The Astrophysical Journal, 605:L81–L84, 2004 April 10

LONG-TERM SOLAR VARIABILITY AND THE SOLAR CYCLE IN THE 21st CENTURY

Boncho P. Bonev,[1,2,3] Kaloyan M. Penev,[4,5] and Stefano Sello[6]
Received 2003 April 27; accepted 2004 February 26; published 2004 March 16

ABSTRACT

We have examined the long-term trends in the solar variability that can be deduced from some indirect data and from optical records. We analyzed the radiocarbon measurements for the last 4500 years, based on dendrochronology, the Schove series for the last 1700 years, based on auroral records, and the Hoyt-Schatten series of group sunspot numbers. Focusing on periodicities near one and two centuries, which most likely have a solar origin, we conclude that the present epoch is at the onset of an upcoming local minimum in the long-term solar variability. There are some clues that the next minimum will be less deep than the Maunder minimum, but ultimately the relative depth between these two minima will be indicative of the amplitude change of the quasi–two-century solar cycle.

Subject heading: Sun: activity

On-line material: color figures

1. INTRODUCTION

The best sunspot activity reconstruction based on optical records covers the period after 1610 and is the group sunspot numbers (GSNs) of Hoyt & Schatten (1998). A first glimpse at these (and previous) data shows significant changes in the amplitudes of the 11 yr solar cycles and even an apparent lack of sunspot activity during most of the 17th century period, known as the Maunder minimum (MM; Eddy 1976). Whether or not these changes are caused by longer solar cycles is a fundamental question concerning the variability of the Sun and has both astrophysical and geophysical significance. This problem is by no means new. Based on optical records, Gleissberg (1944) found a cycle that now bears his name. It is traditionally known as the 80–90 yr Gleissberg period, although increased available data have lead to estimates closer to 100 years (Rozelot 1994; Bonev 1997). The optical records, however, have a limited time coverage, so most of our present understanding of the long-term (LT) solar variability comes from studies of indirect activity proxies, such as measurements on radioisotopes (^{14}C, ^{10}Be), whose accumulation rate in terrestrial reservoirs is modulated by the solar activity, and indices based on auroral records. These data span a millennium timescale. Their applicability in solar studies along with major results have been discussed in the excellent reviews of Siscoe (1980) and Beer (2000).

Along with the aforementioned Gleissberg period, the different indirect data display a quasi–two-century (Suess) cycle. This cycle is examined in detail by Damon & Sonett (1991) and Dergachev (1994); they analyzed data from ^{14}C measurements. Both works emphasize the challenge of disentangling the solar component from a number of other factors (some of which are much stronger), causing variability in the radioisotope rates of accumulation. However, these authors argue for the solar origin of both the quasi-century and two-century cycles. The importance of considering the latter cycle in LT solar activity forecasts has been emphasized by Fyodorov, Klimenko, & Dovgalyuk (1996) and Komitov & Kaftan (2003).

Stuiver & Braziunas (1993) have given evidence that the amplitudes of both signals change with time, a conclusion based on the ^{14}C data. In addition, our preliminary results on auroral activity (Bonev 2000) indicate that the two-century cycle is much more pronounced in the second millennium A.D. than in the first millennium A.D.

This Letter attempts to motivate further in-depth studies of the solar cycles of centurial and bicenturial timescales. We have examined the behavior of these cycles in three independent data sets: radiocarbon, auroral, and optical data. Our main goal is a synthesis of the trends found in the three types of data into a single picture of the LT solar variability. We question whether or not the LT trend in the optical sunspot record is consistent with the peculiarities revealed in the much longer indirect data. In what follows, we describe the time series analyzed, our approach, and findings, and (last but not least) we discuss the implications for the average activity level in the next decades.

2. THE TIME SERIES ANALYZED

The variety of data analyzed with the same technique is a key aspect in this study. We have examined the last 4500 years from the ^{14}C database (Stuiver et al. 1998); the Schove series, based mainly on auroral records (Schove 1983); and the aforementioned GSNs (annual values).

The radiocarbon data (10 yr time step) were obtained by dendrochronology. Their general reliability was ultimately established by 1990 as a result of an extensive comparative study between about 50 laboratories worldwide (Dergachev 1994). The development of the radiocarbon calibration studies is described by Damon & Peristykh (2000).

The Schove series is a reconstruction of the 11 yr cycle amplitudes. We use its continuous part, which is after A.D. 296. Komitov & Bonev (2001) recently discussed the reliability of these data. We mention that Schove was one of the first to provide a clue to the quasi–two-century solar cycle, noticing

[1] Ritter Astrophysical Research Center, Department of Physics and Astronomy, University of Toledo, Toledo, OH 43606; bbonev@lepvax.gsfc.nasa.gov.
[2] Department of Physics and Astronomy, Bowling Green State University, 104 Overman Hall, Bowling Green, OH 43403.
[3] Current address: Laboratory for Extraterrestrial Physics, Code 690, NASA/Goddard Space Flight Center, Greenbelt, MD 20771.
[4] Division of Physics, Mathematics, and Astronomy, California Institute of Technology, 770 South Wilson Avenue, MS 103-33, Pasadena, CA 91125.
[5] Current address: Harvard University, Mail Stop 10, 60 Garden Street, Cambridge, MA 02138; kpenev@cfa.harvard.edu.
[6] Mathematical and Physical Models, Enel Research, Via Andrea Pisano 120, 56122 Pisa, Italy; stefano.sello@enel.it.

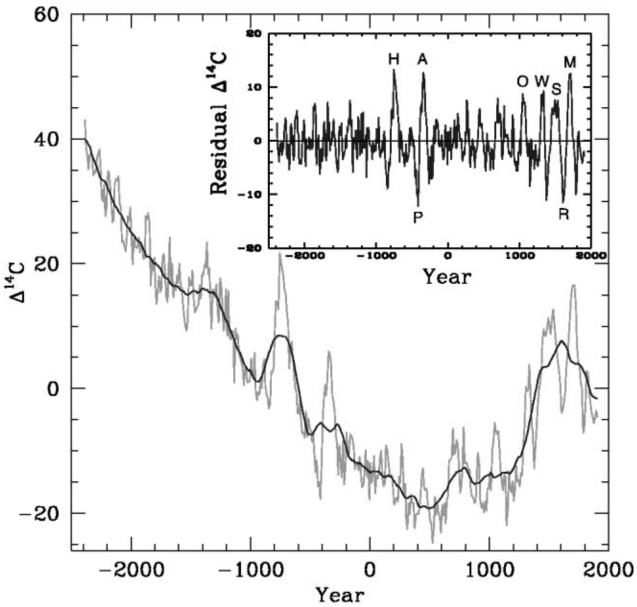

FIG. 1.—Radiocarbon data and zeroth-order Savitzky-Golay smoothing by 25 points; the residuals after the subtraction of the smoothed curve are shown in the inset. The letters indicate extreme phases of the LT solar variability: Maunder minimum (M), renaissance maximum (R), Spoerer minimum (S), Wolff minimum (W), Oort minimum (O), Aristotle minimum (A), Pericles maximum (P), and Homer minimum (H). In general, higher residuals indicate a lower solar activity level. The names of the Aristotle and Pericles minima have been suggested to us by B. P. Komitov (2003, private communication). [*See the electronic edition of the Journal for a color version of this figure.*]

that the auroral activity in the odd centuries has been lower than in the even centuries during the past millennium.

The GSN series is the most detailed reconstruction of the sunspot activity based on direct astronomical observations. The main difference between the traditionally used International Sunspot Number (ISN) and the new index is that the series of the latter display a lower overall activity before the 20th century but a much steeper rise in the LT trend between the MM and the record high 11 yr maximum No. 19 (1957).

3. APPROACH

Our goal is to investigate the stability of the cycles with duration near one and two centuries identified in the indirect data. This problem cannot be assessed by studying the periodograms of the whole data set. Such periodograms are equivalent to "weighted averages" of the present periodicities and consequently do not provide information about the persistence and amplitude changes of the latter. This problem requires "local" methods, developed to investigate nonlinear time series, showing certain varying multiscale quasi-periodic behavior. We have applied two such methods independently: (1) a fast-moving window periodogram algorithm (MWPA) and (2) a multiresolution wavelet analysis (WA).

The MWPA examines consecutively and independently *all* possible 80-point (i.e., 800 yr) subsets of the indirect data described. Periodogram analysis was performed on the data samples consisting of points 1–80, 2–81, etc. The choice of eight-century–long intervals is not unique. However, (1) it is long enough to allow a good detection of periods near two centuries with the cycle-search method we used, and (2) it is much shorter than the whole sample, allowing us to investigate the changes in the variability behavior related to the cycles of interest. The cycle-search method used is called the *T-R periodogram analysis* (TRPA; Komitov 1986; Benson et al. 2003). The essence of this algorithm is equivalent to the *Lomb-Scargle periodogram*, as shown by Scargle (1982, Appendix C). Compared with other methods applied to (evenly spaced) solar data, the TRPA reproduces very closely the results of Rozelot (1994; ISN, *Fourier method*) and Fyodorov et al. (1996; Schove data, *maximum entropy*). Note that in these examples, the length of the investigated series was much larger than the longest periods of interest.

The WA (Torrence & Compo 1998) is a powerful method for investigating unsteady nonlinear time series. It has been successfully applied in analysis of solar and geophysical data (e.g., Ogurtsov et al. 2002; Sello 2003a). This approach is very appropriate to search for multiscale nonlinear variability behavior. Its advantage is that there is not any assumption about the width of the moving window, as the eight century long intervals in the MWPA (even though it is a reasonable choice). WA was used extensively on solar data by Ogurtsov et al. (2002). Our analysis focuses on a shorter time interval, but it is based on a more refined wavelet application that uses a better scale-temporal and amplitude resolution and therefore provides more robust results.

4. ANALYSIS AND DISCUSSION

4.1. *Radiocarbon Residual Data*

Our first task was to disentangle the major trends in the ^{14}C data to a residual level where the solar signal we aim to investigate is dominant. The main source of variability has been of geomagnetic origin, and, in addition, periodicities of millennium and several century-long timescales have been suggested (Damon & Sonett 1991). The length of our data set is too short for reliable detections of such signals, and hence this is out of the scope of the present work. We have approximated the largest timescale variability by applying a low-order Savitzky-Golay smoothing. In this approach, the approximation of the main trend is entirely based on the data, and no assumption about its (complicated) functional form is necessary. The order of the smoothing can be zero (moving average) or one without affecting the results.

The raw ^{14}C data, the smoothed curve, and the residuals after its subtraction are shown in Figure 1. These residuals do not contain periodicities that are comparable to the length of the whole data set. Instead, they are dominated by a cycle near two centuries that likely has a solar origin.

4.2. *Two Regimes of the Long-Term Solar Variability*

Both the WA and the MWPA reveal strong amplitude modulations in the behavior of the quasi–two-century cycle in the ^{14}C series as well as instabilities in the signals of the century-long timescale. Figure 2 shows the corresponding wavelet local power spectrum and three-dimensional periodogram (MWPA) in terms of contour maps. The average power spectrum is also presented for comparison.

We conclude that for the last 4500 years, there have been two distinct regimes of the LT solar variability as inferred from ^{14}C residual data. The first regime (hereafter regime 1) is characterized by the strongest appearances of the quasi–two-century cycle, which is most prominent in the middle of the first millennium B.C. and the second millennium A.D. During these epochs, the variations of century timescale are strongly suppressed. The only other significant periodicity is in the range of 130–150 yr. Ogurtsov et al.

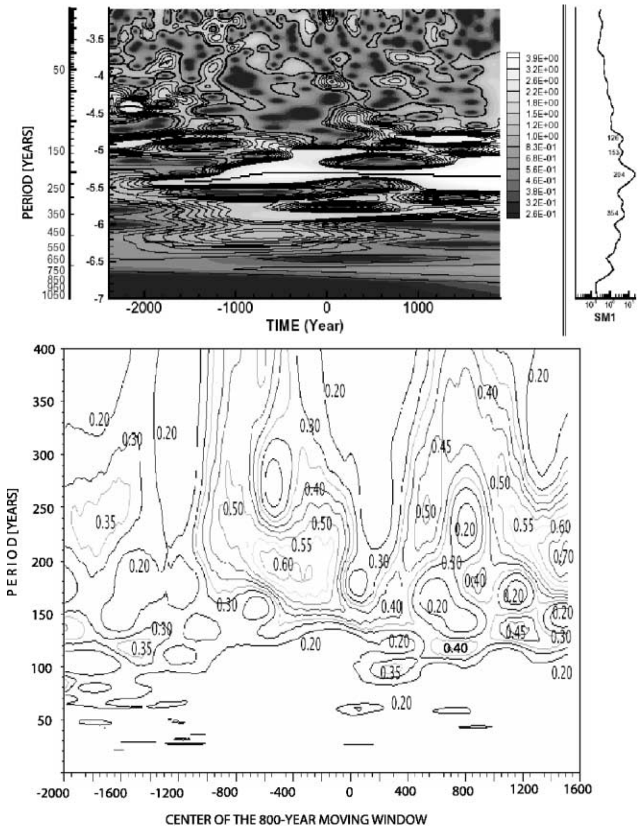

FIG. 2.—Analysis of the ^{14}C data: average Fourier spectrum (*upper right*), wavelet local power spectrum (WLPS; *upper left*), and a three-dimensional periodogram (*lower panel*) from the moving window analysis (MWA) in terms of the contour maps. In the WLPS map, the black contour lines locate the significant power regions against an assumed red-noise background spectrum. This was generated by a suitably tuned univariate autoregressive lag-1 Markov process. The enclosed black tracks indicate the ridge wavelet level corresponding to a 90% significance level (before 1000 B.C. and from A.D. 0 to 1000) and a 99% significance level (from 1000 to 0 B.C., and after A.D. 1000) (see Torrence & Compo 1998). A Tukey-Hamming window was used in the MWA for every 80-point data subset examined. A vertical line on the MWA plot would represent a traditional two-dimensional T-R periodogram (T = period, R = correlation coefficient) for the corresponding moving window. The quantity presented on the MWA plot is the correlation coefficient between the data and a sine wave fitted to them (see details in Benson et al. 2003). This quantity is directly proportional to the cycleâs amplitude. Note that for the period after A.D. 300, a similar variability behavior is seen in the Schove series. [*See the electronic edition of the Journal for a color version of this figure.*]

(2002) proposed that this peak is part of a complex frequency band (50–140 yr) attributed to the Gleissberg cycle. This idea is hard to prove on a purely statistical basis. Moreover, our higher resolution results suggest that the peak around 130 yr in the second millennium A.D. is actually most pronounced near the time when the quasi–two-century cycle has its strongest appearance.

The other distinct regime of the LT solar variability (hereafter regime 2) is characteristically found prior to the first millennium B.C. and during the first millennium A.D. Within these time frames, the quasi–two-century cycle has a significantly lower amplitude than is found in regime 1, while signals on century and subcentury timescales also appear. These signals, however, are more unstable and weaker for most of the examined time frame (as well evidenced by the broken "ridge" wavelet black lines in the upper panel of Fig. 2 prior to the first millennium B.C.). Their generally lower amplitude is expected because the longer periodicities are in principle better pronounced in the ^{14}C data.

We emphasize that the variability in the Schove series is *very similar* to the variability in the radiocarbon data shown in Figure 2 (after A.D. 300). A very strong cycle of ~208 yr reaches peak amplitudes near the time frame as the quasi–two-century period in the ^{14}C series. On the other hand, in the first millennium A.D., this peak is suppressed while multiple but not highly reliable signals of shorter timescale appear. Therefore, although the Schove series might be of more limited reliability (Komitov & Bonev 2001), the main signatures of the nonlinear behavior in the radiocarbon and Schove data are quite similar.

Overall, these results are consistent with those of independent parallel studies on related topics. Very recently, Peristykh & Damon (2003) presented an in-depth study focusing on the narrow periodicity interval around ~90 yr (the "classical" value for the Gleissberg signal). These authors show that this signal has a distinct minimum in the first millennium B.C., a period when, according to this work, century timescale signals are suppressed. B. P. Komitov (2004, in preparation) examined the time variation of the periodogram curves, integrated between 60–130 yr and 170–230 yr. His first integral parameter peaks in the epochs of regime 2, while his second one peaks during regime 1. Finally, our WA results on the quasi–two-century cycle not only confirm but also extend the results of Ogurtsov et al. (2002) for the period after 2500 B.C., because, as mentioned above, we focus on a shorter time interval but use a refined wavelet application with a better scale-temporal and amplitude resolution.

4.3. *Synthesis*

Investigating the extrema in the LT solar variability for the last 10,000 years, Dergachev (1994) emphasized that the time epochs around 2700 and 450 years ago were near grand activity minima, while the periods near 3800 and 1600 years ago were characterized by an overall active Sun. These time intervals of low and high mean solar activity overlap, respectively, with regime 1 and regime 2, as defined in this study. Therefore, for the last 4500 years, there has been a distinct correlation between the overall activity level and the strength of the dominant periodicity. Apparently, the cycle near 210 yr is most pronounced during deep activity "depressions" of Maunder type. Between such epochs, quasi-periodicities of century timescales may develop, but they are generally more unstable.

Does the optical sunspot record support this picture? The data after the MM show both a distinct rise in the mean activity level and a strong quasi-century cycle. We found that the former cycle in the GSN data after 1700 is well pronounced and, together with the 11 yr cycle, is the only prominent periodicity. Of course, the limited length of the optical records limits secure detections of longer signals. With this in mind, we note a near 193 yr signal in the complete GSN data (see also Komitov & Kaftan 2003) and a 150–165 yr one for the data after the MM.

If the hypothesis about the correlation between the quasi–two-century cycle appearance and the overall activity level is valid, the period after the MM has been the beginning of a transition from regime 1 to regime 2. However, at present, the cycle near 210 yr is still important. The influence of the quasi-century cycle on the LT solar variability should also be considered. The middle of the 20th century is significant because during that epoch, the solar activity reached the highest levels ever observed (and possibly the highest in the last ~2000 years, as inferred from the Schove data, smoothed by a five-point running mean [Komitov & Bonev 2001], and moreover by the recent physical reconstruction of the sunspot activity, based on ^{10}Be records, by Usoskin et al. 2003). Dergachev (1994) argued that this epoch has witnessed the end of the rise since the MM and that the solar activity will stay on a high level overall, when *averaged over the next several centuries*. However, *on a shorter timescale*, we expect an LT minimum in the next several decades. We are already in the minimum phase of the quasi-century cycle, whose last primary peak was in the mid-20th century, followed by a secondary one during 11 yr cycle 22. Rozelot (1994), Bonev (1997), and our work conclude that in the optical records, the Gleissberg cycle is near 100 yr, rather than the classical values of 80–90 yr. In the context of a quasi periodicity, this is the most probable, rather than exact, value. Bonev's study predicted the next cycle's minimum in the beginning of this century. The quantitative approach of this paper has shortcomings (not affecting the period value), but its conclusion about the downward trend is valid. This trend is also inferred from analysis of the solar equatorial rotation rate (Javaraiah 2003) and by the forecasts for 11 yr cycle 24, based on the nonlinear dynamics approach (Sello 2001, 2003b) and on the "solar dynamo amplitude" method (Schatten 2003). In addition, the bicenturial cycle is also in its declining phase after a maximum in the middle of the last century (Komitov & Bonev 2001; Komitov & Kaftan 2003). Ultimately, the depth of the next minimum, compared with the MM, will be indicative of the amplitude change in the quasi–two-century periodicity, because this cycle is most pronounced near grand activity minima. However, the general activity rise, well seen in the GSN data, suggests that we might enter a variability regime in which the local long-term minima are less deep than the MM.

Komitov & Bonev (2001) point out some indirect evidence of an LT decline. While the solar precursor method successfully predicted that the 11 yr cycle 23 would be magnetically weaker than cycle 22 (Schatten, Myers, & Sofia 1996), the methods based on the Gnevyshev-Ohl rule in its different versions (Gnevyshev & Ohl 1948; Vitinskii 1997) failed completely, although this rule had been valid for at least 150 years. Some of these forecasts, reviewed by Obridko (1995), implied that cycle 23 would be the highest solar cycle observed (*maximum maximorum*). In fact, there is enough evidence to support the notion that the present cycle is magnetically weaker than cycle 22 (de Toma et al. 2001; Chapman et al. 2001; Livingston 2002; Javaraiah 2003). We consider this "untypical" behavior for an even-odd 11 yr cycle pair related to the LT decline.

To summarize, we expect a long-term solar minimum in this century caused by the declining phases of both the Gleissberg and the near 210 yr solar cycles. The depth of this upcoming minimum with respect to the MM will be indicative of the amplitude change (or lack of) in the quasi–two-century solar cycle. The next decades will provide a unique opportunity to study the Sun and the solar-terrestrial relationships under conditions significantly different from those during the past several 11 yr cycles. It will be interesting to see if the tendency toward no secular increase in the solar magnetic flux over the last two cycles (Arge et al. 2002) will continue in the 21st century.

The authors are grateful to the Quaternary Isotope Laboratory (University of Washington) for providing the scientific community with access to the radiocarbon database. We are also very grateful to Professor John B. Laird, whose mentorship of B. P. B. in 1998 has had a direct impact on this research; to Professor Boris P. Komitov, Professor Dale Smith, and several faculty members at Bowling Green State University and at Ritter Observatory (University of Toledo) for many useful discussions; to Dr. Erika Gibb and to the ApJ production office for helping to improve the clarity of the presentation; and to Ms. Diana Rangelova for technical assistance with Figure 2. B. P. B. gratefully acknowledges partial financial support from the Department of Physics and Astronomy, University of Toledo, and from NASA's GSFC under RTOP 693-344-32-30-07.

REFERENCES

Arge, C. N., Hildner, E., Pizzo, V. J., & Harvey, J. W. 2002, J. Geophys. Res., 107, 1319
Beer, J. 2000, Space Sci. Rev., 94, 53
Benson, J. L., Bonev, B. P., James, P. B., Shan, K. J., Cantor, B. A., & Caplinger, M. A. 2003, Icarus, 165, 34
Bonev, B. P. 1997, Bulgarian Geophys. J., 13, 43
———. 2000, BAAS, 32, 829
Chapman, G. A., Cookson, A. M., Dobias, J. J., & Walton, S. R. 2001, ApJ, 555, 462
Damon, P. E., & Peristykh, A. N. 2000, Radiocarbon, 42, 137
Damon, P. E., & Sonett, C. P. 1991, in The Sun in Time, ed. C. P. Sonett., M. S. Giampapa, & M. S. Matthews (Tucson: Univ. Arizona Press), 360
Dergachev, V. A. 1994, Priroda, 1, 3
de Toma, G., White, O. R., Chapman, G. A., Walton, S. R., Preminger, D. G., Cookson, A. M., & Harvey, K. L. 2001, ApJ, 549, L131
Eddy, J. A. 1976, Science, 192, 1189
Fyodorov, M. V., Klimenko, V. V., & Dovgalyuk, V. V. 1996, Sol. Phys., 165, 193
Gleissberg, W. 1944, Terr. Magn. Atmos. Electron, 49, 243
Gnevyshev, M. N., & Ohl, A. I. 1948, AZh, 25, 18
Hoyt, D. V., & Schatten, K. H. 1998, Sol. Phys., 181, 491
Javaraiah, J. 2003, A&A, 401, L9
Komitov, B. P. 1986, Soln. Dannye, 5, 73
Komitov, B. P., & Bonev, B. P. 2001, ApJ, 554, L119
Komitov B. P., & Kaftan, V. 2003, Geomagn. Aeron., 43, 553
Livingston, W. 2002, Sol. Phys., 207, 41
Obridko, V. N. 1995, Sol. Phys., 156, 179
Ogurtsov, M. G., Nagovitsyn, Yu. A., Kocharov, G. E., & Jungner, H. 2002, Sol. Phys., 211, 371
Peristykh, A. N., & Damon, P. E. 2003, J. Geophys. Res., 108, 1003
Rozelot, J. P. 1994, Sol. Phys., 149, 149
Scargle, J. D. 1982, ApJ, 263, 835
Schatten, K. H. 2003, Adv. Space Res., 32, 451
Schatten, K., Myers, D. J., & Sofia, S. 1996, Geophys. Res. Lett., 23, 605
Schove, D. J. 1983, Sunspot Cycles (Stroudsburg: Hutchinson Ross)
Sello, S. 2001, A&A, 377, 312
———. 2003a, NewA, 8, 105
———. 2003b, A&A, 410, 691
Siscoe, G. L. 1980, Rev. Geophys. Space Phys., 18, 647
Stuiver, M., & Braziunas, T. F. 1993, The Holocene, 3, 289
Stuiver, M., et al. 1998, Radiocarbon, 40, 1041
Torrence, C., & Compo, G. P. 1998, Bull. Am. Meteorolog. Soc., 79, 61
Usoskin, I. G., Solanki, S. K., Schüssler, M., Mursula, K., & Alanko, K. 2003, Phys. Rev. Lett., 91, 211101
Vitinskii, Yu. I. 1997, Astron. Lett., 23, 562