Unpacking and Demystifying ARM Standardized QC

ASR Science Meeting, March 2010

ALWG Translator: Connor Flynn

Pacific Northwest
NATIONAL LABORATORY

Outline

- Instrument data QC challenges
  - Easy – threshold checks applied to data values
  - Harder – indeterminate data.
- Value-Added Processing (VAP) QC challenges
- Challenges in forging a common standard
  - Lack of definition of what constitutes “QC”
  - Mutually exclusive requirements of different users
- The current standard:
  - Includes a concise **definition of “QC”**
  - **Simplifies interpretation**
  - **Facilitates automation**

Pacific Northwest
NATIONAL LABORATORY

Instrument-level flags: “Mentor QC”

Immediate real-time indicators of data quality.

Simple design for automation

- Min / max / delta checks
  - Provided for each field of interest.
  - These checks apply only to the parent field.
  - E.g., qc for data field named “RH” would be named “qc_RH” and would record results of threshold tests of that field alone. For instance valid_min = 0 and valid max = 100 in this case.
- Obviously not the answer to all data quality issues.
  - Doesn’t consider relevant house-keeping data.
  - May be difficult to express useful threshold for some fields. CBH
  - Reflectivity or backscatter, multi-dimension fields in general.

Pacific Northwest
NATIONAL LABORATORY

VAP QC: much broader scope

- As post-processed products VAPs have access to wider range of information to assess data quality.
  - Related house-keeping fields within input data stream
  - Pertinent measurements in other data streams
  - Application of physical or model based constraints
  - Noise estimation
  - And so on…
- But because many VAPs were first-of-a-kind efforts representing unique products QC was not reported in uniform or consistent formats.
- Moreover, some VAPs retrieved quantities for which no direct validation was apparent, so quality assessment was sometimes deferred to later comparisons.

Pacific Northwest
NATIONAL LABORATORY

No consensus on what constitutes QC.

There is an abundance of different types of meta data and data information in ARM data files, but not all are “QC”.

- Status information
- Sentinel values
- Relative “Quality Metrics”

RESOLVED:

- Quality Check: T/F test pertaining to data quality

Pacific Northwest
NATIONAL LABORATORY

How to satisfy competing requirements?

Requirements on QC standard:

- Comprehensive and complete
- Rigid format for automation
- Easy to use/interpret for data users.

Essentially each identified QC test includes:

1. A proscribed form of QC tests as T/F expressions with "T" indicating a failed test. Eg, data < valid_min.
2. An explicit statement of the implications of the test on the data quality. That is, does a failed test indicate the data is merely suspect or that it is truly bad.

In addition we defined a means of "distilling" the detailed QC information into a final summary file.

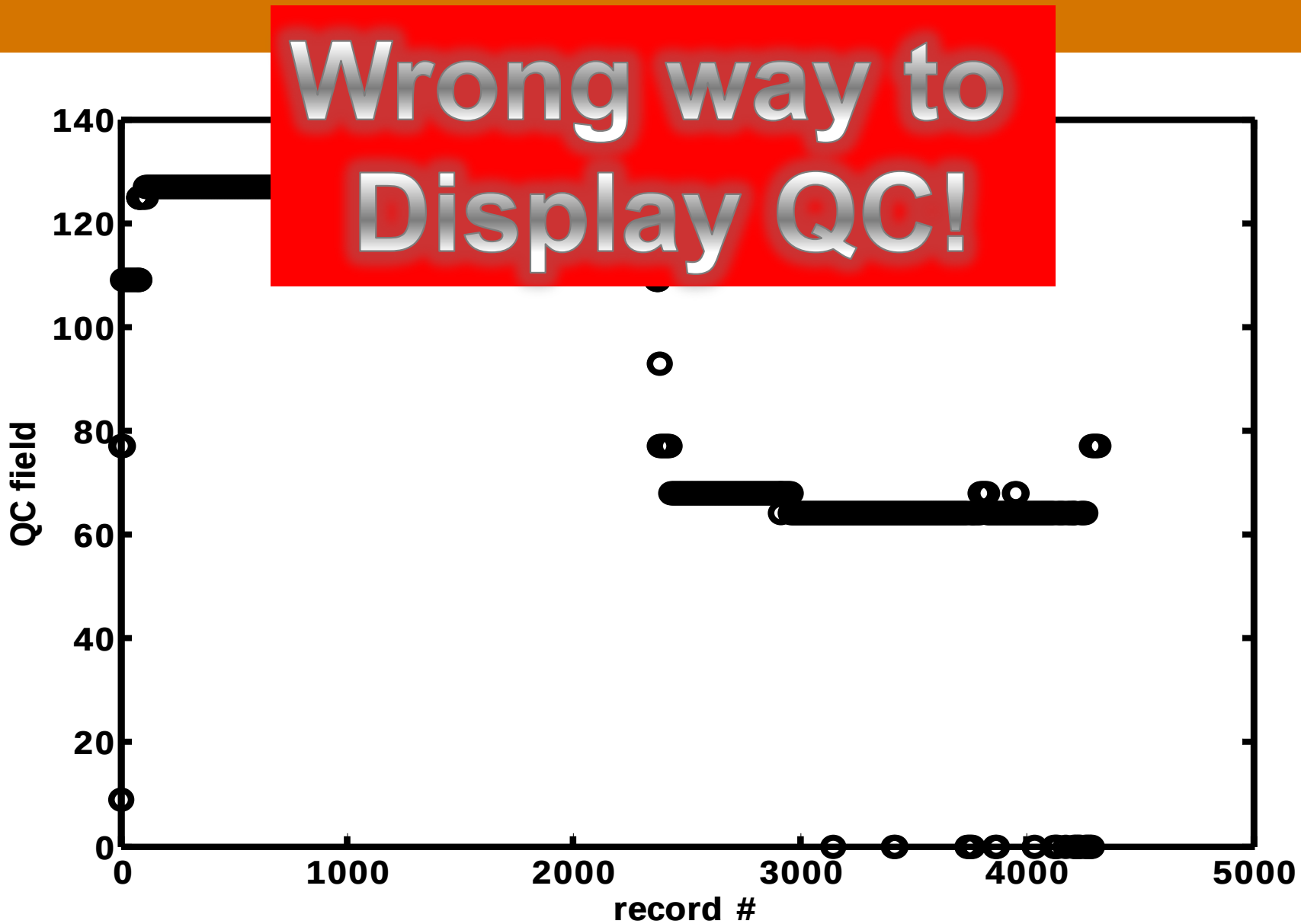
Pacific Northwest
NATIONAL LABORATORY

Example of a QC field definition:

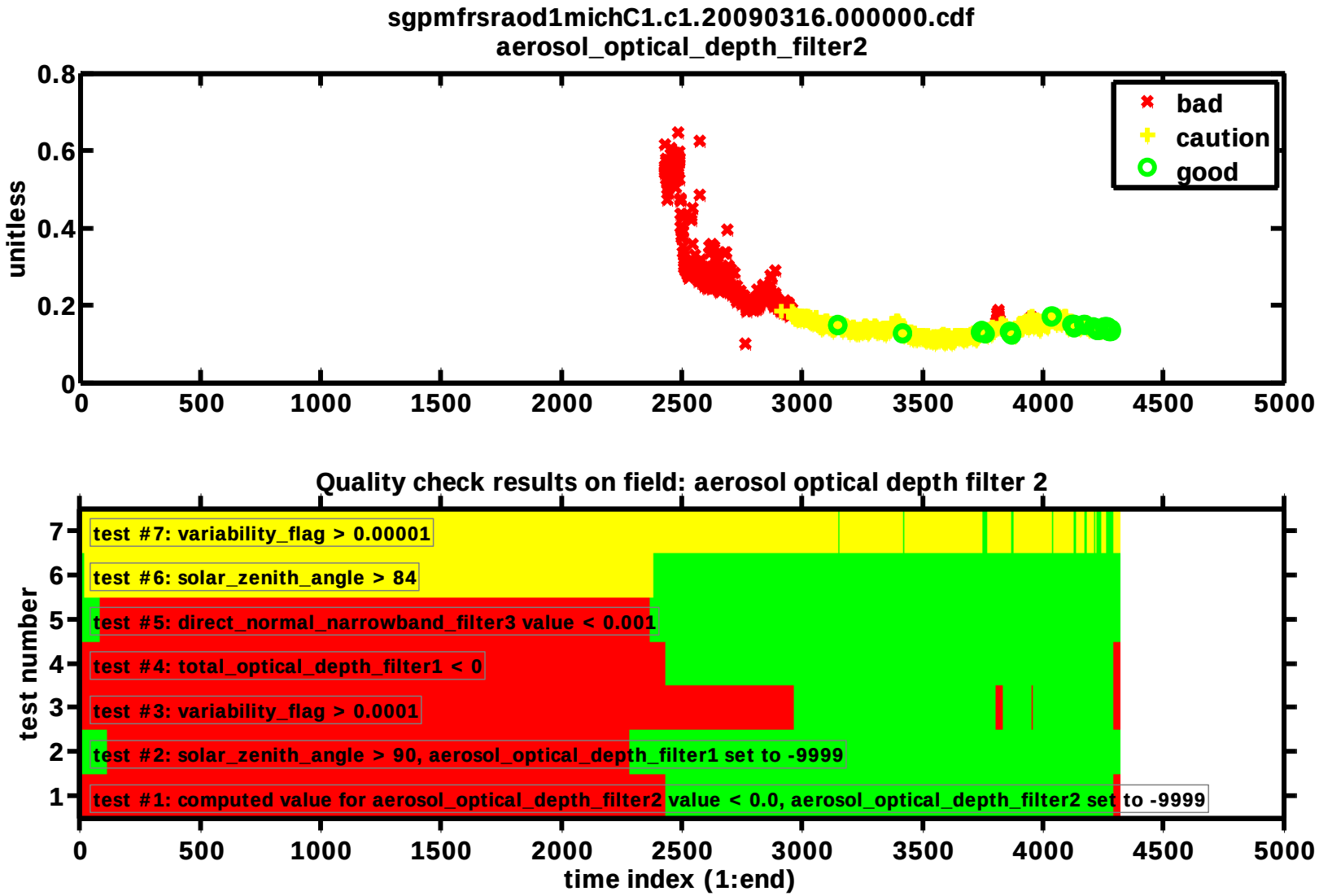
- Description: This field contains bit packed values which should be interpreted as listed. No bits set (zero) represents good data.
- bit_1_description: “aerosol_optical_depth_filter2 value < 0.0”
- bit_1_assessment Bad
- bit_2_description solar_zenith_angle > 90,
- bit_2_assessment Bad
- bit_3_description total_optical_depth_filter1 < 0
- bit_3_assessment Bad
- bit_4_description solar_zenith_angle > 84
- bit_4_assessment Indeterminate
- … etcetera…

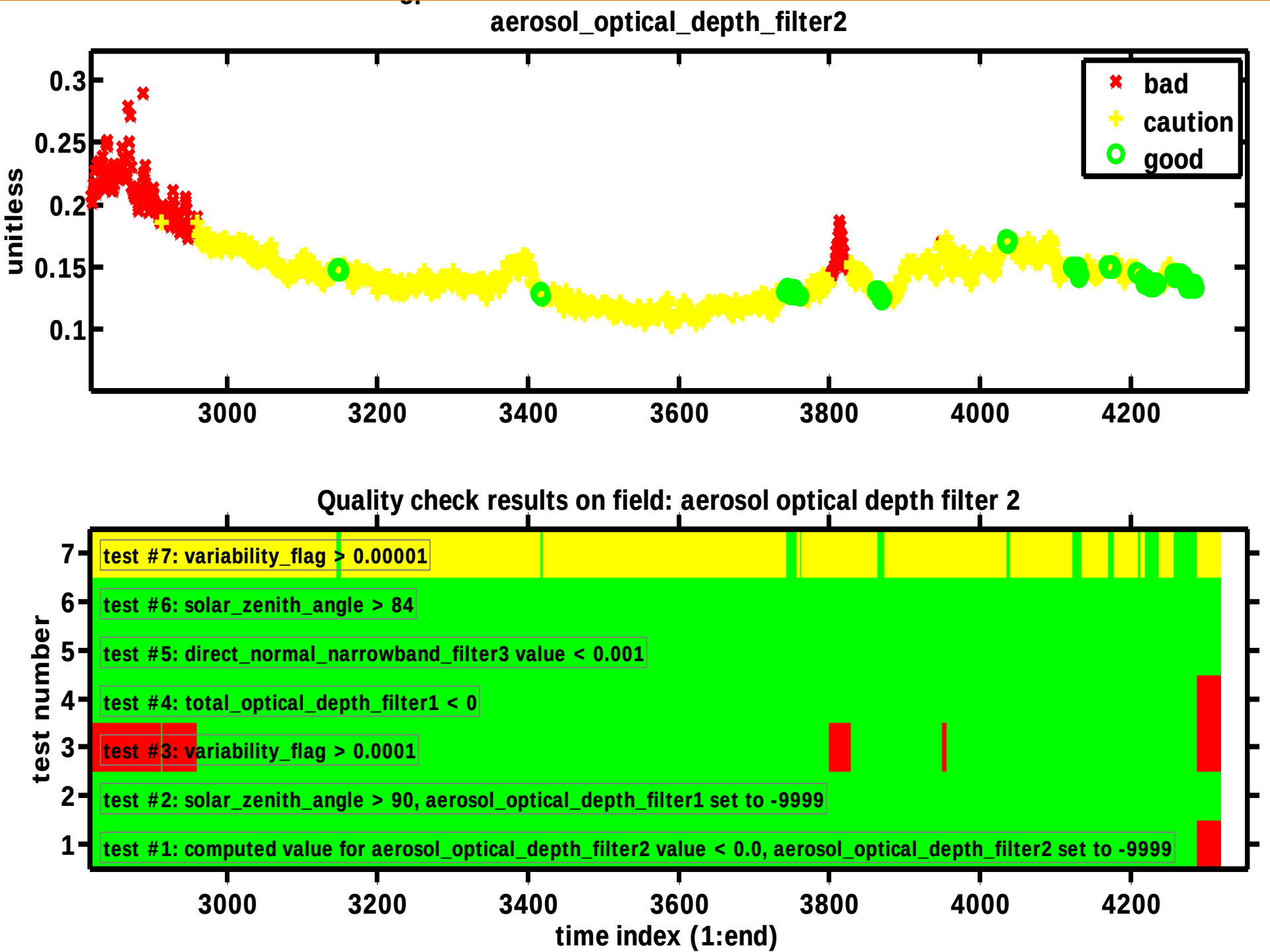
Pacific Northwest
NATIONAL LABORATORY

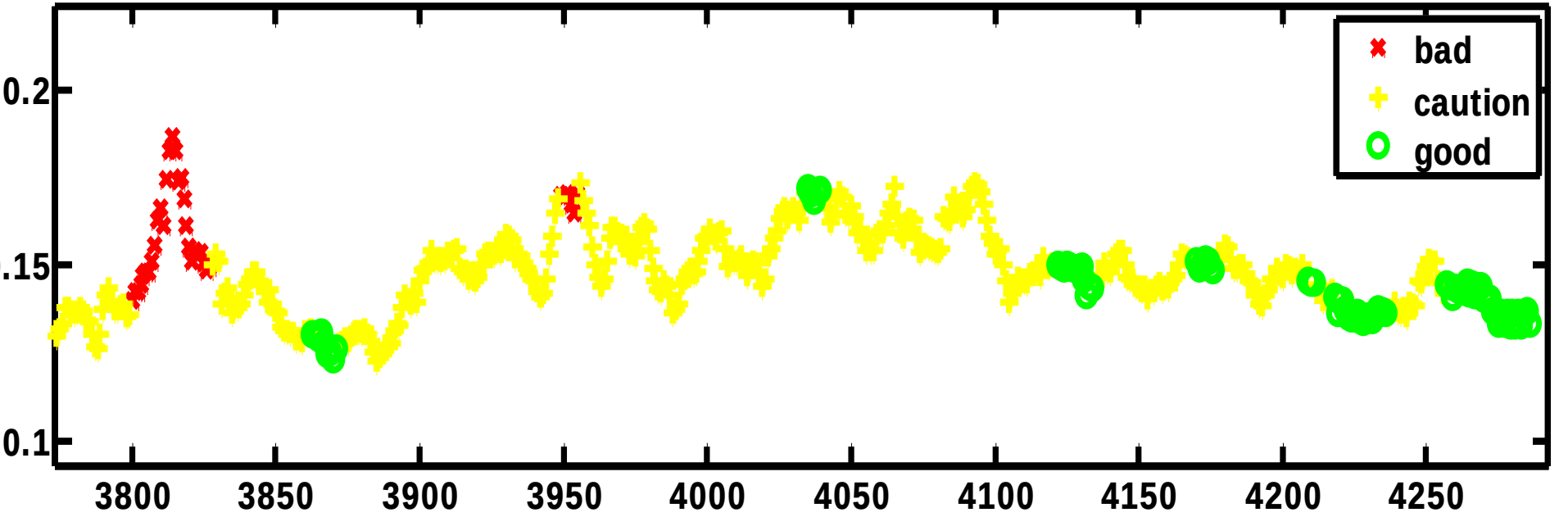
Efficient storage as bit-packed integer, but not really a “number”.

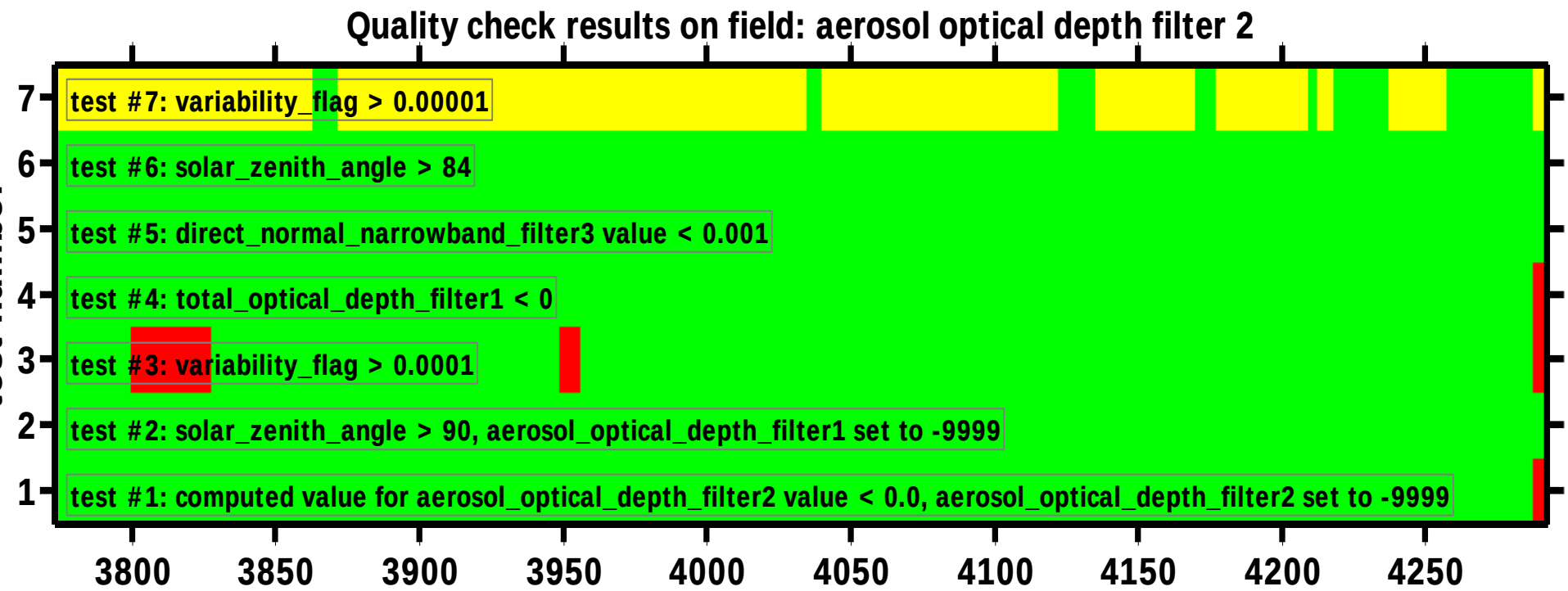


The QC field stores a collection of tests – each with a defined impact on data quality.




aerosol_optical_depth_filter2
unitless
0.3
0.25
0.2
0.15
0.1
3000
3200
3400
3600
3800
4000
4200
bad
caution
good
Quality check results on field: aerosol optical depth filter 2
test number
7
6
5
4
3
2
1
test #7: variability_flag > 0.00001
test #6: solar_zenith_angle > 84
test #5: direct_normal_narrowband_filter3 value < 0.001
test #4: total_optical_depth_filter1 < 0
test #3: variability_flag > 0.0001
test #2: solar_zenith_angle > 90, aerosol_optical_depth_filter1 set to -9999
test #1: computed value for aerosol_optical_depth_filter2 value < 0.0, aerosol_optical_depth_filter2 set to -9999
3000
3200
3400
3600
3800
4000
4200
time index (1:end)


bad
caution
good
0.2
.15
0.1
3800
3850
3900
3950
4000
4050
4100
4150
4200
4250


Quality check results on field: aerosol optical depth filter 2
test #7: variability_flag > 0.00001
test #6: solar_zenith_angle > 84
test #5: direct_normal_narrowband_filter3 value < 0.001
test #4: total_optical_depth_filter1 < 0
test #3: variability_flag > 0.0001
test #2: solar_zenith_angle > 90, aerosol_optical_depth_filter1 set to -9999
test #1: computed value for aerosol_optical_depth_filter2 value < 0.0, aerosol_optical_depth_filter2 set to -9999
7
6
5
4
3
2
1
3800
3850
3900
3950
4000
4050
4100
4150
4200
4250

Brief Demo…

Pacific Northwest
NATIONAL LABORATORY