

Aplicações em meteorologia usando Python:

Material Complementar

Rafael Cesario de Abreu, Natália Machado Crespo,
Rita Yuri Ynoue, Marcia Akemi Yamasoe

Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG-USP)

1 Saldo de radiação

O saldo de radiação Q pode ser descrito da seguinte forma:

$$Q = S + L = S_{UP} + S_{DW} + L_{UP} + L_{DW} \quad (1)$$

Sendo S a componente de radiação de onda curta e L a componente de onda longa. Os subscritos UP e DW referem-se às direções ascendente e descendente, respectivamente. Nesse caso, consideramos um sistema de eixos em que tudo que chega na superfície é positivo e tudo que sai da superfície é negativo. De forma individual podemos representar cada um dos termos indicados no lado direito da Eq. 1 como:

- O termo S_{UP} é dado pela parte que é refletida do que chega na superfície, ou seja, $S_{UP} = -\alpha S_{DW}$, sendo α o albedo da superfície.
- O termo S_{DW} é a irradiância solar que chega na superfície. Consideramos que 54 % da radiação no topo da atmosfera chega de fato na superfície, sendo que os demais 46 % são refletidos por nuvens ou partículas ou absorvidos pelas nuvens, partículas e gases na atmosfera (Peixoto e Oort, 1992). O termo é dado por: $S_{DW} = 0.54(\bar{d}/d)^2 \cos \zeta_0 \epsilon_0$, onde $(\bar{d}/d)^2$ é a distância entre o Sol e a Terra, ζ_0 é o ângulo zenital e ϵ_0 é a constante solar igual a 1360.8 W m^{-2} . O interessante é que com essa equação conseguimos calcular a irradiância no topo da atmosfera utilizando apenas o instante no tempo e as coordenadas geográficas do local. Para mais detalhes sobre o cálculo dessa componente recomendamos a [Apostila da Prof. Dra. Marcia Akemi](#).

- O termo L_{UP} é a radiação de onda longa emitida pela superfície, dada por $L_{UP} = -\epsilon\sigma T^4$, onde ϵ é a emissividade da superfície, $\sigma = 5.67 \times 10^{-8} W m^{-2} K^{-4}$ é a constante de Stefan-Boltzmann e T a temperatura em Kelvin.
- O termo L_{DW} é a radiação de onda longa emitida pela atmosfera para a superfície, dado por $L_{DW} = -\epsilon_a\sigma T^4$, onde ϵ_a é a emissividade da atmosfera. Várias estimativas de ϵ_a são obtidas de forma empírica, como função de temperatura e/ou umidade. Uma das mais simples é a de Swinbank (1963) dada por $\epsilon_a = 9.365 \times 10^{-6} T^2$. Alguns outros métodos são apresentados em Duarte et al. (2006).

A tabela com o albedo e emissividade de algumas superfícies são apresentados na Tabela 1 como exemplo.

Table 1: Emissividade e albedo de diferentes tipos de superfície (Jacobson, 1999).

Tipo de superfície	Emissividade	Albedo
Água líquida	1.0	0.05-0.2
Neve fresca	0.99	0.75-0.95
Neve antiga	0.82	0.4-0.7
Gramma	0.9-0.95	0.16-0.26
Deserto	0.84-0.91	0.2-0.4
Floresta	0.95-0.97	0.1-0.25
Urbano	0.85-0.87	0.1-0.27

2 Análise de Componentes Principais (ACP)

A análise de componentes principais (ACP) ou Principal Component Analysis (PCA) é um método estatístico de redução de dimensionalidade muito popular nas ciências atmosféricas. Com ele é possível representar os principais padrões espaciais e temporais de uma base de dados específica de forma mais concisa. Muitas vezes esses padrões podem ser interpretados fisicamente. Por exemplo, supondo que temos uma variável \mathbf{X} (pode ser temperatura, precipitação ou até mesmo uma combinação dessas variáveis), compostas por n tempos diferentes e p pontos, onde $p = n_y \times n_x$, sendo essas as dimensões na direção y e x , respectivamente (Figura 1).

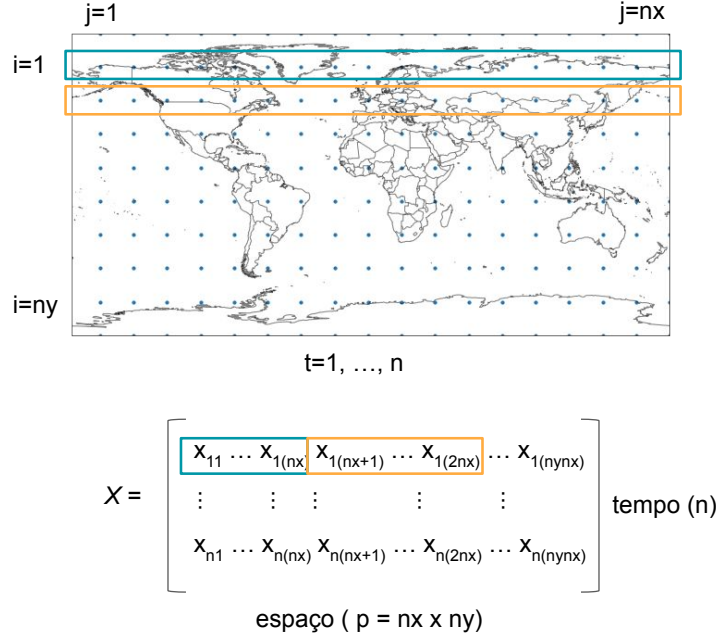


Figure 1: Diagrama da construção da matriz \mathbf{X} para o cálculo da PCA.

Se removermos a média temporal de cada coluna, podemos obter a matriz de covariância da seguinte forma:

$$\Sigma = \frac{1}{n-1} \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ip}x_{i1} & \sum_{i=1}^n x_{ip}x_{i2} & \cdots & \sum_{i=1}^n x_{ip}^2 \end{bmatrix} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} \quad (2)$$

Como Σ é uma matriz quadrada e simétrica, podemos aplicar a decomposição espectral nela, de modo que:

$$\Sigma = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T \quad (3)$$

Onde \mathbf{Q} é uma matriz $(p \times p)$ ortonormal cujas colunas são os autovetores (\mathbf{q}_i) de Σ e $\mathbf{\Lambda}$ é a matriz diagonal com os autovalores de Σ , ou seja:

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p) = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix} \quad (4)$$

Os autovetores \mathbf{q}_i também são chamados de *loadings*; pode-se notar que eles são vetores colunas, com dimensão $p \times 1$, sendo assim possível recuperar um mapa espacial com o padrão

associado a cada componente principal, uma vez que $p = n_y \times n_x$. A primeira componente principal (PC₁) é dada por uma projeção dos dados originais \mathbf{X} no primeiro autovetor:

$$\text{PC}_1 = \mathbf{X}\mathbf{q}_1 = \begin{bmatrix} x_{11}q_{11} + x_{12}q_{12} + \cdots + x_{1p}q_{1p} \\ \vdots \\ x_{n1}q_{11} + x_{n2}q_{12} + \cdots + x_{np}q_{1p} \end{bmatrix} \quad (5)$$

E assim sucessivamente para as demais $p - 1$ componentes. Desta forma, a PC será uma combinação linear das entradas do autovetor associado àquela PC, e terá um tamanho $n \times 1$, ou seja, será uma série temporal de como o padrão definido pelo \mathbf{q}_i varia no tempo. Por construção, definimos que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, e que a variância total de \mathbf{X} , dada por $\text{tr}(\mathbf{\Sigma})$ é:

$$\text{tr}(\mathbf{\Sigma}) = \text{tr}(\mathbf{\Lambda}) = \sum_{i=1}^p \lambda_i \quad (6)$$

Sendo assim, λ_1 representa a componente com a maior variância total da matriz de dados \mathbf{X} , e assim sucessivamente até λ_p , que representa a menor variância total. Dessa forma, com um número $k \ll p$ de componentes podemos representar os principais modos de variância do nosso sinal original \mathbf{X} . Além disso, uma propriedade interessante da matriz de autovetores \mathbf{Q} que é ortonormal, é que temos que o produto escalar $\mathbf{q}_i\mathbf{q}_j = 0$ quando $i \neq j$ e $\mathbf{q}_i\mathbf{q}_i = 1$. Ou seja, os *loadings* são ortogonais entre si e não são linearmente correlacionados. Isso significa que os padrões identificados pelos *loadings* podem representar processos físicos distintos. Para informações mais detalhadas segue como referência Wilks (2011), que possui alguns exemplos de como utilizar ACP nas ciências atmosféricas, e Jolliffe (2005), que aborda alguns temas relacionados à ACP que possam ser de interesse como rotação dos *loadings* e regressão com ACP.

3 Composições

A análise de composições (também conhecida como *composites*) é uma técnica bastante utilizada para determinar características básicas de fenômenos meteorológicos e/ou climáticos. No estudo do clima, a técnica é bastante utilizada pois auxilia no entendimento de fenômenos de grande escala e como esses impactam o globo e/ou regiões específicas. A técnica envolve coletar um grande número de casos de um dado fenômeno; em seguida os casos são compostos juntamente como uma coleção, e então é feita a média dos campos selecionados. Esse campo médio dos eventos selecionados é então comparado com um outro conjunto de eventos que geralmente representam um estado médio como, por exemplo, a climatologia, para verificar

diferenças que sejam significativas. Por exemplo, considerando a variável \mathbf{X} novamente, de forma análoga a seção 2, que pode ser temperatura, precipitação, velocidade do vento etc., selecionamos os k eventos \mathbf{z} no eixo do tempo, onde $\mathbf{z} \in \Theta_1$, ou seja, as datas \mathbf{z} que pertencem ao conjunto definido Θ_1 e podemos comparar eles com os l eventos \mathbf{z} que pertencem ao conjunto Θ_2 :

$$\hat{\mathbf{X}}_{\Theta_1} = \mathbf{E}(\mathbf{X}|\mathbf{z} \in \Theta_1) = \frac{1}{k} \sum_{i=1}^k \mathbf{X}_{\Theta_2,i} \quad (7)$$

$$\hat{\mathbf{X}}_{\Theta_2} = \mathbf{E}(\mathbf{X}|\mathbf{z} \in \Theta_2) = \frac{1}{l} \sum_{j=1}^l \mathbf{X}_{\Theta_2,j} \quad (8)$$

Um exemplo pode ser a coleção de eventos de ciclones ou frentes frias, onde selecionam-se variáveis para caracterizar a estrutura do fenômeno como ventos, temperatura e pressão. Também pode-se fazer uma composição dos campos sinóticos em uma área maior durante um período estendido como, por exemplo, utilizando anos de El Niño no conjunto Θ_1 identificando padrões de precipitação ou temperatura sobre uma determinada região em comparação com os casos de anos neutros Θ_2 .

Um passo além é determinar se os padrões identificados são estatisticamente significativos. Para isso, são utilizados geralmente testes de hipóteses que comparam as médias dos dois conjuntos em cada um dos pontos de grade. Por exemplo, supondo que \mathbf{x}_{Θ_1} é a série de dados do conjunto Θ_1 para um ponto de grade qualquer, e da mesma forma para \mathbf{x}_{Θ_2} , podemos aplicar o teste t para esse ponto, onde a hipótese nula é dada por:

$$H_0 : \hat{\mathbf{x}}_{\Theta_1} = \hat{\mathbf{x}}_{\Theta_2} \quad (9)$$

Para isso é computada a estatística:

$$t = \frac{\hat{\mathbf{x}}_{\Theta_1} - \hat{\mathbf{x}}_{\Theta_2}}{S_p \sqrt{\frac{1}{k} + \frac{1}{l}}} \quad (10)$$

Em que S_p é o desvio padrão conjunto dos dois subconjuntos Θ_1 e Θ_2 , dada por:

$$S_p^2 = \frac{\sum_{i=1}^k (\mathbf{x}_{\Theta_1,i} - \hat{\mathbf{x}}_{\Theta_1})^2 + \sum_{j=1}^l (\mathbf{x}_{\Theta_2,j} - \hat{\mathbf{x}}_{\Theta_2})^2}{k + l - 2} \quad (11)$$

Sob a hipótese nula t tem $k + l - 2$ graus de liberdade. Nesse caso é usada a hipótese de

que \mathbf{x}_{Θ_1} e \mathbf{x}_{Θ_2} possuem a mesma variância S_p , no entanto uma generalização também pode ser usada em que essa hipótese não é levada em consideração:

$$t = \frac{\hat{\mathbf{x}}_{\Theta_1} - \hat{\mathbf{x}}_{\Theta_2}}{\sqrt{\frac{S_{\mathbf{x}_{\Theta_1}}^2}{k} + \frac{S_{\mathbf{x}_{\Theta_2}}^2}{l}}} \quad (12)$$

Onde $S_{\mathbf{x}_{\Theta_1}}$ e $S_{\mathbf{x}_{\Theta_2}}$ são os desvios padrões de \mathbf{x}_{Θ_1} e \mathbf{x}_{Θ_2} , respectivamente. Com o valor de t calculado, podemos consultar a sua distribuição para determinar o p-valor, que nada mais é do que a probabilidade que tenhamos um valor t tão extremo quanto o que encontramos (Figura 2), ou seja:

$$\text{p-valor} = P(|T| > |t|) = P(-|t| > T > +|t|) = 2[1 - P(T < |t|)] \quad (13)$$

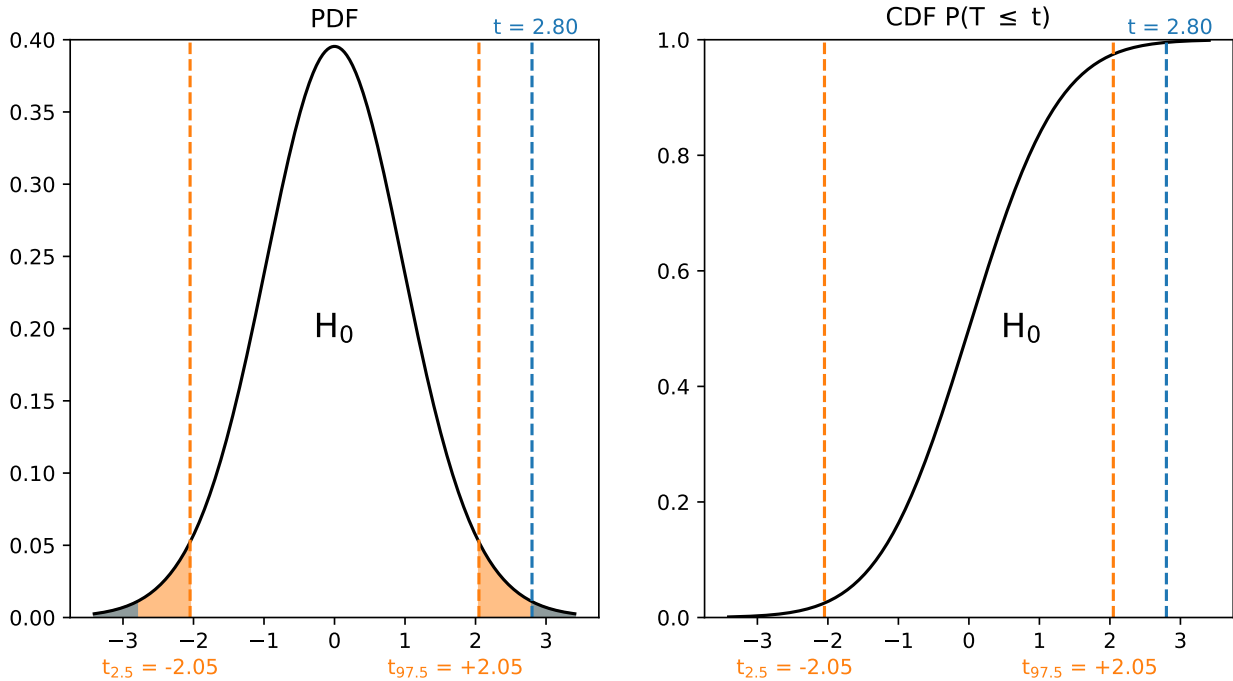


Figure 2: Exemplo de como calcular o p-valor. Ele é dado basicamente pela área em azul na *probability density function* (PDF). A área em laranja representa a área cujo p-valor é de 0.05.

Caso essa probabilidade dada pelo p-valor seja baixa temos mais confiança para rejeitar a hipótese nula, uma vez que dificilmente teremos um valor de t mais extremo do que o encontrado, o que significa que podemos concluir que as médias não são iguais. Normalmente é utilizado um p-valor de 5 % como limite para aceitar ou rejeitar a hipótese nula com um

grau de confiança aceitável, ou seja, se $p\text{-valor} < 0.05$ então rejeitamos H_0 .

Esse teste é então aplicado para cada um dos pontos de grade para identificar as áreas em que as mudanças são mais evidentes. Lembramos que o teste t faz uma hipótese de que os dados do conjunto \mathbf{x}_{Θ_1} e \mathbf{x}_{Θ_2} são provenientes de uma distribuição normal, o que nem sempre é verdade. Em casos em que a distribuição não é normal, é recomendado o uso de testes mais apropriados, como o teste de Mann-Whitney. Para mais informações sobre testes de hipótese recomendamos Morettin e Bussab (2017) e Von Storch e Zwiers (2002) como referências.

References

- Duarte H. F., Dias N. L., Maggiotto S. R., Assessing daytime downward longwave radiation estimates for clear and cloudy skies in Southern Brazil, *Agricultural and forest meteorology*, 2006, vol. 139, p. 171
- Jacobson M. Z., *Fundamentals of atmospheric modeling*. Cambridge university press, 1999
- Jolliffe I., *Principal component analysis*, *Encyclopedia of statistics in behavioral science*, 2005
- Morettin P. A., Bussab W. O., *Estatística básica*. Saraiva Educação SA, 2017
- Peixoto J. P., Oort A. H., *Physics of climate*, 1992
- Swinbank W. C., Long-wave radiation from clear skies, *Quarterly Journal of the Royal Meteorological Society*, 1963, vol. 89, p. 339
- Von Storch H., Zwiers F. W., *Statistical analysis in climate research*. Cambridge university press, 2002
- Wilks D. S., *Statistical methods in the atmospheric sciences*. vol. 100, Academic press, 2011