

Esquema Bayesiano para estimar a distribuição de precipitações máximas anuais com duração subdiária em Portugal Continental

A Bayesian approach for estimating the distribution of annual maximum rainfalls with sub-daily duration in Mainland Portugal

Artur Tiago Silva, Maria Manuela Portela

CEris, Instituto Superior Técnico, Universidade de Lisboa

RESUMO: O presente artigo propõe um esquema, enquadrado no paradigma Bayesiano de inferência estatística, para a estimação de valores de projecto de precipitações máximas anuais com durações subdiárias. Tal esquema pretende combinar a informação contida em amostras, expectavelmente curtas, de registos subdiários da precipitação com informação resultante de longas amostras de precipitações diárias máximas anuais e com o conhecimento prévio que advém de estudos antecedentes sobre precipitações intensas. Exemplifica-se a aplicação do procedimento proposto com base em dados da estação meteorológica de Abrantes. Os resultados obtidos validam a aplicação de técnicas Bayesianas de inferência estatística num contexto de escassez de informação hidrológica em futuros estudos sobre precipitações intensas de curta duração em Portugal Continental.

Palavras-chave: Precipitações intensas subdiárias; análise Bayesiana; Monte Carlo via cadeias de Markov (MCMC).

ABSTRACT: *The present article proposes a methodology for estimating design values of sub-daily extreme rainfalls based on the Bayesian paradigm of statistical inference. The proposed methodology combines information from expectedly short samples of sub-daily rainfall extremes with information from longer samples of extreme daily rainfalls and with prior knowledge obtained from previous studies on extreme rainfalls in Portugal. The application of the procedure is exemplified with rain gauge data at Abrantes. The obtained results validate the use of Bayesian techniques under scarcity of hydrologic information in future studies on extreme rainfalls with a short duration in Mainland Portugal.*

Keywords: *Sub-daily extreme rainfalls; Bayesian inference; Markov chain Monte Carlo (MCMC).*

1. INTRODUÇÃO

Em Portugal Continental as observações de valores extremos de variáveis hidrológicas são, por regra, escassas. É este também o caso das precipitações intensas cujo tratamento estatístico, necessário à subsequente caracterização de cheias, é frequentemente confrontado com amostras de dimensão reduzida ou mesmo insuficiente, designadamente, quando estão em causa precipitações subdiárias. Com efeito e não obstante ser possível dispor, de modo geral, de amostras de precipitações diárias máximas anuais com apreciável dimensão num número muito significativo de postos da rede de monitorização, são as precipitações subdiárias que, em consequência da morfologia do País e das reduzidas áreas das bacias hidrográficas urbanas e periurbanas a que conduz, mais interessam à análise de cheias. Apesar de as reduzidas dimensões das escassas amostras disponíveis destas precipitações permitirem o ajustamento de modelos probabilísticos, tal ajustamento não será, por regra, adequado.

Tais circunstâncias motivaram estudos sobre a caracterização de precipitações intensas em Portugal Continental (e.g. Portela, 2006; Brandão *et al.*, 2001; Correia, 2008), cujos resultados permitem, na ausência ou insuficiência de informação udométrica adequada, a estimação aproximada de precipitações intensas de curta duração.

A maioria das abordagens sobre a análise estatística de extremos hidrológicos adota métodos assentes no paradigma frequencista de inferência estatística, mediante ajustamento de modelos de distribuição de probabilidade a variáveis aleatórias. O paradigma Bayesiano de inferência constitui uma alternativa à abordagem frequencista clássica na medida em que fornece um quadro de análise de valores extremos coerente e prático cujas vantagens incluem a possibilidade de: (i) combinar a informação contida na amostra aleatória com outros tipos de informação, por via de funções de densidade *a priori* sobre os parâmetros dos modelos; (ii) formalizar, quantificar e acompanhar a propagação de incertezas na estimação; e (iii) obter estimativas preditivas da variável sob análise que incorporam as

incertezas decorrentes da aleatoriedade do fenómeno e da estimação dos parâmetros.

Do ponto de vista Bayesiano toda a informação útil susceptível de ser formulada matematicamente (qualquer que seja a sua proveniência, tal como juízos ou experiências individuais, evidências de natureza teórica, técnica ou experimental) pode ser utilizada no modelo de inferência, enquanto do ponto de vista clássico ou frequencista apenas a informação contida nos dados observados é suscetível de ser utilizada.

No presente artigo, propõe-se um esquema de estimação da distribuição de precipitações máximas anuais com duração subdiária em Portugal Continental apoiado em métodos de análise Bayesiana, que pretende combinar as amostras observadas dessa precipitação com informação exógena apresentada em literatura da especialidade.

Importa relevar que a aplicação da abordagem Bayesiana em estudos de engenharia de recursos hídricos requer um domínio sólido dos seus fundamentos teóricos cuja exposição em profundidade está para além da inclusão num artigo, mesmo que de natureza científica. Para tal, recomenda-se a leitura das referências Paulino *et al.* (2013), Robert (2007) e Gelman *et al.* (2013). Ainda assim, incluíram-se, sucintamente, na Secção 2, os fundamentos teóricos da estatística Bayesiana que se afiguraram essenciais para a apreciação do esquema de análise proposto.

O artigo prossegue com a descrição da abordagem de inferência Bayesiana implementada (Secção 3) e com a sua aplicação a um caso de estudo, relativo à caracterização das precipitações intensas subdiárias na estação meteorológica de Abrantes (Código SNIRH - 17H/01C, <http://snirh.apambiente.pt/>) (Secção 4). Por fim, na Secção 5, são discutidos os resultados e sistematizadas as principais conclusões do estudo.

Julga-se que o estudo demonstra a utilidade das técnicas Bayesianas na abordagem de problemas relacionados com a engenharia de recursos hídricos, envolvendo o tratamento estatístico de variáveis hidrológicas.

2. FUNDAMENTOS TEÓRICOS

2.1. A distribuição generalizada de valores extremos, GEV

A abordagem mais comum para analisar a frequência de extremos de uma dada variável hidrológica consiste em ajustar uma distribuição de probabilidade à amostra aleatória de máximos anuais dessa variável. No artigo que se apresenta, a variável objeto de estudo é a precipitação máxima anual com duração t , tendo-se adotado como modelo probabilístico a distribuição generalizada de valores extremos (GEV, *generalized extreme value*), cuja função distribuição de probabilidade, FDP, é dada por (Naghetini e Pinto, 2006):

$$F(x | \theta) = \exp \left\{ - \left[1 - \kappa \left(\frac{x - \beta}{\alpha} \right) \right]^{\frac{1}{\kappa}} \right\};$$

$$\theta = [\beta, \alpha, \kappa]^T; \alpha > 0 \quad (1)$$

em que β , α , e κ , são os parâmetros de posição, escala e forma, respetivamente, armazenados no vetor θ .

A utilização da distribuição GEV é fundamentada pelo primeiro teorema da teoria de valores extremos, designadamente o teorema de Fisher-Tippet-Gnedenko (Fisher & Tippet, 1928; Gnedenko, 1990). A distribuição GEV incorpora as três formas assintóticas das caudas superiores das distribuições de máximos, determinadas pelo sinal do parâmetro de forma, κ . Na parametrização adotada neste trabalho, para $\kappa < 0$ e $\kappa > 0$, a GEV representa as distribuições Tipo II de extremos (cauda superior subexponencial ou pesada) e Tipo III de extremos (cauda superior hiperexponencial e limitada). Para $\kappa \rightarrow 0$, a GEV converge assintoticamente para a distribuição de Gumbel ou Tipo I de extremos.

2.2. A função de verosimilhança

A verosimilhança é uma ideia fundamental para a inferência estatística com recurso a modelos paramétricos. O conceito da verosimilhança é transversal aos quadros frequentista e Bayesiano de inferência.

A função de verosimilhança de um modelo estatístico corresponde à densidade conjunta de uma amostra aleatória, em função dos parâmetros do modelo. Considere-se o vector amostral \mathbf{x} , em que x_1, \dots, x_n são realizações independentes de uma variável aleatória com função densidade de probabilidade, $f(x_i | \theta)$. Nessas condições, a função de verosimilhança é dada por

$$L(\theta) = f(\mathbf{x} | \theta) = \prod_{i=1}^n f(x_i | \theta) \quad (2)$$

Numa abordagem frequentista, o estimador de máxima verosimilhança dos parâmetros, $\hat{\theta}$, corresponde ao conjunto de valores dos parâmetros contido em θ que maximiza a função de verosimilhança. O método de estimação de parâmetros por máxima verosimilhança também permite obter valores aproximados de erros-padrão e de intervalos de confiança para parâmetros, com base em resultados assintóticos sob algumas condições de regularidade. Adicionalmente, é possível aplicar o método Delta (Coles, 2001, p. 33; Davison, 2003, p. 33) para construir intervalos de confiança de qualquer função escalar dos parâmetros, $\psi(\theta)$ (ex., um determinado quantil). Acresce que as estimativas, tanto dos parâmetros $\hat{\theta}$, como das suas funções escalares $\psi(\theta)$, são assintoticamente Normais. De acordo com o método delta, a variância assintótica, V_ψ , da resposta de $\psi(\theta)$, é dada por

$$V_\psi = \nabla \psi^T \Sigma \nabla \psi \quad (3)$$

em que $\nabla \psi$ é o vector gradiente de $\psi(\theta)$ no ponto de máxima verosimilhança. No caso dos três parâmetros da GEV, β , α , e κ , resulta,

$$\nabla \psi = \left(\frac{\partial \psi}{\partial \beta} \quad \frac{\partial \psi}{\partial \alpha} \quad \frac{\partial \psi}{\partial \kappa} \right)^T \quad (4)$$

Σ é a matriz (simétrica) de covariâncias do estimador $\hat{\theta}$,

$$\Sigma = \begin{pmatrix} \text{Var}(\hat{\beta}) & \text{Cov}(\hat{\beta}, \hat{\alpha}) & \text{Cov}(\hat{\beta}, \hat{\kappa}) \\ & \text{Var}(\hat{\alpha}) & \text{Cov}(\hat{\alpha}, \hat{\kappa}) \\ & & \text{Var}(\hat{\kappa}) \end{pmatrix} \quad (5)$$

A matriz Σ é dada pela inversa da simétrica da matriz Hessiana, i.e., da matriz das segundas derivadas da função log de verosimilhança, $\ell(\theta) = \ln L(\theta)$, em relação aos parâmetros,

$$\Sigma = \begin{pmatrix} -\frac{\partial^2 \ell}{\partial \beta^2} & -\frac{\partial^2 \ell}{\partial \beta \partial \alpha} & -\frac{\partial^2 \ell}{\partial \beta \partial \kappa} \\ & -\frac{\partial^2 \ell}{\partial \alpha^2} & -\frac{\partial^2 \ell}{\partial \alpha \partial \kappa} \\ & & -\frac{\partial^2 \ell}{\partial \kappa^2} \end{pmatrix}^{-1} \quad (6)$$

calculada no ponto de máxima verosimilhança. Nos casos em que não é possível ou prático o cálculo analítico, tanto do vector $\nabla \psi$, como da matriz Σ , estes podem ser obtidos mediante aplicação de algoritmos comuns de diferenciação numérica.

O erro-padrão da estimativa $\hat{\psi}(\theta)$ pode, assim, ser aproximado por $\sqrt{V_{\psi}}$.

2.3. O paradigma Bayesiano

No paradigma Bayesiano de inferência estatística, a incerteza associada aos parâmetros, θ , de um modelo estatístico é formalmente descrita por distribuições de probabilidade. O teorema de Bayes constitui um quadro lógico de análise que permite rever ou actualizar algum conhecimento ou crença sobre os parâmetros estabelecidos previamente à análise, à luz de nova informação contida na amostra observada \mathbf{x} . O teorema de Bayes refere que

$$\pi(\theta | \mathbf{x}) = \frac{f(\mathbf{x} | \theta) \pi(\theta)}{\int_{\Theta} f(\mathbf{x} | \theta) \pi(\theta) d\theta} \quad (7)$$

em que $\pi(\theta | \mathbf{x})$ e $\pi(\theta)$ são, respectivamente as densidades *a posteriori* e *a priori* dos parâmetros, e $f(\mathbf{x} | \theta)$ é a função de verosimilhança ou densidade conjunta da amostra aleatória \mathbf{x} (Equação 2). A distribuição *a posteriori* dos parâmetros formaliza a incerteza associada à estimação dos parâmetros do modelo.

Na Equação (7), o integral no denominador representa a verosimilhança marginal da amostra, também denominada constante de normalização ou constante de proporcionalidade, cujo domínio de integração corresponde ao espaço paramétrico, Θ , sendo uma função constante dos parâmetros, θ . Então, é comum representar o teorema de Bayes numa forma simplificada que admite que a densidade *a posteriori* é proporcional ao produto da verosimilhança pela densidade *a priori*:

$$\pi(\theta | \mathbf{x}) \propto f(\mathbf{x} | \theta) \pi(\theta) \quad (8)$$

Na prática, na generalidade das aplicações, o cálculo analítico da constante de normalização é muito complexo, impossibilitando, assim, a definição de $\pi(\theta | \mathbf{x})$. No entanto, pode contornar-se essa dificuldade por meio do método numérico Monte Carlo via Cadeias de Markov, MCMC (*Markov chain Monte Carlo*). Tal método consiste em gerar uma cadeia de Markov de dimensão N_{sim} de parâmetros, $(\theta^{(i)})_{i=1:N_{sim}}$, estacionária, ergódica e convergente em distribuição para a distribuição *a posteriori* dos parâmetros, utilizando, para tal, um algoritmo MCMC apropriado aplicado à densidade *a posteriori* não-normalizada, definida pelo segundo termo da Equação (8).

Existem vários algoritmos MCMC descritos na literatura da especialidade (Robert & Casella, 2004; Gamerman & Lopes, 2006), sendo todos casos particulares do algoritmo Metropolis-Hastings (Metropolis & Ulam, 1949; Metropolis et al., 1953; Hastings, 1970). Neste trabalho utilizou-se o algoritmo *random-walk Metropolis* (Metropolis et al. 1953), implementado no pacote LaplacesDemon (Statisticat, 2013), no software estatístico R (R Core Team, 2013).

A exploração da densidade *a posteriori* $\pi(\theta | \mathbf{x})$ via MCMC permite obter directamente informação

sobre a densidade *a posteriori* de qualquer função escalar dos parâmetros, $\psi(\theta)$. Assim, as propriedades amostrais das cadeias MCMC podem ser utilizadas para quantificar a incerteza de parâmetros, quantis ou qualquer outra função escalar $\psi(\theta)$. É frequente, num contexto Bayesiano, a construção, por aproximação numérica, de intervalos de credibilidade ou de probabilidade, que contêm $100(1 - \alpha) \%$ da probabilidade *a posteriori*, correspondendo às regiões de maior densidade *a posteriori*, ou seja

$$\{ \psi: \pi(\psi | \mathbf{x}) > p^* \} \quad (9)$$

em que ψ é a resposta de $\psi(\theta)$ e p^* é tal que

$$\int_{\psi: \pi(\psi | \mathbf{x}) > p^*} \pi(\psi | \mathbf{x}) \, d\psi = 1 - \alpha \quad (10)$$

A inferência Bayesiana de quantis pode ter por base uma estimativa pontual de parâmetros, por exemplo, as médias ou as medianas amostrais dos parâmetros a que se referem as cadeias MCMC ou as modas *a posteriori* desses parâmetros. Segundo Merz & Thieken (2005), o que resulta dessa abordagem é uma distribuição descritiva que representa a variabilidade do processo sob análise. Contudo, essa distribuição não tem em conta a incerteza na estimação dos parâmetros resultante da variabilidade amostral.

O quadro de análise Bayesiano proporciona uma abordagem alternativa para a estimação de quantis de projecto: a distribuição preditiva de acontecimentos futuros W , dada a amostra finita observada \mathbf{x} . A densidade preditiva, $f(w | \mathbf{x})$, é dada por:

$$f(w | \mathbf{x}) = \int_{\Theta} f(w | \theta) \pi(\theta | \mathbf{x}) \, d\theta \quad (11)$$

Assim, a distribuição de $W | \mathbf{x}$ obtém-se por integração de todas as possíveis realizações de θ , o que resulta numa combinação de variabilidade natural e de incerteza amostral

em uma única estimativa de um quantil de projecto (Merz & Thieken, 2005)

Na prática, muito dificilmente se consegue calcular analiticamente o integral da Equação (11). No entanto, é possível obter uma aproximação numérica da probabilidade de não-excedência de W , com base nas cadeias MCMC geradas (Coles, 2001, p.173):

$$\hat{P}(W \leq w | \mathbf{x}) = \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} P(W \leq w | \theta_i) \quad (12)$$

Segundo Coles & Tawn (1996), a inferência com base na distribuição preditiva constitui a abordagem preferível para a estimação de valores de projecto de extremos de variáveis hidrológicas. Tal se deve à sua interpretabilidade, em conjunto com a sua capacidade de sintetizar conjuntamente a variabilidade do fenómeno natural e a incerteza amostral, e, simultaneamente, explorar fontes adicionais de informação através da distribuição *a priori*.

3. ABORDAGEM DESENVOLVIDA

3.1. Informação *a priori* a partir de relações entre precipitações intensas

Na análise local de frequência de uma variável hidrológica é vulgar dispor-se de conhecimentos exógenos às amostras aleatórias utilizadas que, não obstante poderem fornecer informação relevante sobre aquela variável, raramente são considerados na estimação dos parâmetros do modelo postulado para descrever as probabilidades. Tal se deve, muitas vezes, ao desconhecimento de metodologias formais susceptíveis de serem aplicadas para o efeito ou a dificuldades operacionais inerentes à aplicação dessas metodologias. No paradigma Bayesiano de inferência, tais informações, ao formularem um conhecimento ou convicção/crença sobre a variável X previamente à apreciação da correspondente amostra aleatória, podem ser incorporadas na análise por via da distribuição *a priori* dos parâmetros.

A parametrização do modelo probabilístico constitui uma dificuldade imediata na formulação da distribuição *a priori*: dificilmente

o conhecimento antecedente da variável sob análise incide sobre os parâmetros do modelo, mas antes, sobre a variável hidrológica em si. Essa dificuldade é realçada por Coles & Tawn (1996) e por Viglione *et al.* (2013), que recorreram à opinião de especialistas sobre determinados quantis da variável no entendimento de que esses especialistas estariam mais familiarizados com a escala da resposta do modelo do que com a escala dos parâmetros do modelo.

No âmbito da variável hidrológica em estudo – a precipitação intensa em Portugal Continental – também existem vários trabalhos, conforme sistematizado, por exemplo, em Portela (2006), que podem fornecer informação *a priori* sobre aquelas precipitações. De entre esses trabalhos destacam-se Brandão *et al.* (2001) e Correia (2008).

Na investigação que se apresenta, optou-se por explorar a inclusão num modelo Bayesiano de precipitações máximas anuais com duração subdiária de informação relativa a quocientes entre precipitações intensas com diferentes durações. Para o efeito, utilizaram-se os mapas de isolinhas dos quocientes entre precipitações máximas anuais com durações de 1 e de 6 h – P1ma e P6ma – e a precipitação diária máxima anual – Pdma – obtidos por Brandão *et al.* (2001). A probabilidade de não-excedência considerada foi de 0.99 (Brandão *et al.*, pp. 27). Julga-se razoável admitir que tal informação, em conjunto com estimativas de quantis de precipitação diária máxima anual, Pdma, permite obter informação *a priori* sobre um determinado quantil da precipitação máxima anual com aquelas durações subdiárias, desde que as estimativas de Pdma sejam efectuadas com base em amostras de dimensão adequada (i.e., de algumas décadas),

Com efeito e como antes mencionado, em Portugal, as amostras de registos subdiários da precipitação, quando existentes, referem-se a uns escassos anos, embora por regra respeitem a postos com longas amostras de precipitações diárias máximas anuais que são compatíveis com o ajustamento criterioso de um modelo probabilístico. Através da metodologia proposta pretende-se combinar num modelo para o efeito capacitado informação decorrente das próprias amostras curtas de precipitações intensas subdiárias com informação resultante

das longas amostras de Pdma e com o conhecimento prévio que adveio de estudos detalhados sobre precipitações intensas e que foi sintetizado em mapas contendo a caracterização espacial de quocientes entre essas precipitações.

Convém, nesta fase, especificar a nomenclatura utilizada na designação das precipitações intensas para que resulte clara a compreensão das amostras subjacentes ao estudo. De acordo com tal nomenclatura, que respeita a utilizada por Brandão *et al.* (2001), por precipitação diária máxima anual, Pdma, entende-se o valor máximo de precipitação registado em cada ano hidrológico no período de 24 h, necessariamente com início às 9:00 h da manhã de cada dia. As precipitações máximas anuais com uma qualquer duração (no caso estudado, 1 e 6 h, P1ma e P6ma, respectivamente) referem-se também a valores máximos em ano hidrológico, um por cada ano, mas num qualquer período contínuo com aquela duração.

3.2. Hipóteses-base e descrição do procedimento

No estudo subjacente ao presente artigo propõe-se a construção de uma densidade *a priori* para o quantil com probabilidade de não-excedência de 0.99, ou seja, com período de retorno, T, de 100 anos, da precipitação máxima anual com duração de t horas, $P_{tma,0.99}$. Para o efeito houve que adotar hipóteses-base que, não obstante se revestirem de alguma subjectividade, se afiguraram razoáveis e fundamentadas face a resultados teóricos e a estudos antecedentes disponíveis para Portugal Continental.

Considerou-se, assim, que o mencionado quantil teria uma densidade *a priori* Normal, que se designou por $g(P_{tma})$, com média $\mu_{tma,0.99}$ e desvio-padrão $\sigma_{tma,0.99}$, ou seja:

$$P_{tma,0.99} \sim N(\mu_{tma,0.99}, \sigma_{tma,0.99}) \quad (13)$$

A hipótese de normalidade de quantis decorre da teoria assintótica da amostragem e, como tal, aplica-se também ao quantil da precipitação diária máxima anual:

$$P_{dma,0.99} \sim N(\mu_{dma,0.99}, \sigma_{dma,0.99}) \quad (14)$$

Atendendo ao conhecimento sustentado por investigações antecedentes, referente a coeficientes de repartição de precipitações diárias em precipitações subdiárias, concretamente, a quocientes entre precipitações máximas anuais com duração t e a precipitação diária máxima anual, para alguns períodos de retorno, em que se inclui o de 100 anos, K_{td} (Brandão *et al.*, pp. 27-28):

$$K_{td} = \frac{P_{tma_{0,99}}}{P_{dma_{0,99}}} \quad (15)$$

obtêm-se as seguintes relações em consequência do pressuposto de variáveis aleatórias normalmente distribuídas:

$$\begin{cases} P_{tma_{0,99}} = K_{td} P_{dma_{0,99}} \\ \mu_{tma,0,99} = K_{td} \mu_{dma,0,99} \\ \sigma_{tma,0,99} = K_{td} \sigma_{dma,0,99} \end{cases} \quad (16)$$

Torna-se, assim, necessário elicitare valores para a média e para o desvio-padrão do quantil $P_{dma_{0,99}}$. Para o efeito, propõe-se a estimação desses parâmetros pela análise de frequência clássica, ou frequencista, com base numa amostra longa de precipitações diárias máximas anuais. Concretamente, propõe-se para valores de $\mu_{dma,0,99}$ e de $\sigma_{dma,0,99}$, respectivamente, a estimativa do quantil fornecida pela GEV com parâmetros estimados pelo método da máxima verosimilhança, e o erro-padrão associado, calculado pelo método delta, descrito em 2.2.

Estando reunidos os elementos necessários para construir a densidade *a posteriori* não-normalizada dos parâmetros, $f(x|\theta)\pi(\theta)$, e especificando a densidade conjunta ou verosimilhança, $f(x|\theta)$, de acordo com a distribuição GEV e substituindo $\pi(\theta)$ pela função de densidade Gaussiana $g(P_{tma_{0,99}})$, procede-se à utilização do algoritmo MCMC para geração de amostras da distribuição *a posteriori* dos parâmetros.

Importa notar que a informação sobre os parâmetros contida em $g(P_{tma_{0,99}})$ é parcial, i.e., não é possível especificar a densidade $\pi(\theta)$ na sua totalidade, mas apenas sobre a distribuição de um dos três parâmetros da distribuição GEV, condicional nos restantes dois parâmetros. Com

efeito, Viglione *et al.* (2013), numa aplicação de métodos Bayesianos ao estudo de cheias no rio Kamp, na Áustria, demonstraram para a distribuição GEV que $g(P_{tma_{0,99}}) = \pi(\alpha|\beta, \kappa)$, mas também que $g(P_{tma_{0,99}}) = \pi(\beta|\alpha, \kappa)$ e que $g(P_{tma_{0,99}}) = \pi(\kappa|\alpha, \beta)$. Adicionalmente, esses autores concluíram que a escolha formal do parâmetro sobre o qual incide a distribuição *a priori* não influencia o resultado final da análise. Na prática, a densidade *a priori* utilizada no processo MCMC fornece informação sobre valores razoáveis do quantil $P_{tma_{0,99}}$ e, sendo esse quantil uma função dos três parâmetros, sobre os trios de parâmetros conducentes a tais valores, aumentando as suas probabilidades *a posteriori*.

4. EXEMPLO DE APLICAÇÃO COM BASE NUM CASO DE ESTUDO

A aplicação da abordagem desenvolvida é exemplificada com base em registos de precipitação na estação meteorológica de Abrantes (código SNIRH 17H/01C), localizada no Centro de Portugal.

No Quadro 1 apresentam-se as amostras de precipitação máxima anual na anterior estação com durações de 1 e 6 h, P_{1ma} e P_{6ma} , respectivamente. Na sua constituição não se consideraram os anos hidrológicos de 2006/07 e de 2009/10 por apresentarem falhas de registo que não permitem a determinação de máximos anuais. Resultaram, assim, amostras de dimensão $N = 12$, ou seja, insuficiente para sustentar criteriosamente a análise estatística convencional daquelas precipitações.

Os valores dos quocientes, necessários à construção da densidade *a priori*, $g(P_{tma})$, entre precipitações máximas anuais com durações subdiárias, $P_{tma_{0,99}}$, e a precipitação diária máxima anual, $P_{dma_{0,99}}$, ambas para o período de retorno de 100 anos, retirados de Brandão *et al.* (2001, p. 27), tendo em conta a localização geográfica da estação meteorológica de Abrantes (17H/01C), foram os seguintes:

$$\begin{cases} \frac{P_{1ma_{0,99}}}{P_{dma_{0,99}}} = K_{1d} = 0.39 \\ \frac{P_{6ma_{0,99}}}{P_{dma_{0,99}}} = K_{6d} = 0.72 \end{cases} \quad (17)$$

Quadro 1. Amostras de precipitações máximas anuais com duração de 1 h (P1ma) e de 6 h (P6ma) na estação meteorológica de Abrantes (17H/01C).

Ano hidrológico	P1ma (mm)	P6ma (mm)
2001/02	12.4	29.5
2002/03	15.3	24.6
2003/04	16.9	27.8
2004/05	12.4	19.9
2005/06	18.5	40.7
2007/08	13.4	26.3
2008/09	7.8	16.9
2010/11	0.6	2.4
2011/12	5.1	11.9
2012/13	0.9	3.1
2013/14	22.0	28.8
2014/15	11.8	17.1

A Figura 1 mostra o ajustamento da distribuição GEV à amostra de 67 precipitações diárias máximas anuais, Pdma, apresentada no Quadro 2. A figura contém ainda a representação dos intervalos de confiança para quantis a 95%, obtidos mediante a aplicação do método delta. Para estimar a probabilidade empírica de não-excedência, ou posição de *plotagem*,

utilizou-se a fórmula de Gringorten (1963). Na constituição da amostra de Pdma, teve-se o cuidado de excluir os valores relativos aos anos em que existem registos de precipitação com duração subdiária, conforme decorre dos Quadros 1 e 2, por formalmente tais valores não deverem contribuir para a informação *a priori*. É evidente a quase-linearidade da curva de quantis na Figura 1, o que sugere que a distribuição de Gumbel, forma particular da GEV com $\kappa = 0$, também constitui um modelo plausível para a Pdma, com a vantagem de ter apenas dois parâmetros e, conseqüentemente, uma menor incerteza associada às estimativas de quantis. No entanto, importa notar que tal não significa que modelos GEV com $\kappa \neq 0$ não sejam plausíveis. Na questão Gumbel versus GEV optou-se por seguir a recomendação de Coles (2001, p. 64), que argumenta que o incremento de incerteza resultante da consideração de um parâmetro de forma não-fixa permite uma quantificação mais genuína das incertezas inerentes à extrapolação do modelo para probabilidades de não-excedência associadas a eventos mais gravosos do que os contidos na amostra de base. Deste modo optou-se por prosseguir com a distribuição GEV.

Quadro 2. Amostra de precipitações diárias máximas anuais (Pdma) na estação meteorológica Abrantes (17H/01C).

Ano hidrológico	Pdma (mm)	Ano hidrológico	Pdma (mm)	Ano hidrológico	Pdma (mm)	Ano hidrológico	Pdma (mm)
1931 / 32	53.6	1948 / 49	29.0	1965 / 66	59.0	1982 / 83	36.5
1932 / 33	50.2	1949 / 50	29.2	1966 / 67	28.4	1983 / 84	58.0
1933 / 34	38.0	1950 / 51	35.9	1967 / 68	68.3	1984 / 85	34.5
1934 / 35	33.6	1951 / 52	92.0	1968 / 69	62.5	1985 / 86	45.5
1935 / 36	38.0	1952 / 53	22.2	1969 / 70	44.4	1986 / 87	29.8
1936 / 37	42.0	1953 / 54	44.0	1970 / 71	38.5	1987 / 88	38.0
1937 / 38	62.5	1954 / 55	43.6	1971 / 72	50.6	1988 / 89	53.0
1938 / 39	32.0	1955 / 56	55.2	1972 / 73	47.7	1989 / 90	65.4
1939 / 40	50.0	1956 / 57	50.4	1973 / 74	75.4	1990 / 91	45.2
1940 / 41	42.5	1957 / 58	24.8	1974 / 75	53.7	1991 / 92	24.8
1941 / 42	52.8	1958 / 59	27.4	1975 / 76	52.6	1992 / 93	29.2
1942 / 43	60.0	1959 / 60	31.6	1976 / 77	33.0	1993 / 94	30.0
1943 / 44	18.6	1960 / 61	25.4	1977 / 78	35.2	1994 / 95	32.7
1944 / 45	29.8	1961 / 62	47.2	1978 / 79	45.1	1995 / 96	74.5
1945 / 46	30.0	1962 / 63	43.2	1979 / 80	65.0	1996 / 97	38.8
1946 / 47	25.6	1963 / 64	34.2	1980 / 81	27.3	1999 / 00	40.5
1947 / 48	37.0	1964 / 65	45.2	1981 / 82	61.4		

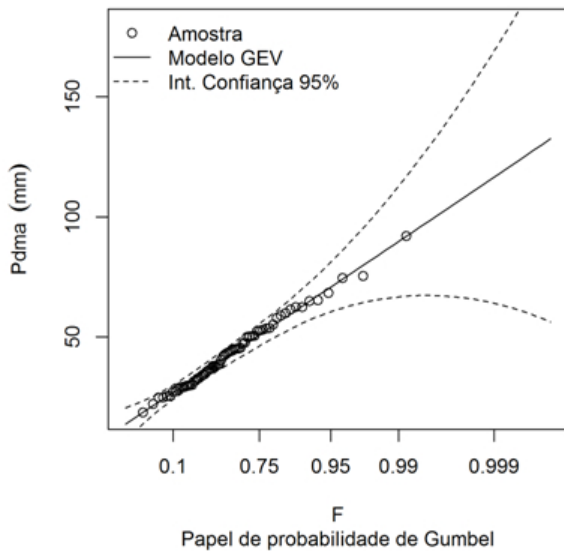


Figura 1. Estação meteorológica Abrantes (17H/01C). Distribuição GEV ajustada à amostra de Pdma e intervalos de confiança a 95% obtidos pelo método Delta.

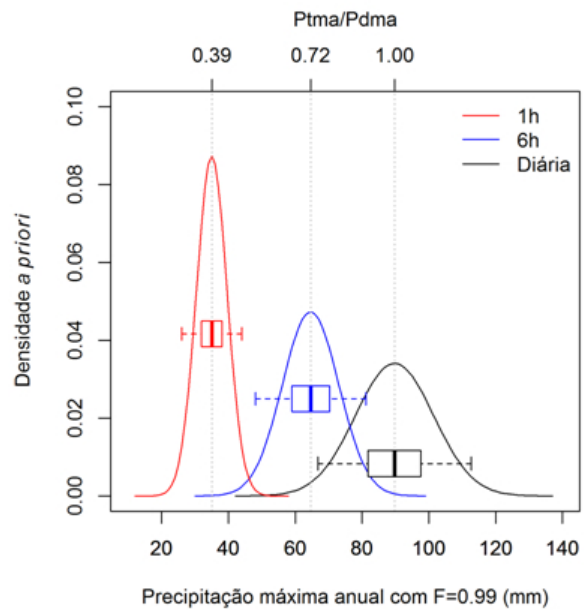


Figura 2. Estação meteorológica Abrantes (17H/01C). Densidades *a priori* do quantil para a probabilidade de não-excedência de 99% da precipitação máxima anual com durações de 1 e 6 h e diária, $P1ma_{0.99}$, $P6ma_{0.99}$ e $Pdma_{0.99}$, respectivamente. Os *boxplots* representam os percentis 2.5, 25, 50, 75 e 97.5%, da densidade *a priori*.

Mediante a implementação do procedimento descrito na Secção 3.2, obtiveram-se as densidades *a priori* para o quantil com $F = 0.99$ de $P1ma$ e de $P6ma$, que estão representadas na Figura 2, em conjunto com *boxplots* que assinalam os respectivos percentis 2.5%, 25%, 50%, 75% e 97.5%. Essas densidades *a priori* representam a informação exógena às amostras do Quadro 1, que se pretende incluir na análise Bayesiana dessas mesmas amostras.

A amostragem da distribuição *a posteriori* dos parâmetros por métodos MCMC, bem como a validação das cadeias geradas (aspectos não explicitados no artigo), recorreu a funções do pacote LaplacesDemon (Statisticat, 2013), no software estatístico gratuito R (R Core Team, 2013).

A título de exemplo, apresentam-se na Figura 3, gráficos de dispersão das amostras de parâmetros *a posteriori* geradas pelo algoritmo MCMC para a variável $P1ma$. Essa figura também apresenta os histogramas *a posteriori* marginais dos parâmetros.

Finalmente, nas Figuras 4a e 4b, apresentam-se os resultados da análise Bayesiana das precipitações máximas anuais com duração de 1 e 6 h, respectivamente, conjuntamente com a representação das amostras dessas

precipitações. Nas figuras, os *boxplots* atribuídos à abcissa $F = 0.99$ correspondem aos apresentados na Figura 2 para cada duração e representam a densidade *a priori* para esse quantil, ou seja, não tendo em conta a informação contida nas amostras de $P1ma$ e $P6ma$. A incerteza na estimação de quantis é quantificada pelos intervalos de credibilidade os quais representam os limites das regiões de maior densidade *a posteriori* de cada quantil que contêm 95% da probabilidade *a posteriori*. As Figuras 4a e 4b representam a distribuição Bayesiana preditiva de realizações futuras das variáveis $P1ma$ e $P6ma$ sob análise. Tal distribuição resume em uma única estimativa: (i) a variabilidade do fenómeno descrita pelas amostras aleatórias do Quadro 1; (ii) a incerteza amostral, mediante a integração da distribuição *a posteriori* no espaço paramétrico, através da Equação (11) e da sua resolução numérica, dada pela Equação (12); e (iii) a informação obtida a partir de conhecimento exógeno à amostra aleatória, através da densidade *a priori*, $\pi(\theta)$.

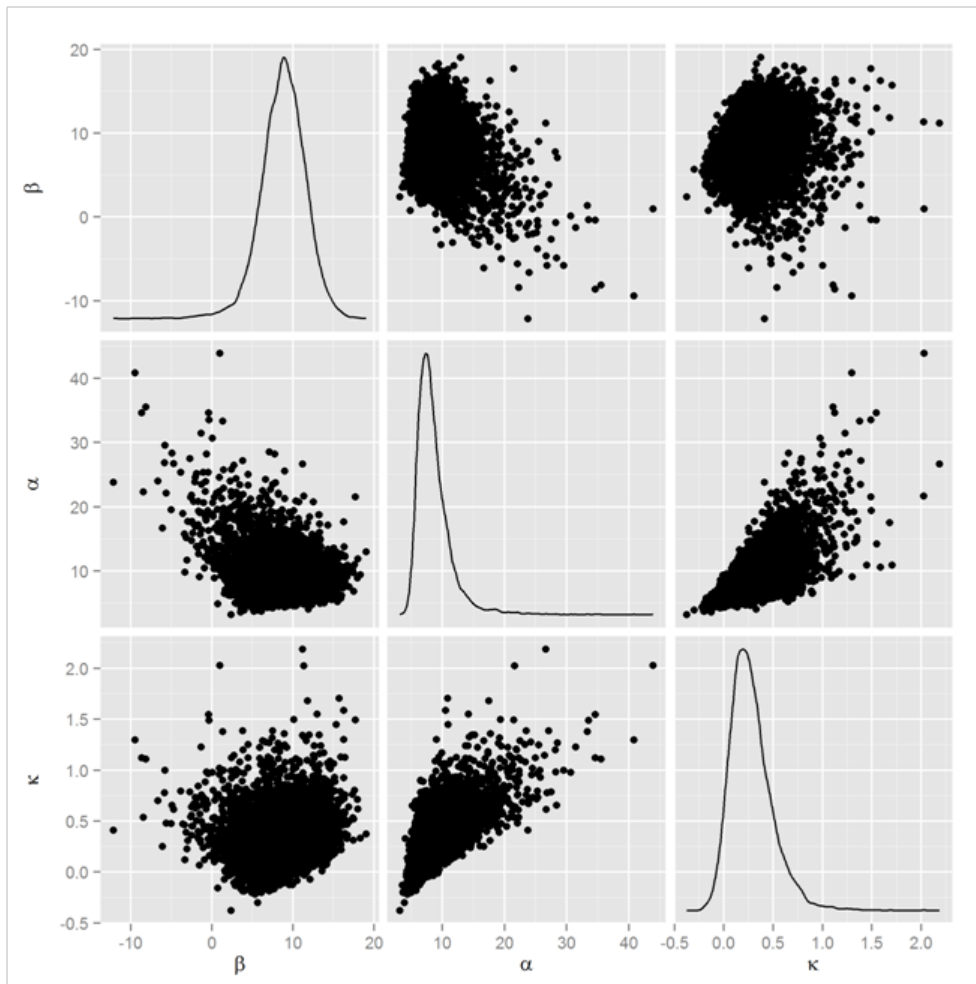


Figura 3. Modelo P1ma: diagramas de dispersão das amostras dos parâmetros α *posteriori* geradas por MCMC e, na diagonal, histogramas marginais dos parâmetros.

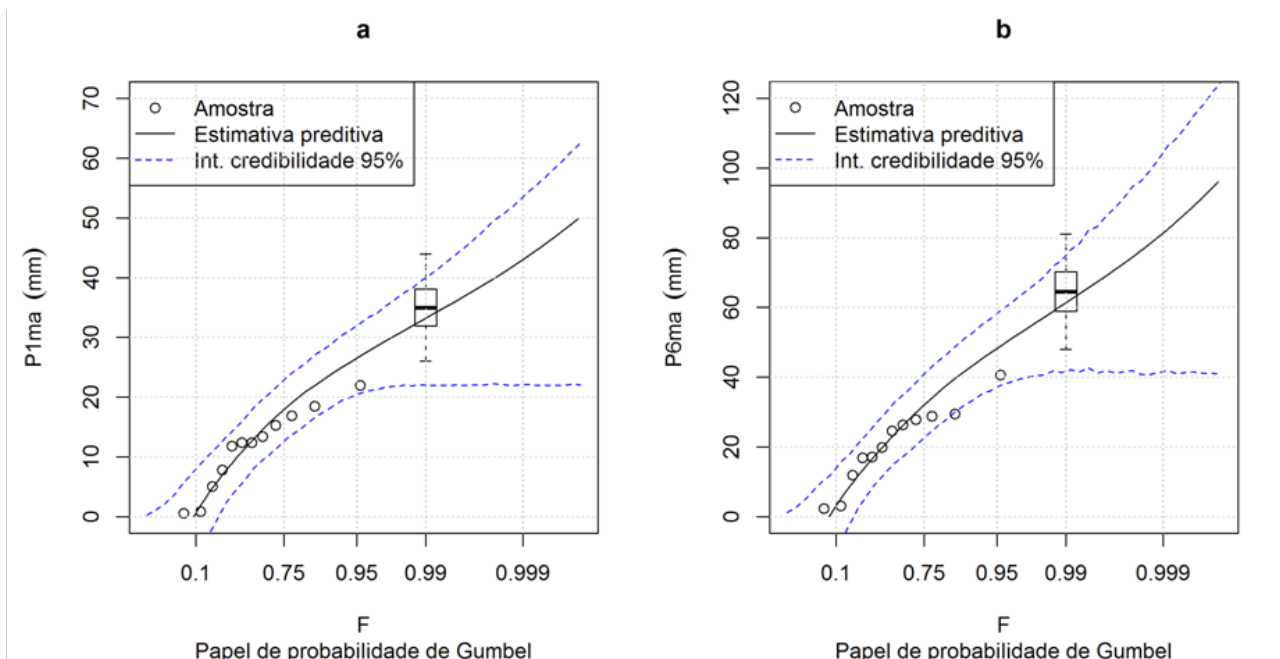


Figura 4. Estação meteorológica Abrantes (17H/01C). Precipitações máximas anuais com duração de 1 h (a) e de 6 h (b). Distribuição preditiva Bayesiana e intervalos de credibilidade a 95% que definem as regiões de maior densidade *a posteriori* dos quantis. Os *boxplots* representam os percentis 2.5, 25, 50, 75 e 97.5% da densidade *a priori* dos quantis com $F = 0.99$.

5. CONCLUSÕES

Apresenta-se um esquema de análise estatística de precipitações intensas com duração subdiária, assente num quadro Bayesiano de inferência. O trabalho foi motivado pela assinalável escassez de dados de precipitações de duração subdiária em Portugal Continental directamente acessíveis ao público.

Os resultados obtidos demonstram a possibilidade de obter uma distribuição de precipitações máximas anuais subdiárias fundamentada numa combinação da informação contida numa amostra, com poucos anos de registos, com informação antecedente sobre um quantil específico. Para o efeito foi necessário admitir algumas hipóteses-base que se afiguraram razoáveis e fundamentadas face a resultados teóricos e a estudos antecedentes disponíveis para Portugal Continental.

A adopção de um quadro de análise Bayesiano terá, porventura, como vantagem mais óbvia relativamente ao quadro de análise frequentista, a possibilidade de formalmente incluir na análise estatística informação exógena à amostra aleatória. Existem, no entanto, outras vantagens na abordagem Bayesiana, designadamente a formalização probabilística das incertezas da estimação de parâmetros através da distribuição *a posteriori*, e a possibilidade de integrar a variabilidade do fenómeno natural e a incerteza amostral em uma única estimativa de projecto dada pela distribuição preditiva.

A investigação apresentada neste artigo abre algumas oportunidades de desenvolvimentos futuros, designadamente: (i) o estudo das implicações da adopção de outro quantil para a incorporação da informação *a priori*; (ii) as implicações decorrentes da eventual não-normalidade dos quantis no que respeita à elicitação da densidade *a priori*; e (iii) a possível extensão da metodologia no sentido da avaliação de dados não-sistemáticos de precipitações intensas.

Os métodos desenvolvidos neste artigo poderão ser utilizados em próximos estudos de análise estatística de precipitações intensas de curta duração em Portugal Continental ou em outras regiões onde seja possível a recolha de informação sobre quocientes entre precipitações com essas durações e com duração diária, ou outra informação sobre algum quantil específico dessas precipitações.

BIBLIOGRAFIA

Brandão, C., Rodrigues, R., Costa, J. (2001). Análise de fenómenos extremos. Precipitações intensas em Portugal Continental. Lisboa, DSRH. http://snirh.pt/snirh/download/relatorios/relatorio_prec_intensa.pdf

Coles, S. G., Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Applied statistics*, 463-478.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

Correia, M.G.P. (2008) Análise da precipitação máxima com média duração em Portugal Continental. Tese de Mestrado. Instituto Superior Técnico, Lisboa, Portugal.

Davison, A. C. (2003). *Statistical Models*, Cambridge University Press.

Fisher, R., e Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24. Cambridge Univ Press.

Gamerman, D., Lopes, H.F. (2006). *Markov chain Monte Carlo: stochastic simulations for Bayesian inference*. CRC Press

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC Press.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *The Annals of Mathematics*: 44(3), 423-453.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97-109.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), 1087-1092.

Metropolis, N., Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, 44(247), 335-341.

- Merz, B., Thielen, A. H. (2005). Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology*, 309(1), 114-132.
- Naghetini, M., Pinto, E. (2007). *Hidrologia Estatística*, CPRM, Belo Horizonte, Brasil. 2007. <http://www.cprm.gov.br/publique/cgi/cgilua.exe/sys/start.htm?infoid=981&sid=36>
- Paulino, C.D.M, Turkman, M.A.A., Murteira, B. (2003) *Estatística Bayesiana*. Fundação Calouste Gulbenkian.
- Portela, M.M. (2006). Estimaco de Precipitaoes Intensas em Bacias Hidrogrficas de Portugal Continental, *Recursos Hdricos*, Vol. 27(1), pp. 15-32.
- RCoreTeam (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementations*. Springer.
- Robert, C., Casella, G. (2004). *Monte Carlo statistical methods*. Springer.
- Statisticat, LLC (2013). *LaplacesDemon: Complete Environment for Bayesian Inference. R packages version 13(04)*
- Viglione, A., Merz, R., Salinas, J.L., Blschl, G. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, 49(2), 675-692.