

Ambiguity of Quality in Remote Sensing Data

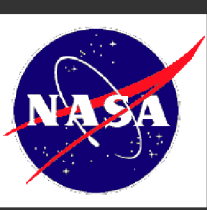


Christopher Lynnes, NASA/GSFC

Greg Leptoukh, NASA/GSFC

Goddard Earth Sciences Data and Information Services Center

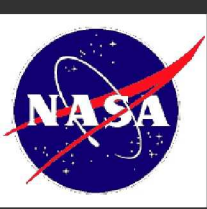
Funded by: NASA's Advancing Collaborative Connections for Earth System Science (ACCESS) and Advanced Information Systems Technology (AIST) programs



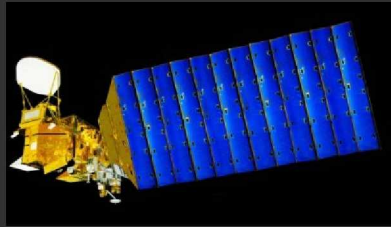
Data Quality Goal



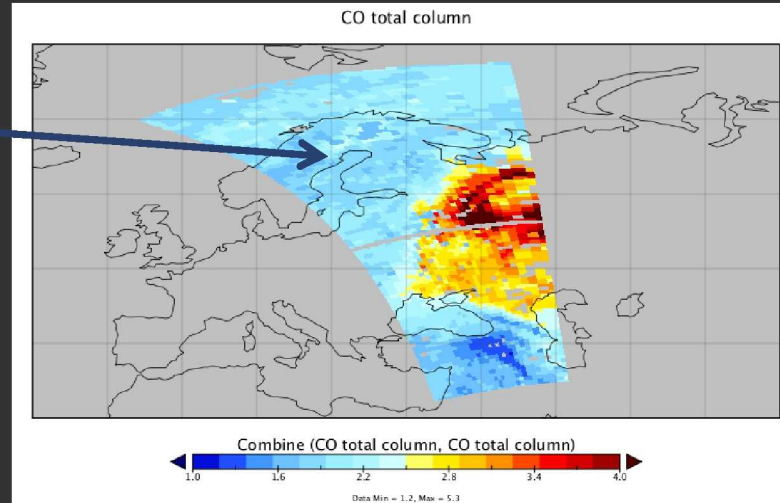
- ❧ Describe data quality so that users understand:
 - ❧ What it means for the data
 - ❧ How to use quality information for data selection and analysis
- ❧ Why so difficult?
 - ❧ Data Quality depends on intended use
 - ❧ Data Quality = Fitness for Purpose



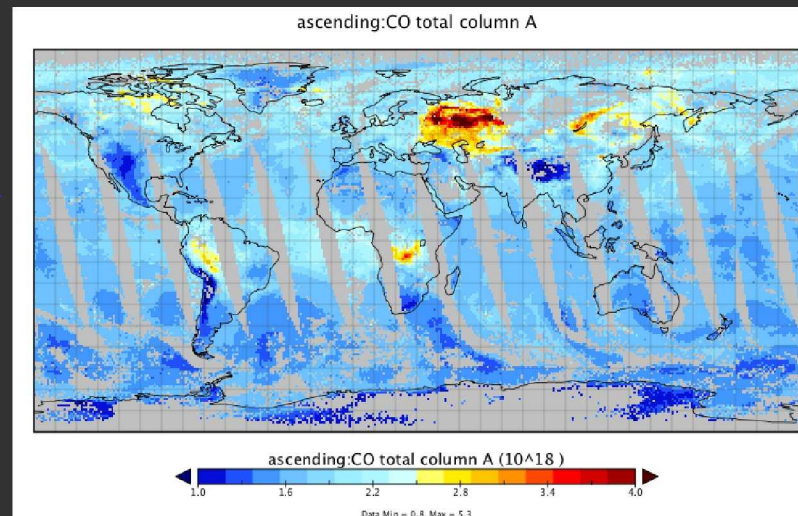
Satellite data swaths and grids



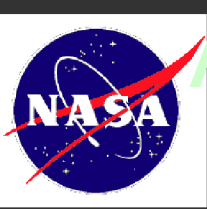
Temporal
and Spatial
Aggregation



Level 2
Swath



Level 3
Grid



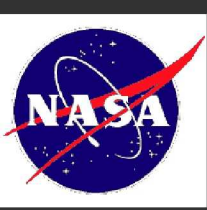
A Survey of Data Quality at Different Levels of Aggregation

Increasing Aggregation →

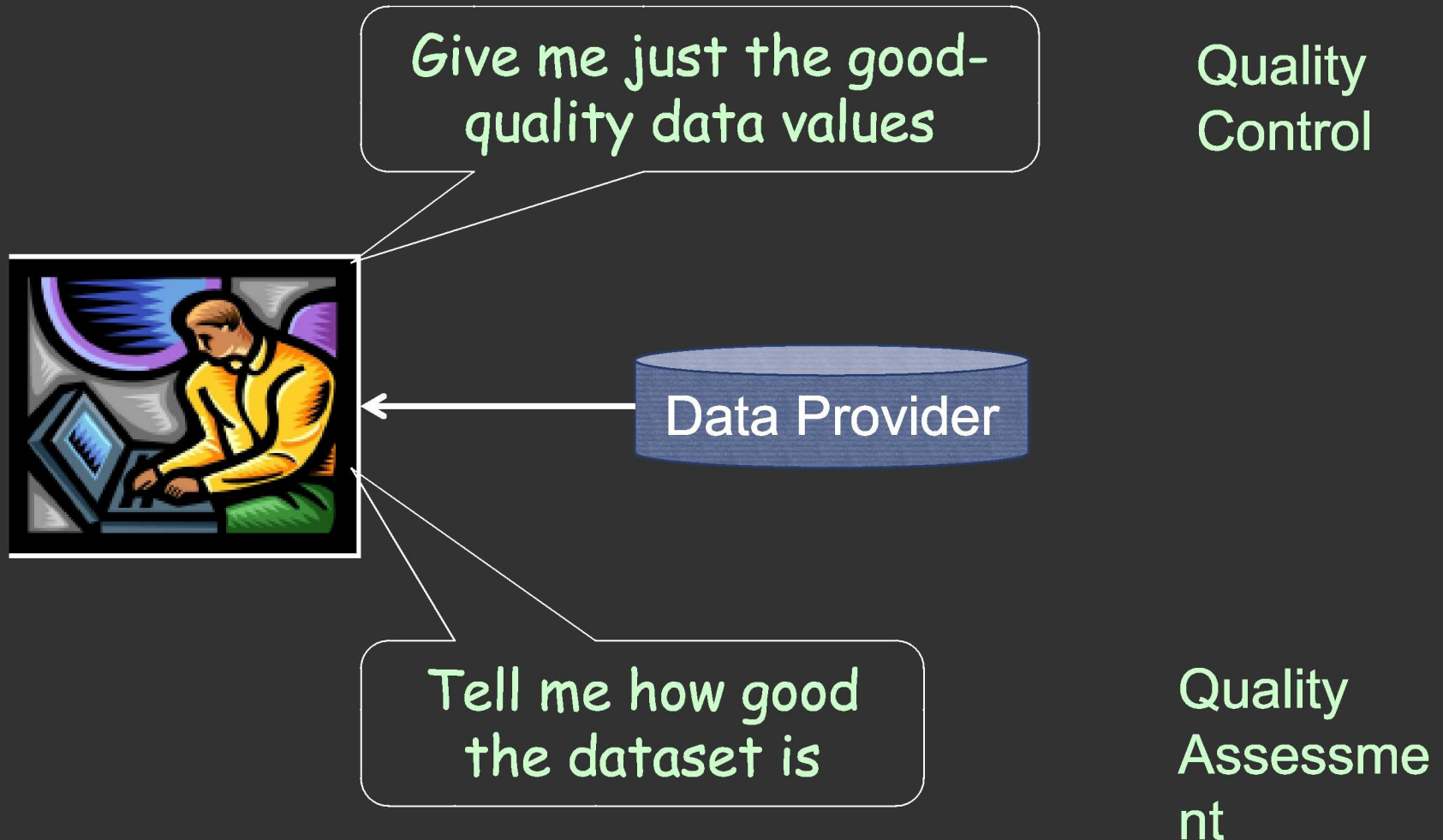
Increasing Aggregation ↓

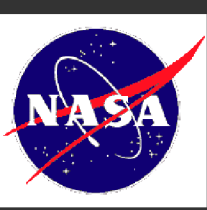
Type of Data	Data-Value Quality	Data-File Quality	Dataset* Quality
Level 2 Swath	Quality Control (QC) flags	Statistical summary of QC info	Statistical accuracy relative to validation sites
Level 3 Grid	•Cell std. dev. •Number of input values	Statistical summary of QC info	??

*Dataset = the assemblage of all the data values of the same type

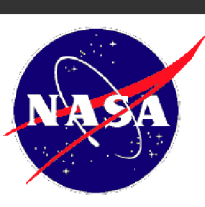


Contrasting dataset-level vs. data-value-level Use Cases





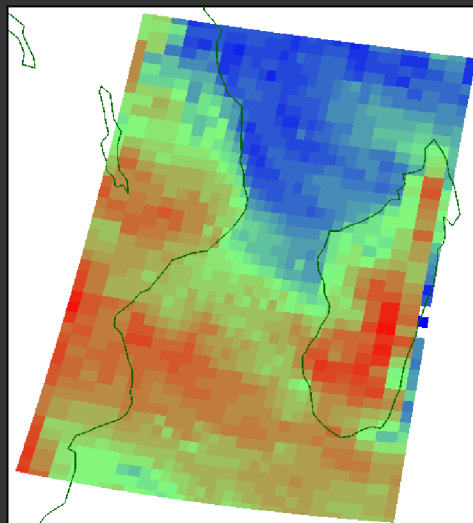
Data-Value Quality Control



The Data Quality Screening Service: a straightforward example of data-value quality control



Original data array

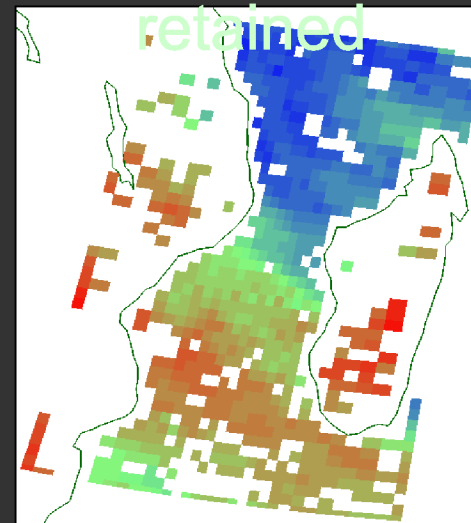


AIRS* Total column precipitable water

Mask based on quality flags

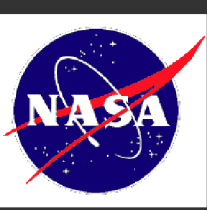


Good quality data pixels retained



Output file has the same format and structure as the input file, with fill values replacing the low-quality data

* Atmospheric Infrared Sounder



Or, maybe not so straightforward...



AIRS* Quality Indicators

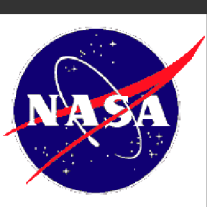
0 Best _____ ? _____
1 Good _____ ?
2 Do Not Use _____

MODIS Atmosphere** Confidence Flags

3 Very Good
2 Good
1 Marginal
0 Bad

* Atmospheric Infrared Sounder

** Moderate Resolution Imaging Spectroradiometer



Match up by recommendation?



AIRS Quality Indicators

0	Best	<i>Data Assimilation</i>
1	Good	<i>Climatic Studies</i>
2	Do Not Use	

MODIS Aerosols Confidence Flags

Ocean

Land

3	Very Good	3	Very Good
2	Good	2	Good
1	Marginal	1	Marginal
0	Bad	0	Bad

*Use these flags in order to
stay within expected error*

Ocean bounds Land

$\pm 0.03 \pm 0.10 \tau$

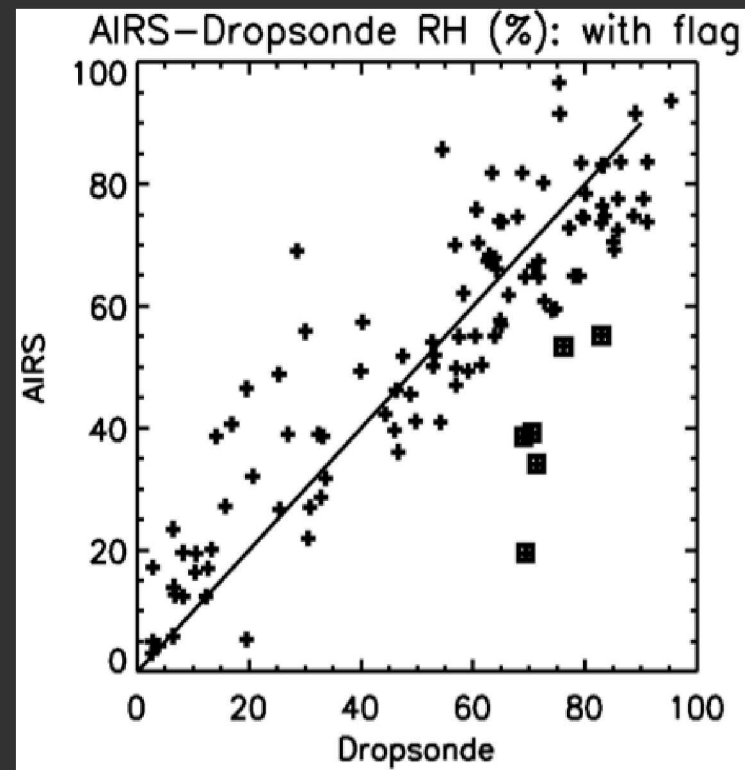
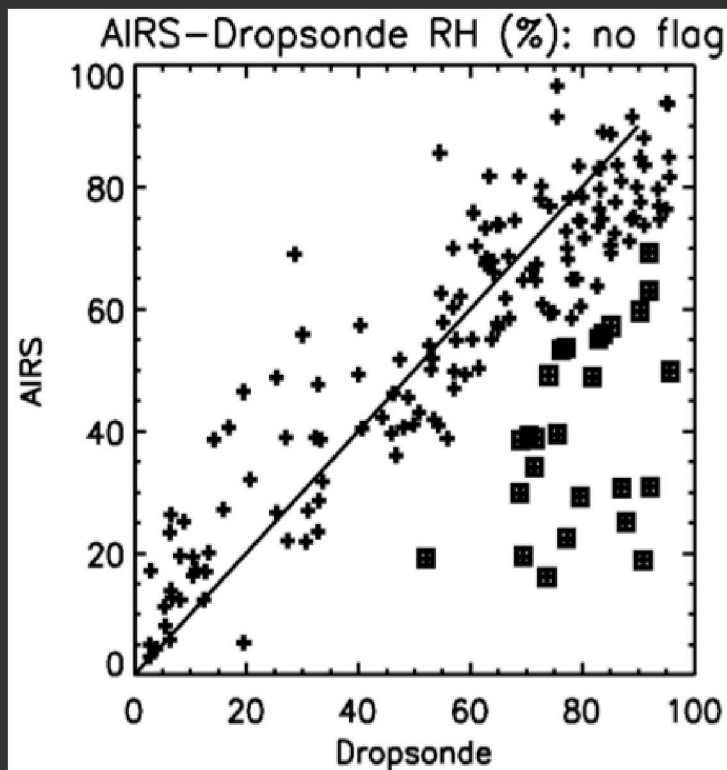
$\pm 0.05 \pm 0.15 \tau$



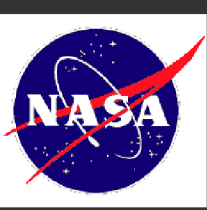
How do quality control indicators relate to dataset quality assessment?



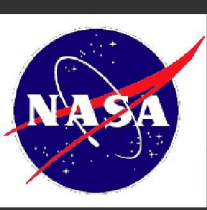
AIRS Relative Humidity Comparison against Dropsonde
with and without Applying PBest Quality Flag Filtering
Boxed data points indicate AIRS RH data with dry bias > 20%



*From a study by Sun Wong (JPL) on specific humidity in the
Atlantic Main Development Region for Tropical Storms*



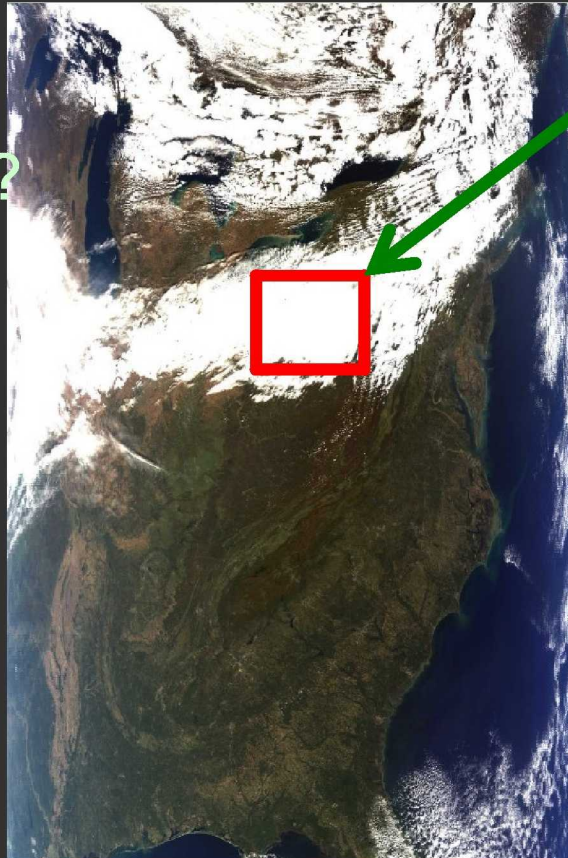
Quality Measures of Aggregated Data



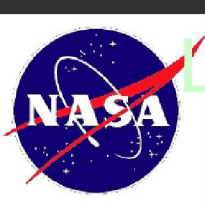
File-Level Quality Statistics may or may not be useful for data selection



Percent Cloud Cover?



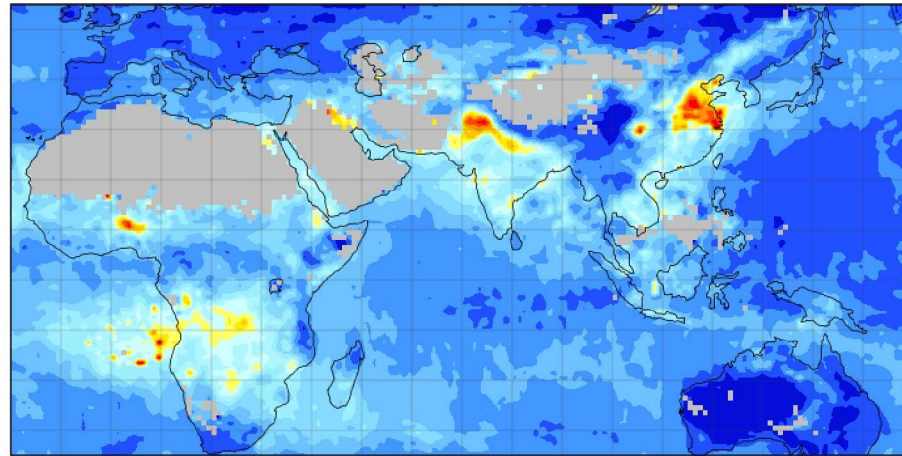
Study Area



Level 3 grid cell standard deviation is difficult to interpret due to its dependence on magnitude

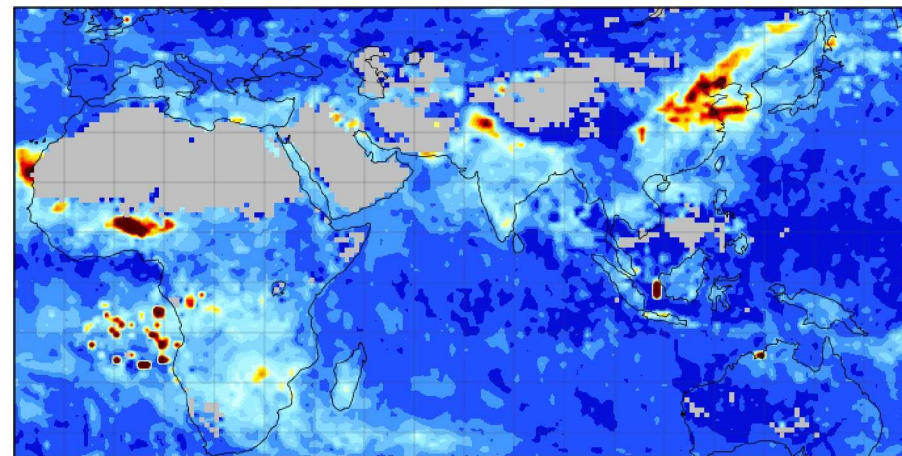
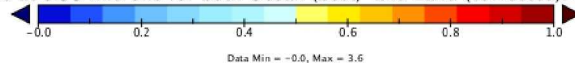
MODIS Aerosol
Optical Thickness
at 550 nm

Aerosol Optical Thickness at 0.55 microns for both Ocean (best) and Land (corrected): Mean of Daily



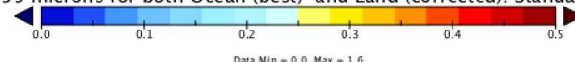
Mean

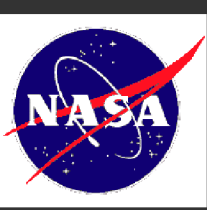
Aerosol Optical Thickness at 0.55 microns for both Ocean (best) and Land (corrected): Mean of Daily Mean (



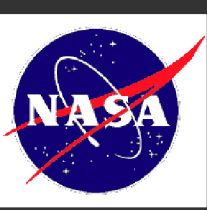
Standard
Deviation

Aerosol Optical Thickness at 0.55 microns for both Ocean (best) and Land (corrected): Standard Deviation of Daily





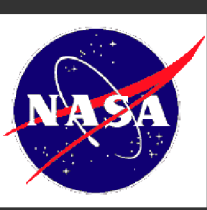
Potential Solution to Data Quality Ambiguity



Solution Part 1: Harmonize Quality Terms



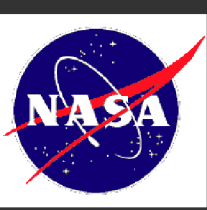
- œ ISO 19115 Data Quality model
- œ Committee on Earth Observation Satellites (CEOS)
 - œ Quality Assurance for Earth Observations (QA4EO)
 - œ Working Group on Cal/Val (WGCV)
- œ Federation of Earth Science Information Partners
 - œ ESIP Information Quality Cluster
- œ Quality Ontology
 - œ Data Quality Screening Service: Data-value QC
 - œ Aerostat: Level 2 Bias
 - œ Multi-sensor Data Synergy Advisor: Bias accuracy



Solution Part 2: Address more dimensions of Quality



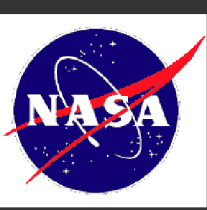
- ∞ Accuracy: bias + dispersion
 - ∞ Accuracy of data with low-quality flags
 - ∞ Accuracy of grid cell aggregations
- ∞ Consistency: spatial, temporal, observing conditions
- ∞ Completeness
 - ∞ Temporal: Time range, diurnal coverage, revisit frequency
 - ∞ Spatial: Coverage and Grid Cell Representativeness
 - ∞ Observing conditions
 - ∞ **N.B.: Incompleteness affects accuracy via sampling bias**



Solution Part 3: Address Fitness for Purpose Directly

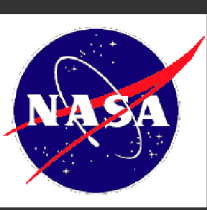


- ⌘ Standardize terms of recommendation
- ⌘ Enumerate more positive realms and examples
- ⌘ Enumerate negative examples

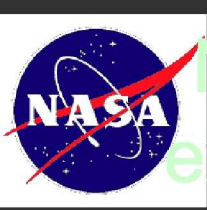


Facets of Quality





Backup Slides

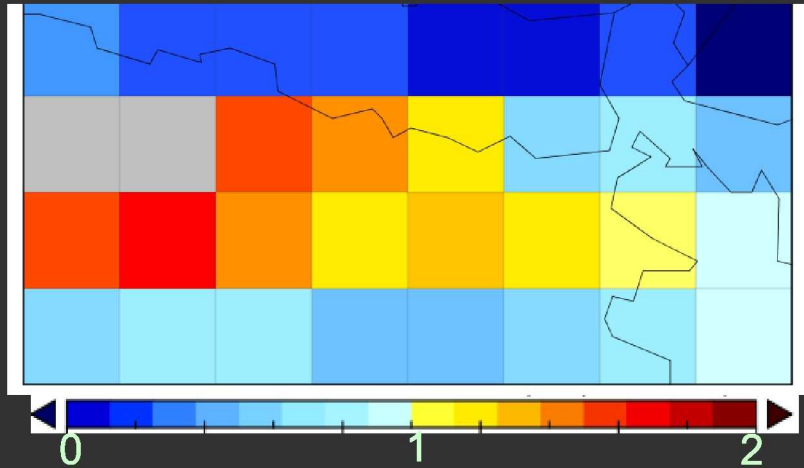


Neither pixel count nor standard deviation *fully* express how representative the grid cell value is

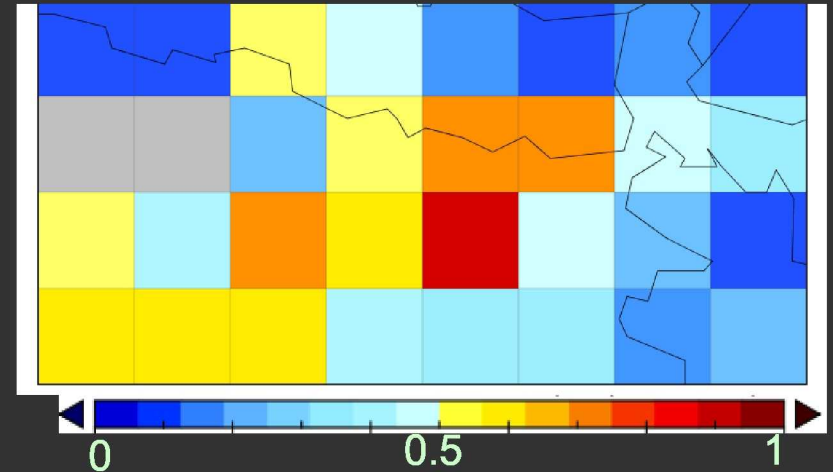


MODIS Aerosol Optical Thickness at 0.55 microns

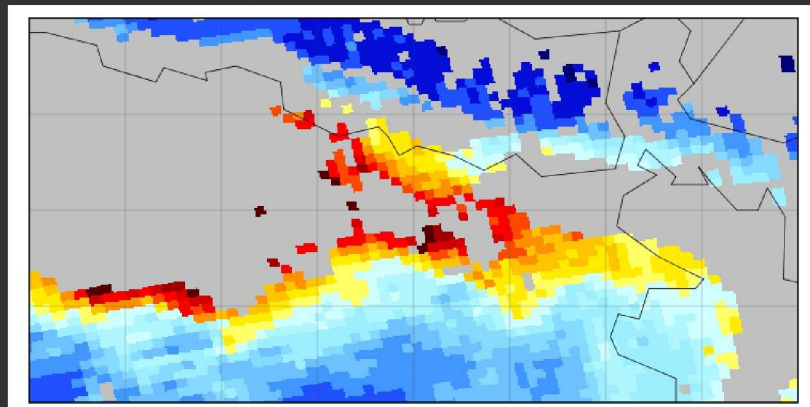
Level 3 Grid AOT Mean



Level 3 Grid AOT Standard Deviation



Level 2 Swath



Level 3 Grid AOT Input Pixel Count

