Customer	**: ESRIN**	**Document Ref**	**: SST_CCI-UCR-UOE-001**
WP No	**: WP21610**	**Issue Date**	**: 04 December 2013**
		Issue	**: 3**

Project : **CCI Phase 1 (SST)**

Title : **Uncertainty Characterisation Report: Sea Surface Temperature v3**

Abstract : This Uncertainty Characterisation Report (UCR) documents the best current understanding of uncertainties (i.e., components of error distributions) in SST_cci's SST observations and products.

Author : ______________________ **Checked** : ______________________

C J Merchant
University of Edinburgh

Paul Spinks
Space ConneXions Ltd

Digitally signed by Craig Donlon
DN: cn=Craig Donlon, o=European Space Agency, ou=EOP-SME, email=craig.donlon@esa.int, c=NL
Date: 2013.12.10 17:05:20 +01'00'

Accepted

C Donlon
(ESA Technical Officer)

Distribution : SST_cci team members

ESA (Craig Donlon)

AMENDMENT RECORD

This document shall be amended by releasing a new edition of the document in its entirety. The Amendment Record Sheet below records the history and issue status of this document.

AMENDMENT RECORD SHEET

ISSUE	DATE	REASON FOR CHANGE
A	3 Oct 2011	Initial Issue
B	3 Oct 2011	Updated to reflect comments by reviewers
C	5 Oct 2011	Cosmetic and editorial changes (by Project Manager)
D	9 Jan 2012	Implemented agreed responses to RIDS raised by ESA.
E	16 Jan 2012	Final version accepted by ESA
1	27 Jan 2012	Accepted
2A	5 Feb 2013	Draft updated from v1.
2B	12 July 2013	Additional sections 3.3.6, 3.2.4 and 5.3 have been added at request of ESA TO.
2C	15 July 2013	Cosmetic and editorial changes (by Project Manager)
2	15 July 2013	Issued
3A	15 Nov 2013	Draft updated and extended from v2.
3	4 Dec 2013	Submitted for acceptance

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Purpose and Scope

The Sea Surface Temperature Climate Change Initiative (SST_cci) is prototyping and demonstrating a climate date record (CDR) for the SST essential climate variable (ECV). Best scientific practice and the results of the SST_cci survey of climate users [RD.171] requires that SST products be characterized with respect to their observational uncertainty.

The purpose of this Uncertainty Characterisation Report (UCR) is therefore to document the best current understanding of uncertainties (i.e., components of error distributions) in SST_cci's SST observations and products.

The scope of the UCR extends across all the sensors from which the SST_cci project retrieves SSTs within its prototype processing chain. These are principally, for the long-term SST system, the Along Track Scanning Radiometers (ATSRs) and the Advanced Very High Resolution Radiometers (AVHRRs). However, many of the generic principles have wider applicability. The Spinning Enhanced Visible and Infra-Red Imager (SEVIRI) and the Advanced Microwave Scanning Radiometer – E (AMSR-E) will also be used within the SST_cci short-term, demonstration system, with additional uncertainty characterisation undertaken within the project.

The scope of the UCR is intended to be sufficient to support the use of SST_cci products in the contexts of climate change detection and attribution and other climate science applications where it is critical to understand signal uncertainty relative to natural variability.

This is version 3 of the UCR. Use cases for uncertainty information have been added. Information about the uncertainty estimates in products has been updated in the light of the SST CCI Product Validation and Intercomparison Report (RD.325, PVIR) and Climate Assessment Report (RD.326; CAR).

1.2 Referenced Documents

The following is a list of documents with a direct bearing on the content of this report. Where referenced in the text, these are identified as RD.n, where 'n' is the number in the list below. The numbering is consistent with a master list of reference documents maintained by the project.

RD.150 Systematic Observation Requirements for Satellite-based Products for Climate: Supplemental Details to the satellite-based component of the "Implementation Plan for the Global Observing System for Climate in support of the UNFCCC (GCOS-92)", GCOS-107, September 2006 (WMO/TD No.1338)

RD.171 CCI Phase 1 (SST) (2010), User Requirements Document, Reference SST_CCI-URD-UKMO-001

RD.191 Bureau International des Poids et Mesures, Guide to the Expression of Uncertainty in Measurement (GUM), JCGM 100:2008, 2008. Available online at http://www.bipm.org/en/publications/guides/gum.html

RD.325 SST CCI Phase 1, Product Validation and Intercomparison Report (PVIR) [Not yet published]

RD.326 SST CCI Phase 1, Climate Assessment Report 9CAR) [Not yet published]

1.3 Definitions of Terms

AATSR	Advanced Along-Track Scanning Radiometer
ARC	ATSR Reprocessing for Climate
ATSR	Along-Track Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
BT	Brightness Temperature
CCI	Climate Change Initiative
CDR	Climate Data Record
CEOS	Committee on Earth Observing Satellites
CLAVRx	Clouds from AVHRR Extended
CMUG	Climate Modelling User Group
ECMWF	European Centre for Medium-Range Weather Forecasts
ESA	European Space Agency
GAC	Global Area Coverage
GCOS	Global Climate Observing System
GDS	GHRSST Data Specification
GHRSST	Group for High Resolution SST
GUM	Guide to Uncertainty in Measurement
L2 / L2P	Level 2 / Level 2 pre-processed
L3 / L3C	Level 3 / Level 3 Collated
L4	Level 4
LUT	Look-up Table
Metop	METeorological OPerational (satellite)
NEDT	Noise Equivalent Differential Temperature
NESDIS	National Environmental Satellite, Data and Information Service, USA
NOAA	National Oceanic and Atmospheric Administration (USA)
NWP	Numerical Weather Prediction
PDF	Probability density function
QWG	Quality Working Group

RT	Radiative Transfer
RTTOV	Radiative Transfer for the Television and Infrared Orbiting Satellite Operational Vertical Sounder
SD	Standard Deviation
SSES	Single Sensor Error Statistics
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SST	Sea surface temperature
WGCV	Working Group on Calibration and Validation
WP	Work Package

2. NOMENCLATURE FOR UNCERTAINTY CHARACTERISATION

2.1 Importance of Careful Usage

Deep understanding of concepts related to error and uncertainty is challenging. The challenge is heightened if uncertainty terms are used loosely. Unfortunately, careless usage is common. An example is the word "error", which is widely used to mean "difference from truth" (correct) and "statistical dispersion of a repeated measurement" (incorrect)[1]. It is not uncommon for both meanings to be intended in the same sentence. The danger is that the underlying concepts corresponding to these different meanings are confused, often with very practical consequences. For example: if you are told "the error in this measurement is 0.27 K", are you to infer that the true value is the measured value minus 0.27 K, or that the measured value is uncertain to within +/-0.27 K? The answer can often be inferred from the context and the conventions of a particular sub-discipline; but it would be better to be confident that the word "error" had been used correctly. (The correct usage is the first.)

2.2 Agreed CCI Guidelines and Other Sources

During the first CCI Colocation Meeting at ESRIN on 15-17 September 2010 an open discussion on uncertainty characterization was held, attended by members of all CCI projects. A "drafting team" was tasked to discuss common issues relevant to uncertainty characterization and to draft relevant useful guidelines, which are followed in this Report, except where explained in footnotes.

The drafting team consisted of the following members:

–*Roland Doerffer (CCI Ocean Colour)*

–*Chris Merchant (CCI SST)*

–*Pierre Defourny (CCI Land Cover)*

–*Martin Schultz (CCI Fire)*

–*Don Grainger (CCI Cloud & CCI Aerosol)*

–*David Tan (CCI CMUG)*

–*Sylvia Kloster (CCI CMUG)*

–*Simon Pinnock (ESA)*

The drafting team took account of definitions in standard documents including "GCOS-107" [RD.150] and "GUM" [RD.191]. Subsequently, it is noted that the draft GCOS update (under review at time of writing) modifies some of the GCOS-107 conventions that had been adopted from GCOS-107 by the drafting team, particularly redefining accuracy and stability (see below).

---

[1] *A third definition is implicit in the nomenclature of "Single Sensor Error Statistics" (SSES) from the Group for High Resolution Sea Surface Temperature (GHRSST). The SSES statistics actually describe the distribution of differences from a validation reference (usually drifting buoys), and thus the "error" in SSES is actually what is better termed as "discrepancy" [RD.191].*

2.3 Definitions of Terms in SST_cci

2.3.1 Describing Error and Uncertainty

A measurement is a set of operations having the object of determining the value of a quantity. Following RD.191 it is helpful to define the term **measurand** as

measurand: particular quantity subject to measurement,

so that the phrases 'true value of a quantity' and value of the measurand are synonymous. For the SST_cci, the measurand is sea surface temperature (for level 2 and higher data).

Very few instruments directly measure the measurand. Generally an instrument reports the effect of a quantity from which the magnitude of the measurand is estimated. As an example, an instrument sensitive to infrared light might be used to measure the temperature of an object.

The process of measurement is inexact. The difference between a measured value and the value of the measurand is called the **error**. Traditionally, the word 'error' has also meant a numerical value that estimates the variability of the error if a measurement is repeated (i.e. a width of the distribution of possible errors). This dual meaning of "error" can lead to confusion or ambiguity. To separate these meanings and avoid confusion the RD.191 definitions are used, i.e.

error (of measurement): result of a measurement minus a true value of the measurand,

uncertainty (of measurement): is a parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

Generally, the "true" value of the error is not known. The likely magnitude of the error can and should usually be estimated however. To do this, the error is often viewed as having one or more random components and one or more systematic components. Following RD.191 the definitions of these terms are:

random error: result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions,

systematic error: mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions minus the true value of the measurand.

In this framework, the random error is variable from measurement to measurement, whereas the systematic error is the same for each measurement. Although it is not possible to compensate for the random error, its effect on the estimate of the measurand can be reduced by averaging over a number of independent repeat observations, if available.

The statistical distribution of random error can be described by a probability density function (pdf) of which the **expected value** (i.e., the average over the pdf) is zero. As the random error often arises from the addition of many effects the central limit theorem suggests that a Gaussian distribution is a good representation of this pdf. Gaussian distributions are fully defined by just two parameters (mean and standard deviation, SD). So, when error distributions are Gaussian, the SD over the pdf of the (unknown) random error of a single observation is equal to the SD that would be obtained from a large number of repeated measurements of the same quantity under the same conditions. This SD quantifies the **random uncertainty** in a single observation.

If the difference between measured value and truth has been usefully estimated, this can be subtracted from the measurement as a correction. This correction reduces the magnitude of the systematic component of the original error (it reduces bias), but there always remains a systematic error, whose sign is unknown but whose likely magnitude we can generally infer. This likely magnitude is the **uncertainty associated with systematic effects** (including the uncertainty associated with the bias correction applied). and should be quantified as the SD of the (estimated / "guess-timated") pdf of the remaining systematic effect.

As explained below, for satellite-retrieved fields in general, and for SST in particular, the neat division of error into random and systematic components is too simple. In reality, there is a spectrum of sources of error with greater or lesser degrees of spatio-temporal correlation between measurements. These correlations matter if averages or trends of satellite measurements are to be formed with appropriate attached uncertainty estimates.

2.3.2 Validation

Validation is defined by the CEOS WGCV as "*The process of assessing, by independent means, the quality of the data products derived from the system outputs*". Within SST_cci we take this to mean the assessment of a measurement *and* of the uncertainty attributed to it. This is principally achieved by **external validation,** i.e. comparison of a measurement to an independent measurement and assessment of their consistency relative to their estimated uncertainties. This independent estimate of the measurand is termed the **validation value**. The discrepancy is then defined as

discrepancy: the difference between the measurement and the validation value.

A small mean discrepancy makes it plausible that systematic errors are small in both measurements and validation values. But it could also result from a fortuitous cancellation of systematic errors that happen to be similar in both data sets.

A small standard deviation of the discrepancies indicates small random errors in both measurements and validation values, *if* there are sound reasons for assuming the errors in the two data sets are independent.

For a small number of measurements it is possible to report individual discrepancies. However, for the large number of measurements typical of satellite remote sensing, validation involves statistically characterising the discrepancies. There are often regimes of instrument behaviour for which uncertainties can be expected to differ, so it is usual to characterize discrepancies for the minimum number of regimes of consistent instrument behaviour. The choice of regimes could come from a cluster analysis of discrepancy (if the difference in regimes causes differences in systematic error), but more commonly comes from knowledge of the measurement process.

The statistical characterization of the discrepancies within a regime is made through three **quality parameters.** Consider the set of n measurements $\{x_1 \pm \delta x_1, x_2 \pm \delta x_2, x_3 \pm \delta x_3, \ldots x_n \pm \delta x_n\}$ of some quantity together with the set of validation values $\{v_1 \pm \delta v_1, v_2 \pm \delta v_2, v_3 \pm \delta v_3, \ldots v_n \pm \delta v_n\}$ made of the same quantity. The quality parameters are then:

(Relative) Bias: the mean value of the discrepancy, i.e.

$$\mathrm{bias} = b = \frac{\sum_{i=1}^{n}(x_i - v_i)}{n}$$

Chi-squared: the goodness of fit between the actual and estimated uncertainties of measurement and validation values, defined by

$$\chi^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - v_i)^2}{\delta x_i^2 + \delta v_i^2}$$

Stability: the change in bias with time defined as

$$\text{stability} = \frac{b(t+\Delta t) - b(t)}{\Delta t}$$

The expectation value of the bias is the sum of the residual systematic errors in the measurement and the validation value. The bias can only be attributed to the measurement if the residual systematic error in the validation value is known a priori. In an ideal case the bias would be zero.

The expected value for χ^2 is unity. A value lower than this indicates the uncertainties attributed to the measurements or the validation values or both are too high. A value greater than unity indicates the uncertainties attributed to the measurements or the validation values or both are too low.

In the ideal case the stability would be zero over any timescale. In remote sensing the stability can display periodicity related to factors such as instrument drift or solar illumination of the satellite - both over an orbit and seasonally. It is suggested that the stability is estimated at the same temporal scale that any trends or variability in the measurements are calculated.

It may be that the quality parameters are independent of the measurement magnitude and conditions of measurement and apply at all locations and times. In that case the three quality values adequately characterize the quality of measurement. More commonly, the quality values vary so a **validation table** is used to summarise the bias, χ^2 and stability for regimes of consistent instrument behaviour.

In some case **internal validation** can be used to check reported uncertainty. Consider the situation where an instrument measures the same quantity under conditions where the reported uncertainty does not vary. Then the variability of the measurements should agree with the reported random uncertainty.

2.3.3 Accuracy and Precision

For the term “accuracy” there seems to be two definitions in common circulation. In RD.150, GCOS considers accuracy to be measured by

> “the bias or systematic error of the data, i.e., the difference between the short-term average measured value of a variable and the truth”

where the average referred to has been sufficient to render the random uncertainty in the measured value negligible.

In contrast, the definition from the GUM [RD.161] is also used, whereby accuracy is

> “the closeness of agreement between the result of a measurement and a true value of a measurand”

and therefore a measurement can be inaccurate either by virtue of a large systematic error or because it has a large random uncertainty.

In SST CCI, will use “accurate” only in a qualitative sense to mean the closeness of agreement between a measured value and the truth (which includes both systematic and random effects).

The scatter that repeated measurements would have (whether fully or partly random in origin) is referred to qualitatively by the term **precision**. In SST CCI, precise measurements are therefore measurements that are “highly repeatable”.

2.3.4 Stability

S**tability** is the degree to which systematic effects in SST measurements are invariant over time (as in RD.150).

2.3.5 Representativity (Sampling Uncertainty)

Fields of satellite SST are not spatio-temporally complete. It is common for climate products to be formed of some spatio-temporal average, and indeed long—term SST datasets from SST_cci will be “L3C” products: daily files consisting of the mean day-time or night-time SST observed in 0.05° latitude-longitude cells on a particular day. The **sampling uncertainty** is the standard deviation of the difference between the average actually observed and the average that would be observed were the cell fully observed. In general, this will be a function of the fraction of the cell that is observed, which may range from a single pixel to complete observation (all possible pixels valid).

Since cells are not in general fully observed (although some are) the standard deviation cannot be directly evaluated. However, indirect means can and will be used to develop a parameterisation of the sampling uncertainty as a function of the fraction of the possible coverage in a cell for which valid SSTs are actually obtained.

A similar issue arises when forming large-scale averages from 0.05° resolution data. A typical example would be the formation of monthly global or hemispheric SSTs from sparse daily data. Again, a parameterisation of the sampling uncertainty will be developed for use within SST_cci tools undertaking such averaging.

2.3.6 Uncorrelated Effects

A useful means of categorising components of the total uncertainty in SST is the degree to which errors are in common between different SST measurements. This is useful, because when gridding / averaging of the data is undertaken, it enables uncertainties to be properly propagated.

The uncertainty from uncorrelated effects is the component of the total uncertainty which is due to random errors (effects that are uncorrelated) between any selected pair of SST measurements (i.e., between any pair of pixels in an image). (In the SST CCI products and earlier SST CCI documents, the quantification of this uncertainty component is referred to as “uncorrelated_uncertainty”.)

When *n* SST measurements are averaged, the uncertainty in the average associated with uncorrelated effects (σ_{mean}) is decreased relative to the uncertainty σ in each individual measurement by the classic “1/√n” rule:

Eq 1. $$\sigma_{mean} = \sigma / \sqrt{n}$$

2.3.7 Synoptically Correlated Effects

It is common in surface remote sensing for a major component of the retrieval (inversion) process to be accounting for the effect that atmospheric variability has on the measurements. Since no inversion is perfect, a component of the error in an SST estimate depends on the atmospheric conditions. Since atmospheric conditions are correlated over "synoptic" spatio-temporal scales, it therefore follows that there is an uncertainty component that is due to effects that are **synoptically correlated**. In other words, there is a component of error that is (to some degree) in common between nearby measurements (i.e., between pixels separated by scales shorter than the synoptic scale).

When *n* nearby measurements are averaged, there is no "1/√n" reduction in this component of uncertainty.

Clearly, there is no clean separation between those pixels that share some synoptically correlated errors and those that are sufficiently separated to have no such correlated component. Instead, such errors decorrelate gradually, probably over scales that depend on the synoptic situation. The determination of actual correlation length scales of synoptically correlated uncertainty in SST estimates is an important area for future research in improving SST uncertainty estimates that cannot be fully addressed within the SST_cci project.

In SST_cci, approximate methods of propagating uncertainty associated with synoptically correlated effects will be applied, based on estimates of correlation length scales in IR SST retrievals.

Uncertainties associated with synoptically correlated effects are aggregated thus*:*

$$\frac{1}{\eta}\sqrt{\sum_{1}^{n}\sigma_i^2}$$

where η is the effective number of synoptic areas in the area being averaged across . η is defined by:

$$\eta = \frac{n}{1 + r^{\alpha}(n-1)}$$

This formulation behaves appropriately in the limit of uncorrelated effects (*r* ~ 0 recovers *n*) and fully correlated effects (*r* ~ 1 recovers 1, i.e., no averaging from fully correlated errors). The appropriate values of the parameterisation remain to be properly determined, and plausible values are proposed below. These values are based on the physical understanding of the retrieval process and atmospheric behaviour. The parameters are

$\alpha = 1$ for which the denominator interpolates the limiting cases linearly

and $r = \exp\left(-1/2\left(\frac{d_{xy}}{l_{xy}} + \frac{d_t}{l_t}\right)\right)$ which models the degree of correlation in a new spatio-temporal box of dimension d_{xy} km and period d_t as exponential. We adopt plausible scales l_{xy} = 100 km and l_t = 1 day. These require to be optimised via further research in future.

2.3.8 Large-scale Correlated Effects

Uncertainty associated with large-scale correlated effects is the component of SST uncertainty that typically persists on scales longer than the synoptic (e.g., monthly to years, 1° to global). Origins of error causing this form of uncertainty include:

- sensor calibration (which may evolve slowly over time and/or have orbital components)
- systematic error in the SST retrieval method (e.g., seasonally persistent regional bias patterns)
- impacts of long-lived stratospheric aerosol (zonal to global, lasting a few years, affecting SSTs estimated from infra-red observations)
- uncorrected forward modelling error in radiative transfer based retrievals

In SST_cci it will be assumed that uncertainty associated with large-scale correlated effects is not reduced by averaging of measurements from a given sensor. If averaging SSTs from several sensors using common wavelengths (e.g., averaging SSTs from similar infra-red sensors), SST uncertainty from sensor calibration uncertainty may average down (if sensors are not cross calibrated); other large-scale correlated effects may not, depending on the details of the channels and retrieval methods. If averaging SSTs from radically different sensors (e.g., an infra-red and passive microwave sensor), the uncertainty associated with large-scale correlated effects will be assumed to reduce accordingly. In some cases, there will be limited information on which to base estimates of the size of large-scale correlated effects, but their inclusion in SST_cci products will be a significant step forward in uncertainty estimation for satellite SST.

3. SOURCES OF UNCERTAINTY

3.1 Estimation (SST retrieval)

3.1.1 Radiometric noise

Radiometric noise is a random effect in the observed signal arising from detector fluctuations. In SST_cci, it is usually expressed as the noise equivalent differential temperature (NEDT) – i.e., the uncertainty in brightness temperatures arising from random effects. The NEDT distribution is generally well represented as a random error (uncorrelated between pixels) with Gaussian statistics. Slight deviations from this are present, e.g. in ATSR imagery when cosmetic fill is used: since one pixel is copied to another pixel location in the image to fill a gap, there is perfect correlation of error between these occasional duplicates. Such deviations are generally negligible for uncertainty estimation purposes.

Information on NEDT should be available associated with measurements from all instruments, although the details of what is available may range from a general literature-based estimate to a per-scan NEDT estimate from the standard deviations of observations of a uniform black-body target. NEDT is often scene temperature dependent, because the Planck function transforms radiance to brightness temperature non-linearly.

In SST_cci, an estimation model for the uncertainty associated with uncorrelated radiometric noise for each sensor is applied within the optimal estimation retrieval used (RD.249).

3.1.2 Uncertainty Associated with Retrieval Algorithm

Inversion algorithms, even in simulation with simulated observations that are "perfectly calibrated" and "noise free", retrieve an SST that is discrepant with the truth. In a simulation study, the simulation-truth can be known, and the standard deviation of the discrepancies from the simulated retrievals is the **algorithmic uncertainty** arising from this retrieval error.

Algorithmic uncertainty thus arises from synoptically correlated effects (in general the retrieval error is a function of the atmospheric conditions), and will be propagated accordingly.

In optimal estimation algorithm used to derive SST in SST_cci, the estimation model for the uncertainty associated with synoptically correlated retrieval error is represented by the uncertainty associated with the propagation of forward model errors (the terms ε_{RT} in tables 2.1 and 2.3 of RD.249).

3.1.3 Uncertainty Associated with Calibration and Forward Model

In SST_cci, the aim is to provide a climate data record for SST that is independent of *in situ* observations, using radiative transfer (RT). In this situation, both the uncertainty associated with sensor calibration, and the uncertainty associated with the forward model (i.e. arising from errors in our ability to simulate brightness temperatures (BTs) of a particular sensor channel) are important. These two effects lead to similar uncertainties associated with SST.

For the period of the ATSR sensors, RT-based retrieval is achievable because experience from the ARC project [RD.184] suggests that

- for AATSR, the ARC line-by-line forward model and sensor calibrations for the 3.7 and 11 μm channels are likely consistent to within 0.03 K
- for ATSR-2, all channels are likely consistent to within 0.05 K
- for ATSR-1, all channels (when available) are likely consistent to within 0.1 K (except perhaps the 12 μm channel towards the end of the routine mission)

These rough estimates are inferred from the range of mean discrepancies compared to drifting buoys observed when ARC radiative-transfer-based coefficients are used to derive SSTs.

In ARC, overlap periods between ATSR2/AATSR and ATSR1/ATSR2 are used as far as possible to adjust the forward modelling of BTs and SST retrieval coefficients to be consistent with AATSR channels 3.7 and 11 μm. This is called “level 1 harmonisation”. In this way, the forward-model-relative-to-calibration errors are reduced for AATSR channel 12 μm and for ATSR1 and ATSR2. The exceptions are for

- ATSR1 channel 3.7 μm, which failed within the first year and doesn’t overlap with ATSR2
- ATSR1 channel 12 μm, whose calibration is susceptible to the detector temperature drift experienced by ATSR1, and for which the overlap period with ATSR2 is not representative of the whole mission

Harmonisation of BTs can only be undertaken to a certain accuracy and precision, because of statistical limitations, and because of real geophysical variations between the times of observations of the paired sensors.

Within SST_cci, multi-sensor matches are used to reference BTs for AVHRR to ATSR BTs (RD.249). The aim is to bring the fast forward modelling of AVHRR sensors (using RTTOV10) to consistency with the AVHRR BT calibration to within 0.1 K, which would then support AVHRR SST retrieval by optimal estimation. This level of consistency has been achieved for some of the AVHRR data (RD.226). Future work on underlying calibration (RD.298) and improved cloud detection for AVHRRs (to improve the quality of inter-sensor matches) should improve results further.

Some error in the sensor calibration relative to the forward modelling capability will remain for all the ATSRs and AVHRRs, mostly likely of magnitude between 0.05 and 0.1 K. This residual error may evolve with time (as sensors age) and may have dependence on the scene temperature, atmospheric water content, etc. The uncertainty associated with these residual errors is large-scale correlated.

3.1.4 Uncertainty associated with skin effect and diurnal models

SST_cci will generate products containing two principal types of SST: the skin SST estimate at the time observed; and the depth SST estimated at a standardized observation time. The latter is derived from the former using a skin effect and diurnal stratification model. The model is to be defined, and is likely to be driven by numerical weather prediction (NWP) fields. The effect is therefore synoptically correlated. The parameterisation of the uncertainty associated with this effect is given in RD.249.

3.2 Sampling (representativity)

3.2.1 Spatio-temporal sub-sampling

In SSTs that are gridded (in the sense of averaged to some spatio-temporal bin or cell) the incomplete sampling of the cell causes an error between the mean measured SST and the (unknown, unobserved) true SST averaged across the cell. The sampling uncertainty (the standard deviation of this error over many instances) is a function of the fraction of the bin or cell sampled by measurements, with sparsely sampled cells having a larger uncertainty than those with near-complete measurement coverage.

Between cells, the sampling error will usually be uncorrelated. (Exceptional circumstances could arise where this is not the case: e.g., if a region of cold SST is unobserved because it is covered in fog.) In SST_cci, estimated sampling uncertainty will be propagated as a component of uncertainty that is uncorrelated between grid cells.

3.2.2 Clear-sky vs. cloudy sky difference

The question has been raised whether SSTs under clear sky conditions differ from SSTs under cloudy conditions. This is a complex question, depending on (at least) the following factors:

- which type of SST (skin or depth) one is considering
- the response (whose nature is scientifically controversial) of the ocean thermal skin to insolation and downwelling infra-red radiance
- the wind speed and the corresponding mixed layer depth, which affects the time scale of solar heating or night-time cooling
- the spatial structure of the cloud field

Whether on climate scales, the observing regime of infra-red sensors (cloud free conditions) is material is not known. This effect can only be noted at the present time, and will not be quantified within this phase of SST_cci.

3.2.3 Sampling biases from false detection of cloud

Cloud detection is effectively the classification of an image into clear and cloud segments. Many cloud detection schemes use lower thresholds on brightness temperatures to detect cloud. All classification schemes have a certain rate of false alarms (or “false detections”), and when schemes include such threshold tests, it has been observed (e.g., in the case of operational ATSR products) that one of the modes of false detection is flagging as cloud areas of anomalously cold water, such as cold core eddies. This introduces a systematic warm representativity error into any area average product. It is difficult to quantify this, and efforts in SST_cci focus on reducing the rate of false alarms by using improved cloud detection.

3.2.4 Microwave fields of view

For microwave SST instruments, the angular resolution is essentially inversely proportional to the observed frequency, for a given antenna design. MW SST sensors using multiple channels therefore have multiple true ground resolutions, with the lowest frequency channels having coarsest spatial resolution. Typically the lowest frequency is used in SST retrieval, but SSTs are returned on finer spatial grids (e.g. 25 km) than the intrinsic resolution (footprint) of the channel(s) at this frequency (e.g. 100 km). The low frequency channels are generally oversampled in space in order that the higher frequency channels are appropriately sampled. Thus, information at 25 km involves some deconvolution of over-sampled SST-sensitive channels, which introduces a level of SST uncertainty.

3.3 Contamination

3.3.1 SST retrieval bias from failure to detect cloud

Infra-red SST retrieval algorithms assume clear-sky conditions. Undetected cloudiness in a pixel perturbs the BTs from what they would otherwise be, and the SST is modified by the propagation of that perturbation through the retrieval algorithm. If the cloud impact is similar to a mode of atmospheric variability with which the retrieval algorithm does cope, the SST error arising may be small. However, in many cases, the BT effect of the cloud is unlike the effect of other atmospheric variability, and an SST error arises. For a single view retrieval, the SST error from this effect is usually negative, and overall, failures to detect cloud can introduce negative bias. For dual view retrievals, either sign of SST measurement/retrieval error is possible, depending on the situation, although cold errors seem more prevalent.

Failures to detect cloud are the main cause of SST retrieval errors that lie outside the (near-Gaussian) distribution of algorithm errors one expects from truly clear-sky observations. In other words, most satellite SST outliers are probably cloud related (for infra-red sensors). Although these errors have a low incidence rate, because the errors can be outliers to the usual distribution and are negatively skewed, they can cause bias in average SSTs.

3.3.2 Failure to detect ice

Particularly during polar night (visible channels not available), relatively warm sea ice can be erroneously interpreted as sea under clear sky. The “SSTs” inferred will often be colder than is physically plausible (i.e., may be below the freezing point of sea water at the salinity in question). As with cloud, the errors from this source are low incidence, but potentially relatively large and negatively skewed.

3.3.3 Elevated aerosol

Typical marine aerosols are represented in RT modelling for defining SST_cci retrieval coefficients. Nevertheless, desert dust outbreaks in particular may cause SST biases for infra-red observations. This is particularly a problem for the northeast tropical Atlantic and Arabian gulf seas. Desert dust indices are available for AATSR and SEVIRI thermal observations, as well as optical depth based aerosol products. These will be used to characterise the sensitivity of SST_cci products to such conditions.

3.3.4 Precipitation

Passive microwave SSTs are available through most cloud, unlike for infra-red sensors. However, the larger water droplets present in precipitating clouds scatter microwaves more efficiently, and lead to erroneous SST retrievals. When precipitation events are not detected, PMW SSTs can therefore be affected by a contamination error, usually cold.

3.3.5 Radio Frequency Interference

Passive microwave SSTs require filtering for errors caused by radio frequency interference. The sea surface can reflect human sources of microwave radiation, both terrestrial and in space, into the PMW sensor footprint, leading to a contamination error.

3.3.6 Proximity effects

Passive microwave sensors have sensitivity to incoming radiance described by an antenna pattern that, depending on the antenna design, may have significant “side-lobes” – sensitivity to radiance with minor peaks outside the main angular field of view (beam width). Where there are large contrasts in atmospheric scattering, emissivity and/or radiating temperature between the main beam and the side-lobe directions, SST retrieval as a representation of the SST in the main field of view will be biased. For this reason, microwave SSTs are not delivered up to the coasts or sea-ice edges. There may be some SST error from this effect along the edge of the provided SSTs. There is always an uncertainty introduced into SSTs from this effect when observing any SST field that is not spatially uniform.

Where atmospheric scattering is significant at IR wavelengths, there is in principle another potential proximity effect. A sharp contrast in emissivity and/or temperature leading to a strong gradient in surface emittance could lead to an excess of scattering in or out of the IR field of view. IR scattering is generally small in clear-sky atmospheres at the relevant wavelengths for SST over the open ocean, but this proximity effect could be relevant to coastal areas affected by Saharan dust, for example. This has not, to our knowledge, been thoroughly assessed.

3.4 Geolocation and collocation

3.4.1 Along Track Scanning Radiometers

The SST_cci project in this phase is exploiting the version 2.0 ATSR archive. It is known that there can be offsets in forward view imagery relative to nadir view imagery of up to ~5 km. Mean forward-nadir offsets for the different sensors have been derived within the ARC project and will be applied (as a shift to the forward view) in SST_cci, as follows:

- AATSR: -1 pixel across-track (xi) and -2 pixels along-track (xj)
- ATSR-2: +1 pixel across-track (xi) and -1 pixels along-track (xj)
- ATSR-1: +3 pixel across-track (xi)

The absolute geolocation of the nadir imagery in the version 2 archive seems precise to of order 2 km for all sensors. Mean nadir offsets are currently being evaluated by the ATSR QWG and their impact will be evaluated within SST_cci when available.

SST_cci will generate L3C products from ATSRs for the long-term climate data record, at 0.05° spatial resolution. The SST uncertainty from the absolute geolocation uncertainty will be assumed negligible compared to sampling uncertainty (see above).

A level 1 reprocessing for the ATSRs is expected during 2013. Adjustments to cone angle assumptions in the L1 processor have been made with the intention of removing the need for the above shifts. This will need to be assessed in due course.

3.4.2 Advanced Very High Resolution Radiometers

Before September 1992, geolocation of the AVHRR GAC L1B data as supplied by NOAA NESDIS is estimated to be within 8 km at nadir and 20 km at swath edge; after September 1992 these uncertainties improve to 4 km and 10 km respectively (http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/klm/html/c2/sec2-3.htm). For the SST_cci project, corrections to the along-track location have been applied to NOAA-11, NOAA-12, NOAA-14 (up to August 2000), and NOAA-15 (up to December 2000). The corrected locations are calculated using the CLAVR-x software using the timing corrections from the AVHRR Pathfinder project (http://yyy.rsmas.miami.edu/groups/rrsl/pathfinder/Processing/proc_app_a.html). No estimate has been made of the improvement in the along-track geolocation after applying these corrections. Timing and location corrections for later satellites are implemented in the NOAA NESDIS L1B processing or on board the satellite.

3.4.3 Passive Microwave Sensors

PMW retrievals simultaneously use channels with different spatial resolution on the surface. All channels are observed through the same antenna, of fixed dimension, and so the field of view is inversely related to the channel frequency. In the case of AMSR-E, for example, the spatial resolution varies from 5.4 km at 89 GHz to 56 km at 6.9 GHz. The spatial sampling interval is driven by the highest-frequency channel, meaning that the footprints of low frequency channels significantly overlap. The SST products to be used within SST CCI have 25 km resolution, a scale that is not resolved by the lower frequency channels (6.9 and 10.7 GHz). Thus the SST in one 25 km observation is in fact somewhat influenced by the SST in adjacent areas.

4. UNCERTAINTY ESTIMATION

4.1 General Concepts

In SST_cci, the following general concepts for uncertainty estimation are adopted

- every SST product (from individual measured values to gridded averages) will have SST uncertainty information attached
- SST products and outputs used in gridding/averaging to create SSTs of lower spatio-temporal resolution require uncertainty components to be attached, in order that uncertainty propagation can be appropriately undertaken; the component uncertainties need to be identified as
    - Uncertainty from uncorrelated effects
    - Uncertainty from synoptically correlated effects
    - Uncertainty from large-scale correlated effects
- within each of the above types of uncertainty (categorised according to how they propagate on averaging), there may be components arising from different origins: e.g., the uncertainty associated with synoptically correlated effects in an SST-depth will be the quadrature combination of the algorithm uncertainty in the corresponding SST-skin and the adjustment uncertainty of the skin-to-depth model
- uncertainties are as important an element of the product as the SSTs themselves
- uncertainties need to be validated

4.2 Uncertainty Algorithms

The SST retrieval algorithms used will require matched uncertainty algorithms. Thus, uncertainty algorithms have been defined for all algorithms that calculate an SST (RD.249).

Each numerical uncertainty value given is a “standard uncertainty” – i.e., is the estimated standard deviation of the error distribution. The units are therefore kelvin.

The objective of the SST CCI is to create SSTs that are unbiased to within 0.1 K on scales of 1000 km. No further corrections are available. The uncertainty associated with large-scale correlated effects is, in effect, a “systematic uncertainty”. It should not be mis-interpreted as a (signed) bias correction.

4.3 Uncertainty Products in SST_cci

SST_cci L2P/L3C products for AVHRR and ATSR will contain

- for each skin SST, the following uncertainty components
    - uncorrelated effects (estimated by propagation of radiometric noise through the retrieval process)
    - synoptically correlated effects (estimated algorithm uncertainty)

    - large-scale effects (type B evaluation of uncertainty)

- for each depth SST, the following uncertainty component will be additionally defined

    - adjustment uncertainty (estimated uncertainty from the skin-to-depth model, to be combined with the synoptically correlated uncertainty if uncertainties are further propagated)

SST_cci L3C demonstration products (SEVIRI and passive microwave) use the same fields as above, where appropriate.

SST_cci L4 products will contain total uncertainty as derived from the analysis procedure, having used SST_cci L2P/L3C uncertainties as inputs. With the currently available techniques, this breaks the chain of formal propagation of uncertainties. The L2P/L3U uncertainties provided do weight the SSTs in the L4 analysis procedure, and are used in the analysis to generate the analysis uncertainty. However, the L4 SST uncertainty associated with the analysis does not account for the decomposition of the uncertainty in the input data by the degree of correlation of the effects. This is essentially because the SST-cci L2P/L3U data are providing this uncertainty information for the first time, and research and experience will be required in future to learn how to use these in the L4 system.

4.4 Validation of Uncertainty Products

The process of SST validation has often been used to generate estimates of uncertainty. An example is the recommended approach of GHRSST to determining SSES.

In SST_cci, validation will be the process of assessing both SST products and the attached SST uncertainty products, using external validation values. The chi-square statistic (section 2.2) is a useful statistic for assessing the degree to which the product uncertainties, when combined with knowledge of the validation value uncertainties, accounts for the observed distribution of discrepancies (as it should).

The principal approach to validation of estimated uncertainties within SST CCI is to examine the distribution of satellite-validation SST differences (“discrepancies”) as a function of the SST CCI uncertainty estimate. For perfectly estimated satellite SST uncertainties compared to validation data with a Gaussian error distribution, the uncertainty validation plot is as Figure 4-1.

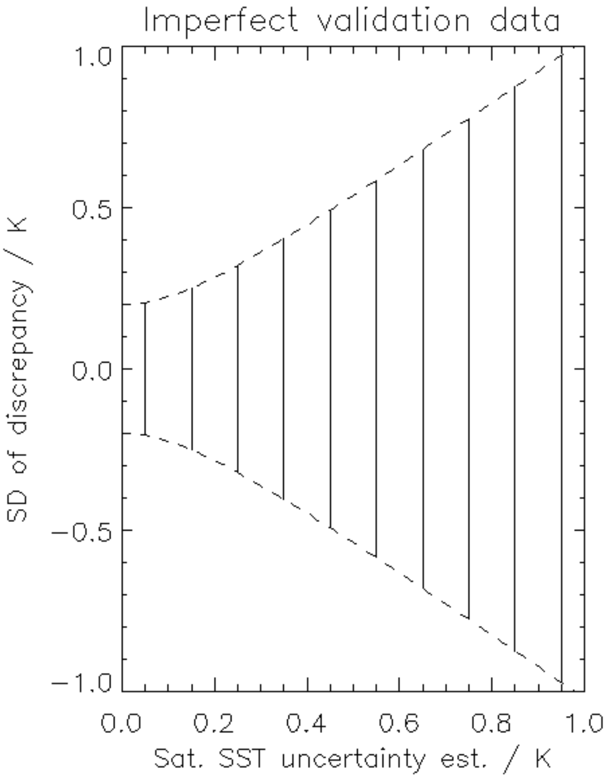

Figure 4-1 Idealised Uncertainty Validation plot, assuming validation against data with Gaussian errors with a standard deviation of 0.2 K. Vertical lines span -1 to +1 standard deviation of discrepancy, for data binned into 0.1 K bins of estimated satellite SST uncertainty. When the satellite SST uncertainty is small, the SD of discrepancy is dominated by the in situ uncertainty. For large satellite SST uncertainty, the SD of discrepancy approaches the estimated uncertainty. The Dotted line gives the locus of the results if the satellite SST uncertainty is perfectly estimated. Deviations from the dotted line indicate biases in uncertainty estimation.

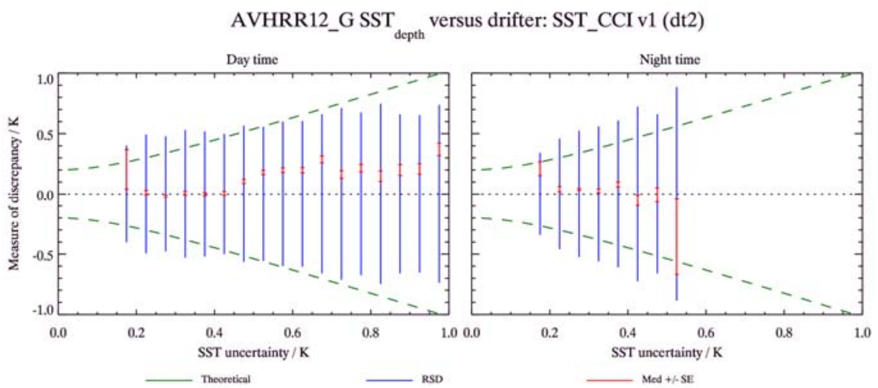

Figure 4-2 Example SST CCI L2P AVHRR validation plot for uncertainty estimates. In addition to the vertical bars for binned discrepancy, the median +/- 1 robust standard error in each uncertainty bin is also shown in red. Where this range is markedly offset from zero, it indicates statistically detectable relative bias in SST within the uncertainty bin. Ideally, the bias should be close to zero independently of the uncertainty estimate, although some dependence is expected in practice. Left panel: day time. Right panel: night-time.

RD.325 gives the uncertainty validation plots found for v1.0 SST CCI products (AVHRR, ATSR and analysis). An example is presented in Figure 4-2. Considering the day time results (left panel) we can see that SSTs with lower uncertainty estimates (around 0.2 K) appear to be over-estimated, whereas cases where the uncertainty is estimated at around 0.8 K actually have higher uncertainty. Accordingly, there is less discrimination of relatively low and high uncertainty cases than the ideal. Nevertheless, the uncertainty estimation is providing useful information: there is a general increase in uncertainty as validated as estimated uncertainty increases; and the low-uncertainty cases (which are the majority) have bias statistically indistinguishable from zero, whereas those with higher uncertainty estimates are more prone to relative bias.

In conclusion, this illustrates that there is some bias in the estimation of uncertainty for SST CCI AVHRR products (the same general message applies to the other AVHRRs). Nonetheless, the uncertainty estimates have the right order of magnitude and a useful (although less than ideal) degree of discrimination of high and low uncertainty cases. Improving the uncertainty model will require further research and development in much the same way that improving SST retrievals does. Similar results were found for SST CCI ATSR L3U products (see RD.325).

This approach has also been applied to validation of the uncertainty estimates associated with the SST CCI analysis product (L4). Unlike the L2P and L3U uncertainty estimates, the analysis uncertainties are not informed by the satellite retrieval context. Instead, as explained below in section 5.3, it is based on a parameterisation that accounts for the degree to which an estimate is driven by observations compared to the analysis background field. As seen Figure 4-3, the resulting estimates appear to be accurate for uncertainty estimates up to 0.5 K, and somewhat underestimated for the less certain data. The degree of discrimination between more and less certain data is excellent.

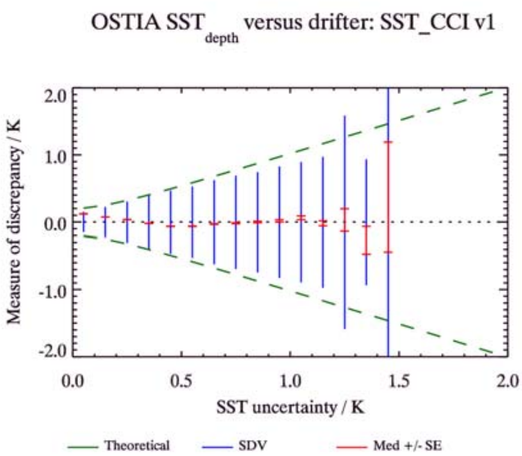

Figure 4-3 Uncertainty validation plot for SST CCI analysis.

Users of SST CCI products are encouraged to use the uncertainty estimates in products, bearing in mind that future product releases will include improvements in uncertainty modelling.

4.5 Other Uncertainty Information

Certain types of uncertainty information are not appropriate to attach to individual SSTs.

The assessed stability of a climate record is the main example of such additional uncertainty information that will be required by some users of SST. This will be available in documentation on the SST_cci products made available to users.

5. GUIDELINES FOR USE OF SST_CCI PRODUCTS ACCOUNTING FOR THEIR UNCERTAINTIES

5.1 Level 2 Products

The SST_cci L2P products contain uncertainties attached to individual SSTs. The motivation for this is that SSTs differ in their uncertainty, even within a single product for a single sensor. SST_cci uncertainty algorithms will model this variation in uncertainty. This gives users of the products the option to weight different SSTs according to their attached uncertainties, if appropriate to their application. Examples of other situations in which the attached uncertainty information should assist users are:

- when propagating SST uncertainties into derived quantities
- when combining SSTs of different origins
- when assessing the significance of differences between SSTs
- in coupled ocean-atmosphere data assimilation
- in SST analysis and re-analysis

Why do different L2P SSTs have different uncertainties? An example is that the uncorrelated uncertainty arising from radiometric noise is expected to be greater near the edge of an AVHRR swath than near the centre. This is because the weights given to the BTs in the inversion process tend to be larger for high satellite zenith angles, and thus the radiometric noise is amplified more in the retrieved SST.

5.1.1 Worked examples of use of uncertainty estimates

Example 1: Is the SST from the L2P product significantly different from a matched, independent buoy measurement?

To compare an L2P SST with a buoy SST, it is appropriate to use the SST-20cm estimate ("sea_surface_temperature_depth", $\mathrm{x_{L2P}}$) and its estimated total uncertainty ("sst_depth_total_uncertainty", $\varepsilon_{\mathrm{L2P}}$). The total uncertainty is a "standard uncertainty" (RD.191, i.e., is an estimate of the standard deviation of the error distribution) from all sources combined. If comparing with a buoy measurement, $x_{buoy} \pm \varepsilon_{buoy}$, one can interpret the significance of the difference, $\mathrm{x_{L2P}} - \mathrm{x_{buoy}}$, compared to the combined measurement uncertainty in the difference, $\sqrt{\varepsilon_{\mathrm{L2P}}^2 + \varepsilon_{\mathrm{buoy}}^2}$, using an appropriate statistical technique, such as a t-test.

Example 2: What is the uncertainty in my calculation of outgoing infra-red IR flux from the sea surface arising from the measurement uncertainty in the L2P SST I am using?

To calculate the thermal emission from the sea surface, one uses an equation along the lines of $\mathrm{E} = \varepsilon\sigma\mathrm{T}^4$, where the appropriate temperature, *T,* in CCI products would be the skin SST ("sea_surface_temperature"), where ε is an estimate of broad spectrum emissivity, and σ is the Stefan-Boltzmann constant. The standard uncertainty in *T* is in the field "sses_standard_deviation", *s*, and may be smaller than the uncertainty in the SST depth uncertainty because there is no contributing uncertainty from skin-to-depth adjustment present in the skin SST measurement. According to the usual method of propagating errors, the uncertainty *E* arising from the SST skin uncertainty is $4\varepsilon\sigma\mathrm{T}^3\mathrm{s}$. (Of course, this is not the only contributing uncertainty.)

5.2 Level 3 Products

The SST_cci L3 products will contain uncertainties attached to 0.05° cell SSTs. The motivation for this is that SSTs differ in their uncertainty, even within a single product for a single sensor. SST_cci uncertainty algorithms will model this variation in uncertainty. This gives users of the products the option to weight different SSTs according to their attached uncertainties, if appropriate to their application. Examples of other situations in which the attached uncertainty information should assist users are:

- when propagating SST uncertainties into derived quantities
- when combining SSTs of different origins
- when assessing the significance of differences between SSTs
- in coupled ocean-atmosphere data assimilation
- in SST analysis and re-analysis

Why do different L3C SSTs have different uncertainties? An example is the different sampling uncertainty between two hypothetical adjacent 0.05° cells in AATSR L3C. One cell may derive from a single clear AATSR pixel, the other may be wholly clear sky and based on >25 pixels. In the latter case the sampling uncertainty is zero (complete coverage) whereas in the first cell it is equal to the standard deviation of SST observed at this resolution across the cell. Being unobserved, that standard deviation is not known, but indirect means allow estimation of sampling effects on cells on an aggregate basis.

5.2.1 Worked examples of use of uncertainty estimates

The Examples 1 and 2 given in §5.1.1 would apply similarly to L3U SSTs. Here, two further examples are given, whose principles can also be transferred to L2P SSTs.

Example 3. What is the best estimate of the average depth SST over an area of 0.25° latitude by 0.25° longitude from the L3U product?

The L3U product is on a grid of 0.05°×0.05°, so the task is to estimate the average depth SST over a cell that includes up to 25 (5×5) SSTs from the product. Each of the 25 SSTs has its own set of uncertainty estimates. The basic choice in forming the average is whether (a) to form a simple mean, or (b) to form a weighted mean reflecting the differing uncertainties. One can argue for (a) if one expects there to be significant true SST variability across the 0.25° cell, which one wants to be evenly represented in the average. One can argue for (b) if one wants the minimum uncertainty in the average SST across an area where the SST variability is negligible (so that the 25 SST can be viewed as repeated measurements of the same SST). Let's assume the choice is to form a weighted mean. The usual expression for the best estimate of the average is $\bar{x} = (\sum \varepsilon_i^{-2} x_i)/(\sum \varepsilon_i^{-2})$, where *i* is an index running over the 25 contributing SSTs, *x* is skin SST and ε is SST uncertainty. This weights the most certain measurements of SST most highly when creating the average across the cell.

The only question remaining is where to source the value for the skin SST uncertainty. To be strictly correct, one should not use the total uncertainty for this. The reason is that this total uncertainty includes the effects of two error components that are highly correlated between the 25 SST values. One component is correlated over large space scales and long time scales, and would include, for example, uncertainty in sensor calibration which changes slowly. A second component is that related to the imperfection of the retrieval process, which is correlated over the length and time scales of the atmosphere. Since we are averaging 25 near-simultaneous SSTs that represent locations no more than 35 km apart, it is reasonable to assume that errors from synoptically correlated effects are nearly perfectly correlated and therefore should not influence the weight given to any particular SST. The best uncertainty estimate for this purpose therefore accounts for the errors that are random between the different SSTs (from instrument noise and from subsampling within each 0.05° cell if there is partial cloud cover). The weights therefore should derive from the field in the L3U product named "uncorrelated uncertainty".

Example 4. What is the total uncertainty in the averaged SST found in Example 3?

The random uncertainty in $\bar{x}$ from the uncorrelated errors in the 25 SSTs being averaged is given by $(\sum \varepsilon_i^{-2})^{-1/2}$, which reduces to the more familiar "$\varepsilon/n^{-1/2}$" reduction in random error if all the cells have equal values of the uncertainty from uncorrelated effects. The total uncertainty, however, includes the additional uncertainty from synoptically correlated and large-scale correlated effects. To estimate these effects, their average value across the cell is sufficient (they differ little, if at all, between the 25 SSTs). Letting ε_{synop} be the average of the field "synoptically_correlated_uncertainty" and ε_{large} be the average of the field "large_scale_correlated_uncertainty". The estimate of total standard uncertainty in the weighted average is then: $\sqrt{(\sum \varepsilon_i^{-2})^{-1} + \varepsilon_{synop}^2 + \varepsilon_{large}^2}$.

5.3 Level 4 Products

The Level 4 files contain an estimate of uncertainty for each analysis value. The OSTIA system uses an analysis quality optimal interpolation approach (Donlon et al., 2012) to estimate this uncertainty.

The first step is to generate a field that describes the degree to which a particular SST value has been contributed by observations rather than the background field. To to this, an optimal interpolation analysis is performed using the same methods as the main SST analysis (Donlon et al., 2012) but with the following differences:

- All observations are given a value of 1
- The background field is set to 0

As with the SST analysis, the uncertainty estimates used in the main SST analysis (which comprise of background and observational uncertainty) are preserved in this analysis, which means that the weight given to the observations in the SST analysis is preserved in this optimal interpolation of ones and zeroes. (The observation uncertainty used is the total uncertainty field provided in the L2 and L3 products.)

The result of the uncertainty analysis is a field on the analysis grid, ε^o, which contains values between zero and one. If the value is closer to one, this implies that the analysis has been heavily influenced by the observations. If the value is nearer zero then this indicates that the observations have input little to the current analysis. Combining ε^o with the background error variance estimates, the formula

$$\varepsilon_i^a = \sqrt{B_i[\alpha + \beta(1 - \varepsilon_i^0)]}$$

is used at grid point i to produce an analysis uncertainty estimate ε^a . Here, α is used to control the smaller errors expected from the analysis as a proportion of the total background error variances B_i and is equal to 0.5, β is used to control the larger errors expected as a proportion of the background errors and is equal to 4. The L4 uncertainty estimates therefore range from a value of √0.5 times the background error standard deviation (when a new observation is given full weight in the analysis) to a value of √4.5 times the background error standard deviation (when there have not been any observations for a long time). As the background error standard deviation can be very small in some locations, a minimum uncertainty value of 0.1 K is also imposed.

No decomposition of uncertainty into components with different correlation structures is available in the L4 product, since it is not at present known how to achieve this. A simple "1/√n" treatment of uncertainty when averaging or otherwise combining SSTs within the L4 product is likely to give highly optimistic estimates of uncertainty.

5.4 Estimating trend uncertainty

Experience in ARC suggests that stability of observation can be assessed with relatively tight confidence intervals for ARC SSTs across the tropical Pacific Ocean. For this, the tropical moored buoy array is used as a reference that is assumed to be stable. For ARC, tropical SSTs over the period 1993 to 2009 are very likely ($p = 0.05$) to be stable to 0.003 K/yr. By referencing the SST_cci products to ARC SSTs (with additional improvements such as more consistent auxiliary information for cloud detection), the aim is to reach a similar decadal stability for both ATSR and AVHRR data. This is a challenge, since AVHRR SSTs are likely to be less stable, because the instruments are generally less stable than ATSRs.

Secure estimates of decadal stability in the extra-tropics are harder to establish, because ocean weather station moorings are less tightly specified than the tropical arrays in terms of temperature accuracy. There is no reason to think that stability in the extra-tropics will be better than in the tropics. Cloud detection problems are known to affect the North Pacific in particular in the ARC data set, and could adversely affect higher latitude stability.

When evaluating climate change trends from the SST_cci products, the stated stability of observation should be accounted for within trend uncertainties, in addition to the trend uncertainty statistic that arises merely from fitting to a finite set of points that display monthly-to-decadal variability.

Stability assessment is reported in the Climate Assessment Report (RD.326). This analysis is summarised below, and uses SST data from moorings which are known to be calibrated before and after deployment. The PVIR (RD.325) also has results tracking the relative change of the SST CCI datasets against drifting buoys, but these are not a focus for stability assessment, because the stability of calibration of the drifting buoy network is not assured.

The methodology for stability assessment is based on the GHRSST Climate Data Assessment Framework (Merchant et al., 2013, RD.317). Briefly, the three SST_CCI datasets (L2P AVHRR, L3U ATSR and L4 analysis) were individually matched to Global Tropical Moored Buoy Array (GTMBA) data for the full time period (1991 – 2010). The high-temporal resolution GTMBA data had a sampling resolution of either 5, 10 or 60 minutes and the highest available time-resolution was always used if multiple resolutions were available. The GTMBA data were matched to the nearest SST_CCI pixel centre and a maximum time difference of 30 minutes was used as a threshold. For the SST_CCI L4 analysis the GTMBA data was a mean of the nearest measurements to 10:30 AM and 10:30 PM as the L4 is a daily mean. No further quality control or filtering was applied to the data prior to analysis.

Following the initial match-up process the monthly median SST_CCI-GTMBA difference for each GTMBA location was calculated. Then for each month of the year and location, the multi-year average of the monthly median SST_CCI-GTMBA differences was calculated. For each month the data were then deseasonalised by subtracting the multi-year average for the appropriate month of the year from each month of the time series. For the L2P AVHRR and L3U ATSR datasets separate multi-year averages were used for day and night. The data were deseasonalised to minimise any potential aliasing of any annual cycle in residual time series following the approach of Merchant et al. (2012, RD.296). Retaining at this point only locations where buoy data were available for > 15 years within the 1991-2010 period, the monthly mean difference across all locations was determined to end up with a single SST_CCI-GTMBA SST time series for each SST_CCI dataset (and for day and night for L2P AVHRR and L3U ATSR). A least squares linear fit to each time series of monthly mean differences was calculated and 95% confidence intervals were determined.

The results from the stability assessment are shown in Figure 5-1. A step-change is apparent from 1995 onwards, which is most likely due to the change between ATSR-1 and ATSR-2. As the ATSRs were used to bias correct the radiances for the AVHRRs the feature is also apparent in the L2P AVHRR time series and the SST_CCI L4 analysis system. Consequently, the 95% confidence interval on the slope of the fit was calculated for two separate periods, 1991 to May 1995 covering the ATSR-1 period and June 1995 to 2010 covering the ATSR-2/AATSR period.

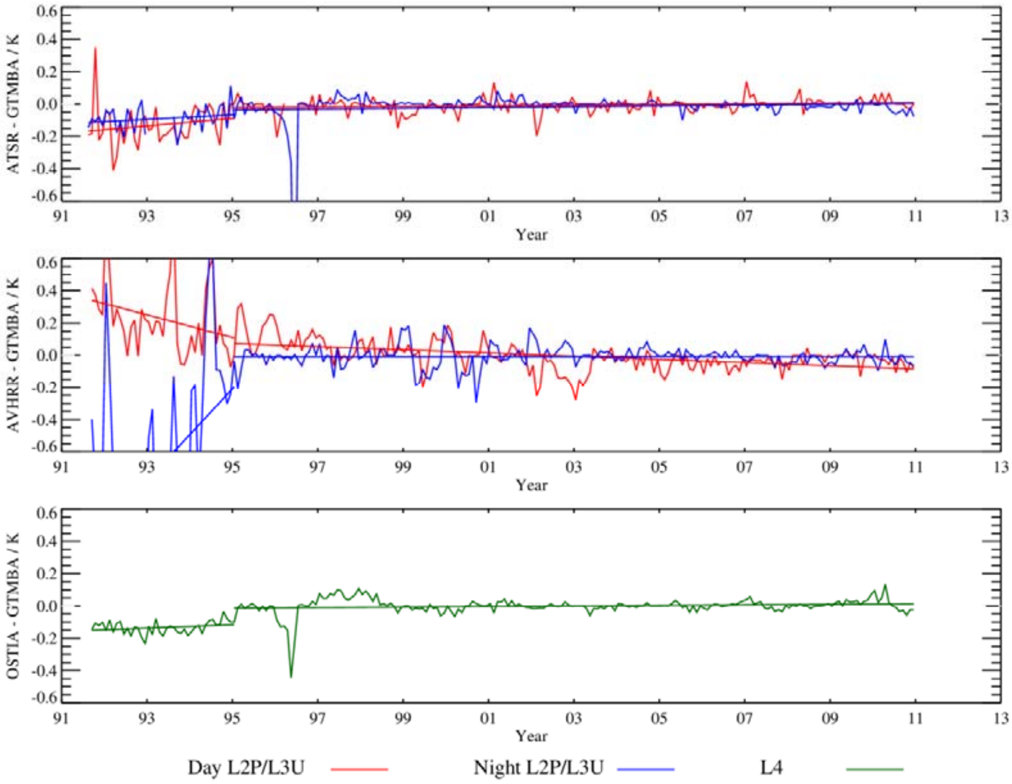

Figure 5-1: Time series of deseasonalised composite monthly mean differences (K) between the SST_CCI products and the GTMBA. Separate day and night time series are provided for the L2P AVHRR and L3U ATSR datasets. Also, plotted are the results of a least squares linear fit for the 1991 to May 1995 and June 1995 to 2010 periods (see text for further discussion).

The resulting confidence intervals for the least squares linear fits to the time series of SST_CCI-GTMBA differences are summarised in Table 5-1.

SST_CCI 95% confidence interval (mK $year^{-1}$) for 1991 - 1995			
	Day	Night	Both
L2P AVHRR	-137.9 < trend < -2.4	105.9 < trend < 462.3	
L3U ATSR	-13.6 < trend < 60.1	-7.4 < trend < 36.8	
L4 analysis			-1.8 < trend < 22.1
SST_CCI 95% confidence interval (mK $year^{-1}$) for 1995 – 2010			
	Day	Night	Both
L2P AVHRR	-12.3 < trend < -7.4	-2.0 < trend < 2.0	
L3U ATSR	0.7 < trend < 3.2	-1.4 < trend < 6.4	
L4 analysis			0.1 < trend < 3.2

Table 5-1: Summary of 95% confidence intervals for least squares linear fits to SST_CCI-GTMBA monthly mean difference time series for 1991 to May 1995 and June 1995 to 2010.

For the SST_CCI L3U ATSR product, the night time trend in the differences to the GTMBA measurements for the 1995-2010 period is comparable to that calculated by Merchant et al. (2012, RD.296). However, the day time stability confidence interval doesn't include zero, and relative to RD.296 is somewhat less stable; nonetheless, the true stability is still likely to be within the GCOS requirement. For the ATSR-1 period, both the day and night trends calculated for the SST_CCI L3U ATSR product have improved stability (based on the most likely relative trend) compared to that reported in Merchant et al. (2012, RD.296), although there is nonetheless likely to be a positive trend artefact that is outside the GCOS target.

Regarding the SST_CCI L2P AVHRR product there is no comparable analysis in the literature for pre-cursor datasets. We note that, as for the L3U ATSR product, the day time stability is poorer than for night time. This may reflect the greater amplification of error in two-channel relative to three-channel SST retrieval that is common to all IR sensors and retrieval methods. The L2P time series in Figure 5-1 appears to have a step improvement in inter-annual stability from 2003/4. The cause of this is not clear, as this does not correspond to a changeover between sensors within the time series.

Being tied to the L3U calibration, the SST CCI analysis product has a stability over the period 1995 to 2010 that likely meets the GCOS requirement (the confidence interval is mostly within the interval -3 to +3 mK/yr).

Users calculating trends from SST CCI products should take account of the above stability information (and the further information in the CAR) when interpreting their results. While the above calculations have only been able to be calculated for the tropical Pacific, the results are the best available information about the trend uncertainty for the datasets as a whole. Thus, if a trend over the period 1995 to 2010 for a particular location were found in the SST CCI analysis (by statistical fitting to a deasonalised/anomaly timeseries) to be 0.03 +/- 0.01 K/yr, this trend, on the basis of the available information, is unlikely to be an artefact of the dataset, since this is an order of magnitude larger than the upper 95% confidence interval reported in Figure 5-1. Confidence in this conclusion will be greater for a tropical location, given the geographical limitation of the possible stability assessment.