Bayesian Inference for Solar Flare Extremes

B. Griffiths¹, L. Fawcett², and A. C. Green³

¹School of Mathematics, University of Bristol, Bristol, UK, ²School of Mathematics, Statistics & Physics, Newcastle University, Newcastle upon Tyne, UK, ³School of Engineering, Newcastle University, Newcastle upon Tyne, UK

Abstract While solar flares are a frequent occurrence, extreme flares are much rarer events and have the potential to cause disruption to life on Earth. In this paper we use Extreme Value Theory to model extreme solar flares, with inference performed in the Bayesian paradigm. The data used have been provided by the National Oceanic and Atmospheric Organisation and consist of recordings of peak flux measurements. After proposing several methods for analysis and selecting our preferred technique—which substantially increases the precision of estimates of key quantities of interest—we improve upon this technique still further by considering the use of informative prior distributions. Doing so, we estimate that a Halloween-type solar event, and a Carrington-type event, might occur once (on average) every 49 (29, 85) and 92 (50, 176) years respectively (95% credible intervals shown in parentheses). These findings are similar to those obtained by Tsiftsi and De la Luz (2018), https://doi.org/10.1029/2018SW001958 and Elvidge and Angling (2018), https://doi.org/10.1002/2017SW001727 however, the confidence intervals obtained in both are substantially wider than those found in our study, lending increased certainty to the estimated time between events of such magnitude in our work. We argue that taking the extremal index into account, even when this measure indicates weak temporal dependence, is beneficial to the analysis.

Plain Language Summary The Carrington event of 1859 induced the most powerful geomagnetic storm on record. It produced stunning light displays as far south as the Caribbean, the likes of which are usually only visible in regions much closer to the Earth's poles. Although a significant portion of the world's telegraph lines were unusable for a short time, the impact on human life was limited. However, it is widely recognized that today the consequences could be far more significant—perhaps causing widespread disruption to satellite communication systems, power outages and impacts upon water distribution systems. Recent attempts to model solar flare extremes have been made using techniques from Extreme Value Theory (EVT), a branch of Statistics providing a framework for "predicting the unpredictable"; for example, estimating the probability of a Carrington-like event happening again. However, such estimates often come with very large error margins, rendering their practical interpretation useless. Using more sophisticated methods from the EVT toolbox, the current paper reports significantly reduced error margins compared to those in the recent literature. For example, we estimate there to be a 95% probability that a Carrington-like event will occur once, on average, every 50–176 years; crude comparisons to recently-published work might put this at once every 20–6,500 years.

1. Introduction
1.1. Contextual Background

Solar flares—explosive outbursts of energy from the Sun—can cause intense disturbances of space weather, with the potential to impact vital technologies that society depends upon. Flares can produce streams of highly energetic protons and electrons—solar particle events—that can affect the Earth's upper atmosphere. For example, such events can impact the Earth's magnetosphere, presenting radiation hazards to spacecraft; ionization in the ionosphere on the sunlit side of the Earth can also occur as a result of the increased level of X-ray and extreme ultraviolet radiation (Liu et al., 2011; Mitra, 1974), potentially leading to temporary blackouts in navigation and communication signals. The strongest flares are almost always correlated with the occurrence of coronal mass ejections (CMEs)—the expulsion of large quantities of plasma and electromagnetic radiation from the Sun. When this ejection is directed toward Earth, the shockwave of traveling mass can cause a geomagnetic storm. Along with strong auroral displays at the Earth's magnetic poles, such storms can also result in damage to satellites and electrical transmission line facilities, with the potential for long-lasting power outages. The review paper of Hapgood et al. (2021) discusses the potential consequences of severe space weather, with reference to societal disruption in the UK.
Understanding the behavior of solar flares is therefore a crucial part of space weather forecasting and prediction. Although such forecasting typically lags behind its terrestrial counterpart (Murray, 2018), cutting edge techniques from Statistics/Data Science are now starting to be explored, including ensemble methods for combining forecasts generated from multiple methods (Guerra et al., 2015). However, given the dependence between the occurrence of CMEs and strong solar flares (Kawabata et al., 2018), it is perhaps surprising to see relatively few attempts at utilizing techniques from extreme value theory—the aim of the current paper. The ability to predict the strength of solar flare extremes is crucial to mitigating the impact of extreme flares and related events, such as CMEs; just as important is the capability to predict the average wait time between events that might cause significant impact to life on Earth. Indeed, such capabilities are seen as a top priority in space weather forecasting (Jonas et al., 2017).

The most common classification system for solar flares, as recorded by the Geostationary Operational Environmental Satellite (GOES), uses letters A, B, C, M and X referring to the peak flux measured in watts per square meter (W m$^{-2}$). The letters refer to flares of magnitude $10^{-8}$, $10^{-7}$, ..., $10^{-4}$ W m$^{-2}$ respectively, and each letter is followed by a number which forms a product with the magnitude. For example, an M5 flare is a flare with peak flux $5 \times 10^{-5}$ W m$^{-2}$.

Solar flares are not rare events, with several per day occurring in a solar maximum. In this paper we are concerned with solar flare extremes, such as those observed during the “Carrington event” and the “Halloween Storms”. The Carrington event of 1859 was recorded by astronomers Richard Carrington and Richard Hodgson independently (published consecutively in Monthly Notices of the Royal Astronomical Society 20, see Carrington [1859] and Hodgson [1859]), where a CME collided with the Earth’s magnetosphere and caused the largest geomagnetic storm ever recorded. The peak flux during this storm has been estimated as an X45 ± 5 flare (see Cliver and Dietrich [2013]), and at the time a significant portion of the world’s telegraph lines were unusable for around 8 hr, with aurorae visible as far south as the Caribbean. The Halloween Storms were a series of solar flares and CMEs, peaking around October 28–29, 2003. They were measured to have had an X35 ± 5 flare (see Cliver and Dietrich [2013]) at their peak, where numerous spacecraft were damaged, aircraft needed to be re-routed, and a power outage occurred for around 1 hr in Sweden (see Pulkkinen et al. [2005]). If a Carrington-type event were to happen today we might see damage to communication systems, satellites and spacecraft (including global positioning systems). There could also be mass power outages, adversely affecting water distribution systems, hospitals and refrigeration for food and medicine (National Research Council, 2008). Such problems could be particularly acute in highly interconnected regions, such as the US eastern seaboard. Thus, preparation for such an event is vital in order to minimize the impact caused.

1.2. Data

In order to predict the magnitude of extreme solar flares, or indeed the expected frequency of events such as the Carrington event or the Halloween Storms, we examine historical data on X-ray fluxes of solar flares consisting of over 70,000 readings from 1981 to 2017 obtained from the GOES (NOAA, 2020). In fact, we analyze the peak X-ray flux (W m$^{-2}$; as used throughout this paper) of each solar flare event, determined by one-minute averages of data recorded by GOES. As described in NOAA (2020), an event is defined based on three parameters—the begin time, the end time, and the peak flux. The begin time of an X-ray event is the first minute of a monotonic increase in flux at wavelengths of 0.1–0.8 nm for over four minutes; the maximum is simply the minute during which the peak flux is the greatest; and the end time occurs when the flux level reaches a point halfway between the peak flux and the level of flux observed before the start of the event.

Figure 1 shows a time series plot of the data collected between 1 January 2003 and 1 January 2004, clearly illustrating the Halloween Storms of 2003. The sample mean of the peak flux measurements is $5.49 \times 10^{-6}$ W m$^{-2}$, which is larger than the third quartile at $3.6 \times 10^{-6}$ W m$^{-2}$, indicative of a very right-skewed data set.

1.3. Statistical Modeling

Previous statistical work to model the occurrence of solar flares includes power law modeling (Lu & Hamilton, 1991; Riley, 2012) and inference based on a lognormal distribution (Riley et al., 2017; Riley & Love, 2017). However, a power law distribution with shape parameter less than 2 has undefined mean and standard deviation. Furthermore, Roodman (2015) suggests a power law distribution is at risk of over-prediciting extreme events. In
our analysis we aim to expand upon work by Elvidge and Angling (2018) and Tsiftsi and De la Luz (2018) who adopt an Extreme Value Theory approach to modeling the most energetic solar flares. These papers estimate that we should see a Carrington-type event, on average, once every 100 and 110 years respectively. However, in both cases these estimates are accompanied by very wide confidence intervals; for example, the 95% confidence interval for the return period for a Carrington-type event, according to Elvidge and Angling (2018), is (30, 900) years.

We aim to increase precision by adopting a more data-rich extreme value analysis. Unlike Elvidge and Angling (2018) and Tsiftsi and De la Luz (2018), who perform inference within a maximum likelihood setting, we choose to work within the Bayesian paradigm. This enables a further increase in precision through the incorporation of informative prior distributions; the Bayesian framework also handles prediction neatly through the posterior predictive distribution (see, for example, Fawcett and Green [2018]), whilst allowing a more intuitive interpretation of confidence intervals. Interested readers can easily reproduce results in this paper through a user-friendly, interactive web-based application, built using the R-Studio package Shiny (RStudio, Inc, 2013). Through this app users can also perform similar extreme value analyses of their own data through a generic data upload feature.

1.4. Structure of This Paper

The structure of this paper is as follows. In Section 2 we introduce Extreme Value Theory, the motivation for using it, and different approaches within. In Section 3 we outline the methods used for performing a Bayesian analysis, and the advantages that such an analysis provides within an extreme value context. In Section 4 we perform our analysis of the data, discussing differences between methods in the approach for dealing with dependence; we also demonstrate the use of informative prior distributions and making predictions via the Bayesian posterior predictive distribution. We conclude in Section 5 with a comparison of our findings to those in the literature; we also make some suggestions for further analysis, and describe the app for performing analyses such as those used in this paper.

2. Extreme Value Theory

Extreme Value Theory (EVT) provides a framework for estimating the probability of extreme events, perhaps more extreme than have ever been recorded to-date. Without any assumption about the underlying distribution of our random variable, the extremal types theorem provides limiting results for the extremes of our process, providing a range of techniques for modeling the tail behavior of our random variable. This is analogous to the central limit theorem, which gives the Normal distribution as a limiting model for the sample mean. Some key references relating to the historical development of the field can be found in Section 2.1; interested readers might
In this work, we examine the role of EVT in the estimation and prediction of solar flares. We are specifically interested in the most energetic flares, since they are those which are most likely to significantly impact upon technology and infrastructure (Kappenman, 2003; see Section 1.1). Working within EVT allows practitioners to prepare for certain strength flares that can be expected to have an impact upon modern global positioning systems and radio communications at low frequencies (Afraimovich et al., 2007; N. Cheng et al., 2019). For example, it is possible to estimate the return period for events such as the Carrington and Halloween events; that is, the average estimated time interval between events of a similar intensity. Inversely, we might estimate the $r$-year return level; that is, the energy of a flare we might expect to see (on average) once every $r$-years (see Section 2.3).

Since we have only ever observed one Carrington-type event (see Section 1)—and in fact, the data we use do not even date far back enough to include this event—we need to use EVT as a tool for extrapolation when estimating the return period for such an event. The same is true of the Halloween event in 2003; this event consisted of a set of flares, the most energetic of which saturated GOES, and so extrapolation beyond our range of observed data is necessary to estimate the associated return period.

2.1. Block Maxima Approach

Suppose we have a sequence of independent and identically distributed (IID) random variables, $X_1, \ldots, X_n$, with common distribution $F$. The block maxima approach focuses on fitting a model to the maximum order statistic $M_n = \max\{X_1, X_2, \ldots, X_n\}$. Typically, it is the case that, for large $n$,

$$\Pr(M_n \leq x) \approx F^{\text{max}}(x)$$  \hspace{1cm} (1)

where $\theta \in (0, 1)$ is known as the extremal index (Leadbetter & Rootzén, 1988) which is a parameter quantifying dependence in our extremes, with $\theta = 1$ for an independent process and $\theta \to 0$ for increasing levels of extremal dependence. Initially concerned with the independent case (i.e., $\theta = 1$), classical EVT sought families of limiting models for $F^n$ for large $n$, without reference to the marginal distribution function $F$; uncertainty in $F$ would, of course, lead to significant uncertainty in $F^n$.

Examining the behavior of $M_n$ as $n \to \infty$ gives rise to the Extremal Types Theorem (also known as the Fisher-Tippett-Gnedenko theorem, see Fisher and Tippett [1928]; Gnedenko [1943]). The theorem states that if there exist sequences of constants $a_n > 0$ and $b_n$ such that, as $n \to \infty$,

$$\Pr\{\frac{(M_n - b_n)}{a_n} \leq x\} \to G(x)$$

for some non-degenerate distribution $G$, then $G$ is one of the three extreme value distributions. In fact, von Mises (1954) and Jenkinson (1955) independently derived a distribution which encompasses all three types of extreme value distribution; the generalized extreme value distribution (GEV), with cumulative distribution function (CDF)
\[ G(x; \mu, \xi, \zeta) = \begin{cases} 
\exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\zeta} \right) \right]^{-1/\xi} \right\}, & \xi \neq 0; \\
\exp \left\{ -\exp \left( -\left( \frac{x - \mu}{\zeta} \right) \right) \right\}, & \xi = 0,
\end{cases} \]  
\tag{2}

defined on \( 1 + \xi(x - \mu)/\zeta > 0 \). The parameters \( \mu \in \mathbb{R}, \zeta > 0 \) and \( \xi \in \mathbb{R} \) are parameters of location, scale and shape respectively. The case where \( \xi = 0 \) is derived by taking the CDF in Equation 2 (where \( \xi \neq 0 \)) and taking the limit as \( \xi \to 0 \) (to obtain the case in Equation 2 where \( \xi = 0 \)). The shape parameter \( \xi \) determines the tail behavior of the process, with \( \xi = 0 \) defining the Gumbel distribution, \( \xi > 0 \) defining the Fréchet and \( \xi < 0 \) defining the Weibull distribution as set out in the Extremal Types Theorem.

A potential difficulty in using the GEV, in practice, is determining the length of blocks from which to derive a set of maxima. Blocks need to be large enough so that the limiting results hold; however, too large and there will be too few maxima from which to make inferences. And of course, the process of extracting maxima for analysis can be extremely wasteful, undoubtedly discarding large values that might inform us about the behavior of the extremes of the process.

### 2.2. Threshold Approach

The threshold-based approach classifies an observation as extreme if it exceeds some pre-determined high threshold \( u \). Provided this threshold is suitably high, and letting \( X \) denote an arbitrary term from \( X_1, \ldots, X_n \), it can be shown that the limiting distribution for threshold excesses \( Y = (X - u) \) (given that \( X > u \)), is

\[ H(y; \sigma, \xi) = \begin{cases} 
1 - \left( 1 + \frac{\xi y}{\sigma} \right)^{-1/\xi}, & \xi \neq 0; \\
1 - \exp \left( -\frac{y}{\sigma} \right), & \xi = 0,
\end{cases} \]  
\tag{3}

defined on \( y > 0 \) and \( (1 + \xi y/\sigma) > 0 \), where \( \sigma = \zeta + \xi(u - \mu) \) and \( \mu, \zeta \) and \( \xi \) are defined as in Equation 2. This is known as the generalized Pareto distribution (GPD; Pickands, 1975). As for the GEV, when we have \( \xi = 0 \) we take the limit of Equation 3 as \( \xi \to 0 \).

The GPD provides a limiting model for excesses over a high threshold \( u \); thus, careful consideration of what constitutes an appropriately high threshold is required. Simple graphical diagnostics are available to aid the choice of threshold (see, for example, Coles [2001]); see also Scarrot and MacDonald (2012) for a comprehensive review of threshold selection methods, including a model formulation enabling automatic threshold identification.

Given the potential for increased precision, owing to the inclusion of more data, in this paper we opt for a threshold-based approach to analysis rather than a block maxima approach.

### 2.3. Return Levels and Return Periods

The goal of most extreme value analyses is often the estimation of return levels or return periods. For the \( r \)-year return level \( z_r \), inversion of the right-hand-side of Equation 1 (i.e., setting equal to \( 1 - (rn) \)) and solving for \( x = z_r \), where \( n \) is the average number of observations per year) leads to the following expression, assuming a GPD for \( F^n \):

\[ z_r = \begin{cases} 
u + \sigma \xi^{-1} \left( \lambda \sigma^{-1} n \right)^{-\xi} - 1, & \xi \neq 0; \\
u - \sigma \log \left( \lambda \sigma^{-1} n \right), & \xi = 0,
\end{cases} \]  
\tag{4}
where \( u_r = 1 - \left[1 - (rn)^{-1}\right]^{1/\theta} \) and \( \lambda_u = \Pr (X > u) \) (which can be estimated as the proportion of observations exceeding \( u \)); see, for example, Coles (2001).

As previously discussed, we may also be interested in the return period of an event of a specified strength. For example, how often (on average) might we expect to see a Carrington-type event? Rearranging Equation 4 for \( r \), where \( z_r \) is the strength of the event of interest, gives return period:

\[
r = \begin{cases} 
   n^{-1} \left[1 - \left(1 - \lambda_u \left\{ \xi \sigma^{-1} [z_r - u] + 1 \right\}^{-1/\theta} \right]^{-1}, & \xi \neq 0; \\
   n^{-1} \left[1 - \left(1 - \lambda_u \exp \left\{ -\sigma^{-1} [z_r - u] \right\} \right)^{1/\theta} , & \xi = 0.
\end{cases}
\]

(5)

2.4. Dependence

A common occurrence in a data set of threshold excesses is dependence owing to temporal clustering. For example, during a solar storm we may expect to see several large solar flares across consecutive time points. Subject to some assumptions relating to long-range dependence (see Coles [2001, Ch. 5]), assuming a GEV for \( F^\theta \) in Equation 1 and then powering by \( \theta \) leads to another GEV, albeit re-parameterized with the extremal index being absorbed into modified location and scale parameters. Working with threshold excesses and assuming a GPD for \( F^\theta \), powering by \( \theta \) (when \( \theta < 1 \)) in Equation 1 does not lead to a re-parameterized GPD, meaning that careful consideration must be given to the estimation of the extremal index here. Ignoring dependence (essentially assuming that \( \theta = 1 \)) can lead to bias in parameter estimates (e.g., maximum likelihood estimates) and inflation of precision (e.g., standard errors that are too small). We now explore two methods for handling this dependence in the context of modeling threshold excesses: one which removes it (Section 2.4.1), and one which accounts for it (Section 2.4.2).

2.4.1. Removing Dependence: Declustering

Probably the most commonly-used method for dealing with dependence in threshold excesses is runs declustering; see, for example, Davison and Smith (1990). Here, a declustering interval \( \kappa \) is chosen; clusters of extremes over \( u \) are then deemed to have terminated once \( \kappa \) consecutive observations fall sub-threshold, and from each cluster the maximum, or “peak” value, is extracted. Inference then proceeds by modeling the set of cluster peak excesses with the GPD, assuming that \( \kappa \) was chosen large enough so that this set of filtered excesses are independent (and so \( \theta = 1 \) in Equation 1). This approach is often referred to as the peaks over threshold (POT) method.

There are, however, issues in identifying a suitable value for \( \kappa \); if it is too small then the cluster peaks might not be far enough apart to safely assume independence; too large and we won’t have enough data from which to perform reliable inference. Furthermore, no matter the choice of \( \kappa \), it is inevitable that an already-reduced data set of threshold excesses is going to be reduced further, which can be very wasteful. Indeed, through an extensive simulation study Fawcett and Walshaw (2012) demonstrate the sensitivity of parameter and return level estimation to the choice of \( \kappa \), with substantial bias in some cases.

2.4.2. Accounting for Dependence Through the Extremal Index

Given the issues surrounding declustering, as alluded to above, we can attempt to account for temporal dependence in our series of threshold excesses through careful consideration of the extremal index in Equation 1. For the remainder of this section, we briefly outline three extremal index estimators (with a more detailed discussion of these estimators being provided in the Supporting Information S1, for interested readers). Although many estimators of the extremal index have been proposed in the literature (see Ancona-Navarrete and Tawn [2000] for a comprehensive survey), two of the methods we choose (labeled \( \theta_1 \) and \( \theta_2 \) in the discussion below) have corresponding likelihood functions and so lend themselves well to Bayesian inference. We choose a third (\( \theta_3 \)) for comparison purposes; although it does not have an associated likelihood function, it has been shown to be a robust estimator of the extremal index and will be used in our analysis to “benchmark” our Bayesian inference schemes for the extremal index.

1. Using a first-order extreme value Markov chain
Here, we assume a first-order Markov structure for the evolution of our process of threshold excesses; that is, we assume the following joint density for our series of flares:

\[
 f(x_1, x_2, \ldots, x_n) = \prod_{i=2}^{n} f(x_{i-1}, x_i) / \prod_{i=2}^{n-1} f(x_i),
\]

where \( f \) denotes a joint or marginal density function, as appropriate. Appealing to bivariate extreme value theory, transformation from GPD to unit Fréchet margins (see, for example, Coles [2001]) gives a range of models to use for contributions to the numerator in Equation 6, the most commonly-used being the logistic family with CDF:

\[
 G(x_{i-1}, x_i) = \exp\left\{-\left(\frac{x_i^{-1/a}}{\eta_{i-1}} + \frac{x_i^{-1/a}}{\eta_i}\right)^{\alpha}\right\}.
\]

Here, \( \alpha \in (0, 1) \) controls the extent of extremal dependence in the process, with independence giving \( \alpha = 1 \) and \( \alpha \to 0 \) corresponding to increasing levels of extremal dependence. Differentiation of Equation 7, with careful censoring when either one or both of \((x_{i-1}, x_i)\) lies below the threshold, gives pairwise contributions to the numerator in Equation 6; univariate contributions to the denominator are given through the GPD, again with censoring for observations falling below the threshold. For a detailed discussion of the use of bivariate extreme value theory in a Markov chain context, see the Supporting Information S1 for this paper.

Through extensive simulations, Fawcett and Walshaw (2012) show that an approximation to the extremal index can then be obtained via a simple polynomial transformation of \( \alpha \); specifically,

\[
 \theta_1 = 0.013 - 0.092\alpha + 1.833\alpha^2 - 0.756\alpha^3.
\]

The major benefit of this method is that the likelihood takes into account all observations, not just the reduced set of threshold excesses (or, even worse, the set of cluster peak excesses in the case of the POT approach), having accounted for the extremal dependence present between consecutive pairs in the process. It is hoped that this will lead to more precise parameter estimates and, correspondingly, return levels and return periods. However, there are underpinning assumptions relating to the choice of model (i.e., logistic, as outlined above), and model order (first-order, as outlined here), that might warrant investigation (see Section 5). Given these potential concerns, we consider an alternative approach to extremal index estimation, as outlined below.

2. Ancona-Navarre and Tawn’s Maxima method

This method of extremal index estimation requires the consideration of the distribution of block maxima, as discussed in Section 2.1. Letting the sequence of (dependent) observations be divided into \( k \) blocks of length \( T \), we denote by \( M_{ij}, i = 1, \ldots, k \), the block maxima for this dependent sequence. An exchangeable sequence with the same marginal distribution as the dependent sequence can be generated by randomizing the index of the observations. For this associated (independent) sequence, denote the block maxima by \( \tilde{M}_{ij} \). As discussed in Section 2.4, both \( M_{ij} \) and \( \tilde{M}_{ij} \) will be GEV-distributed, with the location and scale parameters for the dependent sequence being functions of the extremal index. Thus, Ancona-Navarre and Tawn (2000) propose a joint likelihood function for \((\mu, \xi, \xi, \theta)\), which can then be maximized or used in a Bayesian setting to make inferences for the extremal index \( \theta \). We refer to this estimator of the extremal index as \( \theta_2 \). Inference for the GPD parameters \((\mu, \xi, \xi)\) can then be discarded, with inference for \( \xi \) being carried through to a standard threshold-based analysis to accompany inference for the GPD scale and shape \((\sigma, \xi)\), aiding return level/return period estimation through Equation 4/5 respectively. More details can be found in the Supporting Information S1.

3. Ferro and Segers’ Intervals estimator

The final method we consider for estimating the extremal index was introduced by Ferro and Segers (2003), and is an empirical estimate based on interexceedance times of observations over \( u \). Suppose we have a sample of size \( n \), with \( N \) observations exceeding \( u \), and let \( 1 = S_1 < \ldots < S_n \leq n \) be the exceedance times. The observed interexceedance times are \( T_i = S_{i+1} - S_i \) for \( i = 1, \ldots, N - 1 \). Then, the intervals estimator, as defined by Ferro and Segers (2003), is
where \( a = b = c = 0 \) if \( \max\{T_i; 1 \leq i \leq N - 1\} \leq 2 \), and \( a = b = 1 \) and \( c = 2 \) otherwise. More details on this extremal index estimator can be found in the Supporting Information S1 for this paper.

3. Bayesian Inference

In Section 2 we outlined the general framework for modeling extremes, with a particular focus on the threshold-based approach and methods for handling temporal dependence within. Our preferred modeling strategy is to press all excesses into use (as opposed to using a declustering scheme to obtain a filtered subset of independent threshold excesses), utilizing an established method for extremal index estimation to capture temporal dependence. Through \( \theta_1 \) or \( \theta_2 \), and the GPD as a marginal model for threshold excesses, inference can proceed via optimization of likelihood functions; in this paper, we choose to make use of these likelihoods within the Bayesian setting.

Extremes are scarce, by their very nature, and so any method we can use to augment an extreme value analysis with extra information could be greatly beneficial. Indeed, through careful construction of prior distributions for our model parameters, Bayesian methods can lend an extra source of information to our analysis, one which cannot be incorporated through a standard likelihood approach to inference (as used in, for example, Tsiftsi and De la Luz (2018)).

Bayesian methods also provide us with much more intuitive interpretations of confidence intervals. For example, in the Bayesian context there is a probability of 0.95 that the GPD shape parameter \( \xi \) lies within the bounds of the associated 95% Bayesian confidence interval (often termed credible interval); the same is not true of the corresponding 95% frequentist confidence interval since, in the frequentist setting, \( \xi \) is considered a fixed constant rather than a random variable. Furthermore, maximum likelihood estimation for the GPD shape parameter \( \xi \) is complicated when \( \xi < -0.5 \); when \( -1 < \xi < -0.5 \) maximum likelihood estimates exist but are non-regular, and when \( \xi < -1 \) the likelihood is unbounded and maximum likelihood fails (see R. L. Smith [1985] for more details).

In the Bayesian setting we have no such issue, as the likelihood is not maximized but is instead used as an ingredient in the inference process. Bayesian inference also extends naturally to prediction via the Bayesian posterior predictive distribution, providing a useful and convenient way to summarize Bayesian inference on return levels and return periods (see, for example, Fawcett and Green [2018]).

Given the discussion above, it is interesting to note that there are relatively few publications showcasing applications of extreme value theory within the Bayesian context. Interested readers are referred to Coles and Powell (1996) for a solid review of Bayesian inference for extremes up to that date. Since then, some notable publications include: a Bayesian analysis of rainfall extremes (Coles & Tawn, 1996), with a follow-up paper by E. L. Smith and Walshaw (2003) in which the construction of informative priors is discussed; examples of objective priors for the GEV and GPD models in Beirlant et al. (2004) and Castellanos and Cabras (2007); papers that attempt to exploit meteorological structure in climate extremes through Bayesian hierarchical modeling, including Davison et al. (2012), Fawcett and Walshaw (2006a) and Sang and Gelfand (2009) and some work by Fawcett and Green (2018) extolling the virtues of the posterior predictive return level.

As discussed, a key advantage of performing an extreme value analysis within the Bayesian context is the ability to augment the standard modeling procedure with external sources of information through carefully constructed prior distributions. For example, E. L. Smith and Walshaw (2003) cite up to an 84% reduction in the posterior standard deviation for estimates of rainfall extremes when switching from a Bayesian analysis using uninformative priors to one with priors informed through discussions with a hydrologist; posterior standard deviations in the analysis using uninformative priors were very similar to the standard errors obtained from a classical analysis based on maximum likelihood estimation. Similarly, the analysis of wind speed extremes in Fawcett and Walshaw (2006a) revealed up to an 82% reduction in uncertainty in estimated wind speed extremes when comparing Bayesian and frequentist approaches to inference. Further, Fawcett and Walshaw (2006a) discuss the ease of fitting their extreme value model under the Bayesian approach, relative to fitting the same model under a
classical approach to inference, with modern statistical computing packages having simple syntax and available routines for the fitting of complex models within the Bayesian paradigm.

As far as the authors are aware, there are no current publications utilizing Bayesian methods for the estimation of solar flare extremes.

In the EVT context, if we denote our set of threshold excesses as $y = (y_1, ..., y_N)$, which follow a GPD, then we wish to make inferences on our model parameters (both marginal and dependence), which we denote generically here by the vector $\psi$. Suppose our prior beliefs about $\psi$ are summarized by the probability density function (PDF) $\pi(\psi)$, and we have likelihood $L(\psi|y)$ for the assumed model for our data. Then Bayes Theorem gives the posterior density for $\psi$ as

$$
\pi(\psi|y) = \frac{\pi(\psi)L(\psi|y)}{f(y)}.
$$

The posterior density in Equation 10 represents our beliefs about $\psi$ after having observed the data $y$. Thus, in the context of an extreme value analysis in which data are sparse, including prior information on $\psi$ has obvious appeal. However, collating—and quantifying—this prior information, can be challenging.

It can also be very difficult to obtain an analytic solution for $f(y)$ which, in turn, can lead to difficulties in obtaining the posterior itself. However, Markov Chain Monte Carlo (MCMC) methods circumvent these issues, providing routines for obtaining a sample of observations from the target posterior distribution. MCMC methods are now widely used in many fields of application; the reader is referred to the seminal paper by A. F. M. Smith and Roberts (1993) for full details. In this study we use the Metropolis-Hastings algorithm with component-wise transitions.

### 3.1. Informative Priors

In the absence of any prior information, it is commonplace to choose non-informative (“diffuse” or “objective”) priors for $\pi(\psi)$, for example, for use in a Metropolis-Hastings algorithm. This completely misses one of the major benefits of working within the Bayesian framework, which is the opportunity to work with experts or use past studies to elicit informative priors. In EVT especially, since the data are usually sparse, doing so might be extremely beneficial.

Even for an expert in their field, it may be a far from intuitive task to propose prior distributions for the model parameters directly. Thus, in an extreme value context we might first construct priors for return levels. Here we use two return levels $z_{r_1}$ and $z_{r_2}$, where $r_1$ and $r_2$ are specified periods and $r_2 > r_1$; then we use a Jacobian transformation in order to obtain a corresponding prior for the GPD parameters $(\sigma, \xi)^T$. In our analysis we are assuming independence between the marginal and dependence components a priori, and so a prior for $\theta$ is constructed separately.

We first note that our return levels must be ordered so that $z_{r_1} < z_{r_2}$, hence we construct priors for the differences

$$
z_1 = z_{r_1} - e_1, \quad \text{and} \quad z_2 = z_{r_2} - z_{r_1},
$$

Where $e_1$ is a notional lower end point for the variable (which we consider to be zero in our application). We assume these are independent, and take priors of the form

$$
z_i \sim \text{Gamma}(g_i, h_i), \quad i = 1, 2,
$$

with

$$
\pi(z_i) = \frac{z_i^{g_i-1} e^{-z_i h_i}}{\Gamma(g_i)}, \quad z_i > 0, \quad g_i, h_i > 0,
$$

and
\begin{equation}
\Gamma (g) = \int_{0}^{\infty} x^{g-1} e^{-x} \, dx.
\end{equation}

We then construct the joint prior for \(\sigma\) and \(\xi\) using

\begin{equation}
\pi (\sigma, \xi) = \pi (z_{r1}, z_{r2}) \times J (\sigma, \xi, z_{r1}, z_{r2})
\end{equation}

where \(\pi (z_{r1}, z_{r2})\) is the prior

\begin{equation}
\pi (z_{r1}, z_{r2}) \propto z_{r1}^{a-1} \exp \left\{ -h_1 z_{r1} \right\} \left( z_{r2} - z_{r1} \right)^{b-1} \exp \left\{ -h_2 \left( z_{r2} - z_{r1} \right) \right\}
\end{equation}

and \(J (\sigma, \xi, z_{r1}, z_{r2})\) is the Jacobian transformation

\[ J (\sigma, \xi, z_{r1}, z_{r2}) = \frac{\partial z_{r1}}{\partial \sigma} \frac{\partial z_{r2}}{\partial \xi} - \frac{\partial z_{r1}}{\partial \xi} \frac{\partial z_{r2}}{\partial \sigma} \]

which can be computed using Equation 4.

For the extremal index, we use \(\theta \sim Beta(a, b)\) since it has support over (0, 1); specifically,

\[ \pi (\theta) = \frac{\theta^{a-1} (1 - \theta)^{b-1}}{B(a, b)}, \quad 0 < \theta < 1, \quad a, b > 0, \]

and

\[ B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a + b)} \]

with \(\Gamma(\cdot)\) as defined in Equation 11. We can then adjust \(a\) and \(b\) to assign weight over values we might expect the extremal index to take.

### 3.2. Predictive Return Levels

Another benefit from using Bayesian methods is the natural extension to prediction through the posterior predictive distribution, given by

\[ \Pr (Y \leq y | x) = \int_{\psi} \Pr (Y \leq y | \psi) \pi (\psi | x) \, d\psi \]

where \(Y\) denotes a future extreme. In order to obtain an estimate of the \(r\)-year predictive return level \(z_{r, pred}\) we must solve

\[ \Pr (Z \leq z_{r, pred} | y) = 1 - (rn_y)^{-1} \]

for \(z_{r, pred}\). As Fawcett and Green (2018) discuss, this is arguably the most useful summary for our inference on return levels, as it provides practitioners with a single point estimate which encapsulates uncertainty in parameter estimation. We can use our MCMC sample \(\psi^{(j)}, j = 1, \ldots, M\), to estimate \(z_{r, pred}\) since

\[ \Pr (Z \leq z_{r, pred} | y) \approx \frac{1}{M} \sum_{j=1}^{M} \Pr (Y \leq z_{r, pred} | \psi^{(j)}) \]

which we can set equal to \(1 - (rn_y)^{-1}\) and solve for \(z_{r, pred}\) numerically.
4. Application

In this section, we apply the methods from Section 3 to our series of solar flares introduced in Section 1. After selecting a suitable threshold, we compare the standard POT analysis (as described in Section 2.4.1) to methods which use all threshold excesses accounting for estimation of the extremal index (as described in Section 2.4.2). We work within the Bayesian context throughout, and we conclude this section by investigating the merits of using informative priors (as discussed in Section 3.1) relative to using non-informative priors; we also report predictive return levels (as described in Section 3.2).

4.1. Threshold Selection

We use a combination of mean residual life plots and parameter stability plots (see, for example, Coles [2001] or Scarrot and MacDonald [2012]) to select an appropriately high threshold for classifying observations as extreme. The mean residual life plot for the flares data set is shown in Figure 2. We propose a threshold at around the 99.5% quantile ($u = 1.1 \times 10^{-4}$ W m$^{-2}$, corresponding to an X1.1 class flare); beyond this point, we observe approximate linearity, which is consistent with the threshold stability property of the GPD. Inspection of parameter stability plots supports this choice of threshold.

4.2. Model Fitting

4.2.1. Removing Dependence: Declustering

Following the method outlined in Section 2.4.1, we decluster our series of flares and fit the GPD to the set of independent cluster peak excesses (implementing the so-called POT method). Of course, we must first select an appropriate declustering interval to separate clusters, from which we will extract the (independent) cluster maxima. Using the threshold of $u = 1.1 \times 10^{-4}$ W m$^{-2}$, as selected in Section 4.1, we investigate the autocorrelation between consecutive cluster peaks across a range of values for $\kappa$, paying particular attention to the smallest value of $\kappa$ above which this autocorrelation becomes statistically insignificant. A plot of the $p$-value for the autocorrelation coefficient plotted for each $\kappa \in (1, 50)$ is given in Figure 3, which shows a clear switch to insignificant autocorrelation at $\kappa = 28$ (using the 5% level of significance as our cut-off). Of course, any value for $\kappa > 28$ would also be appropriate, but our aim here might be to select the optimal value for $\kappa$, in terms of maximizing our data usage.

We implement a step-wise Metropolis-Hastings algorithm in order to obtain posterior samples from $\psi = (\sigma, \xi)^T$, using cluster peak excesses obtained from runs declustering with $\kappa = 28$. We work with the logarithm of the

![Figure 2. Mean residual life plot with 95% confidence interval, blue dashed line is the 99.5% quantile ($1.1 \times 10^{-4}$ W m$^{-2}$, corresponding to an X1.1 class flare).](image)
GPD scale parameter, that is, $\eta = \log \sigma$, so we can implement simple random walk proposals for both GPD parameters whilst retaining the positivity of the scale parameter $\sigma$. For now, as priors we choose a completely non-informative specification, in order to quantify the benefits of working with an informative specification later in Section 4.5; specifically, we use 

$$\eta = \log \sigma \sim N(0, 1000) \quad \text{and} \quad \xi \sim N(0, 1000).$$

We use Normal random walk proposals, that is, at each iteration $j$ in the MCMC scheme the proposal distributions are Normal with means $\eta^{(j-1)}$ and $\xi^{(j-1)}$, and standard deviations $\epsilon_\eta$ and $\epsilon_\xi$, for $\eta$ and $\xi$ respectively. After fine tuning the algorithm, and testing convergence using different starting values, a burn-in of 1,000 is removed and samples of 10,000 realizations from the marginal posteriors for $\sigma$ and $\xi$ are obtained. Optimal values of $\epsilon_\eta$ and $\epsilon_\xi$ are found as 0.45 and 0.38 respectively to optimize the mean acceptance probabilities of 0.24 and 0.23 (see Gelman et al. [1997] for more details on optimizing Metropolis-Hastings proposals).

Note that in practice, it is unusual for a diagnostic such as that shown in Figure 3 (top) to be used to aid the choice of declustering interval; more often than not, $\kappa$ will be chosen arbitrarily. Also shown in Figure 3 (bottom) are posterior means, and corresponding 95% credible intervals, for return periods for a Carrington-type event obtained from cluster peaks identified using a range of values of $\kappa$ (we will consider how to obtain estimates of return periods in Section 4.4). Although the increasing width of the credible intervals with increasing $\kappa$ might be obvious, owing to the inclusion of fewer data points as $\kappa$ increases, this plot is nonetheless informative as it
quantifies the extent to which our uncertainty in return period estimation might be influenced by the choice of declustering interval in the POT approach. And of course, although not considered in the present work, there is uncertainty in our choice of threshold \( u \); potentially leading to sensitivity in return level and return period estimates to any \((\kappa, u)\)-interaction.

Better, perhaps, to adopt an approach that avoids declustering altogether; one which explicitly accounts for dependence, for example.

### 4.2.2. Accounting for Dependence Through the Extremal Index

We now consider modeling all excesses over our energy threshold \( u \), as opposed to a filtered set of independent cluster peak excesses. We follow the methodology described in Section 2.4.2, whereby we account for temporal dependence through estimation of the extremal index.

#### 1. Using a first-order extreme value Markov chain

To implement method (1) in Section 2.4.2 we use the same step-wise Metropolis-Hastings scheme as described in Section 4.2.1. However, using the logistic model with dependence parameter \( \alpha \), we now perform inference for the marginal GPD scale and shape parameters and the extremal index (via Equation 8), giving posterior draws for \( \psi = (\sigma, \xi, \theta) \). At this stage, we retain the non-informative prior specification for the marginal GPD parameters that we used in Section 4.2.1. We also use a non-informative prior specification for \( \alpha \), with \( \alpha \sim U(0,1) \); that is, a priori \( \alpha \) is uniformly-distributed over the range \((0, 1)\), meaning that all values in this range are equally likely (hence the prior being non-informative).

Tuning the algorithm through careful selection of the random walk standard deviations \( \epsilon, \eta, \xi, \alpha \), enabled near-optimal acceptance probabilities of around 0.23 for each parameter.

#### 2. Ancona-Navarrete and Tawn’s maxima method

To implement method (2) in Section 2.4.2 we employ two separate MCMC schemes, both utilizing the same step-wise Metropolis-Hastings algorithm as outlined in Section 4.2.1. In fact, the first MCMC scheme is identical but using all threshold excesses, leading to posterior draws for \( \psi_1 = (\sigma, \xi) \). The second involves a full step-wise Metropolis-Hastings scheme for the GEV parameters and extremal index giving parameter vector \( \psi_2 = (\mu, \xi, \theta_2) \), fitting simultaneously to the set of dependent and independent block maxima \( M_{ij} \) and \( \tilde{M}_{ij} \) (respectively) with \( r \approx k, i = 1, \ldots, k \) (see [2] in Section 2.4.2). From these two separate MCMC runs, we can then combine posterior draws for the GPD marginal parameters with those for the extremal index, to make inferences on \( \psi = (\sigma, \xi, \theta_2) \).

Once again, tuning of both schemes through careful selection of the random walk standard deviations returned pleasing acceptance probabilities for all parameters, in the range 0.23–0.25.

#### 3. Ferro and Segers’ intervals estimator

As mentioned earlier, this estimator has been shown to provide robust estimates of the extremal index, and thus is being used in our analysis as simply a point of reference for approaches (1) and (2) described above. Since this estimator is not likelihood-based, we combine our estimate of \( \theta \) obtained directly from Equation 9 with our MCMC output for the GPD scale and shape under method (2) above; however, for each posterior draw for the pair \((\sigma, \xi)\), \( \theta \) will remain constant. Of course, the uncertainty attached to any inference on return levels or return periods will be under-estimated, as our uncertainty in \( \theta \) is not being accounted for.

### 4.2.3. Results

We now present a comparison of posterior densities for the model parameters \( \sigma, \xi \) and \( \theta \) obtained using each of the methods described above. Figure 4 shows density plots constructed from the MCMC samples.

In line with findings in Fawcett and Walshaw (2012), we see over- and under-estimation of the marginal GPD scale and shape parameters (respectively) for the method using the declustered threshold excesses (giving
posterior means for $\sigma$ and $\xi$ of $1.92 \times 10^{-4}$ and 0.395 respectively), relative to the approaches using all threshold excesses (and accounting for dependence; giving posterior means for $\sigma$ and $\xi$ of $1.35 \times 10^{-4}$ to $1.36 \times 10^{-4}$ and 0.449–0.455 respectively). We also see larger posterior variances for the method using declustered excesses, likely owing to the much-reduced data set being used here. Of course, the method used to obtain posterior samples for $\sigma$ and $\xi$ for the approaches that account for dependence is the same, therefore we expect these posterior densities to be similar (the only difference being due to Monte Carlo error).

Inference for the extremal index shows greatest precision when modeling dependence with the extreme value Markov chain, with posterior mean (and standard deviation) for $\theta_1$ of 0.966 (0.010); again, likely due to the fact that this method presses into use more data points than Ancona-Navarrete and Tawn’s method, with posterior mean (and standard deviation) for $\theta_2$ of 0.941 (0.437). Reassuringly, we also see posterior densities for both $\theta_1$ and $\theta_2$ lying very close to Ferro and Segers’ intervals estimator $\theta_3 = 0.964$, giving us faith in our MCMC output for both of these extremal index estimators. Examining the posterior summaries for $\theta$ clearly reveals weak extremal dependence between consecutive solar flare extremes over our chosen threshold. Our analyses in Sections 4.3 and 4.4 present the effects that modeling this dependence (even when weak) has on estimated return levels and return periods; interested readers can also see the results of a simulation study in this paper’s Supporting Information S1, for a detailed examination of the effects of ignoring, and modeling, extremal dependence on return level estimation in the presence of weak extremal dependence.

**Figure 4.** Posterior densities for $\sigma$ (top), $\xi$ (middle) and $\theta$ (bottom) obtained by declustering (blue), logistic (red), Ancona-Navarrete and Tawn (green), and Ferro and Segers (orange) methods.
4.3. Return Levels

As mentioned in Section 2.3, one of the most important aims of an extreme value analysis is often the estimation of return levels. These results enable practitioners to understand and prepare for events we might expect to see (on average) once every $r$ years. For example, a practitioner could use an estimate of the 100-year return level to prepare for the flare we might see once per century. It is often the case that the upper bound of a 95% confidence interval is used as a “most plausible worst case scenario” estimate of this return level.

We now use the posterior samples obtained in Section 4.2 to construct samples from the posterior distributions for return levels. We do this by applying Equation 4 to $\psi^{(i)}$ at each iteration $j$ in our MCMC procedure to give our corresponding value for $A_A^{(j)}$. We repeat this procedure across a range of values for $r$, and plot $z_r$ against $r$ on a double-logarithmic scale to compress the right-hand-side of the plot and give focus to estimates of $z_r$ for large return periods. Such return level plots are shown in Figure 5. It is worth noting that since the declustering method obtains a sample which is assumed to be independent, we set $\theta = 1$ for the return level calculations via Equation 4. Also shown in the plots are predictive return levels, as described in Section 3.2. In Table 1 we report posterior means with corresponding 95% credible intervals, and predictive return levels, for $z_r$ with $r = 10, 100$ and 1,000 years.

The increase in precision that we noted in Section 4.2.3 for the GPD scale and shape parameters, for the methods using all threshold excesses, can again be seen for return levels; posterior standard deviations for the 10, 100 and 1,000 year flares are consistently smaller for the methods using all threshold excesses, compared to the analysis using just the cluster peak excesses. The under-estimation of return levels when using cluster peak excesses,
relative to the methods which model all threshold excesses, can also be seen; especially when comparing predictive return levels. Interestingly, the predictive return levels lie consistently close to the associated return level posterior means. In some environmental applications it is common to see the predictive return level much higher than the posterior mean, and occasionally exceeding the 95% credible interval upper bound (see, for example, Fawcett and Green [2018]). This is likely due to the fact that our fitted GPD has extremely heavy tails (with $\xi \gg 0$).

### 4.4. Halloween and Carrington Event Return Periods

We now apply Equation 5 to $\psi^{(j)}$ at each iteration $j$ in our MCMC procedure to construct posterior samples for the return periods for Halloween and Carrington-type events (strength X35 and X45 respectively). Results are summarized in the right-hand-side of Table 1.

We see that all three of our methods for quantifying dependence give broadly similar findings, with a 95% credible interval for a Carrington-like event of (19.2, 212.7) years when using the extreme value Markov chain with logistic dependence, for example. The estimated return period under the POT approach, using declustered excesses, is considerably higher and, once again, we see a substantial loss of precision here as a result of pressing fewer extremes into use. For comparison, the declustering approach used by Elvidge and Angling [2018] resulted in a 95% confidence interval for the Carrington return period of (30, 900) years; Tsiftsi and De la Luz [2018] report a 95% confidence interval of (20, 6,500) years.

The methods using all threshold excesses clearly lend greater precision to the analysis than the standard POT approach. We should, however, bear in mind that the results using Ferro and Segers’ estimator for the extremal index will likely have posterior standard deviations that are too small, and credible interval that are too narrow, owing to the fixed estimate of $\theta_3$ being used here. Notwithstanding issues around model choice and model order.
We use the same Metropolis-Hastings MCMC scheme as used throughout this paper so far, to obtain posterior

\[ \alpha \sim \text{Beta}(20, 2). \]

Figure 6 shows the prior distributions for \( z_1 \), \( z_2 \) and \( \alpha \).

We use the same Metropolis-Hastings MCMC scheme as used throughout this paper so far, to obtain posterior samples from our parameter vector \( \psi = (\sigma, \xi, \theta_1)^T \) after substitution of our sample for \( \alpha \) into Equation 8.

In Figure 7 we have plots of the posterior densities for the GPD parameters and extremal index using both informative and non-informative priors. Noteworthy is the increase in posterior precision when using informative priors. In particular, the posterior standard deviation for the shape parameter, \( \xi \), when using informative priors is 0.0365, compared to a standard deviation of 0.0773 when using non-informative priors.

Figure 8 and Table 2 summarize return level inference using both the informative and non-informative priors. Comparing the posterior return level densities in Figure 8, we see that the informative priors lead to much narrower credible intervals, which is likely caused by the smaller posterior variance in \( \xi \). We can also see the effect of the shift in \( \xi \), as the return level means using the informative prior are lower than those using the non-informative prior, with posterior means for the 100-year return level of 0.00484 and 0.00652 respectively.

Table 2 summarizes the means and 95% credible intervals for the Halloween and Carrington-type event return periods, obtained using the logistic method with informative and non-informative priors. Note that the posterior means for both events are higher using the informative prior, where we can expect to see a Halloween and Carrington-type event roughly once (on average) every 49 and 92 years respectively, compared to the non-informative analysis with 38 and 69 years respectively. We also see narrower credible intervals when using the informative priors. The results obtained using the informative priors are more comparable to results found in previous studies, however with much narrower credible intervals and thus of more use to practitioners (see Section 5).

### 5. Conclusions

#### 5.1. General Findings

We have seen that EVT provides us with an effective way of modeling extremes of a data set, without needing to know the underlying distribution of our data and focusing only on the tails. This theory is essential when predicting return periods for events such as the Carrington solar storm, which consists of flares larger than any in our
data set and so requires careful extrapolation. The use of Bayesian inference allows us to factor in prior beliefs regarding return levels, provides posterior distributions which are conveniently encapsulated and naturally interpreted through simple summaries (e.g., means, quantiles) taken from posterior samples, and provides a natural extension to prediction within the Bayesian framework.

Modeling under the assumption of a first-order extreme value Markov chain allowed us to estimate return periods for Halloween and Carrington-type events—with posterior means giving (on average) 49 and 92 years respectively when using the logistic model for extremal dependence with informative priors. These results are similar to those obtained by Elvidge and Angling (2018) and Tsiftsi and De la Luz (2018); for example, Carrington return periods were estimated (through maximum likelihood) as 100 and 110 years respectively. While these return period estimates may be similar to our posterior means, a major improvement can be seen when considering the estimation precision. Elvidge and Angling (2018) and Tsiftsi and De la Luz (2018) obtained 95% confidence intervals of (30, 900) years and (20, 6,500) years respectively for a Carrington-type event, which are substantially wider than the 95% Bayesian credible interval found in our analysis of (50,176) years, obtained using the logistic model with informative priors. Similar gains in precision can be seen in our estimated return period for the Halloween event: Our 95% credible interval is (29, 85) years, compared to (10, 100) years in Elvidge and Angling (2018) and (10, 300) years in Tsiftsi and De la Luz (2018).

This analysis therefore highlights the importance of taking into account several key factors; the first of which is properly modeling extremal dependence. The approach taken by Elvidge and Angling (2018) was to remove dependence using declustering, which we have seen reduces the size of the data set and leads to substantially increased estimation uncertainty. Tsiftsi and De la Luz (2018) argued that the extremal index was close to 1 and so did not need to be accounted for. As presented in the Supporting Information S1, the results of a comprehensive simulation study performed by the authors revealed significant bias in return level estimates when extremal

---

**Figure 6.** Prior densities for \( z_1 \) (top), \( z_2 \) (middle) and \( \alpha \) (bottom), constructed using results from Tsiftsi and De la Luz (2018).
Figure 7. Posterior GPD parameter densities using logistic method with informative (solid) and non-informative (dashed) priors.

Figure 8. Return level means and 95% credible intervals (red) with predictive return level (black) using the logistic method with non-informative priors (left) and informative priors (right).
dependence was not properly accounted for. Although this bias reduced as the strength of dependence diminished, this bias remained non-negligible for processes with weak dependence (θ ≈ 0.95), especially in cases with very heavy tails (i.e., when ξ > 0). Furthermore, both Elvidge and Angling (2018) and Tsiftsi and De la Luz (2018) worked within a maximum likelihood setting, missing the opportunity to supplement their analyses with useful prior information. Doing so in our work allowed us to increase estimation precision still further, for both return levels and return periods.

5.2. Further Considerations

While the model proposed here clearly performs well with regard to the precision of return level/period estimates, there might be room for further improvement. For example, one could investigate properties of the sun to determine potential covariates to incorporate into the model. Also, whilst we did consider simple methods for capturing the cyclic nature of our flares series—and found little evidence that doing so could improve estimates of return levels and return periods—in further work we might consider more sophisticated ways for incorporating the complex nature of the solar cycle in our modeling.

Furthermore, when considering bivariate extreme value models for pairwise likelihood contributions in our extreme value Markov chain, we considered only the (symmetric) logistic model, for which the variables within each pair are exchangeable. Further work might investigate the extent of asymmetry in the dependence structure through comparisons with the bilogistic model (see Joe et al. [1992]). In this analysis we also assume a first-order structure for our Markov chain. However, further work might investigate the plausibility of a higher-order dependence; see, for example, Fawcett and Walshaw (2006b). Simple graphical diagnostics, such as simplex plots (as detailed in Coles and Tawn [1991]), might be used to investigate higher-order dependencies. After creating such a plot for the flares data set, we note that the concentration of points on the graph's interior might indicate that a higher-order dependence assumption is more appropriate than the first-order structure used here.

Finally, we would also like to note the method outlined in Section 3.1 to construct the informative prior. We used a simple mathematical approach to construct our prior based on the discussions in another paper. However, our intention in future work is to investigate methods for constructing a prior based on the outcome of elicitation sessions with experts in the field, perhaps drawing upon the expertise of astrophysicists in our department. Our results in this paper might be viewed as a proof-of-concept for showcasing the benefits of using an informative prior, rather than best practice for constructing such priors.

5.3. Shiny Application

The methods discussed in this paper are available as part of an easy-to-use web-based tool, performing Bayesian extreme value threshold-based analyses. This application is part of a web-based framework using the Shiny

---

Table 2

|                | z_{10}    | z_{100}   | z_{1000}  | Halloween (X35) | Carrington (X45) |
|----------------|-----------|-----------|-----------|-----------------|------------------|
| **Informative** |           |           |           |                 |                  |
| Posterior mean  | 0.00183   | 0.00484   | 0.0122    | 48.82           | 91.9             |
| Standard deviation | 1.33 × 10^{-4} | 6.15 × 10^{-4} | 0.00242 | 14.16           | 32.2             |
| 95% credible interval | (0.001 58, 0.002 11) | (0.003 71, 0.006 15) | (0.008 0, 0.017 4) | (29.45, 84.70) | (50.4, 175.7) |
| Prediction      | 0.00183   | 0.00488   | 0.0124    | -               | -                |
| **Non-informative** |           |           |           |                 |                  |
| Posterior mean  | 0.00209   | 0.00652   | 0.0202    | 37.65           | 68.8             |
| Standard deviation | 3.94 × 10^{-4} | 0.00237 | 0.0119    | 24.47           | 55.5             |
| 95% credible interval | (0.001 51, 0.003 00) | (0.003 47, 0.012 39) | (0.007 5, 0.050 1) | (12.80, 102.22) | (19.2, 212.7) |
| Prediction      | 0.00209   | 0.00657   | 0.0213    | -               | -                |

---

Halloween (X35) Carrington (X45)
package on the statistical software package R-Studio, which can be run locally on R-Studio or through our dedicated shiny server. Interested readers should contact the authors to obtain log-in credentials for the application.

The tool allows the user to choose between a selection of pre-loaded datasets (including the GOES peak flux solar flares data discussed in Section 1.2) and manual data upload. Interactive data visualizations can then be used to guide the user through analyses, including those outlined in Section 2, allowing the replication of results from Section 4. A Bayesian MCMC scheme can be performed, accounting for dependence using declustering or the extremal index, allowing for user-specified prior beliefs and parameter tuning. This results in easy to interpret (predictive) return level visualizations and outputs.

Data Availability Statement

The data set of peak flux measurements, recorded by the NOAA from the GOES satellite program, can be found at https://ngdc.noaa.gov/stp/satellite/goes/.

Acknowledgments

We would like to thank the School of Mathematics, Statistics and Physics at Newcastle University for supporting Ben Griffiths’ work through a summer internship in July 2021. We are also indebted to two referees for suggesting considerable improvements to the paper.

References

Afraimovich, E., Zherebtsov, G., & Smol’kov, G. (2007). Total failure of gps during a solar flare on December 6, 2006. Doklady Earth Sciences, 417, 1231–1235. https://doi.org/10.1134/S1028334X07080223

Ancona-Navarrete, M. A., & Tawn, J. A. (2000). A comparison of methods for estimating the extremal index. Extremes, 3, 5–38. https://doi.org/10.1023/A:1009993419559

Beirlant, J., Goegebeur, J., Teugels, J., Segers, J., De Waal, D., & Ferro, C. (2004). Statistics of extremes. Wiley.

Bortot, P., Coles, S. G., & Tawn, J. A. (2000). The multivariate Gaussian tail model: An application to oceanographic data. Journal of the Royal Statistical Society, Series C (Applied Statistics), 49(1), 31–49. https://doi.org/10.1111/1467-9876.00177

Carrington, R. C. (1859). Description of a singular appearance seen in the sun on September 1, 1859. Monthly Notices of the Royal Astronomical Society, 20(1), 13–15. https://doi.org/10.1093/mnras/20.1.13

Castellanos, M. E., & Cabras, S. (2007). A default Bayesian procedure for the generalized pareto distribution. Journal of Statistical Planning and Inference, 137(2), 473–483. https://doi.org/10.1016/j.jspi.2006.01.006

Cheng, L., AghaKouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. Climatic Change, 127, 353–369. https://doi.org/10.1007/s10584-014-1254-5

Cheng, N., Song, S., & Xie, H. (2019). Investigation of solar flares impact on gps/bds/galileo broadcast ionospheric models. Radio Science, 54(1), 91–103. https://doi.org/10.1029/2018RS006591

Clover, E. W., & Dietrich, W. F. (2013). The 1859 space weather event revisited: Limits of extreme activity. Journal of Space Weather and Space Climate, 3(A31). https://doi.org/10.1051/swsc/2013053

Coles, S. G. (2001). An introduction to statistical modelling of extreme values. Springer. https://doi.org/10.1007/978-1-4471-3675-0

Coles, S. G., Heffernan, J., & Tawn, J. A. (1999). Dependence measures for extreme value analyses. Extremes, 2, 339–365. https://doi.org/10.1023/A:1009963131610

Coles, S. G., & Powell, E. A. (1996). Bayesian methods in extreme value modelling: A review and new developments. International Statistical Review, 64(1), 119–136. https://doi.org/10.2307/1403426

Coles, S. G., & Tawn, J. A. (1991). Modelling extreme multivariate events. Journal of the Royal Statistical Society: Series B, 53(2), 377–392. https://doi.org/10.1111/j.2517-6161.1991.tb01830.x

Coles, S. G., & Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. Journal of the Royal Statistical Society, Series C (Applied Statistics), 45(4), 463–478. https://doi.org/10.1037/23096068

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modelling of spatial extremes. Statistical Science, 27(2), 161–186. https://doi.org/10.1214/11-STS376

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. Journal of the Royal Statistical Society: Series B, 52(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

Einmahl, J. J., Einmahl, J. H. J., & de Haan, L. (2019). Limits to human life span through extreme value theory. Journal of the American Statistical Association, 114(527), 1075–1089. https://doi.org/10.1080/01621459.2018.1357912

Elvidge, S., & Angling, M. J. (2018). Using extreme value theory for determining the probability of Carrington-like solar flares. Space Weather, 16(4), 417–421. https://doi.org/10.1029/2017SW001727

Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. North American Actuarial Journal, 3(2), 30–41. https://doi.org/10.1002/10922077.1999.10595797

Fawcett, L., & Green, A. (2018). Bayesian posterior predictive return levels for environmental extremes. Stochastic Environmental Research and Risk Assessment, 32, 2233–2252. https://doi.org/10.1007/s00477-018-1561-x

Fawcett, L., & Walshaw, D. (2006a). A hierarchical model for extreme wind speeds. Journal of the Royal Statistical Society, Series C (Applied Statistics), 55(5), 631–646. https://doi.org/10.1111/j.1467-9876.2006.00557.x

Fawcett, L., & Walshaw, D. (2006b). Markov chain models for extreme wind speeds. Environmetrics, 17(8), 795–809. https://doi.org/10.1002/env.794

Fawcett, L., & Walshaw, D. (2012). Estimating return levels from serially dependent extremes. Environmetrics, 23(3), 272–283. https://doi.org/10.1002/env.2133

Ferro, C. A. T., & Segers, J. (2003). Inference for clusters of extreme values. Journal of the Royal Statistical Society: Series B, 65(2), 545–556. https://doi.org/10.1047/j.1467-9868.2004.00401

Fishers, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society, 24(2), 180–190. https://doi.org/10.1017/S0305004100015681

Gelman, A., Gilks, W. R., & Roberts, G. O. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. Annals of Applied Probability, 7(1), 110–120. https://doi.org/10.1214/aoap/1034625254
Gilli, M., & Kellezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27, 207–228. https://doi.org/10.1007/s10614-006-9025-7

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of Mathematics, Second Series*, 44(3), 423–453. https://doi.org/10.2307/1968974

Guerra, J., Pulkinen, A., & Urtitsky, V. (2015). Ensemble forecasting of major solar flares: First results. *Space Weather, 13*(10), 626–642. https://doi.org/10.1002/2015SW001195

Hagood, M., Angling, M., Attrill, G., Bisi, M., Cannon, P., Dyer, C., & Willis, M. (2021). Development of space weather reasonable worst-case scenarios for the UK national risk assessment. *Space Weather, 19*(4), 1–32. https://doi.org/10.1029/2020SW002593

Hodgson, R. (1859). On a curious appearance seen in the sun. *Monthly Notices of the Royal Astronomical Society*, 20(1), 15–16. https://doi.org/10.1093/mnras/20.1.15a

Holmes, J. D., & Moriarty, W. W. (1999). Application of the generalized pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 81(1–3), 1–10. https://doi.org/10.1016/S0167-6105(99)00056-2

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348), 158–171. https://doi.org/10.1002/qj.49708134804

Joe, H., Smith, R. L., & Weissman, I. (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society: Series B*, 54(1), 171–183. https://doi.org/10.1111.2517-6161.1992.tb01871.x

Jonas, S., Murtagh, W., & Bonadonna, M. (2017). Released for public comment: Space weather benchmarks and operations-to-research plan. *Space Weather, 15*(2), 282. https://doi.org/10.1029/2017SW001603

Kappenman, J. G. (2003). Storm sudden commencement events and the associated geomagnetically induced current risks to ground-based systems at low-latitude and mid-latitude locations. *Space Weather, 1*(3). https://doi.org/10.1029/2003SW000009

Kawabata, Y., Iida, Y., Doi, T., Akiyama, S., Yashiro, S., & Shimizu, T. (2018). Statistical relation between solar flares and coronal mass ejections with respect to sigmoidal structures in active regions. *The Astrophysical Journal*, 869(99), 1–16. https://doi.org/10.3847/1538-4357/aaebfc

Leadbitter, M. R., & Rootzén, H. (1988). Extrremal theory for stochastic processes. *Annals of Probability*, 16(2), 431–478. https://doi.org/10.1214/aop/1176991767

Liu, L., Wan, W., Chen, Y., & Le, H. (2011). Solar activity effects on the ionosphere: A brief review. *Chinese Science Bulletin*, 56(12), 1202–1211. https://doi.org/10.1049/ijn4.010.4226-9

Lu, E. T., & Hamilton, R. J. (1991). Avalanches and the distribution of solar flares. *The Astrophysical Journal*, 380, L89-L92. https://doi.org/10.1086/161180

Meredith, N. P., Horne, R. B., Isles, J. D., & Rodriguez, J. V. (2015). Extreme relativistic electron fluxes at geosynchronous orbit: Analysis of GOES E > 2 MeV electrons. *Space Weather, 13*(3), 170–184. https://doi.org/10.1002/2014SW001143

Mitra, A. P. (1974). *Avalanches and the distribution of solar flares*. Springer.

Pulkkinen, A., Lindahl, S., Viljanen, A., & Pirjola, R. (2005). Geomagnetic storm of 29-31 October 2003: Geomagnetically induced currents and their relation to problems in the Swedish high-voltage power transmission system. *Space Weather, 3*(8). https://doi.org/10.1029/2004SW000123

Riley, P. (2012). On the probability of occurrence of extreme space weather solar events. *Space Weather, 10*(2). https://doi.org/10.1029/2011SW000734

Riley, P., Baker, D., Liu, Y. D., Verronen, P., Singer, H., & Guidel, M. (2017). Extreme space weather events: From cradle to grave. *Space Science Reviews*, 214(21). https://doi.org/10.1007/s11214-017-0456-3

Riley, P., & Love, J. J. (2017). Extreme geomagnetic storms: Probabilistic forecasts and their uncertainties. *Space Weather, 15*(1), 53–64. https://doi.org/10.1002/2016SW001470

Roberts, S. J. (2000). Extreme value statistics for novelty detection in biomedical data processing. *IEEE Proceedings—Science, Measurement and Technology*, 147(6), 363–367. https://doi.org/10.1049/ip-smt:2000841

Roodman, D. (2015). The risk of geomagnetic storms to the grid: A preliminary review. SSRN. https://doi.org/10.2139/ssrn.3635861

RStudio, Inc. (2013). Easy web applications in R [Computer software manual]. [Dataset]. RStudio. Retrieved from http://www.rstudio.com/shiny/

Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *Revista de Estatistica*, 10(1), 33–60.

Smith, A. F. M., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 55(1), 3–23. https://doi.org/10.1111/j.2517-6161.1993.tb01466.x

Smith, E. L., & Walshaw, D. (2003). Modelling bivariate extremes in a region. *Bayesian Statistics*, 7, 681–690.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67–90. https://doi.org/10.1093/biomet/72.1.67

Smith, R. L., Tawn, J. A., & Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2), 249–268. https://doi.org/10.1093/biomet/84.2.249

Thomson, A. W. P., Dawson, E. B., & Reay, S. J. (2011). Quantifying extreme behavior in geomagnetic activity. *Space Weather, 9*(10). https://doi.org/10.1029/2011SW000696

Tsitisi, T., & De la Luz, V. (2018). Extreme value analysis of solar flare events. *Space Weather, 16*(12), 1984–1996. https://doi.org/10.1029/2018SW001958

von Mises, R. (1954). *La distribution de la plus grande de n valeurs*. Selected Papers (Vol. II, pp. 271–294). American Mathematical Society.

Young, I. R., Vinoth, J., Zieger, S., & Babanin, A. V. (2012). Investigation of trends in extreme value wave height and wind speed. *Journal of Geophysical Research, 117*(C11). https://doi.org/10.1029/2011JC007753