GOES-R Ground Segment Performance

The Reproducibility of Research Baseline Results in Implemented Algorithms

Authors: C. Botos[1], E. J. Kennelly[1], H. E. Snell[1], E. Steinfelt[1], T. S. Zaccheo, and R. Khanna[2]

[1]Atmospheric and Environmental Research, Inc., Lexington, MA, United States
[2]GOES-R Ground Segment Project, NASA Goddard Space Flight Center

AMS2013

Overview

- The National Research Council Committee on NASA-NOAA Transition from Research to Operations recommended improved transitional processes for bridging technology from research to operations.
- Research baseline remote sensing algorithms developed by government science teams are implemented in high-performance computing environments to enable near-real time observations and data use.
- Measuring the operational algorithm performance relative to the research baseline is a critical step supporting the GOES-R Mission Objectives.
- The GOES-R program has introduced the Reproducibility Requirement for Ground System (GS) algorithms with the objective of ensuring that product quality is maintained while allowing cost effective implementation optimizations (e.g., for product latency).

Reproducibility Requirement

- GSFPS-2758: The GS shall generate End-Products based on the Government-provided Algorithm Packages such that the comparison of Ground Segment test data outputs and Algorithm Work Group (AWG) test data outputs yields reproducibility based on squared correlation coefficient (r-squared) between these two of at least 0.9995 with no more than 1% of the compared values having error greater than 0.15% from the AWG-provided value for the given data point.

Reproducibility Measurement for Continuous Data

- Continuous data are numerical and can conceptually take on an infinite number of values with no particular label attributed to particular values. Continuous data may be represented using large numbers of discrete values.
- Consider data product where baseline results x and reproduced results y are continuous. In these cases, a squared correlation coefficient r^2 is used to measure how closely the reproduced results correspond to the baseline.

$$r^2 = \frac{\left(\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})\right)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2\sum_{i=1}^{N}(y_i-\bar{y})^2}$$

- The r-squared value describes the correlation of the data but does not prohibit very large individual errors. The outlier requirement sets an upper bound of the number of outliers in the data that exceed a specified threshold D_{thresh}.

$$n = \sum_{i=1}^{N}\begin{cases}1 \text{ if } |x_i - y_i| > D_{thresh} \\ 0 \text{ otherwise}\end{cases}$$

- Together the $r^2 \geq 0.9995$ and $n \leq 1\%$ mandates a high-fidelity implementation of government algorithms that retains product quality.

Reproducibility Measurement for Discrete Categorical Data

- Discrete or categorical data have only finite number of values. Integers, symbols, or labels are associated with each value.
- Consider products containing data elements drawn from a finite set of K discrete elements (categories).

$$x_i \in \{k_1, k_2, k_3, \ldots k_K\},\ y_i \in \{k_1, k_2, k_3, \ldots k_K\}$$

- The classification accuracy is defined as the percentage of instances where x = y relative to the total number of the data points.
- The overall performance of the algorithm can be more completely evaluated in the form of the classification matrix (or confusion table).
  - Consider a matrix W where the rows and columns correspond to the elements of K. Each element of W represents the number of elements labeled with the j^{th} category in Y to those labeled with the i^{th} category in x.
  - The classification accuracy is determined from the diagonal elements as:

$$T_{x=y} = \sum_{i=1\ldots K} w_{i,i}$$

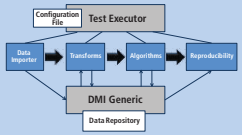
- The interpretation of the Reproducibility requirement for categorical data was addressed by a government-lead Tiger Team as follows:

The GS shall generate End-Products based on the Government-provided Algorithm Packages such that the comparison of GS test data outputs and AWG test data outputs yields reproducibility based on a classification accuracy of at least 99%.

GOES-R Algorithm Test Tools

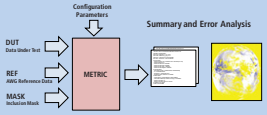
- Algorithm Test Tools (ATT) is a framework for testing AER-developed algorithms and evaluating their end products.
- Driven by command-line and configuration file.
- Has four run modes that may be chained or run individually.
  - Data Importer – Translates AWG-provided data into a common format for use by the tools.
  - Transforms – Alters algorithm input data with the general intention of simulating interference, noise and other data flaws.
  - Algorithm – Instantiates, configures, and runs an algorithm to produce end products for integration and reproducibility testing.
  - Reproducibility – Compares an algorithm's end product to AWG reference data to determine if the product satisfies the Reproducibility Requirement

Overview of Algorithm Test Tools



Reproducibility Input/ Output

- Each Reproducibility Metric:
  1. Ingests DUT and REF data, and the Inclusion MASK
  2. Ingests Configuration Parameters
  3. Produces reports and results specialized for that metric type



Summary of Current Results

- A set of algorithm test tools (ATT) were developed to support testing during the Software Integration and Test (SWIT) development phases.
- ATT facilitates comparison of results from GS-implemented algorithms (DUT=Data Under Test) with government provided reference data sets (REF) providing metrics for reproducibility: R-squared, Outliers, and Classification Accuracy.
- The capabilities of the tools were developed and tested using government provided test data and reference algorithm software addressing both image- and list-based (e.g., Lightning Detection) products and corresponding data quality flags.
- The reproducibility approach has been highly effective for resolving issues with algorithms and test data (e.g., providing insight concerning numerical precision) resulting in a more fully vetted and validated product.
- The reproducibility of Imagery, Lightning, Aerosol, Cloud, Sounding, and Aviation products has been demonstrated during Software Integration and Test.
- Development of Hydrology, Wind, Land and Ocean products is underway.

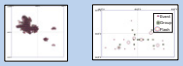
Reproducibility Achieved

- **Cloud and Moisture Imagery**
  - 48 simulated Full Disk, CONUS, an Mesoscale scenes for each of 16 bands

  PASSED

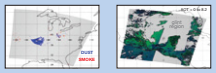
- **Lightning Detection (Events, Groups, Flashes)**
  - 120 simulated and LIS-proxy scenes

  PASSED



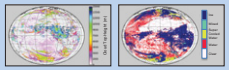
- **Aerosol Product: Detection, Optical Depth**
  - 156 MODIS proxy scenes

  PASSED



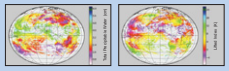
- **Cloud Products: Mask, Phase/Type, Height/ Temperature/ Pressure)**
  - 36 SEVIRI proxy scenes

  PASSED



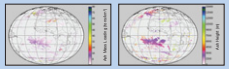
- **Sounding Products: Temperature/Moisture Profiles, TPW, Stability Indices)**
  - 36 SEVIRI proxy scenes

  PASSED



- **Aviation Products: Volcanic Ash**
  - 39 SEVIRI proxy scenes

  PASSED



aer

HARRIS assuredcommunications