MULTI-SENSORS DATA QUALITY TOOLS FOR PRECIPITATION ON THE AMAZON REGION

POUGY, T.A ; [1]CALHEIROS, A. J. P. ; PRAKASH, G. ; CORREA, P.L.P.

[1] National Institute for Space Research (INPE)

Email: {thomazpougy,pedro.correa}@usp.br, alan.calheiros@inpe.br, palanisamyg@ornl.gov

ABSTRACT

The Brazilian National Institute for Space Research (INPE) produces research that helps to understand climate and weather dynamics in Brazil and in the world, with significant impacts on national public and private strategic planning. Among the essential information for the studies are the rainfall data. In this context, ensuring the quality of this data has a direct impact on the reliability of the forecasts and analysis generated from them. Thus, this study, which is a partnership between INPE, the Laboratory of Atmospheric Physics and Polytechnic School of USP and the ARM-DoE (Atmospheric Radiation Measurement Climate Research Facility), aimed to establish computational tools that could deal with the quality of data from rain in accordance with the main international directives. Thus, it was proposed for this study the development of a specific toolkit for data from the Micro Rain Radar (MRR), disdrometers PARSIVEL2 and RD80, and rain gauge that would help researchers from INPE, USP and partners to: standardize the preparation of raw data for internationally accepted formats; processing figures to support quick analyses; analyze and process data quality and, finally, record metadata and quality analysis for publication in international data repositories.

DATA AND METHODS

Different data sources to measure precipitation were used in this study. Among them is a widely applied sensor that is the rain gauge. Two other instrument that measure the distribution of raindrops were also used, the so-called disdrometers. In this case measurements were made with Joss-Waldvogel (RD-80) [1] and Particle Size and Velocity (PARSIVEL2) [2]. In addition to these instruments, a vertical pointing radar was also used, the Micro Rain Radar (MRR). First, a data quality management plan must be carried out. The method proposed by [3,4] approaches data management planning in general, however it will be applied to the specific context of this study for data quality management. Some steps were taken to better define how these data and toolkits will be presented to the scientific community: Documentation; Format; Dimensions and variables; Metadata; Data Preperation; Automation and visualization.

RESULTS

The product processed by the tools must comply with international standards and norms, with a view to facilitating their publication in international repositories. In this context, the reference standard used was the ARM-DoE data management practices and structure.

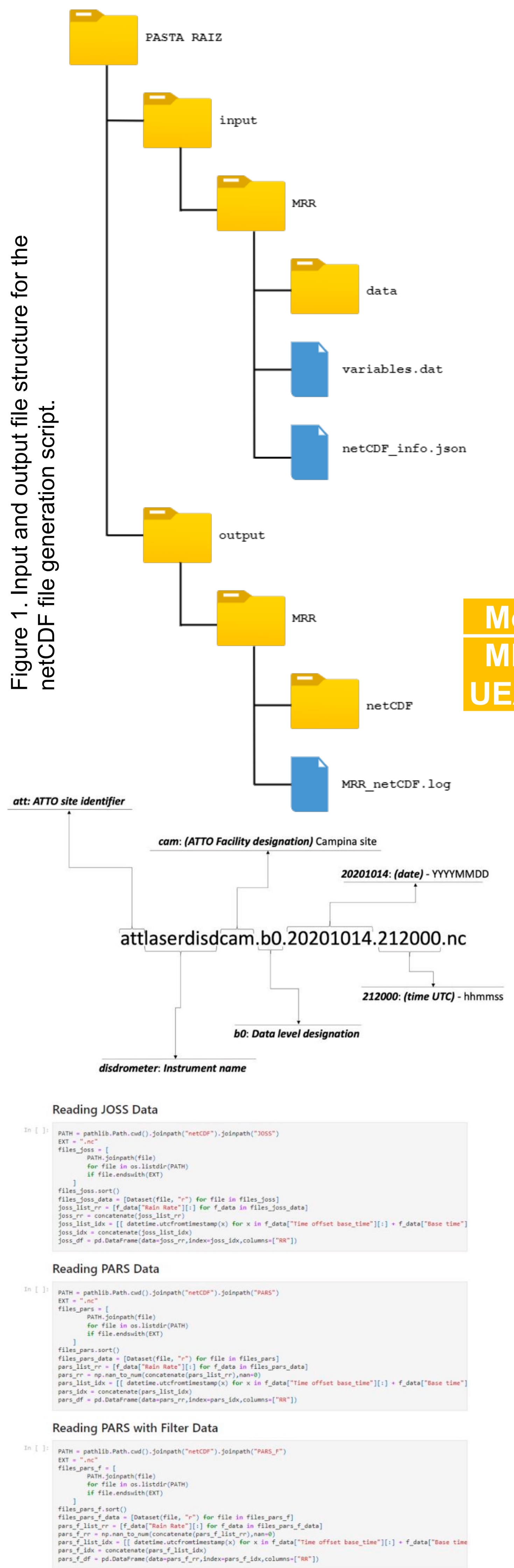


Figure 1. Input and output file structure for the netCDF file generation script.

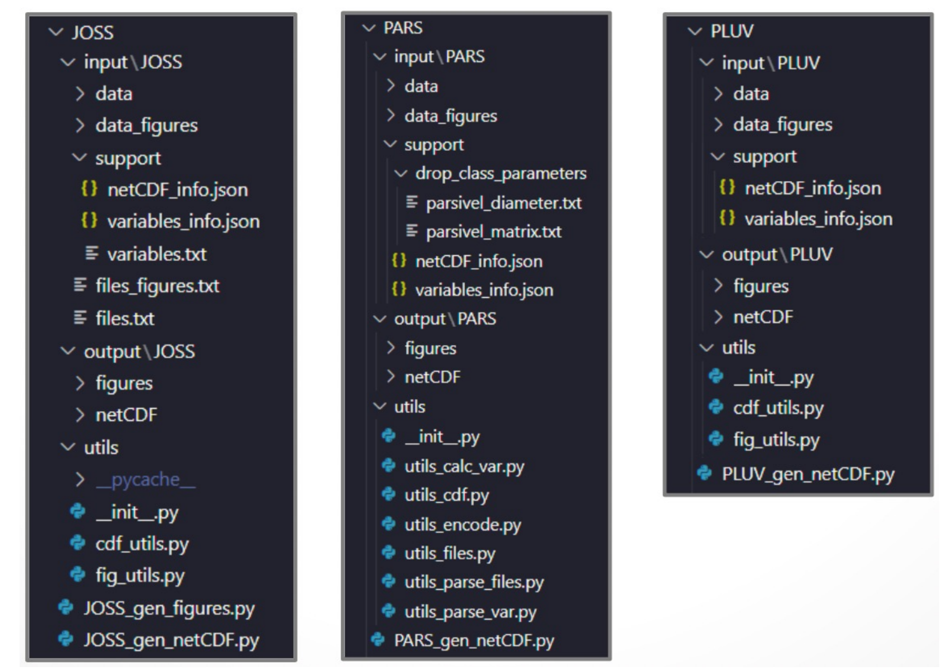
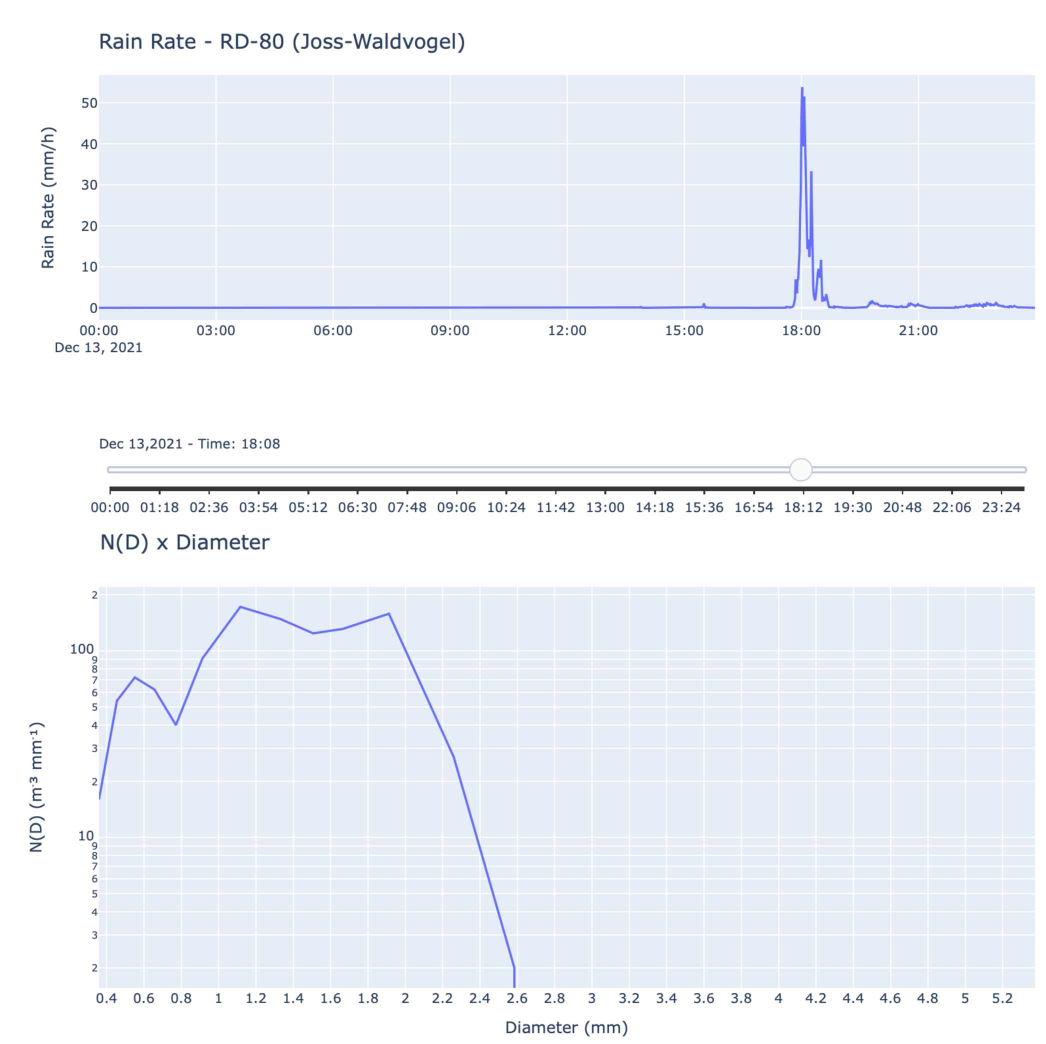


Table 1. Valid MRR data for the GoAmazon experiments.

Month	Jan	Feb	Mar	Aug	Sep	Oct
MRR (%): UEA site	100	99.9	82.69	16.79	100	36.17

Figure 2. RD-80 disdrometer visualization in HTML format.



Following the prototyping and deployment methodology, it was possible to develop a main Python script for converting the MRR/Disdrometer/Rain Gauge data format to netCDF and another secondary script that generates quick analysis figures from the exported netCDFs files. The data conversion script was designed in such a way that its operation depends on the input of files in a folder structure shown in the diagram in Figure 1. In this structure, two complementary files and the raw file (e.g. MRR) are expected as input, whose variables must be extracted and encapsulated in netCDF format. Supplementary files specify the metadata to be included in the file as well as the list of variables to be extracted or defined based on the raw data.

In the data exploration scenario, the analysis of flags associated with missing data was easily implemented, since daily data are created that consider the temporal resolution of the measurements. In Table 1 it is possible to verify how much valid data existed during the GoAmazon [5] Campaign for the MRR measurement period in 2014. For each netCDF created by the algorithms, quick analysis figures are generated. It is important to highlight that for each figure two formats are produced, a static format (png), ideal for controlling data in a visual way, and another interactive format (html) that allows a more detailed exploration of the data and facilitates analysis by researchers (Figure 2, for RD-80 disdrometer).

CONCLUSION AND FUTURE WORK

Based on the results obtained, it was possible to conclude that the proposed tools developed in Python produce data with full equivalence compared to the products of the tools currently in use. Furthermore, the Python script has important advantages over the technology in use, since it is open-source and has a large community. In addition, it offers the possibility of elaborating Python notebooks documents that are interactive and provide a guided execution of the code. This dynamic facilitates the documentation of the tools and leverage their use by researchers with little experience with the Python language.

The quality control flag definition algorithms were built and tested with measurement data from some experiments analyzed here. The analysis showed that they are suitable for use in larger volume data regarding systematic problems, such as lack of measurements and errors resulting from problems associated with the equipment. The results obtained from the intercomparison between the different data sources showed that the equipment is working properly. Furthermore, such metrics could be used to define data quality problems and possible solutions could be implemented in future work on higher-level data.

References:

[1] JOSS, J.; WALDVOGEL, A. Ein spektrograph für niederschlagstropfen mit automatischer auswertung. pure and applied geophysics, v. 68, n. 1, p. 240–246, 12 1967. Disponível em: <https://doi.org/10.1007/BF00874898>. 6

[2] LÖFFLER-MANG, M.; JOSS, J. An optical disdrometer for measuring size and velocity of hydrometeors. Journal of Atmospheric and Oceanic Technology, v. 17, n. 2, p. 130-139, 2000.

[3] Hunziker, D. et al., "Identifying, attributing, and overcoming common data quality issues of manned station observations," Int. J. Clima, vol. 37, no. 11, pp. 4131–4145, Sep. 2017.

[4] BERTIN, P. R. B., VISOLI M. C., DRUCKER D. P., "A Gestão de Dados de Pesquisa no Contexto da E-science: Benefícios, Desafios e Oportunidades para Organizações de P&D", PontodeAcesso, Salvador, v.11, n.2, p. 34-48, ago. 2017.

[5] MARTIN, S. T., et al. Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5), Atmos. Chem. Phys., 16, 4785–4797, https://doi.org/10.5194/acp-16-4785-2016, 2016.

Acknowledgements:

The authors wish to thank the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* - Brasil (CAPES) for their financial support.

Author affiliations:

Pougy, T. (USP,BR), Calheiros, A. (INPE, BR), Correa, P.(USP, BR), Prakash, G. (ORNL, US)