Sea Surface Temperature Error Budget: White Paper

Interim Sea Surface Temperature Science Team (ISSTST)

18 June 2010

Table of Contents

Executive Summary

A number of federal agencies fund research related to the development and use of satellite-derived sea surface temperature (SST) data products. Furthermore, within each agency there may be more than one program addressing SST issues and these programs are often unrelated. In an effort to coordinate NASA funded projects that relate to satellite-derived SST products, the NASA Physical Oceanography Program Manager, Eric Lindstrom, formed an Interim Sea Surface Temperature Science Team (ISSTST) consisting primarily of those in the U.S. research community using or developing SST products, and he funded a workshop (http://www.ssterrorbudget.org/ISSTST/ ; username: isstst ; password: ErrorBudget) to characterize the uncertainty budget of satellite-derived SSTs. The workshop was attended by 42 ISSTST members and 4 observers . Participants at the workshop were divided into 6 groups:

1. The physical basis of SST measurements,
2. Radiative transfer modeling and SST retrieval algorithm development,
3. Calibration, instrument characterization, and measurement validation: pre-launch and on-orbit,
4. Data merging and gridding,
5. The Climate Record: reprocessing, data access, and stability of climate record, and
6. Applications of SST measurements.

Each of these groups produced a report (available at: http://www.ssterrorbudget.org/ISSTST/Workshop.html) summarizing their work. These form the basis for this document.

The workshop and subsequent collaborative work of the ISSTST resulted in contributions in three areas:

1. **Requirements placed on satellite-derived SST products -** These include the traditional requirements imposed by climate change and numerical weather prediction as well as requirements imposed by applications focused on process oriented-studies and other shorter time and space scale uses. The strictest requirements in each of the five categories considered are: a spatial resolution of 100 m, a temporal resolution of 15 minutes, a pixel location uncertainty of 100 m, an absolute SST uncertainty of 0.1K, a relative SST uncertainty of 0.05K and a stability of 0.04K/decade.
2. **A framework for the characterization of the error budget for satellite-derived SST products -** This framework is based on the two major categories of SST products in use today: skin/sub-skin temperatures[1] derived directly from radiances collected by satellite-borne instruments and products derived from these skin/sub-skin temperatures; i.e., obtained indirectly from the radiances. We refer to the latter as 'derived products' in the following. Examples of derived products are skin temperatures resulting from the combination of skin/sub-skin temperatures collected by a suite of satellite-borne sensors and merged into a new product, bulk temperature products corresponding to the temperature some depth beneath the sub-skin, say 1 m, and analysis products obtained using objective analysis techniques to develop complete SST fields from gappy skin temperatures. Physical factors contributing to the uncertainties in the products, skin/sub-skin or derived, fall into three categories: those related to the instrument, I*nstrument Related Errors,* those related to atmospheric correction and/or emissivity effects, R*etrieval Algorithm Related Errors*, and those related to spatial and temporal variability

1. Skin temperatures are those derived from instruments sampling in the thermal infrared portion of the spectrum and subskin temperatures are those derived from instruments samping in the microwave portion of the spectrum.

of temperature in the upper ocean, E*rrors Resulting from Oceanographic Variability*. *Instrument Errors* consist of instrument noise, calibration and geolocation errors and are generally associated with processing from L0 to L1. *Retrieval Algorithm Errors* consist primarily of simulation, input, sampling and classification errors and are generally associated with processing from L1 to L2. *Errors Resulting from Oceanographic Variability* consist of errors in the geophysical models used to represent this variability and variability of the input parameters used by these models. These errors are generally associated with processing from L2 to L3 and L3 to L4, although there are some L2 SST products also impacted by oceanographic variability. In addition to physical factors contributing to the uncertainty of SST products, merging, gridding and analysis operations used to generate L3 and L4 products will also contribute to the error budget of these products. We refer to these as M*erging, Gridding and Analysis Algorithm Errors.* Contributions in this category result from the imperfect statistical prediction of SST in underobserved areas and remaining random errors and biases in the merged data sets from individual sensors.

3. **Recommendations for tasks that need to be undertaken to improve satellite-derived SST products.** These recommendations (approximately 70 total) cover a very broad range of activities and are the prime focus of the SST Error Budget White Paper. Recommendations fall into four general categories: general/administrative recommendations, those that are not SST specific; those that identify specific tasks that, if undertaken, would result in lower SST uncertainties; those that identify research tasks that need to be undertaken and; those that involve enhanced communication between those in the SST community. Recommendations falling into each of these groups are summarized below.
    1. *General/Administrative* - These recommendations encourage the gathering of general information about the reduction of SST errors - lessons learned - and the need to pass this information on to other similar projects.
    2. *Specific tasks* - There are a number of tasks that could be undertaken immediately that would result in an improvement in the uncertainty of SST products with no additional research. These fall into four broad categories:
        1. *Facilities/methodologies* - Tasks in this group range from a recommendation that a centralized pre-launch facility to characterize instrument performance be formed, to a number of tasks related to the development of procedures to be used for instrument characterization both pre- and post launch.
        2. *More/better in situ observations* - The number of recommendations in this group points to the importance that the ISSTST places on *in situ* observations. These recommendations dealt with *in situ* measurements of all types - ship-based infrared (IR) radiometers, drifting buoy measurements, Argo floats and field programs. Specific recommendations were for more high latitude measurements, more ship-based radiometer observations, consistent calibration of ship-based radiometers, increased sampling by Argo floats in the upper few meters of the water column and improved accuracies for drifting buoy measurements.

3. *Reference data sets* - Reference data sets are required to evaluate data products and to tune regression algorithms. Recommendations in this group encourage the creation of such data sets.
4. *Uncertainty characterization/Quantifying of the uncertainty* - All of these recommendations are rather general; they simply state the areas in which the errors associated with an SST product should be improved.

3. *Research Tasks* - The ISSTST recommends a number of actions requiring research. These also fall into four broad categories:
    1. *Improved retrieval algorithms* - The bulk of the research-related recommendations falls in this category and they cover the entire range of retrieval algorithms, specifically, those for regression-based retrievals, physically-based retrievals, hybrid retrievals and special purpose algorithms for lakes, high latitude and near land.
    2. *Data classification* - Recommendations in this group address issues related to identification of pixels for which the retrieved SST value has likely been corrupted. The specific forms of contamination addressed are: clouds, aerosols, sea ice and sun glint. In all cases, the recommendations are made based on the inadequacy of current data classification algorithms. In fact, errors in classification are one of the largest contributers to the SST uncertainty.
    3. *Instrument development* - The development of cheap but accurate ship-based radiometers and the investigation of high-resolution microwave (μ-wave) sensors are recommended.
    4. *Improved merging, gridding, and objective analysis procedures* - Recommendations call for research on improved estimates of SST error and variability, on advanced objective analysis algorithms and implementations, on bias-correction algorithms, on dealing with statistical non-stationarity of SST fields, on systematic validation and evaluation of gridded data sets, on their uncertainty and ways to communicate it, and on 3D objective analyses of gridded temperature data sets for the surface ocean.
    5. *Uncertainties* - Research into methods that might be used to characterize the uncertainty associated with an SST product are recommended.

4. *Communication* - It is clear from the recommendations in this group that the ISSTST sees the need for improved communication between those involved in the development and use of SST products - those responsible for instrument related issues need to communicate their findings to those developing SST products. Those developing products need to communicate with users, and users need to feed their findings back to the product developers and forward to other users.

1. Introduction

A number of federal agencies fund research related to the development and use of satellite-derived sea surface temperature (SST) data products. Furthermore, within each agency there may be more than one program addressing SST issues and these programs are often unrelated. In an effort to coordinate NASA funded projects that relate to satellite-derived SST products, the NASA Physical Oceanography Program Manager, Eric Lindstrom, formed an Interim Sea Surface Temperature Science Team (ISSTST) consisting primarily of those in the U.S. research community using or developing SST products, and he funded a workshop (http://www.ssterrorbudget.org/ISSTST/; username: isstst ; password: ErrorBudget) to characterize the uncertainty budget of satellite-derived SSTs which was held at the University of Rhode Island's Whispering Pines Conference Center in West Greenwich, Rhode Island, 16-18 November 2009. The Workshop was organized by the ISSTST steering committee consisting of Sandra Castro, Peter Cornillon (Chair), Chelle Gentemann, Andy Jessup, Peter Hacker, Alexey Kaplan, Eric Lindstrom, Eileen Maturi, Peter Minnett, and Richard Reynolds. The workshop was attended by 42 ISSTST members and 4 observers. Participants at the workshop were divided into six groups with discussion coordinated by steering committee members as indicated:

1. The physical basis of SST measurements (S.Castro),
2. Radiative transfer modeling and SST retrieval algorithm development (C.Gentemann),
3. Calibration, instrument characterization, and measurement validation: pre-launch and on-orbit (P.Minnett),
4. Data merging and gridding (A.Kaplan),
5. The Climate Record: reprocessing, data access, and stability of climate record (R.Reynolds and P.Hacker),
6. Applications of SST measurements (P.Cornillon and E.Maturi).

Each of these groups produced a report summarizing their work (available at: http://www.ssterrorbudget.org/ISSTST/Workshop.html). These reports form the basis for this White Paper.

This paper is divided into four primary sections and accompanying appendices. The first section begins with a discussion of SST[2] error[3] requirements based on a broad range of applications and is followed with the description of a general framework for the SST uncertainty budget. Section 2 introduces the instrument and retrieval algorithm errors that contribute to the uncertainty budget of determining skin and sub skin SST in satellite coordinates. Section 3 describes the ocean variability and model errors associated with producing gridded or merged SST products. In Section 4 validation of SST products is discussed both from the perspective of formal validation efforts - those associated with the launch of a new sensor - as well as from the perspective of producers of derived data sets and from the perspective of validation performed informally by individual users. In each section, recommendations addressing the SST error

---

2. All acronyms are expanded in Appendix VI.
3. We use *error* in this document to include both statistical uncertainties and errors due to problems in formulation, such as the systematic model error.

budget and its characterization are called out. The overview of these recommendations is given in the Executive Summary.

While seemingly obvious, it is not necessarily self-evident that lessons learned from prior and current satellite missions are carried forward to future sensors at all levels from conceptual design to data processing. This results in inefficiencies and, possibly, repeated errors. It can also lead to a failure to ensure consistent and improving accuracies with time. We therefore begin this document with two recommendations designed to address this issue.

Recommendation: Apply lessons learned from heritage sensors and missions to future missions to ensure consistent and improving accuracies with time.

Recommendation: Document lessons learned as they relate specifically to SST. This might be done by an SST Science Team.

1.1 Requirements placed by applications on the SST error budget

Science applications which benefit from satellite SST products cover the full range of temporal and spatial scales. Time scales associated with ocean processes include diurnal, multi-day, intraseasonal, seasonal, annual, interannual, decadal, and longer term trends. Associated space scales cover the full range - from the meter to the global scale. Phenomena and processes include: diurnal warming and cooling cycle; surface expression of internal waves; regional air/sea interaction processes including synoptic events, intraseasonal processes (especially in the tropical and coastal regions), fluxes near fronts, and convection and mixing in mode and deep water formation regions; open ocean fronts; regional to basin-scale fluxes of heat, evaporation, and precipitation associated with the seasonal cycle, trade wind, and monsoon regimes; seasonal-to-interannual basin- and global-scale processes such as ENSO; decadal variability such as the PDO; and global change-induced trends. For the coastal regions (and in lakes) applications include: upwelling, sub-mesoscale eddies and fronts, jets and filaments, and the impact of biogeochemical, pollutant, and river inflow processes on both dynamics and optical properties. In many cases the effects of coastal processes extend well offshore and influence open ocean conditions. In addition to covering the full range of spatial and temporal scales, the phenomena and processes have associated space/time signatures and amplitudes that vary regionally.

Not surprisingly, the broad range of applications for which SST fields are used results in an equally broad range of requirements with regard to the accuracy of these fields. There have been a number of studies undertaken to date in which requirements have been defined for the accuracy of satellite-derived SST fields for climate change research and as input to atmospheric circulation models. Generation of climate data records (CDRs) poses fairly stringent requirements on the accuracy of derived SST - an absolute temperature uncertainty of 0.1K and stability of better than 0.04K per decade (Ohring et al., 2005). Spatial and temporal averaging can be used to reduce the random errors, i.e., to increase the precision of the retrieval, but averaging does not reduce bias errors nor increase stability. Requirements for numerical weather prediction (NWP) are somewhat less stringent for accuracy, being 0.3K. The retrieved SST fields, however, are required at a spatial resolution of 5 km or better, with a 3-hour temporal resolution (Eyre et al., 2009). That precludes using significant averaging to reduce random errors, especially given the added complication of clouds preventing the retrieval of SSTs using IR radiometers. Another source of information with regard to acceptable SST errors may be found in the NPOESS IORD-II generated by NOAA and the Department of Defense (DoD). It identifies remote sensing parameters used in operational applications (Gudes et al., 2001).

In contrast to the requirements for CDRs, NWP, and operational applications, the requirements for feature and process-related studies have received little attention. Therefore, one of the tasks undertaken at the SST Error Budget Workshop was to augment the list of requirements identified in previous studies with those associated with applications undertaken by meeting participants. These are summarized in Appendix II and tend to focus on feature-related studies. Of interest, although not surprising, is that those interested in feature and process oriented studies imposed fairly stringent requirements on the location of pixels. They also tend to value relative thermal accuracy over absolute accuracy and are not quite as strict with regard to data quality in general. For ocean state estimation and for climate models - also represented by workshop participants - the requirements are similar to those of CDR, except that the modeling applications require a more complete characterization of SST retrieval uncertainties, including a characterization of SST retrieval uncertainties pertaining to the conversion from observables to model prognostic variables, such as averaged temperatures over the upper 1-10 meters. Lake and coastal applications may have higher spatial and temporal resolution requirements than do open-ocean applications. Finally, the computation of air sea fluxes have an absolute SST accuracy requirement comparable to CDR and they require an accurate description of the diurnal cycle.

The most stringent requirements placed on the accuracy of SSTs, both from previous studies and from the analysis of applications undertaken at the workshop, are presented in Table 1.

Table 1. Requirements for SST error budget.

Application	**Source**	**Spatial resolution** (km)	**Temporal resolution** (hrs)	**Geolocation accuracy** (km)	**Absolute accuracy** (K)	**Relative accuracy**
CDR	Ohring et al., 2005				0.1	0.04°K/decade
CDR	Appendix II					0.05°K/decade
NWP	Eyre et al., 2009	5	3		0.3	
Global Operations	NPOESS IORD-II	0.25	3	0.1	0.1	0.05°K
Coastal/Lake Operations	NPOESS IORD-II	0.1	6	0.1	0.1	

Fronts	Appendix II	0.1	0.25	0.1	1	0.1°K
Climate Models	Appendix II	25	24	5	0.2	0.05°K/decade
Lakes	Appendix II	1	3	1	0.3	0.2°K
Air-sea fluxes	Appendix II	10	24	2	0.1	
Mesoscale	Appendix II	1	168		0.1	0.1°K
Submesoscale	Appendix II	0.1	24		0.1	0.1°K
Strictest		0.1	0.25	0.1	0.1	0.05°K 0.04°K/decade

1.2 A Framework for the SST Error Budget

Satellite radiometers are sensitive to skin[4] or sub-skin[5] temperatures depending on the wavelength in which the radiometer is sampling. For both IR and µ-wave instruments, the SST is determined by assuming that the ocean surface is a gray body and that the emissivity is either known or can be estimated as part of the retrieval process. In the IR, the emissivity of the ocean surface is high and wind-speed effects do not introduce a significant dependence (Hanafin and Minnett, 2005; Nalli et al, 2008a, b), whereas in the µ-wave the apparent emissivity is low and it exhibits a strong wind-speed dependence, so an additional source of SST uncertainty is the imperfect correction for wind effects. Similarly, heavy rainfall introduces an error in the µ-wave SST retrievals, and measurements contaminated by rain are flagged as being unreliable. Imperfections in the flagging algorithm can result in elevated SST uncertainties.

In addition to the emissivity, SST retrievals based on satellite-borne instruments also require knowledge of the radiative transfer characteristics of the atmosphere between the pixel of interest and the sensor. Geophysical uncertainties in SST caused by atmospheric processes are very different in the IR and the µ-wave. In the IR, sources of atmospheric uncertainties include residual contamination by radiance emitted from clouds and aerosols in the field of view, and the radiative effects of water vapor. Water vapor is the principal atmospheric gaseous constituent that interacts with the sea surface emission in the spectral regions where atmospheric

4. IR instruments measure radiance from the upper ~10-20 µm, which is referred to as the skin. See Appendix IV for definitions of the various "surface" temperature terms used in this document - interface, skin, subskin, depth or bulk and foundation.
5. Microwave instruments measure radiance from the upper ~100 µm. This is referred to as the sub-skin.

transmission is high where the SST retrievals are made. Atmospheric humidity is very variable in both space and time, and instances where the local atmospheric humidity and temperature profiles deviate from the distribution used in the derivation of the atmospheric correction algorithm will result in higher uncertainties (e.g. Minnett, 1986).

Although satellite-borne sensors measure radiation in the upper 10 to 100 μm of the water column, knowledge of the vertical and horizontal variability of the upper ocean is required to combine data from multiple satellite passes and sensors as well as to generate products representative of the temperature below the upper 100 μm. Processes receiving the most attention in this regard are those related to the diurnal warming and to the skin effect.

Diurnal warming occurs in the upper layer of the ocean during the daytime, as it is heated by solar radiation, and is followed by cooling due to the net loss of heat at night. The amplitude and depth over which this diurnal process occurs is strongly influenced by upper ocean mixing. The most significant warming events are observed only at low wind speeds (e.g., < ~3 m/s), when combined with strong insolation, but warming is clearly present for winds up to 6 m/s. Since 30% of global winds are less than 6 m/s (Donlon et al., 2002), daytime SST measurements should be always suspected of containing a contribution from the diurnal warming.

Because turbulent diffusion is suppressed just below the ocean surface, a laminar sub-layer, known as a "skin layer"[6] and usually less than 1 mm thick, is formed there. Within this layer, the heat exchange with the atmosphere is accomplished by molecular diffusion, requiring large vertical gradients of temperature. The resulting temperature differences across the skin layer are typically a couple of tenths of a degree, but at very low wind speeds they can exceed 0.5K.

Figure 1 is a schematic showing the primary contributors to the error budget for satellite-derived SST products. These products fall into two categories, the two large, gray boxes on the left hand side of the figure: skin and sub-skin temperatures in satellite coordinates and derived SST products. Skin and sub-skin temperatures in satellite coordinates are determined directly from satellite-derived radiances and distributed in satellite coordinates with no re-gridding or averaging. Derived products are determined from skin/sub-skin[7] temperatures either by combining data from different sensors with different spatial and temporal coverage and/or by extrapolating to locations, either in the horizontal or the vertical, that were not sampled. Examples of derived products are skin temperatures resulting from the combination of skin/sub-skin temperatures collected by a suite of satellite-borne sensors and merged into a new product, bulk temperature products corresponding to the temperature some depth beneath the sub-skin, say 1 m, and analysis products obtained using objective analysis techniques to develop complete SST fields from gappy fields of skin temperatures.

---

6. The "skin layer" as defined here includes both the skin and sub-skin temperature regimes.
7. When referring to temperatures inferred from IR and microwave instruments together, we use the expression skin/subskin.

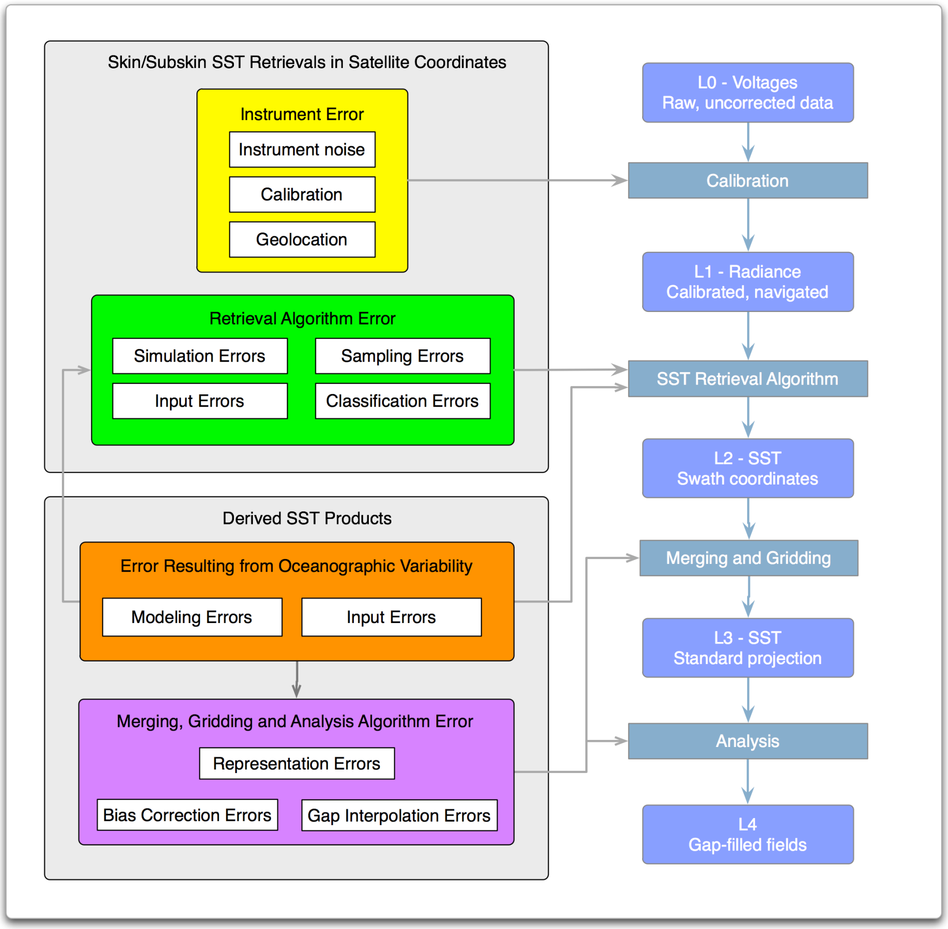


Figure 1. Error budget framework.

The primary distinction between these two categories is that those in the first group require top of the atmosphere (TOA) radiances - measured by space-borne IR and microwave sensors - plus knowledge of the structure of the atmosphere and knowledge of the emissivity of the ocean surface while those in the second category require the skin/sub-skin temperature - the products of the first category - and, possibly, knowledge of the structure of the upper ocean. Put another way, those in the first category deal with the physics of radiometer measurements, the physics of the atmosphere and the physics of the ocean's skin/sub-skin - that which determines its emissivity - and those in the second category deal with the physics giving rise to the variability of temperature in the upper few meters of the ocean. Separating the products into these two categories allows for a well defined separation of the error types contributing to the resulting products. Therefore the organization of this document is based on these two categories of SST products.

As shown in the Figure 1, elements of the SST error budget may also be viewed in the context of the processing steps that carry data through the sequence of five primary data processing levels, from raw voltages (L0) to interpolated gridded SST data sets (L4). Definitions of these five data processing levels as used by NASA and augmented by more recent GHRSST refinements that introduce sub-levels in the L3 are given in Appendix III.

Two groups of errors contribute to the uncertainty in skin/sub-skin retrievals, those associated with the instrument, the yellow rectangle in Figure 1 labeled **Instrument Error**, and those associated with the retrieval algorithm, the green rectangle in Figure 1 labeled **Retrieval Algorithm Error**. *Instrument errors* include contributions from the calibration source(s), the characterization of instrument measurement, stray radiation, measurement geolocation accuracy, etc. Instrument noise is also listed in this block. Instrument errors enter into the processing when processing from L0 to L1[8]. These errors are described in Section 2.1 and Section 4. *Retrieval algorithm errors*, the second group contributing to errors in the skin/sub-skin SST retrievals in satellite coordinates, include all factors involved in the conversion of TOA radiances, L1 products, to skin/sub-skin temperature values, the primary set of products at L2. These errors are described further in Section 2.2. This group consists of *input errors,* those errors associated with ancillary data used by the retrieval algorithm or to develop the retrieval algorithm, *classification errors,* errors resulting from erroneous flagging of the data (e.g., cloud contaminated, in a region of sun glint, etc.), *simulation errors*, errors resulting from the geophysical model used for the retrieval, and *sampling errors*, errors resulting from the averaging within the retrieval footprint.

The **Error Resulting from Oceanographic Variability** group contributes to the uncertainty in derived SST products. These errors enter primarily into the L2 to L3 and L3 to L4 processing steps and are described in Section 3. The nature of oceanographic variability in the upper ocean is discussed in Section 3.1. *Modeling errors* occur in a physical or empirical model when it is used to adjust retrieved SST values for certain geophysical effects, in order to represent the temperature at different depths, different horizontal locations or different phases of the diurnal cycle. *Input errors* are due to uncertainty of the ancillary data input into such models.

Upper ocean variability is a prime contributor to the errors of the L3 and L4 products, but it does so through the algorithms that actually create these products, the purple box of **Merging, Gridding and Analysis Algorithm Error** discussed in Section 3.2. *Representation error* is an error due to the intra-gridbox space-time variability that raises the effective error of individual L2 values when each of them is used as an estimate of the full gridbox average. *Bias Correction Error* is the bias present in the L3 or L4 products after multi-sensor bias correction procedures are applied. *Gap Interpolation Error* contributes uncertainties to the L4 gridboxes which had no direct observations and were infilled by the analysis procedure (Section 3.2.3).

The focus in the previous paragraph was on the L2 to L3 and L3 to L4 processing steps. However, there are cases in which derived products - e.g., temperatures beneath the sub-skin layer - are generated in satellite coordinates, L1 to L2, such as NAVOCEANO operational NLSST and the AATSR Reanalysis for Climate drifter depth SST products. For such products the reported SST no longer corresponds to the location at which radiometric measurements are made; as a result, errors will likely be introduced by the geophysical model of the upper ocean used in the retrieval algorithm. This is indicated by the arrow from the orange box to the L1 to L2 processing step labeled "SST Retrieval Algorithm".

---

8. Note that instrument noise is not introduced by this processing step, L0 to L1, but it must be characterized at this point.

Finally, note that there is an arrow from the oceanographic variability box (orange) to the retrieval algorithm error box (green). This addresses cases for which a skin/sub-skin temperature is retrieved in satellite coordinates using an algorithm affected by the variability in the upper ocean; for example, retrievals based on in situ measures of temperature, generally made at a depth of about 1 m.

2. Skin and sub-skin SST in satellite coordinates

As discussed in the previous section, errors contributing to skin and sub-skin SST retrievals in satellite coordinates fall in two broad groups: instrument error and retrieval algorithm error. These are discussed in the following two major subsections.

2.1 Instrument Error

Pre-launch calibration and characterization of spacecraft radiometers used to derive SST is of vital importance if the resulting SST fields are to be used in a quantitative manner. The pre-launch calibration and characterization should provide everything needed to give confidence in the on-orbit internal calibration procedure, and to quantify the contributions of the instrumental artifacts to the final uncertainty budget of the SST retrieval. Any dependencies of the artifacts on parameters that influence their magnitudes has to be determined in circumstances that mimic, as far as reasonably possible, the operating conditions on-orbit.

Contributions to the SST uncertainty budget are different for IR and μ-wave radiometers, being primarily governed by the physics of the measurement. Thus, they are different for different instrument design, and even between different models of the same instruments. The purpose of the pre-launch calibration and characterization is to establish the magnitudes and dependencies of the instrumental artifacts, and to provide algorithms for their correction in the on-orbit data stream.

2.1.1 Pre-launch calibration and characterization

Uncertainty sources to be quantified in the pre-launch calibration and characterization are very instrument specific. To interpret the pre-launch calibration and characterization measurement, a well-defined and accurate instrument model for each radiometer has to be built. This has to include: the characteristics of black bodies (IR) and hot-cold loads (μ-wave); a model of the detector system and instrument self emission; internal stray radiation, in-band and out-of-band spectral response; optical and electronic cross-talk; and modulation transfer function and co-registration of different detectors (including in the same spectral band for instruments with multiple detectors for each band). The emissivity and temperatures of emissive surfaces in the optical path have to be determined, and temperatures of the focal plane assemblies and filters have to be measured (for IR radiometers).

In the signal chain, the noise levels of the detectors have to be established along with the linearity of their response, and that of the pre-amplifiers. Temperature dependencies of amplifier gains and digitizer linearities have to be known, as does the "preferences" of digitizers for particular states. An important aim of the pre-launch calibration and characterization is to simulate the performance of the sensor on orbit. This has not always been done in the past.

Recommendation: Establish a centralized pre-launch calibration and characterization facility to monitor instrument performance under simulated on-orbit conditions.

The talent and expertise in the scientific community have not always been exploited in the testing process.

Recommendation: Establish a mechanism(s) for the two-way flow of information between the scientific community and the pre-launch calibration and characterization team(s) during pre-launch testing.

For many heritage instruments, comprehensive documentation of the tests performed, and of their results is not available. Transparency of the process, enabling possible reworking of the results in the future requires full documentation.

Recommendation: Fully document the procedures and the data taken, and archive this information in an accessible form.

Recommendation: Generate sensor models for those heritage instruments that lack one, so they can be used to assess the contributions to the error budgets of those instruments. For future sensors, the rigor of the instrument model should not be less complete than those for MODIS and AATSR (for example).

For many heritage IR radiometers the pre-launch calibration was done using laboratory black body calibration targets with uncertain traceability to National Standards. More recent radiometers have been calibrated using laboratory black-body targets that are traceable to National SI standards. The AATSR series have been calibrated using a single black-body target (C. Mutlow, 2009, Pers. Comm.), whereas the MODIS IR bands have been calibrated using a laboratory standard (Guenther et al., 1996). A TXR (Rice and Johnson, 1998) has been developed at NIST for the calibration of laboratory black-bodies used to calibrate satellite radiometers (Rice and Johnson, 1996) and ship-board radiometers for satellite-derived SST validation (Rice et al., 2004).

Recommendation: Perform pre-flight calibration using SI-traceable radiance standards for all future IR radiometers.

In reality the pre-launch calibration and characterization happens late in the launch preparations, usually with cost and time over-runs. Therefore, opportunities to rectify problems have been limited even in subsequent models.

Recommendation: Make every effort to fulfill pre-launch calibration and characterization requirements. Deriving the necessary information post-launch is extremely difficult if not impossible.

One important aspect of the pre-launch calibration and characterization is the opportunity to correct problems as they come to light. Historically, this has not always be done.

Recommendation: Include real-time review of pre-launch calibration and characterization test data.

Recommendation: Design the pre-launch calibration and characterization schedule to accommodate corrections and retesting.

For many past sensors, the important pre-launch calibration and characterization measurements and analyses are not easily accessible.

Recommendation: Convey results to scientific/user community.

2.1.2 On-orbit calibration and verification

Once a sensor is in orbit the accuracy of the calibration needs to be monitored and, if necessary, corrected. This can be achieved in a number of ways including comparisons with modeled radiances; this can give broad indications of significant changes in the on-orbit sensor performance and can track effects such as instrument degradation and orbit drift. Alternatively, detailed analyses of the on-orbit calibration can be undertaken using TOA calibration sources, the use of which is becoming standard practice at many operational agencies through the GSICS. Current sources of TOA spectral radiances include the IR IASI and the AIRS time series. In the future, missions such as CLARREO should be able to provide NIST traceable TOA radiances. Such comparisons can go beyond calibration monitoring and can be used to correct the calibration, something that is likely to be especially important for the heritage sensors.

Recommendation: Continuously monitor post-launch, satellite radiances to estimate biases, their stability, and cross-platform consistency.

Recommendation: Undertake on-orbit calibration monitoring and correction with the most appropriate TOA calibration sources available including RTM and spectral radiances provided by spectrometers such as IASI, AIRS, CrIS or future CLARREO mission.

2.2 Retrieval Algorithm Error

In this section, errors associated with the retrieval algorithm error block of Figure 1 are addressed. There are two sections that address the critical/primary sources of uncertainty for IR sensors and for μ-wave sensors. Special issues related to retrievals in polar regions are discussed in the final subsection related to retrieval algorithm error.

2.2.1 IR SSTs

2.2.1.1 IR Retrieval algorithm

SST retrievals from many IR radiometers (e.g. GOES, AVHRR, MODIS, future VIIRS) have been made using regression algorithms with coefficients determined from match-ups with *in situ* buoy data. Although RTM-based techniques have begun being tested in recent years, regression algorithms still remain the mainstream approach for SST retrievals in the IR. Furthermore, the climate record will likely be generated with these regression-based algorithms as the necessary ancillary data and instrument characterization were not available for many IR radiometers in the past.

For several instruments (eg. GOES, (A)ATSR, AVHRR-Metop) regression coefficients determined from an RTM have been developed. This approach may require a bias adjustment determined through comparisons to *in situ* data. OE techniques using online RTM in conjunction with first-guess SST and atmospheric fields have started being explored in recent years.

Heritage regression techniques have limited potential to resolve large-scale geophysical variations in atmospheric and surface-atmosphere conditions, and result in regional biases. New OE techniques have a better potential to resolve the bias, but they require simulation of NWP fields close in time and space to the observation. The reliance on NWP fields has two major drawbacks: errors in these NWP fields will introduce errors in the SST retrievals, and production of climate records requires consistent processing which may not be possible for historical IR radiometers.

Both regression and physical SST techniques rely on accurate radiances. The regression technique is less demanding and only requires stability of radiances versus absolute accuracy. The physical technique, on the other hand, requires absolute match between RTM and radiances and, therefore, is more demanding, requiring also good knowledge of the spectroscopy of atmospheric gases and the spectral characteristics of the instruments. Radiance bias correction is currently needed before the physical algorithm is applied.

Heritage regression techniques work progressively less accurately as one deviates from typical open ocean conditions, and moves towards coastal areas, lakes, and high latitudes with marginal ice zone conditions. RTM-based techniques may also degrade in those areas, as the prior information is also degraded here.

Recommendation: Improve the error characterization of current standard, regression based-techniques.

Recommendation: Explore alternative regression SST formulations, including regionally-tuned regressions, OE techniques, hybrid regression/physical algorithms.

Recommendation: Address the following open questions related to physically-based retrieval algorithms: Does promised improvement from OE experiments happen operationally? Can accommodation for various types of aerosols be improved? How accurately must the diurnal cycle of atmospheric parameters be and how do errors in NWP spatial/temporal resolution result in errors in SST retrievals? Several fast and accurate RTMs are currently being employed in the NWP community (e.g., RTTOV, CRTM). Their careful evaluation and validation against satellite radiances is needed to establish an RTM model acceptable for physical SST retrievals.

There are over 250 lakes over 500 km^2. Lake SSTs are different from open-ocean SSTs. There are different environmental effects (atmospheric corrections), a different emissivity for fresh versus salt water, and different physical conditions, such as upper layer heating. These different conditions require development of special algorithms to address the challenges that lake SSTs present, including specific validation challenges. Parameterization of environmental effects to reduce their influence on the error for lake SSTs requires more work on atmospheric correction and appropriate emissivity values for lakes.

Recommendation: Perform a radiometer round-robin at a lake to test Lake SST algorithms.

New sensors, e.g. NIRST on SAC-D, may not be appropriate for Oceans but are still very valuable for Land. Since the Ocean SST community is more focused on low resolution data (AVHRR, ATSR, MODIS), higher resolution data are under-utilized or ignored.

Recommendation: Explore use of high resolution IR data for lake SST retrievals.

2.2.1.2 Data classification

SST retrievals in the thermal IR are only made in clear-sky (non-cloudy) pixels. Historically, cloud screening techniques have been based on static thresholds (AVHRR, MODIS, (A)ATSR,and in future, VIIRS,). Recently, RTM-based techniques (Bayesian or dynamic thresholds) have begun being explored (GOES, AVHRR) or tested ((A)ATSR), but they have not become part of a widely accepted community-consensus approach for SST. Additionally, some IR products exclude cloud retrievals whereas some simply flag questionable data. Flagging is the preferred methodology.

There are inherent limitations to the threshold-based cloud clearing techniques employed in many current and future SST processing systems (AVHRR, MODIS, VIIRS). Dynamic thresholding and Bayesian screening techniques already demonstrate improved results due to better accounting for local conditions. However, more work is needed before these techniques can fully realize their potential on a global scale and in long-term perspective.

Recommendation: Continue development of both techniques. Further work to compare the accuracy of threshold-based versus Bayesian cloud screening techniques is required for both current and historical instruments.

Cloud flagging or masking algorithms used over oceans typically fail over land; therefore, development of custom cloud flag for water over land is necessary.

Recommendation: Develop improved cloud detection algorithms that are specific for regions over land. Ideally, these algorithms would be incorporated into global algorithms creating adaptable algorithms depending on land/ocean conditions.

Currently, the mainstream approach is to identify aerosol-contaminated pixels as a byproduct of cloud detection or quality control, and either flag or exclude them from SST analyses. Determining the SST, but flagging questionable data is preferred here as well. The data exclusion approach may be justified in heavy aerosol conditions which are indistinguishable from cloud. However, it is believed that in lower-aerosol cases, prevalent in many geographical areas and seasons, a correction can be applied. Tuning regression algorithms to buoys is one approach, but may result in errors of a few degrees in regions with high spatial and/or temporal variability in aerosol optical properties. RTM-based techniques used in conjunction with first-guess aerosol fields (GOCART, NAAPS) are demonstrating potential in some case studies. Much more work is needed to fully realize the advantages of the RTM-based techniques to correct for aerosols.

Recommendation: Continue research into this and other approaches for aerosols such as adaptable algorithms, bias corrects, and improved screening.

Errors in the land-sea mask also contribute to errors in pixel classification and retrieval errors. At this point, although several different land masks are used for different products, there is little community consensus on a ‘best’ mask. Current SST products either report an ice mask based on external analyses (AVHRR, GOES) or do not report one at all (MODIS).

Recommendation: Improve sea ice masking future products.

Recommendation: Construct an accurate, high resolution land sea mask for SST retrievals, especially in coastal areas and lakes. The mask should identify open-ocean, land, rivers, and, ideally, inter-tidal areas with an indication of expected water coverage at time of observation.

2.2.2 Microwave SSTs

2.2.2.1 Microwave Retrieval algorithm

Between 4 and 11 GHz the vertically polarized BT of the sea-surface has an appreciable sensitivity to SST. In addition to SST, BT depends on sea-surface roughness, atmospheric temperature and moisture profile. Fortunately, the spectral and polarimetric signatures of the surface-roughness and the atmosphere are quite distinct from the SST signature, and the influence of these effects can be removed given simultaneous measurements at multiple frequencies and polarizations. Sea-surface roughness, which is tightly correlated with the local wind, is usually parameterized in terms of the near-surface wind speed and direction.

Microwave SST retrieval algorithms that use all channels are used to simultaneously retrieve SST, wind speed, columnar water vapor, cloud liquid water, and rain rate (Wentz and Meissner, 2000). These environmental variables are calculated using a multi-stage linear regression algorithm derived through comprehensive RTM simulations. SST retrieval is prevented only in regions with sun-glitter, rain, and near land. Any errors in retrieved wind speed, water vapor, cloud liquid water may result in increased uncertainty in the retrieved SST.

Microwave SST retrievals are very sensitive to instrument calibration problems, satellite and instrument mounting pointing accuracy, RFI both near land and in the open ocean due to reflected satellite TV signals, and errors in wind speed and/or direction and some other geophysical fields. Instrument calibration, algorithm development, and product validation must be completed in a consistent, cohesive manner.

Errors in SST retrievals are primarily due to atypical atmospheric conditions that significantly deviate from the statistical mean of the algorithm training set. Since errors in other geophysical retrievals will affect SSTs, it is important to minimize errors in these fields. For example, AMSR-E winds are underestimated during the Arabian monsoons. The effect persists for several months during the summer and impacts SST retrievals.

Recommendation: Continue/expand research into reducing uncertainties through the use of ancillary input data, such as atmospheric profiles. Devote special attention to improved estimates of wind direction, a necessary ancillary parameter to correct for the measurement dependence on both wind speed and direction. Errors in the input wind direction, currently from 6-hourly 100 km NCEP NWP fields, result in errors in SST, usually visible around fast moving atmospheric fronts.

The correction for AMSR-E calibration needs to be revisited now that a significant amount of WindSAT SST data has been produced. At present, differences between MODIS and AMSR-E at high latitudes have not been attributed to an error in either instrument and their cause remains largely unknown. WindSAT improved calibration has the potential to be an independent data source to resolve these differences.

Recommendation: Continue and expand cross-sensor comparisons to improve our understanding of uncertainties in SST retrievals for many sensors.

SST uncertainties for µ-wave SST retrievals are currently estimated using static look-up tables based on environmental parameters (SST and wind speed) and daily GTS buoy match-ups. This methodology has received positive feedback from users but does not adequately describe all sources of uncertainty.

Recommendation: Improve the estimate of uncertainties to also include uncertainties derived during the algorithm training exercises as well as on-going improvements to our understanding of SST uncertainties.

2.2.2.2 Data classification

Currently, µ-wave SST retrievals are not completed in regions with rain, cloud, or near land. The preferred future methodology would retrieve SST but provide a flag for the data that indicates why they are questionable. Regions with sun glitter must be flagged in the SST retrievals. Currently for SST observations, sun angles less than 25° are excluded eliminating almost 20% of the observations for the daytime orbit segment. A sun glitter angle of 10° would exclude only 3% of the observations, which would represent a major improvement.

Recommendation: Investigate corrections that can be applied to reduce the effect of sun glitter, thereby allowing for a smaller sun glitter exclusion angle.

The amount of masked data can be reduced through use of adaptive algorithms that seamlessly switch between retrievals using different channels. This type of algorithm has the potential to reduce land contamination effects, reduce masking of geostationary satellite TV based RFI reflections, and improve our SST retrievals both near and in lightly raining environments. Each of these effects is a significant challenge and will require a specific type of adaptive algorithm to reduce SST uncertainties.

Recommendation: Develop adaptive algorithms to improve retrievals near land, in RFI affected areas, and near rain.

2.2.3 Special retrievals – Polar regions

In polar regions, IR and µ-wave SST retrievals differ and the reasons are not yet well understood. There is little *in situ* data at high latitudes where SST retrieval is complicated by cold air outbreaks coming from sea ice, anomalous air-sea temperature differences, and changes in emissivity.

Recommendation: Increase the number of high latitude in situ observations. Increase high-latitude SST inter-comparisons. Investigate adaptable algorithms that are able to *seamlessly switch to a specialized polar SST algorithm. Improve high-resolution sea ice for accurate flagging of sea ice data.*

The Aqua satellite has carried both IR and μ-wave radiometers, and it is possible that improved SST algorithms utilizing both IR and μ-wave data could be developed.

Recommendation: Research into multi-sensor algorithms.

2.2.4 Surface Emissivity (IR and μ-wave)

For physical retrievals of SST it is necessary to accurately know the emissivity of the sea surface in the corresponding spectral interval. For IR retrievals, the sea surface emissivity is very high and has a relatively weak dependence on view angle. While some research explored variations in the IR emissivity with changing wind speed and roughness (e.g., Harris et al., 1994), later results suggested that the decrease in emissivity with increasing wind speed is largely compensated for by additional reflections of surface-emitted radiation, and assumption of constant emissivity is generally sufficient (Watts et al., 1996). Various approaches to parameterizing the complex effect of surface reflections on effective rough-water emissivity have been tried, including a recent treatment by Nalli et al (2008a) that merits further appraisal.

For μ-wave SST retrievals, the lower emissivity values and stronger dependence on wind speed and roughness make accurate specification more important. The required relative accuracy of the sea surface emissivity in the μ-wave is about 0.05% to enable resolution of SST variations of 0.15 K, while an absolute accuracy of 0.1 to 0.2% is needed (F. Wentz, personal communication). In contrast, for the IR, the sea surface emissivity needs to be known with an accuracy of 0.5% to estimate the SST to within 0.3 K (Wu and Smith, 1997).

Recommendation: Research into the impact of water surfactants on the emissivity, the impact of foam and whitecaps, and potential limitations in the knowledge of the μ-wave emissivity at cold SST values.

2.2.5 Accurate and globally representative in situ SSTs

Accurate SSTs are needed to train empirical regression algorithms, or perform bias correction to both RTM-trained regression and physical OE algorithms. SSTs of the training data set should be globally representative to adequately cover the retrieval space, ideally measured as a skin temperature or, if at depth, a consensus on how to handle diurnal warming and skin effects should be developed and applied in a consistent manner.

Recommendation: Establish community consensus datasets, sampling methodologies and QC procedures to establish unified in situ SST for the use in satellite calibration (bias correction) and validation, which affects SST algorithms.

2.2.6 Future Instruments

There are a number of IR radiometers, useful for retrieval of SST, being launched by a number of different countries. Some of these, but not all, are discussed below. NPP, planned for launch in 2011, will carry the first VIIRS instrument as a successor to AVHRR. Also, the European MetOp series with AVHRR/3 on-board plans two more launches, in 2011 and 2016, with life times of ~5 years each, and will continue 1km resolution AVHRR in orbit into ~2020. GOES-R will carry a new ABI radiometer. The first satellite of the GOES-R series will be launched in 2015. The SEVIRI replacement under MTG will have a higher frequency of sampling of the full disk. MTG is to be available in 2018. MTG IR channels will be quite similar to those of MSG except the 3.9 μm will be redefined to be more suitable for SST retreivals. More visible channels will, in principle, lead to better cloud identification by day. The Sentinel program of ESA will provide circa 2013-2025 a pair of mid-morning dual-view (ATSR-class) radiometers (called Sea and Land Surface Temperature Radiometers) with a ~750 km dual swath and ~1500 km single view swath, 1 km spatial resolution. NASA decadal survey mission, CLARREO, will provide TOA radiances for absolute calibration. For the first time there will be an opportunity to get space-borne IR radiometers calibrated absolutely, as well as inter-calibrated, since the drifting orbit will collocate CLARREO measurements with those from all IR radiometers.

Several ongoing (but data not yet released) or future missions will continue the TMI and AMSR-E μ-wave SST record, including the Chinese FY-3A MWRI, the Japanese GCOM-W AMSR2 (launch February 2012), and NASA's GPM GMI (launch July 2013). The May 2008 launch of the MWRI on FY-3A is the first in a series of satellites designed to provide μ-wave ocean observations from 10.65 to 89 GHz. The L1 data are projected to be available to researchers in early 2010. This radiometer, if well calibrated, could bridge any data gap between AQUA AMSR-E and JAXA GCOM-W AMSR2, thus providing an important dataset both for data continuity and for understanding error characteristics of cool-water SST, a condition that will occur for NASA's GPM GMI measurements. Geophysical data from GCOM-W is a crucial link in the μ-wave climate data record. There are US members on the GCOM-W Science Team and the Japanese have agreed to allow data access for AMSR2. This mission continues the AMSR-E record: it will use an almost identical orbit, have very similar channels to AMSR-E, and an improved hot load for better calibration. SST retrievals from GPM will continue to improve our understanding of the diurnal cycle. The mission is a follow-on to TRMM and will have a similar orbit that slowly precesses through the diurnal cycle.

> *Recommendation: Calibration, algorithm development, and validation of future IR and μ-wave instruments should look for new approaches and strive for consistency with heritage instruments in order to maintain a continuous climate record.*

3. Derived SST Products

This section describes derived SST products and the components of their uncertainty budgets. Majority of these products are L3 and L4 gridded data sets, but there are some derived L2 products as well. The latter are swath-level data sets of temperature that has been corrected to a depth other than that of the sensor's retrieval. In general, derived SST products involve four types of data product modification:

1) adjusting L2 skin/sub-skin data to produce estimates of temperature at different points on the vertical temperature profile and/or at the different phases of the diurnal cycle (applicable to L2-L4 products);

2) gridding[9] the data by either simple remapping of L2 data granules a space grid without combining any observations from overlapping orbits (standard L3, or L3 Uncollated), or collating the data by combining measurements from a single instrument into a space-time grid (L3 Collated), or super-collating multi-sensor data by combining observations from multiple instruments into a space-time grid (L3 Super-Collated);

3) applying a bias-correction procedure based on multi-sensor and/or in situ data (L3 Super-Collated, L4, and some single-sensor L3 products);

4) performing an analysis procedure to interpolate gaps and to improve estimation of available values in gridded data sets (L4).

Models of vertical and temporal variability associated with diurnal warming processes and with the skin-effect are necessary for type 1 adjustments above. Errors of temperature adjustments calculated by these models are among major contributions to the total error of all derived products (L2-L4). These are shown among **Error Resulting from Oceanographic Variability** in Figure 1 as split into two terms: **Modeling Errors** per se and **Input Errors**, which are due to erroneous model parameters and/or ancillary information.

Space-time SST variability is responsible for a host of errors in SST products. Most prominently, variability on scales smaller than the gidbox size but larger than the sensor pixel size, creates **Representation Errors**, a major error source for gridding and analysis procedures (Figure 1) that produce L3 and L4 data sets (step 2 above). Space-time variability estimates at the scales larger than a gridbox plays a central role in producing estimates for missing gridbox values in L4 products (step 4 above), thus it affects **Gap Interpolation Errors** (Figure 1).

In general, the impact of geophysical processes on the SST uncertainty budget depends on the type of SST product. Of the geophysical processes considered in this section, diurnal warming appears to have the largest potential absolute contribution to the SST uncertainty budget (up to O(1K)), but only at specific times (near local noon) and specific environmental conditions (low wind speeds). The skin effect is still not completely “solved” from a physical point of view, but residual uncertainties for satellite SST retrievals are an order of magnitude smaller than that for diurnal warming effects.

---

9. Spatial dimensions of grids in L3 and L4 definitions only include lateral directions, not the vertical one.

Spatial variability effects have an important impact on a wide range of SST products and can also be large particularly in coastal regions, shelf seas, frontal zones, and areas influenced by diurnal warming. Despite its potential impact, the spatial variability of the SST on scales relevant to SST product generation remains relatively poorly characterized. Overall, for the generation of gridded SST products, averaging over the full range of environmental conditions, the expected contribution of geophysical processes to the SST uncertainty budget is expected to be on the order of 0.2-0.3 K.

Section 3.1 presents contributions of near-surface ocean temperature variations to the uncertainty budget of the L2-L4 SST products. Section 3.2 describes the motivation, necessary techniques, errors, and outstanding problems associated with producing L3 and L4 SST products.

3.1 Role of Ocean Variability in Uncertainties of SST Products

Three separate components within geophysical processes at the ocean surface were identified as sources of uncertainty in derived L2-L4 SST products: diurnal warming, skin layer effects, and spatial variability. The first two of these effects create temperature variability that is modeled, with some error, and is used for the SST adjustments of the type 1 listed above. Spatial variability affects the error in gridding and analysis (procedures 2 and 4 above).

3.1.1 Diurnal Warming

Upper layer ocean warming by solar radiation that occurs during the daytime, relative to cooler water temperatures at night, is referred to as diurnal warming. Its amplitude and the depth to which it propagates is strongly influenced by the strength of the upper ocean mixing. The most significant warming events are observed only at low wind speeds (e.g., < ~3 m/s) under strong insolation, but warming is clearly present up to 6 m/s. Considering that 30% of global winds are less than 6 m/s (Donlon et al., 2002), daytime SST measurements should always be suspected of containing an element of diurnal warming.

Although diurnal warming of the ocean's surface has been a subject of research for many years, the exploration of its impacts on the retrieval of satellite SST has been a focus of attention only relatively recently. General investigations have considered characterization of diurnal warming including the evolution of the diurnal heating profile and the frequency, timing, and preferential location of diurnal warming events. The problem is significantly hindered by the lack of information on the continuous time history of all the parameters influencing the warming and a lack of agreement between the empirical and physical models. The production of a meaningful diurnal warming analysis requires significant further development of diurnal warming models, analysis of diurnal warming model errors, and validation against both satellite and *in situ* measurements of diurnal warming.

Diurnal modeling uncertainty is due to the propagation of uncertainties in the required forcing fields through the models as well as deficiencies in the models themselves. As such, diurnal warming contributes to both modeling and input errors included into the *Error Resulting from Ocean Variability* box of Figure 1. Comparisons of simulations of diurnal warming with satellite and in situ observations provide bounds on the current overall uncertainty related to predicting diurnal warming. To reduce uncertainties in diurnal warming predictions, further model improvements are clearly essential. Improved understanding of the role of penetrating radiation and its relationship to available optical properties in the water column is needed. Better predictions are also needed of the warming at arbitrary depths including derivation of the foundation temperature.

Recommendation: Continue/expand research into development of diurnal warming models and analysis of satellite observed diurnal warming

Recommendation: Continue and expand research into the role of penetrating radiation and its relationship to available optical properties in the water column.

To facilitate application of existing and refined models, improved specification of the physical forcing parameters is necessary. The available accuracy and temporal and spatial resolution of model forcing parameters (e.g. wind speed) is a critical limitation in current diurnal warming predictions. Presently, global sub-daily representation of these parameters is largely limited to numerical weather prediction outputs with variable and sometimes ill-defined quality.

Recommendation: Improve specification of diurnal model uncertainty

Diurnal variability in specific geographical regions (e.g., high-latitudes), different reference depths, different product types, different spatial resolutions, and analysis procedures are also inadequately understood. To progress, additional detailed field observations of diurnal warming events are required for model validation purposes. These may entail new process studies and/or technological developments in instrumentation to measure the near surface temperatures and/or temperature profiles.

Recommendation: Perform additional field observations of diurnal warming events.

Finally, improved understanding of uncertainty in diurnal warming and its impacts could benefit from more interactions with additional research communities. Joint studies considering other air-sea interaction processes would help demonstrate their impacts on predictions of diurnal warming and identify the required accuracy. Exploration of feedbacks between diurnal warming and the initiation of atmospheric convection are desired within the applications community and could provide insights into the role and timing of subsequent cloud cover and precipitation. Continued collaboration with the ocean color community would support further advances in understanding the role of turbidity and the influence of optical properties of the water on predicting diurnal warming.

Recommendation: Develop/enhance links with other communities with interests in diurnal warming such as the ocean color and the meteorological communities.

3.1.2 Skin Effect

Temperature gradients arise across the skin layer of the ocean since the turbulence is damped near the interface and the heat exchange with the atmosphere is accomplished through molecular processes. Temperature differences across the skin layer are typically about 0.2K, but can exceed 0.5K at very low wind speeds.

When the target accuracy of satellite SSTs was ~0.5 K, the skin effect was generally ignored, but with recently improved accuracies, and with ever more stringent future accuracy requirements, the contribution of the skin effect to the error budget of satellite-derived SSTs cannot be neglected. While simplified modeling approaches are a good first step, coupling of skin layer processes into general circulation and air-sea interaction models require its more complete description. The precise nature of remaining skin effect dependencies on such parameters as heat flux components, SST, presence of surfactants, and wave processes have not been determined. In particular, it is important to establish whether the behavior of the skin layer during the day can be treated in the same way as during the night.

Recommendation: Undertake additional research into the physics and modeling of the skin effect.

To take advantage of any modeling improvements or more fully utilize the existing models, an improved specification of the input parameters is necessary. The propagation of their errors through models must be better quantified as well. Ultimately, additional studies will be necessary to determine if improved physical models and inputs can result in a significant reduction of uncertainty.

Recommendation: Improve our understanding of how errors in model forcing parameters impact skin model errors

3.1.3 Spatial and/or Temporal Variability in SST

Space-time variability in SST plays important role in the error affecting L2-L4 SST products.

3.1.3.1 Level 2: impact on the retrieval error

Incompatibility between a point sample and a spatial average is bound to introduce uncertainty in the comparison between in situ and satellite SSTs, irrespective of the performance of the retrieval algorithms being used. In general, space-borne sensors measure an integrated value of the emitted radiation over the IFOV of the satellite, whereas the in situ measurements from buoys are sampled at a point and averaged over some short time interval; measurements from radiometers on moving ships can provide a better estimate of the appropriate spatial average value. The mismatch in the sampling capabilities introduces uncertainties (**Sampling Errors** in

Figure 1) in the satellite retrieval when the phenomenon controlling the variable being measured has a length scale that is larger than the field-sampling instrument size but smaller than the IFOV (sub-pixel variability). The additional error of retrieval algorithms that use a regression of situ data on remotely sensed brightness temperatures is created by the SST variability on scales of the accepted spatial and temporal tolerance for the mismatch of satellite and in situ data.

Recommendation: Develop methods to integrate an improved understanding of the SST variability into the SST retrievals error.

An important aspect of the impact of spatial variability on SST uncertainties can be expressed as a need for an improved description of spatial SST variability and its dependence on environmental conditions. A complete description of the variability of SST requires comprehensive measurements on all scales in the spatial domain of interest as well as the temporal evolution of the phenomenon. Efforts to collect high-resolution data of the SST at sub-satellite-pixel scales should continue.

Recommendation: Expand the library of sub-pixel scale SST data sets.

On a longer time scale, development of a high-resolution μ-wave sensor could be considered to facilitate better representation of spatial variability in SST products in regions of persistent cloud cover.

Recommendation: Investigate development of a high-resolution μ-wave sensor.

3.1.3.2 Levels 3 and 4: Contribution to representation error.

Intra-gridbox variability, i.e., physical variability of SST on the the spatial scales between the L2 pixel and the gridbox size and temporal scales between the time span of an L2 granule and the temporal dimension of the gridbox is solely responsible for the representation error of individual L2 values, when each of them is used as an individual estimate of the "true" gridbox average. These representation errors are in general not independent because of autocorrelation of space-time SST variability at these scales. Both variance and autocovariance of this variability are crucial parameters for producing better L3 and L4 products with more realistic error estimates.

Recommendation: To improve estimates of variance-covariance structure for representation error, based on the improved estimates and understanding of intra-gridbox space/time variability.

3.1.3.3 Level 4: A priori estimates of SST autocovariance

SST variability on scales larger than gridbox size does not contribute directly to the error of L4 products. Instead, its autocovariance estimate is used in the OI formalism and, unless available observations have very low error and provide complete grid domain coverage, controls all aspects of the solution: infilled values, smoothness of the analyzed field, and its error. Conversely, inaccurate knowledge of SST autocovariance and the incorrect assumption of its shape has a detrimental effect on all aspects of the L4 product, including its theoretical and actual error.

Recommendation: To improve estimates of the variance-covariance structure between SST averages over typical gridboxes of L4 products on the basis of recent/expected space-time SST variability research efforts.

3.2 Merging, Gridding, and Analysis of Derived SST data sets

Contributions of surface ocean temperature variability to the error budget of SST products have been described in Section 3.1. Here we describe the progression of these and other types of error through merging, gridding, and analysis procedures.

3.2.1 Problem Description and Background

Widening demand for gridded SST data sets. Besides their utility for understanding the role and impact of the oceans on the global climate system, gridded SST products are demanded by a growing number of societal activities and needs related to the use of the world ocean resources. As population and industrial activities increase, there are pressing needs for accurate knowledge and prediction of the state of the ocean. High resolution SST products that are synoptic in space and time as represented by daily or sub-daily gridded fields are important tools for achieving these goals. With respect to societal needs, there are various and diverse requirements related to offshore industrial activities, national security interests, marine ecosystem management, and recreational activities in the public interest. For example, with respect to industry, SST products are proxy indicators of surface circulation that is a crucial factor for planning and operation in oil and gas exploration and extraction and for ship navigation. As a practical example, oil companies operating in the Gulf of Mexico are interested in monitoring intrusions of the northern wall of the Loop Current during which they must curtail their oil platform operations because of a variable surface current on the order of several knots. Other potential needs are related to navigation and ship route planning for the shipping industry, undersea telecommunication cable deployment, national security, and search and rescue operations, since near real time SST data can be utilized to indicate the position and scale of mesoscale eddies and filaments as well as current shear zones. From the perspective of coastal and pelagic fisheries, a gap-free SST product can be utilized for identifying preferred fish habitats, potentially reducing operational costs related to time, fuel and maintenance. Finally, these types of products are usable for tourism planning and promotion, recreational activities encompassing preferred areas for fishing, swimming, surfing, and for other ocean recreational activities in the public interest.

Recommendation: To develop approaches to estimating uncertainties relevant to non-linear statistics appearing in societal applications, e.g. uncertainty (and bias) of SST *gradient magnitudes estimated from Level 4 products and uncertainty in flux estimates based on L4 SST products.*

Target variables for gridded SST data sets. Upper ocean (~10 m) has a complex and variable vertical temperature structure (Donlon et al. 1997). Due to the large vertical gradients in the surface laminar sub-layer, the temperature of its upper 10 μm-wide layer, skin temperature SST_{skin}, is colder by a few tenths of ^{o}C than the water at a few millimeters depth. In practice, the SST_{skin} is measured by infrared radiometers; it is one of parameters governing the air-sea fluxes. The temperature at the base of the laminar sub-layer, the sub-skin temperature $SST_{sub\text{-}skin}$, is close to the temperature measured by microwave radiometers, which detect radiation from an upper centimeter of the water column. $SST_{sub\text{-}skin}$ is a parameter simulated by ocean models that resolve the diurnal cycle. While SST_{skin} and $SST_{sub\text{-}skin}$ are directly measured by remote sensors, SST values corresponding to larger depths have to be estimated on the basis of remote observations of skin and sub-skin depth levels.

The mixed layer temperature (SST_{ml}) is a traditional parameter used in dynamical oceanography. The ocean models that do not include the diurnal cycle, e.g., the majority of climate ocean models, have to use the mixed layer concept. SST_{ml} is the vertical average from the surface to the mixed layer depth (MLD), which is usually defined as the shallowest depth where the net temperature difference ($|\Delta T|$) with the reference level temperature (z=10m) exceeds a certain value, e.g., $0.2^{o}C$.

Observations of SST_{ml} were usually obtained as bulk SST (SST_{bulk}), a term used for a long time in reference to traditional in situ SST measurements. These are taken in the depth range from 1m to 5m by drifting buoys and by ship buckets and in the wider range of depths, from the surface to 10m and 25m, by ship hull sensors and engine room intakes, respectively (Kent and Taylor, 2006). Night-time SST_{bulk} does represent the mixed layer temperature, except in regions where stable haline stratification occurs within the depth range of in situ measurements (Grodsky et al. 2008). In such regions SST_{bulk} will depend on the actual depth of an in situ measurement. Moreover, the MLD definition given above breaks down as well.

Foundation temperature SST_{fnd} is the water temperature at the depth not affected by the diurnal temperature variability (either daytime warming or nocturnal cooling). Defined this way, SST_{fnd} should represent precisely the mixed layer temperature (SST_{ml}). However, its calculation from remotely-sensed parameters has to rely on models or parameterizations of the diurnal cycle.

Cases of shallow (within top few meters) density stratification combined with a possibly deeper propagation of diurnal warming in low-wind condition present a conceptual problem for the uniform applicability of formal definitions and calculation algorithms used for SST_{fnd} and SST_{ml}. "True" values of SST_{bulk} become significantly dependent on the measurement depth. For complicated cases like that, general differences between SST_{bulk}, SST_{fnd}, and SST_{ml} need further examination. Further, the growing variety of SST-related products and their validation needs, combined with advances in ocean modeling and models of the ocean surface layer, suggest that representing the 3-D temperature (SST3D) from the skin layer to the seasonal mixed layer (rather than separate products for each SST type) should become a goal for the future SST product development.

> *Recommendation: Further research towards clarification of SST_{fnd}, SST_{bulk}, and SST_{ml} relationship when there are significant predawn temperature variations in the top few meters of the ocean is needed.*

> *Recommendation: A gradual shift towards simultaneous estimation of the entire surface depth profile in L4 products is warranted; a reasonable first step is to start reporting simulated diurnal variability solutions on which SST depth profile adjustments are based.*

3.2.2 Level 3 Products and Their Uncertainty Characterization

The concept of uncertainty for an SST value only has meaning if the target variable is defined unambiguously, i.e., as the water temperature at a certain position on the vertical profile (e.g., skin SST, foundation SST, temperature at 1 meter depth) or as the average temperature over a certain depth interval, e.g., over the mixed layer with unambiguously defined depth. For gridded data sets, the "true" gridbox values are averages of the target temperature over the space-time gridbox volume.

L3 data sets are gridded by averaging all available observations in each space-time gridbox. Errors of their grid values are due to errors in the L2 products that were used as source data for producing these grid box values. In addition to L2 own error (total error of retrievals), these source data errors also include: **model errors** due to SST value adjustment and gridbox **representation error** (Figure 1). Physical nature of these additional errors are discussed in the beginning of Section 3 and in Section 3.1.

The SST value adjustment errors arise if, for example, a gridded product is expected to represent temperatures at a depth different from the skin or sub-skin depth corresponding to the L2 data source: due to imperfect modeling of the near-surface temperature profile, this adjustment of the SST value will contribute to the error. Similarly, the error will be committed if, for example, a diurnal variability model is used to predict SST at the time shifted from the actual time of observation. Therefore this error is called **modeling error** of the value adjustment; it is generally accompanied by the model **input error** that includes all possible effects of errors in the ancillary data used by the model. At present only specialized, relatively simple models of skin layer and diurnal variability used for this type of SST adjustments.

When individual L2 SST values are used as estimates of space-time grid box mean, to the extent that the measurement footprint and time interval do not coincide precisely with the grid box, the measured SST has an error in representing the grid box – **representation error**. Note that representation error is caused by both spatial and temporal variability, but temporal representation error for depth intervals that feel diurnal variations can potentially be reduced by modeling diurnal cycle and thus replaced by the **modeling error** and **input error** contribution to the temporal average.

A space-time grid box value of an L3-Collated or L3-Supercollated data set is computed as an average (with weights if necessary) of all L2 data available within that grid box. The uncertainty of this average depends on the covariance matrix of the total error (sum of sensor, adjustment, and representation errors) of the data that were averaged. Often the errors of values being averaged are treated as uncorrelated. This assumption simplifies the procedure but will result in underestimating the L3 product error since the errors of adjacent pixels are generally correlated.

> *Recommendation: To start accounting for correlated data error in implementations of data gridding procedures.*

Strongly correlated error components are not reduced by averaging. Their correlations might extend much farther than a single grid box, towards regional or seasonal scales, depending on the mechanism responsible for a correlated error type. Such correlated errors are responsible for **biases** – a major source of error in gridded data sets, which generally cannot be reduced statistically strictly within individual L2 data sets. Most of the biases come from a systematic, non-random component of the retrieval error. Caused by atmospheric or other effects they might have regional and/or significant temporal extent. Inter-sensor bias correction procedures and/or bias correction using in situ data are necessary before merging L2 products together. Most of current bias correction procedures were developed empirically and do not provide rigorous estimates of the remaining biases. If incorrectly implemented, they potentially could increase biases in certain areas. The resulting error of bias correction procedures is identified as **Bias Correction Errors** among components of **Merging, Gridding and Analysis Errors** in Figure 1.

3.2.3 Level 4 Products and Their Uncertainty Characterization

In contrast to L3 products, L4 data sets are outcomes of analysis procedures that provide SST values at all gridboxes, including data gaps, i.e., gridpoints where no L3 estimates could be available. These values are "predicted" on the basis of available SST values at other times and places, using some procedure generally called "analysis". L4 products are computed either directly from L2 data sources or by the analysis of an intermediate L3 data set. In either case, all error types that, as described in the previous section, contribute to the L3 product error enter the analysis procedure as well and affect, in some transformed form, the resulting L4 product too.

Most of currently used analysis procedures that produce gridded SST fields on the basis of irregularly sampled satellite observations employ a formalism based on the Gauss-Markov theorem that simultaneously minimizes the mean squared error of the SST estimates at all times and locations. The resulting estimates have the minimum possible mean squared error among all linear estimates, i.e., estimates formed as linear combinations of the available SST observations. This formalism requires **specification of the spatial and temporal autocovariance functions** (or equivalently, the variances and autocorrelation functions) for both the SST signal and the measurement errors. While these procedures are usually referred to as optimal interpolation (OI), they are in fact optimal only if these variances and autocorrelation functions are rigorously correct. However, for practical reasons, it is usually not possible to use such rigorously correct specifications. Therefore the practical implementations of the OI methodology are referred to hereinafter as Objective Analyses (OA), rather than OI.

The OA implementations do employ OI methodology: assuming a priori a certain **form of spatial and temporal autocovariance for grid box averages of true SST values** (i.e., signal autocovariance function) and utilizing input data error estimates as described in Section 3.2.2, they produce estimates for the SST values in all space-time grid boxes as well as error estimates for them. In addition to infilling data gaps with these estimates, OA also provides similar estimates for gridboxes where SST observations were available. However, since these are usually the most skillful predictors for the SST average over the gridbox that contains them, such L4 values are often close to the co-located L3 values that are produced from the same L2 sources. The L4 product uncertainty for such values is slightly reduced compared to the uncertainty the L3 product, because OA prediction based on SSTs elsewhere do provide additional information. For infilled data, however, OA prediction is the only source of information; in the absence of local data, the uncertainty of these predictions, based on the SSTs in other places and/or at other times, is large: it combines the prediction error with the error of the observed SSTs that were used as predictors. Therefore **Data Gaps Interpolation Error** (Figure 1) is the major source of error in L4 products: the uncertainty of infilled values is usually larger than that of SST observations and always larger than L4 uncertainty for space-time grid boxes where observations were available.

The quality of the OA-based L4 products and reliability of their error estimates depends on **correct specification of prior parameters** in the OI formalism: data error estimates, including representation error estimates, and OI covariance estimates. The latter two statistics are controlled by space-time SST variability on different scales (Section 3.1.3.2-3). The actual non-stationarity of the SST variability creates a major conceptual problem for the use of standard OI implementations in the SST analysis problems. What is the effect of misspecified OI parameters and how to obtain reliable uncertainty estimates in the face of this difficulty remain outstanding issues.

3.2.4 Outstanding Issues

3.2.4.1 A Priori Parameters of OA Procedures and Sub-Optimality

There are many reasons for the sub-optimality of the OI implementations used to produce SST analyses for the L4 products. For example, use of the true autocorrelation function often results in few or no observations within the space and time spans of the estimate, thus resulting in an SST estimate that is close to the assumed background field (e.g., climatology). About 70% of the ocean's surface is cloud covered at any given time, obscuring the sea surface in infrared observations. In order to accommodate the often sparsely distributed infrared satellite observations of SST and to obtain gap-free estimates of the SST field on a regular latitude-longitude global grid at every estimation time, yet to avoid clear homogeneities of the analyzed field (jumps in variance between areas with and without data), the autocorrelation scales used in the analysis procedures are generally prescribed to be much longer than the true autocorrelation scales of the SST field. Many SST analyses now incorporate microwave observations of SST, which are available in all but raining conditions. This allows specification of shorter spatial and temporal autocorrelation functions in the Gauss-Markov procedure, but imposes the resolution limitations by virtue of the fact that the footprint size for microwave is very coarse (~50 km vs. 1-10 km for infrared observations of SST).

A well-known difficulty for the correct implementation of the OI formalism is that the true variance and spatial and temporal autocorrelations of the measurement errors are not really known. But the underlying problem and perhaps the intrinsic difficulty limiting rigorous implementation of the OI formalism is that these parameters, both for the true signal and for the error, in fact, vary geographically and temporally in complicated ways. Many SST analyses incorporate only crude estimates of this geographical variability. For example, the autocorrelation length scales used in the analysis procedures often decrease with increasing latitude, reflecting the decrease of the Rossby radius of deformation that qualitatively characterizes the scales of oceanic variability. Some analysis procedures scale the autocorrelation function inversely with the local SST variance, thus imposing shorter autocorrelation scales in regions of higher SST variance, as is usually observed in eddy-rich regions.

Temporal variability of the signal and error autocovariance functions are often not taken into consideration at all. In reality, they vary predominantly seasonally, but often also vary on event time scales. In eastern boundary current regions, for example, the space and time scales of SST are much shorter during upwelling events. Measurement error characteristics can also vary both seasonally (e.g., from water vapor contamination) and on event time scales (e.g., from aerosol contamination associated with volcanic eruptions or dust storms).

The net effect of using signal and error variances and autocorrelations that are not rigorously correct or incomplete is that the error estimates obtained from the Gauss-Markov formalism are at best only qualitatively correct. Hence a rigorous theoretical assessment of the accuracy of L4 products presents a significant challenge.

> *Recommendation: Research on the impact of the misspecification of a priori OI parameters (signal and error autocovariances) on OA fields and their error estimates.*

3.2.4.2 Advanced OA procedures

At present, several merging algorithms in operation (or considered for operation) are calling for the sequential application of the OI procedures at different scales. The main reason for the multiple interpolation scales is to address the drastically different resolutions in the L2P data sets, e.g., ~50 km for μ-wave and 1 km for the highest resolution IR. Typically the data sets are interpolated at the coarsest scale first, moving hierarchically to finer scales to incorporate higher wave-number components in the data. An advantage of such “multi-resolution analysis” is a proper accounting of the SST energy spectrum by means of a few simple traditionally-implemented steps. For example, the wave-number spectrum of SST tends to exhibit the "power-law" characteristics (a straight line on a log-log plot) with a slope of approximately -2. Reconstruction of such power-law characteristics is difficult to achieve in a single-step procedure.

Some SST analysis procedures address the scale dependencies of the spatial and temporal autocorrelation functions by noting that the finer-resolution SST features often have shorter time scales due to advection. "Motion compensation" or "data registration" approaches can facilitate utility of the irregularly sampled infrared data. For example, a space-time correlation function used in the OI formalism can be parameterized by an advection velocity field.

These techniques represent advanced applications of the OA. While each technique has a potential to reduce analysis errors in the final product, the number of components that potentially contribute to the error budgets increases. For example, in a multi-resolution approach mentioned above, multiple objective interpolation modules can be used hierarchically. In a motion compensation approach, the errors in the advection velocity data would contribute to overall error budget. The merging and OA algorithms employing these and other advanced techniques must be able to account properly for the error contributed by each module and also to take into account correlations between these individual contributions to the total error budget.

Recommendation: Further research into practical ways to deal with the major conceptual problem of OA, the statistical non-stationarity of the true SST field: accounting for non-stationary structures via covariance dependence on location, season, various parameters (e.g. velocities), types of weather/climate state; further development and use of multiscale OI formulation; use of simplified additional dynamical constraints in the OA of SST fields.

3.2.4.3 Inter-Sensor Bias Correction

Sensor biases pose a major problem in merging of SST data sets. An important aspect of inter-sensor biases is due to the regional interferences in the single-sensor correction procedures: retrievals are contaminated depending on the sensor type (IR by clouds and aerosols, μ-wave by precipitation and proximity to land or ice), and atmospheric and other effects that contribute to the contamination are dependent on space and time. The single-sensor bias correction is done typically by using coincident in situ (e.g., buoy- and ship-based) data as the reference SST values. Since the sampling pattern varies widely among the satellites, the set of coincident in situ data would be different from sensor to sensor. Such differences can be significant enough to leave some regional signatures, which are compounded by regionally dependent atmospheric conditions. Figure 2 shows the zonal difference for daily nighttime observations for a three-year period with respect to the daily OI AMSR+AVHRR analysis of (Reynolds et al., 2010) clearly showing the rather large variability of the mean bias between products - this on top of the orbit to orbit variability in bias. Nominal difference in single-sensor SST products is on the order of 0.1°C but can be several times more regionally (such as in sub-polar seas). To reduce such discrepancy, inter-sensor bias correction is the necessary first step of the merging procedure.

At present, inter-sensor bias correction is usually performed using a reference SST data set. A common choice for the reference data set is a combination of in situ data from moored and drifting buoys as well as ship-based SST measurements (Reynolds et al., 2007). Another choice for the reference is a satellite product from the ATSR series of instruments such as AATSR (Le Borgne et al., 2009) whose dual view sensors can reduce biases over other IR products such as AVHRR. Since each of these reference data sets has its own sampling pattern (usually irregular), it has to be interpolated and smoothed at some appropriate spatial scales. Although the interpolation and smoothing procedures would inevitably introduce numerical errors into the merged product, these errors tend to be random in nature and can often be contained. Interpolation scales of approximately 5 degrees in space and 5 days in time are common in practice. With appropriate interpolation for collocation, inter-sensor bias can be halved to a 0.05°C level or less on average, and regional discrepancy (e.g., larger average bias in high latitudes than low latitudes) can be significantly rectified.

Representativeness of the reference SST data is an outstanding issue for the inter-sensor bias correction. For example, because of the relatively narrow swath width, AATSR may not be able to resolve synoptic scales (of a few hundred km). Also, the collection of presently available in situ data is still too sparse to provide adequate sampling for the finest scales measurable by satellites (e.g., 1 km for MODIS). Finally, lack of redundancy in the reference data set can be a problem. For example, use of a single satellite data as the reference can also lead to issues of reliability (in weather forecast operations) and consistency (in climatic data record). To enhance reliability and representativeness, a combination of multiple data types is being used as the reference in some practices. In such an approach, determination of the reference data itself can become a merging procedure. Another approach may be to use ensemble comparison among at least three distinct merged products in order to detect and correct errors a posteriori. Here, the assumption is that the bias-induced error appears only in the clear minority in the ensemble, which may not be true. Further work, in both algorithm developments and experimentation, is needed to establish a reliable procedure in all these approaches.

The use of a common reference in inter-sensor bias correction introduces statistical dependency among the data sets to be merged. The merging algorithm thus must be able to account for such correlations among its components. An outstanding issue in the present merging practice is also to properly account for the contributions from the inter-sensor bias correction procedure in the a posteriori merging error of L3 and L4 SST products.

> *Recommendation: Research efforts into optimizing inter-sensor bias correction, objective choices of data input and reference sets for producing L4 products with the robust bias correction as a goal, intercomparison of ensembles of different L4 products.*

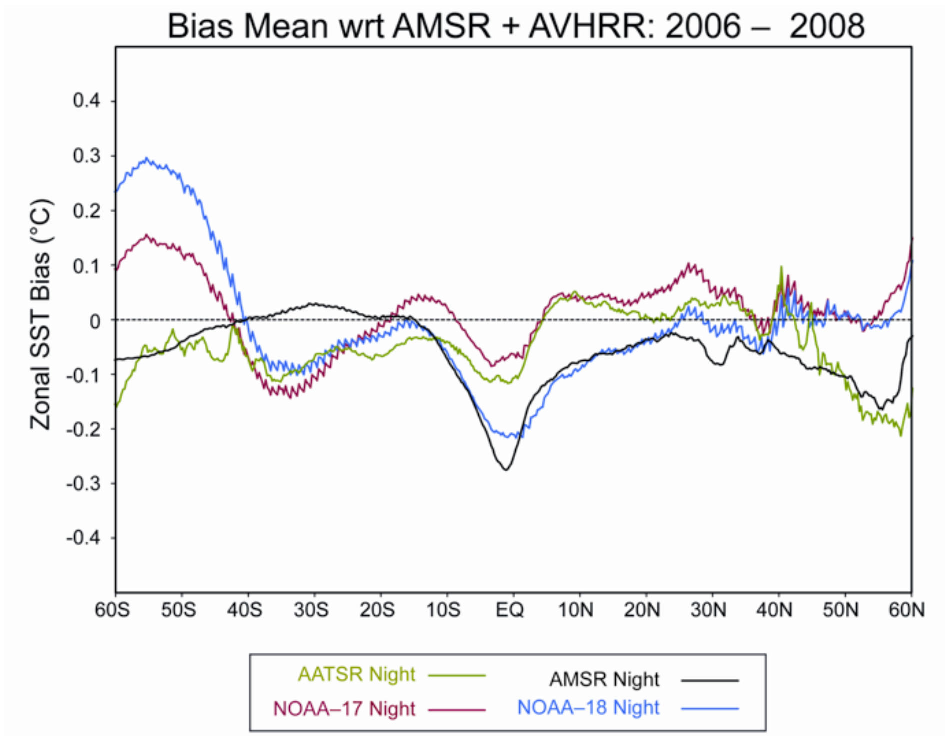


Figure 2. Zonally averaged nighttime satellite bias, for the period 1 January 2006 - 31 December 2008, with respect to the daily OI obtained from AMSR and AVHRR data. The satellite data are from the AVHRR NOAA-17, AVHRR NOAA-18, AATSR and AMSR instruments. Reproduced from Reynolds et al., 2010.

3.2.4.4 Merging pre-satellite and satellite SST record

Since the late 1970's, satellites have provided global SST observations at frequent temporal (daily to sub-daily) and moderate spatial (<1 to 25 km) resolution. However, the length of the satellite observation period is barely adequate for climate studies. Long time series are needed to detect patterns and trends of the climate change, and to distinguish it from decadal variability (e.g., Pacific Decadal Oscillation). The obvious solution is to merge data from the satellite and pre-satellite period. For the pre-satellite period, ship and buoy SST observations are available since the 1700's but statistical analysis can be attempted only after the 1850's due to the limited sampling. These in situ measurements are sparsely distributed in space and/or time, and the methodologies for each ship and buoy drastically differ over space and time. In contrast, there are only a handful of satellite-based SST instruments, and each dataset is massive but has particular biases associated with the instrument design and SST algorithm. Thus, when the two data types are merged in a standard way, satellite biases tend to have disproportionate influence over the resulting SST field, due to the sheer amount of data, and the persistent bias. For example, version 3 of the Extended Reconstruction SST product (Smith et al., 2008) from 1854 to the present was originally developed to include AVHRR data beginning in 1985. However, the AVHRR data contain a small residual bias (.01 to 0.40 C in the global average). This proved unacceptable to the climate monitoring users because it changed their rankings of the warmest months (as compared to in situ data only analysis, Figure II.1), and the satellite data was removed. Thus, at the reduced temporal (monthly) and spatial (2 deg grid) scales of this climate product, the advantages of using satellite data are offset by the unresolved bias.

Recommendation: To develop systematic approaches to the homogenization of analyzed data sets and their error estimates when observational data inputs undergo a major change in data availability.

3.3. Error Propagation through SST Data Processing Levels

3.3.1 Level 1 to Level 2

Application of retrieval algorithms to generate navigated SST products from the satellite brightness temperatures is influenced by geophysical processes to different degrees, depending on the satellite input, desired product, and retrieval methodology. In general, when the desired SST product does not precisely correspond to what the satellite sensor observes, the derived SST product will contain inherent variability due to geophysical processes. For retrievals of the skin SST from measurements sensitive to the skin temperature, there will be no contribution of geophysical effects to the uncertainty budget. If, however, estimates of the diurnal warming and skin layer cooling (both the magnitude and uncertainty) corresponding to that retrieval are derived, then the uncertainties in these processes will be applicable to the computation of those estimates. For satellite retrievals based on regression to subsurface temperatures, although the mean effect of the near-surface temperature profile is removed, variability in the near-surface temperature profile will be a source of random error in the derived temperature. For retrieved skin temperature estimates that are then explicitly referenced to a subsurface value, such as the AATSR Reanalysis for Climate (ARC) L2 drifter depth product, uncertainty in the adjustment for the near-surface temperature profile then contributes to the product uncertainty (both in bias and rms). For validation and generation of regression algorithms, differences between a point measurement or short time average and a spatial average will introduce scatter irrespective of the performance of the retrieval algorithm being used. Subpixel variability is an important factor in comparing point observations with area-averaged satellite observations. Uncertainties in the surface emissivity contribute directly through the retrieval process.

3.3.2. Level 2 to Level 3

While conversion from L2 to L3 generally incorporates the gridding of SST estimates from satellite coordinates onto a regular grid, several distinct sub-levels for L3 products have been defined. The uncertainty contribution from geophysical processes varies between these sub-levels. An L3 "single scene" product entails just the gridding of a single satellite scene or swath. An L3 "collocated" product involves the combination of multiple scenes or passes from a single sensor onto a common grid. An L3 "supercollocated" product then further adds the combination of multiple sensors. For all three types of products, spatial variability in the SST will contribute uncertainty due to the partial data coverage of the gridbox volume. With the collocation of data from different times, temporal variability then contributes to uncertainty as well. When there is an adjustment of the reference depth in the transition from L2 to L3, the near-surface temperature profile effects contribute to the uncertainty too. If, however, products with different effective sampling depths are combined in the supercollocation process, then the existence of gradients in the near-surface profile will be a source of uncertainty in the resulting product.

3.3.3. Level 3 to Level 4

Near surface temperature profile and temporal and spatial variability effects are major contributors to the uncertainty associated with the analyses applied in the formation of L4 products. One portion of the analysis involves the referencing of the individual SST retrievals at different times and effective depths to a representative time and depth. The target depth may be a single value such as the foundation or 1-m value or a depth average or integral value. Similarly, the target time can be a single instant (e.g., predawn) or some time average. For this step in the analysis, uncertainty in the time evolution of the near-surface profile clearly is the primary source of uncertainty. Another L4 product could include a time-resolved analysis where, for example, a single daily foundation temperature analysis is used as the basis for hourly estimates of the skin or 1-m temperature throughout the analysis grid. The primary additional source of uncertainty in this process would be associated with the incomplete knowledge of and inability to accurately predict the time evolution of the near-surface temperature profile. Another component of the analysis is mapping the data to some regular grid with a size typically different from the resolution of the input satellite retrievals. Spatial variability is a main contributor to the representation error in this process and in the error in infilled values (gap interpolation).

4. Validation

There are several aspects to post-launch validation but in each has a common approach using comparisons between retrieved SST and independent measurements with known uncertainty characteristics. Following “first light” there is an intensive period: the commissioning phase or acceptance period. During this activity the basic behavior of the sensor, corrections of instrumental artifacts based on the pre-launch calibration and characterization and instrument model, and corrections for the effects of the intervening atmosphere are determined. This typically lasts 90 - 180 days. Following this phase is a longer period of at least one annual cycle to determine more rigorously the uncertainties of the retrieved SSTs, and to indentify signatures of residual instrumental effects or atmospheric correction errors. Then, throughout the entire mission, comparisons with independent data are conducted to ensure the effects of possible instrumental degradations are identified. At all stages, analysis of the results of the comparisons can lead to improvements to the instrument model, atmospheric correction algorithm or processing procedures which can be applied when the satellite data set is reprocessed to generate more accurate SST fields.

4.1 Post-launch acceptance period

Rapid results are needed following launch to ensure the pre-launch characterization accurately captured the new sensor performance metrics and to establish whether the new sensor is functioning as expected. Experience with current and past instruments has demonstrated significant departures in on-orbit vs. predicted instrument performance. Early distribution of first-look results provides a method of assessing instrument functionality through comparison with current and previously validated SST fields, either from other established and validated satellite sensors, or against analysis fields. Comparisons with in situ SST data, though limited in numbers, may also be useful during the post-launch acceptance period. Another activity is the generation of atmospheric RTM simulations of TOA BTs to serve as references for the sensor-observed BTs and thus provide understanding of their compliance with the results of the pre-launch thermal-vacuum tests.

> *Recommendation: Preparations for the post-launch acceptance tests include implementation of a tested data production and delivery system. This system must include the capability to deliver rapidly data to those interested in participating in evaluating early results. It must be flexible to accommodate quickly the changes that result from the initial evaluations. To achieve these goals, it is necessary to provide sensor characterization information to the broader community involved in the evaluation process well ahead of launch.*

In the initial performance assessment, all aspects of the instrument model, on-board calibration and geolocation, have to be scrutinized for imperfections that would lead to uncertainties in the retrieved SSTs. The corrective actions required depend on the nature of the issues identified, and may differ between μ-wave and IR radiometers.

A test of the geolocation accuracy is done by comparing positions of known features, such as headlands, small islands etc (Ground Control Points) in the imagery produced by the radiometers. Errors can arise from uncertainties in the mirror or antenna pointing knowledge, errors in the knowledge of the satellite location in the orbit and its attitude (pitch, roll and yaw angles), and errors in the knowledge of the mounting of the instrument on the satellite.

For μ-wave radiometers, the accuracy of the radiometer Antenna Temperatures are determined by comparisons with simulated values derived using a radiative transfer model with collocated environmental information to derive top-of-atmosphere BTs. These are transformed into Antenna Temperatures using the instrument- and channel-specific antenna patterns. Since the Antenna Temperatures are dependent on the emission angle at the sea surface, the geolocation and pointing uncertainties must be minimized first.

As the μ-wave antenna rotates, the view at the edge of the earth scene will begin to contain obstructions such as the satellite itself or part of the cold mirror. Additionally, during the scan, the antenna side-lobe pattern may result in contributions from different parts of the spacecraft. Thus, the difference between the measured and simulated Antenna Temperatures are used to determine along-scan biases. The post-launch analyses generally result in an adjustment to the APC that was measured pre-launch. It has sometimes been necessary to derive corrections for hot load thermal gradients and antenna emissivity. Again, RTM simulations are used, this time in conjunction with the measured hot load thermistor temperatures to determine an effective hot load temperature. Finally, the antenna emissivity correction is derived using RTM simulations and often additional information from instrument temperature measurements.

> *Recommendation: To facilitate the post-launch acceptance tests for μ-wave radiometers, develop an analysis environment that includes access to pre-launch calibration and characterization data, on-orbit radiometric data and ancillary spacecraft housekeeping data and an accurate radiative transfer code with appropriate environmental input.*

4.2 Long-term validation & monitoring of SSTs and associated BTs

Comparisons of retrieved SSTs and associated BTs with RTM simulations must also continue for the life of the mission to continuously monitor satellite SSTs and associated BTs for stability and self- and cross-platform consistency (Dash et al., 2010; Liang et al., 2009).

> *Recommendation: Approaches begun in the post-launch acceptance period to validate the SSTs retrieved from the new satellite sensor should continue, including comparisons with other available SST products and data. Associated BTs should continue to be compared with RTM simulations for stability and self-consistency.*

As the mission progresses, the opportunities for direct comparison with independent SST measurements, as opposed to derived fields, increases and a workable number of "matchups" between satellite measurements and validating data is generated. Validating SST retrievals using independent data is the basis of the second and third validation periods. The third, mission-long period is necessary to avoid changes in instrumental performance corrupting the SST time-series. Failure to identify extraneous changes, most problematic being slow degradations, and orbit changes without appropriate correction for diurnal warming, would likely lead to erroneous interpretations of time-series, and regional signatures. Identification of such problems is the first step in deriving appropriate corrections.

Depending on the intervals of time and space between the satellite and validating measurements deemed to be acceptable, order a thousand matchups per month can be achieved for a given wide-swath satellite radiometer. These matchups are limited in the IR to conditions that are identified as being confidently cloud-free, and comprise <10% of all possible matchups (Kilpatrick et al., 2001). The majority of matchups involve comparisons with subsurface measurements from drifting and moored buoys. Over several years of a mission, the current distribution of buoys is adequate to provide data that span most of the geographical and parameter (satellite zenith angles, atmospheric conditions etc) ranges that are influential in the uncertainty budget of the satellite retrievals. There are areas that are under-sampled, such as high-latitudes in both hemispheres, that, if filled, would benefit the validation process.

> *Recommendation: It is important to retain the number of sensors (buoys, profilers) at least at about current levels. A more uniform distribution of drifters would be of benefit. A larger number would provide the necessary data sets in a shorter time.*

A smaller, but significant, contribution comes from ship based radiometers. The advantage of ship-based radiometers is in removing the contribution to the comparison of temperature gradients between the depth of the subsurface measurement and skin layer of the ocean, which is closely related to the source of the radiation detected by the satellite radiometer. These gradients are caused by heat loss to the atmosphere (skin effect) and by heat gain during the day, especially with low wind conditions (diurnal heating) – see 3.1. The uncertainty in wind-speed parameterizations of the skin effect is O(0.1K) (e.g. Minnett et al., 2010), but the uncertainties in corrections for the effects of diurnal heating can be very variable (e.g. Gentemann and Minnett, 2008; Gentemann et al., 2009).

> *Recommendation: Effective interchange between the various segments of the mission and scientific communities should be fostered to ensure the optimum use of scarce resources and facilitate propagation of understanding of the limitations of existing SST products, e.g. location of regions and occurrence of environmental conditions that lead to sub-par performance. This understanding then should lead to deployment of resources, from direct in situ observations to modeling, to reduce SST uncertainties.*

It is becoming apparent that the uncertainties in the buoy temperature measurements are greater than the canonical value of about 0.1K (Emery et al., 2001; O'Carroll et al., 2008; C. Merchant, 2009, Pers. Comm.) and an increase in their accuracies would lead to an improvement in the SST uncertainty budget.

> *Recommendation: Improvements in the accuracies of the buoy temperatures are important; this requires better thermometers, better pre-deployment calibration, and transmission of the signal with 0.01K resolution.*

In recent years >3000 ocean profilers in the Argo program have been deployed (see http://www.argo.ucsd.edu/). The ARGO floats measure profiles of pressure, conductivity and temperature at intervals of about ten days as they rise to the surface, from where they transmit their data via satellite telemetry. The current generation of profilers have sensors that do not measure in the uppermost five to ten meters or so of the water column and so are not ideal for satellite skin SST retrievals. However, a second generation of profilers that take measurements up to the surface could be suitable to contribute to satellite-derived SST validation. With approximately 300 floats breaking surface each day, about 50 can be expected to be reporting their measurements within an hour of the satellite overpass (both night and day), of which perhaps five would be in sufficiently cloud-free conditions for the measurements to be useful for IR SST validation.

Recommendation: Argo profilers be deployed that take temperature and pressure measurements to the ocean surface, and their data be included in SST validation.

In the past decade or so, highly accurate ship-based IR radiometers have been developed for the validation of skin SST retrievals. Compared to buoys they are few, currently numbering less than a dozen worldwide, largely because of cost. Their contribution to SST validation is disproportionately important as the matchups are free of the sources of uncertainty caused by near-surface temperature gradients. In addition, when measurements are taken from a ship underway, averaging along the ship track can produce an average skin temperature along a section of comparable dimensions to the kilometer-scale resolution of IR radiometers on satellites, reducing uncertainties that might arise from small scale horizontal SST variability. Each of the radiometers is self-calibrating in that they have internal black-body calibration targets at known temperatures. To confirm the accuracy of the internal calibration procedures the ship-board radiometers are calibrated in the laboratory before and after each deployment. It is important that these laboratory calibration facilities be of known accuracy, and this can best be achieved by characterization against an SI standard. Such an approach has been adopted throughout the MODIS and AATSR missions through a series of workshops held at the University of Miami at which a common NIST-traceable calibration process has been adopted (Rice et al., 2004). This approach also provides traceability to a National Standard that is a key component to generating Climate Data Records of SST, and providing the mechanism for building times series of satellite-derived SST fields over many satellite missions.

Recommendation: The deployment of ship-based radiometers for validation of satellite-derived skin SSTs should be continued.

Recommendation: The calibration of ship-based radiometers against SI-standards should be continued.

Recommendation: The development of cheap but accurate, high accuracy ship-based radiometers should be encouraged.

4.3 Community consensus in situ data and match-up procedures

Current efforts at validating SST retrievals from a range of satellite sensors are often uncoordinated with the consequence that results are not easily compared. A unified approach to validation procedures would lead to a more coherent set of results from which conclusions could be more readily drawn.

Recommendation: A "Best Practices" guideline should be developed that identifies data sets (NCEP GTS, FNMOC, ICOADS, ...), data types (ships, drifters, tropical & coastal buoys) and QC procedures (more sophisticated & consistent with meteorological community). Unified match-up procedures (space-time windows, selection criteria) should be developed and adopted. Shared databases of skin SSTs and in situ bulk SSTs should be developed and made available to the community. These guidelines should *provide a baseline validation approach, and not be used to restrict additional, innovative approaches.*

4.4 Relating bulk and skin measurements

Given the large number of available buoy temperature measurements compared to skin temperature measurements, subsurface buoy temperatures should not be discarded, but their utility would be increased by a reduction in the contributions to the SST uncertainty budget of the near surface temperature gradients. This can be achieved by modeling the physical processes that control the skin effect and diurnal heating gradients and their evolution.

Recommendation: Ship-based radiometer data should be augmented by measurements of all controlling parameters to enable the development of improved models of skin and diurnal effects that will lead to better propagation of the skin SST retrieval to sub-surface temperatures.

4.5 Parameters influencing retrieval uncertainties

There are some conditions that give rise to larger bias errors or increased scatter in the SST uncertainty. These include anomalous atmospheric water vapor amounts and distributions, aerosols, air-sea temperature differences, particular clouds, coastal effects and high latitude and marginal ice zones.

Recommendation: Parameters influencing SST retrieval uncertainties should be measured, or extracted from other sources (NWP analyses, other satellite data sets), for the validating matchups, to enable improved atmospheric correction algorithms to be developed.

4.6 Radiative transfer simulations

The use of atmospheric radiative transfer simulations provides insight into factors influencing SST retrievals that are not amenable to direct measurement, such as post-launch changes in instrument characteristics or evaluation of how unusual environmental conditions impact SST retrieval, such as volcanic aerosol, anomalous atmospheric water vapor or temperature profiles.

Recommendation: Analyses of SST retrieval uncertainties using atmospheric radiative transfer simulations should be part of the validation process leading to improved retrieval algorithms.

4.7 Uncertainty estimates

It is not immediately obvious how best to form the overall SST uncertainty budget from the component parts. It may be that simple "root sum squares" is not the most appropriate way of determining the characteristics of the SST retrievals.

Recommendation: Research be directed to the best way of constructing accurate SST uncertainty budget.

4.8 How to represent information and render it useful

Historically the uncertainties of the SST retrievals have been represented as a global mean (bias) and a standard deviation. While encapsulating performance in a couple of easily remembered numbers, this approach hides problems, such as regional biases, that can confound the appropriate use of the data. Within the GHRSST project, the error hypercube (a multi-dimensional look up table for retrieval errors) has been developed for both the MODIS, TMI, and AMSR-E SSTs that represent mean and a standard deviation of expected uncertainties as a function of various governing parameters eg. for MODIS: pathlength, environmental regime, wind speed etc. This is a start at improving the description of the inherent characteristics of the error fields.

Recommendation: The concept of the error hypercube should be further developed, perhaps using continuous functions of the controlling parameters and including additional parameters, and applied to all SST retrievals.

4.9 User-based product evaluation

Researchers often learn a significant amount about the quality of the data products they use. This information, which relates to both the overall quality of the product and to the quality of individual data granules[10], is of potential value to the data provider as well as to other users of the data. However, there is no formal mechanism to gather this information and to make it available to others; i.e., the knowledge gained by the user with regard to data quality is often lost because of the lack of a feedback mechanism for this information. In the following we characterize user evaluations of SST data products and we suggest a general framework to capture the information learned by the user and a mechanism to feed this information back to the data provider and forward to future users of the data thus bringing the data user into the validation loop.

User evaluations of a data product are either implicit or explicit. Explicit evaluation occurs when the user systematically examines SST fields to determine their quality prior to using them. For example, Lisan Yu of Woods Hole Oceanographic Institution has been using OISST fields in the generation of a heat flux product that she makes available to the community (http://oaflux.whoi.edu/). The version of the OISST that she has been using is based on AVHRR SST data, and suffers because of the sparse coverage of the underlying SST fields resulting from

10. Data granule as used here refers to the way we generally think about and work with data from a data set; e.g., a latitude, longitude SST field at a given time or an average of the data collected over some period. For many extant archives, the granule is a file such as a NetCDF file.

cloud contamination. In an effort to address this problem Reynolds, who developed the AVHRR-only product, has developed a new OISST based on both AVHRR and AMSR-E data, a product of potential value for Yu's application. Prior to using the new OISST data, Dr. Yu compared the new product with the old product (Figure 3) and discovered a 0.1°C offset between the two. Such evaluations of data products by the scientific user are more the norm than the exception when the researcher intends to use a significant fraction of a data set. Because these evaluations are generally performed on large subsets of a data set, if not on the entire data set, they fall in the category of product evaluation rather than granule evaluation, although they may well uncover granule specific issues. Explicit user evaluations of SST products are not part of a formal cal/val effort associated with the development of a retrieval algorithm so, although most of these evaluations were communicated to the data provider, only in a few cases were they publicized, usually in the form of a journal article; i.e., it is difficult for other users to benefit from this knowledge in that the information is not associated with the data.

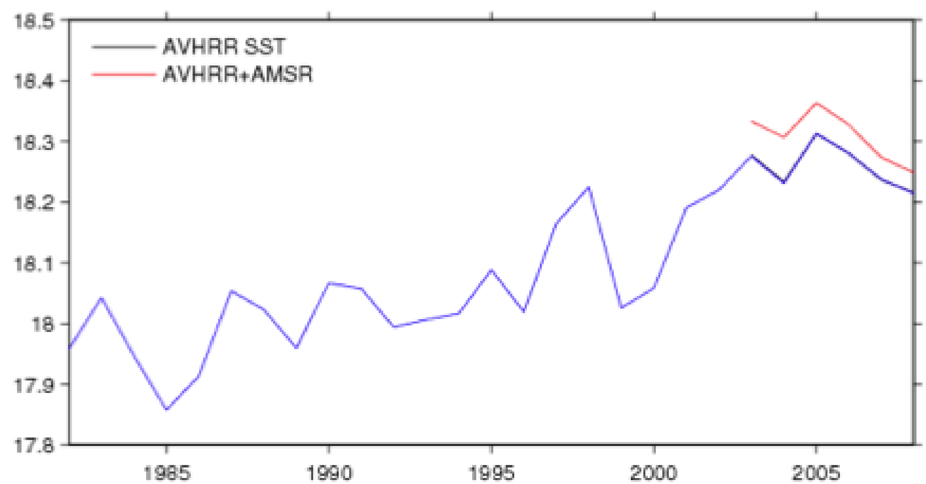


Figure 3. Annual mean SST from Reynold's OISST based on AVHRR only versus that based on AVHRR and AMSER-E. Both for daily, 0.25 degree, V2 products.

Implicit evaluation occurs when the user finds problems with the data through their routine use; e.g., when the user notices that a specific SST field is poorly geolocated - Cape Cod is located in the middle of the Gulf of Maine - or that there are a number of SST values in excess of 50 degrees C that have not been flagged as bad. Again, it is more the norm for the researcher to discover granule specific problems when working with SST products than not. Unlike the explicit evaluations, quality issues discovered in passing tend to relate to per granule issues.

To benefit from both explicit and implicit evaluations by users a means for the user to communicate his or her findings with regard to SST quality to the data provider as well as to future users of the data is required. Annual meetings of an SST Science Team might well serve to address explicit evaluations of data sets, however per granule issues are not likely to be brought forward in a large meeting hence, at a minimum, a mechanism is required to handle these.

Recommendation: Develop a means by which users may easily communicate their SST data quality findings at the granule level and at the data set level to both the provider of the data and to a data quality repository.

Recommendation: Develop a means for users to seamlessly access repositories of data quality developed by both the user and the product developer.

In order to help better understand what we are suggesting here, we provide details for at least one mechanism. There are certainly other ways to capture the knowledge related to data quality gained by users. The details: In that most SST data products are provided via self describing formats, a field could readily be added to each granule that would contain an error reporting URL to the metadata associated with that granule. The user on discovering a problem with a particular granule or with the data set as a whole (the product) would simply enter the URL in a browser which would provide a form in which the user would report the observed problem. As part of the reporting process, the user would specify whether the problem is a data set problem or a granule problem. On submission, an e-mail would be sent to the data provider, and an entry would be made in bug database. When/if the problem is rectified, the bug report in the database would be cleared. For data set issues, the report would also be made accessible to descriptions of the data set; i.e., to entries for the data set in a data catalog or directory. For granule issues, the report would be made accessible to those accessing individual granules. As noted above, this is presented as a general framework, the details of such a system need to be addressed in the future, not only of the reporting mechanism, but also how users access the reported problems, but these issues may be readily addressed with current technology.

4.10 Validation and Evaluation of Derived SST Products of Data Set Producers' Side

4.10.1 Validation of gridded merged products and error estimates

With the growth in demand and number of gridded merged SST products, there needs be a coherent strategy for the validation of both skin and bulk temperature products. The strategy should consist of the following approaches which are not necessarily mutually exclusive:

Inter-comparisons of the merged products. This necessitates using re-gridding schemes that preserve the resolutions of individual products, but at the same time allows for direct comparisons at exact locations.

Subdaily skin temperature products. For daily products that are presented as skin temperatures (not SST_{fnd}), the diurnal cycle becomes problematic. Availability of in situ radiometer data becomes critical for their validation. In regions where diurnal cycle dominates, merged skin SST products should also be validated against geostationary data or other independent observations of the diurnal cycle.

Direct comparisons with independent buoys, both drifting and moored. For validating skin temperature data sets, larger data sets of in situ radiometer data, covering different regions of the world's oceans, are necessary. There are also needs for the quality control of the in situ data itself. Merged products that make use of the in situ data should exclude certain amounts of them for the purpose of independent validation.

Direct comparisons with model output. One example of such efforts is the Observational System Simulation Experiments (OSSEs). Such comparisons, unlike buoy data, will provide error estimates based on the resolution of the models used, with a possibility to quantify the effect on uncertainty of (un)resolved physical and dynamical processes, as well as mesoscale and sub-mesoscale phenomena.

> *Recommendation: To develop systematic approaches to L4 product intercomparison and validation, including validation of uncertainty estimates; judicious approaches to selection of data to use in L4 products vs data to be withheld for validation; to consider more sophisticated validation studies, e.g. OSSE.*

4.10.2 Evaluation of Effective Resolution of Level 4 Products

Every SST analysis procedure can be viewed as a filtering operation. Since the Gauss-Markov procedure is a linear estimator, each SST estimate can be expressed as a linear combination of the observations used by this estimate. It is straightforward to express such estimates in the wavenumber and frequency domains, thus providing an assessment of the filtering properties inherent in each estimate. In general, this filtering varies geographically and temporally, depending on the space-time distribution of the observations used to obtain each estimate. However, the filtering is controlled primarily by the prescribed signal and error spatial and temporal autocorrelations and signal-to-noise variance ratio used in the analysis procedure. The specification of a long spatial autocorrelation scale, for example, results in heavily smoothed estimates of the SST field. Likewise, specification of small signal-to-noise ratio or large-scale or persistent error autocorrelation function also imposes increased smoothing of the SST estimates. The use of different autocorrelation functions and signal-to-noise ratios in different SST analyses thus results in SST fields with varying spatial resolution.

Because of the filtering inherent in L4 products, it is very important to distinguish between grid resolution and feature resolution. SST fields are always produced on finer spatial and temporal grids than can be resolved by the OA procedure used in the analysis. In particular, the 0.1° by 3-hour gridding that is the goal of SST analyses is grossly beyond the feature resolution capability of satellite-based SST analyses, except in rare instances of clear-sky conditions and dense sampling of infrared observations of SST.

> *Recommendation: To develop systematic approaches to clarifying and communicating the actual smoothness of individual L4 products versus their nominal grid resolution; to investigate the influence of these parameters on the data set error.*

References

Dash, P., A. Ignatov, Y. Kihai, and J. Sapper, 2010: The SST Quality Monitor and its use for diagnostics of NESDIS heritage AVHRR products. J. Atmos. Ocean. Tech., in review.

Donlon, C. J., P. Minnett, C. L. Gentemann, I. J. Barton, B. Ward, J. Murray, and T. J. Nightingale, 2002: Towards operational validation of satellite sea surface skin temperature measurements for climate research, J. Climate, 15(4), 353–369.

Donlon, C., I. Robinson, K.S. Casey, J. Vázquez-Cuervo, E. Armstrong, O. Arino, and C. Gentemann, et al., 2007: The Global Ocean Data Assimilation Experiment High-Resolution Sea Surface Temperature Pilot Project. Bull. Amer. Meteor. Soc., 88, 1197–1213.

Castro S. L., G. A. Wick. and W. J. Emery, 2003: Further refinements to models for the bulk-skin sea surface temperature diffrence., J. Geophys. Res., 108,C12, 3377.

Emery, W.J., D.J. Baldwin, P. Schlüssel, and R.W. Reynolds, 2001. Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. J. Geophys. Res., 106, 2387-2405. 10.1029/2000jc000246.

Eyre, J., E. Anderson, E. Charpentier, L. Ferrantier, L. Ferranti, J. Lafeuille, M. Ondras, J. Pailleux, F. Rabier, and L.P. Riishojgaard, 2009. Requirements of numerical weather prediction for observations of the ocean. Community whitepaper to OceanObs09, 21-25 September, 2009, Venice, Italy.

Fairall, C. W., E. F. Bradley, J. S. Godfrey, G. A. Wick, J. B. Edson, and G. S. Young, 1996: Cool-skin and warm-layer effects on sea surface temperature. J. Geophys. Res., 101, 1295-1308.

Gentemann, C. L., C. J. Donlon, A. Stuart-Menteth, and F. J. Wentz, 2003, Diurnal signals in satellite sea surface temperature measurements, Geophys. Res. Lett., 30, 1140, doi:10.1029/2002GL016291.

Gentemann, C. L., P. J. Minnett, P. LeBorgne, and C. J. Merchant, 2008: Multi-satellite measurements of large diurnal warming events, Geophys. Res. Lett., 35, L22602, doi:10.1029/2008GL035730.

Gentemann, C.L., and Minnett, P.J., 2008. Radiometric measurements of ocean surface thermal variability. J. Geophys. Res., 113, C08017. doi:10.1029/2007JC004540.

Gentemann, C.L., Minnett, P.J., and Ward, B., 2009. Profiles of Ocean Surface Heating (POSH):

a new model of upper ocean diurnal thermal variability. J. Geophys. Res., 114, C07017. doi:10.1029/2008JC004825.

Grodsky, S.A., J. A. Carton, and H. Liu, 2008: Comparison of bulk sea surface and mixed layer temperatures, *J. Geophys. Res.*, 113, C10026, doi:10.1029/2008JC004871.

Gudes, S., G. Asrar and P. Pace, 2001: Integrated Operational Requirements Document (IORD) II, National Polar-Orbiting Operational Environmental Satellite System (NPOESS). Technical Report, NOAA, *NPOESS* Integrated Operational Requirements Document (*IORD*) *II*.

Guenther, B., Barnes, W., Knight, E., Barker, J., Harnden, J., Weber, R., Roberto, M., Godden, G., Montgomery, H., and Abel, P., 1996. MODIS calibration: A brief review of the strategy for the at-launch calibration approach. J. Atmos. Ocean Science, 13, 274-285.

Hanafin, J. A. and P. J. Minnett, 2001: Profiling temperature in the sea surface skin layer using FTIR measurements. Gas Transfer at Water Surfaces, 161-166.

Hanafin, J. A. and P. J. Minnett, 2005: Infrared-emissivity measurements of a wind-roughened sea surface. *Applied Optics.*, **44,** 398-411.

Harris, A.R., S.J. Brown and I.M Mason,1994: The effect of windspeed on sea-surface temperature retrieval from space Geophys. Res. Let., 21, No.16, 1715-1718

Katsaros K. B., 1977: Sea-surface temperature deviation at very low wind speeds—Is there a limit? Tellus, 29, 229–239.

Kawai, Y. and H. Kawamura, 2002: Evaluation of the diurnal warming of sea surface temperatures using satellite-derived marine meteorological data. J. Oceanography, 58, 805-814.

Kent, E.C. and P.K.Taylor, 2006: Toward Estimating Climatic Trends in SST. Part I: Methods of Measurement, *Journal of Atmospheric and Oceanic Technology,* **23,** pp. 464–475.

Kirkpatrick, K.A., G.P. Podestá, 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.

Le Borgne, P., A. Marsouin, F.Orain, and H. Roquet, 2009: Operational sea surface temperature bias adjustment using AATSR data. Remote Sensing of Environment, submitted.

Liang, X.-M., A. Ignatov, and Y. Kihai, 2009: Implementation of the Community Radiative Transfer Model in Advanced Clear-Sky Processor for Oceans and validation against nighttime AVHRR radiances. J. Geophys. Res., 114. 10.1029/2008jd010960

Merchant, C. J., M. J. Filipiak, P. LeBorgne, H. Roquet, E. Autret, J.-F. Piolle, and S. Lavender, 2008: Diurnal warm-layer events in the western Mediterranean and European shelf seas, Geophys. Res. Lett., 35, L04601, doi:10.1029/2007GL033071.

Minnett, P. J., 1986: A numerical study of the effects of anomalous North Atlantic atmospheric conditions on the infrared measurement of sea-surface temperature from space. *Journal of Geophysical Research*, **91,** 8509-8521.

Minnett, P. J., 1991: Consequences of sea surface temperature variability on the validation and applications of satellite measurements, J. Geophys. Res., 96, 18475-18489.

Minnett, P.J., Smith, M., Ward, B., 2010. Measurements of the oceanic thermal skin effect. Deep Sea Res. II. Accepted.

Nalli, N.R., P.J. Minnett and P. van Delst, 2008a: Emissivity and reflection model for calculating unpolarized isotropic water surface-leaving radiance in the IR. I: Theoretical development and calculations, Appl. Opt., 47, 3701-3721.

Nalli, N. R., P. J. Minnett, E. Maddy, W. W. McMillan, and M. D. Goldberg, 2008b: Emissivity and reflection model for calculating unpolarized isotropic water surface-leaving radiance in the infrared. 2: Validation using Fourier transform spectrometers. *Applied Optics*, **47,** 4649-4671. doi:10.1364/AO.47.004649

O'Carroll, A.G., J.R. Eyre, and R.W. Saunders, 2008. Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations. J. Atmos. Ocean. Tech., 25 (7), 1197-1207.

Ohring, G., B. Wielicki, R. Spencer, B. Emery, and R. Datla, 2005. Satellite Instrument Calibration for Measuring Global Climate Change: Report of a Workshop. Bulletin of the American Meteorological Society, 86 (9), 1303-1313

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.

Reynolds, R.W., C.L. Gentemann, and G.K. Corlett, 2010: Evaluation of AATSR and TMI Satellite SST Data. *J. Climate*, **23**, 152–165.

Rice, J.P., and B.C. Johnson, 1996. A NIST Thermal Infrared Transfer Standard Radiometer for the EOS Program. The Earth Observer 8, 31.

Rice, J.P., and B.C. Johnson, 1998. The NIST EOS Thermal-IR Transfer Radiometer. Metrologia, 35, 505-509.

Rice, J.P., J.J. Butler, B.C. Johnson, P.J. Minnett, K.A. Maillet, T.J. Nightingale, S.J. Hook, A. Abtahi, C.J. Donlon, and I.J. Barton, 2004. The Miami2001 IR Radiometer Calibration and Intercomparison: 1. Laboratory Characterization of Blackbody Targets. J. Atmos. Ocean. Tech., 21, 258-267.

Saunders, P. M., 1967. The temperature at the ocean-air interface, J. Atmos. Sci., 24, 269-273.

Smith, T.M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). J. Climate, 21, 2283-2296.

Ward, B., R. Wanninkhof, P. J. Minnett, and M. J. Head, 2004: SkinDeEP: A profiling instrument for upper-decameter sea surface measurements. J. Atm. Ocean. Tech., 21, 207-223.

Watts, P.D., M.R. Allen, and T.J. Nightingale, 1996: Wind-speed effects on sea-surface emission and reflection for the Along-Track Scanning Radiometer J. Atmos. Ocean. Tech., 13, No.1, 126-141.

Webster, P. J., C. A. Clayson, and J. A. Curry, 1996: Clouds, radiation, and the diurnal cycle of sea surface temperature in the tropical western Pacific, J. Clim., 9, 1712-1730.

Wentz, F. J., and T. Meisner, 2000: Algorithm theoretical basis document version 2: AMSR ocean algorithm, NASA Goddard Space Flight Cent., Greenbelt, Md. (Available at http://eospso.gsfc.nasa.gov/eos_homepage/for_scientists/atbd/)

Woodcock A. H., and H. Stommel, 1947: Temperatures observed near the surface of a fresh water pond at night. J. Meteor., 4, 102–103.

Wu, X. and W.L. Smith, 1997: Emissivity of rough sea surface for 8-13µm: modeling and verification. Appl. Opt., 36, 2609-2619.

Zappa, C. J. and A. T. Jessup, 2005: High Resolution Airborne IR Measurements of Ocean Skin Temperature. Geosci. Remote Sens. Letters, 2(2), 146-150.

Appendix I. ISSTST members/SST Error Budget Workshop participants

Last Name	First Name	Affiliation
Armstrong	Ed	NASA/JPL
Banzon	Viva	NOAA/NCDC
Barron	Charlie	Navy/NRL
Brix	Holger	UCLA
Carton	Jim	U.Maryland
Castro	Sandra	U.Colorado
Chelton	Dudley	OSU
Chin	Mike	NASA/JPL RSMAS
Clayson	Carol	FSU
Cornillon	Peter	U.Rhode Island
Crosman	Erik	U.Utah
Cummings	Jim	Navy/NRL
Dong	Changming	UCLA
Emery	Bill	U.Colorado
Evans	Bob	U.Miami
Foley	Dave	NOAA/NMFS
Gentemann	Chelle	RSS
Gille	Sarah	UCSD
Grodsky	Semyon	U.Maryland
Grumbine	Robert	NOAA/NCEP
Hacker	Peter	NASA/GSFC
Harris	Andy	NOAA/U. Maryland
Holt	Ben	NASA/JPL
Hook	Simon	NASA/JPL
Ignatov	Shasha	NOAA
Jessup	Andy	U.Washington
Jochum	Markus	NCAR
Kaplan	Alexey	Columbia U.
Kearns	Ed	NOAA/NCDC
Le Borgne	Pierre	MŽtŽoFrance/CMS
Lindstrom	Eric	NASA/HQ
Maturi	Eileen	NOAA/OAR
Mehta	Vikram	CRCES
Menemenlis	Dimitris	NASA/JPL
Merchant	Chris	U.Edinburgh
Minnett	Peter	U.Miami
Mittaz	Jon	NOAA/U.Maryland
Murtugudde	Raghu	U.Maryland
Reynolds	Dick	NOAA/NCDC

Soloviev	Alex	Nova U.
Strub	Ted	OSU
Vazquez	Jorge	NASA/JPL
Wick	Gary	NOAA/ETL
Yu	Lisan	WHOI
Zhang	Rong-Hua	U.Maryland
Casey	Ken	NOAA/NODC
Gramer	Lew	NOAA/AOML
Harrison	Ed	NOAA/PMEL
Hendee	Jim	NOAA/AOML
Markus	Thorsten	NASA/GSFC
May	Doug	Navy/FNMOC
Wentz	Frank	RSS
Xie	Shang-Ping	U.Hawaii/IPRC
Zappa	Chris	Columbia U.

Appendix II. SST Application Requirements

A careful examination of SST applications is important in helping to determine the SST error budget for two reasons: (i) to meet the needs of those using the data and (ii) to evaluate the quality of the data, a critical component in helping to reduce SST error and uncertainty. This appendix aims to answer the following questions for a small sample of representative SST applications: what is the most stringent requirement in terms of (i) spatial resolution, (ii) temporal resolution, (iii) geolocation accuracy, (iv) absolute accuracy, and (v) relative accuracy needed for each particular application and do the errors currently observed in any of the above 5 categories provide insight about the key SST error budget uncertainties for each application.

Fronts (P. Cornillon)
In contrast to the requirements of CDR, NWP, and operational applications, the requirements of feature and process-related studies have received little attention. Of interest, although not surprising, is that those interested in feature and process oriented studies impose some fairly stringent requirements on the location of pixels. They also tend to value relative thermal accuracy over absolute accuracy and are not quite as strict with regard to data quality in general. The most stringent requirements in terms of spatial resolution are 1 km at present but 100 m resolution could also be used, if made available. The cross-frontal SST profile for open ocean and coastal fronts determines the 100-m lower bound. The most stringent requirement in terms of temporal resolution is 1 hr at present (from geosynchronous satellites) but 15 min samples could also be used. The reason for the higher temporal resolution data is not because the frontal features move that fast but because the IR data tend to be obscured by cloud cover fairly often;

the higher the temporal resolution the more views of a given location are available when it is clear. Algorithms that make use of the 3d (lat, lon and time) field are being developped to extract fronts. These algorithms require geolocation accuracy of 1 pixel and relative accuracy of 0.1K. The frontal detection algorithms are relatively insensitive to the actual temperature so an absolute accuracy of 1K suffices. In terms of relative importance, geolocation accuracy is the biggest problem for establishing frontal probability fields. The more the individual SST fields move around the more the probability fields are smeared out. After geolocation the relative SST accuracy of the data is the next most important. It is a close second. Compared to the open ocean, the geolocation information is even more critical in coastal regions.

Climate Models (M. Jochum, S. Gille, D. Menemenlis, and S.-P. Xie)
For evaluation of ocean and of coupled-ocean-atmosphere models, for ocean state estimation, for ocean analysis of mixed layer temperatures and basin-scale heat content, and for ocean acoustics, the requirements are similar to those of CDR, except that the modeling applications require a more complete characterization of SST retrieval uncertainties, including the characterization of SST retrieval uncertainties pertaining to the conversion from observables to model prognostic variables. Specific SST data requests from workshop participants for modeling applications include:

1. L2 SST skin or sub-skin retrieval, as appropriate for each instrument, with quality flags and with standard deviation uncertainty estimate,
2. L4 SST (skin or sub-skin) retrieval, as appropriate for each instrument, with (i) standard deviation uncertainty and (ii) resolved scales, e.g., -3dB width of error covariance matrix in km,
3. empirical model for diurnal variability and for vertical temperature profile in top 10 m as a function of depth, location, and time,
4. L2 daily and depth averaged SST in top 1 m with standard deviation uncertainty estimate,
5. L2 daily and depth averaged SST in top 10 m with standard deviation uncertainty estimate,
6. monthly and 200x200-km averaged uncertainty for items 1, 2, 4, and 5 should be everywhere less than 0.2 K (especially in warm pool and Bay of Bengal), and
7. stability, i.e., temporal consistency, for monthly and 200x200-km averaged temperature for items 4 and 5 should be less than 0.05 K / decade.

A spatial resolution of 25 km is adequate for most contemporary climate modeling applications but as global oceanic, atmospheric, and coupled climate models increase in resolution, the resolution requirements for SST data products will also increase. For climate applications, all of the above are needed in a continuous and consistent manner for as long a time period as possible. μ-wave products have advantage in coverage but are only available after December 1997. Prior to that, IR products can match resolution in theory but suffer from cloud contamination. Prior to IR, we only have ships in open ocean. Therefore long, consistent SST records, suitable for climate studies, is a difficult but important challenge. In coastal ocean regions, increased spatial resolution is required. Coastal ocean regions also experience a larger diurnal cycle, making accurate temporal sampling difficult.

Ocean Models (S.Grodsky, R.Murtugudde, M.Jochum)
The needs of the ocean modeling community are diverse. Traditionally, ocean modelers are interested in the "bulk" temperature data sets whereby the integrated temperature from a certain depth interval in the mixed layer is available. Modelers running coupled ocean-atmosphere models are also interested in the skin SST because it is more closely related to what drives the atmosphere. The general consensus for the modeling community is that to be particularly useful for ocean modelers, L4 products should report an integrated "bulk" temperature in some depth interval (e.g., 1-10 meters), and a diurnal warming model or its output should be available to propagate that temperature to any depth in the mixed layer from the foundation temperature depth to the surface.

Satellite sensors measure SST_{skin} (IR) or $SST_{sub\text{-}skin}$ (microwave) that is the average water temperature in the upper microns /a few centimeters. But, most of ocean models do not have that level of vertical resolution. Some ocean models resolve the diurnal cycle, but many models (including the IPCC4 climate models) do not. These latter models simulate SST that is representative of the upper 10 m layer. Spatial resolution of models is stretched towards the poles, but normally does not exceed 10 km, except for regional applications (for which in-swath rather than gridded SST is more appropriate for validation). In such cases having the spatial resolution of SST products that high is not very important for the intercomparison with models because small-scale processes in models and reality fit each other only in the statistical sense. But providing the error bar for all SST products as well as specifying the depth averaging limits are crucial for validations. Depending on applications, model output is stored at temporal decimations ranging from hours to 1 month.

A sensitive test for current climate model is the seasonal cycle. In particular, to evaluate the Community Climate System Model (CCSM) and to attribute biases (to the OGCM or the AGCM) ocean modelers need a gridded climatology of upper 10 m monthly mean temperature at 1x1 degree resolution. The accuracy (or uncertainty) requirements are dependent on the atmospheric sensitivity, which varies regionally. In the subtropical gyre and higher latitudes, +/- 0.5K is sufficient, whereas in the tropics (especially in the Pacific warm pool) +/- 0.2K is required.

The next generation CCSM will include a parameterization of diurnal warming of SST. Preliminary experiments suggest that it is of utmost importance in the Tropics and in the Pacific warm pool. To evaluate and optimize the parameterization we need at least 3-hr $SST_{sub\text{-}skin}$ for 3 regions: the Pacific warm pool, a subtropical gyre, and the North Pacific. Some regional applications need finer temporal decimation. In particular, 25km 3-hr $SST_{sub\text{-}skin}$ is needed to test model performance of diurnal precipitation and to understand the propagation of diurnal peaks in places like the Bay of Bengal and in the maritime continent.

A wishlist of gridded SST products particularly useful for ocean model validation:

- 25km x 25km x 3h $SST_{sub\text{-}skin}$;
- 50km x 50 km x 1day SST_{bulk} (corrected for diurnal variations);

- 50km x 50 km x 1month SST_{ml} (SST_{fnd})

Lakes and Coastal Ocean (E. Crosman, S. Hook and T. Strub)
Lake and coastal (LAK) applications have higher **spatial** and **temporal resolution** requirements than do open-ocean (OCN) applications. For ecosystem studies and for meteorological applications the **absolute accuracy** is most important, while **relative** and **geolocation accuracy** and **spatial resolution** are very important for feature identification and tracking (e.g., gyres, river inflow). **Geolocation accuracy** is also very important if you process and combine many images from near shore regions. **Temporal resolution** is critical as LAK features form and decay much more rapidly than over the OCN, often on hourly time scales. The observed errors for LAK retrievals are also very different than for the OCN. The split-window algorithms developed for OCN do not work well for LAK, especially for lakes at altitudes well above sea level, and there is generally a lack of high spatial (< 100 m) data with a high revisit. This means we have to worry much more about between sensor differences, e.g., using Landsat and ASTER. Also, because of the rapid variations in surface temperature, validation requirements are also more stringent, requiring LAK matchups to be within 10 min instead of the typical 1 hour for OCN applications. For SST validation over LAK, the absolute accuracy is the most important. The absolute accuracy varies between the ASTER, MODIS, Landsat (5,7) and ATSR (ATSR2 and AATSR) sensors. Each sensor has its own error budget.

Specific examples of LAK applications are Utah's Great Salt Lake (GSL), Lake Tahoe, and the Salton Sea. In the case of monitoring GSL SST for lake-effect snowstorm prediction, and ecological and limnological applications, the requirement is for **spatial resolution** of 1 km, **temporal resolution** of 3 hr, **geolocation accuracy** of 1 km (1 pixel), **relative accuracy** 0.2 K, and **absolute accuracy** of 0.3 K. Cloud/dust contamination and geolocation issues are a large problem (order 3°K) in some individual MODIS and AVHRR images. A systematic cool bias between SST retrievals (order 1K where retrieved SST is cooler than in situ observations) has been noted for MODIS and AVHRR data, but it is unknown how much of this is due to the cool skin versus improper atmospheric correction in dry, elevated climate. Large differences (4-8 K) between SST and subsurface temperature have also been observed when the lake is stably stratified. A specific example of issues with validation is at Lake Tahoe, which encompasses on order 35x17 1-km pixels. With MODIS you can see the same structure you see at ASTER resolutions (90m). This is a testament to the quality of the MODIS data. However, the striping which can be amplified by the algorithm used to produce surface temperature can overwhelm the feature and inhibit the ability to track it so the radiance at sensor product is used instead. Radiometric calibration problems with both Landsat and ASTER have been identified and fixed in the Tahoe retrievals. Large problems with the water surface temperature retrieval algorithms have been observed at the Salton Sea validation site. The ability to validate, matchup, and derive good algorithms for retrieving water surface temperature varies widely between regions. The Tahoe site is the "ideal" site (deep mixed layer, homogeneous surface temperature and emissivity away from shore) for recovering surface temperature but the Salton Sea is the opposite extreme (very shallow surface layer).

LAK applications in the coastal ocean include fronts caused by upwelling, tides and other processes, which are covered by the above discussion of fronts. Another coastal phenomenon is the movement of river plumes by tidal currents. For instance, the plume of warm water emanating from the Columbia River in summer extends and retracts several tens of kilometers with a semi-diurnal period. The instantaneous signal is represented well by temperature resolutions of several tenths of degree K, but quantifying differences in each day's development requires temporal and spatial resolutions of approximately 1-hour and 0.1 km, similar to the requirements for fronts. Other coastal processes include the alongshore movement of coastal trapped waves, with speeds of tens to several hundred kilometers per day. These raise and deepen the thermocline and may produce small signals in SST, requiring hourly sampling and resolution of 0.1 K or less.

Air-sea heat fluxes (S. Gille)
For applications using SST to compute air-sea heat fluxes in the Southern Ocean, the requirements are:

- A spatial resolution consistent with Southern Ocean jet width or eddy scale, i.e., 10-km resolution.
- A temporal resolution consistent with eddy motions. Weekly-averaged fields have been used in the past but daily data with a proper diurnal correction would be preferred.
- Geolocation accuracy consistent with the spatial resolution, e.g., ± 2 km, so that the 10 km resolution means something.
- Absolute accuracy adequate to compute air-sea heat fluxes accurate to 10 W/m^2. This requires that the air-sea temperature difference be accurate to ± 0.5 K, which means that air temperature and ocean temperature should each be accurate to $\pm 0.5/\sqrt{2} = \pm 0.3$ K. However, surface air temperature is hard to retrieve compared with SST, so ideally, one would want SST to be more accurate, e.g., ± 0.1 K) in order to prevent it from corrupting sensible heat estimates.

Extensive comparisons of in situ data and μ-wave SST show that the diurnal cycle (if not properly handled), atmospheric water vapor (at low water vapor conditions), and cloud water (at high cloud water conditions) can contribute biases that exceed 0.5°K.

Mesoscale and submesoscale features - biology/physics (C. Dong and H. Brix)
The first baroclinic Rossby deformation radius (Rd) is used to categorize the mesoscale and submeoscale features. Rd varies geographically from a few kilometers to hundreds of kilometers. This leads to a required spatial resolution that depends strongly on the study region. The approximate requirement for a mesoscale feature in low and mid latitudes is a spatial resolution of 10 km, in high latitudes (the Arctic, for instance) down to 1 km. The required temporal resolution is weekly with an absolute and/or relative accuracy of 0.1°K (depending on the scientific problem at hand). For submesoscale features, the horizontal scale is O(1km), less than Rd. The approximate requirements are a spatial resolution of O(100m), a temporal resolution of 1 day or higher, and an absolute/relative accuracy of 0.1°K. The requirements for harmful algae/ fishery habitats are similar to the mesoscale requirements.

Climate data record - physics (R. Reynolds and P. Hacker)

The National Research Council defines a CDR as "A time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change." The 3-decade long satellite record of SST is sufficiently long to study phenomena and processes with time-scales of days to decades. Consistency and continuity can be achieved by combining and/or merging the various satellite-based products, documenting error sources, and maintaining all relevant data and metadata to enable reprocessing. Figure II.1 shows the annual-averaged Extended Reconstruction SST (ERSST) anomalies from 1880-2008 averaged between 60°S and 60°N (Smith et al., 2008) using a 1971-2000 climatological base period. In the period after the 1940s, the average trend is roughly 0.1K per decade. Thus, to be useful observations must have errors below 0.05K/decade. If the relative bias errors are larger, the signal shown in this figure would be difficult to detect.

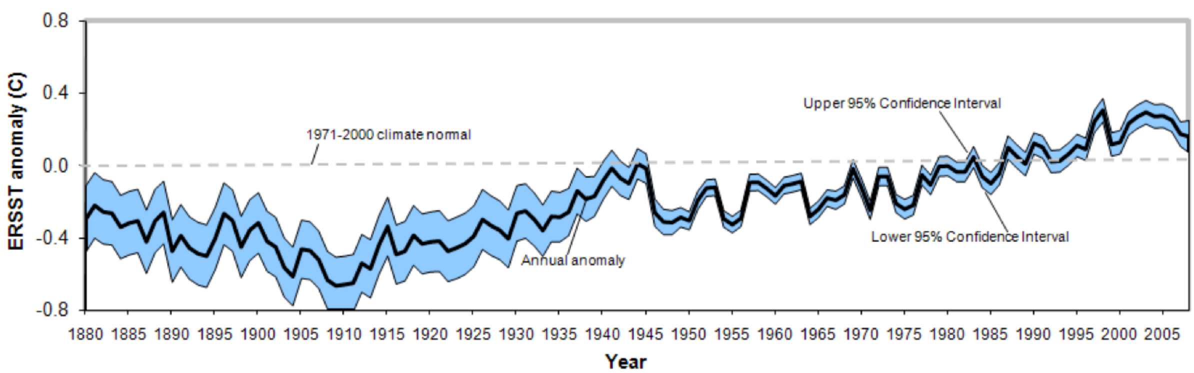


Figure II.1. Annual ERSST.v3b anomaly from 1880-2008 between 60°S and 60°N (solid line) with 95% confidence interval in blue. Note that the data is more reliable after the 1940's. The magnitude of the temperature increase in recent decades is much greater than the uncertainty in the data.

An example of absolute bias errors from different satellite observations can be seen in Figure 2 from Reynolds et al. (2010). This figure shows the zonal difference for daily nighttime observations for a three-year period with respect to the daily optimum interpolation (OI) AMSR+AVHRR analysis of Reynolds et al. (2007). The nighttime biases are especially important because there is almost no diurnal signal to affect them. The zonal biases were only shown between 60°S and 60°N to help ensure that there was sufficient in situ data to define the biases. However, even if the reference OI is not perfect, this figure makes it clear that average zonal differences among satellite products were typically 0.1K and may exceed that value south of 50°S and north of 50°N.

Based on the results of Figures II.1 and 2, we recommend that CDR errors be less than 0.05K per decade and less than 0.05Kover a 1-year period.

Appendix III. Level definitions

Table III.1 shows the definition of data level as we use them in this document. This is a modified version of a NASA table (http://nasascience.nasa.gov/earth-science/earth-science-data-centers/earth-science-data-terminology-and-formats) of their data level nomenclature. The primary distinction is the expansion of the L3 definition to include single sensor/single time, single sensor/multiple time and multiple sensor/multiple time where by single time, we mean one satellite pass or orbit or one geostationary regional scan. "Multiple times" here refers to the merging of several satellite orbits or geostationary regional scans.

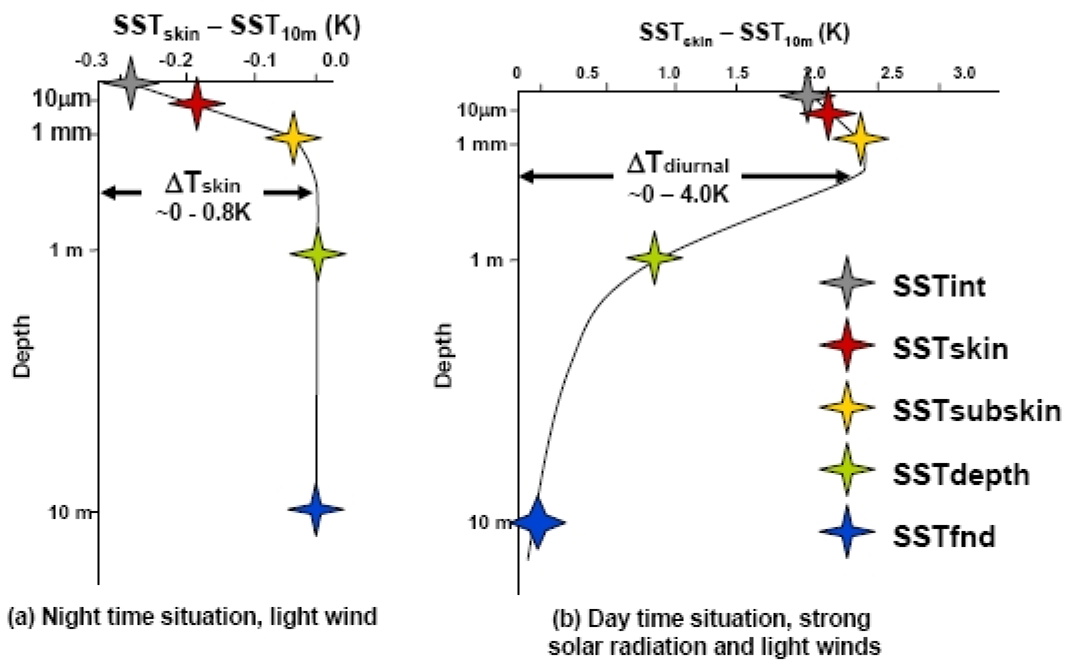
Table III.1. Definition of processing levels as used in this document.

Data_Level	**Description**
Level 0	Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed. (In most cases, the EOS Data and Operations System (EDOS) provides these data to the DAACs as production data sets for processing by the Science Data Processing Segment (SDPS) or by a SIPS to produce higher level products.)
Level 1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform ephemeris) computed and appended but not applied to the Level 0 data.
Level 1B	Level 1A data that have been processed to sensor units (not all instruments have Level 1B data).
Level 2	Derived geophysical variables at the same resolution and location as Level 1 source data.
Level 3	Single sensor/single time Level 2 fields mapped on uniform space-time grid scales
Level 3 Collated	Single sensor/multiple time collated and mapped on uniform space-time grid scales.
Level 3 Super Collated	Multi-sensor/multiple time collated and mapped on uniform space-time grid scales.
Level 4	Model output or results from analyses of lower level data

Appendix IV. Definitions of surface and near-surface temperatures

The following was extracted from the GHRSST web page: http://ghrsst.org/SST-Definitions.html

The figure below presents a schematic diagram that summarises the definition of SST in the upper 10m of the ocean and provides a framework to understand the differences between complementary SST measurements. It encapsulates the effects of dominant heat transport processes and time scales of variability associated with distinct vertical and volume regimes of the upper ocean water column (horizontal and temporal variability is implicitly assumed). Each of the definitions marked in the bottom right of the figure is explained in the following sub-sections.



The hypothetical vertical profiles of temperature for the upper 10m of the ocean surface in low wind speed conditions during the night and day shown in the figure encapsulate the effects of the dominant heat transport processes and time scales of variability associated with distinct vertical and volume regimes (horizontal and temporal variability is implicitly assumed).

The interface temperature (SSTint)

At the exact air-sea interface a hypothetical temperature called the interface temperature (SSTint) is defined although this is of no practical use because it cannot be measured using current technology.

The skin sea surface temperarature (SSTskin)

The skin temperature (SSTskin) is defined as the temperature measured by an IR radiometertypically operating at wavelengths 3.7-12 μm (chosen for consistency with the majority of IR satellite measurements) that represents the temperature within the conductive diffusion-dominated sub-layer at a depth of ~10-20 μm. SSTskin measurements are subject to a large potential diurnal cycle including cool skin layer effects (especially at night under clear skies and low wind speed conditions) and warm layer effects in the daytime.

The sub-skin sea surface temperature (SSTsub-skin)

The sub-skin temperature (SSTsub-skin) represents the temperature at the base of the conductive laminar sub-layer of the ocean surface. For practical purposes, SSTsub-skin can be well approximated to the measurement of surface temperature by a μ-wave radiometer operating in the 6-11 GHz frequency range, but the relationship is neither direct nor invariant to changing physical conditions or to the specific geometry of the μ-wave measurements.

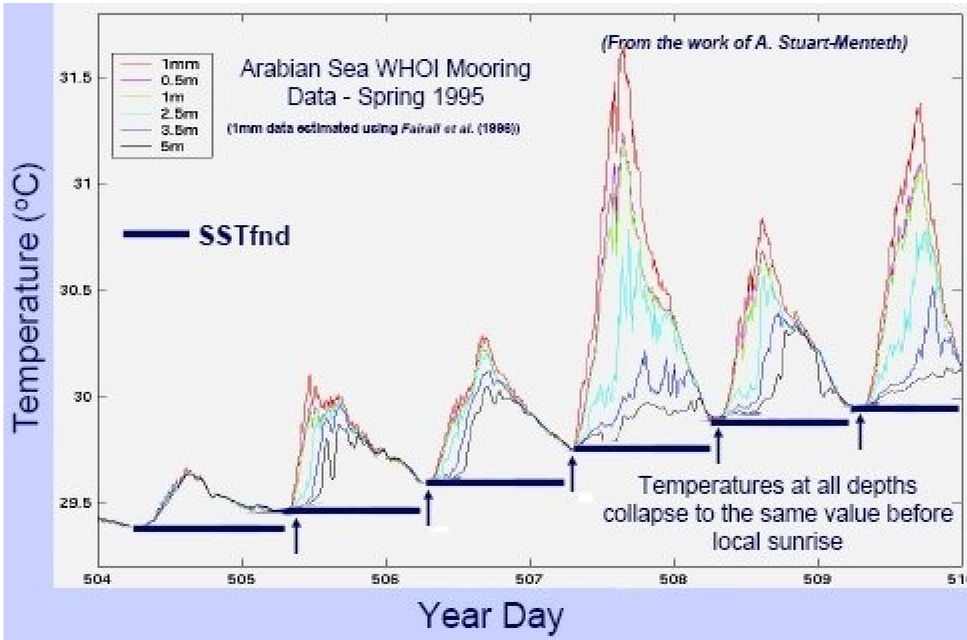
The surface temperature at depth (SSTz or SSTdepth)

All measurements of water temperature beneath the SSTsub-skin are referred to as depth temperatures (SSTdepth) measured using a wide variety of platforms and sensors such as drifting buoys, vertical profiling floats, or deep thermistor chains at depths ranging from 10^{-2} - 10^{3}m. These temperature observations are distinct from those obtained using remote sensing techniques (SSTskin and SSTsub-skin) and must be qualified by a measurement depth in meters (e.g., or SST(z) e.g. SST5m).

The foundation temperature (SSTfnd)

The foundation SST, SSTfnd, is defined as the temperature of the water column free of diurnal temperature variability (daytime warming or nocturnal cooling) and is considered equivalent to the SSTsub-skin in the absence of any diurnal signal. It is named to indicate that it is the foundation temperature from which the growth of the diurnal thermocline develops each day (noting that on some occasions with a deep mixed layer there is no clear SSTfnd profile in the surface layer). Only in situ contact thermometry is able to measure SSTfnd and analysis procedures must be used to estimate the SSTfnd from radiometric satellite measurements of SSTskin and SSTsub-skin. SSTfnd provides a connection with the historical concept of a "bulk" SST considered representative of the oceanic mixed layer temperature and represented by any SSTdepth measurement within the upper ocean over a depth range of 1-20+m. SSTfnd provides a more precise, well-defined quantity than previous loosely defined "bulk" SST and consequently, a better representation of the mixed layer temperature. In general, SSTfnd will be similar to a night time minimum or pre-dawn value at depths of ~1-5 m, but some differences could exist. Note that SSTfnd does not imply a constant depth mixed layer, but rather a surface layer of variable depth depending on the balance between stratification and turbulent energy and is expected to change slowly over the course of a day.

The figure below provides an example of SSTfnd.



Appendix V. Other SST Products

V.1 Potential SST products and errors based on in situ observations

Over the past 30 years, the most useful observations for making CDRs are those obtained from drifters, moorings, Argo, research-ship-based CTDs, Volunteer Observing Ships (VOS), and XBTs. In addition, occasional air/sea interaction process studies (i.e., TOGA/COARE, JASMINE, EPIC, CBLAST, and others) provide valuable data sets for calibration and validation activities, and to quantify the regional SST variability in the ocean in order to better determine error budgets for SST products. In addition, these process studies often include airborne and ship-based radiometer measurements of SST. The near-surface in situ observations are generally made at least O(1m) below the surface, their shallowest depth, although they may be reported as surface observations. Table V.1 lists the measurement techniques together with their characteristics.

Table V.1: In situ SST sampling methods, characteristics, and errors.

Measurement	Measurement Accuracy (K)	Measurement Depth (m)	Spatial Coverage	Temporal Coverage	Sampling Error (K)

Drifters	0.1	1	2000, 1 per 5°x5° box	Daily	0.2-0.3
Moorings	0.1	1-10, + T(z)	O(100) sites	Hourly-Daily	TBD
Argo	0.005	4, + T(z)	3000, 1 per 3°x3° box	Every 10 days	0.1-0.2
WOCE STD CTDs	<0.005	0, 2, 4, 6, 8…	Sections	1-10 stations per day	TBD
VOS	0.2?	Variable	Ship Tracks	4 per day	TBD
XBTs	0.2?	Variable	Ship Tracks	4-10 per day	Var. 0.1-0.5

Near-surface in situ SST observations need to be categorized in terms of their ability to provide an estimate of SST(foundation), or SST(F). This requires inclusion of appropriate metadata on time of day (at least), wind and air/sea flux conditions (at least), and their multi-day history (if available, possibly using atmospheric reanalysis model output). Historically, the in situ data have not been reported in terms of providing added value to satellite SST measurements.

> *Recommendation: A research effort is needed to revisit the data sets listed in Table V.1, identify those data that provide a reliable estimate of SST(F) and associated mixed layer depth and upper ocean T(z) structure.*

This would be a valuable data set for improving the upper ocean diurnal warming model and characterizing the associated errors.

As we move into the future, the Argo data set will likely be the most valuable as a basis for a CDR since the temperature sensors are well calibrated, near-surface measurements of T(z) are made, less sensor drift is likely due to the limited time at the surface, a quality control system is in place, and relatively uniform space/time sampling is planned. The drifter and mooring data will provide additional valuable data. Opportunities for cross-calibration of the three systems will be plentiful during the individual instrument's lifetime. Recommendation:Advancements in the sampling methodology of the Argo floats in the near-surface zone are recommended to enhance the utility of the Argo data in support of satellite-based products.

Moving backward in time before the Argo and drifter period, it is essential to use the other measurement techniques identified in Table V.1, as well as the earlier data sets based on hydrographic bottle data and mechanical BTs. High (research) quality subsets of these data sets have been produced by J. Ried, R. Curry (HydroBase) and others. A similar quality control effort on the T(z) profile data is ongoing at CSIRO by Wijffels and collaborators. The ISSTST should track these activities and incorporate advances into ongoing and future CDR optimization procedures.

In summary, potential candidate data sets useful as CDRs could include: an Argo only product, a drifter only product, a T(z) product based on all historical profile data, a merged in situ data only product, a variety of individual and merged satellite-only products, and a variety of merged in situ and satellite products with various space/time averaging. The optimum product will depend on the requirements of the particular application.

V.2 SST products and errors based on model-based products

At the present time, operational SST products produced at US modeling centers are available as nearly global nowcast and forecast products from NLOM, NCOM and HYCOM models with eddy-permitting and eddy-resolving spatial resolution. Higher resolution regional models as part of IOOS activities as also available. Generally, these products provide no error estimates. At coarser resolution, ocean products from the ECCO group, the HYCOM consortium, and NOAA labs are also available. Currently, these products are probably not appropriate as a basis for CDRs. However, over the coming decade advances in modeling and reanalysis capabilities may result in useful model-based ocean products appropriate for use as CDRs as is currently the situation at the atmospheric research community. The ISSTST should stay abreast of these ongoing efforts and collaborate as appropriate.

Appendix VI. Acronyms

ABI	Advanced Baseline Imager
AIRS	Atmospheric IR Sounder
APC	Antenna Pattern Correction
ATSR	Along-Track Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
CDR	Climate Data Record
CLARREO	Climate Absolute Radiance and Refractivity Observatory
CrIS	Cosmic Ray Isotope Spectrometer
CRTM	Community RTM
DOC	Department of Commerce
DoD	Department of Defense
DVWG	Diurnal Variability Working Group
ENSO	El Niño-Southern Oscillation
ESA	European Space Agency
GCOM-W	Global Change Observation Mission – Water
GHRSST	Group for High Resolution Sea Surface Temperature
GMI	GPM Microwave Imager
GOCART	Goddard Chemistry Aerosol Radiation and Transport (model)
GOES	Geostationary Operational Environmental Satellite

GPM	Global Precipitation Mission
GSICS	Global Space-based Intercalibration System
GTS	Global Telecommunications System
IASI	IR Atmospheric Sounding Interferometer
IFOV	Instantaneous Field of View
IORD-II	Integrated Operational Requirements Document-II
IR	Infrared
ISSTST	Interim Sea Surface Temperature Science Team
JAXA	Japan Aerospace Exploration Agency
Ln	Level n of data processing, n=0, 1, 2, 3 or 4
LST	Local Sun Time
MetOp	Meteorological Operational Satellite
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
MWRI	Microwave Radiation Imager
μ-wave	Microwave
NAAPS	Navy Aerosol Analysis and Prediction System
NASA	National Aeronautics and Space Administration
NIRST	New IR Sensor Technology
NOAA	National Oceanic and Atmospheric Administration
NPOESS	National Polar-orbiting Operational Environmental Satellite System
NPP	NPOESS Preparatory Project
NWP	Numerical Weather Prediction
OA	Objective Analysis
OE	Optimal estimation
OI	Optimal Interpolation
PDO	Pacific Decadal Oscillation
QC	Quality Control
RFI	Radio Frequency Interference
RTM	Radiative Transfer Model
SAC-D	Satélite de Aplicaciones Científicas - D
SEVIRI	Spinning Enhanced Visible and IR Imager

SI	System International
SST	Sea Surface Temperature
TB	Brightness Temperature
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Measuring Mission
TOA	Top-of-Atmosphere
TV	Television
TXR	Transfer Standard Radiometer
URL	Uniform Resource Locator
US	United States
VIIRS	Visible/IR Imager Radiometer Suite
WindSat	