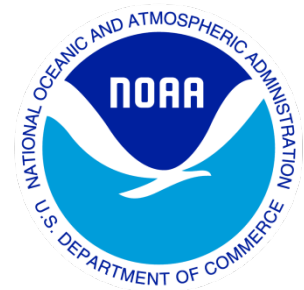

Climate Data Record (CDR) Program

Climate Algorithm Theoretical Basis Document (C-ATBD)

Sea Surface Temperature – Pathfinder



CDR Program Document Number: CDRP-ATBD-0099
Configuration Item Number: 01B-08
Revision 3 / July 20, 2016

A controlled copy of this document is maintained in the CDR Program Library.
Approved for public release. Distribution is unlimited.

REVISION HISTORY

Rev.	Author	DSR No.	Description	Date
1.0	Kay Kilpatrick, University of Miami RSMAS	DSR-104	Initial Submission to CDR Program	09/09/2011
2.0	Kay Kilpatrick, University of Miami RSMAS	DSR-342	Revised all sections	02/05/2013
3.0	Kay Kilpatrick, University of Miami RSMAS Sheekela Baker- Yeboah, University of Maryland CICS & NOAA/NESDIS/NCEI	DSR- 1056	Revised sections 4.2,5.4,6.2 Revised section 3.2, 5.5, 7.1 Changes of organization names: NCDC to NCEI-NC NODC to NCEI-MD Major revisions to section 5 for Amazon Web Service processing. Additional revisions to all sections based comments from the CDR group. Updated Figure 1 and Section 4.3.2. Updated Tables 5-12	07/20/2016

TABLE of CONTENTS

Update this table by right click (Windows) or Control-click (Mac OS-X) and selecting the "Update Field" option from the context menu.

1. INTRODUCTION.....	6
1.1 Purpose	6
1.2 Definitions.....	6
1.3 Referencing this Document	6
1.4 Document Maintenance.....	6
2. OBSERVING SYSTEMS OVERVIEW.....	7
2.1 Products Generated	7
2.2 Instrument Characteristics	7
3. ALGORITHM DESCRIPTION.....	10
3.1 Algorithm Overview	10
3.2 Processing Outline.....	10
3.2.1 Primary sensor data ingestion: ACISEG and ACI programs	12
3.2.2 Level 1 to Level 2 processing: SST Calculation	13
3.2.3 Level 2 Spatial Binning	13
3.2.4 Level 3 Temporal Binning and Mapping	14
3.2.5 Post Processing with GHRSSST Standards	15
3.3 Algorithm Input.....	15
3.3.1 Primary Sensor Data	15
3.3.2 Ancillary Data.....	16
3.3.3 Derived Data	18
3.3.4 Forward Models.....	18
3.4 Theoretical Description	18
3.4.1 Physical and Mathematical Description.....	18
3.4.2 Data Merging Strategy	20
3.4.3 Numerical Strategy	21
3.4.4 Calculations.....	21
3.4.5 Look-Up Table Description.....	21
3.4.6 Parameterization	22
3.4.7 Algorithm Output.....	23
4. TEST DATASETS AND OUTPUTS.....	33
4.1 Test Input Datasets	33
4.2 Test Output Analysis	33
4.2.1 Reproducibility.....	33
4.2.2 Precision and Accuracy	33
4.2.3 Error Budget.....	33
5. PRACTICAL CONSIDERATIONS.....	42

5.1	Numerical Computation Considerations.....	42
5.2	Programming and Procedural Considerations	42
5.3	Quality Assessment and Diagnostics	42
5.4	Exception Handling	44
5.5	Algorithm Validation	47
5.6	Processing Environment and Resources	47
6.	ASSUMPTIONS AND LIMITATIONS	49
6.1	Algorithm Performance.....	49
6.2	Sensor Performance.....	49
7.	FUTURE ENHANCEMENTS.....	51
7.1	Enhancement 1 Latband coeffs	51
7.2	Enhancement 2 -GHR SST compliant SSES bias ad STDEV	51
7.3	Enhancement 3: Level 2 pixel level ice flag.....	51
7.4	Resolving the Binning Problem in SeaDAS.....	52
8.	REFERENCES.....	54
	APPENDIX A. ACRONYMS AND ABBREVIATIONS.....	57

LIST of FIGURES

Figure 1:	PFSST V5.3 Data Processing.....	11
Figure 2:	A comparison of Pathfinder Sea Surface Temperature V5.2 (top panel) and V5.3 (bottom panel) with Canadian Meteorological Center V0.2 Global Foundation Sea Surface Temperature. The binning problem is apparent at high latitudes above 60 N as SST values are continuous across time (see section 7.4).	44
Figure 3:	Hovmöller difference plots for Pathfinder SST and daily optimum interpolated SST (dOISST) for 5.3 during 1981-2014 (upper panel) and 5.2 during 1981-2012.	53

LIST of TABLES

Table 1:	AVHRR Spectral Channels, Resolution and S/N.....	8
Table 2:	Temporal Coverage of sensors used in the PFSST CDR.....	9
Table 3:	Output files Parameter scaling and physical units.....	23
Table 4:	Archived CDR File naming components.....	31
Table 5.A:	Night time validation statistics NOAA- 7	34

Table 6.A Night time validation statistics NOAA- 9 35

Table 7.A: Night time validation statistics NOAA- 11..... 36

Table 8.A: Night time validation statistics NOAA- 14..... 37

Table 9.A: Night time validation statistics NOAA-16..... 38

Table 10.A: Night time validation statistics NOAA- 17 39

Table 11.A: Night time validation statistics NOAA- 18 40

Table 12.A: Night time validation statistics NOAA- 19 41

1. Introduction

1.1 Purpose

The purpose of this document is to describe the algorithm submitted to the National Centers for Environmental Information (NCEI) by Sheekela Baker/University of Maryland/NCEI that will be used to create the Sea Surface Temperature - Pathfinder V5.3 Climate Data Record (CDR), using the NOAA -7, -9, -11, -14, -16, -17, -18, -19 AVHRR sensors. The actual algorithm is defined by the computer program (code) that accompanies this document, and thus the intent here is to provide a guide to understanding that algorithm, from both a scientific perspective and in order to assist a software engineer or end-user performing an evaluation of the code.

1.2 Definitions

Following is a summary of the symbols used to define the algorithm.

Algorithm parameters:

$$T = \text{Brightness temperature} \quad (1)$$

$$\lambda = \text{wavelength of retrieval channels} \quad (2)$$

$$\theta = \text{view zenith angle} \quad (3)$$

Coefficient estimation parameters:

$$r = \text{robustness weight} \quad (4)$$

$$B(u) = \text{bisquare function} \quad (5)$$

1.3 Referencing this Document

This document should be referenced as follows:

Sea Surface Temperature - Pathfinder - Climate Algorithm Theoretical Basis Document, NOAA Climate Data Record Program CDRP-ATBD-0099 Rev.3 (2016). Available at <http://www.ncdc.noaa.gov/cdr>.

1.4 Document Maintenance

This document will be placed under version control as part of the CDR package. A copy will be archived with the CDR documentation package and made available to the public. This C-ATBD will be updated whenever the algorithm or production process has significant changes.

2. Observing Systems Overview

2.1 Products Generated

Generated products are global fields of Pathfinder sea surface temperature (PFSST) derived from the 5-channel Advanced Very High Resolution Radiometer (AVHRR sensors) flown on NOAA's polar orbiting satellites from 1981 to 2014. The Sea Surface Temperature – Pathfinder CDR products are available at a spatial resolution of 4.63km and a daily temporal resolution. Separate fields are produced from daytime only and nighttime only observations. In addition to the core CDR of sea surface temperature, statistics relating to the time of observation, quality of the retrievals, and related ancillary data derived from other data products, are also provided for each geographic bin in the data set.

2.2 Instrument Characteristics

To date fourteen AVHRR sensors have been flown aboard NOAA Polar orbiters. The TIROS-N series satellites were designed to operate in a near-polar, sun-synchronous orbit. The AVHRR's used in this CDR are five or six channel scanning radiometers and utilize a reflective Cassegrain collecting telescope. The sensor records data in five wavebands (Table 1). Channels 1 and 2 record reflected energy, Channel 3 switches to record reflected energy during the day and emitted energy at night for AVHRR/3 afternoon satellites, and Bands 4 and 5 record emitted thermal energy. The IR channels are calibrated in-flight using a view of a stable blackbody and space as a reference. No in-flight visible channel calibration is performed. Although these will vary from instrument to instrument, the design goals for the IR channels were an NE Δ T (Noise Equivalent delta Temperature) of 0.12 K (@ 300 K) and a S/N (signal to noise ratio) of 3:1 @ 0.5% albedo.

The AVHRR cross-track scanning system produces a swath width of 2800 km and each earth location is seen approximately twice per day. The field of view is scanned across the earth from one horizon to the other by continuous rotation of a flat scanning mirror. The spatial resolution instantaneous field of view (IFOV) is 1.1 km at nadir and increases to 5km at the extreme off nadir viewing angles of the swath. The orientation of the scan lines is perpendicular to the spacecraft orbit track and the speed of rotation of the scan mirror is selected so that adjacent scan lines are contiguous at the sub-satellite (nadir) position. Complete strip maps of the earth from pole to pole are thus obtained as the spacecraft travels in orbit. The orbital period is about 102 minutes, which produces 14.1 orbits per day. Because the number of orbits per day is not an integer, the sub-orbital tracks do not repeat on a daily basis, although the local solar time of the satellite's passage is essentially unchanged for any latitude. Instruments with an afternoon crossing times of approximately 1300-1400 (Table 2) predominantly are used in the Sea Surface Temperature – Pathfinder CDR. NOAA-17 is the only morning AVHRR included in the time series as a result of mechanical problems with NOAA-16 early in its life span. Using AVHRR aboard satellites with the same crossing time produces a nearly consistent time of day observation record.

The raw Level 1 datasets from these sensors are available at two spatial resolutions. High resolution, nominal 1km local area coverage (LAC) data is available over limited select

locations around the U.S. and some parts of the world. Reduced resolution, 4km global area coverage (GAC) data used as input to the Sea Surface Temperature – Pathfinder CDR is produced by onboard processors that average selected pixels. To produce this GAC data four out of every five samples along the scan line are used to compute one average value, and data from only every third scan line are processed. As a result, the spatial resolution of GAC data near the sub-point is actually 4.4 km by 1 km with a 1 km gap between pixels across the scan line and a 2km gap between scan lines, and is generally treated as having a 4 km resolution. All of the GAC data acquired during a complete orbit are recorded on board the satellite for transmission to Earth on command.

NOAA maintains two documents detailing sensor and platform attributes and characteristics. The NOAA Polar Orbiter Data User's Guide (PODUG November 1998 version) is a document that describes the orbital and spacecraft characteristics, instruments, data formats, etc. of the TIROS-N, NOAA-6 through NOAA-14 polar orbiter series of satellites and can be found at <http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/podug/index.htm>). The next generation instruments, improved 6 channel AVHRR/3 sensors, are covered by the NOAA KLM User's Guide (April 2007 revision). This document covers the NOAA-K through NOAA-M polar orbiter series of satellites NOAA-15 and later. In addition, the NOAA-N and NOAA-P spacecraft are also described in the NOAA-N,-P Supplement (2009) and can be found at <http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/intro.htm>. The Office of Satellite and Product Operations (OSPO) provides further information and status reports at <http://www.ospo.noaa.gov/Operations/POES/status.html>.

Table 1: AVHRR Spectral Channels, Resolution and S/N

Channel	Spectral Bandpass (micrometers)	Spatial Resolution at nadir (km)	Signal to Noise (S/N) or Noise Equivalent Delta Temperatures (NE Δ T)	IFOV (mr)
1 (Visible)	0.580 - 0.68	1.1	9:1 at 0.5% Albedo	1.39
2 (Near IR)	0.725 - 1.00	1.1	9:1 at 0.5% Albedo	1.41
3A (Near IR)	1.580 - 1.64	1.1	20:1 at 0.5% Albedo	1.51
3B (IR)	3.550 - 3.93	1.1	0.12 K at 300 K	1.51
4 (IR)	10.300 - 11.3	1.1	0.12 K at 300 K	1.41
5 (IR)	11.500 - 12.5	1.1	0.12 K at 300 K	1.30

Table 2: Temporal Coverage of sensors used in the PFSST CDR
 (* indicates the only morning satellite and ** indicates the gap period filled with PFSST 5.0 data from NOAA-9).

Satellite Number	Ascending Node	Descending Node	CDR Start Date use	CDR End Date use
NOAA-7	1430	0230	1981237	1985030
NOAA-9	1420	0220	1985031	1988313
NOAA-11	1340	0140	1988314	1994255
NOAA-9	1420	0220	1994256	1995018
NOAA-14	1340	0140	1995019**	2000285**
NOAA-16	1400	0200	2000286	2000366
NOAA-14	1340	0140	2001001	2001055
NOAA-16	1400	0200	2001056	2002192
NOAA-17	2200*	1000*	2002191	2005158
NOAA-18	1400	0200	2005159	2010365
NOAA-19	1400	0200	2011001	2014365

3. Algorithm Description

3.1 Algorithm Overview

The algorithm description begins in section 3.2 by presenting the processing flow through the main program modules; Level 1 data ingestion, Level 2 product generation, Level 3 aggregation into daily fields. In Section 3.3 we describe the various inputs required during processing, including primary sensor L1b data and ancillary fields. Finally in section 3.3 both the theoretical description and evolution of the Sea Surface Temperature – Pathfinder CDR algorithm are presented.

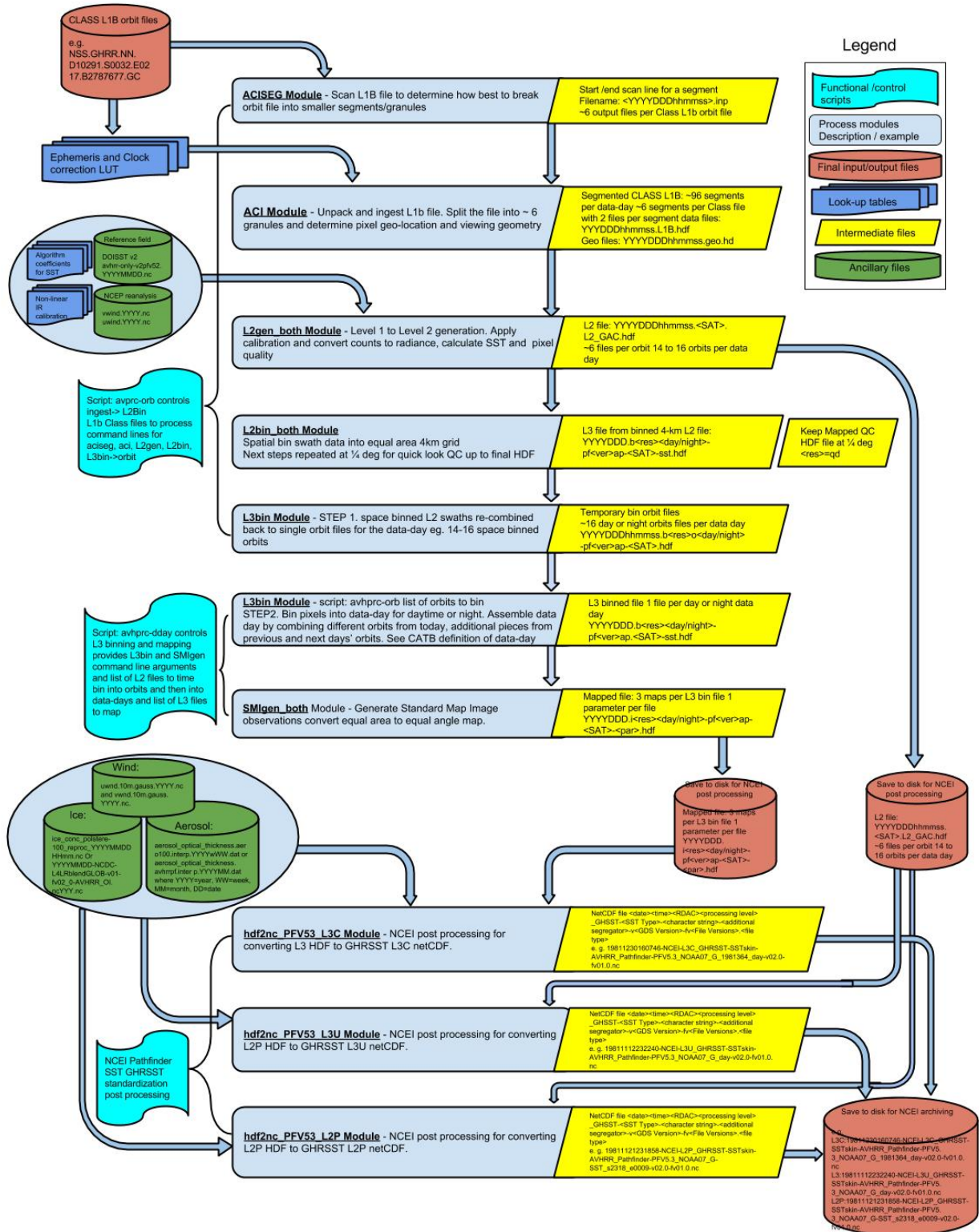
3.2 Processing Outline

Processing of the raw AVHRR channel data to produce the Sea Surface Temperature – Pathfinder CDR dataset occurs in a five-step process:

- Step 1. Ingestions, navigation, and calibration of the level-1 B sensor data
- Step 2. Pathfinder SST calculation and quality level flags
- Step 3. Spatial binning on a 4km global grid
- Step 4. Temporal binning and mapping
- Step 5. Post processing with GHRSSST (Group for High Resolution SST) standards

A detailed flow chart of the processing steps and data flows is presented in Figure 1. Details of the operations performed within each program module are detailed in the subsections below.

Figure 1: PFSST V5.3 Data Processing



A controlled copy of this document is maintained in the CDR Program Library. Approved for public release. Distribution is unlimited.

3.2.1 Primary sensor data ingestion: ACISEG and ACI programs

ACISEG: AVHRR L1B data are received in files containing approximately one orbit of Global area Coverage (GAC) data. An orbit contains about 100 minutes of counts from the AVHRR sensor aboard the spacecraft. For ease of processing, the large orbit file must be split into 6-8 "processing pieces" that separate the ascending and descending parts of the orbit. ACISEG is called approximately 6 times per CLASS L1b file. The L1b file is first scanned to obtain information on the times of the first and last scan lines of the file, and the time and scan line number of any pole crossings. Pole crossings define the transition between ascending and descending data, which are processed separately. The information collected and returned by ACISEG is used by the ingestion routine ACI to split the L1b file into several ascending and descending pieces, and reformat the segments into a file format required by the main l2_gen program. The ACISEG module can be thought of as a preprocessor of the raw Level 1b data that collects information on how best to break the file into small processing segments, extracts information from the header needed for geo-location and calibration, and also evaluates the data for missing or corrupted scan lines. This information is passed to the ACI program module, which performs the actual ingestion and reformatting of the primary input data files. The output of the aciseg module is a group of files named: <YYYYDDDHHMMSS>.inp. Where <YYYYDDDHHMMSS> represents the time of the start scan line of the segment. Specific details on the content of the *.inp output files can be found in the PFSST V5.3 header for aci.f.

ACI: Using the output information provided by the ACISEG, the ACI module begins ingestion by unpacking the L1b class data file, filling missing scan lines, and extracting the running calibration. To improve the accuracy of the navigation in the L1b file header, the navigation is recalculated using an orbital model, and attitude and clock corrections derived from a database of satellite-Earth time offsets data based on University of Miami DOMSAT recorded HRPT file clock differences observed when processing 1km format LAC data. Next the raw digital counts are converted to calibrated units. Calibration for visible *channels 1 and 2* uses pre-launch calibration coefficients to perform a linear counts-to-radiance conversion, followed by a correction for temporal changes using sensor decay rate data and a correction for inter-satellite differences using inter-satellite standardization data to a NOAA-9 reference, both of which use Libyan desert target area data. IR *Channels 3, 4, and 5* use pre-launch calibration data and onboard blackbody (space view and sensor base plate) data to perform a non-linear counts-to-radiance conversion. It should be noted that the final conversion to brightness temperature for these channels, as required by the algorithm, occurs during Level2 processing and not is done in by the ACI ingestion routine. The ACI module outputs the data granules as 2 HDF files; a geo-location file and an AVHRR channel data file that includes the unpacked running calibration from the L1b file as an SDS. The format and content of the HDF output files are compatible with the SEADAS l2_gen software packaged developed and maintained by NASA.

The ACI module is run ~approximately 6 times per L1b file ingested to segment the L1 data into approximately 15 minute granules of 2000 scan lines as specified in the YYYYDDDHHMMSS.inp output file of the aciseg module. The ACI module outputs 2 HDF files per segment. YYYYDDHHMMSS.L1b.hdf contains the channel data and calibration

information and YYYYDDHHMMSS.geo.hdf contains the navigation and viewing geometry for the granule. Files are named for the start of the first scan line in the file. Approximately 84 granules are produced per day (14 orbits with 6 segments per orbit)

3.2.2 Level 1 to Level 2 processing: SST Calculation

L2GEN: This module reads the navigation and radiance data files and applies calibration to convert to brightness temperature, and then applies the SST algorithm to calculate the Pathfinder sea surface temperature value (PFSST). The PFSST atmospheric correction algorithm is a statistical algorithm and relies on coefficients derived from co-located, co-temporal satellite and *in situ* SST observations. Coefficients are estimated on a monthly basis, and for two separate atmospheric regimes (dry and medium to moist atmospheres). Computation of the PFSST requires a first-guess SST. The first-guess reference SST used in version 5.3 of the Sea Surface Temperature – Pathfinder CDR products was derived from Optimum Interpolation $\frac{1}{4}$ -Degree Daily SST Analysis version 2 (DOISSTv2) fields described in Reynolds et al. (2007) and augmented with climatological information for lakes and rivers (see section 3.3.2.1 for additional details). Once the SST has been calculated, a series of tests are used to determine the quality level of the pixel. The results of these tests are aggregated to assign each retrieved SST to quality level 0-7 with 0 being the best quality and 7 being highly suspect. These tests and quality level assignment are discussed in detail later in this document under Section 5.4 Exception Handling. For Sea Surface Temperature – Pathfinder V5.3 processing, only levels 0-6 are retained in the archived CDR since level 7 is highly contaminated with clouds.

3.2.3 Level 2 Spatial Binning

L2BIN: An equal-area grid is defined into which GAC swath pixels are binned. The spatial gridding scheme adopted for PFSST processing is based on the equal-area grid adopted by the International Satellite Cloud Climatology Project (ISSCP). The equal-area grid consists of rectangular bins or tiles, arranged in zonal rows. These bins are approximately 4.63 km per side at the Equator (for brevity, they will be referred to as 4-km bins). This results in 4320 zonal rows of tiles from pole to pole (2160 in each hemisphere). The bins are arranged in rows beginning at 180° longitude and circumscribing the Earth eastward at a given latitude. The number of tiles per row varies with latitude. The rows immediately above and below the equator have 8640 bins; at the poles the number of tiles is always three. Between these two extremes, the number of bins in a zonal row decreases as an approximate cosine function of latitude. Coding details of the spatial gridding algorithm used in the RSMAS binning procedure can be found in Appendix b of the SeaWiFS Technical Memorandum Vol. 23 (Campbell et. al. 1995). Since the area of the bin corresponds to the nominal area of the GAC pixel, generally only 1 AVHRR sample is present in a given bin in a L2 granule. L2bin outputs 1 L2 granule for every L1b granule input to the L2gen module. Naming convention is the same as ACI except the processing level is called <L2_GAC>.

3.2.4 Level 3 Temporal Binning and Mapping

L3BIN: The spatially binned level 2 pieces from Step 3.2.3 are binned into a single daytime or single nighttime file based on the direction of the pass, ascending or descending, and the data-day. To speed processing and reduce the I/O load individual granules are first re-binned into temporary orbits files and the orbits are binned to create the global files. The L3_bin produces 1 day and 1 night file for each data-day. File name:

YYYYDDD.b<res><day/night>-pf<ver>ap-<SAT>-sst.hdf. Where : YYYYDDD = data-day, res = 4k (4 km) or qd (quarter degree), day=1 or night=3, ver=algorithm version (V5.3), sat=N[7-18]. The temporary orbit file naming convention is:

YYYYDDDDHSSMM.o<res><day/night>-pf<ver>ap-<SAT>-sst.hdf with the file name containing the start time of the orbit.

Definition of a data-day: A spatial definition of a data-day is adopted for Pathfinder global SST daily fields. Complete details and the rationale for the data-day definition can be found in Podesta1995 (<http://www.nodc.noaa.gov/sog/pathfinder4km/Data-day.pdf>). Briefly, the start of a data-day is defined as the time at which the satellite crosses the 180° meridian closest to the equator (there are several crossings of this meridian in a day, most of them at high latitudes). The data-day end is the next crossing of the 180° meridian closest to the equator. The time of these crossings varies as a function of the satellite orbit. It should be noted that the NOAA satellites cross the equator at the 180° meridian on the descending orbit approximately 12 hours prior to the ascending crossing. The start and end times for a data-day of descending orbit data, therefore, occur 12 hours prior to the corresponding times for the ascending orbit data-day.

Overlapping swaths: In low latitudes, there is no spatial overlap between consecutive ascending or descending passes in a given data-day, therefore bins in a daily product only have pixels from a single granule. At higher latitudes, on the other hand, there may be spatial overlap between passes. In this case, the selection of data to be binned in time (e.g., bins from two consecutive passes) is performed in the same way as during the spatial binning. That is, only the "best" bins of equal quality level are summed and the equivalent pixel quality level set as a result. More than 1 pixel may be averaged in the bin when pixels of equal quality are present in the overlapping swath.

Standard mapped images (SMI): The data in the bin files are then reformatted from equal-area to equal-angle. The SMI image is a two-dimensional array of an Equidistant Cylindrical projection of the globe. The SMI format presents the data in an equi-rectangular grid, which means the spatial resolution varies with latitude. In contrast, the binned products maintain the data in an equal-area equal angle projection. These mapped images are generated from the Pathfinder main process and saved in HDF formatted files for further post process where additional ancillary data fields are added and the files are converted into netCDF format in compliance with the GHRSSST Data Specifications V2.0 (see Section 3.4.8 for additional details). For each L3 binned day or night input file, 3 map files are produced, 1 map for each product in the bin file; SST, difference from reference, and SST reference. File name: YYYYDDD.i<res><day/night>-pf<ver>ap-<SAT>-<par>.hdf.

Where : YYYYDDD = data-day, res = 4k (4 km) or qd (quarter degree), day=1 or night=3, ver=algorithm version (V5.3), sat= N[7-18], par= sst, sstd, sstref.

In this part of the NASA SeaDAS code (L3BIN), granules containing mixed day and night data are not split properly in Sea Surface Temperature – Pathfinder V5.2 and subsequently V5.3, which leads to errors in the high latitude SST values (see Section 7.4) and requires further investigation and cross checking.

3.2.5 Post Processing with GHRSSST Standards

The NCEI Sea Surface Temperature – Pathfinder CDR post processing includes three modules, which together convert the output of the Pathfinder v5.3 main processes (L3C and L2P HDF files) to GHRSSST Data Specification (GDS) version 2.0 compliant netCDF files and generate png browse image files (see Figure 1).

- hdf2nc_PFV53_L3C Module - NCEI post processing for converting L3 HDF to GHRSSST L3C netCDF. hdf2nc_PFV53_L3U Module - NCEI post processing for converting L2P HDF to GHRSSST L3U netCDF.
- hdf2nc_PFV53_L2P Module - NCEI post processing for converting L2P HDF to GHRSSST L2P netCDF.

Level 2P are preprocess files, Level 3U, uncollated files, and Level 3C, collated swath files. The L3U GHRSSST netCDF files are generated in the post process. They are remapped L2P swath to the Pathfinder v5.3 L3 grids. The binning process is retrieved from the main process and run in this step. The structure of the L3U netCDF files is the same as that of L3C. The only differences are the name of the file and the amount of the available SST grid points. Input and outputs are detailed in the flow diagram (Figure 1.) Wind files are from NCEP/DOE AMIP-II Reanalysis (Reanalysis-2): uwnd.10m.gauss.yyyy.nc and uwnd.10m.gauss.yyyy.nc). Sea Ice input files are from the NOAA/NESDIS/NCEI OISST dataset on 1/4-degree grid: yyyyymmhhdd-NCDC-L4LRblend-GLOB-v01-fv02_0-AVHRR_OI.nc). Aerosol data files (two sources of Aerosol optical thickness/depth are from: NOAA CLASS AVHRRPF (before 1999): yyyy-01-01_00_00_00_yyyy-12-28_23_59_59-AVHRRPF-aotm14a--90_90_0_360-.5.nc and NOAA CLASS Aerosol Optical Thickness (100 KM) (from 1999): yyyy-01-01_00_00_00_yyyy-12-31_23_59_59-AERO100-opt_thick--70_70_0_360-.5.n. The conversion process (GHRSSSTification) of the Level 3C files includes adding new variables: ancillary fields required by GDS 2.0, e.g., global attributes and attributes of the included variables, as well as resetting the attributes to make them GDS 2.0 compliant.

3.3 Algorithm Input

3.3.1 Primary Sensor Data

The primary sensor data used in the Sea Surface Temperature – Pathfinder CDR is the twice per day (~0230 and 1430 local solar time, except for NOAA-17 for which the nominal local solar time is 1000) reduced resolution GAC AVHRR L1b dataset. These files are made

available through NOAA's comprehensive Large Array –Data Stewardship System (CLASS). Files can be obtained via FTP (see: <http://www.class.ncdc.noaa.gov>). NOAA Level 1b Data are provided as raw digital counts that have been quality controlled, assembled into discrete data sets to which Earth location and calibration information have been appended (but not applied). The visible data values may be converted into albedos and the IR data into radiances or temperatures using the appended calibration information present in the file header. Geospatial information includes latitudes and longitudes of 51 benchmark data points along each scan. Other parameters appended in the header are: time codes, quality indicators, solar zenith angles, and telemetry. Time periods are arbitrary subsets of orbits, and may cross orbits (i.e., may contain data along a portion of an orbital track that includes the ascending node, the reference point for counting orbits). Typically consecutive files have 3 to 5 minutes of data overlap between files and each file is approximately 45mb in size. There are two formats versions of L1b data with multiple sub-versions used in the PFSST CDR, v2 prior to Aug. 28 2005 and v3 after this date. Complete details on the file structures and content can be found in the NOAA KLM users guide section 8.3.1.4 GAC Data Sets. <http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/klm/html/c8/sec831-4.htm>.

3.3.2 Ancillary Data

3.3.2.1 First Guess SST and SST Reference field

The first-guess and reference SST field used in processing is a remapped and augmented version of the AVHRR-only NOAA Optimum Interpolation $\frac{1}{4}$ -Degree Daily SST Analysis version 2 (DOISST) product. To create these augmented files the DOISST product is interpolated from the $\frac{1}{4}$ -Degree Daily OI grid to the 4km Pathfinder Version 5.3 grid and values for inland lakes and rivers are filled using the Pathfinder V5.0 monthly harmonic climatology.

Briefly, the original DOISST analysis is read from the NCEI archive (http://data.nodc.noaa.gov/thredds/catalog/ghrsst/L4/GLOB/NCDC/AVHRR_OI/catalog.html). Any pixel that is missing an SST value in the DOISST but has an SST value in the gap-free daily climatology is assigned the value from the daily climatology. This new gap-filled DOISST field is written out as a netCDF file and posted online to <http://data.nodc.noaa.gov/pathfinder/UserRequests/RSMAS/> ([/nodc/www/data.nodc.noaa.gov/htdocs/pathfinder/UserRequests/RSMAS/](http://nodc/www/data.nodc.noaa.gov/htdocs/pathfinder/UserRequests/RSMAS/) on the internal network and archived at <http://accession.nodc.noaa.gov/0071180>) to be downloaded by the Miami team and ingested by SeaDAS. The file naming convention is the same as the standard DOISST product, "avhrr-only-v2.YYYYMMDD.nc.gz". Each daily file contains about 930KB of data. Additional information and documentation for the original DOISST v2 product can be found at <http://www.ncdc.noaa.gov/oa/climate/research/sst/oi-daily.php>

3.3.2.2 Land mask

The land mask used in processing was based on three datasets: The Physical shoreline database from Natural Earth (NE) was used to obtain a high resolution Antarctic shoreline,

the Global Self-consistent Hierarchical High resolution Shorelines (GSHHS) from NGDC for shoreline other than the Antarctic, and the Global Lakes and Wetlands Database (GLWD) from WWF to obtain in-land waters. A polygon mask layer from each of these datasets were merged and converted into grid format using ArcGIS “polygon to raster” tool. The tool converts polygon into a gridded image based on the maximum area method. The criterion for the maximum area was set as “> 50%”. The resulting final land mask has four classes: Land, Ocean, Inland lake and reservoir, and Inland river. The land mask can be obtained from NCEI <ftp://ftp.nodc.noaa.gov/pub/outgoing/Li/OceanMask/>

3.3.2.3 NCEP Reanalysis files

Daily vector wind data is obtained from the NCEP Reanalysis Project. These wind speeds were created by NCEP-DOE Atmospheric Model Inter-comparison Project (AMIP-II) reanalysis (R-2) [Kanamitsu et. al. 2002], and represent winds at 10 meters above the sea surface. Separate files for u and v vectors are available by anonymous ftp (<ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis2.dailyavgs/pressure/>). Data are stored in netCDF format. File naming convention is vwind.YYYY.nc and uwind.YYYY.nc. Each file contains packed daily data for the entire year with a maximum size of 124MB. Additional information relating to these files can be found at <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html>.

3.3.2.4 Sea Ice Fraction

Two sources of sea ice information are used based on availability. Generally sea ice concentration data are taken from the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSISAF) Global Daily Sea Ice Concentration Reprocessing Data Set when available. The data are re-projected and interpolated from their original polar stereographic projection at 10km spatial resolution to the 4km Pathfinder Version 5.3 grid. When the OSISAF data are not available for both hemispheres on a given day, the sea ice concentration data are taken from the sea_ice_fraction variable found in the DOISST (described in 3.3.2.1) product, and are interpolated from the 1/4 Degree DOISST grid to the 4km Pathfinder Version 5.3 grid. Briefly, the daily mean sea ice climatology is generated from the 10km OSISAF Global Daily Sea Ice Concentration Reprocessing Data (See <http://accession.nodc.noaa.gov/0068294>), and binned to 25km spatial resolution. If any single 10km pixel in the bin is water, the entire 25km pixel is classified as water. The 25km ice climatology field is filled with values of 100% sea ice cover over the permanent ice shelves in Antarctica. These permanent ice shelves are defined by the difference between the 25km OSISAF land mask and the 25km Sea Surface Temperature – Pathfinder CDR v5.3 land mask below -60 degrees latitude. File naming convention: ice_conc_<hemisphere>_polstere-100_reproc_YYYYMMDDHHmm.nc for the OSISAF files or if not available then YYYYMMDD-NCDC-L4LRblend-GLOB-v01-fv02_0-AVHRR_OI.nc.

3.3.2.5 Aerosol Optical Depth

Aerosol optical depth (AOD) data are included as ancillary information in the output file but are not required by the SST algorithm. Aerosol optical depth was obtained from two sources: Monthly averaged Aerosol optical depth from the AVHRR Pathfinder Atmospheres

product, AVHRRPF obtained from CLASS, for the period of Jul 1981 to Dec 2000, and weekly averaged Aerosol Optical Thickness (100 KM) (AERO100) for the period of Nov 1998 to present. Data were re-projected and interpolated from their original resolution to the 4km Pathfinder Version 5.3 grid.

File name convention: aerosol_optical_thickness.aero100.interp.YYYYwWW.dat or aerosol_optical_thickness.avhrrpf.interp.YYYYMM.dat where WW=week, M=month.

3.3.3 Derived Data

Not applicable

3.3.4 Forward Models

Not applicable

3.4 Theoretical Description

The subsections below cover the evolution of the heritage algorithms that are the foundation of the Sea Surface Temperature – Pathfinder CDR are presented section 3.4.1. Then PFSST algorithm coefficient generation and the AVHRR matchup database are presented in section 3.4.2. Section 3.4.3 explains assumptions in regard to inter-calibration and ground truth referencing across multiple sensors to create a seamless Sea Surface Temperature – Pathfinder CDR. Sections 3.4.6 describes the generation and format of look-up tables required for IR calibration, clock corrections for increase navigation accuracy Section 3.4.7 describes the method used to parameterize the shape of the algorithm and determined coefficients.

3.4.1 Physical and Mathematical Description

Since 1981, the NOAA series of polar-orbiting spacecraft has carried the second and third generation AVHRR radiometers, instruments with three infrared (IR) channels suitable for estimating SST. These channels are located between 3.5 and 3.9 um and between 10 and 12.5 um, where the atmosphere is relatively transparent. At IR wavelengths the ocean surface emits radiation almost as a blackbody. In principle, without an absorbing and emitting atmosphere between the sea surface and the satellite, it would be possible to estimate SST using a single-channel measurement. In reality, surface-leaving IR radiance is partially attenuated by the atmosphere before it reaches a satellite sensor. Therefore it is necessary to make corrections for atmospheric effects. Water vapor, CH₄, NO₂, and aerosols are the major constituents that determine the atmospheric extinction of IR radiance [Minnett,1990]. Among these constituents, absorption due to water vapor accounts for most of the needed correction [Barton and Chet, 1989].

3.4.1.1 Evolution of SST Algorithms

Several techniques have been proposed over the years to account for the atmospheric absorption of surface IR radiance to improve the accuracy of satellite retrievals of SST. Anding and Kauth [1970] found that the difference in measurement at two properly selected IR channels is proportional to the amount of atmospheric correction required. Using differences in brightness temperatures (BT) measured by an early satellite radiometer, Prabhakara et. al. [1974] estimated SST to reasonable accuracy. Barton [1995] showed that this differential absorption between channels is exploited in all IR SST algorithms and demonstrated that there is a basic form to most algorithms:

$$SST = aT_i + \gamma(T_i - T_j) + c \quad (6)$$

Where T_i and T_j are brightness temperature measurements in channels i and j and a and c are constants. The γ term is defined as

$$\gamma = (1 - \tau_i) / (\tau_i - \tau_j) \quad (7)$$

Where τ is the transmittance through the atmosphere from the surface to the satellite. All AVHRR algorithms share this generalized form, although various modifications have been introduced through the years to improve performance.

The multi-channel SST algorithm (MCSST), NOAA's first operational algorithm, was based on linear differences in brightness temperatures between channels and assumed a constant γ [McClain *et al.*, 1985]. Cornillion *et al.* [1987] incorporated a correction for increased path-length at larger satellite zenith angles. Other improvements in the atmospheric correction involved non-linear formulations where γ is proportional to the brightness temperatures [Walton 1988]. The current operational NOAA algorithm is the nonlinear SST (NLSST) in which γ is assumed to be proportional to a surface SST value that can be obtained in various ways.

The Sea Surface Temperature - Pathfinder CDR algorithm, described by Kilpatrick et al. [2001], is based on the non-linear SST algorithm (NLSST) originally developed by Walton et al. [1988a].

The Sea Surface Temperature - Pathfinder CDR algorithm has the following form:

$$SST_{sat} = a + b T_4 + c (T_4 - T_5) T_{sfc} + d (T_4 - T_5) (\sec(\theta) - 1) \quad (8)$$

Where SST_{sat} is the satellite-derived SST estimate, T_4 and T_5 are brightness temperatures in AVHRR channels 4 (10.8 μm) and 5 (11.4 μm) respectively, SST_{sfc} is a first-guess SST value, and θ is the satellite zenith angle. Coefficients a , b , c , and d can be estimated by regression analysis by either a radiative transfer model or a statistical empirical approach. The PFSST algorithm differs from the NLSST in that coefficients are estimated separately for high and low water vapor regimes and temporally for each month of year.

3.4.1.2 Algorithm coefficients

Ideally a first principal approach to coefficient estimation would be to use a radiative transfer model and a set of vertical atmospheric profiles of temperature and humidity, to simulate at-satellite BTs. However, the use of a radiative transfer model requires that the satellite sensor(s) are very well characterized and that the set of radiosonde observations are properly representative of all atmospheric and oceanic conditions. Typically neither of these core assumptions is consistently met over the 30+ years of multiple sensors aboard the POES satellites. Furthermore, radiosondes measure only atmospheric temperature and water vapor but other factors are known to impact the radiative transfer process including air-sea temperature differences, other gases, surface roughness, and the presence of a variety of aerosols including dust. Due to these limitations rather than using radiative transfer the Sea Surface Temperature - Pathfinder CDR CDR uses an empirical approach to coefficient determination.

The use of an empirical approach allows for errors in both the measured BT and measured surface temperature under a variety of conditions and does not require absolute sensor characterizations. Coefficients are estimated from weighted least squares regression analyses using collocated and coincident *in situ* and satellite measurements (or “matchups”). To improve the accuracy of retrievals between wet and dry atmospheres, coefficient sets are estimated separately for two atmospheric water vapor regimes defined by the difference between AVHRR channel 4 and 5 brightness temperatures (T45). To reduce the influence from potential atmospheric seasonal or episodic events (eg. dust or ash from volcanic eruptions) that may lead to increase bias errors, the PFSST coefficients are estimated on a monthly basis for each year for each T45 water vapor regime.

The statistical approach used in the PF algorithm produces a skin temperature bias by the median skin-bulk temperature difference. IR radiometers including AVHRR measure the skin temperature, the top few microns of the ocean, in contrast to the *in situ* buoy observations that are typically located 10cm to 3m below the surface. The difference between skin and bulk temperature algorithms, as discussed by *Scheussel et. al.* [1990] and *Kearns et. al.* [2000], report median values ranging from 0.1° to 0.3° with variations of -0.5° to 1.8° depending on the wind speed and environmental conditions. *Kearns et. al.* [2000] determined that the median skin to bulk bias in the previous PFSST V4 was $0.14^{\circ} \pm 0.31^{\circ}\text{C}$ based on independent comparisons of coincident observations between PFSST and *in situ* skin measurements with the Marine Atmospheric and Emitted Radiance Interferometer (MAERI). Starting with PFSST V5.3, an average value of 0.17K for the skin/bulk difference determined from MAERI matchups was used to remove the bulk bias so that the PFSST more closely represents a skin temperature. This also aligns the PFSST CDR with other more modern sensors reporting SST physical units as skin temperature. For more discussion on skin, bulk, and foundation temperatures also see GHRSSST science working group reports (<https://www.ghrsst.org/ghrsst-science/sst-definitions/>).

3.4.2 Data Merging Strategy

No data merging is done between AVHRR sensors. Each sensor is individually referenced to *in situ* observations and algorithm coefficients are adjusted to ensure as close agreement as

possible for the available cal/val data set at the time. Generally sensors overlap for a brief period of time and the accuracy of the SST retrieval is compared and documented so as to create a seamless record of observations. The accuracy of the complete time series is continually tracked by comparison to other available satellite and analysis products. As knowledge of the performance improves and upgrades to the processing scheme mature, the dataset should be reprocessed to remove known solvable issues.

3.4.3 Numerical Strategy

See section 3.4.1.1 Evolution of SST algorithms.

3.4.4 Calculations

See section 3.4.1.1 Evolution of SST algorithms

3.4.5 Look-Up Table Description

3.4.5.1 IR calibration tables

In principle, the radiance sensed by an AVHRR IR channel is a function of the temperature of the target (Earth) and the spectral response of the channel. Laboratory calibration of channels 4 and 5 indicated that the response of the detectors is nonlinear. This non-linearity depends both on the AVHRR operating temperature and on the scene radiance as a function of output counts. Brown et. al. [1985, 1993] performed a detailed analysis of AVHRR thermal vacuum test data and found calibration results for various AVHRR radiometers show instrument specific changes in the relative emittance between internal and external calibration targets. Furthermore, several of the AVHRR instruments were found to have been routinely operated outside the range of the thermal-vacuum test results. As a result non-linear corrections could not be interpolated directly from earlier methods. A consistent non-linear calibration methodology was developed based on a second-order polynomial regression with a total calibration accuracy relative to an external calibration standard of less than two digital counts ($\pm 0.2^{\circ}\text{C}$). The procedure to convert AVHRR counts to radiance for the infrared channels is therefore a three-step process. The first step is a linear transformation of counts to radiance, based on the onboard blackbody calibration devices (space view and sensor baseplate). The second step applies a non-linearity correction factor (derived from pre-launch calibration data) to correct radiance values for channels 4 and 5 only. The third and final step involves the use of lookup tables to convert radiance to temperature as a function of the sensor's operating temperature.

3.4.5.2 Clock corrections

The major sources of error in geo-locating AVHRR data are (a) drift in the spacecraft clock (which causes errors in the estimated along-track position), and (b) uncertainty errors in spacecraft and sensor attitude. To minimize error in the along track position estimated by the orbital model, a satellite clock correction factor is applied to the time code imbedded in each scan line. The method used to determine these clock correction factors was as follows. The clock aboard a given satellite drifts continually at a relatively constant rate (e.g., for

NOAA-14, ~9msday-1) compared to the reference clock on Earth. Because of this drift, the NOAA/NESDIS Satellite Operation Control Center prior to 2002 periodically sent a command to the satellite to reset the on-board clock to a new baseline, thereby eliminating the accumulation of a large time offset error between the Earth and satellite clocks. To correct for clock drift between these resets, correction factors were determined from a database of satellite clock time and Earth time offsets collected at the RSMAS DOMSAT receiving station using HRPT observations. During HRPT transmission, both the satellite clock (used to create the embedded time code in each piece) and the Earth clock are simultaneously available. The clock correction bias was determined by (1) visual examination of the Earth/satellite clock differences collected in the database to locate the precise magnitude and timing of clock resets performed by the Satellite Operation Control Center and (2) recorded time differences between the identified reset periods were then filtered to remove spurious noise, and regressed against the corresponding satellite time to determine the clock drift correction. These drift corrections were then applied to all data time-stamped during a given reset period.

File format: ASCII file 6 columns, tab separated values.

Line 1 Header: SAT Start/date Start/time offset End/Date End/time offset

Line 2-last Format: YYYYDDD hhmmss msec YYYYDDD hhmmss msec

After 2002 the clock was corrected by software on-board the satellite so that additional corrections are not required by the PFSST software.

3.4.6 Parameterization

In the current Pathfinder protocol, the shape of the algorithm is parameterized in a three-stage process of coefficient estimation. In the first stage, those matchups passing the cloud and quality control tests are used to estimate a first-guess set of coefficients using a robust MM regression method, in which coefficient estimates are relatively insensitive to large outliers. The statistical theory for this method is based on Rousseeuw and Yohai [1984], Yohai, Stahel and Zamara [1991], and Yohai and Zamar [1998]. The procedure used was implemented in the commercial software package S-Plus as function `lmRobMM()`. The resultant first-guess coefficients are used to compute first-guess SST residuals.

In the second stage, the first-guess residuals are used to compute robustness weights (r), to decrease the influence of matchups with large residuals in the final coefficient estimation. The formula to calculate the robustness weights is:

$$r = B(u) [e / (6 * MAD)] \quad (9)$$

where: MAD is the median absolute difference of the first guess residuals for each period and T45 regime, e is the first-guess residuals for a corresponding MAD, $B(u)$ is a bisquare function where u denotes the function's argument, has a value of $(1 - u^2)^2$ for $|u| < 1$, otherwise it is zero.

The formula produces robustness weights having a value of zero for matchups with first-guess residuals greater than $\pm 6 * \text{MAD}$. The factor of 6 multiplying the MAD was selected so that, if the first-guess residuals have an underlying Gaussian distribution, this threshold is approximately equivalent to rejecting first-guess residuals beyond ± 4 standard deviations. In most cases, MAD values ranged between 0.3° and 0.4°K : this implies that residuals with absolute values greater than 1.8° to 2.4°K were excluded (i.e., had weights equal to zero).

In the third and last stage, both robustness weights and temporal weights are used in a weighted least squares regression. The temporal weights within the 5 month window for each period were for the central month (e.g., month N) assigned a weight of 1.0, for adjacent months (N - 1 and N + 1) the weights were 0.8, and weights of 0.5 were used for the ends of the 5-month window (months N - 2 and N + 2). This third stage of coefficient estimation produces the operational CDR set of coefficients. This procedure is repeated for each period for which coefficients are estimated (usually, a month), and for each T45 regime. Additional details on this protocol can be found in Kilpatrick et. al. 2001.

3.4.7 Algorithm Output

3.4.7.1 Parameters names units and scaling

There are 13 mapped parameters produced as output. Each of the files contains a parameter mapped to a grid with 8640 elements in longitude and 4320 in latitude, plus a vector of length 8640 identifying the longitudes and a vector with 4320 values indicating the latitudes. The parameters are stored either as 16-bit or 32-bit unsigned integers that may be converted linearly ($y = mx + b$) to geophysical units using a scale (i.e., slope= m) and offset (i.e., intercept= b) (see Table 3).

Table 3: Output files Parameter scaling and physical units

Parameter	Data Type	Description
Time [0..0]	integer	units: "seconds since 1981-01-01 00:00:00" axis: "T" note : "This is the reference time of the SST file. Add sst_dtime to this value to get pixel-by-pixel times. Note: in PFV5.3 that sst_dtime is empty. PFV6 will contain the correct sst_dtime values." calendar: "Gregorian"
Lat [0..4319]	integer	_Netcdf4Dimid: 1 standard_name: "latitude" units: "degrees_north" reference_datum: "geographical coordinates, WGS84 projection" grid_mapping: "Equidistant Cylindrical" axis: "Y"
Lon [0..8639]	integer	_Netcdf4Dimid: 2 standard_name: "longitude" units: "degrees_east" reference_datum: "geographical coordinates, WGS84 projection" grid_mapping: "Equidistant Cylindrical" axis: "X"
Sea_surface_temperature: Grid	integer	standard_name: "sea_surface_skin_temperature" grid_mapping: "Equidistant Cylindrical" units: "kelvin" add_offset: 273.15 scale_factor: 0.01 valid_min: -180** valid_max: 4500 _FillValue: -32768 ** note this is the valid min for the geophysical SST parameter not the absolute possible min.

sst_dtime: Grid	integer	<p>long_name: "time difference from reference time" grid_mapping: "Equidistant Cylindrical" units: "second" add_offset: 0 scale_factor: 1 valid_min: -2147483647 valid_max: 2147483647 _FillValue: -2147483648 comment: "time plus sst_dtime gives seconds after 1981-01-01 00:00:00. Note: in PFV5.3 this sst_dtime is empty. PFV6 will contain the correct sst_dtime values."</p>
sses_bias: Grid (empty SDS in PFSST V5.3)	integer	<p>long_name: "SSES bias estimate" grid_mapping: "Equidistant Cylindrical" units: "kelvin" add_offset: 0.0 scale_factor: 0.02 valid_min: -127 valid_max: 127 _FillValue: -128 Note: "Bias estimate derived using the techniques described at http://www.ghrsst.org/SSES-Description-of-schemes.html. Note: in PFV5.3 this sses_bias is empty. PFV6 will contain the correct sses_bias values."</p>

sses_standard_deviation: Grid (empty SDS in PFSST V5.3)	integer	long_name: "SSES standard deviation" grid_mapping: "Equidistant Cylindrical" units: "kelvin" add_offset: 2.54 scale_factor: 0.02 valid_min: -127 valid_max: 127 _FillValue: -128 note: "Standard deviation estimate derived using the techniques described at http://www.ghrsst.org/SSES-Description-of-schemes.html . Note: in PFV5.3 this sses_standard_deviation is empty. PFV6 will contain the correct sses_standard_deviation values.
--	---------	---

dt_analysis: Grid	integer	<p>long_name: "Deviation from last SST analysis"</p> <p>grid_mapping: "Equidistant Cylindrical"</p> <p>units: "kelvin"</p> <p>add_offset: 0.0</p> <p>scale_factor: 0.1</p> <p>valid_min: -127</p> <p>valid_max: 127</p> <p>_FillValue: -128</p> <p>note: "The difference between this SST and the previous day's SST."</p>
wind_speed: Grid	integer	<p>long_name: "10m wind speed"</p> <p>standard_name: "wind_speed"</p> <p>grid_mapping: "Equidistant Cylindrical"</p> <p>units: "m s-1"</p> <p>height: "10 m"</p> <p>add_offset: 0.0</p> <p>scale_factor: 1.0</p> <p>valid_min: -127</p> <p>valid_max: 127</p> <p>_FillValue: -128</p> <p>time_offset: 6.0</p>
Sea_ice_fraction: Grid	integer	<p>long_name: "sea ice fraction"</p> <p>standard_name: "sea_ice_area_fraction"</p> <p>grid_mapping: "Equidistant Cylindrical"</p> <p>units: ""</p> <p>add_offset: 0.0</p> <p>scale_factor: 0.01</p> <p>valid_min: -127</p> <p>valid_max: 127</p> <p>_FillValue: -128</p> <p>time_offset: 12.0</p>

aerosol_dynamic_indicator: Grid	integer	long_name: "aerosol dynamic indicator" (ADI) grid_mapping: "Equidistant Cylindrical" units: "" add_offset: 2.4 scale_factor: 0.02047244 valid_min: -127 valid_max: 127 _FillValue: -128
adi_dtime_from_sst Grid (empty SDS in PFSST V5.3)	integer	long_name: "time difference of ADI data from sst measurement" grid_mapping: "Equidistant Cylindrical" units: "hour" add_offset: 0.0 scale_factor: 1.0 valid_min: -127 valid_max: 127 _FillValue: -128

<p>quality_level: Grid</p>	<p>byte</p>	<p>long_name: "SST measurement quality" grid_mapping: "Equidistant Cylindrical" valid_min: 1 valid_max: 5 _FillValue: 0 flag_meanings: "bad_data worst_quality low_quality acceptable_quality best_quality" flag_values: 1, 2, 3, 4, 5 comment: "Note, the native Pathfinder processing system returns quality levels ranging from 0 to 7 (7 is best quality; -1 represents missing data) and has been converted to the extent possible into the six levels required by the GDS2 (ranging from 0 to 5, where 5 is best). Below is the conversion table: GDS2 required quality_level 5 = native Pathfinder quality level 7 == best_quality GDS2 required quality_level 4 = native Pathfinder quality level 4-6 == acceptable_quality GDS2 required quality_level 3 = native Pathfinder quality level 2-3 == low_quality GDS2 required quality_level 2 = native Pathfinder quality level 1 == worst_quality GDS2 required quality_level 1 = native Pathfinder quality level 0 = bad_data GDS2 required quality_level 0 = native Pathfinder quality level -1 = missing_data The original Pathfinder quality level is recorded in the optional variable pathfinder_quality_level."</p>
-----------------------------------	-------------	--

<p>pathfinder_quality_level: Grid</p>	<p>byte</p>	<p>long_name: "Pathfinder SST quality flag" grid_mapping: "Equidistant Cylindrical" valid_min: 0 valid_max: 7 _FillValue: -1 flag_meanings: "bad_data worst_quality low_quality low_quality acceptable_quality acceptable_quality best_quality" flag_values: 0, 1, 2, 3, 4, 5, 6, 7 comment: "This variable contains the native Pathfinder processing system quality levels, ranging from 0 to 7, where 0 is worst and 7 is best. And value -1 represents missing data."</p>
<p>l2p_flags: Grid</p>	<p>byte</p>	<p>long_name: "L2P flags" grid_mapping: "Equidistant Cylindrical" flag_meanings: "microwave land ice lake river" flag_masks: 1, 2, 4, 8, 16, 32, 64, 128, 256 comment: "Bit zero (0) is always set to zero to indicate infrared data. Bit one (1) is set to zero for any pixel over ocean. Any 4 km Pathfinder pixel whose area is 50% or more covered by land has bit one (1) set to 1. Bit two (2) is set to 1 when the sea_ice_fraction is 0.15 or greater. Bits three (3) and four (4) indicate lake and river pixels, Any 4 km Pathfinder pixel whose area is 50% or more covered by lake has bit three (3) set to 1. Any 4 km Pathfinder pixel whose area is 50% or more covered by river has bit four (4) set to 1"</p>

3.4.7.2 File format

The data are stored in the network common data format (netCDF). Sea Surface Temperature – Pathfinder CDR V5.3 files are compliant with the GHRSSST Data Specifications V2.0 (GDS2 <http://ghrsst-pp.metoffice.com/pages/RGTS/GDS/index.htm>) with the exception of the following variables: The file does not encode time according to the GDS2 specifications, and the sses_bias and sses_standard_deviation variables are currently empty. Full compliance with GDS2 specifications will be achieved in the future Pathfinder Version 6.

3.4.7.3 File naming convention

File naming follows the International Global High Resolution SST program (GHRSSST) GDS 2.0 standard.

<date><time><RDAC><processing level>_GHSST-<SST Type>-<character string>-<additional segregator>-v<GDS Version>-fv<File Versions>.<file type>

The variable components within braces (“<>”) are summarized in Table 4 below.

Table 4: Archived CDR File naming components

Name	Definition	Description
<date>	YYYYMMDD	Center date of data collation window
<time>	HHMMSS	Center time of data collation window GMT
<RDAC>	RDAC (Regional Data Assembly Center) where data are produced	NCEI RDAC code for PFSST products.
<processing level>	Data processing level code	L3C Code for collated observations combined from a single instrument into a space-time grid
<SST type>	A character string identifying the SST product suite	GHRSSST-SSTskin
< character string>	Additional string identifying SST product set. This string is unique within an RDAC	AVHRR Pathfinder-PFV<X.X> Where X.X is the PFSST version
<additional segregator>	Additional string identifier	NOAA-<x>_G_<date>_<xx> Where x is NOAA satellite (7,9,11,14,16,18,19,..) xx == “day” or “night” date== dataday for the file
<GDS Version>	nn.n	Version number of the GDS used to process the file for example

		GDS2.0="02.0"
<File Versions>	xx.x	01.0
<file type>	Netcdf code	nc

Example name:

20141220141545-NCEI-L3C_GHRSSST-SSTskin-AVHRR_Pathfinder-PFV5.3_NOAA19_G_2014354_day-v02.0-fv01.0.nc

20141220014905-NCEI-L3C_GHRSSST-SSTskin-AVHRR_Pathfinder-PFV5.3_NOAA19_G_2014354_night-v02.0-fv01.0.nc

3.4.7.4 Files size

File sizes of the archived and distributed GHRSSST formatted Sea Surface Temperature – Pathfinder CDR files: netCDF: ~27 GiB/year; GHRSSST picture files PNG: ~1.5 GiB/year.

4. Test Datasets and Outputs

4.1 Test Input Datasets

The test input dataset is the same CLASS L1b GAC used during processing a complete description of this dataset is presented in prior section 3.3.1 Primary sensor data.

4.2 Test Output Analysis

4.2.1 Reproducibility

Reproducibility is monitored and ensured by three methods. Extraction of validation matchups during reprocessing, comparison to same day reference DOISST files and other SST climatologies and sensors, and finally trend analysis of the PFSST product with time. For each day, a vector of zonal averages (AVHRR SST minus reference SST) is computed. When the time series for a specific sensor is complete, these vectors are assembled to construct a Hovmöller diagram to ensure consistency of the CDR. At the end of a given sensor's processing run, a Hovmoller diagram is produced that allows trends in latitude and time to be analyzed.

4.2.2 Precision and Accuracy

On a global average the PFSST is reported to have an accuracy of -0.2K and a precision of +/- 0.5K relative to *in situ* buoys [Kilpatrick et. al. 1991]. The precision and accuracy is determined from statistical analysis of the Pathfinder matchup database (PFMDB). The PFMDB includes contemporaneous measurements of co-located satellite brightness temperature, satellite retrieved SST, in-situ observations and radiometer SST, environmental 'observations' from analyzed model or satellite observed fields, and satellite viewing geometry, time and location. The 130 fields contained within the PFMDB are used to explore the error statistics through recursive partitioning, facilitating an understanding of algorithm precision and accuracy as a function of water vapor, latitudinal bands, seasons, aerosol, clouds, sensor characteristics, viewing geometry, mirrors, detectors, and calibrations leading to a more complete understanding of accuracy at the regional level.

4.2.3 Error Budget

The residual analysis from the PFMDB provides complete end-to-end system performance; sensor, atmosphere, and algorithm precision and accuracy. The validated performance is presented in Tables 5-11 for each sensor as a function of year for night data only. Using night only data minimizes the potential difference between the in situ buoy observation and potential diurnal heating to the skin layer measured by the sensor. For completeness, daytime validation statistics are also included (Tables 5B, 6B, ..11B). Only best quality level in situ data from the iQuam2 database (quality flag=0, <http://www.star.nesdis.noaa.gov/sod/sst/iquam/v2/data.html>) are used for Pathfinder

SSTs of good quality flag confidently clear, pixels in the analysis. The bias (satellite – in situ observations) in terms of the mean and standard deviation provide an estimate of the expected errors.

Table 5.A: Night time validation statistics NOAA- 7

Median	Standard Deviation	Count	Year	SAT
-0.399	0.228	497	1981	NOAA-7
-0.121	0.180	548	1982	
-0.279	0.193	620	1983	
-0.270	0.217	1030	1984	
-0.254	0.363	309	1985	

Table 5.B: Day time validation statistics NOAA- 7

Median	Standard Deviation	Count	Year	SAT
-0.400	0.346	842	1981	NOAA-7
-0.251	0.348	1516	1982	
-0.357	0.305	1489	1983	
-0.108	0.354	1882	1984	
0.028	0.589	528	1985	

Table 6.A Night time validation statistics NOAA- 9

Median	Standard deviation	Count N	Year	SAT
-0.338	0.351	3932	1985	NOAA-9
-0.328	0.469	13027	1986	
-0.386	0.514	9814	1987	
-0.409	0.514	9114	1988 (1-312)	
-0.325	0.356	5559	1994 (275-365)	
-0.267	0.413	784	1995 (1-17)	

Table 6.B: Day time validation statistics NOAA- 9

Median	Standard deviation	Count N	Year	SAT
-0.179	0.493	5828	1985	NOAA-9
-0.224	0.545	18116	1986	
-0.241	0.575	14282	1987	
-0.313	0.612	13132	1988 (1-312)	
-0.194	0.367	4311	1994 (275-365)	
-0.081	0.359	601	1995 (1-17)	

Table 7.A: Night time validation statistics NOAA- 11

Median	Standard deviation	Count N	Year	SAT
-0.444	0.477	1542	1988 (313-365)	NOAA-11
-0.399	0.477	11960	1989	
-0.432	0.474	14386	1990	
-0.382	0.544	13127	1991	
-0.326	0.515	18585	1992	
-0.367	0.494	27201	1993	
-0.372	0.467	18736	1994 (1-274)	

Table 7.B: Day time validation statistics NOAA- 11

Median	Standard deviation	Count N	Year	SAT
-0.264	0.541	2097	1988 (313-365)	NOAA-11
-0.282	0.555	17718	1989	
-0.283	0.531	20393	1990	
-0.228	0.617	19256	1991	
-0.137	0.575	25509	1992	
-0.117	0.552	41070	1993	
-0.125	0.536	32141	1994 (1-274)	

Table 8.A: Night time validation statistics NOAA- 14

Median	Standard deviation	Count N	Year	SAT
-0.283	0.468	29075	1995 (18-365)	NOAA-14
-0.286	0.468	48106	1996	
-0.282	0.462	41894	1997	
-0.272	0.446	46814	1998	
-0.313	0.495	68084	1999	
-0.277	0.505	79217	2000 (1-300)	
-0.301	0.470	13418	2001 (1-55)	

Table 8.B: Day time validation statistics NOAA- 14

Median	Standard deviation	Count N	Year	SAT
-0.227	0.516	39472	1995 (18-365)	NOAA-14
-0.208	0.484	64438	1996	
-0.205	0.482	57226	1997	
-0.209	0.479	71180	1998	
-0.199	0.496	99268	1999	
-0.164	0.514	115300	2000 (1-300)	
-0.300	0.503	19240	2001 (1-55)	

Table 9.A: Night time validation statistics NOAA-16

Median	Standard Deviation	Count N	Year	SAT
-0.278	0.462	16555	2000 (301-366)	NOAA-16
-0.301	0.454	82890	2001 (56-365)	
-0.287	0.488	58922	2002 (1-190)	

Table 9.B: Day time validation statistics NOAA-16

Median	Standard Deviation	Count N	Year	SAT
-0.169	0.479	20303	2000 (301-366)	NOAA-16
-0.161	0.445	103949	2001 (56-365)	
-0.116	0.483	72074	2002 (1-190)	

Table 10.A: Night time validation statistics NOAA- 17

Median	Standard Deviation	Count N	Year	SAT
-0.218	0.435	66826	2002 (191-365)	NOAA-17
-0.200	0.418	161460	2003	
-0.228	0.424	176348	2004	
-0.204	0.417	117376	2005 (1-157)	

Table 10.B: Day time validation statistics NOAA- 17

Median	Standard Deviation	Count N	Year	SAT
-0.277	0.433	74761	2002 (191-365)	NOAA-17
-0.255	0.422	186161	2003	
-0.262	0.409	207777	2004	
-0.232	0.413	137035	2005 (1-157)	

Table 11.A: Night time validation statistics NOAA- 18

Median	Standard Deviation	Count N	Year	SAT
-0.186	0.431	186658	2005(158-365)	NOAA-18
-0.198	0.408	394485	2006	
-0.185	0.402	357422	2007	
-0.191	0.409	460083	2008	
-0.177	0.430	509495	2009	
-0.123	0.439	588485	2010	

Table 11.B: Day time validation statistics NOAA- 18

Median	Standard Deviation	Count N	Year	SAT
-0.140	0.456	264147	2005 (158-365)	NOAA-18
-0.145	0.434	545588	2006	
-0.136	0.422	508950	2007	
-0.126	0.425	645625	2008	
-0.099	0.448	726845	2009	
-0.071	0.452	847510	2010	

Table 12.A: Night time validation statistics NOAA- 19

Median	Standard Deviation	Count N	Year	SAT
-0.271	0.396	538644	2011	NOAA-19
-0.240	0.414	418497	2012	
-0.254	0.409	471076	2013	
-0.240	0.411	561327	2014	

Table 12.B: Day time validation statistics NOAA- 19

Median	Standard Deviation	Count N	Year	SAT
-0.205	0.437	767746	2011	NOAA-19
-0.164	0.451	589629	2012	
-0.198	0.454	680180	2013	
-0.169	0.451	790228	2014	

5. Practical Considerations

5.1 Numerical Computation Considerations

Each segment of a GAC orbit is processed pixel by pixel in a serial fashion. No looping is employed throughout the algorithm. The only search operations that are employed are used to select the proper month and T45 difference needed to select the proper SST retrieval coefficients. The other data preparation operation is a bi-linear interpolation of the reference SST field to the latitude and longitude of the satellite observation.

5.2 Programming and Procedural Considerations

To facilitate calculation of the SST product in the Linux processing environment, all AVHRR orbits for a 24 hour period are collected to define a processing element. This element, together with the associated ancillary and calibration data (e.g. SST retrieval coefficients and reference SST field) are then transferred to the appropriate processing computer core together with the associated processing scripts under control of the MAUI cluster control software. Parallelization is achieved by distributing order of 500 days in parallel to the available processors. The i/o to the external disk system is minimized by only needing to send the calibration and ancillary data once per processing day and retaining intermediate data files such as the Level 2 products on storage local to each processing core. Depending on the number of available processors, disk data rates easily can become the governing consideration on the achievable processing rate and thus the job definition and submission strategy must be tailored to the capabilities of the available processing environment.

5.3 Quality Assessment and Diagnostics

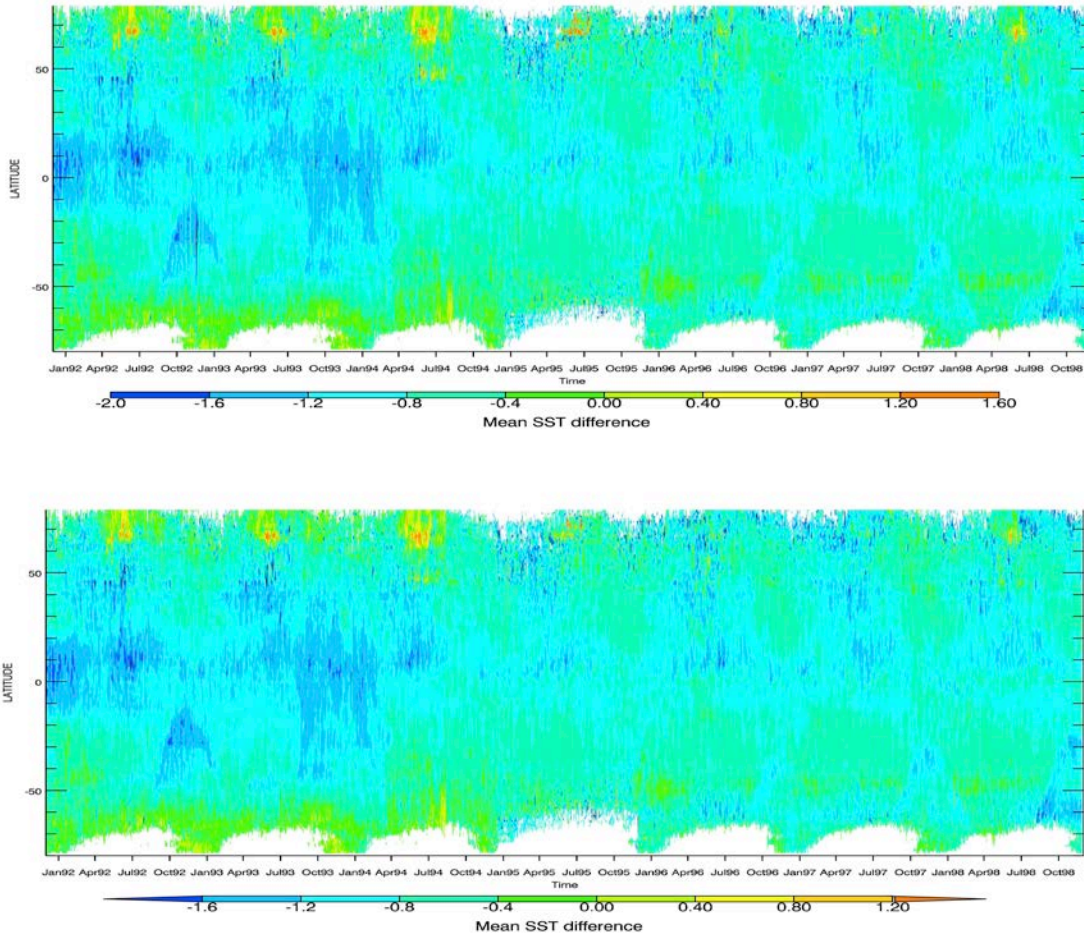
Several quality assurance steps are conducted to insure the integrity of the data. These steps are conducted routinely during the processing and transfer stages of the system. These steps are aimed primarily at assuring file integrity and availability, not at the accuracy of the derived sea surface temperature (SST) values. Validation and calibration methods are presented in a later section (section 5.5). The steps to insure file integrity are as follows:

- ✓ Generate browse images for every file in the dataset. This process verifies the file format integrity for every file by accessing the file and reading the data.
- ✓ Visually inspect every browse image for gross errors, unusual features, etc. This check reveals obvious errors or processing problems.
- ✓ Compute and verify counts of files in each directory structure to insure no files are missing. The counts results are updated hourly.
- ✓ Compute and verify the volumes of each directory to insure reasonable data volumes are present. The results are updated hourly.

- ✓ To ensure consistency as AVHRR data-days are processed, an element of a Hovmöller diagram consisting of the average of all quality level 0 pixels are averaged for each zonal vector in the mapped output file for the difference between the AVHRR satellite SST and the reference SST. At the end of a given sensor's processing run, a Hovmöller diagram is produced that allows trends in latitude and time to be analyzed.

A comparison of Pathfinder SST with daily the Canadian Meteorological Center Global Foundation SST (Figure 2) show good agreement, although differences are expected for a skin SST measure of SST (Pathfinder SST) versus a global foundation SST (Bransett, 2008). The Canadian Meteorological Center Global Foundation SST can be found at <https://data.noaa.gov/dataset/ghrsst-level-4-cmc0-2deg-global-foundation-sea-surface-temperature-analysis-gds-version-25a31f>.

Figure 2: A comparison of Pathfinder Sea Surface Temperature V5.2 (top panel) and V5.3 (bottom panel) with Canadian Meteorological Center V0.2 Global Foundation Sea Surface Temperature. The binning problem is apparent at high latitudes above 60 N as SST values are continuous across time (see section 7.4).



Exception Handling

There are two levels of exception handling during level-2 product generation, initial tests to identify grossly erroneous retrievals, and a second tier which attempt to classify and predict which pixels are likely to have sub-pixel scale problems based on insights gained from analysis of the PFMDB.

The first tier consists of four Initial tests to identify grossly suspect retrievals. If a pixel fails any of these four tests it is automatically assigned to the lowest quality level.

- a. *Brightness temperature range test:* Brightness temperatures for AVHRR channels 3, 4 and 5 must be greater than, or equal to -10°C and less than or equal to 37°C . This test identifies sensor digitizer errors or very cold pixels associated with high cloud tops.

- b. *Uniformity test 2:* This test was similar to that described above, but the threshold was set as 1.2°C. That is, differences between maximum and minimum brightness temperatures must be less than 1.2°C to pass this test. A higher uniformity threshold allows more pixels to pass the test, at the expense of accepting pixels with a higher SST bias.
- c. *Zenith angle test 1:* Satellite zenith angle must be less than 45 degrees to pass this test. At higher zenith angles, radiation emitted by the ocean has to go through a longer atmospheric path before reaching the AVHRR instrument, with consequently higher chances of being attenuated. The received radiance, therefore, is likely to have a lower proportion of radiance originating from the ocean's surface (the signal of interest) and a greater proportion of radiance re-emitted by the atmosphere. The negative side of limiting zenith angles is the loss in geographic coverage.
- d. *Stray sunlight test:* An examination of data stratified by satellite zenith angle and by side of the AVHRR scan line (left and right of nadir) revealed potential problems under certain Earth-Sun-satellite configurations. This flag identifies configurations in which problems may potentially occur. The problem is probably associated with stray solar radiation entering the radiometer and it occurs only in the middle to high latitudes in the Southern Hemisphere. For that reason, in the Northern Hemisphere this flag is always set to 0 (pass). In the Southern Hemisphere, the flag is set to failed when (a) the satellite zenith angle is greater than 45 degrees, and (b) the pixel is located on the Sun side of the AVHRR scan line. For an ascending pass (spacecraft flying from south to north), the Sun side of the scan line is located left of nadir; for a descending pass, the Sun side of the scan line is right of nadir. The latitude in the Southern Hemisphere at which stray sunlight becomes a problem is a function of season. During the austral summer, this problem may potentially reach the mid-latitudes, whereas in austral winter, it is confined to very high latitudes. For speed of processing, we have disregarded the seasonality of the latitude dependence, which may result in "good" pixels being erroneously flagged as failing this test. As this test is later used to define overall quality levels (see below), mid-latitude Southern Hemisphere pixels at high scan angles have the potential of being assigned to the lowest quality level during austral winter.
- e. *Edge test:* Pixels must not be on the first or last scan lines of a piece, or on the first or last 8 pixels in a scan line. Pixels along edges are not surrounded by pixels so that tests based on 3x3 or 5x5 boxes cannot be performed.

The second tier of finer scale tests is applied to pixel that passed the initial tests.

- f. *Reference test:* The absolute difference between the Pathfinder SST for the pixel considered and the reference DOISST field must be greater than -2°C and at night must be less than 5°C. see also aerosol test below.

- g. *SST range test*: the satellite SST must be less than 40°C during the day and less than 37°C at night.
- h. *Aerosol test*: In regions with likely aerosol contamination, Atlantic off of Africa and North Arabian Sea, we impose a more restrict reference test. In side a box defined as lat ≤ 30 and ≥ -10 and a longitude of < 105 and ≥ -105 then the difference from reference mus be ≥ 1.25 and < 5 at night.
- i. *Glint*: daytime pixels in the glint region, $glint > 0.005$, can not be assigned best quality because visible tests to identify clouds can not be performed in glint regions, these pixels can only be assigned to quality 1
- j. *Decision trees*: A family of sensor specific decision trees is used to identify cloud contaminated or suspect retrievals.
- k. *Uniformity test 1*: Maximum and minimum brightness temperature values are calculated for channels 4 and 5, for a 3x3 box centered on the pixel being classified. The difference between maximum and minimum brightness temperatures for both channels must be less than 0.7°C. This test seeks to identify contamination by small clouds, and is based on the assumption that SSTs are relatively uniform at small scales (e.g., 3x3 pixels). The 0.7°C threshold was selected by testing different threshold values in the matchup database. For uniformity thresholds below 0.7°, no significant bias was detected in SST estimates, and the rms of SST residuals was relatively uniform.
- l. *Zenith angle test 2*: Satellite zenith angle must be less than 55 degrees. This is similar to the test in bit 7 of variable mask1, but it allows a larger range of acceptable zenith angle values, with the goal of gaining geographic coverage.

Finally, the results of these second tests are combined into an overall quality level assignment for each pixel, ranging from 0-7 with 0 being the best and 7 the worst quality. Note: versions of the Sea Surface Temperature – Pathfinder CDR prior to V5.2 used a reversed assignment, that is 7 being the best and 0 the worst. The change to quality 0 for best was made beginning with V5.2 to be compliant with logic of binning routines within the SEADAS software package where 0 indicates best clear pixel.

- m. Additional Sensor specific exceptions: pixels are demoted 1 quality level if and of the following conditions are true:
- n. $(T4 - T5) \geq 0$ deg C: In a few cases, the difference between brightness temperatures in AVHRR channels 4 and 5 was negative (indicating that $T5 > T4$). This may suggest an error in the measurements or anomalous atmospheric conditions under which the algorithms are not likely to yield good SST retrievals. For that reason, matchups for which $(T4 - T5) < 0$ deg C were excluded. This test was used for NOAA-14 matchups starting in 1998.

- o. $T_4 \geq 2.5$ deg C. We discovered a problem with the NOAA-14 digitizer that caused problems in T_4 values lower than about 2.5 deg C (the value changes slightly with time). We excluded all matchups for which $T_4 < 2.5$ deg C. This test was applied to NOAA-14 matchups starting with 1999. A more detailed description of the NOAA-14 digitizer problem can be found in Podestá et al. [2005].

5.4 Algorithm Validation

One of the AVHRR Oceans Pathfinder SST program original highlights was that, for the first time, a large validation data set was distributed to accompany the Pathfinder global SST fields. The Pathfinder matchup database (PFMDB) is a multi-satellite, multi-year database of AVHRR and in situ SST matchups. The PFMDB enables both ongoing and retrospective and validation and continued evolution of the algorithm to produce a consistent CDR over a wide range of atmospheric and oceanic regimes.

For future Pathfinder versions, a bias (satellite-in situ) and standard deviation, referred to as single sensor error statistics (SSES), which have been computed for a variety of conditions will be included in the data files. These error statistics are not present in Pathfinder 5.3 data file. The validation hypercube for Sea Surface Temperature - Pathfinder CDR, includes 7 dimensions: time by season (4), latitude bands (5 steps in 20 degree from 60S to 60N), surface temperature (8 increments in 5 degree steps) satellite zenith angle (4 increments), brightness temperature difference as a proxy for water vapor (3 intervals for 11-12 mm SST), retrieved satellite SST quality level, day/night selection (2 intervals).

The validation of the algorithm and product also includes comparisons to SST products from other sensor, and climatologies. Further validation of Sea Surface Temperature - Pathfinder CDR v5.3 data will be provided. The Sea Surface Temperature - Pathfinder CDR dataset is used as the standard reference data set for the Global Climate Observing System (GCOS) SST and Sea Ice inter-comparison project. Inter-comparisons will be made available from the NOAA GHRSSST Long Term Stewardship and Reanalysis Facility (LTSRF) http://www.nodc.noaa.gov/SatelliteData/ghrsst/intercomp_data.html.

5.5 Processing Environment and Resources

Hardware resources: The processing environment supporting Sea Surface Temperature - Pathfinder CDR processing is based on Amazon Web Service (AWS) Elastic Compute Cloud (EC2) environment. The computing nodes are supported by CentOS release 6.6. A typical node configuration is 16-core (with 32 GiB memory) or 32-core (with 64 GiB memory) Intel Xeon CPU E5-2680 v2 running at 2.80GHz. Different jobs may need different optimized configurations to get maximum efficiency. Several nodes are grouped to run the job when the job scale is very large. Local disk volumes of are allocated to hold local input, output and working files. The AWS Simple Storage Service (S3) buckets are allocated to save input/output data generated on AWS. The main process of Pathfinder from reading input AVHRR GAC data to generating L3C and L2P HDF output files costs about 14 hours wall clock for one whole year on

a typical 32-core/cpu node on AWS environment. The post process to generate GDS 2.0 compliant L3C netCDF files costs about 3 hours per year on a typical 8-core/cpu node. The L2P and L3C post process to generate GDS 2.0 compliant L2P and L3U netCDF files costs about 13 hours per year on a typical 16-core/cpu node. The time costs of browse image processes for one whole year on a typical 16-core/cpu node - L3C: 24 hours; L3U: 17 hours; L2P: 15 hours. All input AVHRR GAC data, output L2P and L3C HDF files, output L2P, L3C, L3U netCDF files browse images PNG files are temporarily stored on AWS S3 buckets until processing and archive procedures at NCEI are complete. Disk space requirements are listed below –

AVHRR GAC L1B data: 280 GiB/year; totaling 9.30 Tbytes for 1981-2014.

Ancillary data: 26 GiB/year; totaling ~884 Gbytes for 1981-2014.

Pathfinder L2P data - HDF: 230 GiB/year; netCDF: 260 GiB/year; PNG: 6.2 GiB/year; totaling ~16.50 Tbytes for 1981-2014.

Pathfinder L3U data - HDF: None; netCDF: 32.5GiB/year; PNG: 4.0 GiB/year; totalling ~1.20 Tbytes for 1981-2014.

Pathfinder L3C data - HDF: 40 GiB/year; netCDF: 27 GiB/year; PNG: 1.5 GiB/year; totalling ~2.30 Tbytes for 1981-2014 (includes netCDF and png CDR products).

Software resources: The processing is performed by 3 sets of programs. Ingest of the L1b files uses programs **aci** and **aciseg**, which were originally based on heritage RATFOR code and converted into Fortran, and compiled with the Fortran compiler. The Level 2 programs **l2gen_both** and **l2bin_both**, and the level-3 global binning program, **l3bin**, **smigen_both** (based on the SEADAS Level-3 mapping routine, smigen) are all from SeaDAS 6.4 with changes made by RSMAS of University of Miami. The modified SeaDAS routines were built with the free gcc/g++/gfortran compilers. Libraries employed during processing are HDF4, HDF5 and SeaDAS 6.4 using 64 bit libraries. Specifically used to build PFSST V5.3 binaries on NCEI local server were the following –

Compilers: GNU Fortran V4.4.7, gcc V4.4.7;

External Libraries: Seadas, hdf5, SDPTK5.2.9v1.0,hdfeos;

System Libraries: zlib-devel-1.2.3-1.2.1, libjpeg-devel-6b-36.2.1.

To meet the requirements of different scales of the parallel jobs running on Amazon cloud computing environment, both HTCondor by University of Wisconsin and native Linux utility taskset are employed to manage the multitask jobs on multi-core/cpu system.

6. Assumptions and Limitations

6.1 Algorithm Performance

The main challenge in developing a Pathfinder global SST algorithm is to achieve relatively uniform performance throughout a wide range of atmospheric and oceanic conditions. As Barton [1995] pointed out, SST algorithms assume a first guess of the state of the atmosphere (e.g., a typical shape of water vapor and temperature profiles). A similar statement can be made about typical oceanic conditions (e.g., a certain average structure of the ocean's uppermost layer is assumed in comparisons with *in situ* SST measurements). When conditions deviate from the implicit first guess in atmosphere and ocean conditions, errors arise in SST retrievals. Deviations from implicit first-guess conditions are more likely in a global algorithm than in regionally-tuned algorithms, and this should be kept in mind when evaluating global SST estimates. Furthermore, in the case of statistically-derived global SST algorithms, the first-guess conditions will be the average of conditions at all the matchup locations and times used in coefficient estimation. We stress that this average will be weighted by the relative distribution of matchups, likely to change in time. The performance of an SST algorithm for a given set of atmospheric and oceanic conditions, therefore, depends not only on how close those conditions are to the average state, but also on how well represented are those conditions in the matchup set used to derive the algorithm coefficients.

6.2 Sensor Performance

A variety of assumptions are made concerning sensor performance, radiometric noise, radiometric calibration, spatial and spectra registration, and geophysical effects that can impact accuracy and precision of the CDR. The PFMDB is used to try and identify conditions when these assumptions are violated and retrievals are degraded. A time series (see Tables 5-11) of the residuals indicates that both the accuracy and the precision of the sensors themselves improve with time as indicated by the decreasing standard deviation with the launch of each new sensor. Tracking of the standard deviation also aids in identifying sensor performance degradation as they age on-orbit. Known sensor specific issues are listed below and newly identified known problems that impact the Sea Surface Temperature – Pathfinder CDR are posted on data portal <http://www.nodc.noaa.gov/sog/pathfinder4km/userguide.html> as they are discovered.

Stray sunlight: An examination of data stratified by satellite zenith angle and by side of the AVHRR scan line (left and right of nadir) revealed potential problems under certain Earth-Sun-satellite configurations. A quality flag set during processing identifies configurations in which problems may potentially occur. The problem is probably associated with stray solar radiation entering the radiometer and it occurs only in the middle to high latitudes in the Southern Hemisphere.

NOAA-7: Estimating algorithm coefficients for N-7 was challenging due to the lack of geographic diversity in the *in situ* observations and the eruption of El Chichon in March of 1982. Inclusion of shipboard observations during the coefficient estimation process

improved the time dependent stability of the residuals and provided a modest improvement in the aggregate bias -0.24 compared to -0.4 if only buoy observations were used. Coefficients based solely on *in situ* buoy observations are available for all sensors following NOAA-7. In addition because of the limited matchups we use the cloud tree developed for N-9 and apply it to the N7 data as the limited matchups for N7 precluded development of an efficacious decision tree.

NOAA-11: The June 12, 1991 *eruption of Mount Pinatubo injected about ten cubic kilometers of absorbing aerosols into the atmosphere impacting the accuracy of PFSST retrievals in that latitude band for months. See Kilpatrick et al. [1991].*

NOAA-14: Errors in brightness temperatures for channel 4 in the Advanced Very High Resolution Radiometer (AVHRR) onboard the NOAA-14 spacecraft are examined. The errors involve a low frequency of occurrence for some values, and a corresponding enhancement of frequency for others. Errors appear to be related to the conversion of analog to digital values. Unfortunately, it is not possible to identify and separate erroneous values. The most apparent errors in geophysical products derived from AVHRR's channel 4 occur at low brightness temperatures, therefore sea surface temperatures in high latitudes (below about 6°C) and cloud-related products must be used with caution, as they may have systematic errors as large as 0.5°C. See Podestá et. al. 2003 for details.

NOAA-15: was not included in the Pathfinder 5.3 product set because the behavior of the NOAA-15 Hovmöller diagram for the NOAA-15 time series did not follow the SST error patterns evident in all of the other AVHRR sensors included in the Sea Surface Temperature – Pathfinder CDR v5.3 time series.

NOAA-16: Intermittent scan motor issues on board N-16 result in occasional missing data in the CDR for this sensor as these pixels are flagged as poor or unusable quality. Inconsistencies in the on-board calibration for some scenes have been reported due to thermal stress and gradients on the internal calibration targets (ICT) for N-9 through N-16 when flying through the terminator. Due to these multiple problems, N-16 was only used briefly in the CDR through 2002.

NOAA-17: The CDR switches to N-17 despite the earlier morning crossing time as a result to the poor behavior of N-16. Due to the difference in the morning crossing, and therefore the time of day measurement, users should be aware of the potential impact of diurnal warming in certain seasons and latitude bands when evaluating multi-sensor times series data.

7. Future Enhancements

7.1 Enhancement 1 Latband coeffs

Revised algorithm coefficients as a function of Latitude band, month, and water vapor regime are expected to improve the precision and accuracy of the retrievals. Note that Cloud tree tests for NOAA-7 and NOAA-19 have been updated for v5.3 to be consistent with the rest of the sensors. In v5.2 they were not.

7.2 Enhancement 2 -GHRSSST compliant SSES bias and STDEV

For future Pathfinder versions, a bias (satellite-in situ) and standard deviation, referred to as single sensor error statistics (SSES), will be included in the GHRSSST complaint files. The validation hypercube for Sea Surface Temperature – Pathfinder CDR, is anticipated to include 7 dimensions: time by season (4), latitude bands (5 steps in 20 degree from 60S to 60N), surface temperature (8 increments in 5 degree steps), satellite zenith angle (4 increments), brightness temperature difference as a proxy for water vapor (3 intervals for 11-12 mm SST), retrieved satellite SST quality level, day/night selection (2 intervals).

7.3 Enhancement 3: Level 2 pixel level ice flag

Currently a tiered indirect approach to flagging pixels compromised by sea-ice is being taken during the level-3 bin quality level assignment. To incorporate ice information into the processing first, 25km weekly ice-masks derived from the Special Sensor Microwave Imager (SSM/I) and processed by Remote Sensing Systems, Inc. (RSS), are used to identify regions with ice concentrations exceeding 1% (in other words, if any ice is present the region is flagged as ice covered). These data are available back to July 1987. Prior to 1987, the sea ice information from the Reynolds OISST V2 is used. Because SSM/I sensor data are not valid near land, ice concentrations in coastal regions are not properly identified. These gaps are filled using the Reynolds OISST V2 data at 1-degree resolution. Finally, because some areas appear inconsistent (such as a gap right at the poles in the SSM/I data), an automated approach to "backfilling" the ice mask into interior regions has been developed. To summarize, the first tier uses the SSM/I data, the second tier uses the OISST V2 data, and the third tier fills gaps using an automated algorithm. This tiered ice mask is then used to exclude those pixels from the calculation of the 3-week running mean internal reference that is then used as an internal check (see section 3.2.4 for discussion on *Bin quality and reference field*). Eventually, in Version 6 of the Pathfinder reprocessing efforts, the ice masks will be used to specifically flag the level 2 pixel data as poor quality. The current Version 5 approach is intended as an intermediate step to reduce the likelihood of misclassified pixels (that is, SST pixels with high quality levels but falling on ice mask) and develop an understanding of the impact of the ice masking techniques on the dataset without requiring major changes to the processing system.

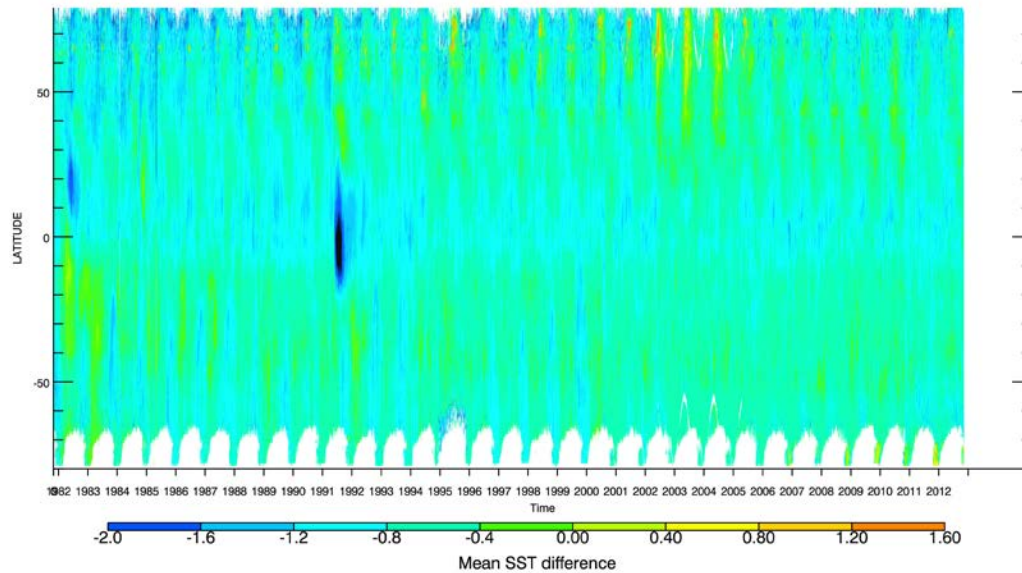
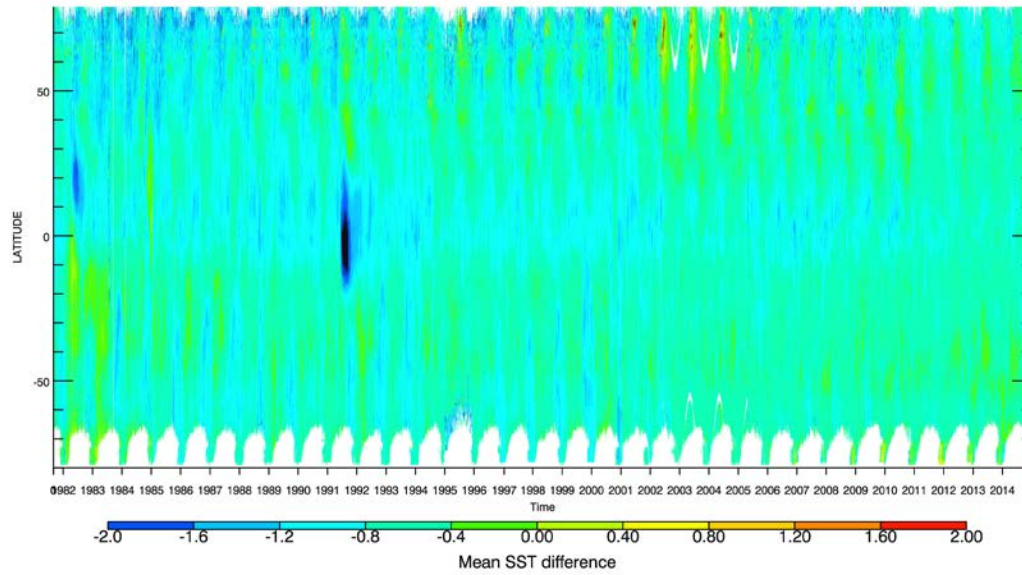
7.4 Resolving the Binning Problem in SeaDAS

Mixed Day and night granules- Granules containing mixed day and night data are not split properly into the appropriate global file in the SeaDAS code module L3BIN. This error has a strong impact on high latitude SST values (see Figure 3).

1. Daytime data is being included in the night only global file, and conversely nighttime data is included in the daytime file.
2. This improper splitting of mixed granules results in high latitude SST binned averages that may be several degrees in error depending on the time of year.
3. The error is most noticeable around the summer and winter solstice for each hemisphere.

Granules over the poles may not get binned and this is still under investigation. On rare occasions a granule appears to be assigned to the wrong data day; the reason for this is still unknown. Since it is a very rare event we assume it is not having a large impact on the quality of the data. Due to the rarity it has also been a rather illusive problem to figure out why it happens. This problem exists in both Pathfinder SST 5.2 and 5.3 in relation to the SeaDas Algorithms, which have not been updated for the AVHRR sensors. This issue will be addressed and updated in the next round of processing.

Figure 3: Hovmöller difference plots for Pathfinder SST and daily optimum interpolated SST (dOISST) for 5.3 during 1981-2014 (upper panel) and 5.2 during 1981-2012.



8. References

- Anding, D. and R. Kauth. 1970. Estimation of sea surface temperature from space, *Remote Sens. Environ.*, 1, 217-220.
- Brasnett, B. (2008), The impact of satellite retrievals in a global sea-surface-temperature analysis, *Q. J. R. Meteorol. Soc.*, 134 , 1745 –1760, doi:10.1002/qj.319.
- Campbell, J., J.M. Blaisdell, and M. Darzi: 1995, Spatial and Temporal Binning Algorithms, NASA Technical Memorandum 104566, Vol 32, appendix A. pp.63-65.
- Cornillion, P., C. Gilman, C.L. Stramma, O. Brown, R. Evans, and J. Brown. 1987. Processing and analysis of large volumes of satellite-derived thermal infrared data. *J. Geophys. Res.* 12,993-12002
- Barton, I.J., 1995. Satellite-derived sea surface temperatures: Current status. *Journal of Geophysical Research* 100: 8777–8790.
- Brown, J. W., O. B. Brown, and R. H. Evans (1993), Calibration of Advanced Very High Resolution Radiometer Infrared Channels: A New Approach to Nonlinear Correction, *J. Geophys. Res.*, 98(C10), 18,257–18,268, doi:10.1029/93JC01638
- Brown, O. B., J. W. Brown, and R. H. Evans (1985), Calibration of Advanced Very High Resolution Radiometer Infrared Observations, *J. Geophys. Res.*, 90(C6), 11,667–11,677, doi:10.1029/JC090iC06p11667.
- M. Kanamitsu, W. Ebisuzaki, J. Woollen, S-K Yang, J.J. Hnilo, M. Fiorino, and G. L. Potter. NCEP-DEO AMIP-II Reanalysis (R-2). 1631-1643, Nov 2002, *Bul. of the Atmos. Met. Soc.*
- Kearns, E.J., J.A. Hanafin, R.H. Evans, P.J. Minnett, and O.B. Brown. 2000. An independent assessment of the Pathfinder AVHRR sea surface temperature accuracy using the Marine Atmospheric Emitted Radiance Interferometer, *Bull. Meteorol. Soc.* 81, 1525-1536.
- Kilpatrick, K. A., G. P. Podesta, and R.H. Evans. 2001. "Overview of the NOAA/NASA Pathfinder algorithm for Sea Surface Temperature and associated Matchup Database." *J. Geophys Res.*, 106: 9179-9198. 13497- 13510.
- McClain, E.P., W.G. Pichel, and C.C. Walton. 1985. Comparative performance of AVHRR-based multichannel sea surface temperatures, *J. Geophys. Res.* 90, 11,587-11601.
- Minnett, P.J. 1990. The regional optimization of infrared measurements of sea surface temperature from space. *J. Geophys Res.*, 95
- Prabhakara, C., G. Dalau, and V.G. Kunde. 1974. Estimation of sea surface temperature from remote sensing in the 11 to 13mm window region, *J. Geophys. Res.*, 79, 5039-5044.

- Podestá, G.P. 1995. SeaWiFS global fields: what's in a day? Chapter 5, pages 34–42. SeaWiFS Technical Report Series, Volume 27. NASA Technical Memorandum 104566, Vol. 27.
- Podestá, G.P., M. Arbelo, R.H. Evans, K.A. Kilpatrick, V. Halliwell and J. Brown. Errors in high-latitude SSTs and other geophysical products linked to NOAA-14 AVHRR channel 4 problems. *Geophys Research Letters*, 30 (11),1548.
- Rao, C.R.N., and J. Chen, 1995: Inter-satellite calibration linkages for the visible and near-infrared channels of the Advance Very High Resolution Radiometers on the NOAA-7, -9, and 11 spacecraft, *International Journal of Remote Sensing*, 16, 1931-1942.
- Rao, C.R.N., and J. Chen, 1996: Post launch calibration of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-14 spacecraft, *International Journal of Remote Sensing*, 17, 273-2747.
- Rao, C.R.N., J.T.Sullivan, C.C. Walton, J.W. Brown, and R.H.Evans. 1993. Non-linearity Corrections for the Thermal Infrared Channels of the Advanced Very High Resolution Radiometer: Assessment and Corrections, NOAA Technical Report NESDIS 69, Department of Commerce, Washington, D.C.
- Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey and M. G. Schlax, 2007: [Daily High-resolution Blended Analyses for sea surface temperature](#). *J. Climate*, 20, 5473-5496.
- Rousseeuw, P.J. and V. Yohai. 1984. Robust regression by means of S-estimators. In: *Robust and Nonlinear Time Series Analysis*, J. Franke, W. Hardle, and R.D. Martin (eds.), Lecture Notes in Statistics, 26, 256-272, Springer-Verlag.
- Schluessel, P., W.J. Emery, H. Grassl, and T. Mammen. 1990. On the skin-bulk temperature difference and its impact on satellite remote sensing of sea surface temperature, *J. Geophys. Res*, 95, 12241-12256.
- Walton, C. C 1988. Nonlinear multichannel algorithm for estimating sea surface temperatures with AVHRR satellite data, *J. Appl. Meteorol.* 27,115-124.
- Walton, C. C., W. G. Pichel, F. J. Sapper, and D. A. May, 1988a. The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with NOAA polar-orbiting environmental satellites. *Journal of Geophysical Research* 103: 27999–28012.
- Yohai, V., W.A. Stahel, and R.H. Zamar. 1991. A procedure for robust estimation and inference in linear regression. In: *Directions in Robust Statistics and Diagnostics, Part II*, Stahel, W.A., and S.W. Wesiberg (eds.), Springer-Verlag, New York.
- Yohai, V., and R.H. Zamar. 1998. Optimal locally robust M-estimates of regression. *J. of Statist. Inf. and Planning*.

Yohai, V., W.A. Stahel, and R.H. Zamar. 1991. A procedure for robust estimation and inference in linear regression. In: *Directions in Robust Statistics and Diagnostics, Part II*, Stahel, W.A., and S.W. Wesiberg (eds.), Springer-Verlag, New York.

Yohai, V., and R.H. Zamar. 1998. Optimal locally robust M-estimates of regression. *J. of Statist. Inf. and Planning*.

Appendix A. Acronyms and Abbreviations

Acronym or Abbreviation	Definition
ACI	AVHRR Class data Ingest program
ACISEG	AVHRR Class data orbit segmenting program
ANSI	American National Standards Institute
AMSRE	Advance Microwave Scanning Radiometer Earth Observing
AVHRR	Advanced Very High Resolution Radiometer
BT	Brightness Temperature
C	Celsius
C-ATBD	Climate Algorithm Theoretical Basis Document
Ch	Channel
CATBD	Climate Algorithm Theoretical Basis Document
CDR	Climate Data Record
CLASS	NOAA's Comprehensive Large Array-data Stewardship System
DOMSAT	Domestic Satellite Earth System Relay Station
GAC	Global Area Coverage
GDS	GHRSSST-PP Data Processing Specification
GHRSSST	Group for High resolution sea surface temperature program
GMT	Greenwich Mean Time
GCOS	Global Climate Observing System
HDF4	Hierarchical Data Format Version 4
HRPT	High Resolution Picture Transmission
K	Kelvin
km	Kilometer
LAC	Local Area Coverage
LTSRF	Long Term Stewardship and Reanalysis Facility
IFOV	Instantaneous field of view
IR	Infrared Radiation
NASA	National Aeronautical and Space Science Agency
NCDC	NOAA National Climatic Data Center

A controlled copy of this document is maintained in the CDR Program Library.
Approved for public release. Distribution is unlimited.

NCEI	National Centers for Environmental Information
NODC	NOAA National Ocean Data Center
NESDIS	National Environmental Satellite, Data and Information Services
NOAA	National Oceanic and Atmospheres Administration
MAERI	Marine Atmospheric and Emitted Radiance Interferometer
POES	Polar Orbiting Environmental Satellite
PODUG	Polar orbital data users guide
PFSST	Pathfinder sea surface temperature
PFMDB	Pathfinder Matchups Database
PRT	Platinum Resistance Thermometers
RSMAS	Rosenstiel School for Marine and Atmospheric Science University of Miami
RSS	Remote Sensing Systems, Inc.
SEADAS	NASA's Seawifs Data Analysis System
S/N	Signal to Noise
SSES	Single Sensor Error Statistics
SST	Sea Surface Temperature
TBUS	TIROS Bulletin United States
TIROS	Television Infrared Observation Satellite