Solar Activity and Geomagnetic Storms: The First 40 Years

E.W. Cliver

Originally published in:
Eos, Transactions, American Geophysical Union, Vol. 75, No. 49,
December 6, 1994, Pages 569, 574-575

This article is the first of a series of three that traces the evolution of our understanding of the relationship between solar and geomagnetic activity from the mid-19th century to the present era.

The origin of solar-terrestrial physics is generally traced to Sabine's recognition in 1852 that geomagnetic activity paralleled the recently discovered sunspot cycle. Perhaps less well appreciated is the slow and uneven pace of progress since then in forging definite links between solar and geomagnetic activity. Thus 50 years after the initial excitement attending Sabine's discovery, the role of the Sun as the fundamental cause of magnetic storms was a matter of contention, and nearly a century would pass before statistics were sufficient to make a convincing case for an association between large solar flares and severe storms.

More recently, M regions remained mysterious until coronal holes were discovered in X ray images in the early 1970s, while energy coupling between the solar wind and the magnetosphere remains a topic of vigorous research. Within the past year, the conventional view of solar flares as the source of great magnetic storms has been strongly challenged, and revisions have been proposed to the "coronal hole as M region" paradigm.

The difficulties faced by 19th century solar astronomers and magneticians in unravelling the relationships between solar and magnetic activity were formidable. Geomagnetic storms can have two distinct solar sources: recurrent high-speed wind streams from coronal holes, and coronal mass ejections (CMEs) associated with eruptive solar flares and disappearing solar filaments (DSFs). Since coronal holes correspond to open solar magnetic field lines, while CMEs originate in closed field regions on the Sun, the chromospheric manifestations of these phenomena are antithetical.

There are other subtleties: not all flares are eruptive; some eruptive events-that is, DSFs-do not look like flares; solar sources of magnetic storms are most geoeffective when they are located near the Sun's central meridian as viewed from Earth; and solar wind disturbances may take from less than 1 day to more than 4 days to propagate from the Sun to the Earth. Adding to the difficulty, the elements of the Earth's magnetic field (declination, horizontal force, vertical force) at any given station exhibit both daily and storm variations that are primarily due to the Sun's ionizing-electromagnetic and corpuscular emissions, respectively. Thus while both the daily and storm variations are modulated by the 11-year solar activity cycle, the underlying physics is different. Finally, the component of the magnetic field of transient or recurrent wind streams that is aligned with the Earth's dipole axis acts as an "on-off" coupling switch, depending on its sign, between the interplanetary magnetic field and the magnetosphere. This variable coupling efficiency produces a semiannual variation in magnetic disturbances with peak activity occurring at the equinoxes.

Virtually none of this was known at the time of Sabine's discovery. Solar flares had not yet been observed, and tools for observing CMEs, sensing interplanetary space, and tracking disturbances enroute from the Sun to Earth were several generations of scientists in the future. At the receiving end, the magnetosphere had not yet been discovered. All that existed in 1852 were a few decades of incomplete solar and magnetic observations and enthusiasm for what was recognized as an important beginning. Given the complexity of the problem and the meager resources for its solution, it should not be surprising that the elucidation of relationships between solar activity and magnetic storms was a long and arduous process.

Prehistory

In 1722 the English scientist Graham devised a compass sensitive enough to determine that the geomagnetic field could undergo large and rapid variations. Nearly 20 years later in Sweden, Celsius and his student Hiorter noted that aurorae were accompanied by disturbances of the magnetic needle and, in collaboration with Graham in London, they determined that magnetic disturbances were not local phenomena [Chapman and Bartels, 1940]. Early studies of magnetic storms, as the rapid variations of the field came to be called, were made by Baron von Humboldt in Germany and Arago in France during the first part of the 19th century.

Systematic observations of Earth's magnetism were given an impetus from the success of the German mathematician Gauss during the 1830s in describing the Earth's magnetic field as an expansion of spherical harmonics. Gauss determined that nearly all of the field originated inside the Earth. His model required widespread observations for its verification and improvement. Thus the Gottingen Magnetic Union, a loose coalition of continental observatories, was formed in 1836 by Gauss and colleagues.

Knowledge of Earth's magnetism was also important for ocean navigation. This, as well as scientific and nationalistic considerations [Cawood, 1979], resulted in the formation of the British Colonial Observatories in 1840 under the direction of Lieutenant Colonel Edward Sabine. Observatories were established at Toronto, Hobarton (Tasmania), the Cape of Good Hope, and St. Helena.

The Birth of Solar-Terrestrial Physics

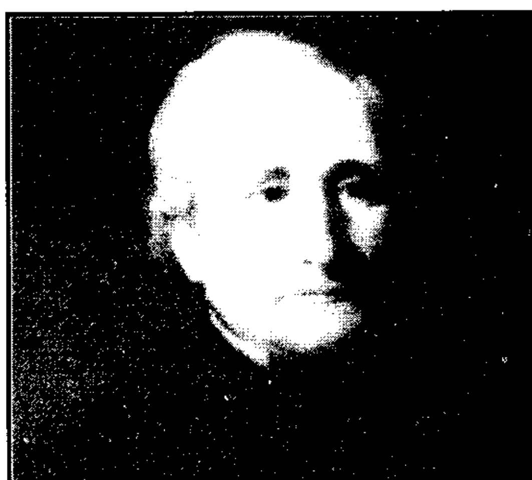

*Fig. 1. Edward Sabine (1788–1883). **Director of the British Colonial Observatories and a founder of solar-terrestrial science. Reproduced with permission of the National Portrait** Gallery, London.*

At roughly the same time that regular magnetic observations were initiated, a systematic examination of the Sun was begun by an amateur astronomer in central Germany. S. Heinrich Schwabe, a pharmacist, began observing the Sun and making counts of sunspots in 1826 at least partly in the hope of discovering intra-Mercurical planets. Schwabe's first publication in 1838 indicated periodic behavior in spot counts, but he did not explicitly call attention to the periodicity of "about 10 years" until late 1843. His result went virtually unnoticed until it was referred to by Humboldt in the third volume of Kosmos in 1851. Schwabe was to continue his patrol of the Sun for nearly 40 years. His diligence was praised by President Johnson of the Royal Astronomical Society in 1857 when the Society awarded him its Gold Medal, for the discovery of the sunspot cycle [Newton, 1958].

A decennial period in the daily variation of magnetic declination had been reported by Lamont from Munich in 1851, but he did not relate it to the sunspot cycle. Such a connection was first made by Sabine (Figure 1) in 1852, who noted that minima in the average rate and size of magnetic disturbances at the widely separated Hobarton and Toronto observatories in 1843 corresponded to a minimum in sunspot numbers, while maxima in 1848 corresponded to a maximum in the decennial sunspot curve. Thus a comparison of the data in Sabine's Tables XIII and XXII (Figure 2) marked the origin of solar-terrestrial physics. Within months of Sabine's discovery it was reported independently by Gautier and Wolf in Switzerland. Wolf refined the average duration of the sunspot cycle to 11.1 years on the basis of reconstructed sunspot numbers obtained from scattered reports dating back to 1610, when telescopic observations of spots were first made.

TABLE XIII.

Years.	Numbers.		Values.		Years.
	Toronto.	Hobarton.	Toronto.	Hobarton.	
1843.........	0·68	0·52	0·55	0·48	1843.
1844.........	0·76	0·81	0·73	0·82	1844.
1845.........	0·72	0·72	0·62	0·67	1845.
1846.........	1·31	1·09	1·26	1·03	1846.
1847.........	1·19	1·36	1·40	1·44	1847.
1848.........	1·37	1·50	1·43	1·60	1848.

Fig. 2. Tables 13 and 22 from Sabine [1852] show that the rates and sizes of magnetic disturbances observed at Toronto and Hobart tracked the sunspot cycle during 1843–1848.

TABLE XXII.

Years.	Groups of spots.	Days free from spots.	No. of days of observation.
1826.........	118	22	277
1827.........	161	2	273
1828.........	225	0	282
1829.........	199	0	244
1830.........	190	1	217
1831.........	149	3	239
1832.........	84	49	270
1833.........	33	139	267
1834.........	51	120	273
1835.........	173	18	244
1836.........	272	0	200
1837.........	333	0	168
1838.........	282	0	202
1839.........	162	0	205
1840.........	152	3	263
1841.........	102	15	283
1842.........	68	64	307
1843.........	34	149	312
1844.........	52	111	321
1845.........	114	29	332
1846.........	157	1	314
1847.........	257	0	276
1848.........	330	0	278
1849.........	238	0	285
1850.........	186	2	308

Growing Pains

The discovery of a sunspot cycle and its reflection in terrestrial magnetism was the cause of great contemporary excitement. In a letter to Faraday in 1852, Herschel, (as cited by Meadows and Kennedy [1982]),remarked on Sabine's discovery, "If all this be not premature we stand on the verge of a vast cosmical discovery such as nothing hitherto imagined can compare with." The anticipated progress, however, was slow in coming. Key observations were lacking, and much of the basic physics had yet to be created. Also, while there was a general relationship between sunspots and magnetic activity, apparent in yearly averages, the correspondence failed when the data were examined for shorter timescales. Thus as early as 1858, Broun from Scotland casually remarked, "It is not the case that the magnetic disturbance coexists always with the spots...". This failure "in the details" of a relationship between sunspots and magnetic storms was to become a central problem of solar-terrestrial physics.

A clue presented itself on September 1, 1859, when R. C. Carrington, the English astronomer who greatly advanced the state of systematic solar observation, first observed a solar flare, while making his daily sunspot drawing (Figure 3a). Fortunately, Carrington's observation was confirmed by Hodgson, an English amateur observing nearby. Carrington compared his observations with the magnetic records from Kew Observatory outside of London and noted that a definite but short-lived magnetic disturbance occurred at the time of the flare while a great geomagnetic storm began about 18 hours later (Figure 3b). Despite the temporal association of the flare and the magnetic varations, Carrington declined to draw a physical connection between the phenomena. Balfour Stewart, the Director of Kew Observatory, was less cautious. He suggested that the short-lived disturbance in the Kew traces at the time of Carrington's observation was a case of "our luminary ... taken in the act" of producing a magnetic response.

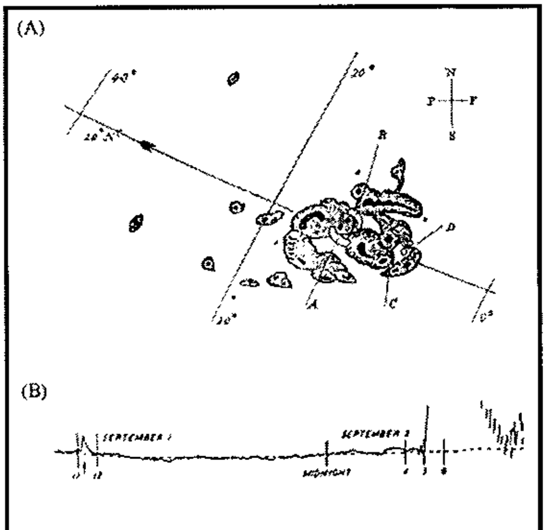

Fig. 3. a) Carrington's 1860 drawing of the first reported flare (the light areas marked A, B, C, and D) observed in white light during his daily sunspot drawing on September 1, 1859. b) The Kew record of the horizontal magnetic force showing the prompt (September 1) and delayed (September 2) effects of the eruptive flare observed by Carrington.

It was a long time, however, before the tantalizing clue provided by this solar-terrestrial event would come to fruition. A basic problem was that no solar instrument was yet available that could image flares in narrow lines such as H-alpha in which they are most prominent. Thus the various compilations of major flares for the 19th century contain only about 10 independent events, observed either in white light or as brilliant reversals of spectroscopic lines near sunspots. Other reversals were frequently observed by spectroscopists but were not systematically reported.

The American astronomer Young noted simultaneous magnetic disturbances during such brilliant reversals in 1872, but the magnetic deflections were small and the associations were later discounted by others as chance coincidences. The fact that the Sun could have both prompt and delayed geomagnetic effects, as manifested by Carrington's flare, would be a source of confusion until well into the following century.

In a series of papers in the American Journal of Science, Loomis tabulated the positions of the widespread aurorae of late August and early September 1859. In a related effort, he determined that the region of the Earth's Northern Hemisphere, where aurorae were most frequent consisted of a somewhat irregular oval-shaped ring centered in northwest Greenland, work that was later improved on by Fritz in Switzerland. Fritz, Loomis, and others established a previously suspected relationship between the sunspot cycle and the occurrence of mid-latitude aurora.

In 1868 the Frenchman Janssen and the Englishman Lockyer demonstrated independently that solar prominences, elevated structures observed at the Sun's limb, could be observed outside of eclipse with spectroscopes. By 1871, Secchi in Italy and others had classified prominences into active and quiescent types. The active or eruptive prominences occasionally displayed velocities of the order of the escape velocity of the Sun. Such observations would eventually provide support for the "corpuscular theory" for the origin of magnetic storms, in which the responsible solar disturbances propagated to Earth as a cloud of charged particles. Their significance was not immediately appreciated, however, because it was thought that the sunspots were the true source of storms.

This is not to say that Sabine's sunspot-magnetic storm link was accepted without reservation. As early as 1863, it was clear to William Thomson, later Lord Kelvin, that direct action of the Sun as a magnet was incapable of causing magnetic storms. He calculated that the Sun's magnetism would need to be 120 times as strong as the Earth's for even a complete reversal of the solar field to cause a small change in magnetic declination at Earth. This critique, appended to a note by Chambers on the same subject, would be given in a more prominent forum by Lord Kelvin nearly 30 years later.

An alternative to a solar cause of storms presented itself in the form of Earth currents that were found to accompany magnetic storms. Because of the close correspondence between the two phenomena, Airy, the English Astronomer Royal, suggested in 1868 that sudden variations in the Earth's magnetic field were caused by the superposed magnetic fields of the transient Earth currents.

In contrast, Stewart thought that the currents responsible for the daily variation of the Earth's field lay in the upper atmosphere. He hypothesized that the rarefied gas there was a good conductor and reasoned that the tidal motions of such gas across the lines of force of the Earth's magnetic field would produce the electromotive force and currents required to explain the daily magnetic variation. In addition to providing the essentially correct explanation for the diurnal variation, Stewart's ideas, published in the ninth edition of the Encyclopedia Britannica in 1878, represent the beginnings of ionospheric research. Later, Schuster, following the harmonic analysis of Gauss, demonstrated mathematically that the currents responsible for the daily variation of the magnetic elements lay mainly outside the Earth.

A comparison by Ellis in 1880 of the daily range of magnetic variation observed at Greenwich (after removal of the larger storms) with Wolf's sunspot numbers for the years 1841-1877 strongly confirmed the earlier studies of Sabine, Wolf, and others showing that 11-year cycles for both solar activity and geomagnetism occurred approximately in phase. In annual averages, the sunspot count and various measures of magnetism, including the occurrence of aurorae, clearly tracked each other. On shorter timescales, however, the correspondence between these two phenomena broke down. Veeder from New York stated the crux of the problem in 1889, "When, however, an attempt is made to ... study these phenomena in detail from day to day, instead of by a system of averaging through extended periods, serious difficulties are encountered. Very often for days together, spots are numerous upon the Sun, and yet there is no aurora and no extraordinary variation of magnetic declination. Likewise, on the other hand, fine auroras are not infrequently seen when the Sun is absolutely devoid of dark spots.... It requires but little observation to show that something else upon the Sun beside the dark spots is concerned in the production of magnetic phenomena..."

Veeder's comparison of spots and aurorae convinced him that spots were most geoeffective at the east limb of the Sun. In a letter to Astronomy and Astrophysics in 1891, Veeder asserted, "I am very decidedly of the opinion that the solar impulses originating terrestrial magnetic phenomena are not conveyed either as heat or light in any form whatever." From a separate letter, "It will be found ere long, I think, that the subject is a very live one..."

As the 1890s began, the relationship between the solar cycle and magnetic disturbances was almost universally accepted [Meadows, 1970]. Quoting from a contemporary textbook [Young, 1888], "One influence of the sunspots upon the Earth is perfectly demonstrated. When the spots are numerous, magnetic disturbances ... are most numerous and violent upon the Earth ... The nature and mechanism of the connection is as yet unknown, but of the fact there can be no question." Such certainty notwithstanding, this point of view would soon receive a serious challenge from the eminent Victorian scientist, Lord Kelvin. Kelvin's challenge and the origin and development of the corpuscular hypothesis for magnetic storms will be recounted in a subsequent article.

References

Cawood, I., The magnetic crusade: Science and politics in early Victorian Britain, *Isis, 70*, 493, 1979.

Chapman, S., and J. Bartels, *Geomagnetism*, Oxford University Press, London, 1940.

Meadows A J., *Early Solar Physics*, Pergammon, New York, 1970.

Meadows, A J., and J. E. Kennedy, The origin of solar-terrestrial studies, *Vistas Astron., 25*, 419, 1982.

Newton, H. W., *The Face of the Sun*, Penguin Books Ltd., Baltimore, 1958.

Young, C. A, *A Textbook of General Astronomy*, Ginn and Co., Boston, 1888.