ASTRONOMICAL INSTITUTE
SLOVAK ACADEMY OF SCIENCES

CONTRIBUTIONS OF THE ASTRONOMICAL OBSERVATORY SKALNATÉ PLESO

• VOLUME LIII •

Number 1

February 2023

ISSN: 1336–0337 (on-line version)

CODEN: CAOPF8

Editorial Office: Astronomical Institute of the Slovak Academy of Sciences
SK - 059 60 Tatranská Lomnica, The Slovak Republic

CONTENTS

EDITORIAL

STARS

GENERAL ASTRONOMY

CELESTIAL MECHANICS

The Contributions of the Astronomical Observatory Skalnaté Pleso
are available in a full version
in the frame of ADS Abstract Service
and can be downloaded in a usual way from the URL address:

https://ui.adsabs.harvard.edu/

as well as from the web-site of
the Astronomical Institute of the Slovak Academy of Sciences
on the URL address:

https://www.astro.sk/caosp/caosp.php

The journal is covered/indexed by:

Web of Science (WoS)
WoS Core Collection: Science Citation Index Expanded

SCOPUS

Index Copernicus International

EDITORIAL

In 2022, the journal Contributions of the Astronomical Observatory Skalnaté Pleso (CAOSP) published 19 articles on 253 pages within three issues. According to the *Journal Citation Reports* database, the articles published in 2021 received 212 citations corresponding to the journal impact factor of 0.435.

Volume 52/1 publishes three regular papers from stellar astrophysics. The first contribution is devoted to the analysis of silicon abundances for a sample of 120 main-sequence stars in the temperature range of 7000-14000 K including also non-magnetic chemically peculiar stars. The second article deals with the possibility of determining the distance of selected open star clusters, provided that the absolute magnitudes of all stars are scattered around a mean value. The third article presents the light and color curves of superluminous SN 2017egm in NGC3191, and of peculiar object AT 2018cow. The authors determined their basic parameters.

Volume 52/2 presents four regular papers, three from stellar astrophysics and one from interplanetary matter. The study of the binary TU UMi (delta Sct variable) presents the orbital period analysis suggesting the presence of a third body or a magnetic activity cycle in the system. The second paper deals with rapidly rotating white dwarfs LAMOST J024048.51+195226.9 and CTCV J2056-3014, while in the third paper, the authors study object PSR J0740+6620 that belongs to the most massive known neutron stars, discovered in 2019. In the fourth article, the authors present multi-color light curves of asteroids whose orbits cross the orbit of Mars and specified their characteristics.

Volume 52/3 introduces twelve selected contributions based on lectures presented at IV Meeting on Astrophysical Spectroscopy: A&M DATA - Atmosphere, held in Andrevlje, Serbia, during May 30 - June 2, 2022. Proceedings of the conference was edited by Vladimir A. Srećković, Milan S. Dimitrijević, and Nikola Cvetanović.

Just a few cosmetic technical changes have been introduced this year: the resulting PDF of a manuscript can now contain hypertext references (both internal and external; e.g. link to author's ORCiD, links to references, etc.).

Tatranská Lomnica, January 19, 2023

Augustín Skopal, Editor-in-Chief

Richard Komžík, Managing Editor

Contrib. Astron. Obs. Skalnaté Pleso **53/1**, 7 – 16, (2023)
https://doi.org/10.31577/caosp.2023.53.1.7

Stochastic nonlinear self-oscillatory model of an accretion disk for the X-ray bursting of the microquasar GRS 1915+105

P.-J. Chen and Z.-Y. Wang

School of Physics and Electronic Engineering, Hubei University of Arts and Science, Xiangyang 441053, China, (E-mail: wangzhiyun@hbuas.edu.cn)

Received: July 4, 2022; Accepted: November 10, 2022

Abstract. The observed light curves of the microquasar GRS 1915+105 show three main characteristics, namely chaotic, stochastic and semi-stochastic (or nonstochastic). In order to probe the origin causing these different features, we propose a stochastic nonlinear self-oscillatory model of an accretion disk, and use its oscillating luminosities with different noise strengths and nonlinear factors to simulate the observational variability, and compare their correlation dimensions calculated by a nonlinear time series analysis. It is shown that when the stochastic noise strength is small and the nonlinear factor is large, the luminosity curves show chaotic behavior. On the contrary, it shows stochastic or semi-stochastic behavior when a stochastic factor dominates. The simulated data exhibits the same nonlinear and random characteristics as the observational sequences, and their corresponding correlation dimensions are also consistent. This indicates that the X-ray bursting of the microquasar should be related to a nonlinear self-oscillation of the accretion disk.

Key words: accretion disk – self-oscillation – X-ray bursting

1. Introduction

It is well known that the microquasar GRS 1915+105 is a prominent black hole system exhibiting a large variable X-ray emission on different time scales and patterns (Belloni *et al.* 1997; Chen *et al.*1997; Paul *et al.* 1997). Based on a large collection of multi-epoch RXTE observations, Belloni *et al.* (2000) classified its observed light curves into 12 temporal states, of which four classes show chaos, three are random, while five others show a deviation from being random (Mukhopadhyay 2004; Harikrishnan 2010). That is, the 12 temporal classes could be grouped into three main types, namely chaotic behavior, stochastic behavior and semi-stochastic (or nonstochastic) behavior (Misra *et al.* 2004, 2006), respectively. Many studies suggest that the complexities of light curves may be caused by the noise fluctuations in a chaotic system involving nonlinear processes, whose evolution can be described by several ordinary differential equations. For example, Massaro *et al.* proposed two nonlinear mathematical models, Fitzhugh-Nagumo (FhN) and Modified Hindmarsh-Rose (MHR) models, and used their solutions to reproduce several different classes of light curves of GRS 1915+105 (Massaro *et al.* 2014, 2020 a, 2020 b, 2020 c). The MHR model described well the transition from stable to bursting states and the occurrence of low-frequency quasi periodic oscillations as a consequence of a transition from an unstable to a stable equilibrium. However, the physical explanation of these nonlinear differential equations is not completely understood.

As an important theoretical model, the oscillations of an accretion disk can interpret a large number of the observational features of black hole binary systems (Kato 2016). In the last decade, the stochastic oscillation of an accretion disk has attracted great attention in astrophysical studies, and has succeeded in explaining some observations. For example, a generalized Langevin equation with white or color noise has been used to describe the vertical oscillations of accretion disks, and the luminosity of stochastic oscillations could reproduce the observations of optical intra-day variability of BLLac objects ((Harko*et al.* 2012, 2014; Leung *et al.* 2011, 2014; Long *et al.* 2016). Based on their model, we have discussed the stochastic resonance phenomenon in power spectral density curves, and considered that the resonance peak is an alternative explanation of the observed low-frequency quasi periodic oscillations (Wang *et al.* 2013 a, 2013 b, 2015, 2016). However, the light curves of this stochastic oscillating model only show stochastic behavior, not present periodicity and chaotic behavior. So Ou *et al.* (2014) developed a chaotic+stochastic oscillating accretion disk model, in which the chaotic factor was introduced into the stochastic oscillating disk. The stable response of this model is completely controlled by the external random and chaotic driving of the system, so it could simulate the light curves showing chaotic behavior of variability, and their correlated dimensions are consistent with that of the observational data.

The oscillations of an accretion disk surrounding a compact object are generally driven by magneto-hydrodynamic turbulence which is an intrinsically nonlinear process. The model for this process should be nonlinear and is expected to show qualitative changes in its behavior as a control parameter is varied. So in this paper, we improve the stochastic oscillation model mentioned above, and suggest a stochastic nonlinear self-oscillatory model of an accretion disk to simulate the light curves of GRS 1915+105. In contrast with linear stochastic oscillation model, self-oscillations of a constrained cycle can generate and maintain a regular mechanical periodicity without requiring a similar external periodicity to drive it (Jenkins 2013). Such constrained cycle oscillations usually occur in non-conservative dissipative systems and have been considered in the dynamics of Cepheids and RR Lyrae stars(Rudd, Rosenberg 1970; Buchler, Kovacs 1987; Das *et al.* 1996). Nakariakov *et al.* (2016) use a limit cycle solution corresponding to the self-oscillations to interpret the undamped quasi-harmonic kink oscillations of coronal loops, and successfully reproduce their observed properties. On some occasions, the complex variability of GRS 1915+105 is characterized by a long series of bursts expected by limit cycles (Castro-Tirado *et al.* 1992). In addition, nonlinear effects in the model may lead to overstable pseudo-periodic, or even chaotic motions. So we will calculate the dissipated energy of this self-oscillating system to simulate the X-ray variability classes of GRS 1915+105, and probe the intrinsic relation between the simulated data with observational variability by comparing their correlated dimensions.

2. Stochastic nonlinear self-oscillatory model of accretion disk

For a thin accretion disk around a black hole in contact with an isotropic and homogeneous external heat bath, the vertical oscillation of unit mass can be described by a generalized Langevin equation, given by Leung *et al.*(2011)

$$\frac{d^2z}{dt^2}+\mu\frac{dz}{dt}+\omega_0^2z=\frac{dF(t)}{M_D}, \tag{1}$$

where z and M_D are the displacement of oscillation and the mass of the accretion disk, respectively. μ expresses a viscosity coefficient and $dF(t)$ is a stochastic excitation. The intrinsic angular frequency ω_0^2 can be written as (Titarchuk, Osherovich 2000)

$$\omega_0=\frac{4.4\pi\times10^3}{m}\{\frac{2-\gamma}{r_{in}^3}\frac{[1-\gamma/((\gamma+1)r_{adj})]}{r_{adj}^{\gamma}r_{out}^{2-\gamma}}\}^{1/2}Hz, \tag{2}$$

where, $m=M_0/M_{\odot}$, M_0 and $M_{\odot}$ are the masses of the central object and the sun. $r_{in}=R_{in}/3R_S$ is the innermost radius of the disk and $R_S=2GM_0/c^2$ is the Schwarzschild radius. $r_{adj}=R_{adj}/R_{in}$ and $r_{out}=R_{out}/R_{in}$ are the adjustment radius and the outer radius of the disk, respectively. The index γ is either 3/5 or 3/4 (Shakura, Sunyaev 1973).

The accretion disk as a whole is subjected to viscous forces mainly from friction between the disk and the external heat bath. If the interaction of the disk with the surrounding moving medium is not stiff but slippery, this friction is a possible way to compensate for the dissipation losses (Nakariakov *et al.* 2016). Further, if we incorporate a cubic nonlinearity into the dissipation, then the damping term of the system may be expressed as $-\mu_0\frac{dz}{dt}+\mu_0az^2\frac{dz}{dt}$ (μ_0 and a are positive constants). The first term is the linear friction force, by which energy is transferred from the external heat bath to the disk, and the second one is the nonlinear dissipation force resulting in the energy loss of the system.

In addition, if we consider that the self-gravitational potential V_z of the disk is $\frac{1}{2}\omega_0^2z^2+\frac{1}{4}\omega_0^2bz^4$, then the nonlinear gravitational term of a disk oscillation may be represented as $\omega_0^2(z+bz^3)$. We also assume random excitation $F(t)=DdW(t)$, where D refers to the noise strength and $dW(t)$, defining the normalized δ-correlated Gaussian noise with the zero mean and unit variance, may be represented by the formal derivative of a Wiener process, $W(t)$. Then we replace the equation (1) of the disk motion with a stochastic nonlinear differential equation, which can be written as

$$\frac{d^2z}{dt^2}-\mu_0(1-az^2)\frac{dz}{dt}+\omega_0^2(z+bz^3)=DdW(t). \tag{3}$$

Self-oscillatory motions described by equation (3) are essentially dissipative, with the dissipative losses compensated by the continuous energy supply of the external heat bath and the radiation luminosity of the disk, representing the energy lost by the disk due to viscous dissipation and the presence of the random force, is given by:

$$L(t)=\mu_0 az^2(\frac{dz}{dt})^2-DdW(t)\frac{dz}{dt}. \tag{4}$$

Obviously, because equation (3) contains nonlinear terms and a random term, the luminosity variables can show different behavior, which should be similar to different temporal states of GRS 1915+105.

3. Light curves of observation sequences and simulated data

We show in this section that the numerical solutions of the nonlinear oscillation system of equations (3) and (4) reproduce the main features of three main variability types of GRS 1915+105. Based on the PUBLIC in the RXTE TOO archive (at heasarc.gsfc.nasa.gov), here we choose three representative observation sequences whose observation IDs are 20402-01-02-02, 10408-01-40-00 and 10408-01-08-00, respectively. Their continuous light curves in the energy range 2-30keV binned at $1s$ intervals are plotted in Figures 1(a),(c) and(e), which show the stochastic behavior, semi-stochastic behavior and chaotic behavior, respectively.

In order to simulate the light curves of three observational sequences, we perform numerical computations of equations (3) and (4) by means of a Runge-Kutta fourth order integration routine, and obtain luminosities with different parameters. For GRS1915+105, we set the same disk parameters as Titarchuk and Osherovich (2000), $r_{in}=1$, $r_{out}=10^4$ and $\gamma=3/5$, $r_{in}=3$, and adopted $m=14$ (the mass range of this source is from $10M_\odot$ to $18M_\odot$ as estimated by Remillard and McClintock (2006)). Then the intrinsic oscillating frequency of the disk can be calculated, $\omega_0^2\approx 1.5$. In the following we will calculate the self-oscillating luminosities of the accretion disk by adjusting the nonlinear term parameter b and the random strength D, with the other three kept fixed, $\mu_0=0.5, a=1.0, \omega_0^2=1.5$.

Figures 1(b), (d) and(f) show the normalized simulated light curves for (b)$D=0.3, b=0.0$, (d)$D=0.2, b=0.2$, and (f)$D=0.1, b=0.23$, which are similar to the light curves of three observational sequences (see Figures 1(a),(c) and(e)). It is very clear that the presence of the noise and the nonlinear

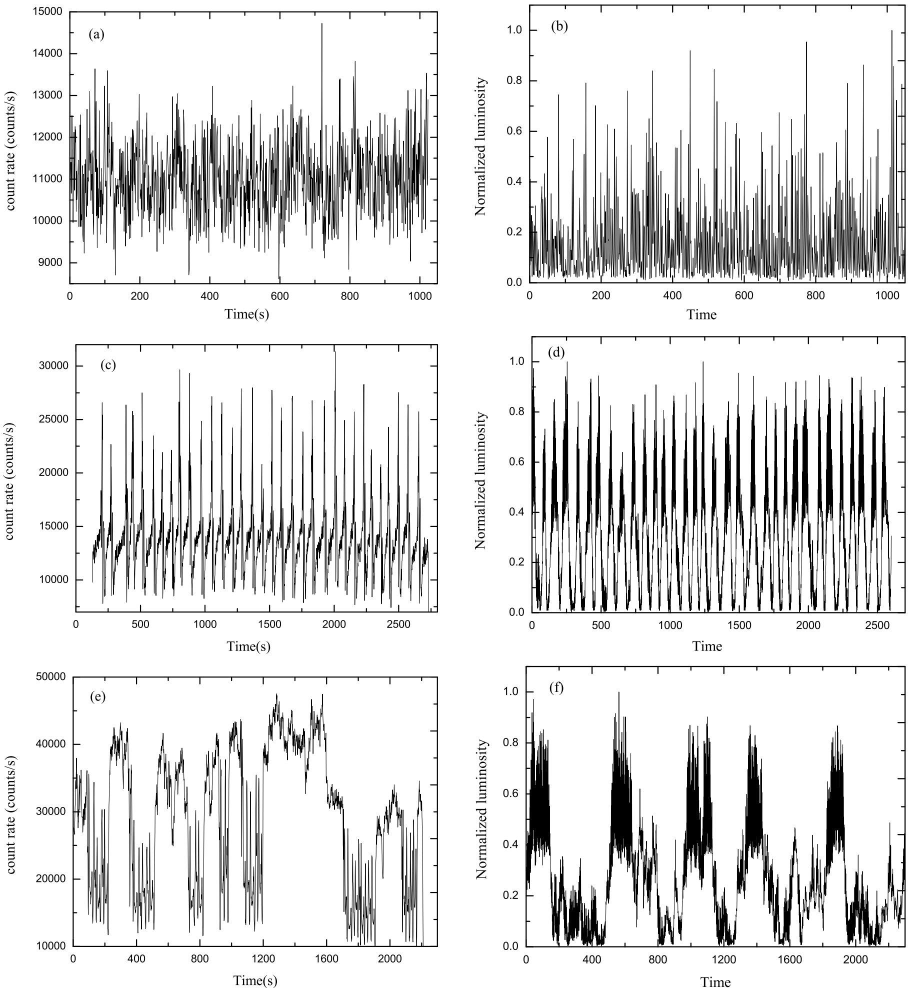

Figure 1. The light curves of the observational sequences and simulated data of GRS 1915+105. The cases (a), (c) and (e) are for the observational sequences 20402-01-02-02, 10408-01-40-00 and 10408-01-08-00, whereas (b), (d) and (f) correspond to the simulated data for D=0.3 b=0.0, D=0.2 b=0.2 and D=0.1 b=0.23, respectively.

term makes the self-oscillating luminosity curves of the accretion disk appear chaotic, random and semi-random. The nonlinear factor makes the system practically appear as chaotic, but noise always suppresses it. Therefore, when the noise fluctuations are much smaller than the variability, the nonlinear factor is dominant, and the chaotic nature can be detectable. Similarly, the random and semi-random appearances are due to the dominance of the noise in the system.

4. Correlation dimension D_2 of light curves

In order to quantitatively compare the similarity between the simulated luminosities and the observed data, we use a nonlinear time series analysis to compute their correlation dimension D_2, which is often used as a discriminating parameter for a hypothesis testing to detect nontrivial structures in the time series.

4.1. Correlation dimension D_2

The conventional method for the calculation of D_2 is the delay embedding method known as the Grassberger-Proccacia (GP) algorithm (Grassberger & Procaccia 1983; Harikrishnan *et al.*. 2006). In this technique, an embedding space of dimension M with delay vectors is constructed by splitting a discretely sampled scalar time series $s(t_i)$

$$\boldsymbol{x}_i = [s(t_i), s(t_i + \tau), ...s(t_i + (M-1)\tau)], \tag{5}$$

where τ is the delay time computed by means of the mutual information method (Kraskov *et al.* 2004), assuming that the vectors are not correlated. The correlation function for the embedded time series is the number of points that are within a distance R from the center, averaged over all the centers, and may be written as

$$C_M(R) = \lim_{N_\nu \to \infty} \frac{1}{N_\nu(N_\nu - 1)} \sum_{i}^{N_\nu} \sum_{j,j\neq 1}^{N_\nu} H(R - |\boldsymbol{x}_i - \boldsymbol{x}_j|), \tag{6}$$

where $N_\nu = N - (M-1)\tau$ is the total number of reconstructed vectors, N is the size of the data set and H is the Heaviside step function. The correlation dimension $D_2(M)$ is defined as

$$D_2(M) = \lim_{R \to 0} \frac{dlogC_M(R)}{dlogR}. \tag{7}$$

In order to test the computational schemes presented here, we first apply the D_2 analysis on the standard chaotic time series and the pure noise from the standard Lorenz attractor and stochastic white noise. Their correlation dimension curves calculated by this method are shown in Figure 2. We can see, as expected, that the $D_2(M)$ values for the Lorenz system saturate at $M = 2$ to $D_2^{sat} \approx 2$, which is close to the known value of 2.04. For the pure uncorrelated stochastic white noise, $D_2(M)$ varies approximately linearly with $M (D_2 \approx M)$ and no such saturation exists. This is consistent with the results of Misra et al. (2006), and confirms that the random data and the low-dimensional chaotic system can clearly be distinguished in this scheme.

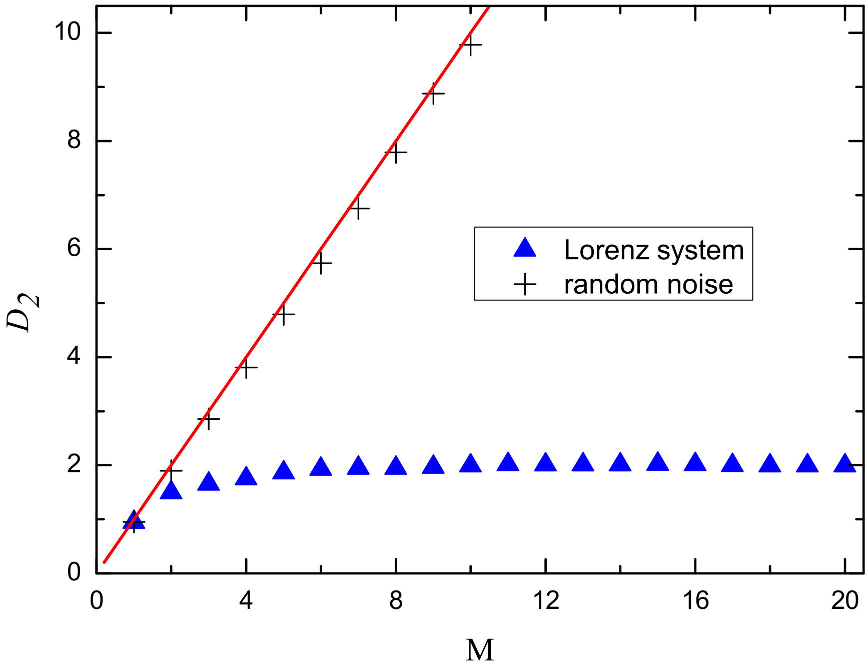

Figure 2. $D_2(M)$ curves for the random noise (crosses) and for a Lorenz system (triangles), the straight line represents a $D_2 = M$ curve.

4.2. Comparison of D_2 between the observed and simulated data

By the scheme mentioned above, we calculate the correlation dimension D_2 of the simulated and the observed data given in section 2. Their $D_2 - M$ curves are plotted in Figure 3.

In this figure, the solid line in each panel indicates the ideal random curve ($D_2 = M$) seen in Figure 2. Figure 3(a) shows the results of observational sequence 10408-01-08-00 and simulated data for $D = 0.1, b = 0.23$, both cases have the same chaotic behavior, namely, their curves show a clear deviation from

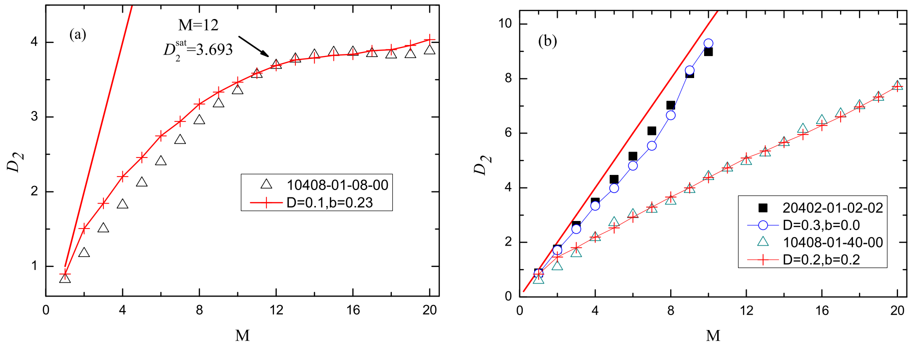

Figure 3. $D_2(M)$ curves for the observation sequences and simulated data.

this ideal random curve, and there is the same saturation correlation dimension of $D_2 = 3.693$ for $M = 12$.

For observational sequences 20402-01-02-02 and 10408-01-40-00, their curves show a different deviation from the ideal random noise (see Figure 3(b)), and D_2 varies approximately linearly with M and no such saturation exists. It can be seen that for the observational sequence 20402-01-02-02, its $D_2(M)$ curve is close to that of the ideal random noise, and is random dominated. While the other one deviates greatly from it and shows semi-stochastic behavior. As expected, the correlated dimensions of these observational variabilities are consistent with the results of the simulated luminosities for different model parameters D and b.

5. Summary and discussion

Many studies suggest that the variability of the micro quasar GRS 1915+105 may be chaotic in nature, this chaotic signature being suppressed in some cases due to the noise dominance when it appears as random or semi-random, but actually this may not be its fundamental characteristic. Therefore, in the present paper we have considered a stochastic nonlinear self-oscillatory equation to describe the vertical oscillations of an accretion disk around a black hole, and obtained the oscillating luminosities showing chaotic, random or semi-random behavior for different noise strengths and nonlinear parameters. Their properties are similar to those of the observed variability of GRS1915+105.

Our results show that the correlation dimensions between observational sequences and simulated data are identical approximately, which implies that the variability of the microquasar may be controlled by the nonlinear self-excited oscillation system; at the same time, it is also affected by the noise with varying intensity, resulting in a variety of variability types. So, as a possible astrophysical application, this model may turn out to be very useful to understand complicated temporal behavior of X-ray bursts.

Acknowledgements. This work has been supported by the Key Program of the Scientific Research Foundation of the Education Bureau of Hubei Province, China (Grant No. D20132603). We thank the anonymous referee for his (her) helpful suggestion.

References

Belloni, T., Méndez, M., King, A. R., van der Klis, M., van Paradijs, J.: 1997, *Astrophysical Journal* **479**, L145, DOI: 10.1086/310595

Belloni, T., Klein-Wolt, M., Méndez, M., van der Klis, M., van Paradijs, J.: 2000, *Acta Astron.* **355**, 271

Buchler, J. R., Kovacs, G.: 1987, *Astrophysical Journal* **318**, 232, DOI: 10.1086/165363

Castro-Tirado, A. J., Brandt, S., Lund, N.: 1992, *IAU Circular* **5590**, 2

Chen, X., Swank, J. H., Taam, R. E.: 1997, *Astrophysical Journal* **477**, L41, DOI: 10.1086/310515

Das, M. K., Chaudhary, A. K., Tavakol, R. K.: 1996, *Astrophysical Journal* **463**, 694, DOI: 10.1086/177283

Grassberger, P., Procaccia, I.: 1983, *Physica D* **9**, 189, DOI: 10.1016/0167-2789(83)90298-1

Harko, T., Mocanu, G.: 2012, *Monthly Notices of the RAS* **421**, 3102, DOI: 10.1111/j.1365-2966.2012.20530.x

Harko, T., Leung, C. S., Mocanu, G.: 2014, *Eur. Phys. J. C* **74**, 2900, DOI: 10.1140/epjc/s10052-014-2900-9

Harikrishnan, K. P. ,Misra, R., Ambika, G.: 2011, *Research in Astron. Astrophys.* **11**, 71, DOI: 10.1088/1674-4527/11/1/004

Harikrishnan, K. P., Misra, R., Ambika, G., Kembhavi, A. K.: 2006, *Physica D* **215**, 137, DOI: 10.1016/j.physd.2006.01.027

Jenkins, A.: 2013, *Physics Reports* **525**, 167, DOI: 10.1016/j. physrep. 2012.10.007

Kato S.: 2016, *Oscillations of Disks*, Astrophysics and Space Science Library, Springer Japan

Kraskov, A., Stögbauer, H.: 2004, *Phys. Rev. E* **69**, 066138, DOI:10.1103/PhysRevE.69.066138

Leung, C. S., Wei, J. Y., Harko, T., Kovacs, Z.: 2011, *J. Astrophys. Astr.* **32**, 189, DOI:10.1007/s12036-011-9026-3

Leung, C. S., Mocanu, G. Harko, T.: 2014, *J. Astrophys. Astr.* **35**, 449, DOI:10.1007/s12036-014-9250-8

Long, G. -B., Ou, J. -W., Zheng, Y. -G.: 2016, *J. Astrophys. Astr.* **37**, 10, DOI:10.1007/s12036-016-9372-2

Massaro, E. , Ardito, A., Ricciardi, P., Massa, F., Mineo, T., D'Aì, A.: 2014, *Astrophys. Space Sci.* **352**, 699, DOI: 10.1007/s10509-014-1924-9

Massaro, E., Capitanio, F., Feroci, M., Mineo, T., Ardito, A., Ricciardi, P.: 2020a, *Monthly Notices of the RAS* **495**, 1110, DOI:10.1093/mnras/staa1124

Massaro, E., Capitanio, F., Feroci, M., Mineo, T., Ardito, A., Ricciardi, P.: 2020b, *Monthly Notices of the RAS* **496**, 1697, DOI:10.1093/mnras/staa1125

Massaro, E., Capitanio, F., Feroci, M., Mineo, T.: 2020c, *Monthly Notices of the RAS* **497**, 405, DOI: 10.1093/mnras/staa1942

Misra, R., Harikrishnan, K. P., Mukhopadhyay, B., Ambika, G., Kembhavi, A. K.: 2004, *Astrophysical Journal* **609**, 313, DOI: 10.1086/421005

Misra, R., Harikrishnan, K. P., Ambika, G., Kembhavi, A. K.: 2006, *Astrophysical Journal* **643**, 1114, DOI: 10.1086/503094

Mukhopadhyay, B.: 2004, in *AIP Conf. Proc.*, ed.: P. Kaaret, F. K. Lamb, & J. H. Swank, Melville, NY: American Institute of Physics, 714
48, DOI: 10.1063/1.1780998

Nakariakov, V. M., Anfinogentov, S. A., Nisticò, G., Lee, D. -H.: 2016, *Astronomy and Astrophysics* **591**, L5, DOI: 10.1051/0004-6361/201628850

Ou, J.-W., Zheng, Y.-G., Zhang, X.: 2014, *Acta Phys. Sin.* **63**, 239801, DOI:10.7498/aps.63.239801

Paul, B., Agrawal, P. C., Rao, A. R., Vahia, M. N., Yadav, J. S., Marar, T. M. K., Seetha, S., Kasturirangan, K.: 1997, *Astronomy and Astrophysics* **320**, L37

Remillard, R. A., McClintock, J. E.: 2006, *Ann. Rev. Astron. Astrophys.* **44**, 49, DOI: 10.1146/annurev.astro.44.051905. 092532

Rudd, T. J., Rosenberg, R. M.: 1970, *Astronomy and Astrophysics* **6**, 193

Shakura, N. I., Sunyaev, R. A.: 1973, *Astronomy and Astrophysics* **24**, 337

Titarchuk, L., Osherovich, V.: 2000, *Astrophysical Journal* **542**, L111, DOI: 10.1086/312935

Wang, Z. -Y., Chen, P. -J., Zhang, L. -Y.: 2013a, *CHIN. PHYS. LETT.* **30**, 099801, DOI: 10.1088/0256-307X/30/9/099801

Wang, Z. -Y., Chen, P. -J., Wang, D. -X., Zhang, L. -Y.: 2013, *J. Astrophys. Astr.* **34**, 33, DOI: 10.1007/s12036-013-9163-y

Wang, Z. -Y., Chen, P. -J., Zhang, L. -Y.: 2015, *Chin. Phys. B* **24**, 059801, DOI: 10.1088/1674-1056/24/5/059801

Wang, Z. -Y., Chen, P. -J.: 2016, *J. Astrophys. Astr.* **37**, 8, DOI: 10.1007/s12036-016-9368-y

Contrib. Astron. Obs. Skalnaté Pleso **53/1**, 17 – 27, (2023)
https://doi.org/10.31577/caosp.2023.53.1.17

The low-frequency carbon radio recombination lines in medium toward S140 nebula

Y.V. Vasylkivskyi, S.V. Stepkin and O.O. Konovalenko

Institute of Radio Astronomy of the National Academy of Sciences of Ukraine (IRA NASU), 4 Mystetstv St., Kharkiv, 61002, Ukraine, (E-mail: vasylkivskyi@rian.kharkov.ua)

Received: February 24, 2023; Accepted: March 14, 2023

Abstract. Low-frequency carbon radio recombination lines (RRLs) serve as an effective probe of cold, partially ionized gas diagnostic. In this paper we report about studies of these lines towards the S140 nebula at UTR-2 near 26 MHz. As a spectrometer, the 4096-channel digital correlometer was used. Low-frequency carbon RRLs were detected for the S140 line of sight and its environs, the line of sight being shifted in 3° by declination to one and other side. For all three directions the similar relative line intensities were obtained. Measured line widths do not coincide with high-frequency RRLs line widths. This can tell us that a low-frequency line forming region is not associated with the S140 nebula itself. By an iterative analysis it was obtained that values of electron temperature $T_e = 50 \div 100$ K and electron density $N_e = 0.01\,\mathrm{cm}^{-3}$ most closely match our experimental data. Due to the fact that the sizes of the line forming region (or regions) exceed the solid angle of the S140 nebula, and due to the mismatch between the widths of low-frequency and high-frequency RRLs, and also based on our analysis of physical conditions of the line forming region it was proposed that low-frequency lines were formed in the local medium associated with HI clouds lying on the line of sight. Further observations and interpretation will help to describe the physical parameters of the medium in different Galactic directions.

Key words: cold interstellar medium – radio spectroscopy – radio recombination lines

1. Introduction

Currently, by methods of experimental radio astronomy a large number of spectral lines from various atoms and molecules have been detected and investigated. Spectroscopy of cosmic matter is one of the most important sources of information about physical parameters and chemical composition of the Universe.

A crucial issue in the problem of cosmic matter investigation is related to interstellar medium (ISM) studies. ISM serves as a star birthplace as well as a repository of their evolution products storage. Due to the continuous energy and matter exchange between stars and the surrounding environment, various ISM phases are possible. For example, the phases of cold neutral gas, warm neutral and ionized gas may exist (McKee & Ostriker, 1977). Its main physical parameters are very different. In cases when their temperature and density are relatively low the atoms can be excited up to high quantum states (such atoms are also often called as Rydberg atoms). They are formed mainly through the processes of ions' and electrons' recombination. Simultaneously with a free electron capture, captured or other electrons in a high quantum level jump into a less excited state. This "jump" produces the photon emission and spectral lines formed by such transitions are called "radio recombination lines" (RRLs) (Gordon & Sorochenko, 2002). The high-frequency RRLs (usually above 1 GHz) mainly trace the warm, dense gas in emission from HII-regions, while the low-frequency lines (usually below 500 MHz) trace a relatively cold and rarefied gas in absorption from diffuse, presumably neutral clouds.

Due to the fact that many astrophysical processes occur differently at low and high frequencies, and some processes do not manifest themselves at high frequencies, low-frequency RRLs provide a unique and powerful method of the ISM diagnostic. For the first time low-frequency RRLs were detected from the carbon in the late 1970s (Konovalenko & Sodin, 1981, and references therein). Detection of carbon lines and not hydrogen ones was due to a lower ionization potential of the carbon compared with the hydrogen ($E_C = 11.2$ eV, $E_H = 13.6$ eV). This makes it possible to almost completely ionize the carbon in cold rarefied regions located away from powerful ionization sources where the hydrogen will be almost all neutral. Decameter lines made it possible to study the ISM physical conditions in line forming regions (e.g. electron temperature, electron density, emission measure etc.) and to refine the kinematic parameters and dimensions of these regions (Konovalenko & Stepkin, 2005).

The partially ionized gas plays an important role in processes of the evolution and energetic of cosmic matter as well as in star formation. In the case of the presence of a strong continuum source it's possible to observe the low-frequency carbon RRLs amplified by stimulated emission. These lines can be used for studying medium regions more distant from hot stars and less dense. Some of the above mentioned objects were already investigated at UTR-2 (Konovalenko & Sodin (1981), Konovalenko (1984a,b), Golynkin & Konovalenko (1991a,b), Konovalenko & Stepkin (2005), Stepkin et al. (2007), Stepkin et al. (2021)) and at other radio telescopes (Erickson et al. (1995), Kantharia et al. (1998), Kantharia & Anantharamaiah (2001), Roshi, D. Anish et al. (2002), Roshi & Kantharia (2011), Asgekar et al. (2013), Oonk et al. (2017), Salas et al. (2017), Roshi et al. (2022)) in previous years. If we take into account a high level of radio frequency interference (RFI) and weakness of received signals, the medium in the Perseus Arm lying on the line of sight toward the bright radio source Cassiopeia A is most suitable for spectroscopic observations at low radio frequencies. A great amount of unique information has been obtained by observations in this direction (see Konovalenko, 1984b; Kantharia et al., 1998; Stepkin et al., 2007; Asgekar et al., 2013; Oonk et al., 2017; Salas et al., 2017). This success led to believing that similar results may be obtained for other Galactic directions. Indeed, this was confirmed by a low-frequency carbon RRLs detection in several Galactic lines of sight (Konovalenko, 1984a; Golynkin & Konovalenko, 1991a,b; Erickson et al., 1995; Kantharia & Anantharamaiah, 2001; Roshi, D. Anish et al., 2002; Roshi & Kantharia, 2011; Roshi et al., 2022). For our studies we chose a very interesting line of sight – the direction toward the S140 nebula.

2. S140 nebula, observational methods and equipment

S140 is a bright rim located near the south-western borders of the cold molecular cloud L1204. S140 with L1204 together form a photodissociation region. Observations of these objects are very important at both high and low frequencies. This direction is also interesting because both the S140 region and the neighboring L1204 cloud represent two examples of different star formation stages. A large amount of data in many ranges of electromagnetic radiation was detected from L1204. This became possible due to intended existence of a strong infrared source (presumably an early type star) besides two known stars of B0V and B2V classes in this cloud. Also, it was proposed that almost all of gas ionizing photons in L1204 come from S140. The high-frequency RRLs observations (C142α and C166α lines) toward S140 were reported in Knapp et al. (1976). Smirnov et al. (1995) described similar observations (C165α – C166α lines) towards S140/L1204. The authors concluded that S140 is part of a larger HII region. Its electron temperature and electron density were preliminary determined as $T_e \sim 75 \div 200$ K and $N_e \sim 0.5 \div 9$ cm^{-3}, respectively.

In the paper by Golynkin & Konovalenko (1991b) the studies of decameter carbon RRLs toward the S140 nebula at UTR-2 near 25 MHz were described (C640α line). Observations were carried out with a 160-channel digital sign correlometer with the frequency resolution of about 1 kHz. Having integration time of hundreds of hours, the authors concluded that the lines were formed in the medium with $T_e > 20$ K and $N_e < 1$ cm^{-3}. It was proposed that traced by decameter carbon RRLs gas can be associated with large volumes of the low-density medium near the S140 nebula. Based on the low electron temperature proposed by authors, they did not consider the mechanism of low-temperature dielectronic-like recombination, which affects the populations of quantum levels of atoms at $T_e \sim 100$ K and, so, the observed line intensities (Watson et al., 1980).

Smirnov et al. (1992) made an attempt to detect the meter C540α RRL in the S140 direction near 42 MHz. No line was found in the obtained spectra, although it was expected that it would have a higher intensity than the decameter C640α line. However, Smirnov et al. (1992) obtained the upper limit of the intensity for RRLs detection in meter wavelengths ($\leq 3.6 \times 10^{-4}$). They also explained a rather large C640α line width obtained by Golynkin & Konovalenko (1991b). It was proposed that the main contribution to the line broadening is made by the high-speed turbulent motions in ISM rather than collisions between Rydberg atoms and electrons. Such a large line width is not consistent with high-frequency carbon RRLs line widths (approximately $2 \div 5$ km s^{-1}). Smirnov et al. (1992) suggested that the observed decameter carbon RRLs line forming region (CII-region) in the S140 direction is not associated with the S140 nebula itself. Observational data of C640α and HI lines were also compared by Smirnov et al. (1992). There was good correspondence between them (see Fig. 3 in Smirnov et al., 1992). Both line profiles' maxima correspond to radial velocities of about 0 km s^{-1}. It was also suggested that decameter carbon RRLs in the S140 direction are formed in multiple diffuse HI clouds located on the line of sight. Such clouds have typical values of $N_H = 6 \div 20$ cm^{-3}, $N_e \sim 0.02$ cm^{-3} and $T_e = 60 \div 80$ K. The carbon in these clouds is completely ionized by UV radiation with $912\ \mathring{A} < \lambda < 1100\ \mathring{A}$.

Over the years the new models of the CII-region based on multi-frequency carbon RRLs data for Cassiopeia A direction were proposed (Kantharia et al., 1998). The best agreement is obtained for $T_e = 75$ K, $N_e = 0.02$ cm^{-3} and the association of CI-regions with neutral atomic HI clouds seems to be most plausible. Also, over the years the spectral equipment of the UTR-2 radio telescope has been improved; in particular, the number of spectral channels of the digital sign correlometer has been increased to 4096, which greatly increases the measurement sensitivity and width of bandpass.

In this paper we will try to discuss the results of high-precision low-frequency carbon RRLs observations performed at the UTR-2 radio telescope toward the S140 nebula at a frequency of 26 MHz. The Ukrainian UTR-2 is the world largest decameter radio telescope. It has an effective area of about 140 000 m^2 and the maximal angular resolution of about $\alpha \times \beta = 30' \times 30'$ at 25 MHz (Konovalenko et al., 2016). Observations reported in this paper were carried out in November 2003 at the North − South array of UTR-2, which has dimensions 1800 m × 54 m and consists of 1440 dipoles. The beam of this array is oriented along the east-west line and has a size of $\alpha \times \beta = 12^\circ \times 40'$ at 25 MHz.

As it was mentioned, the 4096-channel digital sign correlometer serves as a radio telescope back-end (Konovalenko & Stepkin, 2005). The bandpass width during the observations was 1.2 MHz and the frequency resolution was chosen as 1.1 kHz. Spectra, obtained by the Fourier transform of an autocorrelation function of the signal, were recorded on a PC with 4-minute intervals. The obtained data were processed for RFI removal by a special routine. Due to the fact that at the decameter range adjacent high excited quantum levels have a small energy difference, we can consider the neighboring levels as equivalent and fold the lines resulting by transitions from these levels. Regarding this, we fold all lines which fall into the digital correlometer bandpass (stacking 11 lines leads to the increase of sensitivity more than 3 times).

3. Results

In Fig. 1 the averaged spectrum of carbon RRLs of C627α – C637α obtained toward the S140 nebula is presented. Integration time was 108 hours. The line radial velocity is close to 0 $\mathrm{km\,s^{-1}}$, which implies the location of the CII-region within the local spiral arm.

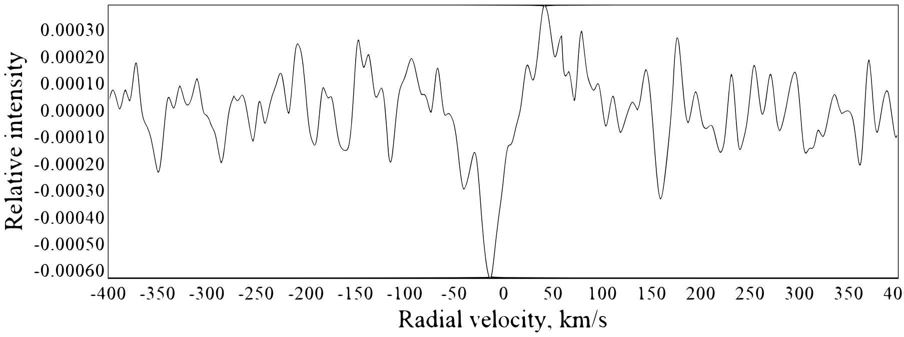

Figure 1. The spectrum of low-frequency carbon RRLs series of C627α – C637α, measured in the S140 nebula direction.

Then, in order for estimate the angular sizes of the medium region where detected low-frequency carbon RRLs were formed, in the same period there were made the additional observations toward the line of sight shifted from the S140 direction by 3 degrees in declination. In Fig. 2 and 3 we present the spectra measured in G105.15+2.8 and G108.48+7.83 directions shifted by three degrees to the side of negative and positive declinations from the S140 line of sight, respectively. Integration times were 46.2 and 72 hours, respectively. Carbon RRLs were detected in both spectra with similar relative line intensities, which have a good correspondence with the measured relative line intensity for the S140 line of sight. This may imply that lines in all three directions were formed in a medium with similar physical conditions. Obviously, they may be formed in the same cloud, or a complex of interstellar clouds. The values of line widths were obtained by fitting of a Gaussian into the line profiles. The line widths for all three directions have basically similar values (for the G105.15+2.8 direction pointed near the Galactic plane the width is slightly larger due to the presence of large gas volumes near the plane with different radial velocities). Obtained line widths in terms of radial velocities are consistent with the line widths measured in other Galactic plane line of sights at slightly higher frequencies, Erickson et al. (1995) and Kantharia & Anantharamaiah (2001). The line widths practically do not change with frequency, therefore the Doppler broadening mechanism plays a predominant role.

In Table 1 the obtained values of main line characteristics are listed.

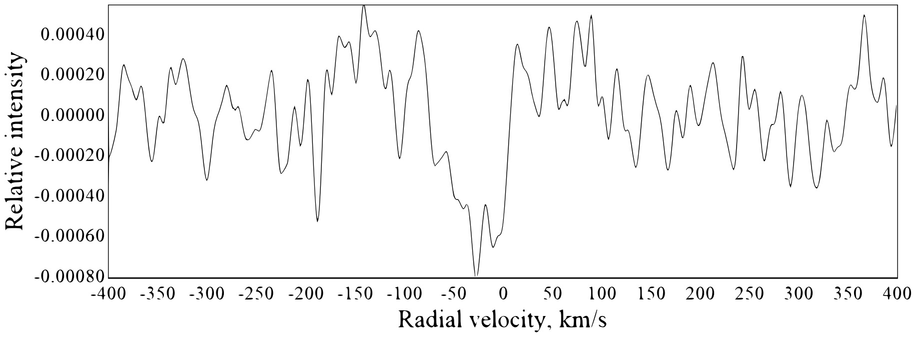

Figure 2. The spectrum of low-frequency carbon RRLs series of C627α – C637α, measured in the G105.15+2.8 direction.

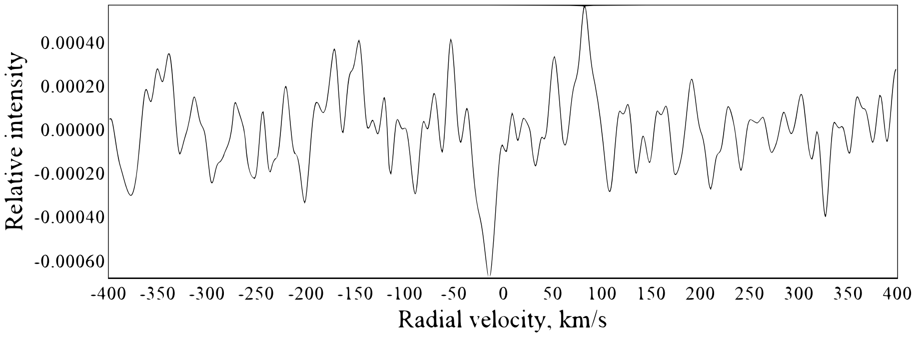

Figure 3. The spectrum, of low-frequency carbon RRLs series of C627α – C637α measured in the G108.48+7.83 direction.

4. Discussion

The line widths obtained in this work (Table 1) are several times smaller than those in Golynkin & Konovalenko (1991b). This may be due to the higher measurement sensitivity secured through the increased number of digital correlometer channels, a wider bandpass and, therefore, a larger number of simultaneously observed transitions as well as the more correct usage of the profile fitting procedure. At the same time, the line widths obtained in this work ($10 \div 28$ km s^{-1}) significantly exceed the widths of the high-frequency lines ($2 \div 5$ km s^{-1}) (Knapp et al., 1976; Smirnov et al., 1995). Smirnov et al. (1992) also described that widths of high-frequency and low-frequency carbon RRLs in the S140 direction do not coincide. This may indicate that the formation of two line types occurs in different, spatially unrelated ISM regions. This implies that low-frequency line forming regions are apparently not directly related with the

Table 1. Obtained low-frequency carbon RRLs characteristics.

Direction	$\Delta T_L/T_C$ $\times\ 10^{-4}$	V_{LSR} [km s^{-1}]	$\Delta\nu$ [kHz]	ΔV [km s^{-1}]	I_L [s^{-1}]
S140	6.2	-16	1.63	18.82	-1.54
G105.15+2.8	8.1	-29	2.47	28.55	-2.92
G108.48+7.83	6.8	-16	0.9	10.38	-1.04

S140 nebula and high-frequency line forming regions.

In Golynkin & Konovalenko (1991b) the lines at the UTR-2 North – South array were observed with a relative intensity of about 8×10^{-4} against the continuum. Relative intensity values obtained in this work are consistent with those of Golynkin & Konovalenko (1991b) and with theoretical calculations by the formula

$$\frac{\Delta T_L}{T_C} = -\tau_L = -\tau_L^* b_n \beta_n, \tag{1}$$

where $\Delta T_L/T_C$ is the relative line intensity, τ_L is the optical depth, τ_L^* is the equilibrium optical depth (optical depth at LTE conditions), b_n is the departure coefficient and β_n is the stimulation emission coefficient,

$$\beta_n = 1 - \left(\frac{kT_e}{h\nu}\right)\left(\frac{d\ln b_n}{dn}\right), \tag{2}$$

where h is Planck's constant ($h = 6.6261 \times 10^{-27}$ erg·s).

Integrated line intensities I_L were experimentally determined (Table 1), the values of which were used for T_e and N_e estimates by the formula

$$I_L = \int\limits_{\nu} \frac{\Delta T_L}{T_C} d\nu \approx -2 \times 10^6 \frac{N_e^2 s}{T_e^{5/2}} b_n \beta_n, \tag{3}$$

where s is the path length in pc. The value of s was taken equal to 5 pc as an average typical value for the size of HI clouds in Galaxy. From six T_e and N_e combinations (T_e = 50 K, N_e = 0.01 cm^{-3}; T_e = 50 K, N_e = 0.1 cm^{-3}; T_e = 50 K, N_e = 1 cm^{-3}; T_e = 100 K, N_e = 0.01 cm^{-3}; T_e = 100 K, N_e = 0.1 cm^{-3}; T_e = 100 K, N_e = 1 cm^{-3}) and corresponding $b_n\beta_n$ values (which are taken from Walmsley & Watson (1982)), our results for the S140 direction are most consistent with values of T_e = 50 and 100 K, N_e = 0.01 cm^{-3} which are consistent with a model by Kantharia et al. (1998). $b_n\beta_n \approx 20$ for T_e = 50 K and $b_n\beta_n \approx 90$ for T_e = 100 K for corresponding $n \sim 630$.

Relative intensities of low-frequency carbon RRLs from directions deviating from the S140 line of sight practically do not differ from those measured towards S140 itself. It should be noted that the relative line intensity is a unit of the line strength used in this work. The real line intensity will differ from the observed relative intensity by

$$\left(\frac{\Delta T_L}{T_C}\right)_{real} = \left(\frac{\Delta T_L}{T_C}\right)_{observed} \frac{\Omega_S}{\Omega_A} \frac{T_B}{T_B + T_F} \frac{T_A}{T_A + T_N}, \tag{4}$$

where Ω_S and Ω_A are solid angles of the source and radio telescope beam, respectively, T_B and T_F are Galactic background brightness temperatures behind and in front of the source and T_A and T_N are the antenna background temperature at the receiver input and the receiver noise temperature, respectively (Golynkin & Konovalenko, 1991b). Since until now the exact values of T_B and T_F have been unknown, it is not possible to determine the real line intensities.

The line radial velocities observed in this work and velocities from Golynkin & Konovalenko (1991b) practically coincide. This suggests that the lines were apparently formed in gas lying in the Perseus and Orion arms.

Based on the presence of lines in spectra in Fig. 2 and 3 with similar widths, relative intensities and radial velocities, we can assume that angular dimensions of the CII-region, where low-frequency carbon RRLs in the S140 direction were formed, are greater than 6° by declination. If we convert this value into linear dimensions units (assuming the distance to S140 of about 1 kpc) by the formula

$$a = \frac{360^\circ}{2\pi} \left(\frac{d}{D}\right), \tag{5}$$

where a is the angular size in degrees, d is the linear size in pc and D is the distance from the observer to the source in pc, we find that linear dimensions of the medium region (or multiple regions) where the detected low-frequency carbon RRLs were formed, is no less than 100 pc. Only giant molecular clouds can correspond to such linear dimensions, but the formation of low-frequency RRLs in these objects is extremely unlikely (Kantharia et al., 1998). A reasonable explanation could be that the detected lines were formed in medium regions lying on the line of sight to S140, but which are located much closer to us. Based on the aforementioned facts and dimensional estimates, the CII-region (CII-regions) cannot be associated with the S140/L1204 complex but with medium regions lying on the line of sight.

The question of particular importance is the nature of CII-region when low-frequency carbon RRL was formed. Golynkin & Konovalenko (1991b) noticed that this CII-region could be associated with the S140 nebula itself. It was also noted that, in principle, there could be a variant of these lines' formation in a local medium lying on the line of sight under the influence of a low-temperature dielectronic-like recombination mechanism (Watson et al., 1980). In this work carbon RRLs detected from all the three directions with almost the same intensities, line widths and radial velocities, and we may suppose that these lines originated in the same gas volume. Based on the analysis carried out by Smirnov et al. (1992) and the analysis of physical conditions carried out in this work, it can be assumed that observed RRLs can be formed in diffuse HI clouds on the line of sight towards S140. HI clouds are widespread in the Galactic Plane, have typical physical conditions' values coinciding with those obtained above ($T_e = 50 \div 100$ K, $N_e = 0.01$ cm^{-3}) and have typical linear dimensions up to 10 pc (usually $4 \div 6$ pc).

5. Conclusions

In this work we present the low-frequency carbon RRLs studies toward the S140 nebula carried out at the UTR-2 radio telescope near 26 MHz. Observations were made using a 4096-channel digital correlometer within the 1.2 MHz bandpass with the frequency resolution of 1.1 kHz.

Low-frequency carbon RRLs were detected in absorption in both the S140 nebula direction and its environs (line of sights shifted from S140 in 3° by declination). In all the three cases the relative line intensities have similar values. The measured line widths are comparable with those for other Galactic plane directions and apparently have a Doppler broadening mechanism. Results obtained for the width of low-frequency RRLs do not agree with the width measurements of high-frequency lines, which virtually eliminates the possibility of low-frequency RRLs formation in a region associated with S140 itself. An analysis of CII-region physical conditions showed that the best fit to our observational data have the values $T_e = 50 \div 100$ K, $N_e = 0.01$ cm^{-3}, which is consistent with the currently most correct multi-frequency model of a CII-region. Taking these facts into account and also the fact that line characteristics for the three line of sights have very close values, we can suppose that detected toward S140 carbon RRLs were formed in a spatially extended medium region with angular sizes no less that 6°. The most reasonable explanation of all of this is the association of a low-frequency lines' forming region with diffuse HI clouds lying in the local medium on the line of sight toward S140. Further RRLs studies in both S140 and other Galactic directions in a wide frequency range will allow us to create a more reliable and comprehensive model of the CII-region where low-frequency carbon RRLs are formed.

Acknowledgements. This work was based on the authors' report at the Young Science Conference on Astronomy and Space Physics which was held in April 2018 in the Taras Shevchenko National University of Kyiv. The authors are grateful to Dr. Olena Shubina for useful advice and discussions that led to a deeper analysis of the obtained results and preparation of this work.

References

Asgekar, A., Oonk, J. B. R., Yatawatta, S., et al., LOFAR detections of low-frequency radio recombination lines towards Cassiopeia A. 2013, *Astronomy and Astrophysics*, **551**, L11, DOI: 10.1051/0004-6361/201221001

Erickson, W. C., McConnell, D., & Anantharamaiah, K. R., Low-Frequency Carbon Recombination Lines in the Central Regions of the Galaxy. 1995, *Astrophysical Journal*, **454**, 125, DOI: 10.1086/176471

Golynkin, A. A. & Konovalenko, A. A., Decameter-Wave Carbon Recombination Lines in the Dust Cloud L:1407. 1991a, *Soviet Astronomy Letters*, **17**, 10

Golynkin, A. A. & Konovalenko, A. A., Radio Recombination Lines of Highly Excited Carbon Near DR:21 and S:140. 1991b, *Soviet Astronomy Letters*, **17**, 7

Gordon, M. A. & Sorochenko, R. L. 2002, *Radio Recombination Lines. Their Physics and Astronomical Applications* (Springer New York, NY), DOI: 10.1007/978-0-387-09691-9, ISBN: 978-0-387-09691-9

Kantharia, N. G. & Anantharamaiah, K. R., Carbon recombination lines from the Galactic plane at 34.5 & 328 MHz. 2001, *Journal of Astrophysics and Astronomy*, **22**, 51, DOI: 10.1007/BF02933590

Kantharia, N. G., Anantharamaiah, K. R., & Payne, H. E., Carbon Recombination Lines between 34.5 and 770 MHz toward Cassiopeia A. 1998, *Astrophysical Journal*, **506**, 758, DOI: 10.1086/306266

Knapp, G. R., Brown, R. L., Kuiper, T. B. H., & Kakar, R. K., Carbon recombination line observations of the Sharpless 140 region. 1976, *Astrophysical Journal*, **204**, 781, DOI: 10.1086/154225

Konovalenko, A., Sodin, L., Zakharenko, V., et al., The modern radio astronomy network in Ukraine: UTR-2, URAN and GURT. 2016, *Experimental Astronomy*, **42**, 11, DOI: 10.1007/s10686-016-9498-x

Konovalenko, A. A., Detection of the excited carbon lines at decametric waves in some galactic objects. 1984a, *Pisma v Astronomicheskii Zhurnal*, **10**, 912

Konovalenko, A. A., Observations of carbon recombination lines at decametric wavelengths in the direction of Cassiopeia. 1984b, *Pisma v Astronomicheskii Zhurnal*, **10**, 846

Konovalenko, A. A. & Sodin, L. G., The 26.13 MHz absorption line in the direction of Cassiopeia A. 1981, *Nature*, **294**, 135, DOI: 10.1038/294135a0

Konovalenko, A. A. & Stepkin, S. V., Radio recombination lines. 2005, in EAS Publications Series, Vol. **15**, *EAS Publications Series*, ed. L. I. Gurvits, S. Frey, & S. Rawlings, 271–295

McKee, C. F. & Ostriker, J. P., A theory of the interstellar medium: three components regulated by supernova explosions in an inhomogeneous substrate. 1977, *Astrophysical Journal*, **218**, 148, DOI: 10.1086/155667

Oonk, J. B. R., van Weeren, R. J., Salas, P., et al., Carbon and hydrogen radio recombination lines from the cold clouds towards Cassiopeia A. 2017, *Monthly Notices of the RAS*, **465**, 1066, DOI: 10.1093/mnras/stw2818

Roshi, D. A. & Kantharia, N. G., Carbon recombination lines towards the Riegel-Crutcher cloud and other cold H i regions in the inner Galaxy. 2011, *Monthly Notices of the Royal Astronomical Society*, **414**, 519, DOI: 10.1111/j.1365-2966.2011.18418.x

Roshi, D. A., Peters, W. M., Emig, K. L., et al., Arecibo-Green Bank-LOFAR Carbon Radio Recombination Line Observations toward Cold H I Clouds. 2022, *Astrophysical Journal*, **925**, 7, DOI: 10.3847/1538-4357/ac35d8

Roshi, D. Anish, Kantharia, N. G., & Anantharamaiah, K. R., Carbon recombination lines near 327 MHz - I. "Diffuse" CII regions in the Galactic Disk. 2002, *Astronomy and Astrophysics*, **391**, 1097, DOI: 10.1051/0004-6361:20020899

Salas, P., Oonk, J. B. R., van Weeren, R. J., et al., LOFAR observations of decameter carbon radio recombination lines towards Cassiopeia A. 2017, *Monthly Notices of the RAS*, **467**, 2274, DOI: 10.1093/mnras/stx239

Smirnov, G. T., Sorochenko, R. L., & Kitaev, V. V., A search for recombination lines at 42 MHz toward S 140. 1992, *Soviet Astronomy Letters*, **18**, 192

Smirnov, G. T., Sorochenko, R. L., & Walmsley, C. M., The S 140/L 1204 complex: radio recombination lines of hydrogen, carbon and sulphur. 1995, *Astronomy and Astrophysics*, **300**, 923

Stepkin, S. V., Konovalenko, A. A., Kantharia, N. G., & Udaya Shankar, N., Radio recombination lines from the largest bound atoms in space. 2007, *Monthly Notices of the RAS*, **374**, 852, DOI: 10.1111/j.1365-2966.2006.11190.x

Stepkin, S. V., Konovalenko, O. O., Vasylkivskyi, Y. V., & Mukha, D. V., Interstellar Medium and Decameter Radio Spectroscopy. 2021, *Radio Physics and Radio Astronomy*, **26**, 314, DOI: 10.15407/rpra26.04.314

Walmsley, C. M. & Watson, W. D., The influence of dielectronic-like recombination at low temperatures on the interpretation of interstellar, radio recombination lines of carbon. 1982, *Astrophysical Journal*, **260**, 317, DOI: 10.1086/160256

Watson, W. D., Western, L. R., & Christensen, R. B., A new, dielectronic-like recombination process for low temperatures and the radio recombination lines of carbon. 1980, *Astrophysical Journal*, **240**, 956, DOI: 10.1086/158309

Contrib. Astron. Obs. Skalnaté Pleso **53/1**, 28 – 45, (2023)
https://doi.org/10.31577/caosp.2023.53.1.28

Appropriate site selection for the astronomical observatory - Erzincan province sample application

A. Yılmaz

Health Services Vocational School, Erzincan Binali Yıldırım University, Erzincan, Turkey (E-mail: abdulvahap.yilmaz@erzincan.edu.tr)

Received: January 26, 2023; Accepted: March 20, 2023

Abstract. In this study, remote sensing methods, a multicriteria decision analysis, and the GIS program were used to determine the most suitable sites for the construction of an astronomical observatory. The study was conducted within the provincial boundaries of Erzincan, where there is not yet an observatory. The study identified and analyzed meteorological, geographic and anthropogenic factors (cloud cover, precipitable water, wind speed, geology and landslides, active fault lines, digital elevation model, urban lighting, mining activities, distance to roads). The areas proposed as a result of the study were plotted on the city map. It is important to reduce the number of eligible areas to be site tested by performing the work prior to the site testing phase, which is one of the phases required for the construction of the observatory. This study will provide energy, time, and cost savings by reducing the number of alternative areas. As a result of the preliminary study we conducted to select the most suitable site for the observatory, the final decision should be made after the proposed areas have been subjected to the site test phases.

Key words: Site selection – Astronomical observatory – Multi-Criteria Decision Analysis – AHP – GIS

1. Introduction

In order for observatories to operate at high efficiency, it is important that the site where they are built provide the most ideal observing conditions. Site selection is an evaluation process to create the most suitable conditions by combining many elements from different sources and consists of the steps of site selection and site testing, Hudson & Simstad (2010) and Koc-San et al. (2013). Site testing is a process that requires many steps, such as sky brightness, atmospheric visibility, precipitable water vapor, and atmospheric extinction coefficient calculation Koc-San et al. (2013). This process will expand in direct proportion to the number of alternatives. It is important to select the site in such a way that the number of alternatives is limited in order to reduce the time, energy, and cost required for the site testing process. The use of multi-criteria decision analysis (MCDA) in conjunction with geographic information systems (GIS) and remote sensing, and the ability to manage complex and voluminous data from multiple data sources, is a powerful tool that provides an efficient process for site selection. In recent years, the development of remote sensing resources for site selection, the use of MCDA and AHP, and their support by a GIS system have attracted the attention of researchers, Koc-San et al. (2013). The Analytic Hierarchy Process (AHP) has become one of the most widely used methods in the decision analysis since its development by Saaty (1988). AHP is a method of decision making in which the weights of criteria are determined by pairwise comparisons. The determination of weights by expert decision makers who evaluate the criteria increases the reliability of the method. There are many studies in various fields using decision analysis approaches along with GIS, AHP and the multi-criteria decision analysis (MCDA). Selection of the most suitable site for a biogas plant, Tulun et al. (2022), selection of sites for solid waste disposal Bilgilioglu et al. (2022), collection areas after earthquakes, Aman & Aytac (2022). The first site selection for an astronomical observatory in Turkey was made in 1989 by Aslan et al. when determining the site for the construction of the TUeBTAK National Observatory (TUG). In this study, field experiments, meteorological measurements and anthropogenic factors were given priority in determining the suitable site, Koc-San et al. (2013). Another study was conducted in 2013 by Koc-San Vd. using GIS, the MCDA method integrated with remote sensing methods and AHP to propose alternative sites for the TUeBTAK National Observatory. The importance of astronomy in the world the increasing growth of Erzincan Binali Yıldırım universty in Erzincan Province, the prediction of scientific studies in the field of astronomy in the future encouraged us to choose the appropriate place for the construction of an observatory within the boundaries of the province. The aim of our study is to identify the most suitable sites for the construction of an observatory in the province. A method combining MCDA with remote sensing technologies and geographic information systems (GIS) was used to select the suitable site, Koc-San et al. (2013). There are many parameters to be considered when selecting an astronomical observatory with the most efficient operating conditions, Jia et al. (2012). Important parameters affecting the selection of suitable sites for astronomical observatories have been published by Walker (1984), Cowles (1989) and Hudson & Simstad (2010). In addition, a study similar to our study was conducted for Antalya province. Considering these studies, the parameters affecting the selection of a suitable site for the observatory were determined by Koc-San et al. (2013). In this study, three main groups were formed as meteorological, geographical, and anthropogenic parameters, 10 factors were created among these main groups and recorded as grid layers in GIS, Koc-San et al. (2013). According to the AHP method, these layers were compared with each other and their weighting values were determined by the experts. Considering these weighting values, the most suitable site was selected for the astronomical observatory in Erzincan, Turkey. In this study, we propose the most suitable optical astronomical observatory for construction by mainly overlapping the maps that we created by digitizing in the system GIS as a result of the data analysis using remote sensing, MCDA and AHP methods for site selection.

2. Workspace and data used

The study area is located in the western part of the Eastern Anatolian region of Erzincan province in Turkey and covers 11.815 km^2, Figure 1. From the data of the General Directorate of State Meteorological Affairs of Erzincan Province, the average number of clear days per year is 116.6. In winter, when the air mass coming from Siberia acts from the east, there may be very severe winter days. The average annual temperature in Erzincan is 10.9 C, with an average of -6.7 C in January, the coldest month, and an average of 31.4 C in July, the warmest month Verileri (2010) and Kaya (2011). The region is surrounded by many mountains whose peaks are over 3000 meters high. High altitude, moisture-free sites are the most suitable for observatory construction, Koc-San et al. (2013).

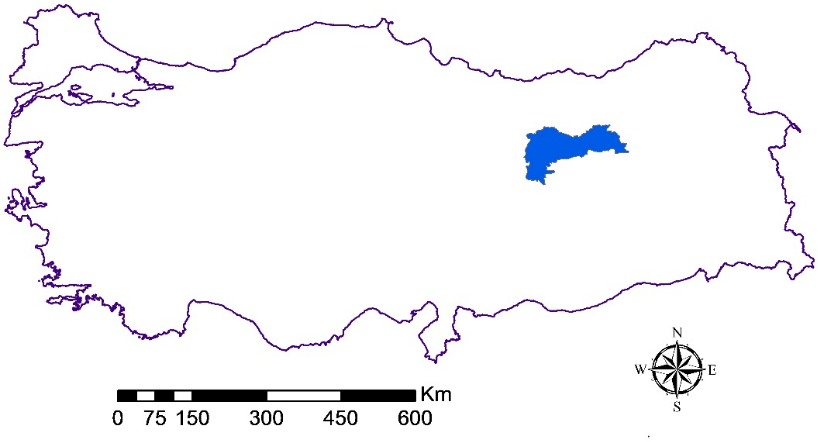

Figure 1. Location of the work area

In this study, datasets from three main areas (meteorology, geography and anthropogenes) were used to determine the most suitable location for astronomical observing conditions. The datasets are explained in detail.

2.1. Meteorological datasets

In our study, the meteorological data set consists of cloud cover, precipitable water, and wind speed parameters. One of the most important factors affecting an astronomical observation is the cloud cover of the area to be observed, Jia et al. (2012). One of the most important criteria affecting atmospheric permeability is the amount of settable water. Cloud cover and precipitation water are the most important parameters in our study, Koc-San et al. (2013). Of the meteorological parameters, cloud cover and collapsing water data are products of Moderate Resolution Imaging Spectroradiometry (MODIS) data, which covers the entire Earth almost daily. MODIS data have resolutions of 250 m (2 bands) and 500 m (5 bands) in the VIS and SWIR bands, respectively, based on different methods, Barnes et al. (1998). In this study, the MOD10A1 products were used for cloud cover and MOD05L2 for stormwater. The MOD10A1 products were obtained from the National Snow and Ice Data Center (NSIDC) and the MOD05L2 products were obtained from the Atmosphere Archive and Distribution System (LAADS) (NASA-EOSDIS, 2022). The edge length of each MOD10A1/MYD10A1 granule is 1200 km (corresponding to about 10), and the projections are equidistant and sinusoidal. MOD10A1 provides a thematic map with HDF extension that includes data on snowpack fraction, snow, cloud mask, terrain, and water classes, Hall & Riggs (2007). MODIS cloud mask data is binary information about cloud cover, i.e., whether a cloud is present or not Strabala (2005). In this study, cloud cover data were obtained from the snow cover fraction layer of the MOD10A1 data, Hall & Riggs (2007). Sedimentable water content, another meteorological data set, is one of the important criteria affecting atmospheric permeability. Sensors have been developed to determine water vapor content (Advanced Very High-Resolution Radiometer (AVHRR), Sobrino et al. (2003), and MODIS, Gao & Goetz (1990) and Kaufman & Gao (1992). Another MODIS product, MOD05L2 (near-infrared and QA infrared), is used to determine water vapor content, Sobrino et al. (2003) and Kern et al. (2008). To obtain the map of the working area of water vapor, the infrared data sets of water vapor obtained from the MOD05L2 product were used to determine the water vapor content (NASA-EOSDIS, 2022). In this study, 4380 granules in HDF format were combined in the program GIS for the spatial information of each parameter cloud cover and precipitable water, and 1095 data in the study area were used (365 3 years), covering a 3-year period from January 1, 2020, to December 31, 2022. Wind speed data were obtained from the Global Wind Atlas 3.0 map, a free web-based application developed and operated by the Technical College of Denmark (DTU). The Global Wind Atlas 3.0 was published in collaboration with the World Bank Group using data provided by Vortex with funding from the Energy Sector Management Assistance Program (ESMAP) (more information: https://globalwindatlas). Our study used wind speed data from the GWA version 3.2 with a spatial resolution of 10 m to account for local variations in average wind speeds. The Global Wind Atlas (GWA) primarily supports wind energy development in the exploration phase and in the preliminary wind resource assessment phase prior to on-site meteorological surveys. The Global Wind Atlas provides wind source mapping at a horizontal grid of 250 m and wind source mapping at 10, 50, 100, 150, and 200 m above the ground level. Users can download GIS data for all layers at a point, in a specific area, or within a country. The speed at which air flows over the water surface affects the evaporation rate. Humidity depends on the amount of vapor in the air. Therefore, a higher wind speed results in minimal evaporation of water and lower humidity. Lower wind speed results in maximum evaporation of water and higher humidity, Ravi & D'Odorico (2005). Wind speed has an indirect effect on site selection for astronomical observatories.

2.2. Geographical datasets

Geographic Data Group; The Digital Elevation Model (DEM) consists of geology, landslides, active fault lines, and data. As altitude increases, sky transparency increases, Hudson & Simstad (2010), components that increase atmospheric scatter are reduced. Therefore, DEM is an important parameter. The Shuttle Radar Topography Mission (SRTM) has a spatial resolution of 90 m 90 m DEM represents the topography of the studied area. DEM Data used in the study were obtained from the USGS Earth Explorer. The data were processed in the ArcGIS 10.8 program, Bal et al. (2021). The construction areas of the buildings should be kept away from geological conditions, landslide zones, active fault lines, and floodplains. Considering that the building to be constructed is an observatory equipped with expensive equipment, it is important to choose the safest construction site. For this reason, maps of geology, landslides, and active fault lines were used in our study. Geology, landslide and active fault maps (for the study area) were drawn separately for each of the geology formation, landslide and active fault parameters using the drawing editor via the soil unit map viewer application of the General Directorate of Mineral Research and Exploration, Akbaş et al. (2011) and Üzer et al. (2013). In Erzincan province, it is divided into four basic tectonic-stratigraphic units of the pre-Eocene age, which have different environmental conditions and are tectonically connected. They are the Munzur Limestone, the Erzincan Nappe, the imen Mountain Nappe and the Kelkit-Rela autochthon from south to north. Of these nappes, the Kelkit autochthonous nappe lies on the imen Mountain nappe in the south and the Erzincan nappe in the north, and the imendaı nappe lies above the Erzincan nappe. The Munzur limestone napi is located below the Erzincan nappe, Yilmaz (1985). The data on the geological maps show that the rock classes of the study area are in seven different formations and are from the Cretaceous, Lower Miocene, Upper Miocene-Pliocene, Oligocene, Eocene, Miocene, Malm Cretaceous, Lower Cretaceous and Upper Cretaceous, Gürsoy (2012). Another criterion for selecting a suitable site is the landslide inventory data. Landslide maps from the General Directorate of Mineral Research and Exploration were used. According to the landslide maps, landslide activity consisted mainly of Miocene clastic neritic limestone, Mesozoic unconfined basic and ultrabasic rocks. In the landslide map prepared for the study area, all landslide activities were marked as areas unsuitable for construction. It is important to build as far as possible from active fault lines, which is another parameter. The urban area is surrounded by the Northeast Anatolian Fault Line and the Ovacık Fault Line, as well as the secondary fault lines associated with these two main fault lines, Barka & Eyidoan (1993). Maps of active faults produced by the MTA were digitized using the GIS program, and maps of active fault lines were produced.

2.3. Anthropogenic datasets

Anthropogenic datasets were created using city lights, distance to roads, and point density of mine sites. City lighting is important for people to continue their activities when they cannot benefit from sunlight. It is one of the most important parameters that negatively affect astronomical observations. The Operational Line Scanning System (OLS) is one of the Defense Meteorological Satellite Program (DMSP) sensors and was originally developed for lunar light cloud observations, Elvidge et al. (1997). Its ability to image in low light has made it the preferred accurate and effective data source for imaging anthropogenic lights on the Earth's surface in cloudless night conditions without moonlight, Zhang & Seto (2011). In this study, the DMSP/OLS data were used to evaluate the effects of urban lights in selecting suitable locations. It is important to determine the positions of astronomical observatories in relation to mine sites because mine sites cause dustiness and affect sky opacity positively and sky observations negatively Koc-San et al. (2013). The locations of the mining sites were taken from the General Directorate of Mining Affairs of Turkey. In order to reduce the construction cost of the astronomical observatory and to make it easy to reach, the distance to the roads was used as a criterion in our study. These comparative data were obtained from the OpenStreetMap. The data for Turkey was obtained using the overpass turbo, Ma et al. (2015).

3. Methodology

In this study, there are five stages: (i) determining the criteria/factors and storing them in the environment GIS, (ii) data analysis, (iii) map evaluation using the MCDA method, (iv) determining the weight of each criterion using the AHP method and (v) combining the criteria. In this study, after analyzing raster data layers and digital data, a model was created in Arc- GIS 10.8 software. The data processing and analysis in this model is shown in the following analysis, Figure 2. In this study, three main parameters consisting of anthropogenic, meteorological, and geographic datasets were used to select the appropriate site, Koc-San et al. (2013). These datasets were obtained from the drawing editor, remote sensing data or digital maps, and a map was created for each dataset using the program GIS. It was ensured that the projection systems (WGS84) of the maps created or used were identical. Each data set used in this study is represented by a layer. Each data layer is converted from a raster layer to a vector layer. Many GIS -based applications are used to analyze these layers. A common field was added to each layer, and the field values were set between 1 and 5, where 5 is the best score and 1 is the worst one. All data layers were converted from vector to raster again according to the generated field data. In the next step, the weights of each criterion were determined by pairwise comparisons of the experts' criteria (45 comparisons for 10 criteria), as required by the AHP model. Finally, the weighting values of each level were entered into the Weight Overly application in the GIS program and the most appropriate location within the boundaries of the study area was selected.

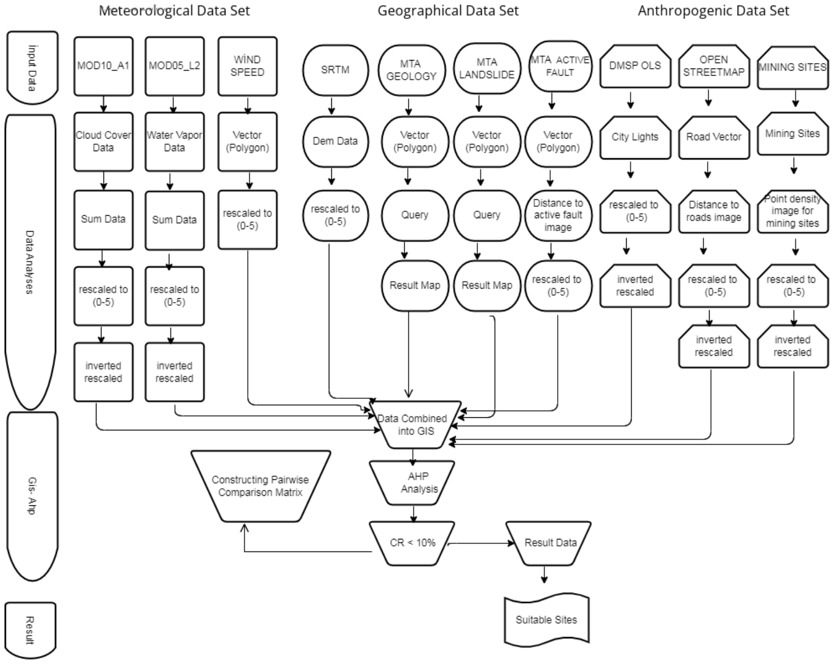

Figure 2. Flow chart of the process steps followed in the study.

3.1. Processing of meteorological data sets

3.1.1. Cloud cover data

Approximately 4380 granules were obtained from MOD10A1 between the years 2020 and 2022, from which 1095 data were obtained. These data were combined into a single layer. The obtained layer was used as spatial cloudiness data. Since areas with low cloud cover are considered ideal observation points, low pixel values with low cloud cover are scaled as 5 points and high pixel values with high cloud cover are scaled as 1 point (see Figure 3(a)). Cloud cover is directly

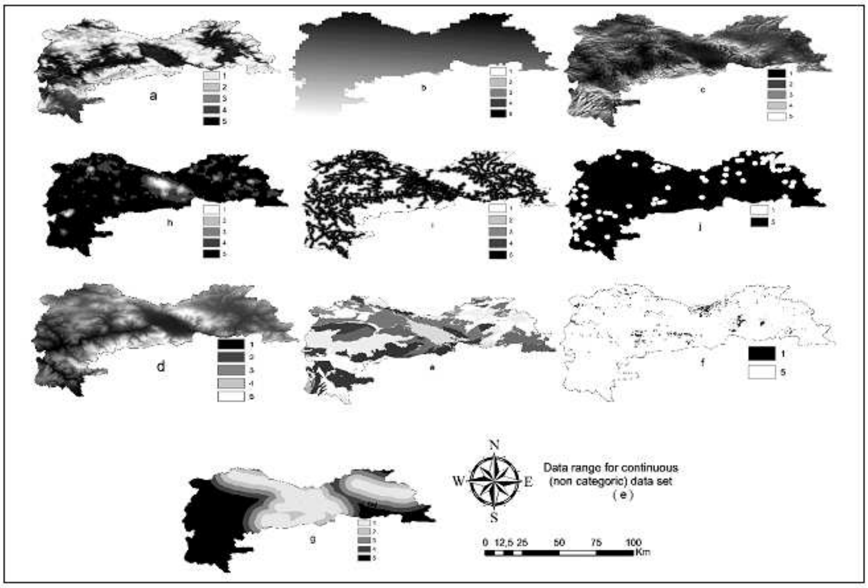

Figure 3. The data sets used; meteorological (cloud cover (a), rainy water (b)), wind speed (c)) geographical (DEM (d), geologically suitable areas (e), landslide (f), distance map of active fault lines (g)) and anthropogenic datasets (city lights (h), distance to roadmap (i) and point density (j) for mining operations).

effective in astronomical observations. The best observation conditions occur during periods of cloud-free weather (no clouds or very few clouds). According to the World Meteorological Organization, total cloud cover is defined as the part of the celestial dome covered by all visible clouds, Mekhaimr (2017). The celestial dome is divided into eight sections, each section is called an arrow, and an estimate is made on how many arrows the clouds cover. It is defined as clear weather or slightly cloudy (< 2 octas), partly cloudy (3-5 octas) and cloudy sky (≥ 6 octas), Mekhaimr (2017). Observation conditions according to cloud cover classifications, 2 octas and below are the most ideal conditions for outdoor observation, the 3-5 octas December is defined as partly cloudy skies and partly offers periods of observation. Between the years 2010-2022, the cloud closure data of the meteorological department of Erzincan province were interpreted according to the above-mentioned octa data classes, and the number of days that can be observed was estimated as 247.5 days per year.

3.1.2. Precipitable water data

Rainfall information was obtained from the MOD05L2 dataset using the water vapor infrared data channel. For the three-year period, the settable water infrared data were processed and converted into a single data layer using the program GIS, and the stormwater data were presented using this layer. The stormwater vapor data are scaled from 1 to 5. In the scaling, 1 represents more settleable water vapor and 5 represents less settleable water vapor (Figure 3(b)).

3.1.3. Wind speed data

Wind speed data are from the Global Wind Atlas database. The data have a spatial resolution of 250 m. The data for Turkey obtained from the Global Wind Atlas are arranged according to the boundaries of the study area. The wind speed data are scaled from 1 to 5. Regions with low wind speeds were scored as 1 and regions with high wind speeds were scored as 5 (Figure 3 (c)).

3.2. Processing of geographic datasets

3.2.1. Geology data

In the study, the geological map of the study area was created in a geojson format using the drawing editor application from the content of the geoscientific map viewer of the General Directorate of Mineral Research and Exploration. The obtained map was digitized using the program GIS and used as a geological data layer. Alluvial fans, sand, dunes and debris cones are marked on the map as unsuitable areas for the construction site, while the other geological infrastructures are marked on the map as suitable areas. In this data layer, a value of 5 represents geologically suitable areas and a value of 1 represents unsuitable areas (Figure 3 (e)), Koc-San et al. (2013).

3.2.2. Landslide inventory data

The layer of landslides was created by the same processes as the layer of geology. There are many active landslides in the study area. Areas with landslides should be excluded when selecting a suitable site Koc-San et al. (2013). During the study, landslide-prone and landslide-free areas were scaled as 1 and 5, respectively (Figure 3 (f)).

3.2.3. Active fault lines data

In this study, active fault lines were identified as vectors using the geoscientific map viewer of the General Directorate of Mineral Research and Exploration. These vectors were converted to raster data, and locations 20000 m and farther from the fault lines were scaled as 5, and locations 5000 m and farther were scaled as 1. The distance map produced was used as active fault line data (Figure 3 (g)).

3.2.4. Digital elevation model

During the study, GTOPO30 data was used for the elevation information, two granules with the boundaries of Turkey were downloaded, these two granules were combined in the application GIS and the study area was cut out from the map. The data range in the study area is between 755 and 3412 m. Accordingly, these data were divided into five data groups on a scale of 1 to 5, with the highest points scaled as 5 and the lowest as 1. (Figure 3 (d)).

3.3. Processing of anthropogenic datasets

3.3.1. City lights data

Data from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) databases within the study area boundaries were used Elvidge et al. (2001). The data have a radiometric resolution of 32 bits. The data sets are scaled from 1 to 5. 1 represents the most illuminated areas, while 5 represents the least illuminated areas (Figure 3 (h)).

3.3.2. Mining activities data

The study area includes chrome, iron, manganese, copper-zinc-lead metal mines and industrial raw materials such as asbestos, barite, gypsum, magnesite, marble, perlite, bricks and sulfur. The dust layer around mining operations affects astronomical observations. For this reason, a grid map has been created to define mining areas. The regions where mining is carried out are divided into 5 scales in the range of 500-2000 m and above. Distances of 2000 m and more to mining sites are scaled by a value of 5, 500 m and less by a value of 1 (Figure 3(j)).

3.3.3. Road information data

Road proximity data were obtained from the OpenStreetMap application. The obtained data were converted from raster to vector data, and the road proximity parameters were defined and converted to raster data again. The raster data are divided into the distance ranges of 100 m and less and 500 m and more. Distances of 100 m and less are scaled by a value of 5, and distances of 500 m and more are scaled by a value of 1 (Figure 3(i)).

3.4. Analytic hierarchy process

The maps created for each of the 10 data in a raster data format were scaled between 1 and 5 and combined with the program GIS according to the weight ratios determined by the AHP method. The AHP method is based on comparing data layers with each other by numerical evaluation on an absolute number scale, Saaty (1980), and is a good way to address seemingly complex problems with a simple hierarchy, Cheng et al. (1999). While the individual criteria in the AHP are compared with each other, their importance levels are evaluated using odd numbers between 1 and 9, although even numbers may be preferred in cases where the determinant is ambiguous. The assigned criteria values appear in the AHP matrix (see tab. 1). The importance of the criteria was determined by expert decision-makers in the fields of astronomical observations, earth sciences, meteorology, and urban planning. When comparing the two criteria, the most important criterion determined by the experts was rated 9. It is rated 1 if the two criteria are equally important.

Table 1. A pairwise comparison scale based on Saaty's Fundamental Scale (Triantaphyllou et al. (1995))

Intensity of Importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
2	Weak or slight	
3	Moderate importance	Experience and judgement strongly favor one activity over another
4	Moderate plus	
5	Strong importance	Experience and judgement strongly favor one activity over another
6	Strong plus	
7	Very strong or demonstrated importance	An activity is favored very strongly over another; its dominance demonstrated in practice
8	Very, very strong	
9	Extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation

The specific AHP procedures are:

1. Determining the importance relationship between the indices and creating the overall AHP structure.

2. Creating the following comparison matrix A with the weights determined by the experts based on the scale from 1 to 9 (see tab. 2), Saaty (1990),

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & . & \dots & . \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \tag{1}$$

   where a_{mn} is the result of comparing the m'th factor with its importance relative to the nth factor, $a_{mn} > 0$, $a_{mn} = 1/a_{nm}$, $a_{mn} = 1$ (when $m = n$).

3. A is the comparison matrix, W and $w_i(i = 1, 2, ..., n)$ represent the corresponding eigenvector of λ_{max} and the weighting value for ranking, respectively and λ_{max} represents the largest eigenvalue of the matrix; in this study, the λ_{max} value was determined to be 10.911.

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & . & \dots & . \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_i \end{pmatrix} \tag{2}$$

$$\lambda_{max} = \frac{1}{n} \sum_{w_i}^{n} \frac{A_i.W_i}{W_i} \tag{3}$$

4. The consistency of the pairwise comparison matrix should be checked using the consistency index (C.I). The consistency check is performed (C.I, Equation 4). In this study, the C.I value was calculated as 0.107,

$$C.I = \frac{\lambda_{max} - 1}{n - 1}, \tag{4}$$

   In addition, to check the consistency of the created matrix, Saaty (1980), the consistency ratio (C.R) value should be determined. The C.R formula is shown below,

$$C.R = \frac{C.I}{R.I}, \tag{5}$$

   Where C.I is the consistency index and R.I is the random inconsistency, Saaty et al. (2008) and Saaty & Tran (2007). The critical value for the C.R value is 0.10. Values above this value are considered inconsistent, while values below this value are considered consistent, Saaty et al. (2008), Saaty & Tran (2007) and Koc-San et al. (2013). For n = 10, the value of the random

Table 2. The AHP pairwise comparison matrix and calculated weights.

	A	B	C	D	E	F	G	H	I	J
A	1	0.333	0.2	0.333	0.200	0.143	0.143	0.125	0.111	0.125
B	3	1	1	0.333	0.333	0.333	0.143	0.125	0.111	0.167
C	5	1	1	0.333	0.333	0.333	0.143	0.125	0.111	0.167
D	3	3	3	1	0.250	0.333	0.143	0.125	0.111	0.200
E	5	3	3	4	1	1	0.167	0.143	0.125	0.333
F	7	3	3	3	1	1	0.333	0.333	0.200	0.500
G	7	7	7	7	6	3	1	1	0.333	0.500
H	8	8	8	8	7	3	1	1	0.333	1
I	9	9	9	9	8	5	3	3	1	1
J	8	6	6	5	3	2	2	1	1	1

discrepancy is R.I=1.49, Saaty & Tran (2007). In our study, the C.R value was calculated as 0.068. The matrix obtained by the experts through pairwise comparisons (see tab. 2) is consistent and can be used in the analysis phase.

$\lambda_{max} = 10.911$, C.I $= 0.107$, R.I ($n = 10$) $= 1.49$, C.R $= 0.068$ (A: distance to roads, B: geology, C: landslide, D: distance to active faults, E: wind speed F: point density of mining sites, G: city lights, H: rainy water, I: cloud cover and J: DEM).

The 10 maps with data layers used in the study were overlaid in the program GIS in relation to the weighting ratios determined by the AHP method, and the most suitable areas (Figure 4, red areas) were proposed for the construction of the astronomical observatory.

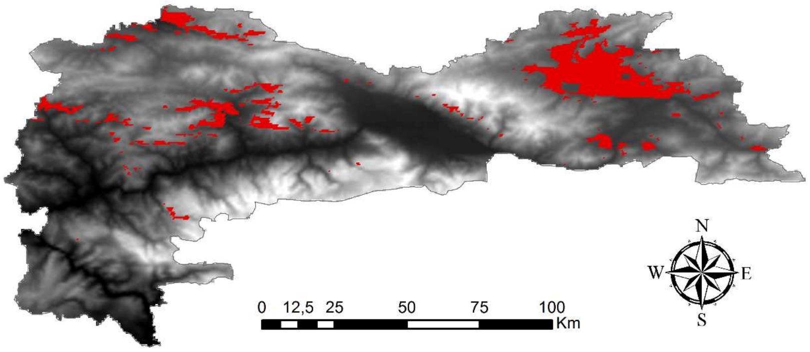

Figure 4. Map of candidate sites for the astronomical observatory covered with DEM (marked in red)).

4. Results and discussions

The most suitable site within the study area was determined by multiplying and overlaying each of the criteria maps (10 maps) obtained after the steps described above with the criteria weights calculated using AHP (Figure 4). The three main criteria identified, meteorological, geographic, and anthropogenic, are shown in red in the most suitable areas (Figure 4) In this study, according to experts, the most important parameters are meteorological datasets that contain information about cloud cover and precipitation water. 3-year image data (between 2020-2022) of these datasets were analyzed. Dealing with over 2000 images provides reliable information in the spatial domain, Koc-San et al. (2013). The 3-year meteorological dataset provides information reliability about the current meteorological condition of the study area and minimizes meteorological changes due to global warming. Statistically, 3 years of data (1095 data from the combination of 4380 granules) is satisfactory for this type of analysis, Koc-San et al. (2013).

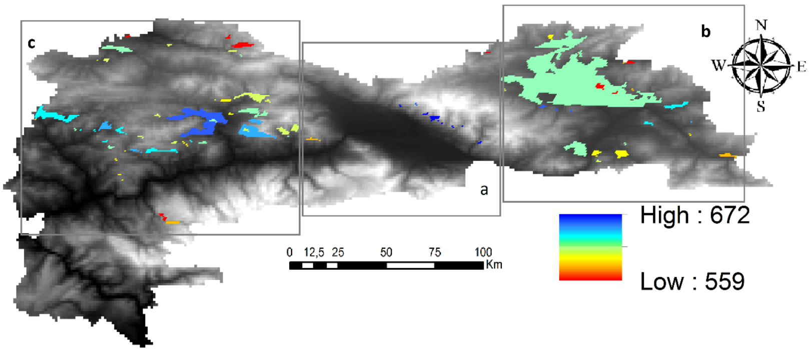

Figure 5. The number of cloud-free days obtained as a result of the 3-year data analysis.

As a result of the analysis of 3-year cloud cover data, the maximum and minimum values of the number of cloud-free days in the study area are shown in Figure 5. The map showing the number of cloud-free days was formed as a, b, c (see Figure 5). Among these classes, the highest number of cloud-free days was determined in the pixels in class a (598-672 days) (see Figure 6, blue pixels). The lowest number of days (559-632 days) was observed in pixels in class c. Although the pixels in class a are ideal for the number of cloud-free days, they are the candidate regions closest to the city center and the university campus. The closest candidate pixel to the city center is at least 8.39 km.

Precipitation water and water vapor is an effective parameter especially for astronomical observatories where observation activities are carried out at in-

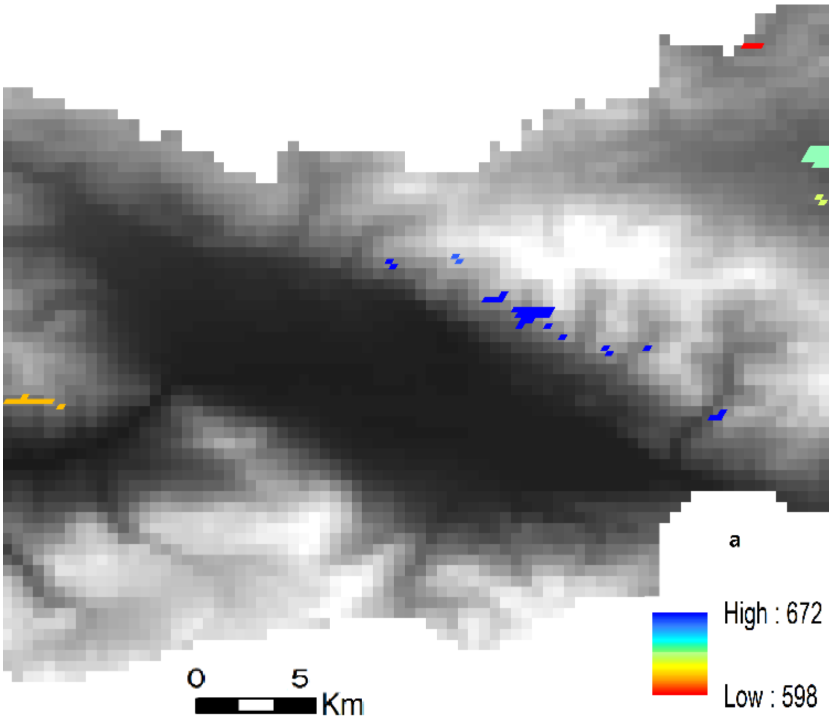

Figure 6. Maps of regions and the number of cloud-free days within the provincial borders of Erzincan.

frared wavelengths. Pérez-Jordán et al. (2015). As the water vapor gets closer to the earth, its relationship with the humidity value increases. Observations are not carried out when the humidity value exceeds 85% Solmaz et al. (2021). The annual average humidity of Erzincan Province was 64.26%, Kaya (2011). These average humidity values mean that observations can be made in general.

The smallest and maximum wind speed values within the provincial borders of Erzincan are shown in Figure 7. Wind speed is the limit value for 11 m/s observations, Solmaz et al. (2021). There are very few areas above the limit value within the provincial borders of Erzincan. However, wind speed values above the limit value are not among the candidate pixels (see Figure 7, red).

5. Conclusion and Recommendations

This study presents an exemplary application of the approach created by using MCDA, remote sensing and GIS technologies in selecting the most suitable site for the astronomical observatory as a result of analyzing complex criteria. This

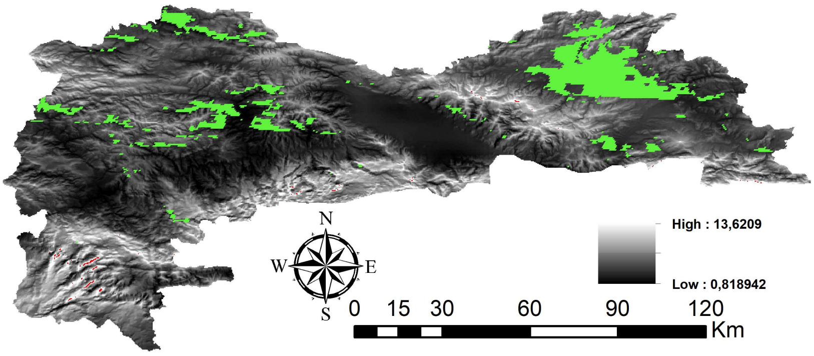

Figure 7. Candidate areas (green) and areas with wind speeds greater than 11m/s (red) on the wind speed map.)

approach is a powerful method for site locating an astronomical observatory. This methodology, supported by field trials, will provide temporal, economic, and energy advantages in determining the location of the observatory. Applicable for areas where the future observatory construction is planned.

Acknowledgements. I would like to thank the experts who contributed to the creation of the decision matrix of the AHP method.

References

Akbaş, B., Akdeniz, N., Aksay, A., et al., 1: 1.250. 000 Ölçekli Türkiye Jeoloji Haritası. 2011, *Maden Tetkik ve Arama Genel Müdürlüğü Yayını, Ankara-Türkiye*

Aman, D. D. & Aytac, G., Multi-criteria decision making for city-scale infrastructure of post-earthquake assembly areas: Case study of Istanbul. 2022, *International Journal of Disaster Risk Reduction*, **67**, 102668

Bal, M., Dandpat, A. K., & Naik, B., Hydrological modeling with respect to impact of land-use and land-cover change on the runoff dynamics in Budhabalanga river basing using ArcGIS and SWAT model. 2021, *Remote Sensing Applications: Society and Environment*, **23**, 100527

Barka, A. & Eyidoan, H., The Erzincan earthquake of 13 March 1992 in eastern Turkey. 1993, *Terra Nova*, **5**, 190

Barnes, W. L., Pagano, T. S., & Salomonson, V. V., Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1. 1998, *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 1088

Bilgilioglu, S. S., Gezgin, C., Orhan, O., & Karakus, P., A GIS-based multi-criteria decision-making method for the selection of potential municipal solid waste disposal sites in Mersin, Turkey. 2022, *Environmental Science and Pollution Research*, **29**, 5313

Cheng, C.-H., Yang, K.-L., & Hwang, C.-L., Evaluating attack helicopters by AHP based on linguistic variable weight. 1999, *European journal of operational research*, **116**, 423

Cowles, K., Site selection criteria for the optical atmospheric visibility monitoring telescopes. 1989, *The Telecommunications and Data Acquisition Report*

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R., Mapping city lights with nighttime data from the DMSP Operational Linescan System. 1997, *Photogrammetric Engineering and Remote Sensing*, **63**, 727

Elvidge, C. D., Imhoff, M. L., Baugh, K. E., et al., Night-time lights of the world: 1994–1995. 2001, *ISPRS Journal of Photogrammetry and Remote Sensing*, **56**, 81

Gao, B.-C. & Goetz, A. F., Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. 1990, *Journal of Geophysical Research: Atmospheres*, **95**, 3549

Gürsoy, Ö. 2012, Cndaki (erzincan Sivas-koyulhisar Arası) Litolojik Farklılık Ve Benzerliklerin Uzaktan Algılama Yöntemleri İle Belirlenmesi, PhD thesis, Fen Bilimleri Enstitüsü

Hall, D. K. & Riggs, G. A., Accuracy assessment of the MODIS snow products. 2007, *Hydrological Processes: An International Journal*, **21**, 1534

Hudson, K. & Simstad, T., The share astronomy guide to observatory site selection. 2010, *Neal Street Design Inc*, **1**

Jia, Y., Yong-qiang, Y., Hong-shuai, W., et al., Processing method of night-time cloudiness for astronomical site selection. 2012, *Chinese Astronomy and Astrophysics*, **36**, 457

Kaufman, Y. J. & Gao, B.-C., Remote sensing of water vapor in the near IR from EOS/MODIS. 1992, *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 871

Kaya, M., Erzincan iklim ve meteoroloji verileri. 2011, *Tesisat Mühendisliği Dergisi*, **34**, 42

Kern, A., Bartholy, J., Borbás, É. E., et al., Estimation of vertically integrated water vapor in Hungary using MODIS imagery. 2008, *Advances in space research*, **41**, 1933

Koc-San, D., San, B., Bakis, V., Helvaci, M., & Eker, Z., Multi-Criteria Decision Analysis integrated with GIS and remote sensing for astronomical observatory site selection in Antalya province, Turkey. 2013, *Advances in Space Research*, **52**, 39

Ma, D., Sandberg, M., & Jiang, B., Characterizing the heterogeneity of the OpenStreetMap data and community. 2015, *ISPRS International Journal of Geo-Information*, **4**, 535

Mekhaimr, S. A., Atmospheric conditions affecting seeing at St. Catherine: Estimation of operational time for NRIAG new telescope. 2017, *NRIAG Journal of Astronomy and Geophysics*, **6**, 5

Pérez-Jordán, G., Castro-Almazán, J., Muñoz-Tuñón, C., Codina, B., & Vernin, J., Forecasting the precipitable water vapour content: validation for astronomical observatories using radiosoundings. 2015, *Monthly Notices of the Royal Astronomical Society*, **452**, 1992

Ravi, S. & D'Odorico, P., A field-scale analysis of the dependence of wind erosion threshold velocity on air humidity. 2005, *Geophysical Research Letters*, **32**

Saaty, T., The analytic hierarchy process (AHP) for decision making. 1980, in *Kobe, Japan*, 1–69

Saaty, T. L., What is the analytic hierarchy process? Mathematical models for decision support. 1988, *NATO ASI Series*, **48**, 109

Saaty, T. L., How to make a decision: the analytic hierarchy process. 1990, *European journal of operational research*, **48**, 9

Saaty, T. L. & Tran, L. T., On the invalidity of fuzzifying numerical judgments in the Analytic Hierarchy Process. 2007, *Mathematical and Computer Modelling*, **46**, 962

Saaty, T. L. et al., Decision making with the analytic hierarchy process. 2008, *International journal of services sciences*, **1**, 83

Sobrino, J., El Kharraz, J., & Li, Z.-L., Surface temperature and water vapour retrieval from MODIS data. 2003, *International Journal of Remote Sensing*, **24**, 5161

Solmaz, A., Aksaker, N., Akyüz, A., et al., ÇÜ Uzay Bilimleri ve Güneş Enerjisi Araştırma ve Uygulama Merkezi (UZAYMER): I. Gözlem Koşulları ve Güncel Projeler. 2021, *Turkish Journal of Astronomy and Astrophysics*, **2**, 1

Strabala, K. I. 2005, *MODIS cloud mask user's guide* (University of Wisconsin–Madison)

Triantaphyllou, E., Mann, S. H., et al., Using the analytic hierarchy process for decision making in engineering applications: some challenges. 1995, *International journal of industrial engineering: applications and practice*, **2**, 35

Tulun, Ş., Arsu, T., & Gürbüz, E., Selection of the most suitable biogas facility location with the geographical information system and multi-criteria decision-making methods: a case study of Konya Closed Basin, Turkey. 2022, *Biomass Conversion and Biorefinery*, 1

Üzer, M., Emre, Ö., Duman, T. Y., et al. 2013, *Açıklamalı Türkiye Diri Fay Haritası, Ölçek 1: 1.250. 000* (Maden Tetkik ve Arama Genel Müdürlüğü)

Verileri, E. İ. U. Y. M., Çevre ve Orman Bakanlığı. 2010, *Devlet Meteoroloji İşleri Genel Müdürlüğü*

Walker, M. F., High quality astronomical sites around the world. 1984, in European Southern Observatory Conference and Workshop Proceedings, Vol. **18**, *European Southern Observatory Conference and Workshop Proceedings*, 3–21

Yilmaz, A., Yukari Kelkit cayi ile Munzur daglari arasinin temel jeoloji özellikleri ve yapisal evrimi. 1985, *Türk jeoloji Kurumu Bülteni*, **28**, 79

Zhang, Q. & Seto, K. C., Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. 2011, *Remote Sensing of Environment*, **115**, 2320

Contrib. Astron. Obs. Skalnaté Pleso **53/1**, 46 – 57, (2023)
https://doi.org/10.31577/caosp.2023.53.1.46

Optimal conditions of the spacecraft acceleration in the gravitational field of planet

M. Vavrukh and D. Dzikovskyi

Department of Astrophysics, Ivan Franko National University, Kyrylo & Methodiy str. 8, 79005 Lviv, Ukraine, (E-mail: mvavrukh@gmail.com)

Received: March 9, 2023; Accepted: March 27, 2023

Abstract. In the approximation of Laplace's sphere influence method, there were established the optimal conditions of the spacecraft acceleration in the gravitational field of the planet, which are determined by the value of the spacecraft velocity on the Earth orbit, as well as by the character of Keplerian trajectory in the sphere of the Sun's influence and the impact parameter relative to the planet. As an example, there was calculated the acceleration of the spacecraft in the Jupiter's field for initial velocities from the Earth's orbit in the range $(40 \div 50)\ \mathrm{km\,s^{-1}}$.

Key words: methods: analytical – celestial mechanics – space vehicles

The mechanism of the spacecrafts acceleration in the gravitational fields of planets has already been used for a long time in the studies of Solar System periphery (Barger & Olsson, 1995; Bartlett & Hord, 1985; Diehl, 1996; Media Relations Office, 1999). The explanation of this effect is based on the usage of the laws of conservation of energy and angular momentum of the body, which moves in the centrosymmetric gravitational field of the Sun and the planet (Barger & Olsson, 1995; Thornton & Marion, 2004). In the strict sense, this is a three-body problem – the spacecraft, the Sun and a planet. To simplify the problem, it is assumed that the motion of the planet and the spacecraft occurs in the same plane, as it is shown in Fig. 1, and orbits of the Earth and the planet are circular. The spacecraft starts from the Earth's orbit (p. A_0), having the velocity $\mathbf{v}_0$, which is orthogonal to the radius vector of the spacecraft in the heliocentric reference frame. Under these conditions, all elements of the trajectory of the spacecraft in the sphere of the Sun's influence are determined by the magnitude of velocity v_0. Having reached the sphere of the planet's influence (point A_1), the spacecraft has the radius vector $\mathbf{r}_1$ and the velocity $\mathbf{v}_i$, which are the initial conditions to describe the motion of the spacecraft in the gravitational field of the planet. The radius vector $\mathbf{r}_1$ is determined by the shape of the heliocentric trajectory of the spacecraft, and its velocity $\mathbf{v}_i$ is found from the law of conservation of energy,

$$\frac{v_i^2}{2} - \frac{GM_\odot}{r_1} = \frac{v_0^2}{2} - \frac{GM_\odot}{a_E}, \tag{1}$$

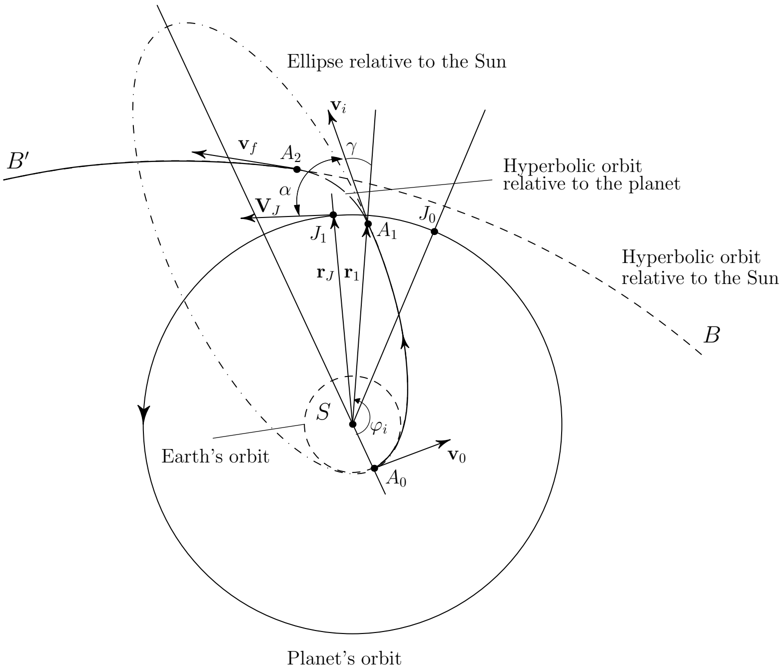

Figure 1. The schematic representation of the spacecraft motion trajectory in the field of the Sun and the planet.

where $a_E = 1$ astronomical unit is the radius of the Earth's orbit.

For definiteness, we will consider that the acceleration occurs in the gravitational field of Jupiter. In order to explore all features of the acceleration effect, we will consider the motion of the spacecraft with the initial velocity in the interval of $40\ \mathrm{km\,s^{-1}} \leq v_0 \leq 50\ \mathrm{km\,s^{-1}}$. This allows us to describe all three types of Keplerian trajectories – elliptical ($v_0 < v_p$), parabolic $v_0 = v_p = (2GM_\odot/a_E)^{1/2} \simeq 42.19\ \mathrm{km\,s^{-1}}$) and hyperbolic ($v_0 > v_p$). Elements of the heliocentric trajectories of the spacecraft for a given velocity v_0 are shown in Tab. 1: the focal parameter p in the astronomical units, eccentricity e, and the module of velocity in the sphere of the planet's influence v_i. Cosine of the angle α between the velocity vector of the planet $\mathbf{V}_J$ and the velocity vector of the spacecraft $\mathbf{v}_i$ is found from the law of conservation of angular momentum,

$$a_E v_0 \ \approx\ r_1 v_i \ \cos\alpha. \tag{2}$$

Fig. 1 schematically illustrates the situation for the elliptical trajectory of the spacecraft, $\alpha \approx \pi/2 - \gamma$, where γ is the angle between the radius vector $\mathbf{r}_1$ and the velocity vector $\mathbf{v}_i$.

From the point A_1 to the point A_2 the spacecraft moves along the transitional hyperbolic planetocentric trajectory in the sphere of the planet's influence. In the point A_2 it reaches the velocity $\mathbf{v}_f$ and leaves the sphere of the planet's influence. Herewith, the angle between the vectors $\mathbf{v}_f$ and $\mathbf{V}_J$ equals $\alpha^{'}$. The further motion of the spacecraft occurs along the hyperbolic heliocentric trajectory whose focus is the point S.

1. Kinematic analysis of the problem

Vectors $\mathbf{V}_J$, $\mathbf{v}_i$ and $\mathbf{v}_f$ denote velocities of planets and the spacecraft in the heliocentric reference frame. It is more convenient to describe the motion of the spacecraft inside the sphere of the planet's influence in the system of mass center, in which the planet is fixed and vectors

$$\boldsymbol{\rho} = \mathbf{r} - \mathbf{R}_J, \;\; \mathbf{u} = \mathbf{v} - \mathbf{V}_J \tag{3}$$

describe the position and velocity of the spacecraft relative to the mass center of the planet. In particular, vectors

$$\mathbf{u}_i = \mathbf{v}_i - \mathbf{V}_J, \;\; \mathbf{u}_f = \mathbf{v}_f - \mathbf{V}_J \tag{4}$$

determine the relative velocity of the spacecraft at the entrance to the sphere of the planet's influence and at the exit from it. The vector

$$\mathbf{c} = \mathbf{v}_f - \mathbf{v}_i = \mathbf{u}_f - \mathbf{u}_i \tag{5}$$

determines the velocity momentum, which is acquired by the spacecraft in the gravitational field of the planet. The difference

$$\Delta E = \frac{m}{2}\,\left(\mathbf{v}_f^2 - \mathbf{v}_i^2\right) = m(\mathbf{c}, \mathbf{V}_J) = m(\mathbf{u}_f - \mathbf{u}_i - \mathbf{V}_J) \tag{6}$$

is the acquired or lost energy of the spacecraft after it passed through the sphere of the planet's influence, depending on the sign of the product $(\mathbf{c}, \mathbf{V}_J)$. Since the energy of motion of the spacecraft in the system of mass center is a constant value, then

$$|\mathbf{u}_f| = |\mathbf{u}_i| = \mathbf{u}. \tag{7}$$

Of course, during the motion of the spacecraft along the transition trajectory, the direction of the vector $\mathbf{u}$ also changes, as well as its value. According to formulae (5) and (7)

$$|\mathbf{c}| = 2^{1/2} u (1 - \cos\beta)^{1/2}, \tag{8}$$

where β is the angle between vectors $\mathbf{u}_i$ and $\mathbf{u}_f$. According to formula (4)

$$u = \left[\mathbf{v}_i^2 + \mathbf{V}_J^2 - 2(\mathbf{v}_i, \mathbf{V}_J)\right]^{1/2} = \left(v_i^2 + V_J^2 - 2 v_i V_J \cos\alpha\right)^{1/2}. \tag{9}$$

Fig. 1 corresponds to the situation when the spacecraft is “catching up” the planet (the trajectory I in Fig. 2). There is possible another situation when the spacecraft is “overtaking” the planet (the trajectory II in Fig. 2). In the case I, during the passage of the spacecraft in the sphere of the planet’s influence there prevails the condition $(\mathbf{a}_1, \mathbf{V}_J) > 0$, where $\mathbf{a}_1$ is the acceleration vector of the spacecraft, which is caused by the gravitational influence from the side of the planet. As the result, the spacecraft accelerates. In the case II the condition $(\mathbf{a}_2, \mathbf{V}_J) < 0$ prevails, which corresponds to the braking. In this work we do not consider the trajectories of type II, which can be used for the correction of the spacecraft trajectory, but not for its acceleration.

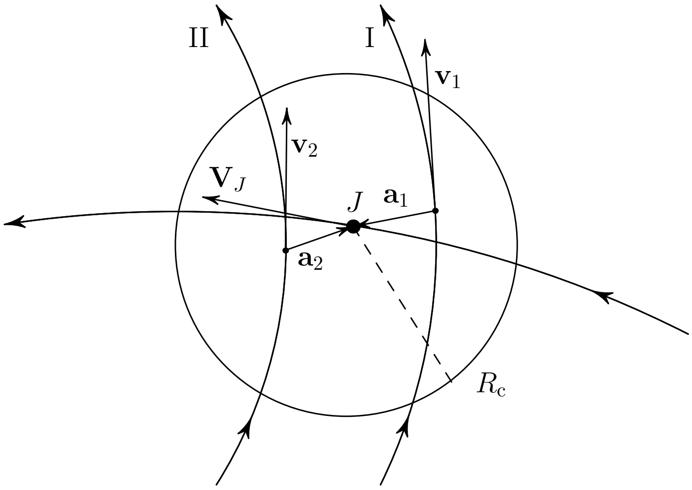

Figure 2. The schematic representation of two types of the spacecraft motion trajectory in the field of the planet.

Bartlett & Hord (1985) used a simplified description of the gravitational planet’s influence on the spacecraft, which is timed as the elastic collision of two point masses, and the radius of the sphere of the planet’s influence is zero. Such a model approach allows us not only to prove the existence of the acceleration effect, but also to estimate its magnitude. However, it is not enough to choose the optimal acceleration conditions, which requires a more detailed description of its motion in the gravitational field of the planet.

Following the method of Barger & Olsson (1995) and Johnson (2003), we consider the vector diagram of velocity shown in Fig. 3. According to formulae (4), (5) and (8), we obtain the relation

$$2(1-\cos\beta)\big(v_i^2 + V_J^2 - 2v_iV_J\cos\alpha\big) = v_f^2 + v_i^2 - 2v_iv_f\cos(\alpha-\alpha^{'}), \qquad (10)$$

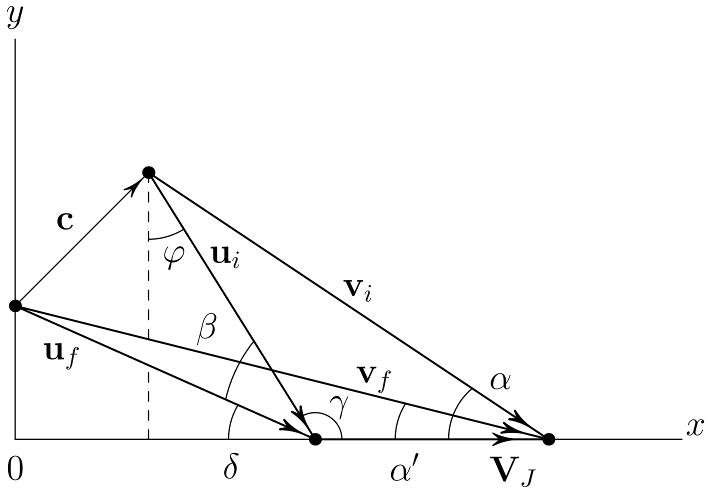

Figure 3. The vector diagram of the spacecraft velocity in two reference frames.

which allows us to determine the value of final velocity v_f. However, in Fig. 3 we see that variables $\alpha, \alpha^{'}$ and β are not independent. To establish the relation between them, we calculate the projections of vectors $\mathbf{v}_f$ and $\mathbf{v}_i$ on the axis of the Cartesian coordinate system according to Fig. 3. As it is shown in the figure

$$\beta + \delta + \gamma = \pi, \;\; \varphi = \gamma - \frac{\pi}{2}. \tag{11}$$

Therefore

$$\begin{aligned} v_f \cos \alpha^{'} &= V_J - u \cos(\beta + \gamma), \\ v_f \sin \alpha^{'} &= u \sin(\beta + \gamma), \end{aligned} \tag{12}$$

as well as

$$\begin{aligned} v_i \cos \alpha &= V_J - u \cos \gamma, \\ v_i \sin \alpha &= u \sin \gamma. \end{aligned} \tag{13}$$

Determining $\sin \gamma$ and $\cos \gamma$ from relations (13) and substituting them in formulae (12), we find that

$$\begin{aligned} v_f \cos \alpha^{'} &= V_J(1 - \cos \beta) + v_i \cos(\alpha - \beta), \\ v_f \sin \alpha^{'} &= \sin \beta (V_J - v_i \cos \alpha) + v_i \cos \beta \sin \alpha, \end{aligned} \tag{14}$$

which allows us to exclude the angle α' from relation (10). Substituting the obtained value in equation (10), we reduce it to the expression

$$v_f^2 = v_i^2 + 2V_J^2(1-\cos\beta) + 2v_i V_J\big[\cos(\alpha-\beta) - \cos\alpha\big]. \tag{15}$$

Such equation for finding the value of the final velocity was obtained in the work of Johnson (2003). In equation (15) the values v_i and α are considered as known. From the condition of the extremum of expression (15) relative to the angle β, we obtain the relation (Johnson, 2003)

$$\sin\beta\big(V_J - v_i\cos\alpha\big) + v_i\sin\alpha\cos\beta = 0, \tag{16}$$

from where we find the extreme value of the angle β

$$\tan\beta_{\text{ext}} = \frac{k_i\sin\alpha}{k_i\cos\alpha - 1}. \tag{17}$$

For convenience, we introduced here the dimensionless velocities $k_i = v_i/V_J$, $k_f = v_f/V_J$. It follows from equation (17) that

$$\begin{aligned} \cos\beta_{\text{ext}} &= \frac{k_i/k_\alpha - 1}{(1+k_i^2-2k_i/k_\alpha)^{1/2}}, \\ \sin\beta_{\text{ext}} &= \frac{(1-k_\alpha^{-2})^{1/2}}{(1+k_i^2-2k_i/k_\alpha)^{1/2}}, \end{aligned} \tag{18}$$

where $k_\alpha = (\cos\alpha)^{-1}$. It is easy to see that

$$\left.\frac{\partial^2 v_f^2}{\partial\beta^2}\right|_{\beta_{\text{ext}}} = -2V_J^2\,\big(1+k_i^2-2k_i/k_\alpha\big)^{1/2}, \tag{19}$$

from where it follows that v_f has the maximum at $\beta = \beta_{\text{ext}}$.

As it was shown from formulae (14) and (16), at β_{ext} the angle $\alpha^{'} = 0$ and v_f takes the extreme value, according to which

$$(k_f)_{\max} = 1 + \big(1+k_i^2-2k_i/k_\alpha\big)^{1/2} = 1 + \frac{u}{V_J}, \tag{20}$$

or

$$|\mathbf{v}_f|_{\max} = V_J + u. \tag{21}$$

Note that instead of the extremum condition (17) we can choose $\alpha^{'} = 0$ in equations (14). From relations (17) and (20) it follows that at $\alpha = 0$

$$\beta_{\text{ext}} = 0, \;\; (k_f)_{\max} = k_i \tag{22}$$

and there is no acceleration of the spacecraft. At $\alpha = \pi/2$

$$\tan\beta_{\text{ext}} = -k_i, \;\; (k_f)_{\max} = 1 + (1+k_i^2)^{1/2} \tag{23}$$

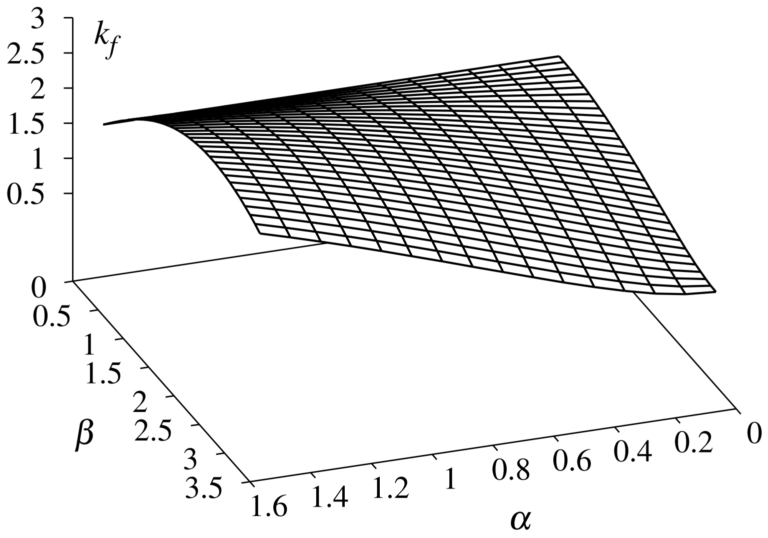

Figure 4. The module of the spacecraft velocity at the exit from the sphere of the planet's influence (in units $|\mathbf{V}_J|$) as a function of angles α and β (in radians).

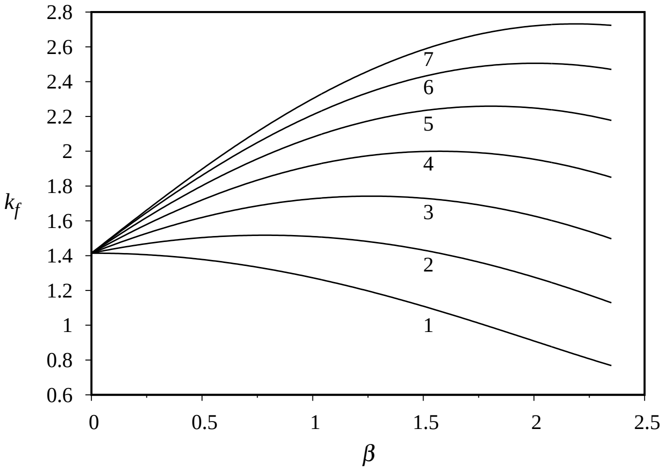

Figure 5. The same as in Fig. 4. Curve 1 corresponds to $\alpha = 0°$, curve 2 – $\alpha = 15°$, curve 3 – $\alpha = 30°$, ..., curve 7 corresponds to $\alpha = 90°$.

and acceleration is maximal.

A general character of the dependence of k_f on the angles α and β according to relations (15) is illustrated in Fig. 4. Herewith α changes in the region $(0 \div \pi/2)$, β – in the region $(0 \div 3/4\,\pi)$, and $k_i = \sqrt{2}$. As it is shown in the Figure, for $\alpha = 0$ there is no extremum, and for $\alpha \neq 0$ it always exists, moreover the extremal value k_f is bigger, the bigger the angle α is. This is also clearly shown in Fig. 5 with a given value of k_f as a function of the angle β at a fixed value of the angle α. In the Figure it is shown that the extremal value of the angle β corresponds to that given by formulae (18), and the value k_f agrees with relation (20).

However, it follows from Fig. 1 and relation (3) that the value of the angle α is close to $\pi/2$, which cannot occur in the case of elliptical motion of the spacecraft in the sphere of the Sun's influence. It is also impossible to implement it with other types of trajectories, because it contradicts the law of conservation of angular momentum (2). To estimate the real value of velocity increase of the spacecraft in the gravitational field of the planet, there should be considered more detailed the initial conditions of the problem in the sphere of the Sun's influence.

2. The motion of the spacecraft in the sphere of the Sun's influence

The radius of the sphere of the Jupiter's influence relative to the Sun we evaluate from the expression (Thornton & Marion, 2004)

$$R_c = a_J \left(\frac{M_J}{M_\odot}\right)^{2/5} \approx 0.33 \text{ a. u.}, \tag{24}$$

where a_J is the orbital radius of the planet. Due to the fact that the impact parameter of the spacecraft relative to Jupiter is much smaller than R_c (as we will see below), let us adopt $r_1 \approx a_J - R_c$ for the calculation of v_i by formula (1). Therefore, in the case of the initial velocity $v_0 = 41$ km s^{-1}, which corresponds to an elliptical trajectory with a semi-major axis $a \approx 9.2$ a. u., we find that

$$v_i = 16.325 \text{ km s}^{-1}. \tag{25}$$

At the orbital velocity of Jupiter $V_J = 13.3$ km s^{-1}, we obtain $k_i = v_i/V_J \approx 1.2274$. If it were possible to get $\alpha \approx \pi/2$, then according to formula (20) we would obtain

$$(k_f)_{\max} = 2.58, \quad v_f \approx 34.31 \text{km s}^{-1}. \tag{26}$$

But with the approximations described above, from equation (2) we find that

$$\cos\alpha \approx 0.5157\ldots, \quad \alpha \approx 60^\circ. \tag{27}$$

Therefore, according to formula (20) we obtain

$$(k_f)_{\max} \approx 2.1138\ldots, \quad v_f \approx 28.1142 \text{ km s}^{-1}. \tag{28}$$

Such a value α corresponds to $\beta_{\text{ext}} \approx 109^\circ.238$.

In the case of the parabolic trajectory of the spacecraft (with initial velocity $v_0 = 42.19 \text{ km s}^{-1}$), we find that

$$v_i \approx 19.118 \text{ km s}^{-1}, \ \ k_i \approx 1.437. \tag{29}$$

According to relation (2)

$$\cos\alpha \approx 0.459\ldots, \ \ (k_f)_{\max} \approx 2.328\ldots, \ \ v_f = 30.962 \text{ km s}^{-1}. \tag{30}$$

The elements of the planetocentric trajectories at different values of the initial velocity v_0 are shown in Tab. 2. As it is shown in the Table, with the increasing initial velocity the angle α increases, and the angle β_{ext} monotonously decreases, approaching $\pi/2$. With the increasing v_0 the ratio v_f/v_0 increases: it is close to 0.7 for elliptical and parabolic trajectories, but for $v_0 = 50 \text{ km s}^{-1}$ the ratio is close to 0.9. As it can also be seen in Tab. 2, with increasing v_0 the module of the vector $\mathbf{u}_i$ increases: for the parabolic trajectory it equals 17.662 km s^{-1} and for the hyperbolic one at $v_0 = 50 \text{ km s}^{-1}$ it is already 31.463 km s^{-1}.

To provide the spacecraft with rotation of the vector of relative velocity in the sphere of the Jupiter's influence by the angle β_{ext}, it is necessary to choose appropriately its position of the heliocentric trajectory relative to the planet. In order for the spacecraft not to be captured by the planet's field, it should move on a parabolic or a hyperbolic trajectory inside the sphere of the Jupiter's influence. In the case of a hyperbolic trajectory,

$$\rho = \frac{p_{\text{H}}}{1 + e_{\text{H}}\cos\varphi}, \tag{31}$$

which is schematically shown in Fig. 6, the focal parameter and the eccentricity are determined by the value of the impact parameter b and the module of the relative velocity $\mathbf{u}_i$:

$$p_{\text{H}} = \frac{u^2 b^2}{GM_J}, \ \ e_{\text{H}} = \left[1 + \frac{u^4 b^2}{(GM_J)^2}\right]^{1/2}. \tag{32}$$

Since the impact parameter is much smaller than the radius of the influence sphere R_c, the angle $\beta_{\text{ext}} = \pi - 2\omega \approx \pi - 2(\pi - \varphi(R_c))$, where $\varphi(R_c)$ is found from relation (31),

$$\varphi(R_c) = \arccos\left[\left(\frac{p_{\text{H}}}{R_c} - 1\right)\frac{1}{e_{\text{H}}}\right]. \tag{33}$$

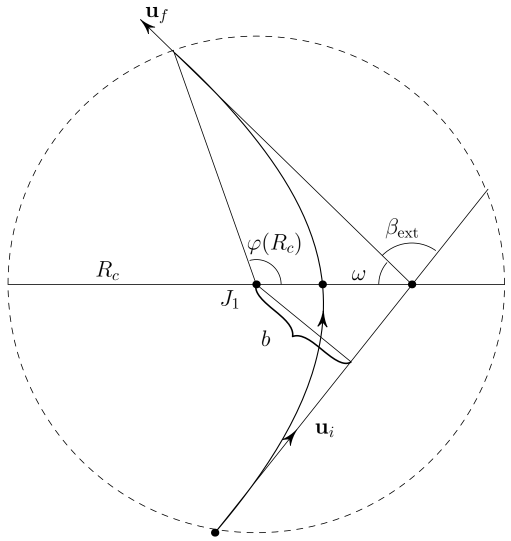

Figure 6. The transitional hyperbolic spacecraft trajectory in the sphere of the Jupiter's influence.

From this it follows the condition for determining the impact parameter

$$2\arccos\left[\left(\frac{p_{\text{H}}}{R_c}-1\right)\frac{1}{e_{\text{H}}}\right]=\pi+\beta_{\text{ext}}. \tag{34}$$

Putting $b = n \cdot R_J$, where $R_J = 71492$ km is the average radius of Jupiter, from equation (34) we find the value of the coefficient n for different initial values v_0. In Tab. 2 there is also shown the distance from the center of the planet to the pericenter of the transition trajectory,

Table 1. The elements of the heliocentric trajectories.

	v_0	p, a. u.	e	v_i	$\cos\alpha$	v_f
1	40	1.7978	0.7978	13.620	0.6031	25.2965
2	41	1.8888	0.8888	16.325	0.5157	28.1142
3	42.19	2.0	1.0	19.118	0.4531	30.9621
4	45	2.2753	1.2753	24.708	0.3740	36.5724
5	50	2.8090	1.8090	32.947	0.3116	44.7533

$$\rho_{\min}=p_{\text{H}}(1+e_{\text{H}})^{-1}, \tag{35}$$

and the module of the relative velocity in the vicinity of pericenter,

$$u(\rho_{\min}) = \left(u^2 + u_p^2(\rho_{\min})\right)^{1/2}, \tag{36}$$

where $u_p(\rho_{\min}) = (2GM_J/\rho_{\min})^{1/2}$ is the parabolic velocity at the distance $\rho_{\min}$ from the center of the planet. All velocities in the Tables are shown in the units $\mathrm{km\,s^{-1}}$.

Table 2. The elements of the planetocentric trajectories.

	u	β_{ext}	$n = b/R_J$	e_{H}	p_{H}, 10^5 km	$\rho_{\min}$, 10^5 km	$u(\rho_{\min})$
1	11.9965	115.087°	8.203	1.1851	37.2970	17.0691	17.3032
2	14.8140	109.238°	6.007	1.2265	30.4986	13.6980	20.3275
3	17.6621	105.220°	4.5477	1.2586	24.8478	11.0013	23.5201
4	23.2724	100.046°	2.8734	1.3050	17.2224	7.4719	29.9467
5	31.4531	95.53°	1.7033	1.3506	11.0542	4.7027	39.4164

3. Conclusions

1. The module of the velocity vector of the spacecraft at the entrance in the sphere of the planet's influence v_i and its orientation relative to the vector of the planet velocity (the angle $\alpha = (\mathbf{v}_i, \mathbf{V}_J)$) are the initial conditions for the description of motion of the spacecraft in the sphere of the planet's influence. These two values are determined by the values of the initial velocity $\mathbf{v}_0$.

2. It follows from the analysis of the velocities diagram that the module of the final velocity of the spacecraft (at the exit from the sphere of the planet's influence) v_f is the function of v_i and angles α, $\alpha' = (\mathbf{v}_f \widehat{,} \mathbf{V}_J)$, $\beta = (\mathbf{u}_i \widehat{,} \mathbf{u}_f)$. The maximal value of v_f can be determined in two ways, which lead to the identical results. In the first case we can choose $\alpha' = 0$, which determines some value of the angle β_{ext}, that is a function of variables v_i and α_i and yields the maximal value of v_f. Another way is to take into account the relation between angles α, α' and β and express α' by α and β. This gives relation (15) with variables v_i, α and β. As it is shown in Figs. 4 and 5, expression (15) has the extremum relative to the variable β for all $\alpha \neq 0$. The extreme value β_{ext} determines the maximal value of the final velocity as a function of variables v_i and α.

3. The velocities diagram determines the possibility of such a value of β_{ext} that provides the maximal value of the final velocity. For realization of this opportunity the spacecraft in the sphere of the planet's influence must move along the appropriate trajectory (31), which is determined by the impact parameter b relative to the center of the planet. This requirement leads to

equation (34), whose root determines the impact parameter and the elements of the hyperbolic trajectory inside the sphere of the planet's influence.

4. As it is shown in Tab. 1, the angle α increases with the increasing initial velocity and goes to $\pi/2$. But the limit $\alpha = \pi/2$ is unattainable, because this would violate the law of conservation of angular momentum. The angle β_{ext} is greater than $\pi/2$ and continuously approaches to the value $\pi/2$ with increasing v_0.

5. The impact parameter $b = nR_J$ has the order $(10^5 \div 10^6)$ km and decreases with increasing v_0.

6. The ratio v_f/v_0, which has the meaning of the energy conservation efficiency, monotonously increases with increasing initial velocity: for $v_0 = 40\,\mathrm{km\,s^{-1}}$ it equals 0.632, and for $v_0 = 50\,\mathrm{km\,s^{-1}}$ it is already 0.894.

7. We have analyzed the case when the spacecraft moves along by the Keplerian trajectory by inertia, in the absence of jet thrust. As it is shown in Fig. 5, there is an obvious possibility of achieving an even bigger final velocity of the spacecraft, if the correction of the trajectory is made before entering in the sphere of the planet's influence in order to increase the angle α. This allows us to decrease the value of the initial velocity.

References

Barger, V. & Olsson, M. 1995, *Classical Mechanics: A Modern Perspective. Second Edition*

Bartlett, A. A. & Hord, C. W., The slingshot effect: explanation and analogies. 1985, *The Physics Teacher*, **23**, 466, DOI: 10.1119/1.2341882

Diehl, R. E., Gravitational Assist. 1996, in *MacMillan Encyclopedia of Physics* (MacMillan Reference), 675–6

Johnson, R. C. 2003, The slingshot effect; https://www.flickr.com/photos/johnsonrc/2878215149/in/album-72157607426633696, last accessed 08 March 2023

Media Relations Office. 1999, Cassini Completes Earth Flyby, August 17, 1999; https://solarsystem.nasa.gov/news/12192/cassini-completes-earth-flyby, last accessed 08 March 2023

Thornton, S. T. & Marion, J. B. 2004, *Classical Dynamics of Particles and Systems. Fifth Edition*

Contrib. Astron. Obs. Skalnaté Pleso **53/1**, 58 – 66, (2023)
https://doi.org/10.31577/caosp.2023.53.1.58

Analytical images of Kepler's equation solutions and its analogues

M. Vavrukh and D. Dzikovskyi

Department of Astrophysics, Ivan Franko National University, Kyrylo & Methodiy str. 8, 79005 Lviv, Ukraine, (E-mail: mvavrukh@gmail.com)

Received: May 17, 2023; Accepted: May 29, 2023

Abstract. Approximate, but highly accurate analytical solutions of Kepler's equation were found by reducing it to an algebraic equation. With the help of this approach and usage of the iterative algorithm there were obtained solutions of a similar equation for the hyperbolic motion near the orbit pericenter.

Key words: methods: analytical – celestial mechanics

1. Introduction

Kepler's equation is one of the main relations in celestial mechanics, as Barker's equation and a similar equation for the case of hyperbolic motion. It is well known from celestial mechanics that the relative motion of two point-like gravitating bodies with masses m_1 and m_2 occurs on Keplerian orbits (Barger & Olsson, 1995; Vallado & McClain, 2001)

$$r = p\{1 + e \cdot \cos v\}^{-1}, \tag{1}$$

where r and v are polar coordinates, the focal parameter p and eccentricity e are determined by the masses of bodies and integrals of motion – angular momentum l and energy $\mathcal{E}$

$$p = l^2\mu^{-2}K^{-1}, \;\; e = \left\{1 + 2\mathcal{E}l^2\mu^{-3}K^{-2}\right\}^{1/2}. \tag{2}$$

Here $K = G(m_1 + m_2)$ is the so-called gravitational parameter, G is the gravitational constant and $\mu = m_1 m_2 (m_1 + m_2)^{-1}$. Using an expression for the angular momentum

$$l = \mu r^2 \frac{dv}{dt} \tag{3}$$

and relation (2), we obtain a well known equation for the time dependence of true anomaly $v(t)$ in the form

$$p^{3/2}K^{-1/2}\int\limits_0^{v(t)} [1 + e \cdot \cos v]^{-2} dv = t. \tag{4}$$

Here, taking into account that $v(0) = 0$, equation (4) determines the time of motion from the orbit pericenter to the point with a fixed value of $v(t)$. The integral in equation (4) is expressed by elementary functions, but it has different images depending on the eccentricity value. The trivial case $e = 0$ corresponds to the uniform motion on a circular orbit of radius p with the angular velocity

$$\omega_p = \frac{K^{1/2}}{p^{3/2}}, \tag{5}$$

and $v(t) = \omega_p t$. At non-zero eccentricities, the integral in equation (4) is calculated with the help of a universal substitution $x = \tan(v/2)$. In the case $e = 1$ and $\mathcal{E} = 0$ the motion occurs on a parabolic orbit, and equation (4) takes the form (Barger & Olsson, 1995)

$$\tan\frac{v}{2} + \frac{1}{3}\tan^3\frac{v}{2} = \frac{2K^{1/2}}{p^{3/2}}t \equiv 2\omega_p t. \tag{6}$$

This equation is known as Barker's equation. When the time varies in the region $-\infty < t < \infty$, the true anomaly changes in the interval $-\pi < v < \pi$.

For $0 < e < 1$, equation (4) takes the form

$$\left(\frac{1-e}{1+e}\right)^{1/2}\tan\frac{v}{2} = \frac{t_*}{2} + \frac{e}{2}\,(1-e^2)^{1/2}\frac{\sin v}{1+e\cdot\cos v}, \tag{7}$$

where $t_* = 2\pi t/T = tK^{1/2}/p^{3/2}(1-e^2)^{3/2}$ (T is the orbital period of an elliptical motion). For $e > 1$, equation (4) has the following form in elementary functions

$$\ln\frac{[\sqrt{e+1} - \sqrt{e-1}\ \tan(v/2)]}{[\sqrt{e-1} + \sqrt{e+1}\ \tan(v/2)]} = t_{\text{H}} - e(e^2-1)^{1/2}\frac{\sin v}{1+e\cdot\cos v}, \tag{8}$$

where

$$t_{\text{H}} = t\frac{K^{1/2}(e^2-1)^{3/2}}{p_{\text{H}}^{3/2}} \equiv \omega_{\text{H}} t. \tag{9}$$

Solving equation (6) relative to $\tan(v/2)$ by Cardano's formulae (Abramowitz & Stegun, 1972), we find an exact solution for the true anomaly,

$$v(t) = 2\arctan\left\{\left[(1+s^2(t))^{1/2} + s(t)\right]^{1/3} - \left[(1+s^2(t))^{1/2} - s(t)\right]^{1/3}\right\}, \tag{10}$$

where

$$s(t) = \frac{3K^{1/2}}{p^{3/2}}t = 3\omega_p t, \tag{11}$$

and $v(-t) = -v(t)$. Function (10) has the asymptotics

$$v(t) \to \begin{cases} 4\omega_p t \ \text{ at } \ |\omega_p t| \ll \pi/2, \\ 2\arctan(6\omega_p\, t) \ \text{ at } \ |\omega_p\, t| \gg \pi/2. \end{cases} \tag{12}$$

For the point which follows a parabolic path, the function $v(t)$ is proportional to the time if the point is near the pericenter; if it is far away – then $v(t) \to \pm\pi$. This behavior is illustrated in Fig. 1, which depicts $v(t)$ as a function of the dimensionless variable $\omega_p t$.

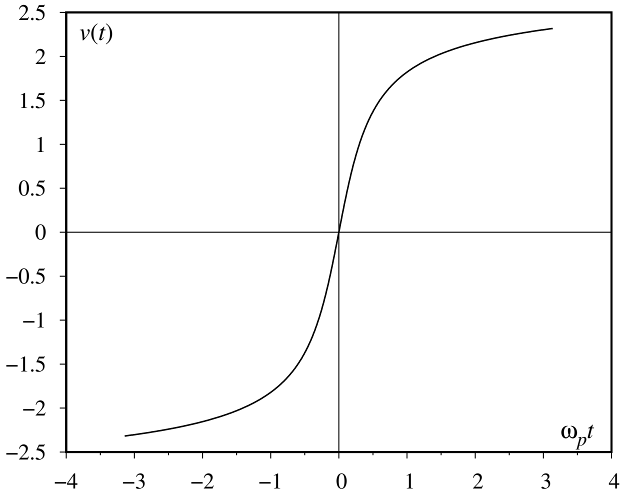

Figure 1. Dependence of the true anomaly $v(t)$ on the dimensionless variable $\omega_p t$.

2. Equation of elliptical motion

Using the substitution

$$\left(\frac{1-e}{1+e}\right)^{1/2} \tan\frac{v}{2} = \tan\frac{E}{2} \tag{13}$$

equation (7) is reduced to Kepler's equation (Vallado & McClain, 2001)

$$E - e \cdot \sin E = t_*, \tag{14}$$

and the additional function E is called an eccentric anomaly. As it is shown from equation (14), $E(-t_*) = -E(t_*)$. Many works are devoted to finding an approximate solutions of this equation. The most famous iterative Lagrange method (Alexandrov, 2003), in which for a zero approximation there is chosen the function $E^{(0)}(t_*) = t_*$ that corresponds to an uniform motion on a circular orbit, and the term $e \cdot \sin E$ is considered as a perturbation. In this way, the solution is represented as an infinite series by powers of eccentricity

$$E(t_*) = t_* + \sum_{k=1}^{\infty} \frac{e^k}{k!} \cdot \frac{d^{k-1}}{dt_*^{k-1}} \left\{ \sin^k t_* \right\}. \tag{15}$$

As it was first shown by Laplace, this series coincides absolutely only in the region $0 < e \le \bar{e} = 0.66274\ldots$ (Alexandrov, 2003).

Usage of the Fourier series leads to the solution in the form (Alexandrov, 2003)

$$E(t_*) = t_* + \sum_{k=1}^{\infty} \frac{2}{k} I_k(ke)\sin(kt_*), \tag{16}$$

where

$$I_k(ke) = \frac{1}{\pi}\int\limits_0^{\pi} \cos\Big[k(z - e\cdot\sin z)\Big]dz \tag{17}$$

are Bessel functions of the first kind with an integer index value (Abramowitz & Stegun, 1972). For the eccentricity values that are close to unity and small values of t_*, series (16) has a weak convergence and requires taking into account dozens of terms, which makes this method cumbersome and irrational.

Usage of the theory of a complex variable allowed the second authors to obtain solutions of Kepler's equation in a finite analytical form (Siewert & Burniston, 1972; Philcox et al., 2021). However, the obtained solutions are too complicated for usage and require the additional calculations. Because of that, they can only be considered as proving the solution existence.

From equation (14) it follows that the function $E(t_*)$ is a periodic function with period 2π. Therefore, variables E and t_* change in the range $(0 \div 2\pi)$ and $E = 0$ at $t_* = 0$, $E = \pi$ at $t_* = \pi$ and $E = 2\pi$ at $t_* = 2\pi$. It also follows from equation (14) that in the interval $\pi \le t_* \le 2\pi$

$$E(t_*) = 2\pi - E(2\pi - t_*), \tag{18}$$

therefore, it is sufficient to find solutions in the region $0 \le t_* \le \pi$.

There can be distinguished two regions of variables for a sufficiently great value of eccentricity, in which the curve behavior $E(t_*)$ has a different character. Namely, in the plane (E, t_*)

$$0 \le E \le \pi, \;\; 0 \le t_* \le \pi$$

it can be selected the region of a rapid change of $E(t_*)$

$$\text{I.} \quad 0 \le E(t_*) \le \frac{\pi}{2}; \;\; 0 \le t_* \le \bar{t}_*(e) \tag{19}$$

and the region of a slow change of $E(t_*)$

$$\text{II.} \quad \frac{\pi}{2} \le E(t_*) \le \pi; \;\; \bar{t}_*(e) \le t_* \le \pi. \tag{20}$$

As it follows from equation (14), for $E = \pi/2$,

$$\bar{t}_*(e) = \frac{\pi}{2} - e. \tag{21}$$

A different behavior of the function $E(t)$ is caused by the different motion velocity of the material point along the orbit – rapid motion in the pericenter region and slow in the apocenter region.

To find analytical solutions of equation (14), we rewrite it in an equivalent form. Using the substitution

$$E = \frac{\pi}{2} - F_1, \tag{22}$$

we obtain an equation in the region I

$$F_1 + e \cdot \cos F_1 = \frac{\pi}{2} - t_*. \tag{23}$$

The substitution

$$E = \frac{\pi}{2} + F_2 \tag{24}$$

allows us to rewrite equation (14) in the region II in the form

$$F_2 - e \cdot \cos F_2 = t_* - \frac{\pi}{2}. \tag{25}$$

To obtain approximate analytical solutions of equations (23) and (25), we use the approximation $\cos F$ in the interval $(0 \div \pi/2)$ in the form of a polynomial of the fourth order

$$f(F) = 1 + a_2 F^2 + a_3 F^3 + a_4 F^4 \tag{26}$$

with coefficients

$$a_2 = -0.503491, \quad a_3 = 0.0111681, \quad a_4 = 0.0327516. \tag{27}$$

We will consider equation (25) in detail and represent it in the form

$$F_2(t_*) - ef(F_2(t_*)) = t_* - \frac{\pi}{2} + e\Big\{\cos F_2(t_*) - f(F_2(t_*))\Big\}. \tag{28}$$

Function (26) is sufficiently accurately approximated by $\cos F$ in the interval $(0 \div \pi/2)$, and deviation does not exceed $3 \cdot 10^{-4}$ even in the vicinity of $F_2 = \pi/2$. Because of that, for equation (28) is applied the method of successive approximations, and the zero approximation is determined by the solution of an algebraic equation

$$F_2^{(0)}(t_*) - ef(F_2^{(0)}(t_*)) = t_* - \frac{\pi}{2}. \tag{29}$$

The first iteration yields a specified solution

$$F_2^{(1)}(t_*) = F_2^{(0)}(t_*^{(1)}), \tag{30}$$

where

$$t_*^{(1)} \equiv t_* + e\bigl\{\cos F_2^{(0)}(t_*) - f(F_2^{(0)}(t_*))\bigr\} \tag{31}$$

and etc.

The solution of the equation of zero approximation is found by Cardano's formulae (Abramowitz & Stegun, 1972). The solution that corresponds to the condition $0 \le F_2^{(0)}(t_*) \le \pi/2$ is determined by the expression

$$F_2^{(0)}(t_*) = c - \left\{c^2 - \frac{1}{2}u + \left[\frac{u^2}{4} - \alpha_0\right]^{1/2}\right\}^{1/2}. \tag{32}$$

Here, we used the following notations

$$\begin{aligned}
&c = -\frac{1}{2}\left\{\frac{1}{2}\alpha_3 - \left(\frac{1}{4}\alpha_3^2 + u_2 - \alpha_2\right)^{1/2}\right\};\\
&u = \{r + (r^2 + q^3)^{1/2}\}^{1/3} + \{r - (r^2 + q^3)^{1/2}\}^{1/3} - \frac{1}{3}b_2;\\
&q = \frac{1}{3}b_1 - \frac{1}{9}b_2^2;\ \ r = \frac{1}{6}\{b_1 b_2 - 3b_0\} - \frac{1}{27}b_2^3;\\
&b_0 = 4\alpha_0\alpha_2 - \alpha_1^2 - \alpha_0\alpha_3^2;\ \ b_1 = \alpha_1\alpha_3 - 4\alpha_0;\ \ b_2 = -\alpha_2;\\
&\alpha_0 = \frac{1}{a_4}\left\{1 + \frac{1}{e}\ \left(t_* - \frac{\pi}{2}\right)\right\};\ \ \alpha_1 = -\frac{1}{ea_4};\\
&\alpha_2 = \frac{a_2}{a_4},\ \ \alpha_3 = \frac{a_3}{a_4}.
\end{aligned} \tag{33}$$

In the region I the solution of the equation of zero approximation is

$$F_1^{(0)}(t_*) = -c + \left\{c^2 - \frac{1}{2}u + \left[\frac{u^2}{4} - \alpha_0\right]^{1/2}\right\}^{1/2}, \tag{34}$$

if in notations (33) we replace $\alpha_1 \to -\alpha_1$ and $\alpha_3 \to -\alpha_3$. In the first iteration, we obtain the expression

$$F_1^{(1)}(t_*) = F_1^{(0)}(t_*^{(1)}), \tag{35}$$

where

$$t_*^{(1)} \equiv t_* + e\Bigl[\cos F_1^{(0)}(t_*) - f\Bigl(F_1^{(0)}(t_*)\Bigr)\Bigr]. \tag{36}$$

The time dependence of functions $F_1^{(0)}(t_*)$ and $F_2^{(0)}(t_*)$ is shown in Fig. 2 in the case $e = 0.6$ and eccentric anomaly in the same approximation $E^{(0)}(t_*)$. The maximal deviation $E^{(0)}(t_*)$ from the solution of equation (14) found by the numerical method does not exceed $5 \cdot 10^{-4}$, and iterative corrections $E^{(1)}(t_*) - E^{(0)}(t_*)$ are negligible.

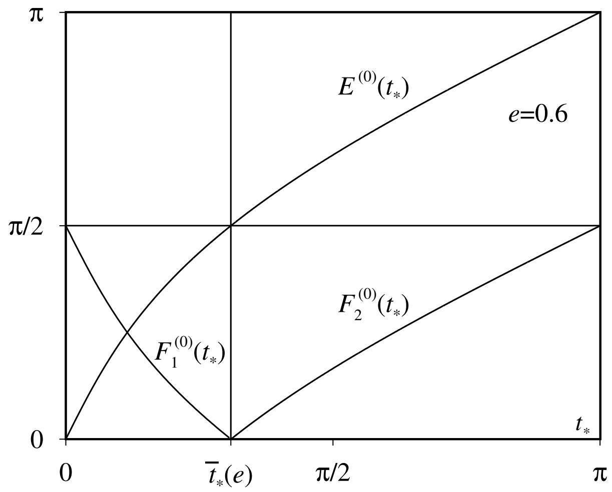

Figure 2. The time dependence of functions $F_1^{(0)}(t_*)$, $F_2^{(0)}(t_*)$ and the eccentric anomaly $E^{(0)}(t_*)$ for $e = 0.6$.

3. Equation of hyperbolic motion

Using the substitution

$$\tan\frac{v}{2} = \left(\frac{e+1}{e-1}\right)^{1/2} \tanh\frac{H}{2} \tag{37}$$

equation (8) is reduced to the form

$$e \cdot \sinh H - H = t_{\text{H}}, \tag{38}$$

which is an analogue of Kepler's equation. Herewith $-\infty < t_{\text{H}} < \infty$, and $-\infty < H < \infty$. Since the function $H(-t_{\text{H}}) = -H(t_{\text{H}})$, then it is sufficient to find the solution of equation (38) in the region $0 \leq t_{\text{H}} < \infty$.

Unfortunately, the function $\sinh H$ cannot be approximated by a polynomial of the fourth order in a sufficiently wide region of change H. Thereby, we will consider the calculation of the eccentric anomaly asymptotics near the pericenter, which is precisely of practical interest. We rewrite equation (38) in the form

$$ef_3(H) - H - t_{\text{H}} = e\{f_3(H) - \sinh H\}, \tag{39}$$

choosing the approximation function in the form

$$f_3(H) = H + aH^3 \tag{40}$$

for $a = 0.188479$. The solution of equation (39) is found by the iterations method, using in the role of zero approximation the root of the equation

$$ef_3(H^{(0)}) - H^{(0)} - t_{\text{H}} = 0, \tag{41}$$

namely

$$\begin{aligned} &H^{(0)}(t_{\text{H}}) = \{[r^2+q^3]^{1/2}+r\}^{1/3} - \{[r^2+q^3]^{1/2}-r\}^{1/3}, \\ &r = \frac{t_{\text{H}}}{2ea}, \ \ q = \frac{e-1}{3ea}. \end{aligned} \tag{42}$$

Specified of this solution we perform by the iterating of equation (39): in the

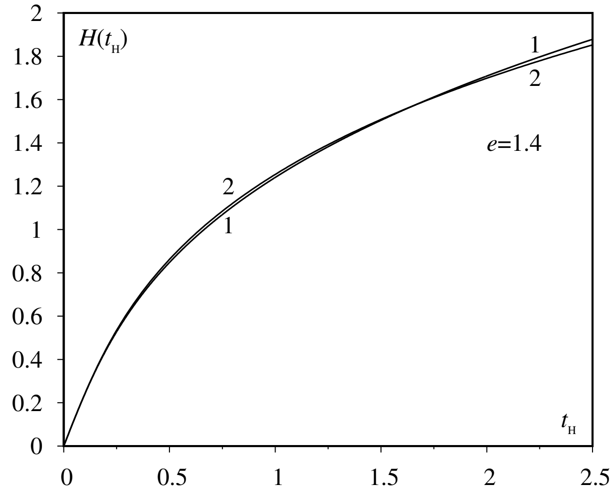

Figure 3. The solution of Kepler's equation in the case of hyperbolic motion for $e = 1.4$ in different approximations. Curve 1 corresponds to approximation (42), curve 2 – a numerical solution of equation (38).

first iteration

$$\begin{aligned} &H^{(1)}(t_{\text{H}}) = H^{(0)}(t_{\text{H}}^{(1)}), \\ &t_{\text{H}}^{(1)} = t_{\text{H}} + e\{f_3(H^{(0)}(t_{\text{H}})) - \sinh H^{(0)}(t_{\text{H}})\}; \end{aligned} \tag{43}$$

in the second iteration, we have

$$\begin{aligned} &H^{(2)}(t_{\text{H}}) = H^{(0)}(t_{\text{H}}^{(2)}), \\ &t_{\text{H}}^{(2)} = t_{\text{H}} + e\{f_3(H^{(0)}(t_{\text{H}}^{(1)})) - \sinh H^{(0)}(t_{\text{H}}^{(1)})\}, \ \text{ and etc.} \end{aligned} \tag{44}$$

Fig. 3 illustrates the function $H^{(0)}(t_{\mathrm{H}})$ (curve 1), as well as the numerically found solution of equation (38) (curve 2) for $e = 1.4$. In the vicinity of $t_{\mathrm{H}} = 2.0$ the relative deviation of these curves equals 0.59%. The deviation of the function $H^{(1)}(t_{\mathrm{H}})$ from the numerical solution does not exceed 0.06% for $t_{\mathrm{H}} = 2.0$, which indicates a rapid convergence of the iterative process.

4. Conclusions

Kepler's equation is one of the main relations of celestial mechanics, which determines the relevance of the problem solving. The classical Lagrange method is the conventional perturbation theory, where the zero approximation is the solution of a linear equation and corresponds to the motion on a circular orbit. The iterative algorithm proposed by us can be called as the renormalized perturbation theory, the zero approximation of which is the solution of an algebraic equation of the fourth order. Such an approximation differs from the solution found by a numerical method no more than $5 \cdot 10^{-4}$. This is already sufficient for practical using, and iterative corrections are negligible.

In the case of hyperbolic motion, the proposed technique is only applicable in the region of the pericenter orbit, because the function $\sinh H$ cannot be approximated by a polynomial of the fourth order in a wide region of change of H with the sufficient accuracy. Thereby, the role of iterations is increasing, but the iterative process is rapidly converging, therefore, it is enough one or two iterations, which have also an analytical representation according to formulae (43), (44).

References

Abramowitz, M. & Stegun, I. A. 1972, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*

Alexandrov, Y. V. 2003, *Celestial Mechanics*

Barger, V. & Olsson, M. 1995, *Classical Mechanics: A Modern Perspective. Second Edition*

Philcox, O. H. E., Goodman, J., & Slepian, Z., Kepler's Goat Herd: An exact solution to Kepler's equation for elliptical orbits. 2021, *MNRAS*, **506**, 6111, DOI: 10.1093/mnras/stab1296

Siewert, C. E. & Burniston, E. E., An Exact Analytical Solution of Kepler's Equation. 1972, *Celestial Mechanics*, **6**, 294, DOI: 10.1007/BF01231473

Vallado, D. A. & McClain, W. D. 2001, *Fundamentals of Astrodynamics and Applications*

PRÁCE ASTRONOMICKÉHO OBSERVATÓRIA
NA SKALNATOM PLESE
LIII, číslo 1

Zostavovateľ: RNDr. Richard Komžík, CSc.
Vedecký redaktor: RNDr. Augustín Skopal, DrSc.
Vydal: Astronomický ústav SAV, Tatranská Lomnica
IČO vydavateľa: 00 166 529
Periodicita: 3-krát ročne
ISSN (on-line verzia): 1336-0337
CODEN: CAOPF8
Rok vydania: 2023
Počet strán: 68

Contributions of the Astronomical Observatory Skalnaté Pleso are processed using
LaTeX 2_ε CAOSP DocumentClass file 3.09 ver. 2021.