

MODELAGEM DA INTENSIDADE DOS VENTOS EM BRASÍLIA-DF POR MEIO DE CADEIAS DE ORDEM VARIÁVEL

Lucas MOREIRA¹

Tatiana Keiko Borges UMEDA²

- RESUMO: Neste trabalho, propomos modelos climáticos de ordem variável para a intensidade dos ventos na cidade de Brasília-DF. Isso significa que para prevermos a intensidade do vento num determinado dia, consideramos o histórico das intensidades dos ventos em dias anteriores. Sendo que o número de dias necessários para fazer uma dada previsão depende do próprio histórico das intensidades do vento nessa cidade. Inicialmente, consideramos medições diárias da velocidade média do vento na cidade de Brasília correspondentes ao período de 21/08/1961 a 15/07/2019. Por meio desse conjunto de dados, fomos capazes de dizer as chances de que num determinado dia ocorresse ventos fortes ou não. Em seguida, dividimos as observações em duas partes, uma correspondente ao período anterior aos anos 2.000 e a outra associada ao período posterior a esse marco. Estimando os modelos de ordem variável correspondentes a estes dois períodos, fomos capazes de mostrar que houve alteração no regime de ventos na cidade de Brasília nos últimos anos.
- PALAVRAS-CHAVE: Cadeias de ordem variável; árvores de contexto; modelos meteorológicos.

1 Introdução

Seguindo Quintino e Moreira (2015), a motivação inicial deste trabalho foi a obtenção de modelos meteorológicos na região do Distrito Federal. Mais especificamente, obtemos modelos para a intensidade dos ventos na cidade de Brasília. Para tanto, os modelos estocásticos considerados foram as Cadeias de Ordem Variável.

¹Universidade de Brasília - UNB, Instituto de Ciências Exatas, Departamento de Estatística, CEP: 70910-900, Brasília, DF, Brasil. E-mail: *lmoreira@unb.br*

²Universidade de Brasília - UNB, Instituto de Química, Caixa Postal, CEP: 70910-900, Brasília, DF, Brasil. *tatianaumeda@gmail.com*

As Cadeias de Ordem Variável têm a riqueza de informação das Cadeias de Markov com a economia objetivada pela Teoria da Informação, ao se considerar somente a porção importante do passado para realizar previsões. Esses modelos foram apresentados originalmente em Rissanen (1983), que denominou de contexto a parcela do passado necessária para prever o próximo símbolo. Como nenhum contexto pode ser representado como sufixo de outro contexto, é possível representar o conjunto de todos os contextos de um processo de ordem variável por meio de uma árvore probabilística com raiz e rótulos.

Considerando as dependências estruturais, esses modelos são mais parcimoniosos, no que diz respeito ao número de parâmetros, e flexíveis que as Cadeias de Markov de ordem fixa. Mais tarde, na literatura estatística, Bühlmann e Wyner (1999), chamaram esses modelos de Cadeias de Markov de Ordem Variável (VLMC).

As Cadeias de Ordem variável podem ser utilizadas em diversas áreas de aplicação. Rissanen (1983) introduz os modelos de ordem variável num cenário da Teoria da Informação. Galves et al. (2012), Moreira e Bomfim (2016) apresentam modelagens de dados linguísticos por meio desses modelos. Moreira e Quintino (2015) propõem modelos meteorológicos de temperatura de memória variável.

Garivier e Leonardi (2011) classificaram os algoritmos de estimação da árvore de contextos em duas famílias. A primeira, deriva diretamente do Algoritmo Contexto, introduzido em Rissanen (1983), e se destaca por realizar a poda de uma árvore dita completa. Isto é, dada uma medida de discrepância para o ramo podado, é determinando se ele deve ser removido da árvore ou não. A segunda família de estimadores se baseia na máxima verossimilhança penalizada. Para cada árvore possível é calculado um critério que combina o ajustamento com a complexidade do modelo. Esse procedimento é derivado do Princípio da Descrição de Comprimento Mínimo (Minimum Description Length Principle - MDL) de Barron. Csizsár e Talata (2006) utilizaram essa abordagem por meio do Critério de Informação Bayesiana (Bayesian Information Criterion - BIC) para estimação da árvore de contextos.

Neste trabalho, analisamos dados diários de velocidade média dos ventos na cidade de Brasília-DF para calcularmos a probabilidade de que em um determinado dia tenha vento forte ou não. Adicionalmente, verificamos que houve uma mudança no regime dos ventos nessa cidade nos últimos dezenove anos, quando comparamos com a série histórica e com o período anterior aos anos 2.000.

Para estimar árvores de contextos através do conjunto de dados analisado, consideramos a versão do Algoritmo Contexto apresentado em Collet et al (2008). Destacamos que toda a computação em Estatística foi realizada no ambiente R (R CORE TEAM, 2020), seguindo a mesma implementação realizada em Moreira e Quintino (2015).

Este trabalho está dividido da seguinte forma: na Seção 2, apresentamos as notações e conceitos básicos. Na Seção 3, apresentamos a metodologia utilizada. Em seguida, a Seção 4, é dedicada à aplicação das Cadeias de Ordem Variável na intensidade dos ventos em Brasília-DF. Na Seção 5, são expostas as conclusões e considerações finais.

2 Revisão bibliográfica

Nesta seção, apresentamos os conceitos e notações que utilizamos ao longo do trabalho. Em particular, definimos formalmente o que é uma Cadeia de Ordem Variável.

2.1 Definições e notações básicas

Considere o alfabeto $\mathcal{A} = \{0, 1, \dots, N-1\}$ com tamanho $|\mathcal{A}| = N$. Dados dois inteiros $m \leq n$ denotamos a_m^n a sequência de símbolos a_m, a_{m+1}, \dots, a_n de \mathcal{A} e \mathcal{A}_m^n o conjunto de tais sequências.

O comprimento da sequência será $l(a_m^n) = n - m + 1$. Caso $n < m$, $a_m^n = \emptyset$ e $l(a_m^n) = 0$.

O conjunto de todas as sequências semi-infinitas e o conjunto de todas as sequências de símbolos de tamanho finito são denotados, respectivamente, por

$$\mathcal{A}_{-\infty}^{-1} = \mathcal{A}^{\{\dots, -2, -1\}} \quad e \quad \mathcal{A}^* = \bigcup_{j=0}^{\infty} \mathcal{A}_{-j}^{-1},$$

em que para $j = 0$ corresponde ao conjunto das sequências vazias \emptyset .

Dadas duas sequências ω e v , com $l(\omega) < +\infty$, denotamos por $v\omega$ a sequência de comprimento $l(v) + l(\omega)$ obtida pela concatenação das duas sequências. Por exemplo, para $v = \dots, v_{-n-2}, v_{-n-1}$ e $\omega = \omega_{-n}, \dots, \omega_{-2}, \omega_{-1}$, a sequência obtida pela concatenação de v e ω será $v\omega = \dots, v_{-n-2}, v_{-n-1}, \omega_{-n}, \dots, \omega_{-2}, \omega_{-1}$. Note que, para o caso em que $v = \emptyset$ obtêm-se $v\omega = \emptyset\omega = \omega$. Analogamente ocorre para $\omega = \emptyset$.

Uma sequência u é dita ser um *sufixo* de ω se existir s , com $l(s) \geq 1$, tal que $\omega = su$ e será denotada por $u \prec \omega$. Caso $u \prec \omega$ ou $u = \omega$, será denotado por $u \preceq \omega$. Dada uma sequência finita ω denotamos por $\text{suf}(\omega)$ o maior sufixo de ω .

Ao longo desse trabalho consideramos o processo $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ estacionário e ergódico sobre o alfabeto $\mathcal{A} = \{0, 1, \dots, N-1\}$. Assumimos que o processo \mathbf{X} é compatível com a probabilidade de transição $p_X(\cdot|\cdot)$, ou seja,

$$p_X(a|\omega) = \mathbb{P}(X_0 = a | X_{-1} = \omega_{-1}, X_{-2} = \omega_{-2}, \dots), \quad (1)$$

para todo $\omega \in \mathcal{A}_{-\infty}^{-1}$ e para todo $a \in \mathcal{A}$. Para $\omega \in \mathcal{A}_{-j}^{-1}$ a probabilidade estacionária definida por essa sequência será denotada por

$$\mu_X(\omega) = \mathbb{P}(X_{-j}^{-1} = \omega). \quad (2)$$

Com intuito de estimarmos a árvore de contextos de um processo \mathbf{X} , dada uma amostra desse processo, consideramos que \mathbf{X} satisfaz as seguintes definições.

Definição 2.1. Dizemos que um processo \mathbf{X} é não-nulo se satisfaz

$$\alpha_X = \inf\{p_X(a|\omega) : a \in \mathcal{A}, \omega \in \mathcal{A}_{-\infty}^{-1}\} > 0. \quad (3)$$

Definição 2.2. Dizemos que um processo \mathbf{X} possui taxa de continuidade somável se

$$\beta_X = \sum_{k \in \mathbb{N}} \beta_{k,X} < +\infty, \quad (4)$$

em que a sequência $\{\beta_{k,X}\}_{k \in \mathbb{N}}$ é definida por

$$\beta_{k,X} := \sup \left\{ \left| 1 - \frac{p_X(a|\omega)}{p_X(a|v)} \right| : a \in \mathcal{A}, v, \omega \in \mathcal{A}_{-\infty}^{-1} \text{ com } \omega_{-k}^{-1} = v_{-k}^{-1} \right\}. \quad (5)$$

A sequência numérica $\{\beta_{k,X}\}_{k \in \mathbb{N}}$ é chamada *taxa de continuidade do processo* \mathbf{X} . Note que, a condição de não-nulidade do processo \mathbf{X} é necessária para que possamos definir a taxa de continuidade do processo por (5). A taxa de continuidade é uma propriedade esperada para o processo \mathbf{X} , pois, desejamos que dois passados coincidindo nos últimos k símbolos tenham a mesma influência na predição do próximo símbolo da sequência, a medida que k cresce.

Rissanen (1983) chamou de *contexto* a porção do passado necessária para predizer o próximo símbolo do processo, sendo o tamanho desta sequência função do próprio passado. Um contexto infinito é uma sequência semi infinita tal que nenhum dos seus sufixos é um contexto. O conjunto de todos os contextos satisfaz a propriedade do sufixo, isto é, nenhum contexto é sufixo de outro contexto. Esta propriedade permite representar o conjunto de todos os contextos (finito ou infinito enumerável) como uma árvore probabilística com raiz e rótulos. Esta árvore é chamada *árvore de contextos* do processo \mathbf{X} . A seguir definiremos de maneira mais formal um contexto.

Definição 2.3. Dizemos que uma sequência $\omega \in \mathcal{A}_{-j}^{-1}$ é um contexto do processo \mathbf{X} se para toda sequência semi-infinita $x_{-\infty}^{-1} \in \mathcal{A}_{-\infty}^{-1}$ tendo ω como sufixo satisfazer

$$\mathbb{P}(X_0 = a | X_{-\infty}^{-1} = x_{-\infty}^{-1}) = p_X(a|\omega), \quad \forall a \in \mathcal{A}, \quad (6)$$

e nenhum sufixo de ω satisfaz a equação 6.

Denotamos por $d(\mathcal{T})$ a profundidade da árvore \mathcal{T} , ou seja,

$$d(\mathcal{T}) := \max\{l(\omega) : \omega \in \mathcal{T}\}.$$

Uma árvore \mathcal{T} é dita *completa* se qualquer sequência em $\mathcal{A}_{-\infty}^{-1}$ pertence a \mathcal{T} ou tem sufixo que pertence a \mathcal{T} . Dizemos que a árvore de contextos é *limitada* se o comprimento do maior contexto é finito. Caso contrário, \mathcal{T} é dita *ilimitada*.

Dizemos que uma árvore é *irredutível* se nenhuma sequência pode ser substituída por um sufixo sem violar a propriedade sufixo. Essa noção foi introduzida em Csiszár e Talata (2006) e generaliza o conceito de árvore completa.

A seguir definiremos de maneira mais formal uma *árvore probabilística de contextos* e uma Cadeia de Ordem Variável.

Definição 2.4. Uma árvore probabilística de contextos em \mathcal{A} é um par ordenado (\mathcal{T}, \bar{p}) que satisfaz

(1) \mathcal{T} é uma árvore irredutível.

(2) $\bar{p} = \{\bar{p}(\cdot|\omega), \omega \in \mathcal{T}\}$ é uma família de probabilidades de transição sobre \mathcal{A} .

Definição 2.5. Dizemos que o processo \mathbf{X} é compatível com a árvore probabilística de contextos (\mathcal{T}, \bar{p}) se satisfaz

(1) \mathcal{T} é a árvore de contextos do processo \mathbf{X} .

(2) Para qualquer $w \in \mathcal{T}$ e $a \in \mathcal{A}$, $p_X(a|\omega) = \bar{p}(a|\omega)$.

Se \mathbf{X} é compatível com a árvore probabilística de contextos (\mathcal{T}, \bar{p}) , dizemos que \mathbf{X} é uma *Cadeia de Ordem Variável* e denotamos a árvore de contextos de \mathbf{X} por $\mathcal{T}_{\mathbf{X}}$. Note que em (4) da Definição 2.2, se $d(\mathcal{T}_{\mathbf{X}}) < +\infty$, então $\beta_{k,X} = 0$ para $k \geq d(\mathcal{T}_{\mathbf{X}})$, ou seja,

$$\beta_X = \sum_{k=0}^{d(\mathcal{T}_{\mathbf{X}})-1} \beta_{k,X} < +\infty.$$

Em alguns casos podemos estar interessados não na árvore de contextos do processo \mathbf{X} mas na utilização desta árvore com uma restrição no tamanho da maior sequência. Seja K esta restrição. Neste caso, chamaremos de árvore truncada no nível $K \geq 1, K \in \mathbb{N}$. Dessa forma, se definirmos $K \geq d(\mathcal{T}_{\mathbf{X}})$, estaremos considerando a própria árvore de contextos do processo \mathbf{X} .

Definição 2.6. Dado um inteiro K , defina a árvore de contextos truncada no nível K por

$$\mathcal{T}_{\mathbf{X}|K} = \{\omega \in \mathcal{T}_{\mathbf{X}} : l(\omega) \leq K\} \cup \{\omega : l(\omega) = K \text{ e } \omega \prec u, \text{ para algum } u \in \mathcal{T}_{\mathbf{X}}\}.$$

Considere $\mathbf{X} = \{Z_t, t \in \mathbb{Z}\}$ um processo tomando valores num alfabeto finito $\mathcal{A} = \{0, 1, \dots, N-1\}$. Seja X_1, \dots, X_n uma amostra aleatória do processo \mathbf{X} . Para toda sequência finita ω , com $l(\omega) \leq n$, denotamos por $N_n(\omega)$ o número de vezes que observou-se a sequência ω na amostra, ou seja,

$$N_n(\omega) = \sum_{t=0}^{n-l(\omega)} \mathbf{1}_{\{X_{t+1}^{t+l(\omega)} = \omega\}}. \quad (7)$$

Para todo elemento $a \in \mathcal{A}$ e para toda sequência finita ω , a probabilidade de transição empírica é dada por

$$\hat{p}_X(a|\omega)_n = \frac{N_n(\omega a) + 1}{N_n(\omega.) + |\mathcal{A}|}, \quad (8)$$

onde,

$$N_n(\omega.) = \sum_{b \in \mathcal{A}} N_n(\omega b). \quad (9)$$

Observamos que a definição 8 implica que $\hat{p}_X(a|\omega)_n$ é assintoticamente equivalente ao Estimador de Máxima Verossimilhança de $p_Z(a|\omega)_n$.

Antes de apresentar o estimador da árvore de contextos que utilizamos nesse trabalho, é necessário definirmos o seguinte operador

$$\Delta_n(\omega) := \max_{a \in \mathcal{A}} |\hat{p}_X(a|\omega)_n - \hat{p}_X(a|suf(\omega))_n|, \quad (10)$$

para qualquer sequência finita $\omega \in \mathcal{A}^*$. O operador $\Delta_n(\omega)$ calcula a distância entre as probabilidades empíricas dado uma sequência ω e a sequência passada $suf(\omega)$.

3 Metodologia

Nesta seção apresentamos a definição estimador de árvores de contextos, proposta por Galves e Leonardi (2008), que utilizamos em nossa aplicação dos modelos de ordem variável. Este estimador baseia-se no operador $\Delta_n(\omega)$ definido na Seção 2.

Definição 3.1 (Galves e Leonardi, 2008). Para todo $\delta > 0$ e $d < n$ a árvore de contextos estimada $\hat{\mathcal{T}}_n^{\delta,d}$ é o conjunto contendo todas as sequências $\omega \in \bigcup_{i=1}^d \mathcal{A}_{-i}^{-1}$ tais que $\Delta_n(asuf(\omega)) > \delta$ para algum $a \in \mathcal{A}$ e $\Delta_n(u\omega) \leq \delta$ para todo $u \in \bigcup_{i=1}^{d-l(\omega)} \mathcal{A}_{-i}^{-1}$.

Note na Definição 3.1 que as constantes $\delta > 0$ e $d < n$ são fundamentais para o estimador, pois, inicialmente é considerada a árvore de contextos maximal. Neste trabalho, utilizamos o valor de $\delta = 0, 1$.

Assim, nesse estimador, cada sequência ω candidata a contexto possui comprimento $l(\omega) = d$, ou seja, $\omega \in \mathcal{A}_{-d}^{-1}$. Em seguida, o estimador reduz o comprimento das sequências ω que não satisfazem o critério de poda, apresentado na Definição 3.1, tomando $suf(\omega)$ como novo candidato a contexto. Este procedimento é repetido até que a condição de parada seja satisfeita para todas as sequências $\omega \in \hat{\mathcal{T}}_n^{\delta,d}$.

4 Aplicações de cadeias de ordem variável

Nesta seção, inicialmente, propomos modelos de ordem variável para a velocidade média dos ventos na cidade de Brasília-DF. Mais especificamente, a partir do histórico das velocidades médias dos ventos nessa cidade, calculamos a probabilidade de que em um determinado dia ocorra ventos fortes ou não. Em seguida, analisamos se houve mudança no regime de ventos nessa cidade após os anos 2.000, quando comparamos com o período anterior a esse marco.

4.1 Modelagem da velocidade e intensidade dos ventos

Na nossa aplicação de Cadeias de Ordem Variável, consideramos medições diárias, em metros por segundo, da velocidade média do vento na cidade de Brasília

Tabela 1 - Codificação das Velocidades Médias dos Ventos

Velocidade Média (em m/s)	Símbolo no alfabeto
< 3,3	0
≥ 3,3	1

correspondentes ao período de 21/08/1961 a 15/07/2019. Essas informações foram obtidas por meio do site (www.inmet.gov.br) do Instituto Nacional de Meteorologia (INMET). Através do conjunto de dados, vemos que a maior velocidade média registrada em Brasília nesse período foi de 10,33 m/s. Pela Figura 1, vemos que esta intensidade de vento corresponde ao valor cinco na Escala Beaufort, que varia de zero a doze. Dessa forma, podemos dizer que Brasília apresenta, em geral, ventos de baixa intensidade.

Grau	Designação	m/s	km/h	nós	Aspecto do mar	Efeitos em terra
0	Calmo	<0,3	<1	<1	Espelhado	Fumaça sobe na vertical
1	Aragem	0,3 a 1,5	1 a 5	1 a 3	Pequenas rugas na superfície do mar	Fumaça indica direção do vento
2	Brisa leve	1,6 a 3,3	6 a 11	4 a 6	Ligeira ondulação sem rebentação	As folhas das árvores movem; os moinhos começam a trabalhar
3	Brisa fraca	3,4 a 5,4	12 a 19	7 a 10	Ondulação até 60 cm, com alguns carneiros	As folhas agitam-se e as bandeiras desfraldam ao vento
4	Brisa moderada	5,5 a 7,9	20 a 28	11 a 16	Ondulação até 1 m, carneiros frequentes	Poeira e pequenos papéis levantados; movem-se os galhos das árvores
5	Brisa forte	8 a 10,7	29 a 38	17 a 21	Ondulação até 2,5 m, com cristas e muitos carneiros	Movimentação de grandes galhos e árvores pequenas
6	Vento fresco	10,8 a 13,8	39 a 49	22 a 27	Ondas grandes até 3,5 m; borrifos	Movem-se os ramos das árvores; dificuldade em manter um guarda chuva aberto; assobio em fios de postes
7	Vento forte	13,9 a 17,1	50 a 61	28 a 33	Mar revolto até 4,5 m com espuma e borrifos	Movem-se as árvores grandes; dificuldade em andar contra o vento
8	Ventania	17,2 a 20,7	62 a 74	34 a 40	Mar revolto até 5 m com rebentação e faixas de espuma	Quebram-se galhos de árvores; dificuldade em andar contra o vento; barcos permanecem nos portos
9	Ventania forte	20,8 a 24,4	75 a 88	41 a 47	Mar revolto até 7 m; visibilidade precária	Danos em árvores e pequenas construções; impossível andar contra o vento
10	Tempestade	24,5 a 28,4	89 a 102	48 a 55	Mar revolto até 9 m; superfície do mar branca	Árvores arrancadas; danos estruturais em construções
11	Tempestade violenta	28,5 a 32,6	103 a 117	56 a 63	Mar revolto até 11 m; pequenos navios sobem nas vagas	Estragos generalizados em construções
12	Furacão	>32,7	>118	>64	Mar todo de espuma, com até 14 m; visibilidade nula	Estragos graves e generalizados em construções

Figura 1 - Escala Beaufort.

Desse ponto em diante, convencionamos que um dia com vento forte é aquele cuja velocidade média do vento foi de pelo menos 3,3m/s. Caso contrário, dizemos que não houve vento forte. Adicionalmente, todos os dias com ventos fortes serão associados ao número 1, enquanto os dias sem vento forte serão associados ao número 0, conforme a Tabela 1.

Neste trabalho, consideramos que a velocidade média diária dos ventos em Brasília-DF é uma Cadeia de Ordem Variável \mathbf{X} , estacionária, ergódica e tomando valores no alfabeto binário $\mathcal{A} = \{0, 1\}$. Assim, se num determinado dia há vento forte, o processo \mathbf{X} assume o valor 1. Caso contrário, \mathbf{X} assume o valor 0.

Inicialmente, nosso intuito é estimar a árvore de contextos $\mathcal{T}_{\mathbf{X}}$ associada \mathbf{X} . Para tanto, consideramos a versão modificada do Algoritmo Contexto, apresentada na Definição 3.1. Os valores dos parâmetros do estimador $\hat{\mathcal{T}}_n^{\delta,d}$ de árvores de contextos utilizados, foram $\delta = 0,1$ e a profundidade $d = 5$. Ou seja, para fazer previsões sobre a velocidade do vento em um determinado dia, observamos as velocidades médias em, no máximo, cinco dias anteriores.

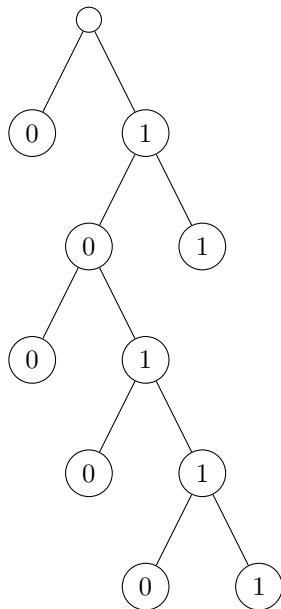


Figura 2 - Árvore de contextos estimada com profundidade $d = 5$.

Apresentamos a árvore de contextos estimada para as velocidades médias dos ventos em Brasília, relativa ao período de 21/08/1961 a 15/07/2019, na Figura 2. A partir da árvore estimada, vemos que se num determinado dia não há vento forte, essa informação é suficiente para prever a velocidade do vento no dia seguinte, pois, 0 é um contexto. Por outro lado, se em um determinado dia houve ventos fortes, então devemos olhar para as velocidades médias no dias anteriores para prever o que acontecerá no dia seguinte. Em particular, se há dois dias consecutivos com ventos fortes, então essa informação é suficiente para prever o que ocorrerá no dia seguinte, pois, 11 é um contexto. Interpretações semelhantes podem ser feitas para os demais contextos estimados 001, 0101, 01101 e 11101.

As probabilidades de transição associadas à árvore de contextos estimada são apresentadas na Tabela 2. Com isso, temos que a probabilidade de que não haja vento forte num determinado dia, dado que no dia anterior também não houve vento forte, é de 82,32%. Analogamente, a chance de que em um determinado dia haja vento forte, dado que nos dois dias anteriores os ventos foram fortes é de 56,43%.

Interpretações semelhantes podem ser feitas para os demais entradas da Tabela 2.

Tabela 2 - Probabilidades de Transições Estimadas da Figura 2. Em que, ω representa o contexto estimado e a denota a intensidade do vento

$\hat{p}_Z(a \omega)$	a	
	0	1
11101	0.4021739	0.5923913
01101	0.6341463	0.3658537
0101	0.5196850	0.4803150
001	0.5811467	0.4188533
11	0.4357013	0.5642987
0	0.8232333	0.1767667

4.2 Comparando períodos

Nessa seção mostramos que houve alteração na árvore de contextos correspondente ao regime de ventos em Brasília dos últimos dezenove anos, quando comparamos com a árvore de contextos do regime de ventos nessa cidade nos anos anteriores aos anos 2.000.

Nas Figuras 3 e 4 apresentamos as árvores de contextos estimadas correspondentes aos períodos de 21/08/1961 a 31/12/1999 e de 01/01/2000 a 15/07/2019, respectivamente. Como as árvores estimadas são distintas, podemos afirmar que há evidências amostrais de que houve alteração na velocidade dos ventos em Brasília após os anos 2.000. Além disso, no período anterior aos anos 2.000, a informação de que houve dois dias com ventos fortes era suficiente para prever a intensidade do vento no dia seguinte, pois, 11 é um contexto. Isso não ocorre após os anos 2.000, pois, nesse período, a informação de que ocorreu dois dias com ventos fortes, não é suficiente para prever qual a intensidade do vento no dia seguinte. No entanto, nos últimos dezenove anos, se houve três dias seguidos com ventos fortes, podemos prever o que ocorrerá no dia seguinte, pois, 111 é um contexto. Interpretações semelhantes podem ser feitas para os demais contextos dessas duas árvores estimadas.

As Tabelas 3 e 4 trazem as probabilidades de transição associadas às árvores nas Figuras 3 e 4. Logo, a chance de que não ocorra vento forte num determinado dia, dado que no dia anterior não ocorreu vento forte é de 81,19% antes dos anos 2.000, e de 84,51%, após esse período. Da mesma forma, pela Tabela 3. é de 57,70% a chance de que ocorra vento forte num determinado dia, dado que nos dois dias anteriores ocorreram ventos fortes, antes dos anos 2.000.

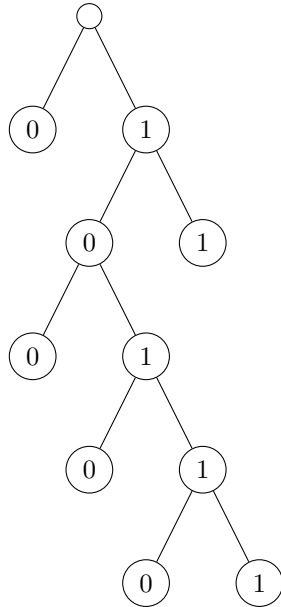


Figura 3 - Árvore de contextos antes dos anos 2.000.

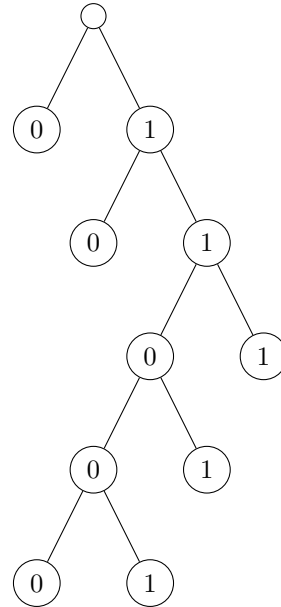


Figura 4 - Árvore de contextos após os anos 2.000.

Tabela 3 - Probabilidades de Transição Estimadas da Figura 3. Em que, ω representa o contexto estimado e a denota a intensidade do vento

$\hat{p}_Z(a \omega)$	a	
	0	1
11101	0.3959732	0.6040268
01101	0.6333333	0.3666667
0101	0.5107914	0.4892086
001	0.5619534	0.4380466
11	0.4230010	0.5769990
0	0.8110386	0.1888610

Tabela 4 - Probabilidades de Transição Estimadas da Figura 4. Em que, ω representa o contexto estimado e a denota a intensidade do vento

$\hat{p}_Z(a \omega)$	a	
	0	1
00011	0.5837321	0.4162679
10011	0.4545455	0.5454545
1011	0.4625000	0.5375000
111	0.4162304	0.5837696
01	0.6037296	0.3951049
0	0.8450832	0.1549168

5 Conclusões

Neste trabalho, consideramos que a intensidade do vento na cidade de Brasília-DF é uma Cadeias de Ordem Variável tomando valores no alfabeto binário $\mathcal{A} = \{0, 1\}$. Sendo que o processo assume o valor 1 se há vento forte na cidade e assume o valor 0, caso contrário. Baseados nessa classe de modelos, propomos modelos meteorológicos de memória variável para prever a intensidade do vento. Para a previsão da intensidade do vento num determinado dia, os modelos propostos levam em consideração a intensidade do vento em, no máximo, cinco dias anteriores.

Inicialmente, consideramos medições diárias da velocidade média do vento na cidade de Brasília correspondentes ao período de 21/08/1961 a 15/07/2019. Por meio desse conjunto de dados, propomos modelos meteorológicos de ordem variável para a intensidade do vento correspondente a esse período.

Em seguida, dividimos as observações em duas partes, uma correspondente ao período anterior aos anos 2.000 e a outra associada ao período posterior a esse marco. Estimamos as árvores de contextos correspondentes a cada um dos dois períodos. Com isso, mostramos que houve alteração no regime de ventos na cidade de Brasília nos últimos dezenove anos. Foge do escopo desse trabalho, indicar as causas dessas mudanças. Porém, destacamos que nos últimos anos a cidade de Brasília passou por um grande aumento da densidade demográfica. Consequentemente, houve grande crescimento na frota de veículos em circulação nessa região, aumentando assim a emissão de gases poluentes. Além disso, observamos um grande aumento no número de condomínios verticais, o que pode afetar a circulação das correntes de ar em Brasília.

Em trabalhos futuros, pretendemos utilizar modelos de contaminação estocástica, como aqueles apresentados em Garcia e Moreira (2015), para propormos modelos meteorológicos de ordem variável baseados em dados contaminados. Uma vez que a contaminação desse tipo de conjunto de dados pode ocorrer por conta de falhas no equipamento de medição de dados climáticos ou por conta da localização geográfica da base de coleta de dados.

Agradecimentos

Queremos registrar que esse trabalho foi parte de um Programa de Iniciação Científica (ProIC), da Universidade de Brasília (UnB) e prestar agradecimentos a Universidade de Brasília, a CAPES e ao CNPq pela contribuição com o nosso estudo e aos revisores e editores pelos comentários e sugestões.

MOREIRA, L.; UMEDA, T. K. B. Model of wind behavior in the city of Brasília using variable order chains. *Rev. Bras. Biom.*, Lavras, v.38, n.4, p.428-439, 2020.

■ **ABSTRACT:** *In this work, we propose scalable models of variable order for the intensity of the winds in the city of Brasília-DF. This means that, in order to predict the intensity of the wind on a given day, we consider the history of intensities from the previous days. The number of days required to make a forecast depends on the history of the wind intensities in that city. Initially, we considered the perfect measurements of the average wind speed in the city of Brasília corresponding to the period from 21/08/1961 to 7/15/2019. Through this data set, we can indicate the chances of a certain number of strong days or not. Then, the dividends are applied in two parts, one corresponding to the period before 2000 and the other associated to the period after this milestone. Estimates of the variable order models corresponding to these two periods, were shown when there was a change in the wind regime in the city of Brasilia in recent years.*

■ **KEYWORDS:** *Chains of variable memory; contexts trees; meteorological models.*

Referências

BÜHLMANN, P., WYNER, A. J. Variable length Markov chains. *Annals of Statistics*, v.27, p.480-513, 1999.

CSISZÁR, I., TALATA, Z. Context tree estimation for not necessarily finite memory processes, via BIC and MDL. *EEE Trans. Inform. Theory*, v.52, n.3, p.1007-1016, 2006.

GALVES, A., GALVES, C., GARCIA, J. E., GARCIA, N. L., LEONARDI, F. Context tree selection and linguistic rhythm retrieval from written texts. *Annals of Applied Statistics*, v.6, n.1, p.186-209, 2012.

GARIVIER, A., LEONARDI, F. Context tree selection: a unifying view. *Stochastic processes and their applications*, v.121, p.2488-2506, 2011.

MOREIRA, L., BOMFIM, A. B. A. Estudo comparativo entre o português brasileiro e europeu: uma aplicação de árvores de Contexto. *Revista Sodebras*, v.12, n.136, p.176-181, 2017.

MOREIRA, L.; QUINTINO, F. S. Cadeias de ordem variável estocasticamente perturbadas: Aplicação em modelos climáticos. *Revista Soderbras*, v.10, p.141-146, 2015.

RISSANEN, J. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, v.11, n.2, p.416-431, 1983.

Recebido em 05.11.2019.

Aprovado após revisão em 15.04.2020.