

# Ajuste da Distribuição Generalizada de Pareto a Densidade de Corrente e Análise da Correlação Com Volatilidade Para Dias Com Lâminas de Corrente no Vento Solar Usando Ferramentas Computacionais

Bruno Vilela Quirino

Orientadores: Tiago F. P. Gomes, Eliseu César Miguel

Outubro 2023

## Resumo

O vento solar, uma corrente contínua de partículas carregadas emitidas do Sol para o espaço interplanetário, desempenha um papel significativo nos estudos de meteorologia espacial. Compreender a dinâmica do vento solar e seus efeitos na Terra é crucial para mitigar potenciais riscos para infraestruturas terrestres e espaciais.

Por meio da análise de séries temporais do vento solar, focamos em eventos de lâminas de corrente, que podem ter consequências severas como interrupções de comunicações, desligamentos de rede elétrica e danos a satélites. Usando a técnica de *curlometer*, comprovamos que para dias com lâminas de corrente nossa distribuição GPD possui parâmetro de forma positivo, enquanto para dias sem lâminas, parâmetro negativo. Com essa informação somos capazes de prever eventos de reconexão magnética, que são os maiores causadores de danos a equipamentos como satélites.

Além disso, ficou comprovado que a nova *proxy* aproximação para a densidade de corrente, denominada volatilidade [19], consegue ser uma melhor aproximação do que a PVI (*Partial Variance of Increments*), que até então era o indicador mais utilizado na área, para dias com lâminas de corrente.

Reunimos todas as ferramentas estatísticas e astronômicas relevantes utilizadas nesse estudo em um pacote Python. Ao combinar e integrar essas ferramentas, oferecemos aos pesquisadores a capacidade de realizar análises mais eficientes e replicáveis, abrindo novas oportunidades para estudos interdisciplinares. A criação desse pacote Python fornece uma plataforma acessível e poderosa para explorar a dinâmica do vento solar e seus impactos na meteorologia espacial, além de ter aplicações cruciais em diversas áreas, incluindo a proteção de satélites, a antecipação de eventos que podem afetar viagens interplanetárias e as operações de satélites no espaço.

## 1 Introdução

O objetivo do campo de estudo da meteorologia espacial é monitorar e prever eventos relacionados à atividade solar e ao meio interplanetário, bem como seu impacto na Terra [8]. Esses eventos incluem erupções solares, ejeções de massa coronal e o vento solar, sendo responsáveis por fenômenos como tempestades magnéticas e subtempestades [4], que podem causar interferências em sistemas de transporte, energia, telecomunicações e navegação [16].

O vento solar é particularmente importante nos estudos de meteorologia espacial, devido à sua interação contínua com a magnetosfera da Terra. Sabe-se que o vento solar é permeado por estruturas coerentes, como tubos magnéticos e lâminas de corrente [17]. Estas últimas constituem o principal fator responsável pela multifractalidade e afastamento da gaussianidade nas estatísticas das séries temporais do campo magnético [10], bem como na turbulência intermitente [13]. Tais fenômenos estão

intimamente relacionados à ocorrência de eventos extremos, que são eventos com alta amplitude em comparação com a média da distribuição. Por exemplo, as interações não lineares do campo magnético no vento solar em pequenas escalas resultam na formação de intensas lâminas de corrente, que podem estar associadas a reconexões magnéticas [14] [15]. Essas reconexões aceleram partículas, causando uma liberação de energia de alta intensidade no plasma espacial [12]. A análise estatística de eventos extremos tem sido amplamente empregada em estudos de fenômenos naturais, como velocidade do vento [7], secas e inundações [6], terremotos [3] e doenças [18]. A teoria dos valores extremos [5] compreende um conjunto de técnicas avançadas e ferramentas estatísticas para modelar as funções de distribuição de séries temporais que exibem eventos extremos.

O primeiro objetivo deste trabalho consiste em reforçar a evidência [19] de que, de fato, o campo magnético no vento solar segue uma distribuição de valores extremos com parâmetro positivo quando há lâminas de corrente, e uma distribuição de Pareto com parâmetro negativo quando não há lâminas. Essa compreensão aprofundada do comportamento do campo magnético possibilitará a melhoria das estratégias de proteção para satélites em órbita, permitindo o desenvolvimento de sistemas mais robustos capazes de resistir às variações e perturbações do campo magnético solar.

O segundo objetivo deste trabalho é demonstrar que a volatilidade, utilizada como *proxy* em [19], é de fato uma boa representação da densidade de corrente, que pode ser medida usando a técnica do *curlometer* [20]. Ao estabelecer a relação entre a volatilidade e a densidade de corrente real, será possível utilizar essa medida indireta para monitorar e prever as variações no campo magnético do vento solar, o que contribuirá diretamente para a proteção de satélites em órbita.

Por fim, o terceiro objetivo deste estudo está relacionado à comparação entre a volatilidade utilizada neste trabalho e a *proxy Partial Variance of Increments (PVI)* existente na literatura. Ao demonstrar que a volatilidade utilizada neste estudo é superior em termos de correlação com a densidade de corrente, estaremos fornecendo uma ferramenta mais precisa e confiável para a avaliação e proteção de satélites contra os efeitos do campo magnético solar.

O ajuste da GPD (Generalized Pareto Distribution) aos dados de ventos solares obtidos pela técnica do *Curlometer* representa um avanço significativo nesta área. Este trabalho não apenas melhora nossa compreensão das complexidades das interações solares-terrestres, mas também possui implicações práticas na prevenção de tempestades solares potencialmente danosas. A natureza imprevisível dos ventos solares requer uma abordagem estatística robusta - um trabalho que não só tem o potencial de enriquecer o campo da ciência solar, mas também de moldar de maneira significativa as estratégias de mitigação de riscos e a previsão meteorológica espacial.

Em paralelo, a descoberta da *proxy* volatilidade para densidade de corrente amplia ainda mais a fronteira deste campo de estudo. Esta descoberta permitirá aos cientistas a criação de modelos mais precisos para análise e interpretação dos dados de ventos solares. Ao fornecer uma métrica adicional para avaliar a densidade do fluxo de partículas carregadas emitidas pelo sol, a *proxy* volatilidade nos dá uma visão mais abrangente dos processos solares.

Na seção 2 iremos definir as metodologias utilizadas, na seção 3 iremos falar um pouco sobre o pacote auxiliar criado, na seção 4 iremos definir os dados, suas particularidades e pré-processamento realizados necessários, na seção 5 iremos discutir os resultados obtidos, e por fim, na seção 6 concluir e apresentar possíveis trabalhos futuros.

## 2 Metodologia

A Teoria de Eventos Extremos (TEE) é um ramo da estatística que se concentra no estudo das distribuições de probabilidade que representam variáveis aleatórias extremas. Esta teoria foi desenvolvida como uma forma de prever o comportamento de grandes variações em uma série temporal que normalmente não são facilmente previstos usando técnicas estatísticas tradicionais, como por exemplo, modelos de árvore de decisão ou regressão logística. A TEE tenta identificar padrões extremos que possam ser usados para prever o comportamento futuro. Os principais objetivos da TEE são a previsão de eventos extremos e a análise de riscos.

## 2.1 Modelos de Excessos Acima de um Limiar (Distribuição Generalizada de Pareto - GPD)

Para capturar o maior número possível de valores extremos em um conjunto de variáveis aleatórias iid (independentes e igualmente distribuídas) com função de distribuição  $F$ , podemos utilizar um modelo de excessos acima de um limiar. Nesse modelo, consideramos um limiar  $u$  acima do qual todos os valores da série são considerados extremos.

A probabilidade condicional da ocorrência de valores extremos acima do limiar é dada pela fórmula:

$$\Pr(X > u + y | X > u) = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0 \quad (1)$$

Essa fórmula nos permite determinar a função de distribuição dos valores acima do limiar, desde que a função de distribuição  $F$  seja conhecida. No entanto, na prática, é necessário fazer certas aproximações para estimar essa distribuição.

O modelo de excessos acima de um limiar é útil para capturar valores extremos sem desperdiçar uma quantidade considerável deles, como acontece no modelo de máximo em blocos. Ao escolher adequadamente o limiar  $u$ , é possível obter uma série de valores extremos que podem ser modelados usando a Distribuição Generalizada de Pareto (GPD).

## 2.2 Modelagem de Valores Extremos Acima de um Limiar

Considere o seguinte teorema [2], [1]:

**Teorema 1:** *Seja  $X_1, X_2, \dots$  uma sequência de variáveis aleatórias independentes e igualmente distribuídas e*

$$M_n = \max \{X_1, \dots, X_n\}. \quad (2)$$

*Suponha que, para valores altos de  $n$ , a função de distribuição acumulada de  $M_n$  ( $F$ ) possua a seguinte forma:*

$$\Pr \{M_n \leq z\} \approx G(z), \quad (3)$$

onde

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (4)$$

*para algum  $\mu, \sigma > 0$  e  $\xi$ . Então, para um valor de limiar  $u$  suficientemente alto, a função de distribuição dos valores extremos acima do limiar ( $X - u$ ), condicionada a  $X > u$ , é aproximadamente*

$$H(y) = \begin{cases} 1 - \left( 1 + \frac{\xi y}{\tilde{\sigma}} \right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp \left( -\frac{y}{\tilde{\sigma}} \right), & \xi \rightarrow 0, \\ \frac{y}{\tilde{\sigma}}, & \xi \rightarrow -1, \end{cases} \quad (5)$$

*definida para  $y > 0$  e  $\left[ 1 + \left( \xi y / \tilde{\sigma} \right) \right] > 0$ , onde*

$$\tilde{\sigma} = \sigma + \xi(u - \mu), \quad (6)$$

*em que  $(\sigma, \mu, \xi)$  é um conjunto de parâmetros da distribuição generalizada de valores extremos (GEV) [5].*

Nesse contexto, se  $(\xi > 0)$   $\xi < 0$ , a distribuição é (não) limitada superiormente, ou seja, possui um suporte (infinito) finito, dado por  $(y = ]0, \infty[)$   $y = ]0, u - (\tilde{\sigma} / \xi)[$  [5].

Também é importante observar pela equação 5 que, se  $\xi \rightarrow -1$  e  $\xi \rightarrow 0$ , então  $H$  é reduzida, respectivamente, para as distribuições uniforme  $U(0, \tilde{\sigma})$  e exponencial com média  $\tilde{\sigma}$ . Portanto, essas distribuições são casos particulares da distribuição GPD.

### 2.3 Inferência Estatística para Distribuição GPD

Inferência Estatística para Distribuição GPD é um método de análise estatística usado para estimar parâmetros de distribuições GPD (generalized Pareto distribution; distribuição de Pareto generalizada). O método usa inferência bayesiana para estimar parâmetros de distribuição GPD  $(\sigma, \xi)$  a partir de amostras de dados reais. O método também pode ser usado para avaliar a qualidade de ajuste de um determinado modelo de distribuição GPD e para prever a probabilidade de ocorrência de eventos futuros. O método Inferência Estatística para Distribuição GPD é amplamente usado na análise de dados complexos, especialmente em áreas como ciência da computação, medicina e finanças.

Sejam as variáveis aleatórias  $(y_i : i = 1, \dots, n)$  (iid), dadas pelos excessos acima do limiar  $u$  ( $y_i = x_i - u$ ) e com distribuição GPD. Neste caso, a função de verossimilhança é dada por,

$$L(\sigma, \xi) = \prod_{i=1}^n h(y_i; \sigma, \xi), \quad (7)$$

onde  $h(y; \sigma, \xi)$  é a densidade de probabilidade conjunta. Trabalhando com o logaritmo de 7, obtemos a função de log-verossimilhança, dada por

$$\begin{aligned} \ell(\sigma, \xi) &= \log L(\sigma, \xi) \\ &= \sum_{i=1}^n \log [h(y_i; \sigma, \xi)] \\ &= \sum_{i=1}^n \log \frac{\partial [H(y_i; \sigma, \xi)]}{\partial y_i}. \end{aligned} \quad (8)$$

Portanto, substituindo (5) em (8), obtemos

$$\ell(\sigma, \xi) = \begin{cases} -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log \left(1 + \frac{\xi y_i}{\sigma}\right), & \xi \neq 0 \\ -n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^n y_i, & \xi \rightarrow 0. \end{cases} \quad (9)$$

dado que  $(1 + \xi y_i / \sigma) > 0$  para  $i = 1, \dots, n$  e  $l(\sigma, \xi) = -\infty$  caso contrário. Os parâmetros  $\xi$  e  $\sigma$  estimados, são aqueles que maximizam (9).

Conforme já mencionado anteriormente, um dos principais objetivos da inferência estatística é a previsão futura. Em se tratando de eventos extremos, este objetivo torna-se uma necessidade, visto o impacto por eles gerado. A exemplo do caso da distribuição GEV, também é possível obter os níveis e períodos de retorno para a GPD. Para isso, seja uma distribuição GPD com parâmetros  $\sigma$  e  $\xi$ , modelando excessos acima de limiar  $x > u$  de uma variável aleatória  $X$ , ou seja,

$$\Pr \{X > x | X > u\} = \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi}. \quad (10)$$

Disso, segue que

$$\Pr \{X > x\} = \varsigma_u \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi}, \quad (11)$$

onde  $\varsigma_u = \Pr(x > u)$ . Logo, o nível  $x_m$  que é em média excedido uma vez a cada  $m$  observações, é a solução de

$$\varsigma_u \left[ 1 + \xi \left( \frac{x_m - u}{\sigma} \right) \right]^{-1/\xi} = \frac{1}{m}. \quad (12)$$

Rearranjando os termos, obtemos

$$x_m = u + \frac{\sigma}{\xi} \left[ (m\varsigma_u)^\xi - 1 \right], \quad (13)$$

dado que  $m$  é suficientemente grande, de modo a garantir que  $x_m > u$ . Tudo isso assume que  $\xi \neq 0$ . Caso  $\xi = 0$ , então podemos escrever a equação 13 da seguinte forma [5]:

$$x_m = u + \sigma \log(m\varsigma_u) \quad (14)$$

De (13) e (14), o gráfico de  $x_m$  em função de  $m$ , em uma escala logarítmica, produz os mesmos comportamentos qualitativos dos níveis de retorno da GEV, ou seja, linearidade se  $\xi = 0$ , concavidade se  $\xi > 0$  e convexidade caso  $\xi < 0$  [5].

## 2.4 PVI

A PVI [11], é uma técnica que foi fortemente validada em diferentes condições, para identificação de lâminas de corrente [11]. A técnica, para eventos extremos, pode detectar reconexão magnética no vento solar[15]. Ela é definida como:

$$\text{PVI}(t, \tau) = \frac{|\Delta B(t, \tau)|}{\sqrt{\langle |\Delta B(t, \tau)|^2 \rangle}} \quad (15)$$

onde  $\Delta B(t, \tau) = B(t + \tau) - B(t)$  são os incrementos avaliados em escala temporal  $\tau$ . A função pode ser computada espacialmente em simulações ou em séries temporais de campo magnético, assumindo a hipótese de Taylor.

## 2.5 Detecção de lâminas de correntes no vento solar - Método de Li

O método de identificação de lâminas de corrente no vento solar de Li [9] é uma técnica usada para detectar e estudar estruturas chamadas de *lâminas de corrente* dentro do vento solar. Essas lâminas de corrente são camadas finas e alongadas no plasma do vento solar, onde o campo magnético do plasma é altamente torcido. O algoritmo baseia-se no fato de que as lâminas de corrente produzem características específicas nas medições do campo magnético. Em termos simples, o algoritmo de Li procura por regiões no vento solar onde o campo magnético muda rapidamente de direção ou onde há uma grande variação na intensidade do campo magnético. Essas regiões são indicativas da presença de lâminas de corrente.

Considere a série temporal de campo magnético  $\vec{B}(t)$ . Então, pode-se definir a densidade de probabilidade de encontrar ângulos de rotação entre dois vetores defasados temporalmente por uma defasagem (lag)  $\tau$ ,  $\vec{B}(t + \tau)$  e  $\vec{B}(t)$ , dentro de um intervalo angular  $\theta$  e  $\theta + \Delta\theta$ , como [9]

$$f(\theta, \tau) \Delta\theta = \frac{N^\tau (\theta < \theta' < \theta + \Delta\theta)}{N^\tau (0 < \theta' < \pi)}, \quad (16)$$

onde  $N^\tau (\theta < \theta' < \theta + \Delta\theta)$  é o número de pares de vetores de campo magnético  $\vec{B}(t + \tau)$  e  $\vec{B}(t)$  cujos ângulos entre eles esteja entre  $\theta$  e  $\theta + \Delta\theta$ . Diante disto, pode-se definir uma função de distribuição [9], dada por

$$F(\theta, \tau) = \int_\theta^\pi d\theta' f(\theta', \tau), \quad (17)$$

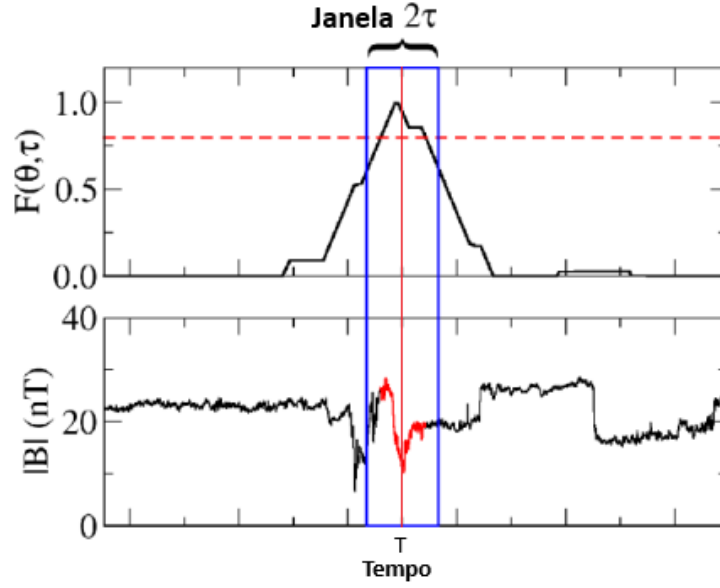


Figura 1: Figura ilustrando a detecção de uma região contendo lâmina de corrente (pontos vermelhos) centrada em  $T$ . Neste caso foi utilizada uma janela deslizante de tamanho  $2\tau$ , delimitada pelas linhas azuis. O limiar  $L$  é representado pela linha tracejada vermelha. O parâmetro  $\tau$  define a escala típica da lâmina de corrente detectada.

que nada mais é que a frequência de ocorrência de medidas de ângulos maiores que  $\theta$ . Logo, se houver lâminas de correntes na série temporal de  $\vec{B}(t)$ , então a quantidade  $F(\theta, \tau)$  deve obedecer uma lei de escala linear com o lag  $\tau$  [9], quando  $\theta$  for maior que um ângulo crítico  $\theta_0$ , ou seja,

$$F(\theta, N\tau) \sim NF(\theta, \tau) \quad \text{quando } \theta > \theta_0. \quad (18)$$

Nesse caso, para cada janela de tamanho  $2\tau$  e centrada em  $T$ , conforme a Figura 1, verificamos a quantidade de ângulos entre  $\vec{B}(t + \tau)$  e  $\vec{B}(t)$  que são maiores que o ângulo crítico  $\theta_0$ . Se esta quantidade for maior que um certo limiar devidamente escolhido  $L$ , então é identificada uma lâmina de corrente em  $T$ .

Vale a pena mencionar que esta técnica somente consegue detectar lâminas de correntes com a mesma escala de tamanho da janela  $2\tau$ .

## 2.6 Técnica de *curlometer*

O *curlometer* é uma técnica utilizada pela Agência Espacial Europeia (ESA) para medir as correntes elétricas no espaço e analisar os processos de transferência de energia entre o Sol e a Terra. Essa técnica baseia-se na medição dos campos magnéticos e elétricos em quatro pontos diferentes no espaço, usando satélites do Cluster Mission da ESA.

A técnica de análise aplicada neste artigo combina diretamente dados simultâneos entre as diferentes espaçonaves para calcular o rotacional do campo magnético. Usamos a lei de Ampère para estimar a densidade média da corrente através do tetraedro formado pela configuração das espaçonaves, usando a aproximação da diferença

$$\begin{aligned} \mu_0 \mathbf{J} \cdot (\Delta \mathbf{r}_i \wedge \Delta \mathbf{r}_j) &= \Delta \mathbf{B}_i \cdot \Delta \mathbf{r}_j - \Delta \mathbf{B}_j \cdot \Delta \mathbf{r}_i \\ \left\{ \text{representando : } \mu_0 \int \mathbf{J} \cdot d\mathbf{s} &= \oint \mathbf{B} \cdot d\mathbf{l} \right\} \end{aligned}$$

com  $\Delta \mathbf{r}_i \equiv \mathbf{r}_i - \mathbf{r}_l$ , e similarmente  $\Delta \mathbf{B}_i \equiv \mathbf{B}_i - \mathbf{B}_l$  (veja Dunlop et al. [1988] para uma derivação). Isso efetivamente estima a corrente média normal à face (1,i,j) do tetraedro 2.

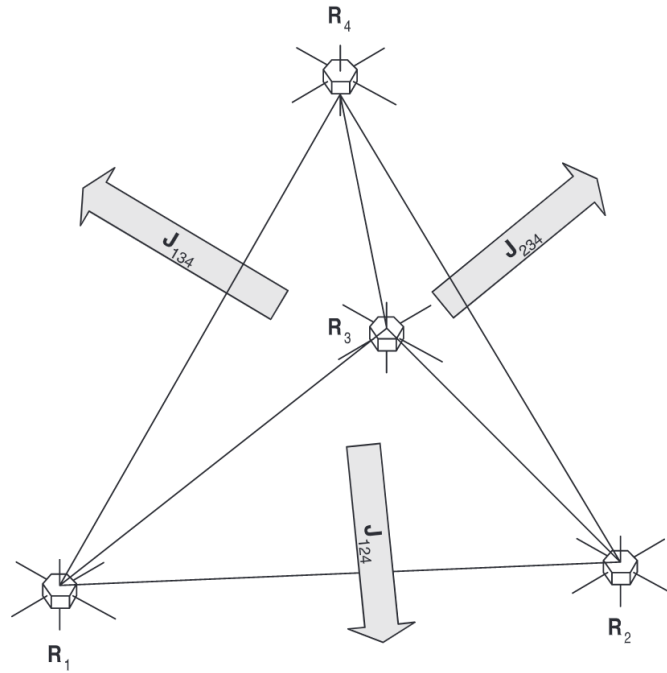


Figura 2: Visualização dinâmica das quatro espaçonaves da missão Cluster da ESA, organizadas em formação de tetraedro.

Como o vetor que define a face é conhecido por  $\Delta r_i \vee \Delta r_j$ , as correntes normais às três faces podem ser facilmente reprojetaadas em um sistema de coordenadas cartesianas. A quarta face fornece informações redundantes por definição, mas pode ser usada como uma verificação de consistência das outras.

Um dos principais erros relacionados à técnica do *curlometer* é a divergência. A divergência é uma medida que indica o quanto um campo, como o campo magnético, se espalha ou se concentra em um ponto específico. Se houver divergência no campo magnético, isso significa que a técnica do *curlometer* pode não ser capaz de medir com precisão as correntes elétricas no espaço.

A divergência ocorre principalmente devido a erros de medição, como ruído nos sinais ou erros de posicionamento dos satélites. Além disso, a técnica do *curlometer* assume que o campo magnético é uniforme entre os satélites, o que nem sempre é verdade. Portanto, é importante avaliar cuidadosamente a qualidade dos dados medidos e aplicar técnicas de correção para minimizar a influência da divergência nos resultados obtidos.

## 2.7 Volatilidade Magnética

A volatilidade magnética definida por [19], é uma medida que aproxima a densidade de corrente. Para a volatilidade magnética, calculamos os desvios padrão em uma janela móvel de tamanho  $w$ , que percorre as séries de tamanho  $N$  dos log-retornos magnéticos  $\Delta r_{mag}(t) = \log \left( \frac{|\vec{B}(t+\tau)|}{|\vec{B}(t)|} \right)$ , que é simplesmente o retorno magnético da série do logaritmo natural do valor absoluto do campo magnético,  $r_{mag}(t) = \log \left( |\vec{B}(t)| \right)$ . Nesse caso, definimos a volatilidade magnética como

$$vol_{mag}(j) = \sqrt{\frac{1}{w-1} \sum_{i=j}^{w+j-1} (\Delta r_{mag}(i) - \mu(j))^2}, \quad (19)$$

onde  $j = 1, \dots, N - w + 1$  rotula a posição da janela móvel de tamanho  $w$  e  $\mu(j)$  é a média de  $\Delta r_{mag}$  dentro desta janela. Para as análises neste artigo, os valores de  $w$  e  $\tau$  são escolhidos de forma a possuírem, no máximo, uma escala de tamanho comparável a das estruturas coerentes mais energéticas, mas que ainda não sofrem dissipação devido ao processo turbulento. Neste caso, procuramos por uma escala de tamanho situada próxima à região intermediária do intervalo inercial, que pode ser identificada por meio do espectro de potências (PSD) [19]. Diante disto, escolhemos para as análises uma escala localizada no centro do intervalo inercial, ou seja,  $\tau = w = 50s$ , já que  $\tau = 1/f$ .

Na próxima seção abordaremos o pacote que implementa essas e algumas outras funções auxiliares, desenvolvido no decorrer da análise, e organizado em um pacote auxiliar que pode ajudar em análises do gênero.

### 3 Pacote Auxiliar

O pacote `pyExtremeHelper` foi criado para oferecer uma solução mais simplificada e eficiente para a análise de eventos extremos na linguagem de programação Python. Este código tem alta aplicabilidade, uma vez que a Python é uma linguagem de programação conhecida pela sua versatilidade e extensa utilização, seja por sua facilidade de uso, ou pelos inúmeros recursos que oferece para a análise de dados. Em um contexto onde a ciência de dados tem ganhado cada vez mais proeminência, desenvolvemos o `pyExtremeHelper` visando contribuir positivamente nesse campo.

O pacote compila todas as funções necessárias nessa análise dentro de uma estrutura coesa e pensada para simplificar tarefas, funções essas que antes só eram encontradas em outros softwares ou linguagens de programação, possibilitando a utilização em conjunto com outras ferramentas python. Este aspecto torna o `pyExtremeHelper` uma ferramenta poderosa para trabalhos de pesquisa, desenvolvimento e análise, pois permite ao usuário focar em questões mais relevantes do que na construção manual de código.

O código do pacote `pyExtremeHelper` está hospedado em um repositório na plataforma GitHub (<https://github.com/vilelabruno/pyExtremeHelper>), lar de inúmeros projetos de código aberto. Esta ação foi tomada com a intenção de proporcionar aos usuários a liberdade de usar o software, além da possibilidade de copiar, modificar e distribuir o software. O software é distribuído sob a licença MIT, conhecida por conceder uma ampla liberdade de utilização.

Além disso, tornar o pacote `pyExtremeHelper` um projeto de código aberto não apenas facilita sua utilização, mas também permite que uma comunidade maior de contribuidores participe da sua evolução. Isso significa que outros pesquisadores e desenvolvedores podem enriquecer o pacote com suas próprias ideias e melhorias, conferindo a ele uma continuada expansão e aprimoramento.

## 4 Dados

Neste estudo, utilizamos dados dos 4 Clusters fornecidos pela ESA, calculando a densidade de corrente usando a técnica de *curlometer* descrita acima. Dentre essas séries, selecionamos 20 com duração de 24 horas nas quais identificamos a presença de lâminas de corrente. Para esta análise, escolhemos quatro séries específicas: 2 com lâminas de corrente: 13 de fevereiro de 2001 e 2 de fevereiro de 2002 (usando o método de Li), e duas sem: 15 de fevereiro de 2013 e 25 de fevereiro de 2015.

### 4.1 Escolha do Limiar e Processo de Declusterização

A técnica POT, conhecida como *peaks-over-threshold* (POT), é utilizada para analisar os pontos de uma amostra que excedem um determinado valor de referência, como explicado anteriormente neste



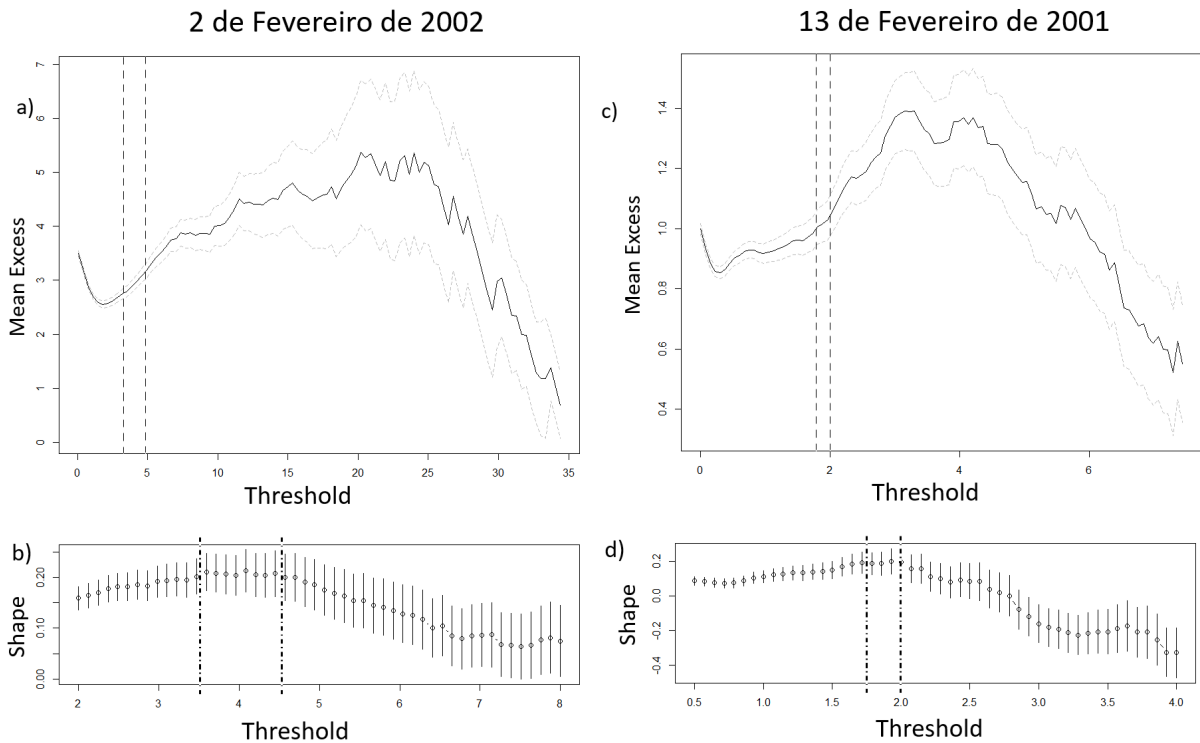


Figura 3: Figuras do *mean excess* (a) e (c) e parâmetros de forma (*shape parameter*) em função dos limiares (b) e (d), para o dia 2 de fevereiro de 2002, para o dia 13 de fevereiro de 2001. Áreas delimitadas nos intervalos para os quais o *mean excess* é aproximadamente linear, dentro de um intervalo de confiança de 95% (linhas cinzas tracejadas) e o parâmetro de forma é aproximadamente constante, dentro das barras de erro.

estudo. Nesse caso, é possível modelar a distribuição de probabilidade GPD (Generalized Pareto Distribution) [1], [2] desses pontos excedentes. No entanto, para uma modelagem precisa, é crucial escolher um valor de referência adequado, que não seja muito alto a ponto de gerar poucos pontos excedentes e aumentar a variância, nem muito baixo a ponto de introduzir um viés na modelagem [5].

Para identificar o valor de referência adequado, são utilizados dois métodos auxiliares. O primeiro método explora a linearidade do *mean excess*, selecionando como valor de referência adequado o menor valor dentro de um intervalo específico, a partir do qual o *mean excess* tem um comportamento aproximadamente linear. O segundo método baseia-se na estabilidade do parâmetro de forma, escolhendo como valor de referência o menor valor dentro de um intervalo no qual o parâmetro de forma é aproximadamente constante. No entanto, é importante ressaltar que a linearidade do *mean excess* e a estabilidade do parâmetro de forma são condições necessárias, mas não suficientes, para a escolha do intervalo de referência. Nem todos os intervalos que satisfazem ambas as condições são adequados para selecionar o valor de referência.

De acordo com a Figura 3, o limiar adequado para as séries de densidade de corrente dos 2 de fevereiro de 2002 é  $u = 3.5$ . Já para o dia 13 de fevereiro de 2001 o limiar adequado foi  $u = 1.7$ . Olhando para os dois dias sem lâminas de corrente, conseguimos ver que o limiar adequado para as séries de densidade de corrente dos 2 de fevereiro de 2002 é  $u = 3.5$ . Já para o dia 13 de fevereiro de 2001 o limiar adequado foi  $u = 1.7$ , conforme pode ser visto em Figura 4.

Um processo de desclusterização se mostra necessário, para obtenção de dados IID, para as séries com lâminas, pois sua autocorrelação é alta. Em contrapartida, para os dias sem lâmina, não foi

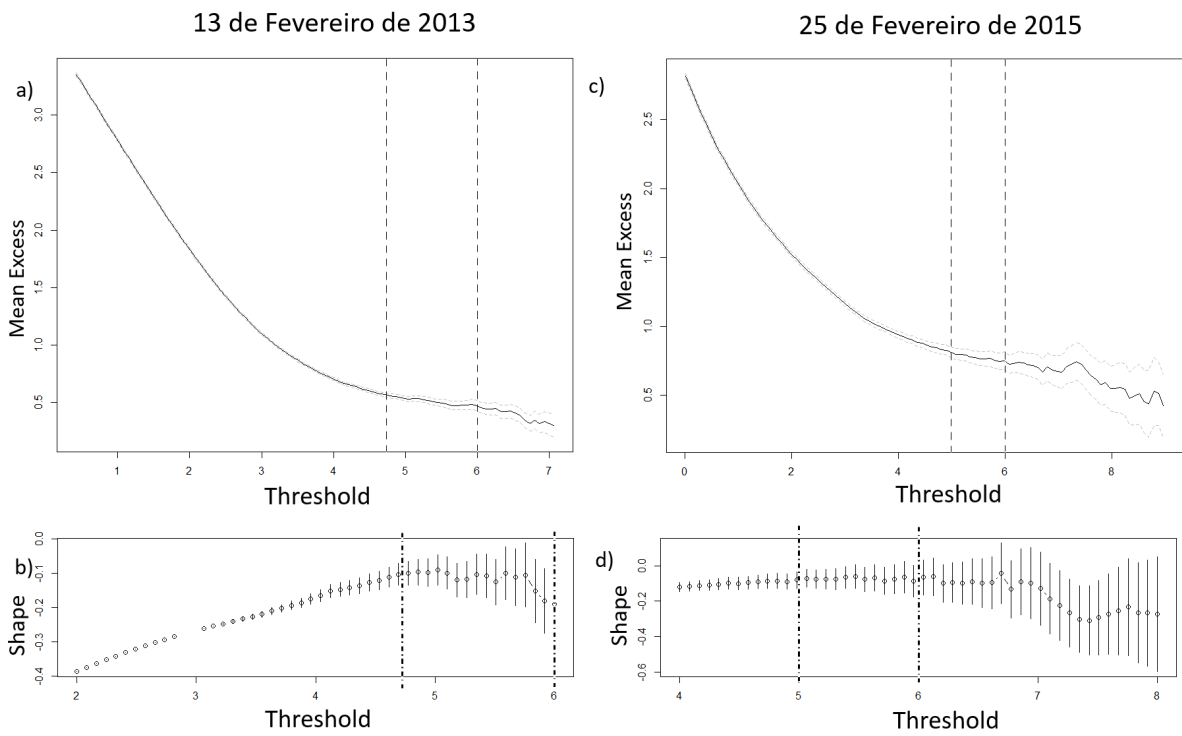


Figura 4: Figuras do *mean excess* (a) e (c) e parâmetros de forma (*shape parameter*) em função dos limiares (b) e (d), para o dia 2015, para o dia 2013. Áreas delimitadas nos intervalos para os quais o *mean excess* é aproximadamente linear, dentro de um intervalo de confiança de 95% (linhas cinzas tracejadas) e o parâmetro de forma é aproximadamente constante, dentro das barras de erro.

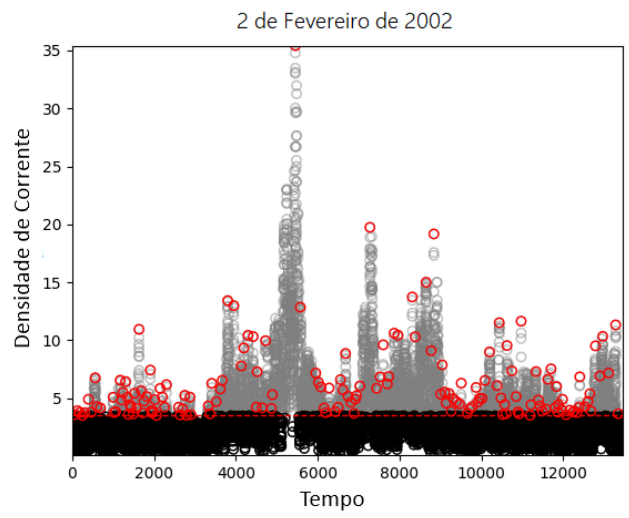
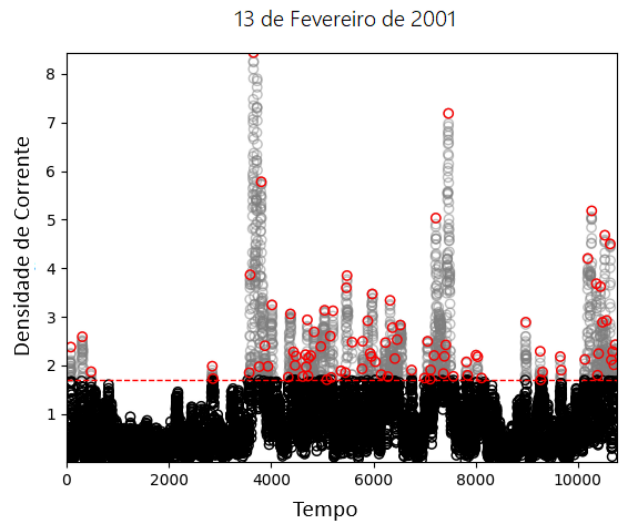


Figura 5: Figuras do processo de declusterização, para a densidade de corrente medida com a técnica de *curlometer* (*curl*) para os dias 13 de Fevereiro de 2001 e para o dia 2 de Fevereiro de 2002. Área tracejada indicada o limiar escolhido, pontos vermelhos são os pontos resultantes do processo de declusterização, pontos cinza são os pontos acima do limiar porém não escolhidos pelo processo de declusterização, e pretos, abaixo do limiar.

necessário a declusterização, já que os dados já são IID. A Figura 5 mostra o processo de declusterização das séries de densidade de corrente, obtidas com a técnica de *curlometer* dos dias 2 de fevereiro de 2002 e 13 de fevereiro de 2001. As novas amostras obtidas após o processo de declusterização são compostas pelas observações destacadas acima dos limiares (linhas tracejadas), que possivelmente são independentes.

A fim de confirmar a eficácia do método de *runs declustering*, foram feitas análises de função de autocorrelação ou *autocorrelation function* (ACF) dos pontos excedentes em todas as séries. A Figura 6, mostra uma forte dependência entre os pontos da amostra para os dias com lâminas de correntes (painéis a esquerda) e antes da aplicação do processo de declusterização, visto que a ACF neste caso é significativamente diferente de zero, dado um nível de significância de 5% (linhas horizontais azuis). Em contrapartida, os correlogramas das séries após a aplicação do método de declusterização (painéis à direita) mostram autocorrelações significativamente iguais a zero, dado um nível de significância de 5%, o que sugere uma independência entre os pontos das amostras.

Já na figura 7, nossa série tem uma autocorrelação muito baixa, não sendo necessário passar pelo processo de declusterização pois os dados já estão IID (independentes e igualmente distribuídos).

## 5 Resultados

Na seção anterior, verificamos que as séries com lâminas de corrente, possuem forte dependência entre seus pontos, quebrando assim um dos pressupostos básicos para aplicação da teoria de eventos extremos [5]. Neste caso, foi necessária a aplicação da técnica de declusterização, a fim de que as séries se tornem aptas a serem analisadas. Além disto, aplicamos técnicas estatísticas para nos auxiliar na escolha adequada do limiar, o que completa as exigências necessárias para uma aplicação confiável da técnica POT.

Nesta seção, iremos apresentar os resultados de fitting após o pré-processamento dos dados, bem como analisaremos como a volatilidade consegue ser uma *proxy* melhor que o PVI para dias com lâminas de corrente.

### 5.1 Fitting GPD

As Figuras 8 e 9 mostram os *diagnostic plots* dos modelos GPD's fitados para as séries temporais declusterizadas da densidade de corrente obtida com a técnica de *curlometer* para os dias 13 de Fevereiro de 2001 e 2 de Fevereiro de 2002. Os QQ-plots dos quantis dos dados empíricos com relação aos quantis teóricos são mostrados nas Figuras 8-(a) e 9-(a). Os eixos-y mostram os quantis empíricos obtidos por meio dos dados observacionais e os eixos-x mostram os quantis teóricos obtidos diretamente da relação da distribuição GPD. Estes QQ-plots revelam um excelente ajuste das séries temporais por uma GPD, conforme também pode ser verificado pela comparação dos histogramas empíricos e modelados, mostrados nas figuras 8-(c) e 9-(c).

Os QQ-plots do quantis dos dados gerados randomicamente pela distribuição GPD em relação aos quantis dos dados empíricos, dado um intervalo de confiança de 95% (linhas tracejadas em cinza), são mostrados nas figuras 8-(b) e 9-(b). Neste caso, os eixos-x representam os quantis empíricos obtidos por meio dos dados observacionais e o eixos-y representam os quantis de uma distribuição GPD selecionada randomicamente. A qualidade dos ajustes estatísticos também pode ser verificada por estas figuras.

Os níveis de retorno com intervalos de confiança ponto a ponto *pointwise* de 95% (linhas tracejadas em cinza), obtidos por uma aproximação normal, são mostrados nas figuras 8-(d) e 9-(d). Nos eixos-x encontram-se, em escala logarítmica, os períodos de retorno, enquanto que os níveis de retorno associados encontram-se nos eixos-y. As curvas nas Figuras 8-(d) e 9-(d) se mostra convexa, e portanto, sem um limite assintótico, sugerindo que a distribuição da série com a presença de lâminas de corrente possui suporte infinito (*unbounded distribution*) [5]. Neste caso, o parâmetro de forma é positivo ( $\xi > 0$ ), conforme será confirmado mais adiante. Em contrapartida, as séries temporais (10-(d) e 11-(d)) para um intervalo de tempo no qual não foram detectadas tais estruturas coerentes (lâminas de

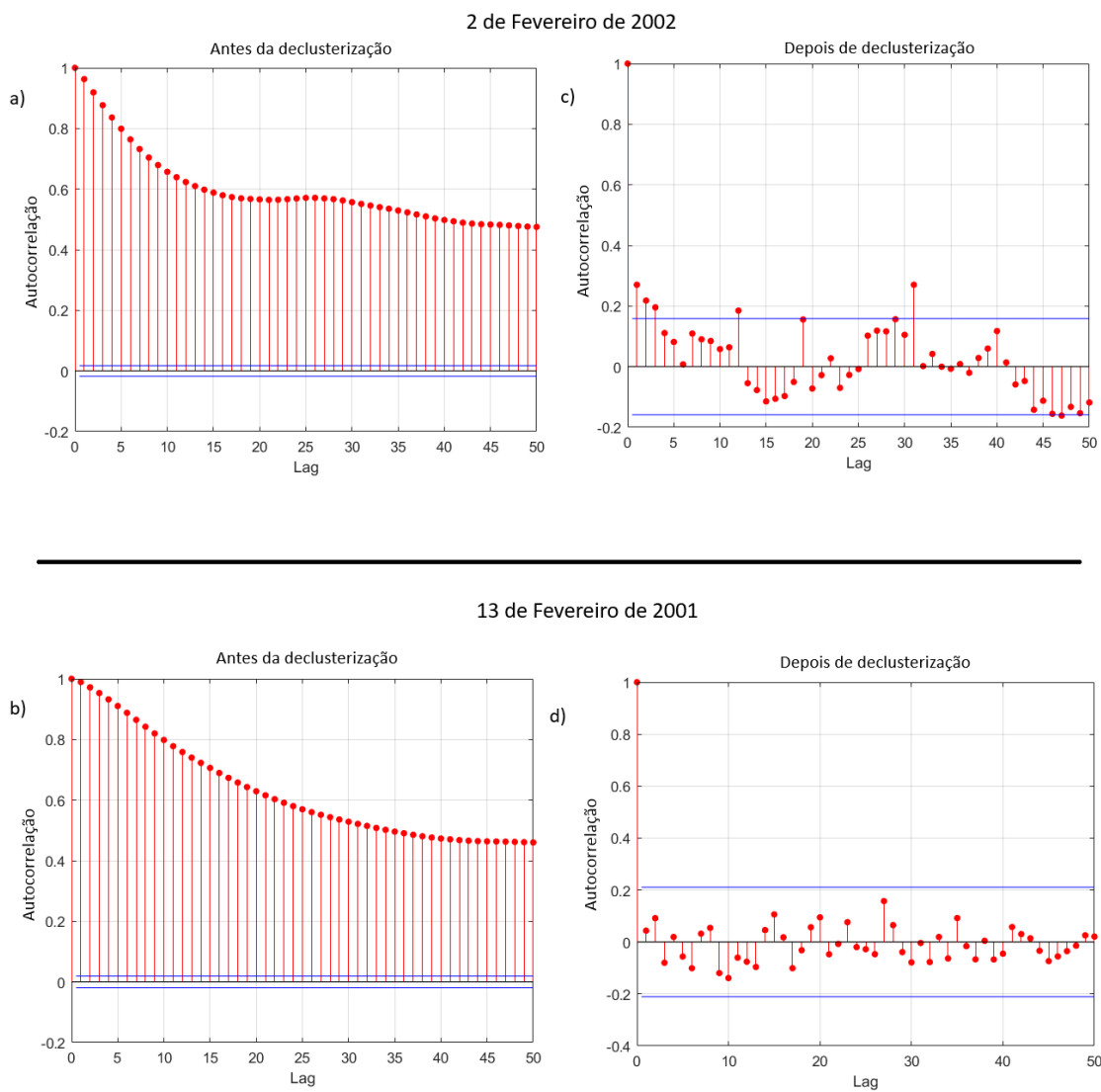


Figura 6: Figuras da autocorrelação das séries de densidade de corrente antes (painéis à esquerda), e após (painéis à direita) o processo de declusterização, para os dias 13 de Fevereiro de 2001 e 2 de Fevereiro de 2002. Área azul indica o intervalo de confiança.

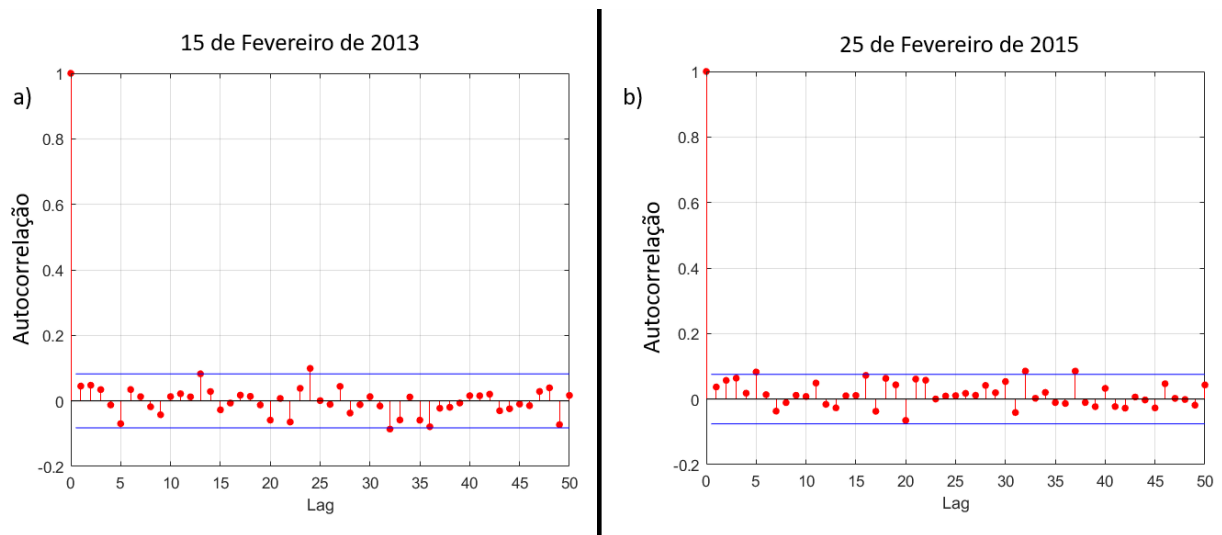


Figura 7: Figuras da autocorrelação das séries de densidade de corrente para os dias 15 de Fevereiro de 2013 e para o dia 25 de Fevereiro de 2015. Área azul indica o intervalo de confiança.

Tabela 1: Profile Likelihood Estimates and Confidence Intervals  
Profile Likelihood

Time Series	$\xi$	
	Value	95% CI
13 Fevereiro de 2001	0.234	(0.0244, 0.5496)
02 Fevereiro de 2002	0.287	(0.2128, 0.5227)
15 de Fevereiro de 2013	-0.078	(-0.1232, -0.0269)
25 de Fevereiro de 2015	-0.12	(-0.1343, -0.0923)

corrente), se mostram côncavas, e portanto, apresentam distribuição com um limite superior assintótico, sugerindo assim, que as distribuições das séries sem a presença de lâminas de corrente possui um suporte finito (*bounded distribution*) [5]. Neste caso, o parâmetro de forma é negativo ( $\xi < 0$ ), conforme também será confirmado mais adiante. Além disto, para todos os casos dos níveis de retorno estudados, os pontos representativos das amostras não extrapolam seus respectivos intervalos de confiança, dando mais evidências de que os ajustes pela GPD das séries analisadas foram satisfatórios.

A tabela 1 apresenta também todos os resultados dos parâmetros estimados no ajuste GPD das séries temporais da densidade de corrente das figuras 13 e 12, bem como todos seus intervalos de confiança, utilizando *profile likelihood*. Os valores de todos os parâmetros são estimados por *profile likelihood*.

Pode-se notar que todas as séries obtidas com *curlometer* que possuem lâminas de corrente (figura 13) apresentam um parâmetro de forma positivo  $\xi > 0$ , enquanto as séries sem lâminas possuem parâmetro de forma negativo.

## 5.2 Correlação volatilidade vs *curlometer*

Nesta subseção, apresentamos os resultados da nossa análise comparativa entre a correlação da volatilidade e a densidade de corrente versus a correlação da densidade de corrente com a PVI. Para avaliar essas correlações, utilizamos o coeficiente de correlação de Spearman, uma medida estatística robusta que não requer que os dados sejam linearmente relacionados, tornando-a adequada para a

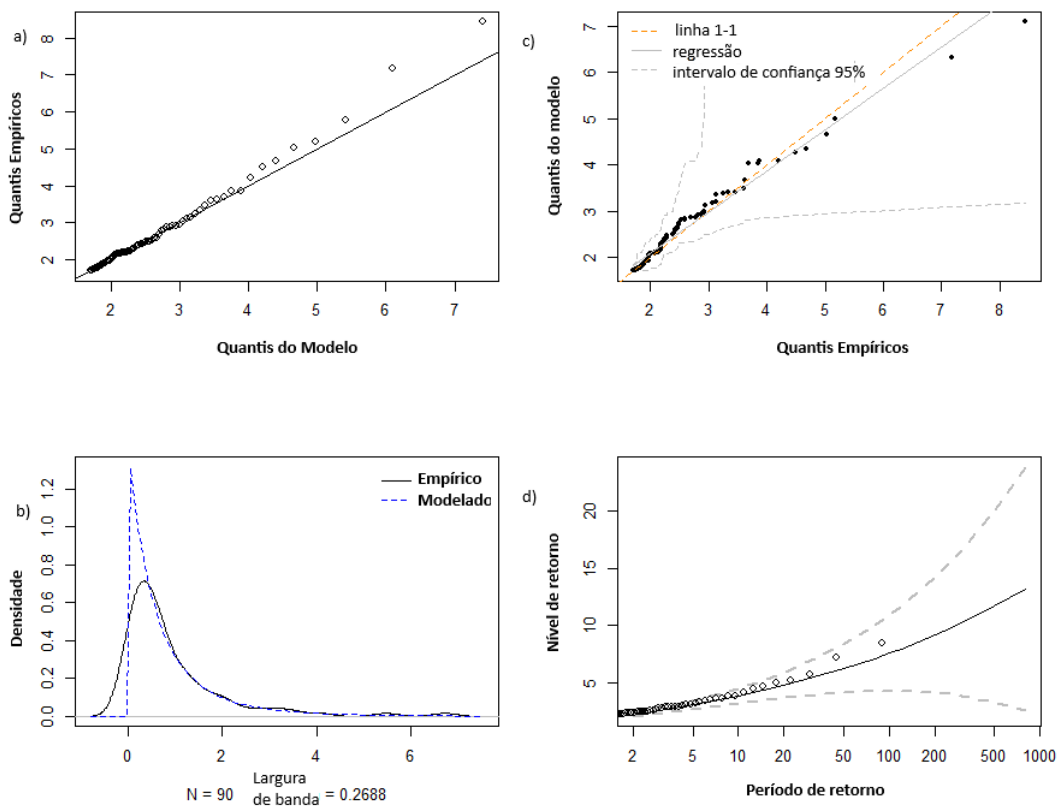


Figura 8: Gráficos de diagnóstico para o ajuste do modelo GPD para a série temporal de densidade de corrente de 13 de Fevereiro de 2001. (a) Gráfico QQ dos quantis de dados empíricos em relação aos quantis de dados ajustados obtidos a partir do modelo GPD; (c) Gráfico QQ de quantis de dados aleatórios, obtidos a partir do modelo GPD em relação aos quantis de dados empíricos. As linhas pontilhadas representam o intervalo de confiança de 95 por cento; a linha tracejada representa a linha 1-1; a linha sólida representa a regressão linear; (b) funções de densidade de probabilidade dos dados empíricos (linha sólida) e do modelo GPD ajustado (linha tracejada); (d) níveis de retorno (linhas sólidas) como função dos períodos de retorno (em escala logarítmica) com intervalo de confiança pontual de 95 por cento (linhas tracejadas)

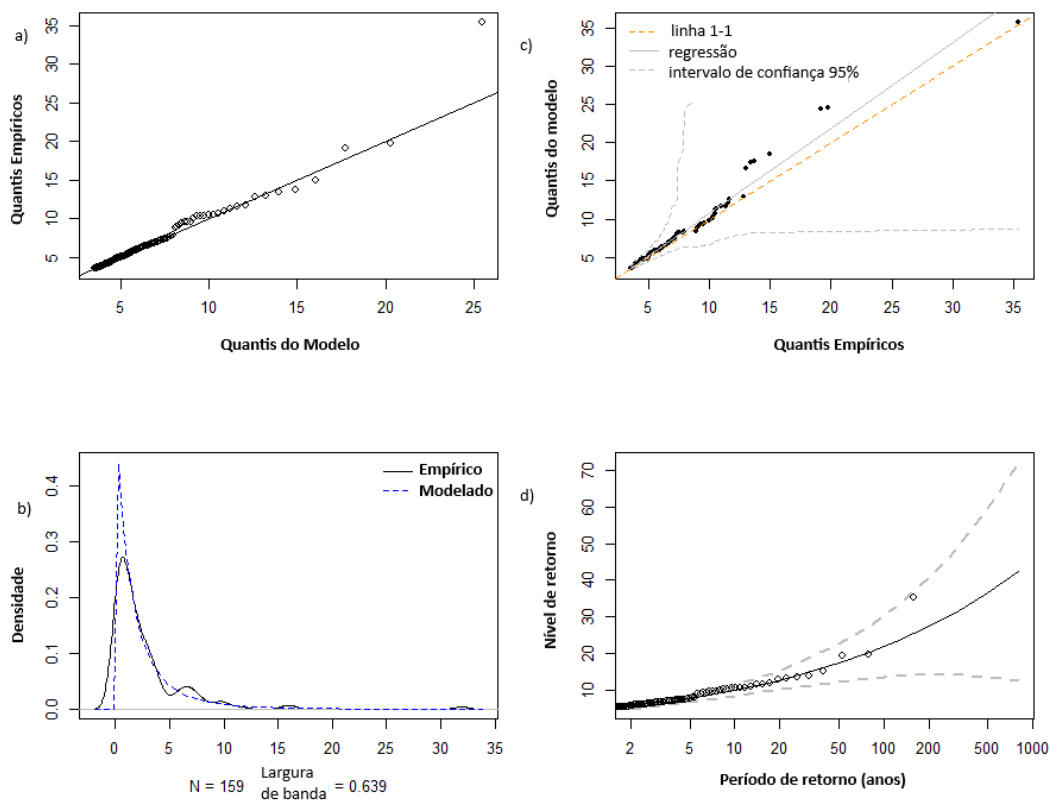


Figura 9: Gráficos de diagnóstico para o ajuste do modelo GPD para a série temporal de densidade de corrente de 2 de Fevereiro de 2002. (a) Gráfico QQ dos quantis de dados empíricos em relação aos quantis de dados ajustados obtidos a partir do modelo GPD; (c) Gráfico QQ de quantis de dados aleatórios, obtidos a partir do modelo GPD em relação aos quantis de dados empíricos. As linhas pontilhadas representam o intervalo de confiança de 95 por cento; a linha tracejada representa a linha 1-1; a linha sólida representa a regressão linear; (b) funções de densidade de probabilidade dos dados empíricos (linha sólida) e do modelo GPD ajustado (linha tracejada); (d) níveis de retorno (linhas sólidas) como função dos períodos de retorno (em escala logarítmica) com intervalo de confiança pontual de 95 por cento (linhas tracejadas)



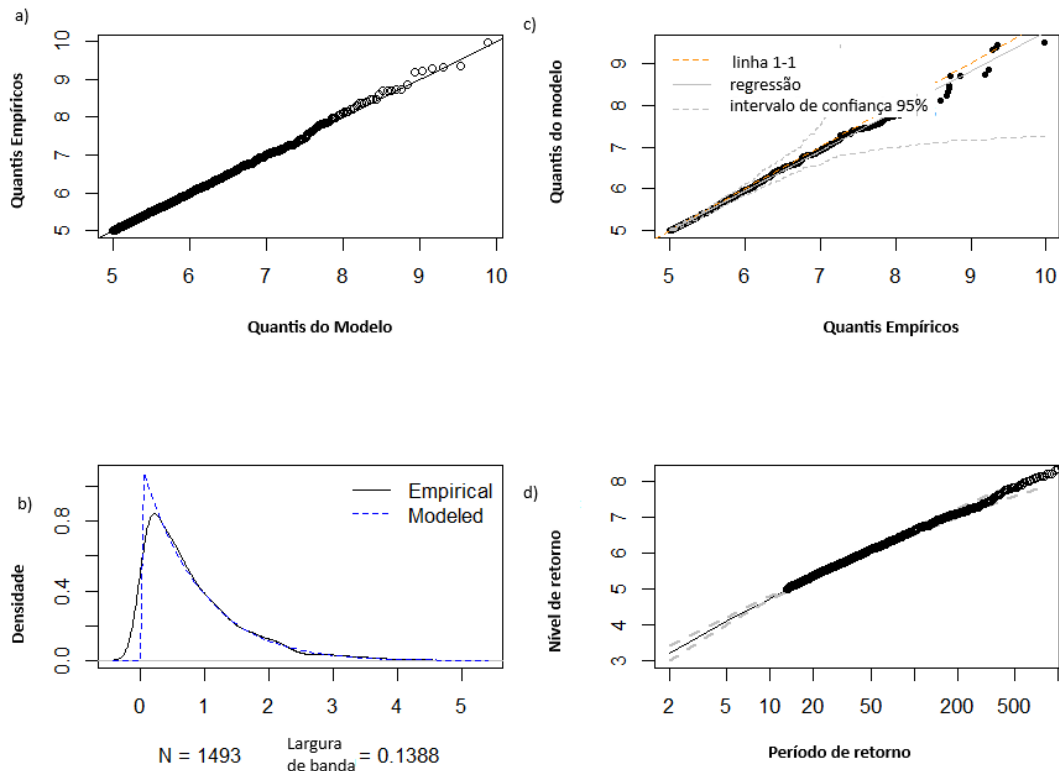


Figura 10: Gráficos de diagnóstico para o ajuste do modelo GPD para a série temporal de densidade de corrente de 15 de fevereiro de 2013. (a) Gráfico QQ dos quantis de dados empíricos em relação aos quantis de dados ajustados obtidos a partir do modelo GPD; (c) Gráfico QQ de quantis de dados aleatórios, obtidos a partir do modelo GPD em relação aos quantis de dados empíricos. As linhas pontilhadas representam o intervalo de confiança de 95 por cento; a linha tracejada representa a linha 1-1; a linha sólida representa a regressão linear; (b) funções de densidade de probabilidade dos dados empíricos (linha sólida) e do modelo GPD ajustado (linha tracejada); (d) níveis de retorno (linhas sólidas) como função dos períodos de retorno (em escala logarítmica) com intervalo de confiança pontual de 95 por cento (linhas tracejadas)

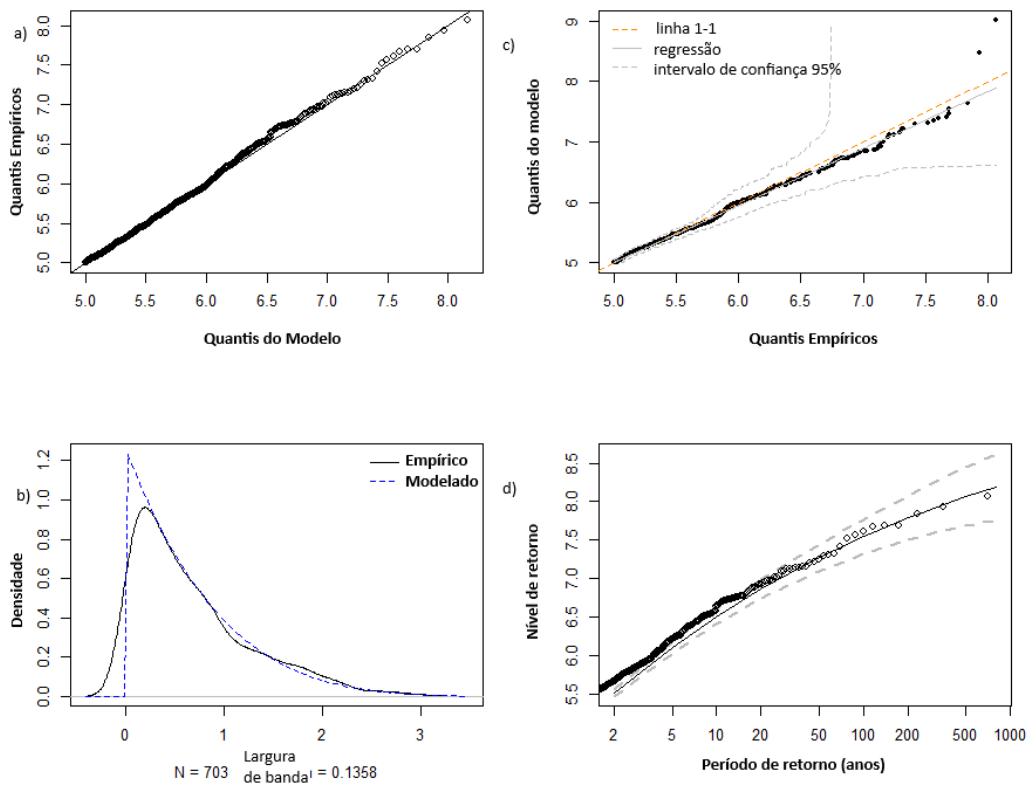


Figura 11: Gráficos de diagnóstico para o ajuste do modelo GPD para a série temporal de densidade de corrente de 25 de fevereiro de 2015. (a) Gráfico QQ dos quantis de dados empíricos em relação aos quantis de dados ajustados obtidos a partir do modelo GPD; (c) Gráfico QQ de quantis de dados aleatórios, obtidos a partir do modelo GPD em relação aos quantis de dados empíricos. As linhas pontilhadas representam o intervalo de confiança de 95 por cento; a linha tracejada representa a linha 1-1; a linha sólida representa a regressão linear; (b) funções de densidade de probabilidade dos dados empíricos (linha sólida) e do modelo GPD ajustado (linha tracejada); (d) níveis de retorno (linhas sólidas) como função dos períodos de retorno (em escala logarítmica) com intervalo de confiança pontual de 95 por cento (linhas tracejadas)

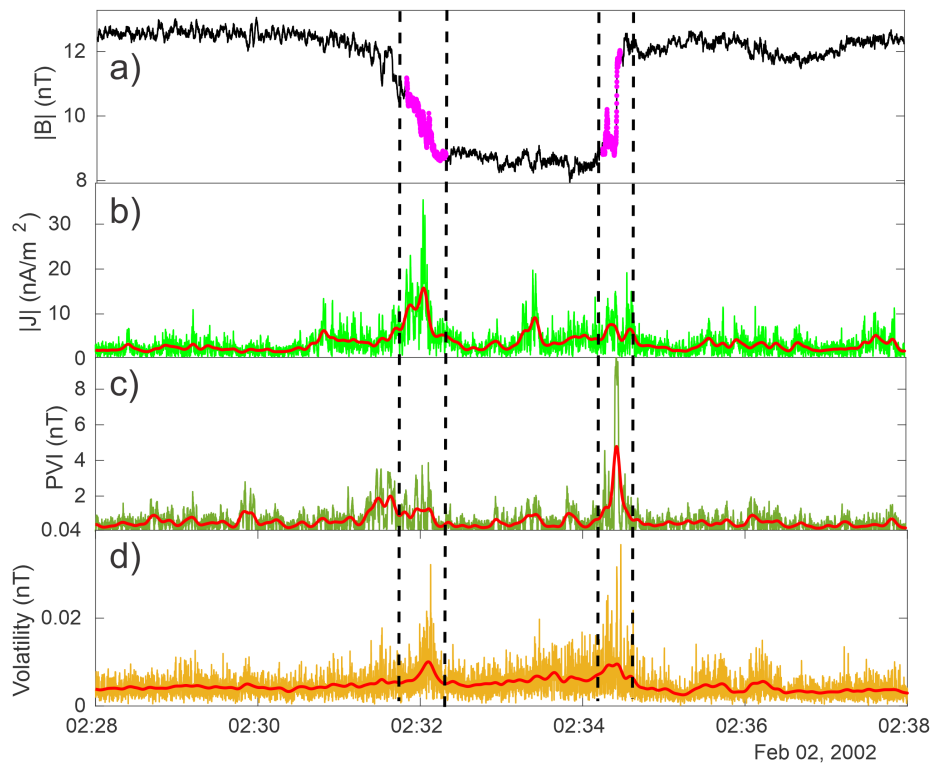


Figura 12: Série temporal com lâminas obtidas com método de Li e suavização Gaussiana para o dia 2 de Fevereiro de 2002, para volatilidade, *curlometer* e PVI

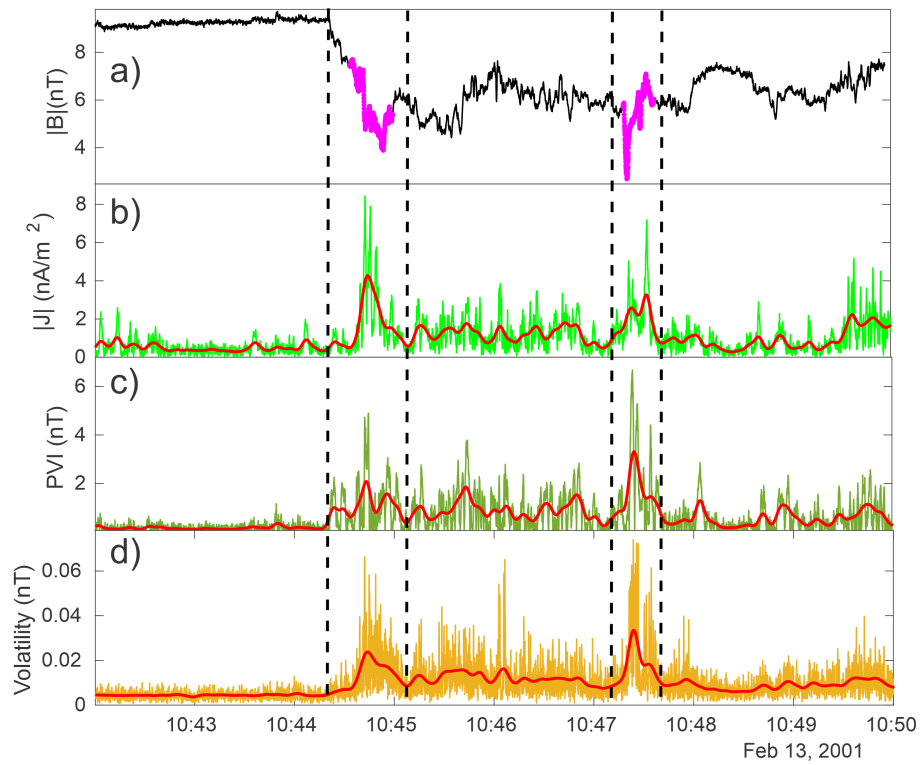


Figura 13: Série temporal com lâminas obtidas com método de Li e suavização Gaussiana para o dia 13 de Fevereiro de 2001, para volatilidade, *curlometer* e PVI

nossa análise.

Para remover o ruído e trabalhar especificamente com as oscilações de baixa frequência provenientes das lâminas de corrente, aplicamos um kernel gaussiano aos nossos dados. Podemos ver os resultados nas figuras 12 e 13. Essa técnica nos permitiu suavizar as flutuações de curto prazo e focar nas variações de longo prazo, que são de interesse para a nossa análise da volatilidade.

Os resultados da nossa análise revelaram uma correlação medida com spearman significativa de 0.64 para o dia 2 de fevereiro de 2002 e de 0.822 para o dia 13 de fevereiro de 2001 entre a volatilidade e a densidade de corrente. Conforme a volatilidade aumentava, observamos um aumento correspondente na densidade de corrente. Esse resultado indica que eventos extremos de volatilidade estão associados a um maior número de eventos de alta densidade de corrente.

Por outro lado, ao explorarmos a correlação da densidade de corrente com PVI, encontramos uma correlação mais fraca de 0.43 para o dia 2 de fevereiro de 2002 e de 0.821 para o dia 13 de fevereiro de 2001, em comparação com a correlação entre a volatilidade e a densidade de corrente. Isso sugere que a volatilidade seja uma aproximação mais apropriada.

Na figura 14 podemos ver que ao redor de *lag* 0 (correlação mais significativa), a volatilidade acaba obtendo uma correlação maior com a densidade de corrente, corroborando para a hipótese de uma melhor aproximação da densidade de corrente.

Em suma, nossos resultados destacam a utilidade da densidade de corrente como uma abordagem promissora para medir eventos extremos de volatilidade. Ao utilizar a correlação de Spearman e remover o ruído com um kernel gaussiano, pudemos identificar uma forte relação entre a volatilidade e a densidade de corrente. Essa descoberta pode ter implicações importantes para a compreensão e a previsão de eventos extremos, contribuindo para uma melhor compreensão da dinâmica da volatilidade e dos fatores que a influenciam.

## 6 Conclusão

Neste estudo, investigamos a ocorrência de eventos extremos em dados do vento solar, recorrendo à técnica de POT e ajustando uma distribuição GPD. O objetivo geral do trabalho era entender melhor as características dos eventos extremos nos ventos solares, especialmente a influência da presença de lâminas de corrente, bem como validar a escolha da volatilidade como uma *proxy* mais precisa do que a PVI para dias com lâminas de corrente. A escolha apropriada do limiar é crítica para a aplicação bem-sucedida do modelo e, neste caso, foi encontrada através de técnicas que buscam por linearidade do *mean excess* e estabilidade do parâmetro de forma. Também foi necessário aplicar um processo de *declusterização* nos dados apresentando alta autocorrelação para garantir sua distribuição independente, pressuposto exigido para estudos de eventos extremos.

Os dados foram analisados para ocorrências de eventos extremos com e sem a presença de lâminas de corrente. Na presença de lâminas de corrente, os dados apresentaram um parâmetro de forma positivo. Isto indica um comportamento volátil ou, em termos estatísticos, pertencente à classe das distribuições de cauda pesada que possui, no extremo superior de sua distribuição, uma queda de probabilidade mais lenta que uma exponencial. Isso implica maior probabilidade de observar valores extremamente altos. O suporte infinito, nesse caso, sugere que não há limite superior para os eventos extremos que podem ocorrer.

Por outro lado, na ausência de lâminas de corrente, os dados apresentaram um parâmetro de forma negativo. Isto aponta para uma distribuição com suporte limitado, ou seja, que existe um valor máximo que o fenômeno em estudo pode atingir. Neste caso, os eventos extremos são menos severos e existe um limite superior para a densidade de corrente.

Em suma, a presença dessas lâminas de corrente impacta a forma como os eventos extremos ocorrem em nossos dados: na presença de lâminas, os eventos podem ser mais severos e sem um limite máximo detectável, enquanto na ausência delas, os eventos são menos severos e há um limite superior.

Considerando os resultados obtidos no estudo atual e a importância do processo de *declusterização* para garantir a independência dos dados, futuras melhorias nesta área poderiam ser extremamente

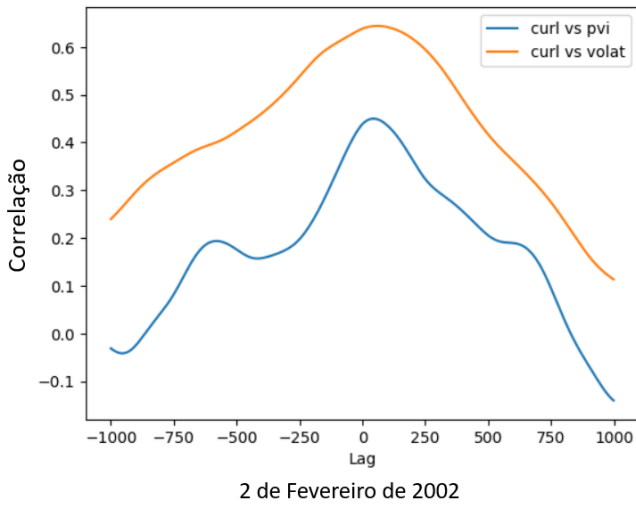
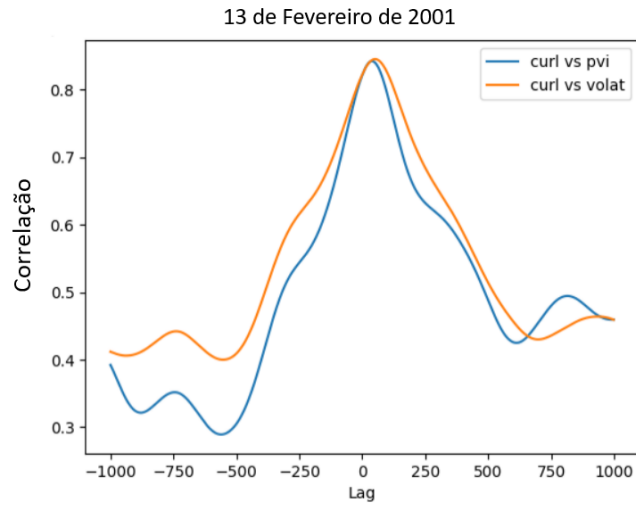


Figura 14: Correlação em diferentes *lag*'s para os dias com lâminas de corrente. Podemos ver uma maior correlação entre a volatilidade e a densidade de corrente, quando comparada com a correlação entre PVI e densidade de corrente.

proveitosas. Embora o método atual de *declusterização* tenha sido eficiente em lidar com a alta autocorrelação nos dados, alternativas mais sofisticadas poderiam ser exploradas. Por exemplo, poderíamos investigar técnicas que levem em conta não apenas as dependências de cada observação com sua mais recente, mas também interações de longo alcance. Estas melhorias poderiam permitir uma modelagem mais eficaz de eventos extremos e aprimorar nossa capacidade de prever e entender tais fenômenos.

A análise da correlação entre a volatilidade e a densidade de corrente mostrou que eventos extremos de volatilidade estão associados com um maior número de eventos de alta densidade de corrente, um resultado que foi observado com forte significância. A correlação da densidade de corrente com a PVI foi encontrada para ser menos acentuada, sugerindo que a volatilidade pode ser uma aproximação mais apropriada para eventos extremos.

Todas as funções utilizadas na análise foram compiladas em um pacote Python abrangente, abarcando desde as funções estatísticas até as relacionadas à astronomia e à física. Essa compilação oferece uma solução completa e acessível para os pesquisadores, possibilitando a realização de análises complexas de forma eficiente e precisa. O pacote, desenvolvido como um projeto de código aberto, possui uma importância significativa para a comunidade, pois permite que pesquisadores, professores e estudantes em todo o mundo tenham acesso às funcionalidades dessas áreas do conhecimento. Além disso, ao ser desenvolvido como código aberto, o pacote fomenta a colaboração, incentivando outros a contribuir para o seu aprimoramento e expandindo ainda mais suas capacidades. Essa abordagem open source também garante a transparência e a confiabilidade do pacote, uma vez que permite a revisão por pares e a identificação de possíveis melhorias. Com o poder e a acessibilidade que o pacote oferece, ele desempenha um papel fundamental no avanço do conhecimento e no compartilhamento de descobertas científicas em uma ampla gama de áreas, impulsionando a colaboração e o progresso científico para benefício de toda a comunidade.

Como conclusões finais, o estudo trouxe evidências que o uso da abordagem de POT fornece uma representação robusta de dados envolvendo eventos extremos, na qual a volatilidade se mostrou mais aproximadamente significativa em relação a eventos de lâminas de corrente. Além disso, a capacidade de identificar e modelar esta relação pode ter implicações valiosas para o entendimento e previsibilidade de lâminas de corrente e reconexões magnéticas.

Em termos de pesquisa futura, seria benéfico expandir essas descobertas, explorando técnicas de declusterização mais avançadas e construindo em cima do entendimento que adquirimos sobre a ligação entre volatilidade e densidade de corrente, além de evolução do pacote aqui desenvolvido para outros indicadores da missão Cluster, implementação de outros algoritmos ou otimização em termos de custo computacional dos algoritmos.

## Referências

- [1] A. A. Balkema e L. Haan, “Residual Life Time at Great Age”, *The Annals of Probability*, vol. 2, n.º 5, pp. 792–804, 1974.
- [2] P. James, “Statistical inference using extreme order statistics”, *The Annals of Statistics*, vol. 3, n.º 1, pp. 119–131, 1975.
- [3] Y. Ogata, “Statistical models for earthquake occurrences and residual analysis for point processes”, *Journal of the American Statistical Association*, vol. 83, n.º 401, pp. 9–27, 1988. DOI: 10.1080/01621459.1988.10478560. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478560>.
- [4] W. D. Gonzalez, J. A. Joselyn, Y. Kamide et al., “What is a geomagnetic storm?”, *Journal of Geophysical Research: Space Physics*, vol. 99, n.º A4, pp. 5771–5792, abr. de 1994. DOI: 10.1029/93ja02867. URL: <https://doi.org/10.1029/93ja02867>.
- [5] S. Coles, *An introduction to statistical modeling of extreme values* (Springer Series in Statistics). London: Springer-Verlag, 2001, ISBN: 1-85233-459-2.

- [6] K. Engeland, H. Hisdal e A. Frigessi, “Practical Extreme Value Modelling of Hydrological Floods and Droughts: A Case Study”, *Extremes*, vol. 7, n.º 1, pp. 5–30, mar. de 2004, ISSN: 1572-915X. DOI: 10.1007/s10687-004-4727-5. URL: <https://doi.org/10.1007/s10687-004-4727-5>.
- [7] L. Fawcett e D. Walshaw, “Markov chain models for extreme wind speeds”, *Environmetrics*, vol. 17, n.º 8, pp. 795–809, 2006. DOI: 10.1002/env.794. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.794>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.794>.
- [8] *THE SUN AND SPACE WEATHER*. Springer Netherlands, 2007. DOI: 10.1007/978-1-4020-5604-8. URL: <https://doi.org/10.1007/978-1-4020-5604-8>.
- [9] G. Li, “Identifying Current-Sheet-like Structures in the Solar Wind”, *The Astrophysical Journal Letters*, vol. 672, n.º 1, p. L65, 2008. URL: <http://stacks.iop.org/1538-4357/672/i=1/a=L65>.
- [10] A. C.-L. Chian e R. A. Miranda, “Cluster and ACE observations of phase synchronization in intermittent magnetic field turbulence: a comparative study of shocked and unshocked solar wind”, *Annales Geophysicae*, vol. 27, n.º 4, pp. 1789–1801, 2009. DOI: 10.5194/angeo-27-1789-2009. URL: <https://www.ann-geophys.net/27/1789/2009/>.
- [11] A. Greco, W. H. Matthaeus, S. Servidio, P. Chuychai e P. Dmitruk, “Statistical Analysis of Discontinuities in Solar Wind ACE Data and Comparison with Intermittent MHD Turbulence”, *The Astrophysical Journal Letters*, vol. 691, n.º 2, p. L111, 2009. URL: <http://stacks.iop.org/1538-4357/691/i=2/a=L111>.
- [12] J. T. Gosling, “Magnetic Reconnection in the Solar Wind”, *Space Science Reviews*, vol. 172, n.º 1, pp. 187–200, nov. de 2012, ISSN: 1572-9672. DOI: 10.1007/s11214-011-9747-2. URL: <https://doi.org/10.1007/s11214-011-9747-2>.
- [13] R. Bruno e V. Carbone, “The Solar Wind as a Turbulence Laboratory”, *Living Reviews in Solar Physics*, vol. 10, n.º 1, p. 2, mai. de 2013, ISSN: 1614-4961. DOI: 10.12942/lrsp-2013-2. URL: <https://doi.org/10.12942/lrsp-2013-2>.
- [14] A. H. Boozer, “Formation of current sheets in magnetic reconnection”, *Physics of Plasmas*, vol. 21, n.º 7, p. 072907, 2014. DOI: 10.1063/1.4890491. eprint: <https://doi.org/10.1063/1.4890491>. URL: <https://doi.org/10.1063/1.4890491>.
- [15] K. T. Osman, W. H. Matthaeus, J. T. Gosling et al., “Magnetic Reconnection and Intermittent Turbulence in the Solar Wind”, *Phys. Rev. Lett.*, vol. 112, p. 215002, 21 mai. de 2014. DOI: 10.1103/PhysRevLett.112.215002. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.112.215002>.
- [16] D. H. Hathaway, “The Solar Cycle”, *Living Reviews in Solar Physics*, vol. 12, n.º 1, p. 4, set. de 2015, ISSN: 1614-4961. DOI: 10.1007/lrsp-2015-4. URL: <https://doi.org/10.1007/lrsp-2015-4>.
- [17] A. C.-L. Chian, H. Q. Feng, Q. Hu et al., “Genesis of interplanetary intermittent turbulence: a case study of rope-rope magnetic reconnection”, *The Astrophysical Journal*, vol. 832, n.º 2, p. 179, 2016.
- [18] M. Thomas, M. Lemaitre, M. L. Wilson et al., “Applications of extreme value theory in public health”, *PLOS ONE*, vol. 11, n.º 7, pp. 1–7, jul. de 2016. DOI: 10.1371/journal.pone.0159312. URL: <https://doi.org/10.1371/journal.pone.0159312>.
- [19] T. F. P. Gomes, E. L. Rempel, F. M. Ramos, S. S. A. Silva e P. R. Muñoz, “Extreme value theory in the solar wind: the role of current sheets”, *Monthly Notices of the Royal Astronomical Society*, vol. 490, n.º 2, pp. 1879–1893, out. de 2019. DOI: 10.1093/mnras/stz2668. URL: <https://doi.org/10.1093/mnras/stz2668>.



- [20] M. W. Dunlop, X.-C. Dong, T.-Y. Wang et al., “Curlometer Technique and Applications”, *Journal of Geophysical Research: Space Physics*, vol. 126, n.<sup>o</sup> 11, nov. de 2021. DOI: 10.1029/2021ja029538. URL: <https://doi.org/10.1029/2021ja029538>.