Proceedings of the VII Bulgarian-Serbian Astronomical Conference (VII BSAC)
Chepelare, Bulgaria, June 1-4, 2010,
Editors: M. K. Tsvetkov, M. S. Dimitrijević, K. Tsvetkova, O. Kounchev, Ž. Mijajlović
Publ. Astron. Soc. "Rudjer Bošković" No 11, 2012, 107-123

BAYESIAN PROBABILITY THEORY IN ASTRONOMY: LOOKING FOR STELLAR ACTIVITY CYCLES IN PHOTOMETRIC DATA-SERIES

ANA BORISOVA[1] , VALERI HAMBARYAN[2] , LASKO LASKOV[3]

[1] *Institute of Astronomy, Bulgarian Academy of Sciences, 72 Tsarigradsko shosse blvd., 1784 Sofia, Bulgaria*
E-mail: ana@skyarchive.org\
[2] *Astrophysical Institute and University Observatory, Friedrich Schiller University of Jena, Schillergäßchen 2/3, 07743 Jena, Germany*
E-mail: vvh@astro.uni-jena.de
[3] *New Bulgarian University, Department of Informatics, 21 Montevideo Str., 1618 Sofia, Bulgaria*
E-mail: llaskov@nbu.bg

Abstract. The application of the Bayesian probability theory in a various astronomical research work over the past decade is discussed in the presented talk. The basic idea of the Bayesian approach to astronomical data is presented with a special attention for its plausibility to the subject of Astroinformatics. In Particular the Gregory-Loredo method for periodic signal detection of unknown shape in time-series with Gaussian errors is tested.

1. BRIEF HISTORY OF THE BAYESIAN PROBABILITY THEORY

At the present astronomical publications the adjective "Bayesian" is increasingly often in use in order to point at specific logical approach and to distinguish the "classical" statistical approach to observational data. As a part of Probability Theory, Bayesian probability theory exploits the idea of the probability as "a measure of state of knowledge" (Jaynes, 2003), rather than a long-run expected frequency of the occurrence of the event. This extended concept for probability comes close to the every-day meaning of this world as insufficient reasoning and is a base for advanced data analysis techniques, such as hypothesises testing, model comparison, subjective reasoning and data mining. At this point of view, following the inductive and deductive logic, Bayesian analysis can significantly improve the parameter estimation, allowing the researcher to assign probabilities to competing hypotheses.

The first fundaments of the Bayesian logic belong to Thomas Bayes (1702 – 1761), who was an English mathematician and Presbyterian minister (Encyclopædia Britannica, 2010). His main work “Essay Towards Solving a Problem in the Doctrine of Chances” (1763), was published in the Philosophical Transactions of the Royal Society. In this essay Bayes published the first version of what latter has become known as “Bayes's theorem”, a discussion of the binomial distribution as well as the first occurrence of a probability logic result involving conditional probability (Dale, 2003). According to Dale (2003), Bayes’ mathematical work include discussions of probability, trigonometry, geometry, solution of equations, series and differential calculus, but also he was interested in electricity, optics, astronomy and celestial mechanics.

The general version of the Bayes theorem, (the theorem that treats the conditional probabilities), and the early Bayesian probability theory were set up and developed by the French mathematician, astronomer, and physicist Pierre-Simon Laplace, (1749–1827). Laplace is well known for his solar system investigations, but he has also demonstrated the usefulness of probability approach to scientific data, especially in celestial mechanics, medical statistics and law sciences (Stigler, 1986). Laplace has also introduced the principle for assignment of the priors, called the principle of insufficient reason (Fienberg, 2006). He used uniform priors, which are the simplest non-informative priors, reasonable in case of insufficient knowledge for setting up the informed priors. Latter on this principle was called by De Morgan (Fienberg, 2006) , the inverse probability as it infers backwards from observations to parameters. After the 1920s, Laplace’s probability principles were argued mostly by Ronald A. Fisher, Jerzy Neyman and Egon Pearson and were substituted by a set of methods latter called frequentist or classical statistics (Fienberg, 2006). Neyman, in his work "Frequentist probability and frequentist statistics" (Neyman, 1977), developed the idea of confidence intervals because "the whole theory would look nicer if it were built from the start without reference to Bayesianism and priors". Actually *Bayesian* appeared as a terminology in the 1930s, and latter on it was used by those who have not accepted the limitations of frequentist statistics (Fienberg, 2006).

According to Daston (1994), “Between 1837 and 1843 at least six authors: … made similar distinctions between the probabilities of things and the probabilities of our beliefs about things." These two different approaches gave rise to the objective and subjective directions in Bayesian theory. In the objective direction the statistical analysis depends only on the data and the assumed models, and is not influenced by subjective decisions. For instance in the early 1920s, John Keynes represent the idea that the probability should be treated as “subjective degree of belief in a proposition”, while the classical approach to probability refers to the frequency of the occurrence of the event. Latter on at 1939 Harold Jeffreys (1939}, published his basic work “Theory of probability” (Jeffreys, 1939), were he developed the Objective Bayesian inference.[1] At 1957^{th} Edwin Jaynes introduced the principle of entropy for priors constructing, and in 1965^{th} Dennis Lindley with his book "Introduction to Probability and Statistics from a Bayesian" promoted the Bayesian methods.[2] The development of the Markov chain Monte-Carlo methods at eighties removed many computational problems in front of the Bayesian statistics. At present Bayesian approach is widely used in different applications for machine learning, data mining, Bayesian network.

2. BASIC CONCEPTS OF THE BAYESIAN PROBABILITY THEORY

In general Bayesian probability theory gives tools for evaluating of the probability of hypothesis, using the prior probability distribution, updated or affected by the relevant data and the available additional information. Thus the probability of a given hypothesis could be updated with the new data releases and might be interpreted as a "state of knowledge" (Jaynes, 2003). Contrary, in the frequentist approach a hypothesis is either accepted or rejected, without assigning to probability. Bayesian approach is based on the Probability theory rules, such as product and sum rules, Boolean algebra and the Bayes theorem. The Bayes theorem gives the rule for the conditional probability P(H|D,I) of the proposition H, given that the proposition D and information I are true (shortly – H given D and I) and its general form, according to Gregory (2005) is:

$P(H|D,I)=P(D|H,I)P(H|I)/P(D|I)$, where

H is a specific proposition or set of hypothesis,

D is the evidence or the data that are observed,

I represent the prior information,

P(H|I) is the prior probability of H given the I, that was assumed before the data became available,

P(D|H,I) is the conditional probability of the data, given the H and I, also called the likelihood function,

P(D|I) is the marginal probability of the data given I: the prior probability of the data under all possible Hypothesizes: $P(D|I)=\sum P(D|H_i,I)\ P(H_i|I)$, it's a normalization factor that ensure the sum of all the probabilities of the hypothesizes to be 1.

---

[1]Wikipedia, History of statistics, http://en.wikipedia.org/wiki/History_of_statistics# Bayesian_ tatistics, Aug. 2010.

[2] Ibid.

$P(H|D,I)$ is the probability of the proposition H (hypothesis) given the proposition D (data) and I (information)

The term in front the $P(H|I)$, $P(D|H,I)/P(D|I)$, describes the influence of the observational data to the probability of the hypothesis. When it is likely to observe the data under the hypothesis, then this factor will be large and it will result in a larger posterior probability of the hypothesis given the data. Contrary, if it is unlikely to observe the data if the hypothesis is true, then the term would reduce the posterior probability for the hypothesis. Thus the Bayes theorem measures the influence of the data on the on the probability of the hypothesis. Bayesian inference, when applied to scientific data analysis, rules the updating hypothesizes given to the new data or experiments by a basic schedule of few steps: (1) setting up the hypothesis space and the prior probabilities; (2) Data models and parameter space definition; (3) Hypothesis testing and (4) calculation of the global likelihood function.

At present, the definition of the hypothesis space in astronomy is based on the observational data, knowledge gained in the previous research or on theoretical consideration. There are two general approaches for assigning prior probability distributions of the parameters: informative, based on the previous evidence or on the expert opinion; and uninformative, based on general or obscure information. Useful practical approach for setting informative priors is to take a normal distribution with expected value based on the previous observation. The simplest rule for setting up uninformative priors is assignment of equal probabilities to all possibilities, but this encounters problems if the prior range of the parameter is infinite. There are also some other reasoning for priors set up: conjugate priors, which provide for the same type of the prior and posterior distribution, reduce the computational problems; the Jeffreys priors that assure that the statement of the prior believe are the same in different scales (such prior distribution is reasonable in time-series analysis to ensure equal results in terms of period and frequency).

Scientific data models in general are described by several parameters and are accounting for observational errors. In this case the Bayesian inference provides the computation of the joint likelihoods and probability distribution functions for each of the parameters. The marginalization (integration or summation of the joint posterior distribution function over the nuisance parameters) procedure gives the marginal probability of the parameter of interest and then the mean, mode values and the credible intervals are easily estimated.

Bayesian model selection answers the question how probable is a model given the data, if we consider a set of models (M_1 and M_2 for instance) independently of the model parameters. The models are evaluated by computing the odds factors (Q_{12}) and marginalization out all the parameters Gregory (2005):

$\mathbf{Q_{12}=P(M_1|D,I)/ P(M_2|D,I)=[P(D|M_1,I)/ P(D|M_2,I)][P(M_1|I)/ P(M_2|I)]}$, where

$[P(M_1|I)/ P(M_2|I)]$ is the prior odds of the model, often taken to be 1, assuming the models are of equal probability

$[P(D|M_1,I)/ P(D|M_2,I)]$ is the Bayes factor (B_{12}), computed by the marginal likelihoods for each model.

3. BAYESIAN PROBABILITY THEORY, LAST DECADE IN ASTRONOMY

Increased impact of the Bayesian inference in astronomical research may be traced in professional literature, in modern statistical application as well as in the publications in the main astronomical journals. A reference into the Amazon.com book store returns several volumes, written by high level professionals in statistical studies and astrophysics:

The work of P.C. Gregory, “Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support” (2005, 2006), (Gregory, 2005) undergoes its third edition at 2010 and is a fundamental book that discusses application of the Bayesian statistics in physical sciences. The book gives detailed and clear exposition of the Bayesian concepts with number of useful examples, numerical techniques for Bayesian calculations, an introduction to Bayesian Markov Chain Monte-Carlo integration and least-squares analysis. In addition it is supported with a Mathematica notebook providing an easy to learning routines.

Modeling disc galaxies using Bayesian/Markov chain Monte Carlo is the subject of the recent book by D. Puglielli “Galaxy Modeling using Bayesian Statistics: A Bayesian/Markov chain Monte Carlo Approach to Modeling NGC 6503” (Puglielli, 2010). A large set of observations for the dwarf spiral galaxy NGC 6503 is examined for fitting with sophisticated dynamical models and the joint posterior probability function for the model parameters is obtained. This approach gives constraints of important properties of the galaxy as its mass and mass-to-light ratio, halo density profile, and structural parameters.

The application of the Bayesian methods in cosmological research is represented in the recent book by Michael P. Hobson and Andrew H. Jaffe “Bayesian Methods in Cosmology” (2010). The contribution of 24 experts in cosmology and statistics makes this book essential and competent guide for researches in cosmology. The book represents precise modeling of the Universe properties and gives a methodology (the basic foundations, parameter estimation, model comparison and signal separation) as well as a wide range of applications such as source detection, cosmic microwave background analysis, classification of galaxy properties.

It also worth mention the publication of the 27th International Workshop on Bayesian Inference and Maximum Entropy Methods, named “Bayesian Inference in Science and Engineering: 27th International Workshop on Bayesian Inference and Maximum Entropy Methods” (2007) by the edition of Kevin H. Knuth, Ariel Caticha, Julian L. Center, and Adom Giffin (2007). For 30 years the MaxEnt workshops have explored the use of Bayesian probability theory, entropy and information theory in scientific, engineering and signal processing applications. Volume No. 27 considers Methods, Foundations and Applications in astronomy, physics, chemistry, biology, earth science, and engineering.

A collection of essays "Blind Image Deconvolution: Theory and Applications" (2007), edited by P. Campisi, and K. Egiazarian expose up to day approaches theoretical fundaments of Blind Image Deconvolution techniques. A special chapter for application of the Deconvolution and Blind image Deconvolution techniques in astronomy is provided by Eric Pantin, Jean-Luc Stark and Fionn Murtagh, which also include Bayesian approach. Bayesian methodology for image deconvolution is also exposed in the book of J.-L. Starck and F. Murtagh, "Astronomical Image and Data Analysis" (2006).

A basic series of Astronomical statistics, "Statistical Challenges in Astronomy", (2003, 2010), Statistical Challenges in Modern Astronomy II" (1997), edited by E. Feigelson, G. Babu, was released after the conferences of the same name. These volumes focus on the topics: Bayesian approaches to astronomical data modelling, the Virtual Observatory impact on present astronomical research, time series analysis, image analysis, statistical modeling of critical datasets and its application in cosmology. Many problems are introduces on the base of large astronomical projects, such as LIGO, AXAF, XTE, Hipparcos, and digitised sky surveys.

In the last decade several software applications were found to be in use for Bayesian astronomical data analysis:

BAYES-ME code of A. Asensio Ramos (2009), for investigations of spectropolarimetric observations with the *Hinode* solar space telescope;

MULTINEST/SUPERBAYES (http://superbayes.org/) - is a robust Bayesian inference tool for cosmology and particle physics;[3]

ARGO: **A**lgorithm for the **R**econstruction of **G**alaxy-traced **O**ver-densities (Kitaura and Enßlin, 2008) – gives methodology, inverse algorithms and numerical optimization for Bayesian reconstruction of the cosmological large-scale structure;

Bayesian Photometric Redshift code BPZ, http://acs.pha.jhu.edu/ ~txitxo/);

ZEBRA: Zurich Extragalactic Bayesian Redshift Analyzer, http://www.astro.phys.ethz.ch/exgal_ocosm/zebra/index.php, combines and extends several of the classical approaches to produce accurate photometric redshifts down to faint magnitudes; it uses template-fitting approach to produce Maximum Likelihood and Bayesian redshift estimates.

The most evident impact of the Bayesian approach to analysis of astronomical data and information could be seen in the publications in the professional journals. In the last decade more than 300 papers reports the application of the Bayesian methods, starting with about 15 papers at 2000 and ended up with more than 50 papers per year at the last two years of the decade. The Bayesian methodology was not only used for precise parameter estimation but also for Model Comparison, Hypothesis testing, Object detection, identification and classification, Image

[3] "SuperBayes code", http://superbayes.org/, Aug 2010

deconvolution. This approach has been applied to rather diverse astronomical topics; the main items are listed below in decreased number of publications:

- Cosmology: Uses the Bayesian inference for data analysis (most of the papers concern the data analysis of the Wilkinson Microwave Anisotropy Probe Observations), Estimation of the Cosmic Microwave Background, The Universe parameters and model estimation, Dark energy studies;
- Gravitational lensing: Uses the Bayesian inference for gravitational lenses detection and modeling;
- Variable stars: Uses the Bayesian inference for variable stars detection, identification and classification, Light and Radial velocity curves analysis - estimation of the light curves properties, SN identification and classification;
- Spectral fitting and deconvolution ;
- Extrasolar Planets: Uses the Bayesian inference for New planet searches (Bayesian Kepler periodogram), Orbit analysis of the Extrasolar planets;
- Solar astrophysics: Uses the Bayesian inference for solar flare predictions, magnetic fields estimation, solar oscillations detection;
- The Galaxy studies: Use the Bayesian inference for the dist and halo kinematics study, Star Formation Ratio estimation, The Velocity Distribution of Nearby Stars, HIPARCOS data analysis.

In principle, the Bayesian statistics is applied for analysis of mostly all astronomical type of data: Spectral (spectra fitting, Radial velocities, spectro-polarimetric data), Photometric, Kinematic (Velocity distribution, Kepler periodogram), and Image (optical, IR, Radio, X-ray data) series analysis.

4. GREGORY-LOREDO METHOD, GAUSSIAN NOISE CASE

In the frames of the Astroinformatics projects of the Bulgarian Academy of Sciences, we are interested of applying Bayesian statistical methods for analysis of the photometric data obtained by the digitization of photographic plates, combined with modern CCD photometry, and with published electrophotometric observations. Such data usually would exhibit random time distribution, wide time intervals with a lack of observations, and also different quality and observational errors. We find out that the Bayesian Gregory-Loredo (GL) method (Gregory, 1999) for time-series analysis with Gaussian error distribution is useful and practical for our research. The method gives robust and relatively fast tool for searching long-term photometric cycles with unknown shape. The GL method employs Bayes approach for signal detection and for the detected signal characteristics estimation. When using the Bayes statistics the first step is to determine the hypothesis space and its priori ranges, then to represent them with suitable models. In general, for stellar photometric data series we have three hypotheses - constant, variable, and periodic magnitude variations.

Time series of photometric data we had obtained, consist of the observed magnitude d_i, taken at the moment $\mathbf{t_i}$ and corresponding errors. So the data model for the observed stellar magnitude consists of model predicted magnitude $\mathbf{d_{pi}}$ plus an error term. The error term, $\mathbf{e_i}$, includes observational error estimated by the observer, $\mathbf{s_i}$ and any unknown noise or signal which is not represented in the model:

$d_i=d_{pi}+e_i$,

In this method we assume the noise variance, $\mathbf{e_i}$, is finite with Gaussian distribution with variance $\boldsymbol{\sigma_i^2}$.

In the GL method periodic models are represented by a signal folded into a stepwise function, similar to a histogram, with **m** phase bins per period plus a noise contribution. In principle, with such a flexible model we are able to approximate a light curve of any shape. Hypotheses for detecting periodical signals represent a class of stepwise periodical models with following parameters:

- **P**, or **ω** - period or angular frequency, in the priori range of (p_{lo}, p_{hi});
- **ϕ**- phase of minimal brightness of the star;
- **m** – number of bins in the priori range from 2 to 12;
- $\mathbf{r_i}$ – light curve value in the i^{th} bin;
- **b** - Noise scale parameter, (defined as $1/\sigma_i^2=1/s_i^2$), the ratio of the variances of ($d_i - d_{pi}$) and that of the observer noise estimates s_i. The priori range of is (0.05, 1.95);

The GL method uses Jeffreys priors distribution for **b** and $\tilde{\omega}$These priors give the good advantages to ensure the same Gaussian distribution for $\mathbf{s_i}$ and $\boldsymbol{\sigma_i}$ as well equal results in the period and frequency scales. The uniform priors are taken for phase and number of bins parameter. Based on the Bayes' theorem, Gregoty and Loredo (1996), and Gregory (1999), gave rules and equations for calculation of the global and marginal likelihood functions, and of the odds ratios of constant, nonperiodic and periodic models. The most probable model parameters are estimated by marginalization of the posterior probability over the priori specified range of each parameter.

We have tested the GL method using modeled datasets, with randomly distributed data-points over 100 years observed period and with data-gaps involved. Results for period detection are shown at Fig.1 (depending on the number of observations) and at Fig. 2 (depending on the amplitude of the light curve). The error bars at the Fig. 1 and 2 represent the deviation of the mode period, calculated by the use of bootstrapping method. Our test shows that the GL method gives reliable and accurate results even with sparse, randomly distributed data.

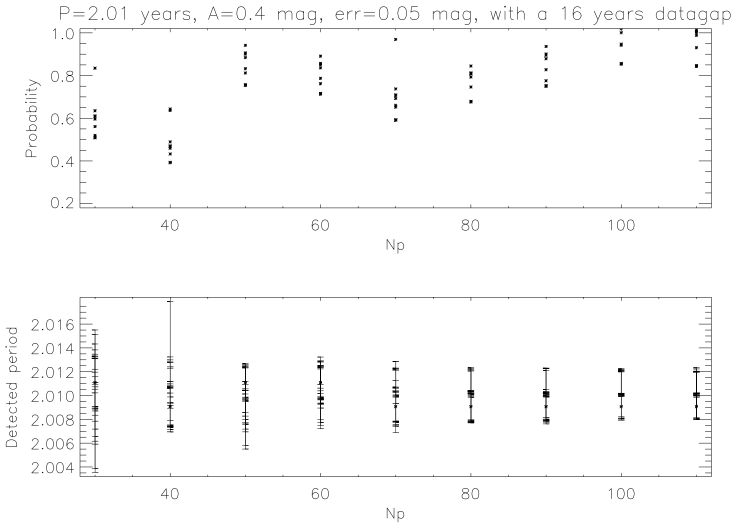


Figure 1. Results from the GL method for period detection: the most-probable detected period (bottom) and the Probability Density Function (top), depending on the number of observations. The modeled time-series are with period P=2.01 years, magnitude amplitude of 0.4 mag and observational error of 0.05 mag, and observed time span of 100 years with involved 16 years data-gap.

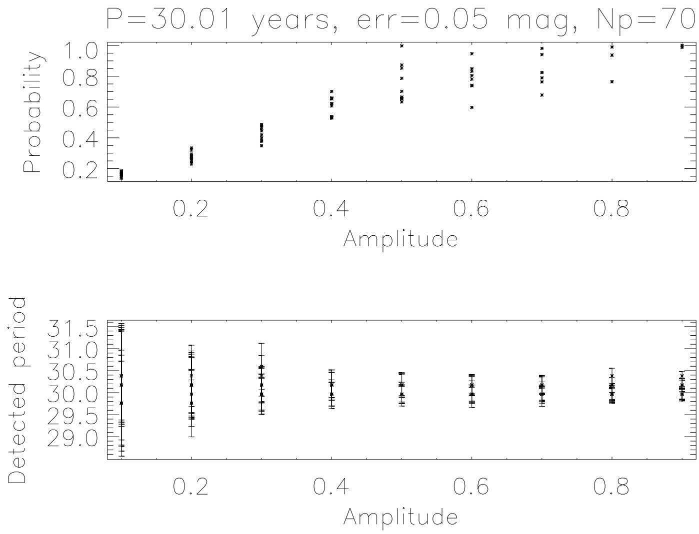


Figure 2. Results from the GL method for period detection, depending on the amplitude of stellar variability: the most-probable detected period (bottom) and the Probability Density Function (top), over modeled time-series, with period P=30.01 years, number of observations is 70, observational error of 0.05 mag and observed time span of 100 years.

5. GREGORY-LOREDO METHOD, APPLICATION FOR ANALYSIS OF THE CF OCT PHOTOMETRIC DATA

We applied the GL method for the analysis of the photometric data we have collected for the bright, southern active giant star CF Oct (HD 196818). Variability of this star was first noticed on the photographic plates from Bamberg Observatory Southern Sky Survey (BOSSS) (Strohmeier, 1967). The GCVS (Samus et al., 2009) mention CF Oct as a RS CVn variable with maximal brightness V=8.27 mag and relatively large photometric variations ~0.3 mag. We have digitized and analysed the early archival observations from the BOSSS (Innis et al., 2004). Photoelectric photometry for the star was published by Innis et al. (1983, 1987), Lloyd et al. (1987) and Pollard et al. (1989). The photometry studies show rotational modulation of 20 d, due to spotted activity. CF Oct is also reported to be a strong, flaring, microwave radio source by Slee et al. (1987), and appears at the ROSAT Bright survey catalogue (Fisher et al., 1998) with 1.12 counts per second in the energy range of 0.1 - 2.4 keV.

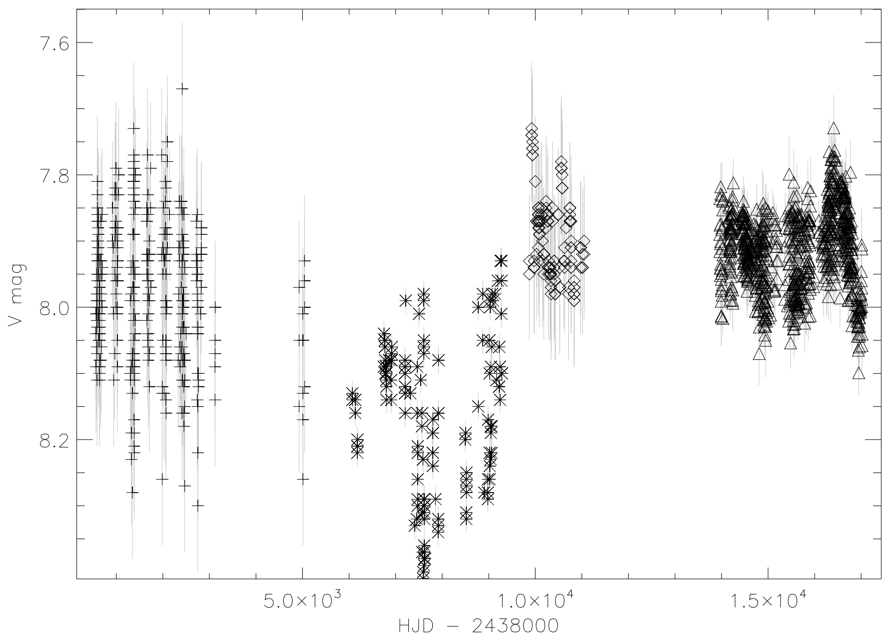


Figure 3. Light curve (with observational errors overploted) of CF Oct for the period 1964 – 2009.

The photometric data in use is collected from Bamberg Observatory Southern Sky Survey (BAM) (Innis et al., 2004), from published photoelectric photometry observations (PHOT) (Innis et al., 1983, 1987; Lloyd et al., 1987; Pollard et al., 1989), from the Hipparcos satellite time-series (HIP),[4] available via the Centre de Donnes astronomiques de Strasbourg (CDS), and from the All Sky Automated Survey (ASAS) data archive (Pojmanski, 2002) (http://www.astrouw.edu.pl/asas/).

The dataset contains data for HJD of the observation, V mag of the star, and corresponding errors. As far the data is taken from different sources, there are significant intervals with lack of data and the data distribution is non-uniform. The resulting V magnitude light curve with overploted errors is presented on Fig. 3, where the crosses represent BAM data, asterisk - PHOT data, diamonds - HIP data and triangles - ASAS data. The dataset statistically is presented in Table 1, with following information: Dataset; N_p - the Number of observations in the set; T_s - the time span of the set in days; HJD in the beginning of the set; V_{min}, V_{max} and <V> - minimal, maximal and mean values of V magnitude respectively.

Table 1. CF Oct photometric data description.

Dataset	Np	Ts	HJD	Vmin	Vmax	<V>
BAM	352	4484	2438560	7.67	8.3	7.98
PHOT	137	3212	2444071	7.93	8.41	8.16
HIP	130	1176	2447873	7.74	7.98	7.90
ASAS	705	3058	2452693	7.67	8.41	7.91

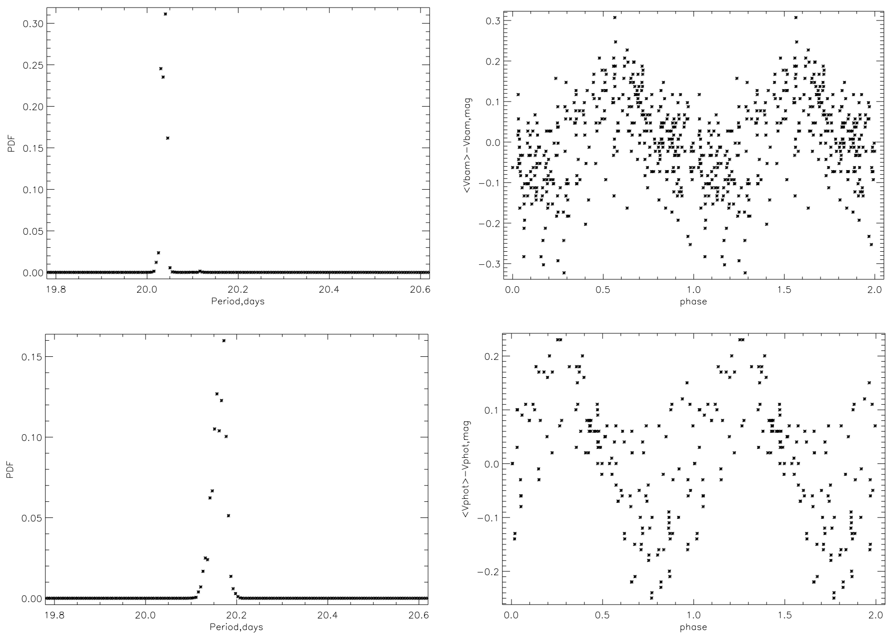
In order to reduce the likelihood of introducing unknown systematic errors into the periodicity study of CF Oct, due to the different observational methods and data reduction procedures employed, we analyze the variations of <V>-V, where <V> is the mean value of V for each dataset, i.e. <V>=(<Vbam>, <Vphot>, <Vhip>, <Vasas>).

With the collected data we are able to study variability of CF Oct in ranges from several days (this limit is set up by the average sampling frequency of our observations) to 15 years (1/3 of the covered observational time span). This is a rather large interval in frequency space, and was examined in several parts while detecting rotational or long-term variability, using a suitable number of frequencies in each case.

Previous periodical analysis (PDM and least-squares method) of photographic and photoelectric observations shows that CF Oct has a well established rotational variability with period near 20 d. Application of the Gregory-Loredo method for rotational variability study, with restricted priori range of model parameters and

[4] ESA,The Hipparcos and Tycho catalogues, ESA SP-1200, 1997.

based on all collected observations would result in more precise period estimation at different epoch. This analysis is relevant as already studied BAM and PHOT data shows slightly different periods (Innis et al., 2004), eventually due to the period estimation errors, phase shifting or more complex periodical modulation. For the separate datasets as well as for all the data combined, we have calculated the joint posterior probabilities for a class of models described by the parameters described above. The prior period P range is selected 19 to 21 days. By marginalization over the nuisance parameters we have computed the posteriori Probability Density Functions (PDF) of the number of bins and the period parameters and calculated their most probable (m_{max} and P_{max}) and mean (m_{mean} and P_{mean}) values. The PDF over the period is then normalized to have an integral over the priori range (19 to 21 d) to be 1. We have also computed the 68 per cent credible intervals (interval that contains 68 per cent of the PDF, and where the PDF is everywhere greater than the one outside the credible interval) for period detection. Table 2 represents our results: the most probable m (m_{max}), maximal probable period (P_{max}), maximal probability for the period ($Prob_{max}$), weighted mean period (P_{mean}) and the credible intervals. Number of bins parameter relates to the complexity of the light curve, to the light-curve shape and probably is connected with the structure of the stellar spots.



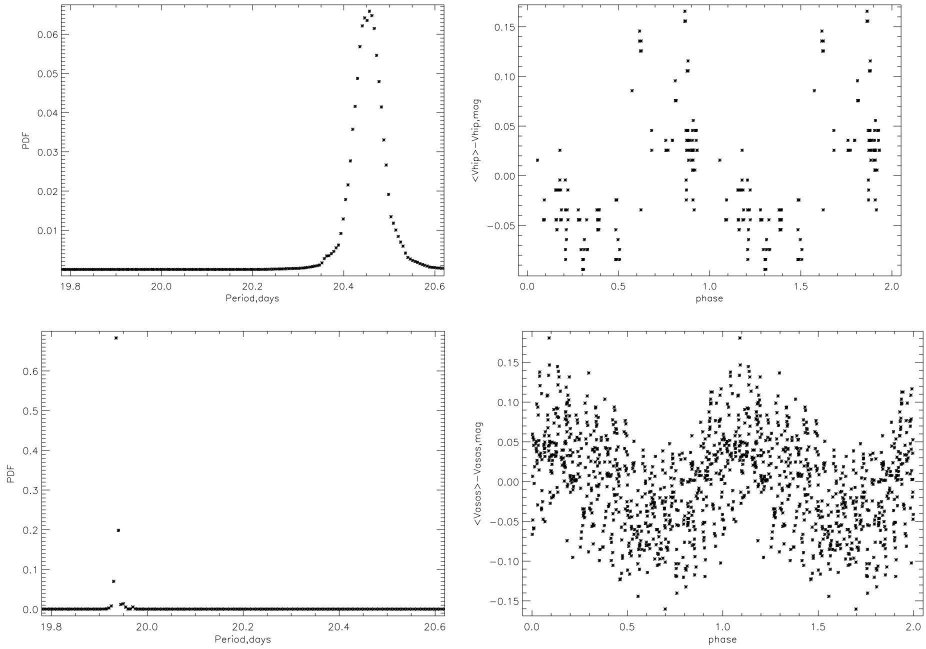


Figure 4. Normalized PDFs and light curves for the different datasets, top to bottom: BAM, PHOT, HIP and ASAS.

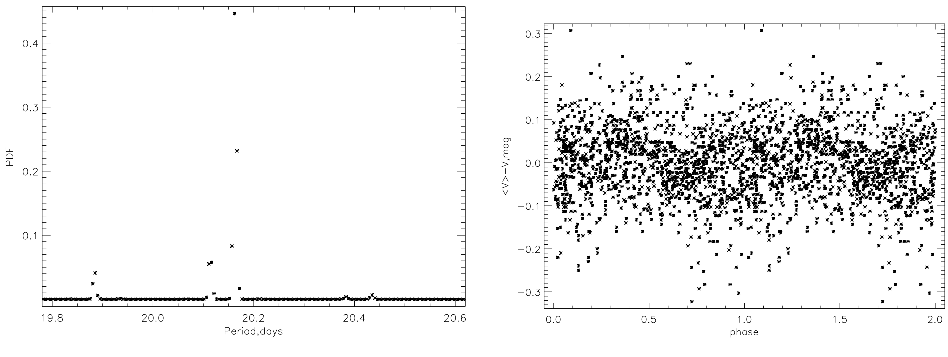


Figure 5. Normalized PDF and light curve for all the data.

Table 2: Derived periods, probabilities and 68 per cent credible intervals.

Dataset	m_{max}	P_{max}	$Prob_{max}$	P_{mean}	cre. int.
BAM	3	20.04	0.31	20.04	0.01
PHOT	2	20.17	0.16	20.16	0.03
HIP	2	20.46	0.07	20.45	0.05
ASAS	3	19.94	0.68	19.94	0.02
ALL	3	20.16	0.45	20.14	0.04

The normalized posterior PDFs for the period detection for the separate datasets as well as for all the data together are given on the left panels of Fig. 4 and Fig. 5. The mean subtracted V light curves, plotted with the most probable periods for each of the datasets and for all the data respectively, and with an epoch set at the beginning of observations at HJD=2438560.4 are presented in the right panels of Fig.4 and Fig. 5.

As it is seen from the PDFs, present Bayesian analysis confirms previous suggestions (Innis et al., 2004) for period changes. The relatively narrow credible intervals (period estimation error) result basically of the prior restriction of the parameters which is based on the previous information about variability of CF Oct and clearly demonstrates general advantage of the Bayesian statistics. Most probable periods obtained for photographic plate and photoelectric photometry data are close to the previous published ones. Scatter in the light curves from different datasets shows that the amplitude of brightness variations, light curve shape and phase of the minimal brightness change with the epoch of observation.

The GL method gives also opportunity for searching for long-term cycles with up to 6000 days length. We have evaluated the hypotheses for constant, for nonperiodic and for periodic signal, and have computed the odds ratios. The result shows that the periodic model is the best to represent the observational data and it is the most probable model. The variable model appears the second probable one and it is very reasonable since periodical models are a special case of variable ones. Searching for long-term periodic modulation over mean-subtracted, magnitude data, reveals three cycles with periods of 3582 d (~9.8 yr), 2432.5 d (~6.7 yr) and 1173 d (~3.2 yr). The marginal probabilities are 0.02, 0.008, 0.0003 and the credible intervals are 300, 150 and 20~d respectively. The normalized (PDF) plot is shown in Fig. 6. Although the shortest cycle has a very low probability and thus is statistically insignificant, the period values of the other two cycles show that there is an evidence for observation of a harmonic signal, more powerful in the longest cycle, with a period of 3583 d and with its two overtones.

6. BAYESIAN PROBABILITY THEORY AND ASTROINFORMATICS?

The astroinformatics appears at a merging area of astronomy with contemporary Information and Communication Technologies and is a consequence by the need of professional astronomers of unified access and tools for analysis of an enormous data volumes produced by multiple sky surveys. At present the Astroinformatics is a part of a common tendency for new sciences, called X-informatics (where X-refers to any scientific discipline), to be formed. The problematic of astroinformatics includes: data management and description, astronomical classification and semantics, data mining, visualization and astrostatistics.

The Bayes theory, for its logical approach can provide a base for creation of robust and practical tools for astroinformatics. The Bayes inference, especially in its objective direction is appropriate for data-mining heterogeneous data series. This approach is highly relevant in modern astronomical research if the analysis of multi-wavelength observational data obtained with various detectors is required, as it is in the astroinformatics research. Particularly the observational data series used in astronomy nowadays often suffer of random time distribution, sparse data, different quality and are usable to be analyzed by the Bayesian statistics . The possibility for updating the analysis results when, given by the Bayes theorem, with new data release and also with additional information coming from theoretical predictions and/or restrictions give the advantage of a deeper and comprehensive, data-driven researches and discoveries. Hypothesis testing in Bayesian theory can be a base for the future development of new ICT tools for updating the “state of the knowledge” in X-informatics sciences, taking in account the complexity of all the available observations, theoretical predictions and previous research experience. Subjective direction in Bayes theory could give also a base for the development of machine learning techniques in astronomy, in special interest in astronomical objects and source classification based on all the available observational data.

Acknowledgements

The authors are thankful to Prof. Ralf Newhauser for hosting the presented research at IAU Jena, Prof. Ulrich Heber for providing the images from the Bamberg observatory plate archive. This work is supported by the project DO-02-275 of the Bulgarian Science Fund as well as by the Deutsche Forschungsgemeinschaft (German Research Foundation).

References

Asensio Ramos, A.: 2009, “Evidence for Quasi-Isotropic Magnetic Fields from Hinode Quiet-Sun Observations”, *ApJ,* **701**, 1032.

Benítez, N.: 2000, “Bayesian Photometric Redshift Estimation”, ApJ, **536**, 571.

Campisi, P., Egiazarian, K. (Eds.): 2007, Blind Image Deconvolution: Theory and Applications, CRC Press Taylor & Francis Group.

Dale, A. I.: 2003, Most honourable rememberance: The life and work of Thomas Bayes, Springer.

Daston, L.: 1994, “How Probabilities Came to be Objective and Subjective”, *Historia Mathematica*, **21**, 330.

Encyclopædia Britannica: 2010, “Thomas Bayes”, Encyclopædia Britannica Online, http://www.britannica.com/EBchecked/topic/56807/Thomas-Bayes, 20 Aug. 2010

Feigelson, E., Babu, G., (Eds.): 1997, Statistical Challenges in Modern Astronomy II”, Springer.

Feigelson, E., Babu, G., (Eds.): 2010, Statistical Challenges in Astronomy, Springer.

Fienberg, S. E: 2006,, When did Bayesian Inference become "Bayesian"?, Bayesian Analysis, 1 (1), 1–40.

Fischer, J.-U., et al. : 1998, “The ROSAT Bright Survey. I. Identification of an AGN sample with hard ROSAT X-ray spectra”, *AN*, **319**, 347.

Gregory, P. C., Loredo, T. J.: 1996, “Bayesian Periodic Signal Detection: Analysis of ROSAT Observations of PSR 0540-693”, *ApJ*, **473**, 1059.

Gregory P. C.: 1999, “Bayesian Periodic Signal Detection. I. Analysis of 20 Years of Radio Flux Measurements of the X-Ray Binary LS I +61 deg303”, *ApJ*, **520**, 361.

Gregory P. C.: 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support, Cambridge University Press, New York.

Hobson, Michael P., Jaffe, Andrew H.: 2010, Bayesian Methods in Cosmology, Cambridge University press.

Innis, J. L., Coates, D. W., Dieters, S. W. B., Moon, T. T., Thompson, K., B.: 1983, “Photometry of the Southern RS CVn Candidate HD 196818”, IBVS, No. 2386.

Innis, J. L., Coates, D. W., Thompson K.: 1997, “Observations of active-chromosphere stars - IV. Photometry and spectroscopy of CF Octantis (HD 196818)”, *MNRAS*, **289**, 515.

Innis, J. L., Borisova, A. P., Coates, D. W., Tsvetkov, M. K.: 2004, “Archival light curves from the Bamberg Sky Patrol: CF Octantis, 1964-76”, *MNRAS*, **355**, 591.

Jaynes, E. T.: 2003, Probability Theory: The Logic of Science, Cambridge University Press.

Jeffreys, H.: 1939, Theory of probability, 1st ed. Oxford: The Clarendon Press.

Kitaura, F. S. and Enßlin, T. A.: 2008, “Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization”, *MNRAS*, **389**, 497–544.

Knuth, Kevin H., Caticha, A., Center, Julian L., and Giffin, A. (Eds.): 2007, “Bayesian Inference in Science and Engineering”, 27th International Workshop on Bayesian Inference and Maximum Entropy Methods, *AIP Conference Proceedings*, **954**.

Lloyd, Evans T., Koen, M. C. G.: 1987, “UBV(RI)c photometry for CaII emission stars. I - Observations at Sutherland”, *SAAO Circ*, **11**, 21.

Neyman, J.: 1977, "Frequentist probability and frequentist statistics", *Synthese*, **36 (1)**, 97–131.

Pojmanski, G.: 2002, *Acta Astronomica*, **52**, 397.

Pollard, M. J., Hearnshaw, J. B., Gilmore, A. C., Kilmartin P. M.: 1989, *JAp&A*, **10**, 139.

Puglielli, D.: 2010, Galaxy Modeling using Bayesian Statistics: A Bayesian/Markov chain Monte Carlo Approach to Modeling NGC 6503, LAP Lambert Academic Publishing.

Samus, N. N., Durlevich, O. V. et al.: 2009, GCVS database, Version 2009, VizieR Online Data Catalog, http://vizier.cfa.harvard.edu/viz-bin/VizieR?-source=B/gcvs.

Slee, O. B., Nelson,G. J.,Stewart, R. T., Wright, A. E.,Innis, J. L., Ryan, S. G., Vaughan, A. E.: 1987, "A microwave survey of southern active stars", *MNRAS*, **229**, 659.

Starck, J.-L., Murtagh, F.: 2006, "Astronomical Image and Data Analysis", Springer.

Stigler, S., M.: 1986, The history of statistics, Harvard University press.

Strohmeier, W.: 1967, "Bright Southern BV-Stars", IBVS, No. 178.