SST Reliability Categories and Reflectance Table Documentation

JACKIE MAY

CHARLIE BARRON

Ocean Data Assimilation and Probabilistic Prediction Section
Ocean Sciences Division

JEAN- FRANÇOIS CAYULA

Peraton, Inc.
Herndon, VA

October 20, 2022

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD-MM-YYYY) 20-10-2022	2. REPORT TYPE NRL Memorandum Report	3. DATES COVERED (From - To)

4. TITLE AND SUBTITLE

SST Reliability Categories and Reflectance Table Documentation

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jackie C. May, Charlie N. Barron, and Jean-François Cayula*

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
9066

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
1005 Balch Blvd
Stennis Space Center, MS 39529

Peraton, Inc.
12975 Worldgate Dr.
Herndon, VA 20170

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/7320/MR--2022/9

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
One Liberty Center
875 N. Randolph Street, Suite 1425
Arlington, VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)

ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
*Peraton, Inc., 12975 Worldgate Dr., Herndon, VA 20170

14. ABSTRACT

The Naval Research Laboratory (NRL) Sea Surface Temperature (SST) software suite categorizes each produced SST retrieval into one of three categories based on the estimated reliability of the data. Reliability categories, from most to least reliable, are respectively labeled "clear" (1), "probably clear" (2) and "questionable" (3). Quantitative reliability estimates in the form of bias and standard deviation are then derived from comparisons to buoys for each category, and assigned to each retrieval. This report documents how the NRL SST software suite categorizes the SST retrievals. Two methods currently coexist to categorize the SST data. The first method is the 20-year-old legacy categorization, which mostly depends on the "field test", a comparison between SST retrievals and the corresponding analyzed SST field. The field test-based categorization is complemented by a series of ad hoc test to recover incorrectly categorized retrievals as the field test performs poorly in dynamic regions. This method has become less effective with the improved detection and discrimination of SST contamination that has been implemented within the NRL SST suite. The second method is the new standard categorization method, which relies on the tests that detect contamination of the SST retrievals within the NRL SST suite, and how well the retrievals pass the contamination tests, to assign reliability categories. For daytime, the "ref2d" reflectance table test is shown to be most valuable to categorize SST retrievals. For nighttime, the 4 μm brightness temperature inter-comparison test appears most effective for that categorization. Additional details on the standard categorization method and on other contamination tests which are used to estimate the reliability of the retrievals are included within this report. Currently the NRL SST suite uses the legacy categorization method to assign reliability categories for GOES-16, GOES-17, and Himawari-8 satellite-derived SST retrievals. The standard categorization is used for S-NPP, JPSS1, Metop-B, and Metop-C. EWS-G1 derived SST retrievals are produced with updated legacy software which is distributed along the NRL SST suite and also uses the legacy, albeit simplified, categorization method. Detailed description of the standard categorization tests, as well as information on the "ref2d" reflectance table, are included within this report.

15. SUBJECT TERMS

SST retrieval classifications
Brightness temperature reflectance

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Jackie May
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	U	21	**19b. TELEPHONE NUMBER** (include area code) (228) 688-4302

This page intentionally left blank.

CONTENTS

FIGURES

TABLES

This page intentionally left blank.

SST RELIABILITY CATEGORIES AND REFLECTANCE TABLE DOCUMENTATION

1 INTRODUCTION

The Naval Research Laboratory (NRL) Sea Surface Temperature (SST) software is designed to produce high quality satellite-derived SST retrievals from brightness temperature data at standard wavelengths by applying nonlinear and multichannel SST equations (Walton, 1998; Cayula, 2014). Brightness temperature and reflectance (visible) data are also used to detect contamination that reduces the accuracy of the SST retrievals (May, 1998; Cayula, 2016; May, 2021). Table 1 summarizes the various sensors, brightness temperature channel wavelengths, and platforms that are handled by the NRL SST software suite. The Naval Oceanographic Office (NAVOCEANO) currently uses the NRL SST software suite v2.8 to produce near real-time SST retrievals from 4 polar-orbiting satellites and 4 geostationary satellites, including EWS-G1 which is processed with updated legacy software. The SST retrievals are provided to the Fleet Numerical Meteorology and Oceanography Center (FNMOC) for assimilation into Navy ocean models, provided to the Physical Oceanography Distributed Active Archive Center (PO.DAAC) for archival and use by the international Group for High Resolution Sea Surface Temperature (GHRSST), and used for thermal analysis or other applications such as front detection.

Table 1: Current sensors, brightness temperatures channels, and satellite platforms being processed by v2.8 NRL SST software suite. The blue rows indicate polar orbiting satellites while the green rows indicate geo-synchronous satellites. The orange row indicates the older generation EWS-G1 satellite, which is processed with updated legacy SST software.

Sensor	Brightness temperature channels (µm)	Satellite platform
Visible Infrared Imaging Radiometer Suite (VIIRS)	3.7, 10.8, and 12.0	Suomi-National Polar-orbiting Partnership (S-NPP) and Joint Polar Satellite System (JPSS-1)
Advanced Very High-Resolution Radiometer (AVHRR/3)	3.7, 10.8, and 12.0	Metop-B and Metop-C (Metop-A until end of mission on 15 November 2021)
Advanced Baseline Imager (ABI)	3.9, 10.3, and 12.3	Geostationary Operational Environmental Satellite (GOES) 16 and GOES 17
Advanced Himawari Imager (AHI)	3.9, 10.4, and 12.0	Himawari-8
Imager	3.9 and 10.8	Electro-optical infrared Weather System – Geostationary satellite (EWS-G1), formerly GOES-13

Real-time reliability was implemented following increased interest in the inclusion of quantitative reliability information to be distributed along with the SST retrievals (Donlon, 2007). Having such information is crucial when, for example, merging SST retrievals from various sources, for optimal interpolation or for data assimilation. Due to being run in an operational environment, the implementation of the reliability scheme has been kept simple, relying on classifying SST retrievals into just three categories: "clear" (1), "probably clear" (2) and "questionable" (3). Another reason for limiting the number of reliability categories to three is that the NRL SST software is designed to only generate high quality data and discard SST retrievals that cannot be readily used, particularly for data assimilation. Quantitative

reliability estimates in the form of bias and standard deviation for each reliability category are updated daily by comparing buoys to the SST retrievals for the previous 30 days of match-up data for each satellite. The bias, standard deviation, and also the legacy root means square information is then appended to each currently produced SST retrieval based on the retrieval's assigned reliability category. The quantitative reliability information for the EWS-G1 SST retrievals is manually updated as EWS-G1 SST processing relies on legacy software.

Currently two procedures coexist to determine the reliability categories: the original, or legacy, categorization and the updated, or standard, categorization. The legacy categorization, which is described in detail in Cayula, 2004, depends principally on the field test: a comparison of the satellite-derived SST retrieval with an analyzed SST field. Originally the analyzed field was a 100-km gridded SST field (K100). This was later replaced by a 10-km gridded SST field (K10) which is produced by merging the past few weeks of neighboring satellite-derived SST retrievals for each K10 cell. SST data are weighed, in a non-linear manner to preserve features, based on relative age, distance and estimated reliability. The field test works well in regions with little variability; however, it performs poorly in dynamic regions, often misclassifying category 1 retrievals as category 2 or 3 due to the K10 representing a static composite SST field from past observations. A series of ad hoc tests that utilize satellite and solar angles as well as both nonlinear and multichannel SST equations are used to recover incorrectly categorized retrievals, but they only very partially alleviate the problem. The issue is more problematic with the NRL SST suite as one of its features is a better retention of oceanic features compared to the previous generation of SST software that it replaces (Cayula, 2017).

Due to the problems with the field test in dynamically active areas, the standard categorization procedure was created as a replacement of the legacy categorization procedure. The standard categorization minimally uses the field test and instead mainly relies on the feature retaining tests that detect contamination of the SST retrievals within the NRL SST suite. How well the retrievals pass multiple contamination tests then determines the quality of the retrievals.

With the development of the standard categorization procedure, the legacy categorization procedure has been superseded and is no longer recommended for use. However, the standard categorization procedure has only been thoroughly validated with the polar-orbiting satellites. Additional evaluation and validation are necessary for implementation with the geostationary satellites. Therefore, while not ideal, the legacy categorization that is within the original reliability procedure from the previous generation of SST software is still used for GOES-16, GOES -17 (ABI sensor) and Himawari-8 (AHI sensor). The reliability procedure for NPP, JPSS1 (VIIRS sensor) and METOP-A/B/C (AVHRR/3 sensor) has been updated to the standard categorization. EWS-G1 (imager sensor) processing relies on legacy GOES SST software with updates because of the new data formats, although the cloud detection was not updated. The reliability estimation procedure is correspondingly based on the legacy reliability categorization but with only the field test applied. Because EWS-G1 is missing the 12.3μm channel, the added ad-hoc tests are not implementable.

The first part of this study presents an overview of the standard reliability categorization detailing which tests and parameters are selected to classify SST retrievals into reliability categories. The second part of this study provides documentation of the "ref2d" reflectance table, a key component of the daytime cloud detection and of the standard reliability categorization. Details on the construction of the reflectance data training set, as well as the creation of a given sensor/satellite "ref2d" reflectance table are included.

2 STANDARD CATEGORIZATION

The standard categorization in the reliability estimation procedure replaces the legacy categorization. Instead of relying mostly on the field test, the standard categorization leverages the calculations from the contamination and operational tests within the SST processing to determine the reliability categories of the SST retrievals.

Section 2.1 describes the initial analysis of the intermediate calculations for selected SST tests based on the processing of S-NPP VIIRS data. Section 2.2 describes the operational standard categorization procedure as it is applied for METOP-B/C (AVHRR/3) SST processing and S-NPP/JPSS-1 (VIIRS) SST processing.

2.1 Analysis and selection of SST for reliability estimation

This section analyzes the intermediate results impact from select SST tests on the estimated reliability of the SST retrievals. The selected SST tests include:
- "ref2d" reflectance
- 4μm inter-comparison
- brightness temperatures difference
- proximity to cloud
- satellite zenith angle

Of most interest is the standard deviation of the SST retrievals as a function of test results because the standard deviation is an indication of reliability. The analysis presented in this study is based on S-NPP VIIRS data and relies on two different evaluation methods: orbital overlap and buoy matchups.

The first evaluation method uses the orbital overlap described in (Cayula et al., 2015). It provides a dense comparison field but results are relative because the SST field is compared to itself. There is also a dependence with latitude as the amount of overlap varies with latitude, with minimum overlap near the equator and large overlap closer to the poles. The analysis for the orbital overlap method is mostly based on S-NPP VIIRS SST data from the entire day of 15 September 2018, including all longitudes and latitudes between 80°N and 80°S.

The second evaluation method is a comparison with in situ buoys. It offers a sparse set of match-up data but error measurements are absolute with respect to the buoy measurements, which provides both bias and standard deviation. However, buoy measurements are not always exact and in this study buoy measurements that differ from the analyzed SST field by more than 3°C are discarded. Buoys not being uniformly distributed across the ocean domain can also introduce bias in the evaluation of the data. The analysis for the buoy matchup method is based on a NAVOCEANO provided data set consisting of matched buoy temperature data to S-NPP VIIRS SST retrievals that span the period from 4 June to 8 July 2020. The data set contains an average of about 3000 observations per day.

2.1.1 *"ref2d" Reflectance Table Test*

The reflectance table test is a daytime only test that compares the satellite observed reflectance values at 0.9μm with corresponding reflectance table (ref2d) values. The ref2d, discussed further in Section 5, is defined as the maximum reflectance that retains 90 percent of atmospheric clear-sky retrievals (derived from a given reference set) for a defined range of solar reflection and satellite zenith angles. SST retrievals with reflectance values greater than the corresponding ref2d values are identified as clouds.

To analyze the use of the reflectance table test for reliability categories, we examine the normalized reflectance difference (*nrefdif*) between the satellite measured reflectance at 0.9μm (*refl*) and the corresponding ref2d value, defined as

$$nrefdif = \frac{refl - ref2d}{ref2d} \qquad (1)$$

The orbital overlap evaluation method is used to determine the reliability of various ranges of *nrefdif*, Table 2. The SST retrievals that are associated with *nrefdif* between 0.4 and 0.8 have the lowest standard deviation, meaning they are the most reliable. SST retrievals that are associated with *nrefdif* > 0.8 have a higher standard deviation, indicating less reliability. This result is expected because regions with high reflectance values often indicate clouds. More unexpectedly, however, is the SST retrievals that are associated with *nrefdif* < 0.4 are also associated with a higher standard deviation.

Table 2: Results of the orbital overlap analysis for defined ranges of normalized reflectance difference (nrefdif) values.

nrefdif Range	count	Standard deviation (°K)
> 1.0	114981	0.39
0.8 to 1.0	150606	0.4
0.6 to 0.8	373878	0.35
0.4 to 0.6	539627	0.35
< 0.4	108043	0.43

2.1.2   *4μm Inter-comparison Test*

The 4μm inter-comparison test is a nighttime only test that compares the observed 3.x μm brightness temperature (which is approximately at 4 μm) with a 4 μm estimate determined as a function of the 10.x μm and 12.x μm brightness temperatures. SST retrievals with large differences between the observed and estimated 4 μm brightness temperature indicate probable contamination. In particular for the VIIRS SST retrievals, when the maximum difference is greater than 0.5°K, the SST retrieval is flagged as contaminated.

To analyze the use of the 4 μm inter-comparison test for reliability categories, we examined *n4μmdif*, the magnitude of the difference between the measured (4_{obs}) and estimated (4_{est}) 4 μm brightness temperatures, normalized by the maximum accepted difference ($maxdiff = 0.5$°K), defined as

$$n4\mu mdif = \frac{|4_{obs} - 4_{est}|}{maxdiff} \qquad (2)$$

The orbital overlap evaluation method is used to determine the reliability of various ranges of *n4μmdif*, Table 3. There is a relatively steady increase of the standard deviation of the SST retrievals as the 4μm inter-comparison difference increases. The standard deviation is lowest (meaning the reliability is highest) for *n4μmdif* under 0.5, which corresponds to 80 percent of the data and a maximum difference of 0.25°K.

Table 3: Results of the orbital overlap analysis for specific ranges of normalized 4μm brightness temperature difference (n4μmdif) values.

n4μmdif Range	Count	Standard deviation (°K)
> 0.9	29855	0.33
0.7 to 0.9	82658	0.32
0.5 to 0.7	130875	0.3
0.3 to 0.5	227702	0.26
< 0.3	809597	0.25

We also use the buoy comparison evaluation method to determine the reliability of various ranges of *n4μmdif*, Figure 1. The buoy comparison evaluation method results agree with those presented for the orbital overall evaluation method, Table 3. As the *n4μmdif* increases, the standard deviation of the SST retrievals also increases. Additionally, as the *n4μmdif* increases, the relative frequency decreases. The relative frequency represents the percentage of SST retrievals for a given range of *n4μmdif* relative to the total number of SST retrievals included.
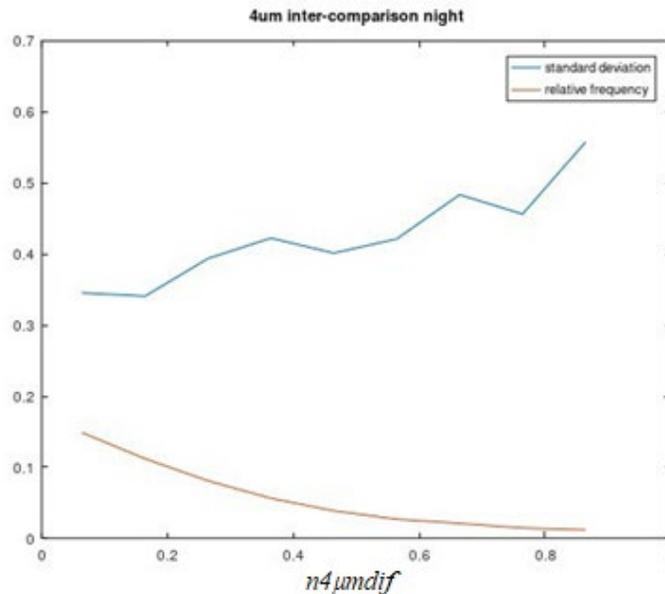
Figure 1: Results of the buoy analysis for n4μmdif values. X-axis is the normalized difference between 4μm BT value and equation. The y-axis is the SST-buoy standard deviation (°K) in the blue line and the relative frequency in the red line.

2.1.3   *Brightness Temperature Difference Test*

The brightness temperature difference test is currently applied during both the daytime and the nighttime. This test compares the difference between the 10.x μm and 12.x μm brightness temperatures with an upper threshold, which is a function of the SST retrieval. SST retrievals that have a brightness temperature difference larger than the defined threshold are flagged as contaminated.

To analyze the use of the brightness temperature difference test for reliability categories, we examined *nbtdif*, the difference between the observed 10.x μm ($10.x_{BT}$) and 12.x μm ($12.x_{BT}$) brightness temperatures, normalized by the maximum threshold ($maxthld$), defined as

$$nbtdif = \frac{10.x_{BT} - 12.x_{BT}}{maxthld} \quad\quad\quad (3)$$

The brightness temperature difference test results are presented separately for daytime and nighttime.

2.1.3.1 DAYTIME

We determined the daytime reliability of various ranges of *nbtdif* using both the orbital overlap evaluation method (Table 4) and the buoy comparison evaluation method (Table 5). Both methods show that the SST retrievals associated with *nbtdif* between 0.2 and 0.8 have the lowest standard deviation, meaning they are the most reliable. The SST retrievals with *n4btdif* greater than 0.8 or lower than 0.2 have greater variability, meaning lower reliability. The amount of SST retrievals with *nbtdif* less than 0.2 represent about 20 percent of the SST retrievals with the orbital overlap method, and only about 3 percent of the SST retrievals with the buoy comparison method. For *nbtdif* greater than 0.8, the proportion of the data for the two methods are reversed. The amount of SST retrievals with *nbtdif* greater than 0.8 represent only about .2 percent of the SST retrievals with the orbital overlap method, and about 9 percent of the SST retrievals with the buoy comparison method.

With the buoy comparison evaluation method, Table 5, the SST retrieval bias for the various ranges of *nbtdif* are also shown with the mean difference. When *nbtdif* < 0.3, the SST retrievals have a negative mean difference, indicating a cold bias in the SST retrievals. A negative mean difference is typically associated with contaminated data, generally from clouds or aerosol. When *nbtdif* > 0.8, the SST retrievals have a more extreme positive mean difference, indicating a significant warm bias in the SST retrievals. This is possibly due to the effect of diurnal warming.

As an additional note here, the VIIRS SST processing also includes a lower brightness temperature threshold test. While the lower threshold was not investigated here; in light of Table 4, this test may need to be reexamined.

Table 4: Daytime results of the orbital overlap analysis for specific ranges of normalized brightness temperature difference (nbtdif) values.

nbtdif Range	Count	Standard deviation (°K)
0.8 to 1.0	1997	0.59
0.6 to 0.8	81068	0.27
0.45 to 0.6	118105	0.3
0.3 to 0.45	341114	0.36
0.2 to 0.3	403162	0.37
0.0 to 0.2	294640	0.41
< 0.0	47049	0.42

Table 5: Daytime results of the buoy match-up analysis for specific ranges of normalized brightness temperature difference (nbtdif) values.

n4btdif Range	Relative frequency	Mean difference (°K)	Standard deviation (°K)
0.8 to 1.0	0.09	0.21	0.55
0.6 to 0.8	0.26	0.06	0.47
0.45 to 0.6	0.32	0.03	0.41
0.3 to 0.45	0.24	0.02	0.42
0.2 to 0.3	0.05	-0.06	0.49
0.0 to 0.2	0.03	-0.24	0.77
< 0.0	0.003	-0.86	0.65

2.1.3.2   NIGHTTIME

While the brightness temperature difference test is now applied to the VIIRS SST processing during both daytime and nighttime, the VIIRS SST data used in this analysis only had this test applied during the daytime. Therefore, to investigated the usefulness of this test for reliability categories during nighttime, we determined the nighttime reliability of various ranges of *nbtdif* using only the buoy comparison evaluation method (Table 6). Similar to the buoy matchup daytime results, the nighttime SST retrievals associated with *nbtdif* within the 0.2 to 0.8 range are the most reliable. Additionally, the nighttime SST retrievals within the 0.8 to 1.0 *nbtdif* range also have a relatively low variability and minimal bias, indicating high reliability. The SST retrievals with *nbtdif* greater than 1.0 represent about 1 percent of the data, and exhibit higher variability (meaning less reliability) but still a low bias.

Table 6: Nighttime results of the buoy match-up analysis for specific ranges of normalized brightness temperature difference (nbtdif) values.

nbtdif Range	Relative frequency	Mean difference (°K)	Standard deviation (°K)
> 1.0	0.01	0.01	0.55
1.0 to 0.8	0.08	-0.01	0.41
0.8 to 0.6	0.28	-0.01	0.37
0.6 to 0.45	0.34	-0.04	0.35
0.45 to 0.3	0.22	-0.06	0.33
0.3 to 0.2	0.04	-0.02	0.42
0.2 to 0.0	0.02	-0.08	0.51
< 0.0	0.003	-0.39	0.52

2.1.4   Proximity to Cloud Test

The proximity to cloud test is applied during both the daytime and the nighttime. This test is used in the SST processing to remove cloud-contaminated pixels at the edge of clear regions. A given pixel is

determined to be cloud-contaminated if at least one point in the 3x3 pixel window surrounding the given pixel is classified as "not clear".

To analyze the use of the proximity to cloud test for reliability categories, we examined the daytime SST retrievals as a function of the pixel cloud classification using the orbital overlap evaluation (Table 7). The reliability of samples that are adjacent to clouds (or more correctly contaminated regions) vs. the reliability of samples within atmospheric clear-sky conditions is measured. The SST retrievals with only clear neighbors (defined by the 3x3 pixel window) have a lower standard deviation than samples with neighboring flagged pixels. While not show, the number of flagged pixels surrounding a sample, from 1 to 8, does not appear to affect the reliability of the statistics.

The proximity to cloud test was also analyzed for reliability categories of the nighttime SST retrievals. However, the results for nighttime processing were inconclusive. This is possibly a consequence of the heavier approach to removing data based on proximity to cloud for nighttime processing than for daytime processing.

Table 7: Daytime results of the orbital overlap analysis for effect of proximity to cloud

Pixel cloud classification	Count	Standard deviation (°K)
At least one "not clear" neighbor	298182	0.52
All clear neighbors	567816	0.44

2.1.5   Satellite Zenith Angle

The satellite zenith angle is also found to affect the reliability of the observations due to multiple reasons, including the path length and footprint size. If the satellite zenith angle is above a defined maximum threshold (sza_{max}), then the SST retrieval is discarded. The value of sza_{max} depends on the sensor and the satellite. It is set to 68° for all the GOES satellites and Himawari; 75° for S-NPP and JPSS1 and, 53°for METOP B/C.

The buoy comparison evaluation method is used to determine the reliability of various ranges of the satellite zenith angle (Figure 2). As the satellite zenith angle increases, the standard deviation also increases, meaning the reliability decreases. Similar results were achieved with the orbital overlap evaluation method; however, those results are complicated because at low latitudes around the equator the amount of orbital overlap is reduced.
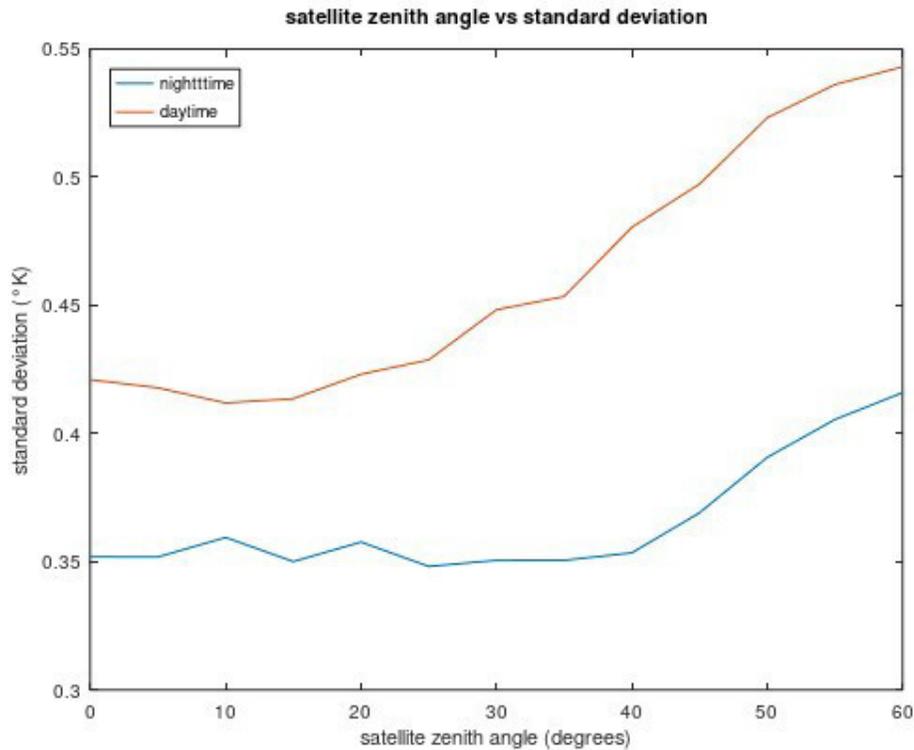
Figure 2: Standard SST-buoy standard deviation (°K) as a function of the satellite zenith angle. Nighttime results are shown in the blue line and daytime results are shown in the red line.

2.2 Operational Implementation

Based on the analysis of the SST tests in Section 2.1, the reliability module in the NRL SST software suite was updated to implement the new standard categorization method for VIIRS and AVHRR SST retrievals. The updated software was provided to NAVOCEANO, the transition partner receiving and operating the NRL SST software suite. NAVOCEANO then operationally tested the updated reliability procedure. Appropriate parameters and thresholds for VIIRS and AVHRR were determined and included in the operational SST processing. The operational daytime and nighttime standard categorization procedure is defined in Figure 3.
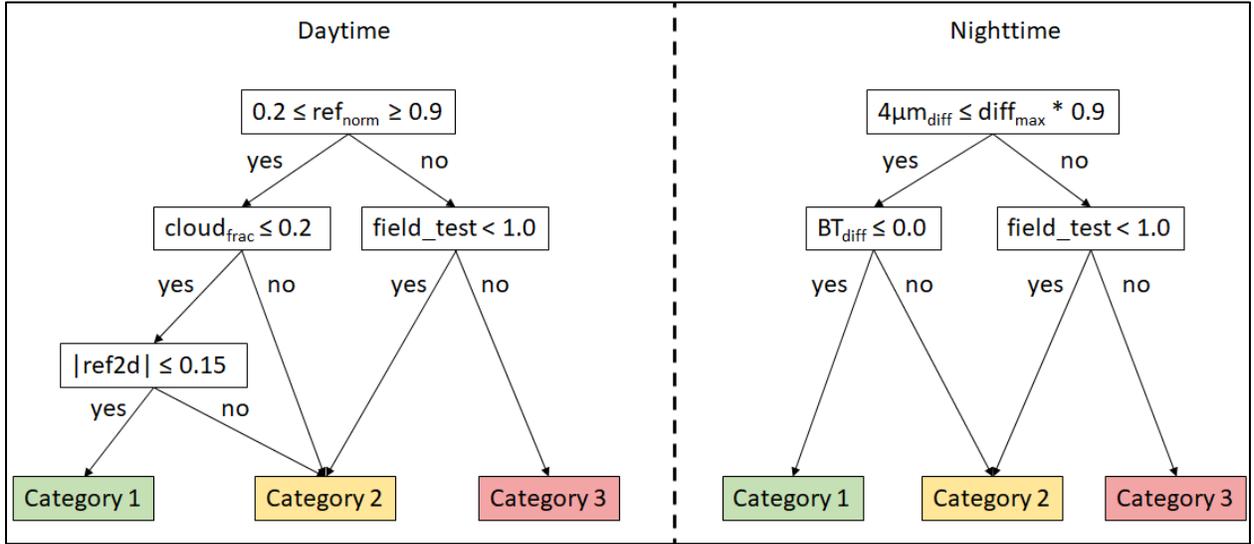
Figure 3: Procedure used for the standard categorization during daytime (left) and nighttime (right).

The field test which is used in certain cases to differentiate between categories 2 and 3 is the absolute difference between the SST field and the analyzed field, currently the K10.

For the daytime, the reflectance test (ref_{norm}) is defined as

$$ref_{norm} = \frac{refl}{ref2d} \qquad (4)$$

where *refl* is the reflectance at 0.9µm and *ref2d* is the reflectance table value determined for each sensor.

The cloud fraction test ($cloud_{frac}$) is defined as

$$cloud_{frac} = \frac{pixels_{cloudy}}{pixels_{total}} \qquad (5)$$

where *pixels_{cloudy}* is the number of cloudy pixels in a 3x3 pixel window centered on the given pixel and *pixels_{total}* is the total number of pixels in a 3x3 pixel window centered on the given pixel.

For the nighttime, the 4µm inter-comparison test ($4µm_{diff}$) is defined as

$$4µm_{diff} = |4_{obs} - 4_{est}| \qquad (6)$$

where 4_{obs} is the satellite measured *3.x* µm brightness temperature and 4_{est} is the estimated 4µm brightness temperature using the 10.x µm and 12 µm brightness temperatures. *diff_{max}* is set to 0.5°K.

The brightness temperature difference test (BT_{diff}) is defined as

$$BT_{diff} = 10.x_{BT} - 12.x_{BT} \qquad (7)$$

where $10.x_{BT}$ is the observed 10.x µm brightness temperature and $12.x_{BT}$ is the observed 12.x µm brightness temperature.

3 REFLECTANCE TABLE

As discussed in the previous sections, the reflectance table is used during the daytime for both SST retrieval processing as well as the determining the retrieval classification. Because it is applied during the daytime, it depends to a large extent on using the data from the visible channels. In particular, the SST retrievals rely on the reflectance at 0.9μm. Essentially, the reflectance table test compares the reflectance (refl) at each latitude / longitude location with a corresponding value in the reflectance table (ref2d), where ref2d is a function of the satellite zenith angle and the solar reflection angle. The SST retrieval passes the reflectance table test when the reflectance is less than or equal to the corresponding ref2d value, defined as

$$refl \leq ref2d \qquad\qquad (8)$$

There is approximately a 10 percent probability that locations that fail the reflectance table test are in fact clear. This is most likely to be true for locations with reflectance values only slightly above the corresponding ref2d value. Therefore, when the SST retrieval and the background field are sufficiently consistent ($(SST - K10) \geq -0.25°K$), a relaxed reflectance table test is applied, defined as

$$refl \leq ref2d * 1.5 \qquad\qquad (9)$$

In effect, we increase the tolerance to retain retrievals as long as those retrievals are also consistent with estimates based on the consensus of prior local measurements. This application of the relaxed reflectance table test reduces the risk that a false positive test incorrectly classifies locations as cloud-contaminated while simultaneously minimizing the effects when contaminated SST retrievals pass through as a consequence of the increased risk of a false negative. While the K10 background field is used here, this test could be modified to use alternate background estimates of SST.

3.1 Construction of a Training Data Set

Building the ref2d relies on an initial set of reasonably clean (low contamination) SST retrievals and associated data in order to determine the range of reflectance values corresponding to clear sky conditions. While the normal cloud detection process attempts to retain fine-scale variations that are associated with features such as ocean fronts, those features which are mostly independent of the reflectance are not needed to build the reflectance table, meaning a coarse set of reliable SST data is sufficient. For this purpose, a comparison test with the analyzed SST field is used with a small tolerance to only retain SST retrievals that are consistent with validated satellite-derived observations. The choice is to limit the produced SST values to a minimum of 0.5°K below the K10 analyzed field and a maximum of 1.0°K above. For this purpose, the following line can be added in the cloud parameter file:

fieldtest   -0.5   1.0   -0.5   1.0

Note the tighter tolerance on the low side is consistent with the expectation that cloud contamination will tend to produce a cold bias in retrievals.

In addition to enabling the field test, the ref2d test must also be turned off. As a reminder, the purpose of the SST retrievals that are produced here is to construct the ref2d itself. The reftbl entry in the cloud parameter file can be commented out to disable the reflectance test.

The ref2d depends on the satellite zenith angles, the solar reflection angles, and the reflectance. The best way to extract this information is to run the SST program in mode 3, which produces extended L2P files and is the only mode that will readily output the solar reflection angle.

Construction of the reflectance needs a reasonable, but not overwhelming, number of observations. The observations should span the full ranges of satellite zenith and solar reflection angles covered by the

corresponding satellite. It is recommended that about 4 days of data be processed, where the days are spread evenly across the months (and thus angles) of available input data.

3.2   Two-step Creation of ref2d

The creation of the ref2d is done in two steps. The first step processes the extended L2P files, which were produced by running the SST program in mode 3. The data that form the training data set are selected and extracted from the extended L2P files with the "lfreflectangle.m" octave script. At NAVOCEANO, the script should be installed in ${OPS}/prod/scripts directory.

The second step uses the extracted data produced from processing the extended L2P files to create the ref2d as a function of the satellite zenith angle and solar reflection angle with the "vistt2dsimple.m" octave script. At NAVOCEANO, the script should be installed in ${OPS}/prod/scripts directory.

The script to produce the ref2d attempts to determine the maximum clear-sky reflectance as a function of the satellite zenith angle and solar reflection angle. Although the field test produces an SST field which is mostly free from contamination, the test itself is not really a contamination test, as such some contamination leakage can be expected. To balance contamination, which results in incorrect higher reflectance maxima and removal of the clear data which depresses the maxima, the top 10 percent of reflectance samples in the training data set are discarded. Thus, the ref2d records the reflectance corresponding to the 90th percentile for the retrievals. These maximum reflectance values are binned as a function of the satellite zenith angle and solar reflection angle.

In the ref2d files, the solar reflection angle spans a range of 0° to 110° with 2° increments. Thus providing 55 values centered at the mid-point of each 2° interval. Similarly, the satellite zenith angles span a range of 0° to 70° range with a 2° increment. Thus providing 35 values centered at the mid-point of each 2° interval.

For reference, the ref2d and sample count for NPP VIIRS are shown in Figure 4 and Figure 5, respectively.
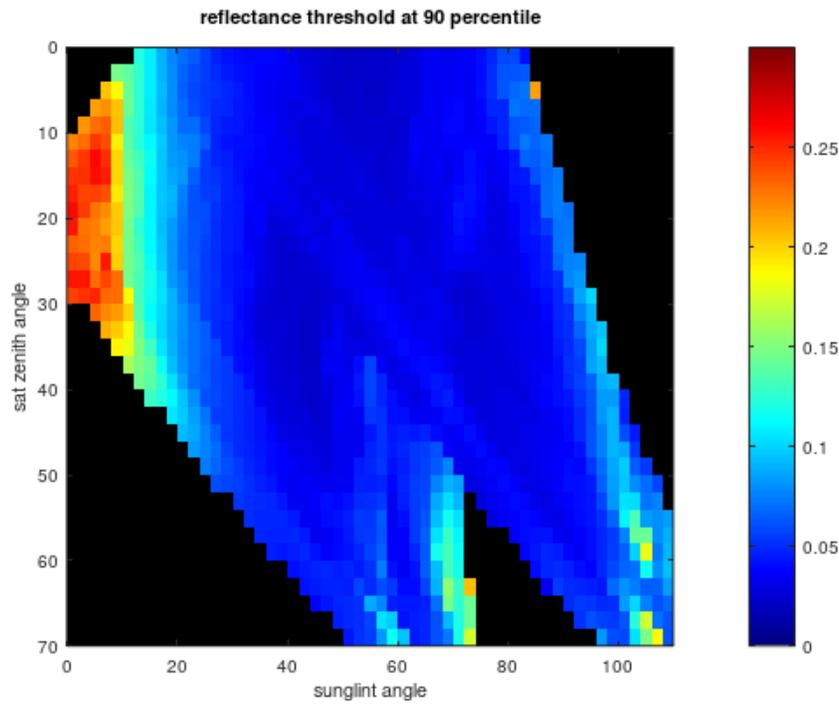
Figure 4: Reflectance table (ref2d) as a function of satellite zenith angle (y-axis) and sun glint angle (x-axis) for NPP VIIRS.
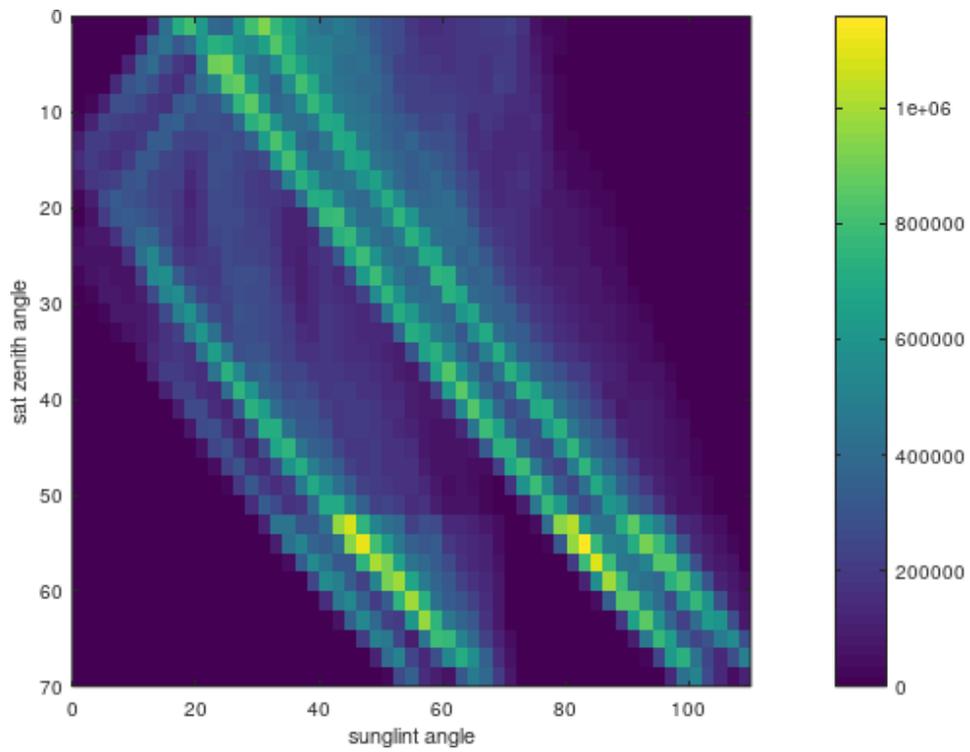

Figure 5: Sample count associated with ref2d in Figure 4.

3.3   One-step Creation of ref2d

For simplification, the "buildref2d.sh" script combines and sets up the execution of the two separate octave scripts described in the previous section. The "buildref2d.sh" script accepts 4 input arguments. However, running the script without arguments displays the help section, as shown below

```
>> buildref2d.sh


Script to build a ref2d table.


Usage:
        buildref2d.sh file_list [ ref2d_file [ stride [ mdir ]]]


Parameters:
    file_list: List of extended L2p files to analyze (mandatory)
    ref2d_file: name of ref2d file (default: refl2d6.txt)
    stride: sampling stride for input data (default: 2)
    mdir: directory containing m-scripts (default: /prod/scripts)
```

The first argument, file_list, is the name of the file that contains the list of all the extended L2P files to be processed. If the L2P files are in directory "/a/b/c", "file_list" could be obtained by typing "ls /a/b/c/*.nc > "file_list".

The second argument, ref2d_file, is the name of the output ref2d file. If the argument is missing, the default is "refl2d6.txt" in the current directory.

The third argument, stride, allows a specific number of pixels to be skipped in both x and y directions. The default "stride" value is 2. The stride argument may be useful to decimate the data when the amount of data is too voluminous. However, it can only be redefined if the second argument, ref2d_file, is present.

The last argument, mdir, defines the location where the octave scripts are installed. At NAVOCEANO, the argument is not necessary if the scripts are installed in their standard location: ${OPS}/prod/scripts.

4   CONCLUSIONS

The legacy and standard reliability categories associated with the SST retrievals produced from the NRL SST software suite v2.8 have been documented in this study. The legacy categorization determines the reliability categories for the GOES-16, GOES-17 and Himawari-8 SST retrievals. This procedure depends heavily on the field test, a comparison of the SST retrievals with the K10 analyzed field, and as such may not work well in dynamic regions. The EWS-G1 SST processing is based on legacy GOES SST software and accordingly also depends on the legacy categorization within the reliability module. However, the reliability categorization for EWS-G1 has been reduced to only the field test because the missing 12.3µm channel prevents the implementation of complementary tests.

The reliability module for S-NPP, JPSS-1 and METOP-B/C SST processing has been updated to implement the new, standard categorization methodology. Compared to the legacy method, the standard categorization depends much less on the field test and instead utilizes tests that detect contamination to determine the reliability categories. The "ref2d" reflectance table test was found to be the most valuable to categorizing the SST retrievals during the daytime, while the 4 µm brightness temperature inter-comparison test was found to be the most valuable during the nighttime.

This study also includes documentation for building the reflectance table.

The NRL SST reliability categorization provides a method to include quantitative reliability estimates with every operationally generated SST retrieval. From the implementation of the standard categorization method, thresholds for the decision rules have not needed updating. Thus, making the methodology robust and low-maintenance, which are major requirements in an operational environment. Future work should include the development of better statistics for regions with dynamic features. The assignment of reliability estimates to currently discarded retrievals would also make the attached reliability estimates more meaningful and allow end-users to better select data according to their needs.

Currently, the standard categorization procedure has only been validated for the polar-orbiting satellites. Future work should include the expansion of the standard categorization procedure to the geostationary satellites to take advantage of the retention of fine-scale variations. Additional evaluation and validation efforts are necessary for this implementation.

REFERENCES

1. Cayula, J.-F., May, D., McKenzie, B., Olszewski, D., Willis, K., 2004: Reliability estimates for real-time sea surface temperature. *Sea Technol.*, **45** (2), 67–73.

2. Cayula, J.-F., Arnone, R.A., Vandermeulen, R.A., 2014. Comparison of VIIRS SST fields obtained from differing SST equations applied to a region covering the northern Gulf of Mexico and western North Atlantic, Proceedings Volume 9111, Ocean Sensing and Monitoring VI; 911110 https://doi.org/10.1117/12.2053435

3. Cayula, J.-F., May, D.A., Arnone, R.A., Vandermeulen, R.A., 2015. Evaluation of VIIRS SST fields through the analysis of overlap regions between consecutive orbits, in: Ocean Sensing and Monitoring VII. Presented at the Ocean Sensing and Monitoring VII, SPIE, pp. 156–166. https://doi.org/10.1117/12.2179606

4. Cayula, J.-F., May, D., Willis, K., 2017. Thermal uniformity analysis of SST data fields. Proceedings of GHRSST XVII, 6-10 June 2016, Washington, DC. https://doi.org/10.5281/zenodo.4700197

5. Donlon, C., Robinson, I., Casey, K. S., Vazquez-Cuervo, J., Armstrong, E., Arino, O., Gentemann, C., May, D., LeBorgne, P., Piollé, J., Barton, I., Beggs, H., Poulter, D. J. S., Merchant, C. J., Bingham, A., Heinz, S., Harris, A., Wick, G., Emery, B., Minnett, P., Evans, R., Llewellyn-Jones, D., Mutlow, C., Reynolds, R. W., Kawamura, H., & Rayner, N., 2007. The Global Ocean Data Assimilation Experiment High-resolution Sea Surface Temperature Pilot Project, *Bulletin of the American Meteorological Society*, *88*(8), 1197-1214. https://journals.ametsoc.org/view/journals/bams/88/8/bams-88-8-1197.xml

6. Martin, M., Dash, P., Ignatov, A., Banzon, V., Beggs, H., Brasnett, B., Cayula, J.-F., Cummings, J., Donlon, C., Gentemann, C., Grumbine, R., Ishizaki, S., Maturi, E., Reynolds, R.W., Roberts-Jones, J., 2012. Group for High Resolution Sea Surface temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE). Deep Sea Res. Part II Top. Stud. Oceanogr. 77–80, 21–30. https://doi.org/10.1016/j.dsr2.2012.04.013

7. May, D. A., Parmeter, M. M., Olszewski, D. S., & McKenzie, B. D., 1998. Operational Processing of Satellite Sea Surface Temperature Retrievals at the Naval Oceanographic Office. *Bulletin of the American Meteorological Society*, *79*(3), 397–407. http://www.jstor.org/stable/26215051

8. May, J.C., Barron, C.N., Cayula, J.-F., 2021. Evaluating SST Retrieval Masking under Volcanic Ash Conditions (Memorandum Report No. NRL/7320/MR—2021/9). Naval Research Laboratory Stennis Space Center.

9. Walton, C. P., Pichel, W., Sapper, J., and May, D., 1998. The Development and Operational Application of NonlinearAlgorithms for the Measurement of Sea Surface Temperatures with the NOAA Polar-Orbiting Environmental Satellites. *J. Geophys. Res, 103*(c12), 27999-28012. https://doi.org/10.1029/98JC02370