Analysis of "On the probability of occurrence of extreme space weather events", by P. Riley[1]

analysis by Stephen Parrott

September 4, 2014: Minor numerical errors corrected. Analysis and conclusions unaffected. ©September 4, 2014.

January 14, 2015: Some of Riley's statistics are ambiguously reported in his paper. Since the author has ignored all requests for clarification, I had to use my best guess to resolve the ambiguities. New evidence since the present paper was posted has caused me to change my best guess. This changes most of the probability estimates reported below.

This new evidence together with revised probability estimates is reported in a reanalysis of Riley's paper in file www.math.umb.edu/∼sp/2ndlook.pdf. It is discussed and can be accessed from the January 14, 2015 entry of the "papers" page on my website www.math.umb.edu/∼ sp .

1 Introduction

This essay analyzes the paper "On the probability of occurrence of extreme space weather events" by Pete Riley [1]. The paper concludes:

> "In this study we have applied a power law probabilistic analysis to assess the likelihood of a space weather event on the scale of, or larger than the Carrington event of 1859. . . . we inferred a probability of $\sim 12\%$ that an event [of that magnitude] would occur over the next decade."

A recent article on the National Aeronautics and Space Administration's (NASA) website,[2] is largely devoted to Riley's article. It gives the strong impression that Riley's 12% estimate can be taken as scientifically determined and reliable. It quotes Riley as saying:

> "Initially, I was quite surprised that the odds were so high, but the statistics appear to be correct," says Riley. "It is a sobering figure."

This gives the impression that from Riley's data, any statistician would arrive at Riley's 12% estimate. This essay will express doubt.

[1] P. Riley, *Space Weather* **10** (2012), S02012

[2] http://science.nasa.gov/science-news/science-at-nasa/2014/23jul_superstorm/

In so arguing, I will have to concentrate on the questionable aspects of Riley's paper. This will give a negative tone to the discssion, which I regret. The problem is that I don't know enough about space weather to highlight what are probably many valuable insights in Riley's paper. As a mathematician, my main reservations concern the paper's mathematics and its application.

There is another reason to focus on the paper's shortcomings more than is usual in scientific writing. Riley's is not an ordinary scientific paper, discussing obscure ideas which are meaningful only to a handful of devotees of some sub-subfield. If his estimate of a probability of 12% of a Carrington-class event in the next decade is taken seriously, an expenditure of hundreds of *billions* of dollars to prepare could be justified. (A study by the National Academy of Science [2] estimated damages from such an event to be in the trillions of dollars.) If there is any uncertainty in the science underlying the estimate or in the execution of that science, it deserves careful scrutiny. I believe that there is such uncertainty.

2 Conclusions

Riley's paper is technical, and the analysis to follow will necessarily be technical. Few will read it in detail. Recognizing this, I state the conclusions first.

Mathematics. The beginning part of Riley's analysis is mathematically incorrect. Althogh the error is easily fixed, it fosters a misunderstanding which may have led to a later incorrect application of the mathematics.

An initial estimate of probability 85% of an event worse than Carrington in the next decade was rejected by the paper as "not credible". This estimate may have been based on an incorrect application of the paper's mathematics—the paper does not contain enough details to be sure. Subsequent estimates may (or may not) have been based on correct applications, but the paper's presentation is again too incomplete to be sure.

Questionable assumption. The paper's analysis is based on an assumption that the data which it analyzes follows a so-called "power law". The paper presents no real evidence for this assumption. The assumption appears to be false for at least one data set which it analyzes, and questionable for several.

Possible errors. Probability estimates which I obtained under the paper's assumptions sometimes differed substantially from those reported in the paper. In consequence, I believe that the numerical results reported by the paper should be independently verified before being used in policy decisions.

3 Overview of Riley's paper

Disturbances on the sun can result in "magnetic storms" on the earth which induce strong currents in power lines and other electromagnetic equipment.

These strong currents can disable power grids and other equipment.

The strongest storm in recorded history occurred in 1859 and is known as the "Carrington event" after the astronomer who reported it. It is said to have induced sparking in telegraph lines strong enough to have caused fires.

Such an event today could conceivably disable large portions of the world's power grids for months or years. A National Academy of Science report [2] estimates the potential damage to the United States in the trillions of dollars.

The importance is obvious, but preparations to alleviate the damage are expensive. The laudable objective of Riley's paper is to produce estimates, or at least educated guesses, as to the probability of a Carrington-class or worse event in the next decade. The paper's best guess is 12%, presented with numerous caveats. The popular press has seized on this estimate as one with firm scientific basis, and it has been widely cited without the caveats.

There are various measures of the strength of a "solar storm". The paper considers four measures. For one of these, X-ray fluxes (section 3.2 of [1]), the size of the Carrington event is not known, so (as the paper notes on its p. 6), meaningful probability estimates cannot be obtained. For that reason, discussion of this measure is omitted below.

The other three measures do yield probability estimates from the authors' power law assumption. These three measures are the strength of coronal mass ejections, the strength of magnetic anomalies ("magnetic storms"), and a historical record over 400 years of nitrate anomalies in ice cores.

The nitrate anomalies are thought by some to be associated with solar eruptions, though (as the paper notes), this is disputed. The interest of the data set of these anomalies is that it is the only data set analyzed by the paper which actually contains the Carrington event, so that the mathematics can be based on its known strength.

Following is a summary of the paper's estimates of the probability of a Carrington-class event (or worse) in the next decade, based on the last three data sets:

Coronal Mass Ejections (CME). This data set spans 15 years. The paper's graphs show that the part describing CME's with speeds between 100 and 700 km/sec cannot follow a power law.[3] If this part is discarded, the remainder consists of two parts, each of which might conceivably follow a power law, a different law for each part.

An initial estimate based on the first part (corresponding to CME's with speeds between 700 and 2000 km/sec.) yielded a probability of 85% of an event as bad as Carrington within a decade. The author rejected this estimate as "not credible".

Despite the failure of the power law assumption to yield a believable estimate for this case, the paper applies a similar method to obtain an estimate based on the remaining observations of CME's between 2000 and

[3]The two graphs, in the paper's Figures (a) and (b) appear inconsistent (as will be discussed later), but both agree that CME's from 100 to 700 km/sec cannot follow a power law.

3500 km/sec. This remaining data consists of only about 20 observations, out of the original 14,735 observations.

The estimate obtained is really two estimates, each based on an auxiliary technical mathematical assumption. The two estimates are reported as probability 12% and probability 8.5%. The author considered the estimate of 12% as more credible and reported it (along with an estimate for another data set) as the paper's main conclusion.

I independently calculated estimates for the same part of the data set and obtained 14% instead of the paper's 12%, and 9% in place of the paper's 8.5%, which are comparable to the paper's figures. They may be of interest in comparison with analyses of the other data sets in which my arithmetic based on the paper's data produced estimates widely different from those of the paper. If I were making some systematic mistake, one might expect all the estimates to differ.

Geomagnetic storms. Roughly speaking, "magnetic storms" are anomalies in the earth's magnetic field "driven by changes in the solar wind", as the paper puts it. Their strength is measured by a technically defined quantity known as "*Dst* ", which is negative for the cases of interest here. The *Dst* of the Carrington event can only be guessed at. An initial estimate cited in the paper's reference "Lakhina et al. 2005" was $Dst = -1760$, but a later estimate, "Siscoe et al. 2006", reduced this to -850.

Because that strength is not known, even if the paper's assumptions and analysis are correct for this data set, the corresponding probability estimates would be expected to vary substantially. Assuming a strength of $Dst\ -850$, the paper reports a probability estimate of 12% for an event worse than Carrington in the next decade. Assuming a strength of $Dst = -1700$, it obtains an estimate of 1.5%, a difference of an order of magnitude. This underscores the uncertain nature of the paper's final estimate of probability 12%.

The paper considers that this data set "appears to follow a single power law distribution" (as opposed, for example, to the CME data set which contains two subsets, each possibly following a different power law). No goodness-of-fit analysis is included to provide quantitative support for this opinion, and the two data graphs in the paper's Figure 8 do not look obviously that way to my eye.

In particular, like the CME graphs, this graph has an apparent "knee" to the right of -300 *Dst* . The paper discounts this knee because it comprises only 6 events. However, the "knee" of the CME Figure 4 graphs comprises only about 20 events, yet the paper based its analysis on the data subset corresponding to that knee.

The point is that the paper makes various arbitrary decisions about what part of the data to ignore. Perhaps some of these decisions are arguably justified, but they do add to the overall uncertainty of the analysis.

Nitrate records. Particles from a solar storm which reach Earth are "generally believed by space physicists" to result in nitrate deposits which are preserved in ice. The paper notes that "ice core chemists are skeptical". Based on this data, the paper reports as 3% the probability of an event worse than Carringon in the next decade.

The paper seems to discount this 3% estimate, which is much lower than some of the others. One of the reasons given is that the data set comprises only 70 events. Yet the 12% estimate based on CME's which is the paper's main conclusion is based on only about 20 events below the "knee".

4 The paper's mathematical methods

The paper's equation numbers range from (1) to (7). In the following, numbers in this range refer to the corresponding equations in the paper. Equations originating in the present review are numbered starting at (101). An equation corresponding to the paper's number n is numbered $(10n)$.

I have quoted the paper's equations exactly as they are written there, even when I felt that they should be altered to conform with usual mathematical notation. In discussing the equations, I have sometimes translated them into more usual notation for clarity.

4.1 Definition of "power law" probability density function (pdf)

Let us start with a quote from the first paragraph of the paper's Section 2 on Methodology:

> "Here we outline the basic tools we will employ to compute the probability of occurrence of an extreme space weather event. A set of events, x, is said to follow a power law distribution if the probability of occurrence, $p(x)$, obeys the following relationship:
>
> $$p(x) = Cx^{-\alpha}, \tag{1}$$
>
> where the exponent α, is some fixed value and C is a constant determined from where the power law intercepts the y axis."

Translated into standard mathematical language, this might read:

> A random variable X taking on positive real values is said to have a *power law distribution* with exponent α if its probability density function (pdf) $p(x)$ is of the form
>
> $$p(x) = \begin{cases} Cx^{-\alpha} & x_{min} < x < \infty \\ 0 & \text{otherwise,} \end{cases} \tag{101}$$

where $\alpha > 1$, $C > 0$, and $x_{min} > 0$ are constants. This means that for all intervals $[a, b]$,

$$P(a \leq X \leq b) = \int_a^b p(x)dx \quad .$$

The paper's statement that "C is a constant determined from where the power law intercepts the y axis" is wrong—for $\alpha > 0$, the graph of $x \mapsto Cx^{-\alpha}$ is asymptotic to the y axis but never intercepts it. The inclusion of x_{min} in the definition for $\alpha > 1$ is essential; otherwise $\int_0^\infty p(x)dx$ will diverge at the lower limit.

The text of the analyses of the various data sets never tell us what is used for the crucial parameter x_{min}, though this can sometimes be guessed from the graphs. The only mention of x_{min} that I have been able to find in the text is on p. 3, where it is stated to be "some appropriate minimun value of x, below which the power law relation breaks down". But this is not the same as the definition in equation (101). Moreover, the x_{min} apparently used by the paper for the Coronal Mass Ejection data set, as inferred from Figure 4(b), is *not* the smallest x such that the power law breaks down below it.

Rescaling the data by defining $\bar{x} := \log_{10} x$, $\bar{y} = \log_{10} y$, changes the graph of a power law $x \mapsto y = Cx^{-\alpha}$ to the graph of $\bar{x} \mapsto \bar{y} = \log_{10} C - \alpha\bar{x}$, which is a straight line with slope $-\alpha$. Most of the paper's graphs of data make this rescaling, which makes it easy to guess whether the data do follow a power law.

The author seems to believe that various quantities to be studied will have power law pdf's, but no real evidence is presented. Assumption of power laws is critical to the paper's analysis, but some of the graphs of data presented which are represented as following power laws don't look unequivocally convincing to me. Examples are Figures 4, 5 and 10.

4.2 Framework of the paper's mathematical analysis

The paper considers various methods of quantifying the size of solar storms, such as X-ray fluxes, or the speed of coronal mass ejections, for which there are historical records. It analyzes these in the following mathematical framework.

It assumes that "events" such as a noticeable coronal mass ejection occur randomly in time and so are described by a Poisson distribution with rate λ: the probability of exactly k events between time t and $t + T$ is given by

$$p(k \text{ events in } [t, t+T]) = \frac{(\lambda T)^k}{k!} e^{-\lambda T} \quad .$$

We remind the reader that the expectation of the number of events in $[t, t+T]$ is easily calculated to be λT. To estimate the rate λ from N data points (events) collected in time interval $[t, t + T]$, one divides N by $T : \lambda \approx N/T$.

For the mathematical model to yield physically meaningful results, "strength of noticeable event" should be identified with the lower limit x_{min} of the power law distribution, (This will be explained in more detail later.) There is internal evidence that the paper may not have always made this identification.

When an event like a coronal mass ejection occurs, we can then ask what is its strength. The paper assumes the strength to be a random variable X with a power law pdf. We are interested in the probability that in a given time Δt, (like a decade) we will see one or more events of strength greater than a given strength x_{crit} (mostly, x_{crit} is taken to be a guess at the strength of the Carrington event). The paper's uses its equation (6) to calculate this probability:

> "... the probability of one or more events greater than x_{crit} occurring during some time Δt:
>
> $$P(x \geq x_{crit}, t = \Delta t) = 1 - e^{-N \frac{\Delta t}{\tau} P(x \geq x_{crit})}, \qquad (6)$$
>
> where τ is the total time span of the data set."[4]

This is the key equation of the paper. Its justification of equation (6) is very abbreviated. I would justify it as follows.

We can view the physical situation of looking for critical events (defined as events x with $x \geq x_{crit}$) as a new Poisson process. Like the old Poisson process, it describes events which occur randomly in time. A new event is determined as follows. First we wait for an event described by the original Poisson process, like a coronal mass ejection (of a predetermined discernible strength or greater). Then we ask if it is at least as large as x_{crit}. If it is, we declare that a new event has occurred. If the old Poisson process had rate λ, this new Poisson process will have rate $\lambda \, P(X \geq x_{crit})$, *assuming that the predetermined "discernible strength" is defined as* x_{min}.

To see this in the context of a numerical example, suppose that we define the predetermined discernible strength to be something less that x_{min}, say $x_{min}/2$. Suppose in the given time interval τ, we see 100 events of strength $x_{min}/2$ or larger. Suppose 80 of these are of strength x_{min} or larger. Then among these 80, about $80 \, p(X \geq x_{crit})$ are expected to exceed x_{crit}. Therefore, the rate of the new process will be estimated at $80 \, p(X \geq x_{crit})/\tau$. But the rate of the old process was approximately $\lambda = 100/\tau$, so the rate of the new process is no longer $\lambda \, P(X \geq x_{crit})$, which is the apparent premise for the paper's (6).

Continuing under the emphasized assumption, we obtain the paper's equation (6) (here numbered (106)) in notation which is more nearly standard and perhaps easier to read:

$$\begin{aligned} &P(\text{event in } [t, t+\Delta t] \text{ larger than } x_{crit}) = \\ &\qquad 1 - P(\text{no event in } [t, t+\Delta t] \text{ larger than } x_{crit}) \end{aligned}$$

[4] A more usual notation for $P(x \geq x_{crit})$ would be $P(X \geq x_{crit})$ where X denotes the random variable corresponding to drawing a value x from the power law distribution (101). In principle, there is nothing wrong with denoting the random variable by x instead of X, but that makes x unavailable for other uses. For example to avoid confusion in writing our equation (101), one would probably want to substitute the x in that equation by another symbol like z, which would make it look unintuitive.

$$
\begin{aligned}
&= 1 - \exp(-\Delta t\, \lambda,\ P(X \geq x_{crit})) \\
&\simeq 1 - \exp(-\Delta t\, \frac{N}{\tau}\, P(X \geq x_{crit})). \qquad (106)
\end{aligned}
$$

The last line estimates $\lambda \simeq N/\tau$ as described above.

Notice that in order to apply this equation as written, one has to know what is to be considered what we are calling "a discernible event" in order to evaluate N. If we change the definition of "discernible event" that will change the definition of N without changing Δt or τ. Hence, if the same value for

$$P(\text{an event in } [t, t+\Delta t] \text{ larger than } x_{crit}) = 1 - \exp(-\Delta t\, \frac{N}{\tau}\, P(X \geq x_{crit}))$$

is to be obtained, the value of $P(X \geq x_{crit})$ must change. But for fixed x_{crit}, this last value depends only on x_{min}, so the definition of "discernible event" must somehow be linked to the definition of x_{min}. As stated above, the physically correct definition is to identify the two.

None of this is discussed in the paper, so we can only guess at the author's definition. Internal evidence to be presented later suggests that an incorrect definition may sometimes have been used.

4.3 A typo in equation (7)

Before passing to the next section, we note that there is a typo in the paper's (7):

> "For Bernoulli distributions, that is, independent events that either happen or not, with a constant probability of occurrence, it can be shown that the probability of occurrence is given by
>
> $$P(x) = \frac{1}{1+\tau} \quad , \qquad (7)$$
>
> where τ is the average time to the event."

I had trouble understanding this because the symbol x is not defined. Finally I realized that the passage should read something like:

$$p(\text{an event occuring in } [t, t+T]\ ) = \frac{1}{1+\tau/T} \quad . \qquad (107)$$

The paper is does not say how this is derived, so perhaps it would be helpful to sketch a derivation. Consider waiting for an event in $[t, t+T]$ as one of N independent Bernoulli trials (a "Bernoulli trial" being an experiment with just two outcomes, "success" or "failure"), corresponding to N successive intervals. Considering one trial as requiring time T, let τ denote the average time from the start of the N trials to and including the first success. Calculate this average time in terms of $q := P(\text{an event occurring in [t, t+T]})$ and solve for q. The calculation is straightforward except that one has to sum the infinite series $\sum_{k=1}^{\infty} k(1-q)^k = (1-q)/q^2$, for which there is a well known trick.

5 The paper's application of the mathematics to the Coronal Mass Ejection (CME) data

5.1 Possibly incorrect analysis of CME data for speeds between 700 and 2000 km/sec

The author's first attempt at analyzing the CME data is described rather sketchily[5] on its p. 6. It appears to have assigned a slope of -3.2 to the part of Figure 4(a) between speeds 700 and 2000 km/sec, and using this calculates 85% for the probability of an event worse than Carrington in the next decade. The author rejected this estimate because he considered it "not credible" (too high), without questioning the power law assumption.

Figure 4(a) clearly cannot represent a power law for speeds between 100 and 700 km/sec because the graph is hook-shaped in this region, whereas a power law would imply a graph which is a straight line. If mathematics that assumes a power law is applied to a pdf which is clearly very different from a power law, one cannot expect meaningful results.

The paper's description of how the 85% estimate was obtained is too vague to evaluate in detail. But I suspect that it may have been incorrectly obtained by a mechanical application of equation (6) to the data corresponding to speeds greater than 100 km/sec. Specifically, it may have been produced by using the incorrect value $x_{min} = 100$. Here is why.

For a power law distribution (101) with exponent α and lower bound x_{min}, one routinely calculates (cf. equation (108) below) that

$$P(X \geq x_{crit}) = \left(\frac{x_{min}}{x_{crit}}\right)^{\alpha-1} \quad .$$

Using the reported slope $-\alpha = -3.2$ of Fig. 4(a), in equation (6) along with the incorrect $x_{min} = 100$ and the uncontroversial $\tau = 15$, $\Delta t = 10$, $x_{crit} := 5000$, gives

$$\begin{aligned} &P(\text{event in } [t, t+\Delta t] \text{ larger than } x_{crit}) = \\ &= \quad 1 - \exp(-\Delta t \, \frac{N}{\tau} \, P(X \geq x_{crit})). \\ &\quad\quad 1 - \exp(-(10/15)N(100/5000)^{3.2-1}) \quad . \end{aligned}$$

Setting this equal to the paper's reported result of probability 0.85 (85%) and solving for N yields $N = 15,562$, which is fairly close to the $N = 14,735$ which the paper reports for the number of events with speed at least 100 km/sec.

However, there seems to be no reasonable way to get the paper's 85% estimate from the correct $x_{min} = 700$. The corresponding calculation for $x_{min} := 700$ is

$$P(\text{event in } [t, t+\Delta t] \text{ larger than } x_{crit}) =$$

[5]For example, it refers to "this slope (-3.2)" without saying whether this is the slope of the pdf of Figure 4(a) or the CCDF of Figure 4(b). An inquiry to the author remains unanswered.

$$1 - \exp(-(10/15)N(700/5000)^{3.2-1}) \quad .$$

This yields $N \simeq 215$, which is far too low since there were on the order of several thousand CME's with speeds at least 700. (See below for an explanation of the qualifier "on the order of".)

At this point, several natural questions may occur to the reader. First, instead of taking this roundabout way of questioning the 85% estimate, why didn't I simply carry out the correct calculation and compare it with the paper's 85%? The reason is technical.

The source of the data lists several "speeds" for each CME. One of these, the so-called "linear" speed, is obtained by fitting a linear polynomial to the observed datum. Another, the "quadratic speed", instead fits a quadratic polynomial.[6] The paper uses quadratic speeds, but one can search the database only on linear speeds. There is not a great deal of difference between the two, but for an airtight comparison with the paper's conclusion of 85%, one should use quadratic speeds, and the only way I know to obtain those seems to be to manually sort all of the thousands of events in the database.

I did carry out the calculation using linear speeds. There were 1719 observations with linear speed larger than 700. Substituting this for N and $x_{min} := 700$ in equation (6) yields an unbelievable probability 99.99% for an event worse than Carrington in the next decade. This indicates that the power law assumption is unrealistic for this data set.

The second natural question is the following. Since the author rejected the 85% estimate as "not credible" anyway, what does it matter if it may have been obtained incorrectly? There are two answers to this second question. First, if the estimate *was* obtained incorrectly, readers should take seriously the possibility that other estimates might have also been incorrectly obtained. The paper's exposition of its mathematical methods is so sketchy that most of the estimates are difficult or impossible to check.

The second answer to the second question is more subtle, at least as important, and somewhat surprising. It is natural to imagine that the use of inappropriate N and x_{min} would have *caused* the "not credible" estimate, but this is not true. *If* the data between 100 and 700 *had* followed the assumed power law, the 85% estimate would have been even *higher*!

To see this, consider the dashed line in Figure 4(a) representing the hypothetical power law pdf that more or less fits the data between speeds of 700 and 2000. This line lies way above the pdf shown in Figure 4(a) for speeds between 100 and 700. If the number N of events in equation (6) were calculated according to this power law (instead of being determined by the observed data reported in the Figure 4(a) histogram), the result would have been a much larger N in equation (6). Since the right side of this equation is monotonic increasing in N, an estimate of the probability of an event as bad as Carrington in the next decade even larger than 85% would have been obtained. This means that

[6]The terms "linear" and "quadratic" in this context refer to the way of estimating the terminal speed of a *single* CME, and have nothing to do with fitting a linear or quadratic polynomial to the entire set of the CME speed data as in the paper's Figure 4.

the original power law assumption was at fault for the "not credible" estimate, rather than an incorrect implementation of it.

5.2 Analysis of the CME data for speeds greater than 2000

5.2.1 The paper's analysis

The paper's Figure 4(a) looks almost like a straight line for speeds greater than 700 km/sec. The only anomaly is two outlying data points above 2000 km/s.

The graph of Figure 4(b) looks noticeably different. It looks like a straight line from 700 to 2000 km/sec, but above 2000 km/sec it looks something like a straight line with a much steeper slope. This indicates that the power law assumption on which the paper's analysis is based is questionable for this data set. If a power law using the slope of the data between 700 and 2000 km/sec is assumed, then Figure 4(b) indicates that this power law fails to predict the observed data for speeds larger than 2000.

But what if the data for speeds greater than 2000 did obey a power law, but a different one than for speeds between 700 to 2000? Is it possible that equation (6) using this new power law might predict correctly the probability of an event worse than Carrington in the next decade?

This possibility doesn't sound like one in which one could have great confidence. If the power law assumption for the data between 700 and 2000 produced a result recognized as "not credible", what reason would there be to think that the prediction based on data between 2000 and 3500 (the largest speed in the data set) might produce a better estimate?

However, the paper soldiers on to obtain an estimate using only the data for speeds greater than 2000 km/sec. This reduced data set contains only about 20 observations (a fact obtained from the original source of the data and not reported in the paper).

Actually, two estimates are obtained, corresponding to two different ways of estimating the slope of the data for speeds greater than 2000. The two different estimates of the slope are the maximum likelihood estimate and the least-squares estimate. (This will be meaningful only to readers who already knoow what these terms mean—to attempt to explain them would take us too far afield.)

The respective estimates of the probability of an event worse than Carrington in the next decade are 12% and 8.5%. The paper considers 12% as the most credible estimate and reports it (along with a similar estimate from another data set) as the conclusion of the paper.

5.2.2 An independent analysis

Because of the vagueness of the exposition of the Coronal Mass Ejection (CME) part of the paper,[7] and the author's lack of response to my inquiries, I carried

[7] For example, paragraph [31] on p. 6 refers to a "slope (-3.2)" without saying of what it is the slope, the α of Figure (a) or the $\alpha - 1$ of Figure (b). The subsequent discussion is

out an independent analysis of the part of the data corresponding to speeds above 2000 km/sec, taken from the original source of the paper's data.[8] The so-called "quadratic speeds" (see the previous subsection for an explanation) were used, just as in the paper. This was feasible because the reduced data set was so small—there were only 20 CME's in the 15 years with (quadratic) speeds larger than 2000. Denoting the estimated slope of Figure 4(a) as $-\alpha$ (so that α is the exponent of the pdf of equation (101)), the results were as follows.

The maximum likelihood estimate (MLE) yielded $\alpha = 5.9$ with a corresponding probability of 14% of an event worse than Carrington in the next decade. This assumes what the paper characterizes as its "guess" of 5000 km/sec as the speed of the Carrington event.

These numbers are not exactly what the paper obtained, but they are close enough that the "slope (-3.2)" of paragraph [31] and the other slopes further on must refer to the estimate of the exponent α of the pdf $p(x) = Cx^{-\alpha}$ (equivalently, to the slope of the "best fit" line to Figure 4(a)) instead of the slope of the complementary cumulative distribution function (equivalently, the slope of the "best fit" line to Figure 4(b)). For example, the reported "slope" $-\alpha = -6.1$ does yield the paper's 12% probability estimate, but if the -6.1 referred to the slope of of the CCDF (Figure 4(b)), a probability estimate of 5% would have been obtained.

The resolution of this ambiguity will become important in the next subsections concerning the paper's analysis of the other data sets. Assuming that the exposition carries the same meaning as for the CME data just analyzed, the probability estimates obtained for the remaining data differ by about an order of magnitude from those given in the paper.

5.3 Analysis of the data on geomagnetic storms

We shall obtain an estimate for $p(X > x_{crit}$ in a decade), the probability of an event worse than Carrington in the next decade, from equation (106) (the paper's equation (6)). The paper reports the data in the same format as for the Coronal Mass Ejection (CME) data analyzed above. Again, there is a Figure 8(a) which is a histogram of the number of events vs. an independent variable.[9] The negative of the slope of the line of best fit to Figure 8(a) (after adjustment for different scales on the horizontal and vertical axes) estimates the the exponent α of the power-law pdf $p(x) = Cx^{-\alpha}$. The MLE estimate of α (the only estimate considered by the paper for this data set) is calculated from its equation (4).

The paper states that:

equally vague. Without this knowledge, it is impossible to check the paper's final probability estimates. An inquiry to the author remains unanswered.

[8] Cited in the caption of its Figure 3 as http://cdaw.gsfc.nasa.gov/CME_ list/.

[9] Roughly speaking, the independent variable is a measure of the deviation of an average magnetic field from its normal value, but the nature of the independent variable is irrelevant to the analysis. The independent variable for the CME data was the speed of a coronal mass ejection.

"The slope of the MLE fit is −3.2."

As with the CME data, the paper does not explicitly say that this is its estimate of $-\alpha$, but that is the natural interpretation. If there were any question that the paper might be referring instead to the slope of the complementary cumulative distribution function (CCDF) graphed in Figure 8(b), this should be resolved by the fact that the syntax was essentially the same for the CME data, and there the meaning was definitively resolved in favor of -3.2 being an estimate for $-\alpha$.

We are going to observe that under this interpretation, the paper's final probability estimates appear badly incorrect. To do this in a way that facilitates checking the arithmetic, we interpolate a brief digression.

Equation (106) requires $p(X \geq x_{crit})$ as input, so it will be convenient to introduce a simple formula for this:

$$P(X \geq x_{crit}) = \left(\frac{x_{min}}{x_{crit}}\right)^{\alpha-1} \quad . \tag{108}$$

The formula just given follows from a routine integration summarized in the paper's equation (3) as

$$P(X \geq x) = \frac{C}{\alpha - 1} x^{-(\alpha-1)} \quad . \tag{103}$$

(The paper states it for $x := x_{crit}$, but x_{crit} can be any value greater than or equal to x_{min}.) Applying this to $x := x_{min}$ gives

$$1 = P(x \geq x_{min}) = \frac{C}{(\alpha - 1)} x_{min}^{-(\alpha-1)} \quad ,$$

whence

$$C = (\alpha - 1) x_{min}^{\alpha-1} \quad .$$

Applying equation (103) with this value of C and $x := x_{crit}$ gives (108). Substituting (108) into (106) gives

$$P(\text{event in } [t, t + \Delta t] \text{ larger than } x_{crit}) = 1 - \exp(-\Delta t \, \frac{N}{\tau} \left(\frac{x_{min}}{x_{crit}}\right)^{\alpha-1}). \tag{109}$$

The input to this equation taken from the paper is:

$$\Delta t = 10, \quad \tau = 2009 - 1964 = 45, \quad N = 746, \quad x_{crit} = 850.$$

The paper never tells us what it uses for x_{min}, but Figure 8(a) strongly suggests it is $x_{min} := 100$, so that is what we shall use. Evaluation of (109) with the indicated inputs gives

$$p(X > x_{crit} \text{ in decade}) = 1 - \exp(-10 \times \frac{746}{45} \times \left(\frac{100}{850}\right)^{3.2-1} = 0.78 \quad ,$$

or 78%, compared to the paper's 12%. The 78% is not credible, given that already more than 15 decades have passed without an event nearly as bad as Carrington.

The value $x_{crit} := 850$ is the paper's best guess at the strength of the Carrington (derived from an estimate in another paper), but the higher value $x_{crit} := 1700$ is also considered. For that value, $p(X > x_{crit}$ in decade) evaluates to 27%, compared to the paper's announced value of 1.5%.

5.4 Analysis of the data on ice core samples

The calculation will follow that of the previous subsection. The text states a "slope" of -2.0, so we use $\alpha := 2.0$ in equation (108).

The other parameters given in the paper are

$$\Delta t = 10, \quad \tau = 1945 - 1562 = 383, \quad N = 70, \quad x_{crit} = 18.8 \times 10^{-9}.$$

The only missing parameter is x_{min}, which as usual is not given. However, Figure 10(a) suggests that it should be between 2×10^9 and 3×10^9, so we shall do the calculation twice for these two values.

For $x_{min} := 2 \times 10^{-9}$ we obtain

$$p(X > x_{crit} \text{ in decade}) = 0.18,$$

or 18%. For $x_{min} := 3 \times 10^{-9}$, this increases to 25%. These are about an order of magitude more than the paper's reported 3.0%. I leave it to the reader to decide if they are credible.

6 Closing remarks

A major conclusion of the paper, as presented in its last paragraph, is

> "Additionally, our analysis has shown that a relatively rich subset of space physics data can be approximated by power law distributions."

I disagree that it has "shown" anything like that, if the word "shown" is used as a synonym for "proved" or "demonstrated" as is usual in mathematics.

Given the nature of the subject matter, the presentation of the conclusions is more important than in more technical scientific work. The popular press has already represented as scientific fact the estimate of a 12% probability of a Carrington-type catastrophe costing trillions of dollars. If the 12% estimate is taken seriously, then one could justify spending hundreds of billions to prepare. But what if the probability is actually much lower? Those who make policy are not likely to delve into the mathematical intricacies underlying the 12% estimate. It is disturbing that it may be becoming uncritically accepted.

I regard the paper's probability estimates as little more than guesses, all based on a questionable assumption that the data does follow a power law. In addition, they are based on other guesses such as the strength of the Carrington event for the CME data set and for the geomagnetic storm data set. The paper does not give us enough information to be sure that its mathematics is correctly applied. Some of it is almost certainly incorrect or misapplied.

There is nothing wrong with making educated guesses. The matter is important and the data insufficient to draw firm conclusions. However, educated guesses should be presented as such, and not as conclusions with which serious disagreement is unlikely. The paper concludes:

> "Our results allowed us to answer a basic question, at least in an approximate way: How likely are [Carrington] events? ... our results overall suggest that the likelihood of another Carrington event ocurring within the next decade is $\sim 12\%$."

This makes it sound as if one might quibble over whether the probability were 10% or 15%, but not over its order of magnitude.

I don't think the paper does convincingly establish even the order of magnitude. The range of estimates obtained by the paper from various data sets using various assumptions vary from 1.5% (p. 8, paragraph [39]) to 85% (p. 6, paragraph [31]).

Because of this, in the paper's present form, I think its probability estimates should not be accepted even as educated guesses. There are just too many uncertainties and unverifiable assumptions in the paper's analysis, as well as probable mistakes.

In conclusion, I want to make clear that I do think society should take seriously the possibility of a Carrington-class event in the coming decades, and prepare according to the best estimates available as to its probability. My criticism of various technical aspects of [1] is made in the hope that they may be helpful in preparing more soundly based estimates in the future. I also hope that it may make people in awe of mathematics more aware of its limitations.

On the other hand, suppose that the true probability of an event worse than Carrington in the next decade is two orders of magnitude below the paper's 12% estimate, on the order of tenths of a percent, which seems conceivable to me. That could still justify spending a few billions of dollars to prepare for a catastrophe that could cost trillions. A billion dollars amounts to about $3 for each inhabitant of the U.S. I would happily pay that amount for insurance.

References

[1] P. Riley, "On the probability of occurrence of extreme space weather events", *Space Weather* **10** (2012), S02012

[2] National Research Council. Severe Space Weather Events–Understanding Societal and Economic Impacts: A Workshop Report. Washington, DC: The National Academies Press, 2008. available at http://books.nap.edu/catalog.php?record_id=12507