MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

aa/bb/cc/dd

APLICAÇÕES DE FERRAMENTAS COMPUTACIONAIS NA QUALIDADE DE DADOS METEOROLÓGICOS OBSERVACIONAIS DE MULTI-SENSORES SOBRE A REGIÃO AMAZÔNICA

Thomaz Assaf Pougy

Relatório Final de Iniciação Científica PIBIC, orientada pelos Dr. Alan James Peixoto Calheiros e Prof. Dr. Pedro Luiz Pizzigatti Corrêa.

URL do documento original:
<http://urlib.net/xx/yy>

INPE
São José dos Campos
2022

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GB)
Serviço de Informação e Documentação (SID)
Caixa Postal 515 - CEP 12.245-970
São José dos Campos - SP - Brasil
Tel.:(012) 3945-6923/6921
Fax: (012) 3945-6919
E-mail: pubtc@sid.inpe.br

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

aa/bb/cc/dd

APLICAÇÕES DE FERRAMENTAS COMPUTACIONAIS NA QUALIDADE DE DADOS METEOROLÓGICOS OBSERVACIONAIS DE MULTI-SENSORES SOBRE A REGIÃO AMAZÔNICA

Thomaz Assaf Pougy

Relatório Final de Iniciação Científica PIBIC, orientada pelos Dr. Alan James Peixoto Calheiros e Prof. Dr. Pedro Luiz Pizzigatti Corrêa.

URL do documento original:
<http://urlib.net/xx/yy>

INPE
São José dos Campos
2022

Dados Internacionais de Catalogação na Publicação (CIP)

Sobrenome, Nomes.

Cutter Aplicações de Ferramentas Computacionais na Qualidade de Dados Meteorológicos Observacionais de Multi-Sensores Sobre a Região Amazônica / Nome Completo do Autor1; Nome Completo do Autor2. – São José dos Campos : INPE, 2022.

xiv + 55 p. ; (aa/bb/cc/dd)

Dissertação ou Tese (Mestrado ou Doutorado em Nome do Curso) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, AAAA.

Orientador : José da Silva.

1. Palavra chave. 2. Palavra chave 3. Palavra chave. 4. Palavra chave. 5. Palavra chave I. Título.

CDU 000.000

ATENÇÃO! A FOLHA DE APROVAÇÃO SERÁ INCLUIDA POSTERIORMENTE.
Mestrado ou Doutorado em Nome do Curso

RESUMO

INPE realiza valorosas pesquisas que subsidiam o avanço do conhecimento científico sobre as dinâmicas climáticas e de tempo no Brasil e no mundo, com impactos significativos no planejamento estratégico público e privado nacional. Nesse cenário, garantir a qualidade desses dados impacta diretamente sobre a confiabilidade das previsões e análises geradas a partir deles. Os dados de precipitação são essenciais no conjunto de informações utilizadas nos estudos supracitados. Assim, propôs-se neste trabalho o desenvolvimento de ferramentas para os instrumentos de referência para precipitação, os disdrômetro, no caso o modelo RD80 (Joss-Waldvogel) e o modelo Particle Size and Velocity (PARSIVEL), e o pluviômetro. Tais ferramentas tem o objetivo de auxiliar pesquisadores do INPE e parceiros na: padronização dos dados brutos para formatos internacionalmente aceitos; processamento de figuras para subsidiar análises; análise e tratamento para a qualidade de dados; e, por fim, registo dos metadados e análises de qualidade para publicação em repositórios de dados internacionais, como o do ARM (EUA) e do instituto Max Planck (Alemanha). Além de melhorias importantes na organização computacional dos códigos, também foram desenvolvidos scripts e bibliotecas em Python que convertem os dados brutos dos instrumentos para o formato netCDF4, em conformidade com as diretrizes de estrutura e qualidade de dados do ARM e Instituto Max Plank, para alguns experimentos de campo no Brasil, no caso para o Amazonian Tall Tower Observatory (ATTO). Produziu-se também arquivos de visualização interativas e estáticas dos dados, que auxiliam principalmente na análise rápida dos dados pelos mentores dos equipamentos e pesquisadores. Outro aspecto importante desta pesquisa foi a elaboração de documentos python do tipo notebook explicativos e pré-organizados para apoiar a exploração e análise dos dados, com destaque para o cálculo de estatísticas para validar a qualidade das medidas (e.g., RMSE, correlações e outros). Por fim, com as ferramentas desenvolvidas, foi possível inicialmente avaliar a performance das medidas dos disdrômetros durante o experimento de campo ATTO. Observou-se que o os disdrômetros apresentam alta correlação com as medidas de taxa de chuva capturadas pelos pluviômetros (0,80 e 0,86). Ademais, foi observado um erro máximo de estimativa de 16 mm/h, indicando que o instrumento apresentou performance satisfatória.

Palavras-chave: Ciência dos dados. Microfísica de nuvens. Qualidade de dados.

COMPUTATIONAL TOOLS APPLICATION IN DATA QUALITY FOR OBSERVATIONAL MULTI-SENSOR METEOROLOGICAL DATA ON THE AMAZON BASIN

ABSTRACT

INPE conducts valuable research that subsidizes the advancement of scientific knowledge about climate and weather dynamics in Brazil and worldwide, with significant impacts on national public and private strategic planning. In this scenario, ensuring the quality of these data directly impacts the reliability of the forecasts and analyses generated from them. Precipitation data are essential in the set of information used in the aforementioned studies. Thus, this work proposed the development of tools for the reference instruments for precipitation, the disdrometer, in this case the RD80 model (Joss-Waldvogel) and the Particle Size and Velocity model (PARSIVEL), and the rain gauge. These tools have the objective of supporting researchers from INPE and partners in: standardization of raw data to internationally accepted formats; processing of figures to support analyses; analysis and treatment for data quality; and, finally, registration of metadata and quality analyses for publication in international data repositories, such as the ARM (USA) and the Max Planck Institute (Germany). In addition to important improvements in the computational organization of the codes, scripts and libraries were also developed in Python that convert the raw data from the instruments to the netCDF4 format, in compliance with the ARM and Max Plank Institute data structure and quality guidelines, for some field experiments in Brazil, in this case for the Amazonian Tall Tower Observatory (ATTO). We also produced interactive and static visualization files of the data, which mainly help in rapid data analysis by equipment mentors and researchers. Another important aspect of this research was the elaboration of explanatory and pre-organized python notebook-type documents to support the exploration and analysis of the data, with emphasis on the calculation of statistics to validate the quality of the measurements (e.g., RMSE, correlation, and others). Finally, with the tools developed, it was initially possible to evaluate the performance of the disdrometer measurements during the ATTO field experiment. It was observed that the disdrometers show a high correlation with the rain rate measurements captured by the rain gauges (0.80 and 0.86). Furthermore, a maximum estimation error of 16 mm/h was observed, indicating that the instrument performed satisfactorily.

Keywords: Data Science. Cloud microphysics. Data Quality.

LISTA DE FIGURAS

Pág.

LISTA DE TABELAS

Pág.

SUMÁRIO

Pág.

1 INTRODUÇÃO

Estudos científicos sobre fenômenos atmosféricos no território brasileiro têm relevância não apenas local como também global. Os resultados desses trabalhos são estratégicos, pois servem de guia para gestão de políticas públicas baseadas em evidências. Uma vez que a acurácia das conclusões tomadas é diretamente dependente da confiabilidade dos dados capturados para condução de análises científicas, cada vez mais, são bem vindas iniciativas que contribuam para disseminar na comunidade científica boas práticas de gestão e de qualidade de dados.

Em especial, no cenário da pesquisa com coleta de dados pluviométricos, no qual este projeto está inserido, garantir a qualidade dos dados é um requisito essencial para a realização de análises posteriores. Esse requisito é um desafio importante, dado o grande fluxo de arquivos recebidos por campanha de coleta de dados e informações operacionais.

Este trabalho de iniciação científica se insere neste contexto de forma a contribuir para comunidade científica do INPE e seus parceiros, fornecendo ferramentas computacionais que suportem as boas práticas de gestão de dados. Com atenção especial à qualidade de dados de instrumentos de medida para captura de dados de precipitação.

É importante ressaltar que este trabalho constitui uma continuação do projeto realizado entre setembro de 2020 a agosto de 2021. Inicialmente, foram propostos algoritmos de preparação de dados para um instrumento de medição (Micro Rain Radar, MRR) e durante o desenvolvimento observou-se a oportunidade de potencializar o impacto geral da pesquisa. Neste caso, explorar a elaboração de ferramentas de tratamento de dados para mais três instrumentos de medida de precipitação, que são complementares ao equipamento inicialmente estudado.

Assim, essa continuação prevê objetivos relacionados aos três instrumentos que serão objeto de estudos, tanto para preparação de dados, quanto para análise de qualidade (*flags* de qualidade).

Por fim, é importante ressaltar que o projeto se insere no contexto de uma parceira com a Escola Politécnica da USP, Laboratório de Física Atmosférica do Instituto de Física da USP e o *Atmospheric Radiation Measurement Climate Research Facility* (ARM/ARM-DoE) (ARM, 2017). O ARM é um programa do governo norte-americano que desenvolve estudos sobre a atmosfera e possui um departamento es-

pecífico para determinar diretrizes para gestão de qualidade dos dados(ARM, 2017). O ARM recebe dados de pesquisas conduzidas pelo próprio instituto ou por instituições parceiras em outros países em seu sistema de *Data Delivery* e agrega ao total cerca de 10PB de dados, sua estrutura de gestão de dados e práticas de qualidade de dados foram utilizadas como base para o desenvolvimento das ferramentas propostas neste estudo.

1.1 Objetivos

Atualmente estamos na era dos dados, big data se tornou uma das palavras mais citadas por institutos de pesquisa e empresas privadas. O INPE realiza junto com seus parceiros uma série de experimentos de campo para entender as características físicas da atmosfera de modo a melhorar suas previsões de tempo e clima, assim como, para validar dados de satélites. Contudo, ainda existe um longo caminho até que esta gama de dados esteja pronta para responder questões científicas. Deste modo, este projeto tem como objetivo propor e implementar ferramentas de gestão, controle e análise da qualidade de dados obtidos por sensores que coletam dados da atmosfera. Neste caso refere-se a informações sobre a chuva na Bacia Amazônica.

Partindo da problemática abordada no projeto de 2021, este trabalho propõem-se a dar continuidade aos objetivos elencados no ano anterior para os quais foi identificado uma oportunidade de complementação que potencializasse o impacto do projeto na comunidade científica do INPE, USP e Brasil.

Os objetivos identificados como elegíveis à complementação foram os objetivos B e C do projeto de 2021.

O objetivo B determinava a elaboração de uma ferramenta para aplicação de *flags* de qualidade de dados e condensação de informações de verificação por meio de palavras binárias. As *flags* seriam construídas para cada linha de dados, com a aplicação de testes de qualidade para cada instante de dados registrado, e o formato de registro dessas informação seria por meio de uma técnica conhecida como Flag by Bit Packing, já utilizada atualmente no ARM — a técnica permite que um conjunto de dados passe por diversos níveis de validação de qualidade e receba flags a cada etapa de processamento para linhas de dado aprovadas ou reprovadas. Essa técnica evita a exclusão de dados entre níveis de validação, evita a criação demasiada de colunas no conjunto de dados e também permite mais controle do usuário sobre os dados que utilizará.

Assim esse trabalho previu a elaboração de ferramentas para implementação de flags de qualidade nos dados de interesse do projeto, seguindo a metodologia supracitada.

Ademais, com relação ao objetivo C do trabalho de 2021, vale destacar que o objetivo foi considerado completo no trabalho realizado, uma vez que foi entregue um toolkit completo para preparação (*data prep*) e visualização (*data viz*) de dados. As ferramentas entregues geram arquivos netCDF a partir dos dados brutos do instrumento MRR e também geram figuras de visualização para apoiar a análise dos dados. Assim, ao longo do desenvolvimento do projeto, foi identificado a oportunidade de complementar o projeto, de forma a desenvolver essas ferramentas de processamento para outros três instrumentos.

Os instrumentos identificados, importantes na análise atmosférica, são instrumentos de estimativa e medida de chuva. São eles o disdrômetro de impacto RD-80 (Joss-Waldvogel), o disdrômetro a laser PARSIVEL^2 e o pluviômetro. Dessa forma, este projeto visa o desenvolvimento de ferramentas de processamento e visualização de dados para os três instrumentos.

Os dados utilizados neste trabalho são baseados no projeto temático GOAmazon (Green Ocean Amazon) (MARTIN et al., 2016) que estuda as características físicas da atmosfera na região amazônica, baseado numa parceria internacional da qual o INPE é um dos líderes. E informações de outros experimentos, como o SOS-CHUVA (MACHADO, 2015).

De modo a seguir os padrões internacionais de qualidade das informações geradas por sensores para ciências atmosféricas, este trabalho será baseado no sistema de coleta, processamento e qualidade de dados do ARM.

Os dados utilizados nesta pesquisa foram aqueles associados a sensores que determinam a distribuição de gotas de chuva na superfície (disdrômetros de superfície e um MRR). Estes dados estão disponíveis nos repositórios do ARM, USP e INPE.

2 DESENVOLVIMENTO

2.1 Revisão Bibliográfica

Uma vez que (i) este trabalho se dá como continuação ao trabalho desenvolvido em 2021 e (ii) os objetivos são similares aos já desenvolvidos, foi possível basear grande parte do trabalho sobre a revisão bibliográfica já realizada. Utilizar a pesquisa produzida em 2021 foi importante para subsidiar a continuidade da utilização das metodologias que guiaram os trabalhos deste projeto.

Nesse contexto, foi reaproveita a revisão a respeito da gestão de qualidade de dados e também sobre dados de precipitação. Por outro lado, foi necessário complementar a revisão para incluir também os conceitos e paradigma de dados dos três novos sensores que são o enfoque deste trabalho.

Os tópicos 2.2 e 2.3 registram os principais pontos da revisão bibliográfica.

2.2 Gestão de Dados e Gestão de Qualidade dos Dados

Garantir boas práticas de gerenciamento em todo o ciclo de vida de dados desde a ingestão até o processamento e análise perpassa por adotar um claras diretrizes de Gestão de Dados. Essa premissa se mostra ainda mais importante no cenário da pesquisa científica, no qual os dados capturado subsidiam conclusões importantes sobre sociedade e meio ambiente, orientando a tomada de decisão de indivíduos e instituições.

Como abordado por Silva (2020), a gestão de dados:

> Refere-se àquelas atividades relacionadas à gestão ativa de dados durante o tempo que continuam a ter interesse acadêmico, científico, administrativo e pessoal, a fim de favorecer sua reprodução, reutilização e agregação de valor, os dados são gerenciados desde a sua criação até que é determinado que eles não são mais úteis, garantindo a sua acessibilidade a longo prazo, sua conservação, sua autenticidade e sua integridade.

Assim, adotar boas práticas de curadoria, gerenciamento e armazenamento dos dados faz parte da definição de uma estratégia robusta de Gestão de Dados a ser adotada.

Como apresentado por Anjos G. A. Dias (2017) no contexto acadêmico brasileiro ainda há uma parcela significativa de pesquisadores, correspondente a 31% da comunidade, que não aplicam plenamente diretrizes robustas de gestão de dados sobre as informações capturadas em suas pesquisas. Dessa forma, iniciativas que consideram

a definição de uma estratégia de gestão de dados como parte do objetivo final de projeto contribuem para disseminar a metodologia de gestão de dados no cenário da pesquisa Brasileira.

Ademais, é importante esclarecer que um dos aspectos a ser considerado na elaboração da estratégia de gestão de dados deve ser a gestão de qualidade de dados. Ela é constituída pelas práticas de análise de qualidade e gerenciamento de metadados de qualidade. Como analisado pela referência Kwon Ohbyung; Lee (2014) o emprego de diretrizes claras de gestão de qualidade de dados em ambientes com abundância de dados é um fator que potencializa a execução de análises mais complexas sobre os dados, incentivando a avaliação de relações múltiplas entre variáveis de origem distintas.

2.3 Dados de precipitação

Os dados utilizados neste trabalho são produzidos por equipamentos de estimativa e medida de taxa de chuva. Os disdrômetros são instrumentos de estimativa de taxa de chuva que registram dados de distribuição de gotas, que por sua vez são utilizados para estimar a taxa de chuva por meio de relações matemáticas previamente estabelecidas (JOEL, 2011) (ISLAM et al., 2012). Já o pluviômetro é o instrumento de referência para medida de taxa de chuva uma vez que registra dados diretos sobre a precipitação em um intervalo determinado.

A escolha dos três instrumentos supracitados como enfoque deste trabalho é oportuna uma vez que permite a condução de validação cruzada entre os instrumentos. Dessa forma é possível avaliar o desempenho e confiabilidade dos dados produzidos pelos instrumentos de estimativa de chuva por meio de estatísticas de validação cruzada que demostrem a distância do dado de estimativa em comparação com o dado da observação (i.e. referência).

O funcionamento dos instrumentos e o formato de dados registrados por cada um deles será descrito com maiores detalhes nas seções 2.3.1, **??**, **??**.

2.3.1 Disdrômetro RD-80 (Joss-Waldvogel)

O disdrômetro RD-80 é um disdrômetro do tipo impacto, equipamento que mede a distribuição de gotas de chuva por meio do impacto delas sobre um anteparo. Ele utiliza o princípio estabelecido primeiramente por Joss J.and Waldvogel (1967), que registra a força aplicada por gotas de chuva sobre um transdutor. Dessa forma, considerando relações entre tamanho, velocidade e formato de gota o instrumento é

Figura 2.1 - Disdrômetro RD-80

Fonte: Retirada de (RD-80..., 2018)

capaz de gerar medidas de estimativa de distribuição de tamanho de gotas (DSD - *Drop Size Distribution*) a partir da informação da força aplicada (KINNELL, 1976). A figura **??** apresenta o equipamento, que consiste em um sensor que fica exposto à chuva e uma unidade de processamento central, responsável por realizar os cálculos e registrar as medidas.

A DSD é o principal dado registrado pelo disdrômetro e é dependente direto da confiabilidade das relações matemática supracitadas, contudo essas relações foram determinadas em laboratório para gotas viajando em velocidade terminal e na prática a influência de fatores ambientais (vento, flutuações atmosféricas e outros) levam a variações de formato, tamanho e trajetória da gota. Essas questões, bem como outras limitações intrínsecas ao instrumento, implicam em imprecisões das medidas registradas (KINNELL, 1976) (ISLAM et al., 2012). O DSD registrado pelo RD-80 é um vetor de 20 posições em que cada uma representa a quantidade de gotas registradas no intervalo de tempo da medida para uma classe de tamanho de gota que varia de 0,359mm a 5,373mm. Uma vez calculada, a DSD é utilizado para calcular a estimativa da taxa de chuva e outros parâmetros associados a microfísica da chuva.

A resolução temporal das medidas do equipamento utilizadas neste trabalho (dados provenientes da campanha GOAmazon) é de 1 minuto.

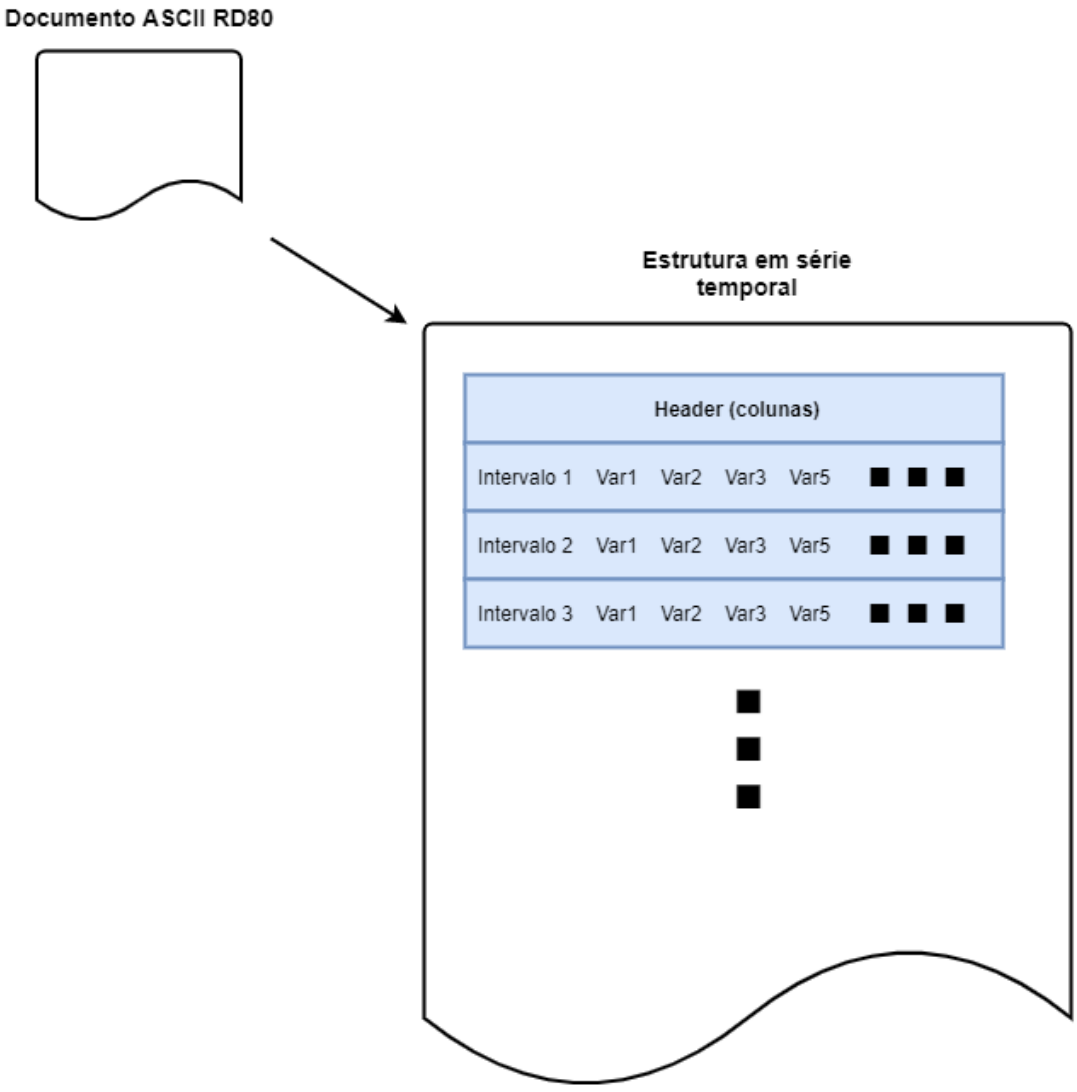
2.3.1.1 Estrutura dos dados do RD-80

Para que seja possível armazenar os dados capturados pelo instrumento em um arquivo netCDF, se faz necessário extrair os dados de cada variável a partir dos

formatos de arquivo específicos do equipamento. Nesse contexto, é fundamental entender a estrutura desse formato específico.

Os arquivos de dados exportados pelo RD-80 são arquivos de texto ASCII, na extensão .trf e .txt, e a estrutura geral é baseada em colunas. Nelas, a primeira linha corresponde ao cabeçalho com o nome das colunas do arquivo e a partir da segunda linha é registrado as medidas para cada intervalo (a primeira e segunda colunas são o índice temporal do arquivo). A Figura 2.2 apresenta um diagrama que resume a estrutura do arquivos de dados do RD-80.

Figura 2.2 - Estrutura do arquivo bruto do equipamento Disdrômetro Joss



Fonte: Elaborada pelo autor.

A partir da referencia DISTROMET (2009) foi possível elaborar a Tabela 2.1 que apresenta as principais informações de cada variável registrada nas colunas do arquivo.

Tabela 2.1 - Variáveis por coluna do arquivo bruto do equipamento Disdrômetro RD-80

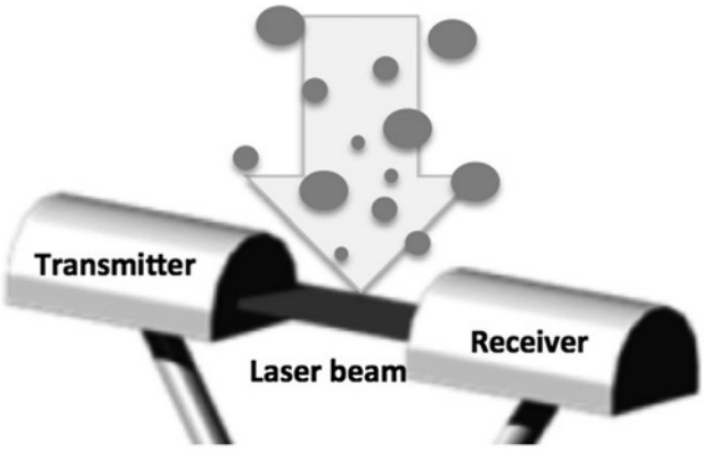
Coluna	**Identificador**	**Significado**
1	YYYY-MM-DD	Identificador da data da medida
2	hh:mm:ss	Identificador do horário da medida
3	Status	Status do equipamento
4	Interval [s]	Intervalo de contagem para as medidas
5 - 25	n1 - n20	Número de gotas contadas no intervalo para os 20 perfis de tamanho de gota
26	RI [mm/h]	*Rain Intesity* ou *Rainfall Rate*
27	RA [mm]	*Rain Amount*
28	RAT [mm]	*Total Rain Amout since start of measurement*

Fonte: Adaptada de DISTROMET (2009)

2.3.2 Disdrômetro PARSIVEL^2

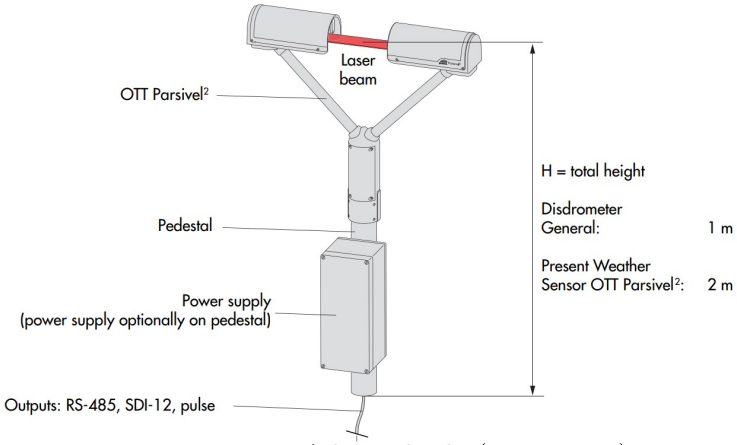
O disdrômetro PARSIVEL^2 consiste em um disdrômetro do tipo laser, equipamento de estimativa do perfil de gotas de chuva por meio do registro da passagem de gotas por uma área coberta com feixes de laser (OTT, 2016). A Figura 2.3.2 ilustra o principio de funcionamento do equipamento e a Figura 2.3.2 apresenta uma perspectiva completa dele.

Figura 2.3 - Princípio de funcionamento do Disdrômetro PARSIVEL[2]



Fonte: Adaptada de (FRIEDRICH et al., 01 Sep. 2013).

Figura 2.4 - Disdrômetro PARSIVEL[2]



Fonte: Adaptada de (OTT, 2016).

Como é possível observar nas Figuras 2.3.2 e 2.3.2, o instrumento é constituído por uma estrutura única que deve ser instalada em local exposto à chuva, esta por sua vez é composta por dois módulos. O transmissor, que emite os feixes de laser para cobrir a área determinada e o receptor que registra o sinal dos feixes. Assim, conforme apresentado por Tokay et al. (01 Jun. 2014), O tamanho das gotas que passam pela área é estimado por meio da atenuação do sinal registrado pelo receptor e a velocidade é calculada a partir da duração da atenuação do sinal (intervalo de

tempo em que a gota passa pelo feixe). Nesse contexto, é importante destacar que o equipamento assume formato esférico para gotas com diâmetro menor que 0,1 mm e para gotas maiores (até 5 mm) assume uma variação linear da variação do eixo vertical do formato elipsar da gota.

A área coberta pelo laser tem aproximadamente 180 mm de comprimento por 30 mm de largura e o feixe do laser por si próprio possui 1 mm de altura.

A principal medida fornecida pelo PARSIVEL^2 é equivalente a DSD do RD-80, porém com maior resolução de classe de tamanho de gota e com a inclusão da informação da velocidade das gotas registradas para o intervalo de tempo da medida. Assim, essa medida do PARSIVEL^2 consiste em uma matriz 32x32 de tamanho de gota *versus* velocidade. Cada posição registra a contagem de gotas de cada classe de tamanho (linhas) que passaram pela área do laser no intervalo de tempo da medida para cada velocidade (colunas). As colunas de tamanho de gota incluem tamanhos de 0 mm a 25 mm e variam em intervalos de 0,125 mm a 3 mm. As linhas de velocidade incluem velocidades de 0 a 20 m/s.

Devido à limitações intrínsecas ao equipamento e características naturais do comportamento de gotas, o PARSIVEL^2 não é capaz de fornecer a medida exata da DSD, assim os parâmetros registrados são uma estimativa da medida da DSD e, portanto, há uma imprecisão intrínseca aos dados (TOKAY et al., 01 Jun. 2014).

Como forma de mitigar a imprecisão das medidas registradas pelo disdrômetro a laser, é possível aplicar uma matriz de correção baseada na filtragem de condições físicas não realísticas (e.g. gotas pequenas com alta velocidade) a cada linha de dado e como resultado, obtendo-se assim dados mais confiáveis.

Assim, se faz necessário que este trabalho preveja que a ferramenta de processamento de dados proposta para o equipamento suporte a aplicação da matriz de correção.

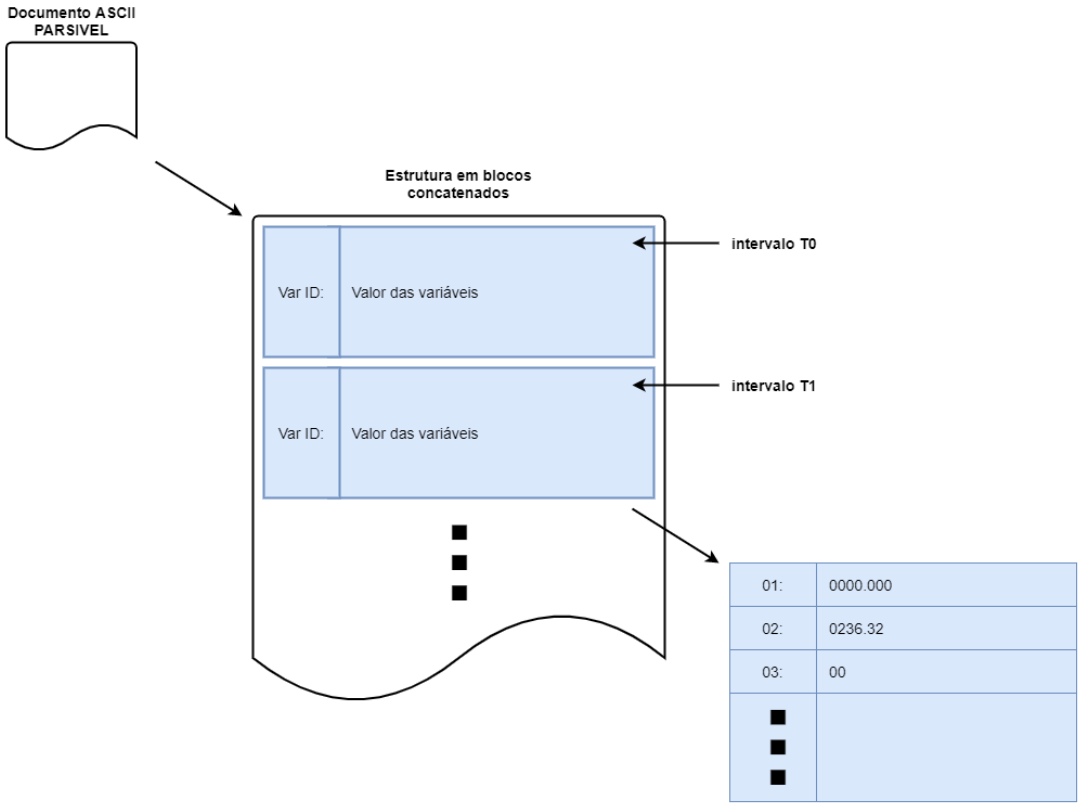
2.3.2.1 Estrutura dos dados do PARSIVEL^2

Para que seja possível armazenar os dados capturados pelo instrumento em um arquivo netCDF, se faz necessário extrair os dados de cada variável a partir dos formatos de arquivo específicos do equipamento. Nesse contexto, é fundamental entender a estrutura desse formato específico.

O arquivo de dados exportado pelo PARSIVEL^2 é um arquivo de texto ASCII sem extensão, e a estrutura geral se dá por meio de blocos de dados. Em cada bloco há

98 linhas e cada qual corresponde a uma variável capturada pelo equipamento no intervalo da medida. Cada linha de variável possui um formato específico, mas para variáveis multi dimensionais, como a DSD citado anteriormente no tópico 2.3.2, a separação de valores para cada dimensão é feita por meio do separador ";". Já a o ponto "."é utilizado como separador decimal. A Figura 2.5 apresenta um diagrama que resume a estrutura do arquivo de dados do PARSIVEL[2].

Figura 2.5 - Estrutura do arquivo bruto do equipamento Disdrômetro PARSIVEL[2]



Fonte: Elaborada pelo autor.

A partir da referencia OTT (2016) foi possível elaborar a Tabela 2.2 que apresenta as informações essenciais das principais variáveis registradas nos blocos de dados do arquivo.

Tabela 2.2 - Variáveis por linha em bloco de arquivo bruto do equipamento Disdrômetro PARSIVEL[2]

Linha	**Significado**
1	*Rain Intesity* ou *Rainfall Rate*
7	Radar Reflectivity
9	Intervalo de contagem para as medidas do bloco
13	Número serial do equipamento
20	Horário das medidas do bloco
21	Data das medidas do bloco
25	Código de erro
93	DSD Bruto

Fonte: Adaptada de DISTROMET (2009).

2.3.3 Pluviômetro

O pluviômetro objeto de estudos deste trabalho é o modelo TB4 Série II, produzido pela Hyquest Solutions. Existem diversos tipos de pluviômetro, que exploram princípios diferentes para aferirem a taxa de chuva diretamente. O instrumento em questão é do tipo balde basculante (*tipping bucket rain gauge*) e conforme descrito por Wang et al. (01 Jan. 2008), consiste em um mecanismo simples que mede a taxa de chuva em um determinado local em incrementos de 0,254 mm (volume do recipiente basculante). A medida produzida pelo instrumento, é, portanto, o registro dos instantes de tempo em que houve pivotamento do recipiente, assim é possível calcular a taxa de chuva em mm/h.

A figura 2.3.3 mostra a montagem do equipamento que fica exposta à precipitação e a figura 2.3.3 apresenta o mecanismo basculante responsável por registrar as medidas.

Figura 2.6 - Montagem do equipamento pluviômetro

Fonte: Retirado de Fondriest (2022).

Figura 2.7 - Mecanismo basculante do pluviômetro

Fonte: Retirado de Fondriest (2022).

2.3.3.1 Estrutura dos dados do pluviômetro

Os dados produzidos pelo equipamento são registrados em arquivos .csv com estrutura simples.

A partir da referencia HYQUEST (2019) foi possível elaborar a Tabela 2.3 que apresenta as principais informações das variáveis registrada nas colunas do arquivo.

Tabela 2.3 - Variáveis registradas no arquivo do pluviômetro

Coluna	**Variável**
1	Índice
2	Data e Horário do evento
3	Evento (mm)
4	Anotações

Fonte: Elaborada pelo autor.

2.4 Metodologia

Dada a continuação do trabalho do ano anterior, foi previsto que este estudo seguisse a metodologia proposta . Assim, a elaboração de cada ferramenta seguiu diretrizes de metodologia bem estabelecida e cada aspecto considerado na produção dos scripts teve práticas e métodos definidos.

2.4.1 Diretrizes gerais

Para guiar a elaboração das ferramentas propostas neste estudo foi considerada uma diretriz geral para alinhar as diversas frentes de trabalho. Os dois pontos principais que compõem a diretriz são apresentados abaixo.

- O produto processado pelas ferramentas deve estar em acordo com padrões e normas internacionais, tendo em vista a facilitação da publicação destes em repositórios internacionais. Nesse contexto, o padrão de referência a ser utilizado são as práticas e estrutura de gestão de dados do ARM-DoE.

- O desenvolvimento das ferramentas seguiu a metodologia em cascata, tradicional para o desenvolvimento de software. Assim, o desenvolvimento das ferramentas que compõem o Toolkit proposto partiu da prototipação da lógica para seguir a sua implementação. Logo, os scripts foram preparado para a sua execução automática.

Os próximos tópicos apresentam questões específicas do desenvolvimento do projeto que se baseiam na diretriz geral supracitada.

2.4.2 Gestão de dados

Para facilitar a posterior distribuição e manutenção dos scripts de processamento de dados desenvolvidos foi elaborada uma metodologia de gestão de dados, a fim de especificar tanto o processamento a qual os dados são submetidos quanto padronizar outros aspectos relacionados como nome de arquivos e variáveis.

Assim, como realizado no projeto passado, o registro da gestão de dados para o cenário específico de cada equipamento foi feito por meio da elaboração de um *Handbook* do equipamento (desenvolvido pelo orientador principal em parceira com o bolsista deste projeto) de forma a suportar as atividades do mentor do equipamento e consumidores dos dados durante e depois de campanhas de coleta de dados. Este

Handbook foi inicialmente desenvolvido para as atividades associadas as medidas realizadas no sítio da campina a 10 km da torre ATTO na região amazônica.

2.4.2.1 Formato de arquivo

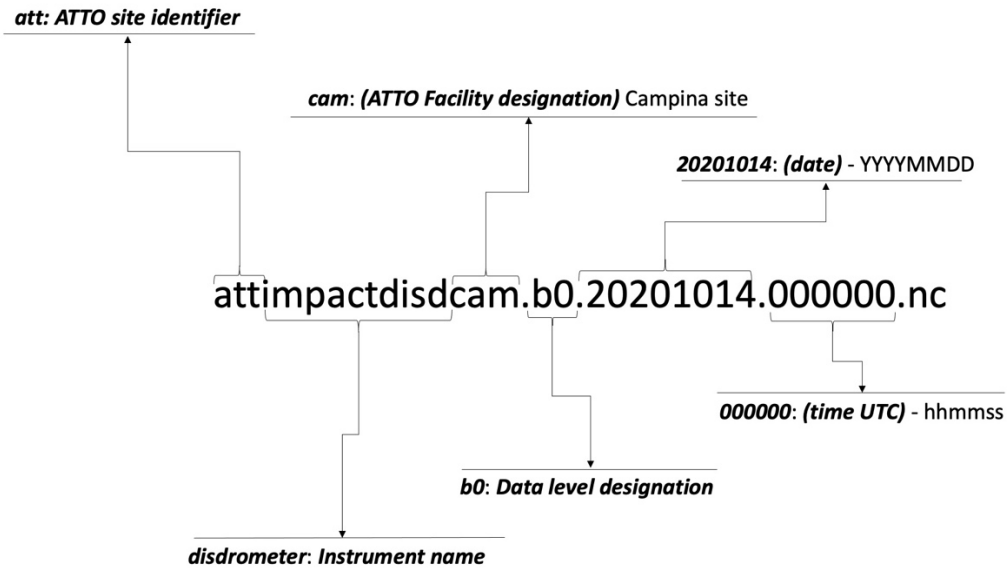
O formato de arquivo utilizado para armazenar e distribuir os dados processados dos instrumentos, seguiu a diretriz determinada no projeto anterior, dessa forma, foi utilizado formato netCDF4. Esse formato foi criado pela Unidata na década 90, possui grande relevância na comunidade científico-atmosférica internacional e implementa uma abstração de dados que modela um *dataset* como uma coleção de variáveis multidimensionais, acompanhadas de seus metadados complementares (REW; DAVIS, 1990).

O formato netCDF4 é o principal padrão de arquivo utilizado para a distribuição de dados no ARM-DoE. Ele permite a integração direta de dados com informações adicionais, característica especialmente importante que garante maior confiabilidade na entrega de metadados.

Ademais, baseado no modelo implementado pela referência ARM (2016) foi elaborado um esquema de nomenclatura de arquivos para cada instrumento que segue a especificação apresentada pelas figuras 2.8, 2.9 e 2.10.

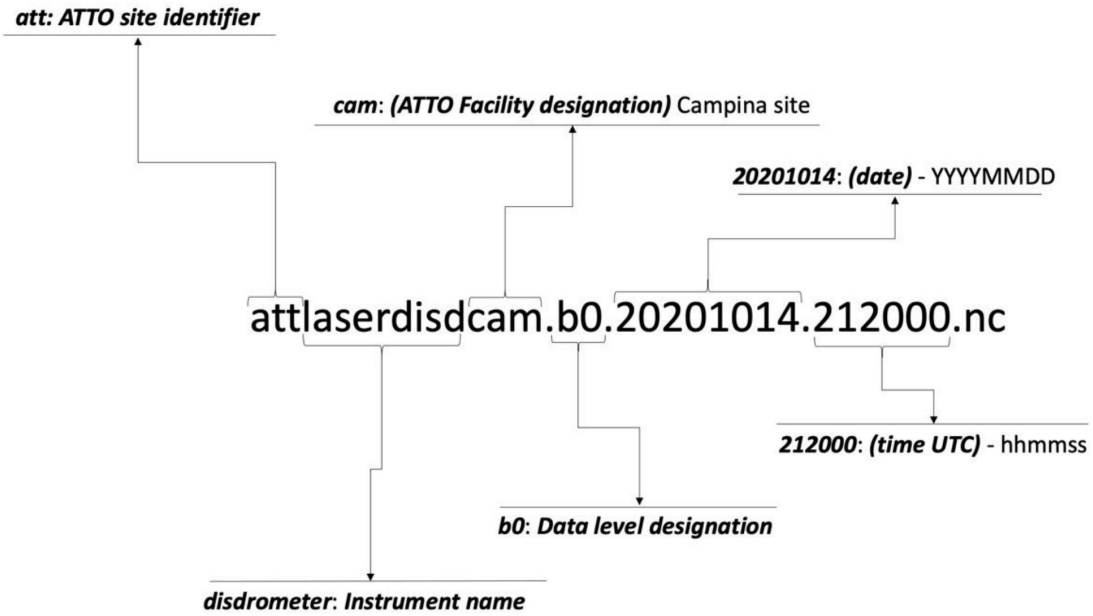
Para a identificação do instrumento no esquema de nomenclatura de arquivos foram utilizadas abreviações iguais aquelas utilizadas no ARM. Para o disdrômetro RD-80 utilizou-se *disdrometer* para o disdrômetro PARSIVEL^2 utilizou-se *laserdisd* e para o pluviômetro utilizou-se *raingauge*.

Figura 2.8 - Esquema de nomenclatura de arquivos para o disdrômetro RD-80



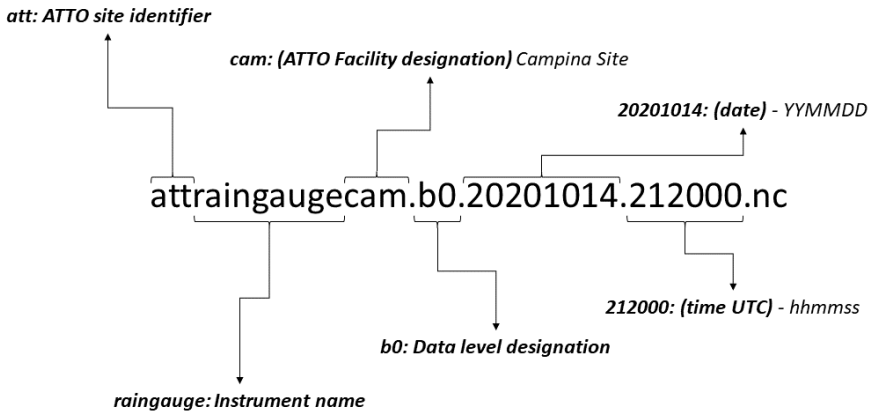
Fonte: Elaborada pelo autor.

Figura 2.9 - Esquema de nomenclatura de arquivos para o disdrômetro PARSIVEL[2]



Fonte: Elaborada pelo autor.

Figura 2.10 - Esquema de nomenclatura de arquivos para o pluviômetro



Fonte: Elaborada pelo autor.

2.4.2.2 Dimensões e variáveis

A definição das variáveis de interesse de cada instrumento a serem incluídas nos arquivos processados e suas informações complementares foram registradas no *Handbook* de cada instrumento, que foi elaborado pelo orientador deste trabalho uma vez que este também atua como mentor dos equipamentos que são objeto de estudos deste projeto, como citado no item 2.4.2. As Tabelas 2.4 e 2.5 apresentam as dimensões das variáveis capturadas pelos instrumentos e as Tabelas 2.6, 2.7 e **??** exibem um destaque das variáveis de interesse e suas unidades.

Tabela 2.4 - Dimensões das variáveis capturadas pelos disdrômetros

Dimensão	**Descrição**
time	Dimensão temporal
drop_class	Dimensão de classe de tamanho de gota e velocidade

Fonte: Elaborada pelo autor.

Tabela 2.5 - Dimensões das variáveis capturadas pelo pluviômetro

Dimensão	**Descrição**
time	Dimensão temporal

Fonte: Elaborada pelo autor.

Tabela 2.6 - Destaque para as variáveis incluídas no arquivo netCDF para dados do disdrômetro RD-80

Variável	**Dimensão**
RI (Rain Intensity)	time
W (Liquid Water Content)	time
EK (Kinectic Energy)	time
N(Di) (Número de gotas por drop_class)	time x drop_class (20)

Fonte: Elaborada pelo autor.

Tabela 2.7 - Destaque para as variáveis incluídas no arquivo netCDF para dados do disdrômetro PARSIVEL^2

Variável	**Dimensão**
Rain Intensity	time
Radar Reflectivity	time
Drop Size Distribution (volume equivalent diameter)	time x drop_class x drop_class (32x32)

Fonte: Elaborada pelo autor.

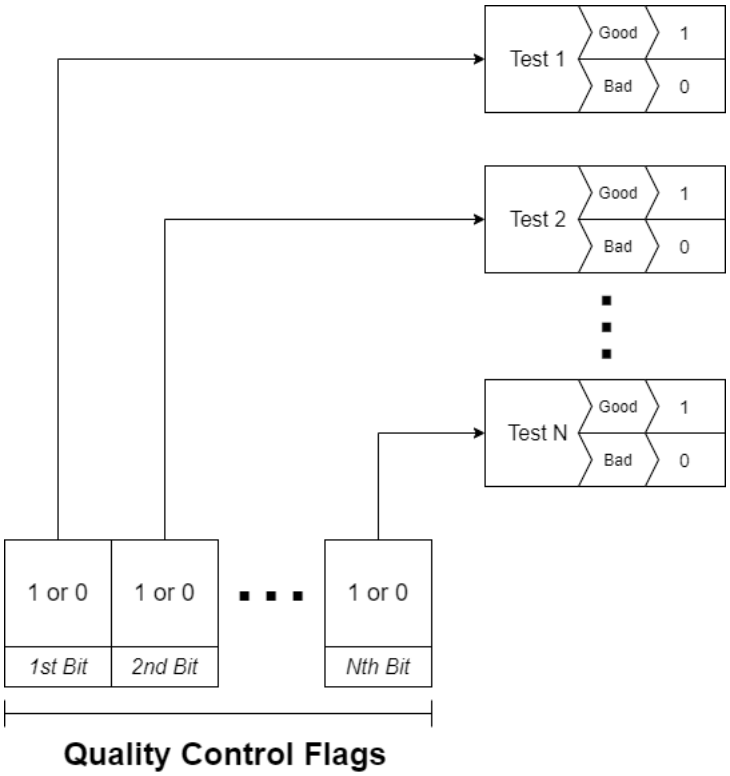
2.4.3 Metadados

Os metadados são dados que trazem informações auxiliares sobre conjunto de dados de interesse. São exemplos de informações complementares relevantes ao dado, a localização, detalhes do equipamento de medida e outros. Os metadados de qualidade por sua vez, são declarações que trazem informações sobre a qualidade de um conjunto de dados (*dataset*) ou sobre um instante/intervalo de medida dentro do *dataset* considerado.

No esquema de gestão de dados proposto neste trabalho são previstos os metadados gerais do arquivo e também metadados de qualidade sobre cada linha de dado registrado. Esse segundo tipo de metadado foi implementado por meio *flags* de qualidade. A distribuição dos metadados de cada arquivo é feita diretamente pelo formato de arquivo netCDF4, que facilita a inclusão desses dados auxiliares no mesmo pacote que carrega os dados principais.

As *flags* de qualidade representam o resultado de um teste de qualidade aplicado linha a linha e foram incluídas diretamente no arquivo netCDF, conforme supracitado. Vale destacar que, baseado nas práticas do ARM DoE, optou-se por fazer a inclusão de múltiplas *flags* por meio da técnica de *bit-packing*. Quando aplicada, esta técnica consiste na inclusão de uma nova coluna de dados e em cada linha desta é inserida uma palavra binária cujos bits registram, individualmente, o resultado de um teste de controle de qualidade aplicado à linha (0 representa reprovado e 1 aprovado). Assim, por exemplo, é possível com uma palavra de 3 bits registrar o resultado da aplicação de 3 testes de qualidade. Note que neste caso, apenas uma coluna é inserida no conjunto de dados, de forma a otimizar o armazenamento e organização dos arquivos. A Figura [2.11 apresenta um diagrama que resume a estrutura de cada *flag* de controle de qualidade.

Figura 2.11 - Estrutura da palavra binária que implementa a *flag* de controle de qualidade



Fonte: Elaborada pelo autor.

2.4.4 Preparação de dados

A preparação de dados (*data prep)* constitui uma ou mais etapas trabalhosas porém imprescindíveis à distribuição final de dados, que serão utilizados para a análise e tomada de conclusões. Essa etapa é responsável por gerenciar tanto os diferentes formatos de dados, específicos de cada instrumento, como também os erros instrumentais que podem se apresentar ocasionalmente em linhas de dados.

Assim diversas lógicas de extração, compilação e registro de informações são aplicadas de forma encadeada sobre os dados brutos dos instrumentos para que se tenha como produto dados em formato padronizado e com qualidade dentro das especificações estabelecidas. A aplicação correta da etapa de preparação assegura que os dados que chegam efetivamente ao processamento e análise estão de acordo com as diretrizes de formato e qualidade.

A garantia dos aspectos supracitados é fundamental para fortalecer o reaproveitamento de código nos *scripts* de tratamento e avaliação, além de potencializar a

otimização da lógica utilizada.

Os instrumentos objeto de estudo deste trabalho, exportam os dados capturados em um formato complexo e ainda, nesses arquivos há a possibilidade de existirem problemas de dados faltantes associados a falhas no sistema de medida. Assim, a metodologia para implementação da lógica para conversão do formato de dados brutos do instrumento em dados netCDF levou em consideração possíveis erros ainda não detectáveis, a partir da suposição do que é esperado de um arquivo sem problemas. Contudo, alguns erros ainda podem ser reportados pelos usuários ao longo da utilização dos dados, sendo estes tratados em novas versões dos algoritmos desenvolvidos.

A preparação de dados foi implementada por meio da elaboração de *scripts* na linguagem Python que recebem dados brutos dos equipamentos e produzem dados padronizados conforme a metodologia apresentada anteriormente. O desenvolvimento das lógicas que compõem cada *script* de preparação de dados se deu inicialmente em uma instância de protótipo para posteriormente ser incluída em um arquivo final para implantação.

2.4.4.1 Prototipação e documentação

A fase de prototipação das lógicas de conversão do formato de dados dos instrumentos para o netCDF neste trabalho se deu de forma equivalente ao desenvolvimento executado no projeto anterior. Assim foram utilizados documentos do tipo notebook editados por meio da plataforma Jupyter para desenvolver os algoritmos de preparação de dados. Essa mesma dinâmica de prototipação também foi utilizada para desenvolver o *scripts* de visualização de dados dos equipamentos.

Ainda em linha com a metodologia utilizada no projeto do ano anterior, este trabalho também previu a elaboração de notebooks explicativos, utilizando-se também a plataforma Jupyter para tal, de forma à apresentar a lógica implementada no código de preparação de dados. Vale destacar que o objetivo desta estratégia é facilitar evoluções posteriores das ferramentas e também incentivar o uso dessa metodologia e linguagem por mais pesquisadores.

2.4.4.2 Implantação

Para a elaboração da versão final dos *scripts* de preparação de dados e geração de figuras de análise rápida foi considerado como principal requisito de funcionalidade que fosse possível a automação da execução das ferramentas por meio do uso de

scripts shell. Assim, para atender à essa funcionalidade, a metodologia de implantação das lógicas na versão final do *script* em Python previu que a elaboração de ferramentas que sejam executadas pelo CLI (*command line interface*).

Ademais, dada a complexidade das ferramentas implementadas e dos cenários de aplicação, existem situações na qual o usuário necessita aplicar lógicas ligeiramente diferentes para os dados. Assim, é necessário que exista uma interface entre usuário e ferramenta que permita a ele selecionar qual lógica será aplicada na execução em questão.

Assim, considerando os requisitos de execução via CLI e necessidade de interface com o usuário concluiu-se que a implementação de *flags* no comando de execução atenderia a ambos requisitos.

Nessa implementação com *flags* de execução, o *script* espera que o comando de execução no CLI seja acompanhado de *flags*, como por exemplo -s ou –standard, que permitem que ocorra a execução condicional da lógica.

Por fim, optou-se por utilizar o paradigma de programação funcional para o desenvolvimento das lógicas.

2.4.5 Controle de qualidade dos dados

O controle de qualidade dos dados visa garantir que aqueles distribuídos aos usuários estejam de acordo com as diretrizes de qualidade especificadas. A metodologia de controle de qualidade dos dados adotada neste trabalho considera que o mesmo se insere em diferentes etapas do processamento dos dados.

Primeiramente, durante a preparação de dados, há a identificação de falhas e aplicação de filtros para erros instrumentais. Em um segundo momento, já durante o processamento efetivo dos dados, há a inserção de informações de validação geral dos arquivos e também de *flags* de controle de qualidade de dados que classificam as linhas de medida com relação a testes de qualidade aplicados sobre elas. Esses testes podem ser simples, que consideram métricas intrínsecas ao arquivo estudado, ou de maior complexidade, que utilizam validação cruzada entre instrumentos de estimativa e referência de taxa de chuva, como é o caso do disdrômetros e pluviômetros, para identificar inconsistências nas medidas.

2.4.5.1 Dados faltantes

Uma vez que existam dados faltantes é necessário definir uma diretriz para tratamento destes nos scripts de processamento. A metodologia adotada neste estudo, segue aquela utilizada no projeto anterior, assim prevê a atribuição da constante -999.99 às variáveis faltantes em linhas de dados afetadas.

2.4.5.2 Validação Cruzada

A validação cruzada é uma etapa significativa na análise de qualidade de dados da pesquisa atmosférica. Ela permite identificar efeitos secundários que impactam a confiabilidade de medidas dos equipamentos estudados.

No contexto deste trabalho, no qual o instrumento objeto de estudos são os disdrômetros e pluviômetros, temos que o pluviômetro é adequado para ser utilizado como referência no cálculo das estatística de validação cruzada com os disdrômetros e assim tomar conclusões com relação ao desempenho das medidas de estimativa deles.

Isso é possível pois os disdrômetros são equipamentos de estimativa de precipitação, uma vez que as medidas registradas em suas variáveis são obtidas de forma indireta através de relações matemáticas para a grandeza efetivamente capturada pelo sensor (e.g. DSD) enquanto o pluviômetro é um equipamento que captura medidas diretamente da precipitação (taxa de chuva). Assim, observar as estatísticas comparativas entre os dados registrados nesses equipamentos em um mesmo intervalo de tempo permite avaliar a qualidade das medidas dos disdrômetros.

Para explorar a validação cruzada, foi definido como diretriz geral a elaboração de documentos Python do tipo *notebook* uma vez que permitem a execução interativa do código e facilita ajustes na lógica de cálculo caso necessário, contribuindo para a flexibilidade no processo de validação.

2.4.5.3 Estatísticas para validação cruzada

Segundo Wilks (2011) e os estudos realizados durante o SOS-CHUVA (CALHEIROS, 2018) alguns índices estatísticos podem ser aplicados para avaliar a comparação de dados que medem variáveis similares. situação que se aplica para a validação cruzada de interesse para este projeto (i.e., estatísticas adequadas para comparar estimativas e dados de referência).

Da mesma forma que foi desenvolvido no projeto anterior, a metodologia definida para implementar o cálculo de cada uma das estatísticas de validação consistiu em desenvolver funções que recebem vetor de dados unidimensionais, um deles contendo as medidas da variável de interesse do instrumento de estimativa e outro contendo as medidas variável do instrumento de referência, e retorna um valor correspondente à estatística calculada para os dados. As estatísticas escolhidas são apresentadas na Tabela 2.8 (WILKS, 2011).

Tabela 2.8 - Estatísticas escolhidas para a validação cruzada

Identificador	**Nome**	**Estatística**	**Cálculo**	**Intervalo válido**	**Valor ideal**
MAE	*Mean Absolute Error*	Erro absoluto médio	$MAE = \frac{1}{N} \bullet \sum_{i=1}^{N} \lvert F_i - O_i \rvert$	$0 \leq MAE \leq \infty$	0
RMSE	*Root Mean Square Error*	Raiz do erro quadrático médio	$RMSE = \sqrt{\frac{1}{N} \bullet \sum_{i=1}^{N} (F_i - O_i)^2}$	$0 \leq RMSE \leq \infty$	0
CORR	*Correlation Coefficient*	Coeficiente de Correlação	$CORR = \frac{\sum (F-\bar{F}) \bullet (O-\bar{O})}{\sqrt{\sum (F-\bar{F})^2} \bullet \sqrt{\sum (O-\bar{O})^2}}$	$-1 \leq CORR \leq 1$	1
BIAS	Bias	Desvio sistemático	$BIAS = \frac{\frac{1}{N} \bullet \sum_{i=1}^{N} F_i}{\frac{1}{N} \bullet \sum_{i=1}^{N} O_i}$	$-\infty \leq BIAS \leq \infty$	1

Fonte: Adaptada de WMO (2017).

2.5 Resultados e análises

Este trabalho, como continuação do trabalho realizado em 2021, tem como objetivos a elaboração de ferramentas de inclusão de *flags* de qualidade de dados e também o desenvolvimento de *toolkits* de preparação de dados e geração de figuras específicos para dados dos instrumentos disdrômetro RD-80, disdrômetro PARSIVEL^2 e pluviômetro.

A elaboração dessas ferramentas perpassa pelo desenvolvimento de ferramentas para a inclusão das *flags* e *scripts* que implementam a conversão de dados brutos dos instrumentos para netCDF bem como a geração de figuras para análise rápida dos dados. A listagem abaixo resume os resultados obtidos para cada um dos dois objetivos deste trabalho de continuação.

- Ferramenta para aplicação de *flags* de qualidade e condensação de informações de verificação de qualidade por meio de palavras binárias.

  Documentos Python do tipo notebook foram elaborados para subsidiar o

acesso de novos pesquisadores às ferramentas e formato netCDF. Nesses documento há seções com código pré-pronto que permite a inclusão de *flags* de qualidade por meio da técnica de *bit-packing*.

- Ferramentas de geração de arquivo netCDF

  Foram elaborados dois *scripts* principais para cada instrumento, um para geração de arquivos netCDF a partir de dados bruto e outro para a elaboração de figuras para consulta rápida. Ainda, foi desenvolvido uma biblioteca de funções importantes ao código e que pode ser reutilizada em evoluções das ferramentas ou novas ferramentas incorporadas ao *toolkit*.

2.5.1 Códigos

Para desenvolver as ferramentas necessárias aos objetivos do projeto foram gerados 12 principais produtos em código: 6 *scripts* em Python (2 para cada um dos três instrumentos), 3 bibliotecas de funções auxiliares (*utils*) que acompanham os *scripts* principais de cada equipamento e implementam os detalhes da lógica de conversão e leitura de dados e 3 documentos Python do tipo notebook. A Tabela 2.9 apresenta as características gerais dos códigos.

Tabela 2.9 - Características gerais dos códigos obtidos

Índice	Documento	Extensão	Tipo	Objetivo
1	JOSS_gen_netCDF	.py	Script python puro	Gerar arquivos netCDF a partir de arquivos brutos do equipamento
2	JOSS_gen_figures	.py	Script python puro	Gerar figuras estáticas e interativas a partir dos netCDFs gerados
3	PARS_gen_netCDF	.py	Script python puro	Gerar arquivos netCDF a partir de arquivos brutos do equipamento
4	PARS_gen_figures	.py	Script python puro	Gerar figuras estáticas e interativas a partir dos netCDFs gerados
5	PLUV_gen_netCDF	.py	Script python puro	Gerar arquivos netCDF a partir de arquivos brutos do equipamento
6	PLUV_gen_figures	.py	Script python puro	Gerar figuras estáticas e interativas a partir dos netCDFs gerados
7	utils (JOSS)	biblioteca	Conjunto de arquivos .py	Contém as principais funções utilização na geração de netCDF e figuras
8	utils (PARS)	biblioteca	Conjunto de arquivos .py	Contém as principais funções utilização na geração de netCDF e figuras
9	utils (PLUV)	biblioteca	Conjunto de arquivos .py	Contém as principais funções utilização na geração de netCDF e figuras
10	quality_flags	.ipynb	Notebook	Incluir flags de qualidade de dados no arquivos netCDF
11	JOSS_explore	.ipynb	Notebook	Subsidiar a exploração dos arquivos netCDF gerados para dados do disdrômetro RD-80
11	PARS_explore	.ipynb	Notebook	Subsidiar a exploração dos arquivos netCDF gerados para dados do disdrômetro PARSIVEL^2
12	validacao_estat	.ipynb	Notebook	Subsidiar a validação cruzada do MRR com outros equipamentos

Fonte: Elaborada pelo autor.

2.5.1.1 *Scripts*

Seguindo a metodologia de prototipação e implantação especificadas nos tópicos 2.4.4.1 e 2.4.4.2 foi possível desenvolver três *scripts* Python principais para a conversão do formato de dados brutos de cada instrumento para netCDF e outros três *scripts* secundários que a partir dos netCDFs exportados geram figuras de análise

rápida para cada instrumento.

Os *scripts* de conversão de dados foram elaborados de tal forma que seu funcionamento depende do *input* de arquivos em uma estrutura de pastas conforme apresentado nas figuras 2.12, 2.13, 2.14. Assim, cada equipamento tem um diretório reservado a ele e dentro dele encontra-se os scripts específicos do instrumento bem como a estrutura de pastas de *input* e *output* de dados, onde espera-se como entrada além dos dados brutos alguns arquivos auxiliares como mostrado na tabela 2.10.

Vale ressaltar que os produtos processados pelos *scripts* também são exportados na mesma estrutura de arquivos supracitada.

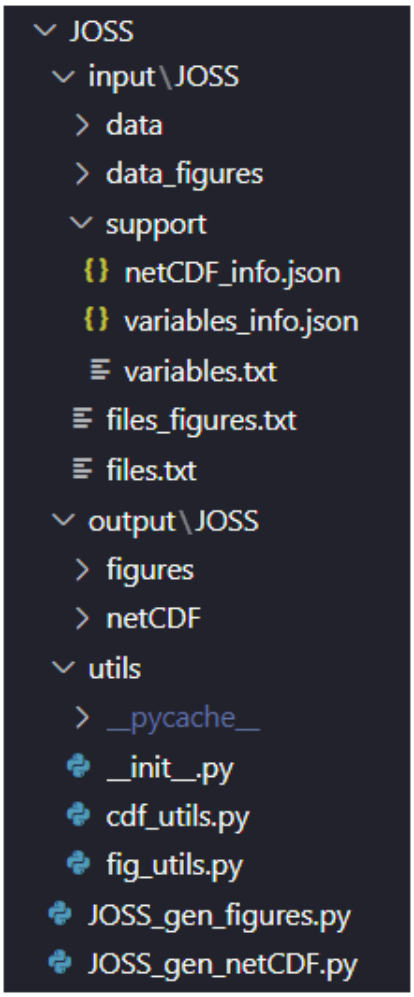
Tabela 2.10 - Arquivos auxiliares esperados para cada *script* de preparação de dados

Arquivos auxiliares em todos os scripts		
Nome	**Localização**	**Descrição**
variables_info.json	\input\support	Contém informações sobre as variáveis presentes no arquivo bruto, bem como outras informações auxiliares sobre ele.
netCDF_info.json	\input\support	Contém informações sobre as variáveis a serem incluídas no netCDF, como nome das variáveis, descrição e outros.

Arquivos auxiliares nos scripts JOSS (referente ao disdrômetro RD-80)		
Nome	**Localização**	**Descrição**
variables.txt	\input\support	Contém o nome das colunas do arquivo bruto do equipamento.

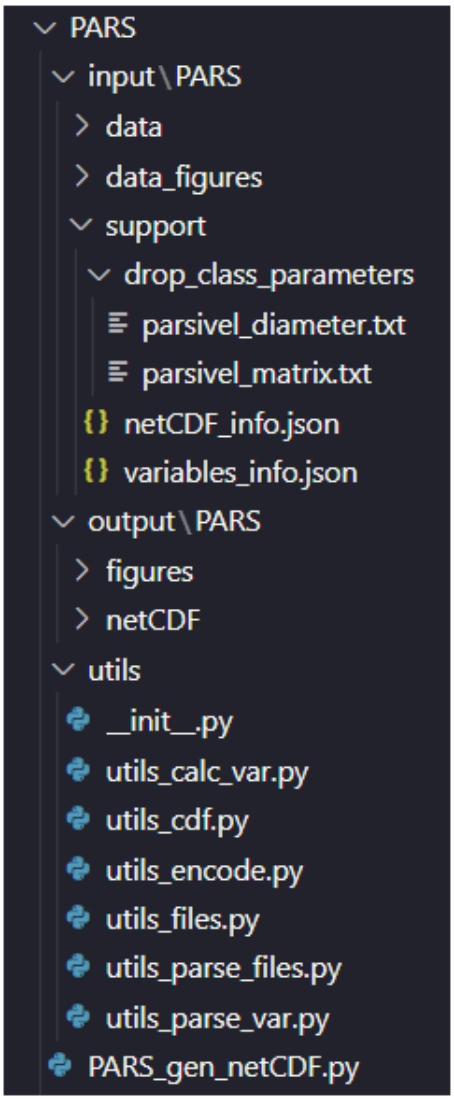
Arquivos auxiliares nos scripts PARS (referente ao disdrômetro PARSIVEL²)		
Nome	**Localização**	**Descrição**
parsivel_diameter.txt	\input\support\drop_class_parameters	Contém a dimensão de cada classe de tamanho de gota
parsivel_matrix.txt	\input\support\drop_class_parameters	Contém a matriz de correção de erro para o equipamento.

Figura 2.12 - Estrutura de arquivos de entrada e saída para o *script* de geração de arquivos netCDF JOSS



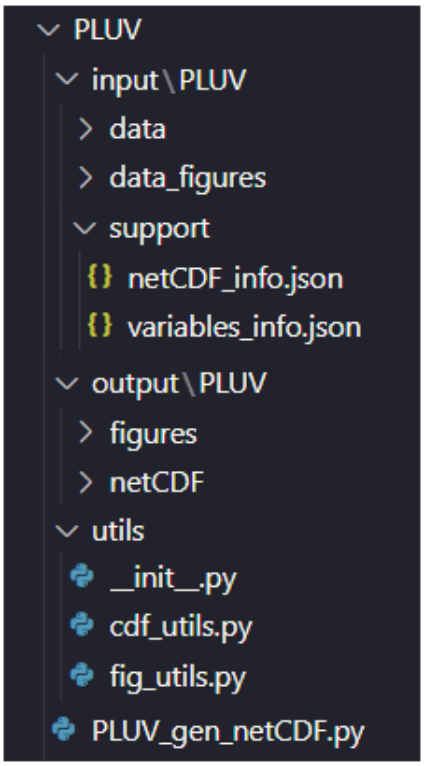
Fonte: Elaborada pelo autor.

Figura 2.13 - Estrutura de arquivos de entrada e saída para o *script* de geração de arquivos netCDF PARS



Fonte: Elaborada pelo autor.

Figura 2.14 - Estrutura de arquivos de entrada e saída para o *script* de geração de arquivos netCDF PLUV



Fonte: Elaborada pelo autor.

A execução do *script* principal é feita pelo terminal (CLI - *Command Line Interface*) utilizando-se Python 3. No comando de execução é necessário inserir as *flags* de execução que permitem especificar quais arquivos serão processados na sessão, além de possibilitar a indicação de inclusão de algumas lógicas adicionais, no caso do PARSIVEL^2. Alguns exemplos de comandos de execução do *script* no terminal são apresentados abaixo.

```
$ .python3 MRR_gen_netCDF.py --help
$ .python3 MRR_gen_netCDF.py --standard
$ .python3 MRR_gen_netCDF.py -l -d "02/08/2022"
```

As *flags* de execução e seus efeitos são listadas a seguir.

- `-h ou --help`

  Imprime no terminal uma mensagem de ajuda explicando as *flags* de execução seus efeitos

- `-s ou --standard`

  Extrai as variáveis de todos os arquivos incluídos na pasta de entrada de dados e encapsula os dados em arquivos netCDF diários.

- `-l ou --list`

  Extrai as variáveis dos arquivos incluídos na pasta de entrada de dados que estejam listado no arquivo files.txt no diretório *support* e encapsula os dados em arquivos netCDF diários.

- `-d "dd/mm/yyyy" ou --date "dd/mm/yyyy"`

  É uma *flag* secundária que tem efeito apenas quando executada em conjunto com outra *flag*.

  Quando executada em conjunto com as flags -s ou -l, não afeta a leitura de arquivos, mas determina para qual data será exportado um arquivo netCDF.

  Quando executada em conjunto com a flag -p, afeta a leitura de arquivos, de forma a determinar quais arquivos serão lidos. Ainda mantém o efeito de determinar para qual data será exportado o netCDF.

- `-p "%Y%d%m" ou --pattern "%Y%d%m"`

  É uma *flag* que exige ser executada junto com a *flag* -d.

  O argumento fornecido na *flag* corresponde ao formato da data (conforme convenção do Python *datetime*[1]) incluída no nome dos arquivos que se encontram na pasta input. Assim, recebe a data de interesse passada como argumento da *flag* -d e para a leitura dos dados busca nos nomes de arquivos (utilizando o formato fornecido), aqueles arquivos correspondentes à própria data de interesse, um dia antes à data de interesse e um dia depois da data de interesse.

  Como a execução ocorre em conjunto com a *flag* -d é exportado arquivo netCDF apenas para a data de interesse.

Uma vez executado, o *script* produz um netCDF para cada dia que possua dados registrados. Nesse contexto vale destacar que o nome do arquivo gerado é elaborado no código seguindo a metodologia especificada no tópico 2.4.2.1.

[1] A linguagem Python possui uma biblioteca padrão para lidar com valores de datas e horários (*datetime values*), esta biblioteca implementa uma convenção de símbolos para extrair *datetime values* a partir de *strings*. A convenção é apresentada de forma resumida por McCutchen (2021).

Ainda, em conformidade com o modelo de dados do formato, as variáveis extraídas do arquivo bruto são incluídas no netCDF cada qual como uma matriz multidimensional, seguindo os nomes e dimensões descritos nas Tabelas 2.4 e 2.5.

De forma a facilitar a manutenção do código e futuras evoluções das ferramentas, o paradigma de programação funcional foi utilizado no desenvolvimento, assim, os processos lógicos de mais baixo nível foram encapsulados em funções. Como forma de organizar a coleção de funções obtidas para o desenvolvimento de cada *script*, elas foram ordenadas em uma biblioteca local denominada *utils*. Esse produto do desenvolvimento do projeto, por si só é estratégico uma vez que potencializa o reaproveitamento de código para futuros trabalhos semelhantes a este.

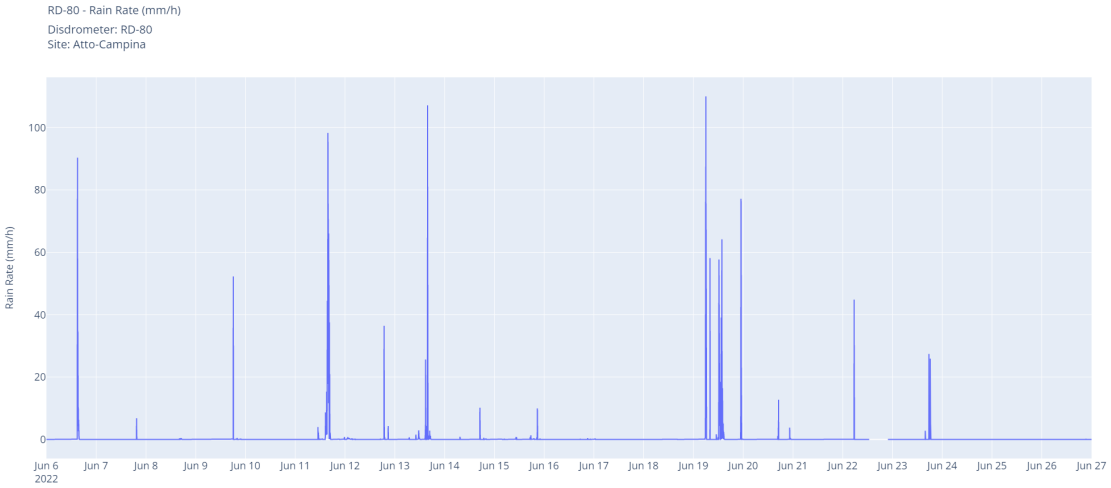
Ademais, os *scripts* que geram as figuras para análise rápida de integridade dos dados também tem seu funcionamento condicionado ao *input* de arquivos na mesma estrutura de pastas especificada nas figuras 2.12, 2.13 e 2.14. Nessas estruturas, é necessário inserir os netCDFs a serem explorados na pasta de entrada de dados para figuras e o produto da execução do *script* é exportado no diretório de saída para figuras.

É importante destacar que para cada figura são produzidos dois formatos, um formato estático (png) e outro formato interativo (html) que permite uma exploração mais minuciosa dos dados e facilita o uso delas em análises.

Nos scripts de geração de figuras dos disdrômetros, para cada netCDF explorado pelos programas são geradas três figuras, conforme apresentado na listagem abaixo. A Figura 2.15 ilustra um gráfico gerado para a variável Rain Intensity (Intesidade da chuva em inglês, a.k.a. Rain Rate) e a Figura 2.16 mostra o conjunto de gráfico gerado para os disdrômetros.

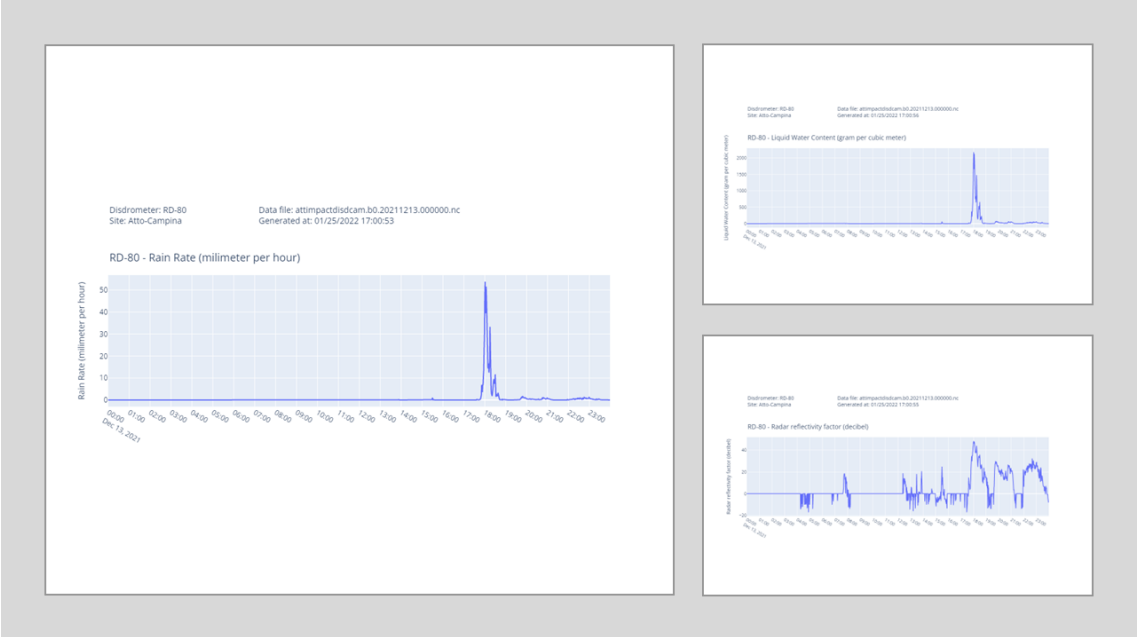
- Gráfico do tipo *line* para a variável RR (*Rain Rate*) x *time*
- Gráfico do tipo *line* para a variável LWC (*Liquid Water Contents*) x *time*
- Gráfico do tipo *line* para a variável Zdb (*Radar Reflectivity Factor*) x *time*

Figura 2.15 - Gráfico elaborado para dados do RD-80 com relação à variável RI (Rain Intensity)



Fonte: Elaborada pelo autor.

Figura 2.16 - Conjunto de gráficos elaborados para dados dos disdrômetros



Fonte: Elaborada pelo autor.

Ademais, no caso do *script* de geração de figuras do disdrômetro RD-80, é gerado também uma figura interativa (html) que permite a análise mais aprofundada dos dados, conforme mostrado na Figura 2.17. Ela apresenta uma interface interativa que mostra os elementos listados abaixo.

- Um gráfico da intensidade de chuva para o intervalo de tempo dos dados lidos
- Um gráfico com a Distribuição de Tamanho de Gotas (DSD) para um determinado instante
- Seletor de instante (*slider)*

Assim é possível selecionar dentro do intervalo de tempo coberto pelo gráfico de intensidade de chuva um instante de tempo de interesse para visualizar a distribuição de tamanho de gotas.

Figura 2.17 - Conjunto de gráficos elaborados para dados dos disdrômetros

Fonte: Elaborada pelo autor.

Seguindo a metodologia especificada em 2.4.4.2, que já foi implementada nos *scripts* de geração de netCDF, a execução das ferramentas de geração de imagens pode ocorrer de três formas distintas, conforme a utilização de *flags* de execução. As *flags* e o funcionamento para cada modo é especificado a seguir.

- `-a ou --auto`
  Executa o *script* (i.e., gera as figuras) para todos os arquivos incluídos na pasta de entrada de dados para figuras, porém acessa os *logs* de geração de arquivo e de geração de figuras e garante que serão geradas figuras apenas para aqueles arquivos que ainda não foram explorados.
- `-f <nome do arquivo netCDF> ou --file <nome do arquivo netCDF>`
  Executa o *script* (i.e., gera as figuras) apenas para o arquivo especificado no comando de execução.
- `-l <nome do arquivo txt> ou --list <nome do arquivo txt>`
  Executa o *script* (i.e., gera as figuras) para os arquivos listados no arquivo files_figures.txt localizado no diretório input.

Por fim, vale comentar que foram utilizadas diversas bibliotecas da linguagem Python para elaborar a lógica dos *scripts* como NumPy e Pandas. O anexo apresenta as bibliotecas necessárias para a execução dos scripts conforme gerado pela ferramenta *pip* de gerenciamento de pacotes Python.

Ainda, é importante mencionar que os códigos continuam em desenvolvimento e manutenção mesmo após a publicação deste relatório, visando a correção de bugs identificados na utilização em situações reais das ferramentas e também contribuições futuras de usuários.

2.5.1.2 Métricas de validação de *script*

A validação das ferramentas propostas neste trabalho de continuação, se deu da mesma forma que fora realizado no projeto inicial. Contudo, considerando que os *scripts* deste projeto possuem a mesma estrutura daqueles produzidos no ano anterior, foi necessário apenas verificar quantitativamente a equivalência dos produtos gerados pelas ferramentas propostas em comparação com as ferramentas em uso. Isso é possível uma vez que a avaliação qualitativa das características de programas em Python *versus* ferramentas vigentes programadas em IDL já foi conduzida no ano anterior e ainda se aplica aos produtos desta continuação.

Ainda, é importante mencionar que apenas as ferramentas de conversão de dados para os disdrômetros realizam cálculos para gerar variáveis de interesse, enquanto o *script* específico do pluviômetro, apenas extrai a variável de taxa de chuva e a encapsula no formato netCDF. Assim apenas os programas dos disdrômetros necessitam de validação quantitativa.

Para explorar essa validação, foi estabelecido um intervalo de tempo para o qual dados brutos dos disdrômetros seriam processados, tanto pelas ferramentas propostas quanto pelas ferramentas vigentes. Para o processamento, foram escolhidos dados da campanha GO Amazon (MARTIN et al., 2016) para o período de 1 de novembro de 2021 até 26 de junho de 2022 e os produtos processados por ambas as ferramentas para os três instrumentos foram comparados entre si.

Foi possível realizar uma comparação direta entre as linhas de dados da variável *Rain Rate* dos arquivos produzidos pelas ferramentas propostas e as ferramentas vigente, nessa comparação foi identificado que a dimensão dos dados lidos a partir de ambos os arquivos era idêntica conforme o esperado (341.280 linhas de dados). Ademais, também foi calculado o *delta* que demonstra o valor absoluto da diferença entre cada linha de dado processado. A Tabela 2.11 apresenta os resultados observados, onde cada posição representa a porcentagem de linhas com *delta* maior que o valor especificado para a coluna. As métricas obtidas demonstram a equivalência satisfatória dos produtos da ferramenta proposta em comparação àquelas em uso.

Tabela 2.11 - Resultados observados para a validação quantitativa das ferramentas de preparação de dados dos disdrômetros

Delta > / **Instrumento**	**10E-2**	**10E-3**	**10E-4**	**10E-5**
RD-80	0,0%	0,0%	0,0%	8,2E-3%
PARSIVEL^2	9,0%	10%	10%	10%

Fonte: Elaborada pelo autor.

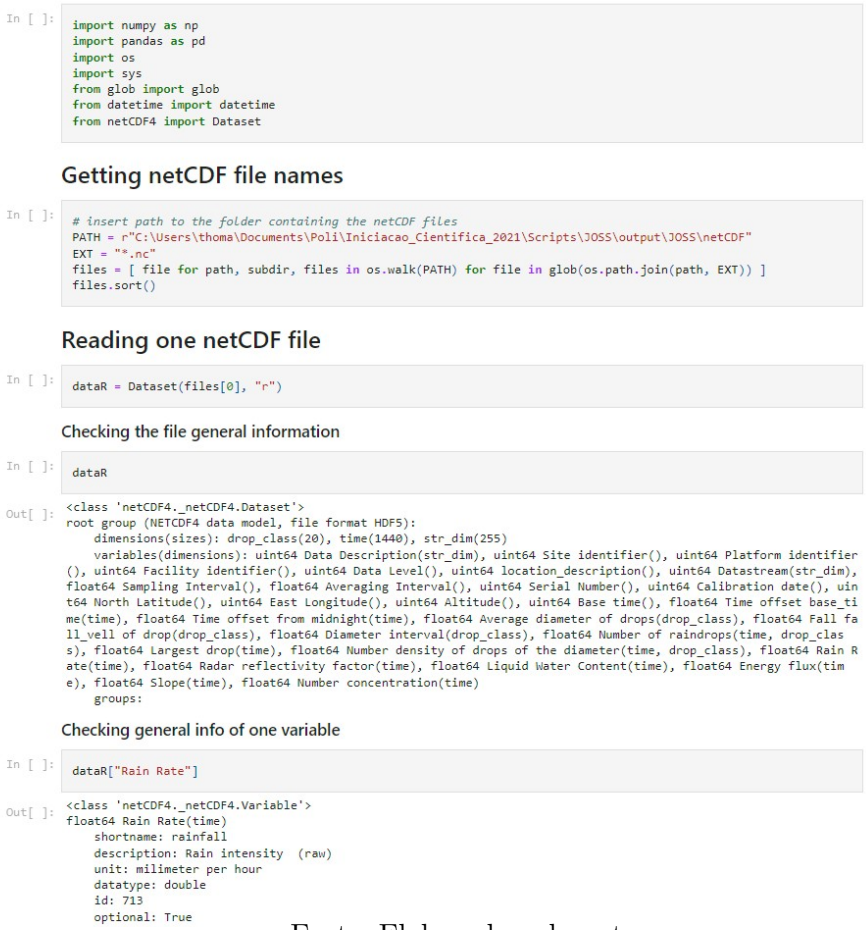
2.5.1.3 Notebook de Exploração de Dados

Seguindo a mesma dinâmica realizada no projeto inicial deste trabalho de continuação, foram elaborados três documentos do tipo notebook (um para cada instrumento objeto de estudos deste projeto) para auxiliarem pesquisadores menos familiarizados com a linguagem Python na exploração dos dados netCDF produzidos pelas ferramentas. Em geral, a distribuição desses documentos visa potencializar o uso das ferramentas pela comunidade do INPE e seus colaboradores.

A plataforma Jupyter foi utilizada no desenvolvimento como descrito na metodologia abordada no tópico 2.4.4.1 e a Figura 2.18 mostra um recorte de um dos documentos

elaborados.

Figura 2.18 - Recorte da interface do notebook de exploração de dados do disdrômetro RD-80



In [ ]:
```
import numpy as np
import pandas as pd
import os
import sys
from glob import glob
from datetime import datetime
from netCDF4 import Dataset
```

Getting netCDF file names

In [ ]:
```
# insert path to the folder containing the netCDF files
PATH = r"C:\Users\thoma\Documents\Poli\Iniciacao_Cientifica_2021\Scripts\JOSS\output\JOSS\netCDF"
EXT = "*.nc"
files = [ file for path, subdir, files in os.walk(PATH) for file in glob(os.path.join(path, EXT)) ]
files.sort()
```

Reading one netCDF file

In [ ]:
```
dataR = Dataset(files[0], "r")
```

Checking the file general information

In [ ]:
```
dataR
```

Out[ ]:
```
<class 'netCDF4._netCDF4.Dataset'>
root group (NETCDF4 data model, file format HDF5):
    dimensions(sizes): drop_class(20), time(1440), str_dim(255)
    variables(dimensions): uint64 Data Description(str_dim), uint64 Site identifier(), uint64 Platform identifier
(), uint64 Facility identifier(), uint64 Data Level(), uint64 location_description(), uint64 Datastream(str_dim),
float64 Sampling Interval(), float64 Averaging Interval(), uint64 Serial Number(), uint64 Calibration date(), uin
t64 North Latitude(), uint64 East Longitude(), uint64 Altitude(), uint64 Base time(), float64 Time offset base_ti
me(time), float64 Time offset from midnight(time), float64 Average diameter of drops(drop_class), float64 Fall fa
ll_vell of drop(drop_class), float64 Diameter interval(drop_class), float64 Number of raindrops(time, drop_clas
s), float64 Largest drop(time), float64 Number density of drops of the diameter(time, drop_class), float64 Rain R
ate(time), float64 Radar reflectivity factor(time), float64 Liquid Water Content(time), float64 Energy flux(tim
e), float64 Slope(time), float64 Number concentration(time)
    groups:
```

Checking general info of one variable

In [ ]:
```
dataR["Rain Rate"]
```

Out[ ]:
```
<class 'netCDF4._netCDF4.Variable'>
float64 Rain Rate(time)
    shortname: rainfall
    description: Rain intensity  (raw)
    unit: milimeter per hour
    datatype: double
    id: 713
    optional: True
```

Fonte: Elaborada pelo autor.

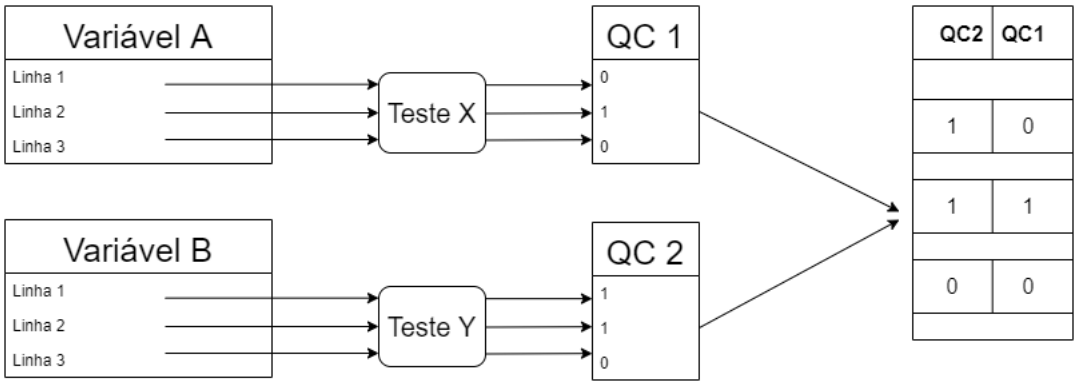
2.5.2 Ferramentas de inserção de *flags* de qualidade de dados

O desenvolvimento das ferramentas de inserção de *flags* de qualidade de dados, se deu por meio de documentos Python do tipo notebook. Essa escolha foi estratégica, uma vez que dado a característica interativa desse tipo de documento garante a flexibilidade necessária ao usuário final, para que possa adaptar e ajustar os testes lógicos que originam as *flags* de controle de qualidade.

O *pipeline* de execução da ferramenta se inicia pela abertura do arquivo netCDF inserido em uma pasta de *input*, seguido pela extração dos dados de interesse para

avaliação de qualidade em uma variável conveniente. Feito isso, a execução segue para a avaliação linha a linha da lógica definida, resultando em um vetor unidimensional com mesmo tamanho da dimensão temporal das variáveis avaliadas. Assim, seguindo a metodologia especificada em 2.4.5 (bit-packing) é possível concatenar cada posição dos vetores de *flags* para formar as palavras binárias a serem efetivamente armazenadas no netCDF, o diagrama 2.19 mostra uma representação da dinâmica supracitada. Por fim, o vetor final é armazenado em uma variável criada no arquivo aberto.

Figura 2.19 - Dinâmica de geração de *flags* de controle de qualidade de dados



Fonte: Elaborada pelo autor.

Para este estudo, de forma a validar as ferramentas propostas, foram aplicados dois testes de qualidade sobre os dados oriundos dos disdrômetros. A Tabela 2.12 apresenta os testes realizados para cada instrumento.

Tabela 2.12 - Testes de controle de qualidade aplicados sobre os dados dos disdrômetros

Variável	**Teste de QC**
Taxa de chuva	> 0.1 mm/h
Número de gotas	> 10

Fonte: Elaborada pelo autor.

2.5.3 Estatísticas de avaliação de dados

Conforme descrito no tópico 2.4.5.2, a etapa de validação cruzada dos dados capturado pelos equipamentos objeto de estudos desse trabalho de continuação é importante para atestar a qualidade das medidas de estimativa registradas pelos disdrômetros, quando comparados com os dados de referência capturados pelo pluviômetro.

Assim, as estatísticas de comparação explicitadas na metodologia foram calculadas utilizando dados da campanha GOAmazon para o intervalo de 15 de abril de 2022 a 30 de abril de 2022. Para o cálculo das estatísticas, foram utilizados dados filtrados por flags de qualidade incluídas com uso das ferramentas conforme descrito em 2.5.2 e foram consideradas 3 duplas de dados de instrumentos, como apresentado na tabela 2.13.

Tabela 2.13 - Duplas de validação cruzada utilizadas no cálculo das estatísticas

Dupla	**Dados de Estimativa**	**Identificador Dados de Estimativa**	**Dados de Referência**	**Identificador Dados de Referência**
1	RD-80	JOSS	Pluviômetro	PLUV
2	PARSIVEL^2	PARS	Pluviômetro	PLUV
3	PARSIVEL^2 com aplicação de matriz de correção	PARS^M	Pluviômetro	PLUV

Fonte: Elaborada pelo autor.

Foram obtidas estatísticas de validação para cada dupla, a tabela 2.14 apresenta os resultados calculados.

A dispersão e correção obtidos mostraram que os dados produzidos pelos algoritmos em Python estão de acordo com o comportamento esperado para medidas de estimativa de taxa de chuva.

O valores da estatística descritiva mostram que os disdrômetros apresentam alta correlação com as medidas de taxa de chuva capturadas pelos pluviômetros (0,80 e 0,86). Ademais, foi observado um erro máximo de estimativa de 16 mm/h, indicando que o instrumento apresentou performance satisfatória.

Tabela 2.14 - Resultados da validação cruzada calculada

Estatística / Dupla	MAE	RMSE	CORR	BIAS
JOSS X PLUV	0,44	4,01	0,80	1,44
PARS X PLUV	17,07	160,72	0,77	18,07
$PARS^{M}$ X PLUV	16,74	156,96	0,86	17,74

Fonte: Elaborada pelo autor.

2.5.4 Distribuição das Ferramentas Desenvolvidas

O desenvolvimento das ferramentas computacionais propostas neste trabalho requisitou uma complexa organização do código e arquivos. Dessarte, se fez necessário a utilização de uma plataforma de versionamento e distribuição de código para organizar os esforços de desenvolvimento e potencializar o compartilhamento dos scripts e notebooks elaborados.

Este trabalho, portanto, optou por utilizar o GitHub como ferramenta de versionamento. Assim, foi criado um repositório público para abrigar os códigos elaborados, e este, será utilizado para a distribuição das ferramentas ao usuários finais. O link para acesso ao repositório é apresentado na listagem abaixo.

```
https://github.com/alancalheiros/INPE-data-processing
```

Ademais, a plataforma provê uma página pública para cada repositório, de forma que o conteúdo exibido, além de listar os arquivos presentes no repositório, inclui textos e instruções que estejam incorporados em um arquivo *markdown*, cujo nome deve ser "README.md", a ser inserido na pasta raiz do repositório. Dessa forma, para o repositório em questão foram incluídas instruções de uso e documentação das ferramentas neste arquivo de forma a facilitar e potencializar sua distribuição.

3 CONCLUSÕES

Acerca da ferramenta de geração de netCDF a partir dos arquivos brutos dos instrumentos disdrômetro RD-80, disdrômetro PARSIVEL^2 e Pluviômetro foi possível atingir plenamente os objetivos estabelecidos. Diante dos resultados obtidos foi possível concluir que as ferramentas propostas desenvolvidas em Python produzem dados com equivalência satisfatória em comparação com os produtos das ferramentas em utilização atualmente.

Além disso, as figuras estáticas e interativas, são fatores de potencialização para o uso da ferramenta em situações reais de uso. O uso das imagens e gráficos é prático e acessível, de forma que seu uso tende a acelerar o processo de análise de dados e identificação de erros pelo pesquisador.

Ademais, os *scripts* em Python apresentam vantagens importantes que facilitam a distribuição das ferramentas por meio de um repositório no GitHub e potencializam seu uso por pesquisadores com menos experiência com a programação. Maiores detalhes podem se encontrados junto ao orientador deste trabalho.

Os algoritmos de avaliação de métricas de qualidade de dados e as ferramentas de inclusão de *flags* de controle de qualidade foram desenvolvidas e testadas com dados de medidas dos disdrômetros. De modo preliminar, foi possível observar que a inclusão das *flags* permitiu selecionar com mais praticidade apenas dados mais confiáveis para utilização em análises. As *flags* sinalizam quais linhas de dados foram afetadas por problemas sistemáticos, como falta de medidas e erros decorrentes de problemas associados ao equipamento.

Em relação às estatísticas de validação cruzada que suportam análises para identificar medidas não-realísticas e outros fatores de segunda ordem que impactam os dados, foi possível elaborar com sucesso um documento do tipo notebook que permite calcular as estatísticas com flexibilidade, permitindo também que antes dos cálculos seja aplicado um filtro aos dados baseado nas *flags* de qualidade geradas pelos outros scripts desse trabalho.

Cabe ressaltar que as ferramentas aqui desenvolvidas para controle de dados dos três instrumentos, se inserem em um amplo grupo de pesquisa que prevê a implementação de ferramentas análogas para outros equipamentos medidos em experimentos na região amazônica, sob a tutela de pesquisadores responsáveis no INPE. Assim, em última instância, as convenções, metodologias e práticas utilizadas neste trabalho

servirão de alicerce para o desenvolver mais programas de gestão de qualidade de dados atmosféricos em conformidade com as diretrizes internacionais para dados, atuando como fator multiplicador para o desenvolvimento de ferramentas computacionais de tratamento de qualidade de dados.

As atividades aqui realizadas foram complexas e exigiram alto empenho e comprometimentos, além de conhecimentos avançados de computação e dos dados aqui analisados.

3.1 Trabalhos Futuros

Nesta seção são explicitadas futuras etapas que tratam de desdobramentos desse projeto de pesquisa que ainda precisam ser testados e implementados seguindo a trajetória especificada pelos órgãos internacionais, como o ARM-DoE.

No contexto do grupo de pesquisa que estabelece a parceira entre Escola Politécnica, LFA-USP e INPE, há espaço para extender as ferramentas para outros sensores de medidas atmosféricas, de forma a dar continuidade de desenvolvimento de ferramentas análogas a essas elaboradas.

Ademais, perpassa por esse processo também, a integração das ferramentas atuais e futuras com uma plataforma de distribuição e gerenciamento de dados atmosféricos que foi proposta pelo grupo de pesquisa para a comunidade científica e está em desenvolvimento atualmente por uma equipe multidisciplinar.

REFERÊNCIAS BIBLIOGRÁFICAS

ANJOS G. A. DIAS, A. A. R. R. L. dos. Dados científicos: As práticas de gestão dos pesquisadores brasileiros na ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIêNCIA DA INFORMAçãO, 18., 2017, Marília, SP, Brasil. Marília, SP, Brasil: ENANCIB, 2017. 5

ATMOSPHERIC RADIATION MEASUREMENT CLIMATE RESEARCH FACILITY. **ARM Data File Standards**. [S.l.], ago. 2016. 1–14 p. 17

ATMOSPHERICRADIATION MEASUREMENT CLIMATE RESEARCH FACILITY. **Management Structure**. 2017. Disponível em: <https://www.arm.gov/about/management-structure>. Acesso em: 19 mai. 2020. 1, 2

CALHEIROS, A. J. P. Relatório sobre os dados do projeto sos-chuva (pluviômetros, disdrômetros, mrr e mp3000a). 2018. Disponível em: <http://chuvaproject.cptec.inpe.br/portal/pdf/relatorios/Rel_dados_sos_chuva.pdf>. 25

DISTROMET LTD. **User Guide for DISDRODATA 2.0**: Data acquisition on personal computer for disdrometer rd-80. Switzerland: DISTROMET LTD, 2009. 9, 13

FONDRIEST. **HyQuest Solutions TB4 Series II Tipping Bucket Rain Gauge**. 2022. Disponível em: <https://www.fondriest.com/hyquest-solutions-tb4-series-ii-tipping-bucket-rain-gauge.htm>. Acesso em: 30 ago. 2022. 14, 15

FRIEDRICH, K.; HIGGINS, S.; MASTERS, F. J.; LOPEZ, C. R. Articulating and stationary parsivel disdrometer measurements in conditions with strong winds and heavy rainfall. **Journal of Atmospheric and Oceanic Technology**, American Meteorological Society, Boston MA, USA, v. 30, n. 9, p. 2063–2080, 01 Sep. 2013. Disponível em: <https://journals.ametsoc.org/view/journals/atot/30/9/jtech-d-12-00254_1.xml>. 10

HYQUEST SOLUTIONS PTY LTD. **INSTRUCTION MANUAL TIPPING BUCKET RAIN GAUGE**: Model tb4/series ii. Australia: HYQUEST SOLUTIONS PTY LTD, 2019. 15

ISLAM, T.; RICO-RAMIREZ, M. A.; HAN, D.; SRIVASTAVA, P. K. A joss–waldvogel disdrometer derived rainfall estimation study by collocated tipping bucket and rapid response rain gauges. **Atmospheric Science Letters**, v. 13, n. 2, p. 139–150, 2012. Disponível em: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.376>. 6, 7

JOEL, J. Insights into parsivel measurements. École Polytechnique Fédérale de LAusanne, 05 2011. 6

JOSS J.AND WALDVOGEL, A. Ein spektrograph für niederschlagstropfen mit automatischer auswertung. **pure and applied geophysics**, v. 68, n. 1, p. 240–246, Dec 1967. ISSN 1420-9136. Disponível em: <https://doi.org/10.1007/BF00874898>. 6

KINNELL, P. I. A. Some observations on the joss-waldvogel rainfall disdrometer. **Journal of Applied Meteorology 1976-may vol. 15 iss. 5**, v. 15, may 1976. 7

KWON OHBYUNG; LEE, N. S. B. Data quality management, data usage experience and acquisition intention of big data analytics. **International Journal of Information Management**, Elsevier Science, v. 34, 06 2014. 6

MACHADO, L. A. T. Previsão imediata de tempestades intensas e entendimento dos processos físicos no interior das nuvens: O sos- chuva (sistema de observação e previsão de tempo severo). 2015. Disponível em: <http://chuvaproject.cptec.inpe.br/portal/pdf/relatorios/Rel_dados_sos_chuva.pdf>. 3

MARTIN, S. T.; ARTAXO, P.; MACHADO, L. A. T.; MANZI, A. O.; SOUZA, R. A. F.; SCHUMACHER, C.; WANG, J.; ANDREAE, M. O.; BARBOSA, H. M. J.; FAN, J.; FISCH, G.; GOLDSTEIN, A. H.; GUENTHER, A.; JIMENEZ, J. L.; PÖSCHL, U.; DIAS, M. A. S.; SMITH, J. N.; WENDISCH, M. Introduction: Observations and modeling of the green ocean amazon (goamazon2014/5). **Atmospheric Chemistry and Physics**, v. 16, n. 8, p. 4785–4797, 2016. Disponível em: <https://acp.copernicus.org/articles/16/4785/2016/>. 3, 38

MCCUTCHEN, W. **Python strftime cheatsheet**. 2021. Disponível em: <https://strftime.org/>. 32

OTT HYDROMET GMBH. **Operating instructions**: Present weather sensor ott parsivel. Germany: OTT HydroMet GmbH, 2016. 9, 10, 12

RD-80 Disdrometer. Jul 2018. Disponível em: <https://ictinternational.com/products/rd80/rd-80-disdrometer/>. 7

REW, R.; DAVIS, G. Netcdf: an interface for scientific data access. **IEEE Computer Graphics and Applications**, v. 10, n. 4, p. 76–82, 1990. 17

SILVA, F. C. C. da. **Gestão de Dados Científicos**. Rio de Janeiro: Editora Interciência Ltda, 2020. 3 p. ISBN 978-65-990252-2-8. 5

TOKAY, A.; WOLFF, D. B.; PETERSEN, W. A. Evaluation of the new version of the laser-optical disdrometer, ott parsivel2. **Journal of Atmospheric and Oceanic Technology**, American Meteorological Society, Boston MA, USA, v. 31, n. 6, p. 1276–1288, 01 Jun. 2014. Disponível em: <https://journals.ametsoc.org/view/journals/atot/31/6/jtech-d-13-00174_1.xml>. 10, 11

WANG, J.; FISHER, B. L.; WOLFF, D. B. Estimating rain rates from tipping-bucket rain gauge measurements. **Journal of Atmospheric and Oceanic Technology**, American Meteorological Society, Boston MA, USA, v. 25, n. 1, p. 43–56, 01 Jan. 2008. Disponível em: <https://journals.ametsoc.org/view/journals/atot/25/1/2007jtecha895_1.xml>. 13

WILKS, D. **Statistical methods in the atmospheric sciences**. Amsterdam Boston: Elsevier/Academic Press, 2011. ISBN 978-0-12-385022-5. 25, 26

WORLD METEREOLOGICAL ORGANIZATION. **Forecast Verification methods Across Time and Space Scales**. 2017. Disponível em: <https://cawcr.gov.au/projects/verification/>. Acesso em: 29 mai. 2021. 26

ANEXO A - BIBLIOTECAS PYTHON NECESSÁRIAS PARA A EXECUÇÃO DAS FERRAMENTAS PROPOSTAS

```
alembic==1.1.0.dev0
appdirs==1.4.4
astroid==2.4.2
attrs==19.3.0
Automat==0.8.0
Babel==2.6.0
backcall==0.1.0
bcrypt==3.1.7
black==20.8b1
bleach==3.1.1
blinker==1.4
certifi==2019.11.28
chardet==3.0.4
click==7.1.2
cloud-init==21.2
colorama==0.4.3
command-not-found==0.3
configobj==5.0.6
constantly==15.1.0
cryptography==2.8
cycler==0.10.0
dbus-python==1.2.16
decorator==4.4.2
defusedxml==0.6.0
distlib==0.3.1
distro==1.4.0
distro-info===0.23ubuntu1
entrypoints==0.3
filelock==3.0.12
Flask==1.1.1
Flask-BabelEx==0.9.3
Flask-Compress==1.4.0
Flask-Gravatar==0.4.2
Flask-Login==0.4.1
Flask-Mail==0.9.1
```

```
Flask-Migrate==2.5.2
Flask-Paranoid==0.2.0
Flask-Principal==0.4.0
Flask-Security-Too==3.4.2
Flask-SQLAlchemy==2.1
Flask-WTF==0.14.2
html5lib==1.0.1
httplib2==0.14.0
hyperlink==19.0.0
idna==2.8
importlib-metadata==1.5.0
incremental==16.10.1
ipykernel==5.2.0
ipython==7.13.0
ipython-genutils==0.2.0
ipywidgets==6.0.0
isort==5.7.0
itsdangerous==1.1.0
jedi==0.15.2
Jinja2==2.10.1
jsonpatch==1.22
jsonpointer==2.0
jsonschema==3.2.0
jupyter-client==6.1.2
jupyter-console==6.2.0
jupyter-core==4.6.3
jupyterthemes==0.20.0
keyring==18.0.1
kiwisolver==1.3.1
language-selector==0.1
launchpadlib==1.10.13
lazr.restfulclient==0.14.2
lazr.uri==1.0.3
lazy-object-proxy==1.4.3
ldap3==2.4.1
lesscpy==0.14.0
Mako==1.1.0
```

```
MarkupSafe==1.1.0
matplotlib==3.4.1
mccabe==0.6.1
mistune==0.8.4
more-itertools==4.2.0
mypy-extensions==0.4.3
nbconvert==5.6.1
nbformat==5.0.4
netifaces==0.10.4
notebook==6.0.3
numpy==1.19.5
oauthlib==3.1.0
pandas==1.2.0
pandocfilters==1.4.2
paramiko==2.6.0
parso==0.5.2
passlib==1.7.2
pathspec==0.8.1
pexpect==4.6.0
pickleshare==0.7.5
Pillow==8.2.0
ply==3.11
prometheus-client==0.7.1
prompt-toolkit==2.0.10
psutil==5.5.1
psycopg2==2.8.4
pyasn1==0.4.2
pyasn1-modules==0.2.1
Pygments==2.3.1
PyGObject==3.36.0
PyHamcrest==1.9.0
pyinotify==0.9.6
PyJWT==1.7.1
pylint==2.6.0
pymacaroons==0.13.0
PyNaCl==1.3.0
pyOpenSSL==19.0.0
```

```
pyparsing==2.4.7
pyrsistent==0.15.5
pyserial==3.4
python-apt==2.0.0+ubuntu0.20.4.5
python-dateutil==2.7.3
python-debian===0.1.36ubuntu1
pytz==2020.5
PyYAML==5.3.1
pyzmq==18.1.1
QtPy==1.9.0
regex==2021.4.4
requests==2.22.0
requests-unixsocket==0.2.0
SecretStorage==2.3.1
Send2Trash==1.5.0
service-identity==18.1.0
simplejson==3.16.0
six==1.14.0
sos==4.1
speaklater==1.4
SQLAlchemy==1.3.12
sqlparse==0.2.4
ssh-import-id==5.10
sshtunnel==0.1.4
systemd-python==234
terminado==0.8.2
testpath==0.4.4
toml==0.10.2
tornado==5.1.1
traitlets==4.3.3
Twisted==18.9.0
typed-ast==1.4.2
typing-extensions==3.7.4.3
ubuntu-advantage-tools==27.0
ufw==0.36
unattended-upgrades==0.1
urllib3==1.25.8
```

```
virtualenv==20.2.2
wadllib==1.3.3
wcwidth==0.1.8
webencodings==0.5.1
Werkzeug==0.16.1
widgetsnbextension==2.0.0
wrapt==1.12.1
WTForms==2.2.1
zipp==1.0.0
zope.interface==4.7.1
```

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.