Better forecasting of extreme geomagnetic storms using non-stationary statistical models

Ting Wang[1], David Fletcher[2], Matthew Parry[1], Craig J. Rodger[3], Andy W. Smith[4], Tanja Petersen[5]

[1]Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand
[2]David Fletcher Consulting Limited, Karitane, New Zealand
[3]Department of Physics, University of Otago, Dunedin, New Zealand
[4]Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Newcastle upon Tyne, UK
[5]GNS Science, Lower Hutt, New Zealand

Key Points:

- Generalized extreme value distribution with constant parameters cannot capture time-varying features of geomagnetic data.
- We use statistical models with parameters varying over time, capturing key features of geomagnetic data, such as effects of the solar cycle.
- We fit these models to New Zealand Eyrewell magnetic observatory data to obtain estimates of return levels over the next 50–500 years.

Corresponding author: Ting Wang, ting.wang@otago.ac.nz

Abstract

An assessment of the risk of extreme geomagnetic storms is critically important for modern society. However, current methods mainly focus on using stationary statistical models to analyze extreme geomagnetic events. These models ignore the non-stationary nature of the data, caused by effects of the solar cycle and the seasons, and thus could provide unreliable estimates of return levels. We propose use of hidden Markov models and generalized additive models, both involving time-varying parameters, in order to capture these features of the data. We use these models to analyze extreme values of the magnitude of the derivative in the horizontal component $R_1(t)$ of geomagnetic observations from the Eyrewell geomagnetic observatory in New Zealand. We use residual diagnostics to check for lack-of-fit of the models, demonstrate that they can successfully model the effects of both the solar cycle and the seasons, and use the best-fitting models to provide more reliable estimates of return levels. From our analysis, the 50-year and 100-year conditional return levels of the extreme magnitude of the derivative in the horizontal component $R_1(t)$ at the Eyrewell magnetic observatory are likely to be within the ranges $500 - 2600$ nT/min and $700 - 4500$ nT/min respectively at solar maximum.

Plain Language Summary

Use of data from magnetic observatories to estimate the possible extreme magnitude of geomagnetic storms in the next 50–500 years is critically important, in order to allow the mitigation of hazards caused by these storms. Current methods of analysis mainly use stationary extreme value models, which cannot capture time-varying features in the data. We use geomagnetic data observed at the Eyrewell magnetic observatory in New Zealand to show that such a stationary model can exhibit strong lack-of-fit and may therefore provide unreliable forecasts of the possible extreme magnitude of geomagnetic storms in the future. We propose use of statistical models that incorporate time-varying parameters, demonstrate that these models can thereby capture important features of the data, particularly the solar cycle, and use them to provide forecasts of the possible future extreme magnitude of geomagnetic storms at Eyrewell.

1 Introduction

Modern society is increasingly dependent on technology, including electrical power grids. These grids can be disrupted and potentially damaged during extremely large geomagnetic storms (e.g., Molinski, 2002; Mac Manus et al., 2022). Being able to forecast the ground impact of the next extreme geomagnetic storm is therefore crucial. Several studies have focussed on using stationary extreme value models to estimate the most likely extreme ground magnetic field variability/rate of change/perturbation expected over a specified time period (e.g., Thomson et al., 2011; Wintoft et al., 2016; Love, 2020; Rogers et al., 2020, 2021). Use of such models involves fitting an extreme value distribution to either "block maxima" or "exceedances over a threshold". The former refers to the maximum value observed within a specified time period ("block"), such as annual maxima (e.g., Nikitina et al., 2016; Wintoft et al., 2016; Love, 2020; Fogg et al., 2023). The latter refers to the use of all observed values that exceed a specified threshold (e.g., Tsubouchi & Omura, 2007; Thomson et al., 2011; Love, 2020; Rogers et al., 2020, 2021). Estimates of the parameters of the extreme value distribution are then used to estimate the return level for a specified period, e.g. the geomagnetic storm severity expected to be exceeded once within the next 100 years.

We focus on modelling block-maxima for the following reasons. Firstly, block-maxima uses equal time-spacing between consecutive observations, which facilitates the use of some of the tools we apply to evaluate the fit of the model (see Section 3.3). Secondly when the original data exhibits serial dependence, the exceedances over a threshold may also be dependent. Modelling exceedances then requires the identification of "clusters" of exceedances, with only the maximum within each cluster being used in the modelling. If the model parameters also need to change over time, i.e. we require a "non-stationary" model, the choice of threshold may also need to change over time.

In fitting models to geomagnetic storm data using the block-maxima approach, a block size is typically chosen to be large enough for the extreme value distribution to provide a good approximation to the true sampling distribution of the maxima (Coles, 2001). The block-size needed to achieve this will increase if there is serial dependence in the observations within a block (Huang et al., 2021). In addition, the block-size should be large enough to make it reasonable to assume that the resulting maxima are independent. Conversely, too large a block-size may make the final sample size (number of maxima) too small to provide precise estimates of the parameters of the extreme value distribution. For example, if we use annual maxima of geomagnetic data, the sample size will often be less than 50, as records of continuous geomagnetic measurements are typically shorter than 50 years. The resulting lack of precision in the estimates of the parameters of the extreme value distribution may then lead to estimates of return levels with very wide 95% confidence intervals. For example, the magnetic observatory at Eyrewell in New Zealand has fewer than 30 years of digital records, and the use of annual maxima, although being required in order to meet the modelling assumptions, leads to 95% confidence intervals for the return levels being too wide for practical use (see Section 4.1).

As well as needing to assume independence between block maxima, or between exceedances, stationary extreme value models assume that the observations are identically distributed through time, i.e. that their distribution is stationary. The exception to this is Elvidge (2020), who used the Hilbert-Huang transform to identify the solar cycles in annual maxima of aa indices over a 150-year period, from 1868. Elvidge (2020) split the time period into "solar maximum" and "solar minimum" segments, corresponding to periods of higher and lower solar activity, and fitted the GEV distribution to the annual aa-index maxima, separately for the two types of period. These maxima were therefore assumed to be independent within each type of period. The only allowance for non-stationarity arose from the parameters of the GEV being different for the two types of period.

Given the existence of time-varying features in geomagnetic data, such as the effects of the solar cycle and the seasons, we would expect use of non-stationary extreme value models, which allow the parameters to vary over time, to lead to more reliable estimates of return levels. In particular, with the availability of digital geomagnetic measurements at a sampling rate of one observation per minute, it is important to use models that can provide a good fit to maxima based on the smaller block-sizes, in order to increase the sample-size. Use of non-stationary models should help in this respect, as they should provide a better fit to such maxima.

In this study, we use hidden Markov models (HMMs) and generalised additive models (GAMs) to analyze geomagnetic field data (one observation per minute) made at the Eyrewell observatory in Canterbury, New Zealand over the period 1994 to 2019, i.e., 26 years. An HMM has parameters changing as a step function and a GAM has parameters varying smoothly. We compare these models and show that they can capture the main features of block-maxima data better than models with constant parameters, particularly when smaller block sizes are used. We provide forecasts of the return levels for different periods, using the best-fitting HMM and GAM models.

2 Data

The Eyrewell magnetic observatory is located in Canterbury, New Zealand. It is a mid-latitude (-43.474 °N, 172.393 °E) observatory that is part of the INTERMAGNET global network of observatories. The geomagnetic data we use in this study consists of observations of the horizontal X (north) and Y (east) components of the Earth's magnetic field recorded every minute from 1994 to 2019 (GNS-Science, 1994). The data have a strong diurnal trend, with both the X and Y components rapidly rising and falling during dawn and dusk (respectively), while being relatively stable during the day. During periods of geomagnetic disturbance, the variance of the observations increases, as larger storms produce larger variation in these horizontal components.

There are gaps in the record, some of which are several days long, caused by equipment maintenance or malfunction. For the purpose of our analyses, we used a method for filling in gaps in the Eyrewell data set that is outlined in the Supplement.

The time derivative of the magnetic field, that is, the first difference ($X_t - X_{t-1}$ or $Y_t - Y_{t-1}$) in the two horizontal directions X_t and Y_t, is related to the electric field driving Geomagnetic Induced Current (GIC, Viljanen et al., 2001; Mac Manus et al., 2017; Rodger et al., 2017; Smith et al., 2024). We model the extreme values of the magnitude of this difference, i.e., the rate of change of the horizontal component of the surface magnetic field, which is given by $R_1(t) = \sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2}$ (Freeman et al., 2019; Smith et al., 2019, 2022).

3 Methods

A common distribution used to model block maxima is the generalized extreme value (GEV) distribution, with probability density function

$$f(x) = \exp\left\{ -\left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\} \frac{1}{\sigma} \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-1/\xi - 1}, \tag{1}$$

where μ, σ, and ξ are the location, scale and shape parameters respectively. This probability density function is undefined if we set $\xi = 0$, but in the limit as $\xi \to 0$ it becomes the probability density function for the Gumbel distribution, which is given by $f(x) = \exp\{-\exp[-(x-\mu)/\sigma]\} \exp[-(x-\mu)/\sigma]/\sigma$. In addition, when $\xi < 0$, the probability density function is only defined for $x \leq \mu - \sigma/\xi$, while for $\xi > 0$, it is only defined for $x \geq \mu - \sigma/\xi$.

A return level z_p is the value of the maxima which has a probability p of being exceeded, and is therefore expected to be exceeded once every $1/p$ periods, where the period is the time-scale associated with the block maxima. For the GEV distribution this return level can be written as

$$z_p = \mu - \frac{\sigma}{\xi}\left\{1 - [-\log(1-p)]^{-\xi}\right\}, \tag{2}$$

and when $\xi = 0$ we have $z_p = \mu - \sigma \log\{-\log(1-p)\}$. See Coles (2001) for a useful and detailed introduction to extreme value analysis. It is important to note that the choice of block-size will affect the estimate and 95% confidence interval for a return level. We use the statistical software R for all calculations and model-fitting (R Core Team, 2024).

3.1 Hidden Markov model with GEV distribution

Hidden Markov models (HMMs) are widely used to model time series that exhibit non-stationarity and dependence (e.g., Wang & Bebbington, 2013; Wang et al., 2017, 2018; Stanislavsky et al., 2020). In an HMM, the distribution of an observation at any time point depends upon the state of an underlying (hidden) Markov chain at that point. In our setting, we again assume that this distribution is the GEV distribution, but we allow the parameters (location, scale and shape) to vary with time, by being dependent upon the state of the Markov chain. Mathematically, we assume that the observed process is $\{X_t, t = 1, 2, 3, \ldots\}$, and the hidden process $\{C_t, t = 1, 2, 3, \ldots\}$ is a first-order Markov chain, where C_t takes values from $\{1, \ldots, m\}$, which may represent periods of solar activity of different intensities. The probability of the Markov chain being in state j at time t, given the previous states is $Pr(C_t = j \,|\, C_{t-1} = i, \ldots, C_1) = Pr(C_t = j \,|\, C_{t-1} = i) = \gamma_{ij}$. The matrix $\Gamma = (\gamma_{ij})_{m\times m}$ is called the transition probability matrix. For a stationary HMM, we have $\pi = \pi\Gamma$, where $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$ is called the stationary distribution of the Markov chain with $\pi_i = Pr(C_t = i)$. Given C_t, the distribution of the observed process X_t depends only on the current state C_t of the Markov chain and not on any previous observations or states, i.e., $f(x_t \,|\, C_t = j, C_{t-1}, \ldots, C_1, X_{t-1}, \ldots, X_1) = f(x_t \,|\, C_t = j)$, where $f(\cdot)$ is the probability density function of X_t. In our setting, we assume that when the Markov chain is in state c_t at time t, X_t has a GEV distribution, with probability density function

$$f(x_t \,|\, C_t = c_t) = \exp\left\{-\left[1 + \xi_{c_t}\left(\frac{x_t - \mu_{c_t}}{\sigma_{c_t}}\right)\right]^{-1/\xi_{c_t}}\right\} \frac{1}{\sigma_{c_t}} \left[1 + \xi_{c_t}\left(\frac{x_t - \mu_{c_t}}{\sigma_{c_t}}\right)\right]^{-1/\xi_{c_t}-1} \tag{3}$$

where μ_{c_t}, σ_{c_t} and ξ_{c_t} are the state-dependent parameters. When c_t takes on the value i at time t, the observation X_t at time t is from a GEV distribution with parameters μ_i, σ_i and ξ_i.

Intuitively, an HMM classifies the observations over time into m distinct classes (e.g., periods of maximum, moderate, and minimum solar activity). The unobserved Markov chain C_t models the transitions between the different intensities of solar activity, with its states representing these intensities.

Let $\theta = \{\gamma_{ij}, \mu_i, \sigma_i, \xi_i : i = 1, \cdots, m; j = 1, \cdots, m\}$. Given the observed block maxima $x_1, x_2, \cdots, x_n$, the likelihood function of the HMM is

$$\begin{aligned} & f(x_1, x_2, \cdots, x_n \,|\, \theta) \\ = & \sum_{c_1, \cdots, c_n = 1}^{m} \pi_{c_1} f(x_1 \,|\, C_1 = c_1; \mu_{c_1}, \sigma_{c_1}, \xi_{c_1}) \prod_{t=2}^{n} \gamma_{c_{t-1}, c_t} f(x_t \,|\, C_t = c_t; \mu_{c_t}, \sigma_{c_t}, \xi_{c_t}) \end{aligned} \tag{4}$$

which includes a sum of m^n terms, each term consisting of a product of $2n$ factors. See Zucchini and MacDonald (2009) for more details about how to calculate this sum. We use the R function `nlm` to numerically minimize the negative log likelihood function and obtain the parameter estimates $\hat{\theta} = \{\hat{\gamma}_{ij}, \hat{\mu}_i, \hat{\sigma}_i, \hat{\xi}_i : i = 1, \cdots, m; j = 1, \cdots, m\}$(Zucchini & MacDonald, 2009; Allen & Wang, 2024). The overall estimate of the return-level for each period $1/p$ is the solution to

$$\sum_{i=1}^{m} \pi_i F(z \,|\, C = i) = 1 - p \tag{5}$$

that is, the $1-p$th quantile of the mixture of the m GEV cumulative distribution functions $F(z \,|\, C = i)$, where π_i is the estimated stationary distribution of the Markov chain, and

$$F(z \,|\, C = i) = \exp\left\{ - \left[1 + \hat{\xi}_i \left(\frac{z - \hat{\mu}_i}{\hat{\sigma}_i} \right) \right]^{-1/\hat{\xi}_i} \right\}. \tag{6}$$

After obtaining the parameter estimates $\hat{\theta}$, we can find the most likely sequence of states of the Markov chain that has given rise to the observed time series, given the fitted model; that is, we can find$\{\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_n\}$ that maximizes the conditional probability $Pr(C_1 = c_1, C_2 = c_2, \ldots, C_n = c_n \,|\, X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, \hat{\theta})$, which can provide the most likely classification of the time series data. This is done by using the Viterbi algorithm (Viterbi, 1967; Zucchini & MacDonald, 2009) and the resulting sequence of states is called the Viterbi path. The return-level for each period $1/p$, conditional on the Markov chain being in state i, is

$$\hat{z}_{pi} = \hat{\mu}_i - \frac{\hat{\sigma}_i}{\hat{\xi}_i} \left\{ 1 - [-\log(1-p)]^{-\hat{\xi}_i} \right\}. \tag{7}$$

We can use this to obtain the return levels conditional on the most likely sequence of states of the Markov chain.

Choosing the appropriate number of hidden states m in an HMM is non-trivial and is an area of active research. A recent study suggests that sample size is a factor affecting which information criterion is more suitable for estimating the number of hidden states (Buckby et al., 2023), with the Bayesian Information Criterion (BIC) being useful for larger samples sizes and Akaike's Information Criterion (AIC) being appropriate for smaller sample sizes (Buckby et al., 2023), where

$$\text{BIC} = -2 \times \text{log-likelihood} + \log(n)k, \tag{8}$$

and

$$\text{AIC} = -2 \times \text{log-likelihood} + 2k, \tag{9}$$

log-likelihood refers to the log likelihood function evaluated at the maximum likelihood estimates of the parameters, k is the number of parameters in the model, and n is the sample size. For the hourly and daily maxima we use BIC to determine the number of hidden states, while for the weekly and monthly maxima we use AIC.

3.2 Generalised additive model with GEV distribution

Generalised additive models (GAMs) are commonly used to smoothly model the relationship between a response variable and one or more predictors. We use GAMs to model the change over time in each parameter of the GEV distribution. Smoothness is achieved using smoothing splines (Wood, 2017), with the degree of smoothness being determined by the data. Mathematically, we assume that the t^{th} maximum (X_t) has a GEV distribution, with probability density function given by the time-varying analogue of Equation 1, namely

$$f(x_t) = \exp\left\{ -\left[1 + \xi_t \left(\frac{x_t - \mu_t}{\sigma_t}\right)\right]^{-1/\xi_t} \right\} \frac{1}{\sigma_t} \left[1 + \xi_t \left(\frac{x_t - \mu_t}{\sigma_t}\right)\right]^{-1/\xi_t - 1} \tag{10}$$

where μ_t, σ_t, and ξ_t are the location, scale and shape parameters for time t ($t = 1, 2, \ldots, n$). For the hourly maxima, time-variation in the location parameter is modelled by the equation

$$\mu_t = a_h\{h(t)\} + a_d\{d(t)\} + a_w\{w(t)\} + a_m\{m(t)\} + a_y\{y(t)\}, \tag{11}$$

where $h(t)$, $d(t)$, $w(t)$, $m(t)$, and $y(t)$ are the *hour* $(0-23)$, *day* $(1-7)$, *week* $(1-52)$, *month* $(1-12)$ and *year* $(1-26)$ associated with time-point t, and $a_h\{\cdot\}$, $a_d\{\cdot\}$, $a_w\{\cdot\}$, $a_m\{\cdot\}$, and $a_y\{\cdot\}$ are smoothing splines (Wood, 2017, Section 5.1). Likewise, the scale and shape parameters are modelled by the equations

$$\log \sigma_t = b_h\{h(t)\} + b_d\{d(t)\} + b_w\{w(t)\} + b_m\{m(t)\} + b_y\{y(t)\} \tag{12}$$

and

$$\xi_t = c_h\{h(t)\} + c_d\{d(t)\} + c_w\{w(t)\} + c_m\{m(t)\} + c_y\{y(t)\}, \tag{13}$$

where $b_h\{\cdot\}$, $b_d\{\cdot\}$, $b_w\{\cdot\}$, $b_m\{\cdot\}$, $b_y\{\cdot\}$, $c_h\{\cdot\}$, $c_d\{\cdot\}$, $c_w\{\cdot\}$, $c_m\{\cdot\}$, and $c_y\{\cdot\}$ are smoothing splines. We specify the model for the scale parameter in terms of $\log(\sigma_t)$, as this parameter must be positive. As there are slightly more than 52 weeks per year, we arbitrarily assign the extra day(s) to week 52.

For daily maxima we use analogous equations to those for hourly maxima, with separate splines for *day*, *week*, *month*, and *year*. For weekly maxima we use separate splines for *week* and *year*, but not *month*, as some weeks occur in more than one month. Finally, for monthly maxima we use separate splines for *month* and *year*. For all maxima, we use cyclic penalised cubic splines for *hour*, *day*, *week* and *month*, in order to allow the splines to match at the end-points of their cycle (Wood, 2017, Section 5.3.2). For *year*, we use thin-plate regression splines, which have good general properties (Wood, 2017, Section 5.5.1).

We use the `mgcv` package (Wood, 2015) to fit the models, including the option that allow coefficients in the model to be set to zero if the data suggest that this improves the fit. This obviates the need to compare models (using AIC, for example) with different degrees of smoothing.

3.3 Checking the fit of a model

We can determine a suitable block-size by checking for goodness-of-fit of the model for the corresponding maxima. We require that the distribution of the maxima be close to what we would expect from the model, and that the maxima do not exhibit any strong dependence. We use residual diagnostics to check these two aspects of model-fit. For the stationary models and the non-stationary GAM models, we use quantile residuals (Dunn & Smyth, 1996), estimated using the *pevd* function in the R package `extRemes` (Gilleland & Katz, 2016). For the HMM models we also use quantile residuals, which are known as "ordinary pseudo residuals" in the HMM literature (Zucchini & MacDonald, 2009; Buckby et al., 2020). The distribution of quantile residuals will be approximately the same as a standard normal distribution if the data are generated from the fitted model. We check this using a quantile-quantile plot (QQ-plot) tailored to the case where the underlying distribution is assumed to be a standard normal (Dodge, 2008). This plot shows the relationship between the sorted residuals and their expected values if the residuals come from a standard normal distribution. It includes a 95% envelope, within which 95% of the points should lie if the data were generated by the fitted model. In addition, we check for any dependence in the block-maxima using a plot of the residual-autocorrelation function (ACF-plot) (Brockwell & Davis, 2016, Section 1.4.1). This shows the autocorrelation in the residuals at each of several lags and also includes a 95% envelope, within which 95% of the autocorrelations should lie if the data were generated by the fitted model.

For both types of model, goodness-of-fit is initially checked by seeing whether at least 95% of the points in the QQ-plot, and 95% of the autocorrelations in the ACF-plot, lie within the corresponding 95% envelope. If this condition is satisfied we then check whether the strength of any of the autocorrelations outside the 95% envelope might raise concerns with the model, e.g. a single strong positive autocorrelation at lag 1.

3.4 Estimates and confidence intervals for return levels

Given the fitted GEV distribution associated with any state (HMM models) or time-point (GAM models), we can estimate a return-level associated with that distribution. As this estimate is a non-linear function of the model parameters (Equation 2), we use single-fit bootstrapping (Fletcher & Jowett, 2022) to calculate a 95% confidence interval for the return-level. This bootstrap-based method uses the large-sample result that the sampling distribution of the estimates of the parameters (location, log-scale, and shape) is multivariate normal (Coles, 2001). It results in a "bootstrap-distribution" of plausible values for the return-level, with the 2.5^{th} and 97.5^{th} percentiles of this distribution providing a 95% confidence interval. For the GAM models, an overall bootstrap distribution is obtained by aggregating the bootstrap distributions associated with each time-point. For the HMM models, an overall bootstrap distribution is obtained by solving Equation 5, replacing the parameters by each set of bootstrap estimates of the parameters. Use of an overall estimate and 95% confidence interval involves the assumption that the data for the period 1994–2019 is representative of geomagnetic behaviour in the future, a point which we return to later.

4 Results

4.1 Effect of block size

We first consider fit of the stationary GEV model to block maxima for a range of block sizes (daily, weekly, monthly and annual). Estimates of the return levels for different periods and block sizes are shown in Figure 1. There was no evidence of lack-of-fit in the QQ-plots, as seen in Figure 2. However, there was strong evidence of autocorrelation in the residuals when using a block size other than annual. In addition, as there are only 26 annual maxima, the 95% confidence intervals for the return levels based on the GEV distribution for these maxima were wide enough to contain negative values, even though $R_1(t)$ cannot be negative (Figure 1). Persistent positive autocorrelation can be caused by a trend in the time series (Brockwell & Davis, 2016, Section 1.4.1). These results suggest that it would be preferable to use non-stationary models, as these are likely to capture the trend and show less evidence of autocorrelation in the residuals for the smaller block-sizes, which should in turn lead to better confidence intervals for the return levels. Note that for the daily, weekly, and monthly block sizes the estimates and confidence intervals in Figure 1 are not to be taken at face value, as the assumptions on which they are based (independent and identically distributed block maxima) were violated.

4.2 HMM results

We fitted HMMs with different numbers of states to the hourly, daily, weekly and monthly maxima of the data from Eyrewell. For both the hourly and daily maxima, use of BIC led to a six-state model. For the weekly and monthly maxima, use of AIC led to a four-state and a three-state model respectively. There was evidence of lack-of-fit for the models fitted to both hourly and daily maxima (see Supplement Figure S1), whereas for the weekly and monthly maxima the residual diagnostic plots in Figure 3 (a,b) (for

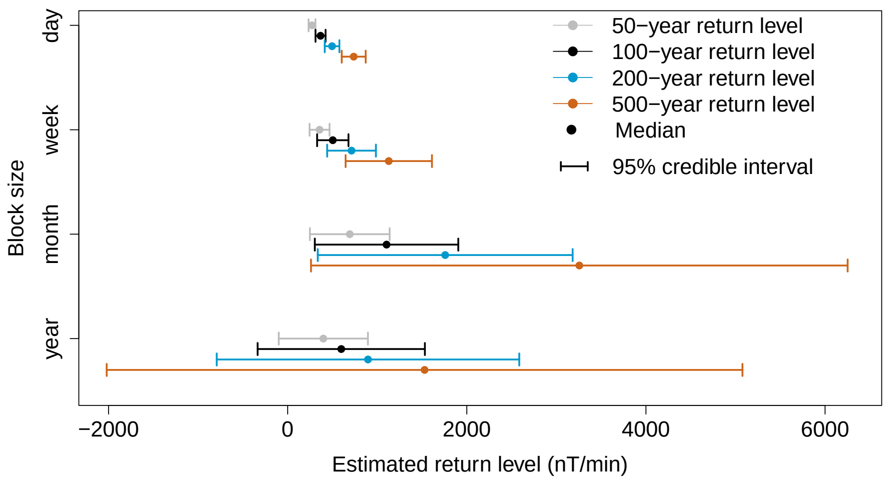


Figure 1. Estimates of return levels for R_1 (nT/min) using the GEV models fitted to block maxima with different block sizes.

weekly maxima) and Supplement Figure S1 (for monthly maxima) suggest that the models provide a good fit to the data.

The estimated parameters for each state along with their 95% confidence intervals are also provided in Table 1. Note that the states are ordered by the value of the location parameter. State 1 most likely corresponds to background noise while state 4 corresponds to the period with strong solar activity. States 2 and 3 have medium location, shape and scale parameters and most likely correspond to periods when solar activity is stronger than the background noise level or weaker than the solar maximum. For the model fitted to weekly maxima, the estimates of each of the parameters over time clearly exhibit non-stationarity (Figure 4 (a–c)). The step functions at each time point in this figure indicate the parameter estimates from the most likely state (obtained using the Viterbi algorithm (Zucchini & MacDonald, 2009)) at that time point. The corresponding estimates of the return levels are shown in Table 2, while those based on monthly maxima are in Supplement Table S1. For both of these maxima, the confidence intervals become wider for the larger block size, as expected, due to the reduction in sample size. Figure 5 shows the estimated return levels over time, conditional on the estimates of the parameters at that time (c.f. Figure 4 (a–c)). The return levels in Figure 5 are seen to be related to solar activity, with very high levels occurring during strong solar activity around 2002 and 2015; and the weaker solar maxima in 2015 compared to

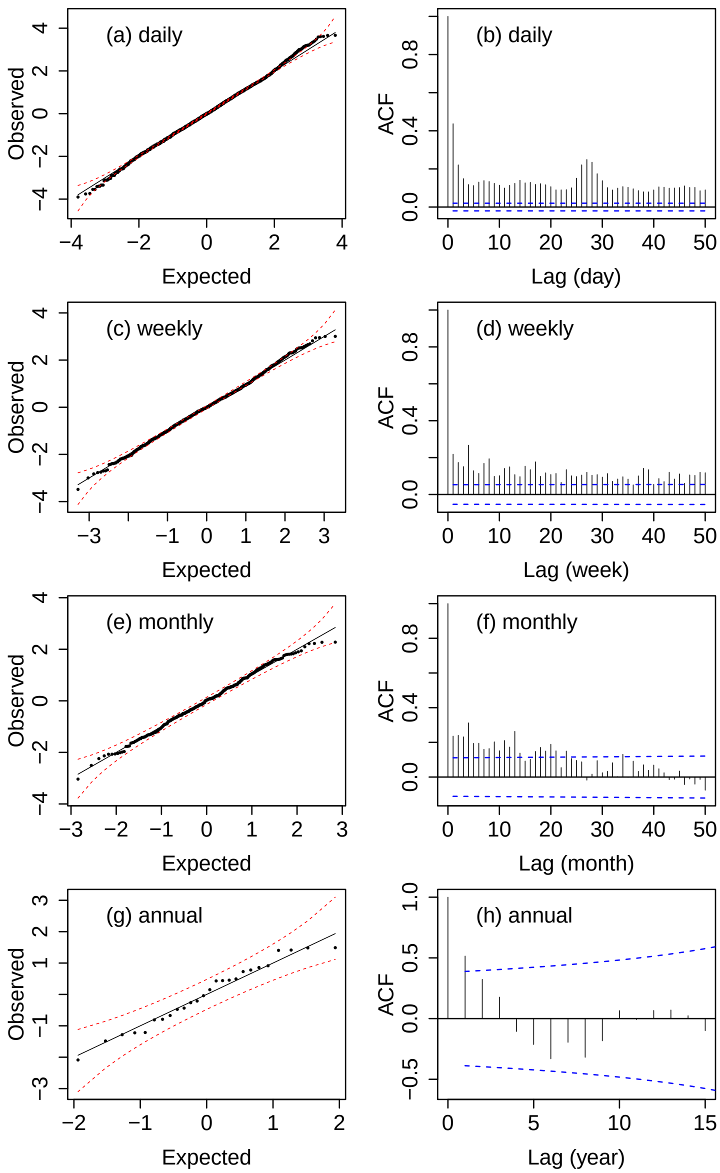


Figure 2. Residual diagnostics for the GEV models fitted to block maxima with different block sizes: (a,c,e,g) QQ-plots of residuals, with black dots showing the quantiles of the residuals (observed) from the model versus the expected quantiles from a standard normal distribution, the diagonal black line indicating a perfect match, and red dashed lines indicating the 95% envelope; (b,d,f,h) ACF plots, showing the autocorrelations, with red dashed lines indicating the 95% envelope.

2002 (e.g., in terms of sunspot number) is also reflected by the mostly lower return levels around 2015 than 2002. This could be particularly important for operational risk forecast, as it may be preferable to use conditional return levels when solar activity is high.

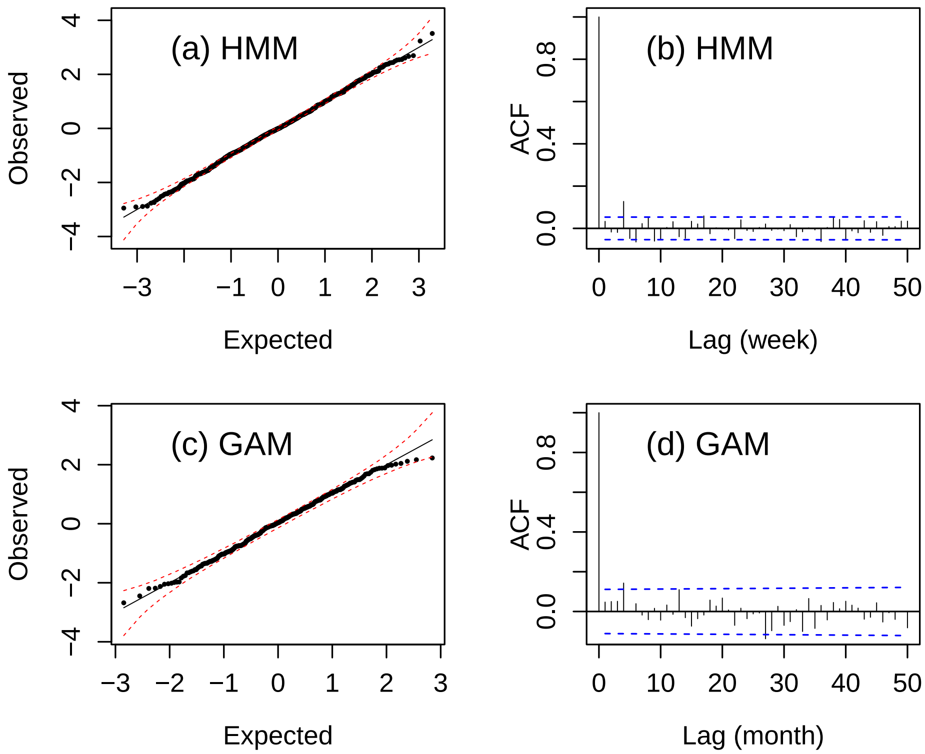


Figure 3. Residual diagnostics for the HMM model fitted to weekly maxima and for the GAM model fitted to monthly maxima. (a,c) QQ-plot of the residuals, with black dots showing the quantiles of the residuals (observed) from the model versus the expected quantiles from a standard normal distribution, the diagonal black line indicating a perfect match, and red dashed lines indicating the 95% envelope; (b,d) ACF plots, showing the autocorrelations, with red dashed lines indicating the 95% envelope.

Although the risk of an extreme event is usually expressed in terms of return levels, it may also be helpful to express it in terms of the probability of the quantity of interest exceeding a specified value. Figure 6 shows the estimated probability of exceeding 500 nT/min in each year, conditional upon the estimate of the most likely state sequence obtained from the HMM model using the Viterbi algorithm for the weekly maxima. This probability was obtained using

$$Pr(M_i > z) = 1 - \prod_{j=1}^{52} Pr(M_{ij} \leq z) \tag{14}$$

where M_i is the maximum in year i, and M_{ij} is the maximum in week j of year i. We can see the high probability of exceedance during the solar maxima around 2002 and 2015, compared to the low probabilities elsewhere.

Table 1. Estimates of the GEV parameters for the best-fitting HMM for weekly maxima, together with 95% confidence intervals.

State i	$\hat{\mu}_i$	$\hat{\sigma}_i$	$\hat{\xi}_i$
$i = 1$	5.37 (5.06, 5.68)	2.05 (1.82, 2.31)	0.27 (0.17, 0.37)
$i = 2$	6.49 (6.10, 6.87)	3.16 (2.80, 3.56)	0.55 (0.45, 0.66)
$i = 3$	7.24 (6.89, 7.59)	2.82 (2.56, 3.11)	0.20 (0.11, 0.28)
$i = 4$	10.46 (9.51,11.42)	6.15 (5.30, 7.10)	0.60 (0.48, 0.73)

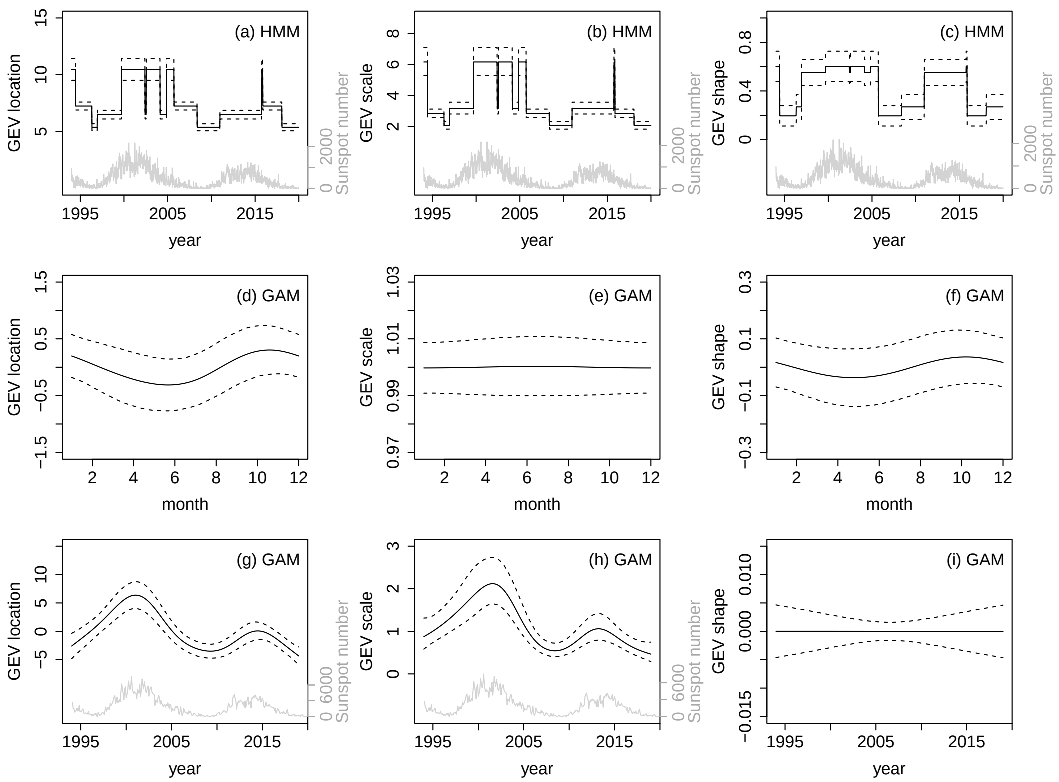


Figure 4. (a–c) Estimates of location, scale and shape parameters over time for the HMM fitted to weekly maxima; (d–i) Effects of both month (d–f) and year (g–i) on the location, scale, and shape parameters for the GAM model fitted to monthly maxima. For the GAM model, the effects for location and shape are additive and centred on zero, while those for scale are multiplicative and centred on one. The dashed lines indicate 95% confidence intervals. The grey lines in (a–c) show the weekly sum of sunspot number and those in (g–h) show monthly sum of sunspot number, with the axis on the right-hand side of each plot (data from `https://www.sidc.be/SILSO/datafiles`).

Table 2. Estimates of return levels for R_1 (nT/min) from the best-fitting HMM for weekly maxima, and the best-fitting GAM for monthly maxima, together with 95% confidence intervals.

Period	HMM (weekly)	GAM (monthly)
50-year	552 (347, 972)	401 (127, 1355)
100-year	830 (495, 1576)	599 (173, 2200)
200-year	1249 (705, 2565)	894 (236, 3585)
500-year	2146 (1126, 4898)	1517 (354, 6871)

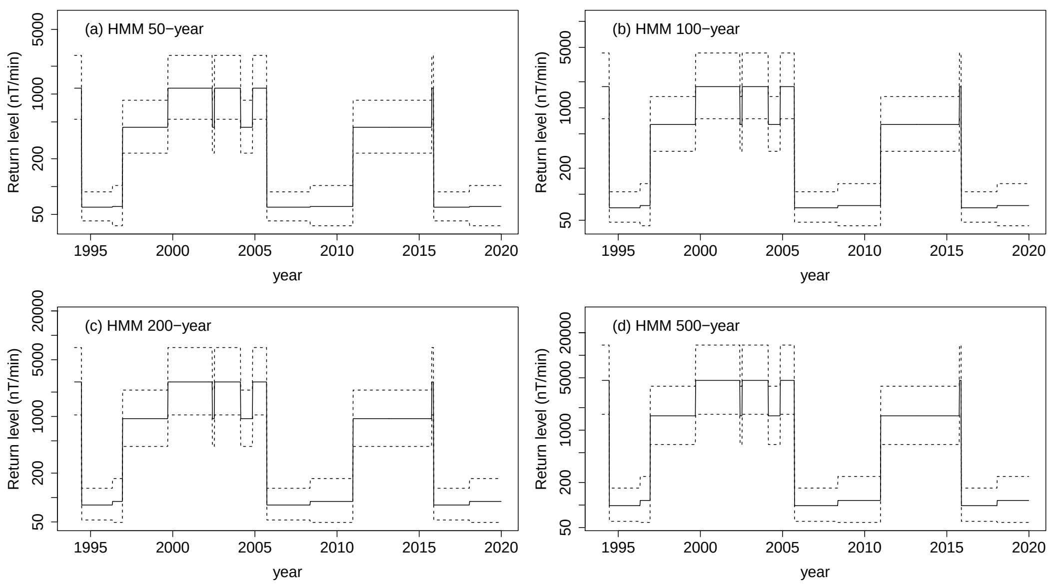


Figure 5. Estimated conditional return levels over time for R_1 (nT/min) using the HMM fitted to weekly maxima.

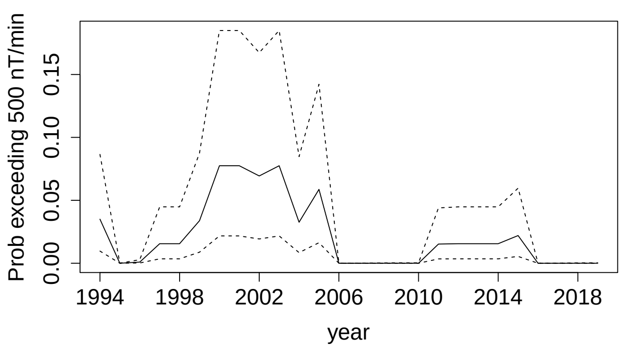


Figure 6. Estimated probability of exceeding 500 nT/min in each year using the HMM fitted to weekly maxima conditional on the Markov chain being in the most likely state sequence obtained using the Viterbi algorithm.

4.3 GAM results

There was evidence of lack-of-fit for the GAM models fitted to the hourly, daily, and weekly maxima (Supplement Figure S2). The GAM fitted to monthly maxima showed no problem with lack-of-fit (Figure 3 (c,d)). For the model fitted to monthly maxima, the temporal variation in the parameters is shown in Figure 4 (d–i). We can see a clear relationship with the solar cycle in the models for location and scale, and a relationship with the seasons in those for location and shape. Unlike the best-fit HMM model (Figure 4 a-c), the best-fit GAM model prefers an effectively constant shape parameter. This is presumably due to the two models using different maxima: the HMM is based on weekly maxima, and is therefore able to make use of many more observations than the GAM model, which is based on monthly maxima. It is generally harder to estimate the shape parameter than the location and shape, and the reduction in sample size when going from weekly to monthly maxima is enough for the GAM model to resort to a very simple, almost stationary, model for the shape parameter (Coles, 2001). When the GAM model is based on weekly maxima, there is a clear relationship with the solar cycle for the shape parameter. However, as that model exhibited lack-of-fit, we do not include those results here.

As both the GAM and HMM models showed no lack-of-fit to the monthly maxima, we can use AIC to compare them. The GAM model had a lower AIC value, with the AIC model weights (Fletcher, 2018) being 0.99 and 0.01 for GAM and HMM respectively. The estimates of the return levels obtained from this GAM model are shown in Table 2. They are similar to those obtained using the HMM fitted to weekly maxima, with the confidence intervals being wider due to the smaller sample size. Figure 7 shows the estimated return levels over time, conditional on the estimates of the parameters at that time (c.f. Figure 4 (d–f)). We can again see the strong solar effect on these return levels, similar to that observed for the HMM fitted to the weekly maxima, with the effects captured by the GAM being smoother than those from the HMM.

4.4 Discussion

The HMM and GAM models that we fitted to the geomagnetic data at Eyrewell appear to capture the main features of the data, with the effects of the solar cycle and the seasons clearly reflected in the temporal variation of the parameters and conditional

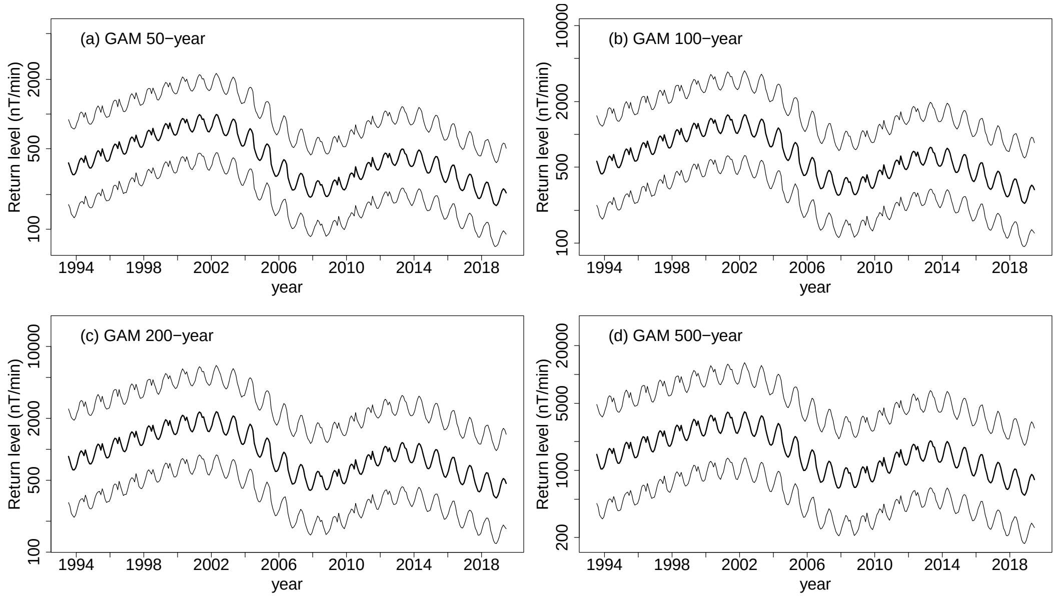


Figure 7. Estimated conditional return levels over time for R_1 (nT/min) using the GAM model fitted to monthly maxima.

return levels of the GAM model (Figures 4 and 7). For the HMM, the time-variation in the parameters and conditional return levels is less smooth than for the GAM model, making it more difficult to see the solar cycles, but we can still see some trend over time, corresponding to the solar maxima and minima. This difference in the temporal variation of the parameters potentially contributes to the different confidence intervals for the return levels (Table 2). The ability to allow for time-varying parameters led to the HMM and GAM models having less evidence of autocorrelation in the residuals, thereby allowing use of a smaller block-size than was possible with the stationary GEV model.

Our analysis indicates that the 50-year, 100-year, 200-year and 500-year return levels of the extreme magnitude of the derivative in the horizontal component $R_1(t)$ at the Eyrewell magnetic observatory are within the following ranges: 120 – 1400 nT/min, 200 – 2200 nT/min, 250 – 4000 nT/min and 350 – 7000 nT/min, respectively. Note that these numbers use the largest range from Table 2.

The length of the digital record available at Eyrewell (1994–2019) is very short, in terms of solar cycles. This is particularly important as the Sun has been relatively inactive for the past 20 years, and the maximum activity only occurs every $\sim$11 years. Our 100- and 200-year return levels are much lower than the UK worst-case scenario (Hapgood et al. (2021): 4000–5000 nT/min for 100-years, 8000–9000 nT/min for 200-years,). Given the period of data that we used, our estimated return levels may represent the lower end of what would be expected, due to the dataset being dominated by a relatively quiet solar cycle. For worst-case operational risk forecast purposes, it could be particularly important to use the conditional return levels (the return levels conditional on the most likely sequence of states of the Markov chain) when solar activity is high. From our analysis, the 50-year, 100-year, 200-year and 500-year conditional return levels of the extreme magnitude of the derivative in the horizontal component $R_1(t)$ at the Eyrewell magnetic observatory are likely to be within the ranges: 500 – 2600 nT/min, 700 – 4500 nT/min, 1000 – 7000 nT/min, 1600 – 14000 nT/min, respectively at solar maximum. Note that these numbers are from the conditional return levels of the best-fitting HMM from Figure 5. An HMM describes discrete stepwise shifts in the trend compared to a GAM which allows for smooth trends. Therefore to capture a few discrete states corresponding to, for example, maximum, moderate and minimum solar activities which is the case for our discussion here, an HMM works better than a GAM.

We have chosen not to compare the estimated return levels from the best-fitting HMM and GAM models with those from the stationary models or the other HMM and GAM models, as these models all showed clear evidence of lack-of-fit. We also note that the confidence intervals of the parameters and return levels have been calculated using bootstrap methods. One could also specify the return level as a parameter in the model and obtain a profile likelihood confidence interval for it (Coles, 2001, Section). This would allow one to constrain the return level to be non-negative, and hence avoid the problem in Figure 1. The formula for the return level in Equation 2 shows that this problem does not arise if the bootstrapped parameter values satisfy the constraint $\mu \geq \sigma/\xi$. Both the HMM for weekly maxima and the GAM for monthly maxima lead to bootstrapped values for the parameters that do satisfy this constraint, and so provide estimates and confidence limits for the return levels that are always positive.

Figure 8 shows the relationship between the weekly total sunspot number and the weekly maxima of R_1. Though there appears to be a positive relation between the weekly maxima of R_1 on log scale (Figure 8(b)) and the weekly total sunspot number, a linear regression between $\log(R_1)$ and the weekly total sunspot number explains only 9% of the total variance in $\log(R_1)$. Future research could incorporate other potential predictors of R_1, such as SymH and AL indices.

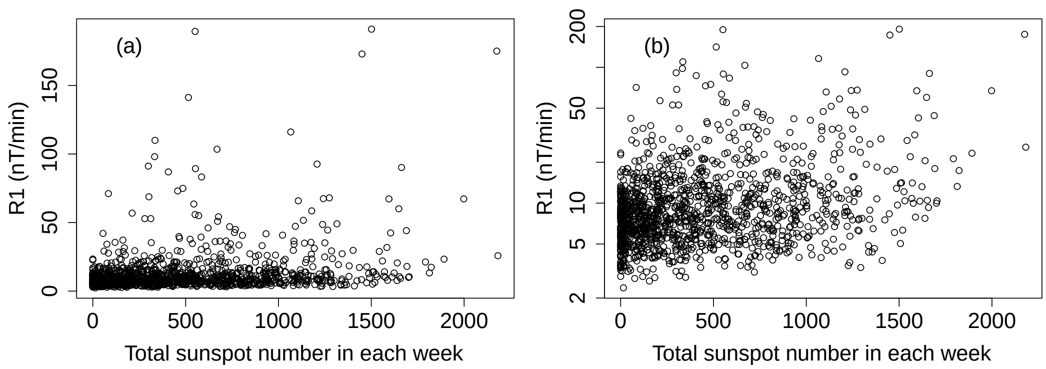


Figure 8. Weekly maxima of R_1 versus weekly total sunspot number. (a) original scale, (b) R_1 on log scale.

5 Conclusion

We have demonstrated that stationary extreme value models cannot capture the non-stationary features of geomagnetic data. We have proposed the use of HMM or GAM models to capture such features, including effects of the solar cycle and the seasons. We fitted these models to extreme values of the magnitude of the derivative in the horizontal component $R_1(t)$ of geomagnetic observations made at Eyrewell magnetic observatory in New Zealand. These data were clearly not independent nor identically distributed, and the non-stationary models therefore provided more reliable forecasts of the corresponding return levels. We have shown that such models can capture non-stationary features, with the GAM clearly modelling the solar cycle and the HMM capturing it more coarsely. With fewer than 100 years of data from most observatories around the world, we recommend using models such as HMMs and GAMs, rather than stationary GEV models for block maxima.

Acknowledgments

This project is supported by New Zealand MBIE Solar Tsunamis Endeavour Programme (UOOX2002). We thank the New Zealand eScience Infrastructure (NeSI) for providing computational resources.

6 Open Research

Geomagnetic data at the Eyrewell magnetometer station can be downloaded from `https://imag-data.bgs.ac.uk/GIN_V1/GINForms2`, using the "Bulk data" option and download the definitive data from observatory EYR (Eyrewell, New Zealand) since 1 January 1994, last accessed in March 2021.

All R code for the analysis and plots is available in Zenodo `https://doi.org/10.5281/zenodo.15347234` (Wang, 2025).

References

Allen, J., & Wang, T. (2024). Hidden Markov models for low-frequency earthquake recurrence. *The American Statistician*, *78*(1), 100–110.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting.* Springer.

Buckby, J., Wang, T., Fletcher, D., Zhuang, J., Takeo, A., & Obara, K. (2023). Finding the number of latent states in hidden Markov models using information criteria. *Environmental and Ecological Statistics*, *30*(4), 797–825.

Buckby, J., Wang, T., Zhuang, J., & Obara, K. (2020). Model checking for hidden markov models. *Journal of Computational and Graphical Statistics*, *29*(4), 859–874.

Coles, S. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). Springer.

Dodge, Y. (2008). *The concise encyclopedia of statistics.* Springer New York.

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and graphical statistics*, *5*(3), 236–244.

Elvidge, S. (2020). Estimating the occurrence of geomagnetic activity using the Hilbert-Huang transform and extreme value theory. *Space Weather*, *18*(8), e2020SW002513.

Fletcher, D. (2018). *Model averaging.* Springer.

Fletcher, D., & Jowett, T. (2022). Single-fit bootstrapping: A simple alternative to the delta method. *Methods in Ecology and Evolution*, *13*(6), 1358–1367.

Fogg, A. R., Jackman, C. M., Malone-Leigh, J., Gallagher, P., Smith, A. W., Lester, M., ... Waters, J. E. (2023). Extreme value analysis of ground magnetometer observations at Valentia observatory, Ireland. *Space Weather*, *21*(7), e2023SW003565.

Freeman, M. P., Forsyth, C., & Rae, I. J. (2019). The influence of substorms on extreme rates of change of the surface horizontal magnetic field in the United

Kingdom. *Space Weather*, *17*(6), 827–844.

Gilleland, E., & Katz, R. W. (2016). extremes 2.0: an extreme value analysis package in r. *Journal of Statistical Software*, *72*, 1–39.

GNS-Science. (1994). Definitive 1-minute Eyrewell (EYR) data [Data set]. *GNS Science, INTERMAGNET*. doi: https://doi.org/10.21420/VTW2-AE75?x=y

Hapgood, M., Angling, M. J., Attrill, G., Bisi, M., Cannon, P. S., Dyer, C., … others (2021). *Development of space weather reasonable worst-case scenarios for the uk national risk assessment.* Wiley Online Library.

Huang, W. K., Monahan, A. H., & Zwiers, F. W. (2021). Estimating concurrent climate extremes: A conditional approach. *Weather and Climate Extremes*, *33*, 100332.

Love, J. J. (2020). Some experiments in extreme-value statistical modeling of magnetic superstorm intensities. *Space Weather*, *18*(1), e2019SW002255.

Mac Manus, D. H., Rodger, C. J., Dalzell, M., Thomson, A. W., Clilverd, M. A., Petersen, T., … Divett, T. (2017). Long-term geomagnetically induced current observations in New Zealand: Earth return corrections and geomagnetic field driver. *Space Weather*, *15*(8), 1020–1038.

Mac Manus, D. H., Rodger, C. J., Ingham, M., Clilverd, M. A., Dalzell, M., Divett, T., … Petersen, T. (2022). Geomagnetically induced current model in New Zealand across multiple disturbances: Validation and extension to non-monitored transformers. *Space Weather*, *20*(2), e2021SW002955.

Molinski, T. S. (2002). Why utilities respect geomagnetically induced currents. *Journal of atmospheric and solar-terrestrial physics*, *64*(16), 1765–1778.

Nikitina, L., Trichtchenko, L., & Boteler, D. (2016). Assessment of extreme values in geomagnetic and geoelectric field variations for canada. *Space Weather*, *14*(7), 481–494.

R Core Team. (2024). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rodger, C. J., Mac Manus, D. H., Dalzell, M., Thomson, A. W., Clarke, E., Petersen, T., … Divett, T. (2017). Long-term geomagnetically induced current observations from New Zealand: Peak current estimates for extreme geomagnetic storms. *Space Weather*, *15*(11), 1447–1460.

Rogers, N. C., Wild, J. A., Eastoe, E. F., Gjerloev, J. W., & Thomson, A. W. P. (2020). A global climatological model of extreme geomagnetic field fluctuations. *J. Space Weather Space Clim.*, *10*, 5.

Rogers, N. C., Wild, J. A., Eastoe, E. F., & Hübert, J. (2021). Climatological statistics of extreme geomagnetic fluctuations with periods from 1 s to 60 min. *Space Weather*, *19*(11), e2021SW002824.

Smith, A. W., Freeman, M. P., Rae, I. J., & Forsyth, C. (2019). The influence of sudden commencements on the rate of change of the surface horizontal magnetic field in the United Kingdom. *Space Weather*, *17*(11), 1605–1617.

Smith, A. W., Rodger, C. J., Mac Manus, D. H., Forsyth, C., Rae, I. J., Freeman, M. P., ... Dalzell, M. (2022). The correspondence between sudden commencements and geomagnetically induced currents: Insights from New Zealand. *Space Weather*, *20*(8), e2021SW002983.

Smith, A. W., Rodger, C. J., Mac Manus, D. H., Rae, I. J., Fogg, A. R., Forsyth, C., ... Dalzell, M. (2024). Sudden commencements and geomagnetically induced currents in New Zealand: Correlations and dependance. *Space Weather*, *22*(1), e2023SW003731.

Stanislavsky, A., Nitka, W., Małek, M., Burnecki, K., & Janczura, J. (2020). Prediction performance of hidden Markov modelling for solar flares. *Journal of Atmospheric and Solar-Terrestrial Physics*, *208*, 105407.

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)

Tsubouchi, K., & Omura, Y. (2007). Long-term occurrence probabilities of intense geomagnetic storm events. *Space Weather*, *5*(12).

Viljanen, A., Nevanlinna, H., Pajunpää, K., & Pulkkinen, A. (2001). Time derivative of the horizontal geomagnetic field as an activity indicator. In *Annales geophysicae* (Vol. 19, pp. 1107–1118).

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, *13*(2), 260–269.

Wang, T. (2025). *GEV-HMM: Forecasting extreme geomagnetic storms using GEV-HMMs and GAMs.* Retrieved from `https://doi.org/10.5281/`

zenodo.15347234

Wang, T., & Bebbington, M. (2013). Identifying anomalous signals in gps data using hmms: An increased likelihood of earthquakes? *Computational Statistics & Data Analysis*, *58*, 27–44.

Wang, T., Zhuang, J., Buckby, J., Obara, K., & Tsuruoka, H. (2018). Identifying the recurrence patterns of nonvolcanic tremors using a 2-D hidden Markov model with extra zeros. *Journal of Geophysical Research: Solid Earth*, *123*(8), 6802–6825.

Wang, T., Zhuang, J., Obara, K., & Tsuruoka, H. (2017). Hidden Markov modelling of sparse time series from non-volcanic tremor observations. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *66*(4), 691–715.

Wintoft, P., Viljanen, A., & Wik, M. (2016). Extreme value analysis of the time derivative of the horizontal magnetic field and computed electric field. *Annales Geophysicae*, *34*(4), 485–491.

Wood, S. (2015). Package 'mgcv'. *R package version*, *1*(29), 729.

Wood, S. (2017). *Generalized additive models: an introduction with R.* chapman and hall/CRC.

Zucchini, W., & MacDonald, I. L. (2009). *Hidden Markov models for time series: an introduction using R.* Chapman and Hall/CRC.

Figure 1.

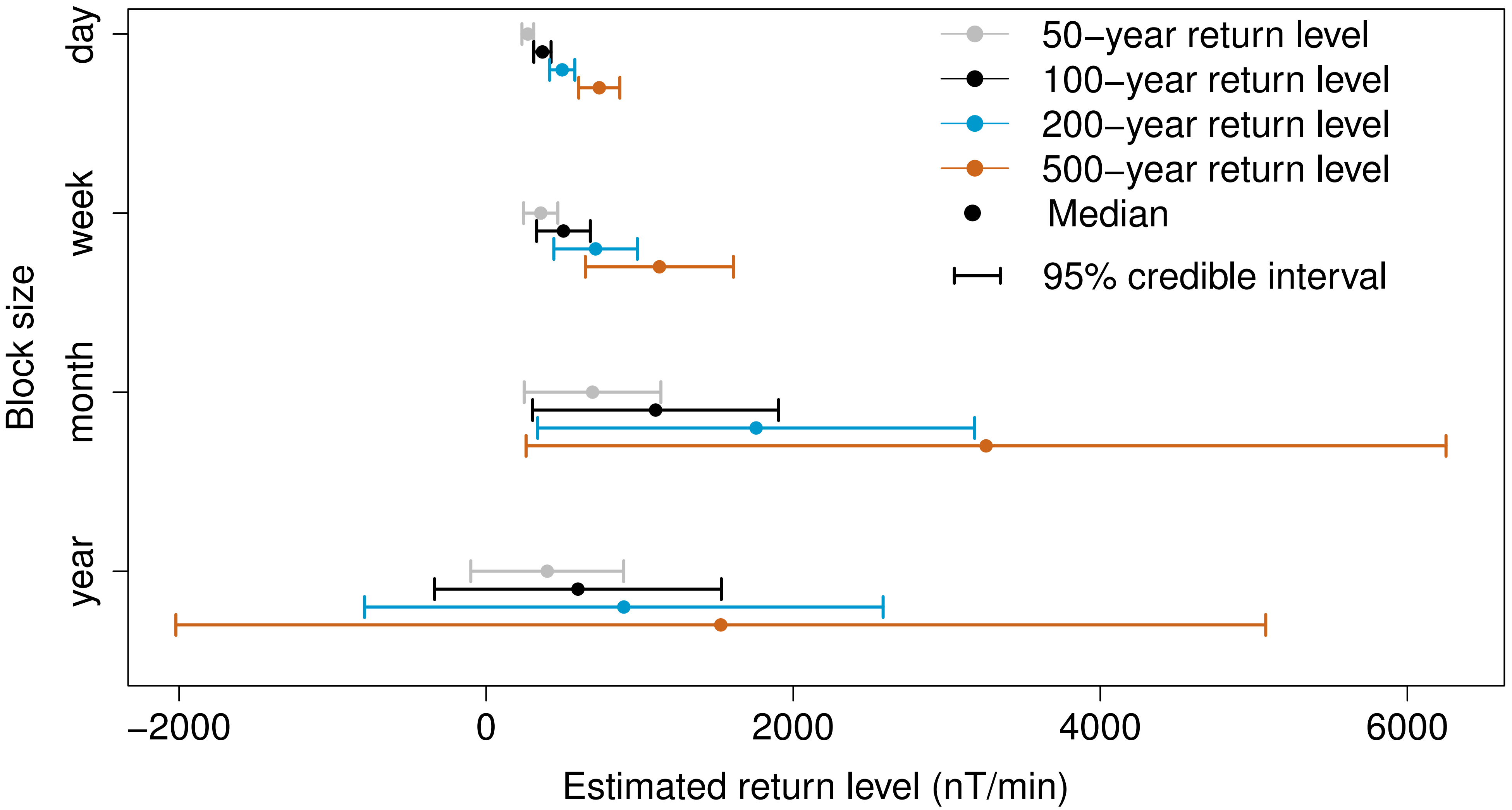
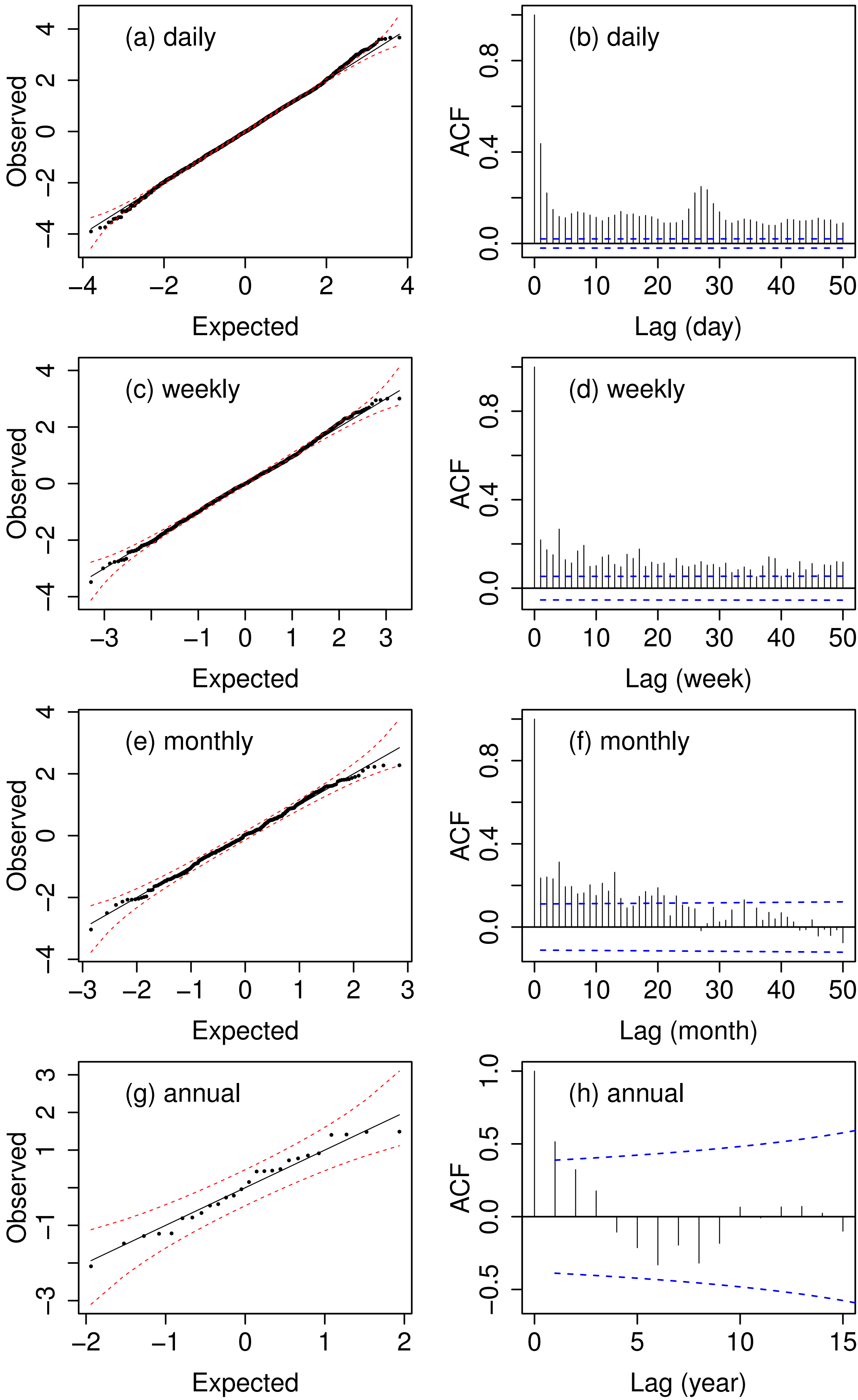

50−year return level
100−year return level
200−year return level
500−year return level
Median
95% credible interval
day
week
month
year
Block size
−2000
0
2000
4000
6000
Estimated return level (nT/min)

Figure 2.



(a) daily
Observed
Expected
(b) daily
ACF
Lag (day)
(c) weekly
Observed
Expected
(d) weekly
ACF
Lag (week)
(e) monthly
Observed
Expected
(f) monthly
ACF
Lag (month)
(g) annual
Observed
Expected
(h) annual
ACF
Lag (year)

Figure 3.

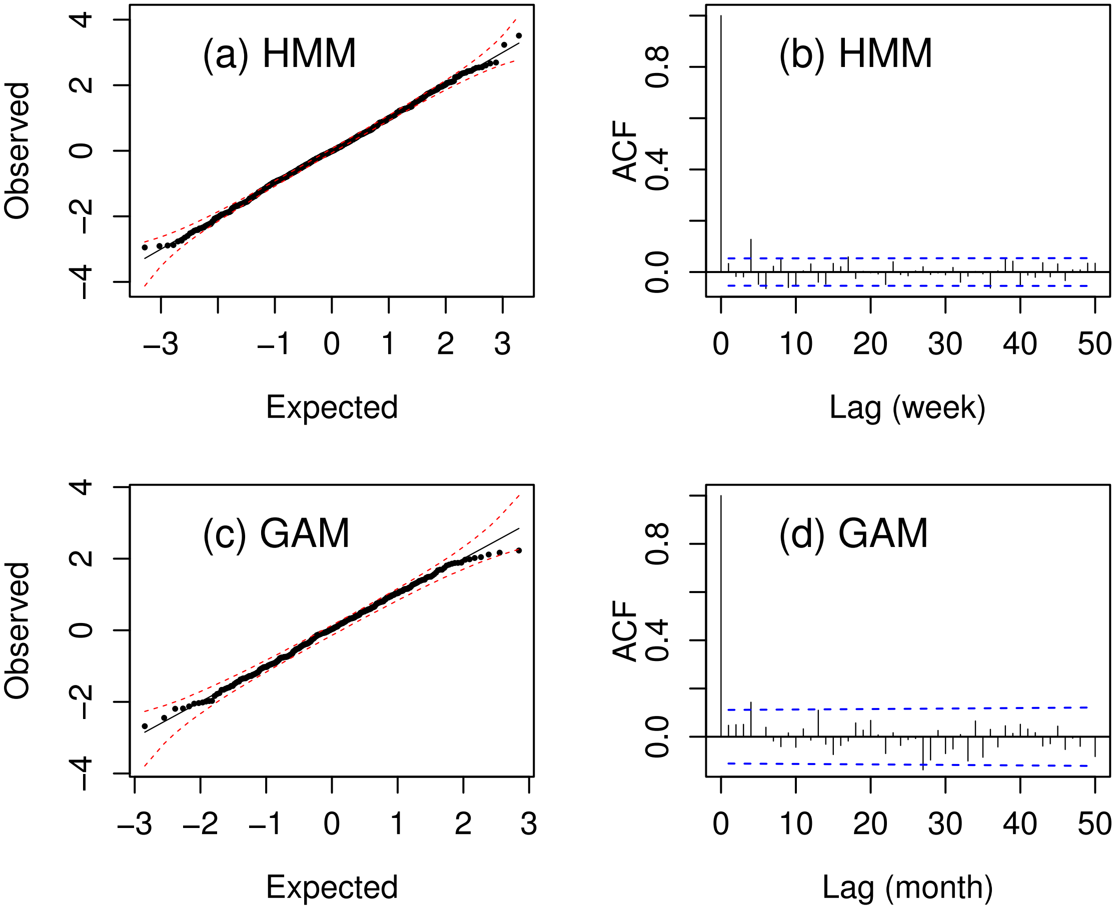

(a) HMM
Observed
Expected
(b) HMM
ACF
Lag (week)
(c) GAM
Observed
Expected
(d) GAM
ACF
Lag (month)

Figure 4.

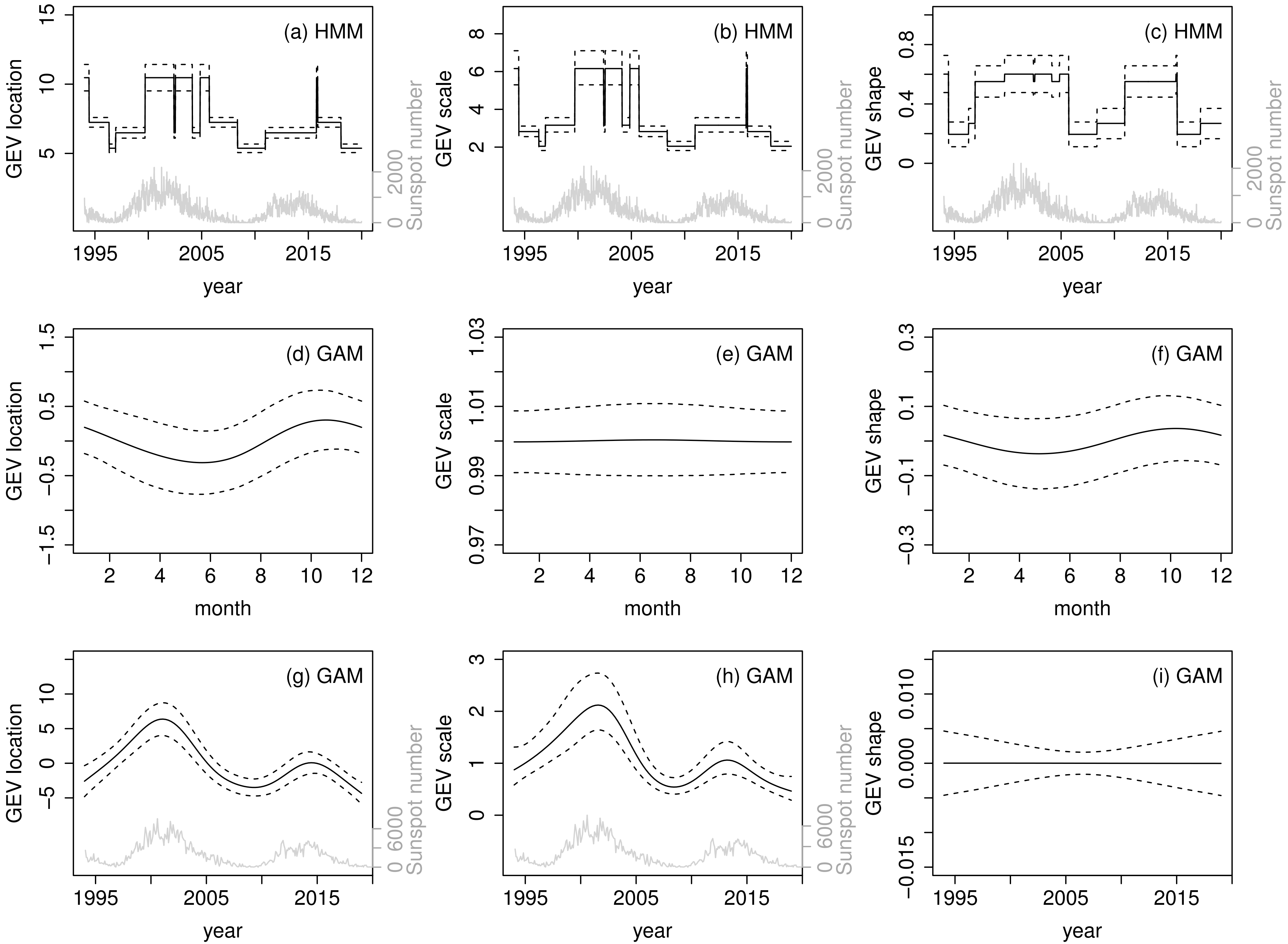
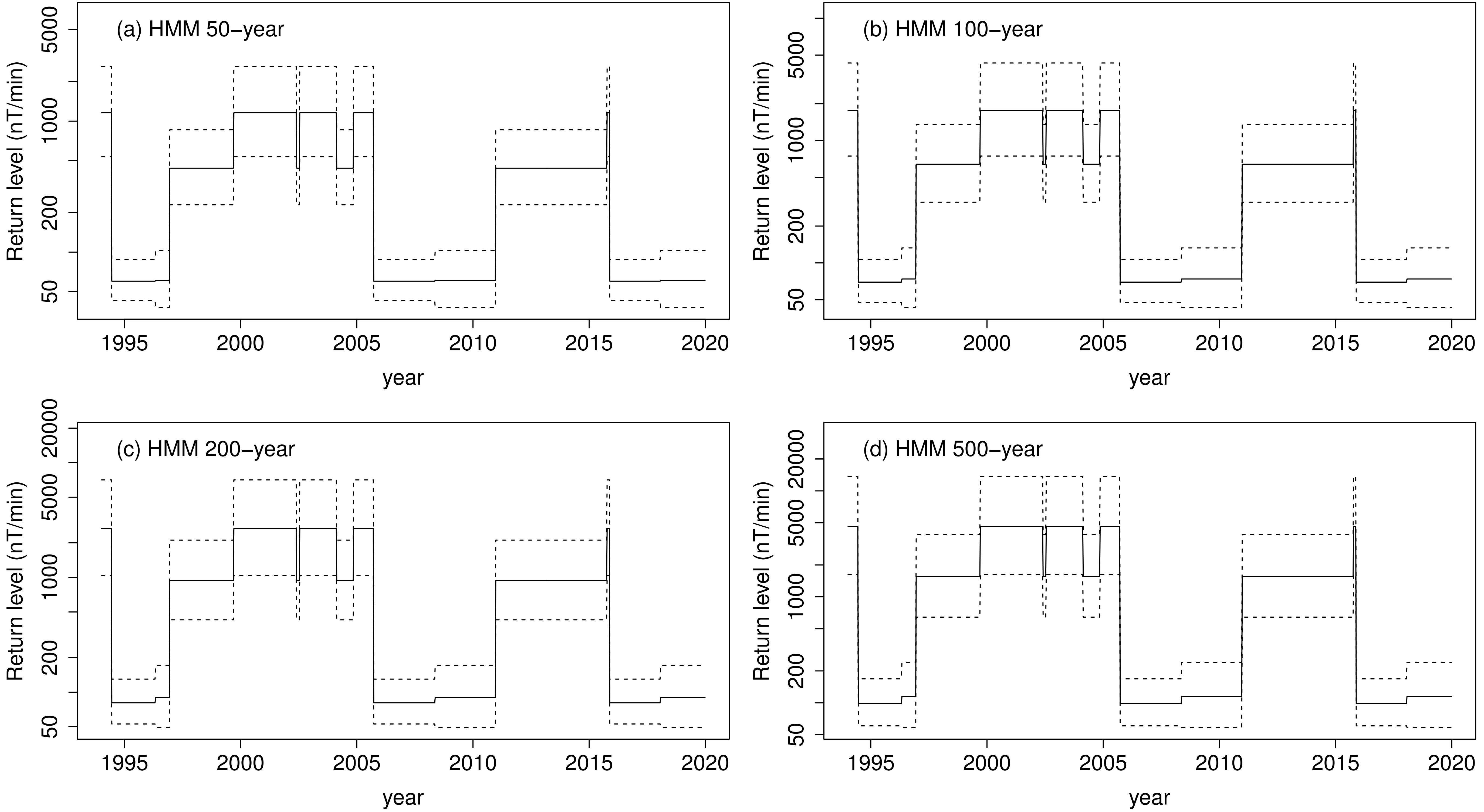
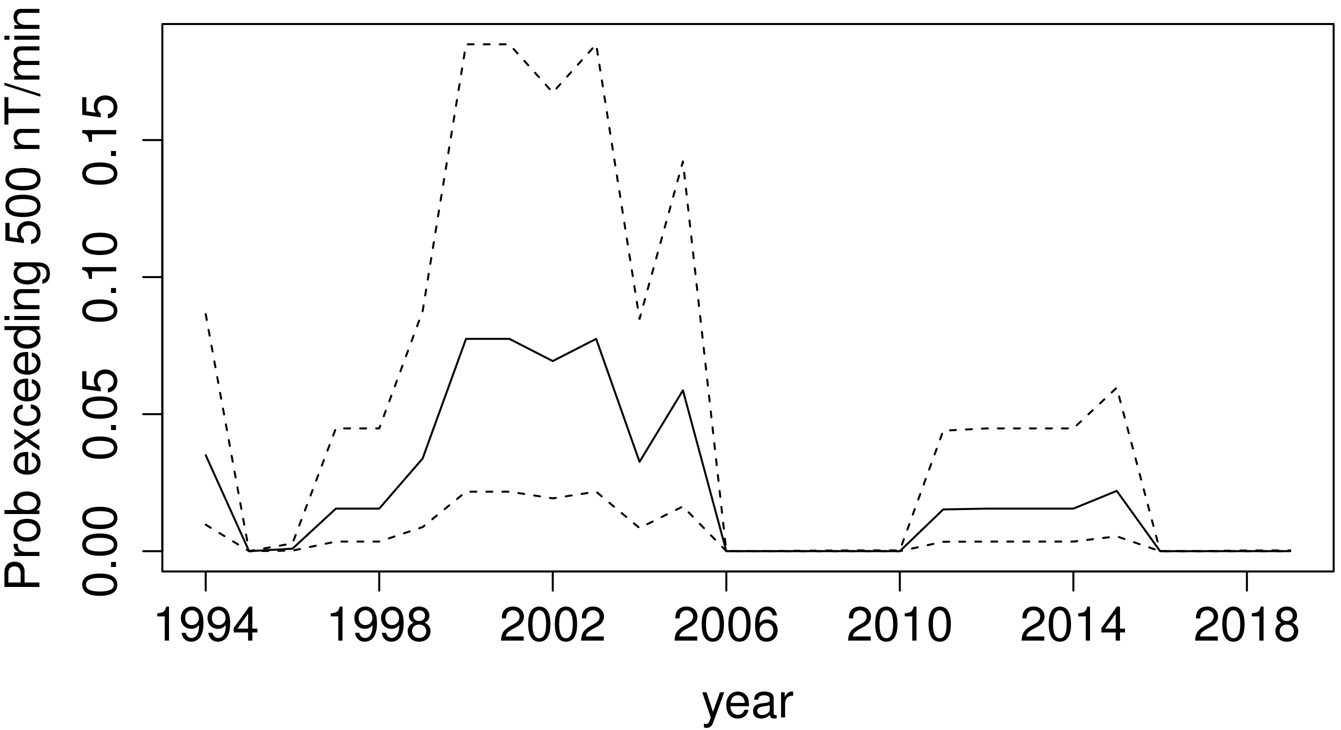

(a) HMM
GEV location
Sunspot number
year
(b) HMM
GEV scale
(c) HMM
GEV shape
(d) GAM
month
(e) GAM
(f) GAM
(g) GAM
(h) GAM
(i) GAM

Figure 5.



(a) HMM 50−year
(b) HMM 100−year
(c) HMM 200−year
(d) HMM 500−year
Return level (nT/min)
year
50
200
1000
5000
20000
1995
2000
2005
2010
2015
2020

Figure 6.



Prob exceeding 500 nT/min
0.00
0.05
0.10
0.15
1994
1998
2002
2006
2010
2014
2018
year

Figure 7.

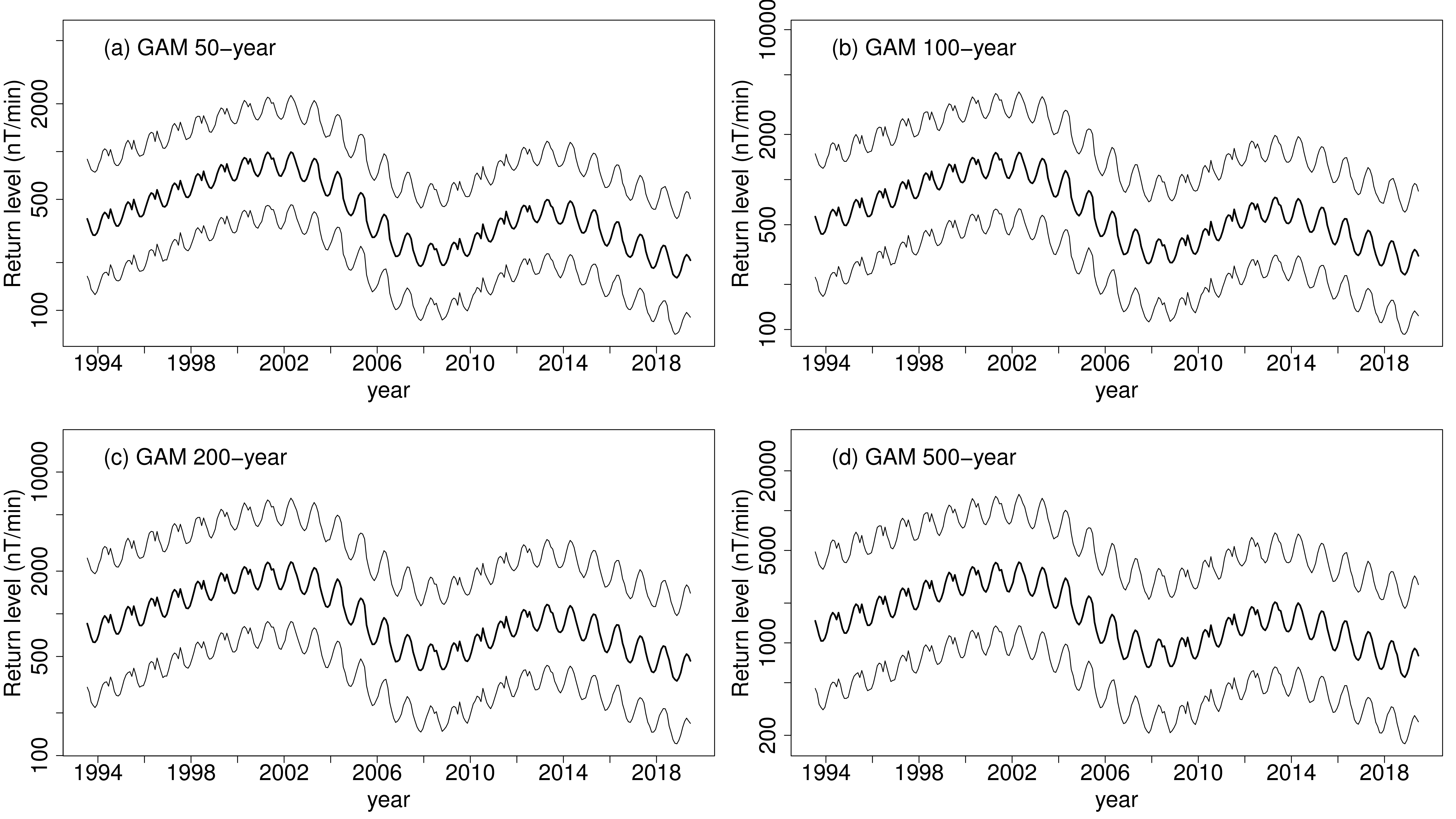
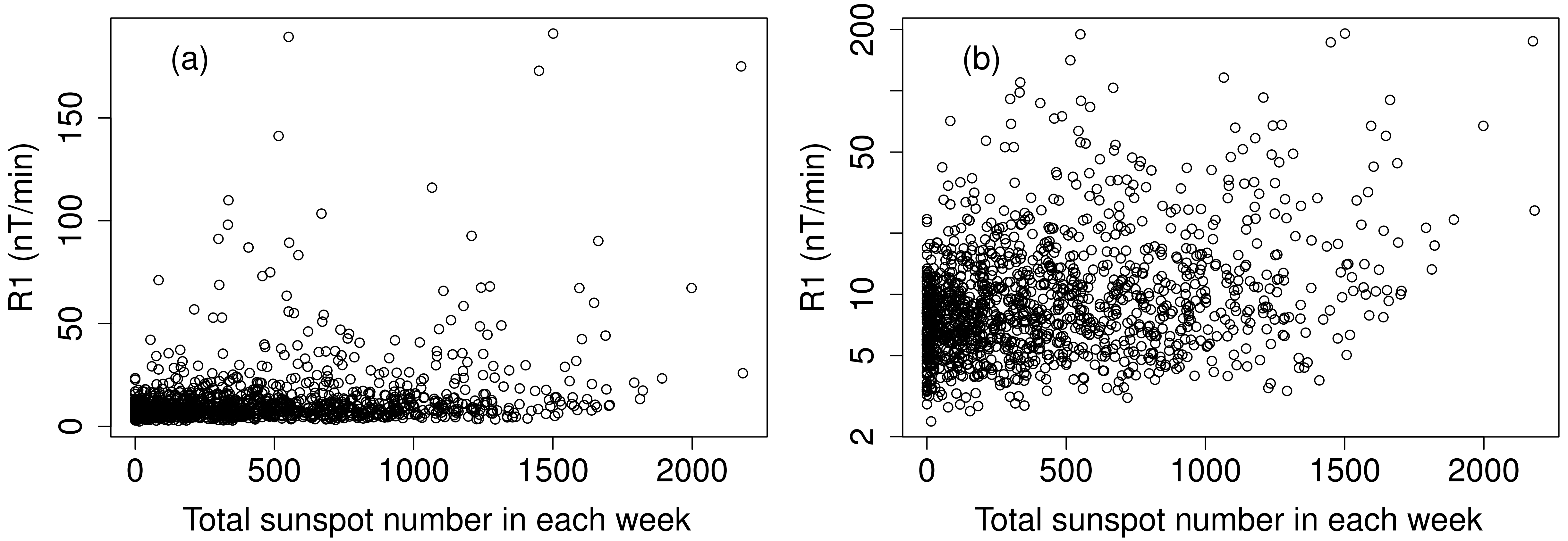

(a) GAM 50−year
Return level (nT/min)
100
500
2000
1994
1998
2002
2006
2010
2014
2018
year
(b) GAM 100−year
Return level (nT/min)
100
500
2000
10000
1994
1998
2002
2006
2010
2014
2018
year
(c) GAM 200−year
Return level (nT/min)
100
500
2000
10000
1994
1998
2002
2006
2010
2014
2018
year
(d) GAM 500−year
Return level (nT/min)
200
1000
5000
20000
1994
1998
2002
2006
2010
2014
2018
year

Figure 8.


(a)
R1 (nT/min)
0
50
100
150
Total sunspot number in each week
0
500
1000
1500
2000
(b)
R1 (nT/min)
2
5
10
50
200
Total sunspot number in each week
0
500
1000
1500
2000