

Base probabilística para geração de pontos aleatórios e o modelo cosseno

Probabilistic basis for creating random points and the cosine model

Otávio Cristiano Montanher

Universidade Estadual de Maringá, Maringá, PR, Brasil

otaviocmontanher@yahoo.com.br

 <https://orcid.org/0000-0002-1103-6995>

RESUMO

Este artigo aborda diferentes técnicas para geração de pontos aleatórios sobre o globo. A longitude (X) e a latitude (Y) são entendidas como variáveis aleatórias contínuas, contidas nos intervalos entre -180° e 180° e entre -90° e 90° , respectivamente. Duas possibilidades são levadas em conta, ambas considerando essas duas variáveis como independentes. A primeira utiliza distribuições uniformes para as duas variáveis (modelo uniforme) e a segunda utiliza uma distribuição uniforme para a longitude e uma distribuição cosseno para a latitude (modelo cosseno). São apresentadas as funções de distribuição de probabilidade para as distribuições marginais de X e Y, e conjuntas - vetor aleatório (X, Y). Também são evidenciadas as propriedades de esperança e variância para as distribuições marginais, bem como demonstra-se a independência das distribuições conjuntas. Considerando todo o globo, observou-se que o modelo uniforme tende a gerar uma densidade de pontos por área menor na região equatorial e maior nos polos, enquanto o modelo cosseno tende a gerar uma densidade variável ao redor de uma média constante. Assim, o modelo cosseno corrige um enviesamento causado pela geometria não-euclidiana inerente ao formato terrestre. Uma demonstração realizada para duas áreas no território brasileiro mostrou que, utilizando a distribuição cosseno, a porção sul do Rio Grande do Sul teria 14,73% a menos de chance na geração de pontos, em comparação com a região equatorial. Essa diferença ocorre por conta da variação da extensão dos paralelos ao longo das latitudes.

Palavras-chave: Geoprocessamento; Probabilidade; Amostragem espacial.

ABSTRACT

This article addresses different techniques for generating random points on the globe. The longitude (X) and latitude (Y) are understood as continuous random variables, contained within the intervals between -180° and 180° and between -90° and 90° , respectively. Two possibilities are contemplated, both considering these two variables as independent. The first one uses uniform distributions for both variables (uniform model), and the second one uses a uniform distribution for longitude and a cosine distribution for latitude (cosine model). The probability distribution functions are presented for the marginal distributions of X and Y, as well as the joint distribution - random vector (X, Y). The properties of expectation and variance for the marginal distributions are also provided, and the independence of the joint distributions is demonstrated. Considering the entire globe, it was observed that the uniform model tends to generate a lower point density per area in the equatorial region and higher at the poles, while the cosine model tends to generate a variable density around a constant average. Thus, the cosine model corrects a bias caused by the non-Euclidean geometry of the Earth's shape. A demonstration carried out for two areas in the Brazilian territory showed that, using cosine distribution, the southern portion of Rio Grande do Sul would have 14.73% less chance in point generation compared to the equatorial region. This difference occurs due to the variation in the extent of the parallels along the latitudes.

Keywords: Geoprocessing; Probability; Spatial sampling.

1. INTRODUÇÃO

A geração de pontos aleatórios é um procedimento de rotina em geoprocessamento, sendo utilizada em contextos variados, como em estudos de saúde pública (KONDO *et al.*, 2014), no levantamento de mudanças de uso do solo (OLOFSSON *et al.*, 2014) e em geomorfologia (CREMON, 2012). Um exemplo de aplicação dessa técnica é na definição de pontos para visita em campo, coleta de dados e/ou amostras. Gerar um conjunto aleatório de pontos nesse caso garante que o processo de amostragem não seja enviesado por questões de comodidade logística, por exemplo, ou por outros fatores que não possuam relevância para o estudo. Constituem exemplos de amostragens que podem ser realizadas de modo aleatório no espaço: coleta de água para análise de parâmetros limnológicos em um reservatório; determinação de parâmetros fitossociológicos em uma unidade de conservação; coleta de solo para obtenção de parâmetros pedológicos em uma área agrícola; seleção de pessoas e/ou de residências para entrevistas em um levantamento socioeconômico, entre outros.

Outro conjunto de aplicações da geração de pontos aleatórios, muito difundido na comunidade de SIG/sensoriamento remoto, é na validação - avaliação da precisão de classificações baseadas em imagens orbitais e outros produtos de sensoriamento remoto, como modelos digitais de elevação. Nesses casos, os pontos aleatórios têm o objetivo de selecionar pixels ou regiões onde se pretende confrontar a classe a que pertencem com o resultado de uma classificação (OLOFSSON *et al.*, 2014). Embora o processo de geração de pontos aleatórios seja normalmente feito após a elaboração e aplicação de modelos, eles também podem ser utilizados no início, definindo quais locais serviriam para o treinamento.

É de fundamental relevância mencionar que não necessariamente toda amostragem espacial deve ser completamente aleatória. Um procedimento comum em amostragens é a geração de pontos a partir de grades regulares que utilizam distâncias pré-determinadas, o que não constitui o tema desta pesquisa. Também pode haver circunstâncias em que seja necessário realizar algum tipo de estratificação (OLOFSSON *et al.*, 2012), aumentando a quantidade de coletas em uma região com maior variabilidade, por exemplo, e diminuindo em áreas mais homogêneas. Este artigo trata, especificamente, de situações em que há o interesse de se gerar pontos aleatórios, ou seja, pressupõe-se que não há interesse em alguma estratificação espacial no processo de amostragem.

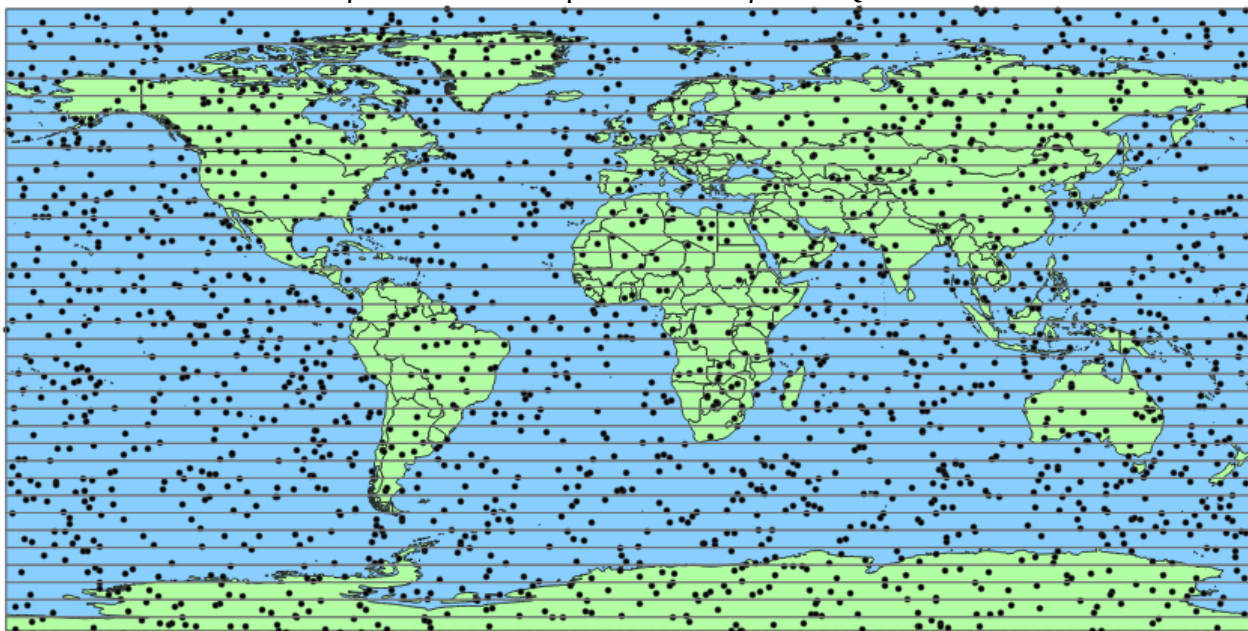
O objetivo é propor um modelo para geração de pontos aleatórios que lide com o sistema de coordenadas geográficas em grandes extensões do globo. Para isso, foram adotados os seguintes passos: i) explicitar o problema que envolve a geração de pontos aleatórios com base em coordenadas geográficas em um meio bidimensional euclidiano; ii) explicitar o modelo tradicionalmente utilizado em termos probabilísticos; iii) apresentar um modelo alternativo e realizar sua descrição probabilística; iv) apresentar uma comparação entre os dois modelos.

Os procedimentos de geoprocessamento foram elaborados com uso do *software* QGIS, um *software* gratuito, de amplo uso nacional (QGIS, 2023). Mais especificamente, o algoritmo utilizado para geração de pontos aleatórios foi o “*Random points in extent*”, presente na versão 3.4 (QGIS, 2020). Os desenvolvimentos matemáticos relacionados com as distribuições de probabilidade dos pontos aleatórios, apresentados neste artigo, seguem princípios básicos de probabilidade que podem ser consultados em livros-texto da área, como em Magalhães e Lima (2005).

1.1. Coordenadas geográficas e pontos aleatórios, qual é a questão?

Para explicar a questão que motivou essa pesquisa, tomou-se o seguinte exemplo: foram gerados 2.000 pontos aleatórios sobre o globo com uso do algoritmo “*Random points in extent*” (**Figura 1**). O mapa-múndi apresentado na **figura 1** apresenta 36 faixas longitudinais, a cada 5° de latitude. Em uma primeira análise, os pontos parecem bem distribuídos sobre o globo, sem haver agrupamentos, isso quer dizer que há aproximadamente a mesma quantidade de pontos em cada uma das faixas latitudinais.

Figura 1: Distribuição de 2.000 pontos aleatórios sobre o globo com base no algoritmo “Random points in extent” presente no software QGIS

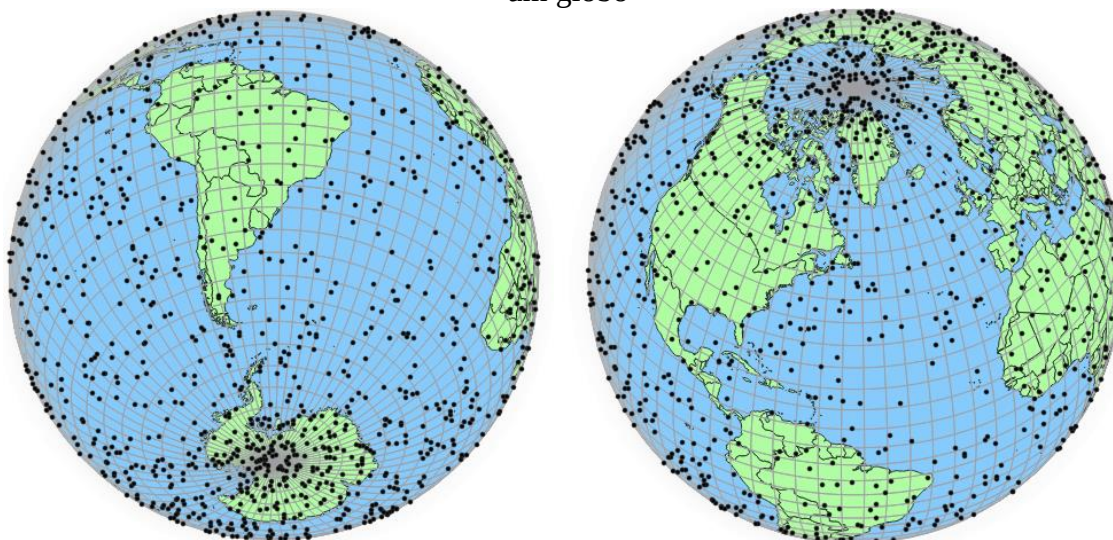


Fonte: Elaborado pelo autor, 2023.

Dada a quase esfericidade terrestre, pode-se afirmar que a região abrangida entre 0° e 5° de latitude possui uma área muito maior do que a região abrangida entre 85° e 90° - a mesma proporção vale para os dois hemisférios. Isso ocorre devido ao fato de que 1° de longitude no Equador equivale a cerca de 111 km e esse comprimento diminui em direção aos polos, chegando a zero exatamente sobre eles (ANDERSON, 1982).

Como as faixas latitudinais possuem aproximadamente a mesma quantidade de pontos, mas áreas bem diferentes, conclui-se que o resultado do modelo estatístico apresentado na **figura 1** para geração de pontos aleatórios privilegia as regiões polares, com maior quantidade de pontos por unidade de área, em relação à região equatorial. Essa afirmação pode ser compreendida visualizando-se a **figura 2**, que mostra a concentração de pontos nos polos e uma menor quantidade de pontos na região equatorial.

Figura 2: Visualização parcial dos mesmos pontos aleatórios apresentados na figura 1, simulando um globo



Fonte: Elaborado pelo autor, 2023.

Em termos matemáticos, a questão abordada nesta pesquisa é que algoritmos como o “*Random points in extent*” tratam o espaço geográfico a partir de uma geometria euclidiana (plano), o que é suficiente para trabalhos em escalas grandes. No entanto, em escalas menores deve-se levar em conta a geometria não-euclidiana, inerente ao formato terrestre (geometria esférica), para a geração de pontos aleatórios. Afirma-se que a Terra apresenta características não-euclidianas por conta da convergência meridiana, que ocorre nos polos (ANDERSON, 1982). Uma introdução aos conceitos de geometria euclidiana e não-euclidiana pode ser consultada em Rocha (2013).

Como a geração de pontos aleatórios é um procedimento frequentemente aplicado em experimentos em escalas grandes e médias, o problema mencionado aqui normalmente não é levado em conta. No entanto, em estudos que envolvam grandes áreas (todo o território brasileiro, um continente, ou todo o globo), a tridimensionalidade terrestre deve ser tratada para a geração não-enviesada de pontos.

2. O MODELO UNIFORME

2.1. Distribuições marginais

Definiu-se a variável X para a longitude, e a variável Y para a latitude, ambas em radianos. Para a produção de pontos aleatórios como os apresentados nas **figuras 1 e 2**, considerou-se que um ponto qualquer sobre o globo é determinado pelas variáveis aleatórias contínuas X e Y. Considerou-se também que todos os valores de X e Y têm a mesma chance de serem selecionados, ou seja, eles seguem distribuições uniformes. Assumiu-se também que a escolha de um valor de X não influencia o valor de Y e vice-versa, ou seja, são variáveis independentes. Assim, tomando-se de forma isolada uma distribuição uniforme para a longitude:

$$f(x) = \begin{cases} \frac{1}{\pi - (-\pi)} = \frac{1}{2\pi}, & -\pi < x < \pi \\ 0, & \text{c. c.} \end{cases} \quad (1)$$

sendo $f(x)$ uma função constante e positiva, e como:

$$\int_{-\pi}^{\pi} \frac{1}{2\pi} dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} 1 dx = \frac{1}{2\pi} [x]_{-\pi}^{\pi} = \frac{1}{2\pi} \cdot 2\pi = 1 \quad (2)$$

a **Equação (1)** descreve uma função de distribuição de probabilidades (f.d.p.). De modo análogo, considerou-se de forma isolada uma distribuição uniforme para a latitude:

$$f(y) = \begin{cases} \frac{1}{\frac{\pi}{2} - (-\frac{\pi}{2})} = \frac{1}{\pi}, & -\frac{\pi}{2} < y < \frac{\pi}{2} \\ 0, & \text{c. c.} \end{cases} \quad (3)$$

sendo $f(y)$ uma função constante e positiva, e como:

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{\pi} dy = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 1 dy = \frac{1}{\pi} [y]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{1}{\pi} \cdot \pi = 1 \quad (4)$$

a **Equação (3)** descreve uma f.d.p. Aplicando-se as propriedades da função de distribuição uniforme, têm-se os seguintes valores de média e variância:

$$E(X) = \frac{-\pi + \pi}{2} = 0 \quad (5)$$

$$E(Y) = \frac{-\frac{\pi}{2} + \frac{\pi}{2}}{2} = 0 \quad (6)$$

$$VAR(X) = \frac{[\pi - (-\pi)]^2}{12} \approx 3,2898 \quad (7)$$

$$VAR(Y) = \frac{\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right]^2}{12} \approx 0,8224 \quad (8)$$

Uma interpretação direta das esperanças é a seguinte: a **Equação (5)** mostra que a média da longitude é no meridiano de *Greenwich*, e a **Equação (6)** mostra que a média da latitude é no paralelo do Equador, sendo resultados muito intuitivos. A **Equação (7)** e a **Equação (8)** mostram os valores das respectivas variâncias, que ao converter para desvio-padrão e para a unidade de graus, resultam em: 103,923° para a longitude e 51,961° para a latitude.

2.2. Distribuição conjunta

A localização de um ponto não é definida apenas pela latitude ou pela longitude, mas a partir das duas. Por isso, para uma descrição teórica mais abrangente da distribuição de probabilidades de pontos aleatórios sobre o globo, é necessário definir a f.d.p. de um vetor aleatório (X, Y). Isso é possível de ser realizado a partir das distribuições marginais devido à independência pressuposta entre X e Y. Assim:

$$f(x, y) = \begin{cases} \frac{1}{2\pi} \cdot \frac{1}{\pi} = \frac{1}{2\pi^2}, & -\pi < x < \pi; -\frac{\pi}{2} < y < \frac{\pi}{2} \\ 0, & \text{c. c.} \end{cases} \quad (9)$$

Sendo $f(x, y)$ uma função constante e positiva, e como:

$$\int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2\pi^2} dy dx \rightarrow \frac{1}{2\pi^2} \int_{-\pi}^{\pi} [y]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dx \rightarrow \frac{1}{2\pi^2} \int_{-\pi}^{\pi} \pi dx \rightarrow \frac{\pi}{2\pi^2} \int_{-\pi}^{\pi} 1 dx \quad (10)$$

$$\frac{1}{2\pi} [x]_{-\pi}^{\pi} \rightarrow \frac{2\pi}{2\pi} = 1 \quad (11)$$

a **Equação (9)** descreve uma f.d.p. Para verificação da independência entre X e Y, tomou-se a covariância:

$$COV(Y) = E(X, Y) - E(X)E(Y) \quad (12)$$

em que, pelas **Equações 5 e 6**, $E(X) = E(Y) = 0$. Assim, a covariância se simplifica em:

$$E(X, Y) = \int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} xy \frac{1}{2\pi^2} dy dx \rightarrow \frac{1}{2\pi^2} \int_{-\pi}^{\pi} x \left[\frac{y^2}{2} \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dx \quad (13)$$

$$E(X, Y) = \frac{1}{2\pi^2} \int_{-\pi}^{\pi} x \cdot 0 dx = 0 \quad (14)$$

assim, a covariância entre X e Y é zero, sendo elas independentes.

3. O MODELO COSSENO

3.1. Distribuições marginais

A abordagem para geração de pontos aleatórios em grandes extensões do globo possui uma diferença no tratamento da latitude apenas, de forma que a longitude permanece tendo uma distribuição uniforme, com as propriedades definidas na **Equação (1)**, na **Equação (5)** e na **Equação (7)**. No modelo proposto também se assume que X e Y são variáveis independentes. O tratamento diferenciado em relação à latitude se baseia no comprimento de um paralelo qualquer, dado por:

$$C = \cos(Y) \cdot E \quad (15)$$

onde C é o comprimento de um paralelo qualquer e E é a extensão do Equador (em unidades de distância). É necessário esclarecer que o modelo proposto pressupõe o formato esférico para a Terra. Uma abordagem mais precisa consistiria no uso de um elipsoide de revolução, cuja solução não está presente neste artigo. Apesar disso, o modelo esférico se aproxima bastante ao modelo elipsoidal, dado o pequeno fator de achatamento, próximo de 1/300 considerando os elipsoides utilizados do Brasil (IBGE, 2016; SAMPAIO; BRANDALIZE, 2018), e à pequena diferença entre áreas, que será mencionada posteriormente. Assim, tomando-se de forma isolada a variável aleatória latitude (Y), é proposto o modelo cosseno:

$$f(y) = \begin{cases} \frac{1}{2} \cos(y), & -\frac{\pi}{2} < y < \frac{\pi}{2} \\ 0, & \text{c. c.} \end{cases} \quad (16)$$

sendo $f(y)$ uma função contínua positiva para o domínio especificado, e como:

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2} \cos(y) dy = \frac{1}{2} [\sin(y)]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{1}{2} \left[\sin\left(\frac{\pi}{2}\right) - \sin\left(-\frac{\pi}{2}\right) \right] = \frac{1}{2} \cdot 2 = 1 \quad (17)$$

a **Equação (17)** descreve uma f.d.p. A esperança de Y é dada por:

$$E(Y) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y \cdot \frac{1}{2} \cos(y) dy \rightarrow \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y \cdot \cos(y) dy \quad (18)$$

que é resolvida por uma integração por partes:

$$E(Y) = \frac{1}{2} [y \cdot \text{sen}(y) + \cos(y)]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \rightarrow \frac{1}{2} \left[\frac{\pi}{2} - \frac{\pi}{2} \right] = 0 \quad (19)$$

de forma que, assim como no modelo uniforme, a média da latitude no caso do modelo cosseno é 0, equivalendo à linha do Equador. Também pode ser demonstrado que:

$$\int_{-\frac{\pi}{2}}^0 \frac{1}{2} \cos(y) dy = \int_0^{\frac{\pi}{2}} \frac{1}{2} \cos(y) dy = \frac{1}{2} \quad (20)$$

o que garante que esse modelo não é enviesado para algum dos hemisférios, ou seja, a chance de um ponto aleatório ser alocado em cada um dos hemisférios (Norte e Sul) é de 50%. Para o cálculo da variância, se utilizou:

$$\text{VAR}(Y) = E(Y^2) - [E(Y)]^2 \quad (21)$$

sendo:

$$E(Y^2) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y^2 \cdot \frac{1}{2} \cos(y) dy \rightarrow \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y^2 \cdot \cos(y) dy \quad (22)$$

que pode ser resolvida por integração por partes em duas vezes:

$$E(Y^2) = \frac{1}{2} \left[y^2 \text{sen}(y) - \int 2y \cdot \text{sen}(y) dy \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \quad (23)$$

$$E(Y^2) = \frac{1}{2} [y^2 \text{sen}(y) - 2(-y \cos(y) + \text{sen}(y))]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \quad (24)$$

$$E(Y^2) = \frac{1}{2} \left[\left(\frac{\pi}{2} \right)^2 - 2 + \left(-\frac{\pi}{2} \right)^2 - 2 \right] = \left(\frac{\pi}{2} \right)^2 - 2 \approx 0,4674 \quad (25)$$

assim:

$$\text{VAR}(Y) = \left(\frac{\pi}{2} \right)^2 - 2 - 0^2 \approx 0,4674 \quad (26)$$

Convertendo-se o valor de variância encontrado para desvio-padrão e para a unidade de graus, encontrou-se o valor de 39,1712°. Comparando-se com o desvio-padrão da distribuição uniforme para a latitude, de 51,961°, pode-se concluir que o uso da função cosseno diminui a variância para a posição latitudinal dos pontos. Isso ocorre devido à maior alocação de pontos próximos à média, que equivale ao paralelo do Equador.

3.2. Distribuição conjunta

De forma análoga ao modelo uniforme para o vetor aleatório (X, Y), a descrição completa

da distribuição de probabilidades de pontos aleatórios sobre o globo a partir do modelo cosseno proposto é:

$$f(x, y) = \begin{cases} \frac{1}{2\pi} \cdot \frac{1}{2} \cos(y) = \frac{1}{4\pi} \cos(y), & -\pi < x < \pi; -\frac{\pi}{2} < y < \frac{\pi}{2} \\ 0, & \text{c. c.} \end{cases} \quad (27)$$

Sendo $f(x, y)$ uma função contínua e positiva para o domínio especificado, e como:

$$\int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{4\pi} \cos(y) dy dx \rightarrow \frac{1}{4\pi} \int_{-\pi}^{\pi} [\text{sen}(y)]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dx \quad (28)$$

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\text{sen}\left(\frac{\pi}{2}\right) - \text{sen}\left(-\frac{\pi}{2}\right) \right] dx \quad (29)$$

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} 2 dx \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} dx \rightarrow \frac{1}{2\pi} [\pi - (-\pi)] \rightarrow \frac{1}{2\pi} \cdot 2\pi = 1 \quad (30)$$

a **Equação (27)** descreve uma f.d.p. Para verificação da independência entre X e Y tomou-se a covariância (**Equação 12**), e como $E(X) = 0$ (**Equação 5**) e $E(Y) = 0$ (**Equação 19**), a covariância se resume em:

$$E(X, Y) = \int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} xy \frac{1}{4\pi} \cos(y) dy dx \rightarrow \frac{1}{4\pi} \int_{-\pi}^{\pi} x \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y \cdot \cos(y) dy dx \quad (31)$$

de modo que, pelas **Equações 18 e 19** demonstrou-se que a integração em Y resulta em zero, então:

$$E(X, Y) = \frac{1}{4\pi} \int_{-\pi}^{\pi} x \cdot 0 dx = 0 \quad (32)$$

assim, a covariância entre X e Y é zero, sendo elas independentes.

4. COMPARAÇÃO ENTRE MODELOS

Realizou-se um experimento em que foram produzidas duas bases de 2.000 pontos aleatórios, uma para cada um dos modelos, sendo eles referidos como uniforme (latitude e longitude seguem distribuições uniformes) e cosseno (longitude segue distribuição uniforme e latitude segue distribuição cosseno). A base do modelo uniforme já foi apresentada na introdução (**Figuras 1 e 2**) para explicação do problema inicial da pesquisa. O modelo cosseno proposto nesse artigo foi implementado em linguagem *Python*, resultando na base apresentada na **figura 3**. Todo o código seria extensivo para apresentação neste artigo, mas algumas partes importantes seguem, com comentários:

```
import numpy    ## Importação de pacote
import math     ## Importação de pacote
```



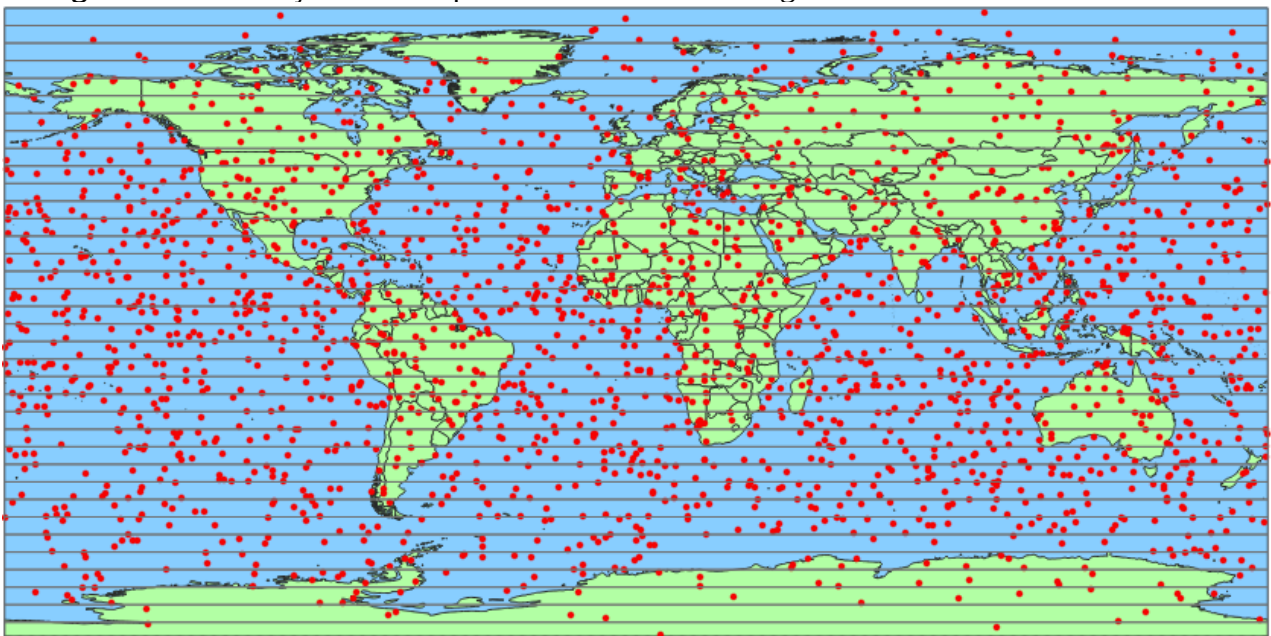
```
import pandas    ## Importação de pacote
x0=-math.pi/2   ## Limita o intervalo mínimo para a latitude
x1=math.pi/2    ## Limita o intervalo máximo para a latitude

def random_custDist():
    samples=[]
    x=numpy.random.uniform(low=x0,high=x1)
    prop=custDist(x)
    if numpy.random.uniform(low=0,high=1) <=prop:
        samples += [x]
def custDist(x):
    return (math.cos(x)/2)
def uniform():
    samples2=[]
    y = numpy.random.uniform(low=-180,high=180)
    samples2 += [y]
df = pandas.DataFrame(samples)
df['Latitude'] = df[0]* (180/math.pi)
```

Cria uma função
Cria um vetor vazio
Define uma variável com dist. uniforme
Obtém o valor da dist. cosseno
Testa a dist. uniforme
Adiciona ao vetor, se satisfeito o teste
Cria uma função
Calcula um valor p/ a dist. cosseno
Cria uma função
Cria um vetor vazio
Calcula um valor p/ a dist. uniforme
Adiciona ao vetor
Cria uma matriz, para exportação
Converte a latitude para graus

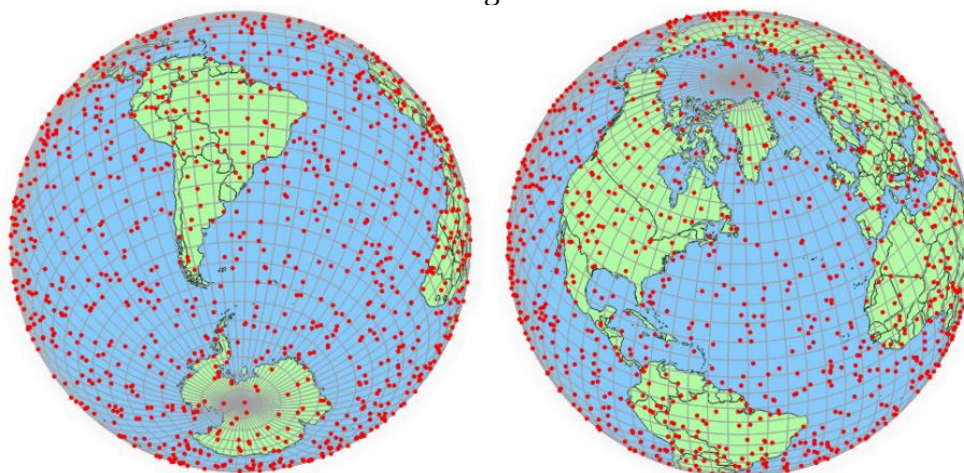
Ressalta-se que foram omitidos detalhes sobre processos de *loop*, exportação, entre outros, mas os pacotes necessários bem como a essência do código estão devidamente apresentados. Uma comparação direta da **figura 3** com a **figura 1** pode mostrar uma aparente aglomeração de pontos na região equatorial, em detrimento dos polos. No entanto, essa percepção é apenas um artefato produzido pela projeção cartográfica, de modo que nesse modelo os polos aparentemente possuem a mesma densidade de pontos em relação ao restante do globo (**Figura 4**). Note que na **figura 4** parece haver uma maior concentração de pontos nas bordas do globo, mas essa impressão também resulta da projeção.

Figura 3: Distribuição de 2000 pontos aleatórios sobre o globo com base no modelo cosseno



Fonte: Elaborado pelo autor, 2023.

Figura 4: Visualização parcial dos mesmos pontos aleatórios apresentados na figura 3, simulando um globo



Fonte: Elaborado pelo autor, 2023.

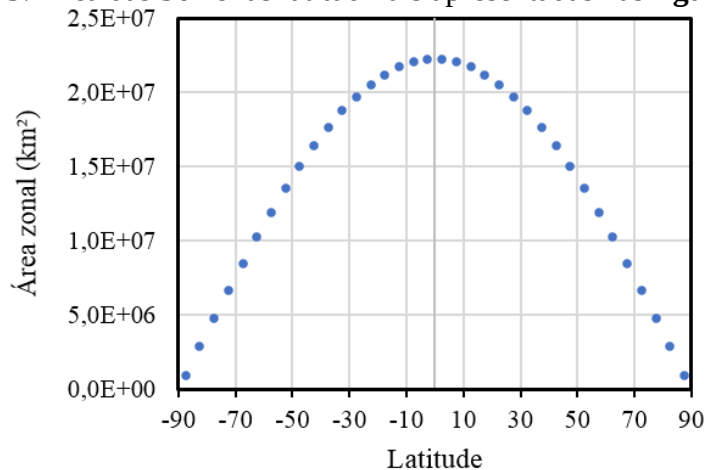
Para uma comparação quantitativa entre os dois modelos, a superfície terrestre foi dividida em 36 faixas latitudinais de 5° (**Figuras 1 e 3**), em que se calculou suas áreas tomando-se o método para calotas esféricas:

$$A = 2\pi rh \quad (33)$$

Onde A é a área, r é o raio terrestre e h é a altura da calota esférica (TAVARES, 2020). Para o raio terrestre adotou-se o valor de 6367,444 km, equivalente à média entre os semieixos maior e menor do elipsoide WGS-84 (IBGE, 2016). O uso desse raio resultou em uma área para a superfície terrestre de 509.495.346,8 km², cerca de 0,112% abaixo da área real da Terra. Assim, primeiramente calculou-se a área total da calota esférica para cada latitude especificada, e a área das zonas latitudinais de 5° foi obtida pela subtração entre calotas. A área superficial das 36 zonas latitudinais pode ser observada na **figura 5**.

Realizou-se também um procedimento de contagem de pontos no interior de polígonos, em que as duas bases de 2.000 pontos foram contabilizadas no interior das zonas latitudinais. Pela área e pela quantidade de pontos calculou-se a densidade de pontos para as duas bases, a partir da razão entre a área e o total de pontos.

Figura 5: Área das 36 zonas latitudinais apresentadas nas **figuras 1 e 3**

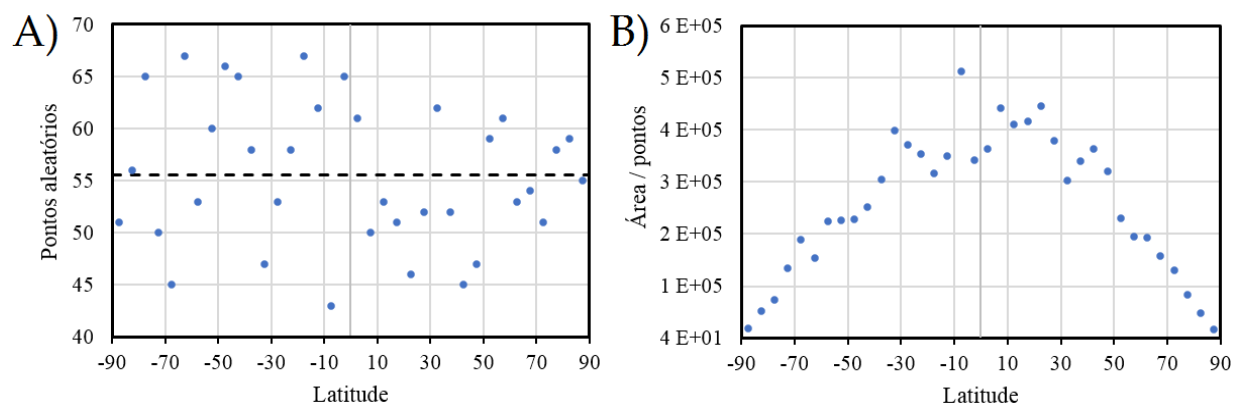


Fonte: Elaborado pelo autor, 2023.

Considerando o modelo uniforme, observa-se que as zonas latitudinais possuem uma quantidade de pontos variável ao redor de uma média constante, indicada pela linha tracejada na **figura 6A**. Por outro lado, a razão área/pontos é claramente maior na região equatorial e menor nos polos (**Figura 6B**). Isso mostra que, gerar pontos aleatórios em mesma quantidade para diferentes latitudes causa uma distorção na densidade de pontos. A linha preta pontilhada na **figura 6A** representa o valor de 55,55 pontos/zona, obtido a partir da razão entre o total de 2000 pontos e as 36 zonas, resultando no valor médio esperado de pontos em cada zona latitudinal. Para averiguar a não-tendenciosidade dos pontos aleatórios, foi realizado um teste-t para amostra simples e média conhecida, considerando-se o valor de 55,55 pontos/zona como a média conhecida. O teste não resultou em diferença significativa entre os dois grupos (valor-p: 0,99).

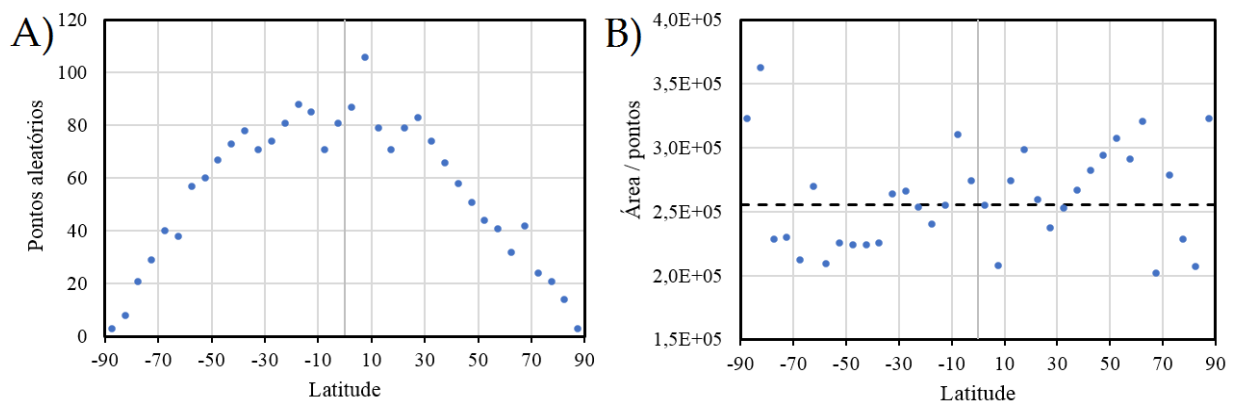
O modelo cosseno proposto resulta em uma distribuição aleatória de pontos com características inversas às apresentadas na **figura 6**. O total de pontos contabilizados em cada faixa latitudinal é menor nos polos e maior na região equatorial (**Figura 7A**), ao passo que a razão área/pontos varia ao redor de uma média constante (**Figura 7B**). A linha preta pontilhada na **figura 7B** representa o valor de 254.747,67 km²/ponto, obtido pela razão entre a área terrestre e a quantidade de pontos, resultando no valor médio esperado de área por ponto. Assim como no caso do modelo uniforme, foi averiguada a não-tendenciosidade dos resultados do modelo cosseno, considerando-se o valor de 254.747,67 km²/ponto como a média conhecida. O teste-t não apontou diferença significativa entre os dois grupos (valor-p: 0,35).

Figura 6: Resultados experimentais para o modelo uniforme. A) Quantidade de pontos em cada zona latitudinal. A linha preta pontilhada indica a média esperada. B) Área em que ocorre um ponto, em média, para cada zona



Fonte: Elaborado pelo autor, 2023.

Figura 7: Resultados experimentais para o modelo cosseno. A) Quantidade de pontos em cada zona latitudinal. B) Área em que ocorre um ponto, em média, para cada zona. A linha preta pontilhada indica a média esperada



Fonte: Elaborado pelo autor, 2023.

Além desses experimentos em nível global foi realizado o cálculo de probabilidades segundo o modelo cosseno para duas zonas localizadas no Brasil (**Figura 8**), a primeira entre 0°S e -1°S e entre -52°O e -53°O (região equatorial). A segunda área se situa entre -31°S e -32°S e entre -52°O e -53°O (Rio Grande do Sul, região subtropical). Observe que as duas áreas possuem 1° × 1° de extensão, mas situadas em contextos distintos de latitude no território nacional.

Figura 8: Localização das áreas experimentais (quadrados alaranjados), em diferentes situações latitudinais



Fonte: Organizado pelo autor, 2023. Fonte dos limites estaduais: IBGE.

Fonte dos limites internacionais: <https://hub.arcgis.com/maps/esri::world-countries-generalized>.

Aplicando-se a **Equação (27)** para a zona equatorial, tem-se:

$$\begin{aligned} \mathbb{P}(X, Y) &= \int_{-0,925}^{-0,907} \int_{-0,01745}^0 \frac{1}{4\pi} \cos(y) dy dx \rightarrow \frac{1}{4\pi} \int_{-0,925}^{-0,907} [\text{sen}(y)]_{-0,01745}^0 dx \\ &\rightarrow \frac{1}{4\pi} \int_{-0,925}^{-0,907} 0,01745 dx \rightarrow \frac{0,01745}{4\pi} \cdot 0,01745 = 2,42395 \cdot 10^{-5} \end{aligned} \quad (34)$$

e para a zona subtropical:

$$\begin{aligned} \mathbb{P}(X, Y) &= \int_{-0,925}^{-0,907} \int_{-0,5585}^{-0,541} \frac{1}{4\pi} \cos(y) dy dx \rightarrow \frac{1}{4\pi} \int_{-0,925}^{-0,907} [\text{sen}(y)]_{-0,5585}^{-0,541} dx \\ &\rightarrow \frac{1}{4\pi} \int_{-0,925}^{-0,907} 0,01488 dx \rightarrow \frac{0,01488}{4\pi} \cdot 0,01745 = 2,06683 \cdot 10^{-5} \end{aligned} \quad (35)$$

Comparando-se a probabilidade de ocorrência de um ponto na região subtropical em relação à região equatorial:

$$\frac{2,06683 \cdot 10^{-5}}{2,42395 \cdot 10^{-5}} = 0,8526 \quad (36)$$

ou seja, a zona de maior latitude possui 14,73% a menos de chance de ocorrência de pontos em relação à zona equatorial. Para se demonstrar que essa diferença é proporcional aos diferentes comprimentos que os paralelos possuem, foi aplicada a **Equação (15)** para se obter o comprimento do paralelo de -31,5°, adotando-se o comprimento de 40075 km para o Equador:

$$C = \cos(-31,5) \cdot 40075 \text{ km} = 34169,6 \text{ km} \quad (37)$$

A proporção entre os comprimentos do Equador e do paralelo de -31,5° é de:

$$\frac{34169,6}{40075} = 0,8526 \quad (38)$$

que é a mesma proporção da calculada entre as probabilidades de ocorrência de pontos nas duas regiões (**Equação (36)**), com pequenas diferenças quando avaliadas em mais casas decimais devido ao fato de que a comparação foi feita entre áreas no caso da **Equação (36)** e entre comprimentos de paralelos na **Equação (38)**. Assim, essa comparação mostra que a menor probabilidade de ocorrência de pontos em maiores latitudes segue exatamente a redução da extensão de seus paralelos.

A proporção de 0,8526 mostra que, em média, para cada 100 pontos aleatórios gerados na região equatorial, seriam gerados cerca de 85 pontos na região de maior latitude do Brasil. Ressalta-se que, caso o modelo uniforme fosse utilizado, as áreas do território de maior latitude teriam uma maior densidade de pontos por área em relação à região equatorial, enviesando o processo de geração de pontos.

5. CONCLUSÃO

O modelo cosseno apresentado foi capaz de gerar uma densidade pontos/área variável, mas ao redor de uma média constante, para todas as regiões do globo, constituindo uma técnica mais adequada para geração de pontos aleatórios em contextos de grandes extensões latitudinais. Considerando todo o território nacional por exemplo, pôde-se demonstrar que há uma diferença não-desprezível entre o Equador geográfico e o sul do estado do Rio Grande do Sul.

Duas sequências podem ser dadas a essa pesquisa, primeiramente, a possibilidade de avançar no modelo teórico, uma vez que o modelo cosseno apresentado supõe um formato esférico para a Terra, o que não condiz com sua real forma. Assim, pode ser desenvolvida uma abordagem com uso de elipsoides de revolução, ajustáveis conforme o modelo terrestre pretendido. Como o cálculo de áreas sobre elipsoides de revolução envolve integrais não triviais, e como a diferença de área e achatamento entre o elipsoide e a esfera é muito pequena, optou-se neste momento pela apresentação deste modelo mais simples. Ademais, outra sequência seria a oportunidade de implementar em SIG, visto que, o algoritmo utilizado nesta pesquisa pode ser adaptado como um *plugin* para o QGIS, por exemplo, ficando disponível para a comunidade usuária. O atual algoritmo, no entanto, não se encontra nesse estágio de desenvolvimento, pois precisam ser elaboradas opções ao usuário, como o distanciamento mínimo entre pontos, e a geração específica para o interior de polígonos e/ou áreas de extensão.

REFERÊNCIAS

ANDERSON, P. S. (ed.) **Princípios de Cartografia Básica**. V. 1. 1982. Disponível em: https://edisciplinas.usp.br/pluginfile.php/4522619/mod_resource/content/1/ANDERSON%201982%20Principios%20de%20Cartografia%20Basica.pdf. Acesso em: 11 out. 2023.

CREMON, E. H. **Caracterização morfológica do sistema fluvial do rio Demini (Amazônia Setentrional) com base em sensoriamento remoto**. 2012. 141 f. Dissertação (Mestrado em

Sensoriamento remoto) – Programa de Pós-graduação em Sensoriamento remoto, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Sistemas de referência**. Rio De Janeiro: IBGE. 2016. Disponível em: https://geoftp.ibge.gov.br/informacoes_sobre_posicionamento_geodesico/sirgas/sisref_2.pdf. Acesso em: 8 jun. 2023.

KONDO, M. C.; BREM, K. D. W.; BARG, F. K.; BRANAS, C. C. A random spatial sampling method in a rural developing nation. **BMC Public Health**, v. 14, n. 338, p. 1-8, 2014.

MAGALHÃES, M. N.; LIMA, A. C. P. **Noções de Probabilidade e Estatística**. São Paulo: Editora da Universidade de São Paulo, 2005.

MORE: **Mecanismo online para referências**, versão 2.0. Florianópolis: UFSC Rexlab, 2013. Disponível em: <http://www.more.ufsc.br/>. Acesso em: 04 jan. 2024.

OLOFSSON, P.; STEHMAN, S. V.; WOODCOCK, C. E.; SULLA-MENASHE, D.; SIBLEY, A. M.; NEWELL, J. D.; FRIEDL, M. A. A global land cover validation dataset, I: Fundamental design principles. **International Journal of Remote Sensing**, v. 33, n. 18, p. 5768–5788, 2012.

OLOFSSON, P.; FOODY, G. M.; HEROLD, M.; STEHMAN, S. V.; WOODCOCK, C. E. W.; WULDER, M. A. Good practices for estimating area and assessing accuracy of land change. **Remote Sensing of Environment**, v. 148, p. 42–57, 2014.

QGIS. **QGIS Documentation: Vector creation**. 2020. Disponível em: https://docs.qgis.org/3.4/en/docs/user_manual/processing_algs/qgis/vectorcreation.html#random-points-in-extent. Acesso em: 5 jun. 2023.

QGIS. **QGIS Geographic Information System**: Open Source Geospatial Foundation Project. 2023. Disponível em: <http://qgis.osgeo.org>. Acesso em: 16 jun. 2023.

ROCHA, R. B. **Geometrias Não-Euclidianas: Proposta de Abordagem Aplicável ao Ensino Básico**. 2013. 74 f. Dissertação (Mestrado profissional em Matemática) – Universidade Federal da Bahia. Salvador, 2013.

SAMPAIO, T. V. M.; BRANDALIZE, M. C. B. **Cartografia geral, digital e temática**. Curitiba: Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas, 2018. 210 p.

TAVARES, M. C. F. P. **Superfícies e sólidos esféricos**. 2020. 128 f. Dissertação (Mestrado profissional em matemática em rede nacional PROFMAT) – Universidade Tecnológica Federal do Paraná – UTFPR. Curitiba, 2020.



Informações sobre a Licença

Este é um artigo de acesso aberto distribuído nos termos da Licença de Atribuição Creative Commons, que permite o uso irrestrito, distribuição e reprodução em qualquer meio, desde que o trabalho original seja devidamente citado.

License Information

This is an open access article distributed under the terms of the Creative Commons Attribution License, which allows for unrestricted use, distribution and reproduction in any medium, as long as the original work is properly cited.