GEOPHYSICAL RESEARCH LETTERS, VOL. 33, L16102, doi:10.1029/2006GL026687, 2006

Statistical modeling of storm-level *Kp* occurrences

K. J. Remick[1] and J. J. Love[1]

Received 24 April 2006; revised 10 July 2006; accepted 14 July 2006; published 22 August 2006.

[1] We consider the statistical modeling of the occurrence in time of large *Kp* magnetic storms as a Poisson process, testing whether or not relatively rare, large *Kp* events can be considered to arise from a stochastic, sequential, and memoryless process. For a Poisson process, the wait times between successive events occur statistically with an exponential density function. Fitting an exponential function to the durations between successive large *Kp* events forms the basis of our analysis. Defining these wait times by calculating the differences between times when *Kp* exceeds a certain value, such as $Kp \geq 5$, we find the wait-time distribution is not exponential. Because large storms often have several periods with large *Kp* values, their occurrence in time is not memoryless; short duration wait times are not independent of each other and are often clumped together in time. If we remove same-storm large *Kp* occurrences, the resulting wait times are very nearly exponentially distributed and the storm arrival process can be characterized as Poisson. Fittings are performed on wait time data for $Kp \geq 5$, 6, 7, and 8. The mean wait times between storms exceeding such *Kp* thresholds are 7.12, 16.55, 42.22, and 121.40 days respectively.

Citation: Remick, K. J., and J. J. Love (2006), Statistical modeling of storm-level *Kp* occurrences, *Geophys. Res. Lett.*, *33*, L16102, doi:10.1029/2006GL026687.

1. Introduction

[2] Because magnetic storms are potentially hazardous to the infrastructure and activities of modern technological systems [e.g., *Allen et al.*, 1989; *Boteler et al.*, 1998], their prediction has long been a goal of the space science community. Traditionally, predictions have been obtained using data-derived, deterministic models [e.g., *Joselyn*, 1995; *McPherron and Siscoe*, 2004; *Wing et al.*, 2005] that produce excellent estimates of magnetic activity up to a few days in advance. When forecasting the occurrence of storms over time scales longer than a few days, however, probabilistic modeling is needed. Here we derive a simple, probabilistic forecast model for storm occurrence based on the historical statistics of *Kp* data.

2. Data

[3] The K index quantifies disturbed magnetic-field activity at an observatory by assigning a 3-hour range of the H or D components to a quasi-logarithmic scale: 0 for the most quiet to 9 for the most disturbed [*Mayaud*, 1980; *Rangarajan*, 1989]. The corresponding planetary magnetic index, *Kp*, is the average of fractional K values at 13 subauroral observatories [*Menvielle and Berthelier*, 1991], giving *Kp* values of 0, 0.3, 0.7, 1.0, 1.3, . . . et cetera. Two of the source observatories are in the United States (FRD/CLH, SIT), two in Canada (MEA, OTT/AGN), seven in northern Europe (BFE/RSV, ESK, HAD/ABN, LER, LOV/UPS, NGK/WIT, WNG), and one each in Australia (CNB/TOO) and New Zealand (EYR/AML). Though the geographic distribution of sites is not even close to uniform, *Kp* is a very robust measure and its utility has been amply demonstrated [e.g., *Thomsen*, 2004] over the 73 years that it has been continuously calculated and archived. Thus *Kp* is ideal for long-term statistical characterization of planetary magnetic disturbances.

3. Theory

[4] We examine the occurrences of magnetic disturbances defined by *Kp* exceeding a given integer value, and the wait times between such occurrences. How are these occurrences statistically distributed? The Poisson process is a classical stochastic model of a series of discrete and independent events realized from a memoryless process [*Cox and Lewis*, 1966; *Blumer*, 1979]. The Poisson probability density function (pdf) is

$$p_p(\mathrm{n}|\lambda t) = \frac{(\lambda t)^{\mathrm{n}} \mathrm{e}^{-\lambda t}}{\mathrm{n}!}. \tag{1}$$

where n is the number of events, λ is the rate of their occurrence, and t is a duration of time. Since the Poisson distribution applies only to the counting of discrete events, the cumulative distribution function (cdf) is the sum of the probabilities over a range of counts, giving the probability that n events, or less, will occur within time t

$$P_p(\mathrm{n}|\lambda t) = \sum_{i=0}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda t} (\lambda t)^i}{i!}. \tag{2}$$

For wait times between events, $n = 0$, giving an exponential pdf

$$p_e(t|\lambda) = \lambda \mathrm{e}^{-\lambda t}, \tag{3}$$

where t is now the wait time between successive events. If the wait-time data are exponentially distributed, the best fit will occur where λ is the inverse of the mean wait time. The cdf is the integral of the pdf over the time considered and gives the probability of successive occurrences. Thus, the probability of the occurrence of a subsequent event before an elapsed time t is

$$P_e(t|\lambda) = \int_0^t p(u)\mathrm{du} = \int_0^t \lambda e^{-\lambda \mathrm{u}} \mathrm{du} = 1 - \mathrm{e}^{-\lambda t}. \tag{4}$$

[1]Geomagnetism Program, U.S. Geological Survey, Denver, Colorado, USA.

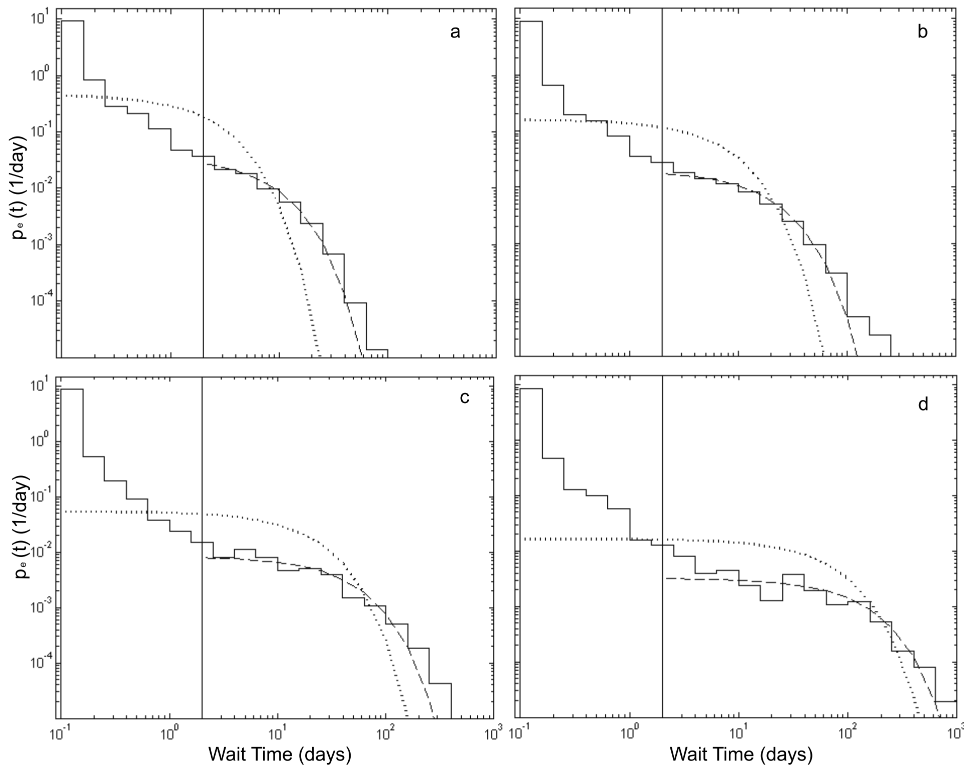

Figure 1. Pdfs of wait times and exponentials determined by all data and, also, data with storm-time clumping removed $T_{\geq 2}$: pdf of the wait times between occurrences (a) for $Kp \geq 5$, (b) with a threshold of $Kp \geq 6$, (c) for $Kp \geq 7$, and (d) for $Kp \geq 8$. Solid line histograms are the data pdfs $p(t_i)$. Dotted lines are the pdfs $p_e(t|\lambda_a)$ determined using all the data. Dashed lines are the pdfs $p_e(t|\lambda_{\geq 2})$ determined using $T_{\geq 2}$. The vertical line denotes a wait time of two days.

These formulae form the basis of the analysis that follows.

4. Results

[5] We first test to see if the wait times between storm level *Kp* have an exponential distribution by comparing the actual wait times with a theoretical, exponential function. Using the *Kp* data between Jan 1 1932 and Dec 31, 2004, the wait times t_i between occurrences of storm level *Kp* were calculated. Separate calculations were made for the cases $Kp \geq 5,6,7$ and 8. $Kp = 9$ was omitted due to the low number of events. In Figure 1 we show the pdfs $p(t_i)$ for the actual wait times; the corresponding cdfs $P(t_i)$ are shown in Figure 2. In Figures 1 and 2 we also show the exponential pdfs $p_e(t|\lambda_a)$ and cdfs $P_e(t|\lambda_a)$ calculated using equations (3) and (4) respectively, where λ_a is one over the mean wait time. A summary is given in Table 1.

[6] With this simplistic but straightforward treatment, the wait-time data distributions are not well fitted by exponential functions. The fits fail the Kolmogorov-Smirnov goodness of fit tests at a 99% confidence level [e.g., *Press et al.*, 1992], but the misfits are visually obvious as well. This failure is due primarily to the disproportionate population of wait times having durations of less than two days. This is not surprising as large storms often have several 3-hour periods with large *Kp* values, and their occurrences are either consecutive, or separated by relatively short durations of time. Multiple occurrences of large *Kp* values within a single storm result in temporal clumping of short wait times. Such clumping is contrary to a model that assumes statistically independent data realized in time from a memoryless process.

[7] Improved fits, also shown in Figures 1 and 2, are obtained by modeling a subset of the data, $T_{\geq 2}$, which contains only data with wait times greater than two days. Instead of considering the occurrences of storm level *Kp*, we are now considering the occurrences of storms having high *Kp* levels. The storms may be of any length, but they must be followed by at least 16 consecutive *Kp* readings below the threshold value. Because we are excluding short duration wait times, some of which are consistent with a Poisson process, the mean wait times determined by averaging the remaining data do not yield the proper occurrence rates $1/\lambda_{\geq 2}$. We therefore treat $\lambda_{\geq 2}$ as a model parameter and estimate its value by maximizing the Kolmogorov-Smirnov goodness of fit to the $T_{\geq 2}$ data subset. As $T_{\geq 2}$ contains only a fraction of the total data, we rescale the data-derived pdf in order to compare it with the new exponential. The resulting fits pass the Kolmogorov-Smirnov test at the 99% confidence level and are visually compelling as well. The mean wait times corresponding to the fitted rates are given in Table 1.

5. Discussion

[8] Wait times of greater than two days are well fitted by an exponential function. This implies, subject to caveats we

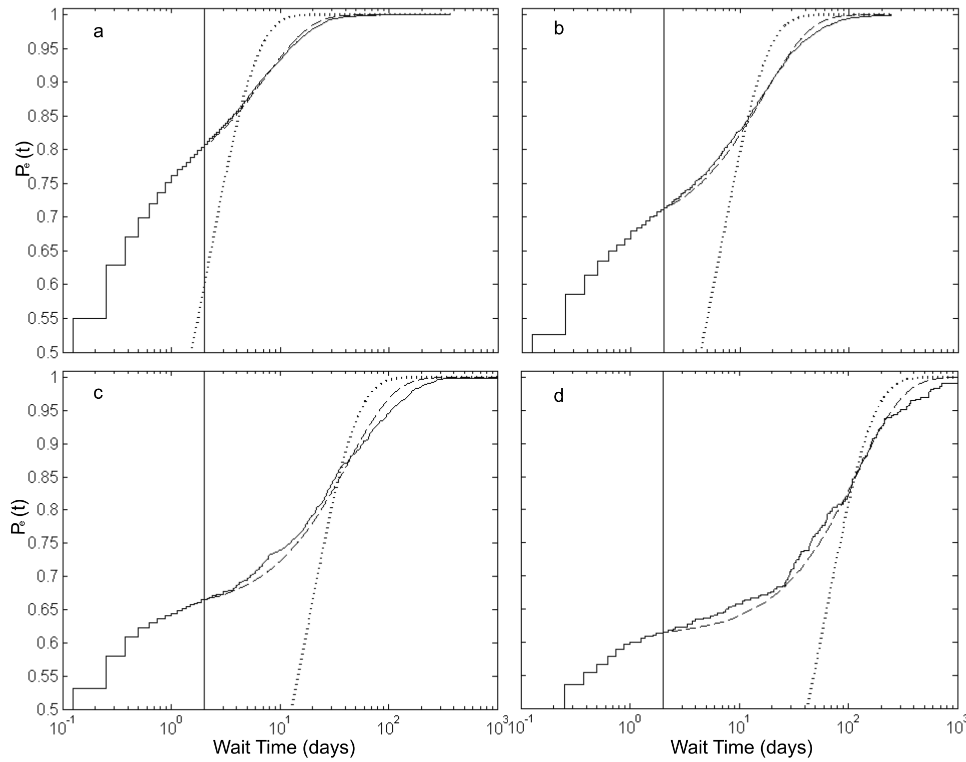

Figure 2. Cdfs of wait times and exponentials determined by all data and, also, data with storm-time clumping removed $T_{\geq 2}$: cdf of the wait times between occurrences with (a) $Kp \geq 5$, (b) $Kp \geq 6$, (c) $Kp \geq 7$, and (d) $Kp \geq 8$. Solid lines are the data cdfs $P(t_i)$. Dotted lines are the cdfs $P_e(t|\lambda_a)$ determined using all the data. Dashed lines are the cdfs $P_e(t|\lambda_{\geq 2})$ determined using $T_{\geq 2}$ and scaled using the factors in Table 1. The vertical line denotes a wait time of two days and shows the percentage of the data that is not included in the $T_{\geq 2}$ subset.

will pursue in future work, that storm occurrence can be considered to be a Poisson process, and this allows for probabilistic prediction. In Figure 3 we show the pdfs (Figure 3a) and cdfs (Figure 3b) for the exponentials determined from $T_{\geq 2}$, whose occurrence rates are shown in Table 1. The probabilities are different from those in Figure 2, as those values where scaled for comparison with the temporally clumped data. The continuous backward extrapolation for wait times less than two days is included as a logical estimate of the probabilities of independent storm occurrences at these wait times. Since the end of a wait time implies a storm occurrence, the wait time cdf shows the probability that the quiet time will end and a storm will occur. Thus the probability of at least one storm

Table 1. Mean and Model Wait Times for Storm-Level *Kp*

	All Data ($1/\lambda_a$), days	$T_{<2}$ ($1/\lambda_{T<2}$), days
$Kp \geq 5$	2.21	7.12
$Kp \geq 5.3$	3.05	9.85
$Kp \geq 5.7$	4.37	11.93
$Kp \geq 6$	6.34	15.85
$Kp \geq 6.3$	8.81	23.01
$Kp \geq 6.7$	12.11	31.32
$Kp \geq 7$	18.50	54.91
$Kp \geq 7.3$	24.72	59.91
$Kp \geq 7.7$	36.09	76.62
$Kp \geq 8$	60.81	121.40
$Kp \geq 8.3$	84.03	164.31

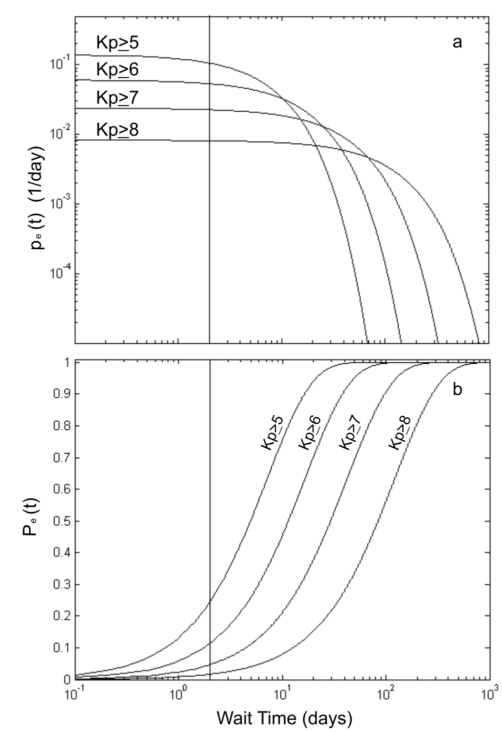

Figure 3. Exponential (a) pdf $p_e(t|\lambda_{\geq 2})$ and (b) cdf $P_e(t|\lambda_{\geq 2})$ for $Kp \geq 5$, 6, 7 and 8.

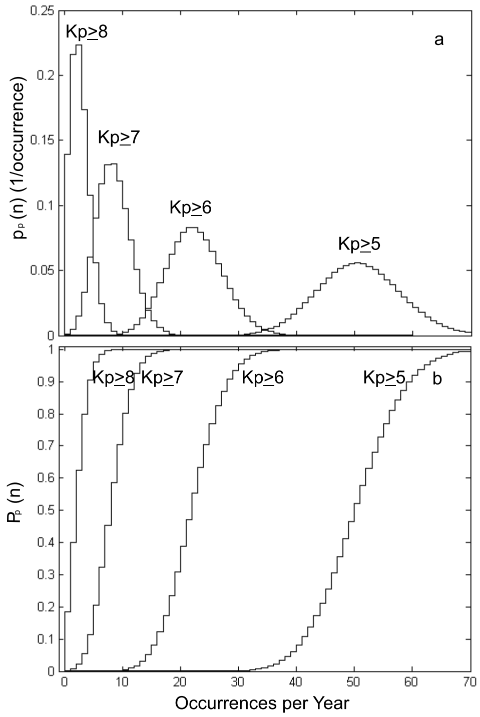

Figure 4. Poisson (a) pdf $p_p(\mathrm{n}|\lambda_{\geq 2}t)$ and (b) cdf $P_p(\mathrm{n}|\lambda_{\geq 2}t)$ for $Kp \geq 5$, 6, 7 and 8 where t = 1 year.

occurring in a given time duration, can be read directly from Figure 3b. Because Poisson processes are memoryless, the clock can be started at any time and need not be timed from the occurrence of the last storm. The rates determined in this work can also be used in the Poisson distribution, equations (1) and (2), to determine probability of a number of events occurring in a given timeframe. We show in Figure 4a the probabilities for the occurrence of discrete numbers of storms in one year, and in Figure 4b the probability that there will be a given number of storms or less, per year.

6. Future Work

[9] The statistical analysis conducted here assumes stationarity; that equation (1) is valid for Poisson processes whose statistical properties do not change over time. However, occurrence rates are known to vary with season and solar cycle phase. Taking these variations into account involves introducing a time dependent rate function $\lambda(\tau)$, where τ specifies the cyclical phase. Other periodic modulations, such as those associated with solar rotation are worthy of additional investigation. By generalizing the model to consider temporal nonstationarity, the statistical nature of large *Kp* storm occurrence will be more fully characterized, and the fits to the data will, almost certainly, be improved.

[10] **Acknowledgments.** We thank the geomagnetic observatory programs of Australia, Canada, Denmark, Germany, New Zealand, Sweden, the United Kingdom, and the United States for the collection of the necessary magnetometer data. We thank the GeoForschungsZentrum in Potsdam for calculating the *Kp* index and the Danish Meteorological Institute–Solar Terrestrial Physics Division World Data Center for Geomagnetism, Copenhagen, from which the *Kp* data were downloaded. Support for the first author was provided by the USGS Mendenhall Postdoctoral Research Fellowship Program. D. M. Perkins is thanked for many useful conversations. We thank two anonymous reviewers for their constructive comments.

References

Allen, J., L. Frank, H. Sauer, and P. Reiff (1989), Effects of the March 1989 solar activity, *Eos Trans. AGU*, *70*, 1479, 1486–1488.

Blumer, M. G. (1979), *Principles of Statistics*, Dover, Mineola, N. Y.

Boteler, D. H., R. Pirjola, and H. Nevanlinna (1998), The effects of geomagnetic disturbances on electrical systems at the Earth's surface, *Adv. Space Res.*, *22*, 17–27.

Cox, D. R., and P. A. W. Lewis (1966), *The Statistical Analysis of Series of Events*, Methuen, New York.

Joselyn, J. A. (1995), Geomagnetic activity forecasting, *Rev. Geophys.*, *33*, 383–401.

Mayaud, P. N. (1980), *Derivation, Meaning, and Use of Geomagnetic Indices*, *Geophys. Monogr. Ser.*, vol. 22, AGU, Washington, D. C.

McPherron, R. L., and G. Siscoe (2004), Probabilistic forecasting of geomagnetic indices using solar wind air mass analysis, *Space Weather*, *2*, S01001, doi:10.1029/2003SW000003.

Menvielle, M., and A. Berthelier (1991), The K-derived planetary indices: Description and availability, *Rev. Geophys.*, *29*, 415–432.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes*, Cambridge Univ. Press, New York.

Rangarajan, G. K. (1989), Indices of geomagnetic activity, in *Geomagnetism*, vol. 3, edited by J. A. Jacobs, pp. 323–384, Elsevier, New York.

Thomsen, M. F. (2004), Why *Kp* is such a good measure of magnetospheric convection, *Space Weather*, *2*, S11004, doi:10.1029/2004SW000089.

Wing, S., J. R. Johnson, J. Jen, C.-I. Meng, D. G. Sibeck, K. Bechtold, J. Freeman, K. Costello, M. Balikhin, and K. Takahashi (2005), *Kp* forecast models, *J. Geophys. Res.*, *110*, A04203, doi:10.1029/2004JA010500.

J. J. Love and K. J. Remick, Geomagnetism Program, U.S. Geological Survey, Denver, CO 80225, USA. (kremick@usgs.gov)