

## STATISTICAL DISTRIBUTION OF SOME ASTEROID PARAMETERS

Mihail-Ioan POP <sup>1</sup>

### Abstract

Probability distributions of several asteroid parameters were obtained by using a Maximum Entropy Method combined with a form of the Simulated Annealing heuristic search algorithm. The parameters of interest were asteroid effective diameter, mass, density, albedo, magnitude and colour indices. The obtained distributions reflect the main asteroid classes in order of occurrence. Moreover, the complex composition and the nature of asteroid surfaces become apparent.

2010 *PACS*: 96.30.Ys, 89.70.Cf, 02.60.Pn, 02.50.Fz.

*Key words*: asteroid, maximum entropy method, simulated annealing, kernel density estimation.

## 1 Introduction

Asteroids are Solar System bodies much smaller than planets which typically orbit the Sun between the orbits of planets Mars and Jupiter. They have formed at the same time as planets, but because of gravitational perturbations induced by Jupiter they failed to accrete in a planet. Nevertheless, they occupy a position that was predicted by the empirical Titius-Bode law which works for planets. The biggest asteroids are Ceres (952 km diameter), Pallas (544 km) and Vesta (529 km). The majority of asteroids are far smaller. There is a corresponding size and mass distribution of asteroids.

Most asteroids have a mixed composition of silicates, metals and organic substances (class C asteroids); there are also asteroids composed mostly of silicates (class S) and a smaller fraction whose composition is dominated by metals, mostly an iron-nickel alloy (class M). Other asteroid classes have been described, but they are less prevalent. The composition of asteroid surfaces determines their electromagnetic reflectivity (albedo) at different wavelengths and their magnitude in visible light. Their composition also determines the mass density.

The statistical distributions of some asteroid parameters have been determined. The raw values were retrieved from NASA's Planetary Data System [10]. The parameters of interest were effective diameter, mass, density, albedo and colour indices. In order to determine their distribution a Maximum Entropy Method was used in combination with a form of the Simulated Annealing heuristic search algorithm, as described below.

---

<sup>1</sup>Faculty of Technological Engineering and Industrial Management, Physics Department, *Transilvania* University of Braşov, Romania, e-mail: mihailp@unitbv.ro

## 2 Computational procedure

Generally, each parameter  $X$  presented as a set of  $N$  raw data  $X_i$ ,  $i = \overline{1, N}$ , spanning an interval  $X_i \in [X_{min}, X_{max}]$ . There is an underlying probability density function  $\rho(x)$  of parameter  $X$  which is assumed continuous and is to be determined. For this, two sets of equidistant points were generated in the interval of  $X$ :  $x_j$ ,  $j = \overline{1, N_p}$  with norm  $\delta = x_j - x_{j-1}$  and  $c_k$ ,  $k = \overline{1, N_c}$  with norm  $\gamma = c_j - c_{j-1}$  respectively. A discrete distribution of probabilities  $p_j$  was assigned to intervals  $[x_j - \delta/2; x_j + \delta/2)$ ,  $j = \overline{1, N_p}$ . Once probabilities  $p_j$  are determined, the probability density function can be approximated as  $\rho(x_j) \approx p_j/\delta$ .

In order to determine probabilities  $p_j$  the Maximum Entropy Method [6, 17] was used. A cost function was built with two types of statistical measures. First, the Shannon entropy was used:

$$H = - \sum_{j=1}^{N_p} p_j \ln p_j. \quad (1)$$

Second, a set of conditions were built with the Gauss "bell" function centered at points  $c_k$ :

$$f_k(x) = \exp\left(-\frac{(x - c_k)^2}{2\sigma^2}\right), k = \overline{1, N_c}. \quad (2)$$

The average of each function was computed in two forms: an empirical average

$$\langle f_k \rangle_{emp} = \frac{1}{N} \sum_{i=1}^N f_k(X_i) \quad (3)$$

and a simulated average

$$\langle f_k \rangle_{sim} = \frac{1}{N_p} \sum_{j=1}^{N_p} p_j f_k(x_j). \quad (4)$$

The distribution of  $X$  must obey the conditions  $\langle f_k \rangle_{emp} = \langle f_k \rangle_{sim}$ ,  $k = \overline{1, N_c}$ . Since the equality is generally not obeyed, a supplementary condition is imposed:  $H = \max$ . The conditions are summed up in the cost function:

$$E_H = \sum_{k=1}^{N_c} (\langle f_k \rangle_{emp} - \langle f_k \rangle_{sim})^2 - k_H \cdot H, \quad (5)$$

where  $k_H$  is a constant. Thus, the above conditions become  $E_H = \min$ . and the distribution determination becomes an optimisation problem. In order to solve it, the Simulated Annealing algorithm [5, 7] was used with a temperature parameter  $T = 0$ .

The conditions used are inspired by the Kernel Density Estimation method [12, 13, 18]. They offer a quasi-local description of the probability density function and, as a consequence, suffer from the main problem of all local estimation methods (*e.g.* histogram estimation): the quality of estimation is directly proportional to the number of raw data used in the local estimation for each point. Too few raw data induce errors in the empirical

conditions, which the method cannot overcome. Nevertheless, the method's nonlocality can be controlled by parameter  $\sigma$ . Higher values of  $\sigma$  reduce the sample error, but instead make the numerical computation of  $\rho$  more time-demanding.

In order to check the goodness of fit, a relative error was computed on the conditions:

$$\varepsilon_k = \frac{|\langle f_k \rangle_{emp} - \langle f_k \rangle_{sim}|}{\langle f_k \rangle_{emp}}, k = \overline{1, N_c}. \quad (6)$$

It can be seen that this is a function of position:  $\varepsilon_k = \varepsilon(x_k)$ . This is very convenient, as it allows the estimation of the quality of the distribution at each point in the interval of  $X$ . Generally, it is good to have a small relative error, typically  $\varepsilon < 0.01$ . Nevertheless, higher values of  $\varepsilon$  are acceptable on portions where  $\rho(x)$  is small relative to a maximum value. This is because very small values of  $\rho$  correspond to a very small number of raw values in the respective area, usually lowering the quality of  $\rho$ . Thus a determination of the distribution up to a multiplicative factor of the order of several units when the distribution is extremely small is not usually significant. In practice, high values of  $\varepsilon$  are still smaller than 1.

In all cases we used  $N_p = 1000$  interval points and  $N_c = 101$  conditions; also,  $\sigma = (X_{min} - X_{max})/30$ , which was found good enough in tests carried on some known distributions. The parameter  $k_H$  was used to control the smoothness of the obtained distribution. Generally, it was set at some successively decreasing values  $k_H = 10^{-2}$ ,  $k_H = 10^{-4}$ ,  $k_H = 10^{-6}$ . The number  $N$  of raw values was different for each parameter  $X$ . The obtained results are presented below. Also, the number  $N$  of raw values used in each case is given in the figures.

### 3 Asteroid size and mass

The effective diameter for several asteroids was previously obtained from their infrared magnitude measured from space during the IRAS mission [14]. These values were used to compute a MEM approximation for the diameter distribution. It was shown to have a peak at  $D \approx 25$  km below which the distribution becomes uncertain, as shown by the relative error. This may be a consequence of the fact that the instruments were unable to detect well asteroids with smaller diameter. Above the peak there is a smooth decrease which can be approximated for  $D > 40$  km with an exponential function of the form:

$$\rho(D) \approx \frac{1}{39 \text{ km}} \exp\left(-\frac{D}{41.4 \text{ km}}\right). \quad (7)$$

From peak to 40 km the distribution decreases with  $D$  faster than this exponential. The distribution of mass computed with data from [1] is supported on two power law functions (two lines in log-log representation). The jumps of the distribution between the two power laws are abrupt. Most of the asteroids have low mass.

The distribution of the asteroid density computed with values from [1] has some peaks at  $\approx 1.3, 2.1, 2.7, 3.4$  g/cm<sup>3</sup> and also at  $\approx 5.6$  and  $7$  g/cm<sup>3</sup>, characterising different asteroid classes. The mass and density distributions are not well determined because of the small number of values used.

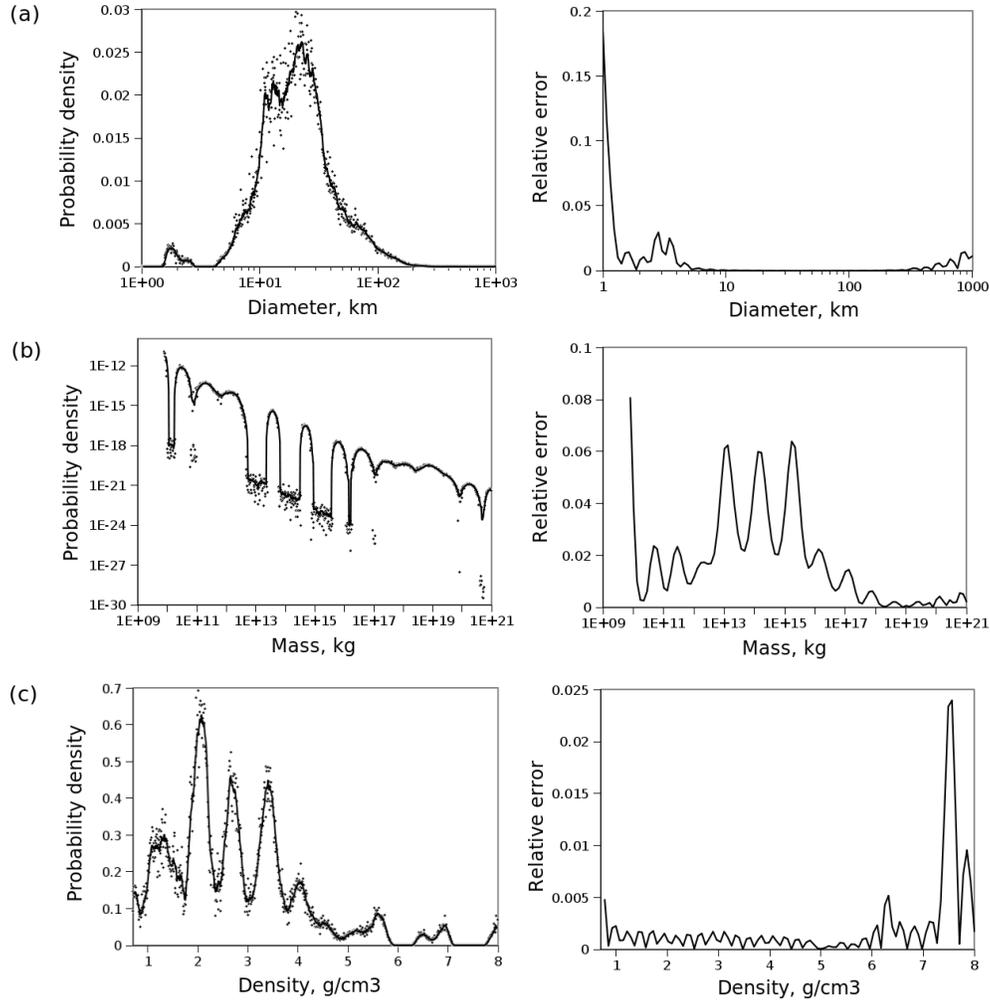


Figure 1: Distributions and associated relative errors of conditions for (a) asteroid diameter ( $N = 2\,228$  measurements used), (b) mass ( $N = 144$ ) and (c) mass density ( $N = 144$ ).

## 4 Asteroid albedo

The albedo of asteroids was measured with a few methods: polarised visible light (data collected from surface telescopes) [3, 8], reflected infrared radiation (data collected during the IRAS space mission) [14] and with the help of radar (microwave radiation, range from 2.38 GHz to 8.57 GHz - surface radars were used) [3, 11]. Different albedo values have been obtained for the same asteroid because radiation is absorbed and reflected as a function of frequency. Albedo data give indications about the nature and composition of asteroid surfaces. Especially significant are data obtained in the microwave spectrum, where metals are particularly reflective. Thus, the microwave albedo can be considered as a measure of the metal composition near the asteroid surface. From an economical point of view,

metal - rich asteroids are important for future space exploitation. Beside iron and nickel, they also contain heavy metals, many of them being scarce on Earth. Polarised - light

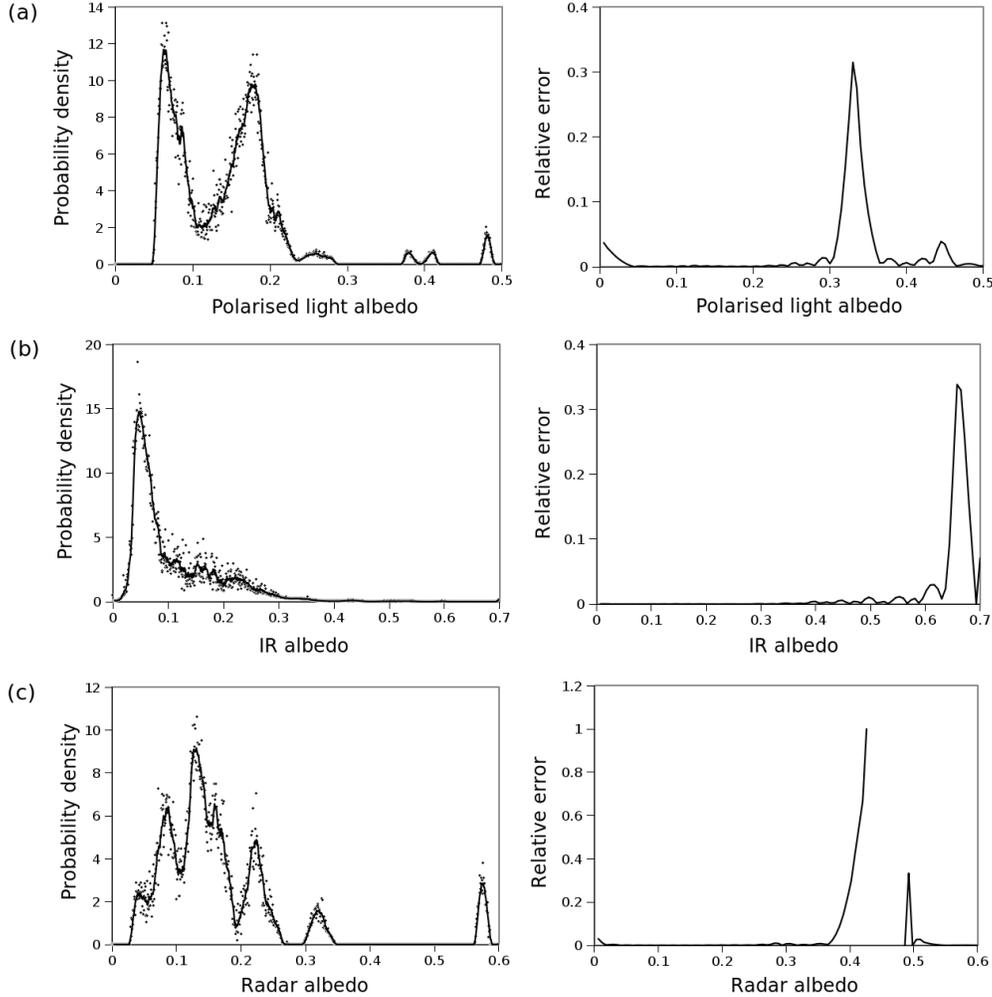


Figure 2: Distributions and relative errors of conditions for asteroid albedos in (a) polarised visible light ( $N = 127$ ), (b) reflected infrared ( $N = 2003$ ) and (c) microwave radiation ( $N = 47$ ).

albedos have two distribution maxima at  $A \approx 0.07$ , corresponding to class C asteroids and respectively  $A \approx 0.18$ , probably corresponding to class S asteroids. The infrared albedos distribution has a maximum at  $A \approx 0.05$ , followed by an exponential decrease. This maximum is probably characteristic to class C asteroids, showing that they form a majority among all asteroids.

The radar albedo distribution is quite complex, with maxima at  $A \approx 0.08$ ,  $0.12$ ,  $0.22$ ,  $0.32$  and  $0.58$ . These may show different metal composition for different asteroid classes. The number of data used here was small, thus the determined distribution is not very

precise. The first maxima accept a smooth concave envelope with a maximum at  $A \approx 0.12$ . It may be that, with more data, the distribution reduces to this envelope.

## 5 Asteroid magnitude and colour indices

Astronomical magnitudes represent a measure of the light flux coming from a body. They are measured on a logarithmic scale, with lower magnitudes representing higher observed flux. The magnitude of asteroids is usually reduced to an absolute magnitude computed for Sun - asteroid and observer - asteroid distances of 1 AU. The standard magnitude used  $M_V$  is measured in the visible spectrum around 550 nm wavelength. Other magnitudes are measured at different wavelengths. They are normally computed relative to a given wavelength magnitude as the difference from that magnitude, yielding colour indices. The most used are indices U-B (difference between magnitudes in ultraviolet and blue spectra) and B-V (blue - green colour index). Colour indices characterise asteroid surface absorbance at different wavelengths and thus the surface nature and composition.

The obtained distribution for asteroid magnitudes with data from [16] is smooth, with a maximum at  $M \approx 15$ . The existence of a maximum is a consequence of the instruments' inability to detect faint asteroids. The ascending portion of the distribution can be approximated with an exponential law. Since magnitude is inversely proportional to asteroid size, it follows that asteroid diameters are distributed exponentially up to a limit.

The colour index distributions obtained with data from [15] have different shapes. The U-B distribution has an almost constant portion with a few peaks between 0.25 and 0.5, while the B-V distribution has two sharp peaks. The asteroid reflectance is higher at higher wavelengths. It follows that asteroid surfaces absorb light preferentially at smaller wavelengths (blue - ultraviolet spectrum). The absorbance is a complicated function of wavelength, as shown by the two distributions. This is an indication of complex chemical substances on the surface of asteroids, probably silicates, carbon and organic substances. Some of them originated with the asteroids, but others may have appeared later under the influence of heat and solar and cosmic radiations.

## 6 Conclusions

Distributions of asteroid parameters have been obtained with a good approximation with respect to the quantity of raw data used. The results show a complex picture of asteroid distributions and composition. The main asteroid classes are reflected in the distributions of mass density and albedos. The radar albedo indicates an average metal composition of around 15 %, which is quite much. Considering that asteroid metals often appear in native or alloyed forms, their extraction needing no chemical reduction processes, this indicates a huge economical potential of asteroids on average. There is even an indication in the radar distribution showing a metal composition approaching 60 %, corresponding to class M asteroids, which would be the first target for extractive space industry. The density and colour indices indicate a complex chemical composition

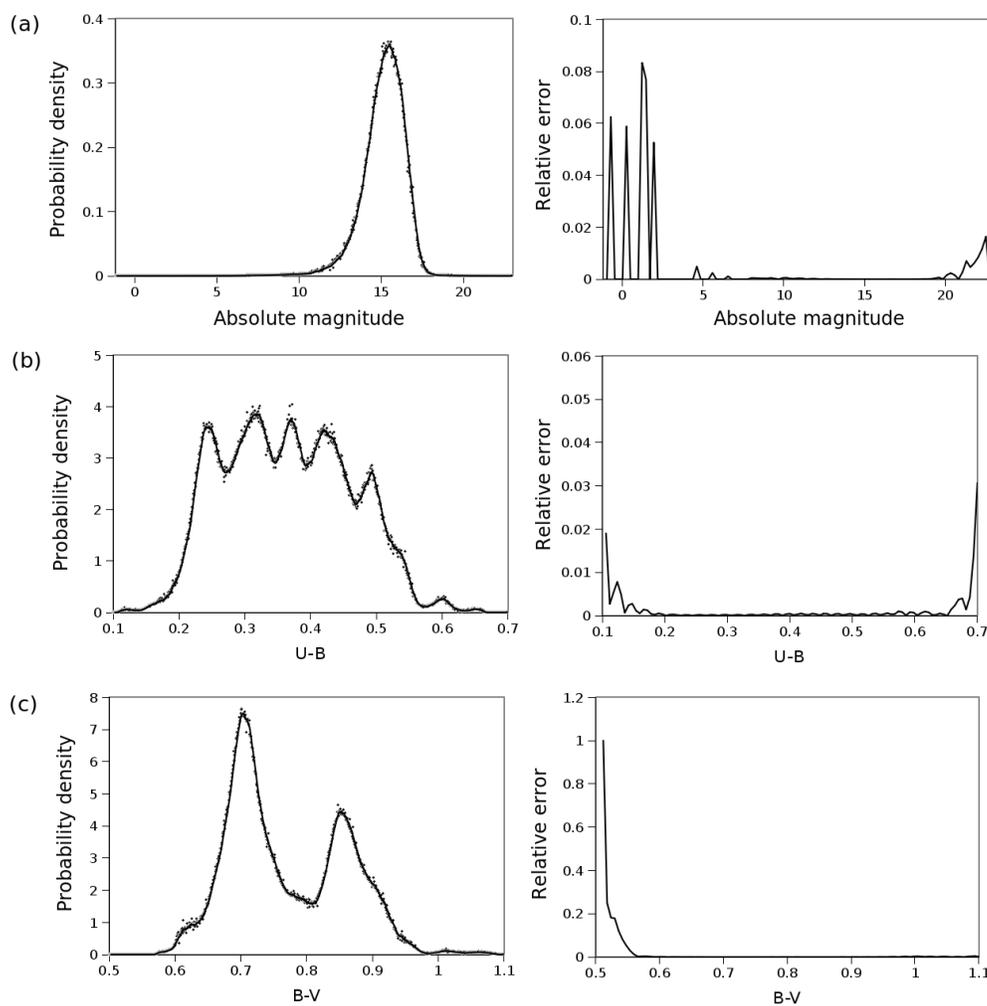


Figure 3: Distributions and relative errors of conditions for (a) absolute magnitude  $M_V$  ( $N = 185\,655$ ), (b) U-B ( $N = 978$ ) and (c) B-V ( $N = 1\,020$ ) colour indices.

of asteroids, partly representative of the environment at the beginning of the Solar System and partly resulting from chemical reactions mediated by the Sun's heat and ultraviolet radiation.

## References

- [1] Baer, J., Chesley, S., and Britt, D., Eds., *Asteroid Masses*, V1.0. EAR-A-COMPIL-5-ASTMASS-V1.0. NASA Planetary Data System, 2009.
- [2] Bus, S.J., Binzel, R.P., *Phase II of the Small Main-Belt Asteroid Spectroscopic Survey*:

- A Feature-Based Taxonomy*, Icarus **158** (2002), 146177.
- [3] Davis, D.R. and Neese, C., Eds., *Asteroid Albedos*, EAR-A-5-DDR-ALBEDOS-V1.1. NASA Planetary Data System, 2002.
- [4] Folescu, C., *Ce este universul?*, Albatros, București, 1988.
- [5] Ingber, L., *Simulated annealing: Practice versus theory*, Mathematical and Computer Modelling, **18** (1993), 29-57.
- [6] Jaynes, E. T., *Information Theory and Statistical Mechanics*, in Statistical Physics, K. Ford (ed.), Benjamin, New York, 181, 1963.
- [7] Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., *Optimization by simulated annealing*, Science, **220** (1983), 671-680.
- [8] Lupishko, D.F. and Vasilyev, S.V., Eds., *Asteroid Polarimetric Database V6.0*, EAR-A-3-RDR-APD-POLARIMETRY-V6.0. NASA Planetary Data System, 2008.
- [9] Mueller, M., Harris, A. W., Fitzsimmons, A., *Size, albedo, and taxonomic type of potential spacecraft target*, Asteroid (10302) 1989 ML, Icarus **187** (2007) 611615.
- [10] NASA PDS Asteroid/Dust Archive <http://sbn.psi.edu/pds/> .
- [11] Neese, C., Benner, L.A.M., and Ostro, S.J., Eds., *Asteroid Radar V16.0*, EAR-A-5-DDR-RADAR-V16.0. NASA Planetary Data System, 2010.
- [12] Parzen, E., *On Estimation of a Probability Density Function and Mode*, Ann. Math. Statist., **33** (1962), 1065–1076.
- [13] Rosenblatt, M., *Remarks on Some Nonparametric Estimates of a Density Function*, Ann. Math. Statist., **27** (1956), 832–837.
- [14] Tedesco, E.F., P.V. Noah, M. Noah, and S.D. Price. *IRAS Minor Planet Survey*, IRAS-A-FPA-3-RDR-IMPS-V6.0. NASA Planetary Data System, 2004.
- [15] Tedesco, E.F, Eds., *UBV Mean Asteroid Colors*, EAR-A-5-DDR-UBV-MEAN-VALUES-V1.2. NASA Planetary Data System, 1995.
- [16] Tholen, D.J., Eds., *Asteroid Absolute Magnitudes V12.0*, EAR-A-5-DDR-ASTERMAG-V12.0. NASA Planetary Data System, 2009.
- [17] Uffink, J., *Can the Maximum Entropy Principle be explained as a consistency requirement?*, Studies in History and Philosophy of Modern Physics **26B** (1995), 223–261.
- [18] Wu, W. B., Mielniczuk, J., *Kernel density estimation for linear processes*, Ann. Statist., **30** (2002), 1441–1459.