

Regressão Espacial Quantílica para Previsão da Velocidade do Vento

Gabriel H. O. Assunção¹ and Marcos O. Prates¹

¹Departamento de Estatística, Universidade Federal de Minas Gerais

5 de março de 2018

Prefácio

O primeiro passo antes de introduzir as ideias envolvidas na elaboração desta Dissertação, é apresentar a teoria que fundamenta a Regressão Quantílica. Sendo que o primeiro estudo que envolvia esta técnica foi publicado por Hogg (1975), porém o artigo que apresentou a Regressão Quantílica formalmente e que inclusive é a principal citação em estudos que envolvem esta técnica foi apresentado por Koenker and Basset Jr. (1978). Além disso, é importante citar o estudo realizado por Yu and Moyeed (2001) que apresentou a Regressão Quantílica Bayesiana, além de introduzir a ideia de utilização da distribuição de Laplace Assimétrica para os erros do modelo, e também o estudo de Hall et al. (1999) que é sempre lembrado nestes estudos como sendo o pioneiro na utilização da Regressão Quantílica no contexto de séries de tempo.

Entrando então nas técnicas que formalizam a Regressão Quantílica, Koenker and Basset Jr. (1978) realizaram o ajuste de um modelo de Regressão Quantílica para cada quantil separadamente, tomando como base o método dos mínimos absolutos de forma ponderada, onde a ponderação ocorre diante do quantil e de seu complementar conforme demonstrado abaixo:

$$\min_{\beta \in \mathbb{R}^k} [\sum_{t \in \{t: y_t \geq x_t \beta\}} q |y_t - x_t \beta| + \sum_{t \in \{t: y_t < x_t \beta\}} (1 - q) |y_t - x_t \beta|],$$

onde y_t é a variável resposta do modelo sendo t correspondente a t -ésima observação, x_t são as covariáveis do modelo, β são os coeficientes da regressão e q o quantil de referência da regressão ajustada.

Diante desta ideia, Yu and Moyeed (2001) mostraram que utilizar a densidade da distribuição de Laplace Assimétrica é equivalente a minimizar a função apresentada por Koenker and Basset Jr. (1978), facilitando na demonstração dos ajustes da Regressão Quantílica, sendo a distribuição de Laplace Assimétrica introduzida como distribuição do erro do modelo de regressão quantílica condicional e sua densidade definida por:

$$f_q(\epsilon_q | \mu, \tau) = \tau q (1 - q) \exp\{-(1 - q)\tau |\epsilon_q - \mu|\} 1[\epsilon_q < \mu] + \tau q (1 - q) \exp\{-q\tau |\epsilon_q - \mu|\} 1[\epsilon_q \geq \mu],$$

onde ϵ_q corresponde ao erro do modelo de regressão quantílico condicional sendo q o quantil de referência do ajuste, μ é o parâmetro de locação da distribuição Laplace Assimétrica e τ é o parâmetro de dispersão da distribuição Laplace Assimétrica.

Dessa forma, definindo a função perda correspondente, sendo que a minimização desta função perda corresponderá justamente a assumir os erros com a distribuição Laplace Assimétrica, sendo possível redefinir a densidade da distribuição de Laplace Assimétrica da seguinte forma:

$$f_q(\epsilon_q|\mu, \tau) = \tau q(1 - q)\exp\{-\tau\rho_q(\epsilon_q - \mu)\},$$

onde ρ_q será a função perda citada, que é definida por:

$$\rho_q(y) = -y(1 - q)1[y < 0] + yq1[y \geq 0],$$

sendo y um conjunto de observações que pode ser denotado como $y = \{y_1, y_2, \dots, y_n\}$.

Portanto, diante da minimização da esperança da função perda pelo $\operatorname{argmin}_{\mu} \sum_{i=1}^n \rho_q(y_i - \mu)$ será obtido o q -ésimo quantil empírico e a partir disso definir a distribuição de Laplace Assimétrica na forma $\epsilon_q \sim \text{AL}(\mu, \tau, q)$, assim ao considerar $\mu = 0$ fica possível definir o valor do quantil da seguinte forma:

$$P_r(\epsilon_q < 0) = q.$$

No contexto Bayesiano, outro ponto importante também apresentado por Yu and Moyeed (2001) é que diante de um conjunto de observações $y = \{y_1, y_2, \dots, y_n\}$, a distribuição a posteriori dos coeficientes β definida por $\pi(\beta|y)$ será dada por:

$$\pi(\beta|y) \propto L(y|\beta)p(\beta),$$

onde $p(\beta)$ é a distribuição a priori dos coeficientes da Regressão Quantílica, sendo comum assumir uma priori imprópria a partir da distribuição Uniforme, e $L(y|\beta)$ é a função de verossimilhança definida por:

$$L(y|\beta) = q^n(1 - q)^n \exp\{-\sum_{i=1}^n \rho_q(y_i - x_i' \beta)\},$$

onde $\mu_i = x_i' \beta$ será o parâmetro de locação.

Resumo

Analisar a velocidade do vento no estado de Minas Gerais é de extrema importância para planejamentos estratégicos da geração de energia através de fontes renováveis. Regressão Quantílica permite realizar a inferência sobre uma variável resposta em diversos quantis, ao contrário da Regressão usual em que o estudo é realizado diante da média do processo. Sendo assim, entender o comportamento da velocidade do vento em diferentes quantis é de suma importância para um melhor entendimento do comportamento da velocidade do vento. Para isso, neste trabalho será proposto um modelo de Regressão Quantílica Espacial para criar mapas dos quantis, provenientes desta regressão, referentes ao potencial eólico no estado de Minas Gerais, e assim, observar pontos específicos em relação à velocidade do vento com a possibilidade de implementação de um parque eólico.

KEY WORDS: INLA, Kumaraswamy, Log-Logística, Previsão da Velocidade do Vento, Regressão Espacial Quantílica, SPDE.

1. INTRODUÇÃO

A energia eólica possui um ótimo custo benefício em comparação com os demais meios de geração de energia, pois além de ser um meio de geração limpo, ou seja, que não emite algum tipo de poluição ou gera prejuízo ao meio ambiente, também é renovável, pois o vento é um bem inesgotável. Bockhorst and Barber (2010) descreveram a importância em realizar estudos envolvendo a previsão da velocidade do vento, com o intuito de reduzir os custos substanciais associados à integração da energia eólica à rede elétrica e, conseqüentemente, reduzir possíveis impactos ambientais oriundos de outras fontes de produção de energia elétrica.

A Regressão Quantílica é uma alternativa à regressão usual que permite entender o comportamento da variável resposta em relação a determinados quantis, pois ao invés da análise na regressão estar relacionada à média da variável resposta, a análise se realizará diante da mediana ou outro quantil de interesse.

Koenker and Basset Jr. (1978) introduziram a regressão quantílica pela necessidade de entender o comportamento dos ajustes de regressão para cada um dos quantis separadamente e assim, permitir uma generalização natural ao modelo linear de certos estimadores robustos de locação. Yu and Moyeed (2001) mostraram que a utilização da densidade da distribuição de Laplace Assimétrica é equivalente a minimizar a função da regressão quantílica apresentada por Koenker and Basset Jr. (1978). A partir daí, inúmeras inovações nessa área surgiram e podem ser vistas, por exemplo, no contexto Bayesiano (Tsionas (2003), Reed and Yu (2009) e Kozumi and Kobayashi (2011)), na utilização da regressão quantílica com regressão Lasso (Li et al. 2010), no ajuste de modelos de mistura (Reich et al. 2010), na utilização de modelos semi-paramétricos (Kottas and Krnjajic (2009) e Stasinopoulos et al. (2017)) e modelos não paramétricos (Hallin et al. 2009). No contexto espacial, Lum and Gelfand (2012) estudaram a regressão quantílica incorporando um campo Gaussiano através da representação estocástica de distribuição Laplace assimétrica. Essa modificação fez com que o modelo proposto fosse capaz de modelar dependência espacial nos quantis.

Assim como a Laplace assimétrica, algumas distribuições podem ser reparametrizadas por seus quantis. A distribuição Kumaraswamy (Kumaraswamy 1980) vem sendo muito utilizada para ajustes referentes a estudos quantílicos, possuindo em sua definição que o

suporte desta distribuição compreende os valores reais entre 0 e 1. Na literatura, diferentes abordagens mostram a flexibilidade e a capacidade dessa distribuição em modelar aplicações quantílicas e.g., Carrasco et al. (2010), Mitnik and Baek (2012), Mahmoud et al. (2015), Taillardat et al. (2016), entre outros.

Outra alternativa à distribuição Kumaraswamy é a reparametrização da distribuição Log-Logística em relação aos seus quantis. No entanto, o suporte da distribuição Log-Logística, diferente da distribuição Kumaraswamy, abrange o conjunto dos números reais positivos. Strupczewski et al. (2005) mostraram que a distribuição Log-Logística possui um desempenho mais adequado, principalmente para os chamados quantis superiores. Aplicações da distribuição Log-Logística em modelagem de problemas ambientais são estudadas por Shoukri et al. (1988), Mkhandi et al. (1996), entre outros.

Com o intuito de analisar a velocidade do vento em diferentes quantis no estado de Minas Gerais, neste trabalho é introduzido um modelo hierárquico que permite uma modelagem espacial quantílica de maneira prática, intuitiva e eficiente. Essa abordagem é diferente da apresentada por Lum and Gelfand (2012) que utiliza a representação estocástica da Laplace Assimétrica para fazer a modelagem espacial. Neste trabalho, serão utilizados modelos baseados nas distribuições Kumaraswamy e Log-Logística, modelando os quantis inferiores a fim de identificar pontos em que velocidade do vento permanece alta durante praticamente todo o período de tempo avaliado, e os quantis superiores para verificar em quais locais a velocidade do vento atinge níveis que são considerados picos de captação de energia eólica por algum período de tempo.

Além do estudo do comportamento da velocidade do vento com a análise da regressão quantílica, este trabalho também apresenta a modelagem espacial do território de Minas Gerais para colocação de novos parques eólicos para geração de energia. Para isso, são utilizadas metodologias de extrapolações, assim como foi apresentado em Jung (2016) que realiza um estudo sobre a produção anual de energia eólica de forma espacial utilizando extrapolações da velocidade do vento para uma área não coberta pela captação dos dados utilizando também a predição com a abordagem de séries temporais.

O formato desta dissertação é desenvolvido da seguinte forma. Na Seção 2, é apresentada a origem dos dados e a forma com que foram coletados, além de uma breve análise descritiva. Na Seção 3, são descritas as metodologias que envolvem os ajustes realizados, percorrendo

sobre as principais distribuições utilizadas, no caso, Kumaraswamy e Log-Logística. Na Seção 4, serão apresentados os passos dos ajustes realizados e a análise dos gráficos obtidos. Concluindo com a Seção 5, enaltecendo sobre as contribuições deste trabalho, além da proposição de futuros novos ajustes.

2. DADOS

Os dados utilizados neste trabalho foram coletados em medições realizadas por torres anemométricas de 10 metros de altura e disponibilizados pelo Instituto Nacional de Meteorologia (INMET) no estado de Minas Gerais (MG) no período de 01 de Janeiro de 2016 a 31 de Março de 2016. O INMET possuía 45 torres anemométricas no estado de MG no período avaliado, sendo importante ressaltar que a cada coleta de dados realizada pelo INMET há mudança na quantidade e inclusive nas torres onde são captados os dados utilizados neste trabalho. A Figura 1 apresenta a localização das torres. As áreas retangulares são áreas consideradas pela Companhia Energética de Minas Gerais (CEMIG) como regiões de alto potencial eólico (Amarante et al. 2010), sendo importante definir neste momento estas regiões para definições ao longo desta Dissertação, assim a área apresentada pelo retângulo em amarelo corresponde ao Triângulo Mineiro. A área indicada pelo retângulo verde corresponde a uma interseção entre a Região Metropolitana de Belo Horizonte, parte inferior do retângulo, e a região do Colar Metropolitano destacando-se a cidade de Sete Lagoas e também a região próxima a cidade de Curvelo, parte superior do retângulo. A área apresentada no retângulo roxo representa a região do Norte de Minas Gerais, mais precisamente regiões no entorno da cidade de Janaúba. Por fim, a área do retângulo azul corresponde a região que integra a cidade de Montes Claros. Sendo importante ressaltar também a região representada por um círculo laranja que não foi definida como região de alto potencial, mas foi uma região relevante na realização desta Dissertação e corresponde ao Sul de Minas Gerais, na região próxima a Itajubá.

Além da velocidade do vento, outras variáveis como direção do vento, umidade relativa do ar, temperatura (medida em graus celsius) e pressão atmosférica também são medidas pelas torres. As variáveis temperatura e pressão atmosférica possuem alta correlação e, por este motivo, são traduzidas em uma única variável que corresponde à densidade do ar, definida

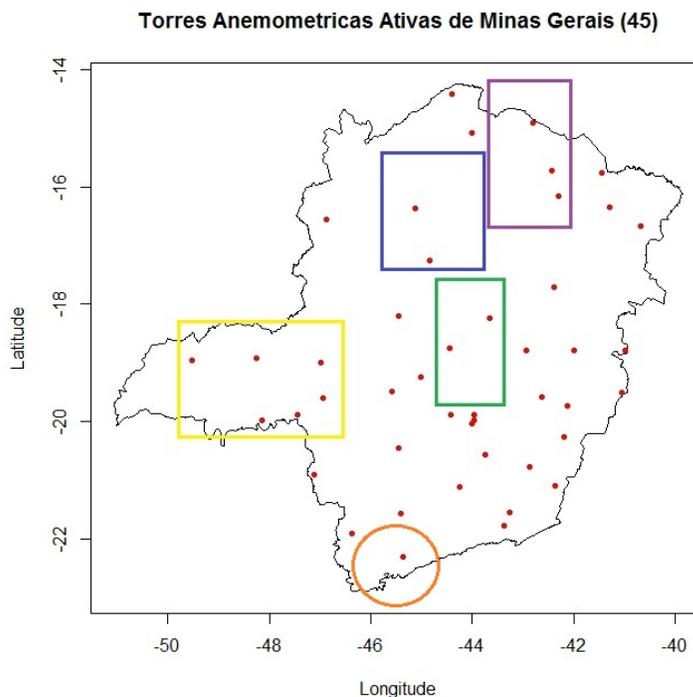


Figura 1: Mapa das Estações de Coleta da Velocidade do Vento

diante da lei dos gases ideais (Clapeyron 1834) como:

$$\text{Densidade} = \frac{\text{Pressao} \times 100}{(\text{Temperatura} + 273.15) \times 287.058}. \quad (1)$$

Modelos numéricos para previsão da velocidade do vento vem sendo utilizados em diversas escalas. O modelo ETA é definido pelo Instituto Nacional de Pesquisas Espaciais (INPE) para fazer a estimativa da velocidade do vento em toda América do Sul utilizando coordenadas verticais. Portanto, além das variáveis medidas pelas torres, optou-se por incluir a velocidade do vento estimada pelo modelo ETA neste estudo. A captação destas medidas é realizada em pontos de 5 em 5 Km no território de MG formando assim um grid de captação, onde também são armazenadas as variáveis de direção do vento, umidade relativa do ar, temperatura (medida em graus celsius) e pressão atmosférica deste modelo numérico que servirão de base para construção da modelagem espacial proposta na Seção 3.

A Figura 2 representa os histogramas dos dados coletados. Diante destes histogramas, é possível perceber que a direção do vento, medida em graus, possui uma concentração em torno de 110 graus. A umidade relativa do ar é medida em porcentagem da quantidade

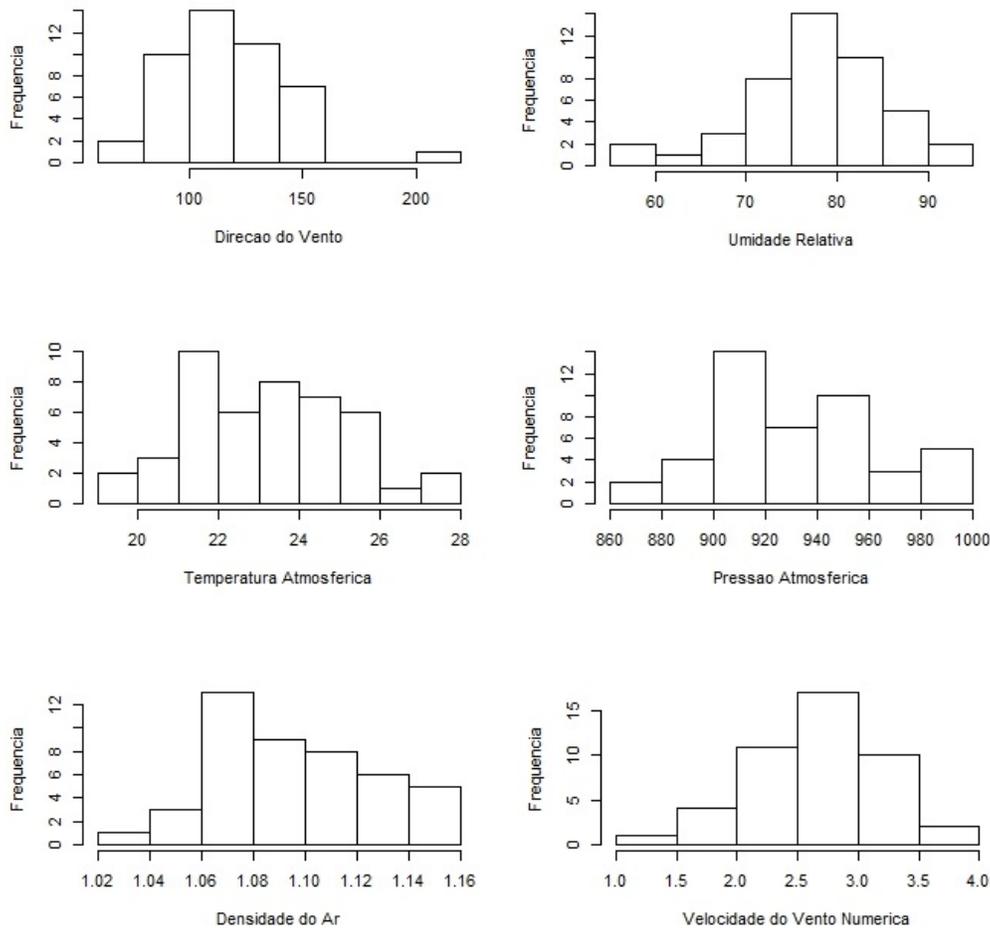


Figura 2: Histogramas das Covariáveis

de água presente no ar. Diante da análise dos dados, percebe-se uma umidade relativa centrada em 80%. A temperatura atmosférica que foi utilizada como função da densidade do ar apresenta um comportamento de variação entre 19°C e 28°C, com maior frequência na faixa entre 22°C e 26°C. A pressão atmosférica também utilizada como função da densidade do ar variou de 860 a 1000 hectopascal, concentrando em 900 hectopascal. A densidade do ar calculada de acordo com a Equação (1) varia de 1.02 até 1.16 kilogramas por metros cúbicos, com uma faixa de concentração em torno de 1.07 kilogramas por metros cúbicos. A velocidade do vento numérica estimada pelo método ETA possui uma variação de 1 a 4 metros por segundo, com uma concentração maior ao redor de 2.7 metros por segundo.

Além disso, este período de referência corresponde ao período da estação do ano do verão,

fator relevante para entendimento do comportamento da velocidade do vento e comparação dos resultados obtidos pelo atlas eólico produzido pela CEMIG (Amarante et al. 2010).

3. METODOLOGIA

Neste trabalho é proposto um modelo hierárquico com modelagem quantílica espacial possibilitando ajustes de maneira prática, intuitiva e eficiente. O modelo é definido da seguinte forma:

$$\begin{aligned}
 y_s | \beta, \phi_s, X_s &\sim \Pi(y_s | \mu_s, \tau_s, q), \\
 g(\mu_s) &= X_s \beta + \phi_s, \\
 \phi_s &\sim GP(0, \Sigma),
 \end{aligned}
 \tag{2}$$

onde y_s corresponde à variável resposta na posição s no espaço; β são os coeficientes das covariáveis; ϕ_s corresponde ao efeito espacial do modelo; X_s é a matriz de covariáveis do modelo; Π corresponde a uma distribuição de probabilidade, algumas de interesse serão vistas com maior detalhe ainda nesta seção; μ_s corresponde a um parâmetro de locação; τ_s corresponde a um parâmetro de dispersão; q corresponde ao quantil de interesse; a função $g(\mu_s)$ é a função de ligação ao preditor linear; e Σ é a função de covariância. O objetivo deste trabalho está não só na modelagem, mas também na predição, ou seja, encontrar a preditiva de $Y_{s_0} | Y$, considerando s_0 como uma nova observação no espaço.

A metodologia *Stochastic Partial Differential Equation* (SPDE, Lindgren et al. 2011) é definida diante da utilização de modelos deterministas a partir de condições aleatórias iniciais e da presença de ruídos, servindo para reduzir a alta complexidade computacional envolvida no âmbito dos processos Gaussianos, onde através de uma equação estocástica parcial realizará a discretização de um processo Gaussiano com a estrutura de covariância pertencente a família Matérn, mantendo uma aproximação adequada e eficiente e ao mesmo tempo garantindo uma superfície suave, possibilitando a obtenção do processo Gaussiano envolvido no ajuste. No contexto geral, sempre em que há o interesse em estabelecer a informação sobre a posição dos dados no espaço, é utilizado algum dos sistemas geográficos, como por exemplo, a latitude e a longitude. Sendo intuitivo pensar que em um modelo de regressão seria possível utilizá-los como covariáveis, porém para a obtenção dos resultados seria necessário utilizar funções complexas não-lineares ou funções não-paramétricas, desta

Mapa de Triangulações e Nova Malha de Minas Gerais

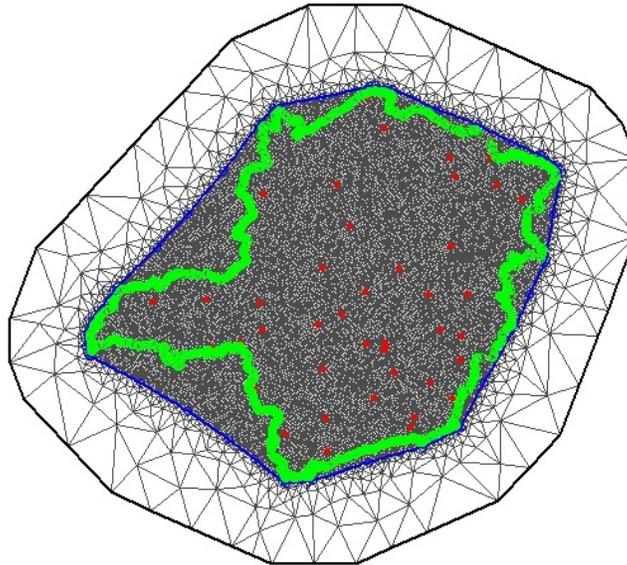


Figura 3: Mapa de Triangulação SPDE do Estado de Minas Gerais

forma o SPDE surge como uma maneira intuitiva de acrescentar o efeito espacial ao modelo. Além disso, segundo a primeira lei da Geografia apresentada por Tobler (1970) em que 'Tudo está relacionado com todo o resto, no entanto coisas próximas estão mais relacionadas que coisas distantes', daí surge a necessidade também da Dependência Espacial. Assim, o SPDE funciona como uma ligação entre Campos Gaussianos e Campos Aleatórios Gaussianos de Markov com o intuito de definir o efeito espacial no modelo de regressão. Neste trabalho, são definidas triangulações para formar uma nova malha na área de análise, baseada em algumas características do local, como por exemplo a escala cartográfica do mapa de referência. Diante da criação desta malha, são definidos os limites espaciais, para então elaborar a nova projeção para realização das interpolações, e identificar a função de covariância Matérn, podendo ser identificados também os parâmetros da distribuição a priori que estarão envolvidos no processo de interação entre SPDE e INLA, que será apresentado a seguir, obtendo o mapa de triangulação definido na Figura 3, onde a linha verde representa o limite territorial do estado de Minas Gerais e os pontos vermelhos representam a localização das torres anemométricas.

Considerando a Figura 3 é importante ressaltar que a precisão das interpolações estão diretamente relacionadas ao tamanho da área dos triângulos gerados, sendo que quanto menor

o triângulo gerado, maior será a precisão da interpolação naquela área, tendo consequentemente um maior custo computacional. Outro ponto importante está relacionado a questão dos vértices destes triângulos que acarretarão na dependência espacial, assim triângulos que possuem vértices em comum influenciam na interpolação de seus 'vizinhos'. Para obtenção das áreas de interpolação destes triângulos foi tomada como base as coordenadas geográficas do modelo numérico ETA de 5 Km, servindo como aproximação para o contorno de Minas Gerais e também para criação da malha que receberá os valores das interpolações.

A utilização do pacote *Integrated Nested Laplace Approximations* (INLA, Rue et al. 2009) surge como uma solução eficiente para inferência Bayesiana que possibilita a obtenção da distribuição a posteriori marginal da variável resposta diante de aproximações de Laplace com a utilização de variáveis latentes e integrações numéricas dos hiperparâmetros. Com a metodologia INLA, além de não ser necessário obter a distribuição a posteriori a partir de uma amostra conjunta desta distribuição, em comparação com métodos que envolvem Monte Carlo via cadeias de Markov, são obtidos resultados precisos com um custo computacional reduzido. No contexto geral, é realizada a aproximação das Marginais Posteriores do Campo Gaussiano Latente a partir de três passos básicos do INLA. O primeiro passo consiste em realizar a aproximação da marginal posterior dos hiperparâmetros utilizando a aproximação de Laplace. Já no segundo passo, será realizada a aproximação de Laplace para obtenção da densidade condicional das variáveis Gaussianas latentes em relação às variáveis observadas e aos valores definidos dos hiperparâmetros, obtendo uma adaptação para as aproximações Gaussianas. Então, no terceiro passo desta abordagem são realizadas integrações numéricas para combinar os resultados obtidos nos dois passos anteriores. O pacote INLA fornece medidas de qualidade de ajuste do modelo, como *Deviance Information Criteria* (DIC), *Watanabe-Akaike Information Criteria* (WAIC) e *Conditional Predictive Ordinate* (CPO), sendo sua principal desvantagem relacionada ao custo computacional diante da quantidade de hiperparâmetros. Neste trabalho, os hiperparâmetros definidos nos passos do pacote INLA tem uma ligação com os processos envolvidos na metodologia SPDE, diante das definições dos parâmetros a priori da inferência Bayesiana relacionado ao modelo ajustado das distribuições quantílicas escolhidas, portanto combinando a metodologia SPDE com os ajustes proporcionados pelo pacote INLA são obtidos os resultados gerados na Seção 4.

Diante das ideias apresentadas sobre a metodologia SPDE e o pacote INLA, é importante

também fornecer uma breve ideia sobre a função de covariância Matérn, visto que sua utilização foi fundamental para definição dos efeitos espaciais a partir de um Processo Gaussiano. Para obtenção destes resultados, foi necessária a definição do parâmetro κ de escala como um valor constante de 0.75 e do parâmetro ν de suavidade como um valor constante de 1, tendo também a utilização da distância Euclidiana na definição da covariância Matérn.

3.1 Distribuições Quantílicas

As distribuições quantílicas são distribuições de probabilidade usuais que possuem flexibilidade em sua parametrização possibilitando a regressão direta em quantis de interesse. A regressão quantílica surge da necessidade em estudar o comportamento de casos 'não-médios', ou seja, casos que envolvem os quantis da variável resposta. As distribuições Kumaraswamy e Log-Logística são distribuições que se caracterizaram como distribuições quantílicas e possuem diversas aplicações conforme apresentado na Seção 1, e por este motivo, foram escolhidas neste estudo.

A distribuição Kumaraswamy (Kumaraswamy 1980) denotada por $Y \sim K(\kappa, \phi, q)$, onde a variável resposta Y deve ter domínio definido nos reais entre 0 e 1, $\phi > 0$ como parâmetro de precisão, $\delta = \frac{1}{\phi}$ sendo parâmetro de dispersão e o valor do quantil $0 < q < 1$ conhecido (Bayes et al. 2017), tem densidade dada por:

$$f(y|\kappa, \delta) = -\frac{\log(1-q)}{\delta \log(1-e^{-1/\delta}) \log(\kappa)} \left\{ y^{-\frac{1}{\delta \log(\kappa)} - 1} \right\} \left\{ 1 - y^{-\frac{1}{\delta \log(\kappa)}} \right\}^{\frac{\log(1-q)}{\log(1-e^{-1/\delta})} - 1}. \quad (3)$$

A distribuição Log-Logística é denotada por $Y \sim LL(\kappa, \delta, q)$ e é definida nos reais positivos e possui a seguinte densidade:

$$f(y|\kappa, \delta) = \frac{q\kappa^{1/\delta}y^{1/\delta-1}}{\delta(\kappa^{1/\delta} + qy^{1/\delta})^2}, \quad (4)$$

onde $\kappa > 0$ é o parâmetro de locação e $\delta > 0$ é o parâmetro de dispersão, tendo também o valor do quantil $0 < q < 1$ conhecido.

3.2 Extrapolação da Velocidade do Vento

Os resultados obtidos pelos modelos apresentados nesta seção são feitos com a utilização de medidas observadas em torres de 10 metros de altura. Porém como um dos objetivos deste trabalho é a averiguação de um local para implantação de um parque eólico em Minas

Gerais, é necessário obter estimativas a 100 metros pois essa é a altura de geração de energia eólica. Na literatura existem alguns trabalhos que apresentam metodologias de realização da extrapolação como Banuelos-Ruedas et al. (2011) e Gualtieri and Secci (2012).

Para realizar a extrapolação da velocidade do vento, foi utilizada a seguinte fórmula:

$$v = v_0 \times \left(\frac{H}{H_0} \right)^\alpha,$$

onde v é a velocidade a 100 metros; v_0 é a velocidade a 10 metros; H corresponde a altura de 100 metros; H_0 corresponde a altura de 10 metros e α é o parâmetro referente ao terreno onde a extrapolação está sendo realizada.

O parâmetro de terreno α é definido a partir de medições do modelo numérico ETA de 5Km com a captação da velocidade do vento a 10 metros e também a 100 metros no período de Outubro de 2017, possibilitando a partir destes dados definidos em cada ponto do estado de Minas Gerais, a obtenção deste parâmetro de terreno em toda cobertura do território avaliado utilizando a seguinte fórmula:

$$\alpha = \frac{\ln(v) - \ln(v_0)}{\ln(H) - \ln(H_0)}.$$

A Figura 4 apresenta o histograma do parâmetro de terreno. Com base nas definições apresentadas por Banuelos-Ruedas et al. (2011), é possível definir que o terreno de Minas Gerais varia de uma área da pastagem ($\alpha = 0.15$) até uma área densamente florestada ($\alpha = 0.25$), porém tendo em grande parte de sua extensão uma área de plantas altas e arbustos ($\alpha = 0.20$). Assim, utilizando esta forma dinâmica de extrapolação da velocidade do vento em cada ponto de acordo com um nível adequado de α será possível obter resultados mais fidedignos considerando a grande diversidade presente no território de Minas Gerais.

3.3 Potência de Geração de Energia Eólica

Um local apropriado para implantação de um parque de geração de energia eólica possui certas especificidades, como pode ser visto na Figura 5. Assim, deve ser realizada uma transformação do valor da velocidade do vento à 100 metros de altura em um índice que defina o potencial de energia eólica no território de Minas Gerais. A ideia da utilização da potência de geração de energia eólica em substituição da velocidade do vento é essencial, já que nem sempre locais com uma grande diferença de velocidade do vento apresentam também uma grande diferença neste potencial de geração.

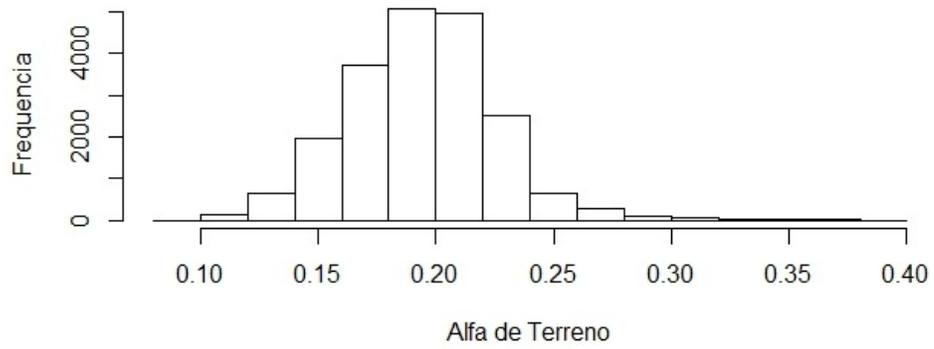


Figura 4: Histograma do Parâmetro Alfa de Terreno

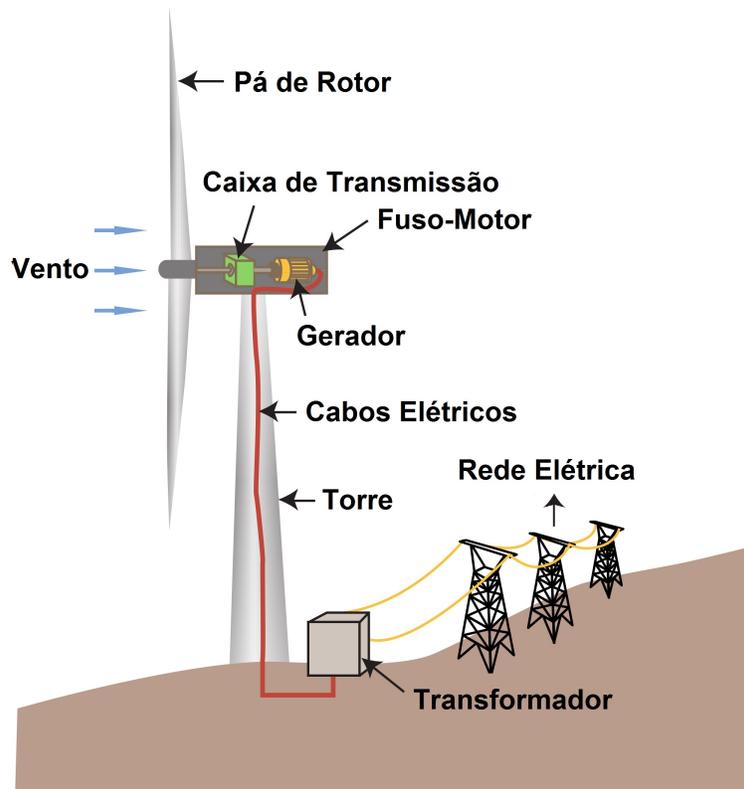


Figura 5: Figura referente ao funcionamento da Torre Eólica

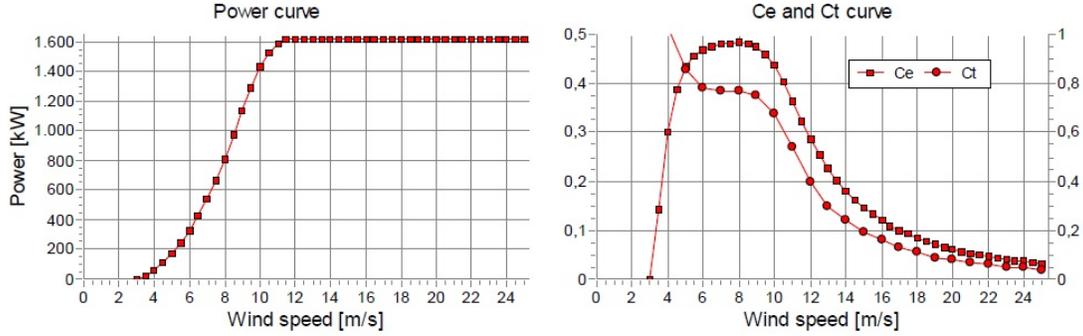
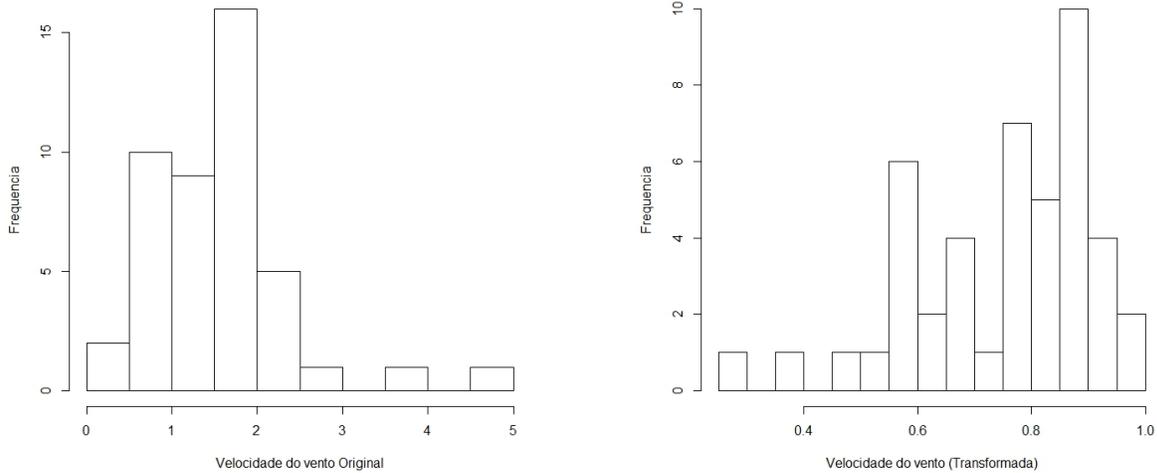


Figura 6: Gráficos da Potência e do Coeficiente Aerodinâmico

Para obtenção dos valores de potência de geração de energia eólica foi utilizada a seguinte fórmula:

$$P_s = \frac{1}{2} \times \rho_s \times A_r \times v_s^3 \times C_p \times \eta,$$

onde P_s é a potência elétrica medida em Watts; ρ_s corresponde a densidade do ar; A_r corresponde a área varrida pelo rotor; v_s é a velocidade medida a 100 metros de altura em metros por segundo, sendo s a localização espacial da torre; C_p corresponde ao coeficiente aerodinâmico de potência do rotor e η é a eficiência do conjunto gerador-transmissor. O parâmetro A_r é considerado constante supondo que em um parque todas as torres de geração de energia eólica possuem as mesmas dimensões. Assim, o valor do diâmetro do rotor das torres foi escolhido como 82.5 metros, e a área varrida pode ser obtida como $A_r = \pi \times \left(\frac{Diametro}{2}\right)^2$. A variável v é obtida diante da abordagem de extrapolação da velocidade do vento apresentada na Subseção 3.2, o C_p é um valor definido para determinadas velocidades como pode ser visto na Figura 6 obtida através de Juliano (2013). Como a velocidade do vento a 100 metros deve estar dentro dos intervalos dos pontos definidos, foi realizado um ajuste de um Spline cúbico obtendo então uma aproximação para o parâmetro C , definindo então o valor do coeficiente aerodinâmico para os valores da velocidade do vento observados. Por fim o valor de η foi obtido diante do texto técnico (Juliano 2013) fornecido pela CEMIG, onde os valores de potência estavam definidos de acordo com determinadas velocidades e coeficientes aerodinâmicos. Assim, foi assumido um valor constante de 0.001 para este parâmetro.



(a) Velocidade do Vento Original

(b) Velocidade do Vento Transformada

Figura 7: Figura referente ao Comportamento da Velocidade do Vento

4. RESULTADOS

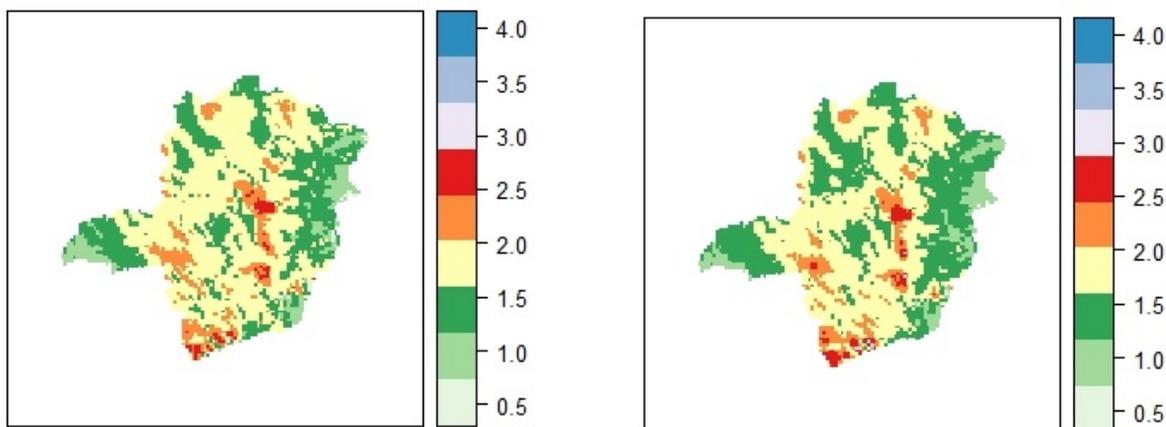
Os resultados apresentados nesta seção foram obtidos com o ajuste dos modelos de regressão na mediana ($q=0.50$) da velocidade do vento a 10 metros considerando também a localização espacial das torres de referência no estado de Minas Gerais. Foi utilizado o pacote R INLA (Lindgren and Rue 2015), estritamente na versão 0.0-1480869339, através do software estatístico R versão 3.3.2 (R Core Team 2016).

A variável resposta velocidade do vento possui valores entre 0 e 5 metros por segundo. Porém, como o suporte da distribuição Kumaraswamy é definido no intervalo entre 0 e 1, a seguinte transformação para a variável resposta, foi realizada considerando a facilidade na obtenção de sua transformação inversa que é fundamental para definição dos resultados finais:

$$g : [0, \infty) \rightarrow [0, 1],$$

$$g(y) = 1 - \exp(-y).$$

Na Figura 7 é possível verificar o comportamento da velocidade do vento a 10 metros utilizada como variável resposta no modelo Log-Logístico e o comportamento da variável resposta transformada para o suporte da distribuição Kumaraswamy.



(a) Distribuição Kumaraswamy

(b) Distribuição Log-Logística

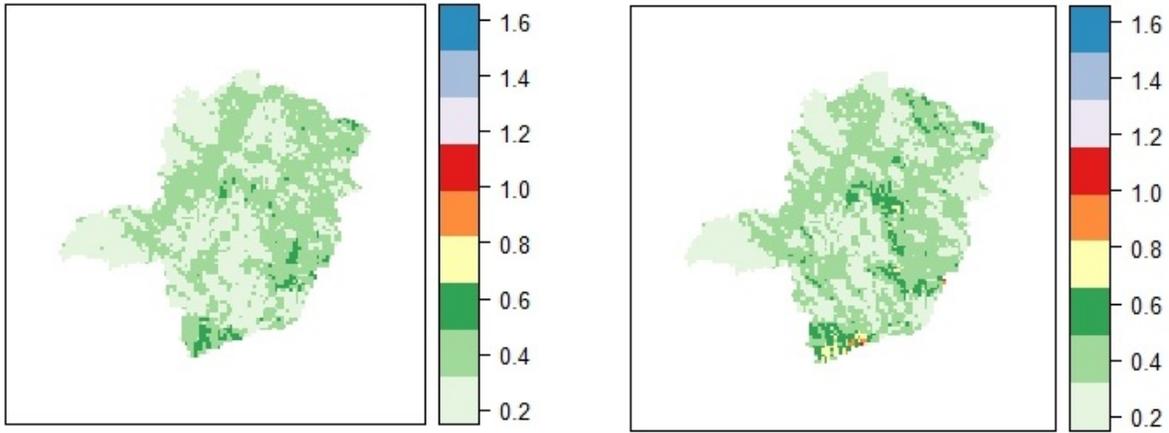
Figura 8: Figura de Níveis da Média da Amostra da Distribuição a Posteriori

Uma observação importante é que a correlação da variável original com a velocidade do vento estimada pelo modelo numérico é de 0.34, ressaltando a importância de manutenção desta estimativa da velocidade do vento como uma covariável do modelo.

Foram elaborados gráficos de níveis para as diferentes distribuições a fim de comparar o desempenho de ajuste utilizando o modelo proposto na Equação (2). A Figura 8 corresponde ao comportamento da média da amostra da distribuição a posteriori da velocidade do vento para as distribuições de interesse. Analisando os resultados apresentados, percebe-se que há uma indicação de um alto nível do valor médio da mediana da velocidade do vento na Região Metropolitana e do Colar Metropolitano de Belo Horizonte. Além disso, é possível perceber alguns pontos com alto nível de velocidade do vento também na região que engloba o Norte de Minas Gerais, região de Janaúba, além de alguns pontos isolados no Sul de Minas Gerais, na região de Itajubá, e outros pontos próximos ao Triângulo Mineiro, nas duas distribuições avaliadas.

Na Figura 9, são apresentados os resultados com relação ao desvio-padrão da amostra da distribuição a posteriori da velocidade do vento. Percebe-se que a distribuição Log-Logística apresenta níveis elevados de desvio-padrão para a mediana da velocidade do vento nas regiões Sul e Leste de Minas Gerais, fato não captado no ajuste pela distribuição Kumaraswamy.

Analisando os resultados obtidos das covariáveis diante dos modelos ajustados, como pode ser observado na Tabela 1 para a distribuição Kumaraswamy, foram significativos, levando



(a) Distribuição Kumaraswamy

(b) Distribuição Log-Logística

Figura 9: Figura de Níveis do Desvio-Padrão da Amostra da Distribuição a Posteriori

Tabela 1: Tabela de Covariáveis do Modelo Kumaraswamy

| Covariáveis do Modelo Kumaraswamy | | | | | |
|-----------------------------------|------------|-------|-----------|-----------|-----------|
| Covariável | Estimativa | SD | HPD-0.025 | HPD-0.500 | HPD-0.975 |
| Intercepto | 26.203 | 6.781 | 12.905 | 26.170 | 39.669 |
| Direção | -0.010 | 0.006 | -0.022 | -0.010 | 0.001 |
| Densidade | -17.388 | 5.139 | -27.670 | -17.343 | -7.383 |
| Umidade | -0.058 | 0.018 | -0.094 | -0.058 | -0.023 |
| Numérica | -0.073 | 0.250 | -0.563 | -0.073 | 0.420 |

em consideração o intervalo HPD das estimativas, o intercepto e as covariáveis de Densidade do Ar e Umidade Relativa, sendo que ambas as covariáveis apresentaram correlação negativa com a velocidade do vento. Realizando a mesma averiguação em relação às covariáveis do modelo, agora para a distribuição Log-Logística, considerando a Tabela 2, assim como no modelo Kumaraswamy, foram significativos, também levando em consideração o intervalo HPD das estimativas, o intercepto e as covariáveis de Densidade do Ar e Umidade Relativa, sendo que ambas as covariáveis apresentaram correlação negativa com a velocidade do vento.

Para analisar o comportamento do modelo em relação ao ajuste para cada torre, foi obtido um gráfico com o valor ajustado da regressão, o intervalo HPD desse ajuste e também uma marcação indicando o valor original da variável resposta apresentados pela Figura 10.

Tabela 2: Tabela de Covariáveis do Modelo Log-Logístico

| Covariáveis do Modelo Log-Logístico | | | | | |
|-------------------------------------|------------|-------|-----------|-----------|-----------|
| Covariável | Estimativa | SD | HPD-0.025 | HPD-0.500 | HPD-0.975 |
| Intercepto | 12.512 | 3.247 | 6.013 | 12.548 | 18.810 |
| Direção | -0.004 | 0.003 | -0.009 | -0.003 | 0.002 |
| Densidade | -8.599 | 2.495 | -13.442 | -8.624 | -3.615 |
| Umidade | -0.028 | 0.010 | -0.047 | -0.028 | -0.009 |
| Numérica | -0.025 | 0.127 | -0.275 | -0.025 | 0.224 |

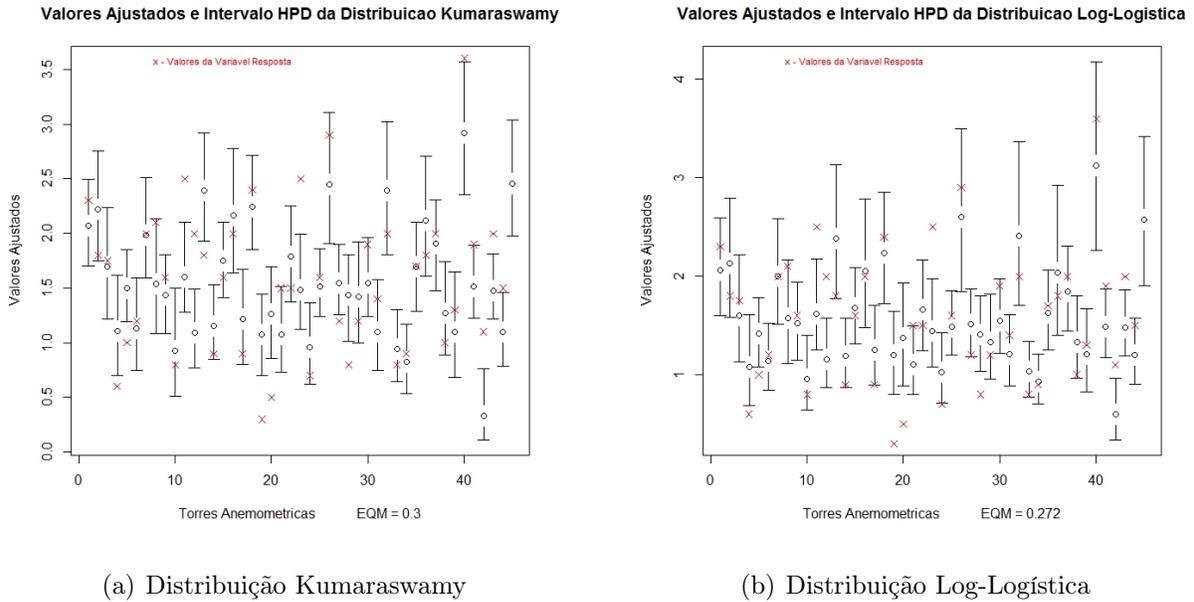


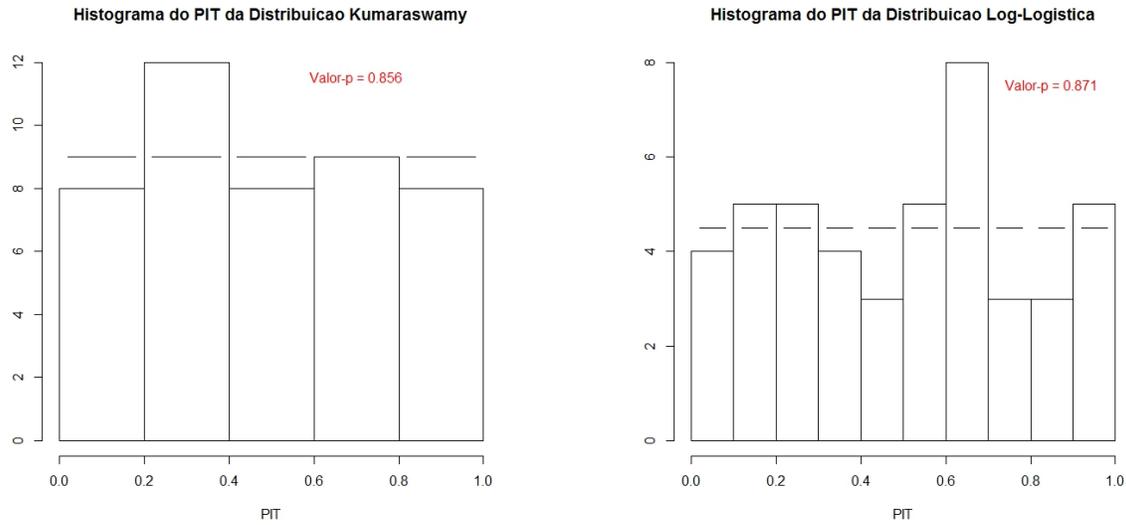
Figura 10: Figuras de Intervalo HPD

Além disso, para comparar as distribuições utilizadas, foi calculado o Erro Quadrático Médio (EQM) do ajuste de cada modelo, definido por:

$$EQM(\hat{Y}) = E(\hat{Y} - Y)^2 = Var(\hat{Y}) + [Vies(\hat{Y})]^2.$$

As distribuições Kumaraswamy e Log-Logística apresentaram EQM de 0.300 e 0.272, respectivamente. Com isso, é possível verificar que o EQM destas distribuições indicam bom ajuste dos modelos, observando uma leve vantagem da distribuição Log-Logística.

Com o intuito de verificar se os modelos ajustados oferecem um bom ajuste aos dados, o teste de uniformidade *Probability Integral Transform* (PIT, Dawid 1984) foi utilizado. O



(a) Distribuição Kumaraswamy

(b) Distribuição Log-Logística

Figura 11: Histograma de Valores do Teste PIT

intuito deste teste é servir como uma medida de qualidade de ajuste do modelo, em que a hipótese do teste é averiguar se o modelo está adequado para realização de análises preditivas. Assim, foram construídos os histogramas apresentados na Figura 11 com os valores obtidos do CPO fornecido pelo pacote INLA dos modelos analisados indicando quais deveriam ser os valores de frequência diante da transformação destas medidas para assumir uma distribuição Uniforme e assim indicar em relação a adequação sobre a realização de análises preditivas. Analisando estes histogramas e os resultados do teste de uniformidade realizado, representado pela linha pontilhada nos histogramas e o p-valor apresentado, onde a hipótese nula deste teste indica que o modelo possui um ajuste adequado, conclui-se que ambos modelos apresentam ajustes satisfatórios, visto que o p-valor do modelo Kumaraswamy e do modelo Log-Logístico foram de, respectivamente, 0.856 e 0.871, que diante de um nível de significância usual de 0.05 não há a rejeição da hipótese nula.

Observando a adequação dos modelos para análises preditivas, foi realizado um teste de predição a fim de comparar o desempenho dos modelos ajustados. Para realização deste teste foram excluídas, de forma aleatória, 10 torres dentre as 45 torres anemométricas presentes no estudo, ajustando o modelo novamente somente com as 35 torres restantes para as duas distribuições de interesse, obtendo ao final do processo uma predição da velocidade do vento

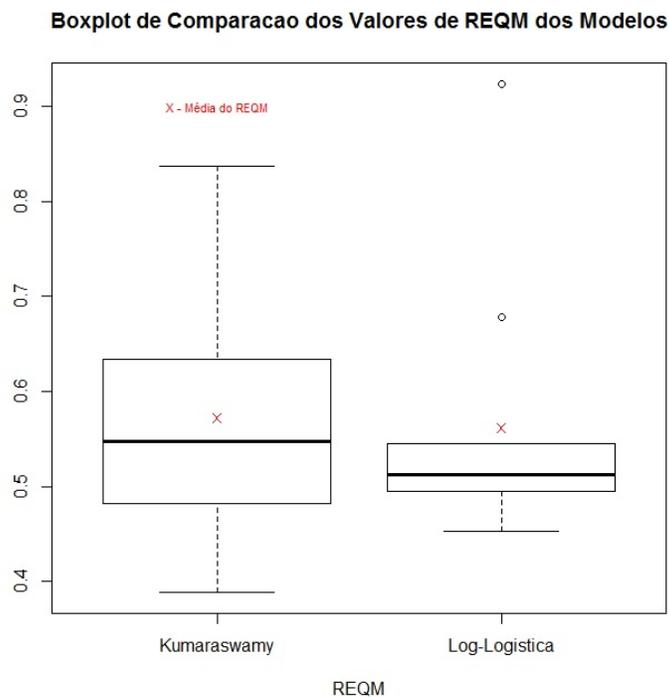
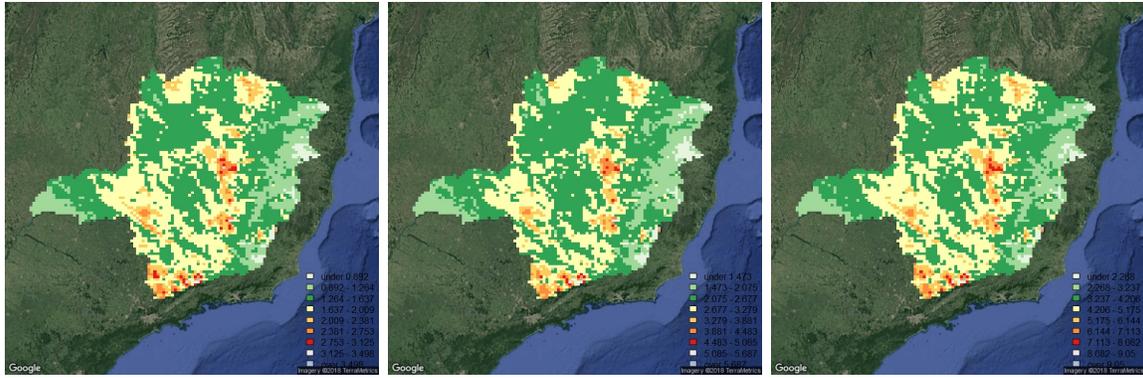


Figura 12: Boxplot dos Valores de REQM de Predição

utilizando as covariáveis das torres excluídas e então foi calculada a raiz do EQM. Este processo proposto para predição foi repetido 10 vezes, devido ao alto custo computacional envolvido, e a partir dos valores da raiz do EQM obtidos em cada repetição foi construído o Boxplot apresentado na Figura 12.

Diante da Figura 12 é possível notar que a distribuição Log-Logística teve um melhor desempenho em relação à predição em comparação com a distribuição Kumaraswamy a partir dos valores obtidos da raiz do EQM, já que sua faixa de concentração e a mediana foram inferiores.

Analisando todos os resultados apresentados, percebe-se um desempenho melhor da distribuição Log-Logística em relação a distribuição Kumaraswamy, não só em termos de poder preditivo, mas também de qualidade de ajuste. Além disso, para a utilização da distribuição Kumaraswamy é necessário realizar uma transformação na variável resposta para mudança do seu suporte, e para obtenção do resultado na escala original, é necessário realizar a transformação inversa. Essa transformação não é necessária na utilização da distribuição Log-Logística. Ressaltando também, o ponto avaliado na literatura por Strupczewski et al.



(a) Extrapolação - Quantil 10% (b) Extrapolação - Quantil 50% (c) Extrapolação - Quantil 90%

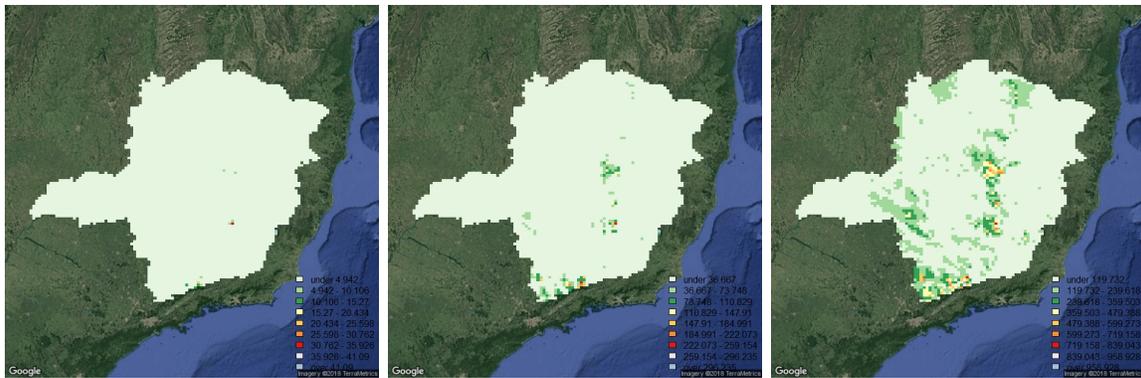
Figura 13: Mapa da Extrapolação da Velocidade do Vento

(2005) de que a distribuição Log-Logística possui um desempenho melhor para quantis superiores em relação a demais distribuições, algo que será analisado na prática nos próximos passos deste estudo.

Portanto, a partir de agora serão apresentados os resultados nos ajustes quantílicos considerando somente a distribuição Log-Logística. Foi realizado o ajuste no quantil 10% (quantil inferior) e também no quantil 90% (quantil superior) para averiguar regiões potenciais de geração de energia eólica. Para estes resultados, serão apresentados mapas para a mediana da amostra da distribuição a posteriori da velocidade do vento, já considerando esta velocidade do vento extrapolada para 100 metros de altura com o auxílio do pacote `spGoogle` (e Silva et al. 2012).

A Figura 13 apresenta os ajustes no quantil inferior ($q=0.10$), na mediana ($q=0.50$) e no quantil superior ($q=0.90$) para realizar a avaliação em diferentes cortes da velocidade do vento. Diante do resultado obtido, é possível perceber que as regiões que apresentam alta velocidade do vento a 100 metros são a Região Metropolitana e do Colar Metropolitano de Belo Horizonte, a região do Sul de Minas Gerais, a região do Norte de Minas Gerais e a região próxima ao Triângulo Mineiro em todos os três ajustes realizados.

Apenas a velocidade do vento pode não ser uma medida adequada para definição da produção de energia eólica, pois como apresentado na Figura 6 a potência gerada não depende de forma linear da velocidade do vento. Assim, serão apresentados os resultados referente ao potencial de geração de energia eólica obtido para a mediana, para o quantil inferior e também para o quantil superior a fim de buscar a melhor região para implantação de um



(a) Potência - Quantil 10% (b) Potência - Quantil 50% (c) Potência - Quantil 90%

Figura 14: Potencial de Geração de Energia Eólica

parque eólico.

Desta forma, A Figura 14 apresenta os ajustes no quantil inferior ($q=0.10$), na mediana ($q=0.50$) e no quantil superior ($q=0.90$) para realizar a avaliação em diferentes cortes da potência de geração de energia eólica. Diante do resultado obtido, é possível perceber que as regiões que apresentam alto potencial de geração de energia eólica para o ajuste na mediana ($q=0.50$) são a Região Metropolitana e do Colar Metropolitano de Belo Horizonte e a região do sul de Minas Gerais. Já com relação ao quantil inferior ($q=0.10$) observa-se que aparentemente não existe região em que a torre funcionaria por quase todo o período de tempo em sua capacidade ideal. Por fim, para o quantil superior ($q=0.90$) é nítido que existem diversas regiões em que existem picos de geração de energia, sendo que os melhores resultados são para a Região Metropolitana e do Colar Metropolitano de Belo Horizonte, a região do Sul de Minas Gerais, a região do Norte de Minas Gerais e a região próxima ao Triângulo Mineiro.

Assim, a melhor região para implantação do parque eólico onde haveriam picos de potencial de produção de energia eólica em períodos específicos, mantendo um nível adequado durante todo o período de avaliação, seria justamente a Região Metropolitana e do Colar Metropolitano de Belo Horizonte, sendo importante ressaltar que os resultados obtidos são para a estação do verão em que a velocidade do vento é, em geral, menor do que nas demais estações do ano.

5. DISCUSSÃO

A energia eólica é um tipo de energia limpa e renovável. Diante de todos os benefícios deste tipo de geração de energia, existem alguns desafios que são comumente discutidos, como a questão da forma de transmissão deste tipo de energia, já que o armazenamento não é possível. Assim, o vento é captado, gerando a energia que será transmitida simultaneamente. Por isso, a captação deve ocorrer em determinados momentos do dia, analisando os picos da velocidade do vento e levando em conta a capacidade máxima das torres de geração de energia eólica. Outro desafio relevante está relacionado à implantação de parques eólicos que deve ocorrer justamente nas áreas onde a velocidade do vento apresenta o padrão necessário para geração da energia de forma adequada. Existe também a questão de ser necessária uma área extensa para alocação das torres de geração de energia eólica e esta área deve ser distante da área urbana, devido aos ruídos gerados, que podem ser prejudiciais ao ser humano.

A grande inovação obtida na realização deste trabalho foi o modelo hierárquico apresentado que possui praticidade diante do uso de distribuições parametrizáveis pelo quantil, possibilitando estudos sobre diferentes cortes da variável resposta. Este modelo foi estruturado de forma intuitiva por possuir uma caracterização semelhante às distribuições usuais de modelos lineares generalizados, permitindo inclusive análises espaciais e preditivas. Assim, através da regressão quantílica espacial foi possível analisar locais nos quais a velocidade do vento permanece em um determinado nível durante praticamente todo o período de tempo avaliado. Para obtenção deste tipo de resultado, foram ajustadas as regressões quantílicas espaciais nos chamados quantis inferiores, neste trabalho definido como 10%. Além de locais onde a velocidade do vento atinge níveis que são considerados picos de captação de energia eólica por algum período de tempo específico. Nesse caso, foram ajustadas as regressões quantílicas espaciais nos chamados quantis superiores, neste trabalho definido como 90%. É possível que haja interseção entre as áreas onde é observado o pico específico e também o nível adequado de vento por um período maior. Isto ocorre porque existem áreas de grande potencial eólico em Minas Gerais. A partir desta análise quantílica e da definição da potência de geração da energia eólica, foram apresentadas regiões adequadas para a implantação do parque eólico, respeitando os princípios ambientais e de custo-benefício levantados anterior-

mente, ficando a critério da CEMIG, responsável pela geração de energia elétrica em Minas Gerais, a escolha da região baseando-se nos resultados apresentados.

A extrapolação da velocidade do vento para 100 metros, altura de funcionamento das torres de geração de energia eólica, foi apresentada de uma maneira dinâmica, já que foi atribuído a cada ponto da extensão do território um nível de terreno específico, fator importante diante da diversidade do território de Minas Gerais. Além disso, a transformação desta velocidade do vento extrapolada no potencial de geração de energia eólica foi fundamental para definição das regiões definidas como ideais, já que a curva de potencial é um indicativo que determina a partir de qual velocidade do vento as torres de geração de energia eólica começam a funcionar e também a partir de qual velocidade do vento a capacidade de produção atinge seu limite.

Como trabalhos futuros há o interesse de acrescentar um fator espaço-temporal e outras covariáveis meteorológicas como radiação e precipitação no ajuste dos modelos, avaliando também os resultados dos ajustes nas demais estações do ano. Além do desejo em realizar estudos de potencial eólico, fornecendo um caráter probabilístico aos mapas determinísticos do Atlas Eólico de Minas Gerais (Amarante et al. 2010), obtidos pela média de um histórico de previsões numéricas, já que os ajustes e resultados realizados neste trabalho possuem um caráter estocástico de predição espacial.

Referências

- Amarante, O. A. C., de Jesus Lima da Silva, F., and de Andrade, P. E. P. (2010), “Atlas Eólico de Minas Gerais,” Companhia Energetica de Minas Gerais (CEMIG).
- Banuelos-Ruedas, F., Camacho, C. A., and Rios-Marcuello, S. (2011), “Methodologies Used in the Extrapolation of Wind Speed Data at Different Heights and Its Impact in the Wind Energy Resource Assessment in a Region,” *Wind Farm - Technical Regulations, Potential Estimation and Siting Assessment*.
- Bayes, C. L., Bazan, J. L., and Castro, M. (2017), “A quantile parametric mixed regression model for bounded response variables,” *Statistics and its interface*, 10, 483–493.

- Bockhorst, J. and Barber, C. (2010), “Gaussian processes for short-horizon wind power forecasting,” *Not found*, 103–108.
- Carrasco, J. M. F., Ferrari, S. L. P., and Cordeiro, G. M. (2010), “A New Generalized Kumaraswamy Distribution,” Tech. rep., Departamento de Estatística, Universidade de Sao Paulo.
- Clapeyron, B. P. E. (1834), “Mémoire sur la puissance motrice de la chaleur,” *Journal de l’École Polytechnique*, 153–191.
- Dawid, A. P. (1984), “Statistical theory: The prequential approach,” *Journal of the Royal Statistical Society*, 147, 278–292.
- e Silva, L. G. S., Prates, M., and Azevedo, M. (2012), *spGoogle: Interacting R with Google maps: Spatial visualization tool*, r package version 0.1-3.
- Gualtieri, G. and Secci, S. (2012), “Methods to extrapolate wind resource to the turbine hub height based on power law: A 1-h wind speed vs. Weibull distribution extrapolation comparison,” *Renewable Energy*, 43, 183–200.
- Hall, P., Wolff, R. C., and Yao, Q. (1999), “Methods for estimating a conditional distribution function,” *Journal of the American Statistical Association*, 94, 154–163.
- Hallin, M., Lu, Z., and Yu, K. (2009), “Local linear spatial quantile regression,” *Bernoulli*, 15, 659–686.
- Hogg, R. V. (1975), “Estimates of Percentile Regression Lines Using Salary Data,” *Journal of the American Statistical Association*, 70, 56–59.
- Juliano (2013), *GE WIND ENERGY GE 1.6 1600 82.5 !O!*, windPRO 3.0.654 by EMD International A/S, Tel. +45 96 35 44 44, www.emd.dk, windpro@emd.dk.
- Jung, C. (2016), “High Spatial Resolution Simulation of Annual Wind Energy Yield Using Near-Surface Wind Speed Time Series,” *Energies*, 9, 344–363.
- Koenker, R. and Basset Jr., G. (1978), “Regression Quantiles,” *Econometrica*, 46, 33–50.

- Kottas, A. and Krnjajic, M. (2009), “Bayesian Semiparametric Modelling in Quantile Regression,” *Scandinavian Journal of Statistics*, 36, 297–319.
- Kozumi, H. and Kobayashi, G. (2011), “Gibbs sampling methods for Bayesian quantile regression,” *Journal of Statistical Computation and Simulation*, 81, 1565–1578.
- Kumaraswamy, P. (1980), “A generalized probability density function for double-bounded random processes,” *Journal of Hydrology*, 46, 79–88.
- Li, Q., Xi, R., and Lin, N. (2010), “Bayesian Regularized Quantile Regression,” *Bayesian Analysis*, 5, 533–556.
- Lindgren, F. and Rue, H. (2015), “Bayesian Spatial Modelling with R-INLA,” *Journal of Statistical Software*, 63, 1–25.
- Lindgren, F., Rue, H., and Lindstrom, J. (2011), “An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach,” *Journal of the Royal Statistical Society*, 73, 423–498.
- Lum, K. and Gelfand, A. E. (2012), “Spatial Quantile Multiple Regression Using the Asymmetric Laplace Process,” *International Society for Bayesian Analysis*, 7, 235–258.
- Mahmoud, M. R., El-Sherpieny, E. A., and Ahmed, M. A. (2015), “The New Kumaraswamy Kumaraswamy Family of Generalized Distributions with Application,” *Pak.j.stat.oper.res.*, XI, 159–180.
- Mitnik, P. A. and Baek, S. (2012), “The Kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation,” *Stat Papers*, 54, 177–192.
- Mkhandi, S. H., Kachroo, R. K., and Guo, S. L. (1996), “Uncertainty analysis of flood quantile estimates with reference to Tanzania,” *Journal of Hydrology*, 185, 317–333.
- R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

- Reed, C. and Yu, K. (2009), “An Efficient Gibbs Sampling for Bayesian quantile regression,” Tech. rep., Department of Mathematical Sciences, Brunel University.
- Reich, B. J., Bondell, H. D., and Wang, H. J. (2010), “Flexible Bayesian quantile regression for independent and clustered data,” *Biostatistics*, 11, 337–352.
- Rue, H., Martino, S., and Chopin, N. (2009), “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations,” *Royal Statistical Society*, 71, 319–392.
- Shoukri, M. M., Mian, I. U. H., and Tracy, D. S. (1988), “Sampling properties of estimators of the log-logistic distribution with application to Canadian precipitation data,” *The Canadian Journal of Statistics*, 16, 223–236.
- Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., and De Bastiani, F. (2017), *Flexible Regression and Smoothing: Using GAMLSS in R*, New York: Chapman and Hall/CRC, 1st ed.
- Strupczewski, W. G., Kochanek, K., Weglarczyk, S., and Singh, V. P. (2005), “On robustness of large quantile estimates of log-Gumbel and log-logistic distributions to largest element of the observation series: Monte Carlo results vs. first order approximation,” *Stoch Environ Res Risk Assess*, 19, 280–291.
- Taillardat, M., Mestre, O., Zamo, M., and Naveau, P. (2016), “Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics,” *American Meteorological Society*.
- Tobler, W. R. (1970), “A Computer Movie Simulating Urban Growth in the Detroit Region,” *Economic Geography*, 46, 234–240.
- Tsionas, E. (2003), “Bayesian Quantile Inference,” *Journal of Statistical Computation and Simulation*, 73, 659–674.
- Yu, K. and Moyeed, R. A. (2001), “Bayesian quantile regression,” *Statistics & Probability Letters*, 54, 437–447.