Lags, hysteresis, and double peaks between cosmic rays and solar activity

R. P. Kane
Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo, Brazil

Received 17 April 2003; revised 18 June 2003; accepted 22 July 2003; published 29 October 2003.

The 12-month moving averages of solar indices and CR (cosmic ray intensity) were compared. The long-term variation had a negative correlation (already known). From the hysteresis plots for five cycles (odd cycles 19, 21, 23; even cycles 20, 22), it was seen that whereas odd and even cycles showed different patterns, the detailed evolution in similar cycles (even or odd) was not alike. The reason for complication is as follows. On medium-term time scales, in some sunspot cycle maxima (notably cycles 22 and 23), there were two peaks in solar indices separated by about 2 years. These were reflected in CR as two distinct minima, some with a lag of a few months but others almost coincident with maxima of solar indices. The implication would be that CR modulation as observed on Earth contains two parts: (1) long-term variation due to magnetic field structure in total heliosphere, determined by convection-diffusion and drift mechanisms (this part has time lags of several months and is characterized by hysteresis phenomenon) and (2) short-term and medium-term variations caused mainly by accumulated shock waves (Forbush effects and associated phenomena) with time lags of a few days, which in the scale of months or a year or two, would show a time lag zero. The second component would be especially important in maximum of solar activity.

INDEX TERMS: 2104 Interplanetary Physics: Cosmic rays; 2124 Interplanetary Physics: Heliopause and solar wind termination; 2162 Interplanetary Physics: Solar cycle variations (7536); 7529 Solar Physics, Astrophysics, and Astronomy: Photosphere; 7534 Solar Physics, Astrophysics, and Astronomy: Radio emissions; KEYWORDS: solar indices, interplanetary parameters, cosmic rays


1. Introduction

Cosmic rays (CR) are energetic particles that are found in space and filter through our atmosphere. The portion of the CR spectrum that reaches the Earth’s atmosphere is controlled by the geomagnetic cutoff (minimum, almost zero, at the magnetic poles to a vertical cutoff of about 15 GV in the equatorial regions). CR are being regularly monitored by ground-based neutron monitors at several locations on the Earth for the last several decades. Observations so far indicate a clear solar cycle effect, with largest reductions in CR neutron monitor intensity during sunspot maximum years, a very good anticorrelation, (long-term variation) [Forbush, 1954; Ahtuwalla and Wilson, 1996, and references therein]. Burlaga et al. [1985] proposed that fast coronal mass ejections (CMEs) contribute to form a propagating diffusion region (heliocentric barrier) further out in the heliosphere, and CR intensity never quite recovers at the Earth’s orbit [Burlaga et al., 1993]. Kota and Jokipii [1983] envisaged the dependence of the CR modulation on the orientation of the solar magnetic dipole moment also so that a complete modulation would involve 2 solar cycles (Hale magnetic cycle of ~22 years) [Hale and Nicholson, 1925].

The structures of the recovery in the 11-year cycle of CR in relation to the state of interplanetary magnetic field have been studied in detail by Jokipii and Thomas [1981] and many others [e.g., Ahluwalia, 2000]. For short-term effects the relationship between solar variations with interplanetary plasma parameters and further with CR Forbush decreases and geomagnetic storms is also discussed in detail in various publications [e.g., Gonzalez et al., 1994, and references therein]. An interesting aspect is the “hysteresis” between solar activity and CR, where CR maxima (minima) do not seem to coincide exactly with solar activity minima (maxima). There is often a lag of a few months, detected more than 40 years ago (e.g., Dorman [1957], Neher and Anderson [1962], and later, many others). The lag has been used to estimate the radius of the CR modulation region. Earlier estimates, based on comparison with coronal green line or by examining CR modulation caused by sudden jumps in solar activity [Dorman et al., 2001a, and references therein] indicated that the radius was very small, about 5 AU, not more than 10–15 AU. Dorman [1975, and references therein] initiated the use of a convection-diffusion model, taking into account the time lag of processes in interplanetary space relative to the processes on the Sun and...
concluded that the modulation region should be much larger, between 50 and 150 AU. Further, Dorman [2001] took into account the drift effects (as depending upon the sign of solar polar magnetic field) according to Burger and Potgieter [1999] and showed different effects for even and odd solar cycles [see also Dorman et al., 2001a, 2001b].

In most of these analyses, correlations are used between CR and sunspot number. There is nothing special about sunspot numbers as such and any other solar index could be used, e.g., coronal green line index used by Pathak and Sarabhai [1970]. In a recent communication [Kane, 2002] it was shown that the evolutions of different solar indices were not always similar. In particular, at solar maximum, often dual peaks were seen and their relative heights were dissimilar for different indices. In this communication, data of some solar indices and cosmic ray neutron monitor intensities at a few selected locations are reexamined and compared, since 1953 up to recent solar maximum interval 1999–2002.

2. Data

The data used are monthly values for (1) several solar indices, (2) interplanetary plasma parameters N (number density), V (wind speed), B (total magnetic field), and (3) average of neutron monitor intensities at Calgary (51°N, 114°W, cut-off rigidity 1.09 GV), Deep River (46°N, 78°W, cut-off rigidity 1.02 GV), Kiel (54°N, 10°E, cut-off rigidity 2.32 GV), Moscow (55°N, 37°E, cut-off rigidity 2.42 GV), Climax (39°N, 106°W, cut-off rigidity 2.99 GV), and neutron monitor intensities at Huancayo (12°S, 75°W, cut-off rigidity 12.92 GV) up to 1992, followed by Haleakala (20°N, 56°W, cut-off rigidity 12.91 GV) for 1993–2002. Most of the data were obtained from the NOAA website ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA. Data for open solar magnetic flux were kindly supplied privately by Dr. Y.-M. Wang.

3. Hysteresis Plots

Figure 1 shows plots of the 12-month moving averages of CR intensity versus 12-month moving averages of sunspot number for (a) solar cycle 19, April 1954–September 1964, only for Climax neutron monitor, (b) cycle 20, October 1964–February 1976, (c) cycle 21, March 1976–February 1986, (d) cycle 22, March 1986–April 1996, and (e) cycle 23 partial, May 1996–June 2002, all for the average of neutron monitor percentage intensities at Calgary, Kiel, Moscow, and Climax (effective rigidity 10–15 GV). Ascending and descending parts of each solar cycle are indicated by full lines and crosses, respectively. Dots indicate annual values for years marked. (f) Cosmic rays versus 2800 MHz solar radio flux F10, for cycle 23.
the sunspots during 1957–1958. CR lingered during 1959 and then recouped rapidly, but the lingering caused a broad hysteresis loop. In the next odd cycle 21, CR continued to fall from 1976 right up to 1982, but sunspots reached a maximum in 1980 and then started falling. During 1980–1982, CR lagged behind sunspots considerably. Thus whereas both cycles 19 and 21 had broad hysteresis loops, the evolution was qualitatively different in finer details. In the correlation analysis of Dorman [2001] and Dorman et al. [2001a, 2001b], these finer detailed differences get lost.

3. In the even cycles 20 and 22, the hysteresis loops are narrow, but in cycle 22, during 1989–1991 when sunspot number was falling, CR lingered, thus causing a broader loop as compared with cycle 20. The lingering near sunspot maximum is not similar for the 4 solar cycles, not even for odd or even cycles. The differences in even, odd cycles are attributed to differences in the drift effects related to the reversal of solar magnetic field in alternate cycles [e.g., Burger and Potgieter, 1999], but it seems that the propagating diffusion region (heliocentric barrier) further out in the heliosphere contributed by fast CMEs (Coronal Mass Ejections) as envisaged by Burlaga et al. [1985, 1993] does not have variations directly comparable to those of sunspots.

4. Figure 1e shows the plots for the odd cycle 23. The fall of CR was slow initially and later picked up, almost like the earlier odd cycles 19 and 21, but in 2000–2001, there is a complicated looped structure. Since the 2800 MHz flux F10 may differ from sunspots, CR was plotted versus F10 as shown in Figure 1f. The looped structure is clearly seen and is unlike anything seen in the previous cycles.

5. The recouping phase of CR has just started. It will probably follow the same pattern as of earlier odd cycles 19 and 21, but it remains to be seen whether there will be a lingering in 2003–2004 resulting into a broadening of the hysteresis loop.

Figure 2. Same as Figure 1, for cosmic rays at Huancayo-Haleakala (effective rigidity 30–40 GV).

4. Plots for Intensities Near Solar Maxima

1. The magnitudes of the maxima of solar indices vary considerably from cycle to cycle. Some sunspot maxima are sharp and single but some have a broad, double-humped structure, long enough to be seen even in 12-month moving averages. (In case of two maxima, the larger ones are indicated by full squares and the smaller ones by dots). Such structures have been reported earlier [e.g., Gnevyshev, 1967, 1977; Feminella and Storini, 1997, and references therein]. The relative strengths of the two humps are different for different solar indices (see details in the work of Kane [2002]).

2. In cycle 19 (top panel), the solar indices (12-month moving averages) have just one maximum in February 1958. The neutron monitor patterns (full line, Climax only;
crosses, Huancayo) are very similar to each other and opposite to those of the solar indices (solar maxima associated with cosmic ray minima, well-known inverse relationship). The magnitudes for Huancayo are about a third of those of Climax (again well-known). However, there is a qualitative difference. Whereas cosmic rays show a major minimum in or slightly before February 1958 (almost coinciding with solar indices), there is a second, minor minimum in cosmic rays in the middle of 1959, i.e., about 15 months later. This raises serious doubts about the concept of delays (lags) between solar indices and cosmic ray intensity. For cycle 19, solar indices do not have a

Figure 3. Plots of 12-month moving averages of solar indices F10 (2800 MHz 10.7 cm solar radio emission) and Rz (sunspot number), and cosmic ray intensities at Climax etc. (full line) and Huancayo (crosses), for intervals of maxima of sunspot cycle 19 (1956–1960, top panel), cycle 20 (1967–1971, middle panel) and cycle 21 (1979–1983, bottom panel, also shows interplanetary parameters N, V, B). Most of the plots show two solar maxima associated with two cosmic ray minima, the larger one indicated by a full square and the smaller one by a dot.
double-peak structure, but cosmic rays have, and there is no lag between the solar maximum and the first cosmic ray minimum.

[17] 3. In cycle 20 (middle panel), solar indices have a very broad maximum, the level remaining almost constant (plateau) for almost three years (1967–1969), falling rapidly after the middle of 1970. The cosmic rays (average of Climax, Moscow, Kiel, Deep River, Calgary, full line; Huancayo, crosses) have a clear minimum in February–March 1969 and a clear recovery thereafter, though the solar...
indices continued to be high till about June 1970 (more than 15 months later). Thus, there is no lag involved between solar indices and cosmic rays (rather, there is a lead).

[18] 4. In cycle 21 (bottom panel), solar indices had a maximum near January 1980, followed by a very slow decrease for the next 24 months, and a faster decrease thereafter. Cosmic rays (average of Climax, Moscow, Kiel, Deep River, Calgary, full line; Huancayo, crosses) had two minima, a first smaller one in April–May 1981 and a second major one near November 1982, about 18 months later. If solar indices are considered as having a maximum in January 1980, the lags would be 17 months for the first (minor) cosmic ray minimum and ~30 months for the second (major) cosmic ray minimum, a fairly confusing situation for calculating the dimensions of the heliosphere based on lags. The bottom part shows variations of the interplanetary parameters N, V, B. All these seem to increase monotonically up to the middle of 1982 and attain maxima there, almost coinciding with the cosmic ray minima.

[19] In cycles 22 and 23, solar indices definitely had double peaks. Figure 4 shows the plots of 12-monthly averages for cycle 22 (1988–1992), with several radio emissions (including F10) in the upper part, some chromo-

Figure 5. Same as Figure 4 but for interval of maximum of cycle 23 (1998–2002).
spheric lines and photospheric parameters (including sunspots) in the middle part, and cosmic rays (average of Climax, Moscow, Kiel, Deep River, Calgary, full line; Huancayo, crosses) and interplanetary N, V, B in the lower part. The following may be noted.

20. All parameters show two maxima, one during June–October (mostly June–July) 1989 and another in the middle of 1991 (two years later). In most of the cases, the first maximum is larger (shown by full square) and the second maximum is smaller (shown by a dot), but in some cases, both are almost equal.

21. In cosmic rays, too, there are two minima. The first minimum was larger and occurred in February 1990, lagging behind solar activity by about 8 months. The second minimum was in August 1991, was smaller than the first maximum for Huancayo, and coincided with the second maximum of solar indices. Thus the first maximum of solar indices and first minimum of cosmic rays involved lags, but the second solar maximum and second cosmic ray minimum did not have a lag. If the second minimum of cosmic rays is compared with the first maximum of solar indices, the lag would be enormous (24 months).

22. The interplanetary parameters plotted at the bottom also show two maxima almost coinciding with the cosmic ray minima, indicating a possible relationship.

23. Figure 5 shows a plot of 12-month moving averages for cycle 23 (1998–2002). Here, the solar indices (upper part of the figure) have a first maximum in a wide range (March–October 2000). A major cosmic ray (lower part of the figure) minimum occurred in May–July 2000, almost coinciding with the solar maximum or perhaps with an uncertain lag of 2–3 months. The solar indices had a very strong and consistent second maximum in February 2002 (about 2 years after the first maximum). In recent months, cosmic rays seem to be undergoing a second minimum. If it is assumed to be in September 2002 (may not be complete yet), it would imply a lag of 7 months with respect to the second maximum of solar indices. The interplanetary parameters N, V, B (bottom part of the figure) show maxima near the first and second maxima of the solar indices. (However, the plots for 2001–2002 are not fully reliable as V, B had missing data for some months in the end of 2001, and values shown by dashes were manipulated).

24. For short-term time scales (hours to days), CR are known to respond to interplanetary magnetic blobs almost simultaneously. However, from the plots of Figure 3, Figure 4, and Figure 5, it seems that even on a medium time scale (1–3 years), some CR variations show almost simultaneity with variations of interplanetary parameters and/or solar indices, while some other CR variations show a lag of few months.

5. Sun’s Open Magnetic Flux

25. Cosmic rays are affected by IMF (interplanetary magnetic field), which has its origin in open magnetic regions (coronal holes) of the Sun. The coronal source surface lies at a heliocentric distance of ~2.5 solar radii and there, the solar magnetic field is approximately radial. Magnetic flux emerges through the photosphere in active regions. The field line loops rise up through the corona as the opposite polarity footprints separate in the differential photospheric rotation and the flux emerges through the coronal source surface. The total open magnetic flux entering the heliosphere by threading this surface is called the coronal source flux. It can be inferred from current-free extrapolations of the observed photospheric field. Wang et al. [2000] derived the long-term variation of the open flux for 1971–1998. Dr. Wang kindly supplied us the data up to 2002. Figure 6 (top plot) shows the 12-month moving averages of the open flux. The other plots are for sunspot number Rz, interplanetary N, V, B, and cosmic ray average (Climax, etc.). The following may be noted.

26. The Sun’s open flux has a long-term variation, but it is not parallel to the sunspot cycle. (The thick line shows a 24-month moving average). During 1969–1980, the flux was almost constant, reached a maximum in 1984 (4 years later than the sunspot maximum of 1980), and another maximum in 1992 (2 years later than the sunspot maximum of 1990). In cycle 23 when sunspot maximum occurred in 2000–2001, the open flux has just come out through a minimum in 2000. Thus no fixed relationship is seen between this open flux and sunspot numbers. Wang et al. [2000] have noticed this and explain that near sunspot minimum, the open flux originates mainly from the large coronal holes, whereas at sunspot maximum, it is rooted in small, lower-latitude holes characterized by very high field

Table 1. Intercorrelations Between the 12-Month Moving Averages of the Various Parameters During 1974–2002

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Open Flux</th>
<th>N</th>
<th>V</th>
<th>B</th>
<th>CR Ave.</th>
<th>Sunspot</th>
<th>aa Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Flux</td>
<td>1.00</td>
<td>0.46</td>
<td>0.45</td>
<td>0.51</td>
<td>0.04</td>
<td>0.11</td>
<td>0.65</td>
</tr>
<tr>
<td>N</td>
<td>0.06</td>
<td>1.00</td>
<td>0.17</td>
<td>0.14</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>V</td>
<td>0.45</td>
<td>0.17</td>
<td>1.00</td>
<td>0.04</td>
<td>0.18</td>
<td>0.24</td>
<td>0.76</td>
</tr>
<tr>
<td>B</td>
<td>0.51</td>
<td>0.04</td>
<td>0.14</td>
<td>1.00</td>
<td>0.04</td>
<td>0.59</td>
<td>0.33</td>
</tr>
<tr>
<td>CR Ave.</td>
<td>-0.41</td>
<td>0.18</td>
<td>-0.04</td>
<td>-0.85</td>
<td>1.00</td>
<td>-0.05</td>
<td>1.00</td>
</tr>
<tr>
<td>Sunspot</td>
<td>0.11</td>
<td>-0.33</td>
<td>-0.24</td>
<td>0.76</td>
<td>-0.85</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>aa index</td>
<td>0.65</td>
<td>0.03</td>
<td>0.68</td>
<td>0.70</td>
<td>-0.59</td>
<td>0.33</td>
<td>1.00</td>
</tr>
</tbody>
</table>
strengths, and thus the open flux is nearly constant between sunspot minimum and sunspot maximum. However, this seems to be true only during 1969–1980. Later, the open flux did show distinct maxima, out of phase with the sunspot cycle, with no fixed lags. Wang and Sheeley [2002] mention major peaks occurring in 1982 and 1991 (more than a year later than the sunspot maxima) and mention that “the open flux approximately follows that of the Sun’s total dipole strength, with a contribution from the magnetic quadrupole around sunspot maximum. Global fluctuations in sunspot activity lead to increases in the equatorial dipole strength and hence in the open flux and the IMF strength lasting typically \( \frac{1}{2} \text{ year} \).” This does not fit with a low correlation between the open flux and IMF-B (see Table 1).

2. The other parameters seem to have some relationship with each other. The crosses under the interplanetary V plot are for the geomagnetic aa indices. Table 1 gives the intercorrelations between the various parameters. The following may be noted.

3. Open flux is poorly correlated with all other parameters. Its maximum correlation is +0.65 with aa indices, and next best +0.51 with interplanetary B. Correlation with Rz is very low, +0.11.

4. Interplanetary N is poorly correlated with all other parameters and so is V, except for a good correlation (+0.68) between V and aa indices (already well-known).

5. Sunspots, cosmic rays and interplanetary B seem to be well correlated, in a general way [Cane et al., 1999].

To check whether the double-peak structure of solar indices is reflected in Sun’s open magnetic flux, the 12-month moving averages near sunspot maxima are plotted on an expanded scale in Figure 7a for cycle 22 (1988–1992) and Figure 7b for cycle 23 (1998–2002). Solar F10, interplanetary N, V, B, and cosmic ray intensities are also plotted. The open flux does show two maxima near the maxima of solar indices in both cycles 22 and 23, but in cycle 22 the second maximum is much larger than the first maximum. In cycle 23, the second maximum has probably not yet reached (August 2002). Thus the open flux does not show long-term variations similar to other indices, but medium-term features like double peaks separated by 1–2 years are seen in all solar indices (including open magnetic flux) almost simultaneously (within a month) and in cosmic ray minima with varying lags (including zero lag).

Active centers on the photosphere lead to coronal fields, Sun’s open flux, and interplanetary fields. Thus the Sun’s open flux and the interplanetary magnetic field should be directly related, as the interplanetary field is simply the extended open flux of the Sun, and these two quantities represent different ways of measuring or obtaining the same quantity. The low correlation between Sun’s open flux and interplanetary fields indicates some kind of a lacuna in the models or misfit between the Sheeley-Wang model and the measurements of the IMF. The lack of coincidence between sunspot maxima and the maxima of the open flux (occurring after more than 1 or 2 years) is disconcerting. Obviously, this open flux could not be the one directly responsible for cosmic ray long-term modulation. The flux must be getting considerably modified in transit to the outer heliosphere and there, falling in line with solar activity, so that its effect on cosmic rays would be with a lag of few months only. As it is, the maximum of Sun’s open flux lags behind sunspots by more than 1 year so that cosmic ray modulation will have no lag at all and perhaps even a lead with respect to the open flux.

Lockwood et al. [1999] developed an alternative method for estimating this open flux by using observations...
of geomagnetic activity. Lockwood [2001] reported that cosmic ray fluxes were highly anticorrelated with their flux values, which is satisfactory. However, the real situation of magnetic field in the heliosphere could be seen only by in situ observations. The Ulysses spacecraft provided very useful information, namely, the radial component of the heliospheric magnetic field is approximately independent of latitude. For the ascending phase of the present cycle 23, Burlaga et al. [2002] reported that the Voyager 1 (V1) observations of the heliospheric magnetic field strength B from 1 to 81 AU during 1978–2001 agreed with Parker’s model; namely, B decreased with increasing distance and had three broad maxima around 1980, 1990, 2000 (sunspot maxima of cycles 21, 22, 23) and minima in 1987 and 1997.

6. Conclusions and Discussion

The 12-month moving averages of solar indices and cosmic ray intensity were compared for 1953–2002. The following was noted.

1. The long-term variation had a negative correlation, with CR minima (maxima) roughly coinciding with solar maxima (minima), a feature already known.

2. From the hysteresis plots for five cycles (odd cycles 19, 21, 23; even cycles 20, 22), it seems that the modulation of CR near solar maximum is not simple. Whereas odd and even cycles show different patterns, the detailed evolution in similar cycles (even or odd) is not alike. The reason for complication is as follows.

3. On medium-term time scales (1–3 years), in some sunspot cycle maxima (notably cycles 22 and 23), there were two peaks in solar indices separated by about 2 years. These were reflected in cosmic rays as two distinct minima. However, some cosmic ray minima were with a lag of a few months with respect to maxima of solar indices, but others were almost coincident with maxima of solar indices.

4. From these observations, one may conclude the following. The observed CR modulation as observed on Earth contains two parts: (1) long-term variation due to characteristics in total heliosphere, determined by convection-diffusion and drift mechanisms described by Dorman [2001] and Dorman et al. [2001a, 2001b] (this part has time lags of several months and is characterized by hysteresis phenomenon) and (2) short-term and medium-term variations caused mainly by accumulated shock waves (Forbush effects and associated phenomena) with time lags of a few days, which in the scale of months or a year or two, would show a time lag zero). The second component would be especially important in maximum of solar activity.

The lags between maxima of solar indices and minima of cosmic rays are often used for estimating the dimensions of the heliosphere. The (almost) coincidences (or very small lags) were probably seen by many earlier workers, who used these for estimating the size of the heliosphere. In the mid-1960s, earlier estimates, based on comparison with coronal green line or by examining CR modulation caused by sudden jumps in solar activity [Dorman et al., 2001a, and references therein] indicated that the radius of the heliosphere was very small, about 5 AU, not more than 10–15 AU. Dorman [1975, and references therein] initiated the use of a convection-diffusion model, taking into account the time lag of processes in interplanetary space relative to the processes on the Sun and concluded that the modulation region should be much larger, between 50 and 150 AU. Further, Dorman [2001] took into account the drift effects (as depending upon the sign of solar polar magnetic field) according to Burger and Potgieter [1999] and showed different effects for even and odd solar cycles [see also Dorman et al., 2001a, 2001b]. In most of these analyses, correlations were used between 11-month moving averages of cosmic rays and sunspot number, and the average lags so obtained were compared with theoretical estimates for different heliospheric sizes. Thus finer features like the double-peak structures were ignored. If these features are considered, and the fact that the lags were almost zero in some cases and varied in a large range (7–20 months) in some other cases, is taken into account, estimates of the heliospheric dimensions may turn out to be different. In any case, the cosmic ray recovery as seen on Earth seems to have two distinct components, one related to the conditions in the large scale heliosphere and another related to conditions nearer the Sun on a medium-term (1–3 year) time scale, and these two components may have different proportions in different cycles, probably also in odd and even cycles.

Acknowledgments. This work was partially supported by FNDCT, Brazil, under contract FINEP-537/CT.

Shadia Rifai Habbal thanks Lev Dorman and another referee for their assistance in evaluating this paper.

References


