Developing Data Services to Provide Data Quality Information for Global Satellite-based Precipitation Products

Z. Liu[1,2], C.-L. Shie[1,3], and D. Meyer[1]

[1]GES DISC

[2]CSISS, George Mason University

[3]Univ. of Maryland at Baltimore County

The 9th Workshop of International Precipitation Working Group

Outline

- Introduction
- Data quality issues
- Solutions and activities
- Summary

Introduction

- Global satellite-based precipitation products have been widely used in research and applications around the world. Compared to ground-based observations, satellite-based measurements provide data on a global scale, especially in remote continents and over oceans.
- The NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) is home to NASA global precipitation product archives including the Tropical Rainfall Measuring Mission (TRMM), the Global Precipitation Measurement (GPM), as well as other global and regional precipitation products.
- Precipitation is one of the top downloaded and accessed parameters in the GES DISC data archive. Meanwhile, users want to easily locate and obtain data quality information at regional and global scales to better understand how precipitation products perform and how reliable they are.

Introduction (cont.)

- As a data service provider, it is necessary to provide easy access to data quality information.
- However, such information normally is not available, and when it is available, it is not in one place and difficult to locate.
- In this presentation, we will present such challenges and activities at the GES DISC to address precipitation data (other datasets as well) quality issues.

Data Quality Issues

Data quality issues are very complex and associated with many things (e.g. observations, algorithms).

In this presentation, our focus is on issues associated with post data production, particularly in these areas:

- Data services
- Value-added products
- User contribution

Data Quality Issues (cont.)

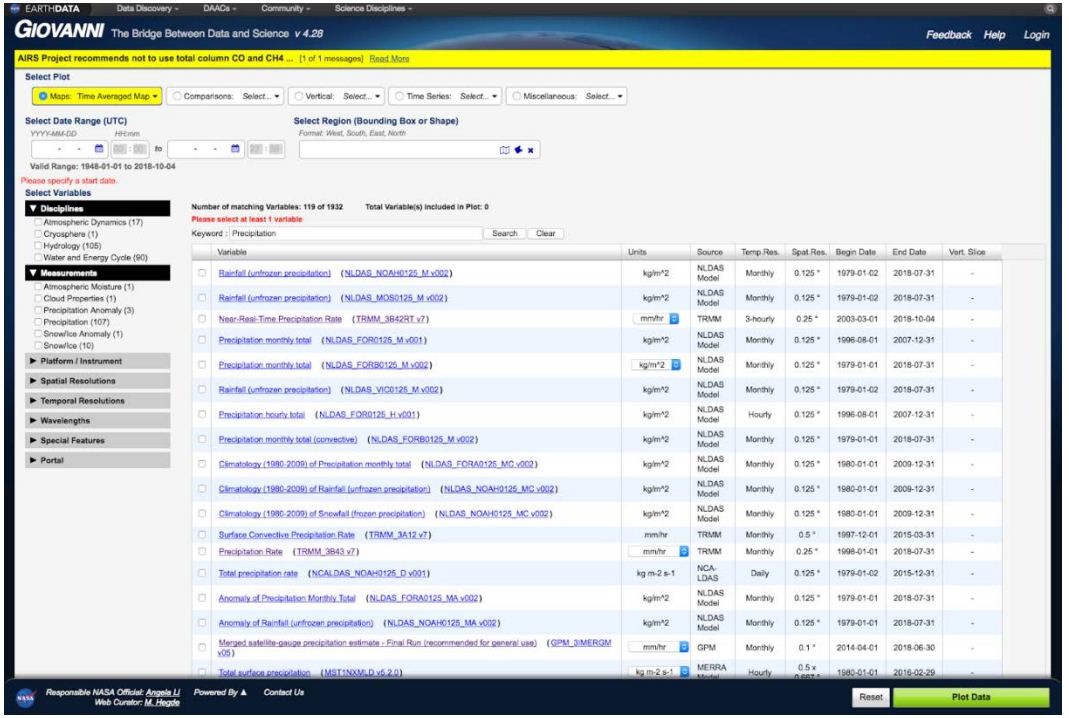
Data Services:

- More user-friendly data services available (e.g. Giovanni, GDS)
- However, users are not clear how the processing at the backend is done
- Errors and known issues associated with processing are not well documented and information often is not available
- Difficult to find and not in one place

Data Quality Issues (cont.)

- A large collection of data services available at NASA data centers such as subsetting, quality screening, re-gridding, reformatting, reprojection, mosaicking, aggregating variables, etc.

- The Geospatial Interactive Online Visualization and Analysis Infrastructure (Giovanni, right), was developed by GES DISC and provides easy access to over ~1900 variables.



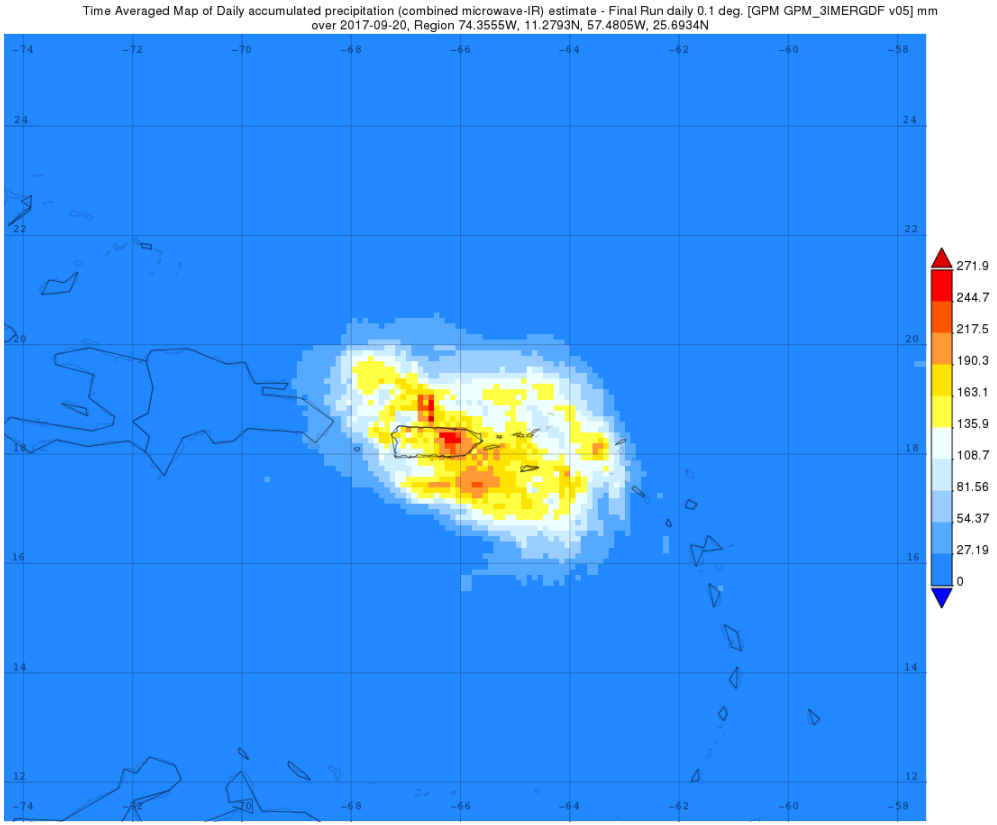
Data Quality Issues (cont.)

Value-added Products:

- More value-added products are available for users with different needs (TMPA daily, IMERG daily, etc.)
- On-the-fly product generation
- Error estimates and known issues are challenging issues

Data Quality Issues (cont.)

More value-added products will be available to meet needs from different users around the world. For example, the TRMM Multi-satellite Precipitation Analysis daily product can be derived from its 3-hourly product. But what is random error for the daily product and many other on-the-fly products?



Question: How can we provide data quality information for value-added products, especially on-the-fly products?

Data Quality Issues (cont.)

User Contribution:

- It is difficult to assess and validate satellite-based precipitation products on a global scale
- Many precipitation users available around the world
- Some of them have their own gauge or radar data
- Information from citizen scientists and crowd sourcing
- Ground validation research papers (see example) and reports in different locations are published each year
- Information is hidden in journal articles or reports
- Needs an information system to capture and harvest the information and make it available to all in one place
- Another challenge: to manage and implement this information as well as improve precipitation product development
- Nonetheless, user contribution can benefit all the stakeholders such as algorithm developers (usually also data producers), data distributors (e.g. data centers like GES DISC), and other users.

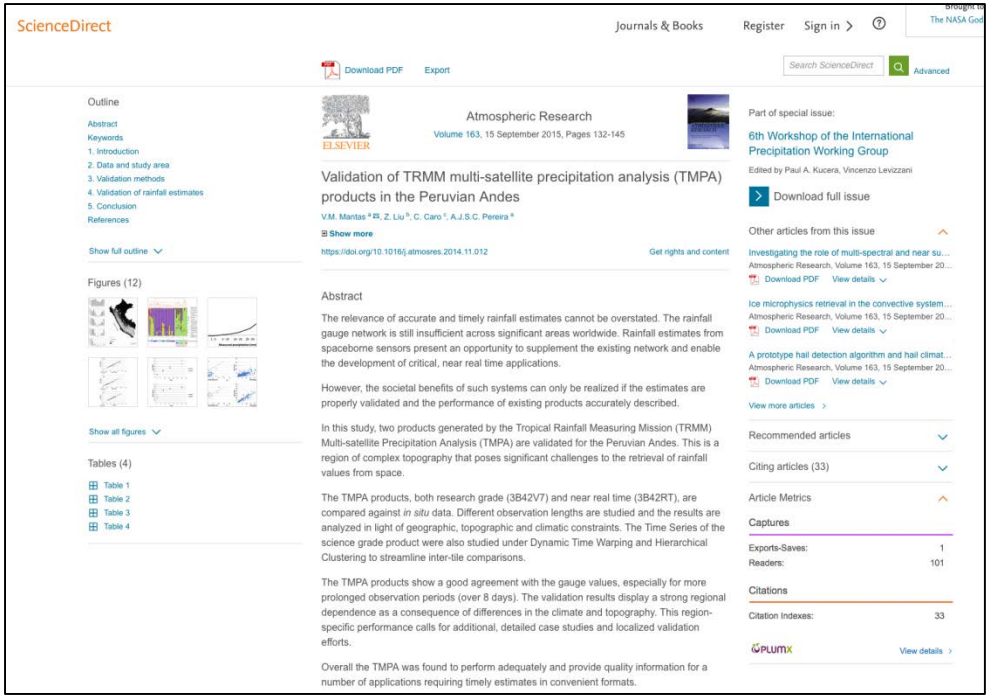
Data Quality Issues (cont.)

• Mantas, V.M., Z. Liu, C. Caro, and A.J.S.C. Pereira. (2015) Validation of TRMM multi-satellite precipitation analysis (TMPA) products in the Peruvian Andes. Atmospheric Research 163, 132-145. (right)

• Prakash, S., A.K. Mitra, A. AghaKouchak, Z. Liu, H. Norouzi, D.S. PaiA preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region, J. Hydrol., 556 (2018), pp. 865-876

• Chen, F., and X. Li, 2016: Evaluation of IMERG and TRMM 3B43 Monthly Precipitation Products over Mainland China. *Remote Sens.*, **8**, 472, doi:10.3390/rs8060472.

• Siuki, S. K., B. Saghafian, and S. Moazami, 2016: Comprehensive evaluation of 3-hourly TRMM and half-hourly GPM-IMERG satellite precipitation products. *Int'l. J. Rem. Sens.*, **38**, 558-571, doi:10.1080/01431161.2016.1268735.



Question: How can we use these results to improve algorithms and applications?

Solutions and Activities

Solutions:

- Data quality standard development (especially for multi-disciplinary research and applications)
- Data services (document algorithms, archive, data processing, operational anomaly, etc.)
- Value-added product development (research on error estimates, document algorithms, known issues, etc.)
- User contribution (system development to collect and disseminate results of their activities in a standard way)

Solutions and Activities (cont.)

Activities:

- Collection of common practices in different disciplines
- Develop plans with stakeholders (algorithm developers and users)
- Continue to establish working groups (WGs) (e.g. the "ESDSWG-Data Quality")
- Develop standards (challenging)

Summary

- Satellite-based precipitation products are widely used in research and applications
- Data quality is a challenging area, especially on a global scale
- Data quality issues from data services at data centers have not been addressed adequately
- Research to better understand error estimates in on-the-fly data products
- Services and tools are needed to capture, document, and deliver data quality information
- User contribution is important for improving global products

Information

- Data information and services: https://disc.gsfc.nasa.gov/
- Giovanni: https://giovanni.gsfc.nasa.gov
- Comments and suggestions: Zhong.Liu@**nasa.gov**