SPACE WEATHER, VOL. 5, S12005, doi:10.1029/2006SW000287, 2007

Statistical modeling of storm level *Kp* occurrences: Solar cycle modulation

J. J. Love,[1] K. J. Remick,[1] and D. M. Perkins[2]

Received 26 September 2006; revised 29 July 2007; accepted 20 August 2007; published 29 December 2007.

[1] We consider the nonstationary, statistical modeling of the occurrence in time of large *Kp* geomagnetic storms over the course of multiple solar cycles. Previous work showed that wait times between storms can be represented by an exponential density function, consistent with the realization of a Poisson process. Here we also assume a Poisson process, but to account for solar cycle modulation of storm likelihood, we assume an occurrence rate given by a parametric constant plus a simple sinusoidal function of time. Parameter estimation is accomplished using maximum likelihood, yielding good fits to the *Kp* data. We find that the relative phase between storms and sunspots depends on storm size. We quantify previous observations that small storms tend to occur during the declining phase of the solar cycle, while large storms tend to occur very close to solar maximum. We predict average wait time between storms and the storm occurrence rate up through the year 2018.

Citation: Love, J. J., K. J. Remick, and D. M. Perkins (2007), Statistical modeling of storm level *Kp* occurrences: Solar cycle modulation, *Space Weather*, *5*, S12005, doi:10.1029/2006SW000287.

1. Introduction

[2] It is an important and well-known observation that the likelihood of geomagnetic storm occurrence is not constant in time; storm occurrence is modulated by solar cycle activity [e.g., *Garrett et al.*, 1974; *Echer et al.*, 2004]. Qualitatively, it has been recognized that storms of various magnitudes are most likely to occur during the declining phase of the solar cycle [*Chapman and Bartels*, 1962, pp. 371–372; *Gorney*, 1990, p. 321], but that the very largest storms tend to occur around the solar cycle maximum [*Chapman and Bartels*, 1962, p. 379; *Gorney*, 1990, p. 321]. Since large magnetic storms are potentially hazardous to the activities and infrastructure of our modern, technologically based society [e.g., *Carlowicz and Lopez*, 2002; *Bothmer and Daglis*, 2007; *Thomson*, 2007], it is important to quantify this observation so that useful statistical predictions can be made [e.g., *Feynman and Gu*, 1986; *Lanzerotti*, 2007]. Previously, *Remick and Love* [2006] showed that storm occurrences, defined in terms of exceedances of chosen *Kp* thresholds, can be described as realizations of a Poisson process. While this allows for probabilistic predictions, a constant occurrence rate was assumed in their analysis. Here we generalize the statistical model, quantifying the effects of the solar cycle. This allows us to make realistic predictions of solar-cycle-modulated, storm occurrence likelihood.

2. Data

[3] The data used in this study are derived from the *Kp* index time series. The *Kp* index quantifies planetary-scale, disturbed magnetic field activity over a 3-h period as measured at 13 subauroral observatories [e.g. *Mayaud*, 1980; *Menvielle and Berthelier*, 1991; *Love and Remick*, 2007]. Individual *Kp* values range from 0, 0.3, 0.7, 1.0, 1.3, et cetera (or, under another notational convention: 0_0, 0_+, 1_-, 1_0, 1_+, et cetera) during quiet times, up to a maximum of 9 for the most disturbed periods. Since its conception [*Bartels*, 1932], *Kp* has provided a temporally uninterrupted measure of magnetic activity over approximately seven solar cycles. Although the index does not have the underpinnings of a physical model, it is widely used in the assessment of space weather conditions [e.g., *NOAA Service Assessment Team*, 2004]. Therefore because of the longevity of its time series and its obvious utility we choose to use the *Kp* index for our statistical analysis of the long-term variation of storm occurrence likelihood. The *Kp* time series used here starts on 1 January 1932, near the end of solar cycle 16, and ends on 31 December 2006, around the completion of solar cycle 23.

[4] There are many ways to define and study magnetic storm occurrences over time. One way, which is appeal-

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

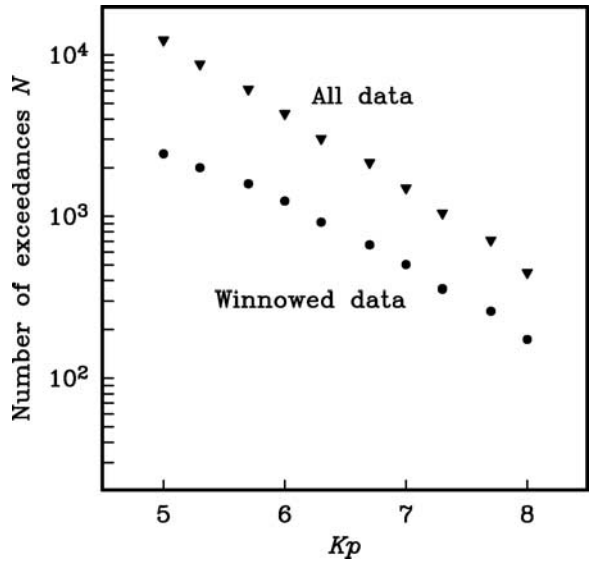

Figure 1. Number of *Kp* threshold exceedances as a function of *Kp* exceedance level. The squares show the number for all the data over the 75 a of the *Kp* time series. The circles show the number of winnowed data, where those with associated wait times of less than 2 d have been removed. The number of winnowed data effectively represent the number of magnetic storms.

ingly simple, is to define a storm occurrence as the moment in time when *Kp* exceeds a chosen threshold. With a long time series of *Kp* data, this definition can be used to construct a population of discrete storm occurrence times that can be analyzed statistically. With this in mind, let us examine the proposed storm definition closely. Consider a series of times defined by a chosen *Kp* exceedance threshold: t_i, t_{i+1}, t_{i+2}, t_{i+3}, ... These occurrence times define an associated set of wait times, that is, durations from one exceedance until the next: τ_i, τ_{i+1}, τ_{i+2}, τ_{i+3}, ..., where $\tau_j = t_{j+1} - t_j$, and which we may also express symbolically as a correspondence $\tau_j \leftrightarrow (t_j, t_{j+1})$. In their analysis of a similar time series, *Remick and Love* [2006] concluded that wait times of 2 d or less primarily represented repeated, causally related, and statistically dependent, intrastorm occurrences, while wait times of greater than 2 d represented interstorm occurrences that could be well described as statistically independent. The need to draw such an arbitrary distinction (2 d) is not surprising. We are attempting to model storm occurrences as a series of discrete events realized in probability over a long period of time, even though over short periods of time the evolution of a storm, or more generically, continuous durations of magnetic activity, are better described in terms of deterministic physics. Inevitably, there is some degree of overlap between the two paradigms and their timescales.

[5] Whereas *Remick and Love* [2006] concentrated only on the wait times τ_j between storm occurrences, without making any explicit modeling of the solar cycle modulation of storm occurrences, here we seek to make such a modeling and this means that we need to consider both wait times τ_j and occurrence times t_j. Since we can express, without ambiguity, each correspondence between these quantities as $\tau_j \leftrightarrow (t_j, t_j + \tau_j)$, then our data set can be completely described by the pairings (τ_i, t_i), (τ_{i+1}, t_{i+1}), (τ_{i+2}, t_{i+2}), ... For consistency with Remick and Love we adopt the 2-d cutoff criterion for the interstorm occurrence wait times. Specifically, suppose, for example, that $\tau_{i+1} < 2$ d, then upon winnowing we have the data pairs (τ_i, t_i), (τ_{i+2}, t_{i+2}), (τ_{i+3}, t_{i+3}), ... For convenience of notation, however, after winnowing we reassign the subscripts identifying the data pairs: (τ_i, t_i), (τ_{i+1}, t_{i+1}), (τ_{i+2}, t_{i+2}), ···. Henceforth, we refer to a winnowed and relabeled data set as $D^{\geq c}$, where c is the wait time cutoff level (2 d for consistency with Remick and Love).

[6] Looking at the data in general terms, in Figure 1 we show the number of occurrences N as a function of *Kp* exceedance threshold, both for the complete *Kp* data set and the winnowed data set. Isolated exceedances are very unlikely, especially for smaller thresholds; once we have a *Kp* exceedance this is usually quickly followed by several more. Therefore enforcement of the 2 d cutoff obviously reduces the number of data considerably. With winnowing, the number of event times for $Kp \geq 5$ ($Kp \geq 8$) is reduced from 12357 (448) to 2440 (174). We take the winnowed occurrence times as those of magnetic storms, and we list their number in the second column of Table 1. In practice it is difficult to consider the statistics of small data sets, and for this reason in this study we do not consider threshold greater than $Kp = 8$; specifically, for the 75-year (a) *Kp* time series there are only 23 storms exhibiting $Kp = 9$ and we have found that this is too few data to make a meaningful estimate of the needed parameters.

[7] In Figure 2 we show the time series of all the (unwinnowed) $Kp \geq 5$ data, the winnowed storm occurrence rates defined for different *Kp* exceedances, and the corresponding sunspot numbers, all since 1932 up to the end of 2006. Displayed in this way, the solar cycle modulation of storm likelihood defined for $Kp \geq 5$ or 6 is not particularly pronounced, and the situation for for $Kp \geq 7$ or 8 is obscured by the statistical jitter that results from the relative rarity of large storms. An objective statistical analysis is required in order to quantify the solar cycle modulation of storm level *Kp* occurrences.

3. Theory

[8] A Poisson process is a probabilistic model that can be used to statistically describe the occurrence in time of discrete random events. The basic properties of a Poisson process are (1) stationarity, the probabilities that describe the process do not change in time, (2) orderliness, the likelihood of two events occurring simultaneously is negligible, and (3) memorylessness, the occurrence likelihood

Table 1. Data and Model Parameter Summary for Different *Kp* Exceedance Thresholds[a]

$Kp\geq$	N	a (per year)	b (per year)	b/a	ω (radians per year)	T_s (years)	φ (radians)
5.0	2440	46.092	14.290	0.310	0.600	10.469	−0.152
5.3	2000	34.640	11.336	0.327	0.603	10.420	−0.012
5.7	1591	25.726	8.980	0.349	0.604	10.405	0.060
6.0	1242	19.149	7.864	0.411	0.598	10.499	0.063
6.3	920	13.628	5.998	0.440	0.597	10.526	0.136
6.7	665	9.571	4.676	0.489	0.599	10.491	0.301
7.0	502	7.075	3.665	0.518	0.596	10.542	0.230
7.3	356	4.944	2.922	0.591	0.599	10.489	0.440
7.7	260	3.523	2.247	0.638	0.595	10.558	0.311
8.0	174	2.286	1.466	0.641	0.592	10.607	0.374

[a]The model phase φ is defined such that time starts on 1 January 2000; thus the rate parameter would be $\lambda(t) = a + b\sin(\omega(t - 2000) + \varphi)$, where t is the year date of interest expressed as a real number.

of an event is independent of the occurrence of previous events [e.g., *Cox and Lewis*, 1966; *Blumer*, 1979; *Ross*, 2003]. The realization of a stationary Poisson process yields wait times τ between temporally ordered, statistical events having positive duration (0, ∞) and which can be described by exponential probability density function (pdf),

$$p_{\mathrm{e}}(\tau|\lambda) = \lambda \mathrm{e}^{-\lambda\tau}, \tag{1}$$

where λ is a constant average occurrence rate. From this density function we determine the wait time mean,

$$\overline{\tau} = \int_0^\infty \tau p_{\mathrm{e}}(\tau|\lambda)d\tau = \frac{1}{\lambda}, \tag{2}$$

as well as the variance of the wait times about the mean,

$$\sigma_\tau^2 = \int_0^\infty [\tau - \overline{\tau}]^2 p_{\mathrm{e}}(\tau|\lambda, c)d\tau = \frac{1}{\lambda^2}. \tag{3}$$

Correspondingly, the Poisson density function can be used to quantify the number of events n occurring with a time window t,

$$p_{\mathrm{p}}(\mathrm{n}|\lambda\mathrm{t}) = \frac{(\lambda\mathrm{t})^{\mathrm{n}}\mathrm{e}^{-\lambda\mathrm{t}}}{\mathrm{n}!}. \tag{4}$$

From this we determine the occurrence rate mean,

$$\overline{\mathrm{n}} = \sum_{\mathrm{n}=0}^{\infty} \mathrm{n}\frac{(\lambda\mathrm{t})^{\mathrm{n}}\mathrm{e}^{-\lambda\mathrm{t}}}{\mathrm{n}!} = \lambda\mathrm{t}, \tag{5}$$

as well as the variance of the occurrence rate,

$$\sigma_\tau^2 = \sum_{\mathrm{n}=0}^{\infty} [\mathrm{n} - \overline{\mathrm{n}}]^2 \frac{(\lambda\mathrm{t})^{\mathrm{n}}\mathrm{e}^{-\lambda\mathrm{t}}}{\mathrm{n}!} = \lambda\mathrm{t}. \tag{6}$$

As a stochastic model the Poisson process has found numerous practical applications, including in the analysis of telephone switchboard traffic, where each incoming telephone call is considered to be an independent statistical event, and in the analysis of radioactive decay, where each emission of a photon or particle from the radioactive source material is considered to be an independent statistical event.

[9] In our case, where we seek to use a winnowed *Kp* data to describe the solar-cycle-modulated statistics of magnetic storm occurrences, the idealized Poisson-process model must be slightly modified. For now, let us put aside the issue of nonstationarity, and consider first the affects of winnowing on the statistical analysis. Assuming that events are well described by a winnowed data set $\mathrm{D}^{\geq c}$, then the wait times between events are defined over a truncated domain (c, ∞), and the pdf of the wait times has the form

$$p_{\mathrm{e}}(\tau|\lambda, c) = \lambda \mathrm{e}^{\lambda c}\mathrm{e}^{-\lambda\tau}, \tag{7}$$

which, we note, is properly normalized,

$$\int_c^\infty p_{\mathrm{e}}(\tau|\lambda, c)d\tau = 1. \tag{8}$$

With (7) we easily calculate the winnowed data (interstorm), wait time mean,

$$\overline{\tau}(c) = \int_c^\infty \tau p_{\mathrm{e}}(\tau|\lambda, c)d\tau \simeq \frac{\mathrm{e}^{\lambda c}}{\lambda}. \tag{9}$$

Note that

$$\overline{\tau}(c) \to \overline{\tau} \text{ as } c \to 0; \tag{10}$$

in other words, the winnowed data, wait time mean approaches the Poisson wait time mean as the level of truncation (c) is reduced. This highlights an important point: we should be careful in interpreting λ. Independent

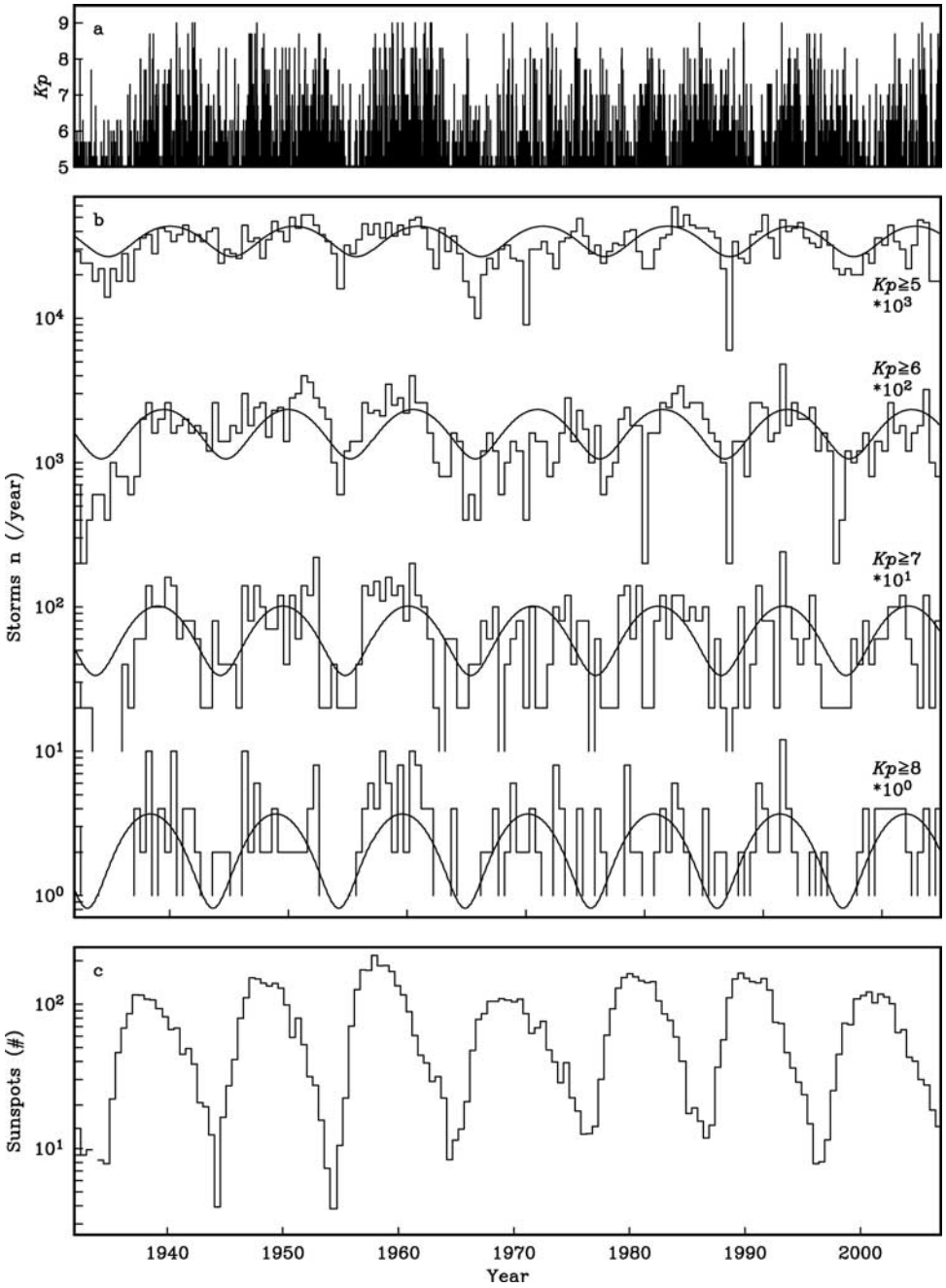

Figure 2. The 75-a time series of data used in this analysis: (a) raw, unbinned, and unwinnowed, $Kp \geq 5$ values, (b) binned occurrence rates (storms per year) as determined for the exceedances $Kp \geq 5$, $Kp \geq 6$, $Kp \geq 7$, $Kp \geq 8$, and the model occurrence rate mean, equation (22); a vertical axis multiplicative separation factor has been applied for presentation clarity, (c) average sunspot numbers. The solid oscillatory line shown for each Kp exceedance corresponds to the model occurrence rate mean, equation (22).

of the actual data being used λ describes the mean occurrence rate for Poisson events, equation (2). Since we have chosen to exclude short wait times, λ is not, strictly speaking, the occurrence rate for our winnowed data, equation (9), although the distinction disappears as the number of excluded events approaches zero. The wait time variance for the winnowed data is easily calculated,

$$\sigma_\tau^2(c) = \int_c^\infty [\tau - \overline{\tau}]^2 p_{\mathrm{e}}(\tau|\lambda, c)d\tau = \frac{1}{\lambda^2}, \tag{11}$$

which we note is identical to the Poisson result (3). The truncated exponential density function (7) has found application, for example, in the analysis of the waiting times for pedestrians wanting to cross a road having a busy flow of automobile traffic, and where the spacing between cars is exponentially distributed; the pedestrians will only cross when there is a minimum spacing (c) between cars [e.g., *Feller*, 1971, p. 378]. In a geophysical context, the truncated exponential density function has been used in the analysis of earthquake occurrence wait times, but where the occurrence times of numerous earthquake aftershocks have been removed since they are not statistically independent of the primary earthquake [e.g., *Kijko and Sellevoll*, 1992].

[10] In contrast to the situation for wait times of winnowed data, the corresponding statistical description of occurrence rate is comparatively more difficult. This is due to the nature of the winnowing itself. Removing (τ_i, t_i) for $\tau_i < c$ has an obvious affect on the distribution of the wait times, but the affect on the accompanying event times t_j is much less straightforward. Indeed, the mathematical expression for what might be called the modified Poisson density function, $p_{\mathrm{p}}(\mathrm{n}|\lambda, \mathrm{t}, c)$, and which corresponds rigorously to the truncated exponential density function $p_{\mathrm{e}}(\tau|\lambda, c)$, is exceedingly complicated; see, for example, *Feller* [1971, p. 469, equation (2.6)] where a non-closed-form expression is given. Therefore we have not made, in the manner of equations (5) and (6), a straightforward calculation of occurrence rate mean and variance for a winnowed data set $\mathrm{D}^{\geq c}$. Fortunately, for the case of the occurrence rate mean the mathematical difficulties are easily circumvented. Consider, for a moment, the proportion of all Poisson events which are retained after winnowing, that is, those for which $\tau \geq c$. This proportion corresponds to a partial cumulative integration over the untruncated density function (1),

$$\int_c^\infty p_{\mathrm{e}}(\tau|\lambda)d\tau = \mathrm{e}^{-\lambda c}. \tag{12}$$

With this and equation (5) we obtain the occurrence rate mean for a winnowed data set,

$$\overline{\mathrm{n}}(c) = \mathrm{e}^{-\lambda c}\lambda \mathrm{t}. \tag{13}$$

We note that

$$\overline{\mathrm{n}}(c) \to \overline{\mathrm{n}} \text{ as } c \to 0; \tag{14}$$

the winnowed data, occurrence rate mean approaches the Poisson, occurrence rate mean as the level of truncation (c) is reduced. Unfortunately, we have not been able to derive an expression for occurrence rate variance. However, on the basis of the basis of the results (9), (11), and (13), we speculate that winnowed data, occurrence rate variance would be of the form

$$\sigma_{\mathrm{n}}^2(c) \simeq \mathrm{e}^{m\lambda c}\lambda \mathrm{t}, \tag{15}$$

where m is an integer; we will not pursue the determination of an explicit expression for $\sigma_{\mathrm{n}}^2(c)$ any further.

[11] Turning now to the matter of nonstationarity, for all but the very largest of magnetic storms the wait time mean $\overline{\tau}$ is relatively short compared to the period of the solar cycle T_s. Therefore when considering a window of time t such that $c \ll \overline{\tau} \ll \mathrm{t} \ll T_s$, the statistics of storm occurrence is only weakly nonstationary. Storm statistics can be described in terms a nonstationary Poisson rate function $\lambda(t)$ whose time-dependence is secular. With this we have, approximately, the nonstationary pdf for wait times between storm occurrences,

$$p_{\mathrm{e}}(\tau_i|\lambda(t), c) \simeq \lambda(t_i)\mathrm{e}^{\lambda c}\exp\{-\lambda(t_i)\tau_i\}. \tag{16}$$

Alternatively, we could, with an equal degree of accuracy, use

$$p_{\mathrm{e}}(\tau_i|\lambda(t), c) \simeq \lambda(t_i + \tau_i)\mathrm{e}^{\lambda c}\exp\{-\lambda(t_i + \tau_i)\tau_i\}. \tag{17}$$

It is reasonable to use an average compromise between these two expressions,

$$p_{\mathrm{e}}(\tau_i|\lambda(t), c) \simeq \frac{1}{2}\lambda(t_i)\mathrm{e}^{\lambda c}\exp\{-\lambda(t_i)\tau_i\} + \frac{1}{2}\lambda(t_i + \tau_i)\mathrm{e}^{\lambda c}\exp\{-\lambda(t_i + \tau_i)\tau_i\}. \tag{18}$$

More accurate and more complicated representations of the nonstationary (nonhomogeneous) occurrence statistics can be obtained by using integrations over time of the rate function [e.g., *Cox and Lewis*, 1966; *Ross*, 2003]. For our purposes, the relative simplicity of (18) is appealing and it is sufficiently accurate for the statistical analysis we conduct here.

[12] We choose to test whether or not the solar cycle modulation of storm likelihood can be modeled by the simple time-dependent Poisson rate function:

$$\lambda(t) = a + b\sin(\omega t + \varphi), \tag{19}$$

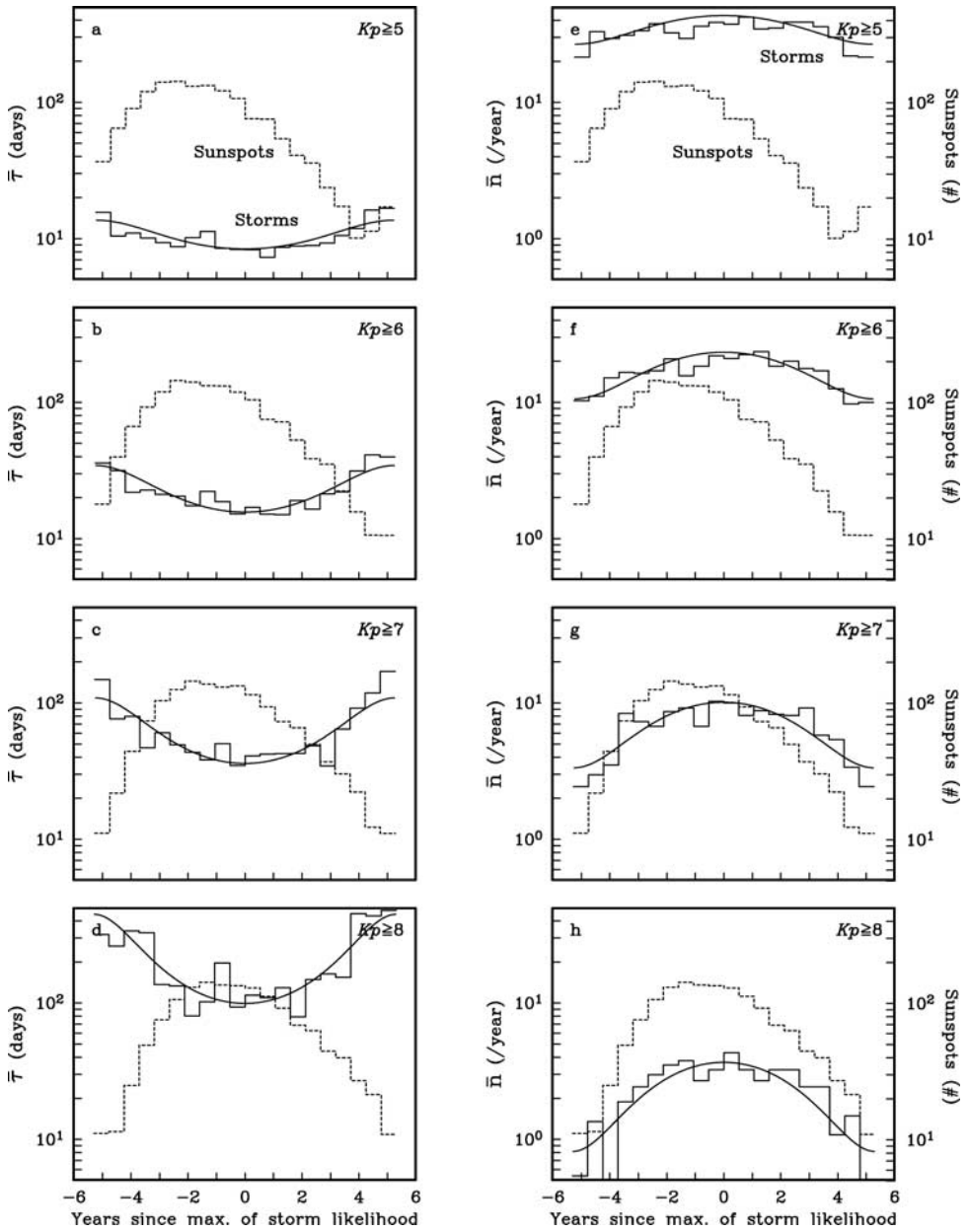

Figure 3. One complete model cycle of the cyclically averaged, phase-binned wait time mean $M_k(\tau)$, equation (24), and occurrence rate mean $M_k(\mathrm{n})$, equation (25), each shown as solid line histograms, centered on the maximum of storm likelihood as determined by the statistical model, for exceedances of (a, e) $Kp \geq 5$, (b, f) $Kp \geq 6$, (c, g) $Kp \geq 7$, (d, h) $Kp \geq 8$. The smooth lines show the model wait time mean $\overline{\tau}(t, c)$, equation (21), and occurrence rate mean $\overline{n}(t, c)$, equation (22). Also shown as dotted line histograms in each window are the average sunspot numbers.

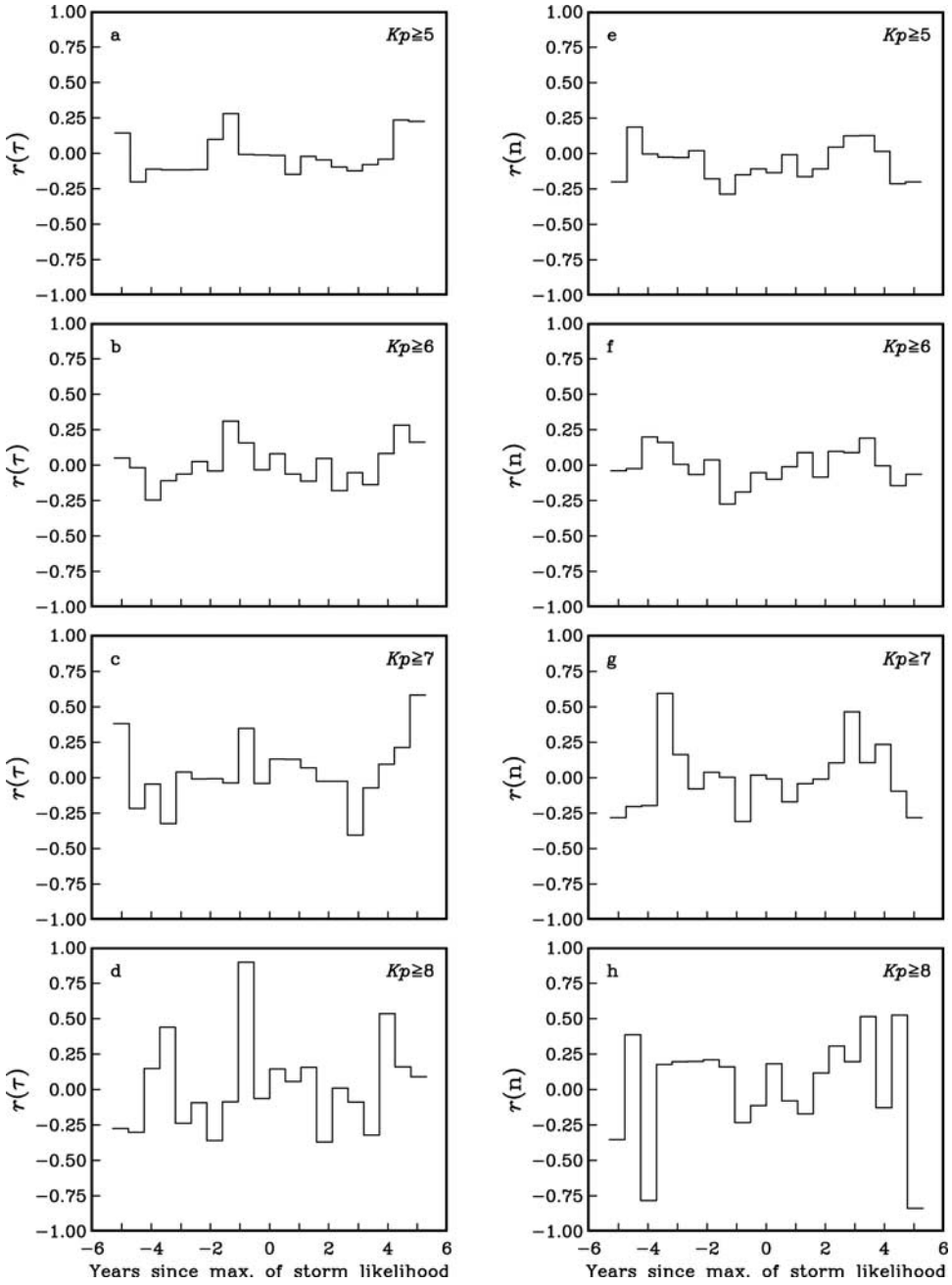

Figure 4. Binned absolute wait time errors $e_k(\tau)$ and occurrence rate errors e_k(n), equations (26) and (27), plotted as histograms over a complete model cycle for exceedances (a, e) $Kp \geq 5$, (b, f) $Kp \geq 6$, (c, g) $Kp \geq 7$, (d, h) $Kp \geq 8$.

with the parameter a being a constant, b the modulation amplitude, ω the solar cycle frequency, and φ the phase. By making this parameterization we are assuming that recent solar cycles are well characterized as periodic. Of course, this is only an approximation, one meant to summarize the majority of the variability of magnetic activity over the course of the past 75 a. Still, we acknowledge the obvious: Kp has exhibited noticeable differences from one solar cycle to another; see Figure 2. Indeed, the waxing and waning of sunspot number, which represents a proxy measure for solar activity that can initiate magnetic storms, also deviates from simple periodicity. These days, complicated programs are often used to try to predict exactly how active the coming solar

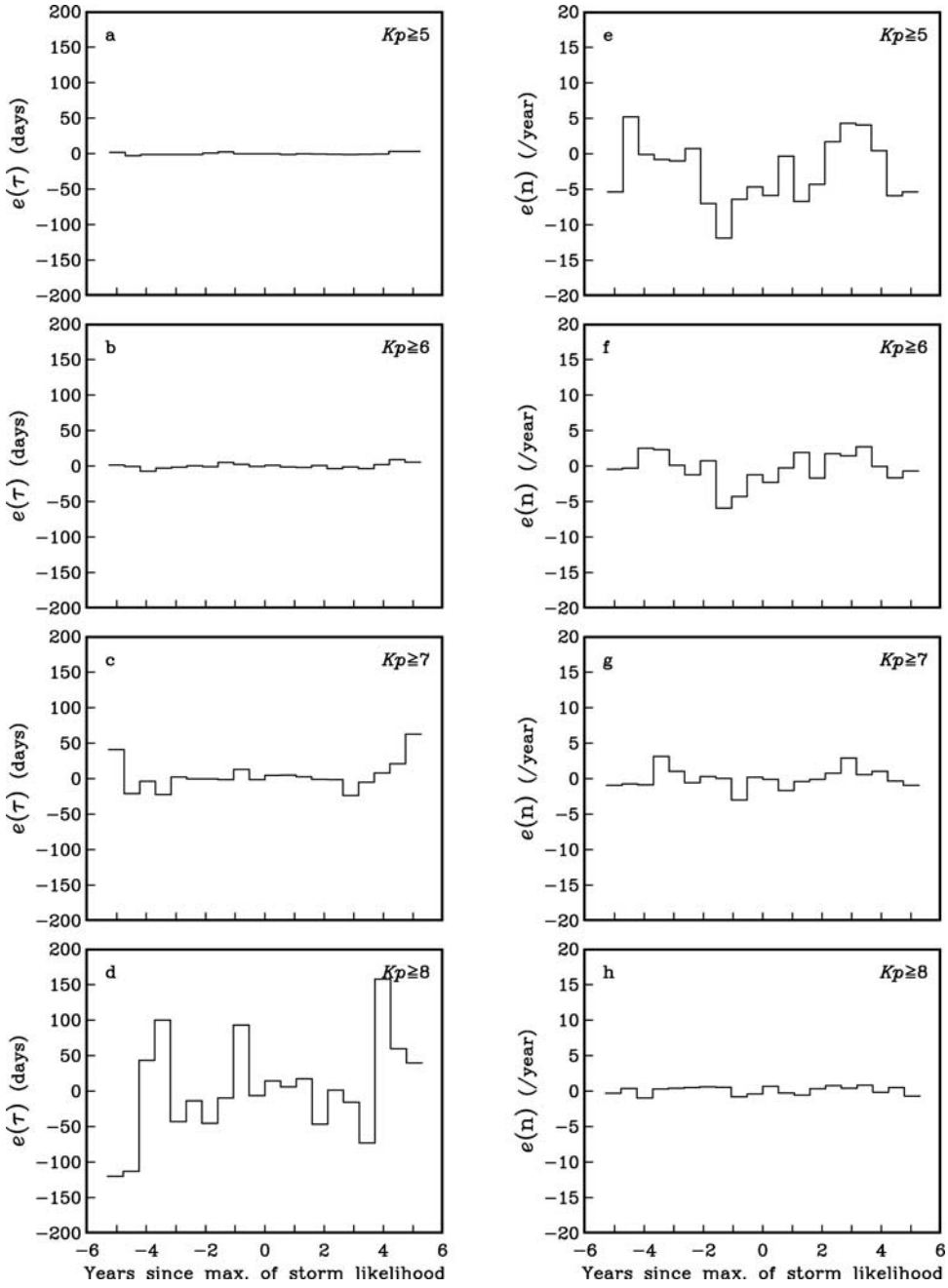

Figure 5. Binned relative wait time errors $r_k(\tau)$ and occurrence rate errors r_k(n), equations (28) and (29) plotted as histograms over a complete model cycle for exceedances (a, e) $Kp \geq 5$, (b, f) $Kp \geq 6$, (c, g) $Kp \geq 7$, (d, h) $Kp \geq 8$.

Table 2. Error Analysis Summary for Different *Kp* Exceedance Thresholds

$Kp\geq$	$\bar{e}(\tau)$ (days)	$R(e(\tau))$ (days)	$\bar{r}(\tau)$	$R(r(\tau))$	$\bar{e}$(n) (per year)	$R(e$(n)) (per year)	$\bar{r}$(n)	$R(r$(n))	$\bar{p}_{KS}$
5.0	−0.099	1.561	−0.013	0.135	−2.458	5.046	−0.064	0.141	0.026
5.3	0.051	1.935	0.001	0.137	−1.364	4.171	−0.043	0.140	0.110
5.7	0.371	3.471	0.015	0.177	−0.667	3.459	−0.023	0.162	0.007
6.0	0.255	3.552	0.008	0.141	−0.315	2.195	−0.009	0.121	0.009
6.3	1.112	6.162	0.022	0.156	−0.128	1.856	−0.006	0.154	0.015
6.7	1.967	11.304	0.031	0.199	−0.041	1.728	−0.002	0.203	0.068
7.0	3.918	19.888	0.039	0.227	0.012	1.361	0.003	0.230	0.087
7.3	4.396	33.429	0.023	0.255	0.021	1.198	0.014	0.233	0.037
7.7	6.876	48.921	0.030	0.306	0.071	1.034	0.021	0.288	0.187
8.0	2.359	67.185	0.022	0.317	0.100	0.561	0.024	0.361	0.284

cycle will be [e.g., *Svalgaard et al.*, 2005; *Dikpati et al.*, 2006]. For a tractable statistical analysis some degree of stationarity is a practical choice. With (19) we are allowing for an extremely simple type of nonstationarity (sinusoidal modulation). The parameters describing rate function modulation are themselves assumed to be stationary; (a, b, ω, φ) do not change over the entire 75-a time span of the *Kp* data. This might seem to be a stringent set of assumptions, but the success (or failure) of our model can be judged, in retrospect, once the data have been fitted.

[13] Our approach is to be contrasted with that of *Wheatland* [2000], who analyzed the nonstationary statistics of solar flares occurrences. He accommodated modulation by dividing the solar cycle into multiple short intervals, and then performing independent parameter estimations using data within each of the intervals. Wheatland's method works well when many data are available within each interval. On the other hand, our continuous model allows us to analyze the relatively fewer number of data recording large magnetic storms. Moreover, by assuming that the modulated rate function is described by a stationary set of parameters we can make time-dependent predictions.

[14] Model parameters are estimated, and the wait time data are fitted, by using the pdf (18) to construct a likelihood function,

$$\mathcal{L} = \frac{1}{N}\prod_{i=1}^{N} p_{\mathrm{e}}(\tau_i|\lambda(t), c), \tag{20}$$

where N is the number of winnowed wait times in $\mathrm{D}^{\geq c}$. $\mathcal{L}$ is maximized by conducting a numerical search over parameter space (a, b, ω, φ) using a simplex computer routine (''amoeba'' [*Press et al.*, 1996, p. 404]). Comparisons of the data with model wait time mean (9) and number (13) can be accomplished using the time-dependent Poisson rate $\lambda(t)$ in place of the stationary Poisson rate λ, which is reasonable assuming, once again, that events occur relatively frequently compared to the timescale of the solar cycle.

4. Results

[15] Numerical values of the model parameters, the result of maximum likelihood estimation, are given in Table 1 for different *Kp* exceedance thresholds. With these, the stationary mean wait time (9), and the modulated Poisson rate function (19), we calculate the model nonstationary, storm wait time mean

$$\overline{\tau}(t, c) \simeq \frac{\mathrm{e}^{\lambda(t)c}}{\lambda(t)}. \tag{21}$$

With the stationary occurrence rate mean (13), we calculate the model nonstationary, occurrence rate mean

$$\overline{\mathrm{n}}(t, c) \simeq \mathrm{e}^{-\lambda(t)c}\lambda(t)\mathrm{t}. \tag{22}$$

Both of these quantities can be compared, now, with the data $\mathrm{D}^{\geq c}$, which we accomplished by averaging the data over multiple solar cycles. We count the number of magnetic storms that fall into phase bins defined by ω and φ. Specifically, we divide the period

$$T_s = \frac{2\pi}{\omega} \tag{23}$$

into a set of N_B bins. (Here we choose $N_B = 20$ so that each bin is slightly more than a half a year in width. It is important to recognize, however, that we are using the frequency ω that comes from the maximum likelihood estimation to define the bin width. The estimation itself is independent of any binning or averaging of the data, and we have not forced any preconceived period or frequency onto the estimation or onto the display of the results.) Then, we accumulate interstorm wait times and storm numbers within each bin according to the phase of the storm time, mod(ωt_i, 2π), with storm time t being in the

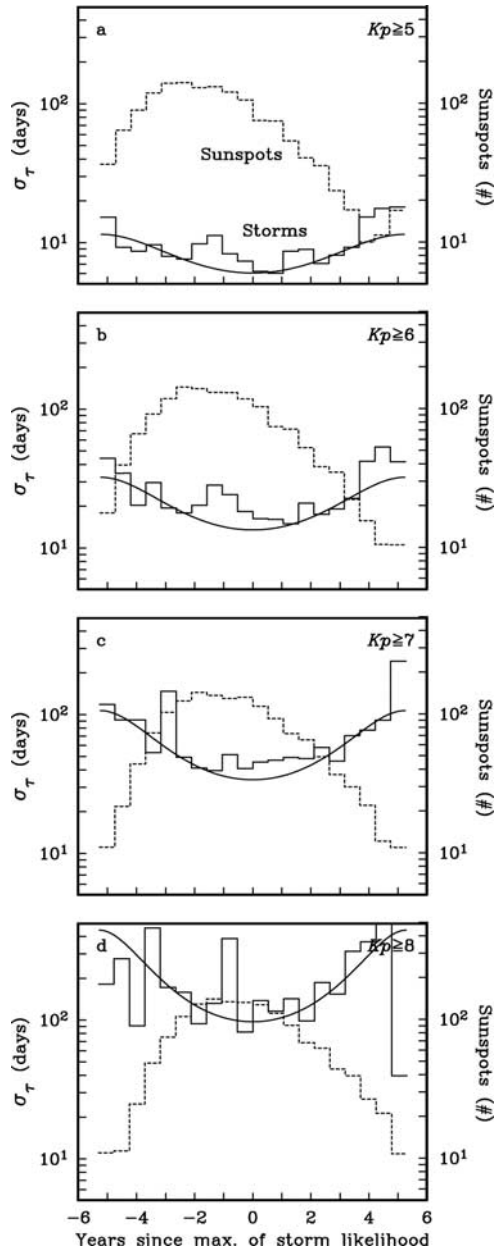

Figure 6. One complete model cycle of the cyclically averaged, phase-binned wait time standard deviation $S_k(\tau)$, from equation (33), shown as solid line histograms, centered on the maximum of storm likelihood as determined by the statistical model, for exceedances of (a) $Kp \geq 5$, (b) $Kp \geq 6$, (c) $Kp \geq 7$, (d) $Kp \geq 8$. The smooth lines show the model wait time standard deviation $\sigma\tau(t, c)$, from equation (32). Also shown as dotted-line histograms in each window are the average sunspot numbers.

same bin as $t + mT_s$, where m is any integer. For the k^{th} bin the storm wait time mean is

$$M_k(\tau) = \frac{1}{N_k} \sum_j \tau_j, \tag{24}$$

where it is understood that the summation is over the N_k wait times within the k^{th} bin. We estimate the storm occurrence rate directly from the data by counting the number of storms within each bin. For the k^{th} bin the storm occurrence rate is

$$M_k(\mathbf{n}) = \frac{N_k N_B}{T_{Kp}}, \tag{25}$$

where T_{Kp} is the total time span of the Kp time series (75 a). In Figure 3 we show, for different Kp exceedances, histograms of the cyclically averaged storm wait time mean and occurrence rate means within each bin, together with the model wait time mean (21) and occurrence rate (22). Here we also show the average sunspot number, binned according to the model parameters ω and φ As expected, storm occurrence rate decreases (increases) as the Kp threshold is increased (decreased); large storms are rarer than small storms, something that is perfectly obvious.

[16] Figure 3 also shows some additional important, and less obvious, relationships. Note, for example, that the modulation for $Kp \geq 5$ (Figures 3a and 3e), which includes many small storms, is relatively flat, b/a is small, and that the maximum likelihood for storm occurrence clearly occurs on the declining phase of the solar cycle. On the other hand, with increasingly large Kp exceedance thresholds the relative amplitude of the modulation b/a increases, and the phase φ shifts so that for the largest storms, $Kp \geq 8$ (Figures 3d and 3h), the likelihood of storm occurrence is more tightly correlated with sunspot number. The amount of time that the maximum of storm likelihood occurrence lags the sunspot maximum decreases from about 3 a for $Kp \geq 5$ to about 1 a for $Kp \geq 8$. This quantitative result is consistent with the qualitative observations made in the past.

[17] We now consider the errors associated with comparisons of our model and the mean quantities of the data $D^{\geq c}$. For each bin k having a center time of β_k we calculate the absolute residual errors,

$$e_k(\tau) = M_k(\tau) - \overline{\tau}(\beta_k, c), \tag{26}$$

$$e_k(\mathbf{n}) = M_k(\mathbf{n}) - \overline{\mathbf{n}}(\beta_k, c), \tag{27}$$

and the corresponding relative residual errors,

$$r_k(\tau) = \frac{e_k(\tau)}{\overline{\tau}(\beta_k, c)}, \tag{28}$$

$$r_k(\mathbf{n}) = \frac{e_k(\mathbf{n})}{\overline{\mathbf{n}}(\beta_k, c)}. \tag{29}$$

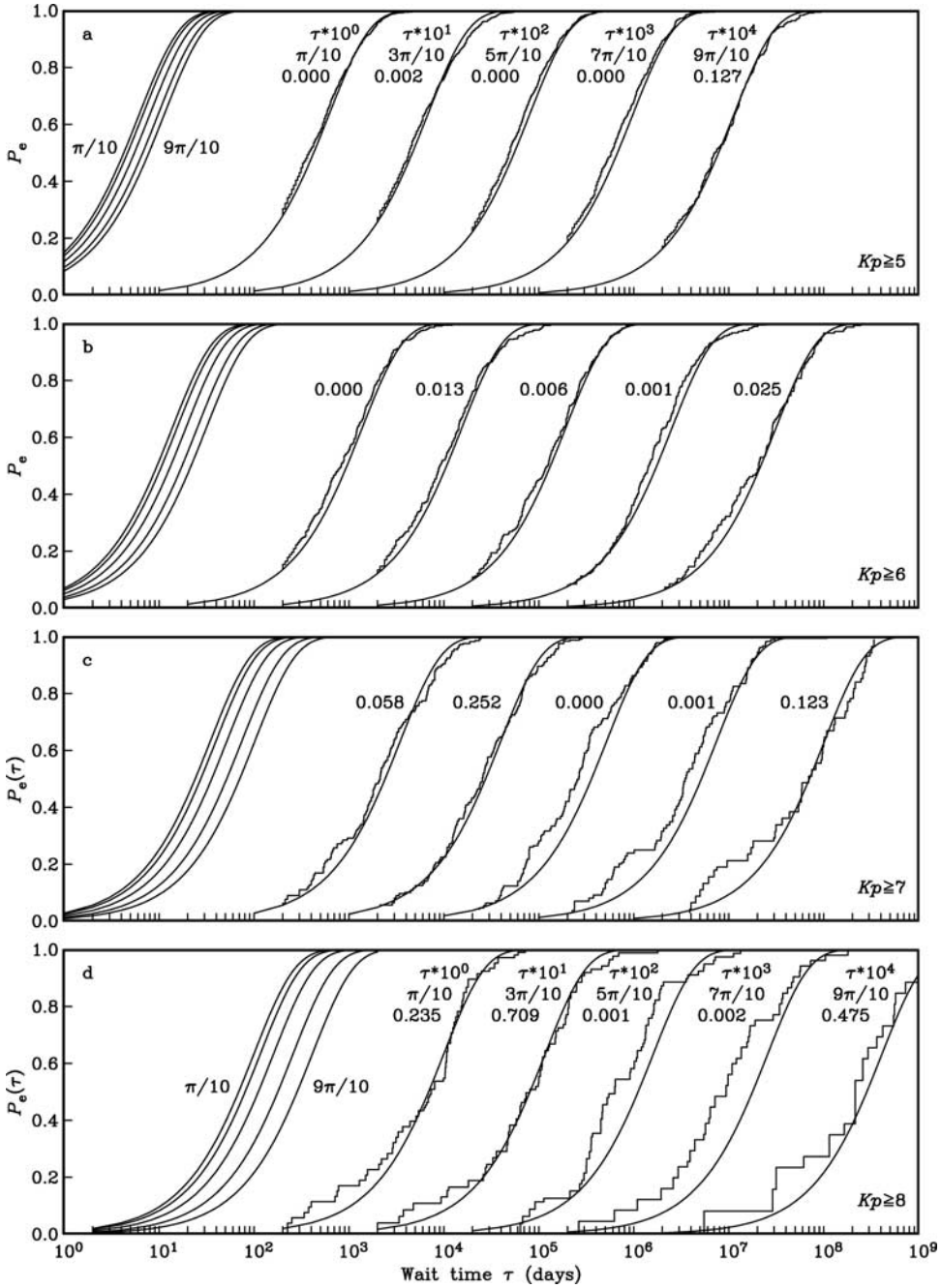

Figure 7. Cdfs of wait time data and the corresponding exponential distributions for data, binned according to their phase within the solar cycle ($\frac{\pi}{10}$, $\frac{3\pi}{10}$, $\frac{5\pi}{10}$, $\frac{7\pi}{10}$, $\frac{9\pi}{10}$), for exceedances (a) $Kp \geq 5$, (b) $Kp \geq 6$, (c) $Kp \geq 7$, (d) $Kp \geq 8$. The smooth lines show the model cdf $P_e(\tau|\lambda(t))$, equation (34). A horizontal axis multiplicative separation factor has been applied for presentation clarity, but on the far left we show, for easy comparison, the cdfs without the application of the separation factor. For each of the five phase bins we have also given the Kolmogorov-Smirnov probability p_{KS}, which we note is greatest for the exceedance (d) $Kp \geq 8$.

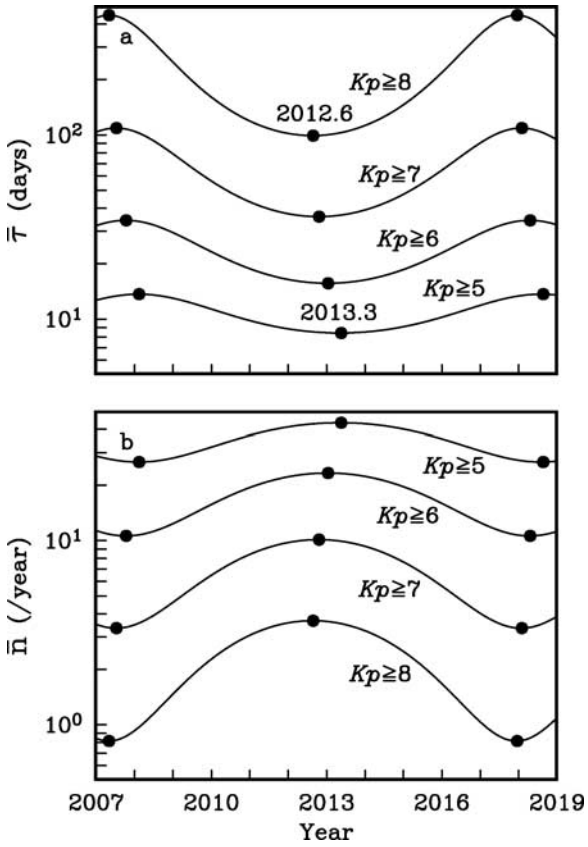

Figure 8. Predictions over the course of the coming solar cycle for (a) the magnetic storm, wait time mean $\overline{\tau}(t, c)$, equation (21), and (b) magnetic storm, occurrence rate mean $\overline{n}(t, c)$, equation (22) for different *Kp* exceedance thresholds.

These quantities are plotted as cyclically averaged histograms in Figures 4 and 5, where we see that both e_k and r_k do not show significant systematic bias or unmodeled coherent variation over the course of a solar cycle modulation. Therefore we conclude that our statistical model is a relatively complete description of the mean data. In Table 2 we give the averages of the residual errors taken over the N_B bins,

$$\left\{ \begin{matrix} \bar{e} \\ \bar{r} \end{matrix} \right\} = \frac{1}{N_B} \sum_k \left\{ \begin{matrix} e_k \\ r_k \end{matrix} \right\}. \tag{30}$$

Since the $\bar{e}$ are much smaller than (say) the rate function parameter a, and since the $\bar{r}$ are much smaller than 1, we conclude that our statistical model is a relatively unbiased description of the mean data. The root-mean-square values of the residual errors,

$$\mathrm{R}\left\{ \begin{matrix} e \\ r \end{matrix} \right\} = \left[\frac{1}{N_B} \sum_k \left\{ \begin{matrix} e_k^2 \\ r_k^2 \end{matrix} \right\} \right]^{\frac{1}{2}}, \tag{31}$$

are given in Table 2. Note that the average absolute error in the rate function, R(e), decreases (increases) as we consider larger (smaller) *Kp* exceedance thresholds. This is expected since there are, for example, many more magnetic storms having $Kp \geq 5$ than $Kp \geq 8$. On the other hand, the average relative error shows the opposite functional relationship, with R(r) increasing (decreasing) as we consider larger (smaller) *Kp* exceedance thresholds. The large number of magnetic storms having $Kp \geq 5$ gives associated averages for each bin which display less statistical jitter than, for example, the smaller number of storms for $Kp \geq 8$. We conclude that our statistical model is a relatively accurate description of the mean data.

[18] Next, we consider the variance of wait times of $\mathrm{D}^{\geq c}$. With the parameters given in Table 1, the stationary value for the variance of the wait times (11), and the modulated Poisson rate function (19), we calculate the model nonstationary, wait time variance

$$\sigma_\tau^2(t, c) \simeq \frac{1}{\lambda^2(t)}. \tag{32}$$

We estimate the storm wait time variance by accumulating, as before, data within bins and then calculating the quantity

$$S_k^2(\tau) = \frac{1}{N_k} \sum_j \left[\tau_j - \overline{\tau}(\beta_k, c) \right]^2, \tag{33}$$

where, once again, β_k is the center time of bin k and the summation is over the N_k storms within the k^{th} bin. Results are shown in Figure 6 for different *Kp* exceedances. The model wait time standard deviation $\sigma_\tau(t, c)$ fairly successfully fits the wait time standard deviation of the data $S_k(\tau)$, but we acknowledge a slight underestimation of the model standard deviation.

[19] As the final part of the presentation of our results, we compare the shape of the wait time data distribution with that of our model. This is comparatively straightforward for the case of stationary statistics. Indeed, *Remick and Love* [2006] did just that, plotting the stationary cumulative exponential distribution against the winnowed wait time data, but without considering the affects of solar cycle modulation. For the nonstationary, solar-cycle-modulated model considered here we compare the cumulative density functions (cdfs) of binned subsets of the wait time data with the modulated model cdfs. First we note that over the course of a modulation period T_s the rate function $\lambda(t)$ attains a maximum (minimum) value of $a + b$ ($a - b$), and values of $\lambda(t)$ intermediate to these extrema occur twice per cycle. Therefore in considering theoretical cdfs, which are a function of the rate function, it is sufficient to consider only half a period $T_s/2$; anything longer would be repetitious. We divide the half period into $N_B = 5$ bins, each denoted by a centered phase: $\frac{\pi}{10}$, $\frac{3\pi}{10}$, etc. We accu-

Table 3. Predictions for the Wait Time Mean $\overline{\tau}(t, c)$ and the Occurrence Rate Mean $\overline{n}(t, c)$ for Magnetic Storms Over the Course of the Coming Solar Cycle and for Different *Kp* Exceedance Thresholds

	$Kp \geq 5$		$Kp \geq 6$		$Kp \geq 7$		$Kp \geq 8$	
Year	$\overline{\tau}$ (days)	$\overline{n}$ (per year)	$\overline{\tau}$ (days)	$\overline{n}$ (per year)	$\overline{\tau}$ (days)	$\overline{n}$ (per year)	$\overline{\tau}$ (days)	$\overline{n}$ (per year)
2007	13	27	34	11	109	3	443	1
2008	14	27	32	11	93	4	317	1
2009	12	30	26	14	66	6	197	2
2010	11	35	21	18	49	7	136	3
2011	9	39	17	21	40	9	108	3
2012	9	43	16	23	36	10	99	4
2013	8	43	16	23	37	10	105	3
2014	9	42	17	21	43	9	126	3
2015	10	38	20	18	55	7	176	2
2016	11	33	26	14	77	5	278	1
2017	13	29	32	11	103	4	422	1
2018	14	27	34	11	105	3	408	1

mulate the wait times τ_j within these bins, assigning a wait time once to the bin for the first associated storm time t_j and once to the bin for the second associated storm time t_{i+1}, and, at the same time, taking account of the fact that storm time t would be in the same bin as $mT - t$, where m is any integer. We sort the wait times within each bin and plot them as a cumulative for comparison with the cdf of the nontruncated, exponential density function,

$$P_e(\tau|\lambda(t)) \simeq \lambda(t) \int_0^\tau e^{-\lambda(t)x} dx = 1 - e^{-\lambda(t)\tau}. \quad (34)$$

In Figure 7, we see that at some points in the solar cycle there are noticeable differences between the data and the model. Still, the overall agreement between the data and the model is reasonably good; most of the statistical signal in the wait time data is being modeled.

[20] For each *Kp* exceedance and for each bin we have calculated the Kolmogorov-Smirnov probability p_{KS} ("ksone" [*Press et al.*, 1996, p. 619]), giving the probability that the data could conceivably have been drawn from a population consistent with the model distribution. We present the individual values for p_{KS} in Figure 7, and in Table 2 we give the average probability $\overline{p}_{KS}$ for each *Kp* exceedance threshold. In all cases $\overline{p}_{KS}$ is relatively small ($\ll 1$). While we would obviously prefer values of $\overline{p}_{KS}$ that approach unity, for our purposes, the Kolmogorov-Smirnov test is overly stringent. Consider, for example, the case for $Kp \geq 8$, for which $\overline{p}_{KS} = 0.284$. This means that there is a 0.716 probability that the data could have been realized from a population given by a hypothetical statistical model that is different, in an unknown way, from our simple model. The Kolmogorov-Smirnov test does not provide any guidance for which of the alternative, better-fitting models, among the infinity that are available, should be used in place of ours. It seems reasonable to conclude that the data $D^{\geq c}$ are the realization of a statistical model that is only slightly different from our model. Although there is certainly room for improvement, since our statistical model is motivated more by simplicity and phenomenology than it is by physics, we consider the results presented here to be an indication of success.

5. Interpretation

[21] One of the most important conclusions of this study concerns the timing of small and large magnetic storm occurrence relative to the phase of the solar cycle. Why would the most likely occurrence of small (large) storms come during the declining phase (maximum) of the solar cycle? One possible explanation is the difference in heliophysical drivers of small versus large magnetic storms. Small storms are associated with both coronal mass ejections and high-speed streams from coronal holes, with high-speed streams tending to initiate storms during the declining phase of the solar cycle [*Echer et al.*, 2004, and references therein]. On the other hand, large storms are primarily associated with coronal mass ejections [*Richardson et al.*, 2001], and these occur most frequently at solar maximum.

6. Predictions

[22] Using the model wait time mean $\overline{\tau}(t, c)$, equation (21) and occurrence rate mean $\overline{n}(t, c)$, equation (22), we can predict the occurrence statistics of magnetic storms over the course of the coming solar cycle; results are shown in Figure 8 and specific numbers are given in Table 3. We offer these results with the hope that the space weather community can make practical use of the statistical model presented here. It appears that 2012 will be a big year for big magnetic storms.

[23] **Acknowledgments.** In addition to data collected by the United States (USGS) geomagnetic observatory program (observatories: FRD/CLH, SIT), the calculation of the *Kp* index relies on data from several other national observatory programs. We thank the observatory programs of Australia (CNB/TOO), Canada (MEA, OTT/AGN), Denmark (BFE/RSV), Germany (NGT/WIT, WNG), New Zealand (EYR/AML), Sweden (LOV/UPS), and the United Kingdom (ESK, HAD/ABN, LER) for the collection of the necessary magnetometer data. We thank the GeoForschungsZentrum in Potsdam for calculating the *Kp* index. The sunspot number data were obtained from the Royal Observatory of Belgium. We thank D. E. McNamara for his review of a draft manuscript, J. C. Green and L. Svalgaard for useful conversations, and J. N. Thomas for assistance with downloading the *Kp* data. We thank M. S. Wheatland and D. J. Thomson for their constructive comments. Support for K.J.R. was provided by the USGS Mendenhall Postdoctoral Research Fellowship Program.

References

Bartels, J. (1932), Terrestrial magnetic activity and its relations to solar phenomena, *J. Geophys. Res.*, *37*, 1–52.

Blumer, M. G. (1979), *Principles of Statistics*, Dover, New York.

Bothmer, V., and I. A. Daglis (Eds.) (2007), *Space Weather Physics and Effects*, Praxis, Chichester, U. K.

Carlowicz, M. J., and R. E. Lopez (2002), *Storms From the Sun: The Emerging Science of Space Weather*, Joseph Henry Press, Washington, D. C.

Chapman, S., and J. Bartels (1962), *Geomagnetism*, vol. 1, Oxford Univ. Press, London.

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

Dikpati, M., G. de Toma, and P. A. Gilman (2006), Predicting the strength of solar cycle 24 using a flux-transport dynamo-based tool, *Geophys. Res. Lett.*, *33*, L05102, doi:10.1029/2005GL025221.

Echer, E., W. D. Gonzalez, A. L. C. Gonzalez, A. Prestes, L. E. A. Vieira, A. Dal Lago, F. L. Guarnieri, and N. J. Schuch (2004), Long-term correlation between solar and geomagnetic activity, *J. Atmos. Sol. Terr. Phys.*, *66*, 1019–1025.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley, New York.

Feynman, J., and X. Y. Gu (1986), Prediction of geomagnetic activity on timescales of one to ten years, *Rev. Geophys.*, *24*, 650–666.

Garrett, H. B., A. J. Dessler, and T. W. Hill (1974), Influence of solar variability on geomagnetic activity, *J. Geophys. Res.*, *79*, 4603–4610.

Gorney, D. J. (1990), Solar cycle effects on the near-earth space environment, *Rev. Geophys.*, *28*, 315–336.

Kijko, A., and M. A. Sellevoll (1992), Estimation of earthquake hazard parameters from incomplete data files. Part II. Incorporation of magnitude heterogeneity, *Bull. Seismol. Soc. Am.*, *82*, 120–134.

Lanzerotti, L. J. (2007), Editorial: Predictions of intense solar activity needed, *Space Weather*, *5*, S03001, doi:10.1029/2007SW000319.

Love, J. J., and K. J. Remick (2007), Magnetic indices, in *Encyclopedia of Geomagnetism and Paleomagnetism*, edited by D. Gubbins and E. Herrero-Bervera, pp. 509–512, Springer, Dordrecht, Netherlands.

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

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

NOAA Service Assessment Team (2004), *Intense Space Weather Storms October 19-November 07, 2003*, U. S. Dept. of Commerce, Silver Spring, Md.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1996), *Numerical Recipes*, Cambridge Univ. Press, Cambridge, UK.

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

Richardson, I. G., E. W. Cliver, and H. V. Cane (2001), Sources of geomagnetic storms for solar minimum and maximum conditions during 1972–2000, *Geophys. Res. Lett.*, *28*, 2569–2572.

Ross, S. M. (2003), *Introduction to Probability Models*, Academic, San Diego, Calif.

Svalgaard, L., E. W. Cliver, and Y. Kamide (2005), Sunspot cycle 24: Smallest in 100 years?, *Geophys. Res. Lett.*, *32*, L01104, doi:10.1029/2004GL021664.

Thomson, A. W. P. (2007), *Geomagnetic hazards, in Encyclopedia of Geomagnetism and Paleomagnetism*, edited by D. Gubbins and E. Herrero-Bervera, pp. 316–318, Springer, Dordrecht, Netherlands.

Wheatland, M. S. (2000), The origin of the solar flare waiting-time distribution, *Astro. Phys. J.*, *536*, L109–L112.

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

D. M. Perkins, Earthquake Hazards Program, U. S. Geological Survey, Denver, CO 80225, USA.