

---

# Climate Data Record (CDR) Program

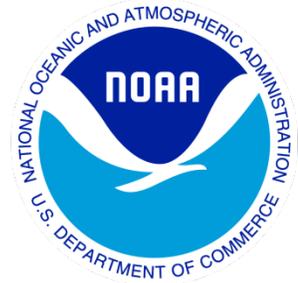
## Climate Algorithm Theoretical Basis Document (C-ATBD)

### Ocean Surface Bundle

[Sea Surface Temperature – WHOI,

Ocean Near Surface Properties,

and Ocean Heat Fluxes]



CDR Program Document Number: CDRP-ATBD-0578  
Configuration Item Number: 01B-27  
Revision 2 / June 03, 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	Carol Anne Clayson, WHOI	DSR-743	Initial Submission to CDR Program	10/24/2014
2	Carol Anne Clayson, WHOI	DSR-1031	Changes to processing for SSM/I time series, inclusion of SSMIS time series. Sections changed: 1.1, 2.1, 2.2, 3.1, 3.2, 3.2.1, 3.2.2, 3.2.3, 3.3.1, 3.3.2, 3.4.1, 3.4.6, 5, 5.3, 5.4.3, References, Appendix.	06/03/2016

## 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.*

<b>1. INTRODUCTION.....</b>	<b>6</b>
1.1 Purpose .....	6
1.2 Definitions.....	6
1.3 Referencing this Document .....	6
1.4 Document Maintenance.....	7
<b>2. OBSERVING SYSTEMS OVERVIEW.....</b>	<b>8</b>
2.1 Products Generated .....	8
2.2 Instrument Characteristics .....	8
<b>3. ALGORITHM DESCRIPTION.....</b>	<b>10</b>
3.1 Algorithm Overview .....	10
3.2 Processing Outline.....	10
3.2.1 Production of Qa, Ta, U10.....	13
3.2.2 Production of SST.....	19
3.2.3 Production of LHF and SHF .....	22
3.3 Algorithm Input.....	22
3.3.1 Primary Sensor Data .....	22
3.3.2 Ancillary Data.....	22
3.3.3 Derived Data .....	23
3.3.4 Forward Models.....	23
3.4 Theoretical Description .....	23
3.4.1 Physical and Mathematical Description.....	24
3.4.2 Data Merging Strategy.....	27
3.4.3 Numerical Strategy .....	27
3.4.4 Calculations.....	28
3.4.5 Look-Up Table Description.....	28
3.4.6 Parameterization .....	28
3.4.7 Algorithm Output.....	29
<b>4. TEST DATASETS AND OUTPUTS.....</b>	<b>30</b>
4.1 Test Input Datasets .....	30
4.2 Test Output Analysis .....	30
4.2.1 Reproducibility.....	30
4.2.2 Precision and Accuracy .....	30
4.2.3 Error Budget.....	30
<b>5. PRACTICAL CONSIDERATIONS.....</b>	<b>39</b>
5.1 Numerical Computation Considerations.....	39
5.2 Programming and Procedural Considerations .....	39

<b>5.3</b>	<b>Quality Assessment and Diagnostics .....</b>	<b>40</b>
<b>5.4</b>	<b>Exception Handling .....</b>	<b>41</b>
5.4.1	Conditions Checked .....	41
5.4.2	Conditions Not Checked .....	41
5.4.3	Conditions Not Considered Exceptions.....	41
<b>5.5</b>	<b>Algorithm Validation .....</b>	<b>42</b>
5.5.1	Validation During Development .....	42
<b>5.6</b>	<b>Processing Environment and Resources .....</b>	<b>42</b>
<b>6.</b>	<b>ASSUMPTIONS AND LIMITATIONS .....</b>	<b>44</b>
<b>6.1</b>	<b>Algorithm Performance.....</b>	<b>44</b>
<b>6.2</b>	<b>Sensor Performance.....</b>	<b>44</b>
<b>7.</b>	<b>FUTURE ENHANCEMENTS.....</b>	<b>45</b>
7.1	Enhancement 1: Reprocessing.....	45
7.2	Enhancement 2: Extend Temporal Record.....	45
7.3	Enhancement 3: Improved Parameter Values .....	45
7.4	Enhancement 4: Reduced Temporary Storage.....	45
7.5	Enhancement 5: Smoother Ice Mask .....	45
<b>8.</b>	<b>REFERENCES.....</b>	<b>46</b>
<b>APPENDIX A.</b>	<b>ACRONYMS AND ABBREVIATIONS.....</b>	<b>48</b>
<b>APPENDIX B.</b>	<b>QUALITY CONTROL FLAGS FOR ATMOSPHERIC PARAMETERS.....</b>	<b>49</b>
<b>APPENDIX C.</b>	<b>QUALITY CONTROL FLAGS FOR SST .....</b>	<b>50</b>
<b>APPENDIX D.</b>	<b>QUALITY CONTROL FLAGS FOR HEAT FLUXES.....</b>	<b>51</b>

## LIST of FIGURES

Figure 1: Top-level processing flow for production of Ta, Qa, and U10 data.....	11
Figure 2: Top-level processing flow for production of SST data.....	12
Figure 3: Top-level processing flow for production the ocean heat flux data.....	13
Figure 4: Processing flow for creation of swath-level fields of Ta, Qa, and U10. ....	15
Figure 5: Processing flow for creation of gridded fields of Ta, Qa, and U10 data. ....	17
Figure 6: Processing flow for creation of final interpolated fields of Ta, Qa, and U10. ....	18
Figure 7: Processing flow for creation of peak diurnal warming SST values. ....	20

Figure 8: Processing flow for creation of final interpolated fields of SST.....	21
Figure 9: Comparisons between SeaFlux satellite-derived dataset and IVAD data for the component fields.....	32
Figure 10: Differences between the SeaFlux v1.0 satellite dataset and the IVAD measurements as a function of the satellite values. ....	34
Figure 11: Mean fields from the 10 years of SeaFlux version 1.0 data for the Ta, SST, Wspd, and Qa values and the associated total uncertainties.....	37
Figure 12: Mean fields from the 10 years of SeaFlux data for the Qs-Qa, LHF, SHF, and Ts-Ta values and the associated total uncertainties.....	38

## LIST of TABLES

Table 1: Processing Environment.....	43
--------------------------------------	----

# 1. Introduction

## 1.1 Purpose

The purpose of this document is to describe the algorithm submitted to the National Climatic Data Center (NCDC) by Carol Anne Clayson at Woods Hole Oceanographic Institution (WHOI) that will be used to create the Ocean Surface Bundle Climate Data Record (CDR). Which consists of the Sea Surface Temperature – WHOI CDR, the Ocean Near Surface Properties CDR, and the Ocean Heat Fluxes CDR. The CDR bundle uses as its main basis the Special Sensor Microwave/Imager (SSM/I), the Special Sensor Microwave/Imager Sounder (SSMIS), and the Advanced Very High Resolution Radiometer (AVHRR). 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.

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

$$T_a = \text{10-m air temperature} \quad (2)$$

$$SST = \text{Sea surface temperature} \quad (3)$$

$$U_{10} = \text{10-m wind speed} \quad (4)$$

$$Q_a = \text{10-m specific humidity} \quad (5)$$

$$Q_s = \text{sea surface specific humidity} \quad (6)$$

$$LHF = \text{surface latent heat flux} \quad (7)$$

$$SHF = \text{surface sensible heat flux} \quad (8)$$

$$CLW = \text{cloud liquid water} \quad (9)$$

$$PW = \text{precipitable water} \quad (10)$$

## 1.3 Referencing this Document

This document should be referenced as follows:

Ocean Surface Bundle - Climate Algorithm Theoretical Basis Document, NOAA Climate Data Record Program [CDRP-ATBD-0578] Rev. 2 (2016). Available at <http://www.ncdc.noaa.gov/cdr/>

## 1.4 Document Maintenance

Synchronization between this document and the algorithms is achieved through version and revision numbers. The version and revision numbers found on the front cover of this document can be compared with the values of VERSION and REVISION in the source file. If the document applies to the algorithm, then these numbers will match. If they do not match, and it is found that the document needs to be updated, then the header comment in the file should be consulted – under its REVISION HISTORY section is a description of the changes for each version and revision from which the necessary updates to this document can be made.

## 2. Observing Systems Overview

### 2.1 Products Generated

The primary generated products are global open-ocean climate data records (CDR) of near-surface air temperature ( $T_a$ ), wind speed (U10), specific humidity ( $Q_a$ ), and sea surface temperature (SST), as well as the resulting fluxes of latent heat (LHF) and sensible heat (SHF). The data records for  $T_a$ , U10, and  $Q_a$  are generated from the V1.0 Fundamental Climate Data Record (FCDR) of brightness temperature ( $T_b$ ) data from the Special Sensor Microwave/Imager instruments (SSM/I) on board the Defense Meteorological Satellite Program spacecraft F08, F10, F11, F13, F14, and F15 provided by Colorado State University, and the V2.0 FCDR of  $T_b$  data from the Special Sensor Microwave/Imager Sounder (SSMIS) on board the DMSP spacecraft F17 and F18 also provided by Colorado State University. The method for determining  $T_a$ ,  $Q_a$ , and U10 is described in more detail in Section 3 and is based on the methodology of Roberts et al. (2010) and Clayson et al. (2014). The SST fields are based on the V1.0 R2 CDR of daily  $1/4^\circ$  Optimum Interpolation Sea Surface Temperature (OISST), with a diurnal warming correction added using the U10 data from this dataset. The peak solar radiation used in the SST diurnal warming algorithm is the Global Energy and Water Experiment (GEWEX) Surface Radiation Budget (SRB)-Release 3.0 dataset (Stackhouse et al. 2011) through February 2000, and then the CERES-Syn1Deg-3H product (Ed3A; CERES 2016) is used. Due to the lag in production of the CERES-Syn1Deg-3H product, the NASA CERES FLASHFlux (Kratz et al. 2014) product is used for the near-real-time data, currently beginning December 2015. Precipitation is from the Global Precipitation Climatology Project (GPCP) v1.2 one-degree daily precipitation data set (Huffman et al., 2001) for the time period from October 1996. Prior to this time Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS) v3.2 precipitation data (Andersson et al. 2010; Fennig et al. 2012) is used and is described in more detail in Section 3. The resulting LHF and SHF fluxes are calculated from these fields of  $T_a$ ,  $Q_a$ , U10, and SST using a neural network version of the Coupled Ocean-Atmosphere Response Experiment (COARE) 3.0 flux algorithm (Fairall et al. 2003; Clayson et al. 2014) and is described in more detail in Section 3.  $T_a$ ,  $Q_a$ , and U10 values are found in the ATMOS files; SST is found in the SST files, and lastly LHF and SHF are found in the FLUX files.

Accompanying the fields of  $T_a$ ,  $Q_a$ , U10, SST, and LHF and SHF are data quality information fields, indicating which grid cells are land or ice masked, lake areas, and where the data is based solely on the neural network-derived quantity or is interpolated. Further information is available in Section 3.

### 2.2 Instrument Characteristics

One set of satellite data used in this algorithm is collected by SSM/I instruments on board the Defense Meteorological Satellite Program (DMSP) satellites F08, F10, F11, F13, F14, and F15. The SSM/I sensor has seven channels at four frequencies, three of which are dual polarized, horizontal (H) and vertical (V): 19.4, 37.0, and 85.5 GHz. The 22.2 GHz

frequency has only a single vertically polarized channel. These satellites are in polar orbits with nominal altitudes ranging from 830 to 860 km with varying earth incidence angles of roughly 53 degrees. The footprint or instantaneous field of view of the sensor varies with frequency (from a 70 x 45 km for the 19.35 GHz channels to 16 x 14 km footprint of the 85.5 GHz channels). The brightness temperatures used in these retrievals based on the SSM/I instruments are from the Colorado State University (CSU) Fundamental Climate Data Records (FCDR). Full information for the brightness temperatures and the SSM/I instrument characteristics can be found in the C-ATBD for that dataset (CDRP-ATBD-0337, Rev. 1 07/11/2013).

Beginning in March 2008, additional data from the SSMIS instruments on board the DMSP satellite F17 are included. Data from the SSMIS instruments on board DMSP satellite F18 begins March 2010. The instrument is a conically scanning passive microwave radiometer sensing upwelling microwave radiation at 24 channels covering a wide range of frequencies from 19 – 183 GHz. Data is collected along an active scan of 143.2 degrees across track producing a swath width on the ground of approximately 1707 km with 12.5 km scene spacing. The channels consist of the following sets: *Environmental sensor* (ENV) – channels 12- 16; *Imager* (IMG) – channels 8-11 and 17-18; *Lower atmospheric sounding* (LAS) – channels 1-7 and 24; *Upper atmospheric sounding* (UAS) – channels 19-23. Of these channels, the SSM/I equivalent channels include all of the environmental channels (12- 16) and imager channels 17 and 18. This product uses the CSU FCDR brightness temperatures the following dual-polarized, intercalibrated channels: 19.4, 37.0, and 91.7 GHz. As with the SSM/I, the 22.2 GHz frequency has only a single vertically polarized channel. Full information for the SSMIS brightness temperatures and the SSMIS instrument characteristics can be found in the C-ATBD for that dataset (CDRP-ATBD-0338, Rev. 2 04/15/2015).

Another set of satellite inputs are from the AVHRR (Advanced Very High Resolution Radiometer) instruments. AVHRR technical documentation, including instrument and operational data formats is available at <http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/intro.htm>. The AVHRR satellites are onboard the NOAA operational polar orbiting TIROS-N series satellites. These satellites are in near-polar, sun-synchronous orbits with orbital periods of about 102 minutes. The NOAA Polar Orbiter Data User's Guides (PODUG) November 1998 revision describes NOAA 14 and earlier, while the next generation instruments covering NOAA 15 and later are covered by the NOAA KLM User's Guide (April 2009 revision). NOAA-N (NOAA 19 after launch) and -P are also described in the NOAA-N, -P Supplement <http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/klm/nnpsupp.htm>. This data is used as input for the Daily 1/4° Optimum Interpolation Sea Surface Temperature (OISST) CDR, which is the base SST that is an input for the diurnally-varying SST in this dataset. Details on the OISST CDR can be found in CDRP-ATBD-0303 (Rev. 2 09/17/2013).

## 3. Algorithm Description

Section 3.1 describes the SeaFlux Ocean Surface Bundle (OSB) CDR V2 algorithm. A processing flow overview is given in Section 3.2. Section 3.3 describes the input datasets as well as ancillary files. The theoretical background and computations are discussed in Section 3.4.

### 3.1 Algorithm Overview

The 10-m air temperature ( $T_a$ ), 10-m specific humidity ( $Q_a$ ), and 10-m wind speed ( $U_{10}$ ) are determined from the input brightness temperatures from the CSU FCDR at swath level using a neural network algorithm. This neural network algorithm (described in further detail in Section 3.4.1) also uses the Daily  $1/4^\circ$  Optimum Interpolation Sea Surface Temperature (OISST) CDR as input. All potentially contaminated pixels including land and ice are flagged and discarded. The data are combined across satellites and binned. Lastly the data are interpolated to fill in missing parameters for each location to produce the ATMOS files.

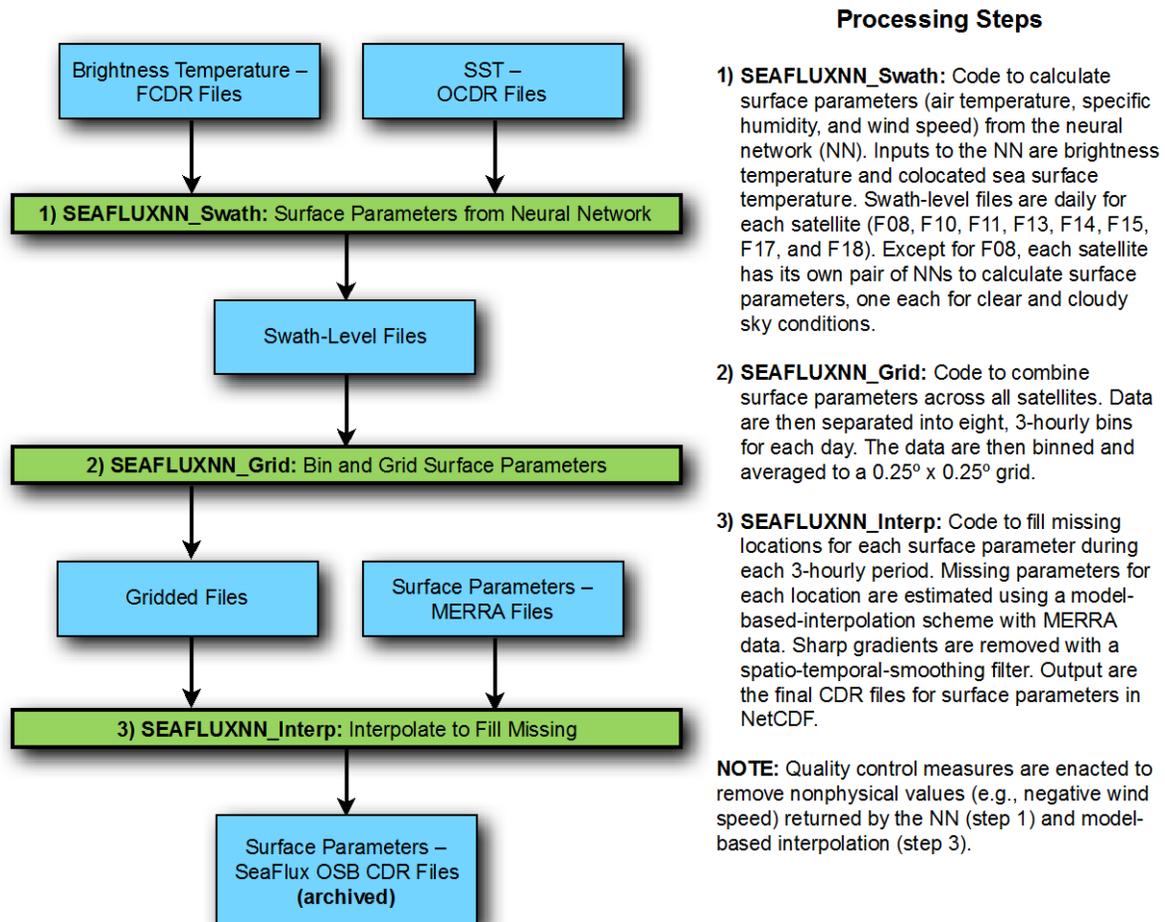
The SST fields takes as input data from several sources: (1) the binned ATMOS files, (2) precipitation data from either the HOAPS dataset or the GPCP product, and (3) solar radiation fields from the GEWEX SRB, the CERES-Syn1Deg-3H, or the FLASHFlux dataset, and calculates the peak warming at each grid point using the algorithm described in more detail in Section 3.4.6. This peak diurnal warming is used to create a curve of daily SST variability, which is added to the OISST data for each day and location. The final diurnally-varying SSTs and the QC flags associated with these data are found in the SST files of this dataset.

Lastly, the ATMOS and SST files are used to calculate the surface fluxes using a neural network version of the COARE 3.0 flux algorithm. The output of the neural network described in Section 3.4.1 are the sensible and latent heat fluxes found in the FLUX files of this dataset.

### 3.2 Processing Outline

An overview of the routine production flow for producing the  $T_a$ ,  $Q_a$ , and  $U_{10}$  fields is shown in Figure 1 and follows the method outlined in Section 3.1. The first main step is to take the input brightness temperatures files and the input SST field and calculate the  $T_a$ ,  $Q_a$ , and  $U_{10}$  fields at the swath-level using the neural network routine. The second main step is to combine the surface parameters across all satellites, and then bin into eight, 3-hourly bins per day and average these into a  $0.25^\circ \times 0.25^\circ$  grid. The last main step is to interpolate the data to fill in missing parameters for each location. As described in Section 3.4.3, the interpolation makes use of the gradient of variability in version two of the Modern Era Retrospective-analysis for Research and Applications (MERRA-2; Bosilovich 2008) output. Each of these main steps is described in detail below.

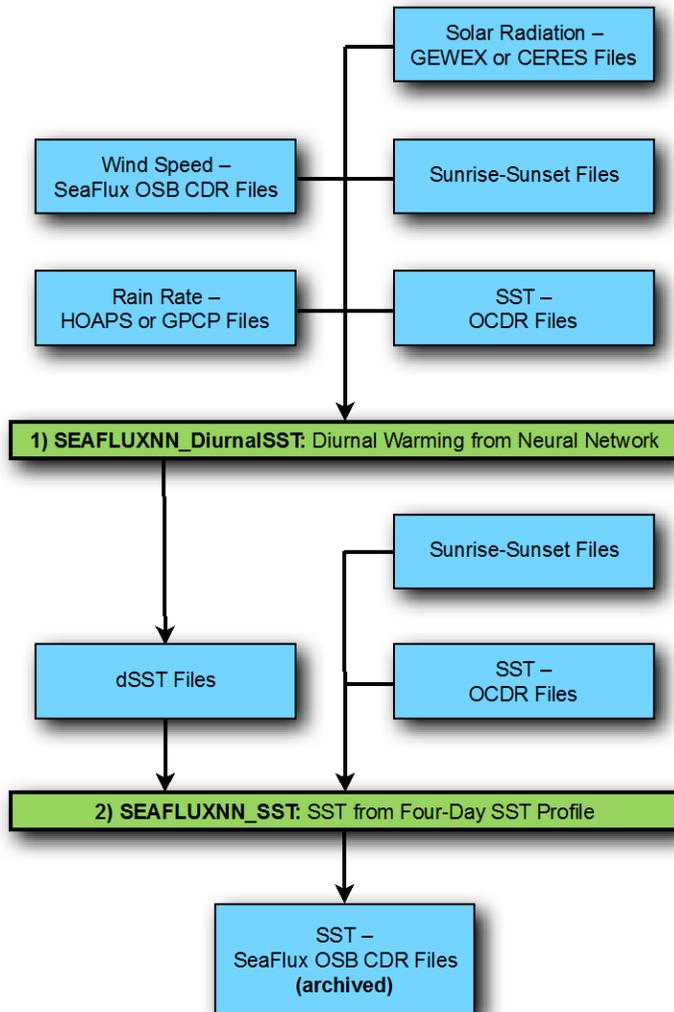
## SeaFlux OSB CDR Processing Flow Chart Ocean Near-Surface Properties



**Figure 1: Top-level processing flow for production of Ta, Qa, and U10 data.**

An overview of the routine production flow for producing the SST fields is shown in Figure 2. The initial step takes as input the binned U10 fields, the HOAPS or GPCP precipitation fields, and the ISCCP solar radiation fields, and calculates the peak diurnal warming at each point in the  $0.25^\circ \times 0.25^\circ$  grid. This peak diurnal warming is used to create a curve of daily SST variability, which is added to the OISST data for each day and location. Further details are below.

### SeaFlux OSB CDR Processing Flow Chart Sea Surface Temperature



#### Processing Steps

1) **SEAFUXNN\_DiurnalSST**: Code to calculate peak diurnal warming for sea surface temperature (dSST) from the neural network (NN). Inputs to the NN are daily-averaged wind speed, daily-accumulated rain rate, solar radiation, length of day from sunrise and sunset times, and sea surface temperature. dSST are produced for each day on the same 0.25° x 0.25° grid as used of the SeaFlux OSB CDR.

2) **SEAFUXNN\_SST**: Code to calculate diurnally-varying sea surface temperature. dSST from step one are compared across four days—two days before and one day after the sample day—superimposing the diurnal cycle onto the foundation SST. Sunrise and sunset times are taken into account along with dSST from the first processing step and SST from satellite observations. Output are the final CDR files for SST in NetCDF using the same spatial and temporal sampling as SeaFlux OSB CDR files: 3-hourly for each day and 0.25° x 0.25° grid.

**NOTE:** Quality control measures are enacted to remove nonphysical values.

Figure 2: Top-level processing flow for production of SST data.

The last step is to use the binned and interpolated data to calculate the surface fluxes LHF and SHF as shown in Figure 3. This is done using a neural net version of the COARE 3.0 algorithm flux algorithm, as described in Section 3.4.1.

### SeaFlux OSB CDR Processing Flow Chart Ocean Heat Fluxes

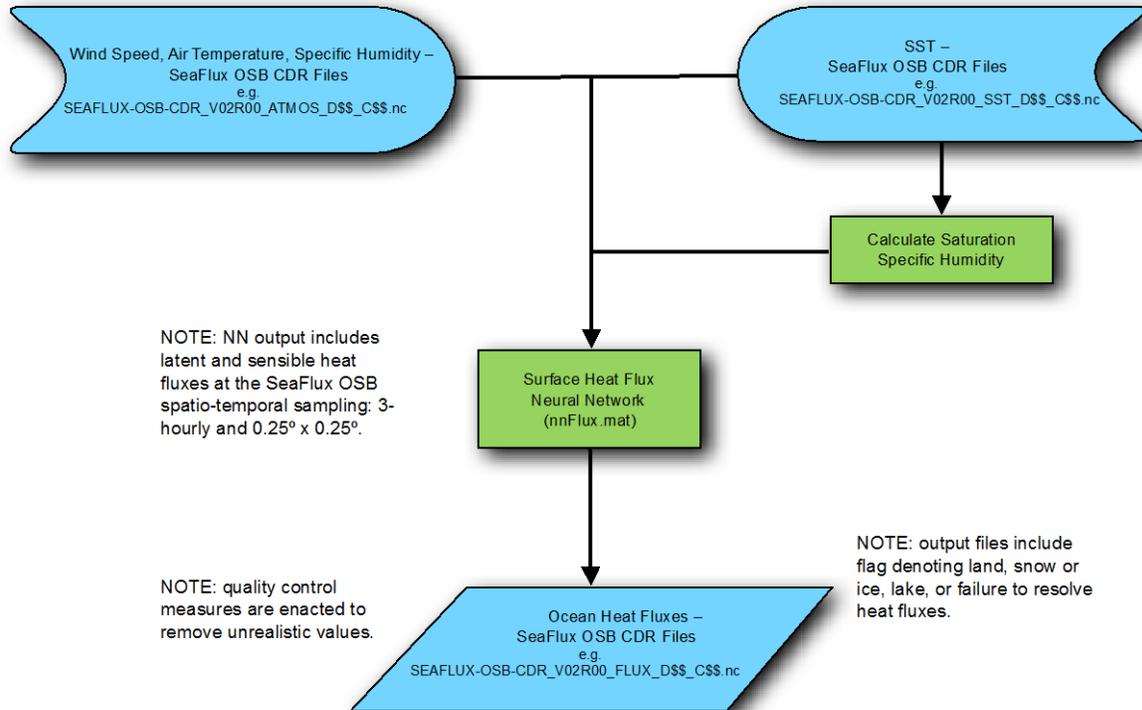


Figure 3: Top-level processing flow for production the ocean heat flux data.

#### 3.2.1 Production of $Q_a$ , $T_a$ , $U_{10}$

The overview of the production of the swath-level fields of  $Q_a$ ,  $T_a$ , and  $U_{10}$  is shown in Figure 4. The  $Q_a$ ,  $T_a$ , and  $U_{10}$  swath-level fields are generated from the NOAA L2A CSU SSM/I  $T_b$  fields (the CSU FCDR brightness temperatures). Pixels are removed by applying the quality flag provided with the brightness temperature data. A first-guess SST from the Reynolds OISSTv2 data is then collocated with the  $T_b$  data. For all of the satellites except F08, cloud liquid water content is calculated using Weng et al. (1997), and pixels that are rain-contaminated (where the CLW is greater than  $0.2575 \text{ kg m}^{-2}$ ) are discarded. The process flow is similar for F08, except that a cloud liquid water mask is NOT applied, as the level of noise in the 85 GHz channel is too high to produce an accurate cloud liquid water mask (W. Berg, personal communication, 2014). For F08, the remaining  $T_b$  data are then weighted according to the neural network weights appropriate to the individual satellite. For satellites F10-F18, pixels are segregated by sky condition: clear sky when CLW is less than  $0.0257 \text{ kg m}^{-2}$ ; cloudy sky when CLW is greater than or equal to  $0.0257 \text{ kg m}^{-2}$ . This yields one neural network for F08 (nnCLRCLD\_F08NoC85.mat) and two neural

networks each for F10-F18 (nnCLR\_F10.mat, nnCLD\_F10.mat, nnCLR\_F11.mat, and so on). For details on the neural net weighting technique, see Section 3.4.1. The resulting files are swath-level, daily, for each satellite of Qa, Ta, and U10.

1) SEAFLUXNN\_Swath

NOTE: satellite F15 data are not used after 31-Jul-2006. See CSU SSMI documentation regarding RADCAL.

NOTE: satellite F17 data are not used after 31-Mar-2016. See NASA satellite status regarding sensor failure.

NOTE: satellite F08 uses a separate neural network because of noise in the 85 GHz channel. Consequently, CLW is not an input to the neural network for F08.

NOTE: satellites F10, F11, F13, F14, F15, F17, and F18 use separate neural networks.

NOTE: sky condition for each pixel is determined by CLW value: clear (CLR) for  $CLW < 0.0257 \text{ kg m}^{-2}$ , cloudy (CLD) for  $CLW \geq 0.0257 \text{ kg m}^{-2}$ .

NOTE: in addition to the brightness temperature, EIA (earth incidence angle) is also included as an input to the neural networks.

NOTE: output files include swath-level, daily estimates of air temperature, specific humidity, and wind speed.

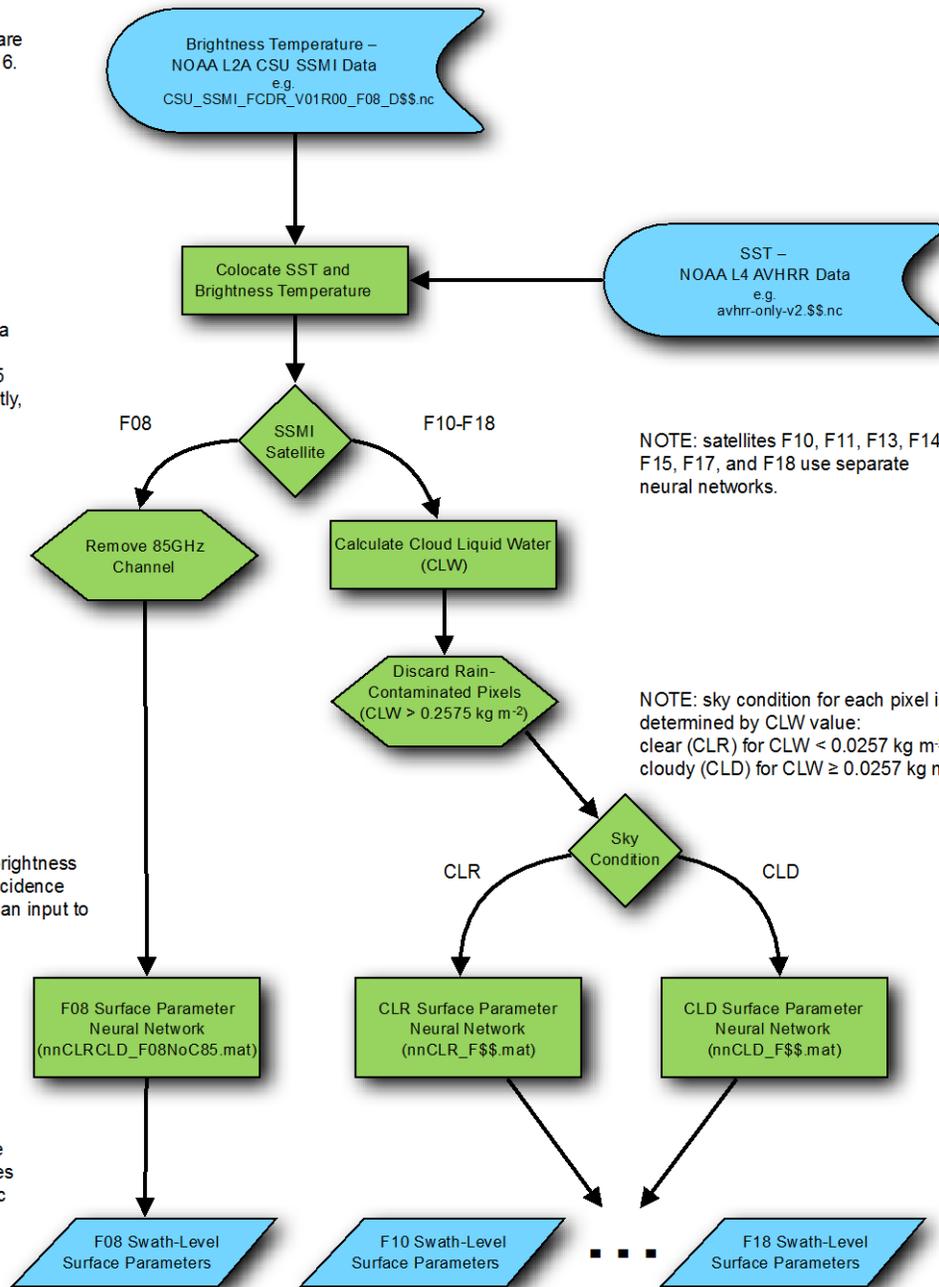
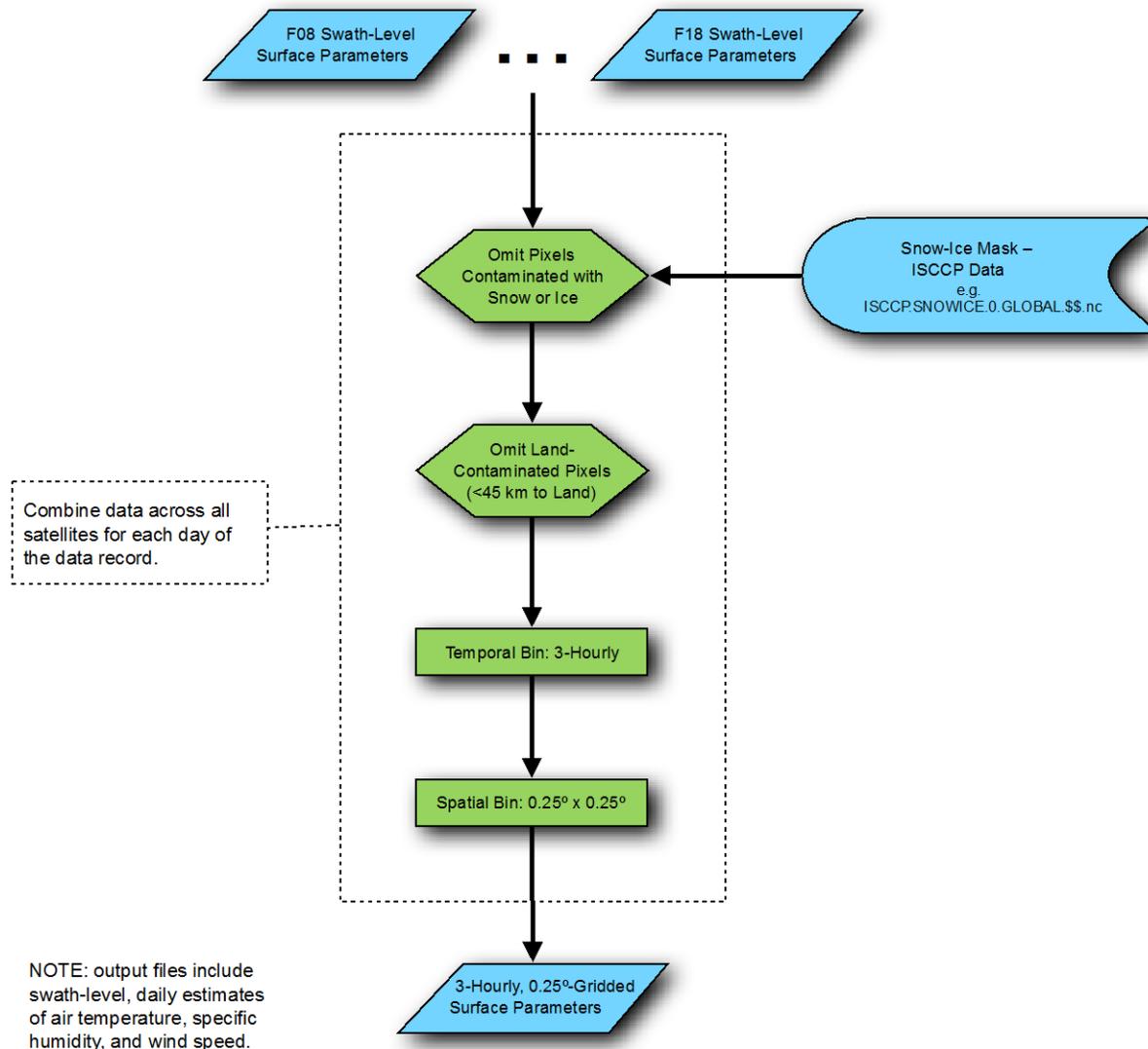


Figure 4: Processing flow for creation of swath-level fields of Ta, Qa, and U10.

The next step for the Qa, Ta, and U10 fields is to take all of the swath-level fields from each satellite and bin the appropriate data into 3-hourly bins. Two types of contamination are identified and the affected swath pixels are discarded: land-contaminated pixels (those within 45 km of land), and snow- or ice-contaminated pixels. A snow-ice-mask provided by the GEWEX. Additional swath pixels are also discarded for nonphysical values for Qa, Ta, or U10. (See Section 3.4.1 for description of nonphysical ranges.) An overview of this procedure is shown in Figure 5. Individual estimates are binned and averaged across all of the appropriate time, location, and satellites onto the 3-hourly temporal (first) and quarter-degree grid (second). The 3-hourly grid contains averages over the periods [00Z – 03Z], [03Z – 06Z], etc.

## 2) SEAFLUXNN\_Grid



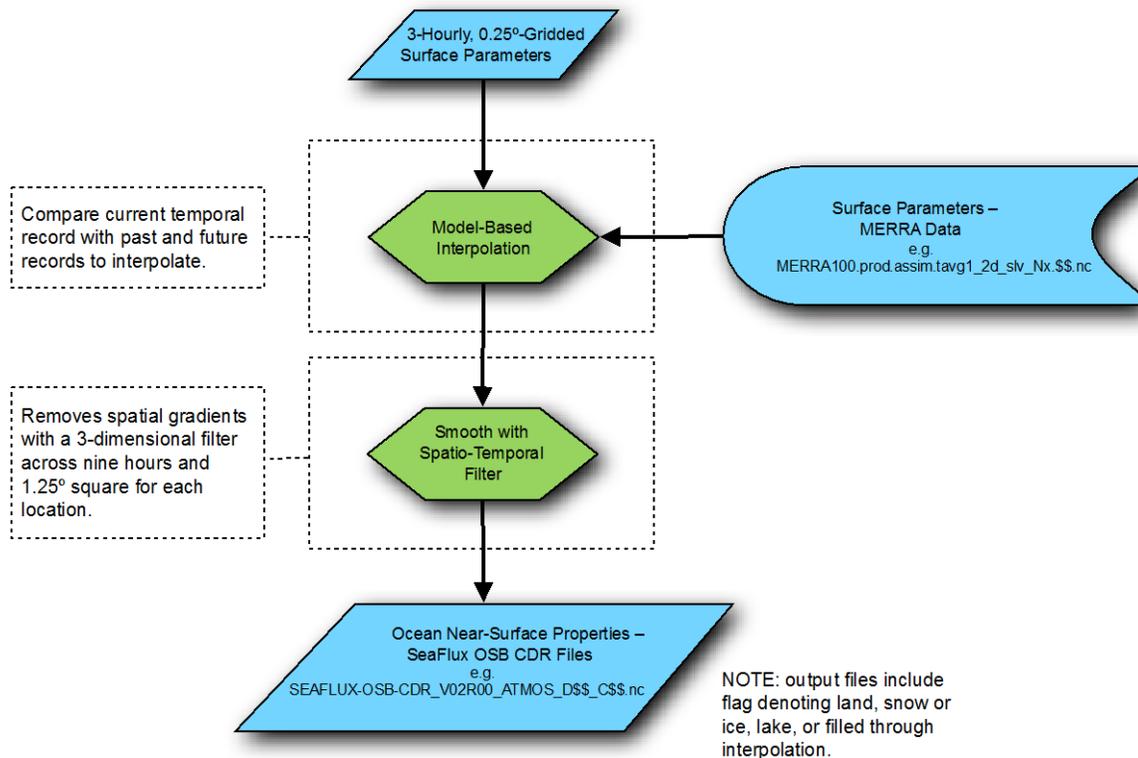
**Figure 5: Processing flow for creation of gridded fields of Ta, Qa, and U10 data.**

The initial gridding results in a large fraction of unobserved locations. This fraction changes as a function of the number of available satellites. The methodology used to provide a global open-ocean dataset is shown in Figure 6. The 3-hourly, gridded surface parameters are first interpolated using the Model-Based Interpolation (MoBI) procedure discussed in Clayson et al. (2014) and further described in Section 3.4.1. As with swath pixels in the previous step, cells filled with the MoBI procedure are discarded for

nonphysical values of  $Q_a$ ,  $T_a$ , or  $U_{10}$  as a quality control measure. (See Section 3.4.1 for description of nonphysical ranges.) Additionally, cells with high wind speeds ( $U_{10} > 45 \text{ m s}^{-1}$ ) that are near ice are flagged as ice and filled as missing to eliminate unverifiable values outside the range of training data for  $U_{10}$ .

After interpolation, to assist in reducing spurious edge gradients between known and analyzed locations due to this increased noise, a Gaussian weighted smoothing is applied across five spatial points and three temporal points. This post-processing results in final analyzed cells that have small departures from the directly observed satellite observation. This completes the processing of the surface parameter fields of  $T_a$ ,  $Q_a$ , and  $U_{10}$  that are part of the CDR, resulting in files called SEAFLUX-OSB-CDR\_V02R00\_ATMOS\_D\$\$\_C\$\$\_nc.

### 3) SEAFLUXNN\_Interp



**Figure 6: Processing flow for creation of final interpolated fields of  $T_a$ ,  $Q_a$ , and  $U_{10}$ .**

### 3.2.2 Production of SST

The base SST is the OISSTv2.0 as described above. The base SST is analogous to a foundation SST (Donlon et al. 2007), which is defined as the depth at which diurnal warming is not present. The next step is determination of the peak solar radiation (dSST) at every grid cell for the entire time period, which follows the processing flow shown in Figure 7. The peak solar radiation every day is determined from the GEWEX SRB Release-3.0 dataset through February 2000, the CERES-Syn1Deg-3H product is used from March 2000 through November 2015, and finally the NASA CERES FLASHFlux (Kratz et al. 2014) product is used for the near-real-time data. The mean daily precipitation is determined from the HOAPS precipitation data set through September 1996, and from the GPCP data set from October 1, 1996. The daily winds are calculated from the interpolated U10 fields described in Section 3.2.1, and a peak diurnal warming is calculated using the methodology described in section 3.4.6 (based on Clayson and Curry, 1996 and Clayson et al. 2014). The SST at each 3-hourly interval is then determined following the methodology shown in Figure 8. At each grid point, sunrise and sunset times have been generated from the Navy's sunrise/sunset algorithm ([http://aa.usno.navy.mil/data/docs/RS\\_OneYear.php](http://aa.usno.navy.mil/data/docs/RS_OneYear.php); see Section 3.4.5). The files are named SEAFUX-OSB-CDR\_V02R00\_SST\_D\$\$\_C\$\$\_nc.

1) SEAFLUXNN\_DiurnalSST

NOTE: GEWEX data are used for Jan-1988 through Feb-2000; CERES data are used thereafter—Syn 1deg product when available and FLASHFlux when not

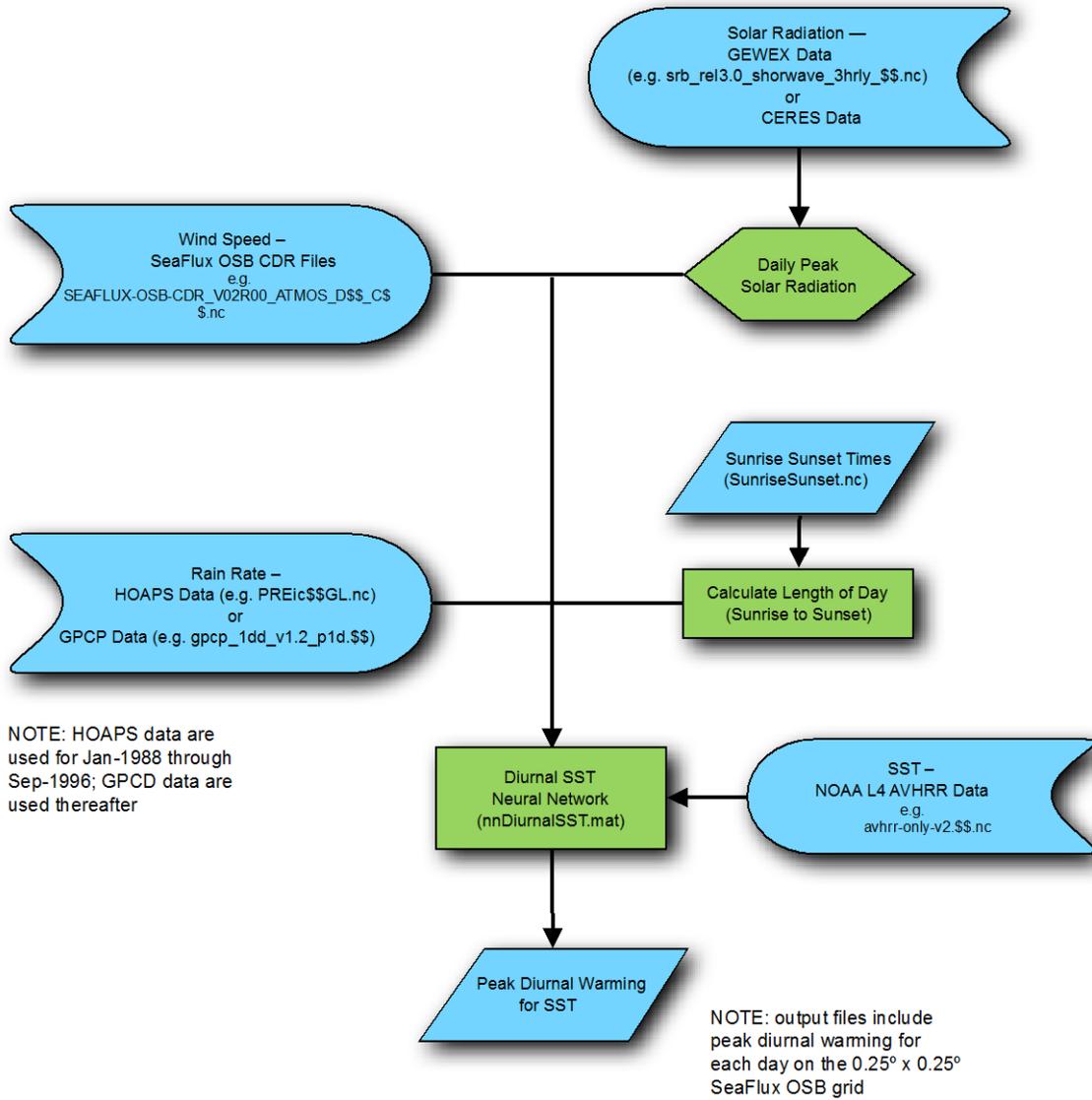


Figure 7: Processing flow for creation of peak diurnal warming SST values.

2) SEAFLUXNN\_SST

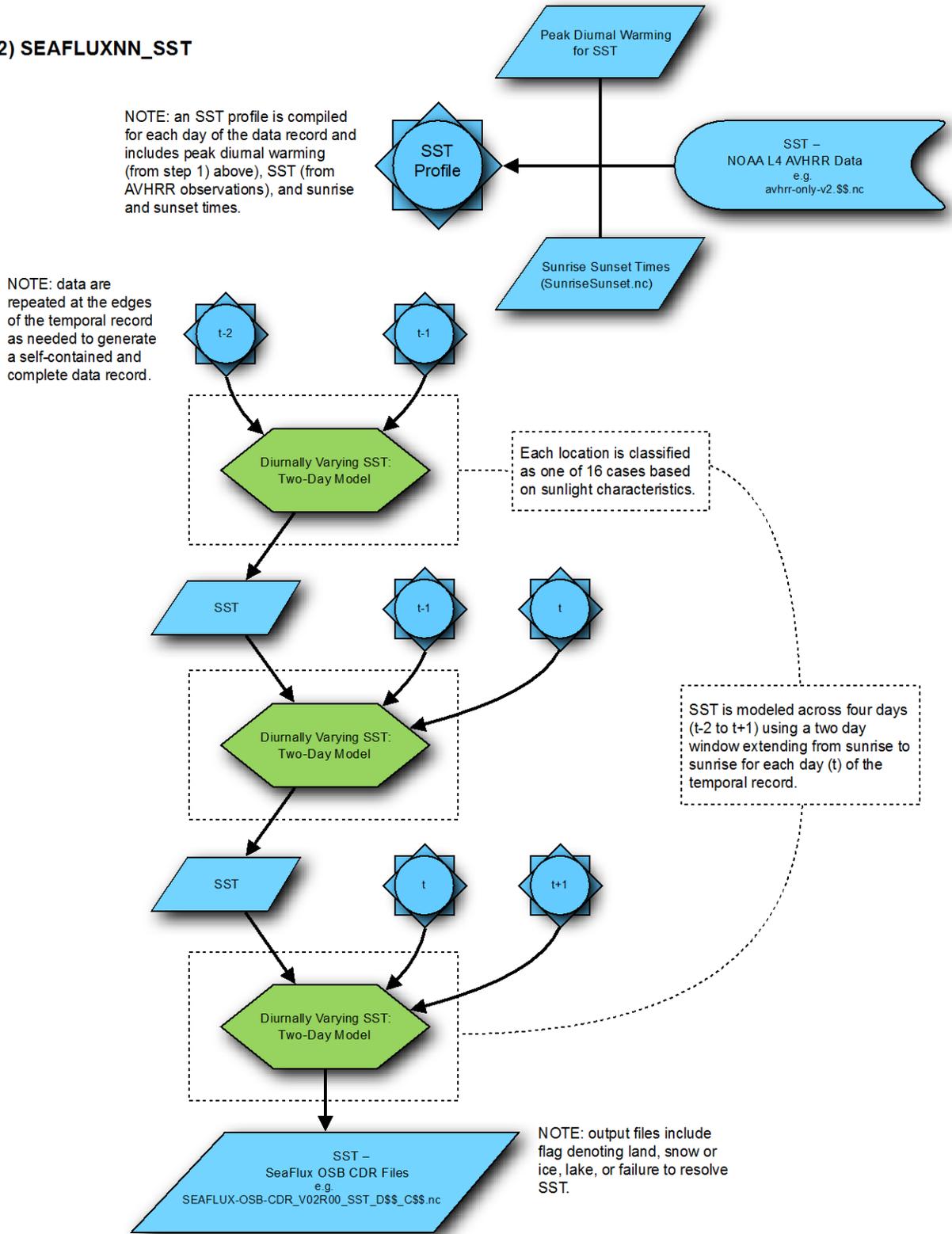


Figure 8: Processing flow for creation of final interpolated fields of SST.

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

### 3.2.3 Production of LHF and SHF

Once the  $Q_a$ ,  $T_a$ ,  $U_{10}$ , and SST fields have been created, the final step is to calculate the fluxes. The flux parameterization used here is a neural network version of the COARE 3.0 algorithm (Fairall et al. 2003) that has been optimized for speed of use for processing. Further details of this algorithm can be found in Section 3.4.1. Prior to input to the neural network,  $U_{10}$  data greater than  $45 \text{ m s}^{-1}$  in the open ocean are set to  $45 \text{ m s}^{-1}$  to eliminate unverifiable values occurring outside the training data range for  $U_{10}$ . See Section 3.4.1 for a description of the heat flux neural network.

Fluxes are available on the same grid and locations as the input fields of  $Q_a$ ,  $T_a$ ,  $U_{10}$ , and SST. The files are named SEAFLUX-OSB-CDR\_V02R00\_FLUX\_D\$\$\_C\$\$\$.nc.

## 3.3 Algorithm Input

### 3.3.1 Primary Sensor Data

One set of satellite data used in this algorithm is collected by SSM/I instruments on board the DMSP satellites F08, F10, F11, F13, F14, and F15. Beginning in March 2008, additional data are included from the SSMIS instruments on board the DMSP satellite F17, with F18 beginning to contribute data in March 2010. The brightness temperatures used in these retrievals based on the SSM/I and SSMIS instruments are from the Colorado State University (CSU) Fundamental Climate Data Records (FCDR). Full information for the brightness temperatures and the SSM/I and SSMIS instrument characteristics can be found in the C-ATBD for that dataset (CDRP-ATBD-0337, Rev. 1 07/11/2013).

Reynolds OISSTv2 sea surface temperature dataset is used both as a first-guess for the neural network retrievals of  $T_a$ ,  $Q_a$ , and  $U_{10}$ , and also as the pre-dawn (and nighttime) value of the SST. This is a CDR, the Daily  $1/4^\circ$  Optimum Interpolation Sea Surface Temperature (OISST) CDR. Details on the OISST CDR can be found in CDRP-ATBD-0303 (Rev. 2 09/17/2013).

### 3.3.2 Ancillary Data

Several ancillary datasets are used in the production of the final surface parameters.

Precipitation is used as an input to the calculation of the diurnal SST warming (see Section 3.4.1). The precipitation used for this dataset is the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS) data, version 3.2 until September 30, 1996. The data is distributed through the CM SAF Web User Interface at ([http://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=HOAPS\\_V001](http://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=HOAPS_V001)). The DOI for this dataset is 10.5676/EUM\_SAF\_CM/HOAPS/V001. The 6-hourly composites are used to create the daily averages. Details about the HOAPS dataset can also be found in Fennig et al.

(2012) and Andersson et al. (2010). From October 1, 1996 until the present, the GPCP one-degree daily precipitation data set (Huffman et al., 2001) is used.

An additional input to the diurnal SST warming calculation is a peak solar radiation value. This is obtained from the GEWEX Surface Radiation Budget (SRB) Release-3.0 data set until February 2000. This data are available from [https://eosweb.larc.nasa.gov/project/srb/srb\\_table](https://eosweb.larc.nasa.gov/project/srb/srb_table). This dataset is described in Stackhouse et al. (2011), and detailed information about the data can also be found at [http://gewex-srb.larc.nasa.gov/common/php/SRB\\_about.php](http://gewex-srb.larc.nasa.gov/common/php/SRB_about.php). The CERES-Syn1Deg-3H product (Ed3A; CERES 2016) is then used from March 2000 through November 2015. Details about this dataset can also be found at [https://eosweb.larc.nasa.gov/project/ceres/syn1deg-day\\_ed3a\\_table](https://eosweb.larc.nasa.gov/project/ceres/syn1deg-day_ed3a_table). Due to the lag in production of the CERES-Syn1Deg-3H product, the NASA CERES FLASHFlux (Kratz et al. 2014) product is used for the near-real-time data, currently beginning December 2015.

The land mask used for this dataset is derived from the Global, Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG), v 2.3.0, produced by the NOAA National Geophysical Data Center, retrieved 04/2014, available from <http://www.ngdc.noaa.gov/mgg/shorelines/shorelines.html>, and as described by Wessel and Smith (1996).

The ice mask used for this dataset is derived from the ISCCP Snowice dataset, available from GEWEX. The data are daily sampled on a 0.25°, equal-area grid and are available from <http://noaacrest.org/rscg/Products/GGEWC/snowice.html>. The version used here has a release date of December 26, 2014. The data were mapped to the SeaFlux OSB CDR grid. The Snowice dataset runs through the end of 2013. After 2013, an ice mask based on the daily-averaged-ice mask for the last four years of available data (2010-2013) is employed.

### **3.3.3 Derived Data**

Not applicable.

### **3.3.4 Forward Models**

Not applicable

## **3.4 Theoretical Description**

The software developed for the OSB CDR processing is a stepwise approach described in Section 3.2. Section 3.4.1 discusses the theoretical basis of the different techniques used to create the final CDRs.

### 3.4.1 Physical and Mathematical Description

**Neural network estimation of Ta, Qa, and U10.** The general form of equations for a neural network can be written as

$$y_h = \tanh\left(\sum_{i=1}^{NIN} w_{ih} x_i + \alpha_h\right) \quad (11)$$

$$z_k = \sum_{h=1}^{NHID} w_{hk} y_h + \beta_k \quad (12)$$

where  $y_h$  is the value of a hidden neuron,  $x_i$  is the input of the  $i^{\text{th}}$  input neuron,  $z_k$  is the output of the  $k^{\text{th}}$  output neuron,  $NIN$  is the total number of inputs,  $NHID$  is the total number of hidden neurons,  $w_{ih}$  is the weight between the  $i^{\text{th}}$  input and  $h^{\text{th}}$  hidden neuron,  $w_{hk}$  is the weight between the  $h^{\text{th}}$  hidden neuron and the  $k^{\text{th}}$  output neuron, and  $\alpha$  and  $\beta$  are bias weight parameters.

For this analysis, the Tbs from all satellites were treated separately without concern regarding consistency across instruments because of intercalibration between satellites (Sapiano and Berg, 2013). Doing so allows for special handling of individual satellites, as with F08 where the noisy 85 GHz channels were omitted, and reduces errors by allowing for segregation based on clear and cloudy sky conditions. Further, there are no significant processing penalties incurred from using separate neural networks for each satellite.

The training data for the neural network is from in situ measurements totaling several million samples. These in situ measurements are obtained from an extensive set of observations that make up the SeaFlux in situ dataset for F08-F15 and the ICOADS-value added dataset (IVAD) for F17 and F18.

The SeaFlux dataset contains observations made from multiple field campaigns over the period 1988 – 2007. Information on the SeaFlux in situ dataset can be found in Curry et al. (2004). The in situ measurements were recorded from several different platform heights, depending on the research vessel or buoy. Measurements consisted of numerous atmospheric parameters such as wind speed, air temperature, specific humidity, and sea surface temperature among others. In an effort to standardize this dataset, log layer profile adjustments are used to adjust all atmospheric parameters to a standard height of 10 meters. Warm-layer/cool-skin adjustments are made to the sea surface temperature so that all SSTs would reflect a true “skin” temperature for purposes of stability calculations. These adjustments are made using the COARE 3.0 algorithm (Fairall et al., 2003). Following typical procedures for neural network training, the input and target data were randomly divided into three subsets: training, validation, and testing.

IVAD data are used to train F17 and F18 because of limited overlap with the SeaFlux data record. F17 begins in March 2008, and F18 begins in March 2010, both after the SeaFlux data record has ended. IVAD observations offer the same heights-adjusted-

atmospheric parameters as the SeaFlux data set, meaning both SSM/I and SSMIS satellites are comparably trained. Information for the IVAD data set can be found in Berry and Kent (2009, 2011).

The neural network structure for each of the F10-F18 satellites includes 10 inputs (seven Tb channels, SST, CLW, and earth incidence angle or EIA) and five targets (SST, Ta, Qa, U10, and PW) that are modeled using two hidden layers. For each satellite, inputs and targets are grouped depending on sky conditions—either clear (CLR) or cloudy (CLD)—depending on the value of CLW: values less than  $0.0257 \text{ kg m}^{-2}$  for CLR and all other values for CLD. The number of neurons in each hidden layer varies for the specific satellite and sky conditions resulting in two neural networks for each of the F10-F18 satellites.

Satellites F17 and F18 make use of the 91 GHz channel of the SSMIS instruments instead of the 85 GHz channel on SSM/I instruments. This change requires that the vertical and horizontal 91 GHz channels are mapped to 85 GHz equivalents prior to calculation of CLW (Yang and Weng, 2008). The mapping is only done to calculate CLW, and Tb data from the 91 GHz channels are used otherwise.

The neural network structure for the F08 satellite includes six inputs (five Tb channels and SST—both 85 GHz channels are omitted along with CLW), the same five targets as the F10-F18 neural network, and two hidden layers. Segregation based on sky conditions, as was done for the F10-F18 satellites, is not possible because CLW calculation depends on the horizontal 85 GHz channel that is omitted due noise. The result is a single neural network for the F08 satellite.

Rather than apply a normalization for the Earth Incidence Angle (see Berg et al. 2013 for a description of the causes and variability of the EIA; Hillburn and Shie 2011), the EIA is used as an input to the neural network.

Training was completed using the Matlab neural network toolbox. The final network structure was chosen based on optimization of error characteristics across nearly 1800 different combinations of hidden layers and neurons, while considering training data for each satellite and both sky conditions (except for F08) separately. Thus all neural networks have different weightings and number of neurons in their hidden layers.

Output from each neural network are compared to nonphysical values for each of the five targets, with values judged unrealistic filled as missing. Any values outside the following ranges are deemed nonphysical:  $[-2, 35] \text{ }^\circ\text{C}$  for SST,  $[-90, 55] \text{ }^\circ\text{C}$  for Ta,  $[0, 100] \text{ g kg}^{-1}$  for Qa,  $[0, 100] \text{ m s}^{-1}$  for U10, and  $\text{PW} < 0$ .

**Flux parameterization.** The flux parameterization used here is a neural network version of the COARE 3.0 algorithm (Fairall et al. 2003) that has been optimized for speed of use for processing. The COARE 3.0 algorithm (and the resulting neural network algorithm) use the standard bulk formulae for calculation of LHF and SHF:

$$LHF = \rho C_e L_v (Q_s - Q_a) U_{10} \quad (13)$$

$$SHF = \rho C_h C_p (T_s - T_a) U_{10} \quad (14)$$

where  $\rho$  is the air density,  $C_e$  and  $C_h$  are the moisture and heat exchange coefficients, respectively,  $L_v$  is the latent heat of vaporization,  $C_p$  is the specific heat capacity of water,  $Q_s$  is surface specific humidity,  $Q_a$  is air specific humidity,  $T_s$  is sea surface temperature,  $T_a$  is air temperature, and  $U_{10}$  is the 10 meter wind speed. Positive fluxes are denoted as from the ocean to the atmosphere. The COARE 3.0 algorithm uses the input variables of SST,  $Q_a$ ,  $T_a$ , and  $U_{10}$  to calculate the moisture and heat exchange coefficients. Technically, the wind term should be  $U_{10} - U_s$  (with  $U_s$  being the surface current); in this dataset we assume  $U_s = 0$ .

The neural network was trained using turbulent latent and sensible heat fluxes computed from the COARE 3.0 algorithm for a range of input wind speeds of 0 to 45 m s<sup>-1</sup>, near-surface specific humidities of 0 to 30 g kg<sup>-1</sup>, saturation specific humidities of 0 to 30 g kg<sup>-1</sup>, sea surface temperatures of -2 to 35 °C, and near-surface air temperatures of -30 to 40 °C. The combined parameter space was sampled using a Latin hypercube to generate random samples that are also uniformly distributed across the entire sample space. The resulting latent and sensible heat fluxes span ranges of -7000 W m<sup>-2</sup> to 7100 W m<sup>-2</sup> and -3100 W m<sup>-2</sup> to 6500 W m<sup>-2</sup>, respectively.

Neural network emulation was tested against fluxes evaluated directly from COARE reproducing the latent heat fluxes with a Gaussian error distribution, a correlation coefficient of 1.000, a bias of 0.00037 W m<sup>-2</sup> and a root-mean-square error of 0.103 W m<sup>-2</sup>. The sensible heat flux also has a correlation coefficient of 1.000, a bias of 0.00003 W m<sup>-2</sup>, and a root-mean-square error of 0.049 W m<sup>-2</sup>. These errors in emulation are less than the stated accuracies of the COARE 3.0 algorithm.

The neural network emulator for the COARE model was also trained using the Matlab neural network toolbox. The final structure for the five input, two output neural network—two hidden layers with 83 neurons in the first and 20 in the second—was found through optimization of error characteristics across nearly 1800 different combinations of hidden layers and neurons.

The heat fluxes produced by the neural network are limited to [-50, 500] W m<sup>-2</sup> for latent heat flux and [-300, 1500] W m<sup>-2</sup> for sensible heat flux. Values outside this range are considered unrealistic and are thus filled as missing.

### 3.4.2 Data Merging Strategy

The Qa, Ta, and U10 fields are determined based on the neural network retrieval strategy at the satellite swath level. The data is binned into 3-hourly bins, collecting all data from all available satellites into the quarter-degree grid. An overview of this procedure is shown in Figure 5. These individual estimates are then averaged across all of the appropriate time, location, and satellites onto the 3-hourly temporal (first) and quarter-degree grid (second). The 3-hourly grid contains averages over the periods [00Z – 03Z], [03Z – 06Z], etc.

### 3.4.3 Numerical Strategy

**Model-Based Interpolation technique.** The Model-Based Interpolation (MoBI) procedure used for interpolation in this dataset is based on the following methodology. This dataset uses version two of the Modern Era Retrospective-analysis for Research and Applications (MERRA-2; Bosilovich 2008) to determine the model-based tendencies that are integrated. The tendencies are produced from consecutive model time steps that take into account the state of the atmosphere and physical and dynamical processes such as radiation and advection. The tendencies are weighted such that the satellite-based observations are matched exactly when available. That is, MoBI is an exact interpolation scheme, tied to the satellite retrieval estimates. The method relies on knowledge at two samples in time of both satellite and model estimates bracketing a single or series of unobserved time steps. Let these two known points in time at the beginning and end of a series be denoted  $t=A$  and  $t=B$ , respectively. Then the difference between the satellite (S) and model (M) estimates,  $\Delta$ , at A and B can be defined as

$$\Delta_A = S_A - M_A$$

$$\Delta_B = S_B - M_B.$$

Forcing the estimated field, X, at the end points to match the satellite observation exactly, the analysis equations at these points become

$$X_A = M_A + \Delta_A$$

$$X_B = M_B + \Delta_B.$$

To estimate the unknown observations at time steps, t, between the known boundaries, the model difference between time t,  $M_t$ , and the model boundary,  $M_A$  or  $M_B$ , can be used to estimate the intermediate point in the forward,

$$X_t = X_A + [M_t - M_A] = M_t + \Delta_A$$

or backward

$$X_t = X_B - [M_B - M_t] = M_t + \Delta_B$$

directions. As shown, the formalism of adding model-driven time evolution to known satellite observations is equivalent to bias-adjusting the model evolution at both the

boundary and unobserved samples. While the model estimates are fully consistent between the model boundary and intermediate samples, the adjustment of the boundaries to the satellite values can result in a mismatch at the boundaries of the forward and backward steps. To smoothly merge between points, a convex combination is used that linearly interpolates over the time interval between the known boundary points as

$$X_t = M_t + \Delta_A + (n/N \Delta_B - n/N \Delta_A) = M_t + (1-f) \Delta_A + (f) \Delta_B$$

where  $n$  is the number of samples from the known time  $A$  over the total number of samples,  $N$ , encompassed from  $A$  to  $B$ . That is,  $f = n/N$  represents the normalized (i.e. fractional) time over the interval. The final analysis equation results in an exact interpolation at the boundaries to the retrieved satellite quantity. For each time sample,  $A$  and  $B$  were limited to 60 days, i.e. the MoBI procedure searched 60 days before and 60 days after for model and observation data to use for interpolation. As formulated, no assumptions of stationarity or estimation or error covariances or structure functions are needed. The only weighted averaging of MoBI occurs through the weighted average of the bias-adjustment to be applied at every time step, which results in an exact interpolation to the satellite observations at each time a satellite observation is available and to a nearly unbiased estimate (with respect to the satellite observations) at intermediate analyzed estimates. Further details can be found in Clayson et al. (2014).

Any nonphysical realizations (see Section 3.4.1) returned by the MoBI procedure are filled with missing and flagged appropriately as having failed interpolation.

### 3.4.4 Calculations

Details on the processing steps involved in the algorithm are provided in Section 3.2.

### 3.4.5 Look-Up Table Description

Only production of SST uses data that has been calculated and stored in a static look-up table. The time of sunrise and sunset for each location of the  $0.25^\circ$  output grid are stored in the NetCDF SunriseSunset.nc. Sunrise and sunset times are used first to calculate the daily peak solar radiation for the peak diurnal warming (Figure 7), and then to produce the two-day model of diurnal variability leading to the final SST CDR (Figure 8).

### 3.4.6 Parameterization

**Sea surface temperature diurnal warming.** The peak diurnal warming for each grid location at each time step is determined by the following methodology. The one-dimensional second moment turbulence closure ocean model of Kantha and Clayson (1994, 2004), modified to include a parameterization of the highly stable surface layer. The model includes the skin surface temperature parameterization of Schluessel et al. (1997), the effects of precipitation, and the effects of the diurnal thermocline. To make use of this skin model, the uppermost level of the model is set at 1 cm. The vertical resolution is then set to every 0.1 m over the upper 50 m. The temporal resolution is 15 minutes.

This model has been run for 100,000 four-day simulations with variable inputs of (a) peak solar radiation with values ranging from 0 to 1300 W m<sup>-2</sup>; (b) daily mean precipitation with values ranging from 0 to 1.0 x 10<sup>-5</sup> m s<sup>-1</sup>; (c) lengths of day with values ranging from 0 to 24 hours; (d) wind speeds ranging from 0 to 10 m s<sup>-1</sup>; and (e) SSTs ranging from -3 to 38 °C. The combined parameter space was sampled using a Latin hypercube to generate random samples that are also uniformly distributed across the entire sample space. The diurnal warming for each of these simulations is taken from the fourth day of simulations as the difference between the maximum and minimum values. This database of 100,000 dSST values and their corresponding winds, solar radiation, precipitation, SST, and length of day are then provided as input to a neural network emulator. This methodology was used by Clayson and Curry (1996) and Clayson and Weitlich (2007) and more details can be found in these references.

Neural network emulation was tested against dSST evaluated directly from the original model reproducing the peak diurnal warming with a correlation coefficient of 0.9562, a bias of -0.0071 °C, and a root-mean-square error of 0.2819 °C. These errors in emulation are acceptable for the model

Peak diurnal warming data come from running the neural network emulator with inputs from (a) GEWEX SRB for peak solar radiation; (b) HOAPS for daily mean precipitation; (c) the sunrise/sunset lookup-table in the file SunriseSunset.nc for length of day (see Section 3.4.5); (d) SEAFLUX OSB CDR Ocean Near-Surface Properties for daily average wind speed; and (e) Reynolds OISSTv2 for SST. Peak solar radiation were found through three-dimensional interpolation of the 3-hourly GEWEX data to the 0.25° output grid for 24 hours. The maximum value at each location for each day was input to the neural network as the peak solar radiation. Daily mean precipitation were interpolated to the 0.25° output grid prior to input to the neural network.

Output of the final SST product are compared to nonphysical values as before with the atmospheric parameters. Any values outside of [-2, 35] °C are judged unrealistic and filled as missing.

### 3.4.7 Algorithm Output

The primary output fields are the 3-hourly, 0.25° x 0.25° Qa, Ta, U10, SST, and LHF and SHF fields. The NetCDF filenames are SEAFLUX-OSB-CDR\_V02R00\_{ATMOS, SST, FLUX}\_D<YYYYMMDD>\_C<YYYYMMDD>.nc, where D<YYYYMMDD> is the date of the data contained in the file and C<YYYYMMDD> is the create date of the file. The total file size for all three data products is 593 gigabytes and can be downloaded from <http://www.ncdc.noaa.gov/cdr/operationalcdrs.html>.

## **4. Test Datasets and Outputs**

### **4.1 Test Input Datasets**

None.

### **4.2 Test Output Analysis**

#### **4.2.1 Reproducibility**

None

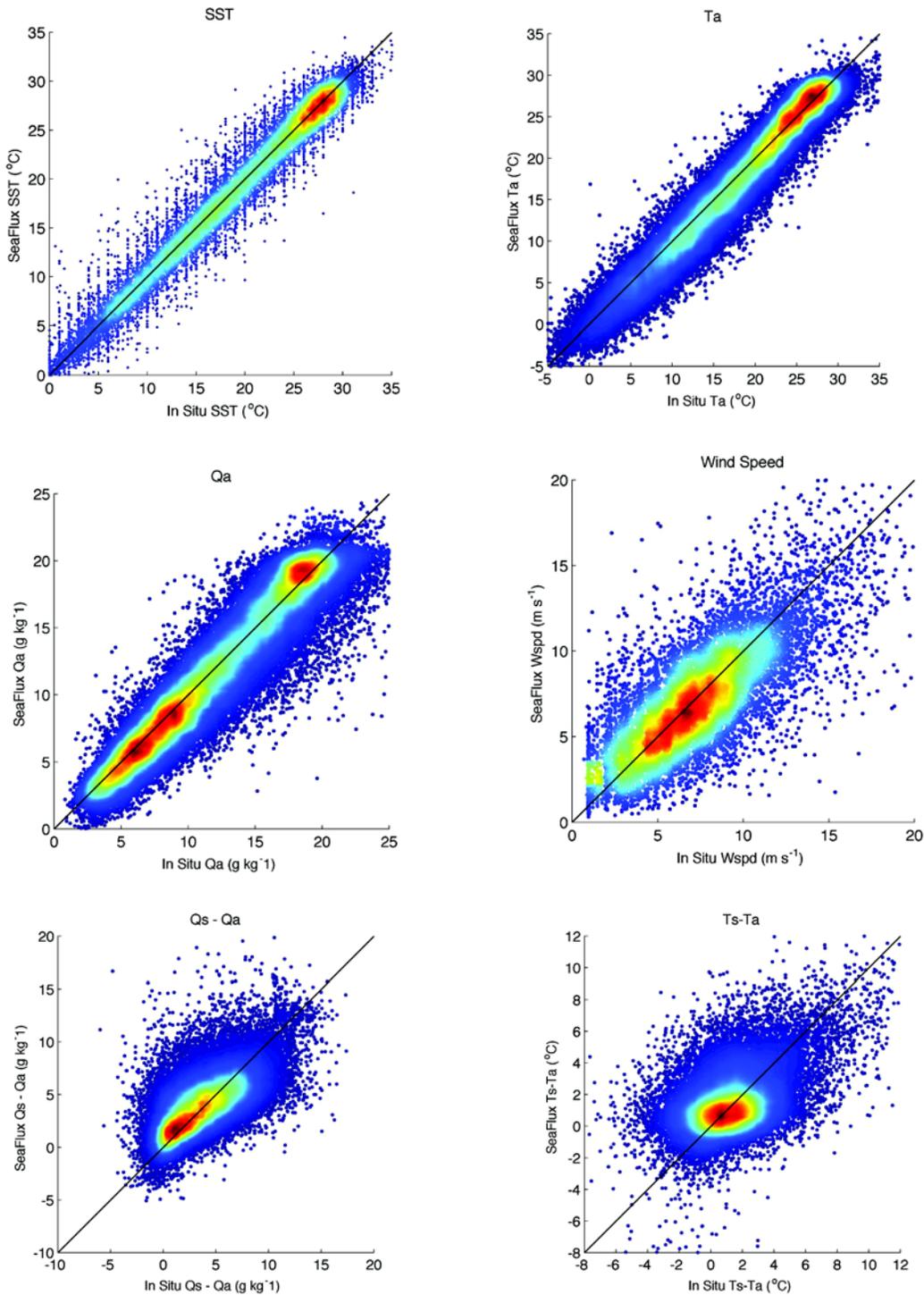
#### **4.2.2 Precision and Accuracy**

See Section 4.2.3 for all estimates of errors.

#### **4.2.3 Error Budget**

In a previous version of this dataset, SeaFlux V1.0, a significant effort was made to estimate uncertainties by comparison with in situ flux measurements from research vessels and the bulk parameters from ships of opportunity data sets. This methodology and results are outlined here. The version of the dataset described in this C-ATBD, the SeaFlux OSB CDR Version 1, retains nearly all of the original methodology with the following exceptions: (1) incorporation of additional SSM/I satellite data to extend the dataset back to 1988 with resulting changes in weights to the neural nets for the retrieval of the swath-level  $T_a$ ,  $Q_a$ , and  $U_{10}$ ; (2) the use of the SeaFlux OSB CDR Version 1 winds for  $U_{10}$  and as input to the fluxes instead of the Cross-Calibrated Multi-Platform Ocean Surface Wind Vectors (CCMP) CCMP winds to provide internal consistency to the dataset; (3) re-calculated weighting for the neural network algorithm of the COARE 3.0 resulting in reduced errors in the extreme flux values; (4) an improved ice mask instead of the Reynolds ice mask; and (5) re-calculated weighting for the diurnal SST parameterization to provide better error characteristics relative to the original model. The errors shown in this version of the C-ATBD relate to the SeaFlux V1.0 dataset, and further details of that dataset and error analysis is discussed in Clayson et al. (2014). There are several potential sources of uncertainty of the estimation process. The first is the random uncertainty, whose influence in the total uncertainty can be reduced by increasing the size of the space/time average. Another is the systematic uncertainty, which does not reduce upon averaging data. Both of these uncertainties are calculated for the dataset and described here, and are summed for the total uncertainty. A further type of uncertainty is the sampling uncertainty. However, the comparisons that are used for calculating the errors use the interpolated data set, and as such represent some of the effects of the lack of sampling the full globe by the satellite sensors. Other sampling issues could include variability that occurs on a less than three-hourly time scale, which is not being resolved by the satellite dataset. These errors would not affect the instantaneous fields, but could possibly reflect on the daily and longer time scale averages.

In order to understand the error characteristics, a set of comparisons is performed with the ship-of-opportunity International Comprehensive Ocean-Atmosphere Data Set (ICODS)-value added dataset (IVAD) based upon the ICODS data containing adjustments for height and known biases (Berry and Kent 2009, 2011). The IVAD data contains only measurements of the bulk parameters; no direct covariance flux estimates are provided, so that the comparisons here are limited to the bulk parameters only. This dataset is provided at the time and location of measurement. For these comparisons the IVAD dataset are binned into the SeaFlux 3-hour bins and the SeaFlux grid. The number of data points available for comparison with the IVAD data set is roughly  $\sim 5.07 \times 10^6$ . Comparisons of the SeaFlux V1.0 dataset with the available IVAD in situ data are shown in Figure 9.



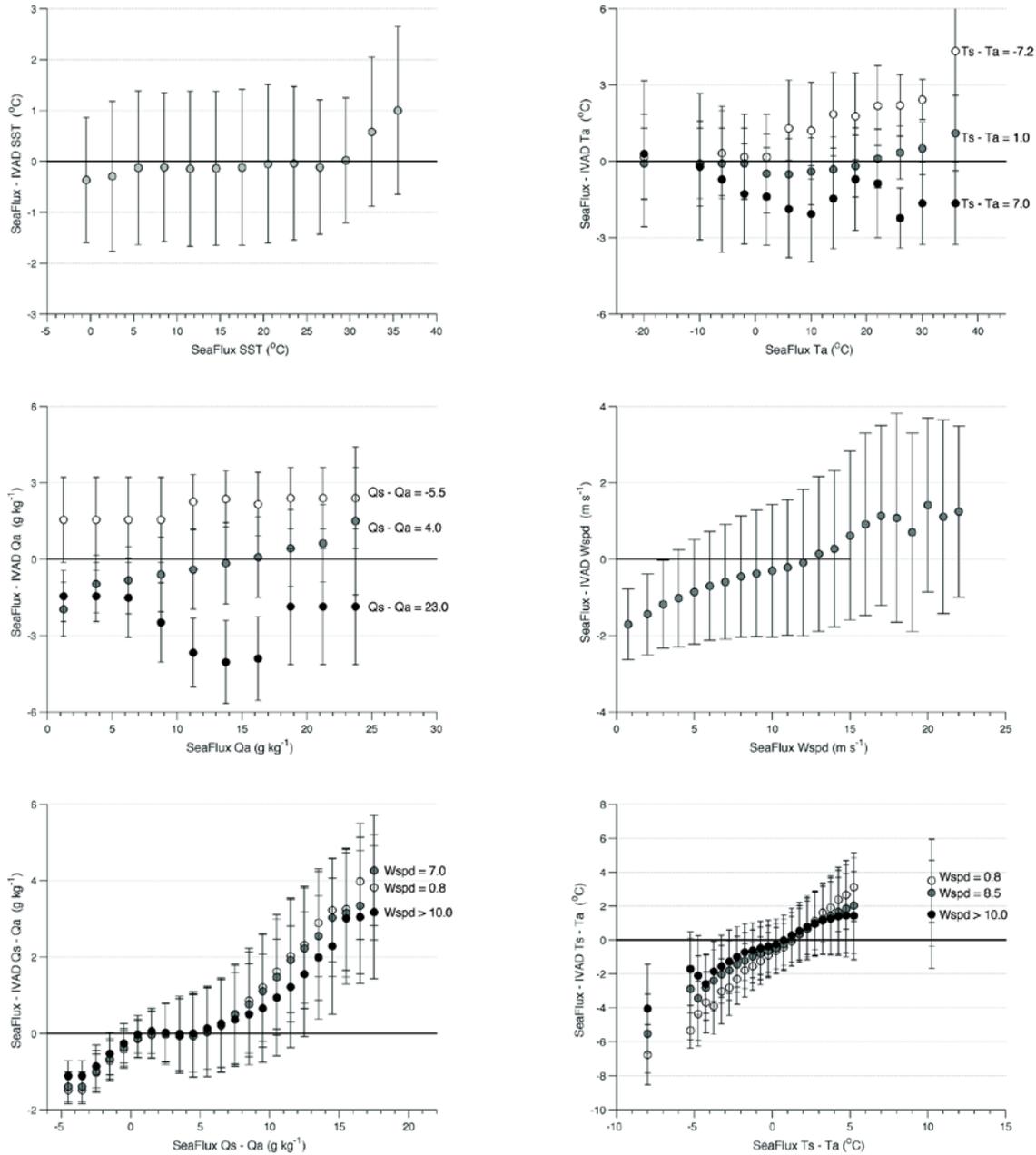
**Figure 9: Comparisons between SeaFlux satellite-derived dataset and IVAD data for the component fields.**

Colors indicate density of data (varying linearly from dark blue representing less than 6000 points to red indicating 60000 or more points). line indicates 1:1

The systematic and random components of the uncertainty for each of the individual components of the fluxes (wind speed, air temperature and humidity, and SST) are calculated first. In order to determine the bias and standard error of each gridded value at each 3-hour time step, all of the matchup data is binned and errors are calculated for each bin. Binning is performed based on the satellite observation of that value, rather than the IVAD, so that an error can be calculated using solely each satellite observation. The error characteristics of most of the components were at least weakly a function of two parameters. For air temperature, error characteristics depended on the air temperature and the air-sea temperature difference. The specific humidity error was a function of both near-surface specific humidity and the air-sea humidity difference. The wind speed errors were correlated with both the estimated wind speed and the air-sea temperature difference. Only the SST errors were computed as solely a function of the SST itself. The systematic uncertainty is here defined as the bias of the satellite observation within each bin, and the random uncertainty is here defined as the standard error within each bin.

The bin intervals were chosen based on the variability of the error characteristics and the amount of data available for that bin. Thus, the bin intervals are very large at the tails of the distributions, such that each bin has at least 150 samples. For example, to attain 150 samples for all values of air-sea temperature differences as a function of wind speed, the highest wind speed bin contains all wind speeds greater than  $22 \text{ m s}^{-1}$ . In addition, for the further calculation of the uncertainties for the fluxes, the air-sea temperature and humidity difference uncertainties were also determined. The air-sea temperature difference error was a function of both the temperature difference and the wind speed; similarly the air-sea humidity difference error was a function of both the air-sea humidity difference and the wind speed. Once the standard deviations for each bin were determined, values greater than three standard deviations from the mean are tagged as outliers and removed from the analysis. The mean bias and standard deviations are then re-calculated. This step is performed in order to provide a more consistent RMS error amongst the bins, in order to remove the types of statistical artifacts as shown by Kent et al. (1998) in comparing satellite and ship winds.

Error characteristics for a representative sample of the ranges for each component are shown in Figure 10. The SST error characteristics are fairly similar across the SST domain, although there is a definite trend towards overestimation of the SST as the SST increases. The SSTs in the IVAD dataset are unadjusted for depth, measurement methodology, or time of day. During nighttime, the difference between a skin temperature and a measurement at some nominal depth in the upper ocean will be relatively close as the ocean is typically substantially mixed (except for the skin effect; Fairall et al. 1996), whereas during the daytime depending on the depth of the measurement there can be a difference of up to several degrees due to diurnal warming. An analysis of the errors at SSTs greater than  $30 \text{ }^{\circ}\text{C}$  shows that mean nighttime differences are only a few tenths of a degree; during the late morning through early evening however the mean differences are roughly  $1.5 \text{ }^{\circ}\text{C}$ . This variability is much reduced at colder temperatures, which may see mean differences of  $0.1 \text{ }^{\circ}\text{C}$  or less across the daytime/nighttime comparisons. For the error analysis the bias value as shown in Figure 10 is kept though it is most likely more conservative than the actual error.



**Figure 10: Differences between the SeaFlux v1.0 satellite dataset and the IVAD measurements as a function of the satellite values. Circles indicate the mean bias in that bin, with standard deviations denoted by the bars.**

The trend in wind speed with increasing wind speed is more apparent. The decrease in standard deviation values at low wind speeds is unsurprising given the hard limit at  $0 \text{ m s}^{-1}$ . The increased error at low wind speeds when comparing satellite winds to ship observations is common; at such low wind speeds convective cells can dominate surface winds with much smaller footprints than the satellite footprint. It should be noted that there are very few (122) wind speeds below  $1 \text{ m s}^{-1}$  in the IVAD dataset available for matchup. This has been noted by others performing similar comparisons (e.g. Mears et al. 2001) and is not necessarily indicative of an actual increased error in the wind fields. Given the trends shown in the data, it is likely then that the magnitude of the biases are overestimated at the low wind speeds and underestimated at very high winds, but without additional data this is the best definition that can be provided.

Error characteristics of both the air temperature and specific humidity were dependent on the retrieved air temperature and humidity and also the temperature and humidity near-surface vertical difference. The  $Q_a$  errors for the bin of humidity difference that contained the mean humidity difference value ( $Q_s - Q_a = 4.0 \text{ g kg}^{-1}$ ) are much less than  $1 \text{ g kg}^{-1}$  for  $Q_a$  values between roughly 5 and  $25 \text{ g kg}^{-1}$ ; as  $Q_a$  drops to zero the error increases, as it does at the  $Q_a$  value increases above  $23 \text{ g kg}^{-1}$ , a very rare occurrence. To the extent that on average the humidity increases with the SST, and given that the SST is an input to the humidity retrieval, it is possible that some of this trend is a result of the similar trend towards overestimation of SST at high SSTs. Little or no trend is apparent in the  $Q_a$  retrieval at the more extreme values of  $Q_s - Q_a$ , which may reflect conditions under which the atmosphere and ocean are less coupled and the input SST had less weight in the retrieval. It should be noted that the error characteristics shown for the  $Q_s - Q_a$  values of  $-5.5 \text{ g kg}^{-1}$  and  $23.0 \text{ g kg}^{-1}$  are well within the tails of the humidity difference distribution and represent less than 0.01% of all the data.

The air temperature errors for the  $T_a$  values near their mean of  $1.3 \text{ }^\circ\text{C}$  have little trend over the temperature scale, which may be due to the fact there is a slight tendency for overestimation of the air temperature at the higher air temperatures (which may again be related to the overestimation of SST trends). Interestingly at air-sea temperature differences at the tails of those distributions the trends are opposite: for extremely stable conditions the retrieved  $T_a$  on average becomes increasingly overestimated, while for unstable conditions the retrieved  $T_a$  becomes increasingly underestimated for increasing  $T_a$ . As with the  $Q_a$  comparisons, these temperature differences represent less than 0.01% of the data, and are shown here to demonstrate the possible range of observed biases.

Similar analyses for the humidity and temperature differences are also shown. The wind speed dependence on the error characteristics of the humidity difference are small but increase as the difference increases. In general, the bias and standard deviation of the  $Q_s - Q_a$  error are within  $1 \text{ g kg}^{-1}$  until  $Q_s - Q_a$  becomes larger than about  $10 \text{ g kg}^{-1}$ . Then there is an increasing overestimation of  $Q_s$ , underestimation of  $Q_a$ , or both, as the difference continues to increase. SST errors have been shown to be on the order of  $1 \text{ }^\circ\text{C}$ , which at a temperature of  $35 \text{ }^\circ\text{C}$  corresponds to an overestimation of roughly  $2 \text{ g kg}^{-1}$ . Conversely,  $Q_a$  has been shown to be more susceptible to underestimation in regions of high  $Q_a$  due in part to saturation of the microwave channels (Roberts et al. 2010).

Over the 10 years of the SeaFlux version 1.0 dataset, the mean global total uncertainty in Ta is 0.35 °C, and the mean global total uncertainty in SST is 0.12 °C.

The wind speed uncertainties are highest in regions with chronically low winds, such as the tropical Indian and western Pacific Oceans, but remain everywhere below approximately 1 m s<sup>-1</sup>. Uncertainties are also higher in regions that experience periods of low wind speeds, as the uncertainty is higher at low wind speeds, even if the overall mean of the wind speeds are higher in these regions. The mean wind speed uncertainty is 0.39 m s<sup>-1</sup>. The specific humidity is below 1 g kg<sup>-1</sup> nearly everywhere, with exceptions in regions that have high Qs-Qa values through some or all of the year, most notably the northern Bay of Bengal and Arabian Seas and over the Kuroshio Current. These are two regions where numerous satellite products significantly underestimate Qa as found in Prytherch et al. (2013). The mean Qa uncertainty is 0.45 g kg<sup>-1</sup>.

The uncertainties in the fluxes are a result of the errors in the input data and errors in the physical model used to compute the turbulent fluxes from these mean values. Using basic sampling theory and propagation of errors (Taylor, 1982), the uncertainties in the fluxes for each grid point at each 3-hour time step take the form:

$$\sigma_F = \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y}\right)^2 \sigma_y^2 + 2r_{xy} \left(\frac{\partial F}{\partial x} \frac{\partial F}{\partial y}\right) \sigma_x \sigma_y}$$

where F is the flux, x and y are input variables, r<sub>xy</sub> is the correlation coefficient between x and y, and s<sub>x</sub> and s<sub>y</sub> are the total uncertainties in x and y, given by:

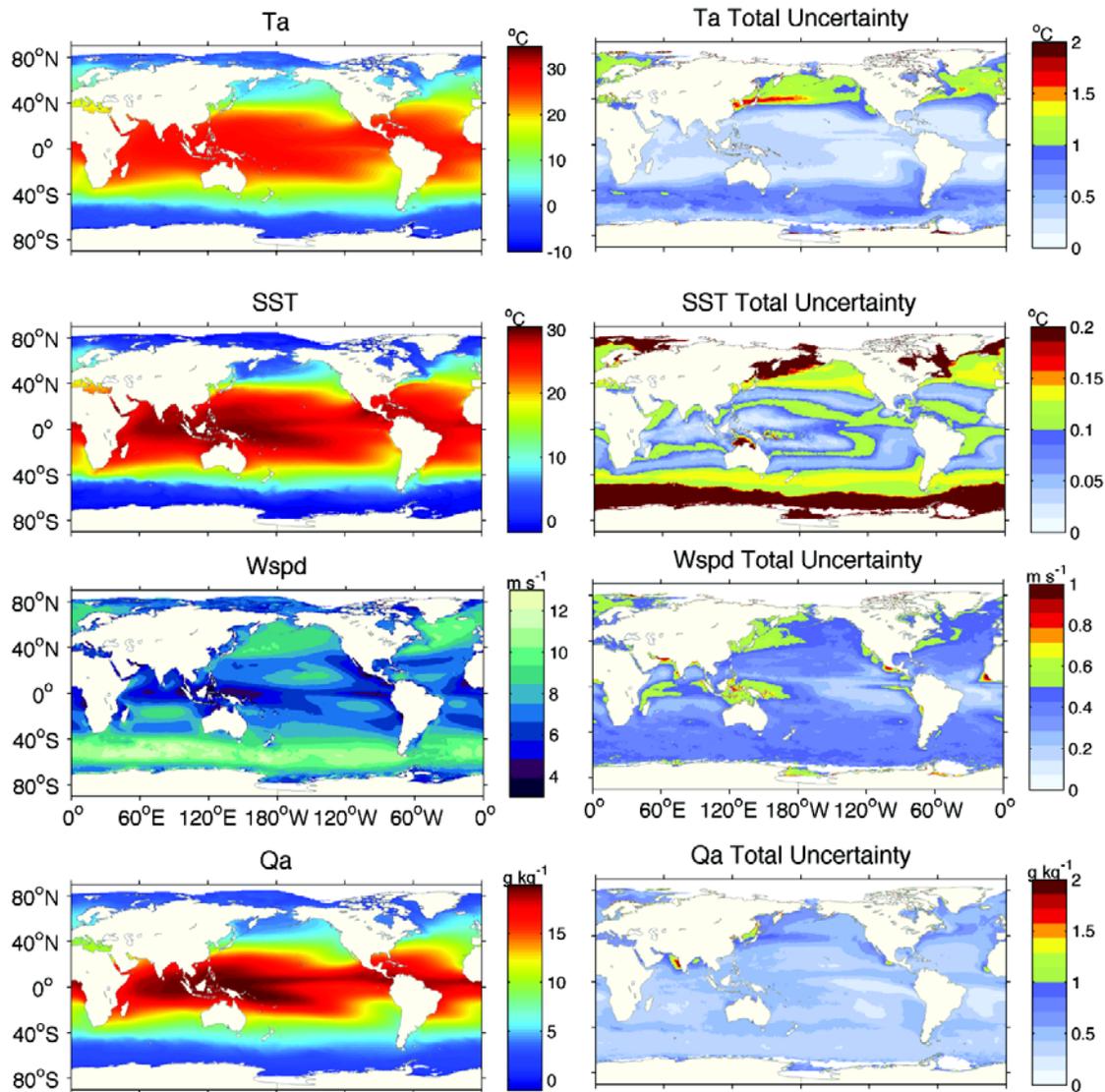
$$\left(\frac{\partial F}{\partial x}\right)^2 \sigma_x^2 \Rightarrow \left(\frac{\partial F}{\partial x}\right)^2 \sigma_{x,sys}^2 + \left(\frac{\partial F}{\partial x}\right)^2 \sigma_{x,ran}^2 [N^{-1/2}]^2$$

where sys and ran refer to the systematic and random components, respectively. N is the number of data points in the collection (for each 3-hour realization at each point is considered one observation). The input variables to the calculation of the LHF are: C<sub>E</sub> (the moisture transfer coefficient), U10 (the wind speed), and (Qs - Qa). The input variables to the calculation of the SHF are C<sub>H</sub> (the heat transfer coefficient), U10, and (SST-Ta). The total uncertainty for each of the fluxes is thus estimated as:

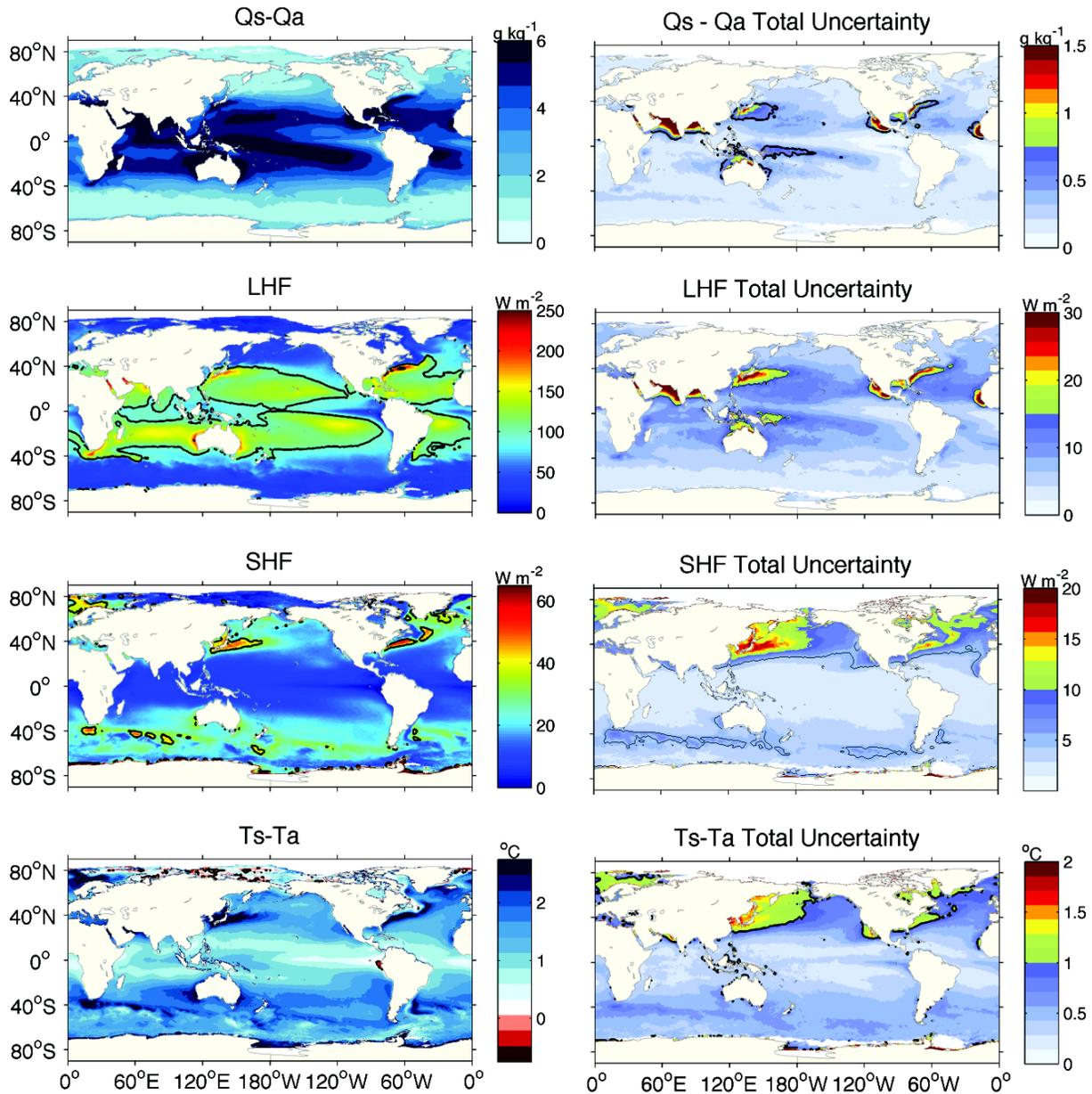
$$\begin{aligned} \sigma_{SHF} = & \left[ (\rho_a C_{pa} U (Ts - Ta) \sigma_{C_H, sys})^2 + (\rho_a C_{pa} C_H (Ts - Ta) \sigma_{U, sys})^2 + (\rho_a C_{pa} C_H U \sigma_{(Ts-Ta), sys})^2 + 2r_{(Ts-Ta), U} \left( (\rho_a C_{pa} C_H)^2 (Ts - Ta) U \sigma_{(Ts-Ta), sys} \sigma_{U, sys} \right) \right. \\ & \left. + (\rho_a C_{pa} U (Ts - Ta) \sigma_{C_H, ran})^2 + (\rho_a C_{pa} C_H (Ts - Ta) \sigma_{U, ran})^2 + (\rho_a C_{pa} C_H U \sigma_{(Ts-Ta), ran})^2 + 2r_{(Ts-Ta), U} \left( (\rho_a C_{pa} C_H)^2 (Ts - Ta) U \sigma_{(Ts-Ta), ran} \sigma_{U, ran} \right) \right]^{1/2} \\ \sigma_{LHF} = & \left[ (\rho_a L_v U (Qs - Qa) \sigma_{C_E, sys})^2 + (\rho_a L_v C_E (Qs - Qa) \sigma_{U, sys})^2 + (\rho_a L_v C_E U \sigma_{(Qs-Qa), sys})^2 + 2r_{(Qs-Qa), U} \left( (\rho_a L_v C_E)^2 (Qs - Qa) U \sigma_{(Qs-Qa), sys} \sigma_{U, sys} \right) \right. \\ & \left. + (\rho_a L_v U (Qs - Qa) \sigma_{C_E, ran})^2 + (\rho_a L_v C_E (Qs - Qa) \sigma_{U, ran})^2 + (\rho_a L_v C_E U \sigma_{(Qs-Qa), ran})^2 + 2r_{(Qs-Qa), U} \left( (\rho_a L_v C_E)^2 (Qs - Qa) U \sigma_{(Qs-Qa), ran} \sigma_{U, ran} \right) \right]^{1/2} \end{aligned}$$

where C<sub>pa</sub> is the specific heat of air, L<sub>v</sub> is the latent heat of vaporization, and ρ<sub>a</sub> is density of air. Over the 10 years of the SeaFlux version 1.0 dataset, the mean total uncertainties for

the LHF and SHF are  $6.2 \text{ W m}^{-2}$  and  $5.2 \text{ W m}^{-2}$ . The spatial distributions of the errors in the surface input fields and the fluxes are shown in Figures 11 and 12. Further details can be found in Clayson et al. (2014).



**Figure 11: Mean fields from the 10 years of SeaFlux version 1.0 data for the Ta, SST, Wspd, and Qa values and the associated total uncertainties. Note the factor of 10 difference in the uncertainties between the Ta and SST fields.**



**Figure 12: Mean fields from the 10 years of SeaFlux data for the Qs-Qa, LHF, SHF, and Ts-Ta values and the associated total uncertainties. Contour lines on the Qs-Qa total uncertainty plot is at  $0.5 \text{ g kg}^{-1}$ . Contour lines on the LHF and SHF plots are at  $100$  and  $35 \text{ W m}^{-2}$ , respectively. Contour lines on the LHF and SHF uncertainty plots are at  $15$  and  $10 \text{ W m}^{-2}$ , respectively. Contour lines on the Ts-Ta plot are at  $0^{\circ}\text{C}$  and on the Ts-Ta total uncertainty plot are at  $1.0^{\circ}\text{C}$ . From Clayson et al. (2014).**

## 5. Practical Considerations

The source code was written in Matlab. Whenever possible, the source code takes advantage of built-in functions. Each driver file denotes Matlab release version used to construct the source code—R2014a in this case—and any toolboxes required to run the code. Toolboxes in Matlab are equivalent to libraries in other languages.

Matlab is (generally) backward compatible. The code is expected to build independent of the platform or release version. The source code, however, will not build if the user does not have the required toolboxes noted in the header for each file.

Matlab toolboxes need to build the source code are: mapping, neural network, statistics, and signal processing.

### 5.1 Numerical Computation Considerations

The various algorithms take advantage of parallelization when possible. When the parallelization toolbox from Matlab, the native development language, is available, the source code will adapt and use the parallelization. Should parallelization be available, but the user's preference is to not parallelize, this feature is easily disabled at the beginning of the source code.

The MoBI scheme—the third processing step for the atmospheric surface parameters—cannot be parallelized because of sequential dependence. Filling missing data consequently takes significant cpu cycles to process the entire data record.

When possible, source code takes advantage of matrix algebra. This is standard practice in Matlab and should pose no problems. Matrix algebra is preferred to improve performance, improve coding logic, and reduce introduction of bugs.

Conversion between different data types—single and double precision, integer, etc.—results in round-off errors in computation. This is expected due to the incorporation of several different observation and model products in producing the final data set. In general, data are handled in their encoded format (usually single precision) until calculations require converting to double precision to combine with other data. The round-off errors are within reason for the algorithm.

### 5.2 Programming and Procedural Considerations

Source codes for this CDR were written following the “General Software Coding Standards” for the CDR program. All code follows standard practices acceptable for the development language—Matlab—and were written to take advantage of built-in functions whenever possible.

The source code for each CDR are broken into phases that correspond to each step of the process flow diagram (see Figure 1, for instance). Source code are organized around a naming structure to assist the user. The atmospheric surface parameters, for example, require three processing steps: Swath, Grid, Interp. Each step has a corresponding

driver file—denoted by the prefix “SEAFUXNN\_”, as in SEAFUXNN\_Swath. Functions associated with each step of the processing carry the prefix denoting the parent driver file—as in Swath\_CLWSSMI. This parent-child relationship allows the user to follow the flow of the source code.

Each driver file has an initialization section where the user can prescribe the input/output directories for the various data files required for each step of processing. This should be the only modification required by the user.

Whenever possible, object-oriented-programming techniques were employed. This includes passing objects between functions, organizing variables, controlling the flow of data, and interfacing with NetCDF. Object-oriented programming, considered standard practice and preferred in Matlab, is especially helpful for code involving many lines, multiple functions, and multiple input/output steps, as is the case for this CDR.

To reduce processing time, the source code was written to loop through each day either in serial or parallel. This approach offers two advantages:

- Users can run multiple instances of the source code simultaneously.
- Output can then be generated for a single day or a specified range.

### 5.3 Quality Assessment and Diagnostics

Sanity checks, both visual and empirical, were employed frequently as diagnostic measures. These measures included:

- Comparison of neural network output with SeaFlux in situ observations.
- Comparison of atmospheric surface parameters with MERRA values.
- Inter-comparison of neural network output and error distribution across the different SSM/I satellites.

To improve neural network performance, the F08 SSM/I satellite was treated separately due to extensive noise in the 85 GHz channel. Performance of the F08 satellite is only slightly degraded compared to the other five SSM/I satellites.

SSM/I Satellite F15 was subject to significant contamination after August 2006 due to the activation of the RADCAL correction. Observations for this satellite were omitted after the correction was activated. For more information, see the CSU-FCDR documentation where the authors note that observations during this period are “unsuitable for climate applications.”

SSMIS sensor failure in the 37 GHz channel occurred for the F17 satellite after March 2016. Observations for this satellite were omitted after failure.

SSMIS satellites F16 and F19 satellites were also considered for inclusion. The F16 satellite suffered calibration issues shortly after launch. The F19 satellite failed after February 2016. Neither satellite was used in creation of this data set.

Appropriate quality flags included with the different data sets were employed during the different stages of processing.

## 5.4 Exception Handling

Matlab includes extensive debugging functions. Most errors are handled through these built-in functions. This is standard practice when programming in Matlab. The source code are written to take advantage of the native debugging by employing coding strategies, such as preallocating matrices using the value “Inf” (infinity) instead of a numeric value to invoke an error if the value is not replaced as expected.

For errors involving a breakdown in logic or calculations, the source code return error-specific information to the user. Aside from input/output considerations and toolbox requirements, the user should not encounter build errors.

### 5.4.1 Conditions Checked

The primary error when building the source code involves input/output directories. Each driver file begins with several lines to specify the input and output directories. Following the comments in the source code, the user should modify accordingly the directory paths based on their system. An error is returned if the user specifies a directory that is empty or for which they do not have write privileges.

Toolbox requirements are also checked by the source code. An error will occur if there is no access to the required toolboxes. The user is encouraged to check the toolbox list above to ensure the code will build.

Memory considerations are handled internally by Matlab. Appropriate information is returned to the user if such errors occur. The source code was written to avoid overloading memory.

### 5.4.2 Conditions Not Checked

Error handling for function and sub-function argument lists is handled internally by Matlab. When appropriate, argument lists are checked as part of a stand-alone function. This should not affect the user because the source code is self-contained, requiring only the augmentation and building of the driver files.

### 5.4.3 Conditions Not Considered Exceptions

Missing data are denoted with various quality flags. The following cases are not considered erroneous:

Locations over land or lakes are considered contaminated and filled as missing.

Satellite locations for brightness temperature within 45 km of land are considered contaminated and are thus discarded.

Locations over sea ice are considered contaminated and filled as missing.

Nonphysical values returned by the neural networks for atmospheric parameters (e.g., negative wind speed; see Section 3.4.1) are filled using the MoBI scheme.

Locations with missing source data or for which the neural network output are outside the operating range (e.g., negative wind speed) are filled using the MoBI scheme (Section 3.4.3). When MoBI fails to interpolate a value, or returns a nonphysical value (see Section 3.4.1), the data are filled as missing with the appropriate quality flag.

Accuracy for high-wind-speed locations (i.e.,  $U_{10} > 45 \text{ m s}^{-1}$ ) near sea ice cannot be verified and are therefore filled as missing because of potential ice contamination. These locations are flagged as sea ice.

Nonphysical values returned by the SST processing (see Section 3.4.6) are filled with missing.

Unrealistic values returned by the FLUX processing (see Section 3.4.1) are filled as missing.

Several steps of the processing rely on collocation of one data set to another. When a suitable collocation cannot be found, the data are marked as missing and filled using MoBI.

## 5.5 Algorithm Validation

Validation of the algorithms used to calculate this dataset can be found in their respective published papers as outlined in the various sections of this document.

### 5.5.1 Validation During Development

The neural network for atmospheric parameters was trained using the extensive SeaFlux in situ observation data record. Neural network structure was chosen based on optimization of error characteristics across almost 1800 different combinations. Performance of the atmospheric parameter neural network was compared with that of Roberts et al. (2012). Comparison included visual comparison of error characteristics and direct comparison of empirical values such as root-mean-square error, bias, and correlation.

Random instantaneous and averaged maps were compared with MERRA spatial patterns for congruity during the processing.

## 5.6 Processing Environment and Resources

Build time for the source code varies depending on parallel processing capabilities. All processing, except for the third processing step for atmospheric parameters (Figure 6), can be completed on a single processor with similar specifications as Table 2 in approximately 10 days.

As noted above, filling missing data with the MoBI scheme (step three in processing the atmospheric parameters; see Figure 1) accounts for the majority of processing time and requires approximately 40 days to complete using 12 processors.

Temporary storage is required during the first and second steps of processing for atmospheric parameters, and during the first step of processing for SST. (See processing flow diagrams in Figure 1 and Figure 2.) Total size of temporary storage is less than 250 GB and can be reduced through file management as each step is completed and files are no longer needed.

**Table 1: Processing Environment.**

<b>Memory</b>	512 GB
<b>Processor</b>	Four hexacore Intel E5-Xeon @ 3.4 GHz
<b>Programming Language</b>	Matlab R2012b (64-bit)
<b>External Libraries</b>	None

## **6. Assumptions and Limitations**

### **6.1 Algorithm Performance**

Data produced in this CDR are constrained by the availability of multiple observations and data products. Reproduction and extension of the temporal record are therefore limited by continued access and service of input data products. Should input data become inaccessible or unsupported in the future, an acceptable replacement would have to be sought. All data used in initial production of the CDR are available and up to date.

### **6.2 Sensor Performance**

Assumptions about the sensors are made by the input data providers, who apply their algorithms to compute brightness temperatures, SST, precipitation, radiation, and ice characteristics.

## **7. Future Enhancements**

### **7.1 Enhancement 1: Reprocessing**

As new and improved versions of the input satellite datasets, including the SSM/I and SSMIS brightness temperatures, are produced, new versions of the data set could be re-run including these improvements.

### **7.2 Enhancement 2: Extend Temporal Record**

Inclusion of the SSMIS series of satellites would extend the CDR temporal record. A new neural-network-weighting algorithm would have to be performed for those satellites. But otherwise the processing stream would remain the same.

### **7.3 Enhancement 3: Improved Parameter Values**

The inclusion of additional satellites, AMSU and AMSR for example, may improve near-surface temperature and humidity. Doing so would require a new neural-network-weighting algorithm. Otherwise the processing stream would remain the same.

### **7.4 Enhancement 4: Reduced Temporary Storage**

At present, production of atmospheric parameters occurs in three steps (see Figure 1). Under this multi-phase processing approach, each phase is manageable in length and runtime. These advantages come at the cost of temporary storage requirements that could be reduced or eliminated in the future.

### **7.5 Enhancement 5: Smoother Ice Mask**

The GEWEX ice mask is available on an equal-area grid. Conversion to the equal-angle (regular) grid used in this CDR resulted in a coarser ice mask than is preferred. Differences are insignificant from an empirical standpoint, but a smoother ice mask would reduce the runtime of the MoBI scheme (step three in Figure 1).

## 8. References

- Andersson, A., K. Fennig, C. Klepp, S. Bakan, H. Grassl, and J. Schulz, 2010: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS-3, *Earth Syst. Sci. Data*, **2**, 215-234, doi:10.5194/essd-2-215-2010.
- Bauer, P. and P. Schluessel, 1993: Rainfall, Total Water, Ice Water, and Water Vapor Over Sea From Polarized Microwave Simulation and Special Sensor Microwave/Imager Data, *J. Geophys. Res.*, 98(D11), 20737-20759.
- Berg, W., M. R. P. Sapiano, J. Horsman, and C. Kummerow, 2013: Improved Geolocation and Earth Incidence Angle Information for a Fundamental Climate Data Record of the SSM/I Sensors. *IEEE Trans. Geosci. Rem. Sens.*, **50 (3)**, 1504-1513.
- Berry, D.I. and E. C. Kent, 2009: A new air-sea interaction gridded dataset from ICOADS with uncertainty estimates. *Bull. Amer. Meteorol. Soc.*, **90**, (5), 645-656. doi:10.1175/2008BAMS2639.1.
- Berry, D.I. and E. C. Kent, 2011: Air-Sea Fluxes from ICOADS: The Construction of a New Gridded Dataset with Uncertainty Estimates, *Int. J. Climatology*, **31(7)**, 987-1001, doi: 10.1002/joc.2059.
- Bosilovich, M., 2008: NASA's Modern Era Retrospective-analysis for Research and Applications: Integrating Earth Observations. Earthzine.
- CERES Science Team, Hampton, VA, USA: NASA Atmospheric Science Data Center (ASDC), Accessed January 15, 2016 at doi: 10.5067/Terra+Aqua/CERES/SYN1degDAY\_L3.003A.
- Clayson, C. A., and J. A. Curry, 1996. Determination of surface turbulent fluxes for TOGA COARE: Comparison of satellite retrievals and in situ measurements. *J. Geophys. Res.*, **101**, 28,515-28,528.
- Clayson, C.A. and A. Bogdanoff, 2013. The effect of diurnal sea surface temperature warming on climatological air-sea fluxes. *Journal of Climate*, **26**, 2546-2556.
- Clayson, C. A., J. B. Roberts, and A. Bogdanoff, 2014: SeaFlux Version 1: a new satellite-based ocean-atmosphere turbulent flux dataset. *Int. J. Climatol.*, revised.
- Curry, J. A., A. Bentamy, M.A Bourassa, D. Bourras, E.F. Bradley, M. Brunke, S. Castro, S.H. Chou, C.A. Clayson, W.J. Emery, L. Eymard, C.W. Fairall, M. Kubota, B. Lin, W. Perrie, R.R. Reeder, I.A. Renfrew, W.B. Rossow, J. Schulz, S.R Smith, P.J. Webster, G.A. Wick, X. Zeng, 2004. SEAFLUX. *Bulletin of the Amer. Meteorol. Soc.*, **85**, 409-424.

- Fennig, K., A. Andersson, S. Bakan, C. Klepp, M. Schroeder, 2012: Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS 3.2 - Monthly Means / 6-Hourly Composites. Satellite Application Facility on Climate Monitoring. doi:10.5676/EUM\_SAF\_CM/HOAPS/V001.
- Hillburn, K. A. and C. -L. Shie (2011), Decadal trends and variability in special sensor microwave/imager (SSM/I) brightness temperatures and earth incidence angle, RSS Technical Report 092811.
- Hollinger, J. Lo, R., and Poe, G., 1987: Special Sensor Microwave/Imager User's Guide, Naval Research Laboratory, Washington, D.C.
- Hollinger, J., 1989: DMSP Special Sensor Microwave/Imager Calibration/Validation, Naval Research Laboratory, Washing, D. C., Vol. 1.
- Hollinger, J., 1991: DMSP Special Sensor Microwave/Imager Calibration/Validation, Naval Research Laboratory, Washington, D. C., Vol. 2.
- Kratz, D. P., Stackhouse, P. W., Gupta, S. K., Wilber, A. C., Sawaengphokhai, P., & McGarragh, G. R. (2014). The Fast Longwave and Shortwave Flux (FLASHFlux) Data Product: Single Scanner Footprint Fluxes. *J. of Appl. Meteorol. Climatol.*, 53(4), 1059-1079. doi:10.1175/JAMC-D-13-061.1
- Mears, C., D. Smith, and F. J. Wentz. 2001, Comparison of SSM/I and buoy-measured wind speeds from 1987 - 1997, *J. Geophys. Res.*, **106**(C6), 11719-11729.
- Sapiano, M. R. P. and W. Berg, 2013: Intercalibration of SSM/I and SSMIS for the CSU FCDR. CSU Technical Report, July, 2013. Available at <http://rain.atmos.colostate.edu/FCDR/>
- Stackhouse, Jr., Paul W., Shashi K. Gupta, Stephen J. Cox, J. Colleen Mikovitz, Taiping Zhang, and Laura M. Hinkelman, 2011: The NASA/GEWEX Surface Radiation Budget Release 3.0: 24.5-Year Dataset. *GEWEX News*, 21, No. 1, February, 10-12.
- Weng, F., N. C. Grody, R. Ferraro, F. Basist, and D. Forsyth, 1997: Cloud liquid climatology from the special sensor microwave/imager. *J. Climate*, 10, 1086-1098.
- Wessel, P., and W. H. F. Smith, 1996: A Global Self-consistent, Hierarchical, High-resolution Shoreline Database, *J. Geophys. Res.*, 101, #B4, pp. 8741-8743.
- Yan, B., and F. Weng, 2008: Intercalibration Between Special Sensor Microwave Imager/Sounder and Special Sensor Microwave Imager, "IEEE Trans. on Geoscience and Remote Sensing", vol. 46 (4), 984-995.

## Appendix A. Acronyms and Abbreviations

Acronym or Abbreviation	Definition
ANSI	American National Standards Institute
AVHRR	Advanced Very High Resolution Radiometer
C-ATBD	Climate Algorithm Theoretical Basis Document
CCMP	Cross-Calibrated Multi-Platform
CDR	Climate Data Record
COARE	Coupled Ocean Atmosphere Response Experiment
CSU	Colorado State University
FCDR	Fundamental Climate Data Record
FOC	Full Operating Capability
GEWEX	Global Energy and Water Experiment
GSHHG	Global Self-consistent, Hierarchical, High-resolution Geography
HOAPS	Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data
IEEE	Institute of Electrical and Electronic Engineers
ICD	Interface Control Document
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
IOC	Initial Operating Capability
ISCCP	International Satellite Cloud Climatology Program
IVAD	ICOADS Value-Added Dataset
MERRA	Modern-era Retrospective Analysis for Research and Applications
MERRA-2	Modern-era Retrospective Analysis for Research and Applications; version two
MoBI	Model-Based Interpolation
NCDC	National Climatic Data Center
NOAA	National Oceanic and Atmospheric Administration
OAD	Operational Algorithm Description
OCDR	Oceanic Climate Data Record
OISST	Optimum Interpolation Sea Surface Temperature
SRB	Surface Radiation Budget
SSM/I	Special Sensor Microwave/Imager
SSMIS	Special Sensor Microwave/Imager Sounder

## Appendix B. Quality Control Flags for Atmospheric Parameters

Flag Value	Flag Meaning	Explanation
0	Observation	Value from neural network
1	Interpolation	Value from model-based interpolation
2	Snow/Ice	Snow or ice contamination
3	Land	Over land
4	Lake	Over lake
5	Failed Interpolation	Model-based interpolation failed

Grid cells for flags with value 2, 3, 4, or 5 are filled as missing where appropriate.

## Appendix C. Quality Control Flags for SST

Flag Value	Flag Meaning	Explanation
0	Observation	Value from neural network
1	Unused	Unused flag
2	Snow/Ice	Snow or ice contamination
3	Land	Over land
4	Lake	Over lake
5	SST Missing	SST unresolved

Grid cells for flags with value 2, 3, 4, or 5 are filled as missing where appropriate. Flags 2, 3, and 4 take precedence over flag 5.

## Appendix D. Quality Control Flags for Heat Fluxes

Flag Value	Flag Meaning	Explanation
0	Observation	Value from neural network
1	Interpolation	Unused flag
2	Snow/Ice	Snow or ice contamination
3	Land	Over land
4	Lake	Over lake
5	High Wind Speed	Wind speed from input file greater than $45 \text{ m s}^{-1}$ ; wind speed set to $45 \text{ m s}^{-1}$
6	Failed Interpolation	Fluxes unresolved

Grid cells for flags with value 2, 3, 4, or 6 are filled as missing where appropriate. Flags 2, 3, and 4 take precedence over flag 6.