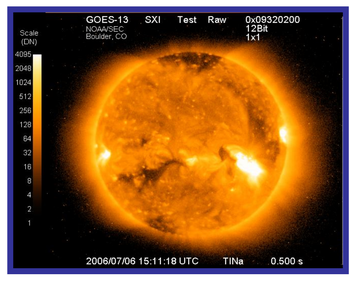
GOES-R Ground Segment Performance
The Reproducibility of Research Baseline Results in Implemented Algorithms

Raj Khanna (rkhanna@noblis.org), Chris Botos, Erik Steinfelt, and Ted Kennelly
GOES-R Ground Segment Project NASA Goddard Space Flight Center & Atmospheric and Environmental Inc

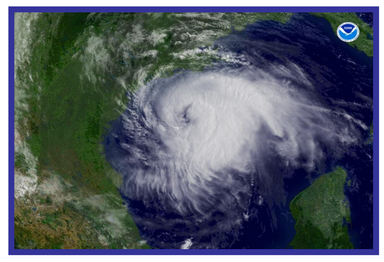
GLM data will augment today's ground based lightning detection systems, capturing synoptic data on in-cloud, cloud-cloud, and cloud-ground lightning

GOES-R Mission Objectives

- ✓ To maintain GOES mission continuity and quality in environmental observations in the 2015-2028 timeframe
- ✓ To provide enhanced environmental data products
- ✓ To improve services and data being provided to Users
- ✓ To be responsive to technology infusion to meet evolving User needs



GOES-R will broadcast five enhanced Space Weather products from the SUVI, EXIS, SEISS, and Magnetometer via GRB



Increased temporal and spatial resolution provided by GOES-R direct broadcast data will help enable more precise forecasting of storms such as Hurricane Ike

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 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 support GOES-R Mission Objectives

Applicable 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

Definitions

- 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.
- Discrete or categorical data have only finite number of values. Integers, symbols, or labels are associated with each value.

Reproducibility Measurement for Continuous Data

Consider data products where baseline results X and reproduced results Y are continuous. In these cases, a correlation coefficient (r) is used to measure how closely the reproduced results correspond to the baseline.

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

Reproducibility Measurement for Discrete Categorical Data*

Consider products containing data elements drawn from a finite set of K discrete elements (categories).

$$x_{n,m} \in \{k_1, k_2, ..., k_i, ..., k_K\}, y_{n,m} \in \{k_1, k_2, ..., k_i, ..., k_K\}$$

A count of each instance that $x_{n,m}=y_{n,m}$ provides a basic measure of how many elements in the verification data match the baseline data.

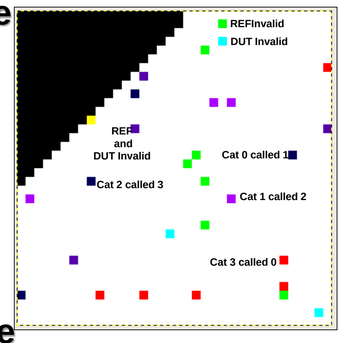
Dividing by the total number of elements, N·M, provides the percentage match. This is an overall measure of the reproduction of the data.

A more detailed view of the differences can be obtained by using a *classification matrix*.

An example of two dimensional samples are shown in these equations.

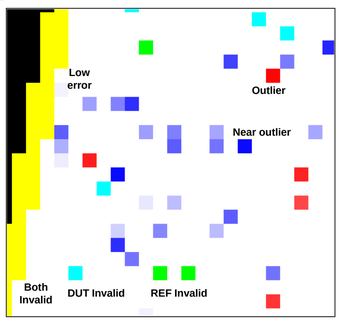
Categorical Error Image

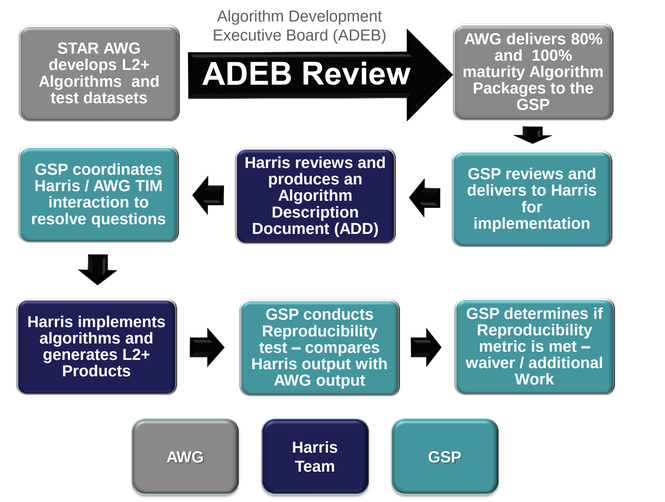
Color	Meaning
White	Non-Error
Red	Classification Error, REF ≠ DUT R,G,B = k(REF, 0, DUT)
Purple	
Blue	
Black	Mask = False LZA > 65 degrees
Yellow	REF and DUT are Invalid
Green	REF is Invalid
Cyan	DUT is Invalid

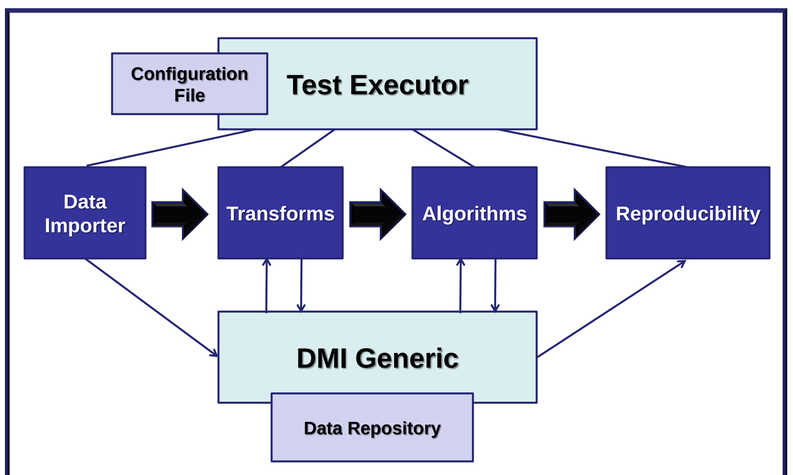


Continuous Error Image

Color	Meaning
White	Nearly zero error
Blue	Non-Outlier Darker as error approaches outlier threshold
Red	Outlier Darker as error increases above outlier threshold
Black	Inclusion Mask is False (Example: LZA > 82 degrees)
Yellow	Both REF and DUT are Invalid
Green	REF is Invalid
Cyan	DUT is Invalid







Algorithm Test Tools

A Classification Matrix Shows the Numerical Distribution of Matching Samples by Category

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 i^{th} category in X.

Baseline \ Verification	k_1	k_2	•••	k_j	•••	k_K
k_1	$w_{1,1}$	$w_{1,2}$	•••	$w_{1,j}$	•••	$w_{1,K}$
k_2	$w_{2,1}$					
⋮	⋮					
k_i	$w_{i,1}$					
⋮	⋮					
k_K	$w_{K,1}$					$w_{K,K}$

Let each element of W be defined:

$$w_{i,j} = count(x_{n,m} = k_i \, and \, y_{n,m} = k_j) \quad \text{for all n=1…N, m=1…M, i=1…K, and j=1…K}$$

Percentages Provide the Relative Frequency in Each Category

Dividing the elements of W by N · M results in a matrix representing the percentage in each category:

$$p_{i,j} = \frac{w_{i,j}}{NM} \quad \text{for all i=1…K and j=1…K}$$

The percentage matrix P is shown on the right.

$p_{1,1}$	$p_{1,2}$	•••	$p_{1,j}$	•••	$p_{1,K}$
$p_{2,1}$					
⋮					
$p_{i,1}$					
⋮					
$p_{K,1}$					$p_{K,K}$

Summing Matrix Values Provides Measurements of Corresponding Matches and Differences

The main diagonal, when i=j, $w_{i,i}=w_{j,j}$ represents the number in each category where the verification data matches the baseline, $x_{n,m}=y_{n,m}$

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

The off-diagonal elements , when i≠j, represent the differences in each category:

$$T_{x \neq y} = \sum_{\substack{i=1...K \\ j=1...K \\ Where i \neq j}} w_{i,j}$$

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

Summary of Current Results

- ATT was developed to support testing during the Software Integration and Test (SWIT) development phase
- 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
- During the SWIT pathfinder activity, ATT was used to support defect correction and to demonstrate the methodology for verification of the reproducibility requirement based on a partial implementation of the GOES-R cloud mask algorithm.

More information on GOES-R may be found at: http://www.goes-r.gov