Predicting Halloween- and Carrington-type solar flares

Lee Fawcett, Thomai Tsiftsi[†], Daniel Juan Pineda[†]

[†]Universidad Nacional Autónoma de México

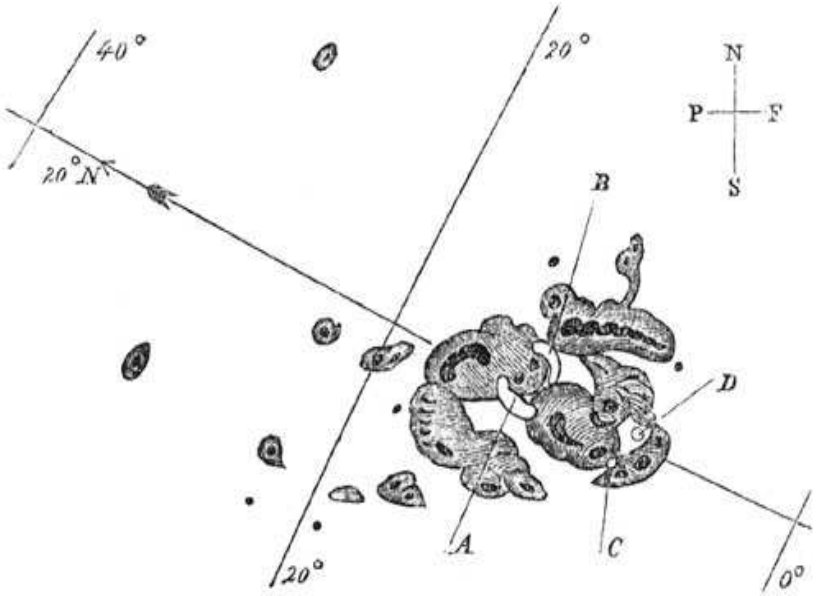
Structure of this talk

- Background / Motivation
  - Carrington event
  - Halloween solar storm
  - Potential consequences
- Data and typical POT analysis
- Improved analysis
  - Threshold selection
  - Dependence
  - Prior specification
  - Results / comparison

Carrington solar storm, 1859

- Powerful **geomagnetic solar storm** during solar cycle 10 (1855-1867)
- Solar **coronal mass ejection** (CME) hit Earth's magnetosphere inducing one of the largest geomagnetic storms on record
- Observed and recorded by **Richard Carrington**
- Aurorae visible in the northern hemisphere as far south as Monterrey, Cuba and southern Japan
- **Telegraph systems** failed across Europe and North America
- Peak fluxes classified as A, B, C, M, X: each letter indicates a flare an order of magnitude larger than the one preceding it
  - A1 flare has a peak X-ray flux of $1 \times 10^{-8}\ \mathrm{Wm}^{-2}$
  - X1 flare has a peak X-ray flux of $1 \times 10^{-4}\ \mathrm{Wm}^{-2}$
  - Carrington event: estimated as X45, i.e. $45 \times 10^{-4}\ \mathrm{Wm}^{-2}$

Carrington solar storm, 1859



The Auroral Display in Boston.

Boston, Friday, Sept. 2.

There was another display of the Aurora last night, so brilliant that at about one o'clock ordinary print could be read by the light. The effect continued through this forenoon, considerably affecting the working of the telegraph lines. The auroral currents from east to west were so regular that the operators on the Eastern lines were able to hold communication and transmit messages over the line between this city and Portland, the usual batteries being discontinued from the wire. The same effects were exhibited upon the Cape Cod and other lines.

The New York Times
Published: September 3, 1859
Copyright © The New York Times

A STORM OF ELECTRICITY

TELEGRAPH WIRES USELESS FOR SEVERAL HOURS.

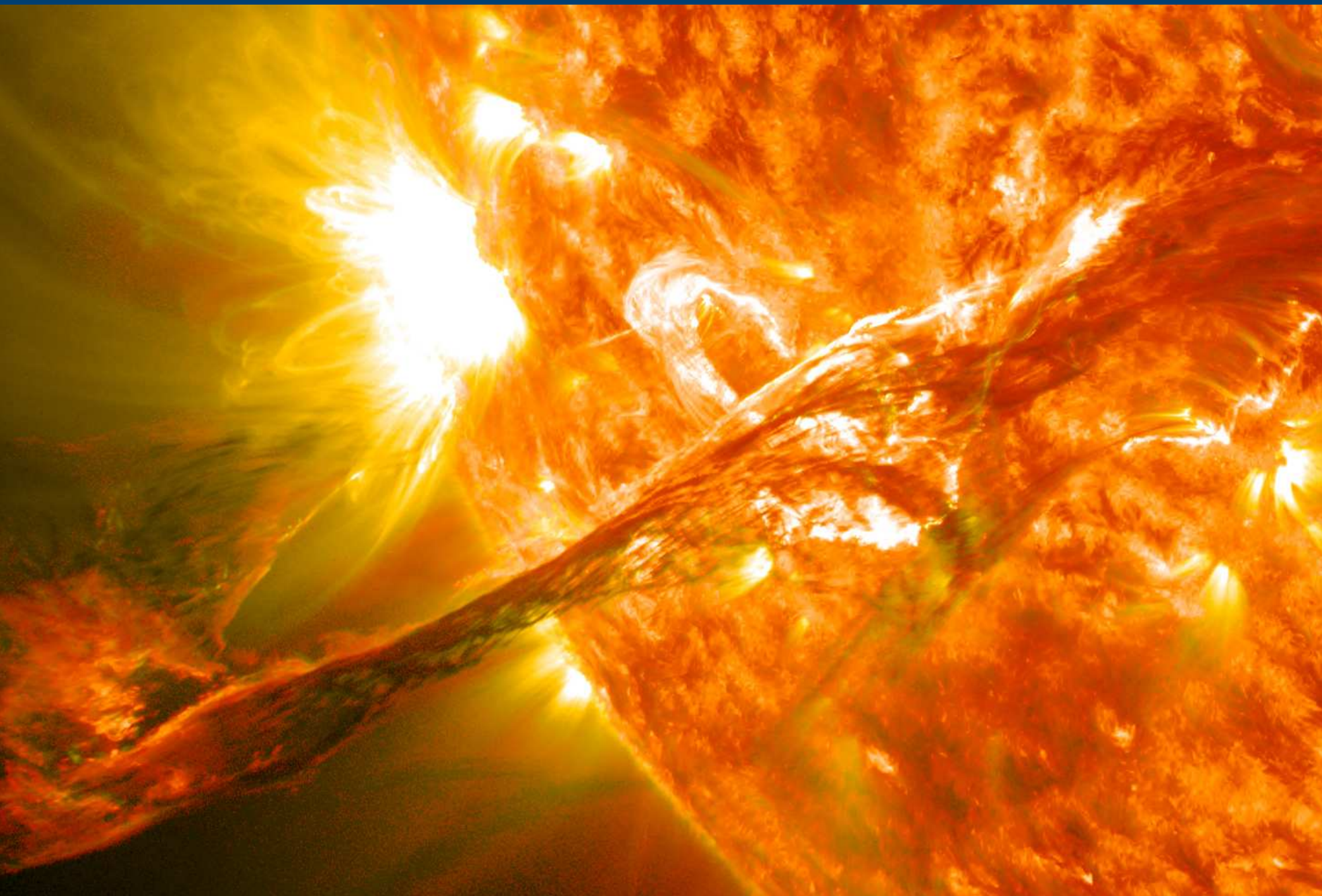
ONE OF THE MOST SEVERE DISTURBANCES FOR MANY YEARS, EXTENDING EVEN TO EUROPE—TELEPHONE WIRES ALSO OBSTRUCTED—BUSINESS DELAYED A GOOD PART OF THE DAY.

Yesterday's storm was accompanied by a more serious electrical disturbance than has been known for years. It very seriously affected the workings of the telegraph lines both on the land and in the sea, and for three hours—from 9 A. M. until noon—telegraph business east of the Mississippi and north of Washington was at a stand-still.

Halloween solar storms, 2003

- Series of solar flares and CMEs occurring in October/November 2003 during solar cycle 23 (1996-2008)
- Generated the largest solar flare ever recorded by the **GOES** system, estimated at X28
- **Satellite systems** and **communications** affected, numerous **spacecraft damaged**, **power outages** in Sweden and Finland
- Aurorae observed in Texas and southeast Mediterranean

Near miss: Carrington-like solar storm, 2012



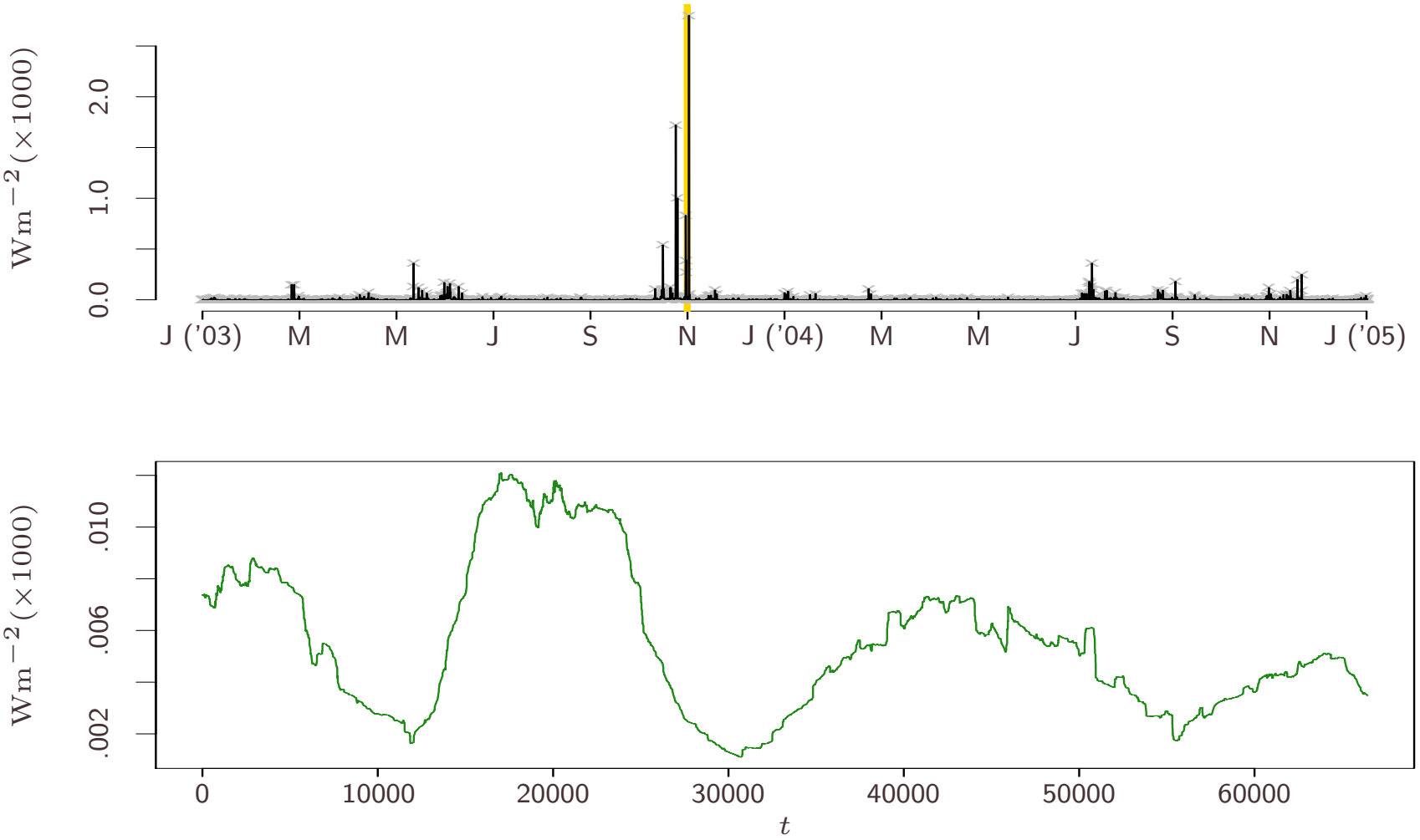
What's at stake?

- Disruption to **global positioning systems**. GPS industry estimated to be a $1 trillion business
- **Satellite communications** at high risk
- Huge disruption to the **electrical grid**: areas such as the eastern US particularly vulnerable as power infrastructure highly interconnected, large cities without power for weeks
- Key questions: When might we see the next Carrington-like solar storm? What is the **return period** of a Carrington-like event? What is the strength of solar flare we might expect to see once per century?

Data

- Extracted from the **Space Weather Prediction Center** (SWPC) online portal
- Recorded by the **Geostationary Operational Environmental Satellite** (GOES) system, operated by the US National Oceanic and Atmospheric Administration (NOAA)
- Data represent solar flares normalised from H_α spectral observations and X-ray fluxes, spanning a period of 43 years: November 1975-October 2017
- Data points fairly regularly-spaced, with around four or five observations per day
- Data exhibit substantial short-term temporal dependence and cyclic variability

Data


$\mathrm{Wm}^{-2}\,(\times 1000)$
0.0
1.0
2.0
J ('03)
M
M
J
S
N
J ('04)
M
M
J
S
N
J ('05)
$\mathrm{Wm}^{-2}\,(\times 1000)$
.002
.006
.010
0
10000
20000
30000
40000
50000
60000
t

Modelling extremes: Starting point

Let $\{X_n\}$ denote a stationary sequence of random variables with common distribution function F, and let $M_n = \max\{X_1, \ldots, X_n\}$.

It is typically the case that, as $n \to \infty$,

$$\Pr(M_n \leq x) \quad \approx \quad F^{n\theta}(x), \tag{1}$$

where $\theta \in (0, 1)$ is known as the **extremal index**; see, for e.g., Leadbetter & Rootzén (1988).

As $\theta \to 0$ there is increasing dependence in the extremes of the process; for an independent process, $\theta = 1$.

Modelling extremes: Block maxima

The **Generalised Extreme Value** (GEV) distribution is the limiting model for F^n, with distribution function (d.f.)

$$\mathcal{G}(y;\mu,\varsigma,\xi) = \begin{cases} \exp\left[-\left(1+\xi(y-\mu)/\varsigma\right)^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left[-\exp\left(-(y-\mu)/\varsigma\right)\right], & \xi = 0, \end{cases}$$

defined on $\{y : 1+\xi(y-\mu)/\varsigma > 0\}$, where $-\infty < \mu < \infty$, $\varsigma > 0$ and $-\infty < \xi < \infty$ are parameters of location, scale and shape.

- It can be shown that $F^{n\theta}$ is also GEV with d.f. $\mathcal{G}(y;\mu^*,\varsigma^*,\xi)$, provided long-range dependence is negligible[1]
- In practice, the GEV is used to model maxima over some convenient calendar unit – usually years

[1] Technically, this requires Leadbetter's $D(u_n)$ condition to hold; see Leadbetter *et al.* (1983).

Modelling extremes: Threshold excesses

Pickands (1975) showed that for $\theta = 1$ and u large, $(X - u | X > u)$ follows a **Generalised Pareto** (GP) distribution with d.f.

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

defined on $\{y : y > 0 \text{ and } (1 + \xi y/\sigma) > 0\}$, where $\sigma = \varsigma + \xi(u - \mu) > 0$ and ξ are the scale and shape.

- Graphical diagnostics available for **selection of threshold** u, but often difficult to interpret
- Extremal index θ can't be absorbed into the GP parameters; careful consideration of **extremal dependence** required

Modelling extremes: Return level estimation

The **r-year return level z_r** can then be obtained by inversion of $\mathcal{G}(z_r)$ or $\mathcal{H}^{\theta}(z_r)$.

For threshold excesses, on equating to $1 - r^{-1}$ this gives

$$z_r = \begin{cases} u + \sigma\xi^{-1}\left[\left(\lambda_u^{-1} w_r\right)^{-\xi} - 1\right], & \xi \neq 0, \\ u - \sigma \log\left(\lambda_u^{-1} w_r\right), & \xi = 0, \end{cases} \tag{2}$$

where $w_r = 1 - \left[1 - (rn_y)^{-1}\right]^{1/\theta}$, λ_u is the rate of threshold excess and n_y is the (average) number of observations per year.

For seasonally varying extremes, we usually solve

$$\prod_{m=1}^{M} \mathcal{H}_m^{\theta_m}(z_r) \quad = \quad 1 - r^{-1},$$

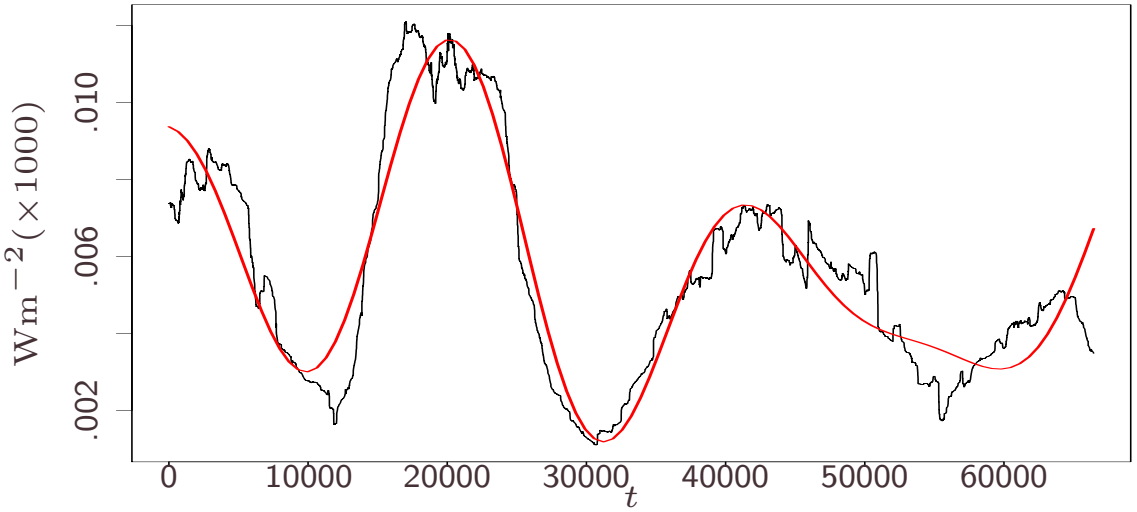
where $\mathcal{H}_m$ is the GP distribution function in season m with parameter set $(\lambda_{u_m}, \sigma_m, \xi_m)$, and θ_m is the extremal index in season m.

Cyclic variation

- Close inspection of the time series suggests non-stationarity **inherited from the solar cycle**, affecting both the frequency and the severity of extreme flares
- After trial-and-improvement, fit sinusoidal waves of the form

$$\sum_{i=1}^{3} A_i \sin \omega_i t + B_i \cos \omega_i t + C$$

to the series of moving averages, subtracting the estimated function before performing extreme value analyses



Typical peaks over threshold (POT) analysis

- Identify independent clusters of solar flare extremes over some **suitably high threshold** u
- Runs declustering, with **cluster termination interval** κ, most commonly used
- Fit the GP distribution to the set of independent cluster peak excesses
- Substitute $(\hat{\lambda}_u, \hat{\sigma}, \hat{\xi}, \theta = 1)$ into (2) to obtain $\hat{z}_r/\hat{r}$:

(u, κ)		$\hat{z}_{50}$		$\hat{r}$ years	
		Post. mean	95% CI	Halloween	Carrington
$u = 90\%$ quantile	$\kappa = 10$	X34.5	(X30.5, X44.5)	38	110
	$\kappa = 15$	X35	(X31.5, X47)	37.5	100
	$\kappa = 30$	X37	(X32, X65)	33	96.5
$u = 95\%$ quantile	$\kappa = 10$	X35	(X30, X46)	38	100
	$\kappa = 15$	X36	(X31, X47.5)	37	97
	$\kappa = 30$	X38.5	(X30, X72.5)	30	79

(As reported in Tsiftsi and De La Luz, 2018)

Typical peaks over threshold (POT) analysis

- Choice of threshold difficult:
  - Just choose a high quantile?
  - Graphical diagnostics – e.g. **mean residual life plot**, **Hill plots**, **parameter stability plots**...
  - Once u is chosen, it is treated as fixed
- Choice of declustering scheme difficult:
  - Runs declustering: Choice of κ?
  - Other filtering schemes available
  - Wasteful – **large uncertainty** on return level estimates
- Prior specification – objective priors, missed opportunity?

Improved analysis: Aims

Here, we have three main aims:

1. Estimate the **extremal index** for use in return level estimation via a Markov chain model for threshold excesses
   - Avoids declustering altogether, and the problems it brings
   - Presses into use information on *all* extremes
2. Capture uncertainty in the **choice of threshold**, and enable a more 'automatic' selection of u
3. Get more out of working within the Bayesian framework
   - Where possible, formulate **informative priors**
   - Make use of the **posterior predictive distribution**

Improved analysis: Mixture model

Suppose our data comprise a sequence of n identically distributed observations $\boldsymbol{x} = \{x_i : i = 1, \ldots, n\}$. We assume the following mixture model:

$$f(x) = (1 - \lambda_u) \cdot f_1(x) + \lambda_u \cdot f_2(x),$$

where

$$\begin{aligned} f_1(x) &= \frac{h(x|\phi, \boldsymbol{x})}{H(u|\phi, \boldsymbol{x})} I_{(-\infty,u]}(x) \quad \text{and} \\ f_2(x) &= g(x|u, \sigma_u, \xi) I_{(u,\infty)}(x). \end{aligned}$$

- Observations sub-threshold follow a non-parametric density h
- Excesses above the threshold follow a GP distribution with parameter set (u, σ_u, ξ) and extremal index θ
- Possibility of a discontinuity at u

Improved analysis: Bulk distribution

For $f_1(x)$ we use a kernel estimator of the form

$$\hat{f}(x) \quad = \quad n^{-1} \sum_{i=1}^{n} K_\phi(x - x_i), \tag{3}$$

where $K_\phi(y) = \phi^{-1} K(y/\phi)$ is chosen to be a unimodal probability density function that is symmetric about zero.

- Kernel function used in (3) not critical; we use a zero mean Gaussian kernel
- GP distribution is used for the upper tail – extrapolation into the upper tails is of no concern
- **Kernel bandwidth likelihood** $L_{\mathsf{K}}(\phi, u|\boldsymbol{x})$, for $A = \{i : x_i \leq u\}$:

$$\left\{ \frac{(1-\lambda_u)}{n^{-1} \sum_{k=1}^{n} \Phi\left(\frac{u - x_i}{\phi}\right)} \right\}^{|A|} \prod_{A} \frac{1}{n-1} \sum_{j=1; j \neq i}^{n} K_\phi(x_i - x_j)$$

Improved analysis: Threshold excess distribution

We assume a first-order Markov structure, giving the likelihood:

$$\prod_{i=1}^{n-1} f(x_i, x_{i+1}; \boldsymbol{\psi}) \Big/ \prod_{i=2}^{n-1} f(x_i; \boldsymbol{\psi}). \tag{4}$$

- Transformation from GP to unit Fréchet margins gives a range of bivariate models to use for the numerator in (4)
- Most commonly-used is the **logistic family** with $\alpha \in [0, 1)$:

$$G(x_i, x_{i+1}) \quad = \quad \exp\left\{-\left(x_i^{-1/\alpha} + x_{i+1}^{-1/\alpha}\right)^{\alpha}\right\}. \tag{5}$$

- **Extreme value Markov chain likelihood** $L_{\mathsf{EV}}(u, \sigma_u, \xi, \alpha | \boldsymbol{x})$:
  - Differentiation of (5), with careful censoring, gives pairwise contributions to the numerator in (4)
  - Univariate contributions to the denominator are given through the GP distribution

Aside 1: Estimating the extremal index?

$$z_r = \begin{cases} u + \sigma\xi^{-1}\left[\left(\lambda_u^{-1} w_r\right)^{-\xi} - 1\right], & \xi \neq 0, \\ u - \sigma \log\left(\lambda_u^{-1} w_r\right), & \xi = 0, \end{cases}$$

where $w_r = 1 - \left[1 - (rn_y)^{-1}\right]^{1/\theta}$, λ_u is the rate of threshold excess and n_y is the (average) number of observations per year.

Aside 1: Estimating the extremal index?

Define (arbitrarily) x_n such that $F^n(x_n) = 1/2$ in (1).

Then, using (1), we can define

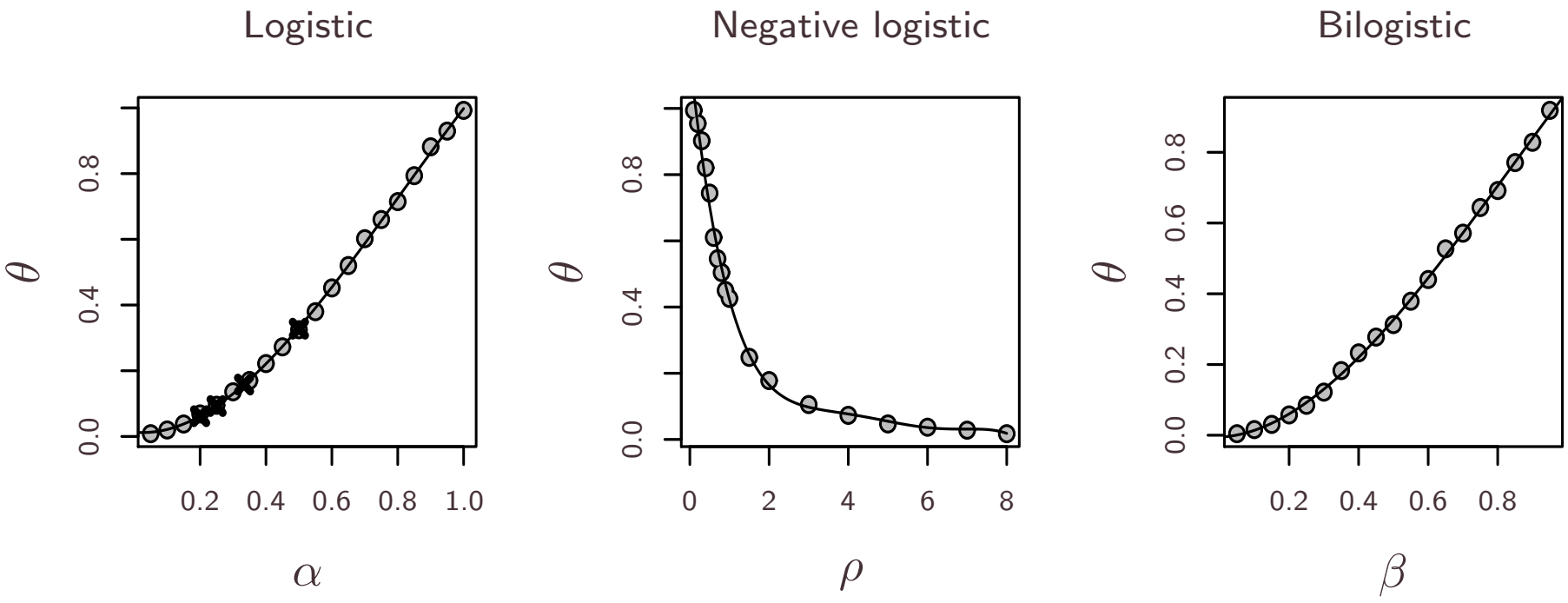
$$\theta_n = -\frac{\log \Pr\left(\max\{X_1, \ldots, X_n\} \leq x_n\right)}{\log 2}, \tag{6}$$

and so $\theta_n \to \theta$ as $n \to \infty$.

We can use (6) to investigate the relationship between θ and α in the logistic model:

- Simulate N first-order Markov chains, each of length n, with logistic dependence α
- The probability in the numerator of (6) is estimated as the proportion of simulated chains whose maximum does not exceed x_n
- We find values for θ_n for $\alpha = \{0.05, 0.10, ..., 0.95\}$, with $N = n = 10^6$

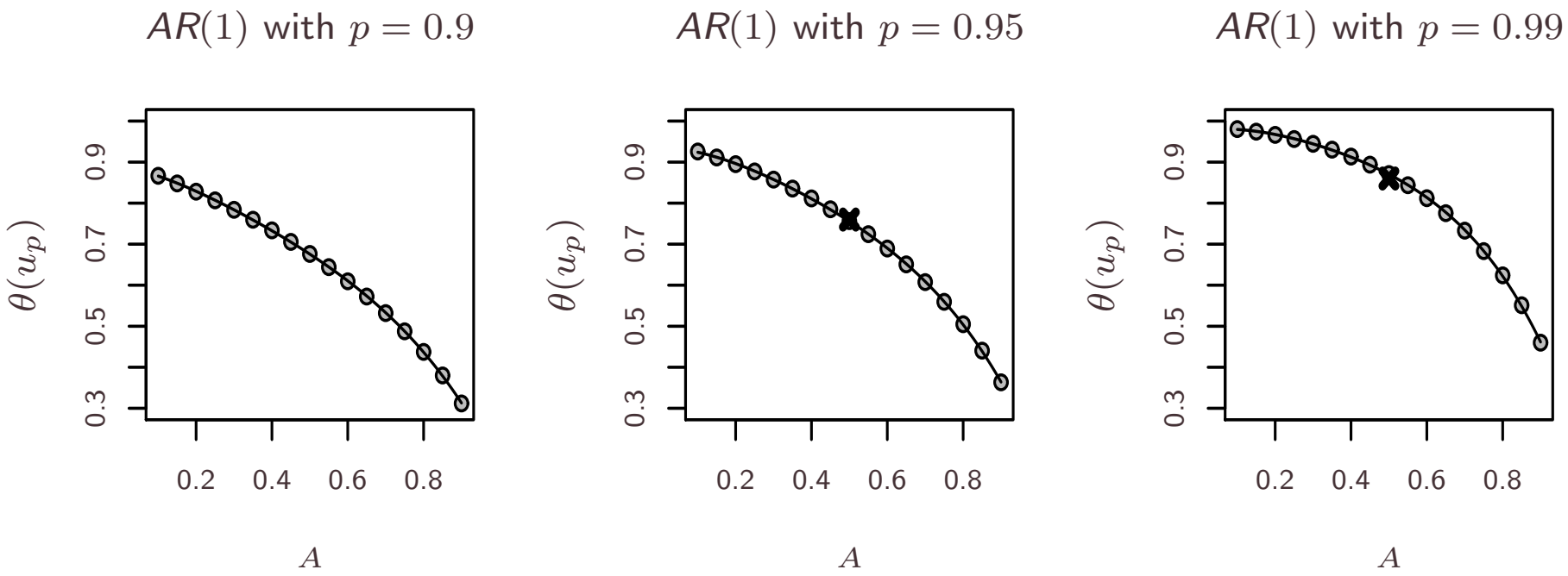
Aside 1: Estimating the extremal index?



Giving, for example,

$$\theta \approx 0.013 - 0.092\alpha + 1.833\alpha^2 - 0.756\alpha^3$$

Aside 1: Estimating the extremal index?



Improved analysis: Likelihood

Combining the Kernel density model for sub-threshold observations with the GP extreme value Markov chain for threshold excesses gives the **extreme value mixture model** likelihood:

$$L(\boldsymbol{\psi}|\boldsymbol{x}) = L_{\mathsf{K}}(\phi, u|\boldsymbol{x}) \cdot L_{\mathsf{EV}}(u, \sigma_u, \xi, \alpha|\boldsymbol{x}),$$

where $\boldsymbol{\psi} = (\phi, u, \sigma_u, \xi, \alpha)$.

Improved analysis: Prior specification

The joint prior for $\boldsymbol{\psi}$ is expressed as

$$\pi(\phi, u, \sigma_u, \xi, \alpha) = \pi(\phi) \cdot \pi(u) \cdot \pi(\sigma_u, \xi) \cdot \pi(\alpha)$$

1. For the **kernel density bandwidth parameter** ϕ we follow Brewer (2000) by specifying the prior for the precision $1/\phi^2$ as inverse gamma:

$$\pi(\phi|d_1, d_2) = \frac{1}{d_2^{d_1}\Gamma(d_1)} \left(\frac{1}{\phi^2}\right)^{d_1-1} \exp\left(-\frac{1}{\phi^2 d_2}\right),$$

   with d_1 and d_2 chosen to give a largely non-informative prior specification.

2. In the absence of any expert prior information regarding the **logistic dependence parameter** α, we assume $\alpha \sim \mathcal{U}(0, 1)$.

Improved analysis: Prior specification

3. Following Behrens et al. (2004) we assume a truncated Normal prior for the **threshold** u:

$$\pi(u|\mu, v^2) = \frac{1}{\sqrt{2\pi v^2}} \frac{\exp\{-0.5[(u-\mu)/v]^2\}}{\Phi[-(e_1-\mu)/v]},$$

where μ is set at some high data percentile; v^2 is set sufficiently large to represent a diffuse prior.

Improved analysis: Prior specification

4. For the **GP parameters** (σ_u, ξ) we attempt to formulate informative priors.
   - We work with a re-parameterised GP distribution: $\eta = \log(\sigma_u - \xi u)$
   - Following Coles & Tawn (2005) we assume

$$\begin{aligned} \tilde{z}_1 &= z_{10} - e_1 \sim \mathrm{Ga}(g_1, h_1) \\ \tilde{z}_2 &= z_{50} - z_{10} \sim \mathrm{Ga}(g_2, h_2), \end{aligned}$$

     where $(g., h.)$ reflect the opinions of an expert[2]
   - Then:

$$\pi(z_{10}, z_{50}) \propto z_{10}^{g_1 - 1} \mathrm{e}^{-h_1 z_{10}} (z_{50} - z_{10})^{g_2 - 1} \mathrm{e}^{-h_2 (z_{50} - z_{10})}$$

   - Multiplying by the Jacobian of the transformation $(z_{10}, z_{50}) \longrightarrow (\eta, \xi)$ gives a prior in terms of (η, ξ)

[2] Victor De la Luz, Laboritorio Nacional de Clima Especial, Instituto de Gefísica, Unidad Michoacán, Unversidad Nacional Autónoma de México

Improved analysis: Posterior sampling scheme

We use a simple Metropolis-Hastings scheme with random walk updates, tuning the innovation variances to optimise efficiency, giving successive draws

$$\boldsymbol{\psi}^{[j]} = (\phi^{[j]}, u^{[j]}, \eta^{[j]}, \xi^{[j]}, \alpha^{[j]})^T, \quad j = 1, \ldots, 10^6.$$

- Chains thinned by a factor of 50 to remove substantial autocorrelation
- Convergence assessed using some standard diagnostics as discussed in Gelman and Rubin (1992)
- For threshold excesses, posterior draws for $(u, \sigma_u = f(u, \eta, \xi), \xi, \theta = f(\alpha))^{[j]}$ used to obtain posterior draws $z_r^{[j]}$ via (2)
- This leads to standard reporting of, e.g., **posterior mean** plus **posterior st. dev.** / **credible interval**

Improved analysis: Predictive inference

- Bayesian framework offers natural extension to prediction through the **posterior predictive distribution**
- Let Y denote a future extreme of our solar flares process. Then we can write

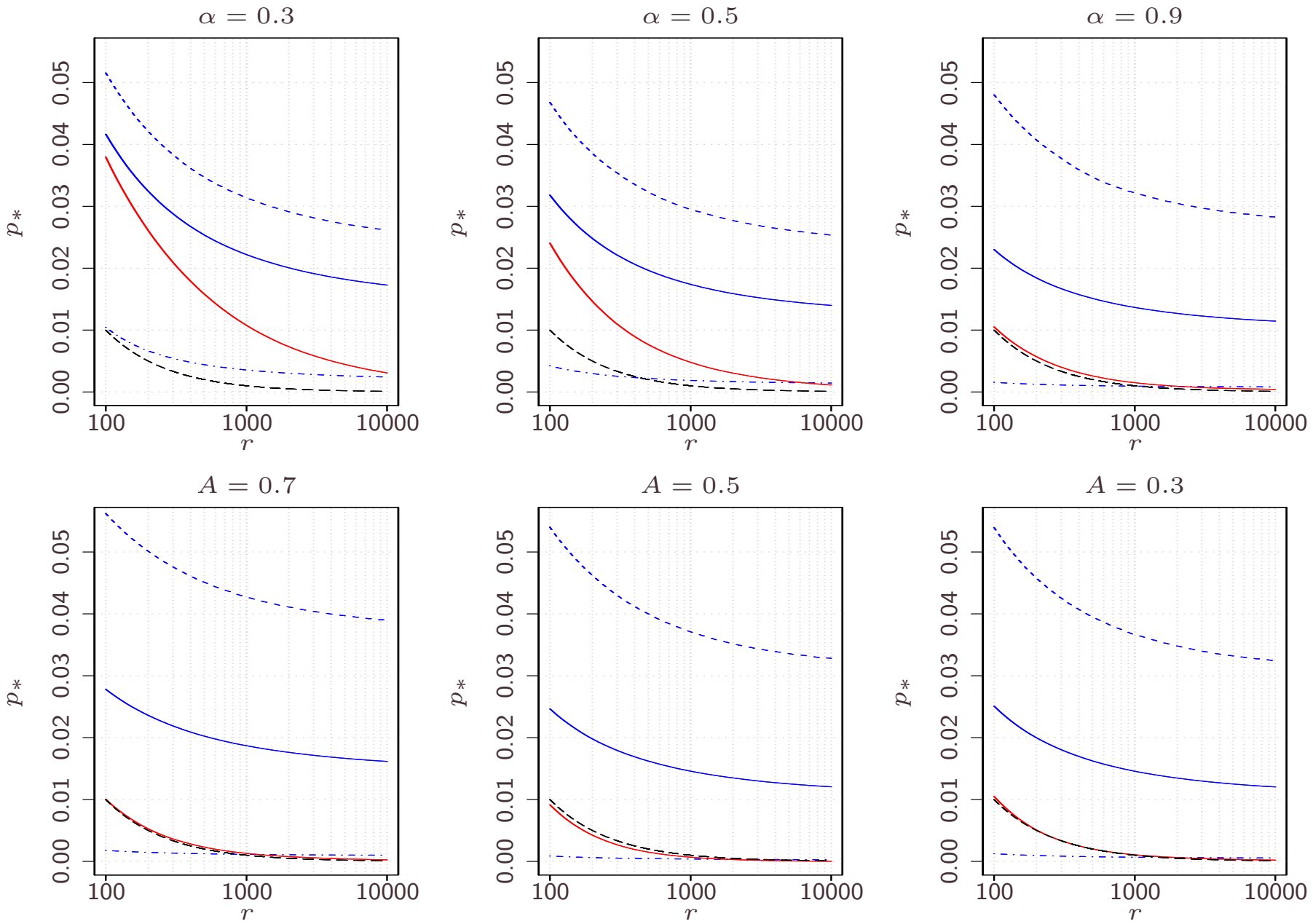
$$\Pr(Y \leq y|\boldsymbol{x}) \quad = \quad \int_{\boldsymbol{\Psi}} \Pr(Y \leq y|\boldsymbol{\psi})\pi(\boldsymbol{\psi}|\boldsymbol{x})d\boldsymbol{\psi} \qquad (7)$$

- Equating to $1 - r^{-1}$ and solving for $y = z_{r,\text{pred}}$ gives a return level estimate which incorporates uncertainty due to parameter estimation
- (7) analytically intractable; approximate using

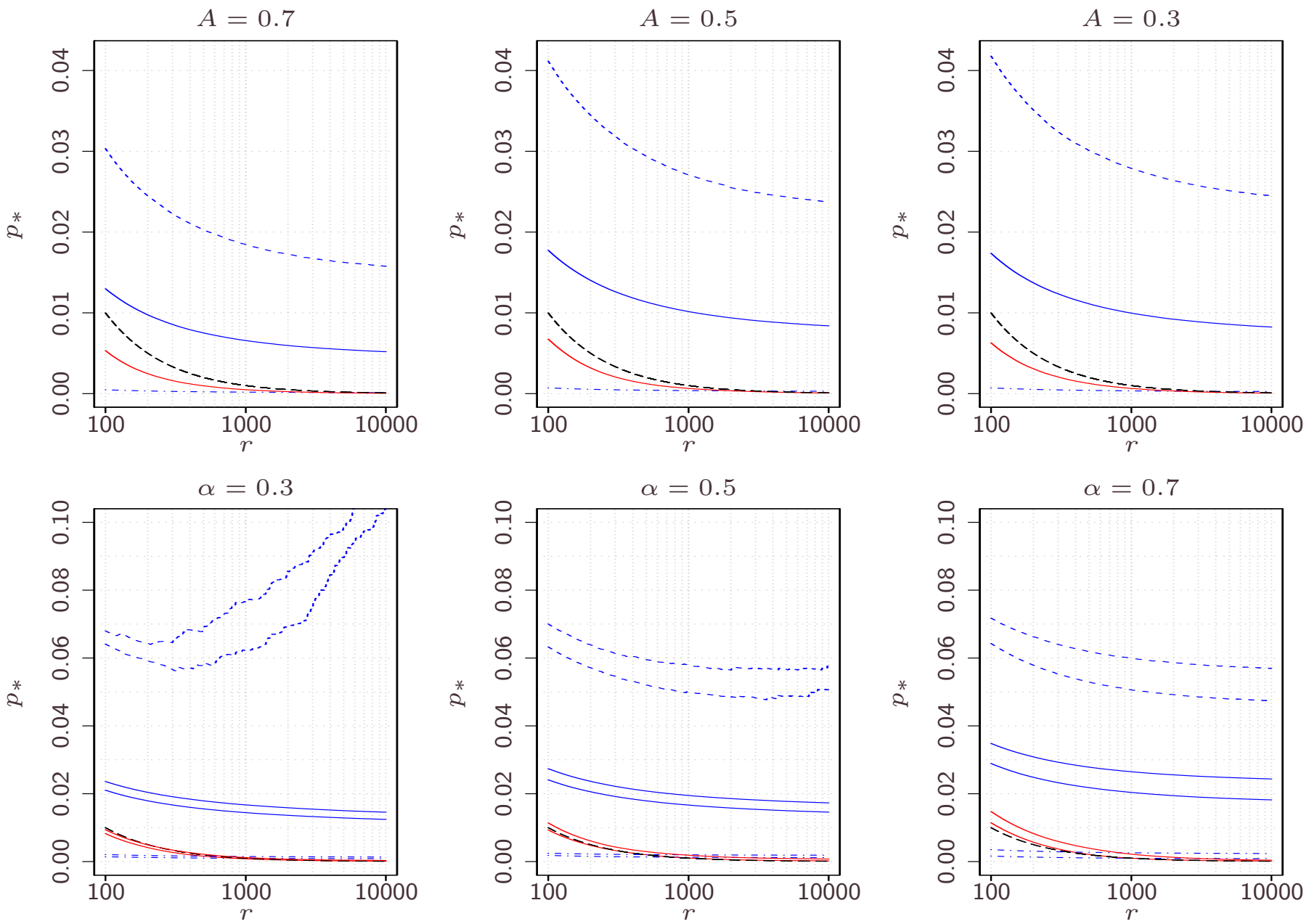
$$\frac{1}{S}\sum_{j=1}^{S} \Pr\left\{Y \leq z_{r,\text{pred}}|\boldsymbol{\psi}^{[j]}\right\} = \frac{1}{S}\sum_{j=1}^{S} \mathcal{H}^{\theta^{[j]}}(z_{r,\text{pred}}|u^{[j]}\sigma_u^{[j]}, \xi^{[j]})$$

where S is the number of MCMC iterations after burn-in

Aside 2: Summarising our return level inference?



Aside 2: Summarising our return level inference?



Improved analysis: Results

	z_{50}	z_{1000}
Post. mean	X35 *X35*	X57 *X54*
95% CI	(X32, X41) *(X31.5, X47)*	(X49, X66) *(X40, 77)*
Predictive	X39	X79

	Halloween	Carrington
Post. mean	37 years *38 years*	93 years *110 years*
95% CI	(31, 45) years *(27, 50) years*	(80, 114) years *(93, 139) years*
Predictive	43 years	112 years

Conclusions

- Fitted a model which **doesn't require a choice of threshold / declustering scheme**
  - Graphical diagnostics for threshold choice are often difficult to interpret / contradictory
  - Uncertainty in threshold choice is accounted for
  - Return level estimates can be sensitive to the choice of declustering scheme
  - Using declustered extremes can result in low precision on return level estimates
- Is the model for dependence the 'best'?
  - We used a model which implies asymptotic dependence
  - Even if this a valid assumption, there are several choices of bivariate extreme value model
  - Model order?
- Unsure about the effect of the bulk distribution
- Grant application submitted to the Newton Fund to support further work