



UNIVERSIDADE D
COIMBRA

Bruno Miguel Charneca Ribeiro

**CARACTERIZAÇÃO DE AEROSSÓIS HUMANOS
COM BASE NA TEORIA DA INFORMAÇÃO**

**Dissertação no âmbito do Mestrado em Engenharia Mecânica, na especialidade de
Energia e Ambiente orientada pelo Professor Doutor Miguel Rosa Oliveira Panão
e apresentada à Faculdade de Ciências e Tecnologia da Universidade de Coimbra
Departamento de Engenharia Mecânica**

Setembro de 2022

1 2



9 0

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

Caracterização de Aerossóis Humanos com base na Teoria da Informação

Dissertação apresentada para a obtenção do grau de Mestre em Engenharia
Mecânica na Especialidade de Energia e Ambiente

Characterization of Human Aerosols based on Information Theory

Autor

Bruno Miguel Charneca Ribeiro

Orientador

Professor Doutor Miguel Rosa Oliveira Panão

Júri

Presidente

**Professor Doutor José Manuel Baranda Moreira da Silva
Ribeiro**

Orientador

**Professor da Universidade de Coimbra
Professor Doutor Miguel Rosa Oliveira Panão
Professor da Universidade de Coimbra**

Vogal

**Professor Doutor Jorge Campos da Silva André
Professor da Universidade de Coimbra**

Coimbra, setembro, 2022

Every bit of learning is a little death. Every bit of new information challenges previous conception, forcing it to dissolve into chaos before it can be reborn as something better.

Jordan B. Peterson

À minha família.

Agradecimentos

Esta dissertação é o culminar de 5 anos de esforço e dedicação, sendo apenas possível com o apoio de várias pessoas e entidades, às quais não poderia deixar passar a oportunidade de dar o devido reconhecimento.

Ao meu orientador Miguel Rosa Oliveira Panão

Pelo acompanhamento, por todos os conselhos e pela partilha de conhecimentos essenciais para terminar esta etapa.

A Coimbra, à Universidade e Departamento de Engenharia Mecânica

A cidade e instituição que me acompanhou nos últimos 5 anos da minha vida, pelas memórias, pelos ensinamentos, pelas oportunidades e pela pessoa que me tornou, um obrigado.

A toda a minha família

Por me apoiarem com tudo o que eu precisei na minha vida e pela confiança e responsabilidade que me foi depositada.

Aos meus amigos e namorada

Por estarem presentes nos pontos mais altos e nos mais baixos, nos momentos de diversão e os de estudo, e acima de tudo por tornarem esta jornada especial, pois sem eles este percurso não teria sido o mesmo.

A todos aqueles que fizeram parte desta jornada única

Aos que posso ter cometido a injustiça de não ter referido, aos que tiveram presentes, aos que o meu caminho já não irá cruzar, obrigado por todos os momentos, ajuda, ensinamentos e pela vossa presença.

Resumo

Desde 2020 que o mundo enfrenta a Covid-19, uma doença respiratória infecciosa que se propaga através de aerossóis humanos, alterando os nossos hábitos e estilos de vida, criando impactos devastadores a níveis socioeconómicos que se prolongam mesmo após um possível fim da pandemia. Todas as alterações necessárias no nosso dia-a-dia para mitigar a propagação desta doença foram o resultado de investigação científica em diversas áreas do saber. O facto da propagação do SARS-Cov-2 estar associada à carga viral transmitida por aerossóis, a sua caracterização serve de definição e melhoramento das estratégias de proteção e planeamento do controlo da propagação do vírus mitigando o seu efeito pandémico. Porém, a informação sobre os tamanhos das partículas aerossóis está ainda dispersa pela literatura e a sua modelação é ambígua. A partir da caracterização dos aerossóis reportada na literatura, esta dissertação contribui com: um estudo da modelação das distribuições de gotas através do ajuste de funções matemáticas de densidade de probabilidade; e um novo método para aferir o ajuste baseado no conceito de entropia diferencial, proveniente da teoria da informação. A entropia diferencial comprova que as melhores funções para caracterizar e modelar as distribuições dos aerossóis são as funções da distribuição Normal Logarítmica. Esse resultado mostra que o mecanismo de produção de aerossóis produzidos por humanos em diversas situações (tosse, espirro, falar, entre outras) é o resultado de uma constante taxa de crescimento da diversidade de tamanho de gotas, como expressão do sentido físico do desvio padrão da função de distribuição Normal Logarítmica. No caso dos aerossóis humanos, os resultados reportados na literatura evidenciam que o aumento dessa taxa de crescimento está correlacionado com um maior valor de diâmetro médio dos aerossóis produzidos e um maior grau de diversidade de gotas do *spray*. Os ajustes apresentados são úteis para os modelos de simulação numérica da dispersão de aerossóis.

Palavras-chave: Aerossóis, Teoria da Informação, Entropia Diferencial, Distribuição Normal-Logarítmica, Diversidade de Tamanhos de Gotas

Abstract

Since 2020, the world has faced Covid-19, an infectious respiratory disease that spreads through human aerosols, altering our habits and lifestyles, creating devastating impacts at socio-economic levels that extend even after a possible end to the pandemic. All the changes necessary in our daily life to mitigate the spread of this disease were the result of scientific research in several areas of knowledge. The fact that the spread of SARS-Cov-2 is associated with viral load transmitted by aerosols, its characterization serves as a definition and improvement of strategies for the protection and planning of the control of the spread of the virus, mitigating its chemical effect. However, the information on the sizes of aerosol particles is still dispersed throughout the literature and their modeling is ambiguous. From the characterization of aerosols reported in the literature, this dissertation contributes with a study of the modeling of droplet distributions through the adjustment of mathematical functions of probability density and a new method to measure the adjustment based on the concept of differential entropy derived from information theory. Differential entropy proves that the best function to characterize and model aerosol distributions is the normal logarithmic distribution function. This result shows that the mechanism of production of aerosols produced by humans in various situations (cough, sneezing, talking, among others) is the result of a constant growth rate of the diversity of droplet size, as an expression of the physical sense of the standard deviation of the Normal Logarithmic distribution function. In the case of human aerosols, the results reported in the literature show that the increase in this growth rate is correlated with a higher average diameter value of the aerosols produced and a greater degree of diversity of *spray* drops. The adjustments presented are useful for numerical simulation models of aerosol dispersion.

Keywords: Aerosols, Information Theory, Differential Entropy, Log-Normal Distribution, Drop Size Diversity

Índice

Índice de Figuras	ix
Índice de Tabelas.....	xi
Nomenclatura	xiii
Símbolos Gregos	xiv
Siglas	xv
1. Introdução	1
1.1. Aerossóis e a sua Caracterização	2
1.1.1. Distribuições de Tamanhos de Gota num Aerossol	3
1.1.2. Diversidade de Tamanhos de Gota num Aerossol	6
1.1.3. Ajuste de uma Função Matemática de Densidade de Probabilidade a um Histograma	7
1.2. Teoria da Informação e Entropia Diferencial.....	11
1.3. Objetivos	14
2. Metodologia	15
2.1. Simulação dos Dados e Ajuste.....	16
2.2. Entropia Diferencial como Critério de Aferição do Ajuste.....	18
3. Apresentação e Discussão de Resultados.....	21
3.1. Caracterização de Aerossóis Humanos	21
3.2. Aferição da Entropia Diferencial como Critério de Ajuste.....	26
3.3. Discussão dos Resultados segundo a Origem Física da Log-Normal.....	29
4. Conclusão	31
Referências Bibliográficas	33
Anexo	35

ÍNDICE DE FIGURAS

Figura 1.1 Comparação entre os resultados obtidos a partir de uma análise usando o método da Teoria da Informação ou não. Ponto de medição ($r = 0$ mm). (a) Tamanho da gota e número de amostras entre a Teoria da Informação e todos os dados medidos. [Adaptado de M. R. O. Panão, 2012].....	2
Figura 1.2 Fluxograma da Caracterização de Aerossóis	3
Figura 1.3 Distribuições de tamanho de gota ponderadas por volume para acetona (vermelho) e água (azul) comparando amplitudes de classe regulares (Linha fina) com variáveis (Linha grossa) para um fluxo volumétrico de $V_{fr} = 5\text{mL}/\text{min}$ e o mesmo número de classes $N_k = 110$ para ambos [Adaptado de Panão, 2021]	5
Figura 1.4 Exemplo do ajuste de uma função cumulativa para os resultados de água destilada, $r=0\text{mm}$ em que a melhor função é a Log-Normal e $r=6\text{mm}$ com a melhor função de ajuste sendo a Gamma. [Adaptado de M. O. Panão <i>et al.</i> , 2020]	8
Figura 1.5 Ajuste de uma função de densidade de probabilidade Log-Normal a uma amostra de dados discretos. [Adaptado de Tolis & Rentizelas, 2011]	9
Figura 1.6 Ajuste de uma distribuição Rosin-Rammler de (Dbouk & Drikakis, 2020) aos dados experimentais para a distribuição de gotas produzidas por saliva em (Xie <i>et al.</i> , 2009).	10
Figura 1.7 Ajuste da distribuição Rosin-Rammler/Weibull indicada por (Dbouk & Drikakis, 2020) e a distribuição de gotas providenciada por (Xie <i>et al.</i> , 2009).	11
Figura 2.1 <i>Flowchart</i> do processamento dos dados	15
Figura 2.2 Categorização de aerossóis humanos.....	16
Figura 2.3 Dados extraídos da literatura do número de gotas e o seu respetivo tamanho de gota para o caso [Espirros, Múltiplos, 15,24 cm]. (Duguid, 1946).....	17
Figura 2.4 Comparação entre a distribuição original de (Duguid, 1946) para espirros múltiplos e a obtida por simulação.	17

Figura 2.5 Comparação entre as distribuições cumulativas para espirros múltiplos (Duguid, 1946).....	19
Figura 3.1 Comparação do diâmetro representativo de 50% da distribuição cumulativa dos dados simulados com o obtido através das funções de ajuste.	21
Figura 3.2 Grau de Polidispersão dos diversos aerossóis humanos de acordo com a distância a que ocorre a captura dos dados.	22
Figura 3.3 Grau de Heterogeneidade dos diversos aerossóis humanos de acordo com a distância a que ocorre a captura dos dados.	23
Figura 3.4 Resultados do teste de KS para cada função ajustada aos dados simulados. Ocorrências de 1-17 com leitura através da Figura 2.2.	24
Figura 3.5 Relação entre o diâmetro característico dos aerossóis humanos simulados com o desvio padrão da distribuição normal logarítmica.	26
Figura 3.6 Comparação da entropia diferencial original com a do aerossol simulado (esquerda) e entre o aerossol simulado e da função contínua testada pelo ajuste (direita)	27
Figura 3.7 Desvio percentual entre a entropia diferencial da distribuição da função contínua de densidade de probabilidade e a do aerossol simulado.	28
Figura 3.8 Distribuição dos ajustes que aproximaram mais a entropia diferencial da função matemática (Log-normal, Gamma e Weibull) à do aerossol simulado.	28
Figura 3.9 Demonstração do efeito de classes regulares vs variáveis em dados simulados nas funções de ajuste Log-Normal e Gamma. LN (2,335; 0,7864) & GM (1,7671; 1,8001).	29
Figura 3.10 Análise da relação da taxa de crescimento da diversidade de tamanho de gota com o grau de heterogeneidade, Diâmetro de Sauter e Brouckere.	30

ÍNDICE DE TABELAS

Tabela 2.1. Valores dos parâmetros característicos das funções ajustadas aos dados simulados fornecidos pelo algoritmo.	20
Tabela 2.2. Entropias Diferenciais dos dados originais; simulados; funções de ajuste e comparação entre simulados e funções de acordo com o teste de KS.	20
Tabela 0.1. Distribuição de tamanhos de gotas grandes(Duguid, 1946).....	35
Tabela 0.2. Distribuição de tamanhos de <i>Droplet-Nuclei</i> (Duguid, 1946)	35
Tabela 0.3. Tabela de distribuição de tamanho de gota composto (Duguid, 1946).....	36
Tabela 0.4. Valor médio do número de gotas medida por pessoa nas distâncias de 1 e 6 cm.(Chao <i>et al.</i> , 2009).....	37

NOMENCLATURA

D – Critério de KS

D_{50} – Mediana da função de distribuição cumulativa

$F(d_k)$ – Distribuição Cumulativa Discreta

F_k – Função Distribuição Cumulativa

H – Entropia de Shannon / Quantidade de Informação

$H_{n,v}$ – Entropia de Shannon Normalizada

H_{dif} – Entropia Diferencial

N – Dimensão da Amostra

N_k – Número de Classes

a – Parâmetro de Forma

b – Parâmetro de Escala

\bar{d} – Diâmetro Médio

$d_{1,0}$ – Diâmetro médio probabilístico / Valor Médio da Distribuição Simples

$d_{3,2}$ – Diâmetro médio de Sauter

$d_{4,3}$ – Diâmetro médio de Brouckere ou Herdan

d_k – Centroide / Diâmetro que representa a classe

d_m – Diâmetro Médio Log-Normal

d_{pq} – Média de diâmetros dos aerossóis

e – Número de Neper

$f_{w,k}$ – Frequência no Histograma de cada Classe k

$f_i(d)$ – Frequência

f_k – Frequência Relativa

n_f – Parâmetro de Forma

n – Ponderado por Número

n_k – Número de gotas em cada classe

p – Função Densidade de Probabilidade

$p(d)$ – Expressão de Rosin-Rammler

p_d – Densidade de Probabilidade

s – Ponderado por Área

v – Ponderado por Volume

SD_v – Desvio padrão ponderado por volume

Δd – Amplitude da Classe

Símbolos Gregos

γ - Constante de Euler-Mascheroni

σ – Desvio padrão da função de distribuição Log-Normal / Taxa de crescimento da diversidade de tamanho de gota

$\Phi(d_k)$ – Valor da Função de Distribuição Cumulativa Contínua

$\psi(a)$ – Função Digamma

Siglas

DTG – Diversidade de Tamanho de Gota

GM – Gamma

IMI – Interferometric Mie Imaging / Interferometria Imagiológica de Mie

KS – Kolmogorov-Smirnov

LN – Log-Normal

PDF – Probability Density Function / Função de Densidade de Probabilidade

PDI – Phase Dopple Interferometer / Interferometria Laser de Fase Doppler

WB – Weibull

1. INTRODUÇÃO

As doenças do foro respiratório, como a Covid-19, provocada pela infeção do coronavírus, SARS-CoV-2, ameaçam a vida e o bem-estar de pessoas em todo o mundo. A gripe espanhola, causou mais de mil milhões de infeções e foi considerada como a pandemia de gripe mais fatal do séc. XX. A Covid-19 infetou mais de 500 milhões de pessoas em todo o mundo e causou a morte a mais de 6 milhões. Anualmente, são reportadas quase 4 milhões de mortes devido a doenças respiratórias infecciosas e 1,5 milhões de mortes devido a tuberculose (Han *et al.*, 2013).

Atualmente, as doenças respiratórias infecciosas podem ser transmitidas pelo ar, através de contactos diretos ou indiretos. O contacto direto ocorre através de gotas de tamanho relativamente grande ($\sim 10-100 \mu\text{m}$), produzidas por tosse, espirros ou fala, cuja carga viral pode infetar caso haja um contacto próximo ($\sim 1 \text{ m}$), ou gotículas ou aerossóis com um tamanho relativamente pequeno ($\sim 1 \mu\text{m}$) que permanecem em suspensão no ar durante longos intervalos de tempo. Já o contacto indireto, pode ocorrer quando as gotas se depositam nas superfícies, infetando-as (Han *et al.*, 2013).

A principal fonte de contágio de doenças respiratórias por transmissão aérea ocorre em ambientes interiores e mal ventilados, aumentando a probabilidade de inalação de gotas produzidas por uma pessoa infetada ao tossir, espirrar ou falar. A caracterização do tamanho das gotas é fundamental para prever como será a sua evaporação ao longo do tempo e alcance, uma vez que gotas de tamanhos diferentes interagem com o ar ambiente de maneira diferente, influenciando o desenvolvimento de estratégias de proteção, incluindo sistemas de ventilação e planeamento do controlo de infeções (Han *et al.*, 2013).

1.1. Aerossóis e a sua Caracterização

Um aerossol é constituído por partículas sólidas ou líquidas em suspensão num ambiente gasoso. As partículas formadas têm tamanhos e velocidades diferentes dependendo do processo de vaporização e técnicas de diagnóstico laser como a Interferometria Laser de Fase Doppler (*Phase Doppler Interferometer*, PDI), em que são medidos parâmetros como o tamanho, velocidade, densidade numérica e fluxo, em pontos dentro do aerossol permitem medir esses parâmetros com suficiente resolução temporal (Panão *et al.*, 2020).

Como a quantidade de gotículas emitidas por uma pessoa não se compara com *sprays* produzidos por atomização de líquidos, o processo de aquisição das características das gotas pode tornar-se moroso pela baixa taxa de aquisição. Nesse sentido, a teoria da informação iniciada a partir do trabalho de Claude Shannon em 1948 permitiu desenvolver um método de aferir a qualidade da informação obtida, distinguindo dados de informação capturada por técnicas de interferometria laser, oferecendo ao experimentalista um critério de paragem na aquisição mais fidedigno (Panão, 2012). Por exemplo, a Figura 1.1 mostra que através da abordagem da Teoria da Informação, amostras maiores não implicam maior quantidade de informação para obter, posteriormente, uma análise estatística fiável dos dados adquiridos.

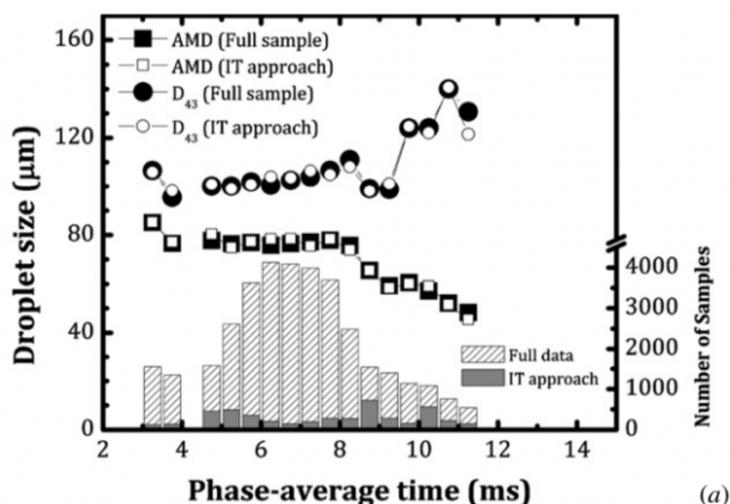


Figura 1.1 Comparação entre os resultados obtidos a partir de uma análise usando o método da Teoria da Informação ou não. Ponto de medição ($r = 0$ mm). (a) Tamanho da gota e número de amostras entre a Teoria da Informação e todos os dados medidos. [Adaptado de M. R. O. Panão, 2012]

Nesta dissertação, iremos abordar a análise estatística associada à caracterização de aerossóis, mais como forma de organizar dados, em que a probabilidade associada a uma certa classe de tamanhos é uma probabilidade de presença, sendo essa linguagem crucial para entender o conceito de variação da diversidade de tamanhos de gotas. Pois, a probabilidade de cada classe representa o grau de relevância dessa classe no aerossol (Panão, 2021a; Panão *et al.*, 2020). O esquema da Figura 1.2 sintetiza a abordagem seguida na caracterização de dados de um aerossol.

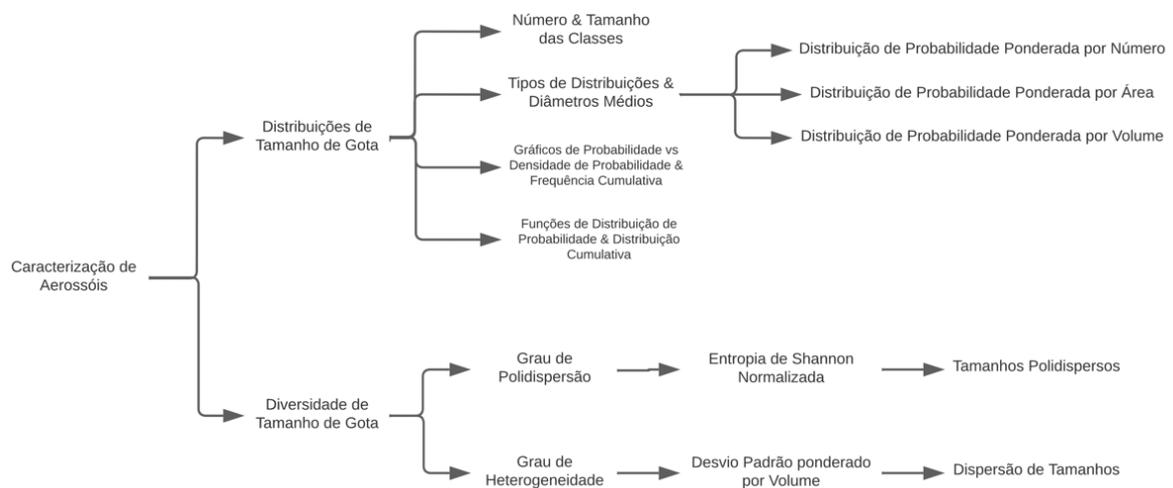


Figura 1.2 Fluxograma da Caracterização de Aerossóis

1.1.1. Distribuições de Tamanhos de Gota num Aerossol

A construção de distribuições de tamanhos de gotas é relevante para a percepção humana dos dados que se pretende apresentar (Sahann *et al.*, 2021), o que implica definir o número adequado de classes N_k mediante o tamanho da amostra e a ordem de valores entre os dados, bem como a decisão de apresentar essa distribuição em frequência, densidade de probabilidade ou em formato cumulativo. Se d_k for a amplitude que representa a classe com uma frequência de relativa f_k , então, considerando todas as classes: $\sum_{i=1}^{N_k} f_i = 1$. Para obter a densidade de probabilidade p_d que permite comparar distribuições construídas de forma diferente (diferente número de classes N_k), sabendo que Δd [μm] corresponde à amplitude de cada classe, $p_d = f_k / \Delta d$ [μm^{-1}]. A função de distribuição cumulativa obtém-se a partir das frequências como $F_k = \sum_{i=1}^k f_i$ variando entre 0 e 1. A primeira questão que se coloca na representação das distribuições de tamanhos de gotas refere-se ao tamanho das classes que pode ser regular ou variável.

1.1.1.1. Número e Tamanho das Classes

De forma a entender o procedimento mais indicado para a construção das classes dos histogramas é preciso abordar a escolha do número de classes e se terão tamanho regular ou variável. Por outro lado, haverá um tamanho mínimo para uma classe? Em Panão (2022), a questão da redundância na parte decimal das medidas das gotas aponta para a irrelevância de se considerar tamanho de classes inferiores a uma unidade da escala usada para os diâmetros. Assim, se a distribuição apresenta os diâmetros em micrómetros (μm), uma classe com tamanho inferior a 1 μm entraria no domínio da redundância da parte decimal associada a cada tamanho de uma gota. Dogan & Dogan (2010), sintetizaram diversas estratégias para a determinação do número de classes baseadas na dimensão da amostra (N), gama dos dados, momentos da distribuição ou intervalo interquartil, mas uma avaliação de Panão (2022), sobre qual o número de classes que mais aproximava o histograma construído com uma simulação de gotas da distribuição conhecida, obtendo a seguinte relação empírica, com $R^2 = 0,995$,

$$N_k(N) = \text{round}\left[\exp\left(3,8256 - \frac{53,726}{N} + 0,099258 \log(N)\right)\right] \quad (1.1)$$

Enquanto as classes de tamanho regular fixam a amplitude de classe (Δd), as classes de tamanho variável mantêm constante a relação entre a amplitude da classe e o seu centroide (d_k).

$$\frac{\Delta d}{d_k} = C \quad (1.2)$$

Panão *et al.* (2020), propõe que C seja,

$$C = N_k^{-1/2} \quad (1.3)$$

Na Figura 1.3 retirada de Panão (2021a), pode-se observar como classes regulares tendem a criar histogramas com maior variabilidade e “ruído” do que as classes com tamanho variável, mesmo em casos de bimodalidade da distribuição.

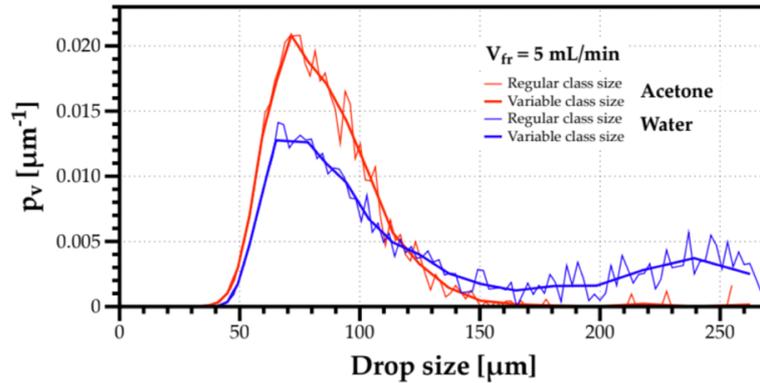


Figura 1.3 Distribuições de tamanho de gota ponderadas por volume para acetona (vermelho) e água (azul) comparando amplitudes de classe regulares (Linha fina) com variáveis (Linha grossa) para um fluxo volumétrico de $V_{fr} = 5 \text{ mL/min}$ e o mesmo número de classes $N_k = 110$ para ambos [Adaptado de Panão, 2021]

1.1.1.2. Tipos de Distribuições de Tamanho de Gota e Diâmetros Médios

Na caracterização de um aerossol, o tipo de distribuição para o tamanho das gotas construído varia de acordo com a questão de investigação (Panão *et al.*, 2020). Isto é, a frequência no histograma em cada classe $k - f_{w,k}$ - estaria afetada de um fator de ponderação que depende da distribuição ser de número ($w = n$), área ($w = s$) ou volume ($w = v$). Por exemplo, se a questão da investigação está mais focada nos tamanhos em si mesmos, não é necessária qualquer ponderação senão o número das gotas medido em cada classe. Logo, a frequência obtida em cada classe k corresponde a

$$f_{n,k} = \frac{n_k}{N} \quad (1.4)$$

Com $f_{n,k}N = \sum_{i=1}^{N_k} n_i$. Mas se a questão de investigação se relacionasse com a área de transferência de calor, faria mais sentido apresentar uma distribuição ponderada pela área de superfície das gotas com frequência de

$$f_{s,k} = \frac{\pi n_k d_k^2}{\sum_i (\pi n_i d_i^2)} \quad (1.5)$$

E se o interesse estivesse focado no volume de *spray*, então, a distribuição deveria ser ponderada pelo volume com frequência em cada classe obtida como

$$f_{v,k} = \frac{\left(\frac{\pi}{6}\right)n_k d_k^3}{\sum_i \left(\frac{\pi}{6}\right)n_i d_i^3} \quad (1.6)$$

Neste sentido, também o primeiro momento destas distribuições adquire um significado físico mais apropriado. Sabendo que esse é obtido em geral como

$$d_{pq} = \left(\frac{\sum_i n_i d_i^p}{\sum_i n_i d_i^q} \right)^{\frac{1}{p-q}} \quad (1.7)$$

O diâmetro médio probabilístico – d_{10} – é um dos parâmetros característicos para se caracterizar um spray. Ao usar este valor na caracterização estamos a assumir que todas as gotas têm o mesmo tamanho. A média de uma distribuição ponderada pela área – d_{32} – é uma medida representativa da relação entre a escala do volume da gota, por exemplo, vaporizada num processo de combustão, e a escala da superfície da gota por onde ocorre a transmissão de calor. O d_{32} é também um parâmetro associado à eficiência da atomização de um *spray* (Panão, 2021). Além dos valores médios para os diâmetros, dada que as gotas possuem diferentes tamanhos e nem todos possuem a mesma probabilidade de presença, importa quantificar essa diversidade de tamanho de gotas.

1.1.2. Diversidade de Tamanhos de Gota num Aerossol

O conceito de Diversidade de Tamanho de Gota (DTG) num aerossol auxilia a resposta a duas questões: 1) “Quantos tamanhos diferentes são relevantes?”; 2) “Quão diferentes são os tamanhos relevantes?”. A palavra “relevantes” remete para a ideia de que a probabilidade de presença de um certo tamanho de gotas sobressair-se relativamente a outros tamanhos. Mediante a resposta às duas questões colocadas, o conceito DTG é medido através de dois graus (Panão *et al.*, 2020):

O grau de polidispersão → Quantifica a variedade de diferentes tamanhos que são relevantes no aerossol.

O grau de heterogeneidade → Quantifica o quão diferentes são os tamanhos relevantes no aerossol.

O principal desafio consiste em escolher os indicadores corretos para medir cada um destes graus de DTG. O indicador mais conhecido e usado para o grau de polidispersão é o *Relative Span*, obtido a partir de diâmetros representativos de uma distribuição cumulativa dos tamanhos de gota. No entanto, Panão *et al.* (2020), argumentam sobre o desuso limitado deste conceito, favorecendo o uso da entropia de Shannon normalizada como método de quantificar o grau de polidispersão.

$$H_{n,v} = \frac{-\sum_{j=1}^{N_k} (p_{v,j} \ln(p_{v,j}))}{\ln(N_k)} \quad (1.8)$$

A terminologia da teoria da informação aplicada a caracterização de aerossóis diz-nos que este valor se anula para casos com todas as gotas do mesmo tamanho e tem valor máximo quando todos os tamanhos têm a mesma relevância, ou seja, a mesma probabilidade de estarem presentes na amostra de gotas do aerossol.

E no caso do grau de heterogeneidade, apontam o desvio padrão ponderado por volume.

$$SD_v = \sqrt{d_{53}^2 - d_{43}^2} \quad (1.9)$$

Depois de se construir uma distribuição de tamanhos de gotas, obter diâmetros médios e quantificar a diversidade de tamanhos de gotas, a próxima secção revê o passo do ajuste de uma função matemática de densidade de probabilidade a uma distribuição discreta (histograma) na caracterização de um aerossol, apresentando no fim a motivação da presente dissertação.

1.1.3. Ajuste de uma Função Matemática de Densidade de Probabilidade a um Histograma

Na caracterização de um aerossol é comum apresentar-se valores médios para efeitos comparativos, mas se a distribuição original apresentar algumas características particulares

como a bimodalidade, essa informação perde-se. Por outro lado, com valores médios ou quaisquer outros momentos de uma distribuição não é possível saber, à partida, qual a escala e forma da distribuição que lhe deu origem. Por isso, o ajuste de uma função matemática de densidade de probabilidade a histogramas que organizam os dados experimentais para os tamanhos das gotas é útil para a sua futura modelação. Aliás, se dos valores médios é impossível saber a distribuição original, quando dispomos de informação suficiente para reconstruir as distribuições de tamanho de gotas a partir de parâmetros característicos das funções de distribuição de probabilidade, podemos sempre obter os valores médios sem perder a informação das distribuições originais. Portanto, abordar a caracterização de um aerossol a partir do ponto de vista da reconstrução das distribuições de densidade de probabilidade, tem vantagens significativas relativamente à divulgação e análise somente dos momentos retirados a partir de distribuições (Panão *et al.*, 2020). O ajuste de uma função matemática a uma distribuição de densidade de probabilidade (Figura 1.5 retirada de Tolis & Rentizelas (2011), ou uma função de distribuição cumulativa como exemplificado na Figura 1.4 (Panão *et al.*, 2020).

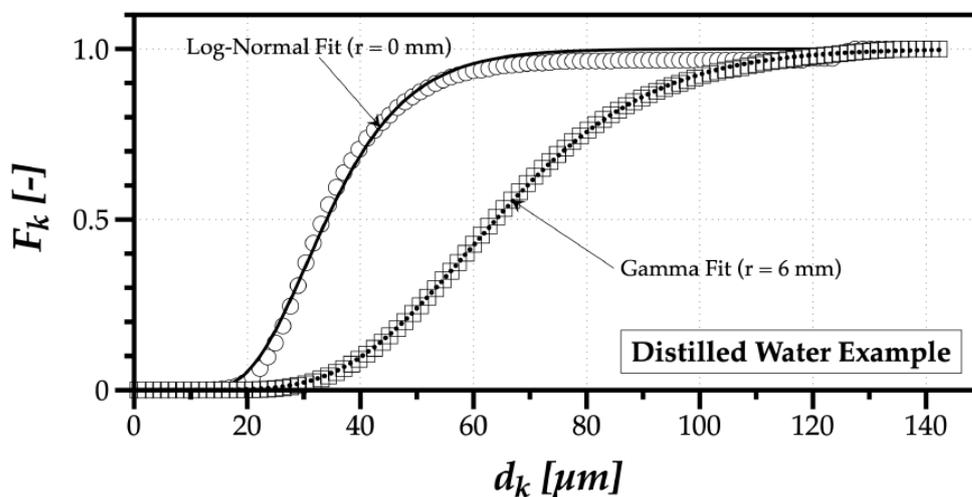


Figura 1.4 Exemplo do ajuste de uma função cumulativa para os resultados de água destilada, $r=0\text{mm}$ em que a melhor função é a Log-Normal e $r=6\text{mm}$ com a melhor função de ajuste sendo a Gamma. [Adaptado de M. O. Panão *et al.*, 2020]

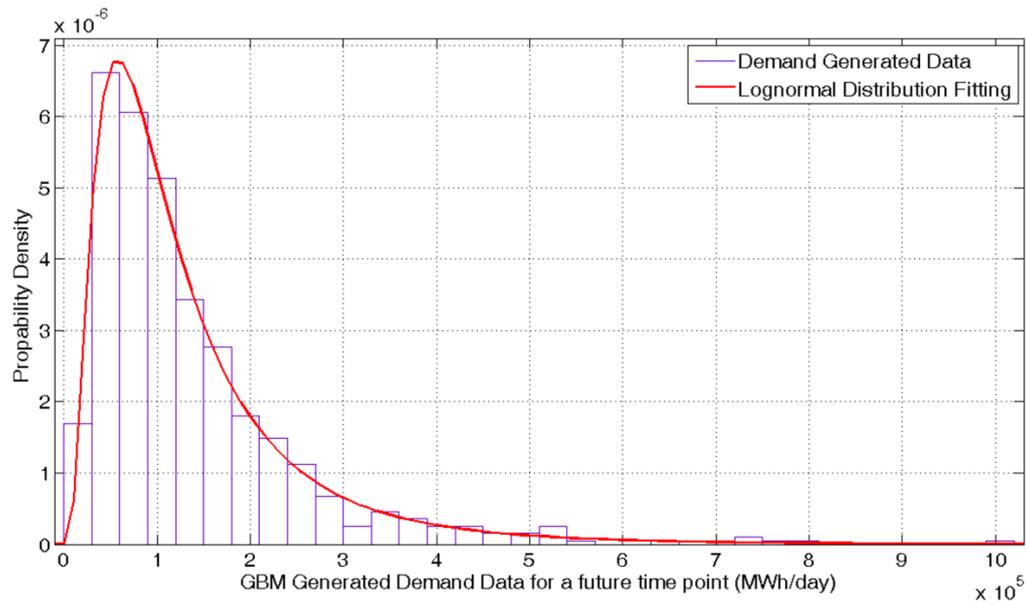


Figura 1.5 Ajuste de uma função de densidade de probabilidade Log-Normal a uma amostra de dados discretos. [Adaptado de Tolis & Rentizelas, 2011]

No caso dos aerossóis humanos, existe alguma informação dispersa relativamente às distribuições de gotas produzidas por diversas fontes de contágio como a tosse, o espirro e a falar (Chao *et al.*, 2009; Duguid, 1946). Não existe ainda um trabalho sistemático de ajuste dos dados experimentais existentes na literatura e qual a função matemática de densidade de probabilidade (PDF – *Probability Density Function*) que melhor os descreve. Aliás, havendo razões físicas para que determinadas PDF se ajustem aos dados de aerossóis, a análise dos parâmetros característicos delas poderia ajudar a entender quais os mecanismos subjacentes à produção de aerossóis. De qualquer modo, o exemplo seguinte fundamenta a necessidade de caracterizar melhor os dados de distribuição de tamanhos de aerossóis, uma vez que são a base de muitos modelos numéricos dedicados ao estudo da dispersão destas gotículas.

Dbouk & Drikakis (2020) é um dos trabalhos mais usados nos estudos sobre a dispersão de gotas que em menos de dois anos conta já com 365 citações. Os autores usam os dados de Xie *et al.* (2009) e apresentam um gráfico de ajuste reproduzido na Figura 1.6.

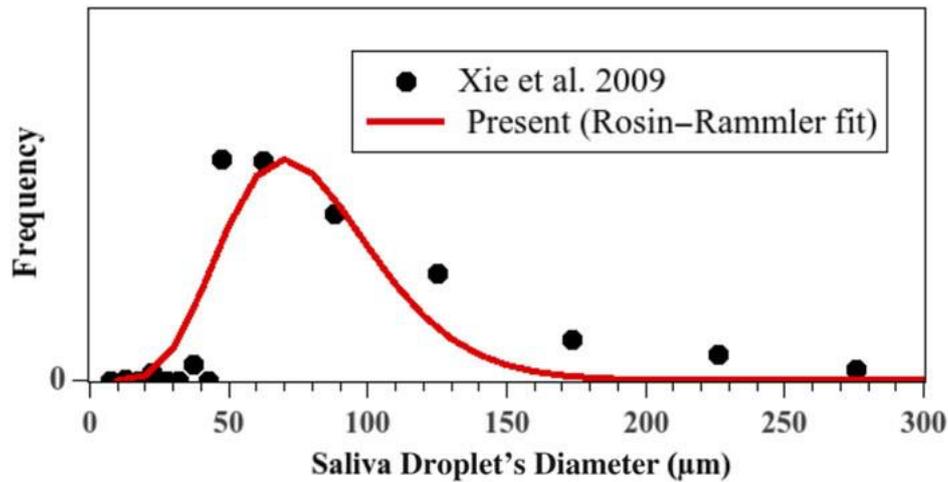


Figura 1.6 Ajuste de uma distribuição Rosin-Rammler de (Dbouk & Drikakis, 2020) aos dados experimentais para a distribuição de gotas produzidas por saliva em (Xie *et al.*, 2009).

Em primeiro lugar, o gráfico usa para as coordenadas y a palavra “Frequency” (Frequência), não apresenta quaisquer valores, e de acordo com a expressão de Rosin-Rammler ou Weibull usada pelos autores para o ajuste

$$p(d) = \frac{n_f}{\bar{d}} \left(\frac{d}{\bar{d}}\right)^{n_f-1} \exp \left[-\left(\frac{d}{\bar{d}}\right)^{n_f} \right] \quad (1.10)$$

Em que \bar{d} corresponde ao parâmetro de escala e n_f ao parâmetro de forma, essa refere-se a uma função densidade de probabilidade, p [μm^{-1}], e não a uma função de probabilidade. Porém, ao reproduzir os resultados da Figura 1.6 usando a distribuição indicada como melhor ajuste pelos autores onde $n_f = 8$ e $\bar{d} = 80$, o resultado real é muito diferente (Figura 1.7).

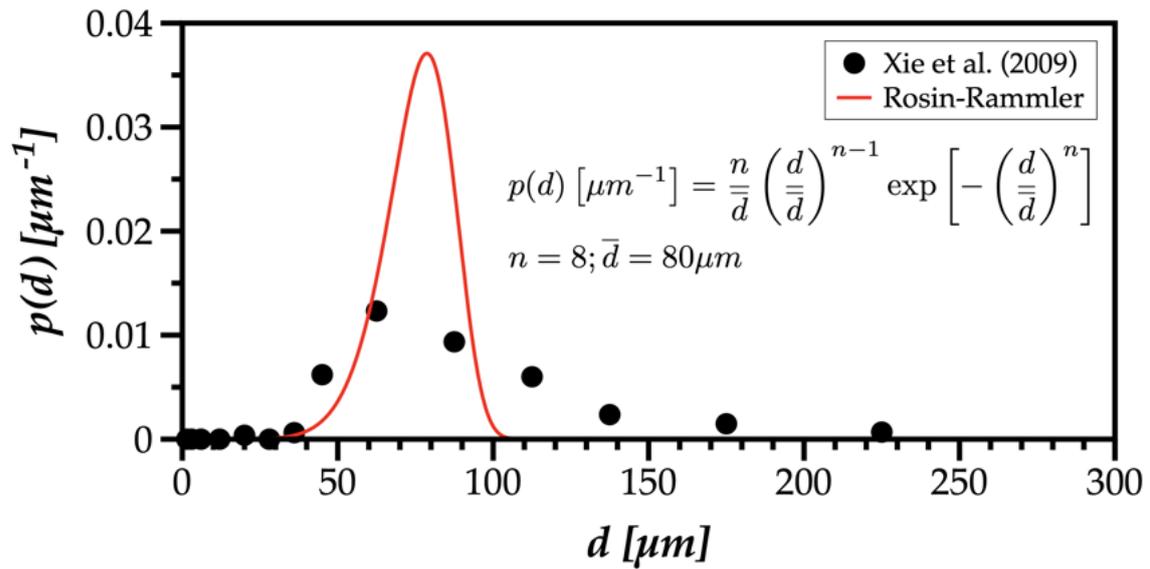


Figura 1.7 Ajuste da distribuição Rosin-Rammler/Weibull indicada por (Dbouk & Drikakis, 2020) e a distribuição de gotas providenciada por (Xie *et al.*, 2009).

Qual a razão dos autores não verificarem o ajuste dos dados experimentais com outras funções como a Gamma ou a Normal-Logarítmica. Este erro demonstra a necessidade de um trabalho mais sistemático de avaliação do ajuste de PDFs aos dados disponíveis na literatura para os aerossóis humanos. Por fim, um ajuste de uma função matemática de densidade de probabilidade à caracterização de um aerossol obtido experimentalmente significa mais do que haver uma boa representação matemática de “dados”. Sabendo que na caracterização de *sprays* o mais importante é garantir que existe “informação” suficiente com os dados disponíveis, distinguindo dados de informação, um dos objetivos desta dissertação será explorar a entropia diferencial para avaliar a qualidade de um ajuste quantitativamente, algo a que a próxima secção se dedica, introduzindo a teoria da informação antes de traçar os objetivos da dissertação.

1.2. Teoria da Informação e Entropia Diferencial

A teoria da informação procura definir matematicamente o conceito de informação e descrever como essa é comunicada entre elementos de um sistema. Interagimos com os frutos da teoria da informação todos os dias, quer nos apercebamos ou não. Um dos mais evidentes seria a *Internet* e a compressão de dados. Isto é, quando um arquivo é comprimido, são removidas informações redundantes, e quase todos os algoritmos bem-sucedidos nessa função estão enraizados na teoria da informação.

A origem da teoria da informação esta em Claude Shannon que no ano de 1948 publicou um trabalho seminal quando estava nos Laboratórios Bell intitulado “*A Mathematical Theory of Communication*” onde quantifica, pela primeira vez a quantidade de informação como uma entropia que adaptada a um *spray* seria

$$H(d) = -K \sum_i f_i(d) \log_2(f_i(d)) [bits] \quad (1.11)$$

Com $f_i(d)$ como a probabilidade em frequência associada a uma configuração de *spray* entre as várias disponíveis e formulada em unidades *bits* pelo facto de a probabilidade ter sido associada aos dígitos binários cujos valores variam entre 0 e 1. Fazendo $K = \ln(2)$, a entropia de Shannon passa a depender do logarítmico neperiano, ou seja $\ln x = \ln 2 \log_2 x$ e a unidade de *bits* transforma-se em *nats*. Neste trabalho, Shannon resolve ainda alguns dos maiores problemas ligados à comunicação de uma mensagem como assegurar o mínimo de informação na presença de ruído, assim como abrir o conceito de informação a uma reinterpretação de conceitos físicos como no caso da entropia ao nível mais informacional. Ainda, toda a variável que possui um carácter aleatório, abre-se a novas interpretações quando é quantificado o “grau de supresa” através da entropia de Shannon, uma expressão aliada à incerteza. Há uma pequena lista de axiomas a considerar, aquando da utilização de $H(d)$ para quantificar a informação, (Princeton University, 2020):

1. Resultados determinísticos não contêm informação
2. Conteúdo de informação aumenta com a diminuição da probabilidade do acontecimento
3. Conteúdo de informação é aditivo para variáveis independentes aleatórias

Se ajustarmos uma função continua a uma distribuição de densidade de probabilidade discreta, qual o reflexo que isso tem sobre a entropia de informação? A uma variável contínua associa-se uma densidade de probabilidade como $p(x_i) = f(x_i)/\Delta x$, que aplicado a equação (1.11) após alteração de base logarítmica para obter a informação em *nats*, resultando em

$$\begin{aligned}
H(x) &= - \sum_i p(x_i)\Delta x * \ln(p(x_i)\Delta x) \\
&= - \left(\sum_i p(x_i)\Delta x * \ln(p(x_i)) + \ln(\Delta x) \sum_i p(x_i)\Delta x \right) \\
&= - \sum_i p(x_i)\Delta x * \ln(p(x_i)\Delta x) + \ln\left(\frac{1}{\Delta x}\right)
\end{aligned} \tag{1.12}$$

Logo, para uma variável contínua positiva como é o caso do diâmetro das gotas de um aerossol, $x \geq 0$, $\Delta x \rightarrow dx \rightarrow 0$ e

$$H(x) = - \int_0^{\infty} p(x) \ln(p(x)) dx + \infty \tag{1.13}$$

O que representa uma dificuldade de análise por tornar-se um parâmetro que diverge para infinito. Como sugere Stone (2015), ao ignorarmos o termo de infinito, obtemos uma medida de entropia chamada de entropia diferencial de uma variável contínua definida como

$$H_{dif}(x) = - \int_0^{\infty} p(x) \ln(p(x)) dx \tag{1.14}$$

Cuja equação corresponde

$$H_{dif}(x^\Delta) = H(x^\Delta) + \ln(\Delta x) \tag{1.15}$$

Com $H(x^\Delta) = - \sum_i f_i(d) \ln(f_i(d))$. Se quisermos analisar se uma distribuição discreta corresponde a uma contínua, o teste de Kolmogorov-Smirnov seria adequado usando o máximo valor absoluto da diferença entre a distribuição cumulativa discreta e os valores dados pela função cumulativa contínua nos centroides de cada classe. Porém, em termos de significado, nada neste teste depende dos parâmetros da função matemática que melhor ajuste produz. Isto é, o teste é “cego” relativamente à PDF. Será que a entropia diferencial produz o mesmo resultado, conferindo-lhe significado?

1.3. Objetivos

Esta dissertação tem como objetivo extrair dados da literatura sobre distribuições de gotas, que caracterizam os aerossóis humanos em situações que induzem contágio de doenças do foro respiratório, como tosse, espirro e fala, e pretende:

1. Encontrar a função matemática de densidade de probabilidade que melhor se ajusta aos resultados experimentais, analisando o seu sentido físico;
2. Caracterizar cada aerossol em termos de DTG e diâmetros característicos;
3. Avaliar o valor informacional interpretativo da entropia diferencial como critério de aferição do ajuste de funções matemáticas às distribuições extraídas.

2. METODOLOGIA

A partir da literatura foram recolhidos dados de aerossóis humanos e, a partir das distribuições discretas cumulativas, foram simuladas amostras de $N = 10^6$ de gotas distribuídas de acordo com essas. A vantagem dos dados simulados será poder variar o número de classes face ao número reportado na literatura. Posteriormente, o ajuste dos dados simulados a funções matemáticas de densidade de probabilidade pretende aferir qual das mais habituais (Log-Normal, Gamma, Weibull) produz melhor ajuste. O teste mais habitual para avaliar se uma função de distribuição contínua se ajusta à função de distribuição discreta é o de Kolmogorov-Smirnov (KS). Porém, dada a elevada dimensão da amostra de dados simulados (N), o valor crítico seria sempre inferior para qualquer nível de significância. Quer isso dizer que o teste KS, por ser somente sensível ao número de dados, não distingue os dados da informação que esses transmitem dos tamanhos das gotas do *spray*. Daí a motivação para desenvolver uma nova metodologia com base na teoria da informação. Contudo, o teste KS pode ser útil para confirmar qual das funções matemáticas se aproxima mais da discreta, entre as funções testadas, por ser a que resulta no menor valor obtido para o critério de KS.

A Figura 2.1 ilustra o tratamento e criação de dados de um ponto de vista holístico de todo o processo. A sublinhado estão os temas-chave a ser analisados de forma a atingir os objetivos traçados.

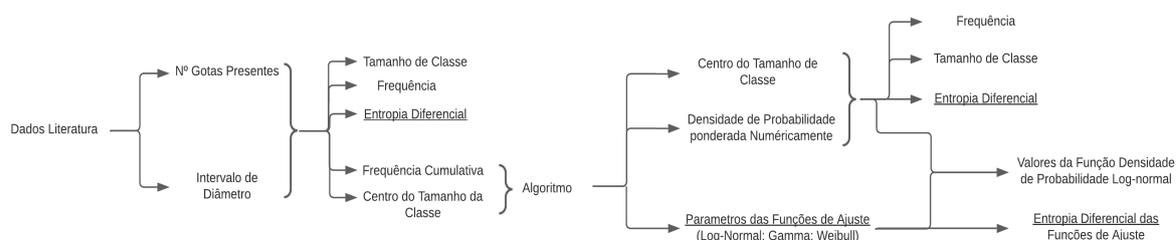


Figura 2.1 Flowchart do processamento dos dados

2.1. Simulação dos Dados e Ajuste

Os dados usados na caracterização dos aerossóis humanos provêm dos trabalhos de (Chao *et al.*, 2009; Duguid, 1946). Em Duguid (1946), as gotas produzidas pelos sujeitos foram medidas por micrometria direta após a sua recuperação do ar por lâminas oleadas. No de (Chao *et al.*, 2009) o tamanho das gotas foi medido usando Interferometria Imagiológica de Mie (*Interferometric Mie Imaging*, IMI), que permite medir o tamanho das gotas na proximidade da boca e evita perdas de amostras no ar.

É necessário categorizar os dados extraídos de acordo com o tipo de ocorrência de aerossol humano (espirro, tosse, falar alto ou contar alto), a posição inicial da boca, (inicialmente aberta ou fechada, quando não especificado, está fechada), se são múltiplas ocorrências ou apenas uma e a distância horizontal à boca a que a medição é feita. A Figura 2.2 serve de apoio para entender a classificação dos aerossóis correspondente aos dados usados na presente caracterização.

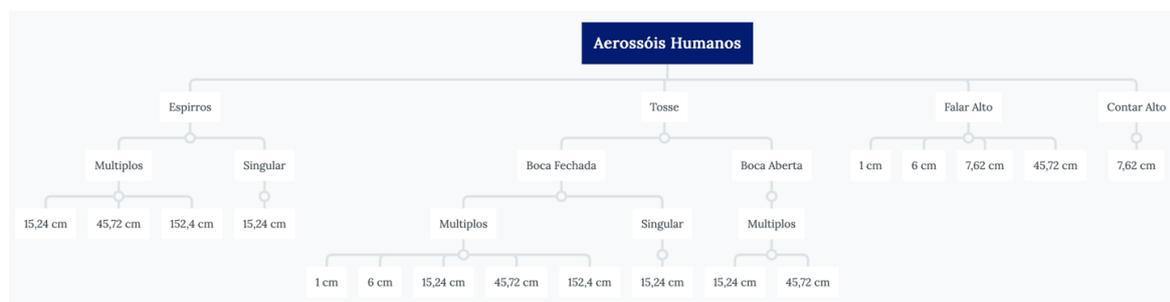


Figura 2.2 Categorização de aerossóis humanos.

Ao longo desta secção, para ilustrar a metodologia, utiliza-se dados de um aerossol representados na Figura 2.3 (Espirros; Múltiplos; 15.24 cm), mas todos os dados extraídos da literatura encontram-se em anexo. A simulação de um *spray* de N gotas com esta distribuição faz-se através do método de simulação de Monte-Carlo utilizando o *MATLAB*.

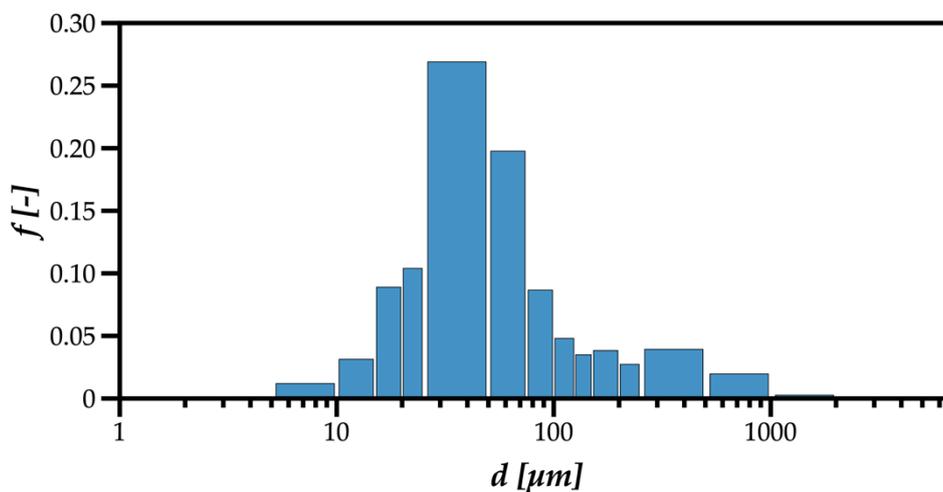


Figura 2.3 Dados extraídos da literatura do número de gotas e o seu respetivo tamanho de gota para o caso [Espirros, Múltiplos, 15,24 cm]. (Duguid, 1946)

É de notar como a distribuição de classes em Duguid (1946), é variável, enquanto a distribuição dos dados simulados é regular e possui um maior número de classes. A Figura 2.4 compara a distribuição original Figura 2.3 com a simulação, mas em densidade de probabilidade.

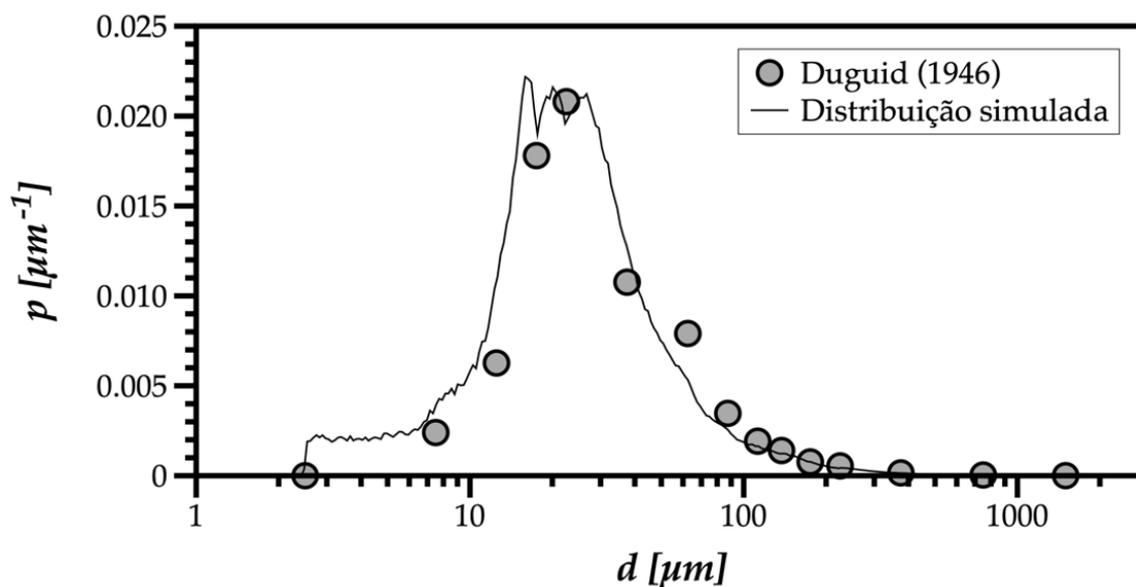


Figura 2.4 Comparação entre a distribuição original de (Duguid, 1946) para espirros múltiplos e a obtida por simulação.

A partir dos dados simulados, as funções de densidade de probabilidade testadas no ajuste são:

- A Log-Normal: $p(d) = \frac{1}{\sqrt{2\pi}\sigma d} \exp\left[-\frac{\ln(d/d_m)^2}{2\sigma^2}\right]$, com d_m como o diâmetro médio e σ o desvio padrão;

- A Gamma: $p(d) = \frac{1}{d \Gamma(a)} \left(\frac{d}{b}\right)^a \exp\left[-\frac{d}{b}\right]$, com a como o parâmetro de forma e b o parâmetro de escala;

- A Weibull: $p(d) = \frac{a}{b} \left(\frac{d}{b}\right)^{a-1} \exp\left[-\left(\frac{d}{b}\right)^a\right]$, com a como o parâmetro de forma e b o parâmetro de escala.

O critério de Kolmogorov-Smirnov (D) é obtido a partir da máxima distância entre o valor da função de distribuição cumulativa contínua para uma classe, $\Phi(d_k)$, e o valor da distribuição cumulativa discreta, $F(d_k)$, como

$$D = \max\{\Phi(d_k) - F(d_k)\} \quad (2.1)$$

A próxima secção descreve a metodologia nova proposta de critério para aferir a qualidade do ajuste das funções matemáticas de densidade de probabilidade usando a entropia diferencial na teoria da informação.

2.2. Entropia Diferencial como Critério de Aferição do Ajuste

De acordo com Stone (2015), conforme o número de classes aumenta, para uma distribuição de densidade de probabilidade discreta, menor é o valor da entropia diferencial definida pela equação (1.15). Isto é, com o aumento do número de classes, o valor da entropia diferencial da distribuição discreta tem tendência a aproximar-se do valor da entropia diferencial da função que é ajustada e descreve a mesma. Assim, será relevante introduzir as entropias diferenciais para as funções de densidade de probabilidade indicadas na secção anterior (Log-Normal – LN, Gamma – GM e Weibull – WB). Essa dedução segue o procedimento definido em (Reza, 1961) e resultam, para a função de distribuição Log-Normal

$$H_{dif}^{LN}(d) = \frac{1}{2} \ln(2\pi e \sigma d_m) \quad (2.2)$$

Com e como o número de Neper; para a função de distribuição Gamma

$$H_{dif}^{GM} = \ln(\Gamma(a)b) + (1-a)\psi(a) + \frac{\Gamma(a+1)}{\Gamma(a)} \quad (2.3)$$

Com $\psi(a)$ como a função Digamma; e para a função de distribuição Weibull

$$H_{dif}^{WB} = \ln\left(\frac{b}{a}\right) + \left(\frac{b-1}{b}\right)\gamma + 1 \quad (2.4)$$

Com γ como a constante de Euler-Mascheroni.

A Figura 2.5 mostra como o *spray* simulado corresponde ao original e apesar de nenhuma das distribuições conseguir descrever bem a distribuição medida a Log-Normal é a que mais se aproxima dessa.

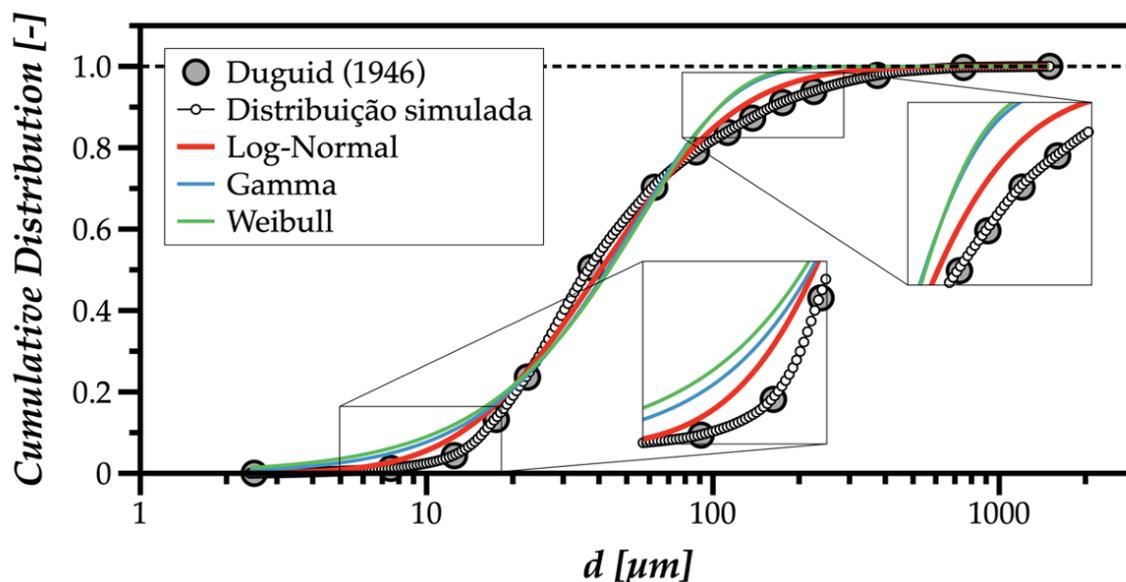


Figura 2.5 Comparação entre as distribuições cumulativas para espirros múltiplos (Duguid, 1946)

A Tabela 2.1 contém os valores dos parâmetros das funções de ajuste – Log-Normal, Weibull e Gamma. São estes os valores usados para o cálculo da entropia diferencial e para a caracterização do aerossol.

Tabela 2.1. Valores dos parâmetros característicos das funções ajustadas aos dados simulados fornecidos pelo algoritmo.

	Log-normal [d_m; σ]	Gamma [a; b]	Weibull [a; b]
[d_m; a]	40.4912	1.7160	1.3581
[σ; b]	0.8806	30.4611	56.7913

De forma a aplicar a entropia diferencial como um critério de aferição do ajuste foram calculados os valores das entropias diferenciais: dos dados originais; dos simulados; das funções de ajuste para cada ocorrência e de seguida comparados. Qualquer diferença que aponte para um desajuste entre a informação contida nos dados quantificada pela entropia diferencial e a entropia diferencial de uma determinada função matemática, significa que a última não descreve a primeira. As unidades da entropia diferencial são o *nats*, dada a dependência da entropia de informação de Shannon baseada no logaritmo natural ou neperiano. Compara-se ainda o resultado entre a simulação e a função de acordo com o teste de Kolmogorov-Smirnov, normalizado pelo valor crítico com um nível de significância de $\alpha = 0,05$ (D_{KS}^α). Como se esperava, para o tamanho da amostra, nenhum dos ajustes passaria o teste de KS, embora o valor menor seja atribuído à distribuição Log-Normal.

Tabela 2.2. Entropias Diferenciais dos dados originais; simulados; funções de ajuste e comparação entre simulados e funções de acordo com o teste de KS.

	Dados Originais	Dados Simulados	Função	$\frac{D_{func}}{D_{KS}^{0.05}}$
	H_{dif} [nats]			$D_{KS}^{0.05} = 0.00136$
Log-Normal	5,248	5,0827	5,0585	35,237
Gamma			5,0002	62,145
Weibull			5,3010	65,526

3. APRESENTAÇÃO E DISCUSSÃO DE RESULTADOS

3.1. Caracterização de Aerossóis Humanos

Sabendo que aos dados simulados foram ajustadas funções a partir de uma distribuição cumulativa, podemos utilizar a mediana de ambos D_{50} [μm] de forma a caracterizar e verificar a qualidade do ajuste. Na Figura 3.1 o significado de os valores entre distribuições serem semelhantes e próximos entre si significa que as aproximações aos dados simulados estão bem feitas. É também possível observar que a função de distribuição Log-Normal, é a que tem os valores de D_{50} mais próximos dos valores dos dados simulados, indicando desta forma que será a mais adequada para modelar e caracterizar os aerossóis humanos.

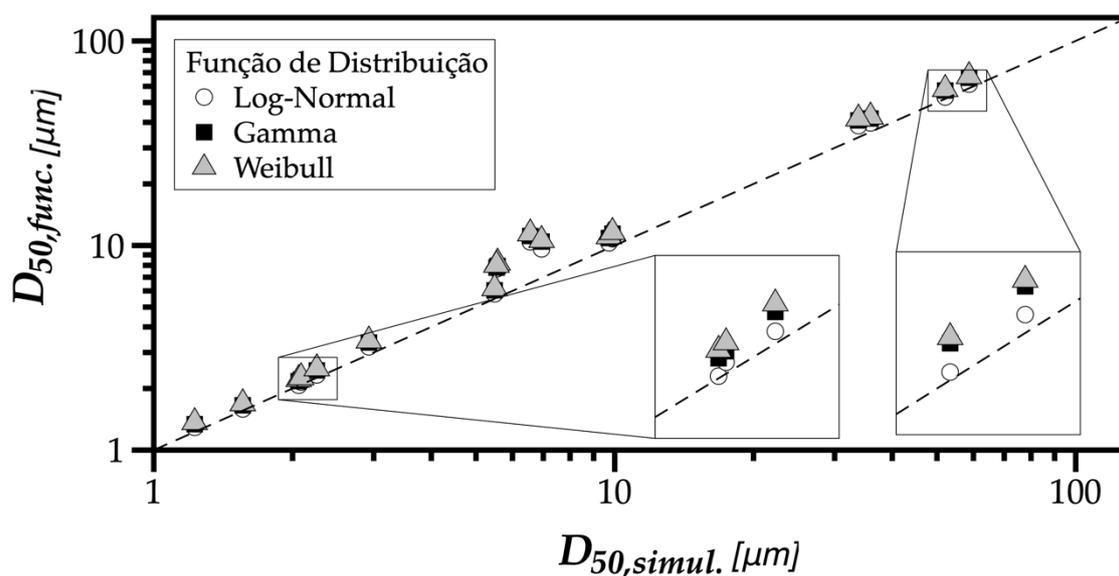


Figura 3.1 Comparação do diâmetro representativo de 50% da distribuição cumulativa dos dados simulados com o obtido através das funções de ajuste.

O conceito de DTG implica caracterizar os aerossóis sobre dois aspetos; i) quantos tamanhos diferentes são relevantes (Grau de Polidispersão) no *spray*; ii) e quão diferentes são os tamanhos relevantes (Grau de Heterogeneidade) no *spray*. O grau de polidispersão é

dados pela entropia de Shannon normalizada e na Figura 3.2 é demonstrado que, a condição “Espirros” são os que têm mais tamanhos diferentes relevantes e quanto mais longe da boca, mais existem. No grau de heterogeneidade o resultado já é diferente como se observa na Figura 3.3, onde o aerossol com maior diferença nos tamanhos relevantes corresponde à condição de “Falar Alto”. Este é o aerossol que pode ser caracterizado com uma maior variedade de diferença nos tamanhos relevantes que o compõe. É, também, possível constatar que quanto maior for a distância de captura, menor será a heterogeneidade do *spray*.

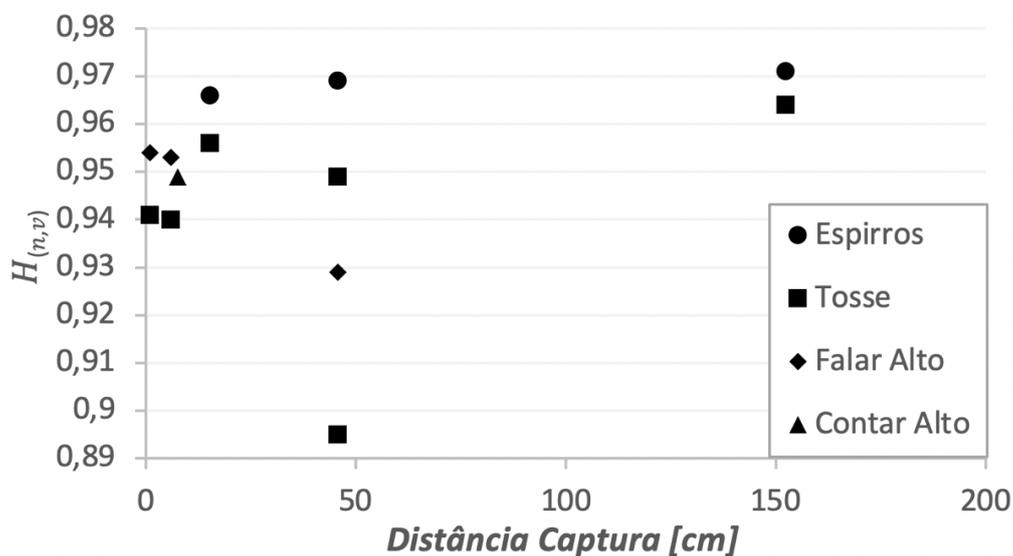


Figura 3.2 Grau de Polidispersão dos diversos aerossóis humanos de acordo com a distância a que ocorre a captura dos dados.

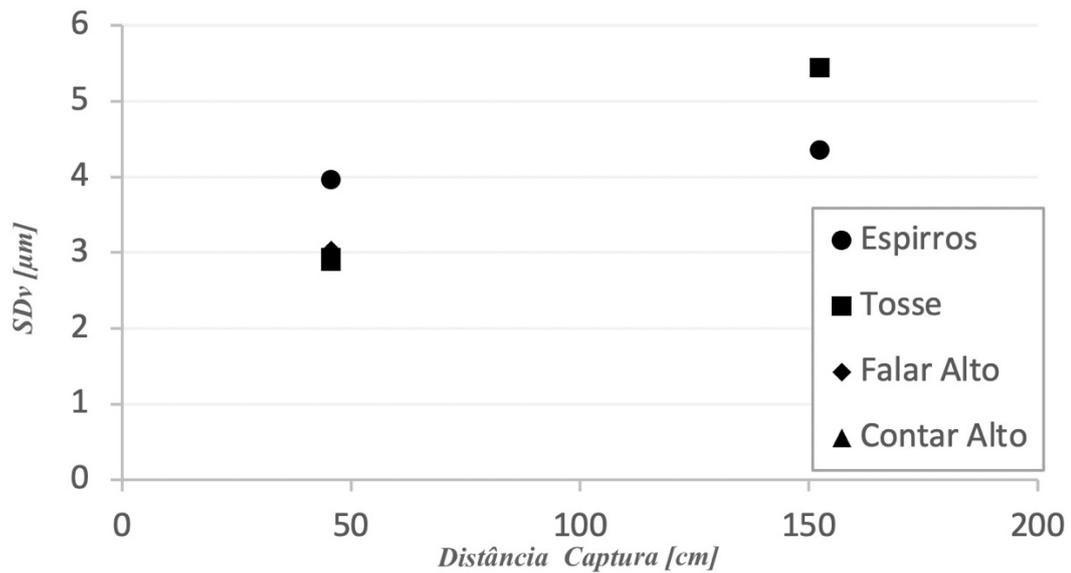


Figura 3.3 Grau de Heterogeneidade dos diversos aerossóis humanos de acordo com a distância a que ocorre a captura dos dados.

Segundo o teste de Kolmogorov-Smirnov (KS), pelo facto de estarmos a simular uma amostra relativamente elevada (10^6), nenhuma função de distribuição cumulativa das três testadas (Log-Normal, Gamma e Weibull) passa o teste quando confrontada com a distribuição cumulativa discreta simulada. Porém, considera-se que a função de densidade de probabilidade que produzir menor valor do teste é a que melhor se ajusta aos resultados experimentais. Nesse sentido, constata-se pela Figura 3.4, que os valores do teste de KS indicam a Log-Normal como a mais adequada para o ajuste.

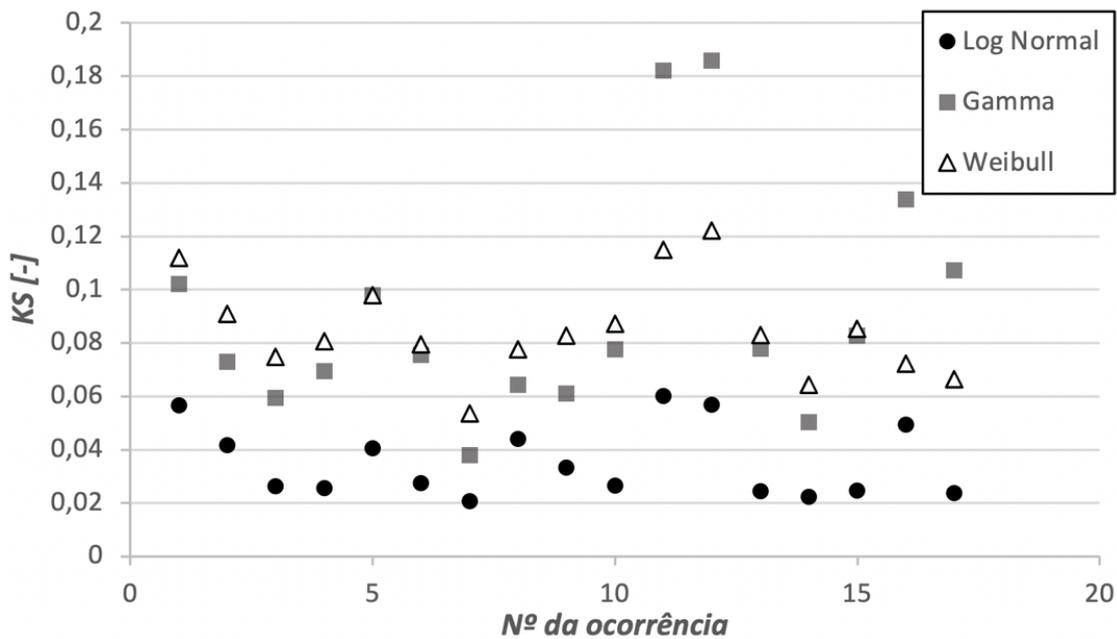


Figura 3.4 Resultados do teste de KS para cada função ajustada aos dados simulados. Ocorrências de 1-17 com leitura através da Figura 2.2.

A razão física que fundamenta a Log-Normal como a melhor função para descrever distribuições de dados discretas de aerossóis refere-se à geração das gotas advir de um crescimento exponencial na diversidade do seu tamanho num aerossol. Considere-se que a relação entre uma variação da diversidade de tamanhos dD das gotas do *spray* e o diâmetro representativo D , depende de uma constante taxa de crescimento de diversidade de tamanho de gota σ numa escala de tempo infinitesimal dt . Assim, a formulação desta relação para o tempo necessário à produção de gotas de uma certa classe k do *spray* expressa-se como:

$$\frac{dD}{D} = \sigma dt \tag{3.1}$$

Que resolvida em ordem ao instante de ocorrência dessa variação da diversidade resulta em

$$t_k = \frac{\ln(D_k) - \ln(D_c)}{\sigma} \tag{3.2}$$

Sendo D_c o diâmetro característico relativamente ao intervalo de tamanho. Posteriormente, considera-se que a probabilidade de um número infinitesimal de gotículas

ser adicionado a uma classe pela medida do seu tamanho corresponde a $\frac{dn_k}{N} = dp_k$, com N como o tamanho total da amostra e dp_k (ou $dp(D)$) como a variação infinitesimal do valor de probabilidade de frequência. Supondo que os instantes correspondentes à formação das gotas a partir de instabilidades nas estruturas hidrodinâmicas da saliva resultam de fenómenos aleatórios e, por isso, podem ser descritos por uma distribuição normal padrão (com a sua média em torno de zero e o seu desvio padrão igual a um) que expressa a natureza aleatória das mudanças na probabilidade de cada tamanho de classe

$$\frac{dp_k}{dt} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t_k^2}{2}\right) \quad (3.3)$$

Introduzindo a equação (3.1) na (3.2) e usando a (3.3), a função de densidade de probabilidade, $f(D) = \frac{dp_k}{dD} = \frac{1}{\sigma D} \frac{dp}{dt}$ resulta em,

$$f(D) = \frac{1}{\sigma D \sqrt{2\pi}} \exp\left(-\frac{(\ln(D) - \ln(D_c))^2}{2\sigma^2}\right) \quad (3.4)$$

Que corresponde à distribuição normal logarítmica. Desta forma, a razão física para as distribuições dos tamanhos de gota de *sprays* serem descritos pela função de densidade de probabilidade Log-normal deve-se a um crescimento exponencial na diversidade de tamanho de gota no *spray*. O desvio padrão ganha, também, um significado físico. Pois, correspondendo à taxa de crescimento da diversidade de tamanho de gota, revela depender do mecanismo de atomização (Sanches *et al.*, 2022).

Ao analisar uma possível relação entre o diâmetro característico de um aerossol d_c ($\equiv D_c$), e a taxa de crescimento da diversidade de tamanho de gota – desvio padrão da distribuição normal logarítmica σ , é possível observar na Figura 3.5 um relacionamento aproximadamente linear entre ambos (note que a escala de d_c é logarítmica). Isto indica que um aerossol com uma taxa de crescimento superior corresponde a um maior valor de diâmetro característico. Por fim, estes resultados serão úteis para a simulação de aerossóis humanos em modelação numérica.

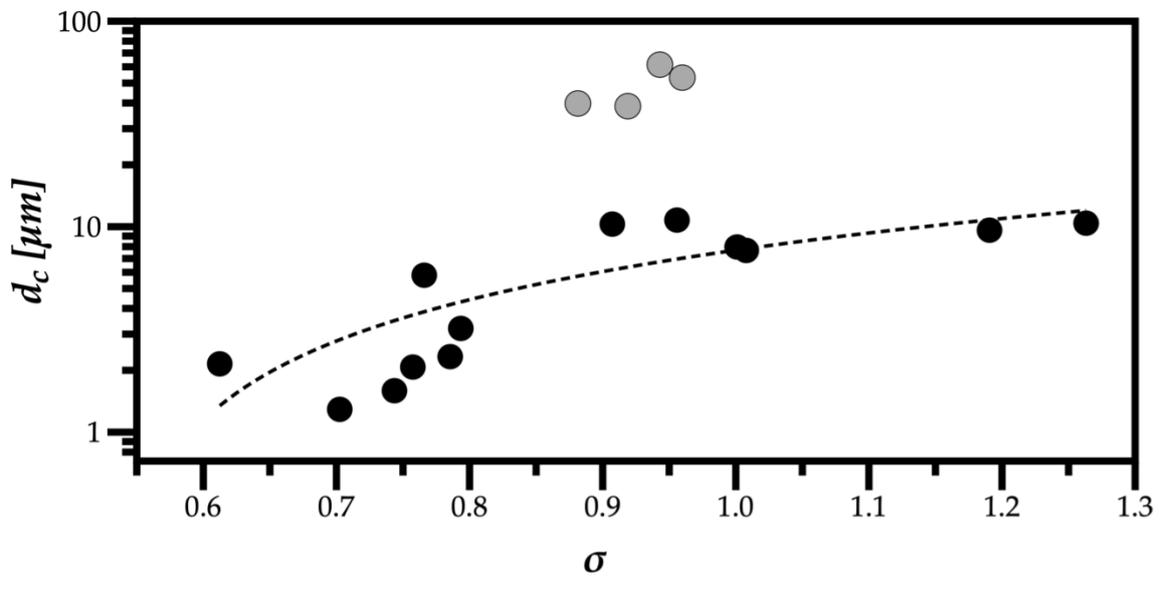


Figura 3.5 Relação entre o diâmetro característico dos aerossóis humanos simulados com o desvio padrão da distribuição normal logarítmica.

De acordo com Stone (2015), a entropia diferencial a que se dedica a próxima secção de apresentação de resultados e discussão, é um conceito peculiar porque tem pouco significado para uma variável contínua, por ser infinita e isso representar que uma variável contínua possui uma quantidade infinita de informação. Porém, a caracterização de *sprays* faz-se com medidas que correspondem a variáveis discretas e, nesses casos, a entropia indica a quantidade de informação finita transportada com as características dessa variável. A precisão de cada medida está limitada pelo ruído. Por isso, ao comparar-se a entropia diferencial de uma variável contínua com uma discreta, a sua proximidade resulta, conceptualmente, da transformação de uma variável contínua descrita por uma função de densidade de probabilidade, numa variável discreta e mensurável descrita por um histograma. É esta a ideia que motiva a análise da entropia diferencial como critério de ajuste feita na próxima secção.

3.2. Aferição da Entropia Diferencial como Critério de Ajuste

De forma a aferir a entropia diferencial (H_{dif}) como um critério de ajuste, a Figura 3.6 compara à esquerda a H_{dif} da distribuição original com o valor obtido para a distribuição do aerossol simulado. À direita compara-se H_{dif} da simulação com a de três funções de

densidade de probabilidade que se pretende avaliar quanto à qualidade do ajuste aos resultados simulados. Como o número de classes das distribuições originais é inferior ao das simuladas, de acordo com o conceito de entropia diferencial detalhado na introdução, o valor de H_{dif} deveria aproximar-se do valor da função de densidade de probabilidade contínua que o descreve, assumindo que existe. Daí que H_{dif} do aerossol simulado seja menor que o valor da distribuição discreta original. Quando os resultados para a H_{dif} do aerossol simulado são comparados com o valor da entropia diferencial obtida para os três tipos de funções de densidade de probabilidade testadas pela qualidade do ajuste, observa-se que todas podem considerar-se alternativas credíveis para caracterizar os aerossóis humanos.

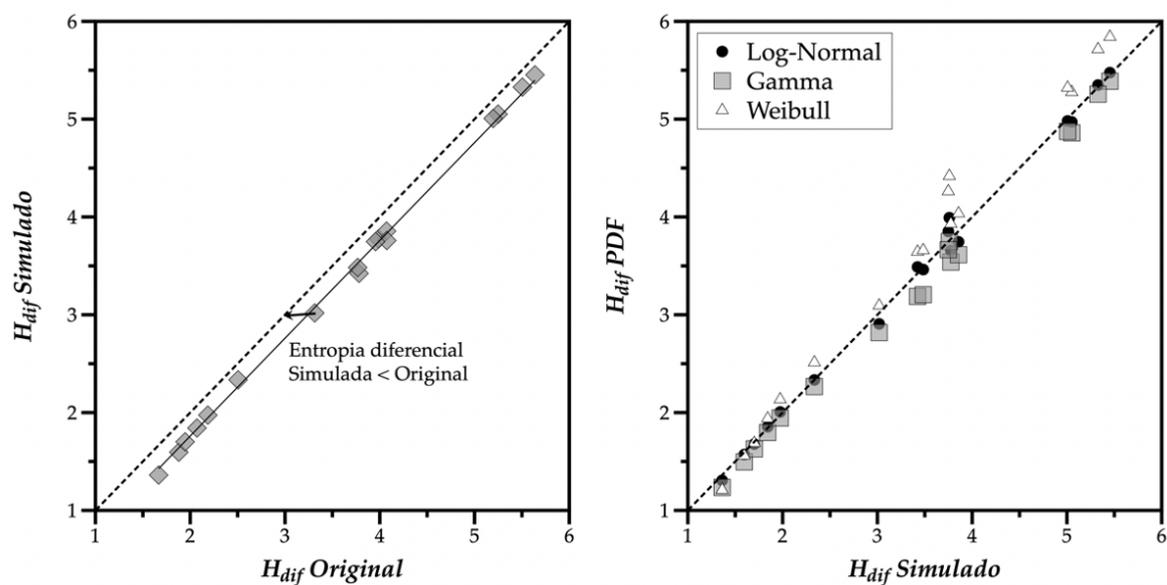


Figura 3.6 Comparação da entropia diferencial original com a do aerossol simulado (esquerda) e entre o aerossol simulado e da função contínua testada pelo ajuste (direita)

Uma análise ao desvio percentual entre a entropia diferencial da função matemática e a da simulação apresentada na Figura 3.7, mostra que a maior parte dos ajustes está concentrada na distribuição Log-Normal. O intervalo de desvio do melhor ajuste é [0,06-3,98%].

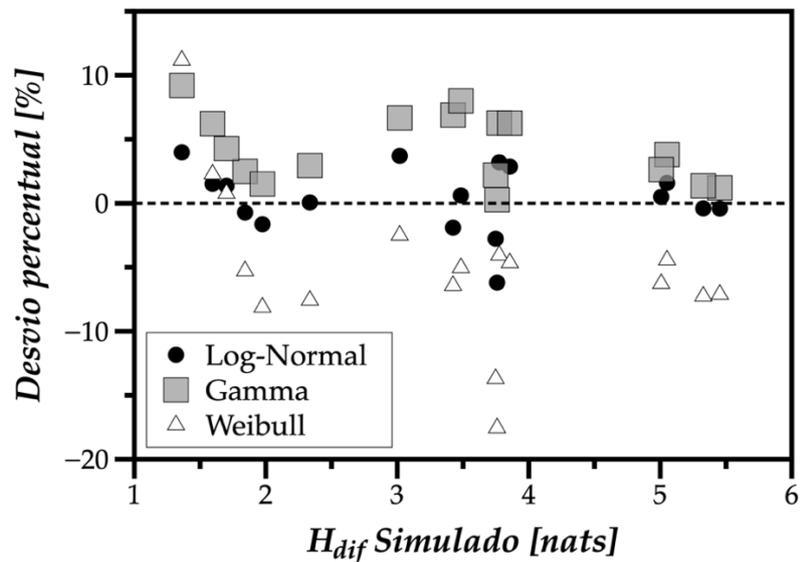


Figura 3.7 Desvio percentual entre a entropia diferencial da distribuição da função contínua de densidade de probabilidade e a do aerossol simulado.

A Figura 3.8 mostra a percentagem distributiva do melhor ajuste pelas funções de densidade de probabilidade que minimizaram este desvio. Mais de dois terços das distribuições de tamanhos de gotas são descritas por uma função de distribuição Log-Normal, embora haja uma percentagem não desprezável (>10%) de casos cujas distribuições mais bem descritas pelas funções de distribuição Gamma e Weibull.

Percentagem de Melhor Ajuste dos Aerossóis Humanos

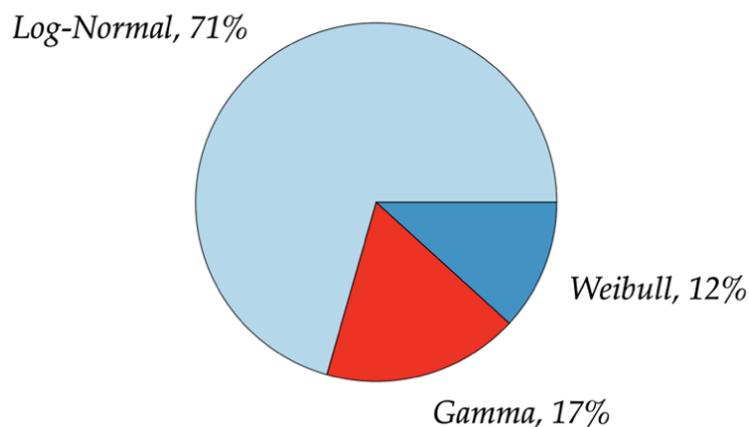


Figura 3.8 Distribuição dos ajustes que aproximaram mais a entropia diferencial da função matemática (Log-normal, Gamma e Weibull) à do aerossol simulado.

Apesar da entropia diferencial estar a demonstrar ser um critério de ajuste adequado, os resultados diferem do valor mínimo obtido no teste KS que aponta a Log-Normal como a distribuição que produz os resultados mais aproximados entre a distribuição cumulativa discreta e a descrita pela função matemática. Assim, o uso da entropia diferencial possui algumas limitações e necessita de mais investigação. A Figura 3.9 demonstra uma dessas limitações. Neste caso, a entropia diferencial demonstra ser menos sensível à variação das funções matemáticas quando as classes têm tamanhos variáveis. Se as classes forem regulares, a convergência entre a simulação e a função matemática obriga ao uso de um maior número de classes relativa às classes de tamanho variável, usando o mesmo critério de paragem.

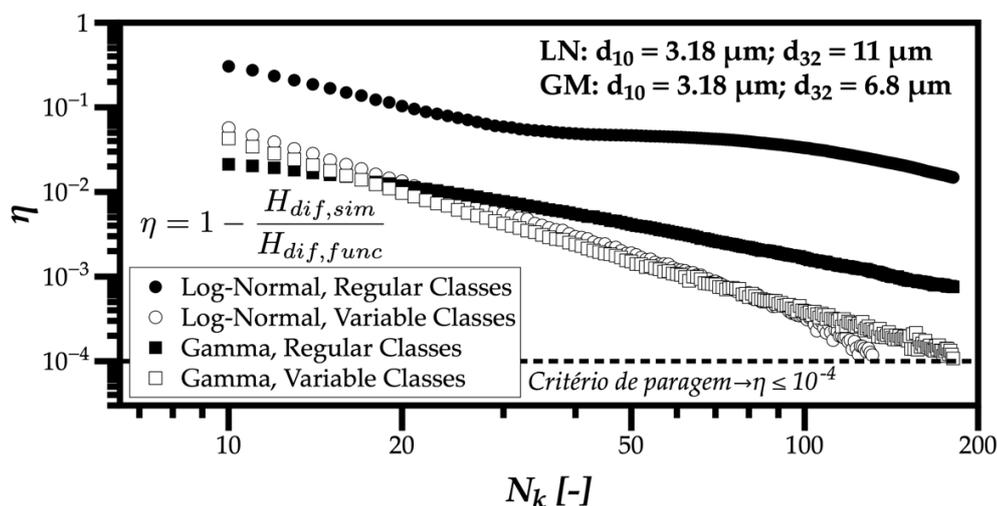


Figura 3.9 Demonstração do efeito de classes regulares vs variáveis em dados simulados nas funções de ajuste Log-Normal e Gamma. LN (2,335; 0,7864) & GM (1,7671; 1,8001).

3.3. Discussão dos Resultados segundo a Origem Física da Log-Normal

Para entender de que forma o desvio padrão – taxa de crescimento da diversidade de tamanhos de gota da função distribuição Log-Normal pode auxiliar na caracterização de aerossóis ele é comparado com valores médios e outros parâmetros – $D_{3,2}$, $D_{4,3}$, Grau de Heterogeneidade – isto é feito com o objetivo de encontrar alguma correlação entre o desvio padrão e algum destes parâmetros. Para este cálculo certos aerossóis humanos foram separados devido à sua diferente origem de captura. Casos como, espirros, múltiplos, 15.24 cm; Tosse, boca fechada, múltiplos, 15.24 cm; tosse, boca aberta, múltiplos, 15.24 cm; Falar

Alto, 7.62 cm, são distribuições de tamanhos de gotas grandes, enquanto todos os outros casos são constituídos predominantemente por núcleos-de-gota (*droplet-nuclei*), que são gotas de pequeno tamanho que ficam em suspensão.

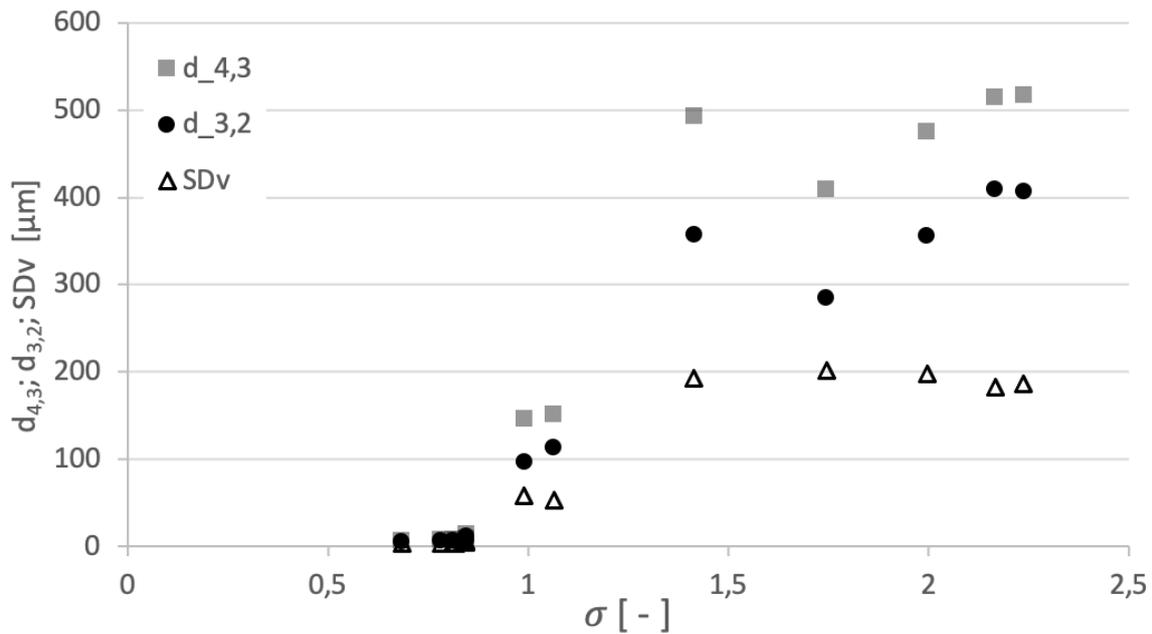


Figura 3.10 Análise da relação da taxa de crescimento da diversidade de tamanho de gota com o grau de heterogeneidade, Diâmetro de Sauter e Brouckere.

À semelhança do que foi observado para o diâmetro característico da função Log-Normal, na Figura 3.10 é possível observar que uma maior taxa de crescimento da diversidade de tamanho de gota, irá estar associada a um maior aumento do diâmetro de Sauter e diâmetro de Brouckere, valores estes relevantes para casos em que a investigação é mais focada na área de transferência de calor ou do volume de *spray*, respetivamente. O grau de heterogeneidade demonstra também que uma maior taxa de crescimento de diversidade de tamanho de gota, traduz-se numa maior diferença entre os tamanhos relevantes no *spray*.

4. CONCLUSÃO

Uma boa caracterização e modelação de aerossóis humanos melhora as estratégias de proteção e planeamento do controlo de propagação de vírus, mitigando o seu potencial efeito na população. Daí surge a importância do trabalho desenvolvido nesta dissertação que apresenta ajustes para aerossóis humanos provenientes de espirros, tosse e falar alto, através da caracterização e modelação das distribuições de gotas usando funções matemáticas de densidade de probabilidade. Procurou-se um novo método de aferir o ajuste entre uma função densidade de probabilidade a uma amostra de dados (simulada realisticamente) de uma variável aleatória contínua, o diâmetro das gotas do spray, através do conceito de entropia diferencial baseado na teoria da informação. O método apresentado comprova a utilização da distribuição normal logarítmica como a que melhor modela os aerossóis humanos. Explorou-se o significado físico do desvio padrão como a taxa de crescimento de diversidade do tamanho de gota e analisaram-se os resultados da sua correlação com diâmetros característicos dos aerossóis.

Ao caracterizar aerossóis através da diversidade de tamanho de gota foi possível observar que a condição de “Espirros” é o aerossol humano com maior grau de polidispersão, ou seja, tem mais tamanhos diferentes que são relevantes quanto mais longe da boca se estiver. A condição “Falar Alto” é o aerossol com um maior grau de heterogeneidade, maior variedade de diferença nos tamanhos relevantes que o compõe, e quanto maior for a distância à boca, menos heterogéneo é o aerossol.

Nenhuma das funções de distribuição cumulativa das três testadas (Log-Normal; Gamma e Weibull) passa o teste de Kolmogorv-Smirnov porque a amostra simulada é muito elevada. Porém, o menor valor para a distância KS indica-nos a função de densidade de probabilidade normal logarítmica como a que melhor se ajusta aos dados discretos. A sua explicação vem do pressuposto relativo à geração das gotas estar relacionado com o crescimento exponencial da diversidade de tamanhos de gota expresso por uma taxa constante – o seu desvio padrão.

A entropia diferencial revela-se como critério viável de aferição do ajuste entre uma função contínua e a distribuição discreta. Isto porque, aponta também a distribuição normal logarítmica como a que melhor descreve as distribuições de tamanho de gota dos aerossóis, como resultado da minimização da distância do teste de KS. No entanto, apesar da sua viabilidade, o resultado não foi integralmente o mesmo que o KS, demonstrando a necessidade de mais investigação.

Por fim, os resultados obtidos para o desvio padrão sugerem que um maior aumento desta taxa se traduz num maior aumento dos diâmetros característicos do aerossol, como os de Sauter e o de Brouckere. O aumento da taxa de crescimento revela ainda um aumento do grau de heterogeneidade, ou seja, uma maior diferença entre os tamanhos relevantes no aerossol. O trabalho futuro passará pelo aprofundamento da entropia diferencial, no desenvolvimento de critérios de definição do número de classes que melhor aproxima a representação discreta de uma distribuição, com a possível função matemática que a poderia descrever.

REFERÊNCIAS BIBLIOGRÁFICAS

- Chao, C. Y. H., Wan, M. P., Morawska, L., Johnson, G. R., Ristovski, Z. D., Hargreaves, M., Mengersen, K., Corbett, S., Li, Y., Xie, X., & Katoshevski, D. (2009). Characterization of expiration air jets and droplet size distributions immediately at the mouth opening. *Journal of Aerosol Science*, *40*(2), 122–133. <https://doi.org/10.1016/j.jaerosci.2008.10.003>
- Dbouk, T., & Drikakis, D. (2020). On coughing and airborne droplet transmission to humans. *Physics of Fluids*, *32*(5). <https://doi.org/10.1063/5.0011960>
- Dogan, I., & Dogan, N. (2010). DETERMINATION OF THE NUMBER OF BINS/CLASSES USED IN HISTOGRAMS AND FREQUENCY TABLES: A SHORT BIBLIOGRAPHY. Em *Journal of Statistical Research*. <https://www.researchgate.net/publication/292152648>
- Duguid, J. P. (1946). The size and the duration of air-carriage of respiratory droplets and droplet-nuclei. *Journal of Hygiene*, *44*(6), 471–479. <https://doi.org/10.1017/S0022172400019288>
- Han, Z. Y., Weng, W. G., & Huang, Q. Y. (2013). Characterizations of particle size distribution of the droplets exhaled by sneeze. *Journal of the Royal Society Interface*, *10*(88). <https://doi.org/10.1098/rsif.2013.0560>
- Panão. (2021a). *Ultrasonic atomization: new spray characterization approaches*. <https://doi.org/>
- Panão. (2021b). Interpreting Liquid Atomization Efficiency. *International Journal on Engineering Technologies and Informatics*, *2*(5). <https://doi.org/10.51626/ijeti.2021.02.00024>
- Panão. (2022). Ultrasonic Atomization: New Spray Characterization Approaches. *Fluids*, *7*(1). <https://doi.org/10.3390/fluids7010029>

- Panão, M. R. O. (2012). Assessment of measurement efficiency in laser- and phase-Doppler techniques: An information theory approach. *Measurement Science and Technology*, 23(12). <https://doi.org/10.1088/0957-0233/23/12/125304>
- Panão, Moita, A. S., & Moreira, A. L. (2020). On the statistical characterization of sprays. Em *Applied Sciences (Switzerland)* (Vol. 10, Issue 17). MDPI AG. <https://doi.org/10.3390/app10176122>
- Princeton University. (2020). *Information Theory Basics*. Youtube. <https://www.youtube.com/watch?v=bkLHszLIH34&t=491s>
- Reza. (1961). *Fazlollah M. Reza - An Introduction to Information Theory-Dover (1961)*.
- Sahann, R., Möller, T., & Schmidt, J. (2021). *Histogram binning revisited with a focus on human perception*. <http://arxiv.org/abs/2109.06612>
- Sanches, M., Moita, A. S. H. O., & Panão, M. R. O. (2022). *Potential contamination of high-speed dental handpiece by spray impact on the mouth surface*.
- Stone, J. v. (2015). *Information Theory A Tutorial Introduction*. <https://rkbookreviews.wordpress.com/>,
- Tolis, A. I., & Rentizelas, A. A. (2011). An impact assessment of electricity and emission allowances pricing in optimised expansion planning of power sector portfolios. *Applied Energy*, 88(11), 3791–3806. <https://doi.org/10.1016/j.apenergy.2011.04.054>
- Xie, X., Li, Y., Sun, H., & Liu, L. (2009). Exhaled droplets due to talking and coughing. *Journal of the Royal Society Interface*, 6(SUPPL. 6). <https://doi.org/10.1098/rsif.2009.0388.focus>

ANEXO

Tabela 0.1. Distribuição de tamanhos de gotas grandes(Duguid, 1946)

LowDk [μm]	HighDk [μm]	Espirros; Múltiplos; 15,24 cm	Tosse; Boca Fechada; Múltiplos; 15,24 cm	Tosse; Boca Aberta; Múltiplos; 15,24 cm	Falar alto; 7,62 cm
0	5	0	0	0	0
5	10	36	24	8	20
10	15	94	119	39	84
15	20	267	337	127	200
20	25	312	346	189	224
25	50	807	767	577	597
50	75	593	468	593	531
75	100	260	285	341	352
100	125	144	160	231	260
125	150	105	125	202	214
150	200	115	115	253	179
200	250	82	96	165	99
250	500	118	113	213	197
500	1000	59	40	52	41
1000	2000	8	5	10	2

Tabela 0.2. Distribuição de tamanhos de *Droplet-Nuclei* (Duguid, 1946)

LowDk [μm]	HighDk [μm]	Espirros; Múltiplos; 45,72 cm	Espirros; Múltiplos; 152,4 cm	Tosse; Boca Fechada; Múltiplos; 45,72 cm	Tosse; Boca Fechada; Múltiplos; 152,4 cm	Tosse; Boca Aberta; Múltiplos; 45,72 cm	Falar alto; 45,72 cm
0,25	0,5	215	49	68	8	0	10
0,5	1	904	729	380	55	43	115

1	2	1931	1586	1238	214	520	455
2	4	1270	1627	1713	726	849	677
4	6	420	574	921	386	362	351
6	8	153	227	404	231	143	213
8	10	64	112	204	155	55	110
10	12	25	52	54	78	20	49
12	14	10	23	15	54	7	16
14	16	5	12	2	37	1	3
16	18	3	6	0	22	0	1
18	20	0	3	1	19	0	0
20	25	0	0	0	11	0	0
25	-	0	0	0	4	0	0

Tabela 0.3. Tabela de distribuição de tamanho de gota composto (Duguid, 1946)

LowDk [µm]	HighDk [µm]	Espirros; Singular; 15,24 cm	Tosse; Boca Fechada; Singular; 15,24 cm	Contar alto; 7,62 cm
1	2	26 000	50	1
2	4	160 000	290	13
4	8	350 000	970	52
8	16	280 000	1 600	78
16	24	97 000	870	40
24	32	37 000	420	24
32	40	17 000	240	12
40	50	9 000	110	6
50	75	10 000	140	7
75	100	4 500	85	5
100	125	2 500	48	4

125	150	1 800	38	3
150	200	2 000	35	2
200	250	1 400	29	1
250	500	2 100	34	3
500	1000	1 000	12	1
1000	2000	140	2	0

Tabela 0.4. Valor médio do número de gotas medida por pessoa nas distâncias de 1 e 6 cm.(Chao *et al.*, 2009)

LowDk [μm]	HighDk [μm]	Falar alto; 1 cm	Tosse; Boca Fechada; Múltiplos; 1 cm	Falar alto; 6 cm	Tosse; Boca Fechada; Múltiplos; 6 cm
2	4	1,7	4	5	3,8
4	8	26,8	55	16,4	17,3
8	16	9,2	20,4	6,2	6,5
16	24	4,8	6,7	4,3	2,8
24	32	3,2	2,5	2,6	1,4
32	40	1,6	2,4	1,9	0,6
40	50	1,7	2	1	0,2
50	75	1,8	2	1,4	0,9
75	100	1,3	1,4	1,2	0,5
100	125	1,7	1,7	1,2	1
125	150	1,6	1,6	0,4	0,7
150	200	1,7	4,4	1	0,6
200	250	1,5	2,5	0,4	0,5
250	500	1,4	2,1	0,6	0,9
500	1000	0,5	1,4	0,1	0,4
1000	2000	0	0	0	0

