Application of Bayesian latent Gaussian models to precipitation data in the context of extreme value theory

H. Kári Hlynsson

Abstract

This article constructs Bayesian statistical models for describing extremes in precipitation using data from South England which ranges over a period of 132 years. Three models are considered: (i) a null model with fixed parameters, (ii) a model involving a linear time trend, and (iii) a Bayesian latent Gaussian model with a piecewise linear time trend. Models are fit to data using MCMC sampling and evaluated using various goodness-of-fit estimators, such as the WAIC and DIC, and empirical quantifiers such as MSE from LOOCV.

1 Introduction

Extreme value theory is a branch of statistics that deals with observations deviating very far from the centre of the underlying distribution they are drawn from. The discipline has proven to be very effective in settings where extreme deviations occur with little warning or precedence, such as natural disasters, financial crashes, and seemingly superhuman athletic achievements. In particular, the generalized extreme value (GEV) distribution has seen great use due to the result of the Fisher-Tippett theorem. In brief, the theorem states that for a sequence of independent and identically distributed random variables $(Y_i)_{i\geq 1}$, the limiting distribution of block maxima of the Y_i is necessarily the GEV distribution. Although frequentist methods have seen considerable success in employing the distribution for various purposes, the Bayesian modelling framework remains especially attractive due to its ability to integrate prior knowledge with ease, and to capture uncertainty in parameter estimates in settings where frequentist derivations of confidence intervals are unknown.

In this article, three Bayesian statistical models of precipitation extremes are constructed. The first model is the most simple; the data are fit to a GEV, the parameters of which are fixed in time. The latter two involve a time trend, either unbroken or split across intervals, which aim to capture the effect of e.g. climate change, which some sources state account for a 2-5% increase in precipitation variability globally per year. Another source maintains that an increase of 1K in temperature correspond to an increase of roughly 7% in saturation vapor pressure, tightly coupled to precipitation intensity, demonstrating the intimate link of global warming to increased precipitation due to intensification of the hydrological cycle. Natural disasters that often accompany periods of intense precipitation, such as floods, can cause severe damage to infrastructure and lead to loss of life. The complexity of the underlying physical mechanisms and limited predictability clearly demonstrate the need for effective statistical models, both for preventing and mitigating the damage sustained by such averse events, and to motivate policymakers to install protective regulations. Thus, the application of Bayesian modelling, especially hierarchical methods, have the potential to dampen the effects felt from these types of events, and to elucidate the underlying mechanisms driving them.

Table 1. Table of summary statistics for the South England precipitation data, grouped by decade. The second column, s_y, denotes the sample standard deviation. 95% confidence intervals are computed for the mean using the standard t-distribution formula.

Decade	No. obs.	$\overline{y}$	s_y	95% CI
1890	9	22.3	5.10	[18.4, 26.2]
1900	10	23.3	7.64	[17.8, 28.8]
1910	10	34.3	6.94	[29.4, 39.3]
1920	10	30.3	12.2	[21.7, 39.0]
1930	10	26.0	11.1	[18.0, 34.0]
1940	10	28.8	5.09	[25.1, 32.4]
1950	10	31.6	12.5	[22.7, 40.5]
1960	10	30.3	9.88	[23.2, 37.4]
1970	10	28.9	7.22	[23.8, 34.1]
1980	10	34.0	7.98	[28.3, 39.7]
1990	10	30.9	7.91	[25.3, 36.6]
2000	10	37.8	13.0	[28.5, 47.0]
2010	10	30.7	7.89	[25.1, 36.4]
2020	3	24.5	4.58	[13.1, 35.9]

Table 2. Five number summary of the columns in the South England data.

	Year	Max. precip. [mm]
Min.	1891	10.35
1st Qu.	1924	22.82
Median	1956	27.95
Mean	1956	29.88
3rd Qu.	1989	36.12
Max.	2022	64.22

2 Exploration of data

The data consist of $n = 48212$ observations of cumulative precipitation in millimetres over a 24-hour period from a catchment in South England. These data are aggregated into yearly maxima, shown as a time series in Figure 1. Henceforth, precipitation or cumulative precipitation will be understood as being recorded over a 24-hour period.

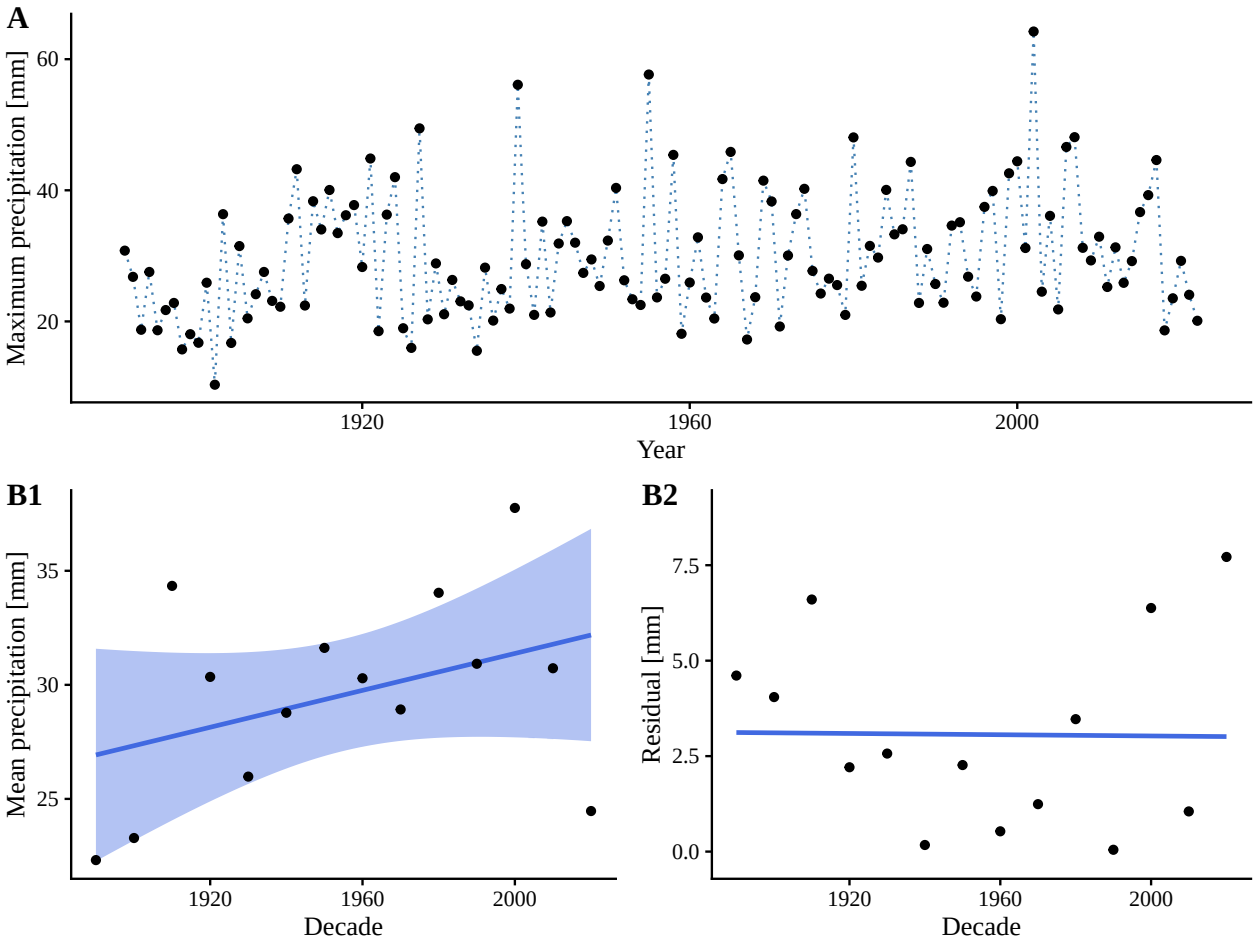

Figure 1. (**A**) Time-series plot of annual maxima of precipitation. **(B1)** Plot of mean maximum precipitation per year by decade. The blue line represents a least squares line and the shaded region represents the 95% confidence interval. **(B2)** Residual plot of mean maximum precipitation per year by decade.

3 Bayesian modelling framework

Throughout the course of this article we will explore three different modelling approaches, all employing a Bayesian framework and MCMC sampling for posterior inference. The notation $\mathcal{M}_j$ will be used to refer to the models.

The first model we consider is a naive fixed-parameter model of the form

$$\begin{aligned} Y_t &\sim \text{GEV}(\mu, \sigma, \xi), \\ \mu &\sim \mathcal{N}(0, \tau_\mu^2), \\ \sigma &\sim \text{Exp}(\lambda_\sigma), \\ \beta &\sim \text{Beta}(4, 4) \end{aligned}$$

where Y_t denotes the maximum annual precipitation at time t.

The second model that is posed is a time-variant model with a linear trend that can be written in

the form

$$
\begin{aligned}
Y_t &\sim \text{GEV}(\mu_t, \sigma, \xi), \\
\mu_t &:= \mu_0\{1 + \Delta(t - t_c)\}, \\
\mu_0 &\sim \mathcal{N}(0, \tau_\mu^2), \\
\Delta &\sim \mathcal{N}(0, \tau_\Delta^2), \\
\sigma &\sim \text{Exp}(\lambda_\sigma), \\
\xi &\sim \text{Beta}(4, 4)
\end{aligned}
$$

where t_c represents the median time of the data set, calculated as $t_c = \lfloor \frac{t_N - t_1}{2} \rfloor$. The last model that we explore is a hierarchical Bayesian model of the form

$$
\begin{aligned}
Y_t &\sim \text{GEV}(\mu_t, \sigma, \xi), \\
\mu_t &:= \sum_{k=1}^{K} \beta_k (t - t_0) \mathbf{1}_{t > t_k}, \\
\sigma &\sim \text{Exp}(\lambda_\sigma), \\
\xi &\sim \text{Beta}(4, 4).
\end{aligned}
$$

In this model, the random effects vector $\boldsymbol{\beta}$ is assigned three different priors;

$$
\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \Sigma_\beta), \quad \Sigma_\beta = \text{diag}(\sigma_\beta), \quad \sigma_\beta \sim \text{Exp}(\lambda)
$$

In other words,

$$
\mu_t := \gamma_t (t - t_0)
$$

Another prior that we will use is the Gaussian random walk. Let