

Análise estatística da distribuição da direcção do vento na região de Castelo Branco

*Armando Mateus Ferreira¹
José Maria Caridad y Ocerin²*

Resumo

A direcção do vento é um importante factor meteorológico a ter em conta na agricultura (localização e orientação de culturas, barreiras corta-vento, na disposição e orientação de estufas, etc), na localização das unidades de conversão de energia eólica e em estudos de impacto social e ambiental de diversas obras de engenharia civil.

Este fenómeno meteorológico é uma variável nitidamente direccionada, cuja análise estatística requer uma metodologia específica.

No presente trabalho são apresentadas algumas metodologias de análise estatística de dados direccionais, bem como a função de distribuição de probabilidades de von Mises, típica deste género de dados estatísticos.

1. Introdução

O conceito de estatísticas direccionais está relacionado essencialmente com observações que são vectores unitários no plano ou num espaço tri-dimensional. O espaço amostral é tipicamente um círculo ou uma

esfera, em que os métodos estatísticos tradicionais de análise de dados uni ou multi-dimensionais não podem ser usados.

A direcção do vento é um exemplo típico de dados susceptíveis de serem analisados com as metodologias da estatística direccionada. Mardia e Jupp (2000) apresentam diversos exemplos de dados tipicamente direccionais, em diversas áreas da ciência, tais como a biologia e ecologia (por exemplo, as direcções preferidas de migrações de aves), medicina (distribuição temporal, ao longo do ano, da incidência de doenças), física (direcções dos eixos ópticos de cristais, distribuição espacial de fontes de radiações cósmicas), meteorologia (distribuição da direcção do vento ou da chuva), astronomia (órbitas dos corpos celestes), etc.

A direcção do vento é a quantificação do rumo do vento, medido num espaço circular, no sentido dos ponteiros do relógio, e cuja referência é o Norte. A unidade de medida mais usada nos registos meteorológicos da direcção do vento é o grau, mas também pode vir expressa em pi-radianos ou em gradus.

A fim de justificar a não validade das metodologias clássicas de análise estatística, consideremos uma amostra teórica constituída por apenas duas observações da direcção do vento registada num determinado local; sejam os seus valores 355° e 5° , respectivamente (Fig. 1). Suponhamos que nos interessaria estimar a direcção média do vento, bem como uma medida de dispersão, por exemplo o desvio padrão.

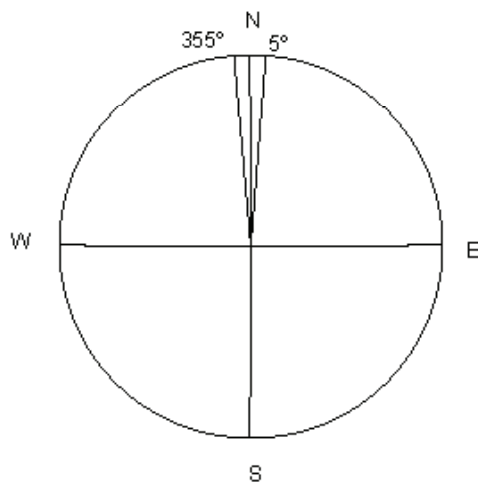


Fig. 1 - Amostra teórica de dois registos de velocidade do vento.

Usando indevidamente as metodologias clássicas de análise estatística, o valor estimado para a média amostral seria

$$\bar{x} = \frac{\sum x_i}{N} = \frac{360}{2} = 180^\circ$$

e o respectivo desvio padrão seria

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{175^2 + (175)^2}{2-1}} = \sqrt{61250} = 247.5^\circ$$

isto é, concluiríamos que a direcção média do vento seria do rumo 180°, exactamente no quadrante oposto da direcção das observações (Fig. 2), e que a direcção do vento seria extremamente dispersa, isto é, muito pouco constante, face ao elevado valor do desvio padrão (como se pode constatar graficamente, a direcção média é 0°, valor que se pode confirmar pela metodologia de seguida exposta). Isto é, estaríamos a chegar a uma conclusão completamente antagónica da realidade exemplificada pela nossa amostra teórica.

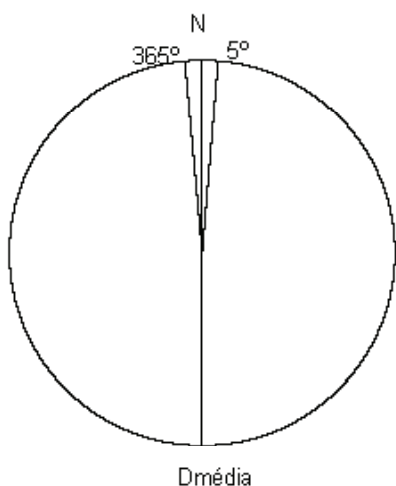


Fig. 2 - A utilização da análise estatística clássica conduz a resultados errados.

Estas conclusões erróneas resultam do uso indevido dos métodos clássicos de estimação estatística, que não podem ser usados com dados tipicamente direccionais.

Feito o enquadramento da necessidade de umas

metodologias estatísticas que permitam a correcta análise e interpretação de dados circulares, passar-se-á de seguida à apresentação de alguns conceitos de estatísticas direccionais, bem como ao ajustamento dos dados a uma função de distribuição habitualmente usada nestas condições.

No presente trabalho utilizaremos os dados da direcção do vento, registadas com a periodicidade de 3 horas (registos às 0, 3, 6, 9, 12, 15, 18, 21 horas UTC), desde Janeiro de 1988 a Dezembro de 1999, na Estação Meteorológica de Castelo Branco. Na tabela 1 apresenta-se um resumo destas observações. Os registos do rumo do vento foram efectuados em graus, arredondando para a dezena de graus mais próxima.

Tab. 1 - Resumo das observações da direcção do vento, no período de 1988 a 1999, na Estação Meteorológica de Castelo Branco

Rumo(°)	Freq.	Rumo(°)	Freq.	Rumo(°)	Freq.
101214	130	363	250	1271	
201143	140	293	260	1727	
30946	150	362	270	2067	
40729	160	313	280	1903	
50726	170	379	290	1391	
60889	180	589	300	981	
70983	190	603	310	571	
801060	200	649	320	504	
901038	210	752	330	448	
100713	220	743	340	576	
110540	230	880	350	811	
120417	240	1228	360	1341	

Número total de observações: 31143
 Número de observações de calma: 3788 12.2%
 Número de observações faltantes: 133

2. Métodos de estatística direccional

2.1 Representação gráfica

Os dados de amostras circulares podem representar-se graficamente, expressando a sua natureza circular. No caso específico da direcção do vento, estes diagramas são geralmente designados por rosa dos ventos.

Na Fig. 3 apresenta-se a rosa dos ventos das frequências dos rumos observados no período em análise, onde se constata que a direcção predominante dos ventos é dos quadrantes oeste e oeste-noroeste (W - WNW), seguido do quadrante este (E).

Em termos de análise estatística, é conveniente apresentar a rosa dos ventos em forma linear, isto é, na forma de um histograma de frequências. Contudo, há que ter em atenção a origem dos eixos na elaboração do histograma, de modo permitir uma análise correcta (Mardia e Jupp, 2000).

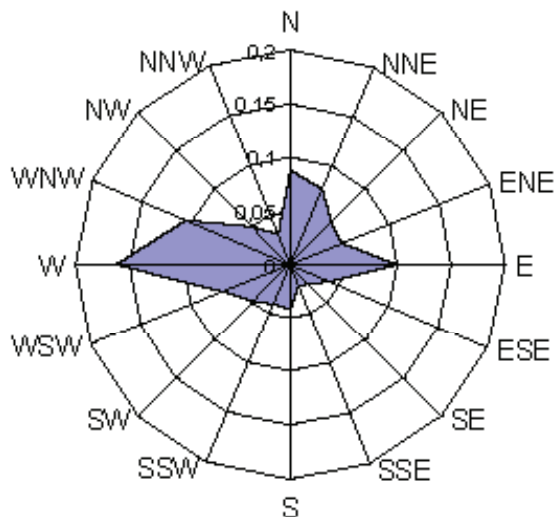


Fig. 3 - Rosa dos ventos para a Região de Castelo Branco (INMG, 1988-1999)

Na Fig. 4 apresenta-se o histograma correspondente à rosa-dos-ventos, considerando a origem dos dados no rumo 0° (Norte). Vejamos que a análise deste histograma conduz a um enviesamento na conclusão errada de que a direcção média do vento se situará entre os 220° e 250°.

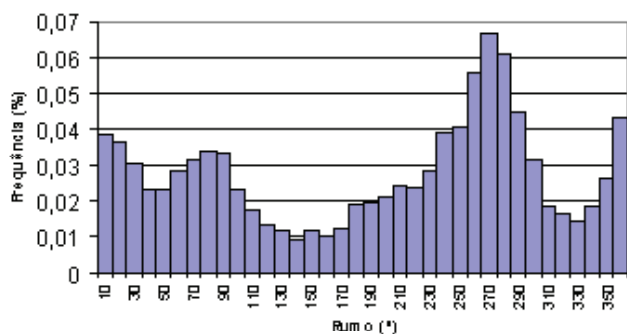


Fig. 4 - Histograma da distribuição de frequências da direcção do vento em Castelo Branco (INMG, 1988-1999)

Repare-se que o gráfico apresenta um mínimo no rumo 157.5° (SSE), e a partir deste ponto vai progressivamente crescendo até atingir o máximo no rumo 270 (W), voltando a decrescer. Usando este mínimo para a origem do histograma, a imagem resultante traduz com maior evidência a distribuição espacial das direcções do vento (Fig. 5).

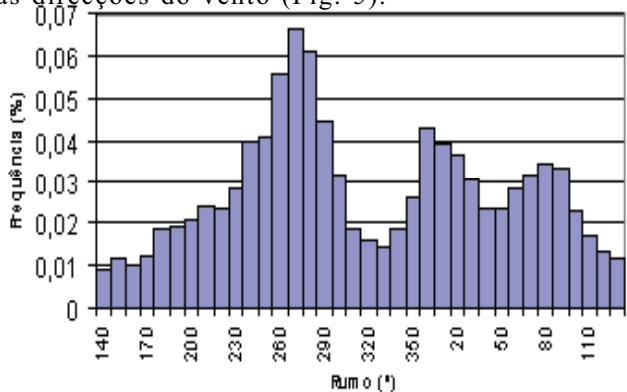


Fig. 5 - Histograma da distribuição de frequências da direcção do vento em Castelo Branco, com alteração da origem (INMG, 1988-1999)

Repare-se que na Fig. 5 é mais evidente distribuição quase bi-modal das frequências da direcção do vento, de acordo com a rosa-dos-ventos (Fig. 3), bem como uma ligeira tendência de assimetria, como será confirmado pelos parâmetros estatísticos de se-

2.2 Estimação dos parâmetros estatísticos

As direcções no plano podem ser entendidas como vectores unitários $\mathbf{x} = (\cos q, \sin q)^T$, ou como pontos no círculo unitário. Assim, cada ponto z pode ser representado por um ângulo ou um número complexo, $z = \cos q + i \sin q$:

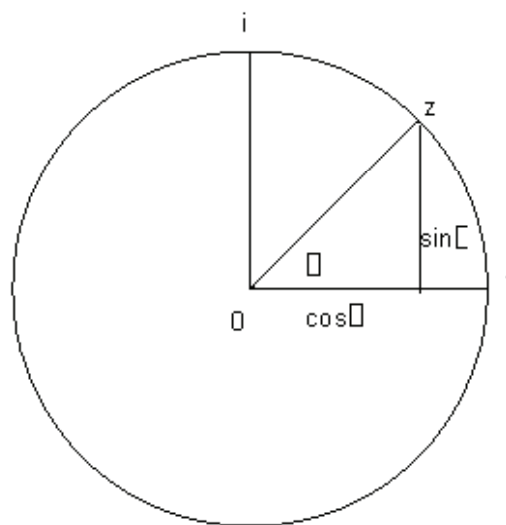


Fig. 6 - Representação da direcção x pelo ângulo θ e pelo número complexo $z = \cos q + i \sin q$

Os ângulos, como já se referiu, costumam exprimir-se em graus; para efeitos de cálculo, geralmente exprimem-se em radianos. A transformação de graus em radianos obtém-se pela expressão $\text{graus} \cdot \pi / 180$. A notação e metodologias de seguida aplicadas são as usadas por Mardia e Jupp (2000) e Zar (1999).

2.2.1 Parâmetros de localização

Suponhamos uma amostra de direcções do vento, medida num determinado local, constituída pelos vectores unitários $\mathbf{x}_1, \dots, \mathbf{x}_n$, com os correspondentes ângulos $q_i, i=1, \dots, n$. A direcção média $\bar{\mathbf{a}}$ é a direcção do vector resultante de $\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n$; isto é, é a direcção do centro de massa $\bar{\mathbf{v}}$ de $\mathbf{x}_1, \dots, \mathbf{x}_n$. Como as coordenadas cartesianas de \mathbf{x}_j são $(\cos q_j, \sin q_j)$ as coordenadas do centro de massa $\bar{\mathbf{x}}$ são (\bar{C}, \bar{S}) , em que:

$$\bar{C} = \frac{1}{n} \sum_{j=1}^n \cos q_j, \quad \bar{S} = \frac{1}{n} \sum_{j=1}^n \sin q_j$$

A direcção média \bar{q} é a solução das equações:

$$\bar{C} = \bar{R} \cos \bar{q}, \quad \bar{S} = \bar{R} \sin \bar{q}$$

em que \bar{R} é o comprimento médio do vector resultante, e é dado por:

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$$

Se a amostra em análise contém os dados agrupados, os cálculos são efectuados considerando que todos os dados num intervalo estão concentrados no ponto médio desse intervalo. Este pressuposto pode afectar o valor de alguns parâmetros estatísticos. Nesta situação, é necessário proceder à correcção do cálculo do valor de \bar{R} . Se os intervalos de agrupamento dos dados têm todos a mesma amplitude h (em radianos), o valor corrigido de \bar{R} é dado por:

$$\bar{R}^* = \frac{\frac{h}{2} \bar{R}}{\sin \frac{h}{2}}$$

Como para $h \approx 45^\circ$ o factor de correcção é inferior a 1.03, esta correcção só é relevante se as classes de agrupamento dos dados são muito amplas.

Das expressões anteriores deduzem-se as seguintes relações:

$$\frac{1}{n} \sum_{j=1}^n \cos q_j = \bar{C}, \quad \frac{1}{n} \sum_{j=1}^n \sin q_j = \bar{S}$$

análoga à relação que se verifica no caso da estatística

linear.

Note-se que \bar{q} não está definido para $\bar{R} = 0$. Quando $\bar{R} > 0$, \bar{q} é dado explicitamente por:

$$\bar{q} = \begin{cases} \arctan \frac{\bar{S}}{\bar{C}} & \text{se } \bar{C} \geq 0 \\ \arctan \frac{\bar{S}}{\bar{C}} + \pi & \text{se } \bar{C} < 0 \end{cases}$$

O coeficiente de assimetria $\frac{\bar{R}_2 \sin 3\bar{q}}{\bar{R}^3}$; o coeficiente de achatamento é $\frac{\bar{R}_2 \cos 3\bar{q}}{\bar{R}^3}$. Estes coeficientes tomam valores próximo de zero se

a amostra é uni-modal, simétrica, e com uma elevação junto à moda semelhante à da distribuição normal. \bar{q}_2 e \bar{R}_2 são a média e o comprimento médio do vector resultante dos ângulos $2q_1, \dots, 2q_n$.

Para a amostra em análise, $n=31143$, $\bar{C} = 0.10642795$ e $\bar{S} = 0.15952106$, de onde resulta que $\bar{R} = 0.19176516$, sendo a direcção média $\bar{q} = 0.98244413$ radianos, ou seja, $\bar{q} = 303.71^\circ$. Repare-se que este valor está de acordo com a análise da rosa-dos-ventos e do histograma apresentado na Fig. 5, enquanto que a Fig. 4 nos induziria erroneamente a supor que a direcção média estaria no quadrante WSW. Os coeficientes de assimetria e de achatamento são respectivamente -0.2734 e -0.0768 , o que significa que a distribuição é ligeiramente assimétrica à direita, com a altura da moda um pouco inferior ao que seria de esperar, comparativamente com uma distribuição normal linear.

A mediana \bar{q} de uma amostra de dados circulares é o ângulo \bar{q} que:

- i) 50% das observações situam-se no arco $[\bar{q}, \bar{q} + \pi]$, e
- ii) Existe maior concentração de observações na vizinhança de \bar{q} do que na vizinhança de $\bar{q} + \pi$.

Pela conjugação destes dois critérios com a amostra em análise, tem-se para esta amostra a mediana o rumo $\bar{q} = 290$. Note-se que, $\bar{q} < \bar{q}$ o que está de acordo com a ligeira assimetria da amostra.

2.2.2 Medidas de concentração e de dispersão

A amplitude de uma amostra circular é o menor arco q_1, \dots, q_n em a totalidade das observações. Sejam $0 \leq q_i \leq 2\pi$ estatísticas ordinais da amostra, tal que $0 \leq q_i \leq 2\pi$. As amplitudes $w_i = q_{(i+1)} - q_{(i)}$ entre as observações $w_n = 2\pi - q_{(n)} + q_{(1)}$ para $i=1, \dots, n-1$ e w_n . A amplitude da amostra circular é:

$$w = 2\pi \max \{w_1, \dots, w_n\}$$

Existem amostras, tais como a amostra em análise, em que existem observações de todos os rumos do círculo, pelo que em tais casos a amplitude $w = 2\pi$.

O comprimento médio do vector resultante $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$ é o comprimento do vector do centro de gravidade ou centro de massa dos vectores unitários $\mathbf{x}_1, \dots, \mathbf{x}_n$, de modo que $0 \leq \bar{R} \leq 1$. Se as direcções q_1, \dots, q_n são muito concentradas, então $\bar{R} \approx 1$; pelo contrário, se as direcções são muito dispersas, então $\bar{R} \approx 0$. Isto é, o parâmetro \bar{R} mede o grau de concentração da amostra de direcções. Para efeitos de descrição e inferência, o parâmetro \bar{R} é mais importante que qualquer outra medida de dispersão.

Para efeitos de analogia com dados não direc-

nais, é conveniente, contudo, definir outras medidas de dispersão.

Uma medida de dispersão circular é dada por $D(\bar{q}) = \frac{1}{n} \sum_{i=1}^n \cos^2(q_i - \bar{q})$. Considerando o afastamento dos ângulos q_1, \dots, q_n em relação à direcção média \bar{q} ,

$$D(\bar{q}) = \frac{1}{n} \sum_{i=1}^n \cos^2(q_i - \bar{q}) = \frac{1}{n} \sum_{i=1}^n \cos^2(q_i) = \bar{R} = V$$

tem-se: Isto é, a variância circular é dada por:

$$V = 1 - \bar{R}$$

O desvio padrão circular, equivalente ao desvio padrão de amostras lineares, é obtido pela expressão:

$$u = \sqrt{2 \log \frac{1}{V}} = \sqrt{2 \log \bar{R}}$$

Para valores pequenos de V, a expressão anterior

$$u \approx \sqrt{2V} = \sqrt{2} \sqrt{1 - \bar{R}}$$

reduz-se a

O coeficiente de dispersão circular é definido pela expressão:

$$\hat{d} = \frac{1 - \bar{R}_2}{2\bar{R}^2}$$

onde \bar{R}_2 é o comprimento médio do vector resultante dos ângulos $2q_1, \dots, 2q_n$.

Na amostra em análise, o valor da variância circular é $\hat{V} = 1 - 0.1918 = 0.8082$ e o desvio padrão circular

$$\hat{u} = \sqrt{2 \log \bar{R}} = \sqrt{2 \log 0.1918} = 1.976$$

$$\hat{d} = \frac{1 - \bar{R}_2}{2\bar{R}^2} = \frac{1 - 0.2046}{2 \cdot 0.1918^2} = 10.81$$

Repare-se que o facto de o coeficiente de concentração (ou comprimento médio do vector resultante) ser relativamente baixo, o desvio padrão $\hat{u} = 1.976$ radianos e o coeficiente de dispersão serem relativamente elevados, são indicações de que a amostra em análise está concentrada em torno da direcção média $\bar{q} = 303.71^\circ$, o que está de acordo com a análise dos histogramas anteriores. Muito embora haja uma nítida tendência de aglomeração em torno da direcção mais predominante (rumo 270°), há contudo alguma concentração de valores dos rumos dos quadrantes Norte e Este, sugerindo que se trata de uma distribuição quase bi-modal, em que as modas estão separadas de aproximadamente 180° .

3. Distribuição estatística da direcção do vento

A função de distribuição de probabilidades mais generalizada em dados de natureza circular, em que os dados apresentam uma simetria em torno de um valor modal é a função de von Mises. Esta distribuição de dados de natureza circular é análoga à função de distribuição normal para dados lineares. Em amostras notoriamente bi-modias, geralmente usa-se uma mis-

tura de funções de distribuição de von Mises (Dobson, 1978; Hill, 1976; Mardia e Jupp, 2000; Upton, 1974).

A função de distribuição de probabilidades de von Mises tem como função de densidade:

$$g(\theta) = \frac{1}{2\pi I_0(k)} e^{k \cos(\theta - \mu)}$$

onde I_0 representa a função Bessel modificada do primeiro género e ordem zero, definida por:

$$I_0(k) = \frac{1}{2\pi} \int_0^{2\pi} e^{k \cos q} dq$$

onde μ e k são a direcção média ($\mu \in [0, 2\pi)$) (e o parâmetro de concentração ($k \geq 0$)), respectivamente. O comprimento do vector resultante A é estimado por \bar{R} em que a função A é definida por:

$$A = \frac{I_1(k)}{I_0(k)}$$

onde I_1 representa a função Bessel modificada de primeiro género e ordem 1, dada por

$$I_1(k) = \frac{1}{2\pi} \int_0^{2\pi} \cos q e^{k \cos q} dq$$

Dada uma amostra aleatória q_1, \dots, q_n , a função de verosimilhança da função de densidade de von Mises é:

$$l(k; q_1, \dots, q_n) = \log \left(\frac{1}{(2\pi)^n I_0(k)^n} \prod_{i=1}^n e^{k \cos(q_i - \mu)} \right) = -n \log(2\pi) - n \log I_0(k) + k \bar{R} \cos(\mu - \bar{q})$$

Como a função $\cos(x)$ tem o máximo para $x = 0$ e o mínimo de máxima verosimilhança de μ é $\hat{\mu} = \bar{q}$ e a estimativa de k é $\hat{k} = A^{-1}(\bar{R})$. Mardia e Jupp (1974) apresentam alguns valores tabelados de $A^{-1}(\bar{R})$ e $A^{-1}(\bar{R})$. Dobson (1978) apresenta diversas expressões para o cálculo aproximado de \hat{k} :

$$\hat{k} \approx 2\bar{R} + \bar{R}^3 + \frac{5}{6}\bar{R}^5, \text{ para } \bar{R} < 0.53$$

$$\hat{k} \approx 0.4 + 1.39\bar{R} + \frac{0.43}{1 - \bar{R}}, \text{ para } 0.53 < \bar{R} < 0.85$$

$$\hat{k} \approx \frac{1}{2} \left(\frac{1}{\bar{R}} + \frac{1}{\bar{R}^2} + \frac{1}{\bar{R}^3} \right), \text{ para } \bar{R} > 0.85$$

$$\hat{k} \approx 0.53 \bar{R}^2 \tan \left(\frac{0.53 \bar{R}}{2} \right)$$

A expressão

$$\left| \frac{A^{-1}(\hat{k}) - k}{A^{-1}(\hat{k})} \right| < 0.032$$

\hat{k} com um erro máximo relativo

Os histogramas apresentados nas figuras 4 e 5 sugerem que a direcção do vento no local e período de tempo considerados seguem uma distribuição bi-modal, com um pico demarcado no rumbo 270° (W) e outro pico menos acentuado no quadrante Este, estando separados por cerca de 180° . Nesta situação,

o modelo estatístico que se ajusta à amostra é uma mistura de duas funções de distribuição de von Mises com o mesmo parâmetro de concentração k^* e modas afastadas de 180° , nas proporções l e $1-l$ (Mardia e Jupp, 2000; Martin et al., 1999). Este modelo estatístico tem a seguinte função de densidade:

$$f(\theta | k, l) = \frac{1}{2\pi I_0(k)} \left[l e^{k \cos(\theta - \mu)} + (1-l) e^{k \cos(\theta - \mu + \pi)} \right]$$

As estimativas \hat{m}^* , k^* e l^* dos parâmetros m , k e l são apresentados em Mardia e Jupp (2000) e Mardia e Zemroch (1975), e são dados por:

$$m^* = \hat{m} = \frac{1}{2} \bar{y}$$

$$1 - \frac{2A}{k^*} = R_2$$

$$1 - \frac{A}{k^*} = \cos \delta + S \sin \delta$$

em que \bar{y} e R_2 são a média e o coeficiente de concentração resultante dos ângulos y_1, \dots, y_n , em que $y_i = 2\alpha_i$.

Os parâmetros anteriormente estimados são $\bar{c} = 0.1064$, $\bar{S} = 0.1595$, $\bar{R} = 0.1918$, $\bar{q} = 0.98244413 \text{ rad} = 303.7^\circ$; a estimativa do parâmetro de concentração é dada

$$k \approx 2\bar{R} + \bar{R}^3 + \frac{5}{6} \bar{R}^5 = 0.3908$$

por

Atendendo à característica bi-modal do histograma, proceder-se-á ao ajustamento de uma mistura de duas funções de distribuição de probabilidades de von Mises, seguindo o algoritmo descrito em Mardia e Jupp (2000) e em Mardia e Zemroch (1975).

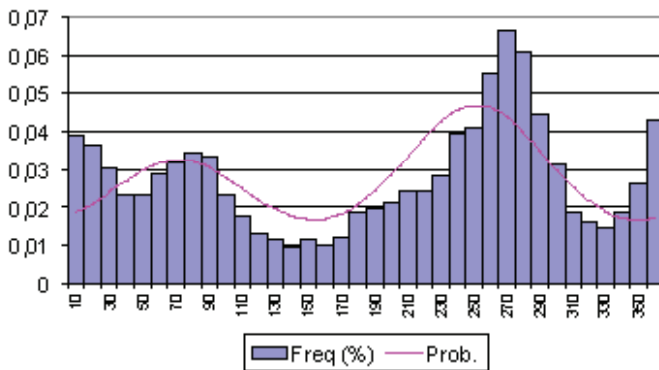


Fig. 7: Histograma de frequências da direcção do vento com a função de distribuição de probabilidades de von Mises ajustada.

Assim, calcularam-se $\hat{m}^* = 1.2533 \text{ rad} = 71.8^\circ$, $\hat{R}_2 = 0.2046$. Usando a expressão

$$1 - \frac{2A}{k^*} = 0.2046, \text{ obtém-se } \frac{A}{k^*} = 0.397$$

pressão. Por interpolação deste valor na tabela 2 (facilmente encontrada na folha de cálculo), obtém-se $k^* = 1.5$. Assim, a proporção de mistura de

funções de densidade de von Mises é de $l = 0.40$.

Tab. 2 - Valores de k , I_0 , I_1 , $\frac{A}{k}$ e $\frac{A}{k}$

k	I_0	I_1	$\frac{A}{k}$	$\frac{A}{k}$
1,0	1,266	0,565	0,446	0,446
1,1	1,326	0,637	0,481	0,437
1,2	1,394	0,715	0,513	0,427
1,3	1,469	0,797	0,543	0,417
1,4	1,553	0,886	0,570	0,407
1,5	1,647	0,982	0,596	0,397
1,6	1,750	1,085	0,620	0,387
1,7	1,864	1,196	0,642	0,378
1,8	1,990	1,317	0,662	0,368
1,9	2,128	1,448	0,681	0,358
2,0	2,280	1,591	0,698	0,349

Na fig. 7 apresenta-se a função de distribuição de probabilidades (mistura de duas funções de densidade de von Mises, centrada na direcção 71.8° , na proporção de, e outra centrada na direcção com a proporção de) ajustada aos dados em estudo. Nota-se que, de um modo geral, esta função traduz com bastante rigor a distribuição real das frequências observadas da direcção do vento da amostra em estudo.

Referências bibliográficas

- Dobson, Annette J. (1978) Simple approximations for the von Mises concentration statistic. *Appl. Statist.*, **27**, 345-347.
- Hill, G.W. (1976) New approximations to the von Mises distribution. *Biometrika*, **63**, 673-676.
- Mardia, K.V. e Jupp, P.E. (2000) *Directional Statistics*. Wiley Series in probability and Statistics, John Wiley & Sons, Chichester.
- Mardia, K.V. e Zemroch, P.J. (1975) Algorithm AS 86: the von Mises distribution function. *Appl. Statist.*, **24**, 268-272.
- Martin, M.; Cremades, L.V. e Santabárbara, J.M. (1999) Analysis and modelling of time series of surface wind speed and direction. *Int. J. Climatol.*, **19**, 197-209.
- Upton, G.J.G. (1974) New approximations to the distribution of certain angular statistics. *Biometrika*, **61**, 369-373.
- Zar, Jerrold H. (1999) *Biostatistical Analysis*. Fourth Edition. Prentice Hall.

¹ Escola Superior Agrária de Castelo Branco, Portugal. armando@eas.ipcb.pt

² Departamento de Estadística, Universidad de Córdoba, Espanha