi^{a-1} \tau^{a-1} \Gamma(1-a) \sin\left(\frac{a\pi}{2}\right) \tag{3.38}$$

where $\Gamma(x)$ is the (complete) Gamma function. That means that under the assumptions of

- stationarity
- zero mean data set
- $f \in [0, \infty]$
- $1 < a < 3$

$SF(\tau) \propto \tau^{a-1}$, meaning that $\beta = a - 1$. For a red noise process, $a = 2$ (sect.3.1.1) which fulfills all the aforementioned assumptions, the slope of the SF is $\beta = 1$. Usually authors blindly pass from the SF slope β to the PSD slope a and *vice versa* without examining if the data set meet the aforementioned requirements (e.g. Takahashi et al. 2000).

3.3.6 Simulations and artificial light curves

Probably the most important issue in the time series analysis is to test how significant is a result coming from a given statistical method. In this section the method of Timmer & Koenig (1995) is reviewed as being the most appropriate one of producing a whole variety of stochastic time series exhibiting the same power-law-like PSD (sect.3.1.1) as that of the observed light curve.

The key point of this algorithm is that it randomizes both the phase and the amplitude of the Fourier transform of the data set (eq.3.11) according to the noisy nature of the process (eq.3.12). This is the most robust method of producing artificial light curves mimicking the first order statistical properties of a given data set. The previous method in use was developed by Done et al. (1992) but has a major flaw in the following sense. The randomization is done only in the phases of the Fourier components leaving the amplitude of each frequency fixed (i.e. deterministic) and equal to the square root of the spectrum at the given frequency. Therefore it does not take into account the randomness of the periodogram according to the χ^2_2 distribution (eq.3.12). This imperfection of the method creates a trend to the simulated light curves caused by the dominating lowest frequency. Actually by not taking into account the random nature of the amplitudes is equivalent of ignoring the intrinsic scatter in the powers of the process.

Based on Timmer & Koenig (1995) one can produce an ensemble of possible light curves having the same spectral index a and the same the first and second statistical moments with the under study data set. The procedure is described briefly here

1. A power-law shape of PSD $\mathcal{P}(f) \propto f^{-a}$ is chosen based on the given data set.
2. For each Fourier frequency f_j two Gaussian distributed random numbers are produced and multiplied by $\sqrt{\frac{1}{2}\mathcal{P}(f_j)}$. Then this result is used as the real and imaginary part of the Fourier transform of the desired data. This emerges from the theory of spectral estimation (Priestley 1981) showing that $P(f_j)$ is a complex Gaussian random number variable[3]: $P(f_j) = \mathcal{N}(0, \frac{1}{2}\mathcal{P}(f)) + i\mathcal{N}(0, \frac{1}{2}\mathcal{P}(f))$ whose variance does not depend on the number of data points.
3. In order to obtain a real valued time series the Fourier components for the negative frequencies are chosen according to $P(-f_j) = P^*(f_j)$ where the asterisk denotes complex conjugation.
4. Finally the desired light curve is obtained by inverse discrete Fourier transform (IDFT) of $P(f)$ from the frequency domain to the time domain.

As it is referred in sect.3.3.4 periodograms measured from real data tend to be biased by aliasing effects. Therefore if there is significant power at frequencies below the lowest frequency probed by the periodogram (reflecting time scales longer than the length of the observations) this can give rise to slow varying trends in the light curve which contribute to the variance of the system. This effect can be taken into account in the simulations by generating a longer light curve, 10 times longer than the under study light curve (e.g. Edelson et al. 2002), from a PSD that extends to very low frequencies (i.e. long time scales) and then truncate it to the desired length. In this way power on time scales much longer than those covered from the resultant light curve segment, is taken into account.

3.4 Nonlinear Methods

The scope of this section is to present in a simple and comprehensible way some of the most recent and robust time series analysis methods that are going to be applied in the following chapters to astrophysical data sets. The main caveat of a direct application of these methods to a given data set is that they have been developed for systems whose evolution equations are known (i.e. the physical system is known). Therefore they are applied in a sense directly to the physical system (i.e. to its equations) and hence its time evolutionary behavior can be predicted up to a certain degree. Additionally, noise processes can be added to the system in the form of pseudorandom numbers originating from a fixed distribution and a complete statistical study of it can be performed by analyzing several realizations. Note here that even though the evolution equations of the system are known the existence of a noise process affects the evolution of the system in such a way that its behavior after some time can not be readily predicted. Another advantage of knowing the evolution equations of the system in advance is that for every realization of the system the sampling pattern, the time resolution and most importantly the number of points can be specified by the needs of the under study problem.

In complete contradiction, the purpose of analyzing astronomical data sets is initially to reveal some characteristics of the evolution equation of the system and not to predict (at least as a first step). Only the outcome of the underlying physical parameters, comprising the physical system, is known in the form of the observed light curve and based on that one should extract as much information as possible concerning the laws (i.e. the equations) that rule them. Moreover the time extension of a given data set is fixed as well as the sampling patterns therefore there is no *a priori* knowledge whether or not the data set can adequately to describe in a representative manner the statistical properties of the system. Finally one should always keep in mind that some of these methods have been developed for continuous dynamical systems and application of them to discrete signals might hide some dangers concerning especially the asymptotic behavior of the system (Noakes 1986; Krivine et al. 2004).

The basic steps of analyzing data coming from nonlinear dynamical systems are the same with those applied to linear systems differing only in the analysis methods followed for each step. In tab.3.1 the basic

---

[3] $\mathcal{N}(0, S^2)$ represents a zero mean Gaussian distribution with variance S^2.

procedures are shown in a condensed form. Roughly speaking the main parameter that one has to define in linear systems is the (PSD) or better the Fourier space of the data set. The Fourier space comprises a robust representation of the system, since it forms an orthonormal basis, of the initial signal in the frequency domain. The advantage of representing a system in this domain is that one can readily discern its dominant components, if there are any, select them and study them separately from the overall system. The peak frequencies are the invariants of the system and based on them in the case of deterministic physical systems (fig.3.1) the time evolution of the system can be studied analytically.

From the other hand in the case of the nonlinear systems end especially for the chaotic ones the representation of the system in the Fourier space is not always if at all possible since the data sets, comprising the “solutions” of the dynamical equations, can not be always expanded into convergent Fourier series. As it is shown by Miles (1984) for the simple case of a weak nonlinear oscillator there are only some frequency ranges that can be probed through Fourier analysis. In the case of nonlinear systems the basic parameter that one has to specify is the phase space through the method of phase reconstruction. The basic characteristics of the reconstructed phase space should remain the same with those of the genuine phase space.

Table 3.1: Time series analysis methods of linear and nonlinear data sets

Procedure	Linear signals	Nonlinear signals
Noise subtraction	Separation of the signal’s continuous spectrum from the “thin-zone” spectrum. For the case of a stationary signal it is convenient to apply a linear filter in a given spectral region.	Separation of the signal’s continuous spectrum from the “thin-zone” spectrum. The methods of principal component analysis (Prin.Comp.Anal) or the radial basis functions (RBFs) are appropriate.
Phase space specification	By making use of the Fourier transformation, the data set is transformed into linear relations of the Fourier parameters: $x(f) = \sum x(t)e^{-2\pi f t}dt$.	By applying the method of delays (MOD) to the data set the genuine phase space can be reconstructed in m embedding dimensions $\vec{X}_i = \{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\}$.
Classification of the signal	The peak frequencies and the resonance frequencies are resolved. These quantities are invariants and independent of the initial conditions.	Based on the reconstructed phase space the invariants of the trajectories are estimated (e.g. the correlation dimension, the Lyapunov exponents). These quantities are independent of the initial conditions.

3.4.1 The phase space

The phase space of a dynamical system is the mathematical space (manifold) whose coordinates are the independent variables describing the evolution of the system as a function of time. For example if a system needs 3 independent variables in order to be fully described then the system has a dimension equals to 3. Based on the discussion of sect.3.1 deterministic systems are characterized by low-dimensionality ($\leq$15) as opposed to stochastic systems characterized by high-dimensionality (>15).

Consider the simple case of an linear autonomous dynamical system (i.e. **A** does not depend on time t) in an m-dimensional Euclidean phase space described from the differential equation

$$\frac{d\vec{y}(t)}{dt} = \mathbf{A}\vec{y}(t) \tag{3.39}$$

where $\vec{y}(t) = \{y_1(t), y_2(t), \ldots, y_m(t)\}$ and **A** is a constant $m \times m$ matrix. The trajectories (i.e. general solutions) of eq.3.39 reveal the evolution of the system as t→ ∞:

- They converge to zero if the real part of all eigenvalues of **A** is negative.
- They diverge to infinity if the real part of at least one eigenvalue of **A** is positive.

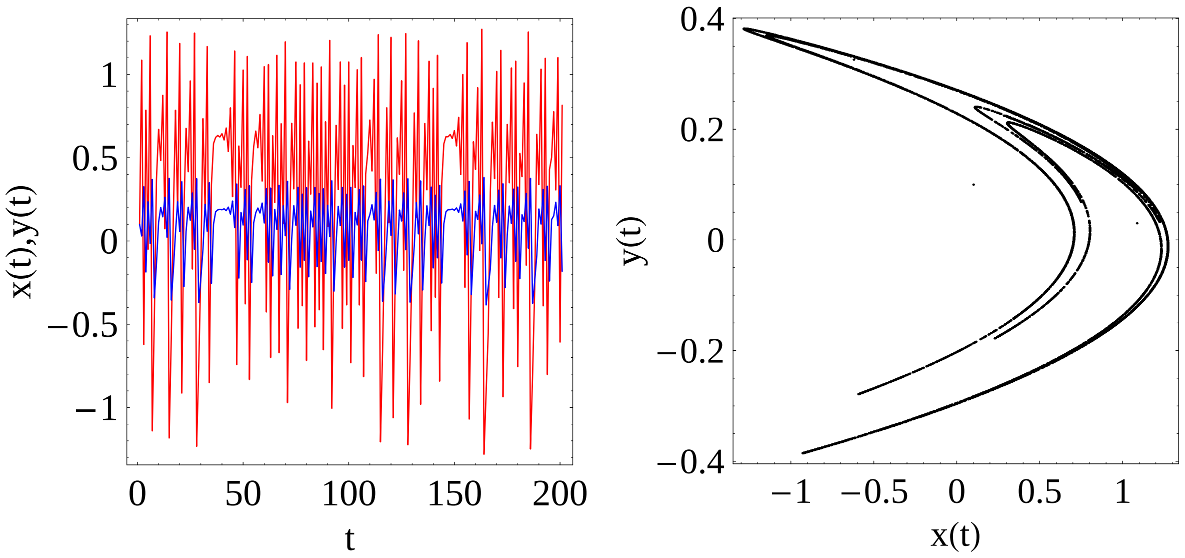


Figure 3.8: [Left panel] The realization of the Hénon dynamical system as described by eq.3.41 with $x_0 = y_0 = 0.1$. The red curve represents the x_t component and the blue the y_t.
[Right panel] The two-dimensional phase space of the Hénon dynamical process.

- They oscillate if all the eigenvalues of **A** form complex conjugate pairs with zero real part.

Note here that in discrete time the analogue of eq.3.39 is the difference equation

$$\vec{y}_{t+1} = \mathbf{A}\vec{y}_t \tag{3.40}$$

For the analysis of nonlinear dynamical systems this linear approach is not adequate to fully describe the evolution of the system (i.e. for all t and for all initial conditions). Linear methods usually help to study the system locally around a known solution and for limited time periods. Therefore the nonlinear terms in the analysis method should also be considered in order to study the system globally.

The Hénon discrete-time dynamical system (Hénon 1976) is probably the simplest two dimensional nonlinear chaotic system and hence it is appropriate to use in order to understand the physical meaning of the phase space. The system is described by the following coupled difference equations for α=1.4 and β=1.3

$$\begin{aligned} x_t &= 1 - \alpha x_{t-1}^2 + y_{t-1} \\ y_t &= \beta x_{t-1} \end{aligned} \tag{3.41}$$

In fig.3.8 is shown the realization of each component of the process (left panel) together with the phase space (right panel) of the system. The dynamical system is described by two equations therefore its phase space can be fully represented in to two dimensions.

Another example of a continuous dynamical nonlinear system is the Lorenz dynamical system (Lorenz 1963) consisting of three coupled differential equations for $\sigma = 10$, $\rho = 28$, $\beta = 8/3$

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y(t) - x(t)) \\ \frac{dy}{dt} &= x(t)(\rho - z(t)) - y(t) \\ \frac{dz}{dt} &= x(t)y(t) - \beta z(t) \end{aligned} \tag{3.42}$$

The realization of the individual system's components together with the phase space of the process are shown in fig.3.9. Since the process consists of three equations each one representing the evolution of each independent variable as a function of time, its phase space can be fully described into three dimensions.

Inside the phase space of each dynamical system a structure is formed representing all the possible states that the system can be found. This structure, theoretically consists of points in the case of discrete

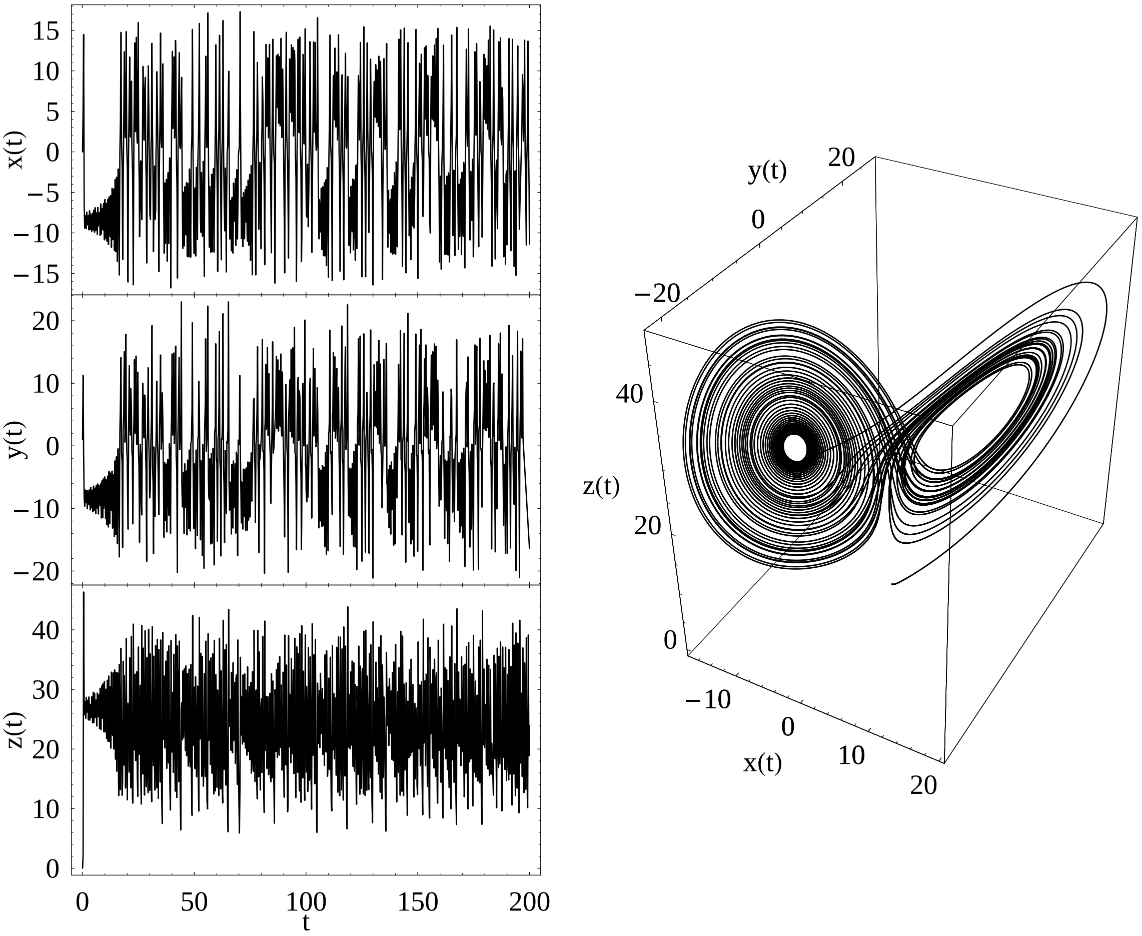


Figure 3.9: [Left panel] The realization of the components of the Lorenz dynamical system as described by eq.3.42 with $x(0) = z(0) = 0, y(0) = 1$.
[Right panel] The three-dimensional phase space of the Lorenz dynamical process.

systems and lines in the case of continuous systems both of them forming a trajectory pattern in the phase space known as *attractor*. Practically since for the continuous systems most of the times an integration step Δt is needed in order to specify the attractor's trajectories the latter consist also of points. The definition of the term attractor together with some ambiguities existing in the literature can be found in Strogatz (1998). Concerning the purposes of this study one can loosely define it as a set to which all neighboring trajectories converge. There are a lot of types of attractors in the literature concerning the dynamical systems having as most representative the point, the limit circle, the toroidal and the strange (or chaotic) attractor. For the rest of this study whatever structure is formed within the phase space it will be called attractor.

3.4.2 The phase space reconstruction

Consider the general case of a dynamical system $\vec{s}$, either linear or nonlinear, evolving to its genuine/physical phase space comprising of a d-dimensional manifold $\mathbf{M}^d$ (fig.3.10). This practically means that $\vec{s}$ can be described by d independent variables each one consisting an axis of the genuine phase space. Every state of the system at a given time instant is a function of the previous states and this can be expressed in a form of a discrete function $\vec{s}_{t+1} = \mathbf{f_t}(\vec{s}_t)$ (general case of eq.3.40). Exactly this function $\mathbf{f_t}$ forms the genuine trajectories of the attractor in the physical d-dimensional phase space. The evolution of this system is mapped through a function h in the Euclidean space $\mathbb{R}$ ($h : \mathbf{M}^d \to \mathbb{R}$) in the form of the observed time series $x_t = h(\vec{s_t})$ for $t = (1, 2, \ldots, N)$. The function h consists practically the projection of the vector components of $\vec{s}$, into a time series data set and this is exactly the stage where also the discretization and the observational noise (sect.3.1.2) attached to the data set. Despite the fact that h maps the information

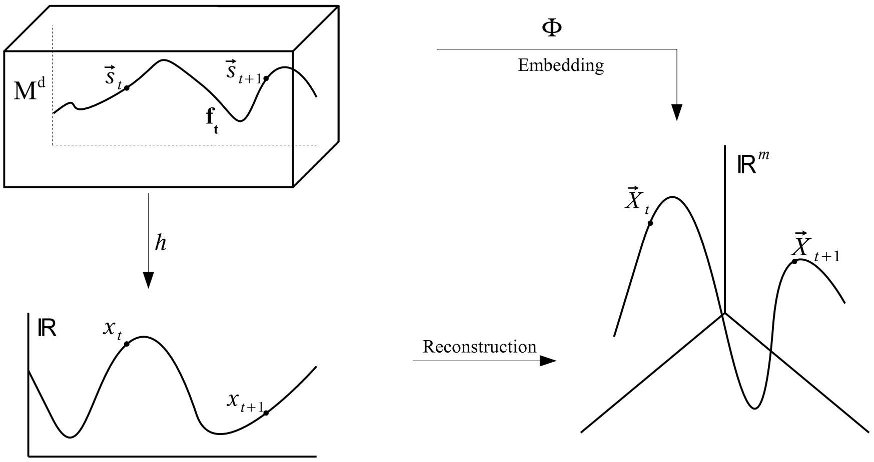


Figure 3.10: Schematic representation of the phase reconstruction. The system is evolved into its physical (i.e. genuine) phase space, being a d-dimensional manifold $\mathbf{M}^d$ and through the function $\mathbf{f_t}$ it forms their its genuine attractor. Via h some of the components of the system are projected in the form of the observed time series, x_t in the set of $\mathbb{R}$. Based on x_t the phase space can be reconstructed in the $\mathbb{R}^m$ by forming the m-dimensional delayed vectors $\vec{X}_t$. In order the reconstruction to be equivalent with the smooth attractor's transformation from $\mathbf{M}^d$ to $\mathbb{R}^m$, through the immersion function $\mathbf{\Phi}$, the dimensionality of the reconstructed phase space should be $m \geq 2d + 1$ as it is dictated from the *delay embedding theorem* (Takens 1981).

only from specific components of $\mathbf{M}^d$, the evolution of the system in time is a result of the ensemble of its components therefore one could say that its course in time towards any direction is a result of d-equations.

The fundamental question that now arises is: Since neither h nor $\mathbf{f_t}$ are known, how is it possible to reconstruct the genuine phase space in order to form the initial attractor of the system, revealing the physics governing its evolution, having as the only tool the observed time series data set? The method which is usually applied in order to answer this question has been developed by Packard et al. (1980) and it is known as *method of delays*. The basic idea is that after applying a smooth transformation (i.e. a map between manifolds which is differentiable and has a differentiable inverse transforming, also called *diffeomorphism*) to the observed time series data set, some of the *geometrical invariants* of the genuine attractor are maintained. These invariants are the correlation dimension, the Lyapunov exponents and the eigenvalues of fixed points.

Based on the MOD an embedding dimension m together with a time delay τ are chosen and the vectors (delay vectors) $\vec{X}_i = \{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\}$ for $i = 1, 2, \ldots, L$ with $L = N - (m-1)\tau$ are formed (tab.3.2). If m and τ are chosen correctly then the $\vec{X}_i$ vectors form in the m-dimensional Euclidean space $\mathbb{R}^m$ an *equivalent representation* of the initial physical phase space. In this case, the original attractor (in d dimensions) is mapped through a function $\mathbf{\Phi}$ to a structure in the reconstructed phase space (m dimensions) in which the former preserves all the aforementioned invariants of the latter but without preserving the geometrical shapes and structures. The only crucial parameter that should be determined in order to achieve a robust reconstruction of the physical phase space (e.g. the one containing the genuine attractor) is the time window $\tau_{\mathrm{w}} = (m-1)\tau$, involving both the embedding dimension and the time delay. In sect.3.4.3 the optimum selection criteria for both m and τ are discussed thoroughly.

The physical concept behind the MOD is that the combination of the future and the past values of a time series data set at a given time instant t contains information about the non-observed independent variables needed to be specified in order to fully describe the system. This combination is achieved through the delay vectors $\vec{X}_i$ whose components, the delay variables $\{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\}$ are going to be used in the same manner as the genuine unknown variables in order to reconstruct a space describing the structure of the phase trajectories (attractor) of the system. In the case of a linear/nonlinear system the delay variables are a linear/nonlinear combination of the "true" independent variable of the system.

In this point the Lorenz dynamical system (eq.3.42) is used in order to check how the MOD is practically working (fig.3.11). Assuming that only the $y(t)$ component is observed (i.e. it is the only one projected and discretized by the unknown function h having no observational noise) the phase space is reconstructed in $m = 2$ and $m = 3$ dimensions respectively. In both cases the time delay has been chosen to be τ=8

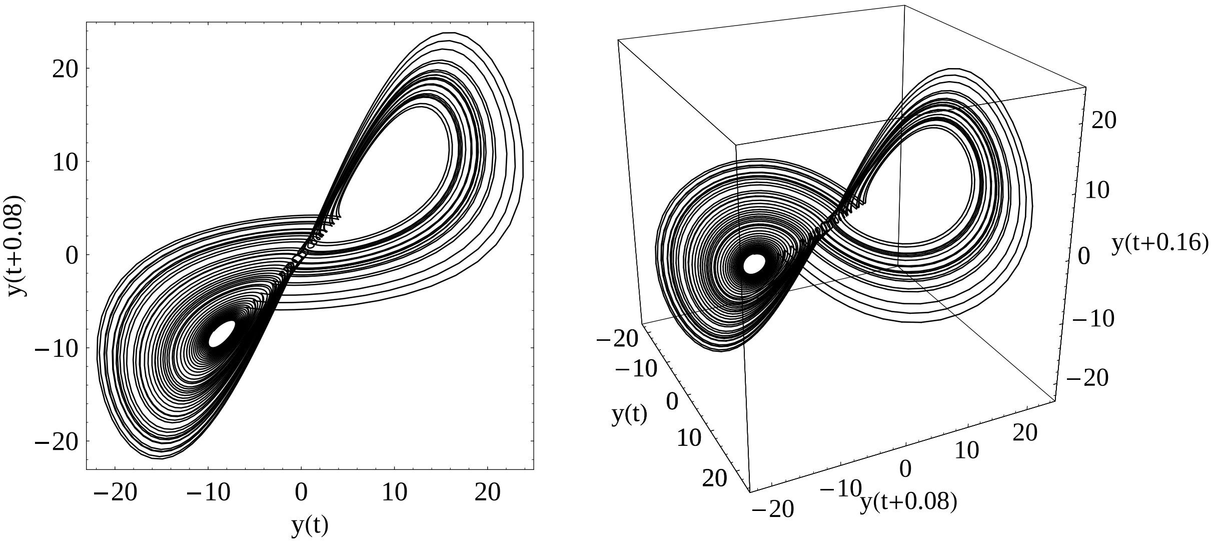


Figure 3.11: [Left panel] The reconstruction of Lorenz dynamical system in two dimensions (m=2) using only the $y(t)$ component with a time delay τ=8 and an integration step δT =0.01.
[Right panel] The reconstruction of Lorenz dynamical system in three dimensions (m=3) using only the $y(t)$ component with a time delay τ=8 and an integration step δT =0.01.

Table 3.2: Examples of several delay vectors $\vec{X}_i$ for the time series $(x_1, x_2, \ldots, x_{10})$

m	τ	Delay variables	L	Delay vectors $\vec{X}_i$
2	1	$\{x_i, x_{i+1}\}$	9	$\vec{X}_1 = (x_1, x_2), \vec{X}_2 = (x_2, x_3), \ldots, \vec{X}_9 = (x_9, x_{10})$
3	1	$\{x_i, x_{i+1}, x_{i+2}\}$	8	$\vec{X}_1 = (x_1, x_2, x_3), \vec{X}_2 = (x_2, x_3, x_4), \ldots, \vec{X}_8 = (x_8, x_9, x_{10})$
3	2	$\{x_i, x_{i+2}, x_{i+4}\}$	6	$\vec{X}_1 = (x_1, x_3, x_5), \vec{X}_2 = (x_2, x_4, x_6), \ldots, \vec{X}_6 = (x_6, x_8, x_{10})$
4	2	$\{x_i, x_{i+2}, x_{i+4}, x_{i+6}\}$	4	$\vec{X}_1 = (x_1, x_3, x_5, x_7), \vec{X}_2 = (x_2, x_4, x_6, x_8), \ldots, \vec{X}_4 = (x_4, x_6, x_8, x_{10})$
4	3	$\{x_i, x_{i+3}, x_{i+6}, x_{i+9}\}$	1	$\vec{X}_1 = (x_1, x_4, x_7, x_{10})$

(see sect.3.4.3). By an "eye inspection" of the results shown in fig.3.11 one can readily distinguish some attractor features resembling to those of the genuine attractor's phase space (fig.3.9, right panel). The fact that in the 3-dimensional reconstructed phase space the attractor is somehow distorted in comparison to the original one is due to the *smooth transformation* that has been applied to the data set. This is expected since the coordinates that have been used for the reconstruction $\{y(t), y(t+0.08), y(t+0.16)\}$ are simply different from the original ones $\{x(t), y(t), z(t)\}$. Despite this distortion the power of the MOD is that the reconstructed attractor (in both $m = 2$ and $m = 3$) conserves some unchanged quantities and these are exactly the geometrical invariants that are going to be used in order to reveal the dynamics of the whole process. The question now is the following: Which one of the two reconstructed phase spaces of fig.3.11 is more appropriate in order to study these invariants? This issue is discussed in sect.3.4.3 where some optimum selection criteria concerning the choice of both the embedding dimension m and the time delay τ are presented.

3.4.3 Selection of the time window

The basic quantity that should be accurately tuned in order the reconstructed phase space to be an *equivalent* version of the original one is the time window $\tau_{\mathrm{w}} = (m-1)\tau$. This means that the selection of the embedding dimension m and time delay τ should be done based on some optimum selection criteria. In the case of m there is a series of topological theorems and definitions that ensure the equivalency of the two spaces but it is aside from the scope of this study to refer to them with all the strict mathematical formalism. Only the physical application of these theories is going to be discussed in the frame work of the time series analysis. Concerning the selection of the τ the methods are more empirical depending each time on the under study data set. The selection of m defines automatically the number of the delay variables, comprising the components of the delay vectors, and the value of τ defines which measurements x_i are going to form these variables (see tab.3.2).

The selection of m

Based on fig.3.10 the m-dimensional phase space is reconstructed by using the observed data set x_k. In order this phase space to be *equivalent* with the genuine d-dimensional phase space some criteria should be fulfilled.

In topology, two algebraic structures (in this case the genuine and the reconstructed phase spaces) are *equivalent* if there is an immersion (embedding function) $\mathbf{\Phi}$ that maps the genuine attractor's trajectories to the reconstructed phase space in such a way that they preserve their orientation. In the case that $\mathbf{\Phi}$ is a continuous and bijective (one-to-one correspondence) function then it is called *homomorphism.* If $\mathbf{\Phi}$ is infinitely differentiable then the transformation between the d-dimensional and m-dimensional phase space is called *smooth.*

The Whitney's theorem states that any differentiable d-dimensional manifold (in this case where the genuine phase space is defined) can be smoothly embedded in an Euclidean space having $2d+1$ dimensions. This immersion should be done by the embedding function $\mathbf{\Phi}$ which is continuous, bijective and smooth: $\mathbf{\Phi} : \mathbf{M}^d \rightarrow \mathbb{R}^{2d+1}$.

The application of this theorem in the field of dynamical systems and by extension to time series was introduced by Takens (1981) in the form of the *delay embedding theorem.* Basically this theorem gives the conditions under which a dynamical system can be robustly reconstructed from a sequence of observations comprising its realization. In order the reconstruction in a given dimension m to be correct it should preserve those properties of the attractor that do not change under smooth coordinate transformations ignoring the fact that the geometric shapes and the structures of the trajectories alter. The only condition necessary to preserve these invariants is that $m \geq 2d+1$. One of the latest developments in the embedding methodologies (embedology) came from Sauer et al. (1991) proving that in order to $\mathbf{\Phi}$ to be continuous, bijective and smooth it is enough $m \geq 2d$.

Consequently the main thing that should be kept in mind concerning the choice of m is that the correctly reconstructed trajectories of the attractor:

- preserve their orientation
- have with the genuine trajectories a one-to-one correspondence
- do not intersect among themselves
- preserve some characteristics quantities unchanged (invariants)

if and only if $m \geq 2d$. Obviously the selection of m automatically sets the number of the components/delay variables of the delay vectors (tab.3.2).

The selection of τ

The parameter of the time delay τ is directly connected with the quality of the reconstruction. Practically its value defines which measurements x_i are going to be chosen in order to form the components (delay variables) of the delayed vectors. Despite the selection of m, if the vector components are not chosen correctly (i.e. if τ is not chosen correctly) then the reconstructed phase trajectories will be distorted (suppressed or stretched). The result is that neither specific trajectory structures can be readily discerned nor the invariants of the system can be correctly estimated. The main considerations concerning the selection of τ are two:

1. Small values of τ force the components of the delay vectors $\vec{X}_i$ to be spatially correlated resulting a distribution of points which "oversamples" only a small part of the whole extension of attractor in the phase space (fig.3.12, left panel). These values are distributed around the diagonal of the phase space hiding all the information about the genuine attractor trajectories.
2. Big values of τ result an inhomogeneous coverage of the phase space, due to the complete lack of any spatial correlation of the delay variables. Therefore it is not possible to extract any information concerning possible attractor's structures corresponding to features of the genuine trajectories (fig.3.12, right panel).

The optimum choice should be the one forming components consisting of independent points up to a given degree with relatively small τ. A first approximation would simply be to consider the ACF of the data set and find for which τ it becomes zero. In this case it is ensured that the data sample is completely released from linear correlations but not from possible higher order correlations. Therefore the use of higher order

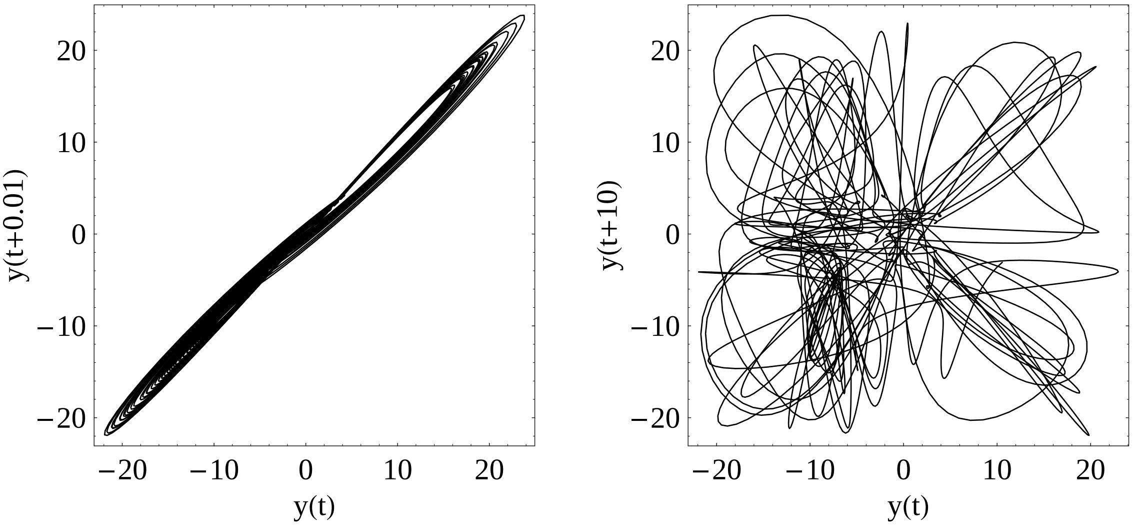


Figure 3.12: A wrong reconstruction of the Lorenz dynamical system in two dimensions (m=2) by using only the $y(t)$ component (see for comparison left panel of fig.3.11).
[Left panel] For a time delay τ=1 and an integration step δT =0.01. There is a suppression of all the points around the diagonal of the phase space due to the spatial correlation of the values.
[Right panel] For a time delay τ=1000 and an integration step δT =0.01. There is a spread of all the points around the space due the total spatial decorrelation of the values.

correlations (Albano et al. 1991) is mandatory in order to check the exact behavior of τ and make the appropriate selection for the MOD.

The usual linear ACF for a zero mean time series y_i, where $y_i = x_i - (\sum_{i=1}^{N} x_i)/N$ with a standard deviation S is given by

$$R_{1,1}(\tau) = \frac{\sum_{i=1}^{N-\tau} y_i y_{i+\tau}}{NS^2} \tag{3.43}$$

The linear ACF is actually a version of the autocovariance function (eq.3.4, in its discrete form where the integration has been replaced by the sum over all N) normalized to the sample variance S^2 (eq.3.19).

$$R_{1,1}(\tau) = V_{y,y}(\tau)/S^2 \tag{3.44}$$

Some of the higher order correlations are

$$R_{2,1}(\tau) = \frac{N\sum_{i=1}^{N-\tau} y_i^2 y_{i+\tau}}{(N-1)(N-2)S^3} \tag{3.45}$$

$$R_{3,1}(\tau) = \frac{N(N+1)\sum_{i=1}^{N-\tau} y_i^3 y_{i+\tau}}{(N-1)(N-2)(N-3)S^4} - 3R_{1,1}(\tau) \tag{3.46}$$

$$\begin{aligned} R_4(\tau) = {} & \frac{N(N+1)\sum_{i=1}^{N-3\tau} y_i y_{i+\tau} y_{i+2\tau} y_{i+3\tau}}{(N-1)(N-2)(N-3)S^4} - \\ & R_{1,1}(\tau)^2 - R_{1,1}(2\tau)^2 - R_{1,1}(\tau)R_{1,1}(2\tau) \end{aligned} \tag{3.47}$$

The skewness (the degree of the asymmetry of a distribution around its mean) and the kurtosis (the peakedness or flatness of a distribution) of y_i is given by $R_{2,1}(0)$ and $R_{3,1}(0) = R_4(0)$ respectively.

For the selection of τ the relations 3.44, 3.45, 3.46 and 3.47 must be plotted as a function of τ and the abscissa of the point where all the functions have an extremum simultaneously for the first time, coincidence time τ_c, gives the optimum value of τ which is appropriate for the MOD (Albano et al. 1991). Note here that this point should not be the one where two or more correlation functions become zero

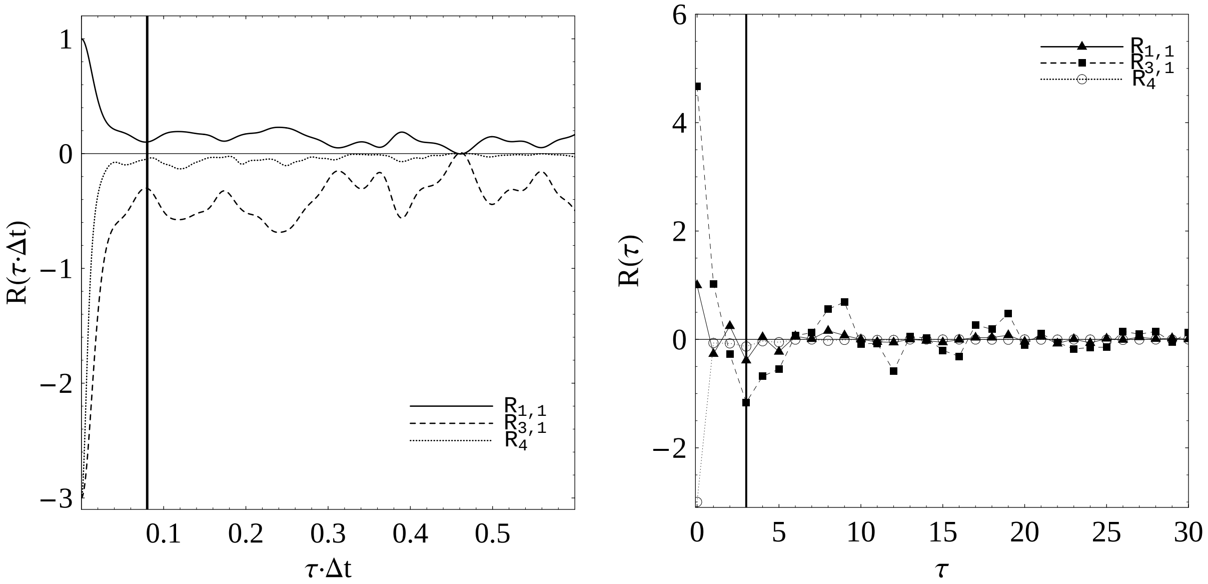


Figure 3.13: [Left panel] The higher order correlations of the data set described in fig.3.11 for the Lorenz system. The vertical thick line indicates the coincidence time $\tau_c = 0.08$ when there is simultaneous occurrence of the extrema for the first time. Since the resultant points are arrayed in a very dense form they have been replaced with the corresponding lines indicating their course.
[Right panel] The same as the previous for the x_t component of the Hénon discrete system (eq.3.41). The simultaneous occurrence of the extrema for the first time happens at t_c=3. Note that the various lines intend only to guide only the eye.

(often the first intersection point happens to have an ordinate around zero) because then at that given τ the data set becomes completely independent (decorrelated) and thus the phase space can not be efficiently reconstructed.

The reconstruction of the Lorenz attractor in 2 and 3 embedding dimensions respectively (fig.3.11, left panel) was done by using a time delay of $\tau = 8$. Based on the aforementioned method since the data set has a τ_c=0.08 (fig.3.13, right panel) that means that the optimum time delay is at $\tau = 8$ ($\Delta t = 0.01$). Therefore the selected delay variables should be separated by 0.08 time units. Application of the same method in 1000 realizations of the x_t component of Hénon system (eq.3.41) yields a coincidence time of $\tau_c = 3$.

3.4.4 Dimensions

The dimension of the attractor in the reconstructed phase space is probably one of the most intently studied invariant quantities (Farmer et al. 1983). As discussed in sect.3.4.1 the number of independent variables needed to be determined (in a form of differential or difference equations) in order to fully describe the evolution of the dynamical system defines the dimension of it and it is always an integer.

Loosely speaking the dimension D is simply connected with the physical quantity of *volume* V (not only in 3-dimensions) the latter being connected with the length L. The general scaling relation for the volume is given by

$$V_D \propto L^D \tag{3.48}$$

For $D = 1$ the length L fully specifies the one dimensional *volume* element which is in this case a *line*. For $D = 2$ (or $D = 3$) the volume element is the *area* (or *3-dimensional volume*) in two (or three) dimensions. The prefix *hyper* is usually used in order to generalize quantities and shapes from 3-dimensions to higher dimensions such as *hypervolume*, *hypersphere*, *hyper-cube* etc. Equation 3.48 becomes more accurate if it is expressed as a function of N boxes of side r, defining a grid, in a given dimension (see fig.3.14).

$$V_D = N_r r^D \tag{3.49}$$

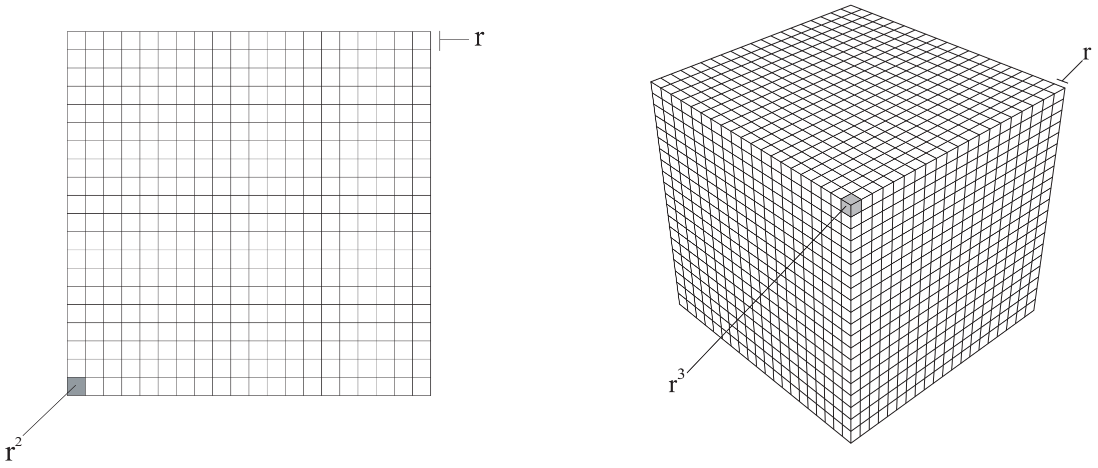


Figure 3.14: [Left panel] A 2-dimensional grid consisting of N_r square-cells (i.e. 3-dimensional degenerated cubes) covering a 2-dimensional phase space. Each cell has a side-length of r and *occupies a 2-dimensional volume* i.e. covers an area, of r^2. All the cells cover an area of $N_r r^2$.
[Right panel] A 3-dimensional grid consisting of N_r cubes covering a 3-dimensional phase space. Each cube has an edge-length of r and occupies a volume of r^3. All the cubes occupy a volume of $N_r r^3$

What is the volume of the attractor in this reconstructed phase space? This quantity is the key parameter that one has to specify in order to pass from the attractor's structure, which can be quantified by the volume, to its dimensionality. In this context, the term "volume" itself is unambiguous since the reconstructed attractor consists of points. In order to elucidate this quantity one needs to think as follows: Consider that the attractor of a data set, x_t, is formed, based on the MOD, into the two-dimensional phase space $m = 2$, which has as delay variables $\{x_t, x_{t+k}\}$ and it is covered by a grid of a certain length r_1 (fig.3.15, left panel). The number of squares including more than one attractor's point are measured, N_r, yielding for the given grid configuration a pair (r_1, N_{r_1}). By increasing the resolution of the grid to r_2 (fig.3.15, right panel), meaning that $r_1 > r_2$, the same counting procedure is performed and a set of (r_2, N_{r_2}) values is again specified. Then by taking the logarithms of both sides of eq.3.49 one can estimate D simply by linear regression based on the method of least squares[4] (Bevington & Robinson 1992)

$$\log N_r = D \log\left(\frac{1}{r}\right) + \text{const.} \tag{3.50}$$

The same exactly procedure is performed in $m = 3, 4, \ldots$ embedding dimensions yielding each time a new value of D. If the number of D converges to a fixed value D_0 as m increases then that means that this is the dimension of the attractor. Most of the times D_0 is a non integer number and it is always $D_0 \leq m$. Having estimated the dimensionality D_0 of the attractor, the dimension (number of independent variables) of the dynamical system equals to the smaller integer greater or equal of D_0, $\lceil D_0 \rceil$.

The aforementioned method concerning the estimation of D_0 it is actually the result expected for the asymptotic limit of eq.3.50, $r \to 0$ (meaning also that $N_r \to \infty$ in the case of continuous dynamical systems with infinite attractor's resolution), as m increases

$$D_0 = \lim_{r \to 0} \frac{\log N_r}{\log\left(\frac{1}{r}\right)} \Bigg|_{m \to \infty} \tag{3.51}$$

For a stochastic process (sect.3.1) D_0 never converges to a given number or it converges practically to unrealistic big numbers of dimensions (grater than 15) meaning that parameters affecting the system's evolution can not be easily (or not at all) analytically manipulated.

There is a big variety of different kinds of dimensions in the literature concerning the dimensionality of a dynamical systems and all of them are directly related with the volume of the attractor in the phase space. The aforementioned quantity D_0 is called *capacity* dimension or *box-counting* dimension. If instead of boxes (*hyperboxes*) the same method is applied for spheres (*hyperspheres*) then the resultant

---

[4] Henceforth the method of least squares is going to be used for linear regression unless otherwise noted.

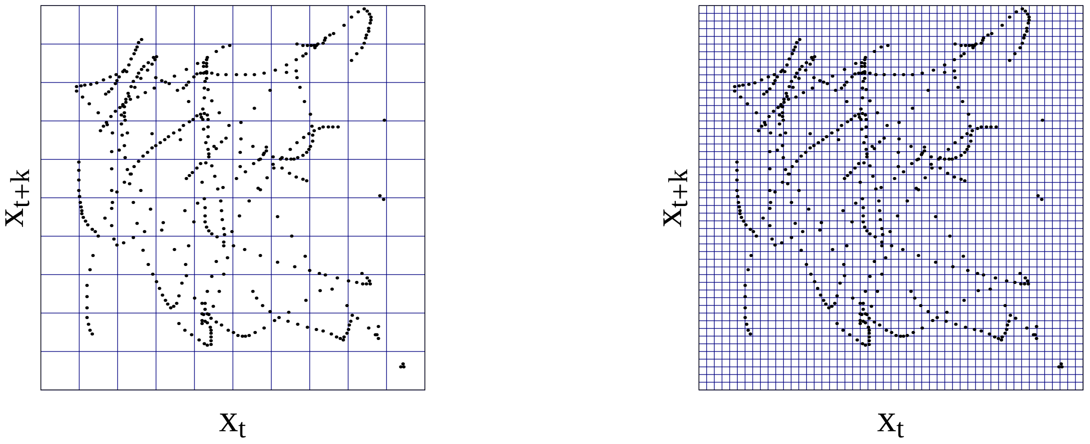


Figure 3.15: [Left-Right panel] The reconstructed attractor (points) in two dimensions, $m = 2$, for two different grid configurations r_1 (left panel) and r_2 (right panel), $r_1 > r_2$.

dimension is called *Hausdorff* dimension (Hausdorff 1918; Duvall et al. 2000). One of the most commonly used quantities for characterizing the dimensionality of the under study system is the *correlation* dimension D_2 (Grassberger & Procaccia 1983a) (see sect.3.4.5) through the estimation of the correlation integral. Application of these methods to the Hénon dynamical system (eq.3.41) yields a capacity dimension of $D_0 = 1.261 \pm 0.003$ (Russell et al. 1980) and a correlation dimension of $D_2 = 1.25 \pm 0.02$ (Grassberger & Procaccia 1983a). Differences between the methods are expected, due to the different counting procedures, but the main result of both methods is that one needs $\lceil D_0 \rceil = \lceil D_2 \rceil = 2$ equations to fully describe the system fact which is true based on eq.3.41. For the case of the Lorenz dynamical system $D_0 = 2.05 \pm 0.01$ and $D_2 = 2.06 \pm 0.01$ (Grassberger & Procaccia 1983a) yielding a consistent result for the number of independent variables needed to describe the system, being $\lceil D_0 \rceil = \lceil D_2 \rceil = 3$.

3.4.5 The correlation dimension

Usually the correlation dimension D_2 is the most commonly used quantity characterizing the dimensionality of a dynamical system (e.g. Brandstater & Swinney 1987; Sato et al. 1988). Due to the efficient and quick algorithms that have been developed for estimating it, firstly by Grassberger & Procaccia (1983a,b) and then by Theiler (1987), most of the researchers prefer it over the other dimensions (sect.3.4.4). The main idea for the computation of the correlation dimension is the same as the one of the capacity dimension.

After reconstructing the phase space of the attractor for a given m (based on the MOD, see sect.3.4.2), m-dimensional *hyperspheres* are then formed having as centers the attractor's points with equal radius r_k. By finding for the ensemble of *hyperspheres* how many of them contain at least one attractor's point (except from the one at the center of each *hypersphere*) the probability of this to happen is estimated. Exactly this probability value is the correlation integral $C_m(r_k)$ and can be readily estimated by comparing the distances of all the reconstructed phase points with the r_k

$$C_m(r_k) = \frac{2\sum_{i=1}^{L}\sum_{j=1,j\neq i}^{L} H(r_k - \|\vec{X}_i - \vec{X}_j\|)}{L(L-1)} \tag{3.52}$$

where $H(x)$ is the Heaviside step function using the convection $H(0) = 1$ (instead of 1/2) and $||\cdot||$ is the Euclidean norm defined in $\mathbb{R}^m$

$$\|\vec{X}_i - \vec{X}_j\| = \sqrt{\sum_{i=k}^{m}(X_{i,k} - X_{j,k})^2} \tag{3.53}$$

The radius of the *hyperspheres*, r_k, at a given m is defined as

$$r_k = R\lambda k \tag{3.54}$$

where R is the statistical range $R = \max[x_i] - \min[x_i]$ of the data set x_i and the parameter λ defines the "scaling region" taking values usually around 0.5%[5]. A thorough discussion concerning the choice of the "scaling region" can be found in (Elner 1988).

A much more neat computational way usually used to estimate the $C_m(r_k)$, based on eq.3.52, is the following

$$C_m(r_k) = \frac{2\sum_{i=1}^{L-1}\sum_{j=i+1}^{L} H(r_k - \|\vec{X}_i - \vec{X}_j\|)}{L(L-1)} \tag{3.55}$$

Methodology

In general the correlation dimension is given by:

$$D_{2,m} = \lim_{r\to 0} \frac{\log C_m(r)}{\log r} \tag{3.56}$$

meaning that

$$C_m(r) \propto r^{D_{2,m}} \tag{3.57}$$

and it is practically calculated for a given m by linear regression over all r_k.

$$\log C_m(r_k) = D_{2,m} \log r_k + \text{const.} \tag{3.58}$$

At a given embedding dimension m, the radius r_k is gradually increasing $k = 1, \ldots, 20$ and each time $C_m(r_k)$ is estimated. Then by linear regression for the pairs $(\log C_m(r_k), \log r_k)$ over all the successive triads $\{(r_1, r_2, r_3), (r_2, r_3, r_4), \ldots, (r_{17}, r_{18}, r_{19}), (r_{18}, r_{19}, r_{20})\}$ 18 values of $D_{2,m}$ come up. In the same fashion these estimates are again binned into 12 successive heptads and from each one of them a mean value and a standard deviation for $D_{2,m}$ is derived. The mean value having the smallest standard deviation consists the correlation dimension of the data set at a given m.

The same procedure is then repeated for successive values of m ($m = 1, 2, \ldots, 20$) and in the case of a non stochastic system the representation of $(\log C_m(r_k), \log r_k)$ at small r_k after some embedding dimension and on exhibits the same slope. This means that $D_{2,m}$ gradually reaches a plateau (i.e. remains unchanged) at a certain value of m_{pl} and on. Hence, $D_{2,m}$ converges finally to a constant value consisting the value of the correlation dimension of the system, D_2. As shown by Ding et al. (1993) the onset of this convergence is starting for $m \geq D_2$ and therefore $m_{\rm pl}$ equals to the smaller integer greater or equal for D_2, $m_{\rm pl} = \lceil D_2 \rceil$. This practically means that for a system which has a correlation dimension of e.g. 2.6 the onset of the plateau will be at $m_{pl} = 3$.

Consider 1000 realizations of the x_t component of the Hénon dynamical system (eq.3.41). The various estimates of the correlation integral at various embedding dimensions m for several radii r_k are shown in the left panel of fig.3.16. Concerning the reconstruction of the phase space the optimum time delay is set to $\tau = 3$ based on the method of higher order correlations (fig.3.13, left panel). In the right panel of fig.3.16 the slopes, i.e. $D_{2,m}$, of the successive estimates are shown as a function of radius r_k (i.e. mean value of every heptad). It can be seen that for small values of r_k ($\sim 10^{-1.35}$) the estimates of the correlation dimension for $m > 1$ converge around a mean value of 1.206 ± 0.003. A more detailed estimation of it can be achieved by following the aforementioned methodology after depicting from each heptad the estimate with the smaller standard deviation for each m. Fig.3.17 shows the various estimates of the $D_{2,m}$ as a function of m together with the plateau. The onset of the plateau is at $m_{\rm pl}$=2 determining a value of the correlation dimension $D_2 = 1.252 \pm 0.003$, which is in accordance with the already well known value of the system's correlation dimension $D_2 = 1.25 \pm 0.02$ (Grassberger & Procaccia 1983a). Therefore this result agrees completely with the fact that the system can be fully described by $\lceil D_2 \rceil = 2$ difference equations. This is exactly the interesting information that one should try to extract initially from a given data set. Are the observed data a result of a low-dimensional process or they consist the outcome of a high-dimensional system where the big number of the contributing parameters classifies it as a stochastic process?

In true physical systems the lack of stabilization for the values of $C_m(r_k)$ can imply apart from the stochasticity either the existence of noise in the data set (dynamical and/or observational, see sect.3.1.2) and/or few number of data points. Concerning the former as shown by Ben-Mizrachi et al. (1984) one should expect a broken power-law behavior of $C_m(r)$ versus r, instead of a simple power-law (eq.3.57).

---

[5] Note that when $C_m(r_k)$ is used as a nonlinear discriminating statistic e.g. in the method of surrogates (see sect.6.3.3 and Theiler et al. 1992) then λk =const.=50%.

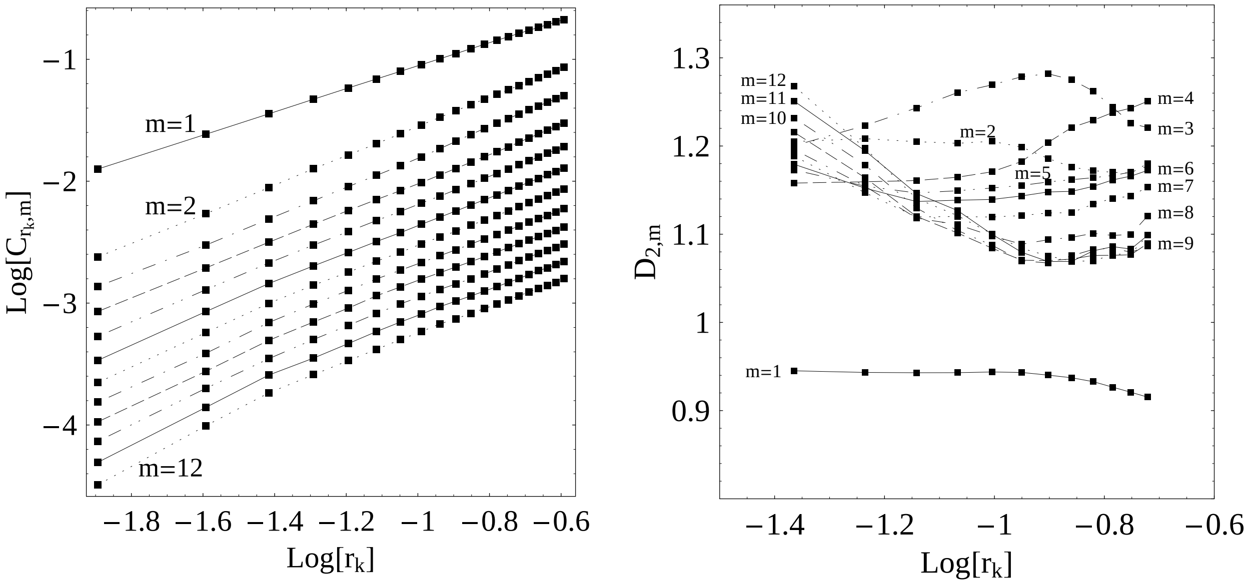


Figure 3.16: [Left panel] The logarithm of the correlation integral $C_m(r_k)$ as a function of the logarithm of the radius $r_k = k \cdot 2.557 \cdot 0.005$ ($k = 1, \ldots, 20$) for $m = 1, \ldots, 12$ (top to bottom plot). The different line-styles correspond to different m in order to be easier to find the correspondence with the plot in the right panel. Every five dimensions the same line-style is repeated.
[Right panel] The correlation dimension $D_{2,m}$ as a function of the logarithm of the radius r_k. Note that for $m > 1$ and for small radius r_k the values cluster around 1.206 ± 0.003.

The first slope should be equal to the embedding dimension and the second one should be equal to the correlation dimension of the data set without the noise influence. The identification of such slopes is a quiet difficult and tricky issue due to the fact that as the radius r increases it tends to occupy the whole extension of the attractor in the reconstructed phase space (due to the finite number of points) yielding a slope for the correlation integral, D_2, equal to zero. For the latter reason, concerning the lack of stabilization, one should always consider the maximum value of the correlation dimension $D_{2,\max}$, that is feasible to be estimated given a data set of N elements. Based on Eckmann & Ruelle (1992) $D_{2,\max} = 2 \log N$ and bigger values should be excluded since they are not meaningful. That practically means that from a data set of 1000 points one can estimate physical meaningful correlation dimensions up to $D_{2,\max} = 6$.

Physical interpretation of the method

The physical concept behind the aforementioned methodology concerning the estimation of D_2 can be understood as it follows (fig.3.18). Consider the genuine phase space of a 2-dimensional dynamical system $\{x(t), y(t)\}$ (fig.3.18, top-left panel) and suppose that only the $x(t)$ component can be observed i.e. only this is projected through the function h (see fig.3.10). From this component the reconstructed phase space of the system in one dimension is then formed (fig.3.18, top-right panel). The line element r_k in one dimension plays the role of *hyper sphere* (circle in 2-dimensions, sphere in 3-dimensions) and the arrows indicate the points that have a distance smaller than the one indicated by the line element i.e. correlated values. These points appear to be *close neighbors* but actually after further immersion of the system into two dimensions only two pairs remain as “true” correlated values (i.e. “true” *close neighbors*). Embedding the system into three dimensions does not change anything since the distances of all the points remain the same and thus the number of the correlated points remains the same as that in two dimensions. This happens simply because during the transition to a bigger dimension the distances of two points (eq.3.53) can either be increased or at the very best remain the same. This means that the probability (i.e. the correlation integral) of having a “false” *close neighbor* diminishes as the embedding dimension becomes bigger.

Therefore through immersions of the data set into successive embedding dimensions m the overall structures of the formed attractor start to remain unchanged at the point when the genuine dimensionality of the system has been reached. The attractor does not “need” to be extended further to more dimensions since all its points are adequately and equivalently described as the ones in the genuine phase space.

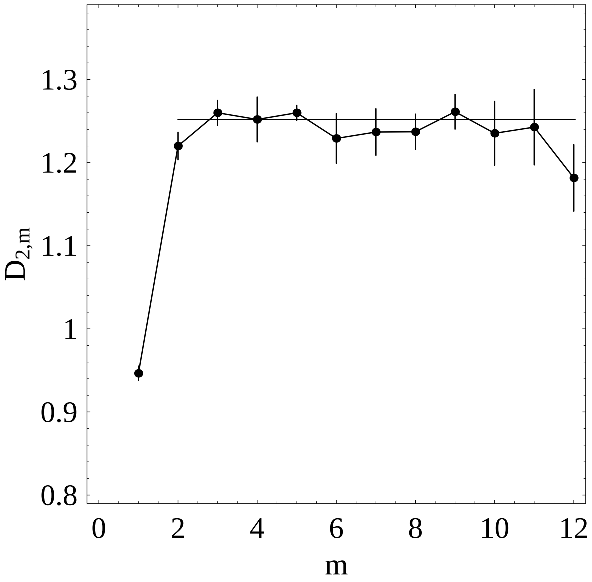


Figure 3.17: The correlation dimension $D_{2,m}$ as a function of the embedding dimension m. The solid line indicates the plateau determining a value of the correlation dimension $D_2 = 1.252 \pm 0.003$, as this yields from linear regression ($\chi^2 = 9.37$ for 10 DOF yielding a NHP=0.497).

The "projection" effects that are responsible for the "false" *close neighbors* (i.e. "false" correlated pairs) gradually diminish and at the end the structure of the attractor stabilizes making it equivalent with the genuine attractor and thus appropriate to extract information about the physical parameters of the system.

3.5 The Principal Component Analysis

The *principal component analysis* (Prin.Comp.Anal.) is an efficient method that can be used in order to reconstruct the phase space of a dynamical system, given a realization of it in a form of time series. The method is based on the computation of an algorithm known as *singular value decomposition* (SVD) used initially in the generalized theory of information by Bertero & Pike (1982). The method was introduced in the field of time series by Broomhead & King (1986) after studying the well known dynamical system of Lorenz (eq.3.42). The method as such has some advantages over the MOD but it suffers as well from some drawbacks. The main advantage of reconstructing the phase space following the Prin.Comp.Anal. is that the method is able to deal with small and noisy data sets (Vautard et al. 1992). Also the selection of an optimum time window t_w is much more flexible than in the case of MOD (sect.3.4.3) since a time delay of $\tau = 1$ (i.e. equal to the sampling rate) is sufficient for almost all the cases. The main disadvantage of the method, as it is claimed by Paluš & Dvořák (1992), is that the SVD is by nature a linear method therefore it can give distorted and misleading results when dominant nonlinear structures are present. Despite this claim up to know the method have been successively used in a wide field of nonlinear sciences (e.g. atmospheric physics, stock market analysis, solitonic physics) yielding robust results. Of course the dynamical noise of a data set is not something that can be always treated linearly but us a first approximation the method can give an idea about the dimensionality of the system especially when there is an *a priori* knowledge/indication of a noisy data set.

The physical interpretation of the method is the undermentioned. Consider a 2-dimensional dynamical system that suffers the influence of a small amplitude white noise process (not necessarily Gaussian). The trajectories in the phase space are mainly driven by the two main components (since the system is 2-dimensional) but the noise process perturbates slightly their course by almost the same displacement towards the two directions. Due to the stochasticity of the noise process no direction towards the two dimensions of the genuine phase space is preferred. As it is discussed in sect.3.4.2 the reconstructed phase space consists actually the projection of a d-dimensional manifold into m dimensions through an embedding function $\mathbf{\Phi}$. Once again the question is: What is d? i.e. what is the dimensionality of the genuine system? Note here that d is equal to the number of axes (independent variables) of the genuine phase space. Based on the Prin.Comp.Anal. the points form a *hyperellipsoid* in the m-dimensional phase space and by

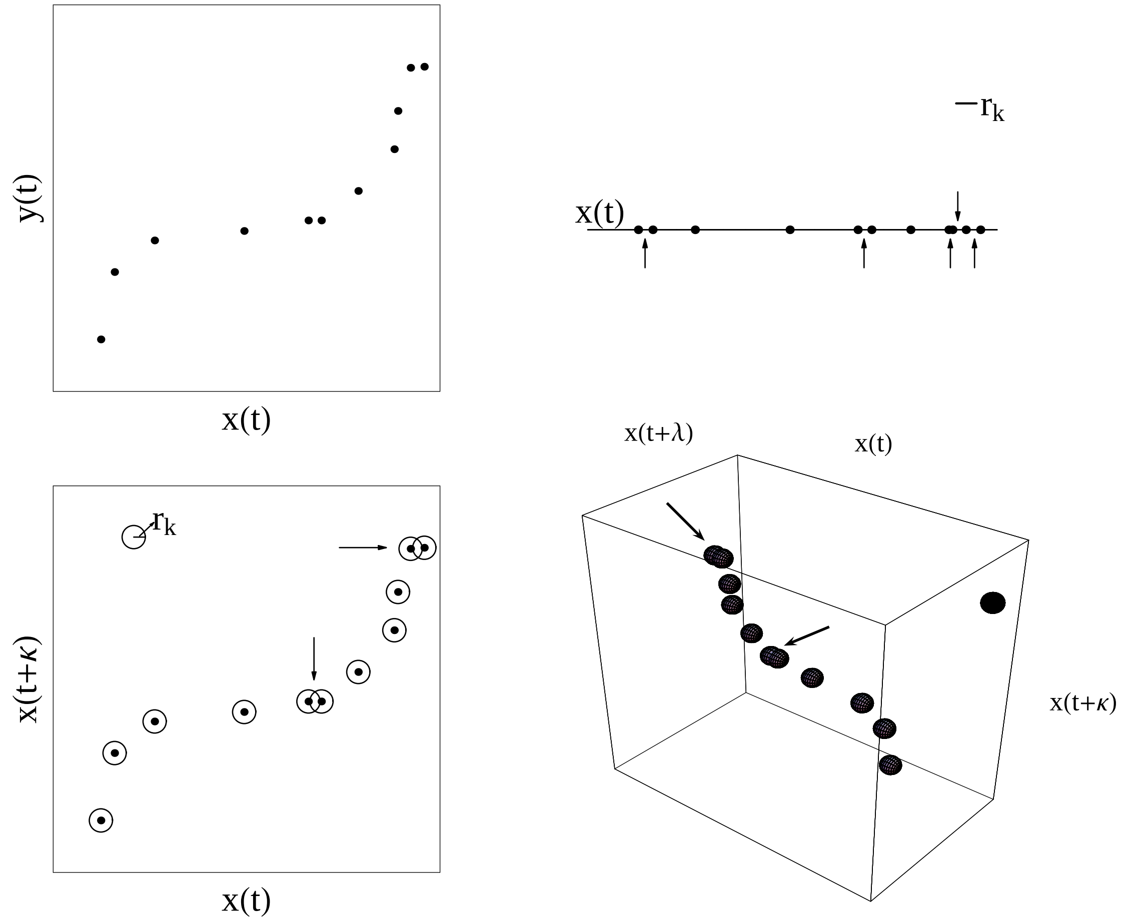


Figure 3.18: [Top-left panel] The genuine phase space consists of two components $x(t), y(t)$ (2-dimensional). [Top-right panel] Reconstruction of the system in one dimension based on $x(t)$. Estimation of the correlation integral for a given radius r_k, being a line element in this space, yields five correlated pair points indicated by the arrows.

[Bottom-left panel] Reconstruction of the system in two dimensions based on $x(t)$. Estimation of the correlation integral for a given radius r_k, yields two correlated paired values indicated by the arrows. The circle at the top left corner of the plot indicate the volume element in the 2-dimensional space, being an *area*.

[Bottom-right panel] Further embedding into three dimensions does not change anything since the number of the correlated pairs remains the same as the one in two dimensions i.e. 2, corresponding to the number of the "true" correlated values in the genuine phase space for the given radius r_k. The sphere at the top right corner of the plot indicate the volume element of radius r_k in the 3-dimensional space.

direct specification of the main *hyperaxes* one can distinguish the major components, consisting the physical interesting part of the system, from the noise components. Consider that the aforementioned 2-dimensional system lacking the influence of the noise component. Then the Prin.Comp.Anal. will result an ensemble of points forming an ellipse having 2 main axes. Consider now the presence of the noise, the resulting pattern will be a *hyperellipsoid* with more than three main axes, two "big ones" representing the two major components and a lot of smaller ones, of almost equal length, representing the noise influence. For higher dimensional dynamical systems one can specify the entire group of the main *hyperaxes* and plot them in decreasing order. Such a plot shows that gradually as the main axes-values are decreasing a plateau is formed indicating the onset of the noise component. Application of the Prin.Comp.Anal. to a white noise process in the m-dimensional phase space will result an m-dimensional *hypersphere*, being the version of a *hyperellipsoid*, having this time equal all the main axes.

3.5.1 Methodology

Consider a time series data set x_t $(t = 1, \ldots, N)$ which is embedded into m dimensions with a time delay τ (usually $\tau = 1$). Then as in MOD the delay vectors $\vec{X}_i = \{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\}$ for $i = 1, 2, \ldots, L$ with $L = N - (m-1)\tau$ are formed.

1. Initially the real $L \times m$ matrix, known as trajectory matrix, is formed

$$T = \frac{1}{\sqrt{L}} \begin{bmatrix} \vec{X}_1 \\ \vec{X}_2 \\ \vdots \\ \vec{X}_L \end{bmatrix} = \frac{1}{\sqrt{L}} \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(m-1)\tau} \\ \vdots & \vdots & & \vdots \\ x_L & x_{L+\tau} & \cdots & x_{L+(m-1)\tau} \end{bmatrix} \tag{3.59}$$

   and the mean value of each column is subtracted from the corresponding column resulting T_0. The matrix T and its transpose may be thought as linear maps between $\mathbf{M}^d$ and $\mathbb{R}^m$.

2. The next step is to find the SVD of the numerical matrix T_0. The decomposed matrix is written into the form

$$T_0 = U \cdot W \cdot V^{\mathrm{T}} \tag{3.60}$$

   - U is an $L \times L$ orthogonal matrix (i.e. consists of orthogonal columns: $U^T U = U U^{\mathrm{T}} = I_L$ where the symbol $^{\mathrm{T}}$ denotes the transpose of the matrix and I_L is the identity matrix of size L) having as columns the rows of the eigenvector matrix of $T_0 T_0^{\mathrm{T}}$ (left eigenvectors of T_0). In the general case where T_0 is not a real matrix U is an $L \times L$ unitary matrix.
   - W is a $L \times m$ matrix with non negative numbers on the diagonal w_i (singular values of T_0) and zeros off the diagonal. The singular values on the diagonal are sorted from the biggest to the smallest value and they are equal to the square root of the eigenvalues of the matrix $T_0 T_0^{\mathrm{T}}$ (or $T_0^{\mathrm{T}} T_0$).
   - V is an $m \times m$ orthogonal matrix having as rows the columns of the eigenvector matrix of $T_0^{\mathrm{T}} T_0$ (right eigenvectors of T_0). The matrix $\Xi = T_0^{\mathrm{T}} T_0$ is called covariance matrix of T_0. It consists an estimation of the mean time correlation (i.e. not spatially) between all the pairs in the m-dimensional phase space. In the general case where T_0 is not a real matrix then in eq.3.60 instead of V^{T} the conjugate transpose of V appears, V^*.

   From this procedure only the diagonal elements of W, w_j and the matrix V are of interest.

3. The singular values of T_0, w_j are then normalized

$$\sigma_j = \log \left(\frac{w_j}{\sum_{j=1}^{m} w_j} \right) \tag{3.61}$$

4. The dot product of each m-dimensional row-vector element of matrix V, $\mathbf{v_j}^{\mathrm{T}}$ (the $\mathbf{v_j}$ consists a column element of the matrix V^{T}) , with the m-dimensional delayed vector $\vec{X}_i$ is then estimated

$$y_{i,j} = \mathbf{v_j}^{\mathrm{T}} \vec{X}_i \tag{3.62}$$

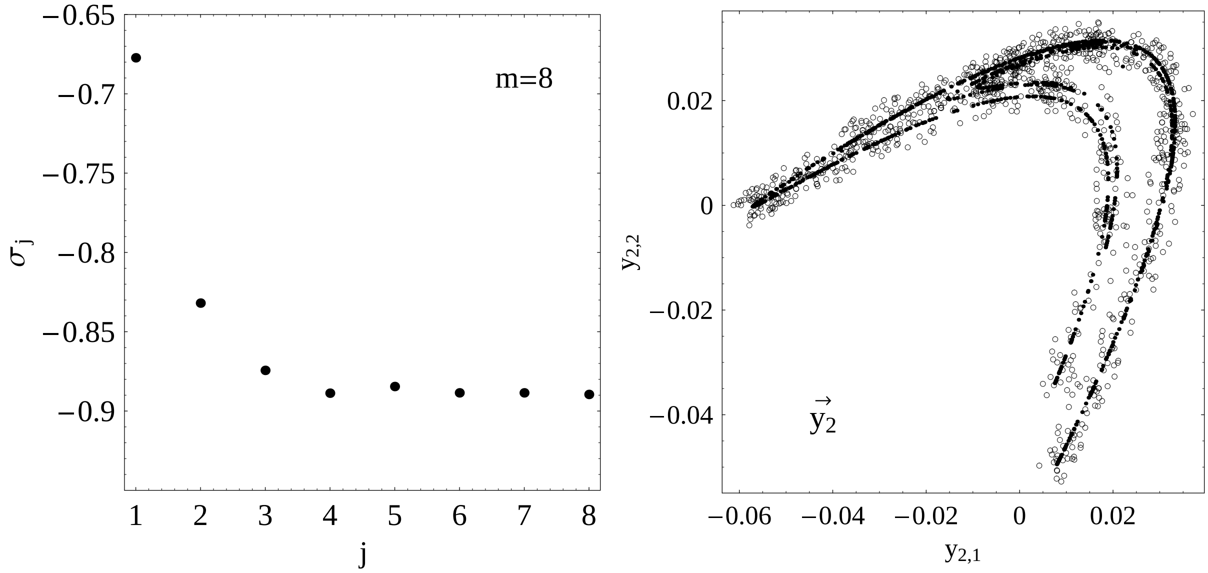


Figure 3.19: [Left panel] The logarithm of the normalized singular vectors for $m = 8$. A plateau after $j \sim 2-3$ is formed revealing that the dimensionality of the system is between 2 and 3. The onset of the plateau is due to the noise level.
[Right panel] The reconstructed Hénon system into $m = 2$ dimensions based on 1000 realizations of the x_t component with noise (open circles) and without noise (filled circles).

with $i = 1, \ldots, L$ and $j = 1, \ldots, m$, representing the value of the $\vec{X_i}$ vector in the j dimension of the m-dimensional reconstructed phase space. Then the singular vectors can be formed at a given embedding dimension m

$$\vec{y_m} = \{y_{1,m}, y_{2,m}, \cdots, y_{L,m}\} \tag{3.63}$$

comprising the reconstructed phase space of the attractor at the given m. The $\vec{y_m} = VT_0$ consist actually the projection of T_0 (i.e. phase trajectory) in the basis of the singular vectors of the m-dimensional space.

The first singular values w_j, $(j = 1, \ldots, q)$ are large and they characterize the main trajectory in the reconstructed m-dimensional phase space of the attractor. The other $m - q$ singular values represent the noise level which needs theoretically infinite number of components (i.e. dimensions) in order to be fully represented due to its stochastic nature.

Consider once again 1000 realizations of the x_t component of the Hénon dynamical system (eq.3.41) but this time an additional white noise component is added coming from a uniform distribution in the range $[-0.1, 0.1]$. The time delay is set to $\tau = 1$ and the embedding dimension is set to 8 since the genuine dimensionality of the system is only 2. The results of the Prin.Comp.Anal. are shown in the left panel of fig.3.19. It can be seen that after $j \sim 2-3$ a plateau is formed being indicative for the existence of the noise. The dimensionality of the Hénon system can therefore being derived to be between 2 and 3 being consistent with the genuine dimensionality of the system. The right panel of fig.3.19 shows the reconstructed attractor (open circles) based on the two dimensional singular vectors, $\vec{y_2}$. In the same plot the $\vec{y_2}$ vectors are plotted for the x_t component but this time without the noise influence (filled circles). The bulk structure of the attractors coincide for scales grater than ~ 0.02 length units. Despite the lack of trajectory definition, concerning the noisy version of the x_t component, the Prin.Comp.Anal. is able to distinguish the deterministic part of the Hénon system from the noisy component.

3.6 Long-Term Memory Analysis

3.6.1 The existence of cycles and the Hurst exponent

The first analysis method usually applied in a time series data set coming from a given source is the PSD (sect.3.3.1). The application of this method indirectly implies the existence of patterns in the data set that can be broke down into a sum of sinusoids with different frequencies and amplitudes. Unfortunately there is no intuitive reason for believing that the underlying radiation mechanism of an AGN producing the observed AGN flux is structured on the basis of periodic cycles. This can be readily seen from the form of their PSD exhibiting no significant peaks (sect.3.1.1).

In the case that a multicomponent physical system, such the one of an AGN (sect.1.2.1), does not exhibit any periodic cycles at all within its realizations, possible nonperiodic cycles might still be present. These nonperiodic patterns usually demonstrate the *same statistical behavior* without having neither the same time duration nor fixed occurrence times in comparison to the classical periodic systems. The term "same statistical behavior" implies that the system retains a kind of memory (long-term memory) concerning its past variability behavior, which can be expressed through a statistical quantity (e.g. the variance) and continues its course in time accordingly to the existence tendency.

The existence of a long-term memory embedded in a time series data set is usually studied through the estimation of the Hurst exponent H (Hurst 1951) based on the method of *rescaled range analysis*, (R/S) (Mandelbrot 1972). Roughly speaking the method consists of the following steps. The data set is partitioned into non-overlapping subsets of the same length N_A and then the cumulative deviation $\mathcal{D}_A$ is estimated for each one of them. Finally the statistical range R_A and the standard deviation S_A for all the subsets are computed and a mean value of their ratio $(R/S)_{N_A}$ is derived. The same exactly procedure is then repeated having another partitioning configuration of different N_A. Then H can be estimated by linear regression as it consists the slope of the straight line passing through the points $((R/S)_{N_A}, N_A)$ in a *log-log* representation.

Under the null hypothesis of no long-term dependence $H = 0.5$, meaning that the data set is simply the realization of a random walk process (i.e. a sequence of discrete steps of fixed length, sect.3.1.1). For data sets exhibiting positive long-term dependence (persistency) a value of $H > 0.5$ is expected in contrast to antipersistent behavior in which $H < 0.5$. Time series characterized by persistency (antipersistency) scale faster (slower) than the square root of time (sect.3.1.1) meaning that the system covers more (less) "distance" than a random walk process. A persistent behavior implies that an increasing or decreasing trend will probably continue the same course in the next time instant Δt in contrast to an antipersistent behavior which tends to reverse itself around the mean. The Hurst exponent H gives exactly the probability of a time series to continue the same course (increasing or decreasing) during the next time step Δt. Note here that the existence of a long-term memory does not imply any periodic behavior but rather a circular dependence of similar statistical behavior (i.e. persistency or antipersistency) in time scales having unequal duration. Exactly these nonperiodic cycles are characterized by an average time duration but within these cycles the various flaring events occur in various time scales.

3.6.2 Methodology

Estimation of the Hurst exponent through the rescaled range analysis

Consider a stationary time series data set (sect.3.1) consisting of N measurements, x_i, measured at discrete times t_i with $(i = 1, \dots, N)$ separated by Δt time units.

Starting from the beginning, the data set is partitioned into A non-overlapping subsets of N_A successive number of points. This procedure is performed for $N_A = 2, \dots, N$ and the last points which can not form a complete subset are simply dropped. The total number of subsets for a given N_A is equal to the greatest integer less than or equal to N/N_A, therefore $A = 1, \dots, \lfloor N/N_A \rfloor$. At the end there are $N-1$ groups each one consisting of A subsets of N_A points. The total number of points within each group is $A \times N_A$ and should be the closest possible to N.

Initially the arithmetic mean, the standard deviation (see eq.3.19) and the cumulative deviation respectively for every subset (among the A) consisting of N_A points within a group, is estimated

$$\overline{x_A} = \frac{\sum_{n=1}^{N_A} x_{n,A}}{N_A} \tag{3.64}$$

$$S_A = \sqrt{\frac{\sum_{n=1}^{N_A} (x_{n,A} - \overline{x_A})^2}{N_A - 1}} \tag{3.65}$$

$$\mathcal{D}_A = \sum_{n=1}^{k} (x_{n,A} - \overline{x_A}) \;\text{ for }\; k = 1, \dots, N_A \tag{3.66}$$

Then, the statistical range of $\mathcal{D}_A$, R_A, is estimated and it is normalized to the S_A. That gives the rescaled range statistic for a given subset of a given group

$$\frac{R_A}{S_A} = \frac{1}{S_A}\left(\max[\mathcal{D}_A] - \min[\mathcal{D}_A]\right) \tag{3.67}$$

Finally the mean value of the rescaled range statistics is computed for all subsets A within the same group and the same exactly computations are performed for all the $N-1$ groups

$$\left(\frac{R}{S}\right)_{N_A} = \frac{\sum_{A=1}^{\lfloor N/N_A \rfloor}\left(\frac{R_A}{S_A}\right)}{\lfloor N/N_A \rfloor} \quad \text{for} \quad N_A = 1, \ldots, N-1 \tag{3.68}$$

The Hurst exponent H for the initial data set is then computed based on the following relation

$$\left(\frac{R}{S}\right)_n = Cn^H \tag{3.69}$$

and by taking the logarithms of both sides this yields

$$\log\left(\frac{R}{S}\right)_n = \log C + H\log n \tag{3.70}$$

The plot of $(R/S)_n$ versus n in a *log-log* representation (i.e. eq.3.70) consists of a steep curve followed at the point $(N_p, (R/S)_{N_p})$ by a plateau in case that the system exhibits indeed a long-term memory. The abscissa of this point N_p (inflection point or local maximum) defines the mean time duration (i.e. $N_p \times \Delta t$) when the long-term memory of the system starts to dissipate. The corresponding group, consisting of subsets each one being $N_p \times \Delta t$ time units long, displays the biggest deviations from the mean and therefore it will be the one with the dominant trend. Practically the Hurst exponent H is equal to the slope of the linear regression model fitted to the ensemble of points $\{(2, (R/S)_2), (3, (R/S)_3), \ldots, (N_p, (R/S)_{N_p})\}$.

The aforementioned methodology is known as *classical R/S analysis* since it is the original one proposed by Mandelbrot (1972) for the estimation of the Hurst exponent. In general there is a big variety of methods aiming to a more robust and less biased estimation of H (e.g. Lo 1991; Moody & Wu 1991; Hauser 1997). All methods different from the aforementioned analysis only in the normalization factor of R_A, S_A (eq.3.67). As it was shown from Davies & Harte (1987) the conventional R/S analysis using a Hurst regression can be biased towards accepting a long-term dependence hypothesis even when the true process is first order autoregressive (AR) (e.g. Priestley 1981). Moreover another crucial issue is that the measurements of a data set might exhibit short-range dependences (i.e. autocorrelations) fact that make them depended. Usually small data sets have statistical different properties from their parent distributions in the sense that they might have significant different S_A from their parent distribution (heteroscedasticity) (Lancaster 1968; Levenbach 1973). A final matter is that the various estimates $((R/S)_{N_A}, N_A)$ might be distributed around non-Gaussian distributions converging very slowly as $N_A \to \infty$ to Gaussian distributions. All these issues give raise to biases concerning the estimations of the coefficients of the linear regression model. In order to incorporate these biases Lo (1991) proposed the *adjusted R/S analysis* which uses instead of S_A (eq.3.65) the quantity $S_{\mathrm{L}A}$

$$S_{\mathrm{L}A} = \sqrt{\hat{\gamma}(0) + 2\sum_{i=1}^{q}\left(\left(1 - \frac{i}{q+1}\right)\hat{\gamma}(i)\right)} \tag{3.71}$$

where $\hat{\gamma}(i)$ is the sample autocovariance function (eq.3.4) but this time divided by $N_A - 1$ instead of N_A (see sect.3.3.3). The reason is simply that for $i = 0$ this relation should be identical to the one of the standard deviation S_A (eq.3.65).

$$\hat{\gamma}(i) = \frac{\sum_{j=1}^{N_A - i}(x_{j,A} - \overline{x_A})(x_{j+i,A} - \overline{x_A})}{N_A - 1} \tag{3.72}$$

and

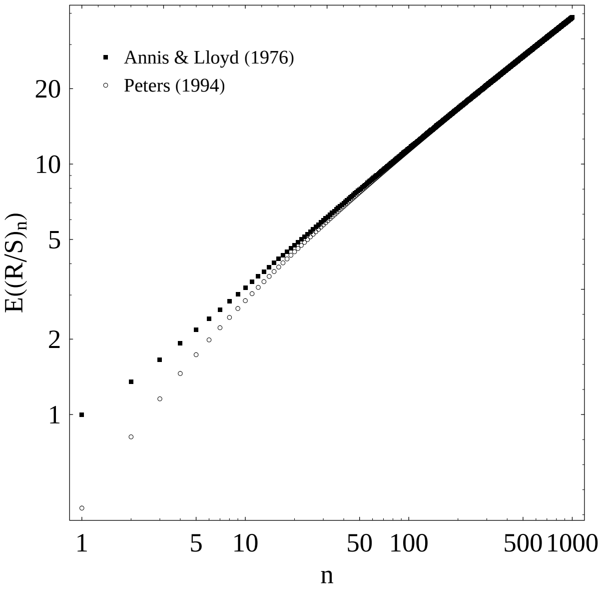


Figure 3.20: The expected values of $(R/S)_n$, $E\left((R/S)_n\right)$, as they are estimated from eq.3.77 (Anis & Lloyd 1976) and eq.3.79 (Peters 1994).

$$q = \left\lfloor \left(\frac{3N}{2}\right)^{1/3} \left(\frac{2p}{1-p^2}\right)^{2/3} \right\rfloor \tag{3.73}$$

with p being the first order autocorrelation coefficient of the data $p = \hat{\gamma}(1)/\hat{\gamma}(0)$. Note here that eq.3.71 for $q = 0$ gives the usual estimate of the standard deviation eq.3.65.

V statistic

The V statistic (Peters 1994) gives the probability of having a long-term behavior embedded in the under study data set. A one dimensional random walk process is known to have an average distance between N successive steps equal to $\sqrt{2N/\pi} \approx \sqrt{N}$ (sect.3.1.1).

In the general case of having a stationary data set consisting of N points the term average distance can be expressed by means of eq.3.68 which is actually expressing the biggest deviation of the subsets, consisting of N_A points, from the mean value. If the data set is a realization of a random process then $(R/S)_{N_A}$ is expected to scale as $(R/S)_{N_A} \propto \sqrt{N_A}$. Therefore the variable V_{N_A} is defined having an arithmetic mean value u which is estimated from all the $N-1$ groups where

$$V_{N_A} = \frac{(R/S)_{N_A}}{\sqrt{N_A}} \tag{3.74}$$

$$u = \frac{\sum_{N_A=2}^{N} V_{N_A}}{N-1} \tag{3.75}$$

and the probability distribution function of V (Lo 1991) is given by

$$F_V(u) = 1 + 2\sum_{l=1}^{\infty}(1 - 4l^2u^2)e^{-2(lu)^2} \tag{3.76}$$

having a mean value and standard deviation of 1.25 and 0.27 respectively. After estimating u the probability of having in the data set a long-term memory behavior is computed from eq.3.76.

The expected value of R/S

Apart from the fact that a small data set might not be representative of the statistical properties of the underlying mechanism, its moderate length can also affect the result of H i.e. overestimating/underestimating it. In this sense the outcome of the method is up to a certain degree biased due to the limited number of observations. Therefore in order to have an idea about the degree of this bias the R/S method is estimated for a random walk process with equal length with the under study data set, yielding the expected values of $(R/S)_n$, $E\left((R/S)_n\right)$ for every n. These values then can be directly compared with the ones produced directly from the true data set.

The $(R/S)_n$ statistic for a random walk process (eq.3.70) is an asymptote such that the value of the Hurst exponent is gradually converging to $H = 0.5$ as the length of the $N-1$ groups is increasing $(N \to \infty)$. Thus, in order to reject or accept the null hypothesis (i.e. random walk) the estimated values of the $(R/S)_n$ statistic, coming from the data set, should be compared to their expected value $E\left((R/S)_n\right)$, coming from a random walk process having the same number groups (i.e. points). However in order to estimate correctly $E\left((R/S)_n\right)$ one should take into account the aforementioned biases which will result at the end $H \neq 0.5$ even for the null hypothesis due to the limited length of the data set.

The expected value of $(R/S)_n$ was firstly estimated by Anis & Lloyd (1976)

$$E\left((R/S)_n\right) = \frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}} \tag{3.77}$$

where $\Gamma(x)$ is the (complete) Gamma function.

Peters (1994) showed that eq.3.77 performs poorly for $n < 20$ and hence he multiplied it with an empirically derived correction term

$$E\left((R/S)_n\right) = \frac{n-0.5}{n} \frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}} \tag{3.78}$$

Based on the Stirling's approximation $n! \approx \sqrt{2\pi n}(n/e)^n$ for $n \to \infty$, eq.3.78 reads

$$E\left((R/S)_n\right) = \frac{n-0.5}{n} \sqrt{\frac{2}{n\pi}} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}} \tag{3.79}$$

From fig.3.20 it can be seen that eq.3.77 for $n < 20$ overestimates $E\left((R/S)_n\right)$. The overestimation factor, simply by subtracting the two plots, goes as $\sim n^{-0.39}$.

Chapter 4

X-RAY OBSERVATIONS OF MRK 421

This chapter contains a detailed description of the PCA and ASM data sets of Mrk 421 which are going to be used in the next chapters in order to study the short and long term variability behavior of the source in the X-ray band. Some cautions concerning the data analysis procedures should be taken since the observations were obtained with different electronic configurations of the RXTE satellite throughout the years. Another important issue that is discussed here is the fact that ASM instrument is not in general as sensitive as the PCA and thus there are a lot of accusations in the literature, based on rather superficial criteria, concerning its reliability. A direct comparison between the two instruments is conducted via their data sets in order to check whether the ASM observations of Mrk 421 are indeed representative of the source state.

4.1 The Proportional Counter Array Data Set

All the archival data of Mrk 421, obtained by the PCA instrument (sect.2.1.2) on board RXTE since its launching date (December 1995), are selected and analyzed in a strictly homogeneous way. The whole data set consists of 14 campaigns and has an actual observational time of about 1.5 Msec (92468 observations of 16 sec) covering a time period of 3200 days. Tab.4.1 shows the exact time window of each observation together with the total observing time within it.

Table 4.1: Time coverage and total duration of the PCA observations of Mrk 421

	Earliest obs.		Latest obs.		Total obs. time
Proposal number	Date	Time	Date	Time	(in ksec)
10341	1/3/96	06:08:00	1/3/96	06:46:56	1,856
10345	15/3/96	13:18:24	19/4/96	02:01:04	57,808
20341	2/4/97	03:30:24	3/6/97	06:34:40	63,312
30261	24/3/98	00:52:16	13/4/98	07:37:36	33,520
30262	18/4/98	11:36:00	8/5/98	20:58:40	153,168
30269	26/2/98	13:27:44	25/7/98	22:40:48	86,064
40182	5/2/00	03:33:52	8/5/00	22:23:34	31,584
50190	24/01/01	01:04:48	6/2/01	16:10:08	52,624
60145	18/3/01	03:30:24	1/4/01	09:07:44	291,616
70161	2/12/02	04:33:04	14/1/03	13:10:08	110,912
80172	26/02/03	14:49:36	6/3/03	11:24:32	231,472
80173	20/02/03	03:05:52	13/05/04	05:41:04	261,136
90138	10/05/04	09:10:40	21/05/04	04:43:44	30,816
90148	18/04/04	03:26:08	22/12/04	12:17:36	73,600

Table 4.2: Time extension of the gain Epochs and the scaling factors

(a) For 2–5 keV and 2–10 keV

			Scaling factors	
Gain Epoch	Start time	End time	2–5 keV	2–10 keV
1	Launch	21/3/96-18:33:00	1.024±0.002	1.027±0.002
2	21/3/96-18:33:00	15/4/96-23:05:00	1.004±0.004	1.015±0.002
3	15/4/96-23:05:00	22/3/99-17:38	1.000	1.000
4	22/3/99-17:39	13/5/00 00:00:00	0.904±0.012	0.934±0.008
5	13/5/00-00:00:00	2006	0.857±0.003	0.892±0.005

(b) For 2–15 keV and 5–10 keV

			Scaling factors	
Gain Epoch	Start time	End time	2–15 keV	5–10 keV
1	Launch	21/3/96-18:33:00	1.028±0.001	0.993±0.003
2	21/3/96-18:33:00	15/4/96-23:05:00	1.018±0.002	1.020±0.003
3	15/4/96-23:05:00	22/3/99-17:38	1.000	1.000
4	22/3/99-17:39	13/5/00 00:00:00	0.938±0.001	1.030±0.002
5	13/5/00-00:00:00	2006	0.891±0.001	0.836±0.002

4.1.1 Homogeneity of the data set

Owing to operational constrains most of the PCUs are often switched off. The PCU3 and PCU4 are most of the time not in operational mode due to electrological problems occurred during October–November 1998. The automatic procedures that turn off these two PCUs in case of contingent breakdowns were not fully re-enabled after an earlier spacecraft computer crash. The PCU1 experienced a sequence of four break downs in March 1999 and in order not to cause a total break down of the detector is also switched off. Finally after May 2000 the pressure in the antico propane layer of PCU0 (fig.2.3) began decreasing resulting an increased background rate and a different gain. The xenon layer of PCU0 is still functional but the background rates are increased. The contamination events caused by the low-energy electrons entering the detector are now much more significant. This is particular relevant for variability studies of relatively week sources such as Mrk 421. Therefore in order to form the most homogeneous observational sample only those observations obtained by the PCU2 are selected and analyzed following the procedures described in detail in sect.2.2.

Despite the fact that all the selected observations were taken with the same detector, the electronic configurations of the PCA have been changed five times since 1996 and hence the gain settings of the instrument are different for these five epochs. These electronic adjustments (e.g. voltage modifications), made by GSFC/NASA, are mainly conducted in order to bring all the PCUs around the same sensitivity level after the detector's discharge problems. As a consequence all the source count rates, obtained during these $\sim$ 9 years, have to be rescaled to a common gain Epoch if one wants to compare possible different variability stages of the source in a statistical manner and make statements concerning the overall variability level of the source.

The rescaling of the X-ray detectors, such as the PCA, to a common gain level is usually done by comparing the flux level of a constant X-ray source throughout the years (e.g. Crab Nebula, Jahoda et al. 2006). Therefore the observations of the supernova remnant Cassiopeia A (Cas A), obtained during these five epochs, are analyzed and compared with respect to their X-ray count rate level. Initially the five Cas A data sets are reduced and a mean count rate with an error is derived for each epoch. Then the ratio of each epochal mean values to that of Epoch 3 is computed and the resultant rescaling factors are then used in order to calibrate the various observations from the data set of Mrk 421 to the latter gain Epoch (tab.4.2 (a) and (b)). The reason of choosing the gain settings of Epoch 3 as the optimum ones is that during this period the detector was in fully operational mode (apart from the last 6 months) and the response of the detector as well as the background model were performing better than any other epoch.

Note here that neglecting the gain corrections concerning the observation obtained before Epoch 3 results an underestimation of the count rate (in the 2–10 keV) band of $\sim 3.0\pm0.5\%$. For those observations taken after Epoch 3 the count rates are overestimated (6.0–10.0)±0.3%, depending on the source state. Tab.4.2 contains the time duration of these five epochs together with the scaling factors for different energy bands and tab.4.3 shows the gain Epochs for all the observations.

Another issue needed to be taken under consideration is the choice of the correct parametrized background model file. In sect.2.2.3 is mentioned that the selection of a background file ("faint" or "bright") strictly depends on the source state. During these ten years of observational coverage of Mrk 421, the source exhibits both "faint" and "bright" states therefore it is not optimum to use the same background file in order to produce the ensemble of scientific products. All the observations have been reduced with both background models and after checking the mean count rate it was then decided about the choice of the most appropriate model. Concerning Mrk 421 four out of the fourteen campaigns are characterized as "bright" and the remaining 10 as "faint" (tab.4.3).

It should be stretched out the fact that the background subtraction process is the dominant source of systematic errors during the reduction procedure. Since before choosing the appropriate background model the data are reduced using both models an overestimation factor can be derived for the case of using a "faint" model for a "bright" source state. This factor is negligible for the overall light curve 0.5±0.3% (lasting on average 90–100 ksec) but can be significantly increased for individual short lasting flaring events (lasting 1–2h) reaching up to the value of 8±0.4%.

Table 4.3: The gain Epochs and the background models for the PCA observations of Mrk 421

Proposal number	Gain Epoch	Background model file
10341	Epoch 1	Faint
10345[a]	Epoch 1,2,3	Faint
20341	Epoch 3	Faint
30261	Epoch 3	Faint
30262	Epoch 3	Faint
30269	Epoch 3	Faint
40182	Epoch 4	Faint
50190	Epoch 5	Bright
60145	Epoch 5	Bright
70161	Epoch 5	Faint
80172	Epoch 5	Faint
80173	Epoch 5	Bright
90138	Epoch 5	Faint
90148	Epoch 5	Bright

[a] During this observation the gain Epoch was changed three times.

4.1.2 Light curves

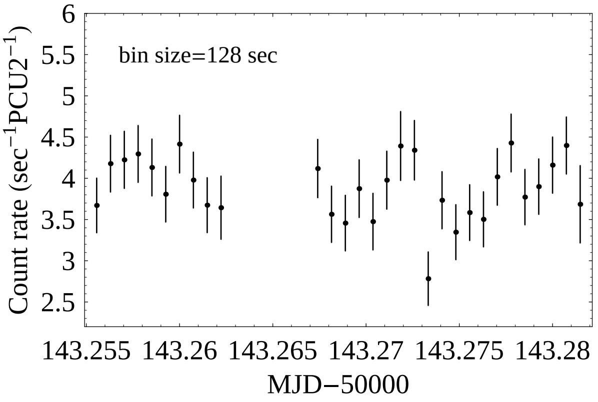
All the observing periods of Mrk 421, in the 2–10 keV energy range are shown in fig.4.1. In general the sampling of these public available data is highly uneven. The original observations were conducted having a wide variety of sampling patterns and duration due to the different scientific purposes. Mainly these proposals are part of multifrequency campaigns having sometimes very special and particular configurations concerning their observing plans. Simultaneous coverage of observations performed by Čerenkov arrays, optical telescopes, radio receivers as well as general XBL monitoring programs show the copiousness of the observing modes. At the same time some other proposals are triggered instantaneously either from the ASM or from ground-based Čerenkov telescopes whenever the X-ray or γ-ray flux of a source, exceeds some limit values.

The combined version of all the observations is shown in fig.4.2 characterizing fully the source states during these nine years. Morphologically, based on this plot there are 8 observing periods separated roughly by one year. Two huge flares separated by three years are the most dramatic events reaching up a count rate of 155 counts $\sec^{-1}$PCU2^{-1}. During the other periods the source seems to be in a relative quiescent state with a mean count rate around 30±8 counts $\sec^{-1}$PCU2^{-1}.

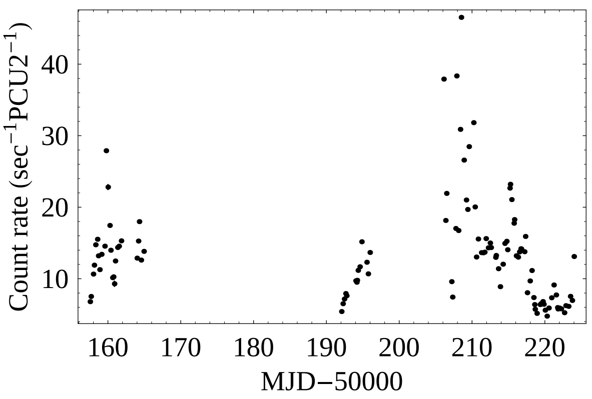
4.2 The All-Sky Monitor Data Set

More than 350 sources have been observed by ASM during the last 10 years. Galactic sources such as pulsars X-ray binaries and supernova remnants as well as extragalactic sources such as nebulae, Seyfert galaxies and blazars. Concerning the extragalactic objects ASM can be employed as a trigger whenever a source is on a flaring state since the data sets are available on a daily basis. Among the extragalactic sources

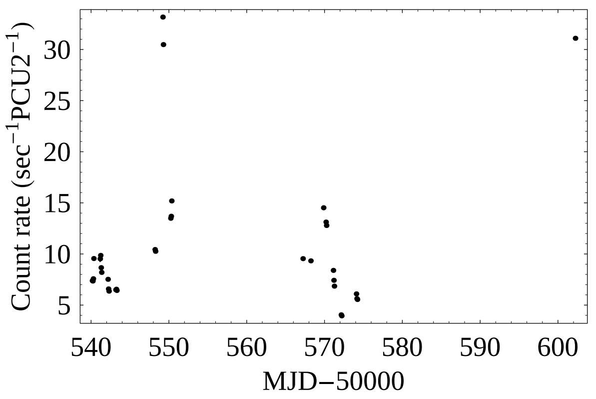
10341



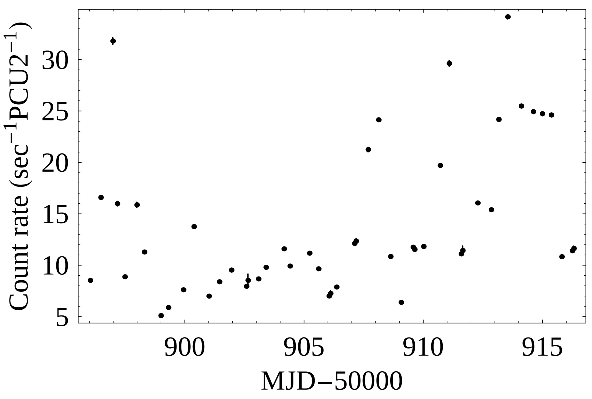
10345



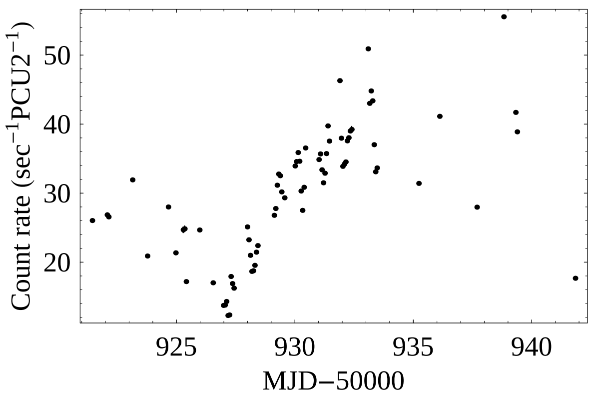
20341



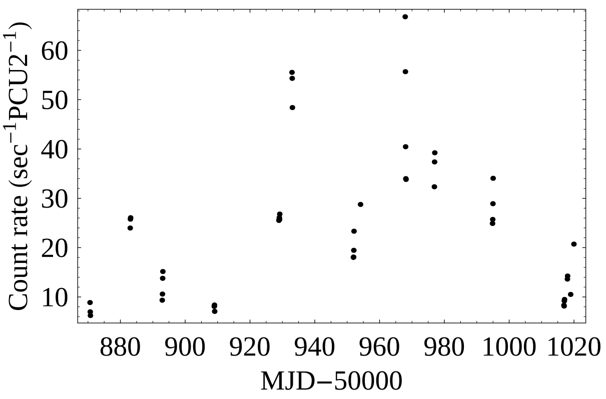
30261



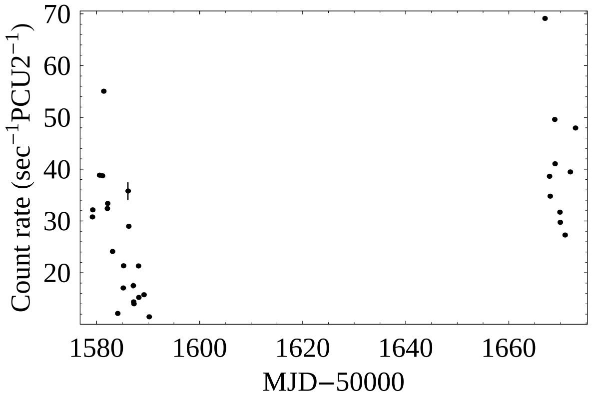
30262



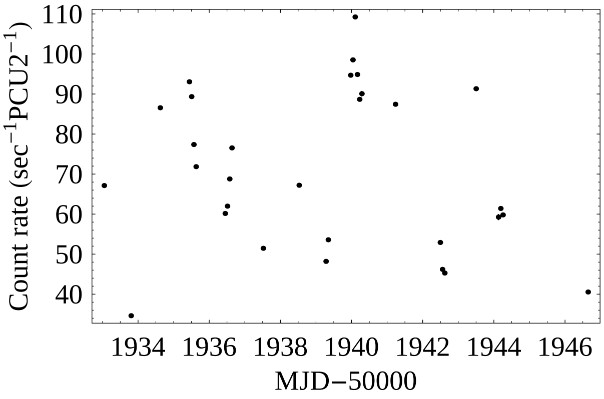
30269



40182



50190



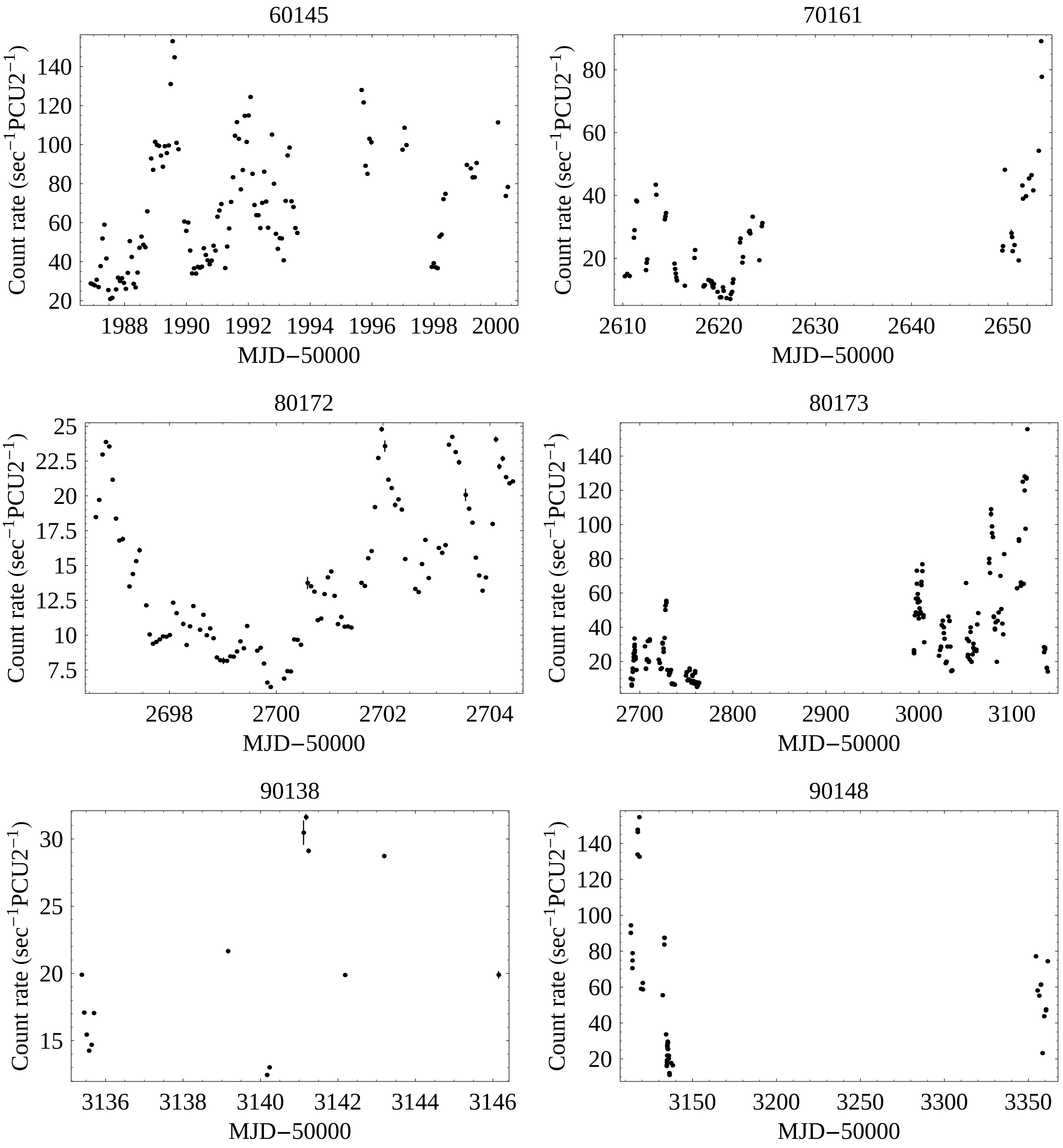


Figure 4.1: The individual observations of Mrk 421 during the period 1996-2005. All the count rates are given in the energy band of 2–10 keV and the bin size is 90 min (∼ 1 RXTE orbit). The only exception is the observation 10341 for which a bin size of 128 ksec has been used since it has a total time duration of ∼ 1.9 ksec i.e. less than an orbit.

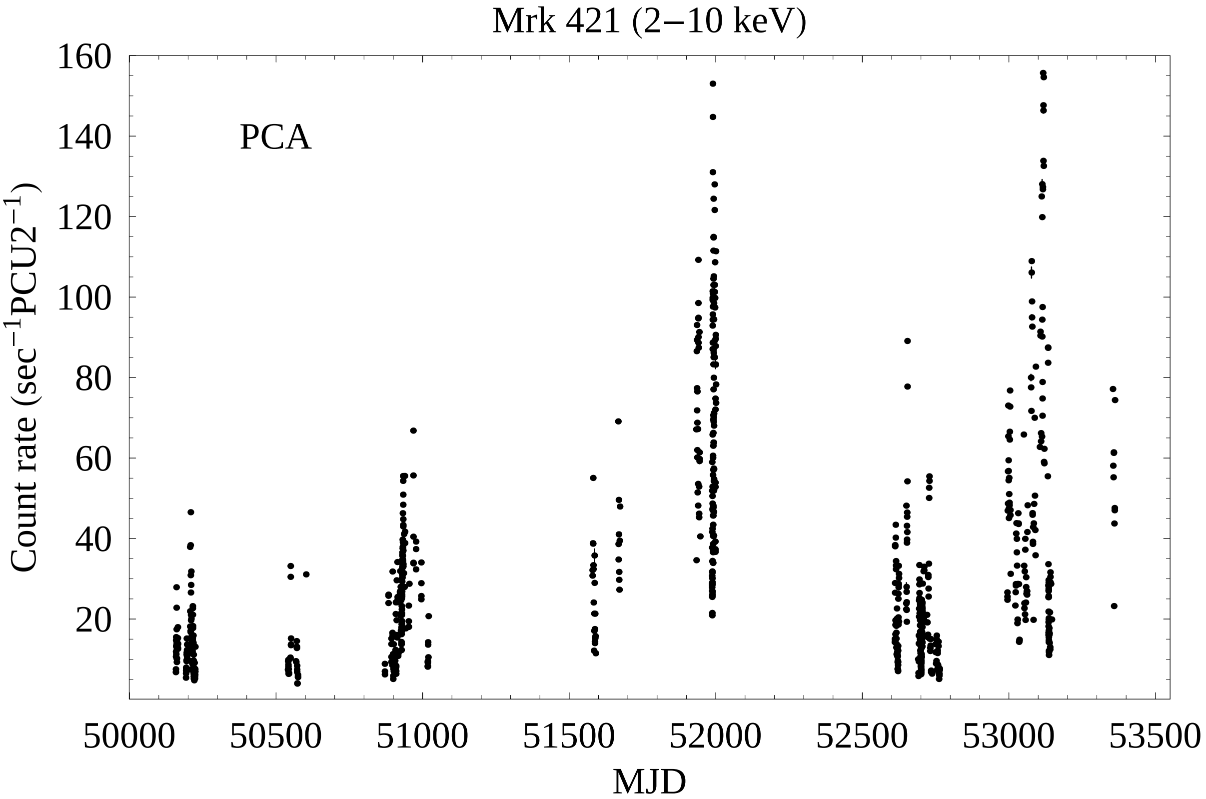


Figure 4.2: The combined observations of Mrk 421 during the period 1996–2005. The count rate is given in the energy band of 2–10 keV and the bin size is 90 min (~ 1 RXTE orbit).

Mrk 421 is one of which ASM has been used extensively either for triggering multifrequency campaigns or for its direct measurements on a daily or on a *dwell-by-dwell* basis (sect.2.3). The overall variability behavior of the source during the past ten years can be seen in fig.4.3 in which each point consists the average of 15 daily average values each one of them coming from a set of dwell observations fulfilling some criteria (sect.2.3). Five flaring events can be readily discerned embedded in an overall quiescent state around 0.3 ± 0.2 counts sec^{-1}. Up to now no statistical analysis of any kind have been conducted for the given data set. In chap.6 the first thorough higher order time series analysis is conducted yielding interesting results.

4.3 Comparison of the All-Sky Monitor and the Proportional Counter Array Observations

4.3.1 Morphological comparison

Irrespective of the nominal accuracy of the ASM, derived for bright-stable X-ray sources in the center of the FOV and not being situated very close to the Galactic bulge (see sect.2.1.4 and fig.2.7), the best assessment of its reliability concerning variable and low flux sources, such as Mrk 421, comes from direct comparison of ASM fluxes with measurements taken with more sensitive narrow-field instruments such as PCA.

In fig.4.4 the common time period of the two light curves (fig.4.2, fig.4.3) is shown in a comparative way with the dashed lines indicating the correspondence of four source events between the two detectors. This plot depicts two basic features concerning the PCA and the ASM data sets. Firstly, it shows that the PCA data set consists mainly of observations taken during high energy source states. Even though the periods that were characterized as relative quiescent (sect.4.1.2) they seem to be taken during burst activity. Only the first two PCA periods can be characterized as “truly” quiescent based on this sample. Secondly, the relatively good “by eye” correspondence between the measurements of two detectors seems prominent in order to check for further reliability of the ASM data set in smaller time scales.

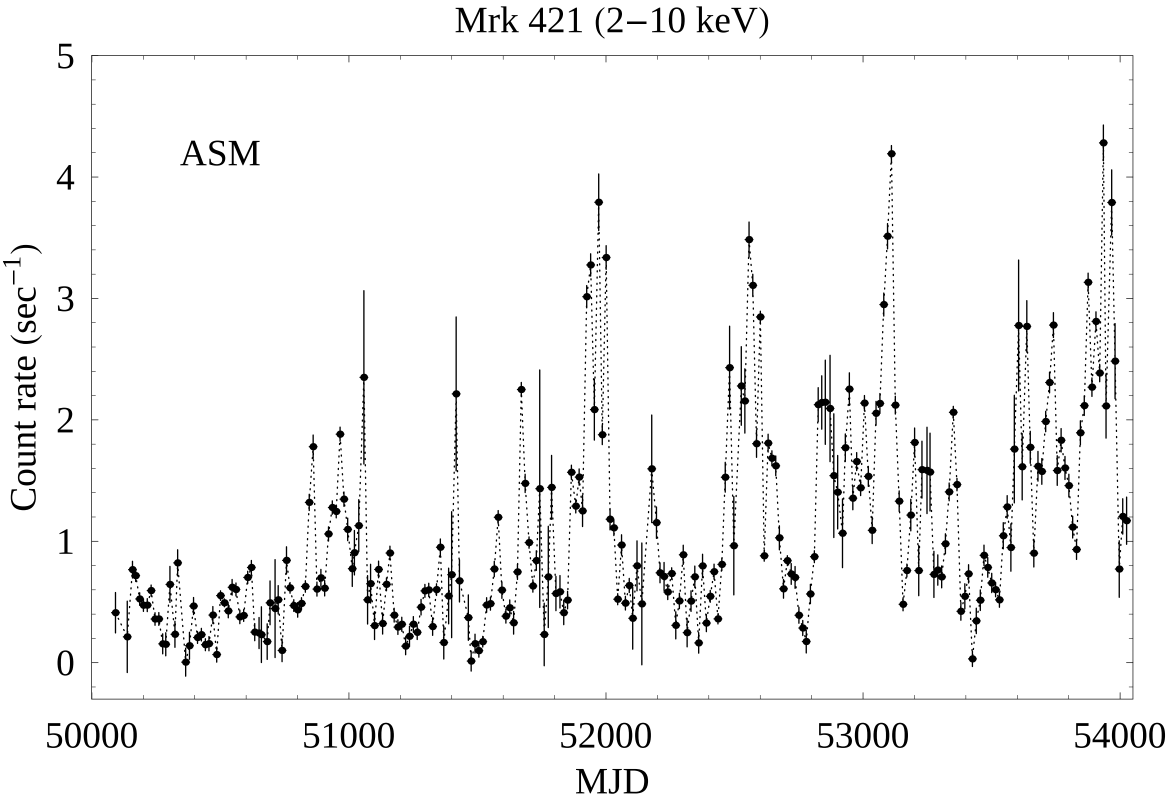


Figure 4.3: The light curve of Mrk 421 as registered from ASM during the period 1996–2006. The count rate is given in the energy band of 2–10 keV and the bin size is 15 days. The horizontal dashed line indicates the apparent "quiescent" state of the source around 0.3±0.2 counts sec^{-1}.

4.3.2 Quantitive comparison

The first think that one should check based on a statistical study is up to which degree the dwell observations can be considered reliable since the ASM suffers from large systematics (e.g. enhanced noise from bright sources in the FOV). Up to now the use of ASM data for detailed temporal analysis is usually disfavored (Kataoka et al. 2002; Uttley et al. 2002) since for a given PCA observation it is not possible to establish any apparent (i.e. "by eye") correlation between the two data sets. Due to the fact that ASM is not as sensitive as a pointing instrument the reliability of the observed variability patterns depends strongly on the properties of the observed source such as position and number of sources within the ASM FOV.

For the case of Mrk 421 from a total number of 37240 dwell observations (fig.4.3) ∼ 1800 do not fulfill all of the filtering criteria (sect.2.3). For the majority of dwells the number of sources in the ASM FOV is continuously below 16 for 3% of them the field becomes unusually crowded 8 sources compared to the usual 1–2 sources contributing to increased background events. Also the long-axis angle θ is for 2% of the dwells out of the given range $-41.5° < \theta < 46°$. Therefore individual dwell events can indicate erroneously flaring source activity which is actually attributed to the background.

The best way to check about the reliability of the ensemble of the dwell observations of Mrk 421, is to correlate them with observations taken from a pointing instrument such as PCA. Especially for the case of this source this procedure is the most robust method since its PCA archive is the biggest one among all the AGN and therefore one can find a significant big number of strictly simultaneous data pairs taken with both instruments. This work has been already performed by Emmanoulopoulos et al. (2005) in which all the 90 min dwell observations have been gathered together and compared with the simultaneous PCA data during the period 15/3/96–6/3/03 (fig.4.3). During these 2500 days a total of 33677 dwell observations were obtained by ASM and 69621 measurements of 16 sec (∼ 1.1 Msec) were obtained by PCA.

Initially both data sets are averaged in daily bins forming a set of 168 data points and plotted versus each other (fig.4.5). Despite the overall correspondence of the intensities measured by both instruments, there is considerable scatter and significant outliers. Fitting the data to a straight line, $y = ax + b$ and taking into account only the measurement errors of ASM (in the case of the PCA these are practically negligible ∼ 0.1 counts sec^{-1}PCU2^{-1}) yields $a = 0.041 \pm 0.002$ and $b = 0.087 \pm 0.047$ with a poor reduced χ^2

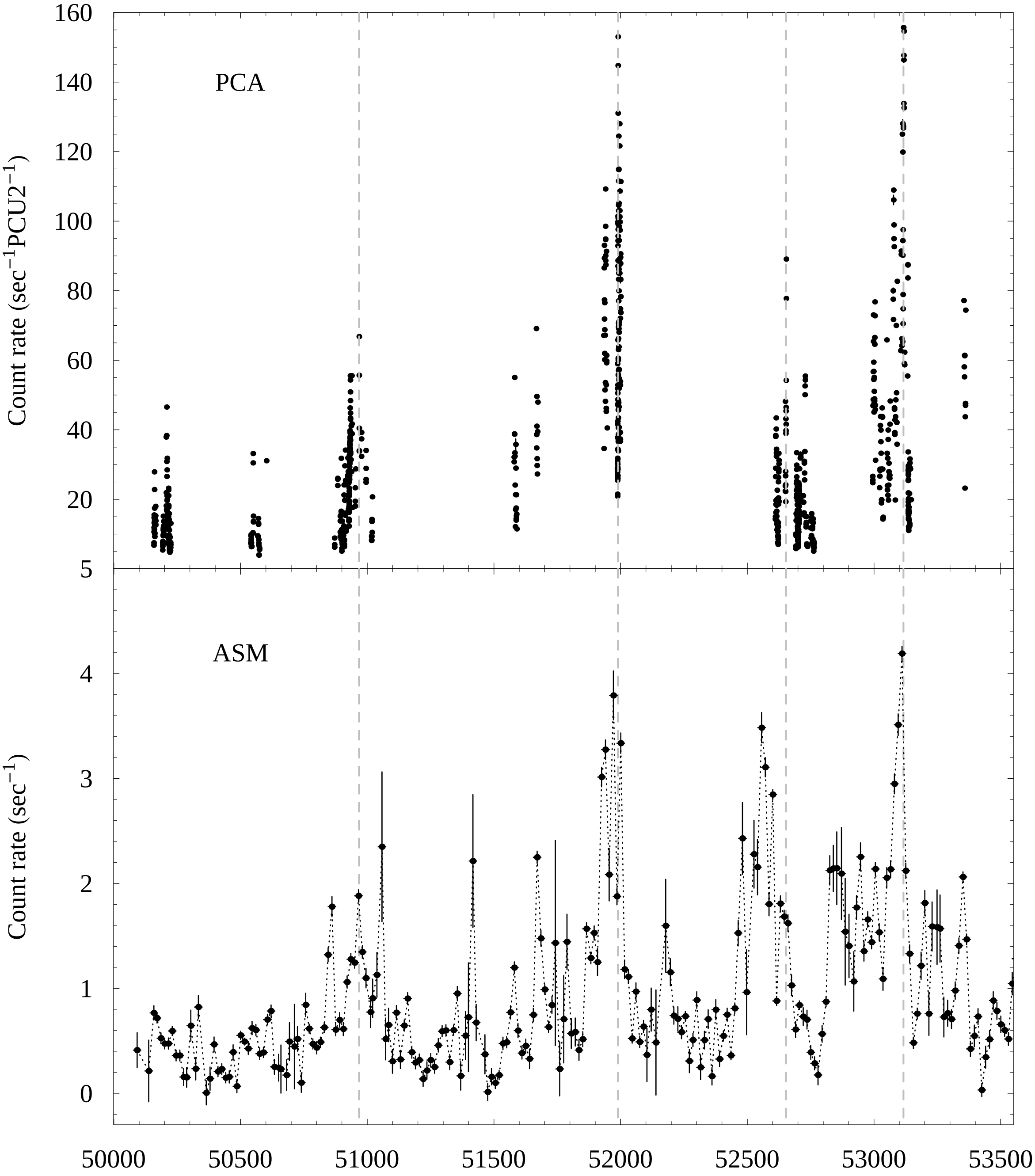


Figure 4.4: The common observations between PCA and ASM for Mrk 421 during the period 1996–2006. The dotted lines indicate the correspondence between the PCA and the ASM for four flaring events. Details for the PCA and ASM points are given in fig.4.2 and fig.4.3 respectively.

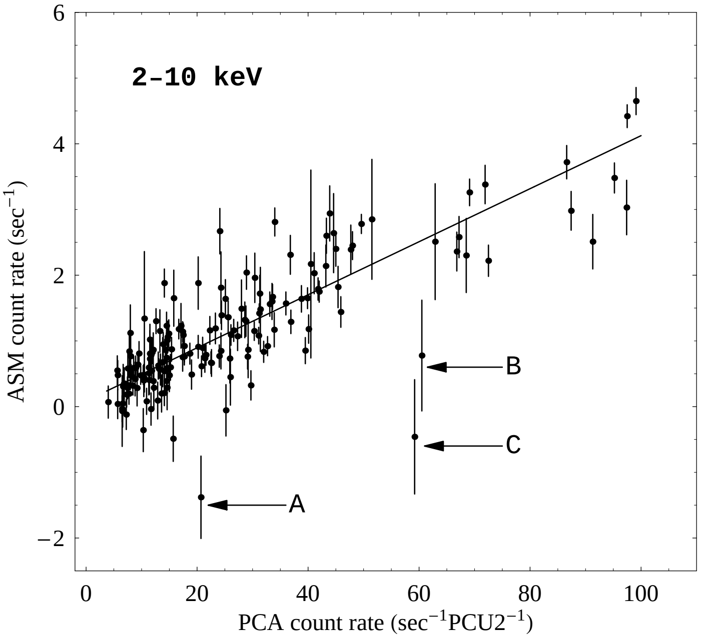


Figure 4.5: The ASM versus the PCA daily average measurements. The errors associated with the PCA measurements are ~ 0.1 counts sec^{-1}PCU2^{-1}. The arrows A, B, C indicate three badly affected values from the binning.

of 3.24 for 166 DOF and a NHP of 0. Among the most significant outliers in fig.4.5 are the points indicated by the arrows A, B, C having a deviation of 3σ, 3.7σ and 1.6σ respectively. The reason of these deviations is that within these days there is no direct overlap between the observations of the two instruments as seen in fig.4.6. During the period 15/3/96–6/3/03 the number of dwells within a day changes as shown in fig4.7 hence the probability of having exactly only one dwell observation within a day is $\sim 10\%$.

In order to overcome the difficulty of the problematic binning a more precise method is followed. All the 16 sec PCA observation lying within the 90 minutes ASM dwell observations are binned forming a sample of 173 data points. The time overlap between the measurements ranges between 88% and 100% and the time offset during the selection criteria is set to 0.0001 day (or 8.64 sec). In fig.4.8 the measurements of the two detectors are plotted the one versus the other together with two linear fit models, $y = ax$. The first model (solid line) takes into account only the ASM measurement errors $\sigma_{\rm ASM}$ and the second one considers additionally the PCA measurements errors, $\sigma_{\rm PCA}$, (Fasano & Vio 1988). The values of a coming from the two methods are consistent within the errors 0.042 ± 0.001 and 0.041 ± 0.002 respectively, therefore the first method is going to be used further on. The reduced χ^2 yields a value of 1.871 for 172 DOF having a NHP of 2.64×10^{-11}.

4.3.3 Reliability limit of the ASM dwell observations

As the source count rate augments the scatter around the fitted line becomes smaller since the detector's response is better. Thus the query is: what is the lowest ASM count rate limit that one can trust? In order to answer this question the following analysis is performed.

Successively by excluding each time the first data point with the smallest ASM count rate (from left to the right of fig.4.8), subsets are formed containing fewer points in the lower ASM count rate regime. Every time the reduced χ^2 value between the remaining values and the model is estimated (fig.4.9). The latter quantity is expected to reach a plateau, since the fit is driven every time from the points with higher count rate, depicting the fact that above the given ASM count rate the detector is linearly responding to the source count rate (fig.4.9). The determination of the plateau level is done by fitting successively a constant line to all points each time excluding the point with the highest ASM count rate (from the right to the left of fig.4.9). The fit quality starts to a given point to worsen indicating the onset of the plateau. The outcome for Mrk 421 is that a dwell ASM observation above 2.0 ± 0.1 counts sec^{-1} can be a significant detection since that is the lowest-limit-value of the detectors' detection range.

Usually in observing campaigns, especially the ones triggered from ground-based Čerenkov telescopes, due to the lack of simultaneous X-ray coverage of the observed source the use of the ASM detector is

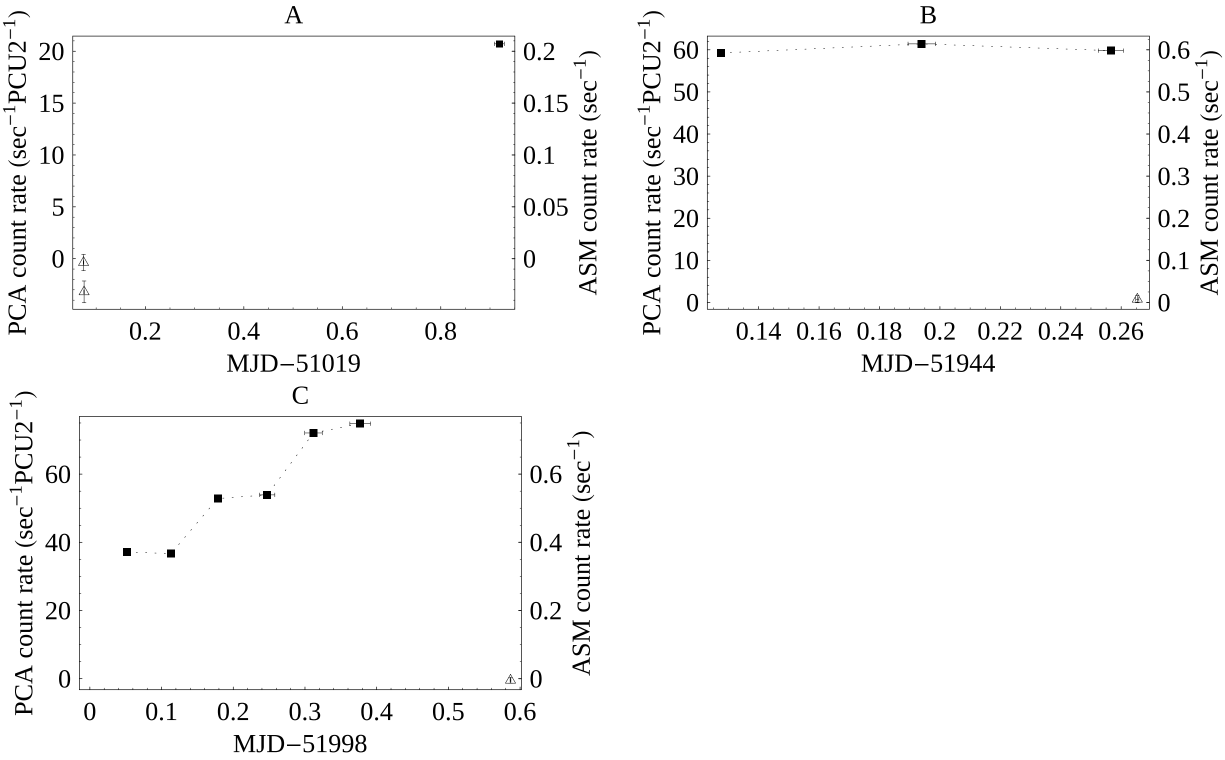


Figure 4.6: A time-zoom on the points A, B, C of fig.4.5 indicating that the two data sets do not overlap in time. The squares represent the PCA observations binned in 5440 sec (the PCA count rate is shown on the left axis) and the triangles represent the ASM dwell observations lasting ∼ 90 sec (the ASM count rate is shown on the right axis).

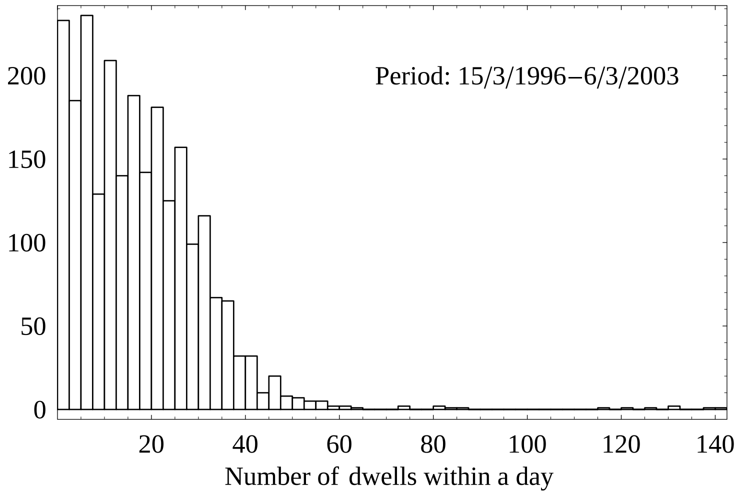


Figure 4.7: The distribution of the number of the dwell observations within a day for the period 15/3/96–6/3/03.

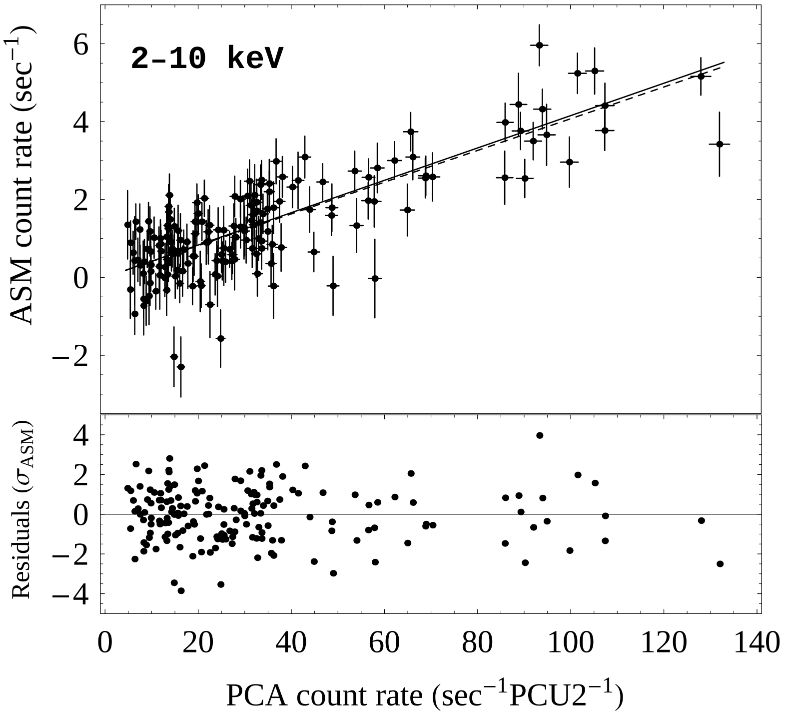


Figure 4.8: [Top panel] The ASM versus the PCA simultaneous measurements with a time offset of 0.0001 day. Both lines represent a linear fit model of the form $y = ax$. The solid line is the result of the linear fit taking into account only the ASM measurement errors σ_{ASM} (yielding an a of 0.042±0.001) and the dashed line shows the result of the fit considering additionally the measurement errors coming from PCA σ_{PCA} (yielding an a of 0.041±0.002). For the same exposure time of 90 min, the mean value of the ratio $\sigma_{\mathrm{ASM}}/\sigma_{\mathrm{PCA}}$ is of the order of 2 (1.85 ± 0.05).
[Bottom panel] The residuals of the linear fit, taking into account only the ASM measurement errors σ_{ASM}, in units of σ_{ASM}.

favored. Before using the ASM data one should perform a similar analysis in order to check the reliability of the instrument for the given source. The parameters described in sect.2.3 change from source to source therefore the result should be a function of them.

A direct application of the aforementioned method is performed for the case of Mrk 421 in (Aharonian et al. 2005a). During April 2004, a coordinated multi-wavelength campaign monitored the activity of Mrk 421 in the radio, the optical, the X-ray, and the γ-ray regimes (Cui et al. 2005). The source was seen to be active in X-rays where observations with the PCA were performed. These observations were not simultaneous with the observations with the H.E.S.S. array, but by combining the PCA with the ASM data a good temporal coverage overlapping with the H.E.S.S. observations can be achieved. An average of 4 counts sec^{-1} was detected by the ASM during the first weeks of April. This is sufficiently high to probe the activity of Mrk 421 during the individual ASM dwell observations.

The Whipple 10 m Čerenkov telescope was observing Mrk 421 simultaneously with the RXTE satellite (Cui et al. 2005) starting generally within a few hours after the H.E.S.S. pointings. In fig.4.10, the different observations are combined such that the observed count rate is normalized to the average flux count rate during the time between MJD 53107 and MJD 53116. The preliminary Whipple light curve is derived from the count rate which is not corrected for different zenith angles of observations. The ASM light curve and the 1σ uncertainty band is obtained by calculating the sliding average over five dwells. The horizontal dashed line indicates the lower level up to which we can rely on PCU measurements.

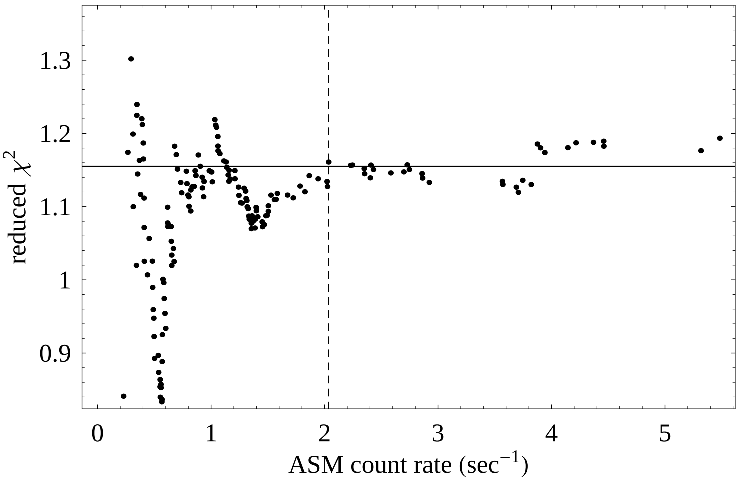


Figure 4.9: The values of reduced χ^2 after removing successively the first ASM point from the strictly simultaneous PCA-ASM data set. The x-axis denotes the first value of the ASM count rate for each data set. The horizontal solid line represents the plateau and the vertical dashed line represents the abscissa of this onset in units of ASM count rate.

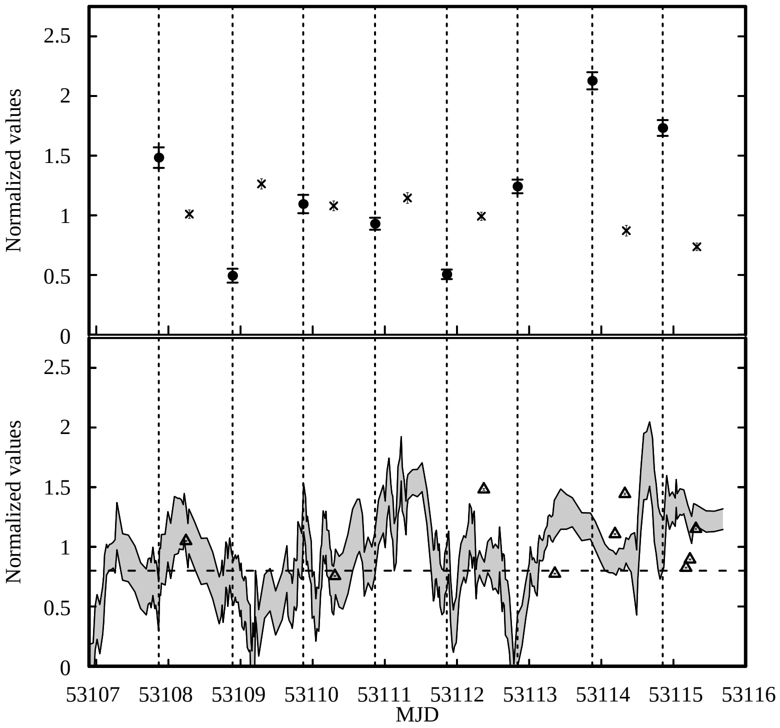


Figure 4.10: The relative changes of the flux with respect to the average observed during the multi-wavelength campaign in April 2004 (included in Aharonian et al. 2005a).
[Top panel] H.E.S.S. flux above 2 TeV (filled circles), Whipple (crosses, preliminary count rate without correction for varying zenith angles Cui et al. (2005). According to previous observations, the energy threshold is estimated to be approximately 400 GeV (Krennrich et al. 2002)).
[Bottom panel] PCU (open triangles, count rate), ASM (1σ error band, see text for details). The dashed horizontal line indicates the lowest reliable count rate concerning the ASM (the Whipple and RXTE PCU data are taken from Cui et al. 2005).

Chapter 5

THE STRUCTURE FUNCTION AND THE TIME SCALES

The serious misapplication of the structure function (SF) (sect.3.3.5) in the field of blazar astronomy during the last 15 years has created the fallacy that "characteristic" time scales do exist in the data sets of these objects. The basic idea is that the abscissa of the "break" point (see point B in fig.3.7) indicates a time scale which is somehow directly connected with a source property. This thought has created confusion among astrophysicists who associate this time scale with several physical quantities without giving most of the times any clear justification about why this should be the case. There are a lot of ambiguous terms connected with the abscissa of the "break" point[1] such as *characteristic time scale*, *hidden time scale*, *quasi periodic time pattern* etc. that have nothing to do with either the physics of the source or with any true time property embedded in the data set itself. These time scales are supposed to be directly connected with the linear size of the emission region in the case of the BL Lac objects (e.g. Wagner et al. 1996; Takahashi et al. 2000) or with the physical source time properties such as Keplerian, thermal and viscous time scales for the case of the Seyfert galaxies (e.g. Czerny et al. 2003). This phenomenon appears only in the field of astrophysics probably due to the fact that standard texts in time series analysis are rarely consulted. Sometimes also vague methodologies are employed in order to consolidate a pseudo-accuracy of the results in order to appear more "convincing". The main drawback from this misuse of the SF is that fictitious time scales, coming usually from completely random data sets, are accepted creating the idea that well defined time patterns do exist in these sources. In this chapter the long-look X-ray observations of Mrk 501, obtained by ASCA on March 2000, (Tanihata et al. 2001) are reviewed since they comprise the first X-ray data set from which a 1 day "characteristic" time scale is claimed through the use of the SF. Based on simulations it is shown that this time scale is simply an artefact which depends on the duration of the observations and which is unrelated to any physical source property.

5.1 The Fake "Characteristic" Time Scale of One Day

Besides Mrk 421, two equally well studied BL Lac sources (sect.1.2.1) are Mrk 501 (z=0.033) and PKS 2155-304 (z=0.117) both of them detected in the VHE γ-rays by Quinn et al. (1996) and Chadwick et al. (1999) respectively. At the end of 1990s the X-ray satellite ASCA (Tanaka et al. 1994) performed three "long look", uninterrupted and continuous observational campaigns for these objects covering the X-ray band ($\sim$ 2–10 keV). Initially Mrk 421 was observed for 7 days during April 1998 (Takahashi et al. 2000; Tanihata et al. 2001) and then Mrk 501 and PKS 2155-304 during March and May 2000 respectively for 10 days each (Tanihata et al. 2001).

In order to elucidate the matter of the breaks in the SF the ASCA data set of Mrk 501 is chosen, since it is the biggest among the three consisting of 191 observations and having a time resolution of 4096 sec

---

[1] Henceforth the abscissa of the "break" point is going to be referred simply as a break.

(fig.5.1). The given campaign is claimed to have a "characteristic" time scale of almost a day as a result of the SF analysis (fig.5.2). Exactly these observations consolidated the idea that something special is happening in these sources in the X-ray regime during this time scale, which is directly connected with the light crossing time in the emission region[2]. Based on the algorithm presented in sect.3.3.6, 2000 thousand purely random artificial light curves, are produced having the same length[3], the same PSD, the same mean value, 6 counts $\sec^{-1}$ and the same standard deviation, $S = 0.19$, as the light curve of Mrk 501. The are two questions that should be answered: How do their SFs look like? In case that their SFs exhibit breaks when do these breaks occur?

In order to estimate the PSD of the original data set, its periodogram is computed (sect.3.3.1) together with the corresponding errors due to measurement uncertainties (see below) (fig.5.3, red points). After binning the logarithm of the periodogram (sect.3.3.2) its slope can be then robustly specified (see e.g. Ferrero et al. (2006)) through linear regression based on the least squares method. For the computation of the binned logarithmic periodogram $n = 16$ sets are formed each one having $M = 12$ observations and the estimates of the logarithm of the PSD, $\log(\mathcal{P}(f_{\text{avg},i}))$ (eq.3.15) at the geometric mean frequencies $f_{\text{avg},i}$ (eq.3.14) are then derived (fig.5.3). The double error bar lines indicate firstly the variance of the binned logarithmic periodogram at the various frequencies being 0.026 (eq.3.16) and secondly the errors due to the measurements uncertainties. The latter have been computed by using the method described in sect.3.1.3 by producing 1000 artificial light curves from the original light curve of Mrk 501. Firstly all the periodograms were estimated and then the logarithmic estimates of them were calculated by using exactly the same procedure as the one applied to the original light curve (i.e. same n, M). Based on the distribution of the periodogram values at the various frequencies f_j, the errors of the periodogram estimates due to measurement uncertainties can be derived. The uncertainty of the $\log(\mathcal{P}(f_{\text{avg},i}))$ at each frequency $f_{\text{avg},i}$ is taken as the standard deviation of the distribution of the 1000 logarithmic periodogram estimates at $f_{\text{avg},i}$. Finally the slope of the PSD of the original light curve is estimated to be $a = 1.68 \pm 0.07$ and this together with its mean value and standard deviation comprise the basic ingredients for the production of the 2000 artificial light curves.

The basic idea behind this experiment is to check whether or not the claimed "characteristic" time scale of one day can occur in light curves of the same length which are completely random in nature meaning that both their phases and amplitudes are random. The surrogated light curves have exactly the same variability power for all the time scales (i.e. all frequencies) as the light curve of Mrk 501 (fig.5.1) since they have been produced based on the genuine PSD. After producing the artificial light curves their corresponding SFs are estimated. The SFs are derived following exactly the same methodology as in Tanihata et al. (2001) (eq.1, eq.2 and eq.3) in order not to introduce systematic deviations due to different formulas. For a light curve consisting of data points $f(i)$ which have an uncertainty of a standard deviation $\sigma_f(i)$ the SF is given

$$SF(\tau) = \frac{1}{N(\tau)} \sum w(i)w(i+\tau)[f(i+\tau) - f(i)]^2 \tag{5.1}$$

where $N(\tau) = \sum w(i)w(i+\tau)$ and $w(i) \propto f(i)/\sigma_f(i)$. The main difference between eq.5.1 and eq.3.23 is the weighting factor $w(i)$ which is considered to be an efficient modification of the usual formula incorporating the statistical errors. Note that the introduction of $w(i)$ is lacking any justification since it appears more like an intuitional term rather than a rigorous mathematical quantity. Since at this point the interesting thing is the result of the method and not the method itself the same methodology is applied. Note here that the uncertainties of the given data set are so small that scarcely affect the SF results. The measurement errors, $\sigma_f(i)$, are of the order of 2% in comparison to the measurements, $f(i)$, therefore the resulted SF estimates after ignoring the errors differ on average only by 1.7%.

Since the artificial light curves do not have any errors the genuine statistical errors from the original light curve are annexed to the surrogated values. For the particular data set this can be done since the errors and the count rate of the genuine Mrk 501 data set are not correlated. The linear correlation coefficient is 0.018, meaning that the probability of a value like this or bigger yielding from a random sample of 191 uncorrelated measurements is 0.80 (Bevington & Robinson 1992). More than 70% of the measurements uncertainties are around 0.020 ± 0.05 therefore it can be considered that they are drawn from a Gaussian distribution with the corresponding mean value and standard deviation. In fig.5.4 two of those artificial light curves are shown with their corresponding SFs.

---

[2] The time t that the light needs to travel a distance R going from the one side of the emission region to the other is $t = R/c$ where c is the speed of light (causality argument). If the source has a Doppler factor δ and a redshift z then in the observer's frame $t = (R/c)(1+z)/\delta$ (e.g. Wagner et al. 1996).

[3] Initially the light curves are 10 times longer and afterwards they are chopped to the desired length of 191 points (see sect.3.3.6).

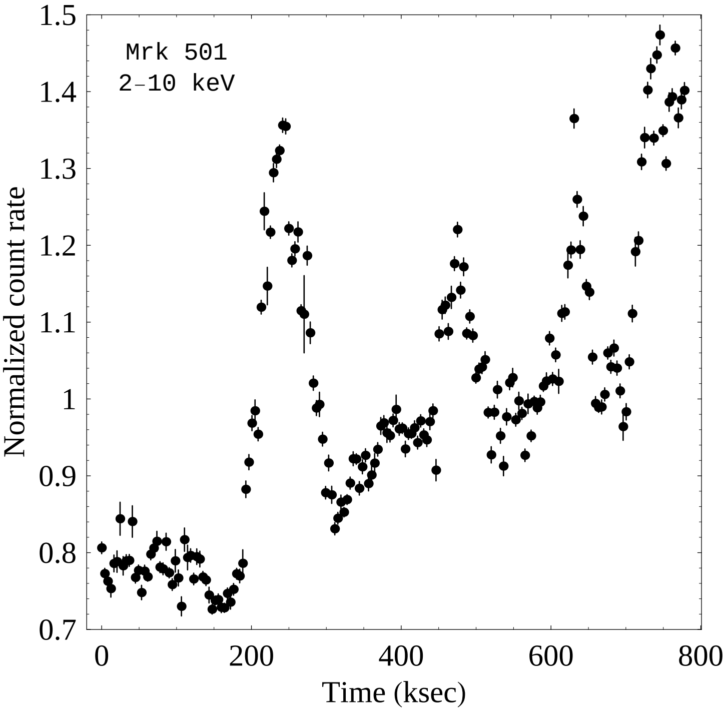


Figure 5.1: The X-ray light curve of Mrk 501 as registered by ASCA during March 2000 (Tanihata et al. 2001) with binning time of 4096 sec. The light curve is normalized to its mean being 6 counts sec^{-1} (the measurements are taken from fig.1 of Tanihata et al. 2001).

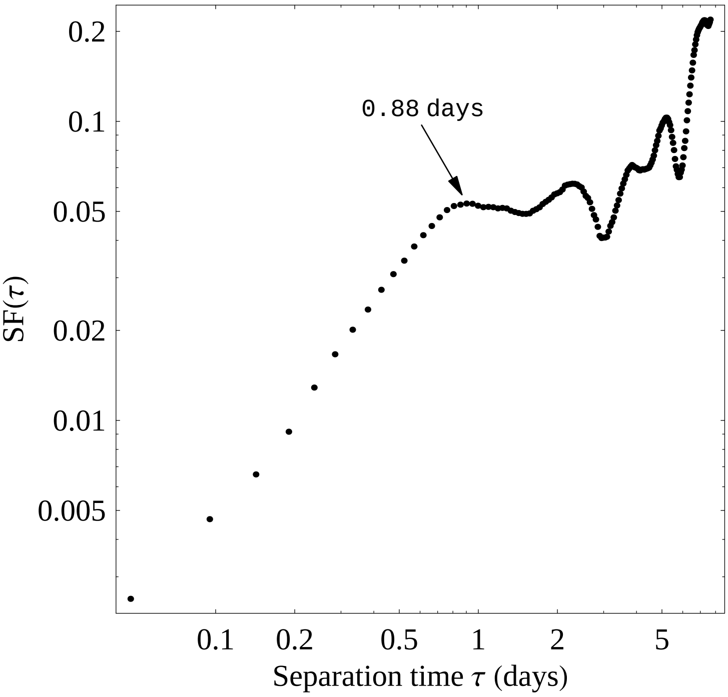


Figure 5.2: The SF of Mrk 501 as computed from the ASCA light curve shown in fig.5.1. The points have a separation time of 4096 sec or 0.0474 days. The "characteristic" time scale occurs at 0.88 days (i.e. ∼ 1 day) (in perfect agreement with the top panel of fig.4 in Tanihata et al. 2001).

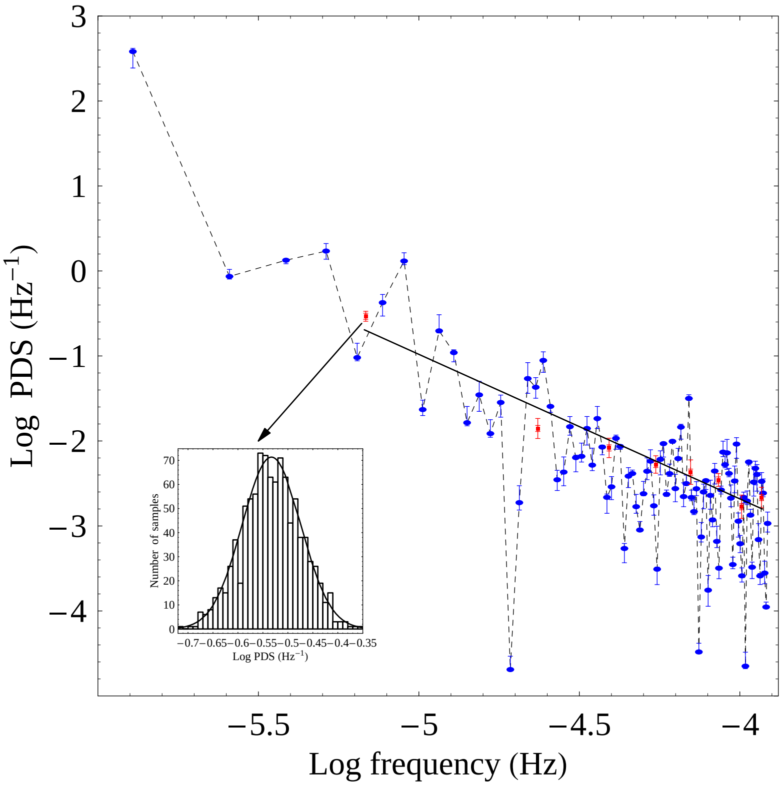


Figure 5.3: The logarithm of the PSD of Mrk 501 as a function of the logarithm of the frequency. The dashed line shows the estimate of the periodogram (blue points) as it is computed from eq.3.11. The errors of the periodogram have been derived based on the methodology presented at sect.3.1.3 by producing 1000 artificial light curves from the original light curve of Mrk 501. The red points represent the binned estimates of the periodogram (eq.3.15) for $M = 12$ and $n = 16$. The short error bars indicate the variance of the binned logarithmic periodogram at a given frequency being 0.026 (eq.3.16) and the longer ones the uncertainties based on measurements errors estimated from the 1000 aforementioned simulated light curves. The arrow shows an example about how the distribution of the binned logarithmic periodogram looks like for the first frequency bin at 6.83 μHz (or $\sim$ 146.4 ksec) as it comes up directly from the simulated light curves. The solid line represents the linear fit to the logarithmic periodogram estimates yielding a power-law index of -1.68 ± 0.07 yielding a χ^2 of 3.52 for 7 DOF with NHP=0.833.

It seems that in the literature there is no explicit methodology concerning the specification of the SF break. Most authors employ arbitrary “eye selection criteria” in order to specify it (sometimes referring to it as observed break) without giving any explicit description of the methodology (e.g. Hughes et al. 1992; Lainela & Valtaoja 1993; Takahashi et al. 2000; Tanihata et al. 2001; Czerny et al. 2003). In order to determine the SF break the following methodology is applied to the ensemble of artificial light curves. For large values of τ (i.e. comparable to the time extend of the data set) the SF is expected to be equal to $2S^2$ where S^2 is the sample variance (sect.3.3.5). Based on that for every artificial data set the ACF, $R_{1,1}(\tau)$, is computed and the value $\tau_{\rm uncor}$ is specified. Then for every SF an interpolated version of it is produced and the time $t_{\rm break}$ when its value equals for the first time to $2S^2$ is specified. If $0.8\tau_{\rm uncor} \leq \tau_{\rm break} \leq 1.2\tau_{\rm uncor}$ and the distribution of the following SF points (i.e. SF($\tau > 1.2\tau_{\rm uncor}$)) has a maximum (most probable value) around the value of $2S^2$ then that means that a clear plateau is formed. If the inequality is valid but the distribution of points do not have a maximum around $2S^2$ then that means that there is a break followed by an increase or decrease in the values of the SF.

It should be mentioned here that this procedure does not aim to introduce a methodology for specifying breaks in the SF but rather to mimic the “eye selection criteria”. The issue that is tested here is how representative and characteristic are the breaks of a completely random sample sharing the same statistical properties with the under study light curve.

From the 2000 light curves 1883 exhibit a break, based on the aforementioned criteria, whose distribu-

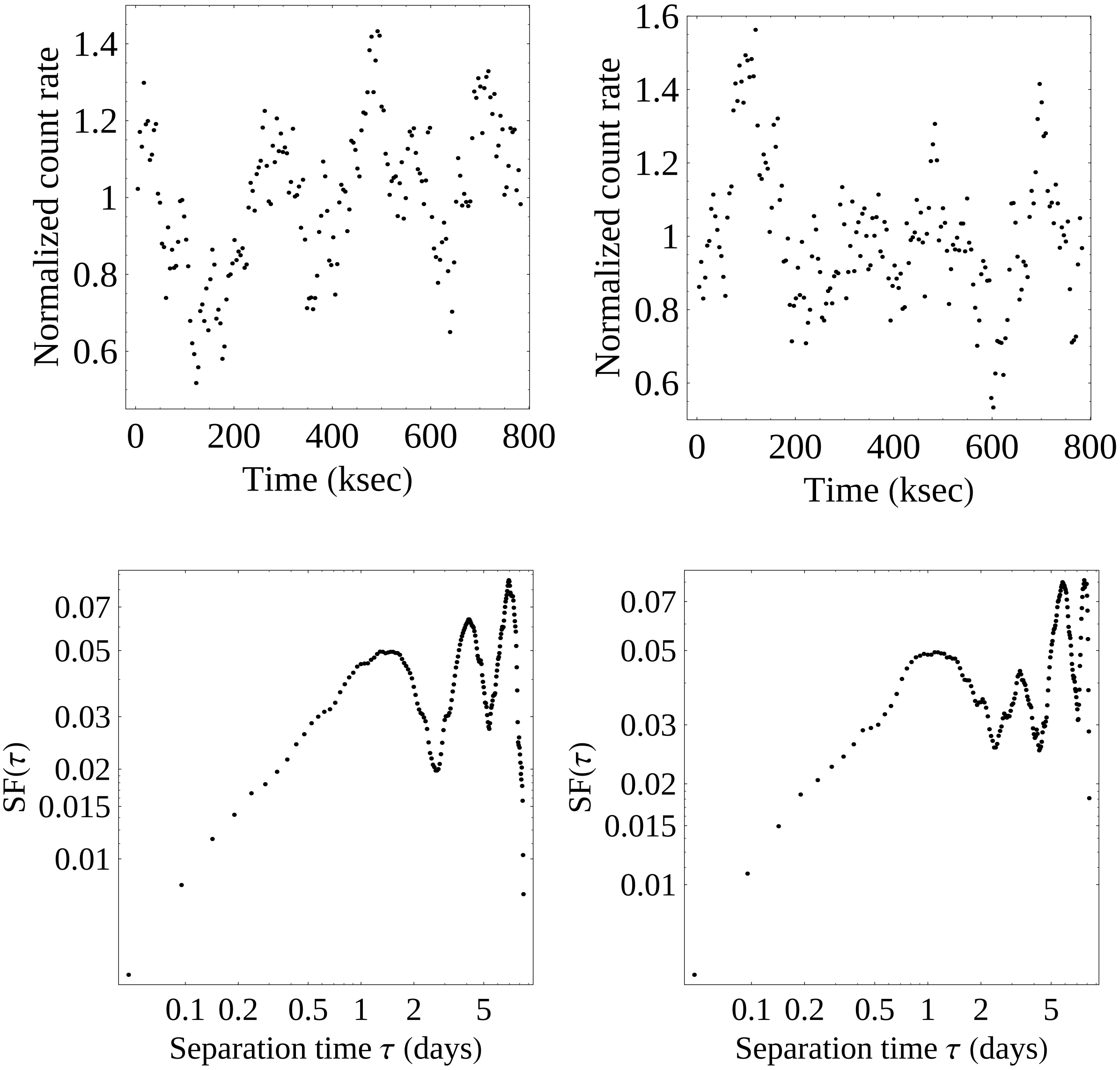


Figure 5.4: [Top panels] Two artificial light curves sharing the same second order statistical properties with the light curve of Mrk 501 (fig.5.1). The measurements errors are taken from the genuine light curve. [Bottom panels] The SFs of the two artificial light curves.

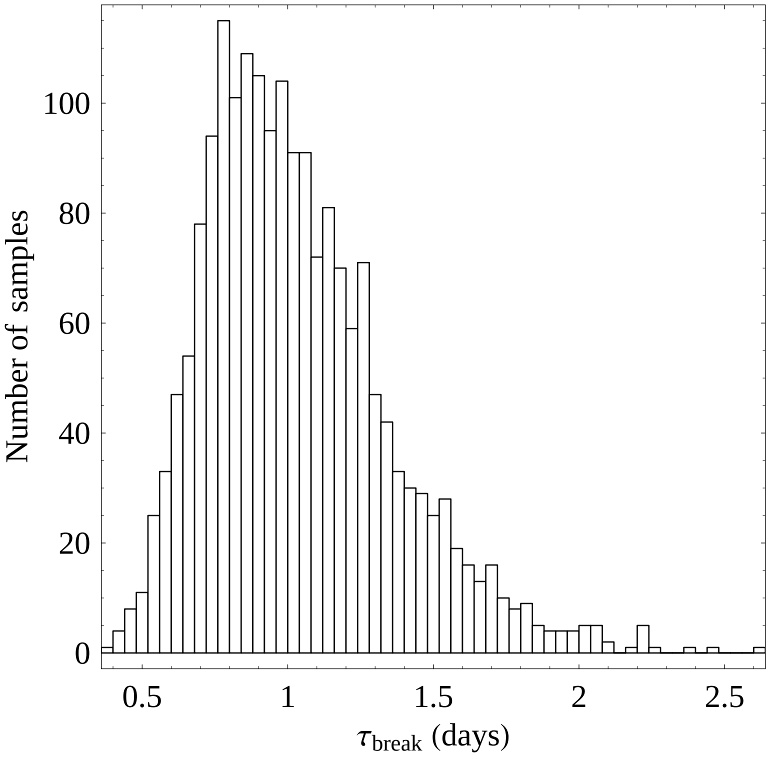


Figure 5.5: The distribution of the SF breaks coming from the 1883 artificial light curves for the case of the ASCA data of Mrk 501. The mean value and the standard deviation of the distribution are 0.923±0.007 and 0.320 days respectively. The histogram bins have a length of 0.04 days.

tion is shown in fig.5.5. The distribution of the SF breaks has a mean value of 0.923±0.007 days and a standard deviation of 0.320 days. It is clear from this plot that for a purely random process having the same PSD the same second order moments and the same length as the ASCA light curve of Mrk 501, a SF break of 0.88 is quiet possible to occur. Performing a simple hypothesis test between the mean value of the distribution and the observed SF break 0.88 days, yields the result that up to significance level of 2.4% the latter value can be considered that it is drawn from the former distribution.

This example proves robustly that there is nothing characteristic concerning the time scales in the flares of the light curve of Mrk 501 as registered by ASCA in March 2000. This can simply be the result of a random process having the same second order statistical properties and the same length as the observed light curve. Concerning the length of the light curve it is expected that longer data sets from Mrk 501, having the same statistical properties, will exhibit breaks in their SFs at longer time scales depicting exactly the fact that this break has nothing to do with any physical properties of the flares whatsoever.

The need of a long-term light curve for Mrk 501 necessitates the use of ASM (sect.2.1.4), similarly as in the case of Mrk 421 (sect.4.2). The ASM light curve of Mrk 501 is retrieved for the period January 1996 till January 2001 (1770 days) binned into 15 days time bins (in total 118) and normalized to its mean being 0.62 counts sec^{-1} (fig.5.6). The sample was chosen as such in order to include the ASCA observations taken during March 2000 and at the same time to have as much as possible the same statistical properties. The ASM light curve has a standard deviation of $S = 0.21$ and its PSD, as it is estimated from the logarithmic periodogram, has a power-law index of -1.62 ± 0.05. Once again 2000 artificial light curves are produced having a length of 118 points (bins of 15 days) and their SF is estimated. Following the same selection criteria, from the 2000 samples 1956 exhibit breaks whose distribution is shown in the left panel of fig.5.7. The mean value and the standard deviation of their distribution is 399.81±0.25 days and 11.27 days respectively. The SF of the original ASM light curve is also shown in the right panel of fig.5.7 having a maximum at 402 days.

This is exactly what is expected from the simulations of random data sets having the same second order statistical properties as the ones of the ASM light curve of Mrk 501 during this period. A hypothesis test, similar to the one performed for the ASCA data set, yields that up to a significance level of 1.7% the observed break can be considered as a result of the aforementioned distribution.

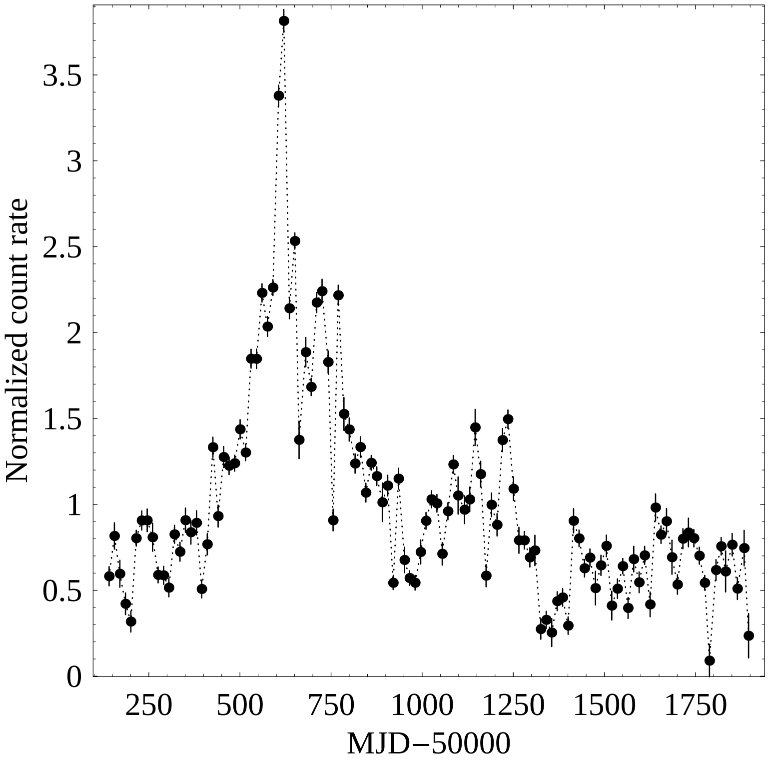


Figure 5.6: The X-ray light curve of Mrk 501 as registered by ASM during the period January 1996 till January 2001 in bins of 15 days. The light curve is normalized to its mean being 0.62 counts $\sec^{-1}$.

5.1.1 The SF and the shot model

In Tanihata et al. (2001) an effort is made to interpret the SF features (i.e. the slope and the break) tuning "structural" flare parameters such as the rise and decay times as well as the separation time of them. Simulated light curves are then produced and the resulted SFs are examined. The light curves are regarded as a superposition of various flares or *shots* occurring randomly, following a Poisson distribution with average number of shots during a time unit, r. Each shot is considered to have a simple triangular shape, with rise time scale $\tau_{\rm r}$, decay time scale $\tau_{\rm d}$, and occurrence time of the shot $t_{\rm p}$. The intensity of each shot is also set randomly, but the number density follows a power-law distribution $n \propto \alpha^{-1.5}$. Having this,

$$f(t) = \begin{cases} \alpha(t - (t_{\rm p} - \tau_{\rm r})), & \text{for } t \leq t_{\rm p} \\ -\alpha(t - (t_{\rm p} + \tau_{\rm d})), & \text{for } t \geq t_{\rm p} \end{cases} \tag{5.2}$$

the following cases are considered

- Symmetric shots: $\tau_{\rm r} = \tau_{\rm d} = \tau$
    1. Varying times $\tau = 10$ and $\tau = 100$, and fixed average number of shots per time unit $r = 1$.
    2. Varying number of shots per unit time $r = 0.1$, $r = 1$ and $r = 10$, with fixed $\tau = 50$.
    3. Randomly varying times τ between $\tau_{\rm min} = 10$ and $\tau_{\rm max} = 100$ with average number of shots per time unit $r = 1$.
- Nonsymmetric shots: $\tau_{\rm d} = 100$ and $\tau_{\rm r} = 1$, $\tau_{\rm r} = 10$, $\tau_{\rm r} = 100$ with fixed average number of shots per time unit $r = 1$.

The main results from the simulations are that the time scales of individual shots, τ, determine the location of the break in the SF (fig.5.8, left panel), result which is independent on the shot occurring rates r. When τ takes a range of time scales a deviation from a slope of 2 happens at $t_{\rm min}$ and the break occurs at $\tau_{\rm max}$. Finally when the rise and the decay timescales of the shots are different then the break occurs at the timescale which is the shortest of the two.

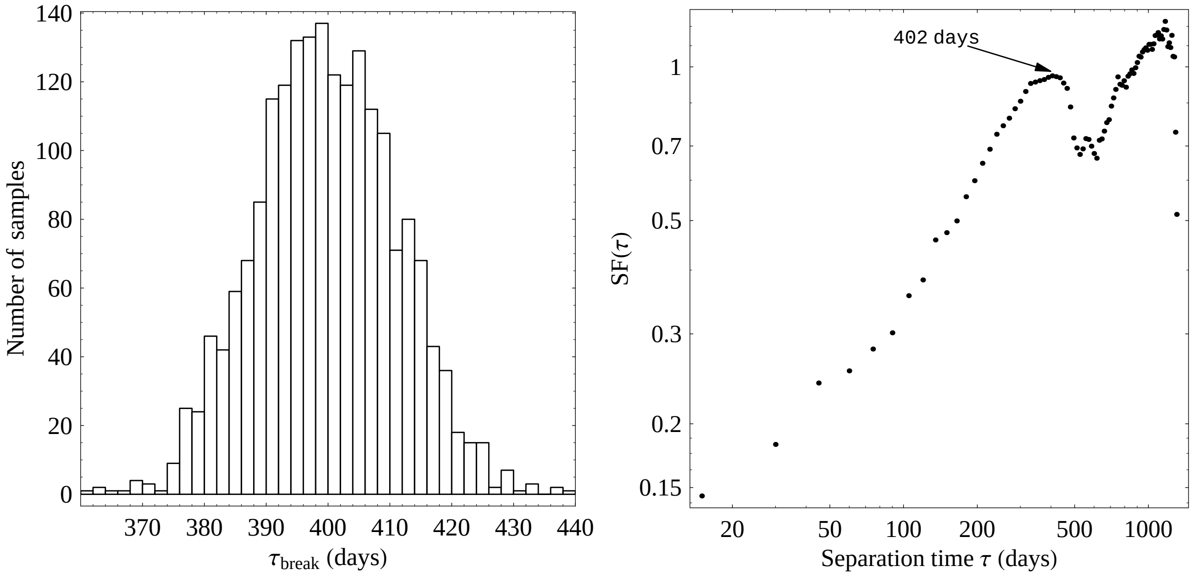


Figure 5.7: [Left panel] The distribution of the SF breaks coming from the 1956 artificial light curves for the case of the ASM data of Mrk 501. The mean value and the standard deviation of the distribution is 399.81±0.25 days and 11.27 days respectively. The histogram bins have a length of 15 days.
[Right panel] The SF of Mrk 501 as computed from the ASM light curve shown in fig.5.6. The points have a separation time of 15 days. The break denoting a "characteristic" time scale occurs at 402 days.

For the case of Mrk 501 the basic claim is that since its SF has a clear break $\sim$ 10 ksec (fig.5.2) its light curve should consist of shots with time scales of $\tau = 10$ ksec. This scenario (i.e. fixed τ) can be readily rejected based on the PSD of the data set. Since the shots have the same duration (i.e. the same τ) the Fourier components are going to depict this timescale, represented by the corresponding frequency f, and each sinusoid has only to adjust its amplitude since the intensity of each shot is not the same. For $\tau = 10$ and $\tau = 100$ a feature with a frequency $f = (10/2000)^{-1}$ and $f = (100/2000)^{-1}$ respectively is expected to be appeared in the PSD of such data sets (fig.5.8, right panel). The genuine PSD of Mrk 501 (fig.5.3) does not exhibit any indication of any periodic pattern of this nature therefore symmetric flares with fixed time scales is not a physical possible scenario.

In addition to the last point there is more fundamental problem concerning the *shot model* scenario. The model can not reproduce long lasting flaring events which produce breaks in the SF of the order of years. From fig.5.6 it can be seen that variations of $\sim$ 400 days do exist in the light curve of Mrk 501 resulting a break in the SF at $\sim$ 400 days (fig.5.7, right panel). Shot models consisting of flares with time scales of the order of $\sim$ 10 ksec can not reproduce so extended long-term variations appearing as breaks in the SF.

5.1.2 The SF and its linear properties

One of the issues that it is of primary importance is the fact that SF makes use only of the first two statistical moments of the data set (i.e. mean value and variance) (sect.1.1) ignoring all the higher statistical moments that possibly exist. The latter can be very insightful concerning the nature of the process responsible for the observed variability. Consider the following linear and nonlinear dynamical systems respectively

$$\frac{dx(t)}{dt} = 0.9x(t) + n(t) \tag{5.3}$$

and

$$\frac{dy(t)}{dt} = 8 - 0.54y(t)^{1/3} - (1.5 - n(t))\sqrt{y(t)} \tag{5.4}$$

where $n(t)$ is a standard zero mean white noise component coming from a Gaussian distribution having a standard deviation of 7 for the linear system and 1.75 for the nonlinear system. For the nonlinear case the

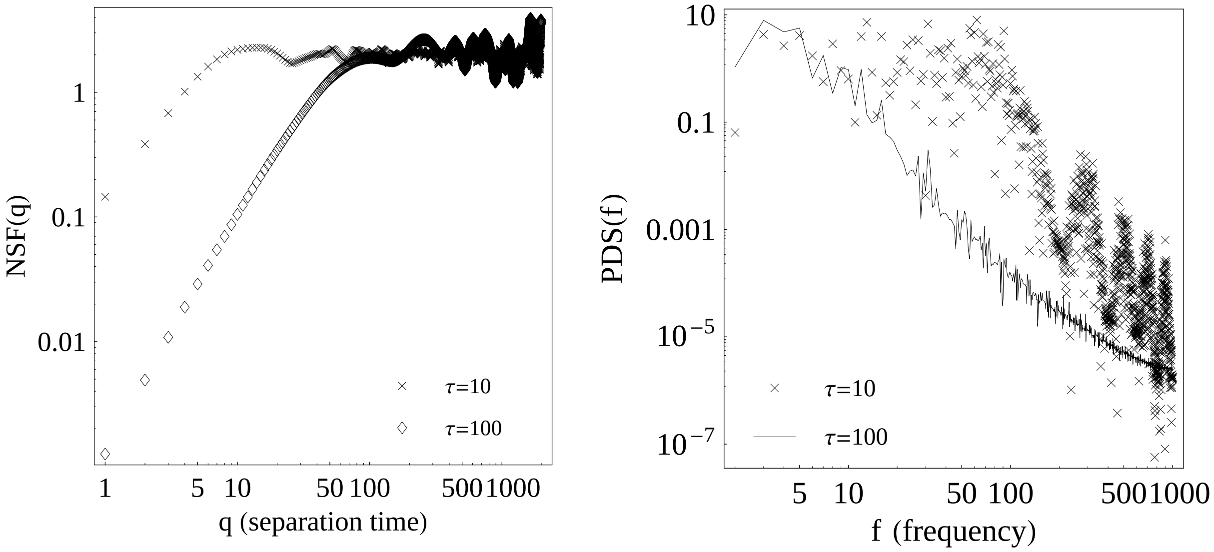


Figure 5.8: [Left panel] Reproduction of the results of Tanihata et al. (2001) for the case of symmetric flares with $\tau = 10$ and $\tau = 100$. Instead of the SFs the NSFs (eq.3.31) are plotted in order to avoid the proposed arbitrary scaling. The breaks occur respectively at $q = 10$ and $q = 100$.
[Right panel] The PSD in the case of $\tau = 10$ shows a clear periodic pattern with a frequency $f = 200$. For the case of $\tau = 100$ a solid line is plotted among the various estimates in order to guide the eye. A periodic pattern with frequency 20 can be discerned (5 peaks every 100 frequency units).

values of $y(t)$ are rescaled, by adding the constant value of 80, to have the same mean value and variance with those of the linear light curve being 100 and 300 respectively. Moreover measurement errors are added to both data sets by simply assuming Poissonian noise (i.e. for the linear $\propto x(t)^{-1/2}$ and for the nonlinear $\propto y(t)^{-1/2}$). The first 1000 realizations of the two processes, after integrating their evolution equations (eq.5.3, eq.5.4), are shown in fig.5.9 together with their histograms.

To allow a meaningful comparison of the SFs coming from the two systems the NSFs (eq.3.31) are estimated and plotted together (fig.5.10, left panel). The errors of the NSF points were computed based on the methodology presented in sect.3.1.3 and reflect how the measurement errors affect the various NSF estimates. Phenomenologically both the NSFs increase following the same course and their breaks occur at the same time, $t = 46.7$ time units. If these two data sets were light curves that had been observed from two different BL Lac objects, based on the SF method one would conclude that they exhibit exactly the same time properties and a "characteristic" time scale of ~ 47 time units. These statements are in a way misleading since they tend to equate the parameters which are tuning the variability behavior of the two systems. Moreover the breaks of the NSFs do not give any real information about the systems since their evolution equations do not favor any time scale of the order of ~ 47 time units. In order now to test up to which significance level the two NSFs can originate from the same parent distribution, up to the break point (null hypothesis H_0), the Kolmogorov-Smirnov test is applied (Press et al. 1992). This yields for the first 100 NSFs estimates a maximum value of absolute difference between the two cumulative distribution functions $D = 0.163$ (fig.5.10, right panel). Therefore H_0 is accepted at a significance level of 1%[4] since the critical value is $D_{c,0.01} = 1.63 N_{\rm e}^{-1/2} = 0.230$ where $N_{\rm e}$ is the effective number of points[5].

The main outcome from this example is that in case that two time series data sets share the same statistical properties, up to the second order terms, but they originate from completely different physical systems i.e. different variability processes, the SF method can not discern any difference concerning the properties of the two processes. Erroneously the two realizations appear as they come from the same underlying process and also that possible deviations between them can be understood in the framework of statistical fluctuations.

---

[4] The significance level 1% represents the probability of rejecting H_0 while it is correct.

[5] For two data sets consisting of N_1 and N_2 points respectively, $N_{\rm e} = N_1 N_2/(N_1 + N_2)$.

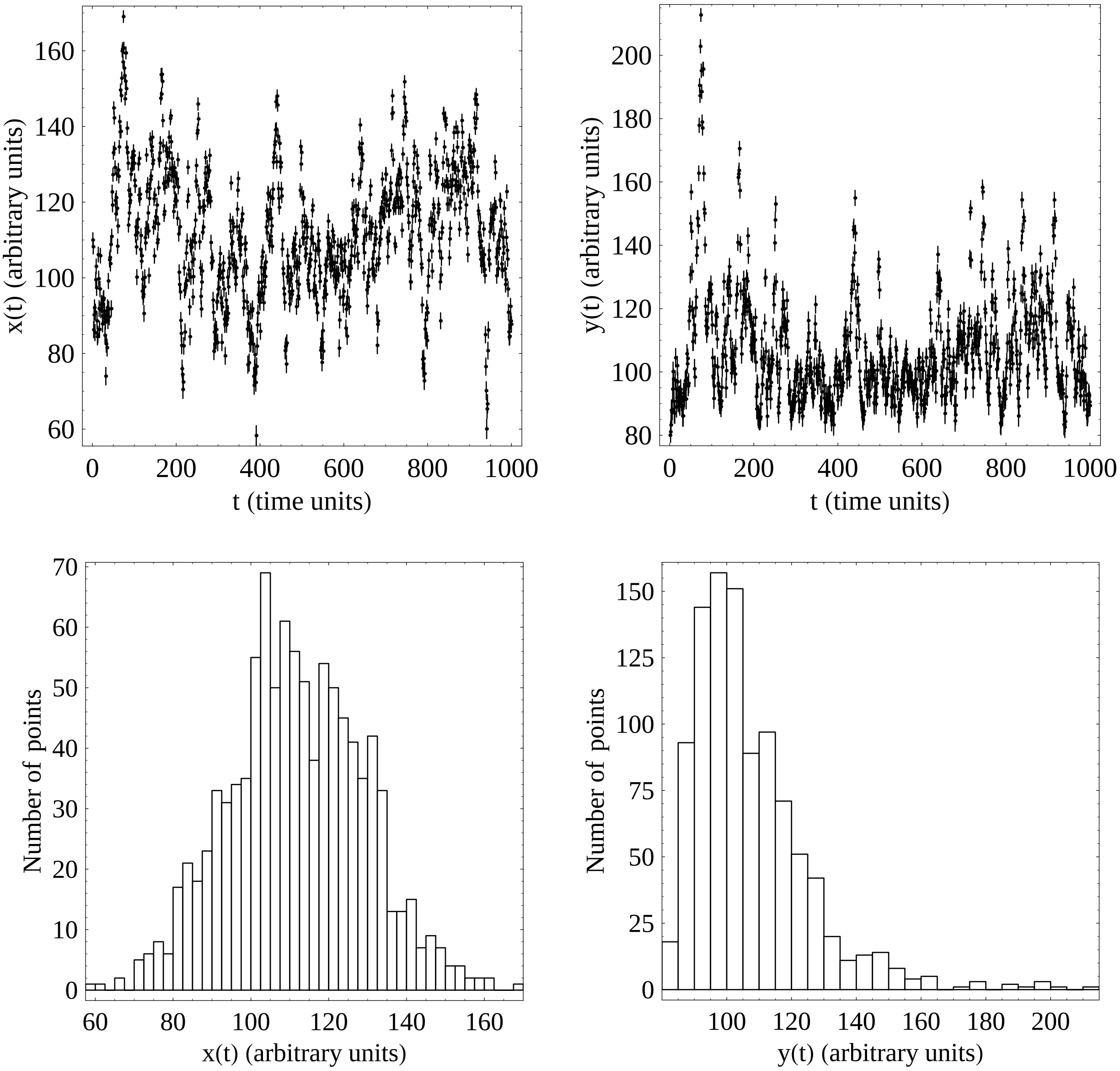


Figure 5.9: [Top panels] Time series generated by the linear (left) and the nonlinear (right) models given by eq.5.3, eq.5.4 respectively. Each one consists of 1000 points separated by 1 time unit.
[Bottom panels] The distributions of the measurements for the two systems (left, linear and right, nonlinear). The first two statistical moments (i.e. mean value and variance) are by construction tuned to 100 and 300 respectively.

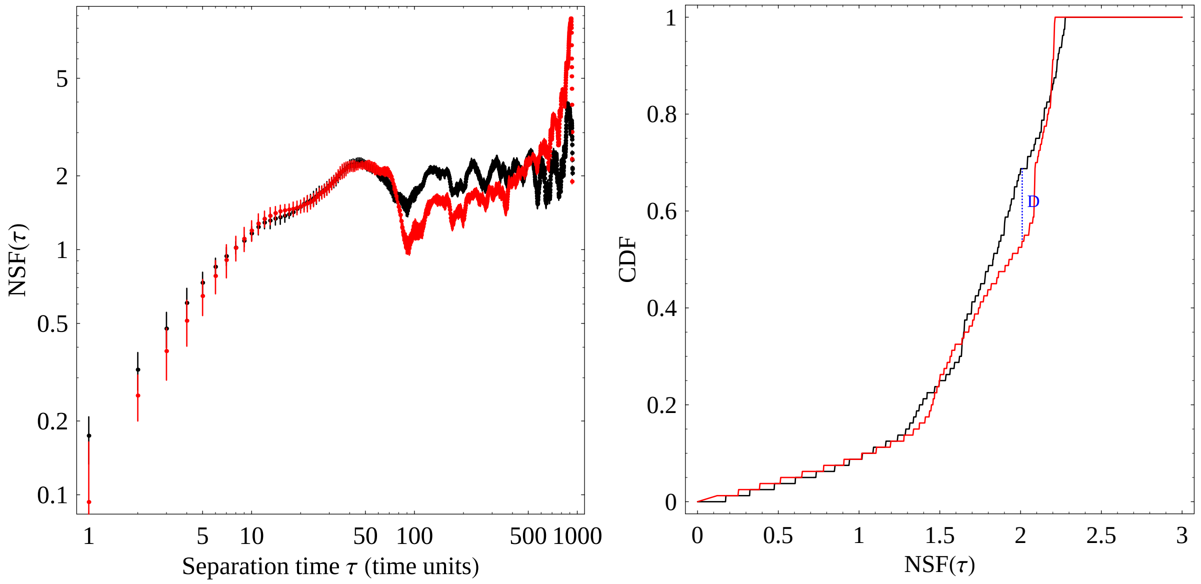


Figure 5.10: [Left panel] The NSF for the two systems as it is computed for the measurements shown in the top panel of fig.5.9. The black points represent the linear system and the red points the nonlinear system. [Right panel] The CDFs of the first 100 NSF estimates for the two systems, linear (black points) and nonlinear (red points). The maximum value of absolute difference between the two CDFs is indicated by the blue line $D = 0.163$.

5.2 First Order Time Series Analysis Modeling

The previous sections show clearly that there is nothing characteristic in the light curve of Mrk 501 which can be directly connected with any physical property of the source. The question now is, what information can be extracted from the ASCA data set using only the first two statistical moments of it? The answer to this question comes after applying some of the standard linear time series analysis methods (i.e. well established methods based on exact mathematical theorems) adjusted to the needs of astronomical data sets. The latter, for the case of Mrk 501, focuses only to the matter of the measurements errors (sect.3.1.3) since the data points are equidistant in time.

In any type of statistical analysis the art of model fitting requires the existence of a "bank" of standard types of models from which the analyst selects the particular one best fitting to the data. Exactly the same strategy applies equally well (or should be applied) to time series analysis. Before considering the problem of model selection it is firstly optimum to formulate a general class of models which appear to offer plausible descriptions for a broad range of different types of series. One of the first ideas was that a time series x_t may be regarded as being made up of three types of variations *classical decomposition model* (Chatfield 2003)

- A *trend* m_t describing the long-term behavior of the series.
- A *seasonal* component $s_t(T)$ describing the cyclical variations of the system having a known period T.
- A *noise component* n_t purely random in nature that is stationary (sect.3.1).

$$x_t = m_t + s_t(T) + n_t \tag{5.5}$$

The light curve of Mrk 501 (fig.5.4) suggests that the time series data set might be nonstationary (sect.3.1) since an increasing trend can be readily distinguished. In fig.5.11 the successive estimates of the mean normalized count rate, M_{NCR}, indicate that as time goes on M_{NCR} is increasing. Therefore the data set is characterized as nonstationary since it does not have a well defined mean value.

A seasonal component is not present in the ASCA data set of Mrk 501 as it can be seen from the periodogram estimation (fig.5.3, blue points), meaning that there is no dominant harmonic component present in the data set. In order to obtain a more quantitative estimate of the latter statement the periodograms

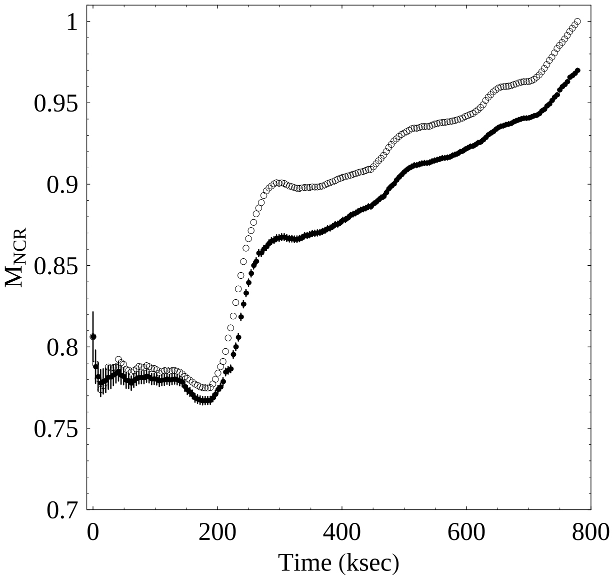


Figure 5.11: The filled circles represent the successive estimates of the mean normalized count rate $M_{\rm NCR}$ by including each time the next data point following the last entry, of the ASCA light curve of Mrk 501 (fig.5.4). These estimates do not converge to unity because the normalization of the light curve, being its mean value (6 counts sec^{-1}) estimated by Tanihata et al. (2001), is not taking into account the measurements errors. By ignoring the measurements errors the values of $M_{\rm NCR}$ converge finally to unity (open circles). Concerning the morphology of both curves, the values of $M_{\rm NCR}$ are characterized by an increasing trend with respect to time depicting the nonstationary behavior of the data set.

of the 2000 artificial light curves, produced in sect.5.1, are shown together with the periodogram estimates of the original ASCA light curve (fig.5.3) in fig.5.12 (the former with black points and the latter with blue points). Based on the distribution of the 2000 periodograms, for every frequency bin, f_j, a significance level (red-dotted lines) around its mean (red-solid line) can be established depicting the probability of having a given amplitude value, as a result of a nonharmonic process. By construction the simulated light curves do not have any harmonic components in their structure therefore the distribution of their periodogram amplitudes, at a given f_j, yields the probability of getting a given amplitude as a result of a purely random noise process. The red lines concerning the mean and the significances are the result of the interpolation of the corresponding points (mean, $+1\sigma$,$+2\sigma$, -1σ, -3σ, -5σ) as they come directly from the distribution of the periodogram amplitudes at the various f_j. No information can be extracted for the values of the interpolated quantities between two adjacent frequency bins. Since the highest periodogram amplitude of Mrk 501's light curve is only two standards deviations ($+2\sigma$, where "+" denotes that is situated on the right side of the mean value) away from the expected (mean) value, one concludes that there is no harmonic component embedded in the data set.

After resolving a trend in the data set and rejecting the existence of a harmonic component, how one can specify the noise component? As a first try one can simply fit to the data set x_t (fig.5.1) a straight line and check the behavior of the residuals. Based on the usual linear regression method a linear fit of the form $y(t) = \alpha t + \beta$ yields $\alpha = 0.00058 \pm 0.00004$ and $\beta = 0.754 \pm 0.018$. The residuals, $x_t - y(t)$ of the fit together with their ACF, $R_{1,1}(\tau)$, (for the first 100 estimates, till $\tau = 409.6$ ksec) are shown in the left panel of fig.5.13. The errors of the residuals come from error propagation (i.e. $\sqrt{(\delta x_t)^2 + (\delta y(t))^2}$) and the errors of the ACF are estimated based on the procedure described in sect.3.1.3 by varying normally the residual values 1000 times.

As it is proven in Priestley (1981) for a white noise process the distribution of the $R_{1,1}(\tau)$ for $\tau > 0$ can be very well approximated by a Gaussian distribution having a mean value 0 and standard deviation $1/\sqrt{N}$ where N is the number of points in the data set. Hence around 95% of the sample autocorrelations should fall between the bounds $\pm 1.96/\sqrt{N}$ (dashed lines in the ACF plot of fig.5.13) since 1.96 is the 0.95 quantile of the Gaussian distribution. The ACF of the linear fit residuals has 53 value out of the range of $\pm 1.96/\sqrt{N} = \pm 0.142$ therefore the residuals can not be characterized as white noise. Therefore a simple linear model seems to be inappropriate to describe the data set.

Nevertheless the ACF of the residuals can be used as a very insightful tool in order to choose the appro-

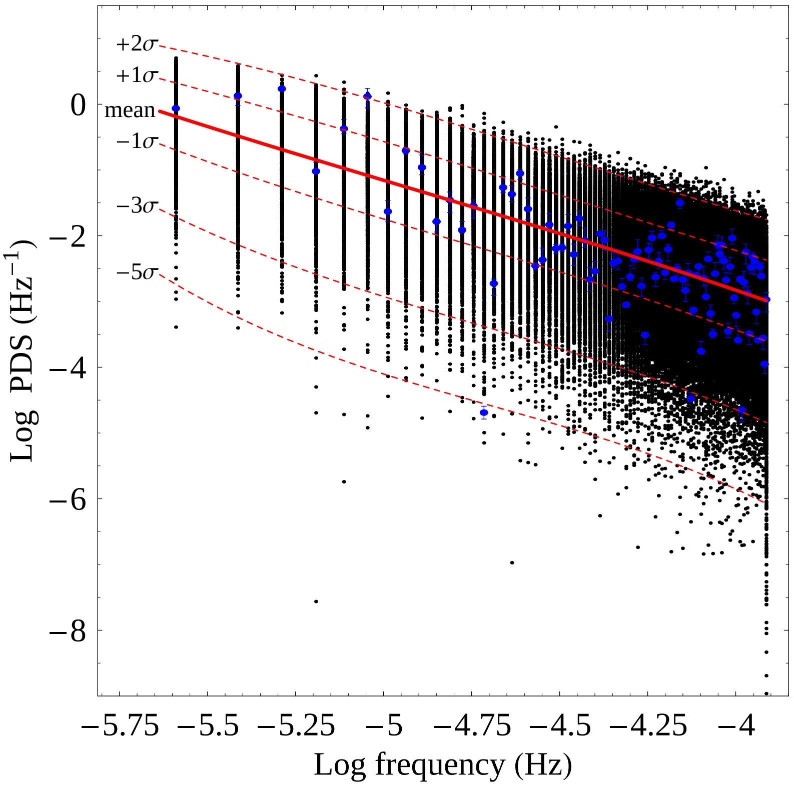


Figure 5.12: The logarithm of the PSD as a function of the logarithm of the frequency for the 2000 artificial light curves (sect.5.1) is plotted with black points. Blue points represent the PSD of Mrk 501 as it is shown in fig.5.4. Based on the distribution of the 2000 periodograms in every frequency bin, f_j, a significance level (red-dotted lines) around its mean (solid red line) is established. The line between two adjacent bins is the result of linear interpolation and is used only to guide the eye along to the various frequencies.

priate model for the data set in the following sense. The first autocorrelations seem to decay geometrically[6] since the ratio $r = R_{1,1}(\tau+1)/R_{1,1}(\tau)$ for the first time lags, $\tau = 1, \ldots, 4$ can be well fitted, through linear regression, by a constant line $r = 0.932 \pm 0.051$ (fig.5.14, left panel) suggesting that an AR(1) model of $r \approx 0.93$ might be appropriate.

To check the appropriateness of such a model the paired values from the original light curve (fig.5.1) are formed: $(x_1, x_2), (x_2, x_3), \ldots, (x_{190}, x_{191})$ and plotted together (fig.5.14, right panel). This plot suggests that a linear relationship between the values of the data set does indeed exist. A linear model fit of the form $x_{i+1} = rx_i$, by taking into account the errors in both directions[7] (Fasano & Vio 1988), yields an $r = 0.971 \pm 0.020$ suggesting an AR(1) model of the form

$$x_{i+1} = 0.971 x_i + n_i \tag{5.6}$$

is appropriate for the data set. Note that the value of r derived from this fit is consistent with the empirical one derived by the ratio of the ACF ($r = 0.932 \pm 0.051$).

The nature of the noise component, n_i, is specified through the residuals of this fit $n_i = x_{i+1} - 0.971x_i$ (fig.5.15, left panel). In order n_i to be consistent with white noise its ACF should have 95% of its estimates inside the range ±0.142. In order to estimate the ACF of n_i together with its uncertainties firstly the errors of n_i are estimated based on error propagation from all the involved parameters (x_i, x_{i+1}, r) and then the ACF uncertainties are estimated based on 1000 simulations (sect.3.1.3). The first 100 ACF (fig.5.15, right panel) estimates of n_i give 6 measurements outside this range depicting the fact that the noise component is consistent with a white noise process having a variance 0.0025 ± 0.0011. Therefore the fitted AR(1) model (eq.5.6) is the appropriate model for the ASCA light curve of Mrk 501.

---

[6] When the ratio of all the consecutive terms of a time series data set equals to a constant: $y_{t+1}/y_t = r$ then the terms y_t are of the form $y_t = y_0 r^t$ and the data train comprises the simplest case of geometric time series.

[7] The uncertainties on both axes should be considered since both of them contribute with the same weighting factors in the fit.

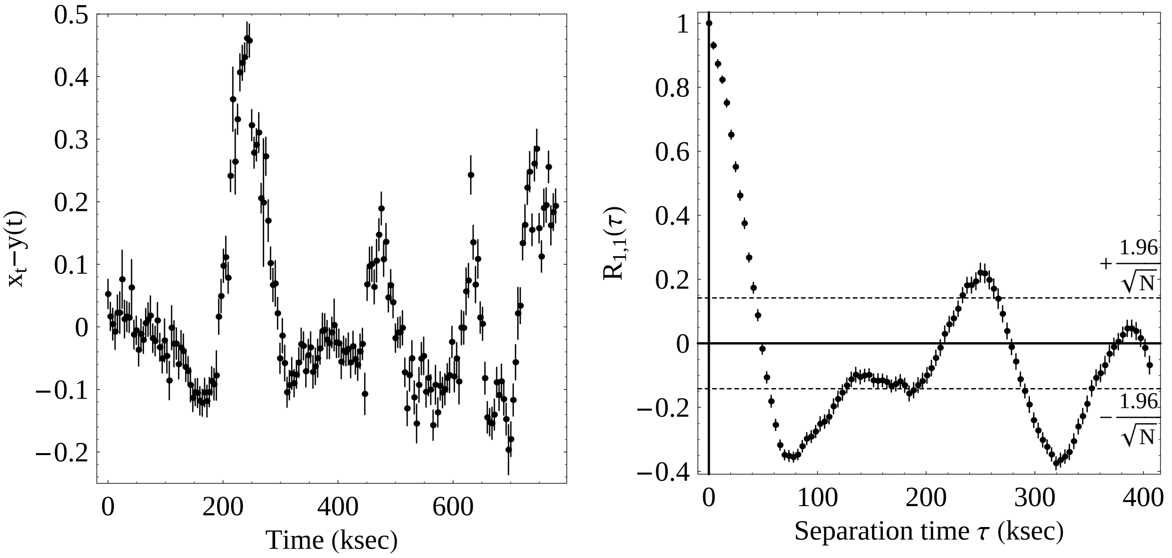


Figure 5.13: [Left panel] The residuals $x_t - y(t)$ after subtracting the linear fit $y(t) = \alpha t + \beta$ ($\alpha = 0.00058 \pm 0.00004$, $\beta = 0.754 \pm 0.018$) from the observed ASCA data set x_t shown in fig.5.1.
[Right panel] The ACF of the residuals together with the boundaries (dashed lines) denoting the area ± 0.142, where 95% of the estimates should have been falling if the residuals were white noise.

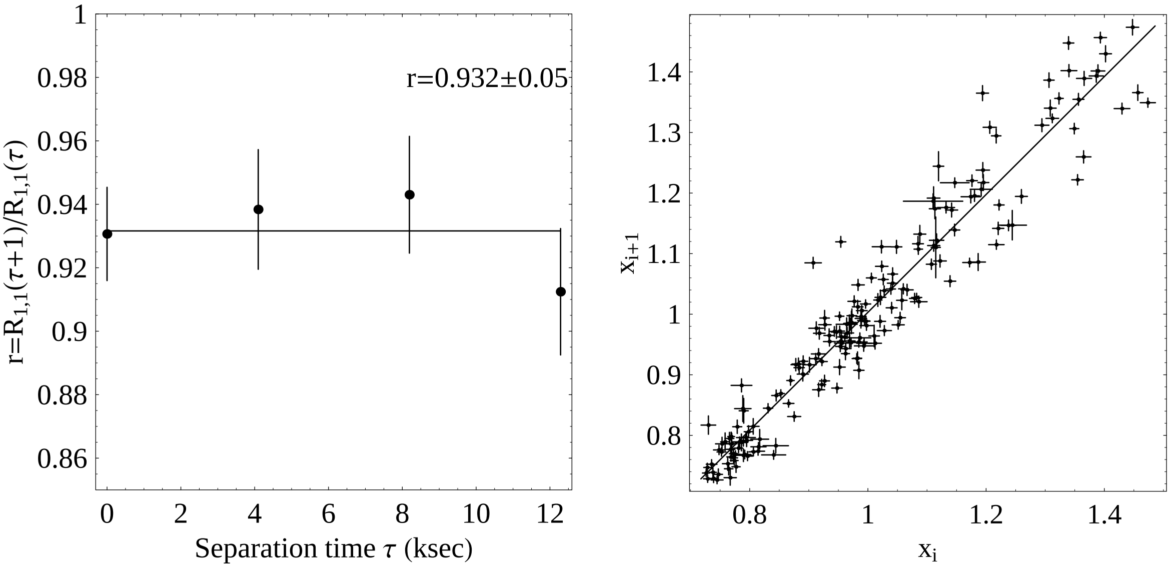


Figure 5.14: [Left panel] The ratio r for the first five ACF estimates $\tau = 1, \ldots, 4$ can be fitted by a constant line $r = 0.932 \pm 0.051$.
[Right panel] The linear relationship between the values x_i and x_{i+1} for $i = 1, \ldots, 190$ is fitted with the linear model $x_{i+1} = r x_i$ with $r = 0.971 \pm 0.020$. The fit takes into account the errors on both axes (Fasano & Vio 1988).

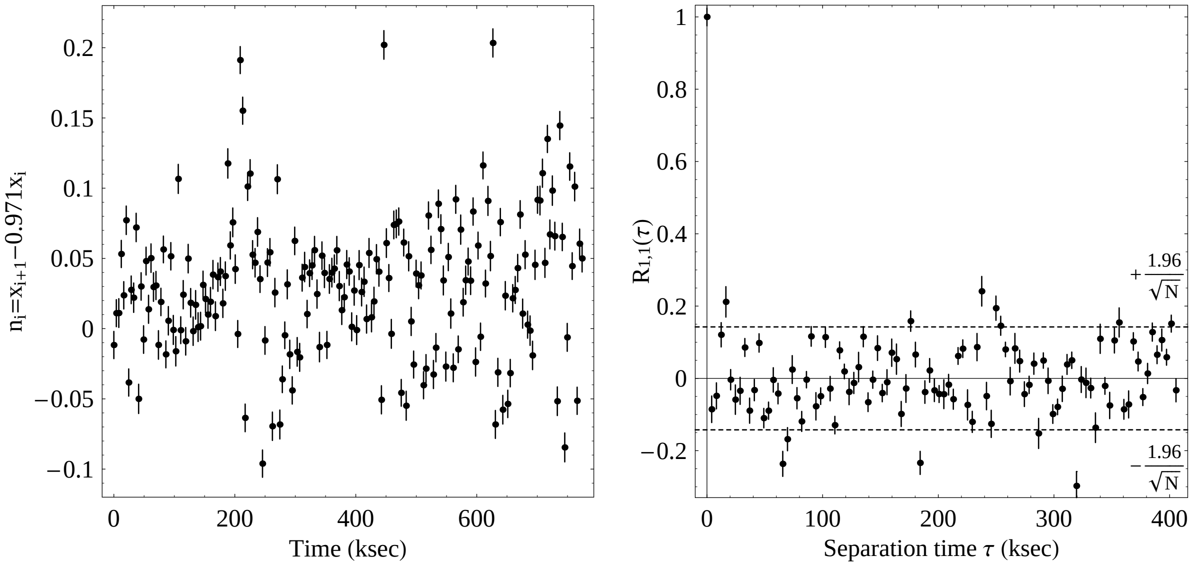


Figure 5.15: [Left panel] The residuals n_i estimated by $n_i = x_{i+1} - 0.971x_i$.
[Right panel] The ACF of n_i. The dashed lines indicate the bounds ±0.142, where 95% of the estimates should have been falling in the case of a white noise. Six out of the one hundred estimates fall outside this range fact that verifies that n_i consists is a purely noise component.

Chapter 6

TIME SERIES VARIABILITY STUDIES OF MRK 421

In the following sections the X-ray time properties of Mrk 421 are examined thoroughly through a direct application of the time series analysis methods, presented in chap.3, to the two data sets presented in chap.4. The first data set consists of the PCA observations, having a temporal resolution of a few minutes, and the second one consists of the 15-day time-averaged ASM observations, both of them covering a time period of 9 years. The former is used in order to check whether or not the short-term variability behavior of the source is characterized from any time property which is preserved throughout the years. The latter data set is used in order to unveil the dimensionality of the source on a long-term basis and determine the number of parameters being responsible for the observed variations.

6.1 Power-Law Spectra and Stationarity

As mentioned in sect.3.1 the AGN light curves exhibit PSD functions (sect.3.3.1) which can be well represented by power-laws $\mathcal{P}(f) \propto f^{-a}$ with $1 < a < 3$. Initially, the lack of any characteristic time feature in the shape of their PSDs indicates that none of the processes happening inside the source, as they are seen through the jet (sect.1.2.2), have any periodic time properties. Moreover this also demonstrates that none of the existing time scales, coming from different source components, prevails over the others.

The existence of a significant *peak* in the PSD at a given frequency f_c is a direct indication for the existence of a harmonical time behavior embedded in the source with a period of f_c^{-1}. These kind of peaks have been observed by Mittaz & Branduardi-Raymont (1989), in the X-ray band using EXOSAT (sect.2.2.3), for the case of the Seyfert galaxy NGC 6814, having a period of 12 ksec. As shown by Done et al. (1992) this periodic component is responsible for 75% of the total source variability. Additionally "breaks" in the PSD show that the observed variability can be explained in the sense of two dominant components and as such the "break" frequency can be used to estimate some physical parameters of the source. This kind of "broken" PSD has been seen by Papadakis et al. (2002) in the case of the narrow-line type I Seyfert galaxy (sect.1.2.1) Ark 564, using ASCA X-ray observations. The break occurs at $\sim 2 \times 10^{-3}$ Hz and from this estimate one can derive the black hole mass and the accretion rate in the source.

In contrast, for the case of BL Lac objects there are no particular "fixed" patterns in their PSD which can be used as handles to disentangle the properties of the various emission regions. Since these objects point their powerful jets towards the observer their light curves comprise the summation of the emission coming from the various regions inside these jets. Examination of specific details in their light curves is not insightful at all since each light curve is only one realization of the underlying process[1], stochastic

---

[1] The term *underlying process* refers to the ensemble of the emission activity occurring within the jet. It comprise the result of all the radiation processes as they are mapped through the jet in the observed light curve and it characterizes the variability behavior of the source.

or chaotic (i.e. high- or low-dimensionality, see sect.3.1 and sect.3.2), drawn from the ensemble of light curves that can be generated from the source. That means that two light curves from the same source might have completely different statistical characteristics, as they are described by the first two statistical moments, even if the physical parameters describing the variable emission process remain the same. Every new realization looks different from the previous ones i.e. different mean values and variances, and these changes can simply reflect the statistical fluctuations which are inherent in the nature of the variability process as opposed to genuine changes in the nature of the process itself (Vaughan et al. 2003).

The characterization of individual flaring events is in a sense meaningless since phenomenological arguments do not help to gain any information concerning the physical source parameters. In order to give an example, some phrases from Cui (2004) for the case of Mrk 421 are presented in the following itemization form

- The light curve phenomenology is complex.
- There is hardly any time period that resembles a “quiescent state”.
- Smaller flares are nearly always present.
- The source is by no means quiet.
- Two larger flares are recognizable.
- Superimposed on them are many “spikes”.
- Coexistence of flares on different time scales.
- Coexistence of flares on a wide range of scales.
- The flares seem to occur at all time scales.
- The seemingly scale-invariant nature of flaring activities.

All the aforementioned statements do not give any interesting information concerning any physical or any time property of Mrk 421. Through this kind of unnecessary phenomenological description the analysis does not conclude to any final statement concerning the most important questions dealing with the variability properties of the source: ARE ALL THE TIME PROPERTIES IN THE DATA SETS COMPLETELY RANDOM? ARE THERE ANY TIME PROPERTIES PRESERVED THROUGHOUT THE OBSERVATIONS?

For the last 20 years individual light curve studies have been conducted extensively without any major progress or any final well defined results, despite the fact of the large amount of available data. Different observations, for the same source, give different source parameters being completely inconsistent among them. For the case of BL Lac objects a crucial model parameter is the Doppler factor δ (eq.1.1) characterizing the bulk velocity of the source Γ. For the case of Mrk 421 the values for this parameter range from $\delta \gtrsim 10$ (Gaidos et al. 1996) up to 20–30 (Maraschi et al. 1999) and even larger $\delta > 50$ (Krawczynski et al. 2001). Remarkably direct *Very Long Baseline Array* (VLBA) observations of the parsec-scale radio jet of the source reveals apparent blob speeds of only $(0.10 \pm 0.02)c$ (Piner & Edwards 2005). The main reason firstly for this parameter divergence and secondly for this discrepancy between models and observations might be that all the models use some parameters derived directly from the light curve. Such a parameter is the time scale of fastest flares which is considered to be an upper limit for the linear dimensions of the source. Of course this parameter can change from one data set to another not just due to physical reasons i.e. different source size, but also due to statistical fluctuations.

Physical processes with red noise PSD have intrinsically the ability to change their statistical moments even if the physical properties of the underlying process itself remain unchanged throughout the observing period. Therefore since the statistical moments change with time, based on the definition of stationarity in sect.3.1, these processes appear always nonstationary even though they might have genuinely a well defined mean value and variance (Vaughan et al. 2003). The top panel of fig.6.1 shows an artificial light curve, 3000 points long, produced from a well defined PSD with a spectral index $a = 2$ (sect.3.3.6). Since the PSD does not vary with time the variability power as a function of temporal frequency remains the same and thus the system is considered to be emerged from a genuine stationary process. Therefore it would have been reasonable to expect the resulting time series, consisting a realization of this process, appear stationary but this is not the case. The artificial light curve is partitioned into 150 segments each one consisting of 20 consecutive points, yielding 150 estimates for the variance, S^2, of the process (fig.6.1, middle panel).

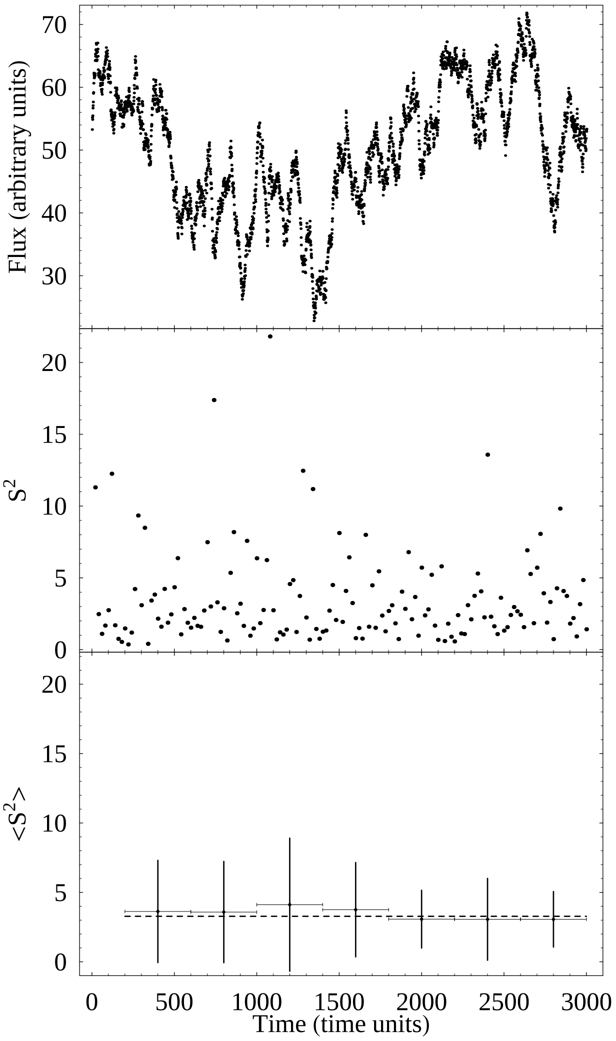


Figure 6.1: [Top panel] Artificial light curves 3000 points long created by a PSD with a spectral index $a = 2$.
[Middle panel] Estimates of the variance S^2 measured from segments (150) of 20 consecutive points. The estimates are not consistent with the constant line fit $y = 3.478 \pm 0.265$ since χ^2=1571.32 for 149 DOF corresponding to a NHP of 0.
[Bottom panel] Average variance $\langle S^2 \rangle$ measured by binning the individual variances into groups of 20 consecutive estimates. The dotted line represents the constant line fit $y = 3.277 \pm 0.133$ having a χ^2=0.093 for 6 DOF and NHP=1 (after Vaughan et al. 2003).

It is clear that S^2 is not consistent with being constant since a constant line fit yields $y = 3.478 \pm 0.265$ having a χ^2=1571.32 for 149 DOF (NHP=0). Therefore the variance changes with time fact that has nothing to do with measurement errors whatsoever, since the simulation has zero errors. It is evident that significant variations in the variance of a light curve (i.e. significant variability) are not sufficient to claim that the variability process is nonstationary i.e. the physical properties of the source have changed. The simulated process has a constant PSD, representing a system where the physical parameters responsible for the variability do not change with time, but through its realization the latter appear to be different.

On the other hand it is also possible that the underlying process responsible for the observed variability does indeed change with time (e.g. the PSD changes) in which case the variability process can be characterized as genuine nonstationary. In this case the physical parameters responsible for the variable process in the source change significantly with time and thus each variance level corresponds to a different set of parameters.

As the purpose of time series analysis is to gain insight into the process through statistical approach, not through the features of specific realizations, a more robust approach is needed to determine whether the data are produced by a genuine stationary or nonstationary process. It is more insightful to consider whether the expectation values of the variance $\langle S^2 \rangle$ are time-variable (Bendat & Piersol 1986; Vaughan et al. 2003). The bottom panel of fig.6.1 shows the $\langle S^2 \rangle$ estimates together with their standard deviation as they are derived after binning the individual variance estimates S^2 (fig.6.1, middle panel) into groups of 20 consecutive estimates. The estimates are consistent with a constant fit line $y = 3.277 \pm 0.133$ with χ^2=0.093 for 6 DOF having NHP=1, fact that reveals the true nature of the process being stationary. The deviations

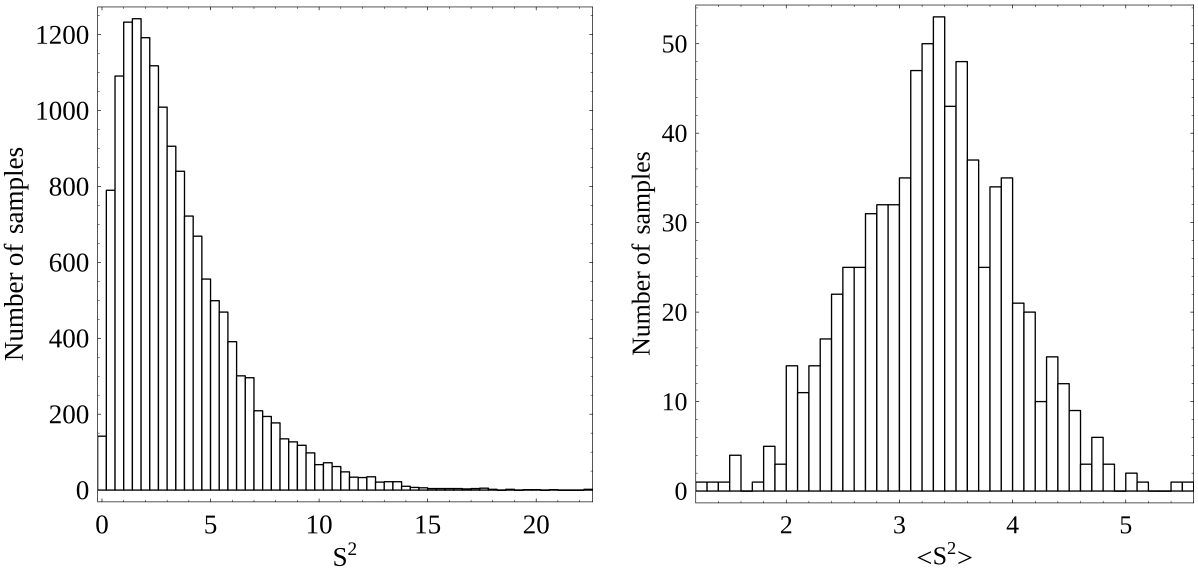


Figure 6.2: The distribution of S^2 (left panel) and $\langle S^2 \rangle$ (right panel) as it is estimated from the 300000 point long artificial light curve. In total there are 15000 estimates of S^2, each one calculated from 20 consecutive points, and 750 estimates of $\langle S^2 \rangle$, each one calculated from 20 consecutive S^2 estimates.

in the estimates of $\langle S^2 \rangle$ are governed by the normal Gaussian statistics in contrast to those of S^2. This can be seen in fig.6.2 where the distributions of S^2 and $\langle S^2 \rangle$ is shown as they result from an extended version of the same simulation 300000 points long.

6.2 Mrk 421 and Stationarity

6.2.1 General methods for estimating the mean variance

The point that should be elucidated is whether or not the observed erratic variability of Mrk 421 is the fluctuational outcome of the same radiation mechanism, preserving the same physical properties throughout the entire observing time period, or the result of true changes in the physical variability parameters of the source. As discussed in sect.6.1 it is more insightful to consider the expectation values of the statistical moments of a specific realization and check whether or not they vary with time. Significant variations of these quantities which are not expected up to a certain confidence level by Gaussian fluctuations, unveil that the observed variability is the outcome of a true nonstationary emission process caused by different physical properties of the emission region. The basic concept behind this idea is that different emission states of the same variability process or different variability emission regions within the jet of the BL Lac, should have different statistical properties reflecting exactly the fact of this alteration.

As discussed by Vaughan et al. (2003), if the process responsible for the flux variability in a source is stationary then its PSD should be constant in time. Therefore, light curves taken at different time periods should have the same PSD, estimated through their binned logarithmic periodogram (sect.3.3.2), at a given confidence level. That means that the expectation value of variance will be the same from epoch to epoch (eq.3.17) but individual variance estimates will fluctuate following eq.3.20. In Papadakis & Lawrence (1995) another method suitable for testing whether large AGN data sets display evidence for genuine nonstationarity is suggested, again by comparing PSD derived from light curves obtained at different time periods. In particular they form the test statistic $\mathcal{S}$ based on the ratio of two normalized periodograms and if this deviates significantly from its expected value for stationary data ($\langle \mathcal{S} \rangle = 0$) then the hypothesis that the data set is stationary can be rejected at some confidence level.

Both methods aim to check whether the mean variance of the process (eq.3.17) differs significantly from observation to observation by comparing the PSDs of the corresponding observations. This indirect methodology of checking whether or not the mean variance of the process changes as a function of time over a given period can be quiet complicated and sometimes uncertain especially when the data sets are irregularly sampled. There is no straightforward way to estimate the PSD of a “gappy” data set without making a number of additional simulations (Uttley et al. 2002) in order to check the spectral behavior of the data set in time scales equal to the gap duration (e.g. the arbitrary leakage in adjacent frequencies). Also one should also include the measurements uncertainties in order to check how they affect the PSD estimates in the various frequencies again through simulations (sect.3.1.3).

Ideally all the aforementioned procedures, involving comparison of the various PSD, at the end should be consistent with the $\left\langle S^2 \right\rangle$ estimates. Of course a most robust and direct way to specify the mean variance $\left\langle S^2 \right\rangle$ of a light curve is directly from the data set itself without making any simulations at all. Given sufficient number of data one can simply test whether the expectation values of the variance, estimated directly from an ensemble of light curves, at the various observed epochs are consistent with a stationary process as it is done in the simulation in sect.6.1 (fig.6.1). The basic quantity that one has to specify is the standard deviation of the mean variance $\left\langle S^2 \right\rangle$ (i.e. the error bars of fig.6.1) consisting an estimate of the expected scatter. Ideally this uncertainty can be obtained directly from the data set, by measuring the standard deviation of multiple mean variances. In the case of not sufficiently large data sets a shape for the PSD is assumed and after producing a number of simulated light curves (sect.3.3.6) the distribution of the variances, S^2, expected for a stationary process can be formed. This allows to estimate the limits within which S^2 is expected to be distributed under the assumption of a stationary underlying process (Vaughan et al. 2003).

6.2.2 Stationarity analysis of the PCA data set

In order to check about genuine intrinsic changes in the variability behavior of Mrk 421 (i.e. nonstationarity), the PCA data set (see sect.4.1 and fig.4.2) is chosen since it consists the biggest collection of data ever taken for a BL Lac from PCA (1.5 Msec). As such it can be used in order to estimate directly the mean variance $\left\langle S^2 \right\rangle$ for every observing period together with an error coming from the various variance estimates S^2. There is no need to compare the PSD from the various observing periods since their estimation might be quite tricky and ambiguous due to the numerous gaps. Additionally for a given observation there is no need to make any simulations in order to estimate the expected scatter of the variance S^2 coming from a stationary process with the same PSD as a given observation. Both the expected value of the variance $\left\langle S^2 \right\rangle$ and its error can be directly and unambiguously specified by the data set itself.

The artificial light curve used in sect.6.1 is an idealized version of a real data set in the sense that it does not have flux uncertainties. The data set of Mrk 421, as every real light curve, contains measurements uncertainties adding more variance to the data set. In order to estimate the genuine *intrinsic* variance of the source the *excess variance* is employed (Nandra et al. 1997) consisting the outcome of the variance after subtraction of the measurements errors.

Assume a light curve comprising a time series of fluxes x_i with errors $\sigma_{\mathrm{err},i}$ measured at discrete times t_i, with $(i = 1, 2, ..., N)$. Then

$$\sigma^2 = S^2 - \overline{\sigma^2_{\mathrm{err}}} \tag{6.1}$$

where S^2 is the flux variance of the light curve given by eq.3.19. and $\overline{\sigma^2_{\mathrm{err}}}$ is the mean square error

$$\overline{\sigma^2_{\mathrm{err}}} = \frac{1}{N}\sum_{i=1}^{N} \sigma^2_{\mathrm{err,i}} \tag{6.2}$$

The normalized excess variance σ^2_N is given by

$$\sigma^2_N = \sigma^2/\overline{x}^2 \tag{6.3}$$

The 14 archived observations (tab.4.1 and fig.6.3) are separated into 8 periods each one covering roughly 120 days on a yearly basis. Then for each period the count rate is binned in 64-sec bins and the latter are collected in groups lying within one satellite orbit ($\sim$ 5440 sec, tab.2.1). Therefore for each period there are 5440-sec time spans consisting of successive 64-sec binned observations. In order to estimate σ^2_N for each orbit and then its mean value with the corresponding error within each observing period it should be first ensured that there is a large number of points able to yield a statistical robust estimate of both the excess variance (within an orbit) and its mean value (within a period). Therefore, two selection criteria are applied to the data set choosing only those periods consisting of more than 20 orbits each one consisting of more than 20 consecutive 64-sec bins. Note here that since data gaps within an orbit reduce the statistical significance on the estimated value of σ^2_N, orbits with data gaps shorter than 1280 sec (20×64) are rejected. The final sample consists of 5 periods as the only ones fulfilling the aforementioned selection criteria.

After selecting these optimum periods for each orbit within them the σ^2_N is calculated from a sample of 20 successive 64-sec data bins. For those orbits consisting of more than twenty 64-sec observations a sliding

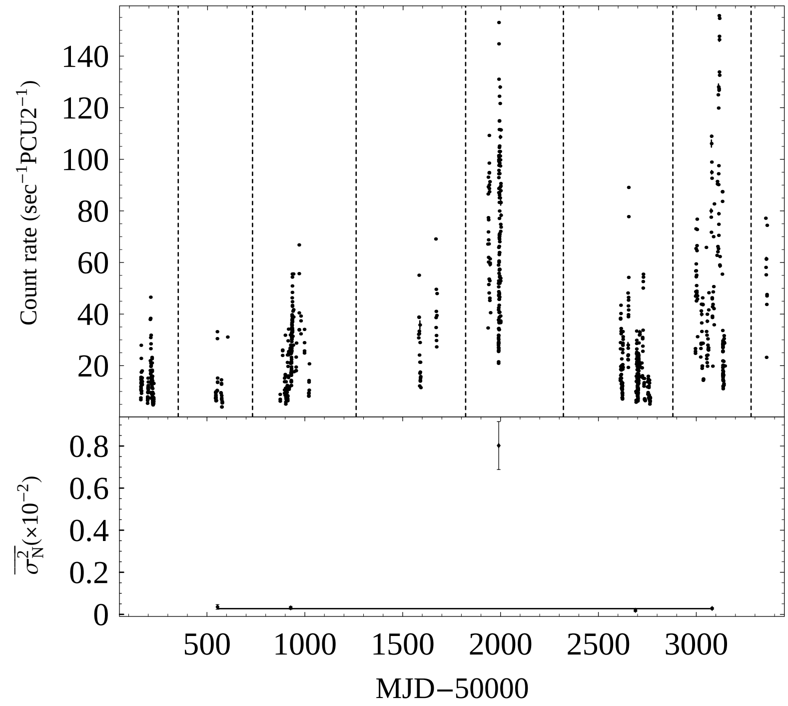


Figure 6.3: [Top panel] The light curve of Mrk 421 as shown in fig.4.2. In combination with the bottom panel it shows which periods are finally included in the estimation of $\overline{\sigma_N^2}$.
[Bottom panel] The $\overline{\sigma_N^2}$ for the five campaigns. A constant line fit for the four campaigns (occurring at MJD 50554, MJD 50928, MJD 52690 and MJD 53081) yields $y = (2.72 \pm 0.28)10^{-4}$ with reduced χ^2 of 0.815 for 3 DOF.

window of 20 successive observations is used in order to estimate a mean value of σ_N^2 within each orbit. Finally for each observing period one can readily compute an overall mean of value of normalized excess variance, $\overline{\sigma_N^2}$ together with an error (fig.6.3) based on the estimated orbital mean values of σ_N^2 making use of the usual weighted mean/standard deviation formulas (e.g. Bevington & Robinson 1992).

From fig.6.3 it can be seen that four out of the five campaigns, occurring at MJD 50554, MJD 50928, MJD 52690 and MJD 53081, appear to have the same statistical properties concerning the mean value of σ_N^2. Differences of the mean values can be well attributed to random fluctuations expected from Gaussian statistics. A constant line fit to these measurements yields a value of $(2.72 \pm 0.28) \times 10^{-4}$ which actually represents a measure of the mean value and the standard deviation of the parent distribution where these fluctuations originate. The $\overline{\sigma_N^2}$ for the third campaign, MJD 51991, differs significantly from the other four in the sense that its deviation above the mean value ($\sim$ 6.8 standard deviations) can not be interpreted in the frame of random fluctuations. During this campaign the variability properties of the source change dramatically indicating that the physical properties of the underlying process are different from the ones of the other four campaigns. In this sense one can readily distinguish two intrinsic variability states of the source yielding a nonstationary behavior.

This is the first time that something like this is deduced directly from observations for a BL Lac object and especially for the case of Mrk 421 can be used in order to distinguish at least two variability states of the source. This of course does not exclude the existence of other states which can be simply not manifested in the given time period. These states are characterized from statistically significantly different variability properties therefore it is reasonable to consider that the variable emission comes either from two different emitting regions within the jet or from the same emitting region in which its physical parameters responsible of the emission vary between two stages. Based on the first scenario since Mrk 421 is a BL Lac object its jet points towards us forming a small angle with the line of sight, 1° (Piner et al. 1999; Piner & Edwards 2005) and hence numerous jet components (at least two) with different physical properties being situated at different regions contribute to the observed X-ray light. Based on the second scenario the emission region is the same but its physical properties (e.g. geometrical dimensions, magnetic field) change between at least two levels.

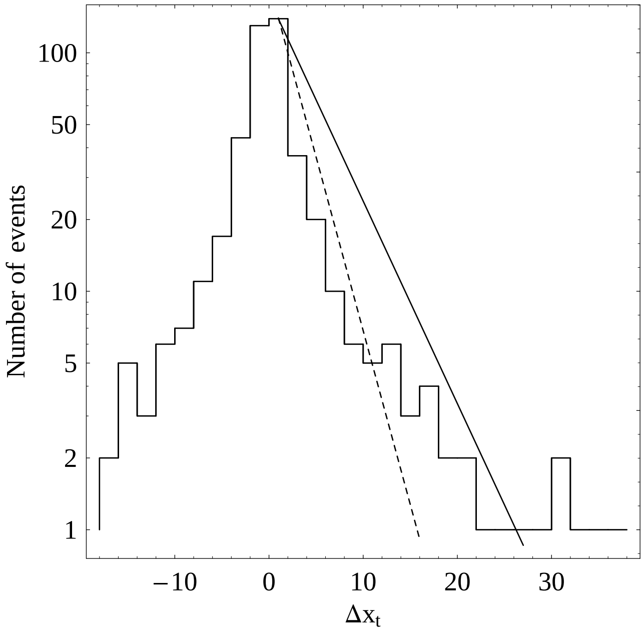


Figure 6.4: The histogram of the increments Δx_t, having a bin size of 2. In order to specify the slope of the wings of the distribution a linear regression model is fitted (solid line) having a slope of -0.196 ± 0.013. The dashed line represents the same fit, having a slope of -0.335 ± 0.009, in the case that the increments, Δx_t, having a mean value of -0.095 and standard deviation of 6.920, were distributed Gaussian.

6.2.3 Intermittent behavior

The existence of a nonstationarity variable behavior is the first direct indication (i.e. through observations) that Mrk 421 has the ability to undergo transitions between at least two energetic levels characterized by different variability properties. This intermittent dynamical behavior (see sect.3.2.1) allows the source to change significantly and suddenly its long-term statistical properties due to genuine alteration of the source physical parameters. The significance of this change can be established based on the value of the normalized excess variance during this period, derived in the previous section, being 6.8 standard deviations above its usual long-term values.

Concerning now the time duration of the transition the only secure statement that can be made should come from the ASM light curve (fig.4.4) since the time coverage of these observations is much more representative for the overall source behavior. In fact fig.4.4 shows that the period where the value of the normalized excess variance deviates, MJD 51991, coincides with an epoch of high source activity. This high source state lasts around a year posing an upper limit of equal duration for the transition time. Additionally, from the ASM light curve one would expect that at least during two more periods MJD 52690 and MJD 53081, where the PCA normalized excess variance estimates are low, could have been parts of intermittent events. Unfortunately for these periods the PCA observations sample only the left wings of the flaring events making it impossible to distinguish whether these states belong to a truly intermittent event or simply to the usual quiescent source state.

In case that these events are part of an intermittent behavior then an indirect evidence can be arisen from the histogram of the increments $\Delta x_t = x_{t+\tau} - x_t$ (fig.6.4), where x_t represent the data points of the observed long look PCA light curve and $\tau = 5440$ (Paladin & Vulpiani 1987; Vio et al. 1992). The histogram shows that big differences in two adjacent values of the data set can occur more often than expected from Gaussian statistical fluctuations. The tail of the distribution can be well fitted by an exponential power-law $\propto \exp(-0.196\Delta x_t)$ with a much flatter index from the one expected from the corresponding Gaussian distribution $\propto \exp(-0.335\Delta x_t)$. The parameters of the Gaussian distribution are the mean value and the standard deviation of the increments, -0.095 and 6.920 respectively. The existence of broad exponential tails implies a probability of large increments which is much larger than the one expected from Gaussian distribution. That means that in the PCA light curve there are indications of such events, occurring more often than expected, which they might belong to bigger intermittent flares.

6.3 The Short-Term Variability Behavior of Mrk 421

6.3.1 Nonlinearity and burst activity

Based on the analysis of the previous sections (sect.6.2.2, sect.6.2.3) it is evident that Mrk 421 exhibits nonstationary variable behavior representing genuine differences in its emission properties. Following that, the next step needed in order to probe more into the time properties of the underlying variability process is to check whether or not the physical mechanism responsible for the emission can be considered as either linear or nonlinear in its dynamical representation (sect.3.2).

As it is discussed in sect.3.2 dynamical systems described by nonlinear equations are quiet difficult or even impossible to predict their behavior since they appear through their realization as random. Therefore on certain occasions they can be confused with simple linear stochastic systems which are by construction purely random processes. An early and important finding of quasar variability is the recognition of its unpredictable variability nature which was thought initially to be purely stochastic (for the case of 3C 273 by Manwell & Simon 1968). A shot-noise process (sect.5.1.1) was considered to be the appropriate model for explaining the temporal behavior of the source. Based on this random event model, shots occur following a Poisson distribution having a fixed average number of events per unit time, connected with the variance of the light curve, and a fixed time duration (see also Terrell & Olsen 1970). Unfortunately shot-noise processes have PSD which do not change (within the statistical fluctuations) with time and thus the mean variance $\langle S^2 \rangle$ remains the same with time (sect.6.2.1).

In this setting, sudden bursts of very large amplitude embedded in the light curves of AGN can not be interpreted by linear stochastic processes, in which *shot-noise* processes belong to, comprising the most important limitation of these models (Vio et al. 1991). This kind of burst behavior has been seen for several sources such as 3C 446 (Barbieri et al. 1990), OJ 287 (Valtaoja et al. 2000), 3C 345 Kidger (1989), 3C 390.3 (Leighly & O'Brien 1997), NGC 4051 (Green et al. 1999). For the last three sources it has been shown from the same authors (except for 3C 345 shown by Vio et al. 1991) that a nonlinear physical mechanism is responsible for their erratic flux variability proving profoundly that methods from the nonlinear theory must necessary be employed in order to understand the number of parameters affecting the evolution of the system (sect.3.1). This is probably the most fundamental parameter that one has to specify since it characterizes the nature of the underlying process being stochastic, in the case that the luminosity variations depend on numerous physical parameters, or deterministic when few parameters convolved in a nonlinear way designate the course of the system.

6.3.2 Nonlinearity tests

Various models and tests have been developed to trace possible nonlinearities embedded in a data set. In general there are three big categories of tests that one can choose, depending on the case:

- frequency domain test: This test was developed by Subba Rao & Gabr (1980) and it was improved by Hinich (1982). The only restriction that the data set should fulfill is that it should consist from equidistant measurements.

- time domain tests: They are divided into two categories *portmanteau tests* and the *Lagrange multiplier tests*. The former consists of two subcategories. Firstly the χ^2 *tests* applied mainly by Maravall (1983); Luukkonen et al. (1988); Li & McLeod (1986); Davies & Petruccelli (1986); Lawrance & Lewis (1985). It seems that these tests are more appropriate to test the adequacy of a given model (linear or non-linear) based on the study of residuals than to detect a nonlinear behavior embedded in the data set. Secondly the *additivity tests* which are based on Tukey's one DOF test for non-additivity (Keenan 1985; Tsay 1986). These tests detect second order nonlinearities but they are not appropriate for data sets with errors.
  The *Lagrange multiplier tests* consist of the *likelihood ratio test*, the *Wald test* and the *score test* (Saikkonen & Lukens 1988; Buse 1982; Guegan & Pham 1992; Guegan & Wandji 1996). These tests allow to find certain nonlinearities such as self-exciting threshold autoregression (SETAR) (Ghaddar & Tong 1981; Clements et al. 2003) and exponential autoregression (EAR or ExpAr) (see Lawrance & Lewis 1980).

- non-parametric tests: These test are constructed using kernel methods and are able to detect nonlinearities and measure their strength but they can not describe their detailed structure (Hjellvik & Tjostheim 1995).

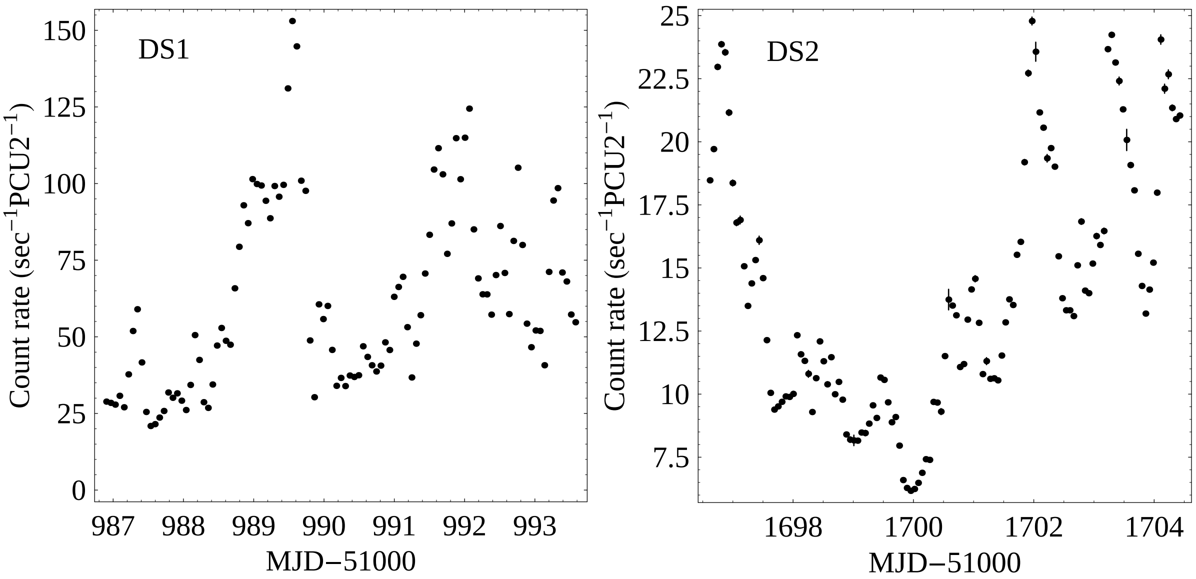


Figure 6.5: The two longest and most densely sampled data sets DS1 (left panel) and DS2 (right panel) for Mrk 421 coming from the PCA data set (tab.4.1). The former consists part of the proposal 60145 and the latter is the complete data set of the proposal 80172 (see fig.4.1).

- statistical tests: These are the most robust tests for identifying nonlinearities in a time series data set. Firstly one specifies some linear process comprising the null hypothesis and then generates surrogate data sets which are consistent with that. Finally a discriminating statistic is computed for the original time series and for each one of the surrogate data sets. If the values between the original data set and the ensemble of surrogates differ significantly then the null hypothesis is rejected and nonlinearity is detected. This method was first developed by Theiler et al. (1992) and from then on it has been updated in several aspects making it more sensitive for data sets that there is an *a priori* knowledge of the behavior of the system that they come from (e.g. pseudoperiodic behavior Small et al. 2001) or for unevenly sampled time series[2] (Schmitz & Schreiber 1999). This is the method that is going to be used in order to trace possible nonlinearities in the long and the short-term variability behavior of Mrk 421. More details about the method are presented in sect.6.3.3.

6.3.3 Linear or nonlinear source behavior? The method of surrogates

In order to check for a possible non linear dynamical behavior embedded in Mrk 421 for small time scales such as days the method of surrogates Theiler et al. (1992) is applied to two of the most densely sampled PCA data sets DS1 and DS2, consisting of 107 and 125 points respectively (fig.6.5). In total there are 5 and 15 missing points respectively in DS1 and DS2 filled with linear interpolated values between adjacent observations. Since these data sets form almost two uninterrupted sequence of measurements this makes them the most appropriate samples for short-term variability studies. Moreover they are obtained during two different time periods where the variability activity of the source differs significantly. DS1 maps part of the intermittent event occurred around MJD 51991 having a mean count rate of around 65 counts sec^{-1}PCU2^{-1} and DS2 maps the source behavior during a relatively quiescent period around MJD 52690 having a mean count rate of around 15 counts sec^{-1}PCU2^{-1}.

The method of surrogates consists mainly of two parts. Initially a linear process is specified having the same first two statistical moments (mean, variance) identical with the observed time series and therefore the same PSD. This linear process consists the null hypothesis H_0, preserving all the linear properties of the original data set and based on that all the surrogated data sets are produced. Then a discriminating statistic is calculated for both the original data set and the surrogate time series and for the latter a mean value with an error is derived. Finally H_0 is accepted or rejected based on the significance of the difference of the given statistic between the real and the artificial light curves, yielding a detection or not of nonlinearity.

---

[2] The time series should consist of more than 1000 points.

The surrogate data sets were produced by randomizing both the phase and the amplitude of the Fourier transform of the original data according to sect.3.3.6. It should be stretched again the fact that this method can produce an ensemble of non deterministic linear time series exhibiting the same power spectrum as the real data set taking correctly into account the intrinsic scatter in the powers following a χ^2 distribution with 2 DOF, χ_2^2 (eq.3.12).

The discriminating statistic should not be derived from the first and second order statistical properties of the data sets since by definition the surrogates have been created in such a way to preserve these linear properties being identical with those of the original data. The best choice is to choose a nonlinear statistic which can serve two purposes at the same time, rejecting H_0 and characterize the nature of nonlinearity (low-dimensional chaotic process as opposed to stochastic process). Therefore the value of the correlation integral $C_m(r_k)$ (eq.3.52) at a given radius r_k in the m-dimensional reconstructed phase space (sect.3.4.2) is computed which is one of the primarily quantities used in order to specify the dimension correlation dimension D_2 (sect.3.4.5) of a dynamical system.

It is important to note that in the context of the surrogate data sets $C_m(r_k)$ is not calculated in order to derive D_2 but instead to be used as a nonlinear estimator for establishing a measure of the difference between the original and the artificial data sets for a given embedding dimensions m. The reason is that $C_m(r_k)$ is derived for a given fixed value of *hypersphere*-radius, r_k, (eq.3.54) and it is not the result of an asymptotic convergence as it is dictated from eq.3.56. Therefore since the r_k is fixed for all the embedding dimensions D_2 can not be derived from eq.3.58. In case that nonlinearity is detected then the things are easier concerning the derivation of D_2 since the optimum time delay has been already chosen and the only remaining issue is the estimation of $C_m(r_k)$ at the various r_k, $k = 1, \ldots, 20$.

Application of the method to the two data sets (DS1 and DS2) requires the specification of their PSD spectral index, a, in order to produce 2000 artificial light curves ($N = 2000$) for each one of them. The PSDs are estimated after binning the logarithm of the periodogram (sect.3.3.2) taking also into account the measurements errors (sect.3.1.3) yielding $a_1 = 1.82 \pm 0.26$ and $a_2 = 2.10 \pm 0.18$. In order to reconstruct the phase space of DS1 and DS2 the optimum time delay is chosen based on the method of higher order correlations (sect.3.4.3). The simultaneous occurrence of the extrema (coincidence time, $\tau_{\rm c}$) in eq.3.44, eq.3.46 and eq.3.47 is for DS1 at $\tau_{\rm c,1}$=6 bins of 5440 sec and for DS2 at $\tau_{\rm c,2}$=10 bins of 5440 sec (fig.6.6). Based on these values the time window $\tau_w = (m-1)\tau_{\rm c}$ is defined and by applying the MOD (sect.3.4.2) the phase space of the system is reconstructed into m successive embedding dimensions ($m = 1, \ldots, 20$). Finally for each m the value of $C_m(r_k)$ is estimated for a given fixed radius $r_k = 0.5R$ (eq.3.54). For the case of DS1, $R_1 = 132/$ and for DS2, $R_2 = 18/$. In order to estimate the uncertainty of $C_m(r_k)$ coming from the errors of the data points the same computations are repeated 1000 times (sect.3.1.3) having fixed the values of $\tau_{\rm c,1}$, $\tau_{\rm c,2}$, R_1 and R_2.

The correlation integral is then estimated for the surrogates of each data set, following exactly the same procedure i.e. same $\tau_{\rm c}$ and r_k, for all the embedding dimensions $m = 1, \ldots, 20$. Then for each m a mean value $\mu_{\rm H_0}$ and a standard deviation $\sigma_{\rm H_0}$ of the correlation integral can be estimated from the ensemble of the surrogates. The top panels of fig.6.7 show the estimates of $C_m(r_k)$ for DS1 and DS2 (black points) together with the interpolated estimates of 10 randomly selected surrogates. The interpolation aims only to guide the eye since the embedding dimension m takes only integer values.

For each one of the two data sets (DS1 and DS2) there are two values for the correlation integral, one coming directly from the data set itself, $C_m(r_k)$, and one coming from the surrogates, $\mu_{\rm H_0}$. If these values different significantly then nonlinearity is present in the corresponding data set. In order to establish a measure of the significance $\mathcal{S}_m$ concerning the difference between the values of $C_m(r_k)$ and $\mu_{\rm H_0}$, at a given embedding dimension m, the following quantity is defined

$$\mathcal{S}_m = \frac{|C_m(r_k) - \mu_{\rm H_0}|}{\sigma_{\rm H_0}} \tag{6.4}$$

Under the assumption of Gaussianity, concerning the distribution of the correlation integral at a given m for the N surrogates, one can compute the errors on the significance by the standard error propagation

$$\Delta\mathcal{S}_m = \sqrt{(1 + \mathcal{S}_m^2/2)/N} \tag{6.5}$$

Note here that for the case of multiple realizations of the observed process, $N_{\rm D}$, it is possible to compare the two distributions (observed data and surrogate) directly using either the Kolmogorov-Smirnov(Press et al. 1992) or the Mann-Whitney test (Mann & Whitney 1947). From the observed process a mean value of $C_m(r_k)$ can be computed together with an uncertainty $\sigma_{\rm D}$, yielding an error in the significance

$$\Delta\mathcal{S}_m = \sqrt{(1 + \mathcal{S}_m^2/2)/N + (\sigma_{\rm D}/\sigma_{\rm H_0})^2/N_{\rm D}} \tag{6.6}$$

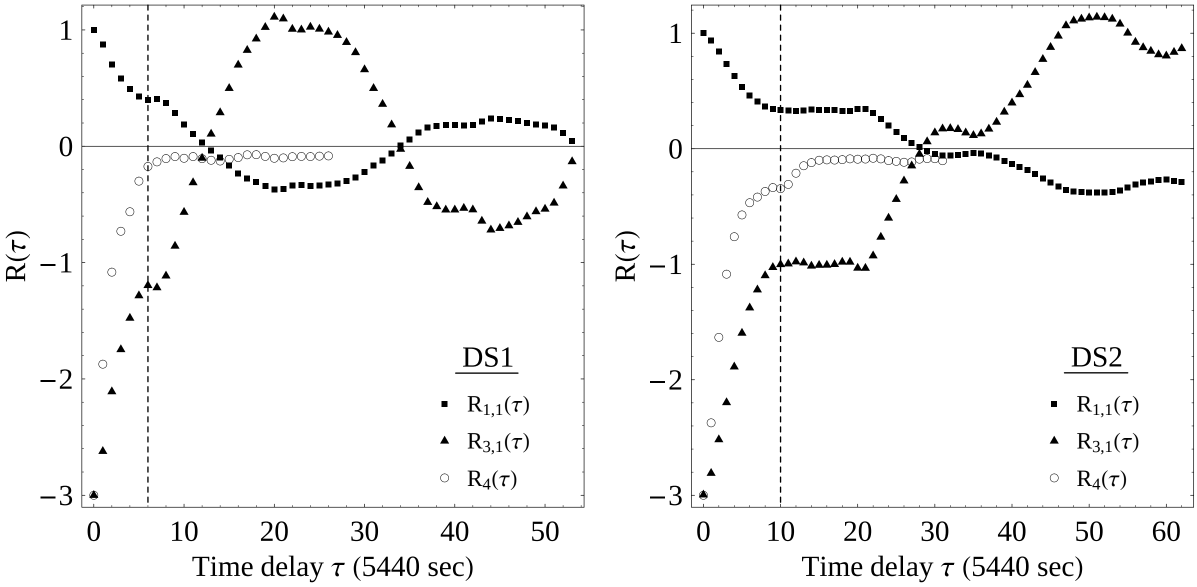


Figure 6.6: The higher order correlations $R_{3,1}(\tau)$, $R_4(\tau)$ together with the ACF $R_{1,1}(\tau)$ for DS1 (left panel) and DS2 (right panel). The vertical dashed lines indicate the simultaneous occurrence of the extrema for all the correlations. For DS1 $\tau_{\mathrm{c},1}$=6 and for DS2 $\tau_{\mathrm{c},2}$=10.

For $\sigma_{\mathrm{D}} = 0$ one gets eq.6.5.

The bottom panels of fig.6.7 shows $\Delta\mathcal{S}_m$ for the two data sets DS1 and DS2. The assumption concerning the Gaussianity of the surrogated values of the correlation integral at a given m has been checked and confirmed by forming the histogram for each m. Therefore one can estimate the probability of observing a significance $\mathcal{S}_m$ or larger (at a given m) if the null hypothesis is true based on the relation, $P_{\mathrm{H}_0}(\geq \mathcal{S}_m) = \operatorname{erf}\left(S_m/\sqrt{2}\right)$, where erf indicates the error function encountered in integrating the Gaussian distribution. There is no indication of any nonlinear behavior embedded in both the data sets since none of the estimated deviations are significant. Thus DS1 and DS2 can be considered as the output of a linear process. This of course does not exclude the possible existence of a nonlinear dynamical behavior manifesting itself in much longer time scales. The only secure statement that can be made is that the given data sets for the given time extension do not exhibit any significant sign of nonlinearity. At the same time one should note that the number of points plays a major role in the whole analysis since the more the points the better and more complete representation of the phase space can be achieved.

6.3.4 Is there any memory?

In this section the possible existence of a “memory trend” embedded in the data sets DS1 and DS2 (fig.6.5) is going to be examined via the estimation of the Hurst exponent, H (sect.3.6). Once again it should be noted that in this context the term “memory” is used in order to indicate a circular dependence of similar statistical properties without implying any periodicity. In case that such nonperiodic patterns do indeed exist in the data set then they correspond to a statistical tendency which can not be picked up from the PSD due to the nonharmonical nature of the events.

Before applying the methodology presented in sect.3.6.2 it should be ensured that the data sets are stationary (Peters 1994) in order to have a statistical meaningful values of sample mean (and variance). The point here is that despite the fact that in a time series one can always find a mean value[3], in the case that the values of the data set continuously increasing with time then the mean value is not representative of the behavior of the data set. For a nonstationary data set x_t the *backwards difference operator* $\nabla x_t = x_t - x_{t-1}$ is appropriate to make it stationary (Brockwell & Davis 2002).

After filtering the two data sets with the stationary operator the estimation of the Hurst exponent is straightforward. The two methods, presented in sect.3.6.2 which differ with respect to the normalization of the statistical range R_A for a given subset (eq.3.67), are applied to the two data sets. The first method

---

[3] In the pathological situation where the parent distribution is a Cauchy distribution then the estimation of a mean value is completely meaningless and wrong since by definition it can not be defined.

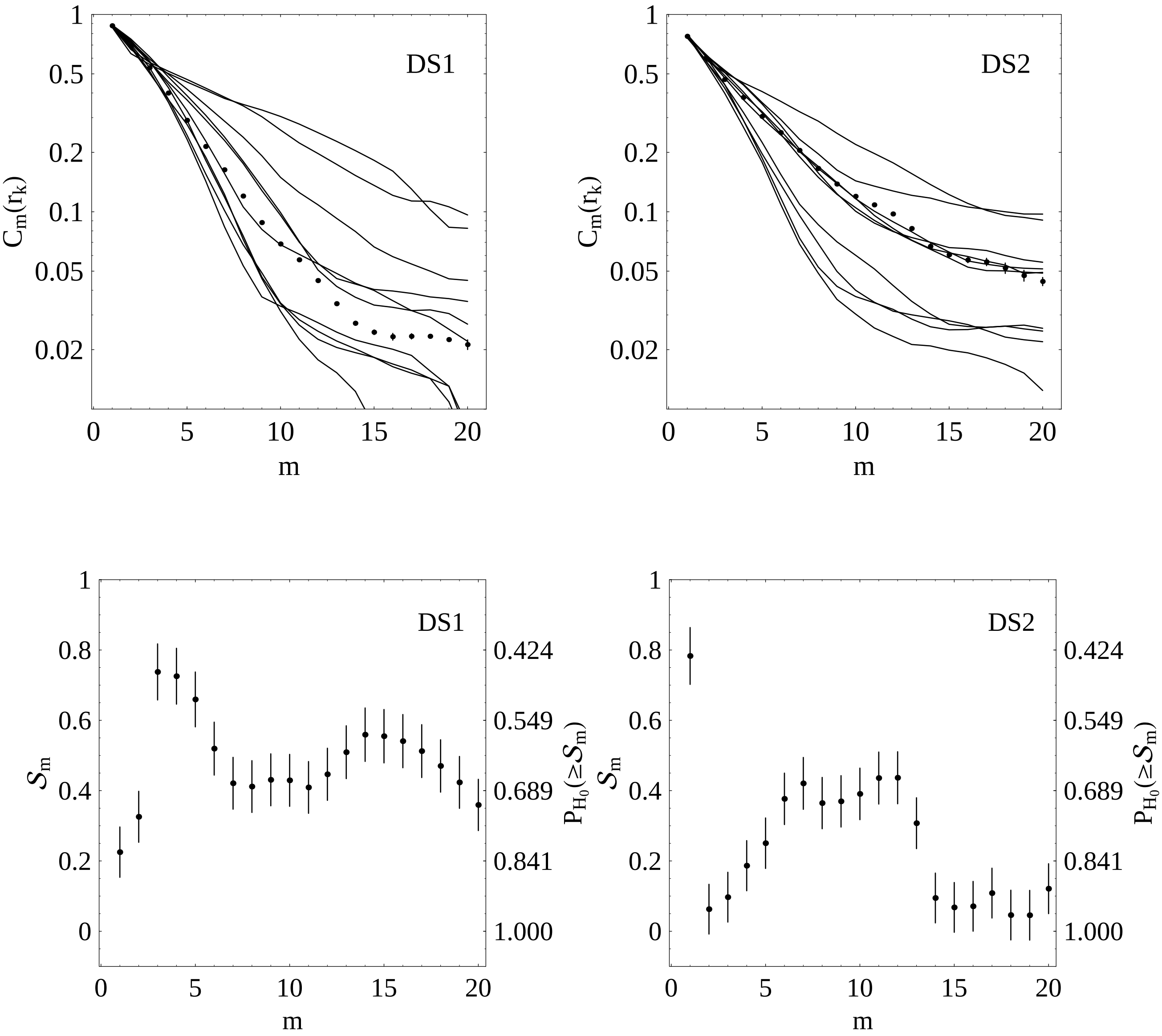


Figure 6.7: [Top panels] The values of $C_m(r_k)$ for DS1 and DS2 (black points) together with 10 randomly chosen surrogates from an ensemble of 2000. The correlation integral is estimated for $\tau_{c,1} = 6$ bins of 5440 sec and $r_k = 66$ (DS1) and for $\tau_{c,2} = 10$ bins of 5440 sec and $r_k = 9$ (DS2). The errors for the estimates of the correlation integral for the original data sets have been computed based on the procedure described in sect.3.1.3 for 1000 light curve variations.
[Bottom panels] The significance (in units of σ_{H_0}) concerning the difference between the value of the correlation integral coming from the original data set, $C_m(r_k)$, and the ensemble of the 2000 surrogates, μ_{H_0}, as it is estimated from eq.6.4. The error bars are estimated based on eq.6.5. The right axes shows the probability of observing a significance of $\mathcal{S}_m$ or larger if the null hypothesis is true.

uses for normalization the usual standard deviation, S_A (eq.3.65) and the other one the $S_{\mathrm{L}A}$ (eq.3.71). In order to take into account the measurements errors the two methods are applied to 1000 artificial light curves whose points represent Gaussian random values distributed based on the original data sets DS1 and DS2 (see sect.3.1.3). The left panels of fig.6.8 contain the results based on the classical R/S analysis (first method) (Mandelbrot 1972) and the right panels contain the results after applying the adjusted R/S method (second method) (Lo 1991).

Concerning the first method for DS1 (fig.6.8, top-left panel) there is an indication of a persistent behavior since $H = 0.602 \pm 0.008$ with a probability of having a long-term memory trend, based on the V statistic (eq.3.76) being 29.5% for $V = 1.09$. The time scale when the system's memory starts to dissipate is around 4.66 days (N_p=74 and each time bin is 5440 sec or 6.30×10^{-2} days), indicated by the arrow. In the same plot the gray points indicate the estimates coming from a random walk process (eq.3.79) in order to be compared with the results. The differences between the two populations of points are significant fact that indicates that based on this classical R/S method the data set can not be considered as simple random walk. For the case of DS2 (fig.6.8, bottom-left panel) the situation is similar since a persistent behavior with a Hurst exponent of $H = 0.640 \pm 0.009$ is favored and the memory dissipation starts around 5.29 days.

Concerning the adjusted R/S method for DS1 (fig.6.8, top-right panel) the results differ significantly from those derived with the classical method. An antipersistent behavior appears to be embedded in the data set since $H = 0.426 \pm 0.010$ but this time the measurements do not differ significantly from a purely red noise process above 1.5 days. Nevertheless the time scale when the system's memory starts to dissipate is around 4.91 days comparable with the one derived previously. Similarly in the case of DS2 (fig.6.8, bottom-right plot) an antipersistent behavior seems to be the case since $H = 0.400 \pm 0.013$ with a memory of around ~ 5.1 days.

Considering the fact that the adjusted method suffers from less biases such as short range dependence and heteroscedasticity, with respect to the classical method, one should rely on these results. The main concern that one may have before accepting the results of the adjusted method is that the time scales concerning the memory dissipation (4.91 days and 5.1 days for DS1 and DS2 respectively) are too long in comparison to the total length of the data sets (6.7 days for DS1 and 7.9 days for DS2) or in other words the data sets are too small in order to be able to unveil a 5 day nonperiodic cycle as a standard time property of Mrk 421. Peters (1994) states that the time series should be long enough in order to contain at least 8 potential cycles of finite memory periods. Based on that DS1 and DS2 are capable to reveal potential memory cycles of the order of $\sim 0.8 - 1$ days. Also Aydogan & Booth (1988) suggest that each group should consist of more than 15 number of subsets, $\lfloor N/N_A \rfloor$. For DS1 the latter relation is fulfilled for $n = 6$ and for DS2 for $n = 7$ corresponding roughly to 0.4 days. Therefore based on these authors if a memory trend was embedded in the data set it should have been of the order of half a day in order to be revealed.

It seems that data sets coming from different variability states of Mrk 421 tend to have the same time and structural "memory" properties. The fact that different methodologies alter the outcomes for both data sets in the same way raises some skepticism. This could well be attributed to the small time extent of the data sets. It is not possible, based on any statistical method, to reveal a time property embedded in a system being of the order of the data set's duration. Even data sets originating from purely periodic systems must cover in time more than half the system's period in order to be able to recover their harmonics in the PSD. Therefore bigger data sets should be employed in order to consolidate the existence of any "memory trend" in the variability process of Mrk 421.

6.4 The Long-Term Variability Behavior of Mrk 421

6.4.1 Linear or nonlinear?

The study of the long-term variability behavior of Mrk 421 is performed based on the 15-day binned ASM light curve (fig.4.3) consisting of 256 points. Based on a purely phenomenological base the "burst-like" flares superimposed to the quiescent state ($\sim$ 0.3 counts sec^{-1}) predispose for the existence of a nonlinear underlying process. Once again the method of surrogates is going to be used (sect.6.3.3) in order to check possible deviations from linearity. The crucial point here is that one has really to take into account the measurement uncertainties because they are almost twice as large as those of DS1 and DS2 (these are the PCA data sets used in sect.6.3.3). This can be readily seen from fig.4.8 where for the same exposure time $\sigma_{\mathrm{ASM}}/\sigma_{\mathrm{PCA}} \approx 2$.

Initially the parameters that should be estimated are the spectral index a of the PSD and the optimum

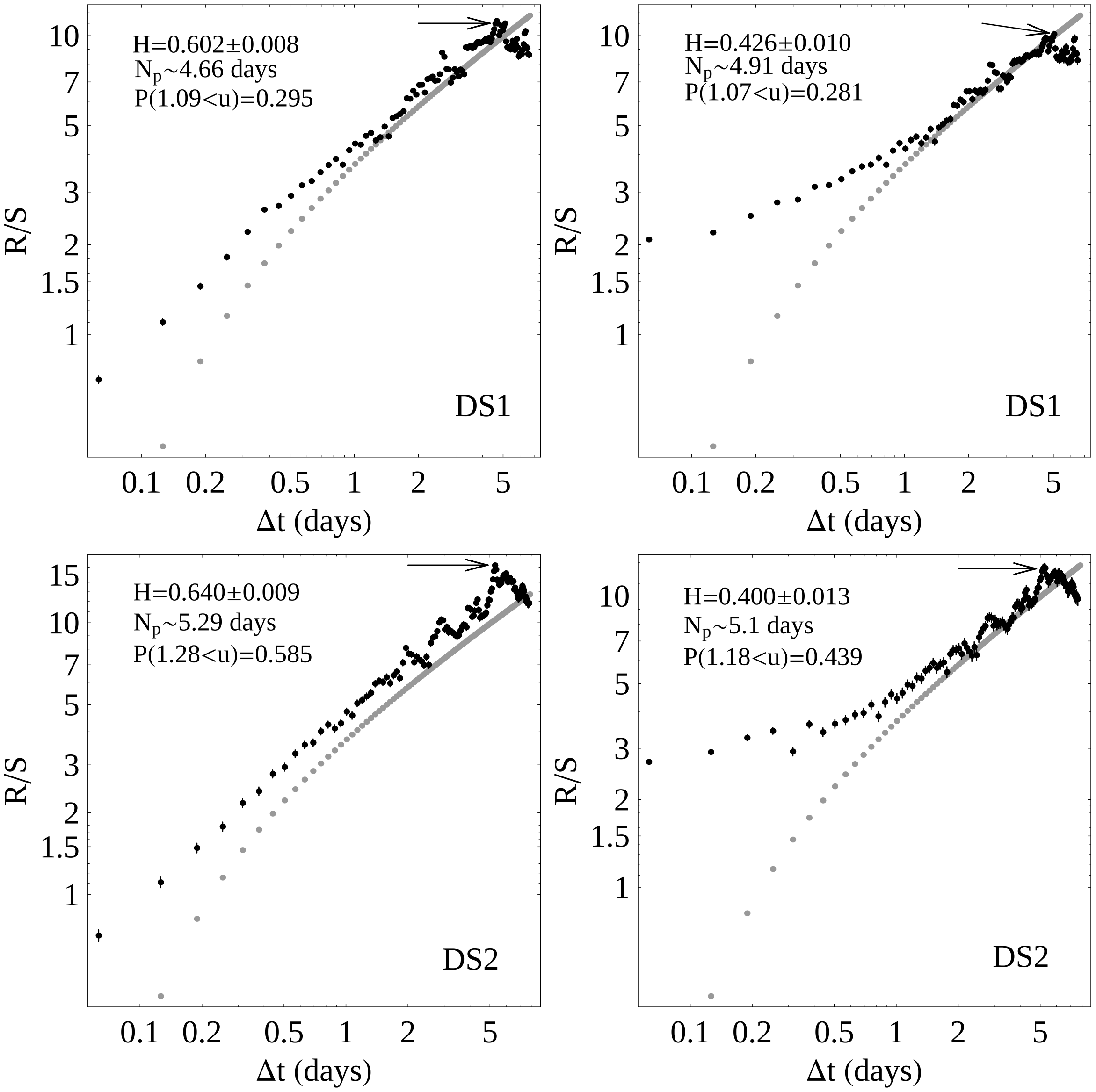


Figure 6.8: [Left panels] The (R/S) method based on the classical methodology proposed by Mandelbrot (1972). The black points represent the estimates for the data set, together with their errors. The grey points indicate the expected values of the rescaled range, $E\left((R/S)_n\right)$ coming from a random walk process as given by eq.3.79. The arrows show the onset of the plateau, defining the mean time duration when the memory of the system stars to dissipate.
[Right panels] The adjusted (R/S) method as proposed by Lo (1991).

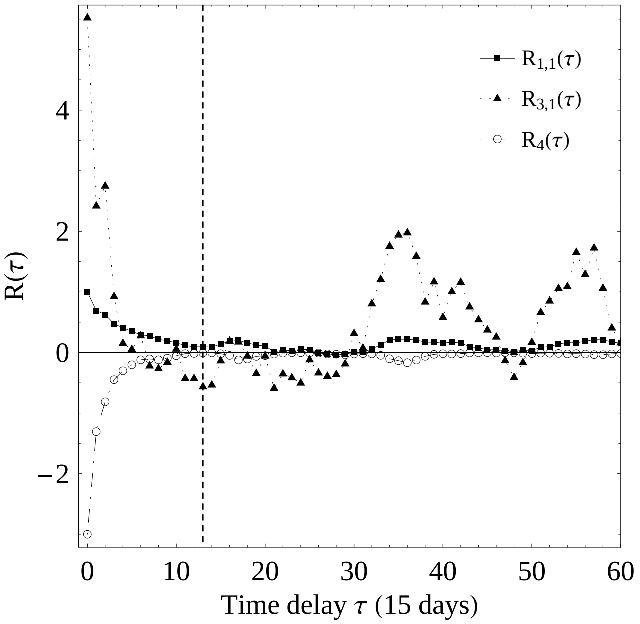


Figure 6.9: The higher order correlations $R_{3,1}(\tau)$, $R_4(\tau)$ together with the ACF $R_{1,1}(\tau)$ for the ASM data set. The vertical dashed line indicates the simultaneous occurrence of the extrema for all the correlations yielding at $\tau_{\rm c} = 13$ bins of 15 days (195 days).

time delay $\tau_{\rm c}$. The first quantity is used in order to produce the surrogate light curves and the second in order to reconstruct the phase space of the system. The binned logarithmic periodogram (sect.3.3.2) yields a spectral index of $a = 1.81 \pm 0.18$ which is then used in order to produce 2000 surrogates retaining the linear properties of the ASM data set. Based on the method of higher order correlations (sect.3.4.3) the simultaneous occurrence of the extrema is at $\tau_{\rm c}$=13 bins of 15 days (195 days) (fig.6.9). The phase space of the original data set and that of the surrogates is reconstructed into m successive embedding dimensions ($m = 1, \ldots, 18$) and for each one of them the correlation integral, $C_m(r_k)$, is estimated (fig.6.10, left panel) for a given fixed radius $r_k = 0.5R$, having a statistical range $R = 4.28$ (eq.3.54). From the right panel of fig.6.10 the detection of nonlinearity corroborates the initial phenomenological arguments. A significance of more than $5\sigma_{\rm H_0}$ for all embedding dimensions m provides a strong indication that the long-term behavior of the source should be considered that originates from a nonlinear process.

6.4.2 Dimensionality and the correlation dimension

Following the positive detection of nonlinearity, the next parameter that one has to specify is the number of the components affecting the evolution of the system. This is a fundamental step that probes into the dimensionality of the underlying process (sect.3.4.1, sect3.4.4) and tries to answer the question: Is this variable behavior a result of few physical parameters (i.e. less than 15) characterizing a low-dimensional system or it comprises the outcome of an extravagant number of parameters resulting a stochastic system (high-dimensionality)?

In order to answer this question the correlation dimension D_2 (sect.3.4.5) is going to be estimated by using the correlation integral, $C_m(r_k)$ (eq.3.52) at given embedding dimensions m ($m = 1, \ldots, 20$) for an ensemble of radii r_k (eq.3.54), $k = 1, \ldots, 20$. In order the system to have a low-dimensional behavior, meaning small number of equations/parameters that affect it, the correlation dimension as a function of m should gradually increase and at certain point it should reach a plateau indicating the dimensionality of the system, usually at $m < 15$. Contrarily, for a stochastic process one expects $C_m(r_k)$ to be continuously increasing as m increases depicting the fact that numerous parameters affect the evolution of the system therefore more dimensions (i.e. equations) are needed to describe the system.

A very crucial issue concerning the ASM data set is the existence of noise. In order to take that into account the method presented in sect.3.1.3 is followed for deriving the distribution of $D_{2,m}$ at a given m but this time for 100 artificial light curves due to computational time restrictions. The optimum time delay needed for the immersion of the ASM data set into the various embedding dimensions m, has been

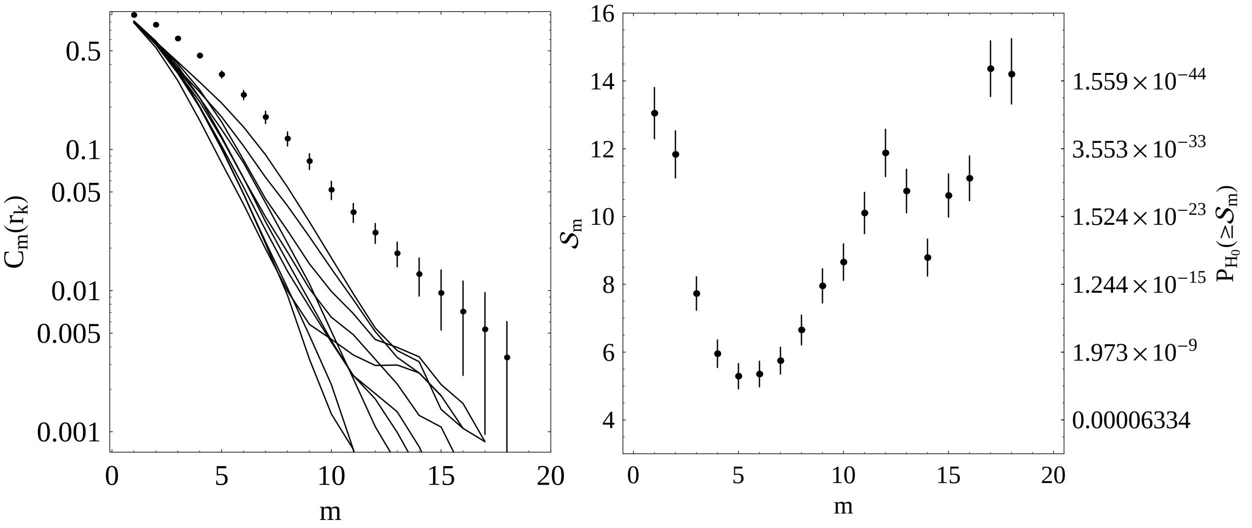


Figure 6.10: [Left panel] The values of $C_m(r_k)$ for the ASM data set (black points) together with 10 surrogates drawn from the 2000 artificial light curves. The correlation integral is estimated for $m = 1, \ldots, 18$ with τ=13 bins of 15 days and $r_k = 2.14$. The errors for the estimates of the correlation integral for the original ASM data set have been computed based on the procedure described in sect.3.1.3 for 1000 light curve variations. The various estimates of $C_m(r_k)$ coming from the 10 surrogates are shown only till $m = 17$ for plotting reasons.
[Right panel] The significance (in units of σ_{H_0}) concerning the difference between the value of the correlation integral coming from the original data set, $C_m(r_k)$, and the ensemble (2000) of the surrogates, μ_{H_0}, as it is estimated from eq.6.4. The error bars are estimated based on eq.6.5. The right axes shows the probability of observing a significance $\mathcal{S}_m$ or larger if the null hypothesis is true.

already chosen for the method of surrogates (sect.6.4.1) and it is equal to τ_{c}=13 bins of 15 days (fig.6.9). Additionally in sect.6.4.1 the statistical range has been also estimated and it is equal to $R = 4.28$.

The correlation integral as a function of the phase radius r_k for $m = 1, \ldots, 20$ together with the various slope estimates (i.e. $D_{2,m}$) coming for the successive heptads (see sect.3.4.5) are shown in fig.6.11. Note here that the plots do not contain the error bars since they are used only for visual inspection of the dynamical behavior of the system[4]. For embedding dimensions $m > 5$ one can see that the are no estimates for the $C_m(r_k)$ at small radii. This happens because the correlation integral at a given m (eq.3.52) compares actually the radius r_k with the distances between all the phase points (by constructing nominal *hyper*-spheres of radius r_k around each point) and checks how many of these points are closer than r_k (i.e. included in the *hyper*-spheres). If all the distances are grater than r_k then $C_m(r_k) = 0$. As the radius becomes bigger the probability of having more and more points inside the *hyper*-spheres is increasing and when it becomes bigger from the smallest existing distance, between all the phase points then $C_m(r_k) \neq 0$. The fact that as m increases less estimates towards the smaller radii are feasible, is a first indication of possible dynamical noise (sect.3.1.2) embedded in the data set of the ASM detector. For small embedding dimensions several “false” correlations (due to “false” *close neighbors*) are counted during the calculation of the correlation integral. As the dimension becomes bigger some of the “false” *close neighbors* disentangle and they are not anymore included in the counting procedure resulting sometimes a complete absence of correlated pairs if at the same time there are no “true” *close neighbors* among the phase points. Practically that means that the data set is comprised from an ensemble of values which are spatially (i.e. in the genuine phase space of the system) unrelated. This effect can be caused by noise component which “fills” in a way the process during its realization (i.e. light curve) with uncorrelated values.

Usually a rough idea about the dimensionality of the system can be obtained from the plot of $D_{2,m}$ versus the phase radii r_k as in the case of the Hénon system (fig.3.16, right panel). Unfortunately for the ASM data set this does not seem to be the case since there is no obvious indication of any convergence of the $D_{2,m}$ estimates at small r_k (fig.6.11, right panel). The most robust statement concerning the dimensionality of the system can be derived from the final estimates of the correlation dimension $D_{2,m}$

---

[4] The errors have been estimated and they are going to be used later in order to derive the uncertainty of $D_{2,m}$ at a given m (fig.6.12).

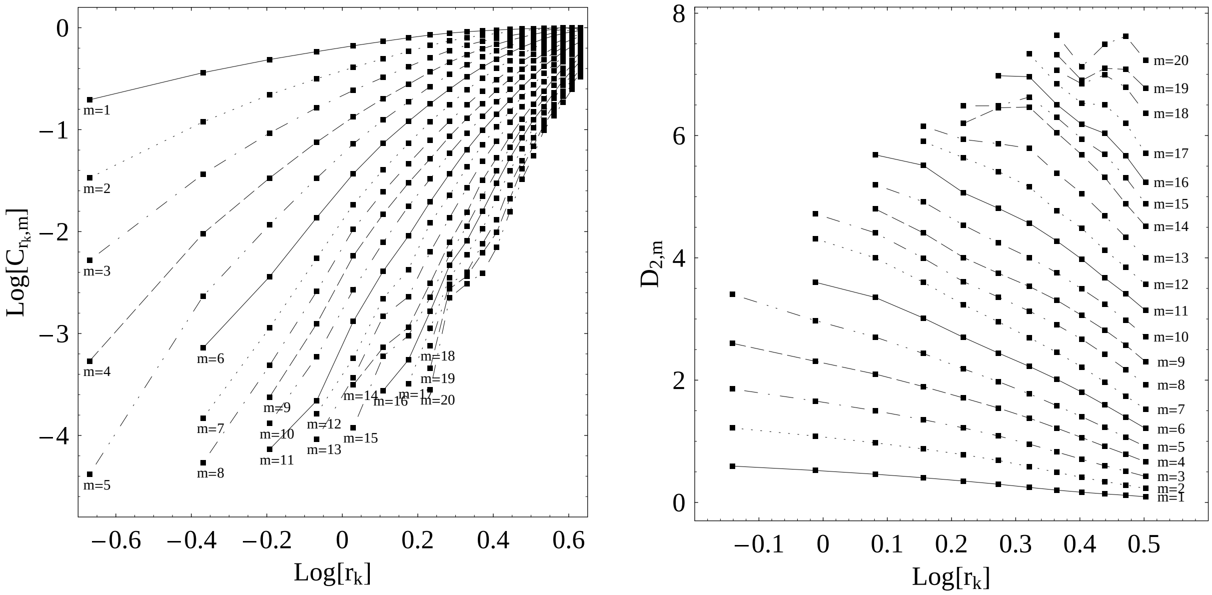


Figure 6.11: [Left panel] The logarithm of the correlation integral $C_m(r_k)$ as a function of the logarithm of the radius $r_k = k \cdot 4.28 \cdot 0.005$ ($k = 1, \ldots, 20$) for $m = 1, \ldots, 20$ (top to bottom plot). The different line-styles correspond to different m in order to be easier to find the correspondence with the plot in the right panel. Every five dimensions the same line-style is repeated.
[Right panel] The correlation dimension $D_{2,m}$ as a function of the logarithm of the radius r_k.

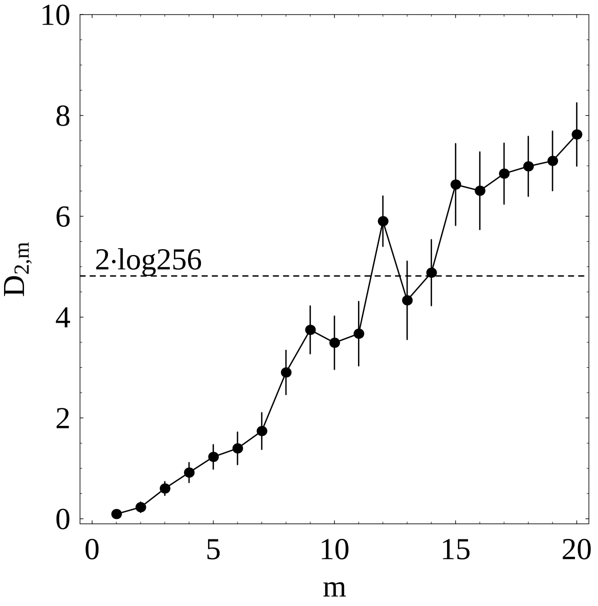


Figure 6.12: The correlation dimension $D_{2,m}$ as a function of the embedding dimension m. The dashed line depicts the limit of the meaningful estimates of $D_{2,m}$ based on the length of the data set Eckmann & Ruelle (1992).

at a given embedding dimension m (fig.6.12). The error bars include the errors due to measurements uncertainties, as they are derived from the simulations. Since the data set has a length of 256 points based on the formula provided by Eckmann & Ruelle (1992) the estimates of $D_{2,m}$ are meaningful till the value $2 \cdot \log 256 = 4.816$ corresponding to $m = 14$. Up to this dimension there is no well defined plateau fact that reveals the large number of parameters affecting the system. A constant line fit between $m = 9$ to $m = 14$ has $\chi^2 = 15.718$ for 5 DOF with a very low NHP of 0.0077. Therefore no claim of stabilization of the correlation dimension can be made concerning the ASM data set for Mrk 421 at least for $D_{2,m} < 4.816$.

Ignoring momentarily the meaningfulness of the estimates above $m = 14$ it seems that there is an apparent stabilization of the correlation dimension for $m \geq 15$. A constant line fit gives $D_{2,m} = 7.00 \pm 0.15$ having a $\chi^2 = 1.70$ for 5 DOF with a very high NHP of 0.89. If this plateau had been derived from a data set of more than $\sqrt{10^7} = 3162$ points (after the formula of Eckmann & Ruelle 1992) then that would imply that the behavior of the system is the outcome of only 7 major components. Unfortunately this is not the case but still this small indication can trigger a further investigation concerning the dimensionality of the system by employing the method of Prin.Comp.Anal. (sect.3.5). Since the length of the ASM data set affects the outcome of the correlation dimension the Prin.Comp.Anal. is supposed to overcome this problem in case that something does indeed exist above $m = 15$. Additionally this method is less influenced by the noise components being another source of confusion for the estimation of the correlation dimension as it is shown and explained in this section.

6.4.3 Dimensionality and the Prin.Comp.Anal.

In the previous section it was shown that the ASM observations of Mrk 421 compose a noisy data set whose number of measurements does not allow the derivation of any meaningful estimate of the correlation dimension above ~ 5. Since the Prin.Comp.Anal. (sect.3.5) is able to specify the dimensionality of a system[5], even in the case of small and noise data sets (Vautard et al. 1992), the method is employed in order to investigate further the number of components affecting the time evolution of Mrk 421. A short note here concerning the term "small" data set, which is sometimes completely ambiguously and vaguely understood, is that it refers to data sets which have enough measurements to allow a full statistical description of their parent distribution. For example an ensemble of 30 randomly Gaussian distributed measurements is enough to describe the statistical properties of their parent distribution by deriving the first two statistical moments. Specification of higher order statistical moments of complicated parent distributions may require lengthier data sets consisting of more than 300 points. Despite the fact that a data set might be 100% representative of the parent distribution some methods need more data points than others in order to yield a statistical meaningful estimate of a quantity. In this regard the Prin.Comp.Anal. is much more robust from the correlation integral, for specifying the dimensionality of a system, but it suffers at the same from precision lacking.

The results of the method are shown in fig.6.13 yielding a pretty clear picture about the dimensionality of the system[6]. Firstly, it seems that the complexity of the system is stochastic (as opposed to deterministic) and secondly there is a strong indication that the realization of the underlying variability process is severely affected by the noise component. The first ascertainment is attributed to the fact that there is no stabilization of the normalized singular values since their estimates are continuously decreasing without forming any plateau revealing clearly that numerous components affecting the evolution of the system. The second argument can be justified from the fact that the singular value estimates up to $i = 22$ they do not become zero. That means that a new *hyper*-axis is formed in the *hyper*-ellipsoid every time that an immersion in a higher dimension takes place. A schematic description of the 22-dimensional reconstructed phase space is the following: There are in total $L = 256 - (22 - 1) = 235$ phase points, each one requiring 22 components (i.e. numbers) in order its position to be fully specified. The lengths of the major axes of the nominal *hyper*-ellipsoid consist the singular values and the vectors defining the axes are the singular vectors. The fact that the singular values decrease as the immersion increases does not mean that the first components are more dominant than the last ones. Each estimate takes place in a different dimension therefore only the formation of a plateau is considered to be the distinctive feature depicting exactly those dimensions which do not alter the geometry of the formed attractor.

For the ASM data set of Mrk 421 there is no particular elongation towards any direction favoring any particular group of singular vectors. Usually in true physical systems plots similar to fig.6.13 show

---

[5] As shown in sect.3.5 the Prin.Comp.Anal. method provides a rougher estimation of the system's dimensionality from that of the correlation dimension, D_2, e.g. for the Hénon system $D_2 = 1.252 \pm 0.003$ and the result of the Prin.Comp.Anal. is between 2 and 3.

[6] Once again the errors represent the weight of the measurement uncertainties to the various estimates. They are estimated from 100 artificial light curves following the same methodology described in sect.3.1.3.

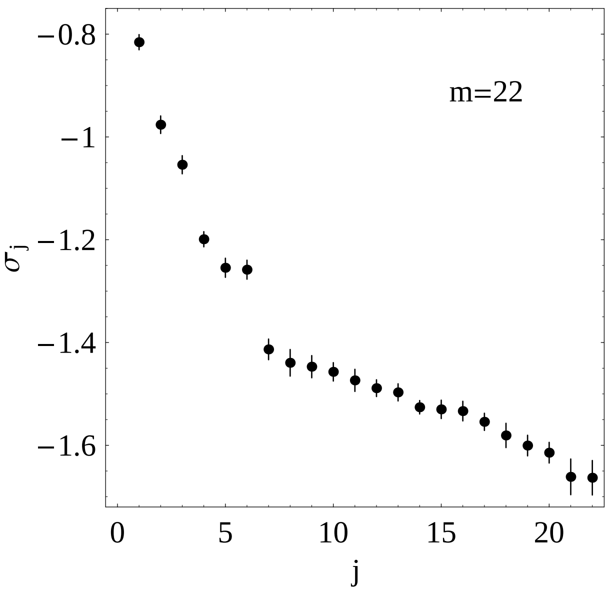


Figure 6.13: The logarithm of the normalized singular values for $m = 22$. Note the sudden transition of the estimates around the 6th and the 7th singular values.

deviations from a smooth decay. These deviations are checked by examining the form of the components of the singular vectors corresponding to the respective singular values (Broomhead et al. 1987). In the case of Mrk 421 there is a "suspicious" feature between $j = 6$ and $j = 7$. If this transition is due to a passage from a band of high significant singular vectors (i.e. leading the course of the system) to a group of low significance singular vectors then the form of the latter will differ significantly from those of $j < 7$. Theoretically the latter should have the shape of orthogonal polynomials[7] of $j - 1$ order. The plot of the four successive singular vectors (eq.3.63) situated around the transition region ($y_{5,22}, y_{6,22}, y_{7,22}, y_{8,22}$) is shown in fig.6.14. Since their form is complicated and random without any indication of orthogonality the conclusion is that the sudden transition is not a result of a different significance level between the singular values but probably is caused due to nonlinearities which are embedded in the data set. As discussed in sect.3.5 since the Prin.Comp.Anal. is based on the SVD which is a linear method therefore distortions can arise from strong nonlinear behavior such as the observed sudden burst activity.

The results coming from the Prin.Comp.Anal. are in accordance with those coming from the correlation dimension (sect.6.4.2). It is evident that there is no deterministic mechanism governing the underlying process responsible for the variability of Mrk 421. Numerous parameters are involved and affect the outcome of every realization which is observed in the X-rays. Trying to understand individual light curve features is not insightful at all since the observed photons consist the outcome of an additive process as it is seen through the jet. There is nothing characteristic which is "hidden" in the data set that can point towards any specific direction concerning some leading physical parameters of Mrk 421.

---

[7] Just as Fourier series provide a convenient method of expanding a periodic function in a series of linearly independent sinusoidal terms, orthogonal polynomials provide a natural way to solve, expand, and interpret solutions to many types of important differential equations. Orthogonal polynomials have a given order and morphologically the bigger the order of a polynomial the more complicated its representation.

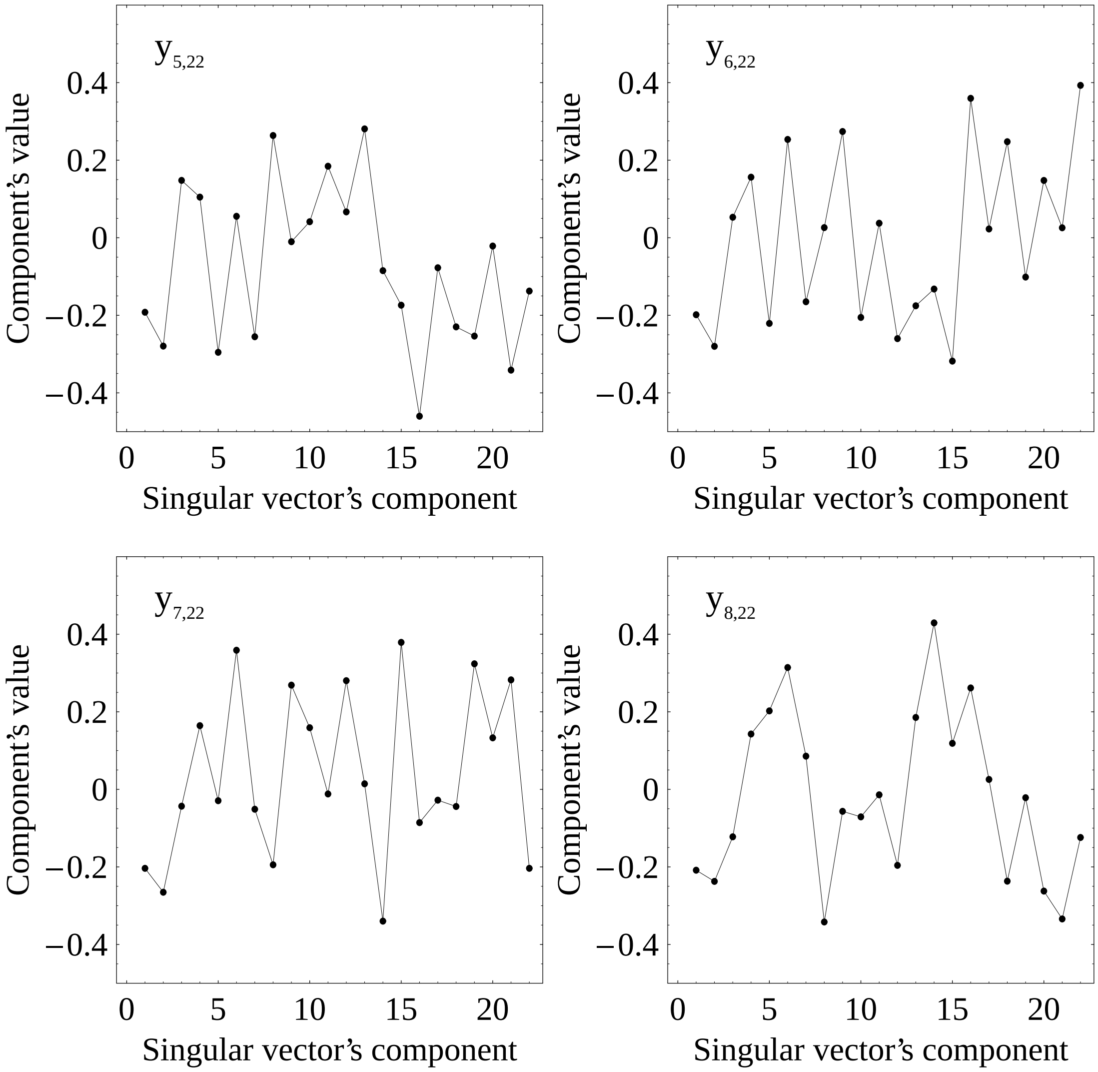


Figure 6.14: The components of the four successive singular vectors ($y_{5,22}, y_{6,22}, y_{7,22}, y_{8,22}$) situated around the transition region between $j = 6$ and $j = 7$. Their form is complicated and random without any indication of orthogonality.

Chapter 7

The H.E.S.S. MULTIWAVELENGTH CAMPAIGN OF PKS 2155-304

As discussed in sect.1.2.1 the BL Lac objects are of primary interest to VHE γ-ray astronomy since some of them are able through the IC process to scatter photons to "exotic" energies above 100 GeV. The time series analysis methods, presented in the previous chapters for the X-ray light curves, can not be easily applied to light curves originating from the VHE regime due to observational restrictions. Big uncertainties combined with poor source coverage yield time series data sets that can be treated only in a phenomenological fashion with very general and rough analysis methods (e.g. correlation analysis). Among all the BL Lac objects probably the most prominent candidate for conducting variability studies is the HBL object (see sect.1.2.1) PKS 2155-304. This southern hemisphere source is regularly observed by a new generation Čerenkov array H.E.S.S., being located in Namibia. The sensitivity levels of H.E.S.S. in combination with the convenient zenith angle of the source, brings PKS 2155-304 in the first place of the most extensively observed and well-studied VHE sources. Since the emission of the BL Lac objects covers a vast energy range of the electromagnetic spectrum (more than 15 decades in frequency), copious multifrequency campaigns, usually following the observing plan of the VHE observatories, are conducted covering simultaneously the X-ray, the optical and sometimes the radio energy bands. The results of a completely new and unpublished multifrequency campaign conducted during the period of August–September 2004, concerning PKS 2155-304 are presented in this chapter.

7.1 Previous Observations of PKS 2155-304

The HBL object PKS 2155-304 (z=0.117) was firstly discovered in the X-rays by the HEAO-1 (see, sect.2.2.3) satellite (Schwartz et al. 1979; Griffiths et al. 1979). Nowadays it is firmly established as one of the brightest extragalactic X-ray sources in the sky since it can be detected on a regular basis e.g. ROSAT (Brinkmann et al. 1994), *Beppo satellite per astronomia X* (BeppoSAX) (Chiappetti & Torroni 1997), RXTE (Vestrand & Sreekumar 1999) and Chandra (Nicastro et al. 2002). The first γ-ray emission in the energy range of 30 MeV–10 GeV was detected from PKS 2155-304 by the *energetic gamma ray experiment telescope* (EGRET) detector onboard *Compton gamma ray observatory* (CGRO) mission (Vestrand et al. 1995).

The first evidence of VHE emission from the source was reported by Chadwick et al. (1999) for energies above 300 GeV using the University of Durham Mark 6 atmospheric Čerenkov telescope, with a significance of ~ 6 standard deviations above the background (σ). Since then a lot of observations were conducted from other Čerenkov telescopes without yielding any significant detection. The CANGAROO experiment tried to detect the source during 1999 (Nishijima et al. 2001), 2000 and 2001 (Nakase & Cangaroo Collaboration 2003) yielding only upper limits. The first robust detection of PKS 2155-304 was made by H.E.S.S. at a 45σ significance level, corresponding to 169 half an hour runs between 2002 and 2003 (Aharonian et al.

2005b). From then on, during all the H.E.S.S. observations, the source is detected on a permanent basis. At the same time simultaneous multifrequency observations covering the energy bands of the X-ray, the optical and the radio regime are performed aiming to map the overall behavior of the source across the whole electromagnetic spectrum. The first campaign as such was carried out during October–November 2003 (Aharonian et al. 2005c) and one more was conducted during August–September 2004 the results of which are going to be discussed here.

7.1.1 Variability properties in VHE

In the first claimed detection of the source Chadwick et al. (1999) there was an indication of monthly variations concerning the flux of the source (5 monthly averaged points with the biggest deviation between two points being 4 standard deviations) but no variations on daily time scales was observed.

During the first robust detection of the source (Aharonian et al. 2005b) the VHE flux was found to be variable in time scales of months, nights and even hours but without any evidence of spectral variability with respect to time.

Finally the first multifrequency campaign of PKS 2155-304 (Aharonian et al. 2005c) was interesting in the sense that no correlations could be established between any of the observed wave bands (optical, X-rays, γ-rays).

7.2 The Observing Campaign during August–September 2004

7.2.1 The VHE observations

The H.E.S.S. detector

The H.E.S.S. experiment (Hinton 2004) consists of four atmospheric Čerenkov telescopes operating stereoscopically. Each telescope consists of a tesselated 13 m diameter (107 m^2 surface area) mirror, focusing the Čerenkov light from the showers of secondary particles, created by the interaction of γ-rays in the atmosphere onto a camera in the focal plane. Each camera consists of 960 photomultipliers with a pixel size of 0.16°, providing with a FOV of 5°. The array is located in Khomas highlands in Namibia ($-23^\circ\ 16'$, $16^\circ\ 30'$, 1835 m AMSL). The angular resolution of the stereo system is better than 0.1° per event. The energy threshold of H.E.S.S. is about 100 GeV (at zenith) with spectral measurements possible above ~ 150 GeV with an energy resolution of 15%. For the data set of PKS 2155-304 the time averaged spectrum has an energy threshold of 228 GeV.

The VHE data

The H.E.S.S. observations start basically with a small observing period between 14–20 of July 2004 and they continue to the main observing period between 6–26 August 2004 and 6–11 September 2004. The ensemble of the observations consist of 105 observing runs each one lasting on average 28 min (fig.7.1, left panel).

The data were taken in the *Wobble* mode where the source direction is positioned $\pm 0.5^\circ$ in declination relative to the centre of the FOV of the camera during observations. This allows for both on-source observations and simultaneous estimation of the background induced by charged cosmic rays. The data reported here are selected and analyzed with the "standard analysis" (described in sect.4 of Aharonian et al. 2005b). Concerning the background a *ring* of radius of $\sim 0.5^\circ$ around a trial source position is used in order to provide the background estimate (fig.7.1, right panel). This yields a total of 12122 on-source events and 40934 off-source events with an on-off normalization of 0.090 corresponding to a significance of $\sim 102\sigma$ (eq.17 in Li & Ma 1983).

The photon index Γ derived from the time-averaged spectrum (see below the power-law fit) is used as a fixed parameter to estimate the integral flux above 200 GeV for each run. This integrated flux takes into account the effective area and threshold variations due to the source movement through the sky. The overall light curve (fig.7.2, top panel) is inconsistent with a constant flux since a constant fit has a $\chi^2 = 582.50$ for 104 DOF yielding a NHP=0. On a run-by-run basis within a night there is no strong evidence for significant variations. From the whole data set only four nights are slightly deviate from a constant flux MJD 53225, MJD 53229, MJD 53232 and MJD 53233 (fig.7.3, top panels).

The methods used here for the energy reconstruction of each event and for determining the overall spectrum are described in sect.6 of Aharonian et al. (2005b). The measured time-average spectrum is fitted by three models (fig.7.4):

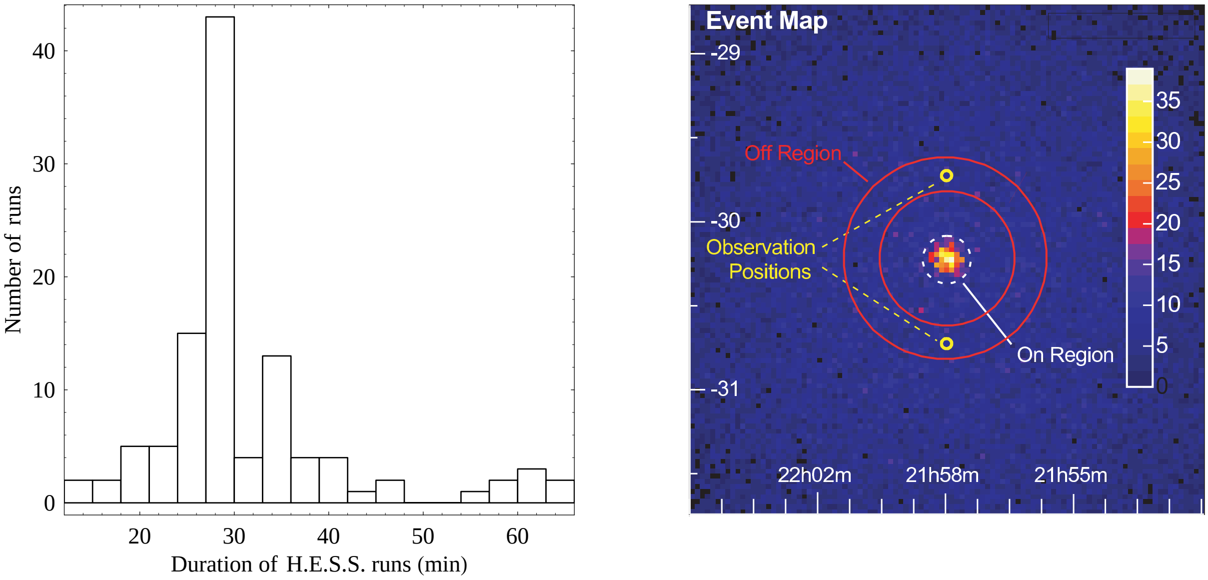


Figure 7.1: [Left panel] Histogram of the duration of the H.E.S.S. runs.
[Right panel] Schematic representation of the *ring*-model on an event map of 5 hours of H.E.S.S. observations of PKS 2155-304 during 2002–2003 (Aharonian et al. 2005b) (taken from Berge et al. 2006).

- The first model is a power-law of the form

$$\frac{dN}{dE} = I_0 \left(\frac{E}{E_0}\right)^{-\Gamma} \tag{7.1}$$

where I_0 is the flux normalization in 1 TeV and Γ is the photon index. The fit yields $\Gamma = -3.56 \pm 0.06$ and $I_0 = (2.22 \pm 0.13) \times 10^{-12}$ cm^{-2}sec^{-1}TeV^{-1}, having a χ^2 fit of 27.47 for 10 DOF yielding a NHP of 0.002. This photon index is used in order to estimate the integral flux above 200 GeV on a run basis as shown in the top light curve of fig.7.2.

- The second model is a broken power-law of the form

$$\frac{dN}{dE} = \begin{cases} I_0 \left(\frac{E}{E_0}\right)^{-\Gamma_1}, & E \geq E_{\text{break}} \\ I_0 \left(\frac{E_{\text{break}}}{E_0}\right)^{\Gamma_2-\Gamma_1} \left(\frac{E}{E_0}\right)^{-\Gamma_2}, & E < E_{\text{break}} \end{cases} \tag{7.2}$$

where E_{break} is the energy corresponding to the spectral break in units of TeV and Γ_1, Γ_2 are the spectral indices below and above the energy break. The fitting procedure is performed in a recursive way by fixing the value of E_{break} within the range of 0.2 TeV to 4.7 TeV in steps of 0.01 TeV. For every step the NHP of the fit is estimated and the one having the biggest value is then selected. The optimum NHP is 0.82 ($\chi^2 = 4.36$, 9 DOF) for $E_{\text{break}} = 0.39$ TeV yielding $I_0 = (4.47 \pm 0.46) \times 10^{-12}$ cm^{-2}sec^{-1}TeV^{-1}, $\Gamma_1 = 3.00 \pm 0.09$ and $\Gamma_2 = 3.79 \pm 0.04$.

- The third model is a power-law with exponential cut-off of the form

$$\frac{dN}{dE} = I_0 \left(\frac{E}{E_0}\right)^{-\Gamma} e^{-\frac{E}{E_{\text{cut}}}} \tag{7.3}$$

where E_{cut} is the cut-off energy in TeV. The fit yields $I_0 = (4.57 \pm 1.26) \times 10^{-12}$ cm^{-2}sec^{-1}TeV^{-1}, $\Gamma = -3.14 \pm 0.16$ and $E_{\text{cut}} = 1.39 \pm 0.54$ TeV, having a χ^2 fit of 15.29 for 9 DOF yielding a NHP of 0.08.

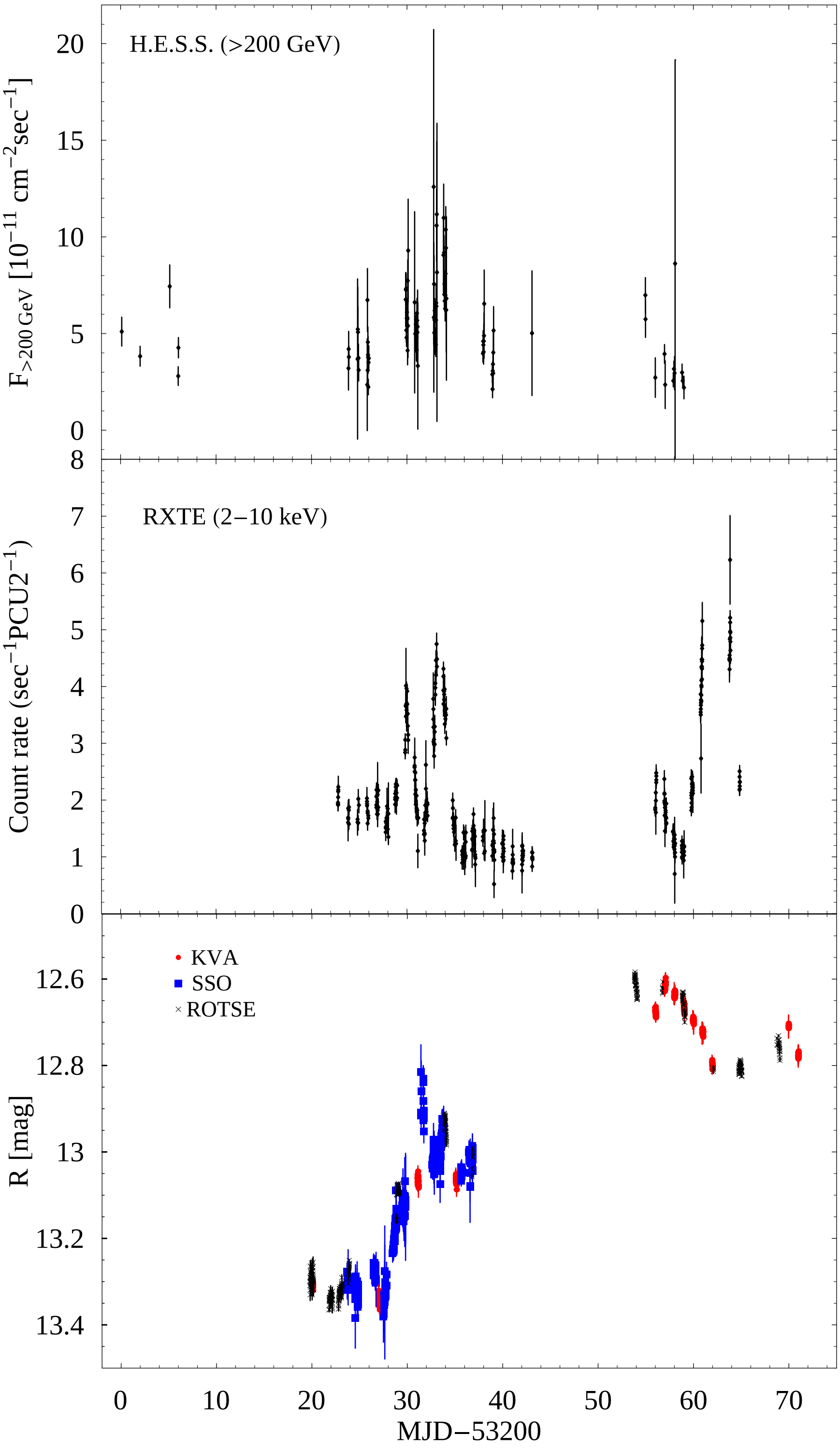


Figure 7.2: The light curves of PKS 2155-304 during the multiwavelength campaign of 2004. [Top panel] The H.E.S.S. light curves above 200 GeV binned in run lengths, each one lasting ∼ 28 min. [Middle panel] The RXTE light curve, coming from PCU2, between 2–10 keV with a bin size of 512 sec. [Bottom panel] The R-band magnitude registered from the optical telescopes of KVA, SSO and ROTSE.

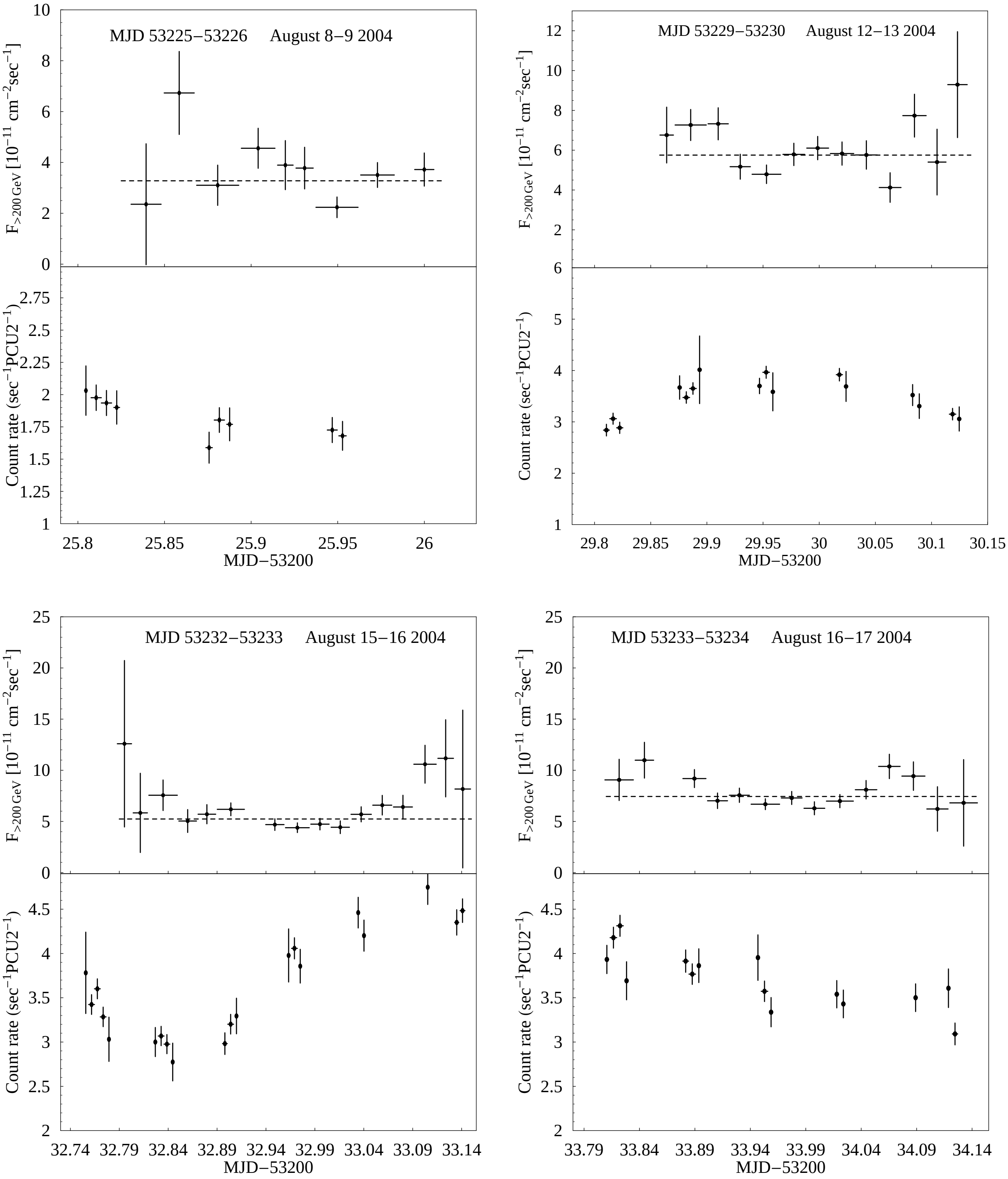


Figure 7.3: Intranight variations in the VHE and X-ray band. The top panels show the flux variations in the VHE regime above 200 GeV and the bottom ones show the X-ray variations in the 2–10 keV energy range. For these four nights in the VHE regime the χ^2 chance probability to a constant flux fit (indicated by the dashed line) gives for MJD 53225, MJD 53229, MJD 53232 and MJD 53233: 0.052, 0.023, 0.030 and 0.032 respectively.

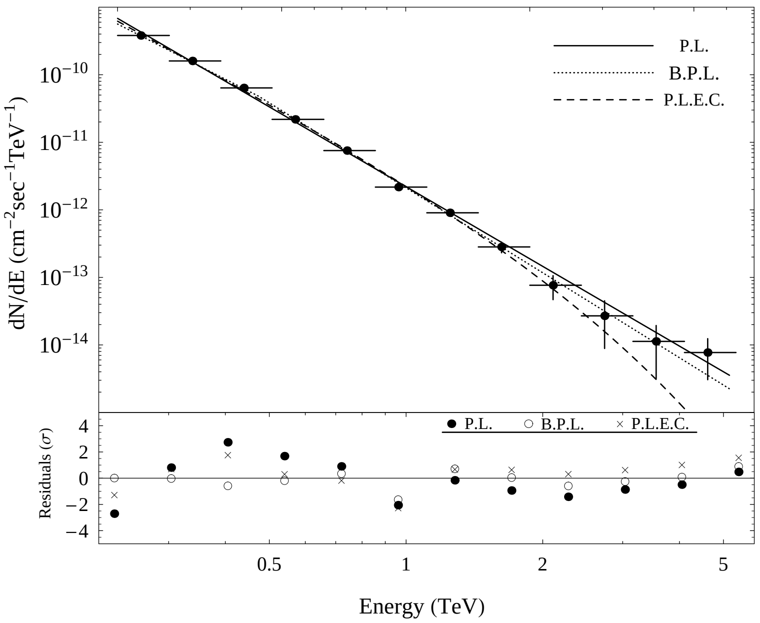


Figure 7.4: [Top panel] The VHE time-averaged spectrum of PKS 2155-304 as registered by H.E.S.S. The various lines represent the best fits to a power-law (P.L.), a broken power-law (B.P.L.) and a power-law with exponential cut-off (P.L.E.C.).
[Bottom panel] The residuals (measurements−model) of the various fits in units of (σ) i.e. the statistical error of each measurement.

7.2.2 The X-ray observations

The RXTE satellite

The X-ray data were obtained by the RXTE satellite (sect.2.1) during the periods of 5–26 August and 8–16 September. From all the onboard satellite experiments only PCA (sect.2.1.2) data are used for the analysis since it is the most sensitive instrument for AGN studies in comparison to HEXTE (Cui 2004). During the campaign only PCU0 and PCU2 were in fully operational mode among the five PCUs due to contingent breakdowns (see sect.4.1.1). Moreover since PCU0 has no longer the front antico propane layer (sect.2.2.5 and fig.2.3) it is much more susceptible in background events induced by particles therefore only the PCU2 data were used. In fact similar “fake” flaring events as the ones reported by Xue & Cui (2005) (fig.2.8) are also present in this data set coming from the PCU0 (fig.7.5).

The X-ray data

The total duration of the PCA observations for these two months is 152.352 ksec distributed in $\sim$ 3 thirty-minute runs within each night. Concerning the observations of the ASM detector (sect.2.1.4), obtained during this period, they are checked but they do not seem to be very interesting since they are marginally consistent with being constant (fig.7.6).

The X-ray light curve from the whole campaign is derived following the analysis described in detail in sect.2.2. Concerning the background estimation (sect.2.2.3) the “faint” background model file was used since the source count rate is less than 40 counts sec^{-1}PCU2^{-1}. The light curve is shown in the middle panel of fig.7.2 in bins of 512 sec. The very good time coverage between the VHE and the X-ray observations can be also seen in fig.7.3.

The overall X-ray spectrum is estimated based on the procedure described in sect.2.2.4. It is fitted by a power-law with absorption having a $N_{\rm H\,I}$ of 1.69$\times$10^{20}cm^{-2}. This yields a flux in the 2–10 keV energy range of $F_{2-10{\rm keV}} = (2.80 \pm 0.01) \times 10^{-11}$erg cm^{-2} sec^{-1}. The same spectral model with the same $N_{\rm H\,I}$ for the months of August (92 ksec) and September (60.5 ksec) yields $F_{2-10{\rm keV}}$ =(2.52$\pm$0.02)$\times$10^{-11} erg cm^{-2} sec^{-1} and $F_{2-10{\rm keV}}$ =(3.21$\pm$0.02)$\times$10^{-11}erg cm^{-2} sec^{-1} respectively. The results of the spectral fits are shown in tab.7.1 and the overall spectrum in fig.7.7.

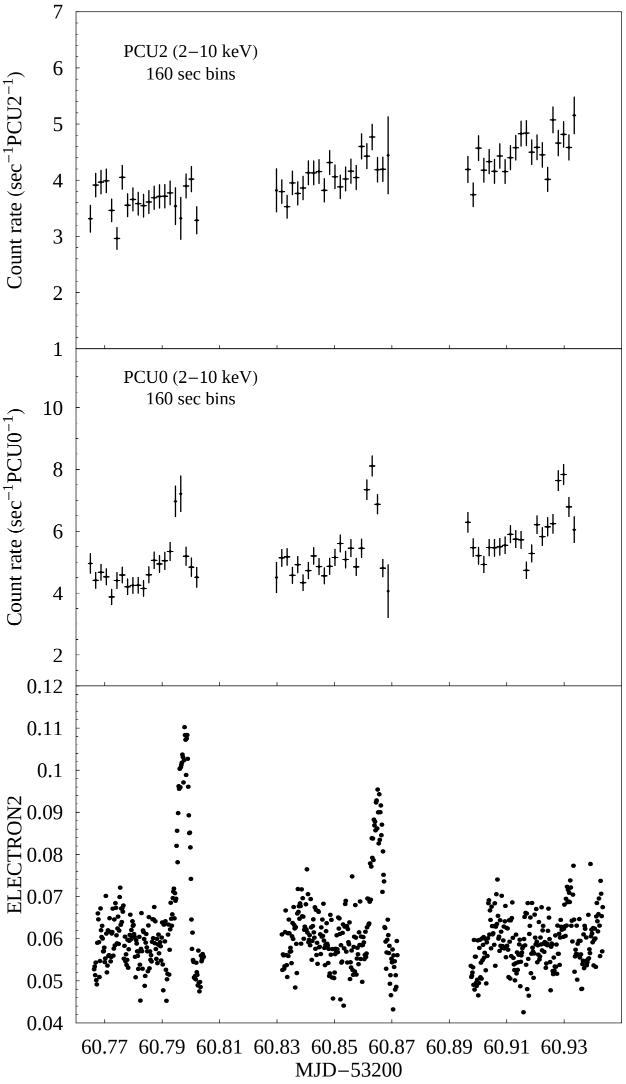


Figure 7.5: [Top panel] The count rate of PCU2 during 3 orbits on MJD 53260. Correctly the detector does not respond to the background events shown in the bottom panel.
[Middle panel] During the same observing period PCU0 registers three flares corresponding to the background events shown in the bottom panel.
[Bottom panel] The background events during the observed period indicated by the ELECTRON2 parameter (sect.2.2.2).

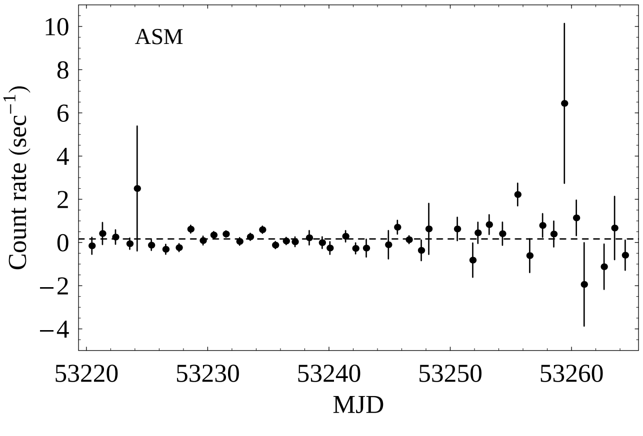


Figure 7.6: The ASM light curve for PKS 2155-304 for the period 5 August–16 September in daily bins. A constant fit, shown by the dashed line, yields y=0.164±0.055 having a χ^2 of 76.49 for 43 DOF and a NHP of 0.001.

Table 7.1: Spectral fits: power-law with $N_{\rm H\,I}$=1.69×10^{20}cm^{-2}

Month	Spectral index	Norm. ×10^{-2} *	χ^2 (for 38 DOF)
August	3.06±0.04	$4.29^{+0.24}_{-0.23}$	49.8
September	3.03±0.04	$5.31^{+0.30}_{-0.28}$	43.6
Total	3.05± 0.02	$4.72^{+0.15}_{-0.14}$	117.56

*in units of photons keV^{-1}cm^{-2}sec^{-1} at 1 keV.

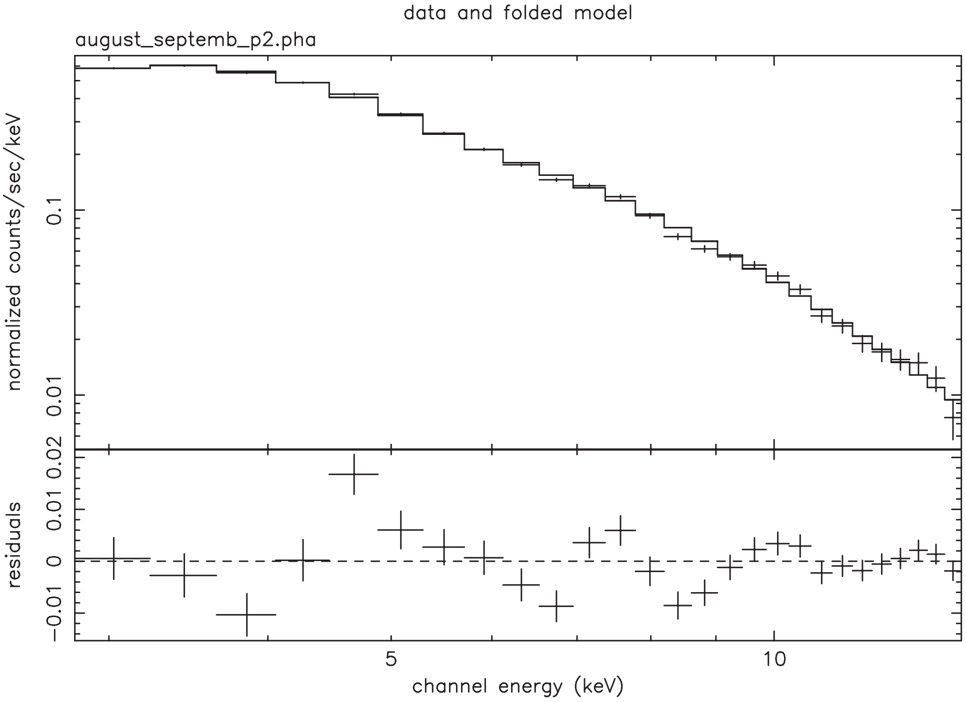


Figure 7.7: [Top panel] The X-ray spectrum derived from the summed August and September 2004 data using PCU2, fitted by a power-law model having a fixed absorption N$_{\rm H}$=1.69×10^{20}cm^{-2} (folded model). [Bottom panel] The residuals between the data and the model.

A broken power-law fit with the same absorption of $N_{\rm H\,I}$=1.69×10^{20}cm^{-2} does not seem to ameliorate significantly the goodness of the fit. For the August data set it yields an integrated flux in the 2–10 keV band of $(2.44^{+0.02}_{-0.12})\times10^{-11}$erg cm^{-2}sec^{-1}. For the September data set the fitting procedure can not trace a minimum in the χ^2 space in order to derive the errors for the fitted parameters. Therefore the break-energy, $E_{\rm br}$, is fixed to the resulted August fit value yielding an integrated flux of $(3.16^{+0.04}_{-0.06})\times10^{-11}$erg cm^{-2}sec^{-1}. The total flux for both months (after fixing $E_{\rm br}$ to the August's value) is $F_{2-10\,{\rm keV}}$ = (2.72±0.02)×10^{-11}erg cm^{-2}sec^{-1}. The results of the spectral fits are shown in tab.7.2.

Table 7.2: Spectral fits: broken power-law with $N_{\rm H\,I}$=1.69×10^{20}cm^{-2}

Month	Spectral index, Γ_1	Spectral index, Γ_2	$E_{\rm br}$ *	Norm. ×10^{-2} †	χ^2/DOF
August	$2.91^{+0.11}_{-0.16}$	$3.16^{+0.09}_{-0.07}$	$4.71^{+0.90}_{-0.70}$	$3.50^{+0.55}_{-0.66}$	42.2/36
September	2.95±0.01	3.09±0.07	4.71 (fixed)	$4.71^{+0.67}_{-0.60}$	41.38/37
Total	2.93±0.06	3.13±0.04	4.71 (fixed)	$3.98^{+0.32}_{-0.30}$	101.92/37

*in units of keV.
†in units of photons keV^{-1}cm^{-2}sec^{-1} at 1 keV.

7.2.3 The optical observations

The optical telescopes

Optical photometric observations of PKS 2155-304 were performed during August–September 2004 with three telescopes from three different observatories. The difference among the longitudes of the telescopes involved allowed to reduce the gaps in the data trains due to bad weather and daylight. The telescopes are:

- the *Kungliga Vetenskapsakademien telescope* (KVA telescope), at the Observatorio del Roque de Los Muchachos, located in La Palma, Spain ($-16^\circ\,26'\,60.00''$, $28^\circ\,46'\,0.12''$, 2330 m AMSL). The telescope is operated remotely through internet connection by the staff of the Tuorla observatory in Finland. It is composed of a 60-cm f/15 Cassegrain reflector and it is equipped with a CCD polarimeter devoted to polarimetry only. Moreover it has a 35-cm f/11 SCT auxiliary telescope equipped with BVRI filter wheel and a CCD which are devoted to multi-band photometry. The latter was used for the observations of PKS 2155-304 being the southernmost observations ever performed with the KVA since the telescope's declination lower limit is $\delta \sim -30^\circ$. A total of 439 R-band photometric data points were obtained during 13 nights between August 3 and September 23. Due to its low declination, the target could be followed for less than 4 hours per night and the number of frames per night varied from 9 to 50, with an average of $\sim$ 38. During 4 nights in September (Sep. 8–9, 9–10, 12 and 22–23), simultaneous polarimetric observations were also performed (Ciprini et al. 2006).
- the 40-inch (101.6 cm) telescope of the ANU at *Siding Spring observatory* (SSO) near Coonabarabran in Australia ($-31^\circ\,16'\,24.24''$, $149^\circ\,3'\,45.00''$, 1135 m AMSL). The telescope has an f/8 setup and it was used for BVRI multi-band photometry using a single-CCD detector. PKS 2155-304 was observed from August 6 to August 19 for 11 nights yielding a total of 1188 photometric measurements (B-band: 157, V-band:128, R-band: 727, I-band:176) varying from 38 to 160 frames per night with average number 85. Due to the optimum observatory's geographical latitude the source could be followed for $\sim$ 10 hours per night depending on the weather conditions.
- the *robotic optical transient search experiment* III (ROTSE-III) telescope (Akerlof et al. 2003), located at SSO near Coonabarabran, New South Wales, Australia ($-31^\circ\,16'\,24.24''$, $149^\circ\,3'\,39.96''$, 1110 m AMSL). The ROTSE-III array is a worldwide network of four robotic automated telescopes built for fast ($\approx$6sec) response to γ-ray burst (GRB) triggers from satellites such as Swift. Each telescope consists of a mirror of 45-cm f/1.9 and has wide FOV ($1.85^\circ \times 1.85^\circ$) imaged on a Marconi 2048×2048 pixel back-illuminated thinned CCD. It is operated without filters in order to collect as much as light as possible from GRBs afterglows. A total of 424 photometric data points were obtained during 13 nights between August 2 and September 21.

Data analysis and calibration of the optical data

The KVA and the SSO data are analyzed with aperture photometry on de-biased and flat-fielded frames. The former are analyzed adopting an aperture of 7.5 arcsec, whereas the latter are analyzed adopting an aperture of 5.4 arcsec. Instrumental magnitudes are then extracted, night by night, for the PKS 2155-304 and for the reference stars nos. 2, 3, 4, and 5, whose finding chart and absolute calibration can be found in Hamuy & Maza (1989). The calibration of the instrumental magnitudes of the BL Lac object is then performed by using as calibrators the reference stars nos. 2, 3 and 4. Those stars are suitable for calibration, because their brightness were comparable with that of PKS 2155-304 in the period considered, and they are relatively constant with time. For instance, in the R-band: $R_2 = 11.67 \pm 0.01$;$R_3 = 12.47 \pm 0.02$;$R_4 = 13.42 \pm 0.02$, and $R_{\mathrm{PKS}\ 2155-304} = 12.6 - 13.4$. Moreover the average rms scatter of the mag differences between each of the calibration stars (2, 3 and 4) and all the others is variable in the range (0.011–0.014) mag for the complete (KVA and SSO) data set. The calibration is performed by simply adding to the BL Lac's magnitude a zero-point magnitude m_0 is computed, frame by frame, in the following way:

$$m_0(i) = \sum_i \frac{[m_{\mathrm{stand}}(i) - m_{\mathrm{instr}}(i)]/\sigma^2_{\Delta m}(i)}{\sigma^2_{\Delta m}(i)} \tag{7.4}$$

with the index i indicating the reference stars chosen for the calibration, and the quantity $\sigma^2_{\Delta m}(i)$ being the variance of the difference $\Delta m(i) = m_{\mathrm{stand}}(i) - m_{\mathrm{instr}}(i)$ between the absolute calibration and the instrumental magnitude of the i-th star: $\sigma^2_{\Delta m}(i) = \sigma^2_{m_{\mathrm{stand}}} + \sigma^2_{m_{\mathrm{instr}}}$.

The error on the calibrated magnitudes is computed using the stars 3 and 4, having magnitudes closer to the BL Lac's magnitude, as follows. The larger between the two following quantities:

1. the rms scatter of the mag difference $\Delta m_{3,4} = m_3 - m_4$.
2. the absolute value of the deviation of the mag difference $\Delta m_{3,4}$ from its mean value over the data set is taken, and then combined, frame by frame, with the instrumental errors on stars nos. 3 and 4, and with the instrumental error on the PKS 2155-304.

This is a quite conservative estimate of the error, because it takes into account both the instrumental errors and the stars' scatter. Finally before combining the calibrated KVA and SSO light curves a thorough check concerning possible systematic offsets is performed. These offsets may arise from the use of different photometric routines and different apertures. To this purpose, the two photometric routines are both run on a selected set of SSO frames, and then calibrated. The two final light curves of PKS 2155-304 are consistent with each other within the errors. However, a systematic offset $\Delta R_{\mathrm{KVA-SSO}} = 0.026 \pm 0.023$ mag exists between the two light curves. This offset is therefore applied to the KVA R-band light curve when assembling the total (KVA+SSO) light curve. Since the ROTSE observations are obtained without any filter, the data are simply scaled to much the (KVA+SSO) light curve by a multiplicative factor 0.99.

7.3 Multiwavelength Correlation Analysis

Among the various energy bands it is usual to check for correlated variability behavior through correlation analysis. From the first multifrequency campaign of PKS 2155-304 (Aharonian et al. 2005c) the interesting result was that no correlations could be established between any of the observed wave bands i.e. optical, X-rays and γ-rays. Since this campaign provides the lengthiest data set of PKS 2155-304, obtained simultaneously in the aforementioned energy bands, these observations can be used in order to check the correspondence of the flux variations across the electromagnetic spectrum.

For the correlation analysis between the γ-rays and the X-rays the following procedure is followed. All the H.E.S.S. runs are separated in 1 min slices (on average 30 per run, fig.7.1, left panel) each one defining a time window. Then the 16 sec X-ray observations falling within these slices are binned together. Finally, the adjacent bins are merged and for the corresponding γ-ray fluxes and X-ray count rates the mean value together with the uncertainty is derived yielding 61 strictly simultaneous pairs lasting on average 14 min (fig.7.8, left panel). The correlation plot between these 61 values is shown in (fig.7.8, right panel) indicating with the open circles four VHE flux estimates having an error bigger than a factor of 4 from all the other VHE measurements. The linear correlation coefficient, after ignoring these values[1], is $r = 0.48 \pm 0.07$ yielding a probability of obtaining a value like this from an uncorrelated population $P(0.48, 55) = 0.00015$. A linear fit of the form $y = ax + b$, after taking into account the measurement uncertainties in both axes (Fasano & Vio 1988), yields $a = 0.135 \pm 0.014$ and $b = 0.126 \pm 0.027$.

Concerning the correlation analysis between the γ-ray fluxes and the optical data, all the optical observations obtained within a H.E.S.S. run are averaged and their errors are derived. A total of 32 paired values is formed exhibiting no correlation since $r = 0.29 \pm 0.09$ and $P(0.29, 30) = 0.1$ (fig.7.9, left panel). In the same setting, the correlation analysis between the X-rays and the optical band, the latter is averaged based on the 512 sec binning scheme of the former, yields an ensemble of 21 paired values having a linear correlation coefficient $r = -0.29 \pm 0.07$ (fig.7.9, right panel).

---

[1] Taking into consideration these four values the linear correlation coefficient r is equal to 0.27 ± 0.11 with $P(0.27, 59) = 0.033$.

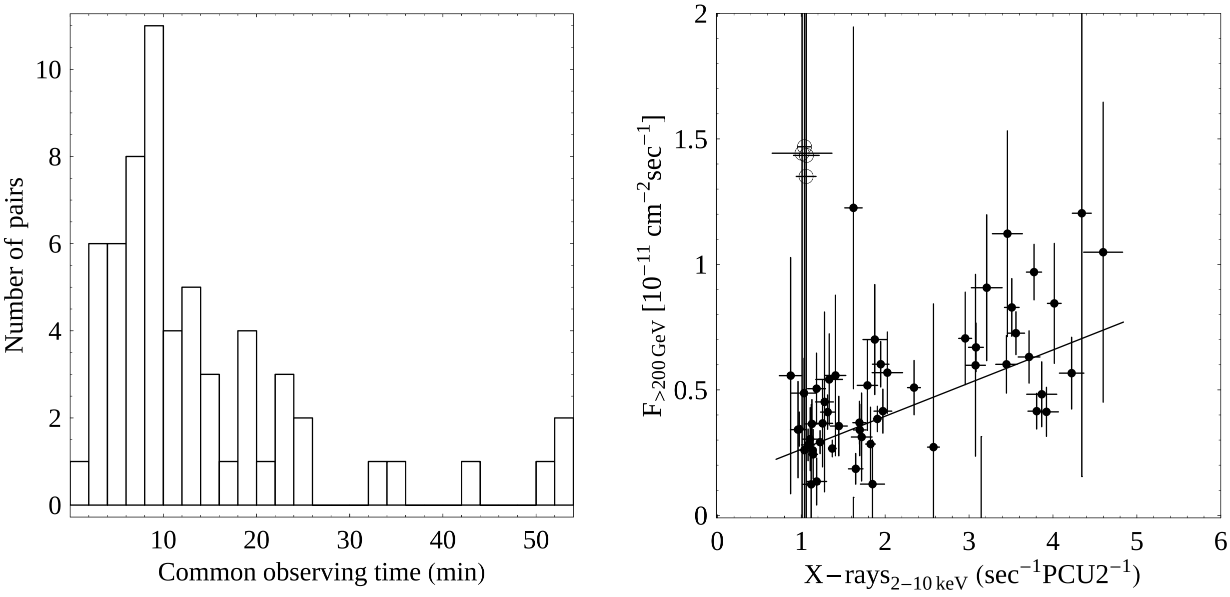


Figure 7.8: [Left panel] The distribution of the common observing time for the 61 points.
[Right panel] The correlation plot between the VHE band and the X-rays consisting of the most strictly simultaneous observations between the two bands (61 points). The open circles indicate the H.E.S.S. observations having errors bigger than a factor of 4 from the other VHE measurements. The correlation coefficient after ignoring the latter values is $r = 0.48 \pm 0.07$. The solid line represents a linear fit of the form $y = ax + b$, taking into account the error of both measurements, yielding $a = 0.135 \pm 0.014$ and $b = 0.126 \pm 0.027$.

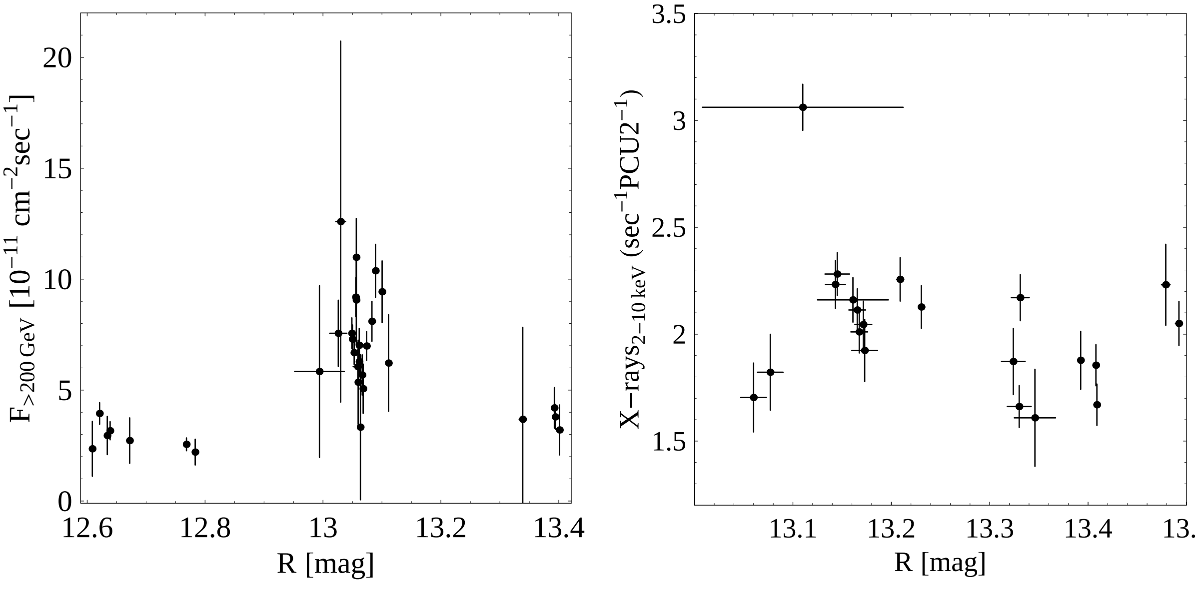


Figure 7.9: [Left panel] The correlation plot between the VHE band and the optical band. The correlation coefficient for the 32 values is $r = 0.29 \pm 0.09$.
[Right panel] The correlation plot between the X-rays and the optical band. The correlation coefficient for the 21 values is $r = -0.29 \pm 0.07$.

Chapter 8

DISCUSSION AND CONCLUSIONS

8.1 Physical Implications

Through the investigation of the timing properties of the BL Lac object Mrk 421 several points concerning both the source itself and the analysis methods were elucidated. One of the main outcomes of this work is that the examination of individual light curve features for the derivation of source parameters seems to be away from the physical reality. The short-term light curves (i.e. lasting around 10–20 days) are purely stochastic in nature meaning that the parameters, determining the shape of the various features, is supernumerary and all of them contribute with the same weight to the observed outcome. Since the emission of the BL Lac objects is seen through their jet, the flaring events comprise the summation of the radiation originating from different source regions within the jet excluding any kind of causal behavior.

8.1.1 The linear source size

Almost all the emission models, trying to explain the radiation properties of the BL Lac objects, consider a quantity which is known as *linear source size.* This automatically poses a question concerning which is the source that one is referring to. The usual method followed for the derivation of the linear source size relies on the object's light curve by estimating the time needed for the flux to be increased by a factor of 2 (doubling time). Following the definition of Harris et al. (2006) (used for the case of the HST-1 knot in M 87) and taking into account both the relativistic beaming (sect.1.2.2) and the redshift of the source z, the doubling time in the observer's frame is given by

$$DT = \left(\frac{\delta}{1+z}\right)\frac{1}{y-1}\Delta t \tag{8.1}$$

where δ is the Doppler factor (eq.1.1), y is the fractional increase of the flux from a value I_1 to a value I_2, $y = I_2/I_1$ and Δt is the time duration of the event. This definition comes from Burbidge et al. (1974) and can be found in several forms in the literature taking into account the measurements uncertainties based on some point selection criteria (e.g. Zhang et al. 1999; Fossati et al. 2000). The size of the emitting region can be derived based on the causality argument (see footnote sect.5.1) determining an upper limit for its linear dimension: $R < cDT$. That means simply that for a given time Δt the larger the amplitude of the transition (i.e. the larger the y) the smaller the DT, yielding a more compact source (i.e. smaller R). The fundamental assumption behind the causality argument is that the radiation should originate from the same emitting source something which is not the case for the BL Lac objects whose emission originates from numerous jet regions exhibiting no dominant leading component for the short-term variations.

Ignoring for the moment the relativistic effects, consider the following simplified scenario. The emission originating from three spatially different jet regions of a BL Lac object reaches the observer almost at the

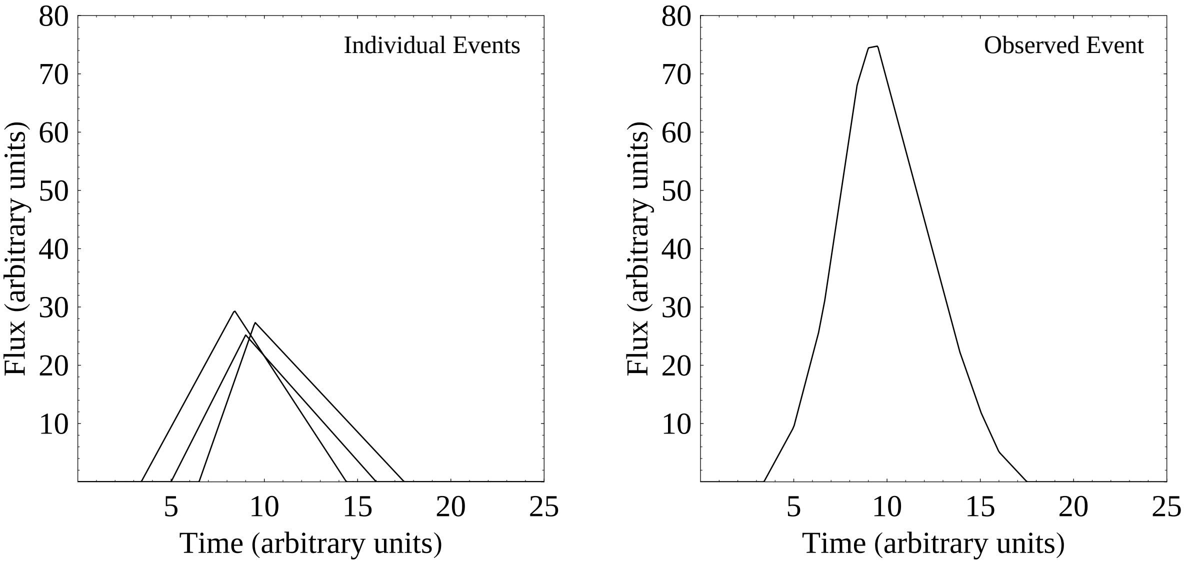


Figure 8.1: [Left panel] Three flaring events of triangular shape reaching the observer approximately at the same time. The events have approximately the same time duration and for each one of them the light-crossing time argument applies: $R_{real} < cDT_{\mathrm{real}}$.
[Right panel] The observer sees the three events merged into one flaring event having almost the same duration and a flux equal to the sum of the three fluxes. In this case the light crossing argument is meaningless.

same time, appearing as one flaring event to the observer (fig.8.1). These regions are homogeneous (i.e. they do not have clumps) share the same radiative properties (i.e. they radiate through the same radiation mechanism) and have the same size. It is natural to consider that the causality argument can be applied to EACH event individually giving an upper limit for the size of EACH source, which is almost the same for this case. The doubling times for these flares is of the order of $DT_{\mathrm{real}} = 0.3$ time units corresponding to a linear size of $R_{real} < cDT_{\mathrm{real}}$. The observed flaring event, coming from the superposition of the three, has almost the same Δt (≈ 6) with the individual flares but now due to the summation of the events, the amplitude is more than a factor of 3 bigger than the average amplitude of individual flares, yielding a $DT_{\mathrm{obs}} = 0.08$ time units. Estimating a linear dimension corresponding to the value of DT_{obs} is equivalent with merging the emitting power of the three regions into one region having a linear dimension R_{real}. The underlying process, appearing as one component to the observer, it actually consists of three components which are spatially separated and their time evolution is unrelated. Complicating the phenomenology of the aforementioned scenario favors towards accepting even larger source sizes. More components distributed around the same time instant will increase dramatically the y parameter. Similarly different source sizes will add up to the flux keeping Δt approximately equal to the duration of the biggest event (i.e. bigger R).

8.1.2 The brightness temperature

Obviously, trying to extract information concerning the physical conditions prevailing within the BL Lac objects, by light curve examination, can act quiet misleadingly. In very compact sources characterized by very high photon densities, catastrophic cooling via SSC emission (sect.1.2.3) can produce high energy radiation much higher than the observed (Hoyle et al. 1966). The onset of this catastrophic radiation losses (inverse Compton catastrophe) limits the maximum observed brightness temperature of the source[1] to the limit of $T_{\mathrm{B}} \lesssim 10^{12}$ K (Kellermann & Pauliny-Toth 1969). There are numerous claims in the literature for violation of this limit from several *intraday variable* (IDV) sources (Qian et al. 1991; Quirrenbach et al. 1992; Kedziora-Chudczer et al. 1997) with the most famous one that of the BL Lac object S5 0716+714 (Quirrenbach et al. 1991; Wagner et al. 1996; Ostorero et al. 2006; Agudo et al. 2006). To avoid this exceedance exceptional high Doppler factors (eq.1.1) ($\delta > 10$ or even $\delta \sim 100$) are usually invoked in order to lower the intrinsic brightness of the source below the theoretical limit.

---

[1] This condition yields from the requirement that the photon energy density in the emission region should be smaller than the energy density of the magnetic field.

The brightness temperature, $T_{\rm B}$, for a compact stationary source of an angular size θ having an observed flux density F_ν at a frequency ν is given by

$$T_{\rm B} = \frac{F_\nu c^2}{2k\nu^2\theta^2} \tag{8.2}$$

where k is the Boltzmann's constant. If the source is at a distance $d_{\rm L}$ (luminosity distance) from the observer then $\theta \sim R/d_{\rm L}$ with R being its linear size which is connected with the doubling time (eq.8.1) as $R < cDT$. Thus the last equation now reads

$$T_{\rm B} = \frac{F_\nu d_{\rm L}^2}{2k\nu^2 DT^2} \tag{8.3}$$

The most important determinants in this relation are the flux density F_ν and the quantity DT. For the case of S5 0716+714 the doubling time are considered to be of the order of $\sim$ 1 day at the GHz radio band, yielding from eq.8.3 brightness temperatures of the order of 10^{14}–10^{19} K. In order to lower down these values to the limit of the 10^{12} K, the relativistic transformation of eq.8.3 is considered yielding in the observers' frame a brightness temperature of $T_{\rm B} = \delta^3 T_{\rm B}{}'$ with $T_{\rm B}{}'$ being the intrinsic brightness temperature of the source having as a maximum value 10^{12} K. This yields unrealistic Doppler factors[2] that sometimes can reach values up to 100 (Qian et al. 1991; Wagner et al. 1996).

Estimation of the brightness temperature from the flares of the BL Lac objects is expected to have ultra high values based on the summation effect of the flux originating from the various jet regions. Consider the scenario of 100 jet regions producing equal number of triangular shots (eq.5.2) reaching the observer at a given time period. The shots have all their physical parameters random (i.e. occurrence time, rise time, decay time, intensity) within some fixed limit values. The brightness temperature of each region individually can be estimated from its shot being: $T_{\rm B} \propto \lambda$ where $\lambda = F_\nu/DT^2$ (eq.8.3). For this example the biggest value of λ is $\lambda_{\rm Indi.Max} \approx 572$ ($F_\nu = 57.2$, $\Delta T = 5.5$ and $DT = 0.1$), corresponding to the biggest value of brightness temperature $T_{\rm B,Indi.Max}$ ($< 10^{12}$ K) existing within the jet. $T_{\rm B,Indi.Max}$ corresponds to the most compact and most active jet region where the causality argument is valid. Estimating λ from the observed light curve gives a value of $\lambda_{\rm obs} = 2.03 \times 10^8$ ($F_\nu = 1825$, $\Delta T = 5.2$ and $DT = 0.003$) yielding a brightness temperature of $T_{\rm B,Flare} \approx 36000 \cdot T_{\rm B,Indi.Max}$ being 4 orders of magnitude bigger than the maximum existing value in the jet. The value of $T_{\rm B,Flare}$ can well be above the catastrophic limit of $< 10^{12}$ K since it is not related to any physical temperature within the jet.

Therefore the observed light curve appears to violate the catastrophic limit, something which is not the case since the observed flare is not the result of a coherent event. It is completely misleading to extract any information from it concerning the values of $T_{\rm B,Flare}$ and DT since this is not the realization of a single process happening within one jet region. The underlying emission process consists of several components whose variability activity is merged through the jet axes appearing as one component. Of course in more realistic light curves these daily variations (similar to the one shown in the right panel of fig.8.1) are usually superimposed on much larger flaring events. Even in the case that the former were originating from a single jet region the estimated brightness temperature would be still wrong since F_ν in eq.8.3 does not correspond to the event's flux but to the overall flux (event plus large flaring event).

8.2 The Value of the Long-Term BL Lac Light Curves

Statistical analysis of big data sets, covering extended time periods of the order of years, can be really insightful on revealing interesting source properties. In the case of Mrk 421 (chap.6) it was shown that around the period of MJD 51991 the source exhibited a completely different variability behavior from the usual one (fig.6.3). This is denoted by the significant different value of the mean excess variance indicating that the physical parameters of the underlying variability process during this period were greatly modified. Going to the ASM light curve (fig.4.4) it can be seen that this time period corresponds to an exceptional high source activity lasting around a year. Despite the fact that the remaining three ASM bursts (indicated by the dotted lines in fig.4.4) do not have the same good coverage from the PCA instrument, the intermittent behavior of the whole PCA light curve (fig.6.4) discloses that this burst-like activity of Mrk 421 is consistent with that of an intermittent system. Additionally the detection of nonlinearity in the long-look ASM light curve (fig.6.10) further supports the idea of intermittency, since this kind of behavior is favored by

---

[2] Based on *Very Long Baseline Interferometry* (VLBI) observations the maximum observed values of the Doppler factor are between 20 and 30 (Bach et al. 2005).

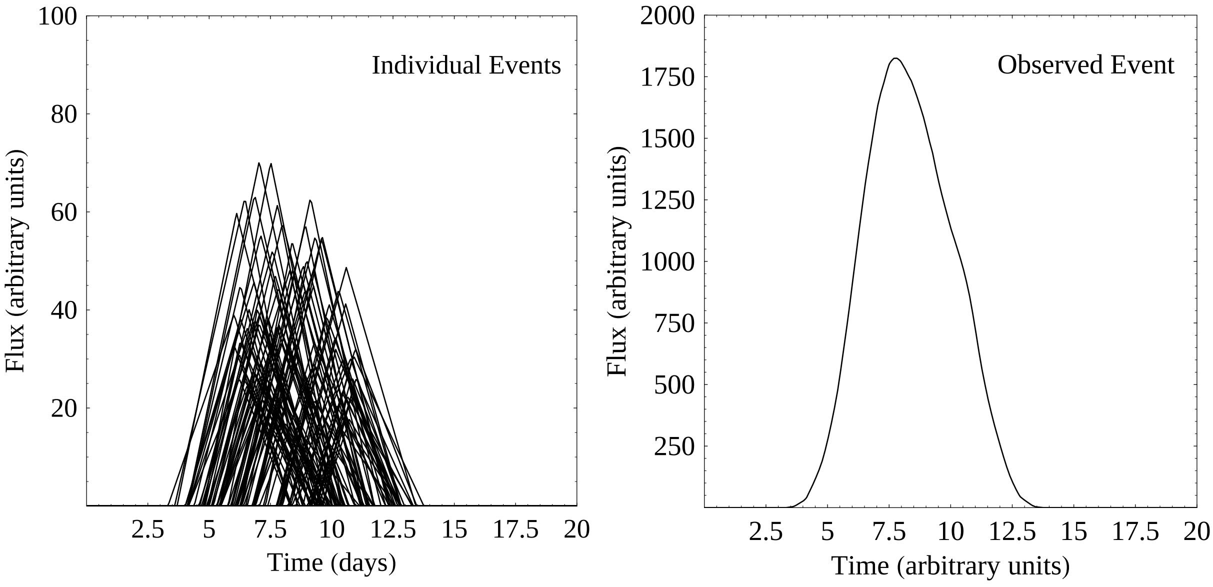


Figure 8.2: [Left panel] One hundred flaring events of triangular shape reaching the observer approximately at the same time. The physical parameters of the events (defined by eq.5.2) are uniformly distributed around some fixed values. Each shot originates from a different jet region and within it the brightness temperature has a given value smaller than 10^{12} K. The biggest value from these shots is the one corresponding to $\lambda_{\mathrm{Indi.Max}} \approx 572$.
[Right panel] The observed light curve yielding from the superposition of the individual events has a $\lambda_{\mathrm{obs}} = 2.03 \times 10^{8}$ being four orders of magnitude bigger than the existing maximum value within the jet.

nonlinear systems. Finally, through an attempt of characterizing the dimensionality of Mrk 421 (through the estimation of the correlation dimension and through the Prin.Comp.Anal.), the speculation of dominant source components leading the underlying variability process can be readily rejected strengthening at the same time the idea that the noise component plays a major role in the time evolution of the system.

Throughout the years the physical sciences have been developing based on well defined evolutionary scheme concerning the study of the various physical phenomena. Initially a general model is conceived trying to interpret the physical behavior of a given phenomenon based on some general laws. Then by successively adding more components to the given model one can create a better and more precise representation of the under study phenomenon which probes deeper into its properties. Take for example the field of thermodynamics, in early 1800s the first thermodynamical laws were created explaining the relation between the macroscopic properties of a system e.g. how the temperature of a system is affected by the pressure changes under a certain volume (classical thermodynamics). Following that the connection between the macroscopic and the microscopic properties came through the statistical thermodynamics. Through this approach the gas consists of moving particles that interact among them (including quantum phenomena) and hence one can describe the relation between the macroscopic properties in more detail as well as what is happening in smaller length scales and smaller time scales.

Nowadays the study of the BL Lac variability phenomenon is done following a completely opposite way. People try to derive from small data sets (lasting some days) source parameters knowing that these objects exhibit variations of the order of years. One has firstly to understand and interpret the overall behavior and then try to explain what is happening in smaller time scales. Trying to understand the physics of the system based on a single daily light curve is similar to studying the ideal gas law based on the evolution equation of one particle.

One of the most standard concluding phrases in the field of BL Lac astronomy dealing with variability is the following: “more observations are needed in order to elucidate the variability properties of the BL Lac objects”. Maybe it would be more insightful firstly to analyze the already existing data archives, consisting of thousands of measurements, in a homogeneous way. Sometimes these archives cover periods of more than 50 years mapping in a very representative way the long-term behavior of the sources. Of course this does not exclude variations in time scales of centuries but this is the best that one can do up to now. Only through the study long-term light curves one might find some deterministic signatures connected with some real physical source parameters.

List of Figures

List of Tables

Bibliography

Abramowitz, M. & Stegun, I. A. 1970, Handbook of Mathematical Functions (New York: Dover Publications, INC., |c1964, 9nth print.)

Agudo, I., Krichbaum, T. P., Ungerechts, H., et al. 2006, A&A, 456, 117

Aharonian, F., Akhperjanian, A., Beilicke, M., et al. 2003, A&A, 410, 813

Aharonian, F., Akhperjanian, A. G., Aye, K.-M., et al. 2005a, A&A, 437, 95

Aharonian, F., Akhperjanian, A. G., Aye, K.-M., et al. 2005b, A&A, 430, 865

Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., et al. 2005c, A&A, 442, 895

Akerlof, C. W., Kehoe, R. L., McKay, T. A., et al. 2003, PASP, 115, 132

Albano, A. M., Passamante, A., & Farrell, M. E. 1991, Physica D: Nonlinear Phenomena, 54, 85

Anis, A. A. & Lloyd, E. Y. 1976, Biometrika, 63, 111

Arnaud, K. A. 1996, in ASP Conf. Ser. 101: Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes, 17–+

Aydogan, K. & Booth, G. G. 1988, Southern Economic Journal, 55, 141

Bach, U., Krichbaum, T. P., Ros, E., et al. 2005, A&A, 433, 815

Barbieri, C., Vio, R., Cappellaro, E., & Turatto, M. 1990, ApJ, 359, 63

Bautista, M. & Kallman, T. 1999, in Bulletin of the American Astronomical Society, 1495–+

Ben-Mizrachi, A., Procaccia, I., & Grassberger, P. 1984, Phys. Rev. A, 29, 975

Bendat, J. S. & Piersol, A. G. 1986, Random Data: Analysis and Measurement Procedures (New York: John Wiley & Sons, |c1986, 2nd ed.)

Berge, D., Funk, S., & Hinton, J. 2006, ArXiv Astrophysics e-prints

Bertero, M. & Pike, E. R. 1982, Optica Acta, 29, 727

Bevington, P. R. & Robinson, D. K. 1992, Data reduction and error analysis for the physical sciences (New York: McGraw-Hill, |c1992, 2nd ed.)

Blackburn, J. K. 1995, in ASP Conf. Ser. 77: Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes, 367–+

Błażejowski, M., Blaylock, G., Bond, I. H., et al. 2005, ApJ, 630, 130

Boldt, E. & Garmire, G. 1975, in X-Rays in Space - Cosmic, Solar, and Auroral X-Rays, Volume 1, 1146–1155

Brandstater, A. & Swinney, H. L. 1987, Phys. Rev. A, 35, 2207

Brinkmann, W., Maraschi, L., Treves, A., et al. 1994, A&A, 288, 433

Brockwell, P. J. & Davis, R. A. 2002, Introduction to Time Series and Forecasting (New York: Springer-Verlag, |c2002, 2nd ed.)

Broomhead, D. S., Jones, R., & King, G. P. 1987, J. Phys. A: Math. Gen., 20, L563

Broomhead, D. S. & King, G. P. 1986, Physica D: Nonlinear Phenomena, 20, 217

Burbidge, G. R., Jones, T. W., & Odell, S. L. 1974, ApJ, 193, 43

Buse, A. 1982, Am. Stat., 36, 153

Catanese, M. & Sambruna, R. M. 2000, ApJ, 534, L39

Chadwick, P. M., Lyons, K., McComb, T. J. L., et al. 1999, ApJ, 513, 161

Chatfield, C. 2003, The Analysis of Time Series: An Introduction (USA: Chapman & Hall/CRC, |c2004, 6th. ed.), Texts in Statistical Science

Chiappetti, L. & Torroni, V. 1997, IAU Circ., 6776, 2

Ciprini, S., Lindfors, E., Nilsson, K., & Ostorero, L. 2006, in ASP Conf. Ser. 350: Blazar Variability Workshop II: Entering the GLAST Era, ed. H. R. Miller, K. Marshall, J. R. Webb, & M. F. Aller, 75–+

Clements, M. P., Franses, P. H., Smith, J., & van Dijk, D. 2003, Journal of Forecasting, 22, 359

Cui, W. 2004, ApJ, 605, 662

Cui, W., Blazejowski, M., Aller, M., et al. 2005, in AIP Conf. Proc. 745: High Energy Gamma-Ray Astronomy, ed. F. A. Aharonian, H. J. Völk, & D. Horns, 455–461

Czerny, B., Doroshenko, V. T., Nikołajuk, M., et al. 2003, MNRAS, 342, 1222

Davies, N. & Petruccelli, J. D. 1986, The Statistician, 35, 271

Davies, R. B. & Harte, D. S. 1987, Biometrika, 74, 95

Dickey, J. M. & Lockman, F. J. 1990, ARA&A, 28, 215

Ding, M., Grebogi, C., Ott, E., Sauer, T., & Yorke, J. A. 1993, Phys. Rev. Lett., 70, 3872

Done, C., Madejski, G. M., Mushotzky, R. F., et al. 1992, ApJ, 400, 138

Duvall, P., Keesling, J., & Vince, A. 2000, J. London Math. Soc., 61, 748

Eckmann, J. P. & Ruelle, D. 1992, Physica D: Nonlinear Phenomena, 56, 185

Edelson, R., Turner, T. J., Pounds, K., et al. 2002, ApJ, 568, 610

Elner, S. 1988, Phys.Lett.A, 133, 128

Emmanoulopoulos, D., Pühlhofer, G., & Wagner, S. 2005, in AIP Conf. Proc. 745: High Energy Gamma-Ray Astronomy, ed. F. A. Aharonian, H. J. Völk, & D. Horns, 475–480

Engle, R. F. 1982, Econometrica, 50, 987

Farmer, J. D., Ott, E., & Yorke, J. A. 1983, Physica D: Nonlinear Phenomena, 7, 153

Fasano, G. & Vio, R. 1988, Bulletin d'Information du Centre de Donnees Stellaires, 35, 191

Ferrero, E., Wagner, S. J., Emmanoulopoulos, D., & Ostorero, L. 2006, A&A, 457, 133

Fossati, G., Celotti, A., Chiaberge, M., et al. 2000, ApJ, 541, 153

Gaidos, J. A., Akerlof, C. W., Biller, S. D., et al. 1996, Nature, 383, 319

George, I. M., Warwick, R. S., & Bromage, G. E. 1988, MNRAS, 232, 793

Ghaddar, D. K. & Tong, H. 1981, Appl. Statist., 30, 238

Giommi, P., Angelini, L., Jacobs, P., & Tagliaferri, G. 1992, in ASP Conf. Ser. 25: Astronomical Data Analysis Software and Systems I, ed. D. M. Worrall, C. Biemesderfer, & J. Barnes, 100–+

Grassberger, P. & Procaccia, I. 1983a, Phys. Rev. Lett., 50, 346

Grassberger, P. & Procaccia, I. 1983b, Physica D: Nonlinear Phenomena, 9, 189

Green, A. R., McHardy, I. M., & Done, C. 1999, MNRAS, 305, 309

Green, A. R., McHardy, I. M., & Lehto, H. J. 1993, MNRAS, 265, 664

Griffiths, R. E., Briel, U., Chaisson, L., & Tapia, S. 1979, ApJ, 234, 810

Guegan, D. & Pham, T. D. 1992, Statistica Sinica, 2, 157

Guegan, D. & Wandji, J. N. 1996, Statistics & Probability Letters, 29, 201

Hamuy, M. & Maza, J. 1989, AJ, 97, 720

Harris, D. E., Cheung, C. C., Biretta, J. A., et al. 2006, ApJ, 640, 211

Harwit, M. 2006, Astrophysical concepts (New York : Springer Science+Business Media, LLC, |c2006, 4th ed. (Astronomy and astrophysics library))

Hausdorff, F. 1918, Mathematische Annalen, 79, 157

Hauser, M. A. 1997, Empirical Economics, 22, 247

Hayashida, K., Inoue, H., Koyama, K., Awaki, H., & Takano, S. 1989, PASJ, 41, 373

Hénon, M. 1976, Comm.Math.Phys., 50, 69

Hinich, M. J. 1982, Journal of Time Series Analysis, 3, 169

Hink, P., Pelling, M., & Rothschild, R. 1992, in EUV, X-ray, and gamma-ray instrumentation for astronomy III; Proceedings of the Meeting, San Diego, CA, July 22-24, 1992 (A93-29476 10-35), p. 140-150., 140–150

Hinton, J. A. 2004, New Astronomy Review, 48, 331

Hjellvik, V. & Tjostheim, D. 1995, Biometrika, 82, 351

Hoyle, F., Burbidge, G. R., & Sargent, W. L. W. 1966, Nature, 209, 751

Hughes, P. A., Aller, H. D., & Aller, M. F. 1992, ApJ, 396, 469

Hurst, H. E. 1951, Transactions of the American Society of Civil Engineers, 116, 770

Jahoda, K., Markwardt, C. B., Radeva, Y., et al. 2006, ApJS, 163, 401

Jahoda, K., Swank, J. H., Giles, A. B., et al. 1996, in Proc. SPIE Vol. 2808, p. 59-70, EUV, X-Ray, and Gamma-Ray Instrumentation for Astronomy VII, Oswald H. Siegmund; Mark A. Gummin; Eds., 59–70

Kataoka, J., Tanihata, C., Kawai, N., et al. 2002, MNRAS, 336, 932

Kedziora-Chudczer, L., Jauncey, D. L., Wieringa, M. H., et al. 1997, ApJ, 490, L9+

Keenan, D. M. 1985, Biometrika, 72, 39

Kellermann, K. I. & Pauliny-Toth, I. I. K. 1969, ApJ, 155, L71+

Kidger, M. R. 1989, A&A, 226, 9

Kolmogorov, A. N. 1941a, Doklady Akad. Nauk CCCP, 30

Kolmogorov, A. N. 1941b, Doklady Akad. Nauk CCCP, 32

Krawczynski, H., Sambruna, R., Kohnle, A., et al. 2001, ApJ, 559, 187

Krennrich, F., Bond, I. H., Bradbury, S. M., et al. 2002, ApJ, 575, L9

Krivine, H., Lesne, A., & Treiner, J. 2004, ArXiv Condensed Matter e-prints

Lainela, M. & Valtaoja, E. 1993, ApJ, 416, 485

Lancaster, T. 1968, Journal of the American Statistical Association, 63, 182

Lawrance, A. J. & Lewis, P. A. W. 1980, J. R. Statist. Soc. B (Methodological), 42, 150

Lawrance, A. J. & Lewis, P. A. W. 1985, J. R. Statist. Soc. B (Methodological), 47, 165

Lawrence, A. & Papadakis, I. 1993, ApJ, 414, L85

Leighly, K. M. & O'Brien, P. T. 1997, ApJ, 481, L15+

Levenbach, H. 1973, Journal of the American Statistical Association, 68, 436

Levine, A. M., Bradt, H., Cui, W., et al. 1996, ApJ, 469, L33+

Li, T.-P. & Ma, Y.-Q. 1983, ApJ, 272, 317

Li, W. K. & McLeod, A. I. 1986, Biometrika, 73, 217

Lo, A. W. 1991, Econometrica, 59, 1279

Lorenz, E. N. 1963, J. Atmos.Sci., 20, 130

Luukkonen, R., Saikkonen, P., & Teräsvirta, T. 1988, Biometrika, 75, 491

Macomb, D. J., Akerlof, C. W., Aller, H. D., et al. 1995, ApJ, 449, L99+

Mandelbrot, B. B. 1972, Annals of Economic and Social Measurement, 257

Mann, H. B. & Whitney, D. R. 1947, Annals Math. Stat., 18, 50

Manwell, T. & Simon, M. 1968, AJ, 73, 407

Maraschi, L., Fossati, G., Tavecchio, F., et al. 1999, ApJ, 526, L81

Maravall, A. 1983, Journal of Business & Economic Statistics, 1, 66

Miles, J. 1984, PNAS, 81, 3919

Miller, H. R., Carini, M. T., & Goodrich, B. D. 1989, Nature, 337, 627

Mittaz, J. P. D. & Branduardi-Raymont, G. 1989, MNRAS, 238, 1029

Moody, J. & Wu, L. 1991, in Proceedings of the Third International Conference: Neural Networks in Financial Engineering, ed. A. Refenes, Y. Abu-Mostafa, J. Moody, & A. Weigend (London: World Scientific), 537–553

Morrison, R. & McCammon, D. 1983, ApJ, 270, 119

Nakase, T. & Cangaroo Collaboration. 2003, in International Cosmic Ray Conference, 2587–+

Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1997, ApJ, 476, 70

Nicastro, F., Zezas, A., Drake, J., et al. 2002, ApJ, 573, 157

Nishijima, K., Asahara, A., Bicknell, G. V., et al. 2001, in International Cosmic Ray Conference, 2626–+

Noakes, J. L. 1986, IMA J Math Control Info, 3, 293

Ostorero, L., Wagner, S. J., Gracia, J., et al. 2006, A&A, 451, 797

Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. 1980, Phys. Rev. Lett., 45, 712

Paladin, G. & Vulpiani, A. 1987, Phys.Rep., 156, 147

Paluš, M. & Dvořák, I. 1992, Physica D: Nonlinear Phenomena, 55, 221

Papadakis, I. E., Brinkmann, W., Negoro, H., & Gliozzi, M. 2002, A&A, 382, L1

Papadakis, I. E. & Lawrence, A. 1993, MNRAS, 261, 612

Papadakis, I. E. & Lawrence, A. 1995, MNRAS, 272, 161

Papadakis, I. E., Samaritakis, V., Boumis, P., & Papamastorakis, J. 2004, A&A, 426, 437

Pence, W. 1999, in ASP Conf. Ser. 172: Astronomical Data Analysis Software and Systems VIII, ed. D. M. Mehringer, R. L. Plante, & D. A. Roberts, 487–+

Pence, W., Xu, J., & Brown, L. 1997, in ASP Conf. Ser. 125: Astronomical Data Analysis Software and Systems VI, ed. G. Hunt & H. Payne, 261–+

Peters, E. E. 1994, Fractal Market Analysis: Applying Chaos Theory to Investment and Economics (New York: John Wiley & Sons, |c1994)

Piner, B. G. & Edwards, P. G. 2005, ApJ, 622, 168

Piner, B. G., Unwin, S. C., Wehrle, A. E., et al. 1999, ApJ, 525, 176

Platt, N., Spiegel, E. A., & Tresser, C. 1993, Phys. Rev. Lett., 70, 279

Pomeau, Y. & Manneville, P. 1980, Communications in Mathematical Physics, 74, 189

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in FORTRAN. The art of scientific computing (Cambridge: University Press, |c1992, 2nd ed.)

Priestley, M. B. 1981, Spectral Analysis and Time Series: Probability and Mathematical Statistics, Vol. 1-2 (London: Academic Press, |c1981)

Provenzale, A., Smith, L. A., Vio, R., & Murante, G. 1992, Physica D: Nonlinear Phenomena, 58, 31

Punch, M., Akerlof, C. W., Cawley, M. F., et al. 1992, Nature, 358, 477

Qian, S. J., Quirrenbach, A., Witzel, A., et al. 1991, A&A, 241, 15

Quinn, J., Akerlof, C. W., Biller, S., et al. 1996, ApJ, 456, L83+

Quirrenbach, A., Witzel, A., Kirchbaum, T. P., et al. 1992, A&A, 258, 279

Quirrenbach, A., Witzel, A., Wagner, S., et al. 1991, ApJ, 372, L71

Racine, R. 1970, ApJ, 159, L99+

Rebillot, P. F., Badran, H. M., Blaylock, G., et al. 2006, ApJ, 641, 740

Russell, D. A., Hanson, J. D., & Ott, E. 1980, Phys. Rev. Lett., 45, 1175

Saikkonen, P. & Lukens, R. 1988, Scandinavian Journal Of Statistics, 15, 55

Sato, S., Sano, M., & Sawada, Y. 1988, Phys. Rev. A, 37, 1679

Sauer, T., Yorke, J. A., & Casdagli, M. 1991, Journal of Statistical Physics, 65, 579

Schlegel, E. M. 2006, Appendix F: The XTE Technical Appendix, Tech. rep., NASA/GSFC, web:`http://rxte.gsfc.nasa.gov/docs/xte/appendix_f.html`

Schmitz, A. & Schreiber, T. 1999, Phys. Rev. E, 59, 4044

Schwartz, D. A., Griffiths, R. E., Schwarz, J., Doxsey, R. E., & Johnston, M. D. 1979, ApJ, 229, L53

Simonetti, J. H., Cordes, J. M., & Heeschen, D. S. 1985, ApJ, 296, 46

Small, M., Yu, D., & Harrison, R. G. 2001, Phys. Rev. Lett., 87, 188101

Stella, L. & Angelini, L. 1992, in ASP Conf. Ser. 25: Astronomical Data Analysis Software and Systems I, ed. D. M. Worrall, C. Biemesderfer, & J. Barnes, 103–+

Strogatz, S. H. 1998, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering (USA: Addison-Wesley Publishing Company, |c1994, 9nth print.), Perseus Books, Reading, Massachusetts

Subba Rao, T. & Gabr, M. M. 1980, Journal of Time Series Analysis, 1, 145

Takahashi, T., Kataoka, J., Kii, T., et al. 1996, Memorie della Societa Astronomica Italiana, 67, 533

Takahashi, T., Kataoka, J., Madejski, G., et al. 2000, ApJ, 542, L105

Takens, F. 1981, in Lecture Notes in Mathematics, Vol. 898, Dynamical Systems and Turbulence, ed. D. A. Rand & L. S. Young, Warwick 1980 (Berlin: Springer-Verlag), 366–381

Tanaka, Y., Inoue, H., & Holt, S. S. 1994, PASJ, 46, L37

Tanihata, C., Urry, C. M., Takahashi, T., et al. 2001, ApJ, 563, 569

Taylor, B. G., Andresen, R. D., Peacock, A., & Zobl, R. 1981, Space Sci. Rev., 30, 479

Teräsranta, H., Wiren, S., Koivisto, P., Saarinen, V., & Hovatta, T. 2005, A&A, 440, 409

Terrell, J. & Olsen, K. H. 1970, ApJ, 161, 399

Theiler, J. 1987, Phys. Rev. A, 36, 4456

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Doyne Farmer, J. 1992, Physica D: Nonlinear Phenomena, 77

Timmer, J. & Koenig, M. 1995, A&A, 300, 707

Tsay, R. S. 1986, Biometrika, 73, 461

Ulrich, M.-H., Maraschi, L., & Urry, C. M. 1997, ARA&A, 35, 445

Urry, C. M. & Padovani, P. 1995, PASP, 107, 803

Uttley, P., McHardy, I. M., & Papadakis, I. E. 2002, MNRAS, 332, 231

Valtaoja, E., Teräsranta, H., Tornikoski, M., et al. 2000, ApJ, 531, 744

Vaughan, S., Edelson, R., Warwick, R. S., & Uttley, P. 2003, MNRAS, 345, 1271

Vautard, R., Yiou, P., & Ghil, M. 1992, Physica D: Nonlinear Phenomena, 58, 95

Vestrand, W. T. & Sreekumar, P. 1999, Astroparticle Physics, 11, 197

Vestrand, W. T., Stacy, J. G., & Sreekumar, P. 1995, ApJ, 454, L93+

Vio, R., Cristiani, S., Lessi, O., & Provenzale, A. 1992, ApJ, 391, 518

Vio, R., Cristiani, S., Lessi, O., & Salvadori, L. 1991, ApJ, 380, 351

Wagner, S. J., Witzel, A., Heidt, J., et al. 1996, AJ, 111, 2187

Xue, Y. & Cui, W. 2005, ApJ, 622, 160

Xue, Y., Yuan, F., & Cui, W. 2006, ApJ, 647, 194

Zhang, Y. H., Celotti, A., Treves, A., et al. 1999, ApJ, 527, 719

Author Index

Index

Acknowledgments

While it is not possible to identify everyone who contributed in one way or another to this work (that would need several pages) I will make a try to mention some of the individuals that were mostly related with my working environment during these four years that I spent in Heidelberg but please note that NO ONE has been forgotten!

Initially, I would like to express my gratitude to my supervisor *Stefan Wagner* who supported me all these years by offering me guidance and a lot of opportunities to present my work in Germany and abroad. I am also grateful to him for introducing me to one of the biggest astrophysical collaborations in the field of VHE astronomy the H.E.S.S. Additionally I would like to say that I am feeling deeply indebted to all the members of the E.N.I.G.M.A. network (HPRN-CT-2002-00321), where my funds came from, who created an ideal environment for a beginner like me to express thoughts and see "from the inside" how a scientific community is actually working.

Then I would like to thank sincerely *Iosif Papadakis* for dedicating a vast amount of his time, during my visit in the university of Crete, in order to introduce me to the modern analysis methods of X-ray astronomy. Throughout these years I had plenty of fruitful discussions with him concerning several aspects of my work. I also want to special thank *John Kirk* who was able during our conversations to explain to me in a precise and accurate way, several issues dealing with theoretical astrophysics. It would be a great omission if I did not acknowledge the persons who gave me the incentive, during my university studies in Athens, to work on this subject. *Apostolos Mastichiadis* was the person who through his lectures gave me an overview of the basic concepts in high energy astrophysics and *Kanaris Tsinganos* was the person who, apart from his lectures on plasma astrophysics, proposed me to join the group of Stefan Wagner in Heidelberg. Therefore without their contribution this work would not have even begun.

Many thanks also to all the people in the LSW for helping me in several ways during the last years and for creating a pleasant working atmosphere: The "computer guru" *Otmar Stahl* who had a solution to all my problems (originating mainly from my UNIX inexpertness ☺), *Gerd Pühlhofer* who elucidate to me several aspects concerning the H.E.S.S. system and all the persons of Stefan Wagner's group (Marcus, Elisa, Bagmeet, Stephanie, Tam, Giovanna and Peter) for being so cooperative during these years. Of course not to forget *André* with whom we were "traveling" together, almost every day for the last year, with the bus 21 "auf den Königstuhl", *José* who shared with me the same views on several subjects during our network meetings and *Thomas* Mäd*ler* for his "Mad" parties!

Luisa it is now your time to be acknowledged! Thanks a lot for being so polite and so tolerant these three years that we spent together. As I have already told you are going to be the "comparison unit" for all my future office mates. Since I know that you get too emotional when you read these kind of statements here is something in order to cheer you up: "This is a very important plot whatever it means" !!! ☺

A big thanks should also go to the persons that helped me to settle in during my first months in Heidelberg: *Dina Theochari*, *Dimitris Kampanaros* and the very nice couple and colleagues *Konstantina Manolakou* and *Dieter Horns*.

In this point I must also say a big Vielen Dank to my girlfriend *Henrike* for sharing with me this "*rrr*ealy" nice last year. Thanks a lot for supporting me with all your strength and for surprising me with your unexpected acts. Ich spreche zwar nicht so "Hochdeutsch" aber ich glaube, das wirst Du verstehen "Vielen Dank dafür alles zu tun, dass jeder Tag ein neuer und spezieller Tag ist".

Last BUT OF COURSE not least I would like to say a big "ΕΥΧΑΡΙΣΤΩ" to my marvelous family who supported all my efforts and all my decisions during these years in an unconditional and unique way. I am grateful to my father *Διαμαντής* my mother Ε*υθυμία* and my sister Ε*λένη* just for being there whenever I needed them.

Ίσως είναι η πρώτη φορά που σας λέω τόσο άμεσα "ευχαριστώ πάρα πολύ" για όλα όσα μου προσφέρατε όλα αυτά τα χρόνια.