

## CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ  
ISSN on-line 2317-4536 / ISSN impresso 1413-9022 - v.50, p.22-54, 2021  
DOI: 10.12957/cadest.2021.64202

# CENÁRIOS DE VENTOS HORÁRIOS PARA O PLANEJAMENTO DA EXPANSÃO DO SISTEMA ELÉTRICO BRASILEIRO

Jorge Machado Damázio  
UERJ – Universidade do Estado do Rio de Janeiro  
CEPEL- Centro de Pesquisas de Energia Elétrica  
[damazio@ime.uerj.br](mailto:damazio@ime.uerj.br)

Maria Luiza Viana Lisboa  
UERJ – Universidade do Estado do Rio de Janeiro  
CEPEL- Centro de Pesquisas de Energia Elétrica  
[mlisboa@ime.uerj.br](mailto:mlisboa@ime.uerj.br)

### **Resumo**

*Apresenta-se uma metodologia para a definição de cenários de ventos horários em regiões de implantação de parques eólicos visando apoiar estudos de planejamento para a evolução do parque gerador do sistema elétrico brasileiro. A metodologia utiliza modelagem estocástica autorregressiva para a construção de uma amostra com elevado número (milhares) de cenários anuais visando posterior aplicação de técnicas de discretização de variáveis contínuas para seleção de um conjunto menor de cenários (dezenas) que serão explicitados nos modelos de planejamento. O artigo apresenta uma aplicação a partir de um histórico de 41 anos de registros de ventos horários em área promissora para instalação de aerogeradores no estado da Bahia.*

**Palavras-chave:** Séries Temporais de Ventos Horários, Modelos Auto-Regressivos, Discretização e Variáveis Aleatórias Contínuas.

## 1. Introdução

No parque de geração do setor elétrico brasileiro é crescente a participação da geração eólica. Sendo a participação desta fonte ainda considerada baixa, o seu tratamento nos modelos de otimização do planejamento energético pode ser feito de forma aproximada considerando valores médios mensais de geração. À medida em que cresce a participação da geração eólica, cresce também a necessidade de se levar em conta a variabilidade temporal de curto prazo (intradiária) dos ventos e, portanto, da geração da fonte eólica, de forma a evitar a superestimação da capacidade do sistema de acomodar esta variabilidade.

Os modelos de otimização utilizados em estudos estratégicos de longo prazo (horizontes de 30 anos ou mais) de sistemas energéticos de forma geral adotam implementações de programação linear (MESSNER; STRUBEGGER, 1995) ou linear mista (ANTUNES *et al.*, 2004; Loulou *et al.*, 2005) para cálculo de trajetórias do sistema ao longo do horizonte maximizando superávit ou minimizando custos (demandas inflexíveis). No Brasil, citamos o modelo MATRIZ (LISBOA *et al.*, 2012) desenvolvido para estudos integrados de sistemas energéticos e o modelo MELP (LISBOA *et al.*, 2007) desenvolvido para estudos do sistema elétrico e de seus acoplamentos com a cadeia do gás. Uma característica comum a estas implementações é a incorporação de simplificações de representação espacial, temporal e técnica (COLLINS *et al.*, 2017) tendo em vista a adequação computacional dos modelos em termos de tempo de processamento. Na representação temporal, basicamente estes modelos utilizam intervalos anuais ou plurianuais para os balanços entre oferta e demanda, podendo serem representados balanços sazonais e de patamares cronológicos ou não de curvas de carga. Com o aumento da penetração das fontes renováveis<sup>1</sup>, tem-se buscado a melhoria da representação das variações intradiárias das gerações das fontes renováveis e das cargas do sistema buscando uma aproximação mais realista da operação.

A consideração mais acurada da variabilidade intradiária da fonte eólica sem prejuízo da adequação computacional representa um importante desafio para os modelos de planejamento energético de longo prazo. Até à primeira década deste século, as aplicações dos modelos de planejamento energético de longo prazo incorporavam ao problema equações e restrições buscando refletir impactos das variabilidades de curto

---

<sup>1</sup> Este artigo se concentra na fonte eólica, mas também a variabilidade temporal de curto prazo da fonte solar deve ser representada.

prazo, tais como: limites para a participação das renováveis; requisição de quantidades fixas de backup por capacidade instalada; e/ou balanços de capacidade de ponta. Com as perspectivas de aumento da participação da fonte eólica, duas classes de abordagens passaram a ser utilizadas para melhor capturar os desafios econômicos e técnicos (COLLINS, et al., 2017): (i) integração do modelo de longo prazo com modelos de operação de curto prazo (“soft link approach”); (ii) melhoria da representação temporal no modelo de longo prazo. A metodologia de definição de cenários temporais de ventos horários descrita neste trabalho visa apoiar abordagens da segunda classe.

A metodologia apresentada se baseia no uso do modelo AR(p) (BOX *et al.*, 2008) para a construção de uma amostra com elevado número (milhares) de cenários independentes entre si de ventos horários ao longo de um ano para posterior uso de técnicas de discretização de variáveis contínuas para a seleção de um conjunto menor de cenários que serão explicitamente representados nos modelos de planejamento. A metodologia foi implementada em linguagem R e sua aplicação exposta utilizando registros de ventos horários em área promissora para instalação de aerogeradores no estado da Bahia.

## 2. Modelagem AR(p) para construção de amostra grande de cenários

A modelagem AR(p) é aplicada em duas etapas: na primeira etapa aplica-se a modelagem para o vento diário; na segunda etapa utiliza-se um modelo de desagregação dos cenários de ventos diários obtidos na primeira etapa em cenários de ventos horários.

### Modelagem AR(p) para cenários diários de ventos

Seja  $y_{t,d}$  o vento médio (m/s) num dia  $d=1, \dots, 365$  e ano  $t=1, \dots, N$  na localidade em estudo, onde  $N$  é o número de anos. A modelagem adota o esquema multiplicativo:

$$y_{t,d} = MV_d \cdot \gamma_{t,d} \quad \gamma_{t,d} \geq 0 \quad (1)$$

onde:

$MV_d$  é o valor esperado do vento médio diário no dia  $d$  na localidade em estudo

$\gamma_{t,d}$  é um ruído adimensional não-negativo para o dia  $d$  e ano  $t$  valendo:

$$E[\gamma_{t,d}] = 1 \quad (2)$$

$$\sigma[\gamma_{t,d}] = cv[\gamma_{t,d}] = cv[y_{t,d}] = CV_d \quad (3)$$

Para distribuição de probabilidades do ruído  $\gamma_{t,d}$  foi considerado o uso da distribuição de Weibull que tem sido utilizada para modelar velocidade de vento em diversos casos (CARRILO *et al*, 2014; REUK, 2020) embora (TULLER; BRETT, 1984) tenham mostrado que na prática a distribuição de Weibull garante apenas um ajuste aproximado às distribuições de frequência de velocidades de ventos observadas. Na aplicação, descrita a seguir, não foram obtidos bons ajustes com a distribuição de Weibull em alguns meses, tendo-se adotado a distribuição GEV (JENKINSON, 1955) cuja distribuição acumulada é dada por:

$$GEV(y; \theta, \tau, \delta) = \exp \left[ - \left\{ 1 + \delta \left( \frac{y - \theta}{\tau} \right) \right\}_+^{-1/\delta} \right]; \quad -\infty < \theta, \delta < \infty \text{ e } \tau > 0 \quad (4)$$

$$y \geq \theta - \left( \frac{\tau}{\delta} \right), \quad \delta > 0;$$

$$-\infty < y < \infty, \quad \delta = 0;$$

$$y \leq \theta - \left( \frac{\tau}{\delta} \right), \quad \delta < 0$$

$$\{x\}_+ = \max(0, x);$$

sendo  $\theta$ ,  $\tau$  e  $\delta$  respectivamente parâmetros de locação, escala e forma.

Para a estrutura de dependência temporal do ruído  $\gamma_{t,d}$ , adotou-se o modelo AR(p) escrito como:

$$\tilde{\gamma}_{t,d} = \mu + \sum_{j=1}^p \phi_j \tilde{\gamma}_{t,d-j} + u_d \quad d=1,365 \quad (5)$$

onde:

$\tilde{\gamma}_{t,d}$  é a transformação do ruído  $\gamma_{t,d}$  em ruído de distribuição normal padrão utilizando a transformação da probabilidade integral:

$$\tilde{\gamma}_{t,d} = \Phi^{-1}(GEV(\gamma_{t,d}; \theta_d, \tau_d, \delta_d)) \quad (6)$$

$\Phi^{-1}(\cdot)$  é a inversa da distribuição acumulada da distribuição normal padrão.

$u_d$  é um ruído de distribuição normal com média zero e variância  $\sigma_u^2$  e sem estrutura de dependência temporal

$\mu$  é uma constante

$p$  é a ordem do modelo

$\phi_j$  são os coeficientes auto regressivos.

Dado um conjunto de  $n$  anos de dados de ventos diários  $y_{t,d}$  (medidos ou calculados por reanálise), a calibração da modelagem de cenários de ventos horários é feita pelos seguintes passos:

- 1- estima-se os valores esperados de cada dia,  $\widehat{MV}_d$ ,  $d=1,365$ , por suavização das médias dos ventos diários de cada dia.
- 2- Usando-se (1) obtém-se a série histórica de ruídos  $\gamma_{t,d}$
- 3- Para cada dia estima-se os parâmetros  $(\theta_d, \tau_d, \delta_d)$  por máxima verossimilhança.
- 4- Através de (6) transforma-se a série histórica de ruídos  $\gamma_{t,d}$  em série histórica de ruídos  $\tilde{\gamma}_{t,d}$  normal padrão.
- 5- Com a série  $\tilde{\gamma}_{t,d}$  calibra-se a ordem  $p$  e os parâmetros  $\mu, \phi_j$  e  $\sigma_u^2$  usando-se a metodologia de Box&Jenkins.

#### Construção de amostra aleatória de $N$ cenários de ventos diários:

Considerando que cada cenário corresponde a um ano, a amostra é construída usando para cada ano  $t$  os seguintes passos:

- 1- Utilizar o modelo AR( $p$ ) para gerar uma sequência de ruídos normal padrão  $\tilde{\gamma}_{t,d}$ ,  $d=1,365$  com estrutura de dependência temporal.
- 2- Usando a inversa de (6), obtém-se a sequência de ruídos GEV,  $\gamma_{t,d}$ ,  $d = 1,365$ .
- 3- Usando-se (1), obtém-se a sequência de ventos diários  $y_{t,d}$ ,  $d=1,365$ .

#### Desagregação dos cenários diários em cenários horários

A modelagem de desagregação também adota o esquema multiplicativo:

$$y_{t,d,h} = 24 \cdot y_{t,d} \cdot \xi_{t,d,h} \quad t = 1, \dots, N; d = 1, 365; h = 0,23 \quad (7)$$

onde:

$y_{t,d,h}$  é vento na hora  $h$

$\xi_{t,d,h}$  é o coeficiente de desagregação adimensional do dia  $d$  e ano  $t$  para a hora  $h$ .

Com a constante 24 em (7), os coeficientes de desagregação atendem às duas restrições seguintes:

$$0 \leq \xi_{t,d,h} \leq 1 \quad (8)$$

$$\sum_{h=0}^{23} \xi_{t,d,h} = 1 \quad (9)$$

Para a modelagem dos coeficientes de desagregação em (7) considerou-se distribuições Normais calibradas pelo método dos momentos para cada mês  $m=1,12$  e para cada hora do dia. Adicionalmente, considerou-se ainda uma classificação binária dos dias dentro de cada mês conforme o vento médio diário seja maior ou que menor que um valor de fronteira para o mês, denotado por  $L_m$ :

$$X_{t,d} = 1 \text{ se } y_{t,d} > L_m \text{ (dia de vento alto no mês } m) \quad (10a)$$

$$X_{t,d} = 0 \text{ se } y_{t,d} < L_m \text{ (dia de vento baixo no mês } m) \quad (10b)$$

Desta forma em cada mês definem-se distribuições dos coeficientes de cada hora para os dias de ventos médios diários altos e para os dias de ventos médios diários baixos.

A estrutura de dependência temporal do coeficiente de desagregação foi modelada de maneira análoga ao adotado na modelagem do ruído adimensional do vento diário através do modelo AR(p) escrito como:

$$\tilde{\xi}_{t,d,h} = \alpha + \sum_{j=1}^p \phi_j \tilde{\xi}_{t,d,h-j} + v_d \quad (11)$$

onde:

$\tilde{\xi}_{t,d,h}$  é a transformação de  $\xi_{t,d,h}$  em ruído padronizado normal obtido por:

$$\tilde{\xi}_{t,d,h} = \frac{\xi_{t,d,h} - \bar{\xi}_{m,h,X_{t,d}}}{s\tilde{\xi}_{m,h,X_{t,d}}} \quad (12)$$

$v_d$  é um ruído de distribuição normal com média zero e variância  $\sigma_v^2$  e sem estrutura de dependência temporal

$\alpha$  é uma constante

$p$  é a ordem do modelo

$\phi_j$  são os coeficientes auto regressivos.

$\bar{\xi}_{m,h,X_{t,d}}$  é a média dos coeficientes de desagregação históricos do mês  $m$ , hora  $h$  e classificação binária  $X_{t,d}$

$s\tilde{\xi}_{m,h,X_{t,d}}$  é o desvio-padrão dos coeficientes de desagregação históricos do mês  $m$ , hora  $h$  e classificação binária  $X_{t,d}$ .

Dado um conjunto de  $n$  anos de dados de ventos horários (medidos ou calculados por reanálise) a calibração da modelagem de desagregação é feita pelos seguintes passos:

- 1- Usando-se (7) obtém-se a série histórica de coeficientes de desagregação  $\xi_{t,d,h}$
- 2- Adota-se como limites  $L_m$  as medianas dos ventos diários de cada mês  $m$ .
- 3- Para cada mês, para cada trinca (dia, hora, classificação  $X$ ), estima-se a média e o desvio padrão dos coeficientes de desagregação históricos.
- 4- Através de (12) transforma-se a série histórica de coeficientes de desagregação em série histórica de ruídos normal padrão  $\tilde{\xi}_{t,d,h}$ .
- 5- Com a série  $\tilde{\xi}_{t,d,h}$  calibra-se a ordem  $p$  e os parâmetros  $\alpha, \phi_j$  e  $\sigma_v^2$  usando-se a metodologia de Box&Jenkins.

#### Construção de amostra aleatória de $N$ cenários de ventos horários:

A amostra de ventos horários é construída pela desagregação da amostra de ventos diários. Supondo que cada cenário corresponde a um ano  $t$ , a desagregação de cada cenário é feita pelos seguintes passos:

- 1- Utilizar o modelo  $AR(p)$  para gerar uma sequência de ruídos normal padrão  $\tilde{\xi}_{t,d,h}$ ,  $d=1,365$ ;  $h=0,23$  com estrutura de dependência temporal.
- 2- Usando a inversa de (12), obtém-se a sequência de coeficientes  $\xi_{t,d,h}$ ,  $d = 1,365$ ;  $h=0,23$ .
- 3- Em cada dia: (i) calcule a soma dos coeficientes; (ii) dividir os coeficientes pelo valor obtido no item (i).
- 4- Usando-se (7), obtém-se a sequência de ventos horários  $y_{t,d,h}$ ,  $d = 1, 365$ ;  $h=0,23$ .

### **3. Seleção de Cenários**

A partir da aplicação da modelagem  $AR(p)$  descrita no item anterior é possível construir amostras contendo o número que se queira de cenários de ventos na escala horária durante um ano. Como exemplo, vamos tomar como premissa a necessidade de amostras capazes de estimar com acurácia probabilidades  $\pi$  de ocorrência anual de eventos desfavoráveis, podendo  $\pi$  se situar no entorno de 1% a 10%. Seja  $N$  o número de cenários anuais na amostra e  $\hat{\pi}$  a estimativa para  $\pi$  obtida por  $n/N$ , onde  $n$  é o número de cenários com ocorrência do evento desfavorável. O valor esperado de  $\hat{\pi}$  é igual a  $\pi$  e a variância é igual a  $\pi(1-\pi)/N$ . Considerando a distribuição normal como aproximação para a distribuição amostral de  $\hat{\pi}$ , o intervalo de confiança a nível de significância 5%

para  $\pi$  é centrado em  $\hat{\pi}$  com largura igual ao dobro de  $1.96(\pi(1-\pi)/N)^{1/2}$ . O número de cenários da amostra necessário para fazer a largura do intervalo de confiança igual a  $\beta\pi$  é dado por:

$$N_{nec} = \frac{3.84(1-\pi)}{\pi\beta^2} \quad (13)$$

Considere as variações de  $N_{nec}$  em (13) com  $\beta$  para  $\pi=1\%$ ,  $3\%$ ,  $5\%$  e  $10\%$ . Para  $\pi=5\%$ , os valores de  $N_{nec}$  variam entre 811 e 7300 conforme a largura requerida para o intervalo de confiança. Considerando ainda  $\pi=5\%$  e adotando  $\beta=0.2$ , correspondente à uma largura de intervalo de confiança de  $1\%$ , encontramos por (13),  $N_{nec} = 1825$ . Com uma amostra deste tamanho, uma alternativa de expansão será considerada como atendendo de forma adequada o critério de confiabilidade de  $5\%$  quando nas simulações a frequência de cenários com ocorrência de eventos desfavoráveis não mitigados estiver entre  $4.5\%$  e  $5.5\%$ . Caso esta frequência nas simulações suplante  $5.5\%$ , alguns investimentos deverão ser antecipados ou novos investimentos precisarão ser incorporados à alternativa de expansão; caso a frequência não alcance  $4.5\%$ , alguns investimentos poderão ser postergados. Para efeito de comparação, considerando ainda  $\pi=5\%$ , mas adotando  $\beta=0.1$ , correspondente a um intervalo de confiança com largura de  $0.5\%$ , (13) fornece  $N_{nec}=7300$ . Neste caso, as alternativas são consideradas adequadas com frequências de cenários anuais não mitigados nas simulações entre  $4.75\%$  e  $5.25\%$ .

O exercício acima pode ser tomado como indicador da necessidade de uma amostra com pelo menos 2000 cenários para a análise da capacidade do sistema elétrico de acomodar a variabilidade de curto prazo da fonte eólica. Com efeito, nos estudos de simulação da operação do sistema elétrico Brasileiro para avaliações de garantia de suprimento utilizam rotineiramente 2000 cenários hidrológicas (GT MODERNIZAÇÃO DO SETOR ELÉTRICO, 2019; EPE, 2018; LISBOA *et al.*, 2007). Por outro lado, o uso explícito de 2000 cenários em modelos de otimização da trajetória de longo prazo da configuração do sistema compromete a convergência dos algoritmos num tempo de execução prático. A busca da conciliação entre os dois objetivos demanda a aplicação de alternativas metodológicas.

Diversos algoritmos alternativos de amostragem têm sido aplicados visando aliviar as limitações práticas do método de Monte Carlo simples na análise de eventos raros. Stedinger (1997) analisa o uso as técnicas de amostragem por importância,



amostragem por hipercubo latino e amostragem estratificada em análises de segurança de barragens. (GIL *et al.*, 2015) analisa os resultados de um modelo de expansão de sistema hidrotérmico obtidos com subconjunto de cenários anuais hidrológicos selecionado do registro histórico total através de algoritmo de maximização das correspondências entre os momentos do histórico e do subconjunto selecionado. Penna (2009) compara o uso da amostragem por hipercubo latino com a aplicação de técnicas de agrupamento (MACQUEEN, 1967) na construção de cenários de afluências hidrológicas em modelos de otimização da operação de sistemas hidrotérmicos de potência. (OLIVEIRA *et al.*, 2010) apresentam técnicas para redução da cardinalidade de árvores de cenários de afluências hidrológicas mensais obtidas por Monte Carlo visando também aplicações em modelos de planejamento da operação de sistemas hidrotérmicos.

### Técnicas de Discretização de Variáveis Contínuas

A abordagem adotada neste trabalho utiliza técnicas de discretização de variáveis aleatórias contínuas (CHAKRABORTY, 2015; BARBIERO; HITAJ, 2021; DREZNER; ZEROM, 2015). Estas técnicas de forma geral são algoritmos para divisão dos domínios das variáveis contínuas num número finito de intervalos. Denotando por  $W$  a variável contínua, e seu suporte como o intervalo  $S = [\inf, \sup] \subseteq \mathbb{R}$ , o algoritmo de discretização de  $W$  fornece uma partição de  $S$  formada por um conjunto de  $k$  intervalos  $P = \{ (h_0, h_1), (h_1, h_2), \dots, (h_{k-1}, h_k) \}$ ,  $\inf = h_0 < h_1 < h_2 < \dots < h_{k-1} < h_k = \sup$ . Para cada intervalo  $i = 1, \dots, k$ , a discretização fornece um valor representativo,  $w_{di}$ ,  $h_{i-1} \leq w_{di} \leq h_i$ , e uma probabilidade  $p_i$ . O resultado define uma variável aleatória discreta  $W_d$ , com suporte dado pelo conjuntos  $w_{di}$ ,  $i=1, \dots, k$  e massas de probabilidades associadas  $p_i$ ,  $i=1, \dots, k$ . A distribuição da variável aleatória  $W_d$  deve se aproximar da distribuição da variável aleatória contínua  $W$  de acordo com algum critério, existindo diversas abordagens conforme as características priorizadas da distribuição contínua (HAMMOND; BICKEL, 2015). Algumas abordagens procuram correspondências entre momentos ou quantis das distribuições  $W$  e  $W_d$ . Outras abordagens se voltam para a minimização de “distâncias” entre  $W$  e  $W_d$  (BARBIERO; HITAJ, 2021).

Dentro da abordagem da preservação de momentos, uma aproximação discreta com  $k$  pontos fornece  $2k-1$  parâmetros ( $x_1$  a  $x_k$  e  $p_1$  a  $p_{k-1}$ ) que podem ser calibrados de forma a preservar  $2k-1$  momentos resolvendo um sistema de  $2k$  equações com a primeira

equação representado a soma de  $p_i = 1$  (GOLUB; WELSCH, 1969). Drezner e Zerom (2015) propõe um algoritmo considerando uma primeira aproximação  $Wd^{(1)}$  onde os representantes  $wd_i^{(1)}$  são obtidos pelo valor esperado condicionado de cada intervalo e as probabilidades  $p_i^{(i)}$  obtidas integrando a distribuição contínua. Com esta primeira aproximação, a preservação da média é garantida, enquanto a variância é sub-estimada. Os autores propuseram um algoritmos para cálculo de uma segunda aproximação  $Wd^{(2)}$ , próxima de  $Wd^{(1)}$ , que preserva a média e a variância de  $W$  através da solução numérica de um problema de otimização não-linear cujas variáveis de decisão são os intervalos e seus representantes, mantendo as probabilidades obtidas por integração da distribuição contínua.

A abordagem de preservação de quantis proposta em (ROY; DASGUPT, 2001) parte de uma distribuição discreta  $X_d$  com suporte de inteiros usando  $Pr\{X_d = i\} = F_X(i+\delta) - F_X(i - [i - \delta])$ , recomendando-se escolher  $\delta$  de forma a acertar média e desvio padrão. Observou-se  $\delta = 0.5$  para distribuições simétricas. Barbiero (2012) generalizou o método utilizando um intervalo padrão  $(-g, +g)$ , com  $g$  positivo, fixado a priori. Este intervalo é dividido em  $k-1$  intervalos igualmente espaçados cujos limites são dados por  $x_i = -g + \frac{2g(i-1)}{k-1}$   $i = 1, \dots, k$ . Os  $k$  pontos da aproximação discreta de  $W$  são obtidas usando a inversão da normal padrão nos pontos  $x_i$ ,  $u_i = \Phi(x_i)$ . Os limites dos intervalos no suporte de  $W$  são obtidos aplicando a inversa da distribuição de  $W$  como  $h_i = F_W^{-1}(u_i)$ . Os representantes obtidos pela média dos limites de cada intervalo,  $\frac{h_{i-1} + h_i}{2}$ , e as probabilidades pelas diferenças de probabilidades acumuladas nos limites dos intervalos.

Como técnica de discretização, neste trabalho utilizou-se uma abordagem de preservação de quantis bastante flexível, onde o usuário escolhe uma partição do intervalo de probabilidades acumuladas  $(0,1)$ . Desta forma o usuário pode optar por fazer um maior detalhamento de uma das caudas, deixando a representação da outra cauda de uma forma menos precisa. Os limites dos intervalos no suporte de  $W$  são calculados pela inversa da distribuição de  $W$  nos limites da partição do intervalo de probabilidades definida pelo usuário. Os representantes são definidos pelos valores esperados condicionados de cada intervalo e as probabilidades obtidas pelas diferenças das probabilidades acumuladas. Observa-se que a média de  $W$  será preservada em  $Wd$  e a variância sub-estimada.

A estratégia de modelagem aqui projetada consiste nos seguintes passos:

- 1- Construção de uma amostra original de ventos na escala horária com cerca de 2000 cenários anuais equiprováveis.
- 2- Definição de atributo de cenário considerando quantil de uma métrica da flutuação diária do vento na escala horária.
- 3- Discretização da distribuição de densidade de probabilidade contínua empírica do atributo correspondente à amostra original.
- 4- Para cada ponto da distribuição discretizada, selecionar entre os cenários da amostra original, o cenário cujo valor de atributo mais se aproxima do ponto em questão.
- 5- Para cada cenário selecionado, apresentar o perfil horário do vento do dia cuja métrica de flutuação diária corresponde ao atributo do cenário.

A seguir apresenta-se uma aplicação da metodologia de geração de amostra original com 2000 cenários.

#### 4. Aplicação

Para a aplicação utilizou-se um histórico de 41 anos (1979-2018) de ventos horários para a altura de 100m de um ponto na região das Serras Azul e do Açuruá, uma das áreas promissoras para instalação de aerogeradores no estado da Bahia (ESTADO DA BAHIA, 2018). Os dados são oriundos da base de dados de reanálises ERA5 (DEE, 2020) e sofreram ajustes estatísticos por uma equação de primeiro grau para compatibilizar com medições realizadas de 2012 a 2014 numa torre localizada na altitude de 970m.

##### Ajuste Estatístico

O ajuste estatístico aplicado aos ventos da base ERA5 foi feito utilizando uma equação de 1<sup>o</sup> grau para cada dia:

$$V_{ajuste} = \alpha_d * V_{ERA5} + \beta_d \quad d=1,365 \quad (14)$$

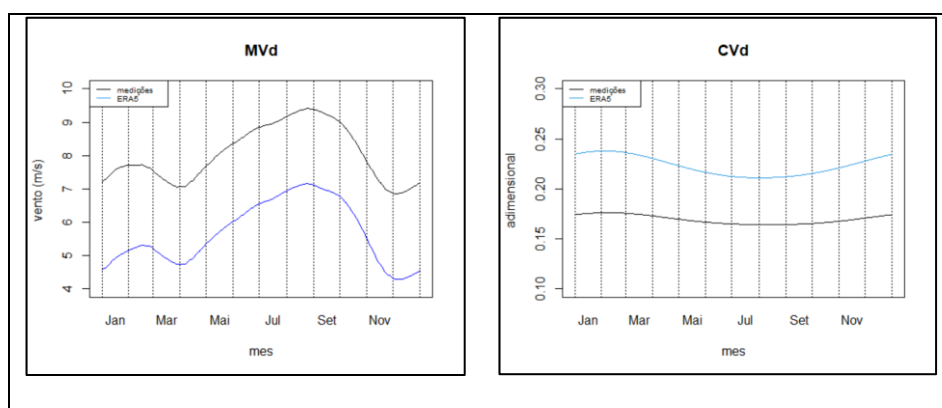
Os coeficientes  $\alpha_d$  e  $\beta_d$  foram calculados por:

$$\alpha_d = \frac{CV_d(ERA5)*MV_d(ERA5)}{CV_d*MV_d} \quad (15)$$

$$\beta_d = MV_d - \alpha_d * MV_d(ERA5) \quad (16)$$

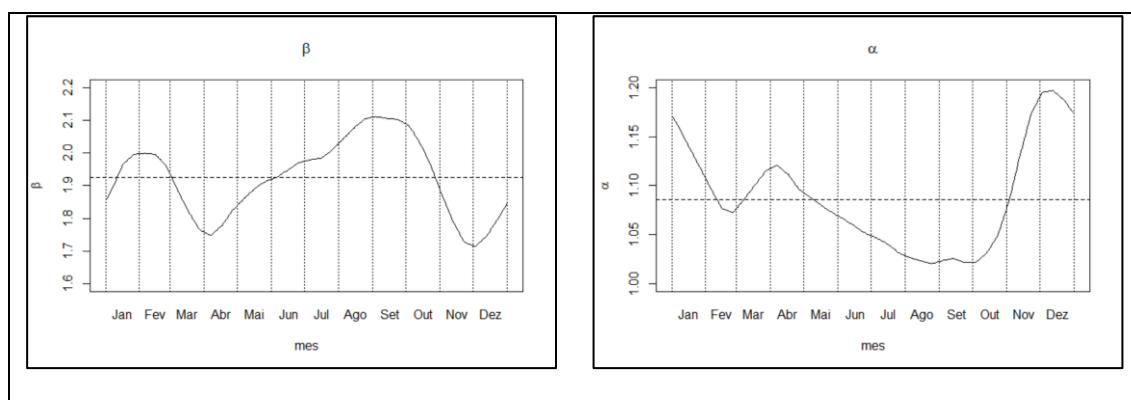
onde  $CV_d$  e  $MV_d$  são valores suavizados de coeficientes de variação e médias dos ventos diários a 100m para cada dia  $d$  das medições e  $CV_d(ERA5)$  e  $MV_d(ERA5)$  são os dados equivalentes da base ERA5 no período 2012-2014. A figura 1 apresenta as comparações  $MV_d \times MV_d(ERA5)$  e  $CV_d \times CV_d(ERA5)$ .

Figura 1.  $MV_d$  e  $CV_d$  para medições e dados ERA5



A figura 2 ilustra os coeficientes da equação de ajuste obtidos por (15) e (16). O coeficiente de viés ( $\beta$ ) flutuou ao longo do ano com coeficiente de variação de 6% em torno de 1.93 m/s. O coeficiente de ajuste da variabilidade ( $\alpha$ ) flutuou ao longo do ano com coeficiente de variação de 5% em torno de 1.09. Vale observar que o uso destes coeficientes em (14) garante que as evoluções de média e coeficientes de variação suavizadas dos ventos médios diários ERA5 ajustados no período 2012-2014 se sobreponham às correspondentes curvas das medições na torre.

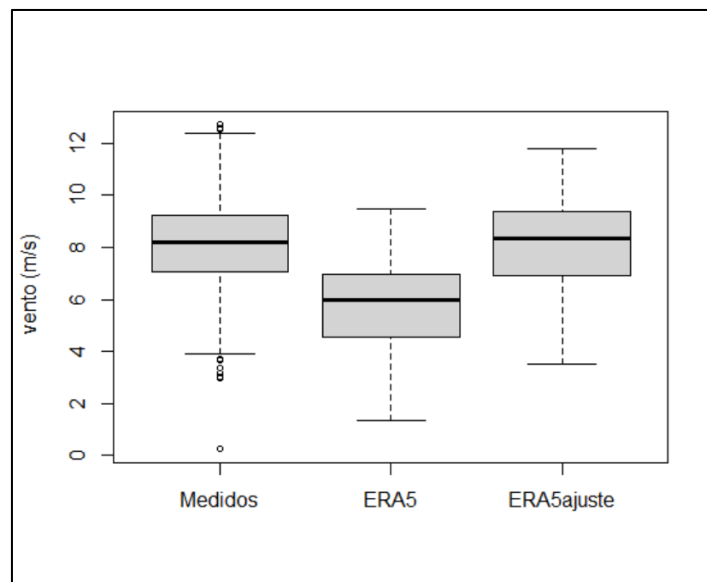
Figura 2. Coeficientes da equação do ajuste estatístico dos dados ERA5



A figura 3 ilustra o resultado do ajuste em (14) comparando os box-plots dos ventos médios diários de todos os dias para o período 2012-2014 com os correspondentes

box-plots dos dados da base ERA5 originais e ajustados, onde se ressalta a importância do ajuste de viés.

Figura 3. Efeito do Ajuste Estatístico nos Ventos Médios Diários ERA5. 2012-2014



### Modelagem de Ventos Diários

A equação (14), com os coeficientes obtidos por (15) e (16), foi aplicada à totalidade dos ventos diários da base ERA5 de 1979-2018 resultando numa série de ventos diários ajustados.

A série de ventos diários ajustados foi utilizada para estimar valores de  $\widehat{MV}_d, d = 1,365$  (por médias móveis com janela calibrada até se alcançar uma boa suavização, tendo-se optado por uma janela de 24 dias) e calcular a série de ruídos  $\gamma_{t,d}$  usando-se (1). A partir desta série de ruídos, os parâmetros das distribuições GEV para cada dia  $d=1,365$  foram estimados por máxima verossimilhança usando todos os dados dentro de janelas de 24 dias no entorno do dia sendo estimado. Os parâmetros de máxima verossimilhança foram obtidos usando o pacote R extRemes (GILLELAND; KATZ, 2016). Os valores esperados das distribuições ajustadas variaram dentro do intervalo de 2% no entorno de 1.000 (0.9824 a 1.0163) e os coeficientes de variação reproduziram essencialmente os coeficientes de variação dos dados dentro das janelas suavização (figura 4). As bondades de ajuste medidas pela estatística de Kolmogorov-Smirnov variaram de 1.8% a 7.6% (figura 4).

Os parâmetros de forma  $\delta_d$  flutuaram no entorno de -0.3569 (-0.4631 a -0.2431, indicando distribuições para ventos médio diário limitadas superiormente. A figura 5 compara os limites superiores para os ventos das GEV's obtidas com os valores de  $\widehat{MV}_d$  e os ventos diários ajustados do ano de 1979.

Figura 4. Ajuste de Distribuição GEV. CV e K-S.

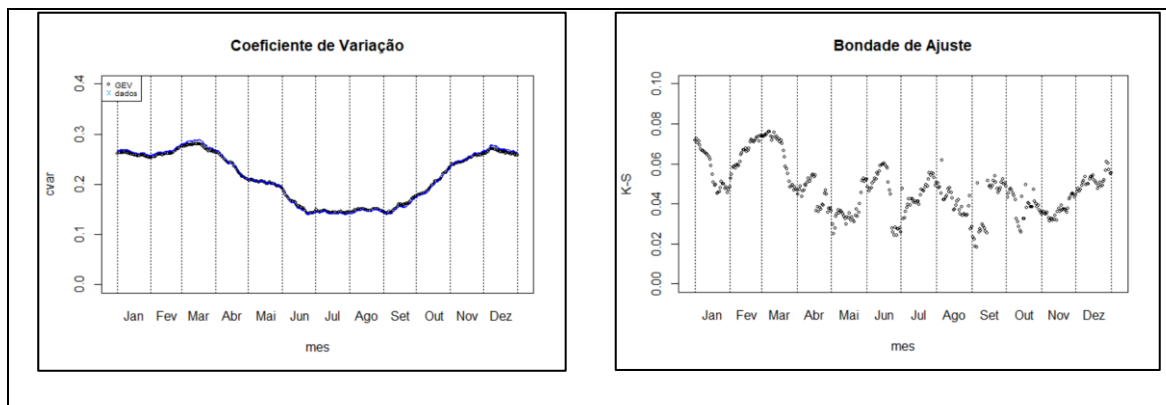
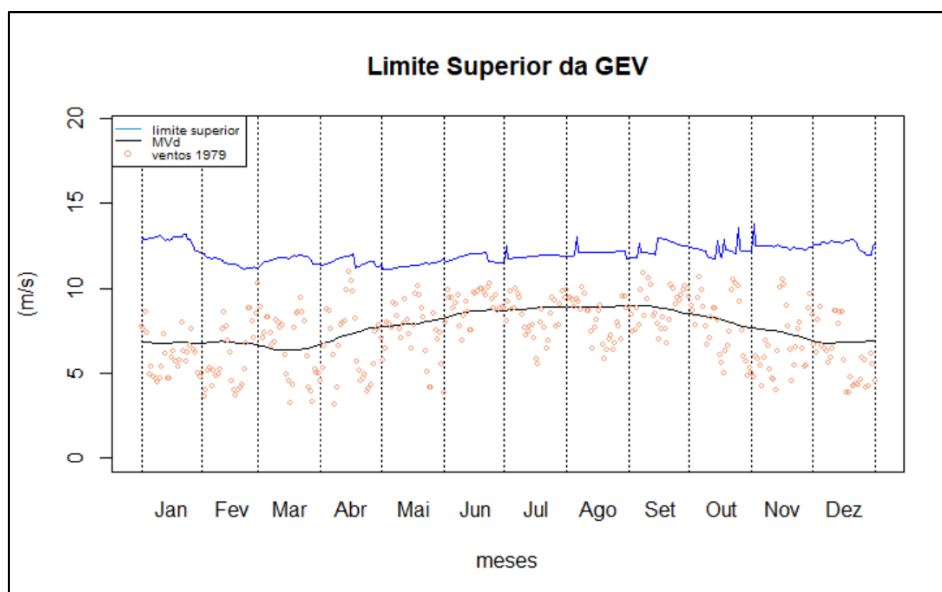


Figura 5. Ajuste de Distribuição GEV. Limites Superiores.



Através de (6) transformou-se a série de ruídos  $\gamma_{t,d}$  em série de ruídos  $\tilde{\gamma}_{t,d}$  normal padrão que, utilizada no pacote R FitAR (MCLEOD; ZHANG, 2008) resultou na modelagem AR(p) com ordem  $p=3$  e as estimativas dos parâmetros do Quadro 1.

Quadro 1. Estimativas do modelo AR(3) para a série histórica de 41 anos de ruídos  $\tilde{Y}_{t,d}$ 

\$Fi
phi(1)   phi(2)   phi(3)
0.9693781 -0.3962428 0.1284025
\$MU
[1] 2.138643e-17
\$VaR
[1] 0.4220295

### Amostragem Experimental

Visando a verificação da formulação matemática e da sua implementação gerou-se uma amostra experimental de 100 cenários de um ano.

O Quadro 2 compara sumários estatísticos das séries gerada e histórica, podendo-se verificar a proximidade dos valores à exceção da estatística Mínimo. O histórico é formado pelos ventos diários ajustados pela equação (14) e, portanto, está limitado inferiormente pelos coeficientes de viés ( $\beta$ ) da figura 3.

Quadro 2. Sumários Estatísticos dos ventos diários históricos e gerados

summary(gerado)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	6.519	7.874	7.742	9.117	12.346
summary (historico)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.986	6.425	8.043	7.736	9.170	12.43

A figura 6 compara as curvas de médias e coeficientes de variação suavizadas pela janela de 24 dias dos dados ERA5 ajustados com médias e coeficientes de variação de cada dia da amostra experimental. A figura 7 compara box-plots e distribuições acumuladas e compara correlogramas de ventos diários padronizados<sup>2</sup> dos dois conjuntos de dados. A média e o coeficiente de variação dos dados na amostra experimental (7.742 m/s e 0.239) coincidiram na a segunda casa decimal com média e coeficiente de variação dos dados ERA5 ajustados (7.736 m/s e 0.240). As frequências de velocidades de ventos menores que 4,0 m/s das duas amostras foram muito próximas (3.09% na amostra experimental e 2.95% nos dados ERA5 ajustados). Em resumo, o modelo reproduz de forma satisfatória as características da amostra histórica analisadas. A única discrepância encontrada entre a amostra gerada e a amostra histórica foi o conjunto de ventos gerados

<sup>2</sup> Vento diário menos média do dia dividido pelo desvio-padrão do dia.

menores que 3 m/s que não aparecem na amostra histórica, ressaltado na figura 7 pelo conjunto de outliers do box-plot da amostra gerada. Conforme já comentado na análise da estatística de Mínimo, a não ocorrência no histórico de ventos abaixo de 2.986 m/s em muito está relacionada com a metodologia adotada para o ajuste estatístico dos dados. Deve-se ressaltar que o maior interesse energético está nas faixas de ventos maiores que 4 m/s, para a qual a modelagem reproduziu o histórico com muito boa precisão.

Figura 6. Cenários Ventos Diários. Verificação Estatística. Médias e CVs.

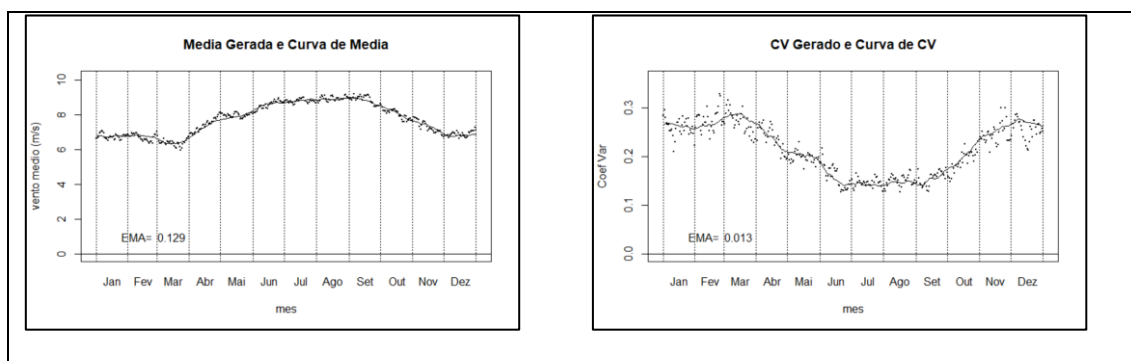
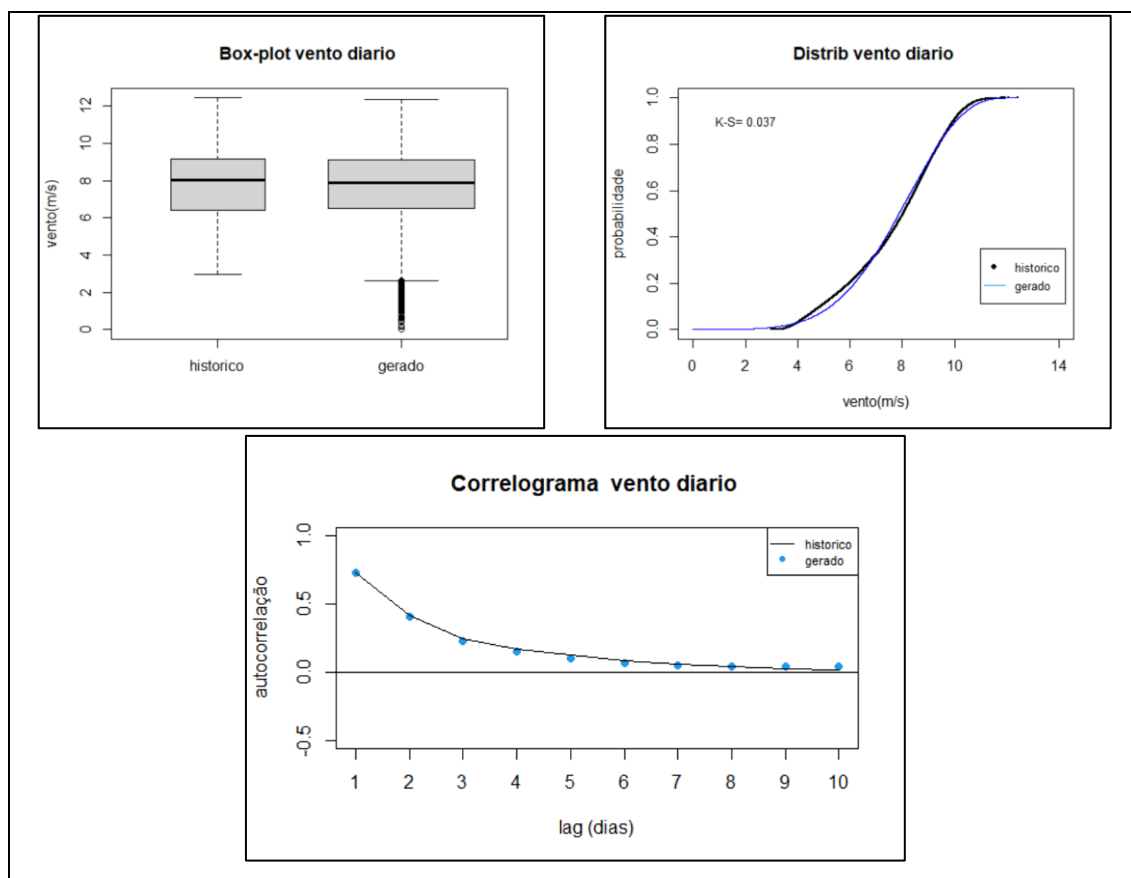


Figura 7. Cenários Ventos Diários. Verificação Estatística. Box-Plots e Distribuições e Correlograma de ventos diários padronizados.





### Modelagem da Desagregação Horária de Ventos Diários

A equação (7) foi aplicada aos ventos horários dos 41 anos da base ERA5 ajustada para a obtenção dos coeficientes de desagregação históricos  $\xi_{t,d,h}$ .

As medianas dos ventos médios diários do mês definiram os limites  $L_m$  para a classificação binária  $X$  em (10a) e (10b). A figura 8 ilustra perfis horários de médias e percentis dos coeficientes de desagregação horária nas duas classificações para janeiro, abril, julho e outubro.

Distribuições GEV e NORMAL foram calibradas para as duas classificações em cada mês para cada hora do dia por máxima verossimilhança (para a calibração da GEV utilizou-se o pacote R *extRemes* (GILLELAND; KATZ, 2016)). As calibrações da NORMAL foram de forma geral um pouco melhores. A figura 9 ilustra médias e piores estatísticas K-S obtidas para cada mês das duas distribuições, podendo-se notar a superioridade da distribuição NORMAL.

Através da transformação em (12) obteve-se a série histórica de ruídos normal padrão  $\tilde{\xi}_{t,d,h}$  com a qual foi calibrado o modelo autorregressivo utilizando também o pacote R *FitAR* e resultando na modelagem AR(p) com ordem  $p=44$  e as estimativas dos parâmetros do Quadro 3.

### Amostragem Experimental

A construção da amostra experimental seguiu os passos de construção de amostra de ventos horários.

Passo 1: Gerou-se série com  $41 \times 365 \times 24$  ruídos normais padrões utilizando o modelo AR(44) ajustado.

Passo 2: A série foi padronizada foi transformada em série de coeficientes de desagregação pela inversa de (12).

Passo 3: Cada sequência de 24 horas passou pelo processo de correção para somar 1, obtendo-se 41 cenários de coeficientes de desagregação.

Passo 4: Os coeficientes de desagregação de cada dia dos 41 cenários gerados foram utilizados para desagregar a série histórica de ventos diários da base ERA5 ajustada

com base na equação (7), obtendo-se um conjunto de 41 cenários gerados anuais de ventos horários.

A figura 10 compara os perfis horários de médias e percentis dos coeficientes de desagregação históricos e gerados, podendo-se verificar a boa representação da modelagem sob este aspecto.

O Quadro 4 apresenta sumários estatísticos dos dois conjuntos de ventos horários (históricos e gerados), observando-se valores muito próximos para todas as estatísticas à exceção do máximo e mínimo.

A média dos ventos horários gerados e históricos foram coincidentes na sexta casa decimal: 7.734174 m/s. Os desvios-padrão foram bem próximos: 2.182 (gerado) e 2.171 (histórico). A figura 11 compara as distribuições e os box-plots dos ventos horários históricos e gerados, onde pode-se observar o maior alcance na cauda da direita da amostra gerada.

Definiu-se como “hora OFF” as horas com vento médio horário menor que 4.00 m/s ou maior que 11 m/s. Duas variáveis relacionadas a “hora OFF” foram analisadas: Percentual Anual de horas OFF e Duração de sequências de horas OFF. A tabela 1 compara médias, e variâncias das variáveis nas duas amostras. As figuras 13 e 14 comparam as distribuições de frequências e box-plots das duas amostras para as duas variáveis.

Pode-se concluir que a modelagem reproduz com boa precisão a distribuição anual do número de horas OFF do histórico. Em termos de duração de sequências de horas OFF, a reprodução da distribuição do histórico também pode ser considerada boa, mesmo observando-se uma diferença estatisticamente significativa de -17 minutos entre as médias das durações das sequências de horas OFF geradas e históricas.

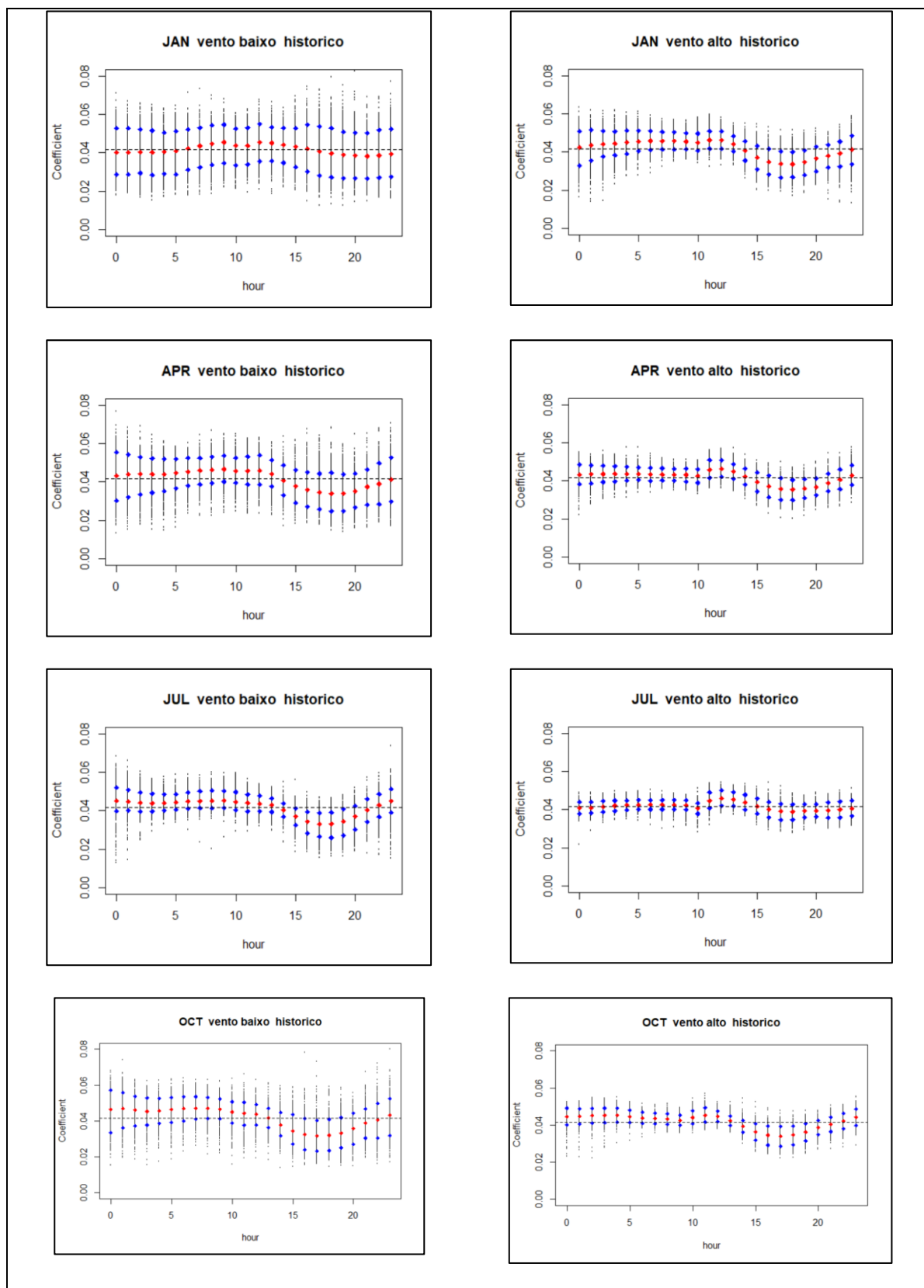
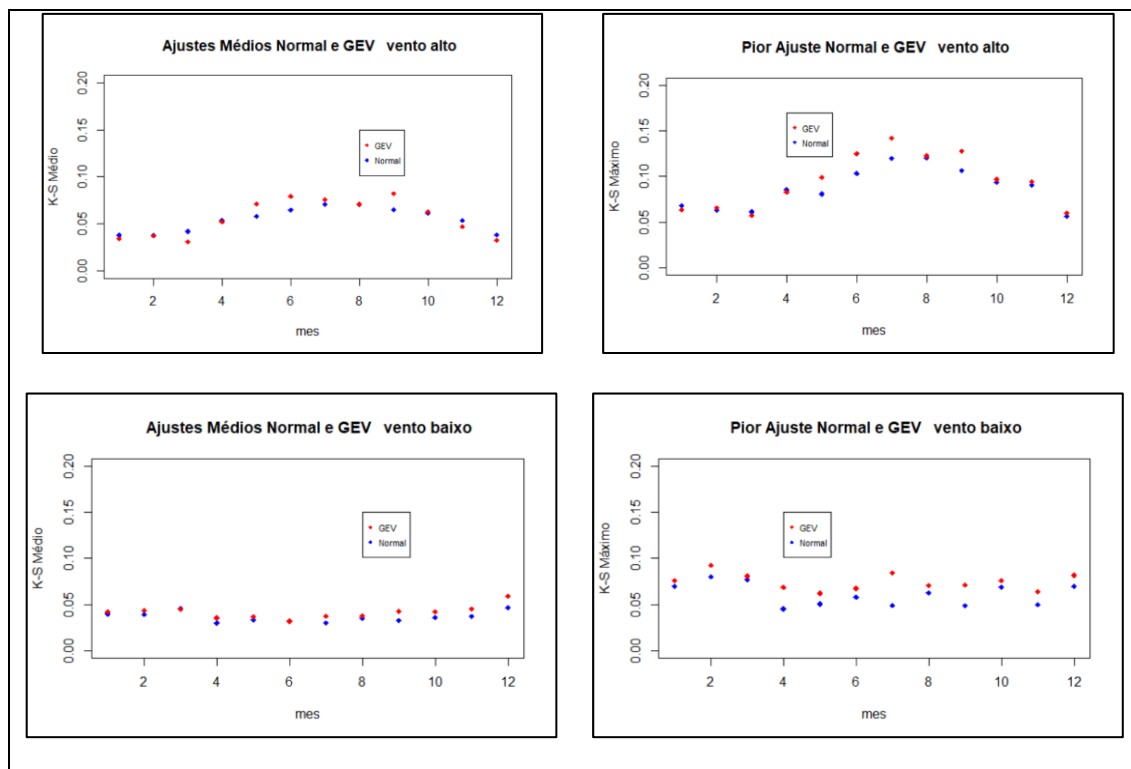
Figura 8: Perfis Históricos horários de Médias e Percentis 10% e 90% de  $\xi_{t,d,h}$ 

Figura 9: Comparação dos Ajustes de GEV e Normal para Coeficientes de Desagregação

Quadro 3. Estimativas do modelo AR(44) para a série histórica de 41 anos de ruídos  $\tilde{\xi}_{t,d,h}$ 

\$Fi

phi(1)	phi(2)	phi(3)	phi(4)	phi(5)
1.070881e+00	-3.335009e-01	9.710084e-03	2.857832e-03	-2.025071e-02
phi(6)	phi(7)	phi(8)	phi(9)	phi(10)
-1.596278e-02	-2.602614e-02	-2.017796e-02	-1.972604e-02	-1.852659e-02
phi(11)	phi(12)	phi(13)	phi(14)	phi(15)
-1.890530e-02	-1.083476e-02	-7.197151e-03	-4.234005e-02	-1.157642e-02
phi(16)	phi(17)	phi(18)	phi(19)	phi(20)
-7.649651e-03	-2.212008e-02	-1.461536e-02	-6.413875e-03	-8.381878e-03
phi(21)	phi(22)	phi(23)	phi(24)	phi(25)
3.938463e-03	-3.784037e-03	1.294617e-04	6.439774e-02	-8.090071e-02
phi(26)	phi(27)	phi(28)	phi(29)	phi(30)
1.176802e-02	-1.837551e-02	-7.179043e-03	-5.807427e-03	-7.184153e-03
phi(31)	phi(32)	phi(33)	phi(34)	phi(35)
-5.424128e-03	-4.581465e-03	-1.907704e-03	-6.907310e-03	-1.568383e-02
phi(36)	phi(37)	phi(38)	phi(39)	phi(40)
1.298795e-02	-3.053788e-03	-1.417403e-02	3.486773e-03	-6.934276e-03
phi(41)	phi(42)	phi(43)	phi(44)	
-4.222620e-03	-6.319159e-05	3.874647e-04	-8.115067e-03	

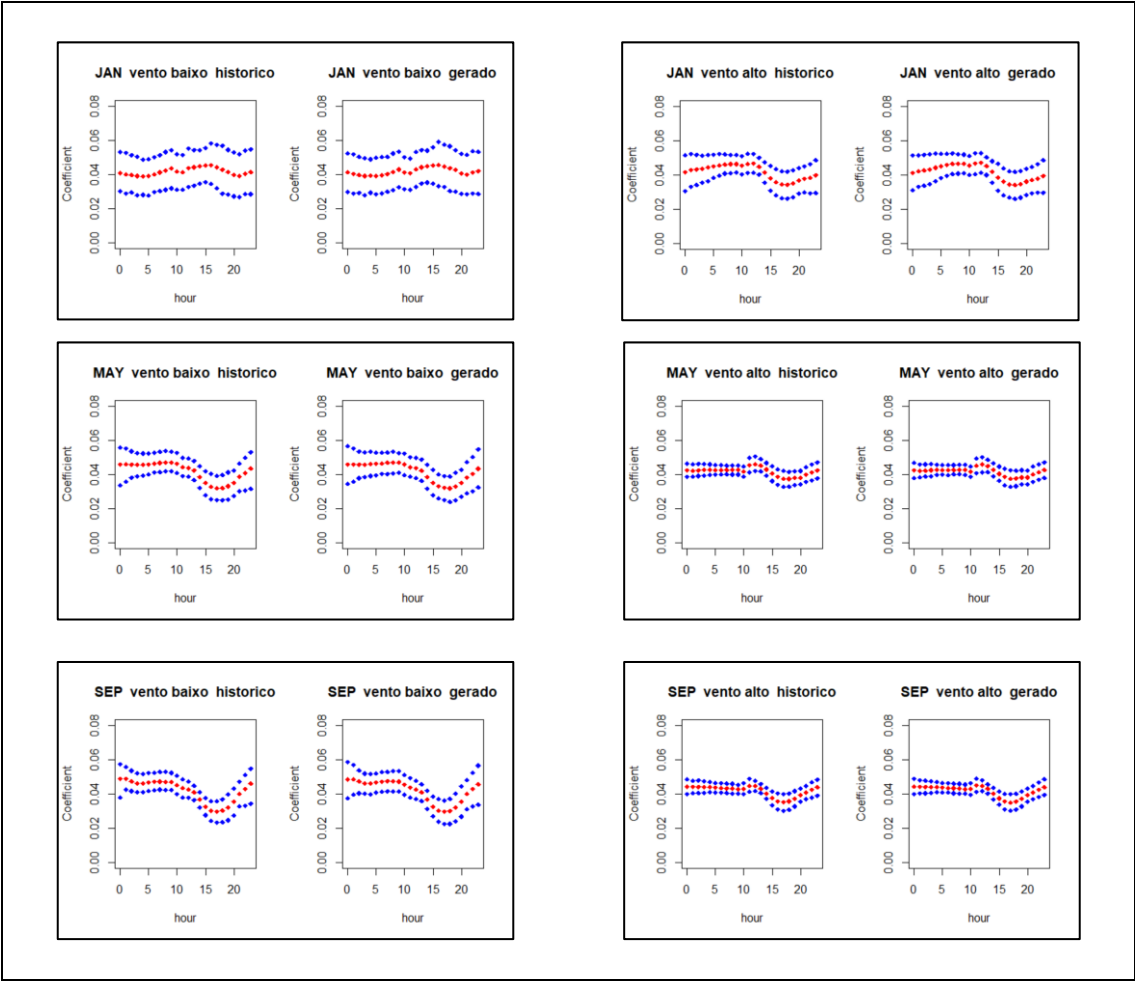
\$MU

[1] -1.18984e-17

\$VaR

[1] 0.2422092

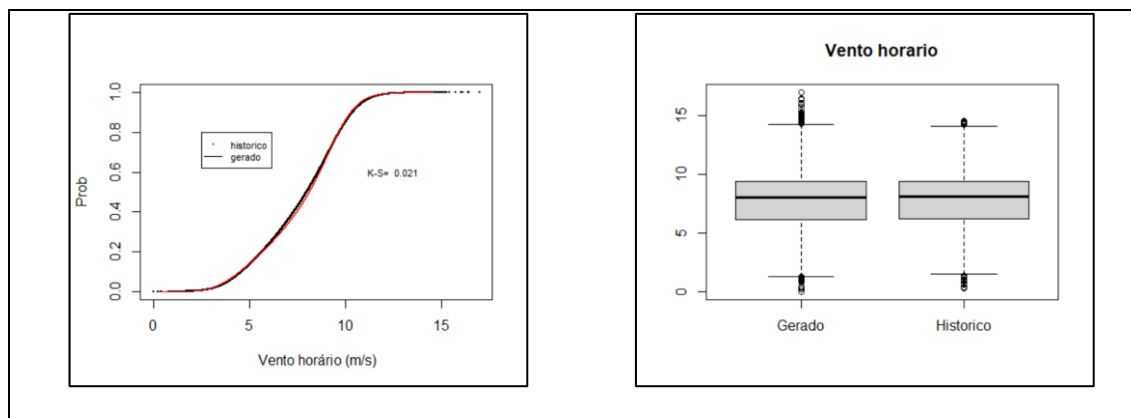
Figura 10: Perfis horários Históricos e Gerados de Médias e Percentis 10% e 90% de  $\xi_{t,d,h}$



Quadro 4. Sumários Estatísticos dos ventos horários históricos e gerados amostra experimental

summary(gerado)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	6.136	7.994	7.734	9.405	16.982
summary(historico)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.270	6.185	8.116	7.734	9.385	14.590

Figura 11. Distribuições e Box-Plots de Ventos Horários. Histórico e Gerado



A figura 12 compara os correlogramas das séries de ventos horários padronizados histórico e gerado.

Figura 12. Correlogramas de vento horário

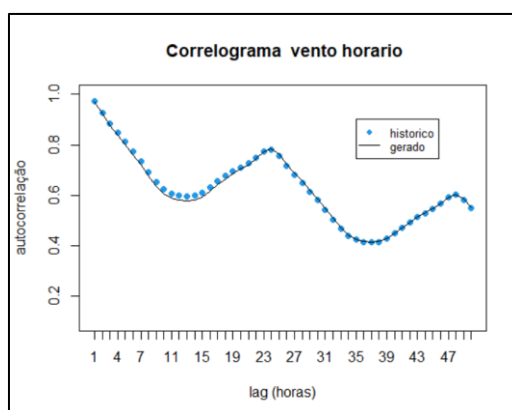


Tabela 1 – Estatísticas de Horas OFF de Ventos Horários. Amostras Gerada Experimental e Histórica

Variável	Média			Variância		
	ger	hist	Teste t Sig	ger	hist	Teste F Sig
% anual de horas OFF	10.04	9.88	0.7	2.877	3.893	0.34
Duração de Sequências de horas OFF	3.54	3.82	1.5e-08	11.09	11.15	0.79

Figura 13. Distribuições e Box-plots. % anual de horas OFF. Amostra Gerada Experimental e Histórica

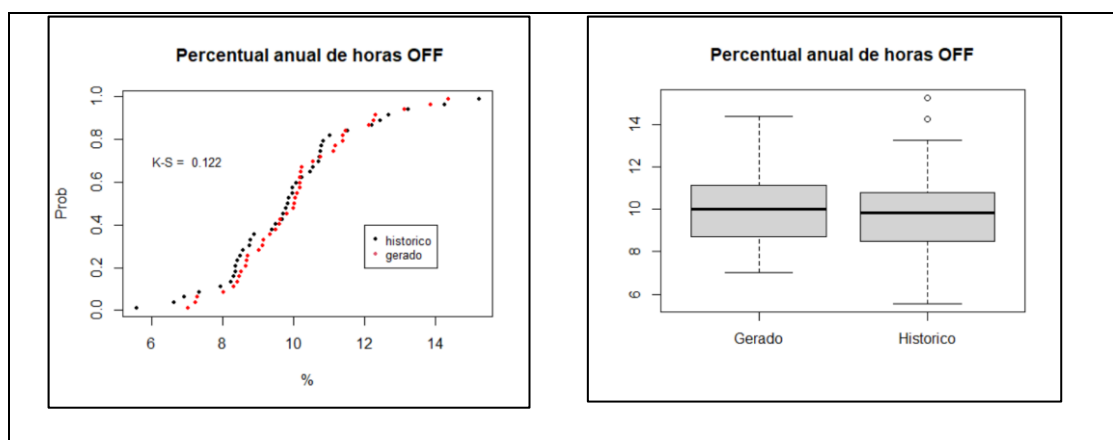
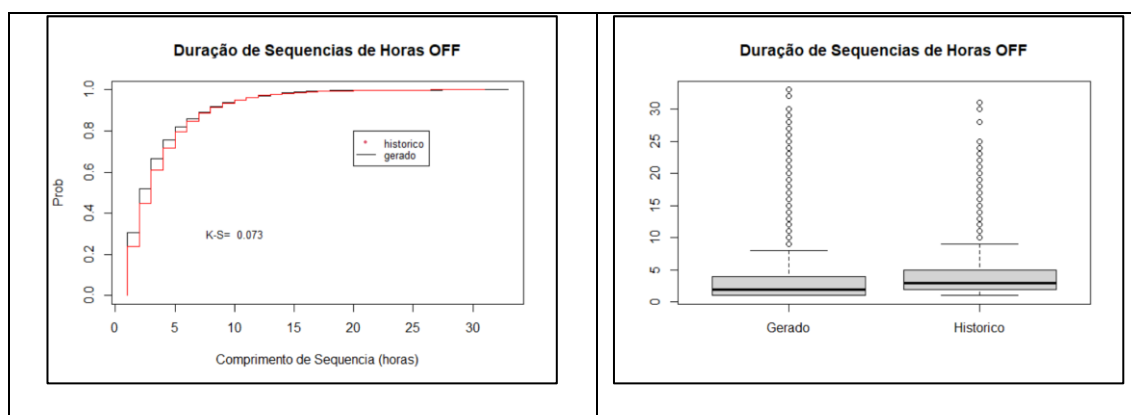


Figura 14. Box-plots e Distribuições. Duração de sequencias de horas OFF das duas amostras. Amostra Gerada Experimental e Histórica



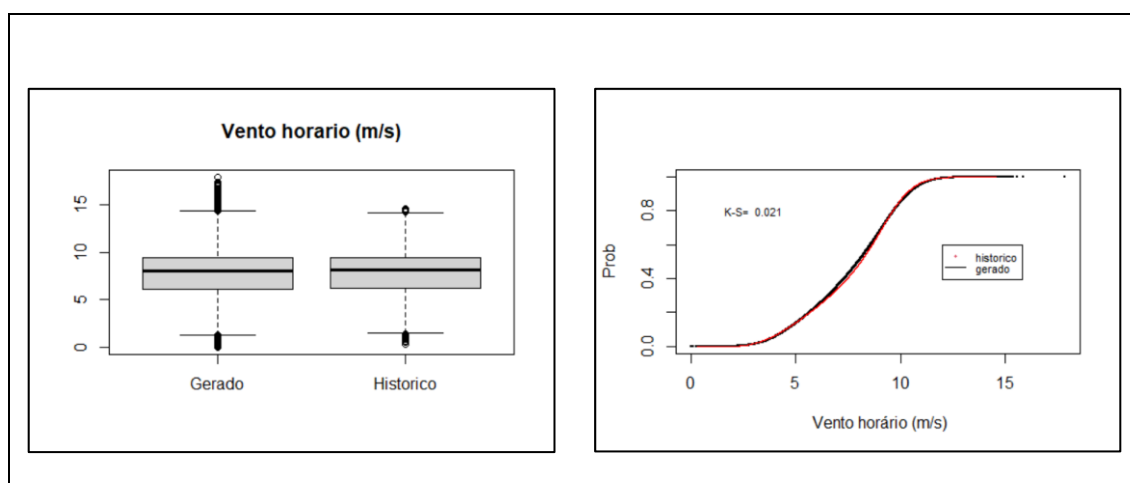
### Construção da amostra original

O procedimento de construção da amostra experimental de coeficientes de desagregação apresentado acima foi repetido 47 outras vezes, totalizando uma amostra de 1968 (48 x 41) cenários de ventos horários, todos eles obtidos por desagregação de cenários de ventos diários obtidos pela metodologia descrita anteriormente. O Quadro 5 apresenta sumários estatísticos dos dois conjuntos de ventos horários (históricos e gerados), observando-se valores praticamente coincidentes e, relação à amostra experimental do Quadro 4 para todas as estatísticas, sendo a maior diferença no máximo. A média dos ventos horários gerados e históricos foram coincidentes na sexta casa decimal: 7.734174 m/s. Os desvios-padrão foram bem próximos: 2.182 (gerado) e 2.171 (histórico). A figura 15 compara os box-plots e as distribuições dos ventos horários históricos e da amostra gerada original.

Quadro 5. Sumários Estatísticos dos ventos horários históricos e gerados amostra original

summary(gerado)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	6.133	7.993	7.734	9.407	17.887
summary(historico)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.270	6.185	8.116	7.734	9.385	14.590

Figura 15. Distribuições e Box-Plots de Ventos Horários. Histórico e Gerado Amostra Original



Adicionando-se o histórico de ventos horários à amostra original obtida no item anterior, forma-se um conjunto original estendido com 2009 (49 x 41) cenários anuais de ventos horários.

### Definição de Atributo

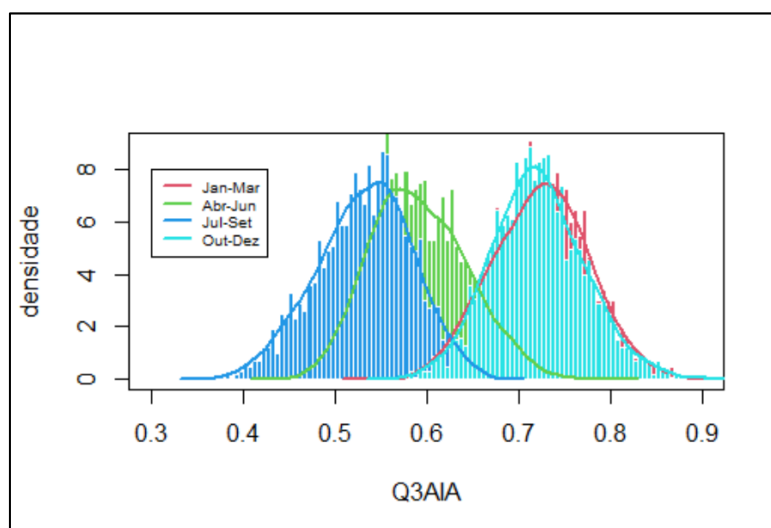
Para atributo de cenário considerou-se as amplitudes intradiárias de vento horário para todos os dias de um trimestre padronizadas pelo vento médio do dia e posterior cálculo do terceiro quartil (Q3AIA) destas amplitudes padronizadas de cada ano. Desta forma se obtém, para cada trimestre, uma amostra com  $N=2009$  de Q3AIA. O quadro 6 apresenta os sumários estatísticos das amostras de Q3AIA de cada trimestre. Nota-se uma variação sazonal das estatísticas de Q3AIA ao longo do ano. As estatísticas diminuem gradualmente de Jan-Mar até Jul-Set, voltando aos níveis de Jan-Mar já em Out-Dez. A figura 16 mostra as distribuições (histogramas e densidades empíricas) de Q3AIA para cada trimestre. As distribuições dos trimestres Out-Dez e Jan-Mar são muito próximas.



Quadro 6. Sumários Estatísticos das amostras de Q3AIA para cada trimestre

summary(Jan-Mar)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.5480	0.6869	0.7257	0.7241	0.7600	0.8784
summary(Abr-Jun)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4478	0.5551	0.5882	0.5921	0.6277	0.7900
summary(Jul-Set)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.3709	0.4987	0.5355	0.5324	0.5684	0.6649
summary(Out-Dez)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.5737	0.6868	0.7196	0.7216	0.7538	0.9050

Figura 16. Histogramas e Densidades Empíricas de Terceiros Quartis de Amplitudes Intradiária Anuais na Amostra Original (N=2009 anos)



### Discretização da Distribuição de Q3AIA e Seleção de Cenários

As figuras 17a e 18a comparam os histogramas de Q3AIA da amostra original com histogramas das distribuições discretas de Q3AIA correspondentes a intervalos de probabilidades acumuladas com 4 intervalos equiprováveis e 4 intervalos não equiprováveis, correspondentes às probabilidades acumuladas  $\{0, 2/4, 3/4, 3.5/4, 1\}$ , detalhando mais a cauda da direita. As figuras 17b e 18b apresentam para cada um dos pontos da distribuição discreta de Q3AIA os correspondentes perfis horários de vento de dias dos cenários selecionados. As figuras 19a e 19b consideram 8 intervalos equiprováveis e as figuras 20a e 20b consideram 8 intervalos não equiprováveis,

correspondentes às probabilidades acumuladas {0, 2/8, 4/8 ,6/8, 7/8, 7.25/8, 7.5/8, 7.75/8, 1}. A tabela 2 compara os desvios-padrões das distribuições discretas e contínuas de Q3AIA.

Tabela 2 – Desvio-padrão das distribuições de Q3AIA

N		Jan-Mar	Abr-Jun	Jul-Set	Out-Dez
2009	equiprováveis	0.052	0.052	0.052	0.051
	equiprováveis	0.048	0.048	0.048	0.047
4		(93.2%)	(93.2%)	(93.2%)	(92.4%)
	não-equiprováveis	0.046	0.047	0.045	0.045
		(88.4%)	(91.1%)	(87.3%)	(89.2%)
8	equiprováveis	0.050	0.050	0.050	0.049
		(97.3%)	(97.5%)	(97.4%)	(96.6%)
	não-equiprováveis	0.050	0.050	0.050	0.049
		(95.7%)	(96.6%)	(95.2%)	(96.0%)

\*entre parêntesis razão entre desvio-padrão da distribuição discreta e da distribuição contínua.

Figura 17a. Histogramas de Terceiros Quartis de Amplitudes Intradiária Anuais na Amostra Original (n=2009 anos) e Amostra Reduzida (n=4). Partição de probabilidades acumuladas equiprováveis

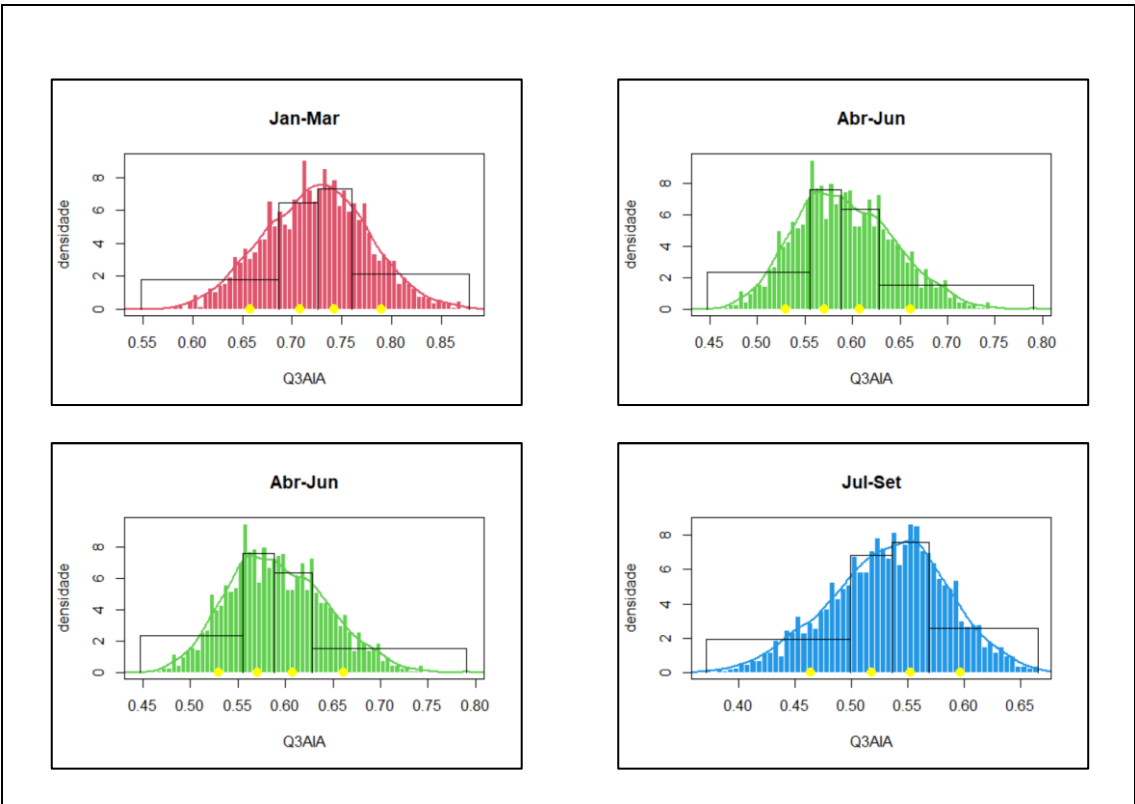


Figura 17b. Perfis de Ventos Horários Representativos  $n=4$ . Partição de probabilidades acumuladas equiprováveis

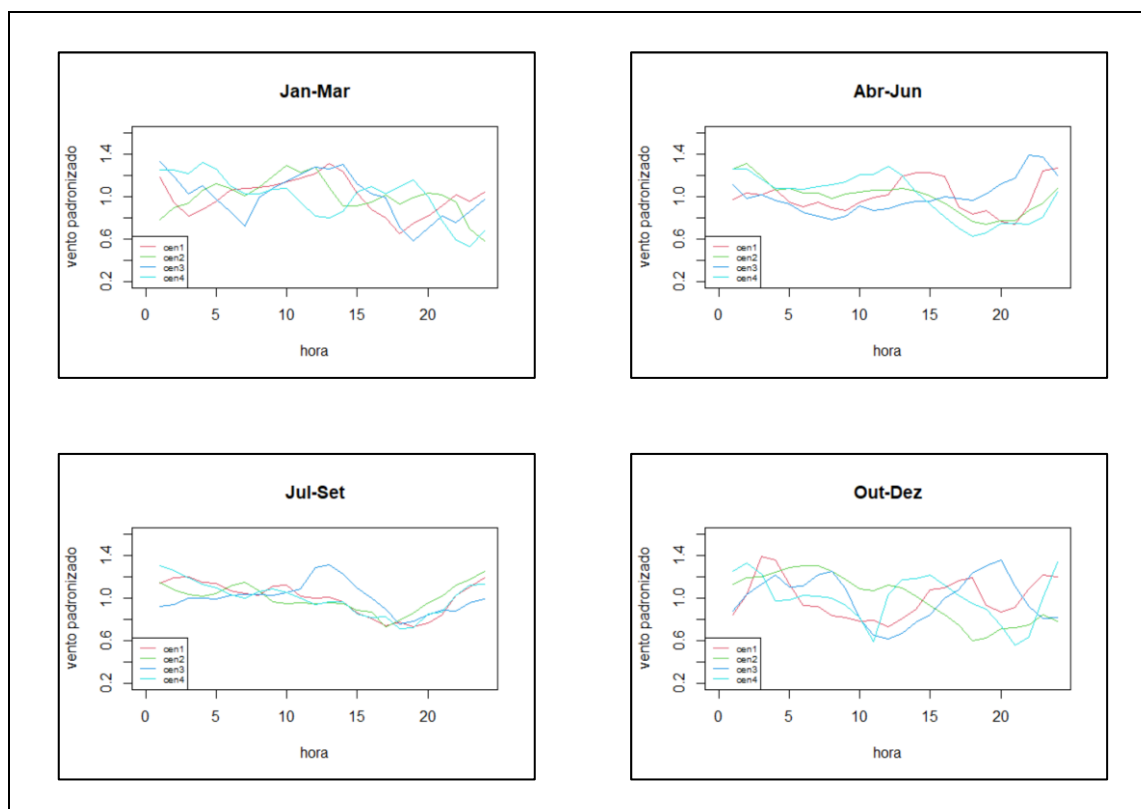


Figura 18a. Histogramas de Terceiros Quartis de Amplitudes Intradiária Anuais na Amostra Original ( $n=2009$  anos) e Amostra Reduzida ( $n=4$ ). Partição de probabilidades acumuladas  $\{0, 2/4, 3/4, 3.5/4, 1\}$

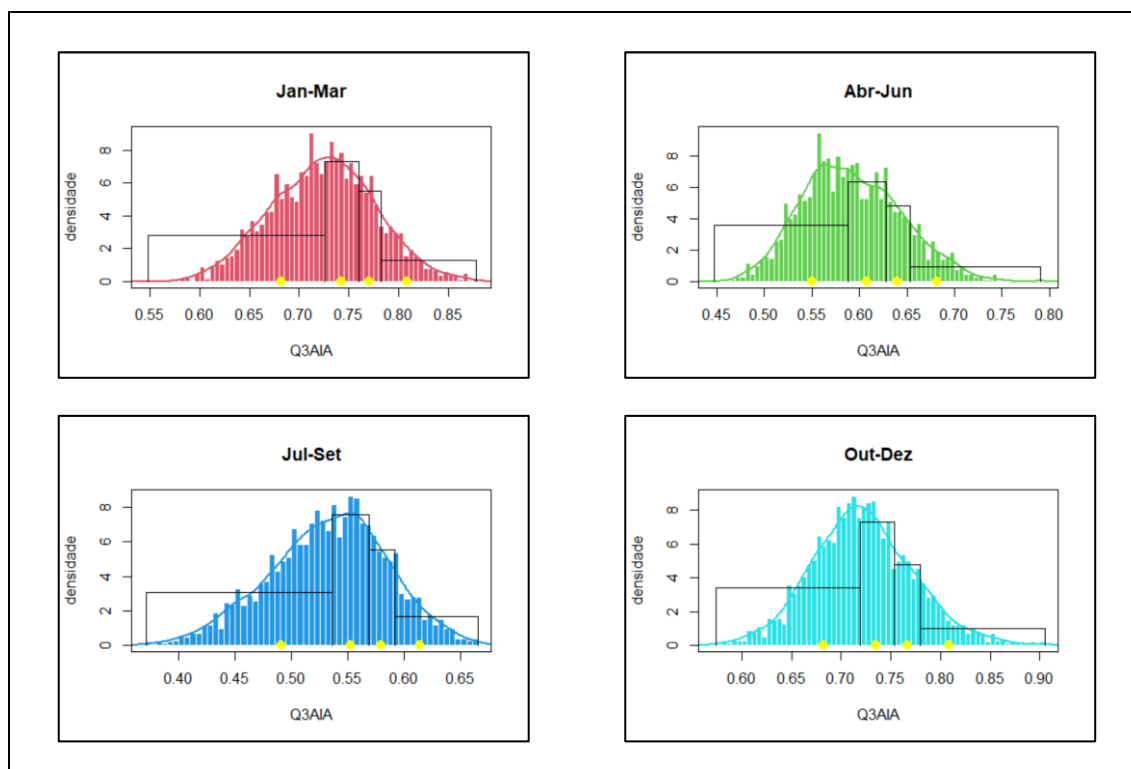


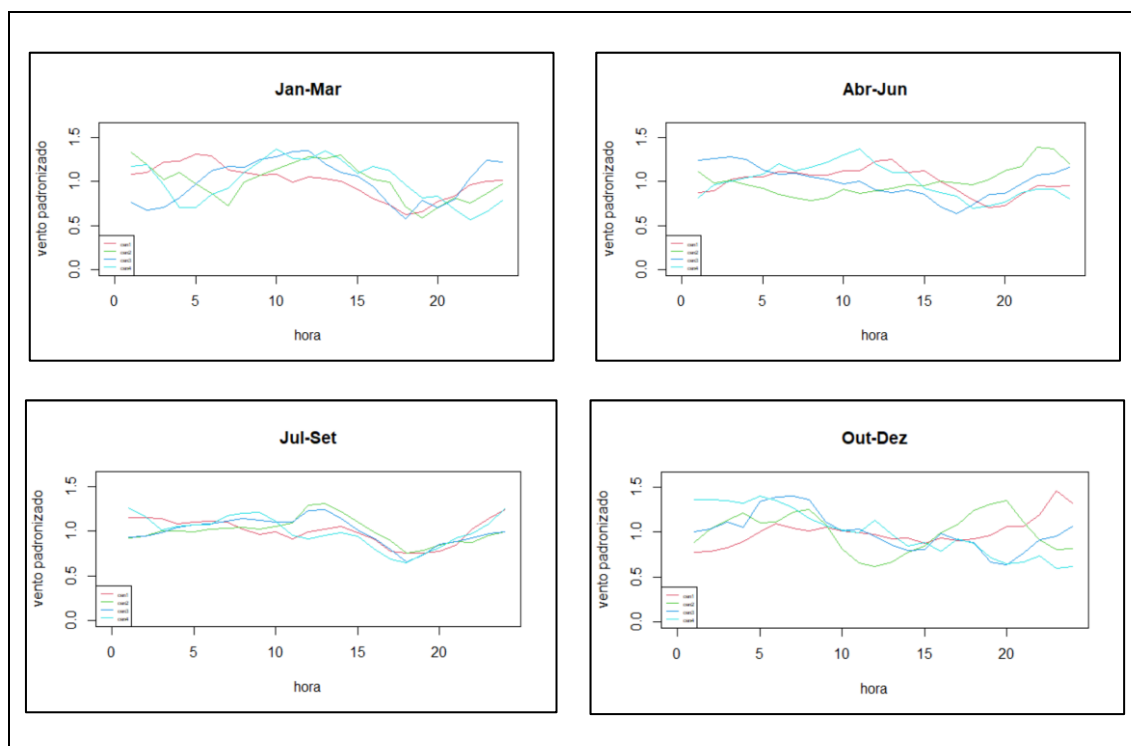
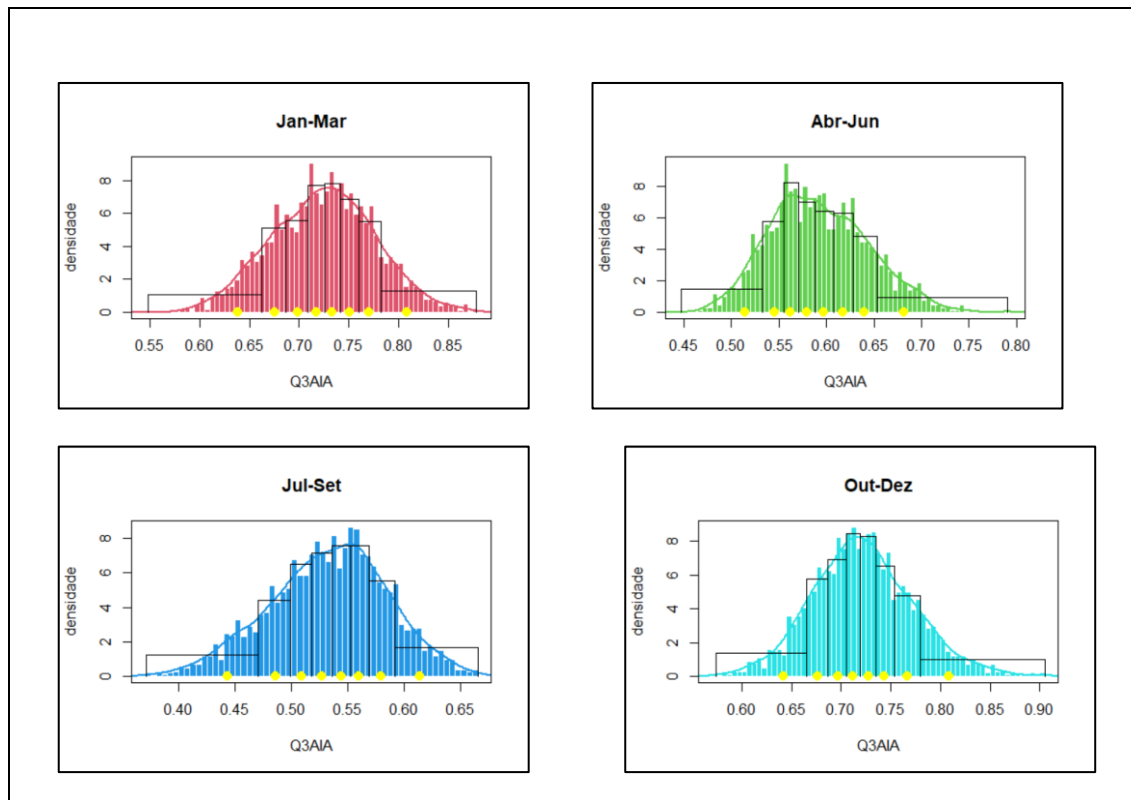
Figura 18b. Perfis Ventos Horários Representativos  $n=4$ . Partição de probabilidades  $\{0, 2/4, 3/4, 3.5/4, 1\}$ Figura 19a. Histogramas de Terceiros Quartis de Amplitudes Intradiária Anuais na Amostra Original ( $n=2009$  anos) e Amostra Reduzida ( $n=8$ ). Partição de probabilidades acumuladas equiprováveis

Figura 19b. Perfis de Ventos Horários Representativos  $n = 8$ . Partição de probabilidades acumuladas equiprováveis

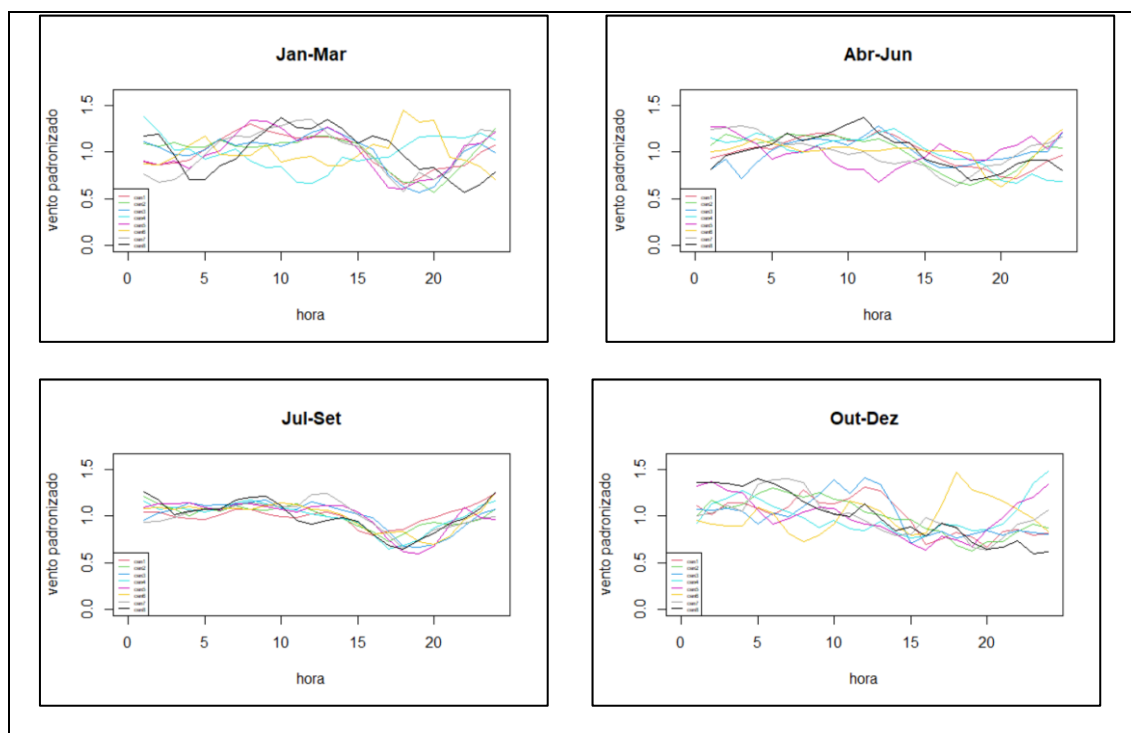


Figura 20a. Histogramas de Terceiros Quartis de Amplitudes Intradiária Anuais na Amostra Original ( $n=2009$  anos) e Amostra Reduzida ( $n=8$ ). Partição de probabilidades acumuladas nos pontos  $\{0, 2/8, 4/8, 6/8, 7/8, 7.25/8, 7.5/8, 7.75/8, 1\}$

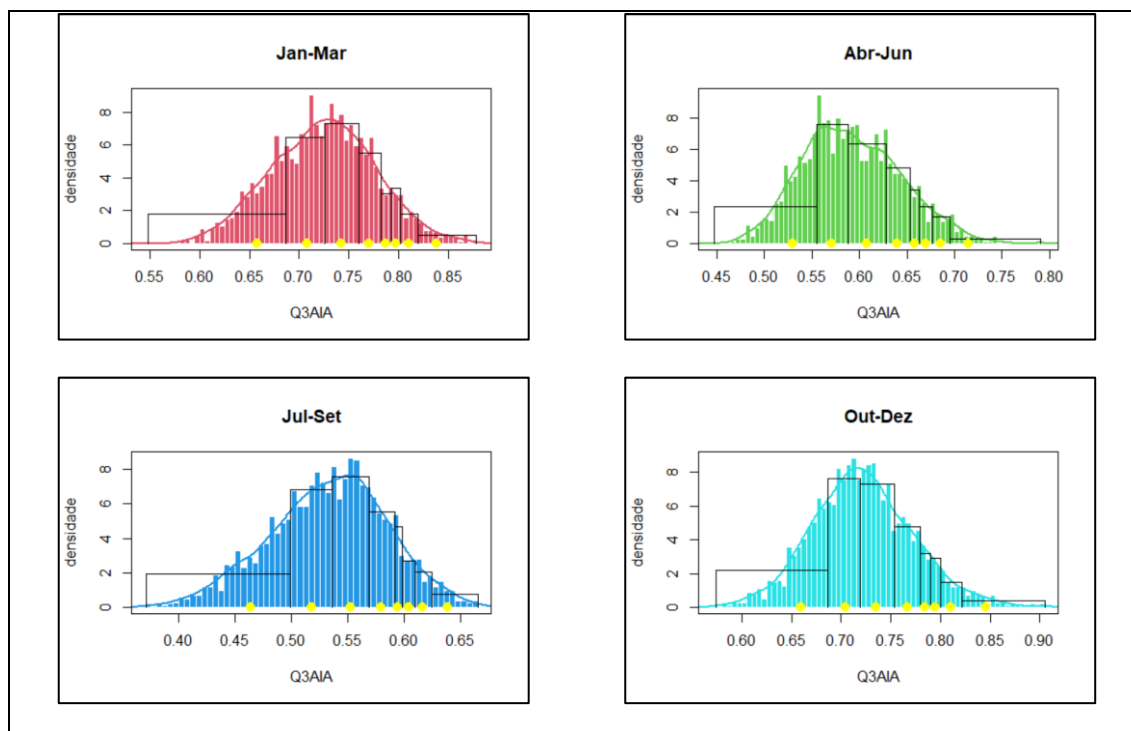
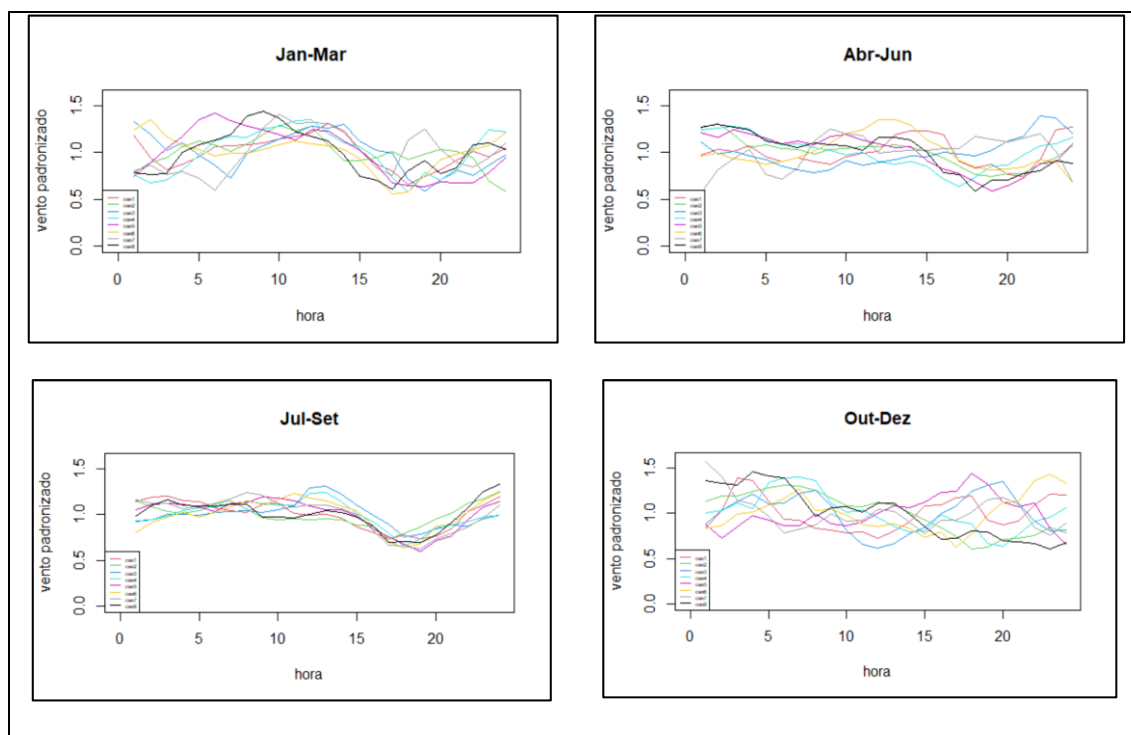


Figura 20b. Perfis de Ventos Horários Representativos  $n = 8$ . Partição de probabilidades acumuladas nos pontos  $\{0, 2/8, 4/8, 6/8, 7/8, 7.25/8, 7.5/8, 7.75/8, 1\}$



## 5. Conclusões

A necessidade de representação da variação intradiária da produção de fontes eólicas em estudos de planejamento da evolução no longo prazo de parques geradores exige, sob o ponto de vista computacional, o desenvolvimento de metodologias para definição de um número limitado de perfis diários de flutuação intradiária dos ventos. Estes perfis são utilizados nos modelos de planejamento em restrições de atendimento da demanda para garantir uma expansão adequada, com a incorporação de tecnologias para mitigar as flutuações da oferta eólica. A metodologia apresentada neste artigo emprega uma modelagem estocástica autorregressiva para a geração computacional de um número muito grande de séries anuais de ventos horários, com posterior uso de técnicas de discretização para seleção de um conjunto menor de séries anuais.

O estudo de caso apresentado, feito a partir de um histórico de 41 anos (1979-2018) de ventos referentes à localidade de Itaguaçu, BA, mostra a potencialidade da metodologia. A modelagem autorregressiva proposta, caracterizada pela abundância de parâmetros para calibração, mostrou-se flexível o suficiente para preservar os principais aspectos do fenômeno. Também a técnica de discretização sugerida para redução da

cardinalidade da amostra gerada mostrou-se adequada para fornecer amostras reduzidas de perfis diários para serem fornecidas para o modelo de planejamento.

## Referências

- ANTUNES, C.; MARTINS, A.; BRITO, I. A multiple objective mixed integer linear programming model for power generation expansion planning. **Energy**, 29, 613-627, 2004.
- BARBIERO, A. A general discretization procedure for reliability computation in complex stress-strength models. **Mathematics and Computers in Simulation**, 1667-1676, 2012.
- BARBIERO, A.; HITAJ, A. Approximation of continuous random variable for the evaluation of the reliability parameter of complex stress-strength models. **Annals of Operations Research**, 1-26, 2021;
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis. Forecasting and Control**. Hoboken, New Jersey EUA: John Wiley & Sons, 2008.
- CARRILO, C.; CIDRÁS, J.; DÍAZ-DOURADO, E.; OBANDO-MONTAÑO, A. An Approach to Determine the Weibull Parameters for Wind Energy Analysis: The Case of Galicia (Spain). **Energies**, 7, 2676-2700, 2014.
- CHAKRABORTY, S. Generating discrete analogues of continuous probability distributions0 A survey of methods and constructions. **Journal of Statistical Distributions and Applications**, 6, 2015.
- COLLINS, S.; DEANE, J. P.; PANOS, E.; PIETZCKER, R.; DELARUE, E.; PÁDRAG, B. G. Integrating short term variations of the power system into integrated energy system models: A methodological review. **Renewable and Sustainable Energy Reviews**, 76, 839-856, 2017.
- DEE, D. **The Climate Data Guide: ERA5 atmospheric reanalysis**. (N. C. Staff, Editor) Acesso em 16 de 12 de 2020, disponível em Climate Data Guide: <https://climatedataguide.ucar.edu/climate-data/era5-atmospheric-reanalysis>, 2020.
- DREZNER, Z.; ZEROM, D. A Simple and Effective Discretization of a Continuous Random Variable. **Communications in Statistics - Simulation and Computation**. doi:10.1080/03610918.2015.1071389, 2015.
- EPE. Plano Decenal de Expansão de Energia 2027. Brasília: MME. Estado da Bahia. **Atlas Eólico da Bahia**. Fonte: [http://www2.secti.ba.gov.br/atlasWEB/apresentacao\\_p2.html](http://www2.secti.ba.gov.br/atlasWEB/apresentacao_p2.html), 2018.
- GIL, E.; ARAVENA, I.; CÁRDENAS, R.. Generation Capacity Expansion Planning Under Hydro Uncertainty Using Stochastic Mixed Integer Programming and Scenario Reduction. **IEEE Transactions on Power Systems**, 30, 1838-1847, 2015.
- GILLELAND, E.; KATZ, R. W. An Extreme Value Analysis Package in R. **Journal of Statistical Software**, 72, 1-39, 2016.
- GOLUB, G.; WELSCH, J. Calculation of Gauss quadrature rules. **Mathematics of Computation**, 6, 221-230, 1969.
- GT MODERNIZAÇÃO DO SETOR ELÉTRICO. **Relatório do Grupo Temático Critérios de Garantia de Suprimento**. Brasília: MME, 2019.
- HAMMOND, K. R.; BICVKEL, J. E. Reexamining Discrete Approximations to Continuous Distributions. **Decision Analysis**, 1-31, 2013.

HAMMOND, R. K.; BICKEL, J. E. Discretization Methods for Continuous Probability Distributions. Em J. J. Cochran, **Wiley Encyclopedia of Operations Research and Management Science** (pp. 1-13). New York: John Wiley & Sons, 2015.

JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society**, 81, 158-171, 1955.

LISBOA, M. L.; DAMAZIO, J. M.; SABOIA, C. H.; MACEIRA, M. E. Matriz Modelo para Estudos Integrados de Planejamento de Longo Prazo para o Sistema Energético Brasileiro. **Anais**, XII Simpósio de Especialistas em Planejamento da Operação e Expansão Elétrica: Cigrè Brasil, 2012.

LISBOA, M. L.; MARZANO, L.; SABOIA, C.; MACEIRA, M. E.; MELO, A.; VENTURA, A.; . . . CUNHA, S. H. Validação do modelo MELP para utilização no Plano Nacional de Energia. **Anais**, XIX Seminário Nacional de Produção e Transmissão de Energia Elétrica. Rio de Janeiro: Cigrè Brasil, 2007.

LOULOU, R.; UWE, R.; KANUDIA, A.; LEHTILA, A.; GARY, G. (2005). Documentation for the TIMES Model. **Energy Technology Systems Analysis Programme (ETSAPP)**. Paris, France: IEA-ETSAP.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. **Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol 1** (pp. 281-297). Berkeley, CA, USA: University of California Press, 1967.

MCLEOD, I.; ZHANG, Y. Improved Subset Autoregression With {R} Package. **Journal of Statistical Software**, 2008.

MESSNER, S.; STRUBEGGER, M. **User's Guide for MESSAGEIII**. Laxenburg, Austria: IIASA, 1995.

OLIVEIRA, W. L.; SAGASTIZÁBAL; C. P., MACEIRA P. M.; DAMAZIO, M. J. Optimal Scenario tress reduction for stochastic streamflows in power eneration planning problems. **Optimization Methods and Software**, 917-936, 2010.

PAUL, BRUGGE, R.; FIELDING, K.; FUENTES, M.; KALLBERG, P.; KOBAYASHI, S.; . . . SIMMONS, A. **The Era-Interim archive Version 2.0**. Reading, UK: European Centre for Medim Range Weather Forecasts, 2011.

PENNA, D. D. Definição da Árvore de Cenários de Afluências para o Planejamento da Operação Energética de Médio Prazo. **Tese de Doutorado**, Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Energia Elétrica. Rio de Janeiro, Brasil: PUC, 2009.

REUK.. **Wind Speed Distribution Weibull**. Fonte: The Renewable Energy Website: <http://www.reuk.co.uk/wordpress/wind/wind-speed-distribution-weibull/>, 2020.

ROY, D.; DASGUPT, T. A Discretizing Approach for Evaluating Reliability of Complex Systems Under Stress-Strenght Moel. **IEEE Transactions on Reliability**, 145-150, 2001.

STEDINGER, J. Evalutation and Presentation of Dam Failure and Flood Risks. **Journal of Water Resources Planning and Managemente**, 123, 216-227, 1997.

Tuller, S. E.; Brett, A. C. The Characteristics of Wind Velocity that Favor the Fitting of a Weibull Distribution in Wind Speed Ananlysis. **Journal of Climate and Applied Meteorology**, 23, 124-134, 1984.

XAVIER, V. L. Resolução do Problema de Agrupamento segundo o Criétrio de Minimização da Soma de Distâncias. **Dissertação de Mestrado**, UFRJ, COPPE/Programa de Engenharia de Sistemas e Computação. Rio de Janeiro: UFRJ, 2012.

ZAINO, A. N.; D'ERRICO, J. Optimal Discrete Approximations for Continuous Outcomes with Applications in Decision and Risk Anlysis. **Journal of the Operational Research Society**, 379-388, 1989.



## HOURLY WIND SCENARIOS FOR THE BRAZILIAN ELECTRIC SYSTEM EXPANSION PLANNING

### ***Abstract***

*The paper presents a methodology for the definition of hourly wind scenarios in regions where wind farms are being implemented aiming to support planning studies for the evolution of the Brazilian generation system. The methodology uses autoregressive stochastic modeling to build a sample with a high number of scenarios (thousands) and discretization techniques of continuous variables to select a smaller set of scenarios that will be used in the planning models. The article presents an application based on a 41-year of hourly wind historical record in a promising area for the installation of wind turbines in the state of Bahia.*

**Keywords:** *Hourly Wind Speed Time Series, Autoregressive Stochastic Modeling, Discretization of Continuous Variables.*