GHRSST compliant AVHRR SST products over the Australian region

Christopher Griffin, Helen Beggs, Leon Majewski

August 3, 2017, version 1, overview

Abstract

The Australian Bureau of Meteorology produces a Group for High Resolution Sea Surface Temperature (GHRSST) compliant Sea Surface Temperature (SST) dataset over the Australian region, based on direct measurements from National Oceanic and Atmospheric Administration (NOAA) Polar Orbiting Environmental Satellites (POES) Advanced very High Resolution Radiometer (AVHRR) instruments. This document discusses features, function, performance and operational details of this system.

Contents

1 Product Matrix 4

1.1 GHRSST 2.0r5 Compliant Product types . . . 4
1.2 SST types . . . 13
1.3 Product domains and time coverage . . . 18
1.4 Ancillary Information . . . 18
1.5 Product versions . . . 24

2 Estimating SSES for GHRSST compliant L2 product 24

2.1 Determination of quality level . . . 27
2.2 SSES for L2P Class product - a look-up table approach - fv01 . . . 27
2.2.1 Binning based on *in situ* measurements . . . 28
2.2.2 Validation of SSES - a lookup table approach - fv01 . . . 31
2.3 SSES for L2P Class product - an empirical model approach - fv02 . . . 31
2.3.1 Model fitting performance . . . 44
2.3.2 Validation of SSES - an empirical model approach - fv02 . . . 46
2.4 Comparing fv01 to fv02 . . . 46

3 Quality control of matches between *in situ* and satellite SST 51

3.1 Selection criteria for favourable conditions for *in situ* to satellite SST matches . . . 51
3.2 Selection of *in situ* data sources . . . 52
3.3 LatQC: Quality control of *in situ* Temperature measurements . . . 53
3.4 L4QC: Quality control of *in situ* measurements using level 4 analysis SST . . . 59
3.5 2 channel SST regression quality control. . . . 66
3.6 Data sets for fitting and quality control . . . 68
3.7 L3U class product and the computation of L3U from L2P product . . . 70
3.8 L3C class product and the computation of L3C from L3U product . . . 74
3.9 L3S class product and the computation of L3S from L3C product . . . 77

4 Application of SSES 79

4.1 A reassessment of quality using SSES . . . 79

5 Computation of retrieval sensitivity 81

5.1 Computation of retrieval sensitivity from *in situ* measurements . . . 81

6 Interpretation and use of l2p_flags 82

7 SSES model fitting 82

7.1 SSES determination - fv02 - detailed procedure . . . 91
7.2 Performance of the SSES model . . . 92

8 SST model fitting 94

8.1 Verification data set . . . 94
8.2 Fit data set . . . 94
8.3 Model types . . . 94
8.4 Model Regression . . . 96
8.5 Model performance . . . 97

8.6 Hybrid Day and Night models. 97
8.6.1 Correcting brightness temperatures for satellite view angle 97
8.7 Fixed retrieval regression - fv01 98
8.8 Progressive retrieval regression - fv02 100
8.9 NOAA-11 performance 103
8.10 NOAA-12 performance 103
8.11 NOAA-14 performance 103
8.12 NOAA-15 performance 103
8.13 NOAA-16 performance 103
8.14 NOAA-17 performance 103
8.15 NOAA-18 performance 103
8.16 NOAA-19 performance 103

9 GHRSST Compliant file format. 103
9.1 File naming convention 120
9.2 File field description 120
9.3 Issue description 123

A Use of SSES for quality assessments 125
A.1 L3U class product and the computation of L3U from L2P product 125
A.2 L3C class product and the computation of L3C from L3U product 130
A.3 L3S class product and the computation of L3S from L3C product 133
A.4 A new definition of quality level from SSES 135
A.4.1 Standard Deviation component to quality 137
A.4.2 Bias component to quality 138
A.4.3 Overall quality assessment from Bias and Standard Deviation 138
A.4.4 Determination of characteristic parameters 139
A.4.5 Using quality reassessments to provide best quality merges from multiple sources145
A.4.6 Generalizing to include other quality factors in the assessment 147
A.4.7 Generalizing for measurement distributions that are expected to be asymmetric and/or fat tailed 149

1 Product Matrix

The Australian Bureau of Meteorology has a reasonably complete archive of Advanced very High Resolution Radiometer (AVHRR) direct receptions over multiple National Oceanic and Atmospheric Administration (NOAA) satellite missions, from the 1990's to the present. The Group for High Resolution Sea Surface Temperature (GHRSST) provides products for interested researchers and users. These products are intended to record historical Sea Surface Temperature (SST) measurements, as well as reanalysis products, as a contribution to the Integrated Marine Observing System (IMOS, www.imos.org.au). We have produced four types of satellite measurement, at the resolution of the sensor or at 0.02 degree, over the Australian region, from multiple satellite platforms. Our products contain ancillary information and estimates of biases and uncertainties based on measurement correlations with *in situ* devices. All of the products are compliant with GHRSST 2.0r5 network Common Data Format (netCDF) file format.

1.1 GHRSST 2.0r5 Compliant Product types

There are four major types of AVHRR data for which we produce GHRSST Data Specification (GDS) version 2.0r5[29] compliant data records.

SST measurements, Sensor Specific Error Statistics (SSES), quality, ancillary and other pixel information are provided for each valid pixel in each image file as follows,

$$\{T_{\text{satellite}}, t, q, \mu, \sigma, n, \texttt{ancillary}, f_{\text{L2p}}\} \tag{1}$$

where $T_{\text{satellite}}$ is the retrieved sea surface temperature measurement, t is the time of the measurement, q is the quality level of the data (an integer in the range 0 to 5) based on an estimate of proximity to cloud or other forms of atmospheric interference, μ is an estimate of the bias in $T_{\text{satellite}}$ by regression against *in situ* measurements, σ is an estimate of the standard error of $T_{\text{satellite}}$ by regression against *in situ* measurements, and n is the indicative number of degrees of freedom of *in situ* measurements provided in order to estimate μ and σ. `ancillary` represents other ancillary information that is added to the file [29], and f_{L2p} indicates additional information that may further impact the interpretation of the quality or applicability of the data.

In GHRSST 2.0r5 format, these data take on the variable names shown in tables 1 and 2. See table 10 for a description of the ancillary fields, and table 16 for a description of f_{L2p} masks and bits. For further detailed description of these products and how to apply the SSES and their use, see sections 2 and A.1. Each of the four major types of data has a corresponding GHRSST product class, as listed below:

GHRSST product L2P Geolocated, cloud-cleared, ungridded AVHRR SST swaths.

These are direct translations from raw satellite image data, received from Australian Direct Reception stations, (see figure 9 for indicative locations of these stations at the time of writing), and merged into continuous swaths representing each satellite pass. SST is retrieved from brightness temperatures, after cloud detection and removal, bad line and bad pixel removal and navigation are performed. Each pixel has its own latitude and longitude based on the navigational corrections to the expected satellite orbit.

Raw brightness temperature measurements from the AVHRR sensor are used to determine SST, using an algebraic function that includes satellite zenith angle. SSES are then estimated based on matching the determined AVHRR SST with *in situ* buoy SST. Providing errors of

Parameter name	Symbol	L2P		L3U	
`sses_count`	n	n_P	Indicative of the number of *in situ* measurements made under similar viewing conditions	n_U	Indicative number of best quality L2P pixels merged when converted to a fixed grid.
`sses_bias`	μ	μ_P	Indicative median bias for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing conditions.	μ_U	Indicative gridded median bias for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing and merging conditions.
`sses_standard_deviation`	σ	σ_P	Indicative standard deviation for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing conditions.	σ_U	Indicative gridded standard deviation for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing and merging conditions.
`l2p_flags`	f_{L2p}	$f_{\text{L2p},P}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions.	$f_{\text{L2p},U}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location.
`quality_level`	q	q_P	Quality level as a measure of proximity to detected cloud in kilometres. $q = 0$ is also used to indicate invalid data for other reasons.	q_U	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.
`sea_surface_temperature`	$T_{\text{satellite}}$	$T_{\text{satellite},P}$	Retrieved sea surface temperature	$T_{\text{satellite},U}$	An average of the retrieved sea surface temperature based on all of the measurements contributing to the gridded location.

Table 1: Association between field names in GHRSST compliant files and symbols used in this text, with a short description of the intent of the parameter and symbol, for L2P and L3U files.

Parameter name	Symbol	L3C		L3S	
`sses_count`	n	n_C	Indicative number of good quality L3U measurements merged to L3C	n_S	Indicative number of good quality L3U measurements merged to L3S
`sses_bias`	μ	μ_C	An estimate of the median bias of the platform and sensor over the time window of the L3C file.	μ_S	An estimate of the median bias over all measurements over the time window of consideration.
`sses_standard_deviation`	σ	σ_C	Indicative uncertainty over the time window, including contributions from natural variation as they affect the estimate of the mean SST.	σ_S	Indicative uncertainty over the time window, including contributions from natural variation as they affect the estimate of the mean SST.
`sst_count`	n_w	$n_{w,C}$	Number of measurements merged to L3C.	$n_{w,S}$	Number of measurements merged to L3S.
`sst_mean`	T_w	$T_{w,C}$	Unweighted mean measured sea surface temperature.	$T_{w,S}$	Unweighted mean measured sea surface temperature.
`sst_standard_deviation`	σ_w	$\sigma_{w,C}$	Unweighted standard deviation of measured sea surface temperature.	$\sigma_{w,S}$	Unweighted standard deviation of measured sea surface temperature.
`l2p_flags`	f_{L2p}	$f_{\mathrm{L2p},C}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location over the time window.	$f_{\mathrm{L2p},S}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location over the time window.
`quality_level`	q	q_C	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.	q_S	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.
`sea_surface_temperature`	$T_{\mathrm{satellite}}$	$T_{\mathrm{satellite},C}$	Estimate of the sea surface temperature characteristic of the time window.	$T_{\mathrm{satellite},S}$	Estimate of the sea surface temperature characteristic of the time window.

Table 2: Association between field names in GHRSST compliant files and symbols used in this text, with a short description of the intent of the parameter and symbol, for L3C and L3S.

the bias and uncertainty of the AVHRR SST at the time and place of measurement. This is the highest resolution product (∼ 1km at nadir and ∼ 4km at the edge of swath) with the least post-processing. SST and SSES from a typical L2P file are shown in figure 1.

The cloud clearing algorithm is based on a variant of Cloud Advanced Very High Resolution Radiometer Extended (CLAVRX) algorithm[25].

The SST measurements are intended to be representative of the time of observation in the coordinate system of observation, thus SSES represent bias and standard deviation associated with the instantaneous state of the sensor measurements when compared to *in situ* measurements. L2P files are assigned a `file_quality_level` as outlined in table 3, depending on the mission and instrument status at the time of measurement, from 0 (worst) to 3 (best) based on the mission status as reported on the National Oceanic and Atmospheric Administration Polar Operational Environmental Satellites (NOAA POES) status page[18] and internal Australian Bureau of Meteorology reception quality records, on a platform and date basis.

If `file_quality_level` is not 3, the `issue` entry in the `history` metadata will contain additional information that will allow the source of the potential issue to be determined.

GHRSST product L3U Gridded best quality L2P AVHRR product.

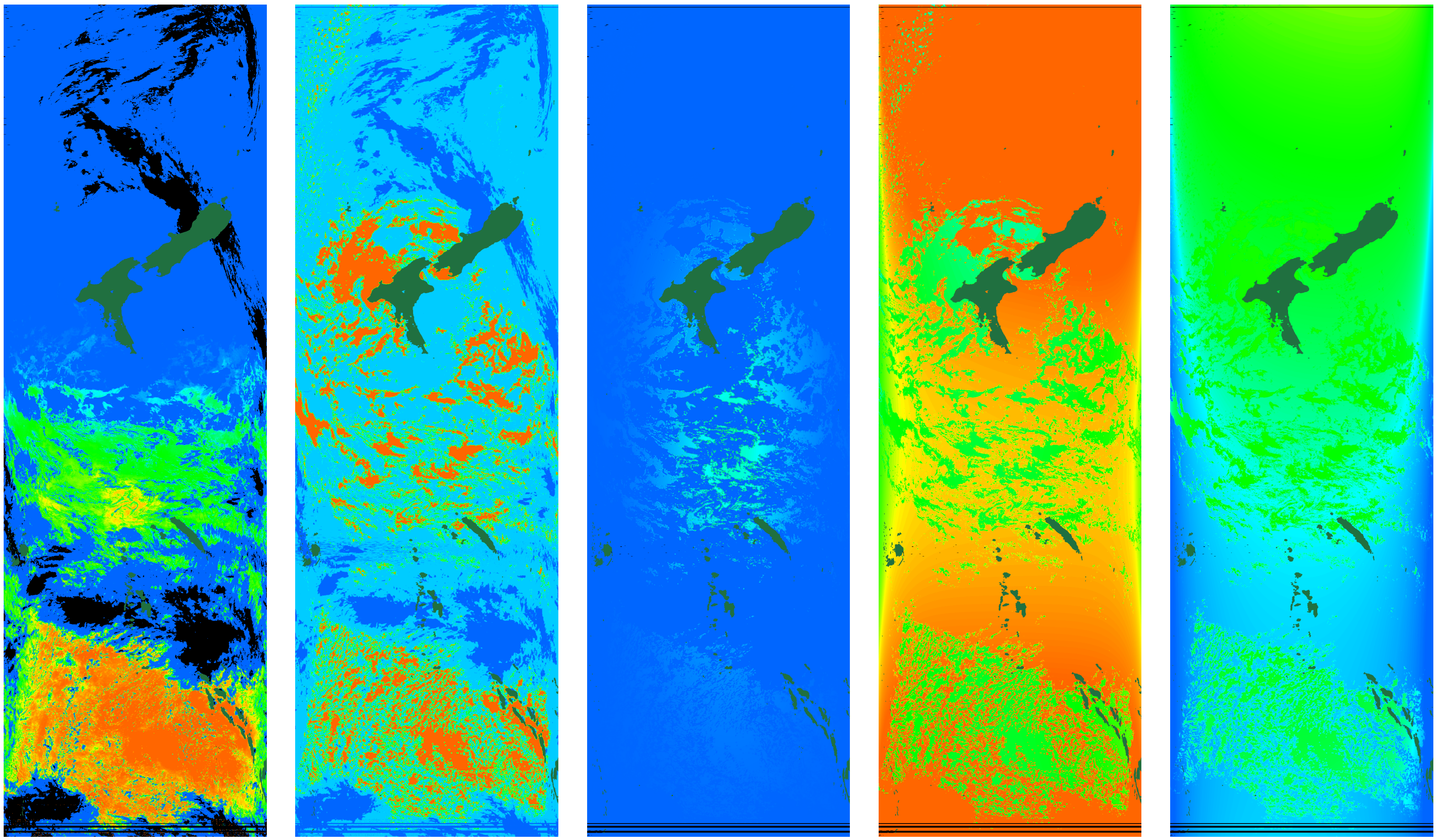
The gridded product is provided for convenience and ease of use, since the grid is standardized and uniform in latitude and longitude. The grid used for this product, equally spaced by 0.02 °, while uniform in a cylindrical equidistant coordinate system, is coarser than the resolution of L2P products. Thus, best quality measurements from multiple L2P pixels are merged into a single L3U pixel, or spread over multiple L3U pixels, as required. Using a common grid layout for all satellite passes makes processing and use simpler. SSES are derived from L2P sensor specific error statistics and intended to represent the bias and uncertainty of the sensor at the time and place of measurement, in much the same way as the L2P product. There is one L3U file produced for every L2P file. These represent the highest resolution product on a fixed grid, with the least post-processing. SST and SSES from a typical L3U file are shown in figure 2. L3U `file_quality_level` is inherited from L2P `file_quality_level`, and reduced if the L2P to L3U processing experiences any issues.

GHRSST product L3C Single sensor/platform, fixed time period composites.

Single AVHRR sensor composites of L3U files over day or night periods for one or more days. These data follow the same grid layout as the L3U files, but typically provide a higher degree of coverage since they consist of multiple passes of the same satellite. SSES are derived from the L3U statistics and are intended to represent the typical deviation of the stated SST value from the *in situ* value, allowing for the fact that the time scale over which the data are considered valid is no longer instantaneous. There is a one day L3C day file, and a one day L3C night file produced for every satellite for every day, in addition to three day files. SST and SSES from a typical L3C file are shown in figure 3. L3C `file_quality_level` is inherited from the minimum of the L3U `file_quality_level` indications, with further reduction if L3U to L3C processing has issues.

GHRSST product L3S Multiple sensor/platform, fixed time period composites.

L3S files are multiple platform, AVHRR sensor composites of L3C files. These data follow the common grid layout for all L3U files, but merge multiple satellite L3C composites together.



SST, $T_{\text{satellite},P}$

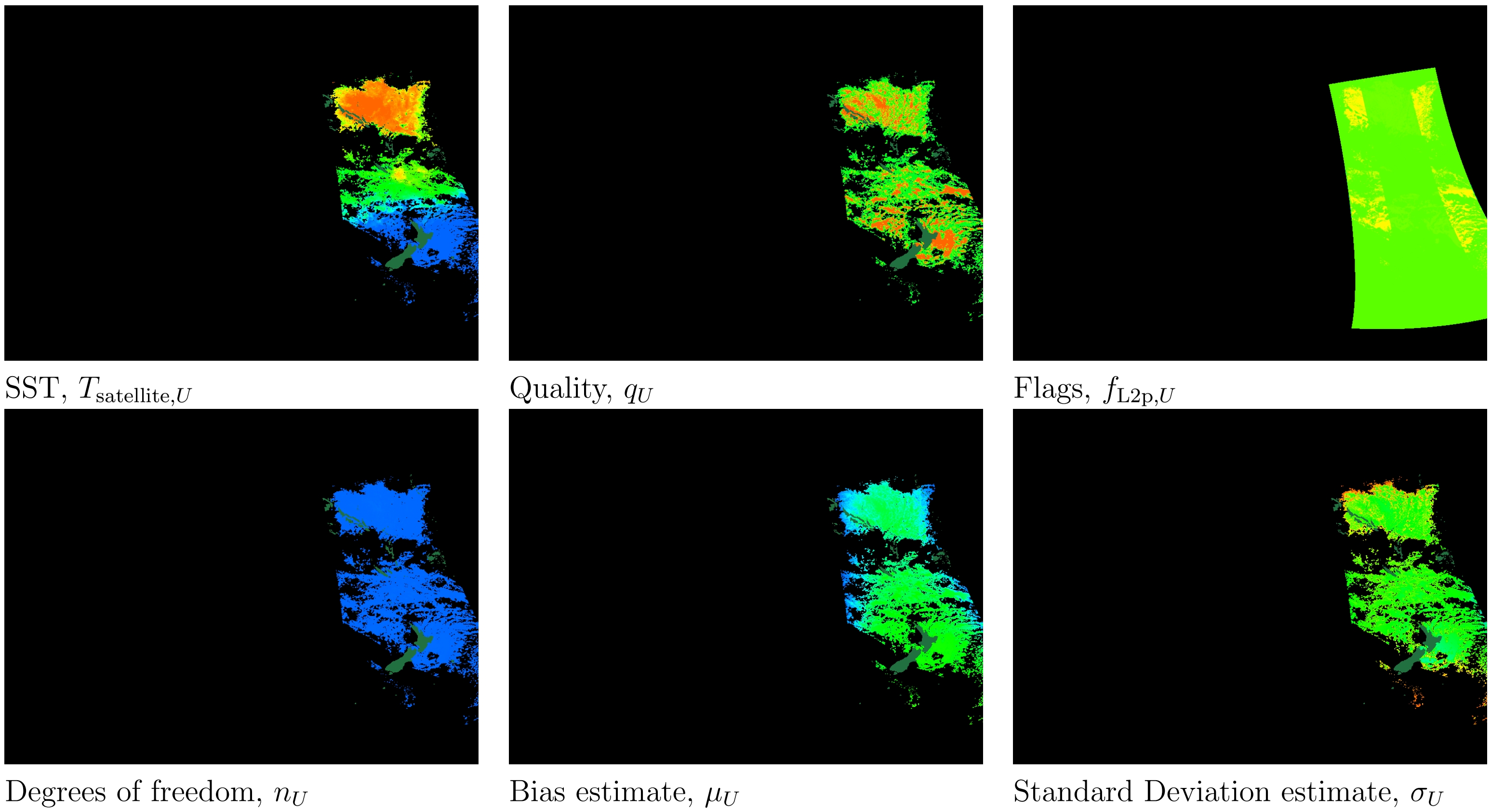
Quality level, q_P

Number of degrees of freedom, n_P

Standard deviation, σ_P

Bias, μ_P

Figure 1: A typical L2P file contains fields for SST ($T_{\text{satellite},P}$), quality level based on proximity to cloud (q_P), SSES number of degrees of freedom (n_P), standard deviation (σ_P) and bias (μ_P), in addition to navigation information and ancillary fields. Note regions where the reception may have either dropped out or was of very poor quality are removed from the data set, making use of fill values (shown in black). The data above is a single ascending pass from NOAA-11 with reception commencing at 03:46, April 1st, 1992 UTC. See table 1 for further information about where to find these parameters within L2P files.



SST, $T_{\mathrm{satellite},U}$

Quality, q_U

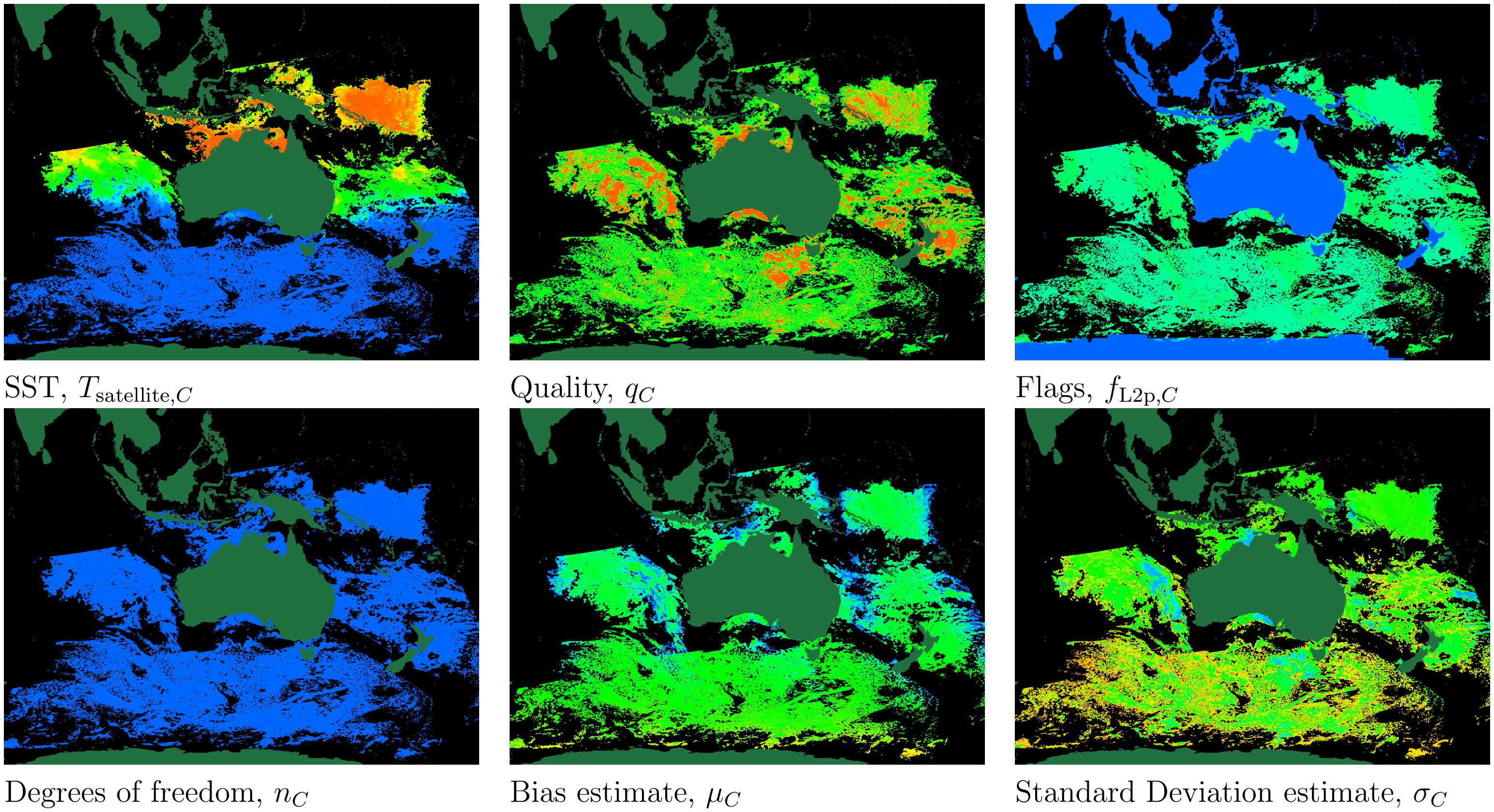
Flags, $f_{\mathrm{L2p},U}$

Degrees of freedom, n_U

Bias estimate, μ_U

Standard Deviation estimate, σ_U

Figure 2: Sample L3U data set over the Australian domain. A typical L3U file contains fields for Sea surface temperature ($T_{\mathrm{satellite},U}$), quality level based on proximity to cloud (q_U), processing flags ($f_{\mathrm{L2p},U}$), SSES number of degrees of freedom (n_U), bias (μ_U) and standard deviation (σ_U), in addition to navigation information and ancillary fields. Note regions where the reception may have either dropped out or was of poor quality (L3U files do not contain pixels with quality less than 2) are removed from the data set, making use of fill values (shown in black in the above images). The data above is a single pass from NOAA-11 with reception commencing on 03:46, April 1st, 1992 UTC. See table 1 for further information about where to find these parameters within L3U files.



SST, $T_{\mathrm{satellite},C}$

Quality, q_C

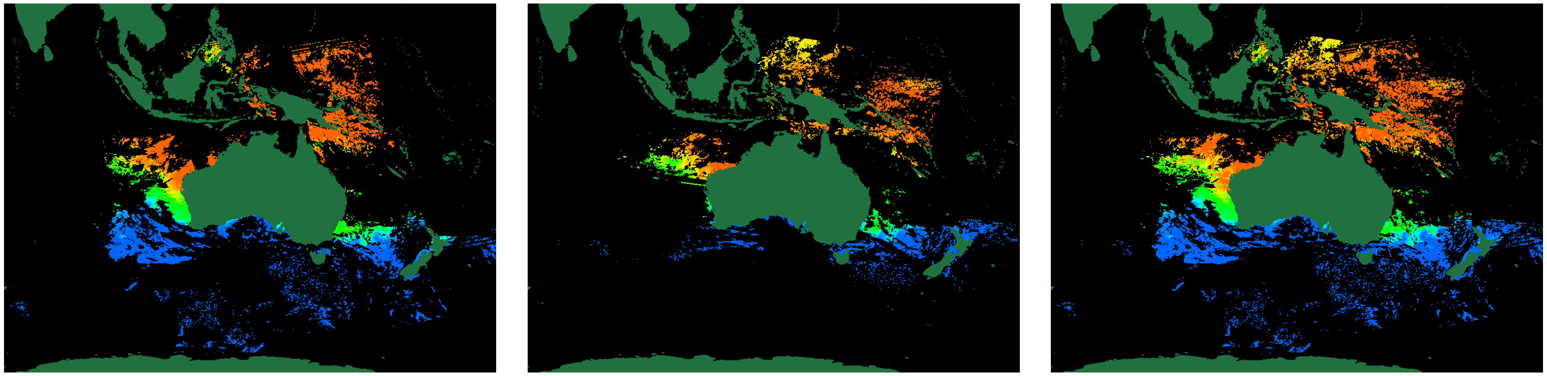
Flags, $f_{\mathrm{L2p},C}$

Degrees of freedom, n_C

Bias estimate, μ_C

Standard Deviation estimate, σ_C

Figure 3: Sample L3C one day, daytime data set over the Australian domain. A typical L3C file contains fields for sea surface temperature ($T_{\mathrm{satellite},C}$), quality level based on proximity to cloud (q_C), processing flags ($f_{\mathrm{L2p},C}$), SSES number of degrees of freedom (included only if there are pixels derived from more than one source), (n_C), standard deviation (σ_C), and bias, (μ_C), in addition to navigation information and ancillary fields. Note regions where the reception may have either dropped out or was of poor quality (L3C files do not contain pixels with quality, $q < 2$) are removed from the data set, making use of fill values (shown in black). The data above is a single day, daytime composite from NOAA-11 with a characteristic time of 03:20, March 31$^{\mathrm{st}}$, 1992 UTC at the central longitude of the image. See table 2 for further information about where to find these parameters within L3C files.



$T_{\text{satellite}}$, day, 03:20 UTC

$T_{\text{satellite}}$, night, 15:20 UTC

$T_{\text{satellite}}$, day and night, 09:20 UTC

Figure 4: Sample L3C data set over the Australian domain, showing three different conditions of composite - a day only composite, a night only composite, and a day and night composite. The data above is a single day, daytime composite from NOAA-11 with a characteristic time of 03:20, March 3$^{\text{rd}}$, 1994 UTC at the central longitude of the image. Night composites have a characteristic time of 15:20 UTC, while day and night composites have a characteristic time of 09:20 UTC. See table 2 for further information about where to find these parameters within L3C files.

`file_quality_level`	Meaning
0	Satellite was not considered functional or able to make any reasonable measurements contributing to the file.
1	Satellite was considered functional, but measurements in the file are suspect due to degraded functionality of one or more of the satellite components, processing functions, or component files, if this is a derived file.
2	The file contains a mixture of `file_quality_level` data. The satellite was considered functional. However, the measurements are a mixture of degraded and reasonable functionality, due to a possible issue in processing. Had this issue not occurred, `file_quality_level` would have been 3.
3	Satellite was considered functional and all the measurements reasonable based on the quality designation and other flags stated in the file. If processing was completed successfully on a satellite that is expected to be of acceptable condition at the time of reception, `file_quality_level` will be 3.

Table 3: Designation of `file_quality_level`. `file_quality_level` is typically determined by the expected quality of the incoming satellite per mission and instrument status at the time of measurement, depending on the status as reported on the NOAA POES status page [18] and Australian Bureau of Meteorology reception quality, on a platform and date basis. The occurrence of issues during processing further reduces the `file_quality_level`. For processing that requires merging multiple files, the resulting `file_quality_level` is the minimum of the `file_quality_level` of the component files, degraded by 1 if the processing had further issues. In our processing framework, `quality_level` and `file_quality_level` are considered independent assessments of quality, with `file_quality_level` used to determine if the entire file has sufficient quality data, from which the `quality_level` is used as a relative measure of this quality.

Daily, there is one L3S day file, one L3S night file, and one day and night composite file, which is intended to be used as a daily foundation SST (see section 1.2 for further information). SSES are derived from L3C statistics and are intended to represent the typical deviation of the stated SST value from the *in situ* value, allowing for the fact that the time scale over which the data are considered valid is extended and the sensors are not identical. Only L3C files from platforms capable of providing a minimum `file_quality_level` are merged into L3S files, but the resulting L3S `file_quality_level` assessment could be further reduced by issues in processing, or lower than expected `file_quality_level` of source L3C files. SST and SSES from a typical L3S file are shown in figure 5.

Scaling over different time scales is managed by making use of two sets of statistical data, a weighted SSES and a raw SSES. The weighted SSES consists of weighted SST ($T_{\mathrm{satellite},S}$) which is intended to be corrected with the weighted bias (μ_S), weighted degrees of freedom (n_S) and weighted standard deviation (σ_S) estimates. These are assessed based on the assumption that discrepancies in measurement are primarily due to a combination of satellite measurement, retrieval, cloud clearing and other sensor specific items. During aggregation, weights that include the quality, degrees of freedom and standard deviation explicitly are used. On the other hand, the raw SSES mean ($\mu_{w,S}$), raw standard deviation ($\sigma_{w,S}$) and raw degrees of freedom ($n_{w,S}$) are collected without quality weighting, under the assumption that all of the measurements are accurate and reliable, and reflect the natural temperature variation of the ocean, used to categorize what might be possible over the time period in question.

The raw SSES can then be used to provide an upper estimate of the deviation of the ocean temperature over the aggregation set (or period of time). Since the weighted SST is representative of the temperature over the aggregation set, which is effectively an average, the raw variation can be included in the SSES assessment to reflect the uncertainties in the average SST over the same aggregation set due to observed ocean temperature variations. This contribution is scaled by $n^{-\frac{1}{2}}$, i.e. as an error of the mean. As the aggregation set gets larger, the error of the mean will diminish, whereas purely sensor and algorithm specific errors will not.

Figure 6 illustrates the typical suite of SSES related fields that may be present in an L3S file. Missing fields are assumed trivial values - for example, if all valid values of $n_{w,S}$ are 1, then $n_{w,S}$ and $\sigma_{w,S}$ may not be provided in the file.

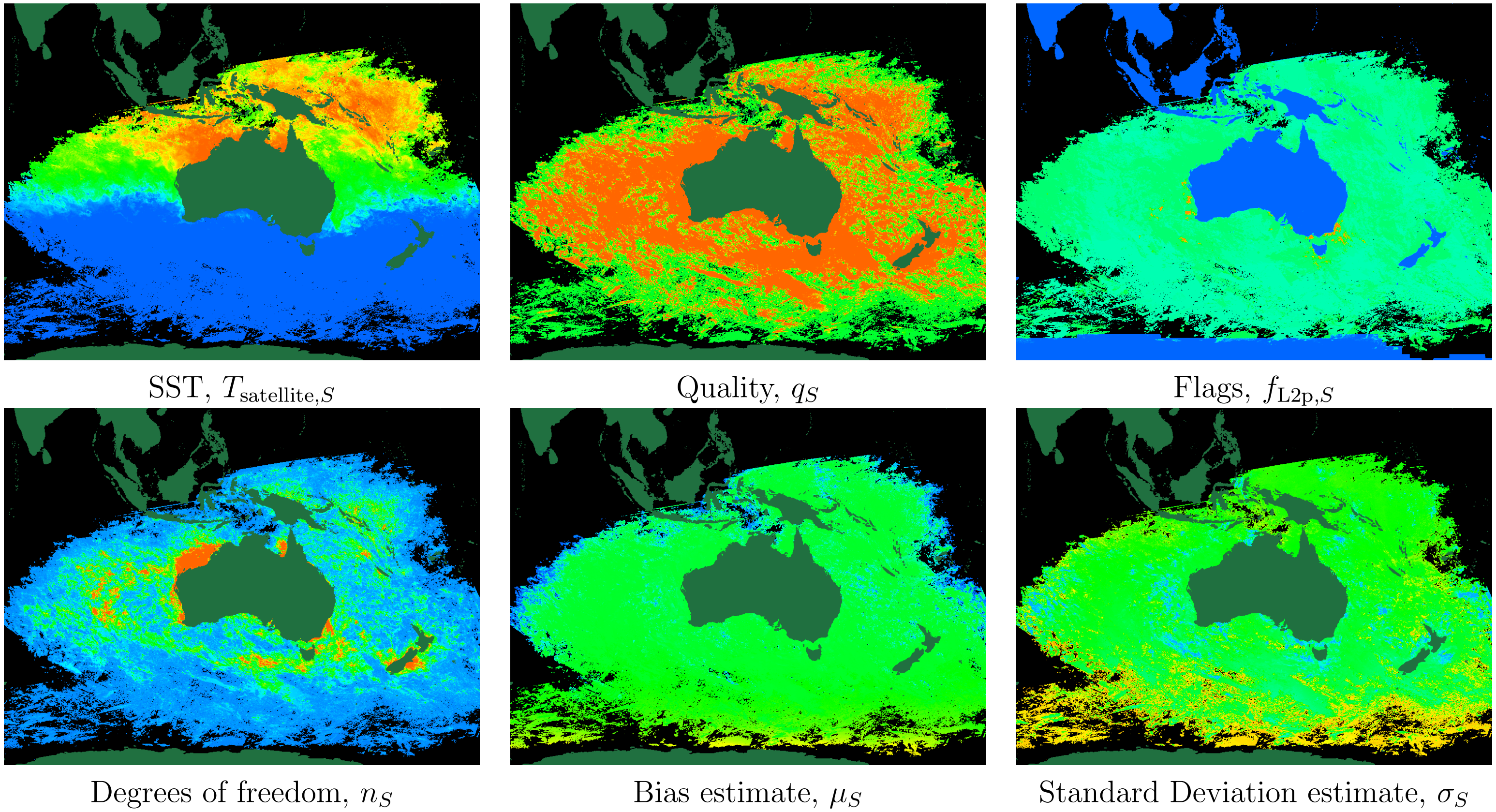
While the raw mean temperature $\mu_{w,S}$ and the sea surface temperature $T_{\mathrm{satellite},S}$ show great similarity, the degrees of freedom and standard deviation assessments are quite different. The raw degrees of freedom ($n_{w,S}$) counts the number of measurements made over the time period, or the number of observations made at the composite level. This could also be used as a proxy for the degree of clear sky over the period of aggregation. Whereas, the weighted degrees of freedom (n_S) includes quality weighting and an assessment of relative numbers of buoy measurements that went into the error assessment, and this is thus an indication of the strength of the retrieval. Similarly, the weighted standard deviation (σ_S) represents a buoy to sensor and retrieval uncertainty in addition to the sensor uncertainty, whereas the raw $\sigma_{w,S}$ represents an upper bound in the variation in the measured sea surface temperature under the assumption that all measurements were made with high accuracy and thus is a proxy for an upper bound of the geophysical variation over the period of aggregation.

The impact of the weighting scheme is demonstrated in figure 7. The weighted temperatures tend to be slightly more smooth even though the algorithm does not contain spatial averaging, because the temperatures are intended to be representative of the time period - see for example the extent of the yellow patches in these two diagrams. However, because no spatial averaging is employed, most of the fine shape detail is preserved. See section A.3 for more information about the composition and merging method employed.

The current GHRSST standard does not require degrees of freedom information or raw SSES to be provided in level 2 or level 3 files, so these should be considered experimental, and are labelled as such. However in section A.1 it is made clear that this information is essential if a consistent formulation of SSES is to be developed, especially when aggregation is required.

1.2 SST types

The GHRSST specification mentions several types of SST, characterizing the ocean temperature at different nominal depths near the surface. These are summarized in figure 8, reproduced from [28]. We are primarily concerned with two types of product, with the method of generating the SST from satellite brightness temperatures outlined in each case as follows,



SST, $T_{\text{satellite},S}$

Quality, q_S

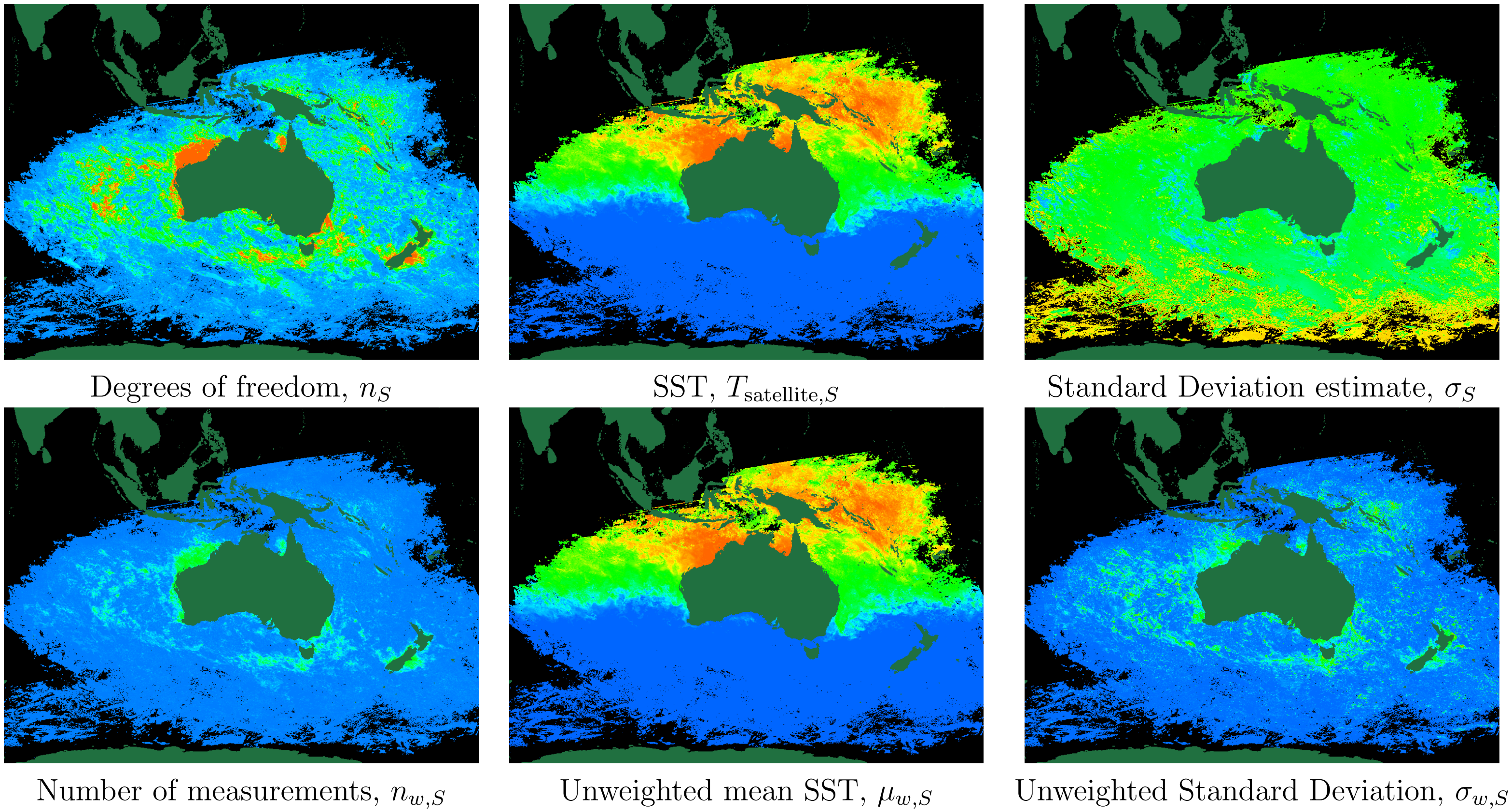
Flags, $f_{\text{L2p},S}$

Degrees of freedom, n_S

Bias estimate, μ_S

Standard Deviation estimate, σ_S

Figure 5: Sample L3S data set over the Australian domain. A typical L3S file contains fields for Sea surface temperature ($T_{\text{satellite},S}$), quality level based on proximity to cloud (q_S), SSES number of degrees of freedom (n_S), standard deviation (σ_S), and bias (μ_S), in addition to navigation information and ancillary fields. Unweighted number of degrees of freedom, bias, and standard deviation may also be provided if there is significant good quality over lap between neighbouring swaths. Figure 6 provides an example. Note regions where the reception may have either dropped out or was of poor quality (L3S files do not contain pixels with incoming quality q_C less than 2) are removed from the data set, making use of fill values (shown in black). The data above is a fourteen day, day time composite from NOAA-11 with a characteristic time of 15:20, April 7th, 1992 UTC at the central longitude of the image. The coverage over a 14 day time period is typically near complete over the Australian region. See table 2 for further information about the parameters contained within L3S files.



Degrees of freedom, n_S

SST, $T_{\mathrm{satellite},S}$

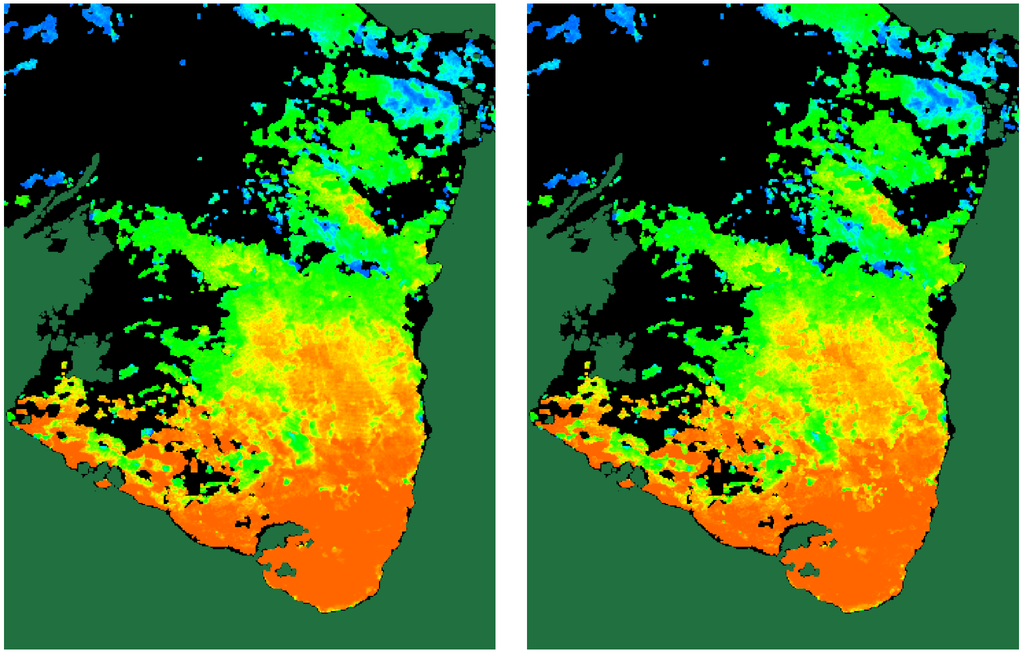
Standard Deviation estimate, σ_S

Number of measurements, $n_{w,S}$

Unweighted mean SST, $\mu_{w,S}$

Unweighted Standard Deviation, $\sigma_{w,S}$

Figure 6: L3S files contain two sets of SSES parameters, so that error contributions due to population variation over the characteristic time period can be isolated from those derived by weighting measurements based on relative surety. The data above is a fourteen day, daytime composite from NOAA-11 with a characteristic date of 15:20, April 7th, 1992 UTC at the central longitude of the image. The upper row is the weighted information, whereas the lower row contains the unweighted information. The unweighted mean temperature is $\mu_{w,S}$ whereas the weighted sea surface temperature is $T_{\mathrm{satellite},S}$. Degrees of freedom assessments count the number of measurements made over the time period $n_{w,S}$, and a quality assessment that include the buoy measurements that went into the error assessment, n_S. The weighted standard deviation is σ_S, whereas $\sigma_{w,S}$ represents the unweighted estimate.



Weighted SST, $T_{\text{satellite},S}$

Unweighted SST, $\mu_{w,S}$

Figure 7: A small region of interest in the Gulf of Carpentaria illustrating the difference between weighted ($T_{\text{satellite},S}$) and unweighted ($\mu_{w,S}$) sea surface temperature. The data above is a fourteen day, daytime composite from NOAA-11 with a characteristic date of 15:20, April 7$^{\text{th}}$, 1992 UTC. The unweighted sea surface temperature tends to show more local variation due to difference in quality and σ, whereas the weighted region reproduces a more homogeneous but still varying representation, especially to the south east of the gulf where the number of measurements over this period is somewhat higher than in the west. Warmer measurements can generally be attributed to a higher weighting of good quality day measurements, or measurements that are less likely to be cloud contaminated due to poor detection.

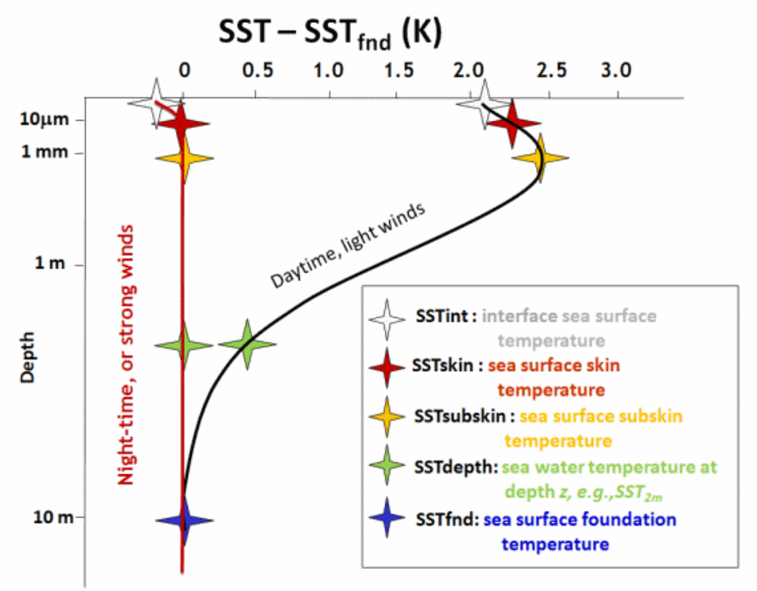


Figure 8: Different types of SST based on a hypothetical vertical profile. Reproduced from the GHRSST website[28].

Channel	Nadir resolution (km)	Wavelength (μm)	Typical use
1	1.09	0.58–0.68	Daytime cloud and surface mapping
2	1.09	0.725–1.00	Land-water boundaries
3A	1.09	1.58–1.64	Snow and ice detection
3B	1.09	3.55–3.93	Night cloud mapping, sea surface temperature
4	1.09	10.30–11.30	Night cloud mapping, sea surface temperature
5	1.09	11.50–12.50	Sea surface temperature

Table 4: AVHRR channels and wavelength, reproduced from [19]. We use channels 3B, 4 and 5 for SST retrieval. Within this text, channel 3 is used to refer to channel 3B.

Skin SST - SSTskin

The AVHRR infra-red radiometer on NOAA POES measure brightness temperatures in the top 10-20μm of the ocean, based on the wavelength sensitivity. See table 4, reproduced from [19], for the radiometer details. Ocean temperatures that are observed by the satellite thus correspond to the very top or skin layer of the ocean. When comparing with *in situ* ocean measurements from temperature sensors located up to 10m below the skin, a uniform cool skin correction of -0.17K should be added to the *in situ* measurements, to compensate for a mean systematic bias between skin and sub-skin.[27] It is worth noting that if the ocean is not well mixed, which can be characterized by a low ($<$ 6m/s during the day and $<$ 2m/s at night) wind speed, or the surface is particulary rough (characaterized by a wind speed of $> 20m/s$), *in situ* and skin measurements may still exhibit a sizable discrepancy, even after this systematic correction is applied. For the purposes of validation it is common to not include *in situ* measurements which may exhibit such discrepancies.[6, 10]

Foundation SST - SSTfnd

The upper layer of the ocean may experience some warming during the day and cooling at night, particularly in shallow water that is exposed to the sun for long periods of time during the day, or when the water is not well mixed near the surface. This can be quite large (of order 5K or more [26]). When mixing occurs, the heat absorbed from the sun will generally be better dissipated in the water column, and the diurnal variation on the ocean surface is not as marked. In this sense, the upper layer warms and cools with respect to a thermal foundation or baseline, at some depth (typically 10m) below the surface, which is typically immune to daytime fluctuations. This baseline is the foundation SST defined by GHRSST[28] in figure 8. In order to reproduce this from satellite surface measurements, it is desirable to restrict ourselves to measurements where we suspect high surface mixing will ensure that the ocean is properly mixed, during times when the incident thermal radiation is not high and has not been high for a long period of time. In practical terms, we choose a daytime wind speed range of 6 to 20m/s, and a nighttime wind speed range of 2 to 20m/s.[6, 10] Estimates of foundation SST are thus produced from skin SST L3C values by considering a merge of day and night observations, subject to the appropriate range of wind speed. The resulting merge is systematically offset by adding the cool skin correction -0.17K[6, 10] to further compensate for the observation that ocean skin temperatures are on average cooler than sub-skin temperatures.

1.3 Product domains and time coverage

Two product domains are currently provided and are summarized in table 5, along with the periods of time that are also supported. The actual dates of coverage for each satellite platform are outlined in table 9. The coverage by satellite over time is shown graphically in figure 10. Products are derived from images received at Australian satellite reception stations, illustrated in figure 9, and with the exception of Antarctic reception, processed in near real time. Due to limitations in data transmission, reception from Antarctic stations is currently not included in near real time processing, and is periodically added to the data record on a batch basis. For most of the year, Hobart reception is sufficient to provide SST measurements that extend close enough to the ice edge for most practical purposes.

L2P (swath), L3U (gridded) and 1 day L3C files are available for all of the satellites over the time periods indicated, and merged multi-day multi-platform L3S files are available based on data from the (multiple) satellites indicated. Each satellite platform is considered independently, however, harmonization of the measurements at L2P, L3U or L3C level, is provided by a common *in situ* database, which underpins the retrieval and SSES estimation algorithms. L3S data sets from multiple platforms consider these SSES estimates and compensate accordingly, thus the L3S data set can be considered a harmonization of the same instrument over multiple platforms.

1.4 Ancillary Information

Each GHRSST compliant SST record contains ancillary information, which may aid in the interpretation or use of the SST data provided. A description of each of the ancillary information is summarized in table 10.

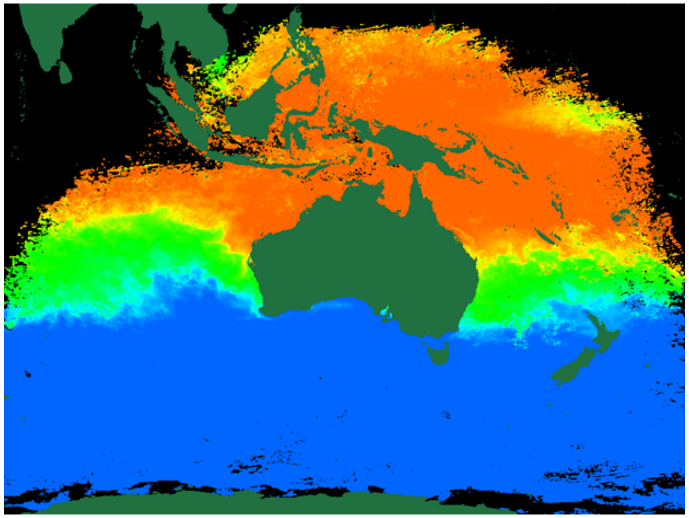
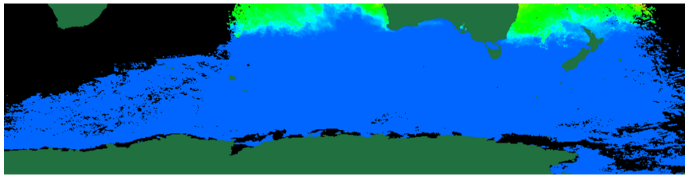
Domain name	Continental Australia	
Extent	$\theta_{\text{lat}} = 69.99°S$ to $19.99°N$ $\phi_{\text{lon}} = 69.99°E$ to $170.01°W$	
Resolution	$\theta_{\text{lat}} = 0.02°$ $\phi_{\text{lon}} = 0.02°$ 6000×4500 pixel	
Coordinates	cylindrical equidistant	
Time period	1992 January 1^{st} to date	
Platform	NOAA-09, 11, 12, 14, 15, 16, 17, 18, 19	
Received by	see figure 9	
Products	L3U, L3C, L3S. See table 6.	
real time	within 3 hours	
archive	usually within 3 days	
Domain name	Southern Ocean	
Extent	$\theta_{\text{lat}} = 77.49°S$ to $27.51°S$ $\phi_{\text{lon}} = 2.51°E$ to $157.51°W$	
Resolution	$\theta_{\text{lat}} = 0.02°$ $\phi_{\text{lon}} = 0.02°$ 10000×2500 pixel	
Coordinates	cylindrical equidistant	
Time period	1992 January 1^{st} to date	
Platform	NOAA-09, 11, 12, 14, 15, 16, 17, 18, 19	
Received by	see figure 9	
Products	L3U, L3C, L3S. See table 8.	
real time	not available from Antarctic reception stations	
archive	batch processed when Antarctic reception data becomes available	

Table 5: Currently supported product domains, with approximate extent of coverage based on monthly merged SST during all received satellites in February 2008. Coverage gaps south of the Indian subcontinent, in the southern Indian ocean and in the equatorial western Pacific ocean are due to lack of reception by Australian reception stations. L2P files are ungridded and available over both regions.

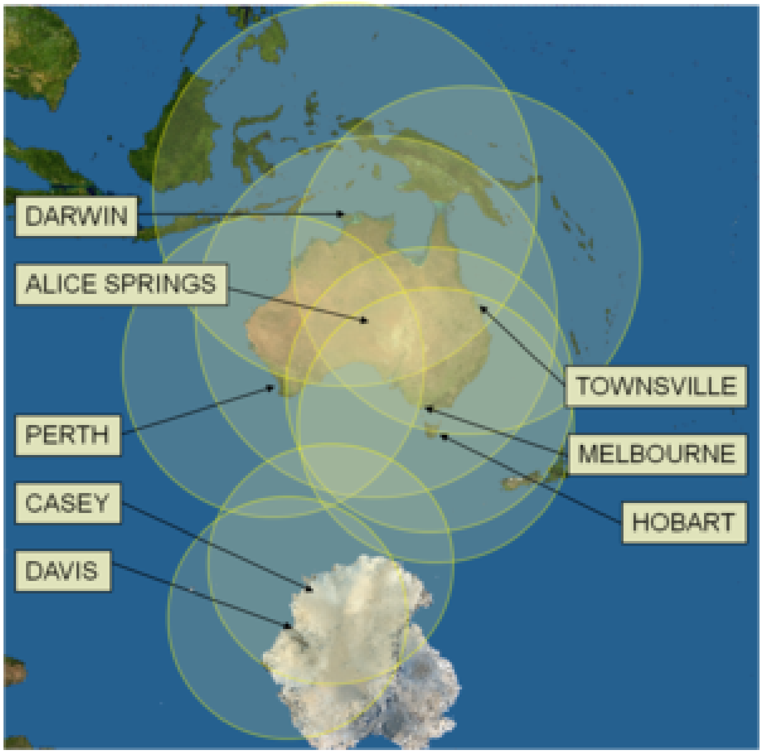


Figure 9: Australian satellite receptions stations, reproduced from the IMOS SST web portal[12].

Type	SST type	Time period	File name template
L2P	SSTskin	per swath	`{YYYYMMDDhhmmss}-ABOM-L2P_GHRSST-SSTskin-AVHRR{nn}_D-Des-v02.0-fv{fv}.nc`
		ungridded	`{YYYYMMDDhhmmss}-ABOM-L2P_GHRSST-SSTskin-AVHRR{nn}_D-Asc-v02.0-fv{fv}.nc`
L3U	SSTskin	per swath	`{YYYYMMDDhhmmss}-ABOM-L3U_GHRSST-SSTskin-AVHRR{nn}_D-Des-v02.0-fv{fv}.nc`
			`{YYYYMMDDhhmmss}-ABOM-L3U_GHRSST-SSTskin-AVHRR{nn}_D-Asc-v02.0-fv{fv}.nc`
L3C, 1 day	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3C_GHRSST-SSTskin-AVHRR{nn}_D-1d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3C_GHRSST-SSTskin-AVHRR{nn}_D-1d_night-v02.0-fv{fv}.nc`
L3C, 3 day	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3C_GHRSST-SSTskin-AVHRR{nn}_D-3d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3C_GHRSST-SSTskin-AVHRR{nn}_D-3d_night-v02.0-fv{fv}.nc`
L3S, 1 day	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-1d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-1d_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-AVHRR_D-1d_dn-v02.0-fv{fv}.nc`
L3S, 3 day	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-3d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-3d_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-3d_dn-v02.0-fv{fv}.nc`
L3S, 6 day	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-6d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-6d_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-AVHRR_D-6d_dn-v02.0-fv{fv}.nc`
L3S, 14 day	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-14d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-14d_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-AVHRR_D-14d_dn-v02.0-fv{fv}.nc`
L3S, monthly	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-1m_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-AVHRR_D-1m_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-AVHRR_D-1m_dn-v02.0-fv{fv}.nc`

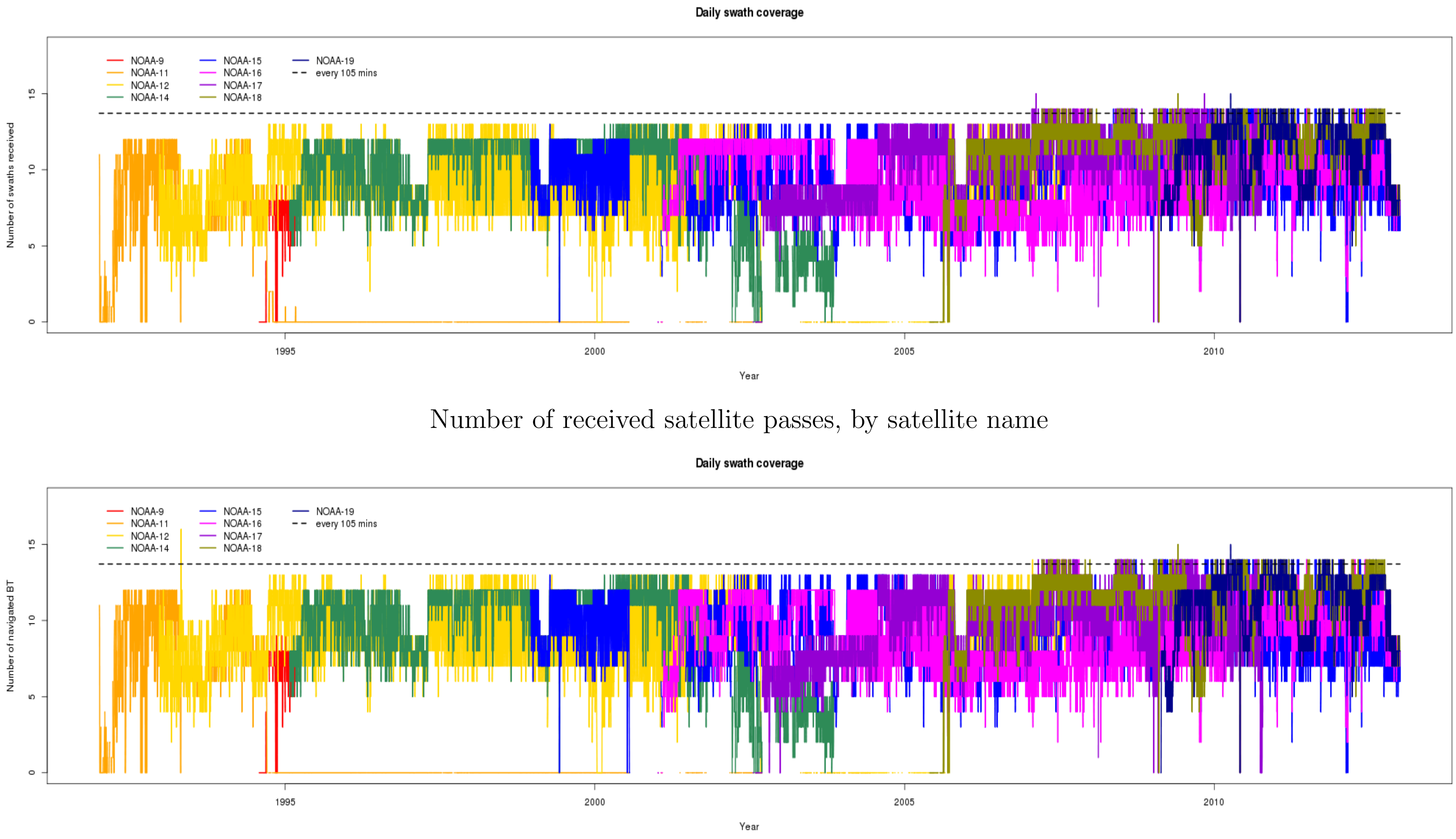
Table 6: Currently supported Australian region NOAA AVHRR based products. The file names contain variable information comprising a characteristic date, `{YYYYMMDDhhmmss}`, platform number `{nn}`, and file version `{fv}`. `Asc` and `Des` describe the orbit as ascending or descending. Note L2P files are navigated swath files and thus do not specify the domain of coverage explicitly in the file name.

Type	SST type	Time period	File name template
L3U VIIRS	SSTskin	per swath	`{YYYYMMDDhhmmss}-ABOM-L3U_GHRSST-SSTskin-NPP_VIIRS-Pol-v02.0-fv{fv}.nc`
L3C, VIIRS 1 day	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3C_GHRSST-SSTskin-NPP_VIIRS-1d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3C_GHRSST-SSTskin-NPP_VIIRS-1d_night-v02.0-fv{fv}.nc`
L3S, 1 day, Mixed	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-1d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-1d_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-MIXED-1d_dn-v02.0-fv{fv}.nc`
L3S, 3 day, Mixed	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-3d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-3d_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-3d_dn-v02.0-fv{fv}.nc`
L3S, 6 day, Mixed	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-6d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-6d_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-MIXED-6d_dn-v02.0-fv{fv}.nc`
L3S, 14 day, Mixed	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-14d_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-14d_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-MIXED-14d_dn-v02.0-fv{fv}.nc`
L3S, monthly, Mixed	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-1m_day-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTskin-MIXED-1m_night-v02.0-fv{fv}.nc`
	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-MIXED-1m_dn-v02.0-fv{fv}.nc`

Table 7: Currently supported Australian region based products from sources other than NOAA AVHRR. The file names contain variable information comprising a characteristic date, `{YYYYMMDDhhmmss}`, and file version `{fv}`. `Pol` describes the orbit as Polar. Note the lack of L2P files indicates data that is not based on direct reception over the Australian region.

Type	SST type	Time period	File name template
L2P	SSTskin	per swath	`{YYYYMMDDhhmmss}-ABOM-L2P_GHRSST-SSTskin-AVHRR{nn}_D-Des-v02.0-fv{fv}.nc`
		ungridded	`{YYYYMMDDhhmmss}-ABOM-L2P_GHRSST-SSTskin-AVHRR{nn}_D-Asc-v02.0-fv{fv}.nc`
L3U	SSTskin	per swath	`{YYYYMMDDhhmmss}-ABOM-L3U_GHRSST-SSTskin-AVHRR{nn}_D-Des_Southern-v02.0-fv{fv}.nc`
			`{YYYYMMDDhhmmss}-ABOM-L3U_GHRSST-SSTskin-AVHRR{nn}_D-Asc_Southern-v02.0-fv{fv}.nc`
L3C, 1 day	SSTskin	day	`{YYYYMMDDhhmmss}-ABOM-L3C_GHRSST-SSTskin-AVHRR{nn}_D-1d_day_Southern-v02.0-fv{fv}.nc`
		night	`{YYYYMMDDhhmmss}-ABOM-L3C_GHRSST-SSTskin-AVHRR{nn}_D-1d_night_Southern-v02.0-fv{fv}.nc`
L3S, 1 day	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-AVHRR_D-1d_dn_Southern-v02.0-fv{fv}.nc`
L3S, monthly	SSTfnd	day & night	`{YYYYMMDDhhmmss}-ABOM-L3S_GHRSST-SSTfnd-AVHRR_D-1m_dn_Southern-v02.0-fv{fv}.nc`

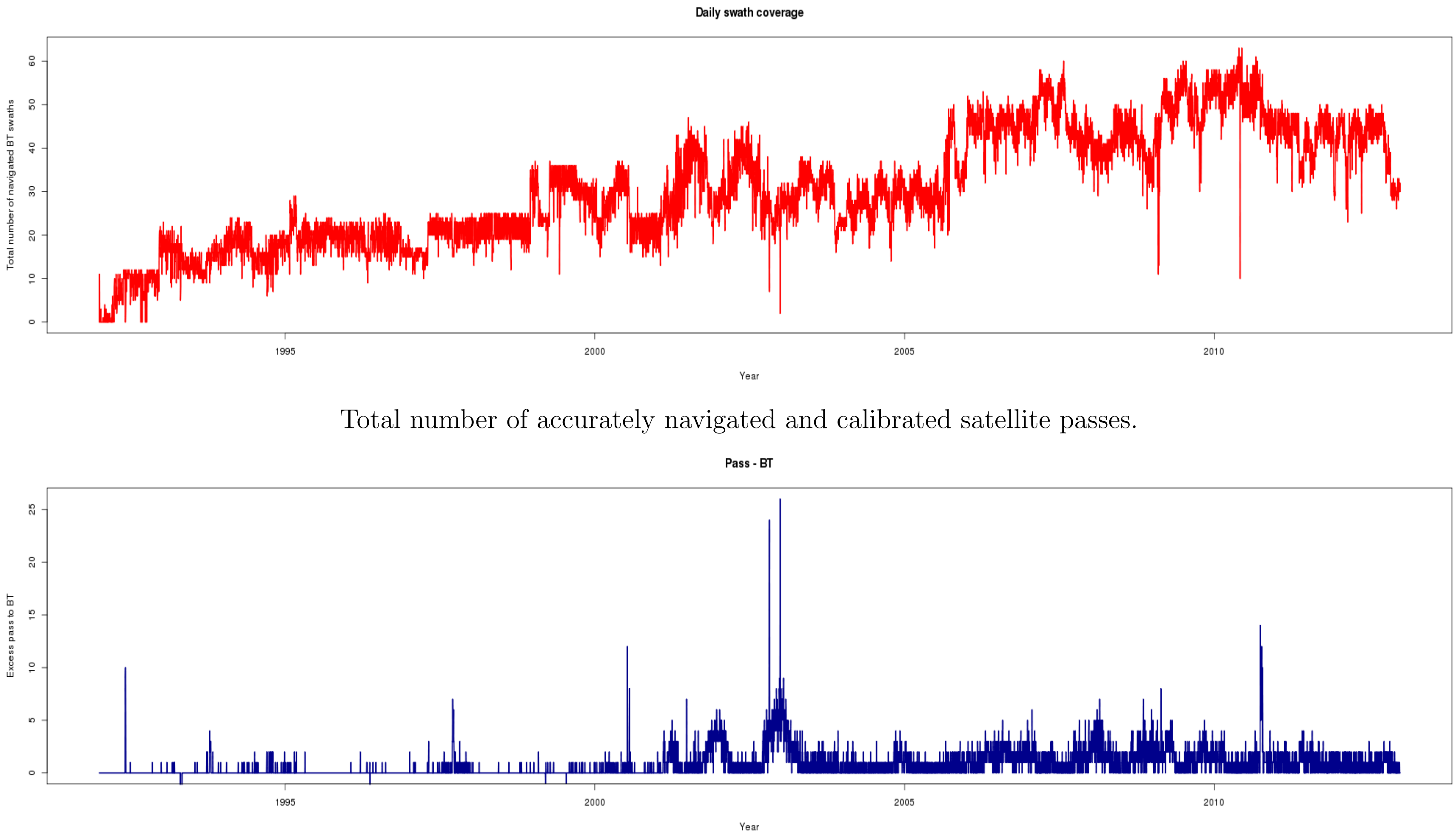
Table 8: Currently supported Southern Ocean products. The file names contain variable information comprising a characteristic date, `{YYYYMMDDhhmmss}`, platform number `{nn}`, and file version `{fv}`. `Asc` and `Des` describe the orbit as ascending or descending, and `Southern` indicates that the resulting grid corresponds to the Southern domain. Note L2P files are navigated swath files and thus do not specify the domain of coverage explicitly in the file name. The `history` global metadata in L2P netCDF files contains an entry `isSouthern` which indicates if there is potentially coverage over the Southern domain. See table 17 for futher information.



Number of received satellite passes, by satellite name

Number of accurately navigated and calibrated satellite passes, by satellite name

Figure 10: Time and platform daily coverage extends over nine platforms and 21 years. Over every time period multiple platform data is available.



Total number of accurately navigated and calibrated satellite passes.

Total number of archived satellite passes unable to be converted to brightness temperatures.

Figure 11: Time and platform daily coverage extends over eight platforms (Although some NOAA-9 data exists, it was decided that there is too little to provide a properly calibrated data set), and 21 years. Total number of navigated and received swaths on a daily basis. Total number of passes that were not converted to brightness temperatures due to format, transmission, lack of data, navigation or other technical issues is also shown.

Platform	Unnavigated First Date (YYYY-MM-DD)	Unnavigated Last Date (YYYY-MM-DD)	Navigated First Date (YYYY-MM-DD)	Navigated Last Date (YYYY-MM-DD)
NOAA-11	1992-01-01	1996-03-23	1992-01-02	1994-09-21
NOAA-12	1992-12-19	2000-11-30	1992-12-19	2000-11-30
NOAA-14	1995-01-27	2003-11-14	1995-01-28	2003-11-14
NOAA-15	1998-12-16	to date	1998-12-16	to date
NOAA-16	2000-09-22	2014-06-05	2001-02-06	2014-06-05
NOAA-17	2002-07-18	2010-10-14	2002-09-15	2010-10-14
NOAA-18	2005-05-29	to date	2005-08-17	to date
NOAA-19	2009-02-18	to date	2009-02-22	to date

Table 9: Satellite coverage by platform for fv02 version product. Navigated data sets are used to generate the GHRSST data set. Un-navigated data sets may be recovered in future extending the possible coverage of the GHRSST data set as indicated. The NOAA-09 archive is too sparse to accurately regress against *in situ* measurements, and has been excluded from the scope of the current work.We have NOAA-12 direct reception records and navigation up until 2007-08-09, with a gap of 3 years between 2002-09-05 and 2005-09-05. However, technical issues with night data have meant that data after 2000-11-30 is considered substandard.

1.5 Product versions

There are two product versions which are currently produced by the Australian Bureau of Meteorology. These are called file version 1 (fv01) and file version 2 (fv02).

fv01 The fv01 product represents the legacy product, which is a fixed (non-adaptive), regression retrieval, with bin based SSES estimations[**paltaglou**], covering NOAA-15,16,17,18, and 19, platforms. This product version is currently produced in real time for ABOM real time systems and is published by IMOS for at least the last year of activity (currently data is available from 1st January 2015). Some fields and metadata that is discussed in this manual may not be present in the fv01 product.

fv02 The fv02 product represents a long time series product, and is substantially more complete than the fv01 product, produced with a variable (adaptive) regression based retrieval with modelled SSES estimations, covering NOAA-11 to 19 platforms. This is the default product covered by the discussion in this manual, unless it is explicitly stated otherwise. This product version is currently produced in delayed mode, and for long term archives, and is published by IMOS in the historical archive. The historical archive is updated when reception files become available from all of the Australian reception station stations, and is current to 31st December 2014. From the point of view of fields and metadata, the fv02 product is considered a superset of fv01.

A summary of the differences between the two products is shown in table 11

2 Estimating SSES for GHRSST compliant L2 product

There are two approaches made for the construction and determination of SSES:

Ancillary Field Name	Description and Use
`dt_analysis`	$T_{\text{analysis}} - T_{\text{satellite}}$. The difference between a level 4 foundation SST from the previous UTC calendar day and the observed SST. The level 4 foundation SST reflects the SST at 10m below the surface and is thus expected to be considerably more stable than skin SST measurements. This may be an indication of diurnal warming events, which will show as positive values, and possibly aid in the determination of cloud that was not correctly identified, which will show as negative values.
`sst_dtime`	The time of the observation. Add `sst_dtime` to the global coordinate `time` to give the number of seconds since 00:00:00, January 1st 1981. For aggregated level 3 files, the SST is assumed linearly interpolated, and the time is the weighted average of the individual times.
`wind_speed`	The 10 meter wind speed. The wind speed is estimated from re-analysis numerical weather prediction models. If the wind speed is low, the surface of the ocean is not well mixed, and there is expected to be a larger discrepancy with *in situ* measurements and a larger diurnal effect. If the wind speed is high, there may be some uncertainty as to what satellite and *in situ* measurements are actually measuring due to large surface waves and spray.
`wind_speed_dtime_from_sst`	The time difference between the reported wind speed and the observation, in hours. This may be used to estimate the accuracy of the stated wind speed.
`sea_ice_fraction`	A number between zero and one that represents the fraction of coverage by sea ice on the previous day. The presence of sea ice may be problematic for cloud clearing and SST measurements. Sea ice forms a dynamic coastal zone in the southern ocean which is of scientific interest, being able to identify where this is facilitates studies of this.
`aerosol_dynamic_indicator`	An indication of the amount of atmospheric contaminants (dust, ash, etc) on the previous day. It is possible that such contaminants, which create spurious and short lived atmospheric conditions, interfere with the SST retrieval.
`adi_dtime_from_sst`	The time difference between the reported aerosol and the observation, in hours. This may be used to estimate the accuracy of the aerosol measurement.
`satellite_zenith_angle`	The angle from the normal at which the observation is made, in degrees. Observations made far from the normal (large satellite zenith angle) tend to be coarser in resolution and exhibit poorer navigation. Retrievals additionally require a more complicated atmospheric compensation. In spite of this, we keep observations over the full range of zenith angles to maximize coverage. SSES are adjusted based on zenith angle, and will compensate for these effects somewhat with enlarged error estimates. It may be worth considering removing observations with a large zenith angle for some applications.

Table 10: Summary of ancillary data sources provided in GHRSST compliant files, with a short description of how each might be useful.

Item	fv01	fv02
SST Retrieval	Regression, fixed coefficients tuned over the first 2 years of platform operation. Separate algorithms for day and night with a large number of terms.[**paltaglou**]	Regression, adaptive coefficients tuned over a rolling 2 year window, updated monthly. Separate standard algorithms for day and night as well as a three channel unified day and night algorithm.
SSES Generation	Lookup table based using a 60 day rolling window.[**paltaglou**]	Modelled using a one year rolling window, updated every five days.
Time coverage	2000-01-01 to present. Coverage is incomplete over some periods, although the L3S daily composites form a close to complete record.	1992-01-02 to the end of the most recent batch process. Coverage is complete up to navigation and reception issues, and subject to the availability of data from all Australian reception stations.
Spatial coverage	Includes reception from Australian continental reception stations.	Includes reception from both continental and Antarctic stations. Coverage is enhanced over time periods where earlier and later platforms overlap.
Platform coverage	NOAA-15,16,17,18,19. Earlier dates only include NOAA-15 in L3S composites.	NOAA-11,12,14,15,16,17,18,19. For any given date, all retrievals from the relevant set of active platforms are included in composite L3S files.
Fields	Default set of fields	`sea_surface_temperature_day_night`, `sses_quality` included in L2P files for enhanced diurnal studies and time relevant quality assessment (respectively).
Metadata	Default, but tends to be inconsistent in some comment fields.	Default.
Coordinate systems	Common coordinate systems are applied. However, Level 3 Latitude coordinates are not uniformly monotonically increasing.	Level 3 Latitude coordinates are monotonically increasing.

Table 11: Differences between fv01 and fv02 product

look-up table approach A legacy approach based on the fv01 product, using a look-up table based on 60 days of *in situ* measurements prior to the measurements under consideration is used to estimate the error statistics based on *in situ* measurements made under similar conditions.

modelling approach An empirical modelling approach which uses one year of *in situ* measurements to construct a model for SSES based on measurements made under similar conditions. This approach is employed in the fv02 product.

Both of these approaches require an assessment of measurement quality which is determined based on proximity to identified cloud in L2P files, then carried from level 3 product generation as will be discussed in the following sections.

2.1 Determination of quality level

The quality level, q is determined by computing the distance of each pixel in kilometers to identified cloud. Best quality pixels, $q = 5$ are greater than or equal to 5 kilometers from the nearest cloud. Since at nadir, 1 pixel is approximately 1 kilometer, the kilometer distance corresponds closely to the pixel distance. Away from nadir, however, the pixel distance can be considerably smaller than the kilometer distance. The kilometer distance, d, is computed using a spherical approximation, under the assumption that the changes in θ_{lat} and ϕ_{lon} are small when one travels from pixel X_1, to X_2, as follows,

$$d^2(X_1, X_2) = 40681004 \left((\theta_{\text{lat},X_1} - \theta_{\text{lat},X_2})^2 + (\phi_{\text{lon},X_1} - \phi_{\text{lon},X_2})^2 \cos^2 \theta_{\text{lat},X_1} \right) \tag{2}$$

The quality level, q_j at location X_j is thus defined in terms of the set of pixels that are flagged cloudy,

$$\texttt{cloudy} = \{ i : X_i \text{ is cloud contaminated} \} \tag{3}$$

$$q_j = \min \left(\text{int} \left(\sqrt{ \min_{i \in \texttt{cloudy}} d^2(X_j, X_i) } \right), 5 \right) \tag{4}$$

The regions of the image fully on cloud correspond to $q = 0$, whereas fully clear (5 kilometres or greater from identified cloud) regions have $q = 5$. The boundaries between these two regions correspond to $q = \{1, 2, 3, 4\}$. The number of pixels with $q < 4$ increases with decreasing quality, and because of the boundary nature of $q = \{1, 2, 3, 4\}$, the number of $q = 4$ pixels is considerably less than the number of $q = 5$ pixels. The cloud detection is a variant of the CLAVRX algorithm[25].

2.2 SSES for L2P Class product - a look-up table approach - fv01

In order to estimate the SSES for SST on raw measurements, we consider the deviation from drifting buoy (excluding Argo floats) temperature measurements with a simple fixed $0.17K$ cool skin correction, δ_i, over a period of 60 days prior to the time of measurement,

$$\delta_i = \Delta T_{i,sI} + 0.17 = T_{i,\text{satellite}} - T_{i,\text{insitu}} + 0.17 \tag{5}$$

The deviations are quality controlled for favourable conditions, per section 3.1, then binned in one of 48 bins based on binning in three dimensions,

- Day (solar zenith angle < 90°), night (solar zenith angle $\geq$ 90°), and both (no restriction on solar zenith angle), corresponding to the different time of day conditions in which the brightness temperature information that is available and retrieved, three categories in total.

- Quality level, $q = \{2, 3, 4, 5\}$, 4 categories in total.

- Satellite zenith angle θ_z, corresponding to $0° < \theta_z \leq 30°$, $30° < \theta_z \leq 50°$, $50° < \theta_z \leq 60°$ and $60° < \theta_z \leq 90°$, 4 categories in total.

as outlined in section 2.2.1. This results in SSES estimates for each bin which are applied on a bin by bin basis to the target data.

2.2.1 Binning based on *in situ* measurements

The binning process is critically dependent on the number of available measurements for each bin. Over a 60 day period the total number of favourable measurements may not be very large and varies somewhat over time, as does the error associated with the determination of the SSES. The historical SSES for each bin are used to compensate for the lack of data as follows:

- A number of measurements $n < 5$ is considered to be insufficient to determine accurate SSES information. This can be qualitatively understood if one considers the minimum sample size in order to do a test for normality such as the Jarque-Bera test[13]. In this case, the maximum of the most recent historical values $(\mu_{\text{def}}, \sigma_{\text{def}})$, and the day or night SSES (provided day or night statistics have $n \geq 5$), are used, with the number of degrees of freedom forced to $n = 1$,

$$\begin{aligned}
\mu_b &= \max\left(\mu_{\text{def}}, \mu_b|_{n\geq 5}\right) & (6)\\
\sigma_0 &= \sqrt{\sigma_{\text{def}}^2 + \mu_{\text{def}}^2} & (7)\\
\sigma &= \max\left(\sigma_{\text{def}}, \sigma_b|_{n\geq 5}\right) & (8)\\
n &= 1 & (9)
\end{aligned}$$

  Historical values are computed on the first day of the month, corresponding to the difference between *in situ* and satellite measurements over 366 days prior (or as much data as the record permits, if less than 366 days). The typical performance of this historical trend using the fv01 method of retrieval given by Paltoglou *et al*(2010)[**paltaglou2010**] are shown in figure 12. This shows among other things, that the bias correction can be of the order of $0.5K$ for poor quality, $q = 2$ observations, but is considerably smaller for good quality, $q = 5$, observations. Moreover, the bias shows reasonable short term stability, and is correlated for different quality levels.

  In the absence of historical values or sufficient data, default values which represent typical performance are applied as listed in table 12.

  μ_{def} in this table becomes increasingly negative at lower quality due to the interference with cloud or partial cloud pixels which biases satellite SST to be lower than the *in situ* SST as can be seen in figure 12.

- If there are a sufficient number of measurements to have some assurance of accuracy, $n > 32$, then the SSES are computed directly from these measurements by attributing the deviation

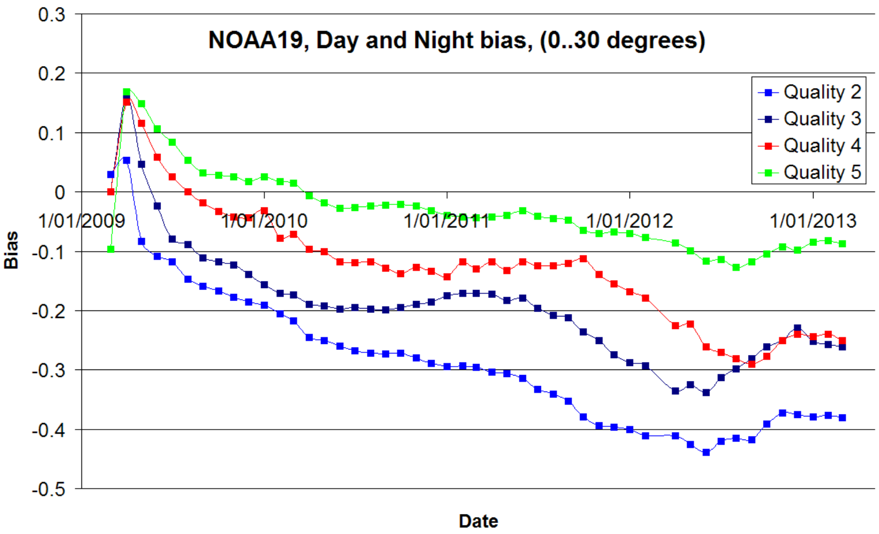


Figure 12: NOAA-19, monthly bias at the centre of the swath over a rolling one year window, for various quality levels, using a fixed retrieval equation tuned on *in situ* data near the start of the mission.

σ_{def} θ_z / q	2	3	4	5
0° to 30°	0.9	0.8	0.7	0.6
30° to 50°	0.9	0.8	0.7	0.6
50° to 60°	0.9	0.8	0.7	0.6
60° to 90°	0.9	0.8	0.7	0.6

μ_{def} θ_z / q	2	3	4	5
0° to 30°	-0.5	-0.35	-0.2	0
30° to 50°	-0.5	-0.35	-0.2	0
50° to 60°	-0.5	-0.35	-0.2	0
60° to 90°	-0.5	-0.35	-0.2	0

Table 12: Default values for SST bias and standard deviation

entirely to the sensor specific error,

$$\mu_b = \frac{1}{n} \sum_{i=1...n} \delta_i \tag{10}$$

$$\sigma_0^2 = \frac{1}{n} \sum_{i=1...n} \delta_i^2 \tag{11}$$

$$\sigma^2 = \sigma_0^2 - \mu^2 \tag{12}$$

$$\delta_i = T_{i,\text{satellite}} - T_{i,\text{insitu}} + 0.17 \tag{13}$$

The critical value of $n = 32$ was chosen based on the number of degrees of freedom required to ensure less than a 25% error in a χ^2 distributed variance, within the confidence range set as the equivalent of a single standard deviation confidence on a typical normal distribution. ie. This corresponds to a 1-σ error in σ^2 of $\leq 25\%$.

- If there are sufficient number of measurements to have some reasonable estimate made, but there may be questions about the accuracy, $5 \leq n \leq 32$, the variance is determined as a linear interpolation of the estimated variance and historical variance as a function of n, while the bias is a linear interpolation of the estimated bias and historical bias if the historical bias is greater than the estimate made.

$$n = \text{count}(\delta_i) \tag{14}$$

$$\mu = \frac{1}{n} \sum_{i=1...n} \delta_i \tag{15}$$

$$\sigma_e^2 = \frac{1}{n} \sum_{i=1...n} \delta_i^2 \tag{16}$$

$$\begin{aligned} \mu_b &= \mu, \ |\mu| \geq |\mu_{\text{def}}| \\ &= \xi\mu + (1 - \xi)\mu_{\text{def}}, \ |\mu| < |\mu_{\text{def}}| \end{aligned} \tag{17}$$

$$\sigma = \sqrt{\xi\left(\sigma_e^2 - \mu^2\right) + (1 - \xi)\sigma_{\text{def}}^2} \tag{18}$$

$$\sigma_0 = \sqrt{\xi\sigma_e^2 + (1 - \xi)\left(\sigma_{\text{def}}^2 + \mu_{\text{def}}^2\right)} \tag{19}$$

$$\delta_i = T_{i,\text{satellite}} - T_{i,\text{insitu}} + 0.17 \tag{20}$$

$$\xi = \frac{(n - 5)}{(32 - 5)} \tag{21}$$

ξ is a fraction which interpolates linearly between the $n = 5$ sample default $(\mu_{\text{def}}, \sigma_{\text{def}})$ and the $n = 32$ sample default (μ, σ_e). The interpolation is applied to the mean and the variance. The appearance of 0.17 in the definition of δ_i is the systematic bias that exists between skin sea surface temperature and *in situ* sea surface temperature, which results in warmer *in situ* easurements on average[6, 10].

All of the points on the swath are then assigned a set of three parameters: degrees of freedom, bias and uncertainty estimate, $\{n, \mu_b, \sigma\}$, based on their time of day, (day or night), quality level q, and view angle, θ_z, and this information is stored in the SSES fields designated in table 19.

2.2.2 Validation of SSES - a lookup table approach - fv01

SSES bias correction is applied by subtracting the estimated bias from the SST. The estimated bias is computed from historically recent data. In operational fv01 systems, the definition of historically recent is 60 days prior to the day before the day that the satellite measurement was made as a UTC date. The impact of the correction can be seen by comparing valid, surface mixed (significant, but not high amounts of surface wind), *in situ* sea surface temperature with corrected, and uncorrected SST.

Figures 13 through 16 show the impact of applying the bias correction to L2P measurements, at various quality levels.

The corrected SST (solid line) are clearly closer to the expected systematic cool skin bias than the uncorrected SST for all platforms over the period for which the method was applied. Robust standard deviations are not greatly affected by the change, which in shows that this approach adds little noise to the corrected SST, and in many cases there is a slight improvement.

Due to the statistical nature of this evaluation, we need to choose an aggregating window that is appropriate to reduce sources of noise. The impact of this choice for quality level 4, for example, is illustrated in figure 17 and 18.

The robust standard deviation of the variation of the median bias as a function of the window size, for quality level 4, for all platforms, is shown in figure 19. The optimal window size would be expected to be around 60 days, since this is the period over which the lookup table statistics are collected and biases determined.

However, since when sampling from any distribution, the standard error decreases as sample sizes increase, the robust standard deviation of $T_{i,\text{satellite}} - T_{i,\text{bias}} - T_{i,\text{insitu}}$ will decrease as the window period is increased (which corresponds to larger samples), up to window time periods that remain smaller than characteristic stability times of $T_{i,\text{satellite}} - T_{i,\text{bias}} - T_{i,\text{insitu}}$. This trend is broadly observed in the NOAA-18 data and to a lesser extent in NOAA-19 data. NOAA-15, 16 and 19, are relatively immune to window size changes (so long as the window is bigger than 20 days), with NOAA-16 perhaps showing a trending increase with window size consistent with the notion of continuous instability.

The chart shows that a window period of the order of 30 days seems like an adequate trade off between having frequent assessments, and accurate assessments.

2.3 SSES for L2P Class product - an empirical model approach - fv02

The approach of the previous section required binning in three variables (time of day, sun zenith angle and quality) and made use of a limited number of *in situ* measurements arbitrarily chosen to be subject to minimum variation. The bins were considered independent of each other, thus estimates were allowed to vary dramatically from bin to bin. This effect is made worse during periods where the number of *in situ* measurements was relatively small, as it was prior to the mid 1990s. Furthermore the accuracy of *in situ* measurements has improved dramatically in recent times due to the advent of drifting buoys. Prior to drifting buoys, *in situ* SST measurement datasets are dominated by moored buoy measurements, which are stationary and tend to bias measurements to the coastal regions, where the correlation with skin measurements may be influenced by other physical factors, such as tides, and interference by organisms. Uncertainties such as these can lead to very great differences in empirically determined biases and standard deviations, and this lack of continuity across bin boundaries could be considered a significant limitation and give reason to doubt the estimates made.

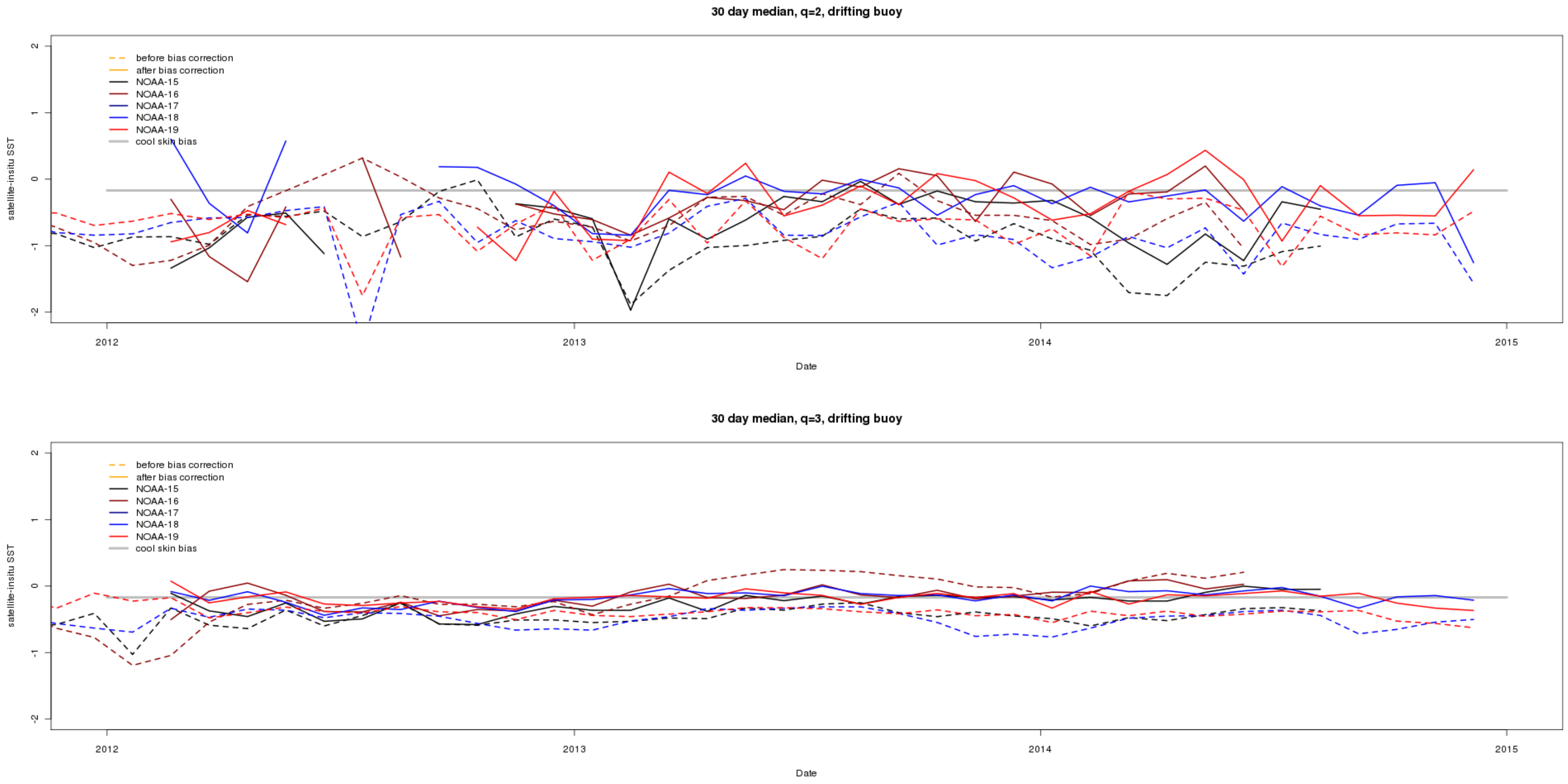


Figure 13: Impact of applying SSES bias correction (fv01 - lookup table approach) to the 30 day median of $T_{i,\text{satellite}} - T_{i,\text{insitu}}$, $q = 2, 3$. Dotted lines are before correction, solid lines are after correction.

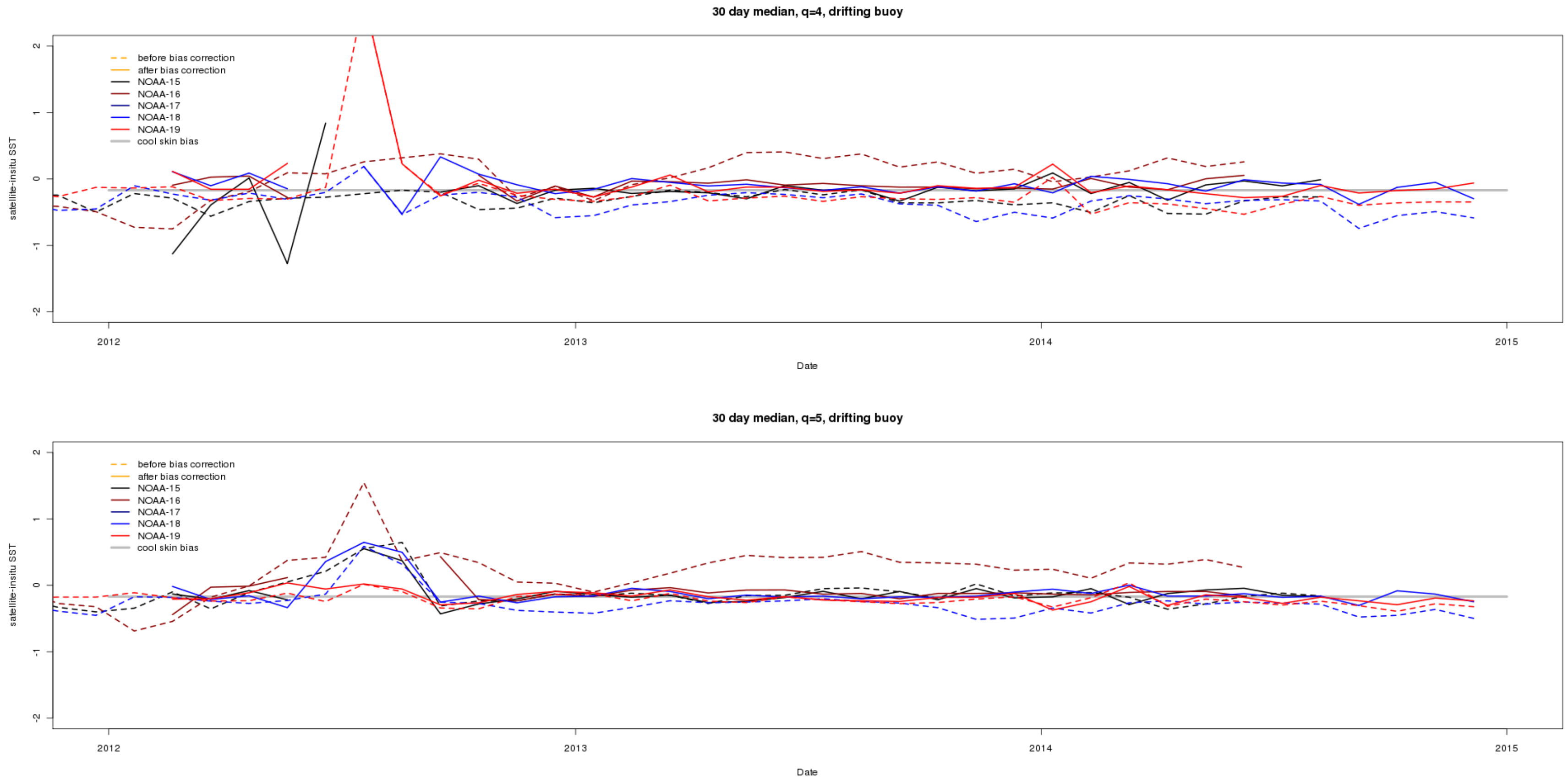


Figure 14: Impact of applying SSES bias correction (fv01 - lookup table approach) to the 30 day median of $T_{i,\text{satellite}} - T_{i,\text{insitu}}$, $q = 4, 5$. Dotted lines are before correction, solid lines are after correction.

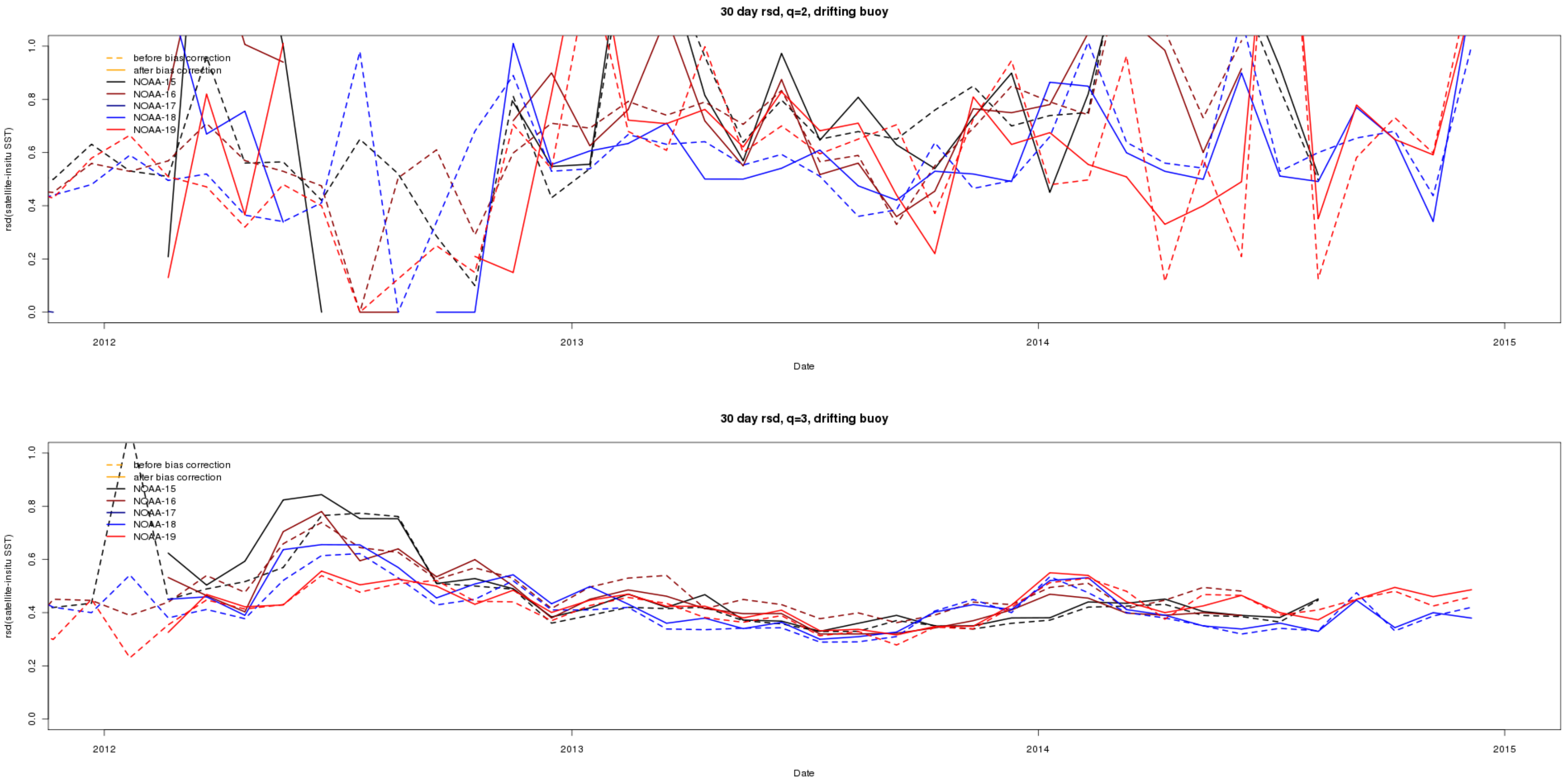


Figure 15: Impact of applying SSES bias correction (fv01 - lookup table approach) to the 30 day rsd of $T_{i,\text{satellite}} - T_{i,\text{insitu}}$, $q = 2, 3$. Dotted lines are before correction, solid lines are after correction.

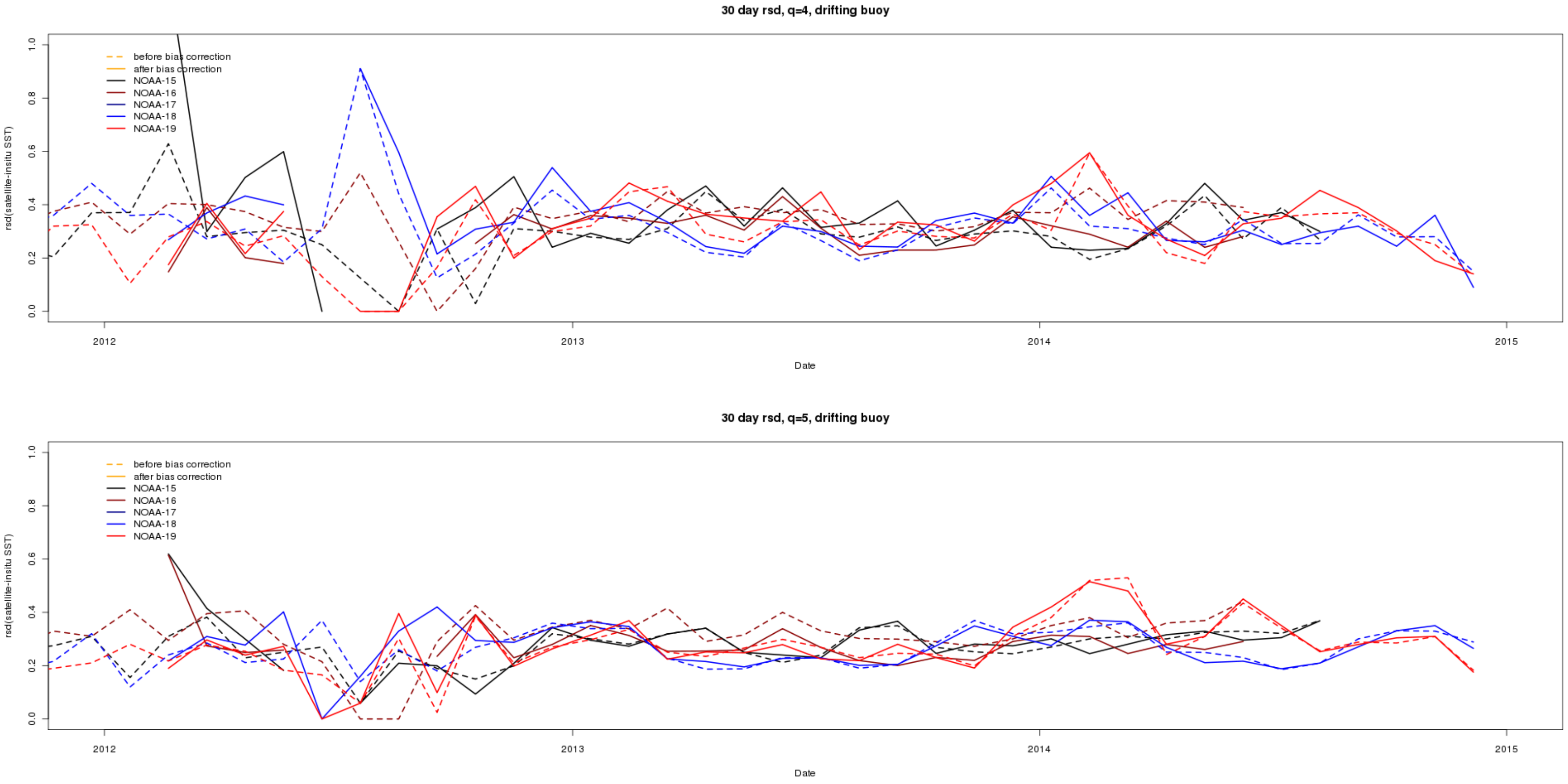


Figure 16: Impact of applying SSES bias correction (fv01 - lookup table approach) to the 30 day rsd of $T_{i,\text{satellite}} - T_{i,\text{insitu}}$, $q = 4, 5$. Dotted lines are before correction, solid lines are after correction.

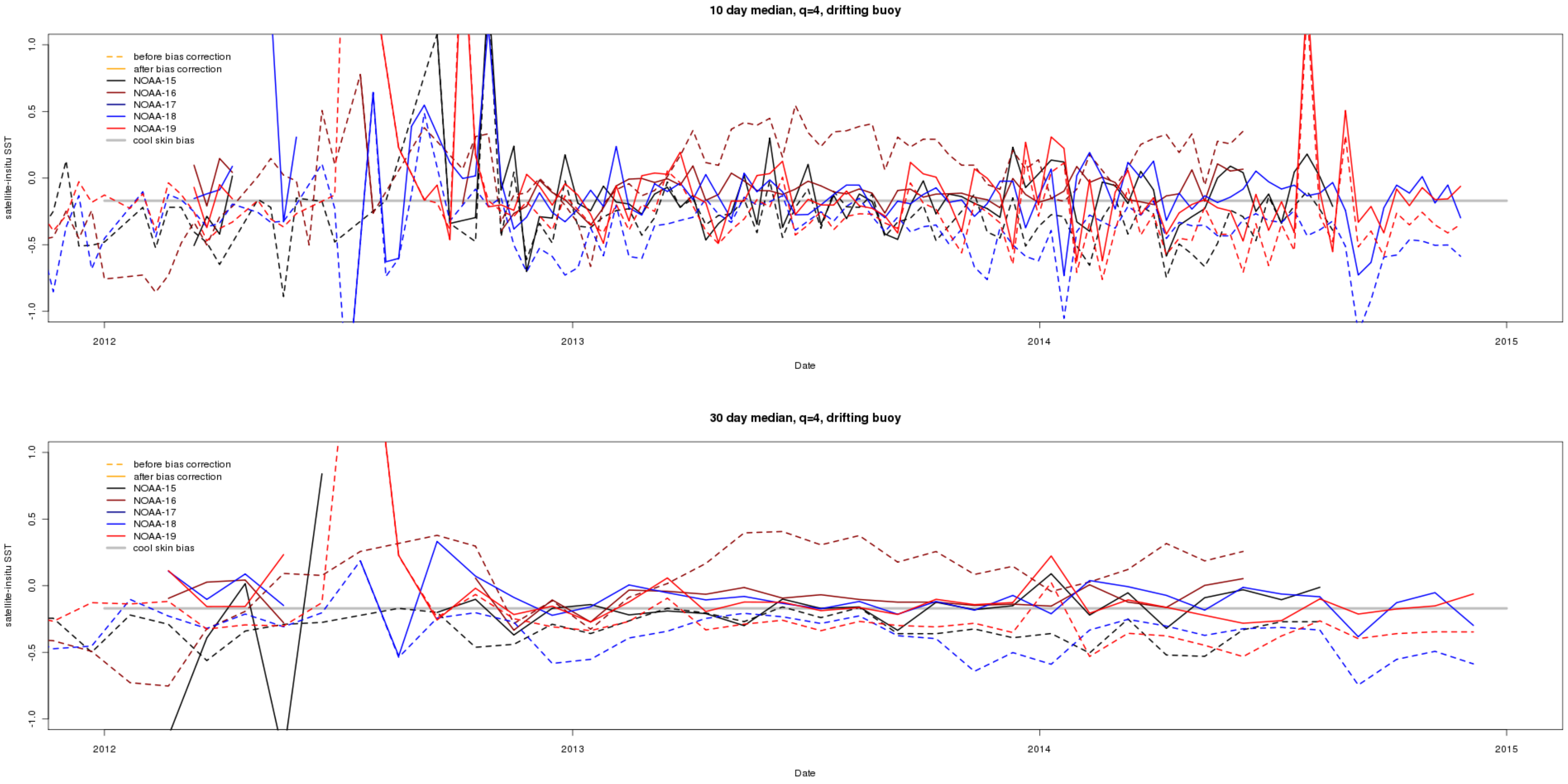


Figure 17: Impact of applying SSES bias correction (fv01 - lookup table approach with 60 day sample) to the median of $T_{i,\text{satellite}} - T_{i,\text{bias}} - T_{i,\text{insitu}}$ with 10 day and 30 day averaging windows. Dotted lines are before correction, solid lines are after correction.

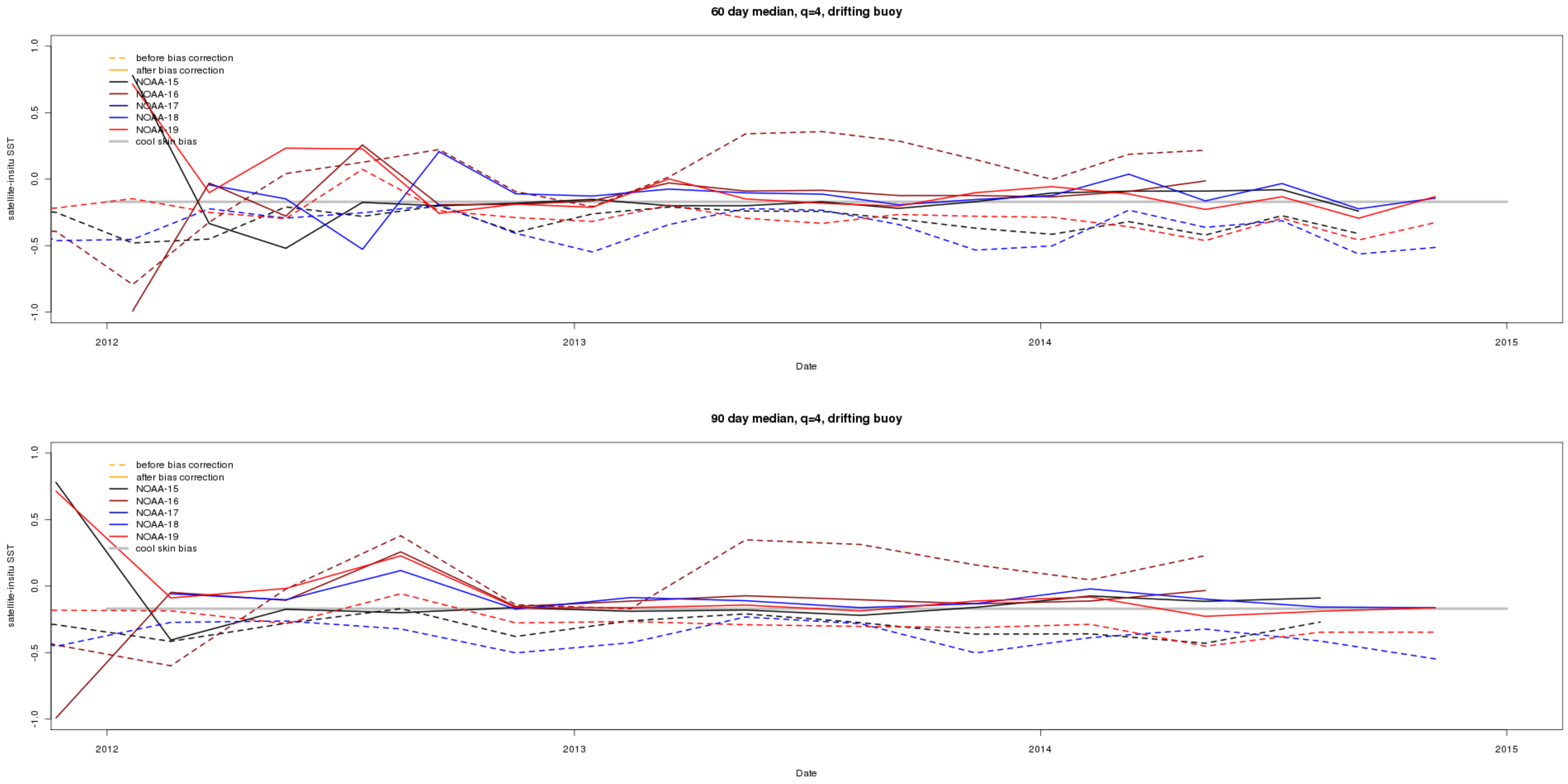


Figure 18: Impact of applying SSES bias correction (fv01 - lookup table approach with 60 day sample) to the median of $T_{i,\text{satellite}} - T_{i,\text{bias}} - T_{i,\text{insitu}}$ with 60 day and 90 day averaging windows. Dotted lines are before correction, solid lines are after correction.

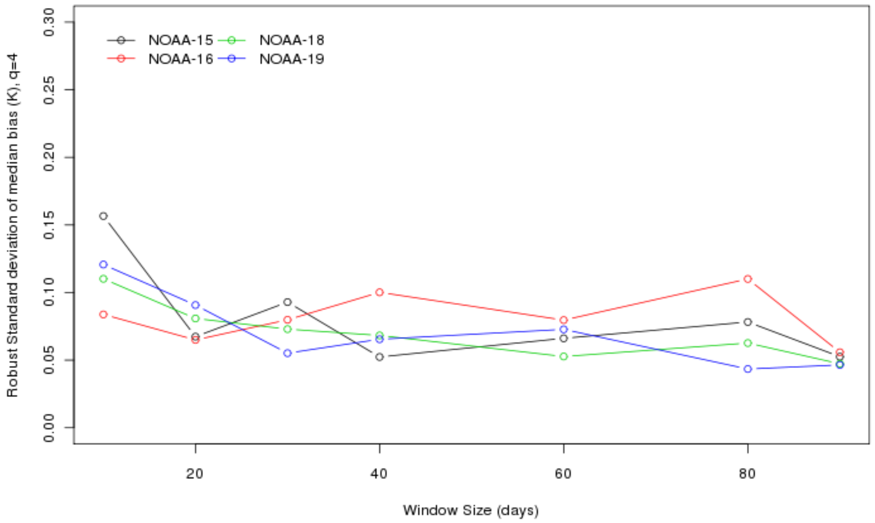


Figure 19: Impact of the window size can be assessed by computing the robust standard deviation of the difference between *in situ* and bias corrected SST, $T_{i,\mathrm{satellite}} - T_{i,\mathrm{bias}} - T_{i,\mathrm{insitu}}$, as a function of the window size, for lookup table based SSES on fixed retrieval system (fv01). The chart above corresponds to quality level 4 measurements, over the time period from 1st January 2012 to 31st December 2014, a period with reasonable coverage of matched data in the fv01 processing system.

In more concrete terms, over a 60 day period, the number of *in situ* matchups in each bin may be fewer than 20 or less, as shown in figure 21, which will ensure that all of the lookup table values exhibit large small-sample fluctuations and high sensitivity to outliers. Moreover, it is expected that the biases and standard deviations will most likely be slowly varying or smooth functions over the fundamental parameters, rather than show disjoint behaviour that is enforced by arbitrary choices in binning boundaries or large statistical fluctuations exasperated by a lack of measurements. It is desirable therefore to have an estimation method that will provide a smooth functional estimate, (employing the requirement of continuity to constrain any modelling) based on a larger sample which is quality controlled, but less rigidly so. Our belief is the combination of the requirement for smooth behaviour and the larger data set aids in reducing errors and compensates for less stringent quality control. Furthermore, a best estimate of the errors should include both swath dependent anomalies and geographical anomalies, essentially independently, and allow for slow time variation, which is more appropriate on the seasonal scale for geographical anomalies than for the satellite.

We consider an empirical model for the number of degrees of freedom, n, median bias, μ, and standard deviation, σ, which is seperable in swath $\{n_{\mathrm{swath}}, \mu_{\mathrm{swath}}, \sigma_{\mathrm{swath}}\}$, and geographical components $\{g_n, g_\mu, g_\sigma\}$, as follows,

$$\begin{aligned} n &= n_{\mathrm{swath}} g_n && (22)\\ \mu &= \mu_{\mathrm{swath}} + g_\mu && (23)\\ \sigma &= \max\left(\sigma_{\mathrm{swath}} g_\sigma, \sigma_0\right) && (24) \end{aligned}$$

We choose the median δ_i as the basis for our model, because the distribution of the difference between *in situ* and satellite measurements is generally asymmetric, with more cool satellite mea-

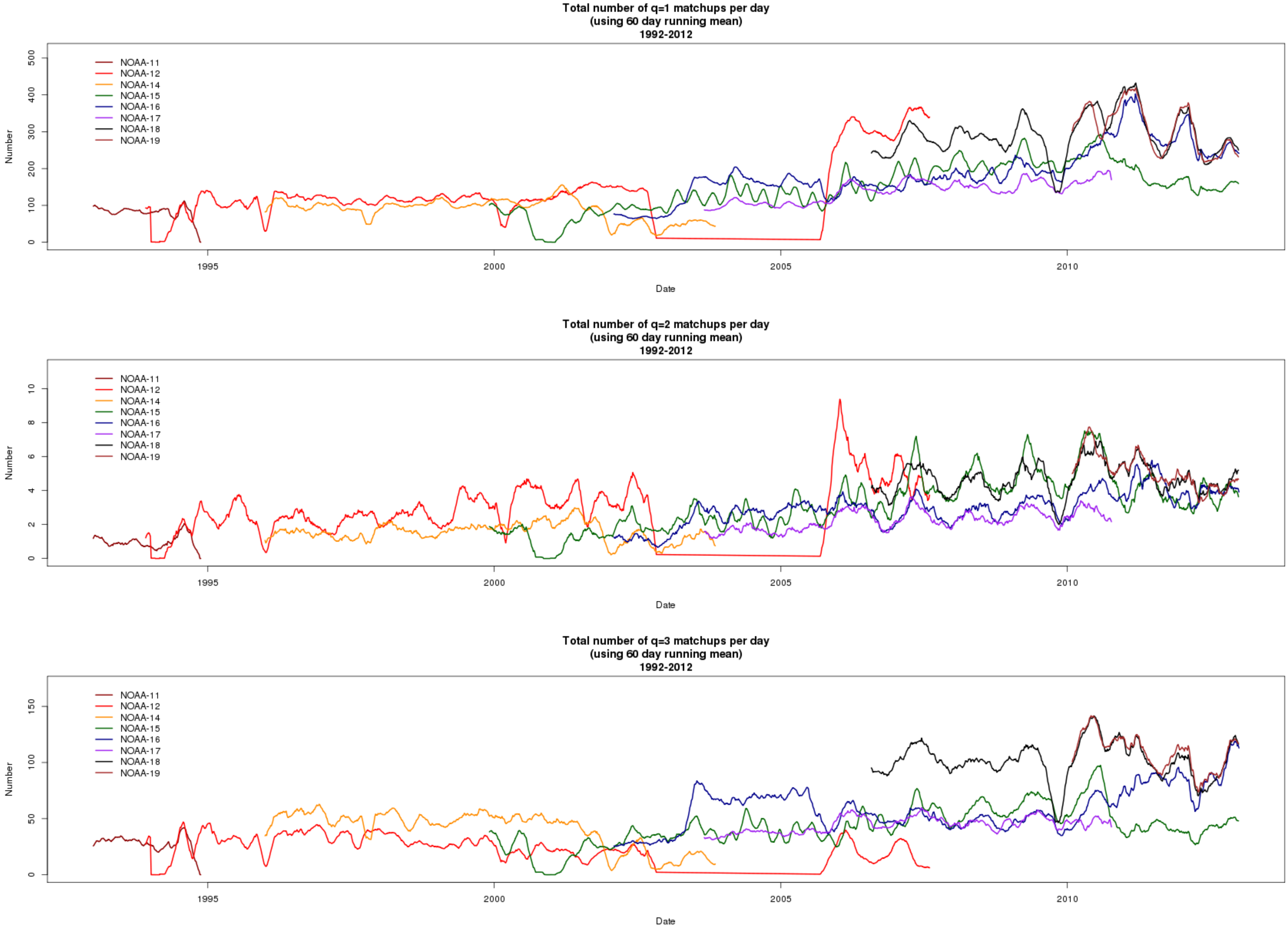


Figure 20: Typical number of *in situ* measurements per day, of various satellite SST quality, from all sources, for all target bins, before quality control, $q = \{1, 2, 3\}$. Based on a running 60 day mean.

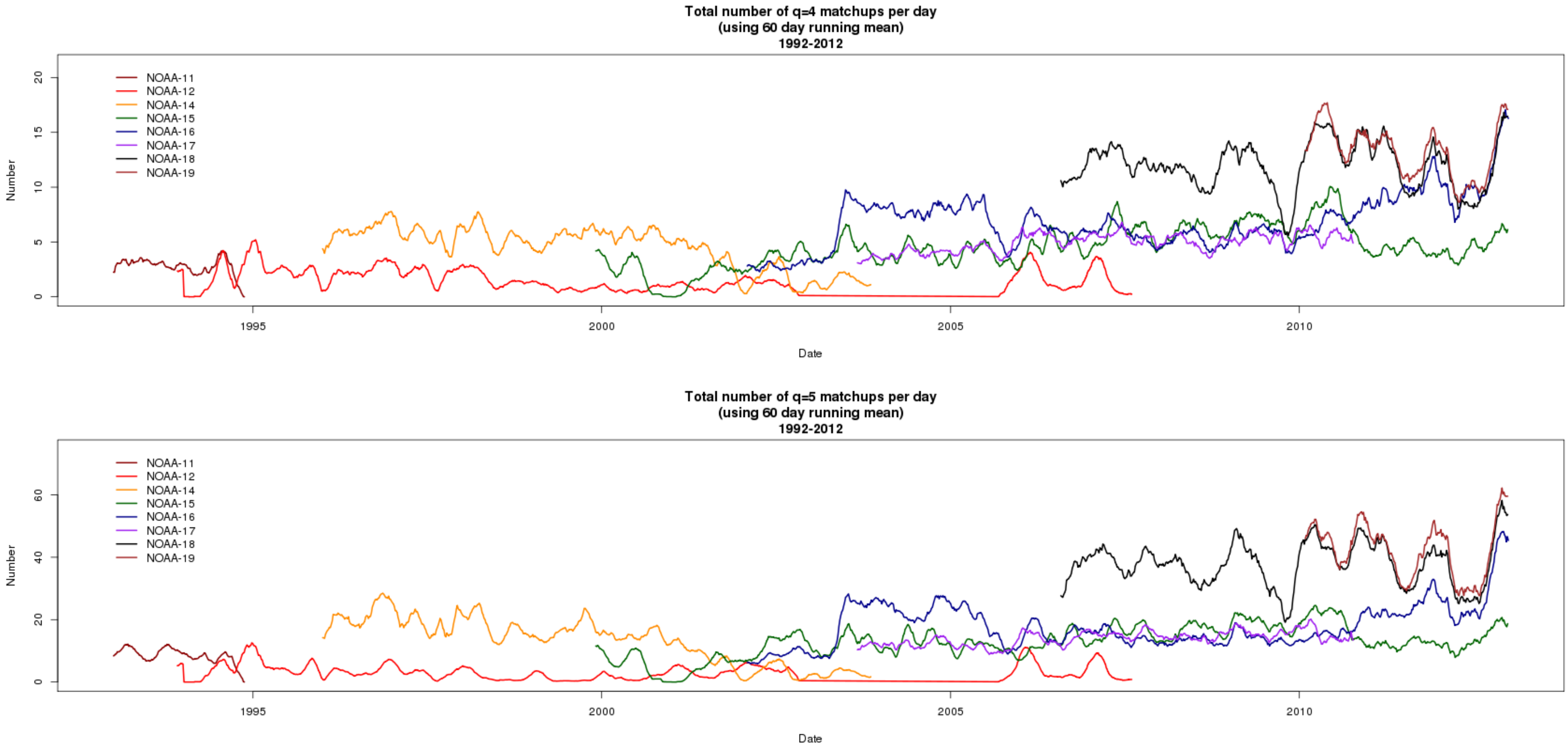


Figure 21: Typical number of *in situ* measurements per day, of various satellite SST quality, from all sources, for all target bins, before quality control, $q = 4, 5$. Based on a running 60 day mean.

surements which becomes more pronounced at lower quality level, due to errors in classification and detection of cloud. See figure 22 for typical distributions of the difference between *in situ* and satellite measurements, δ_i.

σ_0 is the minimum realistic uncertainty for a measurement. It is a fixed constant that reflects the best expected sensor performance. For AVHRR sensors we assume a value of 0.23K[1][15]

A significant cause of the uncertainty in AVHRR retrieval is poorly detected atmospheric contamination, such as cloud, where the temperature observed suffers from atmospheric interference rather than being representative of the sea surface. Targeting the median thus provides a more robust representative value that will be close to the mean for high quality pixels and less sensitive to contaminated pixels at lower quality level.

In our basic determination of $\{n_{\text{swath}}, \mu_{\text{swath}}, \sigma_{\text{swath}}\}$, we consider functional dependencies which depend on the first and second harmonics of the day night cycle, an interaction between the day / night cycle and the quality level, and variation over the satellite field of view, modelling systematic biases that relate to the field of view, using the three dimensions,

θ_z The satellite zenith angle at the point of observation. Angular dependence on the amount of atmosphere between the sensor and the sea depends on $(\sec\theta_z - 1)$, which is the standard form used to introduce this dimension.

θ_s The diurnal angle, or sun zenith angle at the point of observation, corrected so that angles prior to midday are negative. The first two harmonics of the diurnal cycle are represented by the four terms $\cos\theta_s$, $\sin\theta_s$, $\cos^2\theta_s$ and $\cos\theta_s\sin\theta_s$

q The quality level, defined as the distance to detected cloud, with a maximum of 5. Since the closer proximity to cloud results in a reduction of SST, q is introduced as a number between -1 (lowest quality) and 0 (best quality), using the normalized quality term, $\left(\frac{q}{5} - 1\right)$.

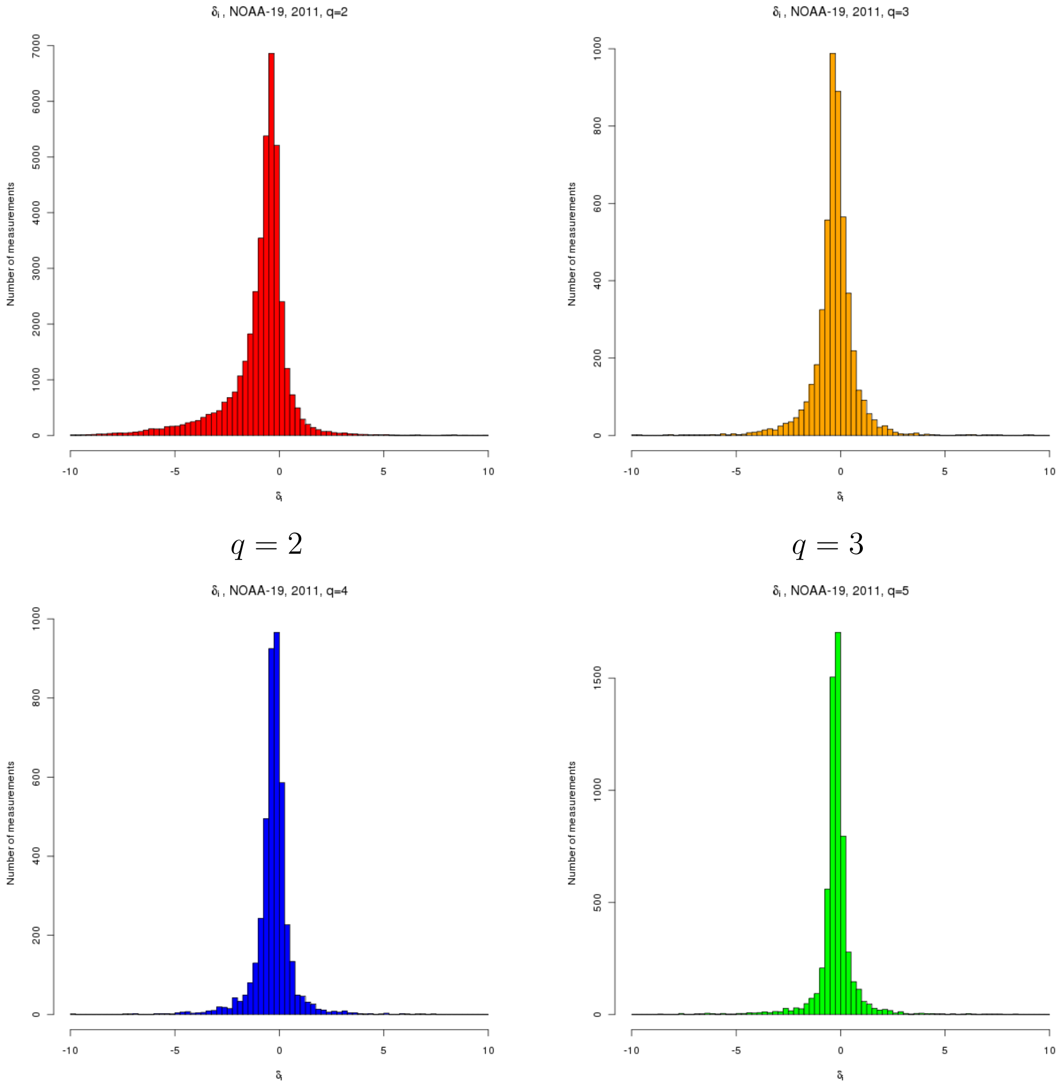
These pose a natural generalization to the binned approach described in the previous section, where the parameters have assumed dependencies on time of day, view angle and quality level. The advantage in this approach however is the possibility of smooth variation over the time of day and field of view. The precise nature of the terms to use were decided based on exploratory single variable correlation studies and principle component analysis.

This swath determination is corrected by $\{g_n, g_\mu, g_\sigma\}$, which represent interactions with latitude, longitude, quality and time referenced to the current time t_0 at which the model is considered optimal, and in the process introducing the following additional three additional dimensions in addition to the quality level,

$t - t_0$ The Julian date, in days, offset from the time at which the model is considered optimal which is usually the time at which the last *in situ* measurement was recorded. We assume that this applies linearly, and represents a very low frequency drift.

θ_{lat} The latitude. We choose a polynomial representation of this dimension, since it allows the coupling between this and other dimensions to be introduced more simply. Coupling between latitude and longitude can be modelled with a single $\theta_{\text{lat}}\phi_{\text{lon}}$ term, rather than worrying about two terms that consider the shift in the amplitude and phase of a harmonic function in *lat*, for example.

[1] In [15] several validation studies of the AVHRR sensor are discussed, indicating a typical standard deviation of $0.23K$ for buoy measurments[20] and a minimum standard deviation of $0.24K$ for AVHRR to radiometer validations[17].



$q = 2$ $q = 3$

$q = 4$ $q = 5$

Figure 22: Distribution of the difference between *in situ* and satellite measurements from all *in situ* sources, for different quality levels, where the difference is less than $10K$. Data is for NOAA-19, 2011, using the fixed regression SST retrieval discussed in section 8.7. NOAA-19 performance typifies the best performing platform over the entire time period. Cold tails can be clearly seen for $q = \{2, 3\}$. $q = 5$ has a slight skewed to warm side, which is likely related to the non-linear geographical terms used in retrieval model.

ϕ_{lon} The longitude. Since we are dealing with only a small section of the globe, and we wish to introduce dimensional coupling more simply, we use a polynomial representation of this dimension.

When time dependence is modelled, a relatively large set of historical data can be used in the analysis, permitting better statistical estimates, while weighting more recent measurements to allow sensitivity to recent trending behaviour. The initial period of the platform is linearly regressed with a time independent model, using linear least squares method, to ensure that the artefacts due to a reduced data set are minimized.

In its entirety, the model can be represented as shown in equations 25 to 30,

$$\begin{aligned}
\log\, n_{\text{swath}} &= a_0 + a_1 \cos\theta_s + a_2 \sin\theta_s + a_3 \cos\theta_s \sin\theta_s \\
&\quad + a_4\left(\frac{q}{5}-1\right)\cos\theta_s + a_5\left(\frac{q}{5}-1\right)\sin\theta_s + a_6 \cos^2\theta_s \\
&\quad + a_7(1 - e^{-(\sec\theta_z - 1)}) \qquad (25) \\
\mu_{\text{swath}} &= b_0 + b_1\left(\frac{q}{5}-1\right) + b_2(\sec\theta_z - 1) + b_3\cos\theta_s + b_4 \sin\theta_s \\
&\quad + b_5\left(\frac{q}{5}-1\right)(\sec\theta_z - 1) + b_6\left(\frac{q}{5}-1\right)\cos\theta_s + b_7\left(\frac{q}{5}-1\right)\sin\theta_s \\
&\quad + b_8\cos\theta_s \sin\theta_s + b_9(\sec\theta_z - 1)\cos\theta_s + b_{10}(\sec\theta_z - 1)\sin\theta_s \\
&\quad + b_{11}\left(\frac{q}{5}-1\right)^2 + b_{12}(\sec\theta_z - 1)^2 + b_{13}\cos^2\theta_s \qquad (26) \\
\sigma^2_{\text{swath}} &= c_0 + c_1\left(\frac{q}{5}-1\right) + c_2(\sec\theta_z - 1) + c_3\cos\theta_s + c_4 \sin\theta_s \\
&\quad + c_5\left(\frac{q}{5}-1\right)(\sec\theta_z - 1) + c_6\left(\frac{q}{5}-1\right)\cos\theta_s + c_7\left(\frac{q}{5}-1\right)\sin\theta_s \\
&\quad + c_8\cos\theta_s \sin\theta_s + c_9(\sec\theta_z - 1)\cos\theta_s + c_{10}(\sec\theta_z - 1)\sin\theta_s \\
&\quad + c_{11}\left(\frac{q}{5}-1\right)^2 + c_{12}(\sec\theta_z - 1)^2 + c_{13}\cos^2\theta_s \qquad (27)
\end{aligned}$$

$$\begin{aligned}
\log g_n &= \alpha_0 + \alpha_1\theta_{\text{lat}} + \alpha_2\phi_{\text{lon}} \\
&\quad + \alpha_3\theta^2_{\text{lat}} + \alpha_4\phi^2_{\text{lon}} + \alpha_5\theta_{\text{lat}}\phi_{\text{lon}} \\
&\quad + \alpha_6\theta_{\text{lat}}\phi^2_{\text{lon}} + \alpha_7\phi^3_{\text{lon}} + \alpha_8\theta_{\text{lat}}\phi^3_{\text{lon}} + \alpha_9\theta^2_{\text{lat}}\phi^2_{\text{lon}} \qquad (28) \\
g_\mu &= \beta_0 + \beta_1\theta_{\text{lat}} + \beta_2\left(\frac{q}{5}-1\right) + \beta_3(t_0 - t) + \beta_4(t_0 - t)\left(\frac{q}{5}-1\right) \\
&\quad + \beta_5\theta_{\text{lat}}\left(\frac{q}{5}-1\right) + \beta_6\theta_{\text{lat}}(t_0 - t) \\
&\quad + \beta_7\theta^2_{\text{lat}} + \beta_8\theta^2_{\text{lat}}\left(\frac{q}{5}-1\right) + \beta_9\theta^2_{\text{lat}}(t_0 - t) \qquad (29) \\
\log g_\sigma &= \gamma_0 + \gamma_1\theta_{\text{lat}} + \gamma_2\left(\frac{q}{5}-1\right) + \gamma_3(t_0 - t) + \gamma_4(t_0 - t)\left(\frac{q}{5}-1\right) \\
&\quad + \gamma_5\theta_{\text{lat}}\left(\frac{q}{5}-1\right) + \gamma_6\theta_{\text{lat}}(t_0 - t) \\
&\quad + \gamma_7\theta^2_{\text{lat}} + \gamma_8\theta^2_{\text{lat}}\left(\frac{q}{5}-1\right) + \gamma_9\theta^2_{\text{lat}}(t_0 - t) \qquad (30)
\end{aligned}$$

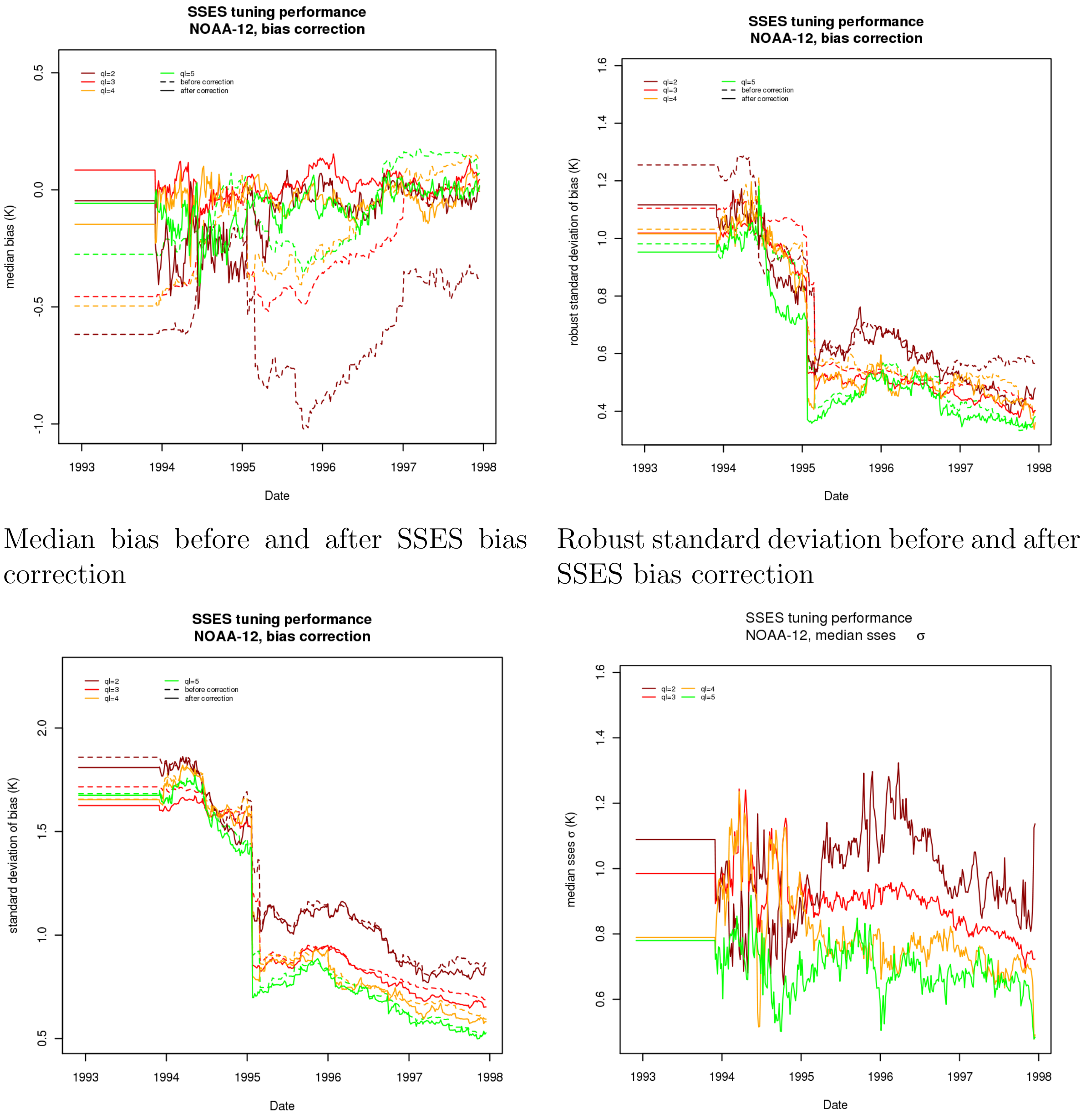
The parameters $\{a_i, b_i, c_i, \ldots \alpha_i, \beta_i\gamma_i \ldots\}$ (we use lower case English letters for swath fitting, and Greek letters for geographical fitting), are fitted progressively using standard linear least squares methods, with different weight schemes, in an attempt to provide a consistent functional form which rationalizes specific expected behaviours of the parameters and reduces the impact of outliers on the fitting process which assumes Gaussian residuals. The term $(1 - e^{-(\sec\theta_z - 1)})$ was determined based on exploratory studies of single variable correlations and the dominance of $(\sec\theta_z - 1)$ as a source, together with a more detailed investigation of the functional dependence of the correlation. See section 7 for further details of the SSES fitting algorithm. In the geographical fit for g_n, cubic and higher terms in latitude are removed since they tend to drive rapid divergence in the empirical relationship.

In its entirety the model has 66 free parameters, less than half the number in the binning approach outlined in section 2.2 (if we include degrees of freedom in our binning model we have effectively 144 parameters for the binning approach).

The number of parameters can be further reduced by ignoring the time based parameters and running a static model which is frequently updated, if required. During the initial period of each satellite record where less measurements are available, this method is used, and the model is fitted on *in situ* measurements made and matched after the satellite pass. Where the time is included as a model variable, the model parameters are fixed by historical data, then assumed persistent for a short time afterwards, over which SSES are estimated (there is no use made of future measurements for the current processing).

2.3.1 Model fitting performance

During the model fitting process, the effectiveness of the model can be evaluated by applying the model to the *in situ* to satellite matches, and considering the median residual as well as the change in standard deviation, after the bias correction is made. If the bias correction is good, there should be a reduction in standard deviation, and a shift in median bias towards zero. Figure 23 shows the result of using the fitted SSES for bias correction to the SST at the point of fitting the model for NOAA-12, at all quality levels. NOAA-12 was chosen for illustration because mission extended over a time period where the number of *in situ* measurements changed dramatically, from a regime which was dominated by less accurate coastal moored buoys to one that was dominated by drifting buoys. The initial constant values reflect the use of the first years data for the first year of observations, for the first year, SST matches from the future are used to tune the model, however after the first year, a rolling historical one year window is used. The plots show dotted lines for biases and standard deviations without bias correction, and solid lines indicate biases and standard deviations with bias correction. The bias model adds a clear correction to the median bias for all quality levels, over time, with especially large contributions in the lower quality biases. The robust standard deviation of the bias shows modest improvements throughout, even for good quality observations. The standard deviation appears to show a smaller improvement, however this is dominated by outlying measurements to a greater extent than the robust standard deviation. Finally the median SSES standard deviation is less erratic at later times when the *in situ* measurements are more stable, but varies in nominally the same range, indicating, at least broadly, that the satellite sensor variability has remained somewhat stable. Referring to table 13, the Fitting fv02 data set is used to fit the SSES model, whereas the larger Verification fv02 data set is used for the verification that is shown in figure 23.



Median bias before and after SSES bias correction

Robust standard deviation before and after SSES bias correction

Standard deviation before and after SSES bias correction

Median SSES standard deviation

Figure 23: NOAA-12. The impact of modelled SSES, 1993 to 1998, updated every 5 days, as an example or how the correction makes some improvement to early satellite retrievals.

2.3.2 Validation of SSES - an empirical model approach - fv02

SSES bias correction is applied by subtracting the estimated bias from the SST. The estimated bias is computed from a model tuned on historically recent data. In operational fv02 systems, the definition of historically recent is 1 year prior to the day before the day that the satellite measurement was made as a UTC date. The impact of the correction can be seen by comparing valid, low wind, *in situ* sea surface temperature with corrected, and uncorrected SST.

Figures 25 and 27 shows the impact of applying the bias correction to L2P measurements, at various quality levels.

The corrected SST (solid line) are clearly closer to the expected systematic cool skin bias than the uncorrected SST for all platforms over the period for which the method was applied. Robust standard deviations are not greatly affected by the change, which in shows that this approach adds little noise to the corrected SST, and in many cases there is a slight improvement.

Due to the statistical nature of this evaluation, we need to choose an aggregating window that is appropriate to reduce sources of noise. The robust standard deviation of the variation of the median of $T_{i,\text{satellite}} - T_{i,\text{bias}} - T_{i,\text{insitu}}$ as a function of the aggregating window size, for quality level 4, is shown in figure 28 for more recent platforms.

When sampling from any distribution, the standard error decreases as sample sizes increase, the robust standard deviation of $T_{i,\text{satellite}} - T_{i,\text{bias}} - T_{i,\text{insitu}}$ will decrease as the window period is increased (which corresponds to larger samples), up to window time periods that remain smaller than characteristic stability times of $T_{i,\text{satellite}} - T_{i,\text{bias}} - T_{i,\text{insitu}}$. This trend is broadly observed in the NOAA-18 data and to a lesser extent in NOAA-19 data. NOAA-15, 18 and 19, are relatively immune to window size changes (so long as the window is bigger than 20 days), with NOAA-16 perhaps showing a trending increase with window size consistent with the notion of continuous instability.

The chart shows that a window period of the order of 30 days seems like an adequate trade off between having frequent assessments, and accurate assessments. In addition, the general trend is a decrease in robust standard deviation with window size, which shows that the data source instabilities are able to be accounted for in the SST retrieval and SSES estimation process.

2.4 Comparing fv01 to fv02

While it is difficult to compare the empirical and table based SSES models directly, since the two data streams that use them have different SST retrieval methods - fv01 (which uses the table estimation method) uses a fixed retrieval, whereas fv02 (the empirical model method) is adaptive - it is clear in this analysis that over the similar time periods considered, the fv02 algorithm combination provides superior performance to fv01 combination. The lower values of robust standard deviation of median bias over all sampling time windows indicates that the variations are better handled with the modelling approach. In addition the decrease of robust standard deviation of median bias with window time period is consistent with expectations of stability, which comes from a combination of an adaptive SST retrieval with an adaptive modelled SSES estimate.

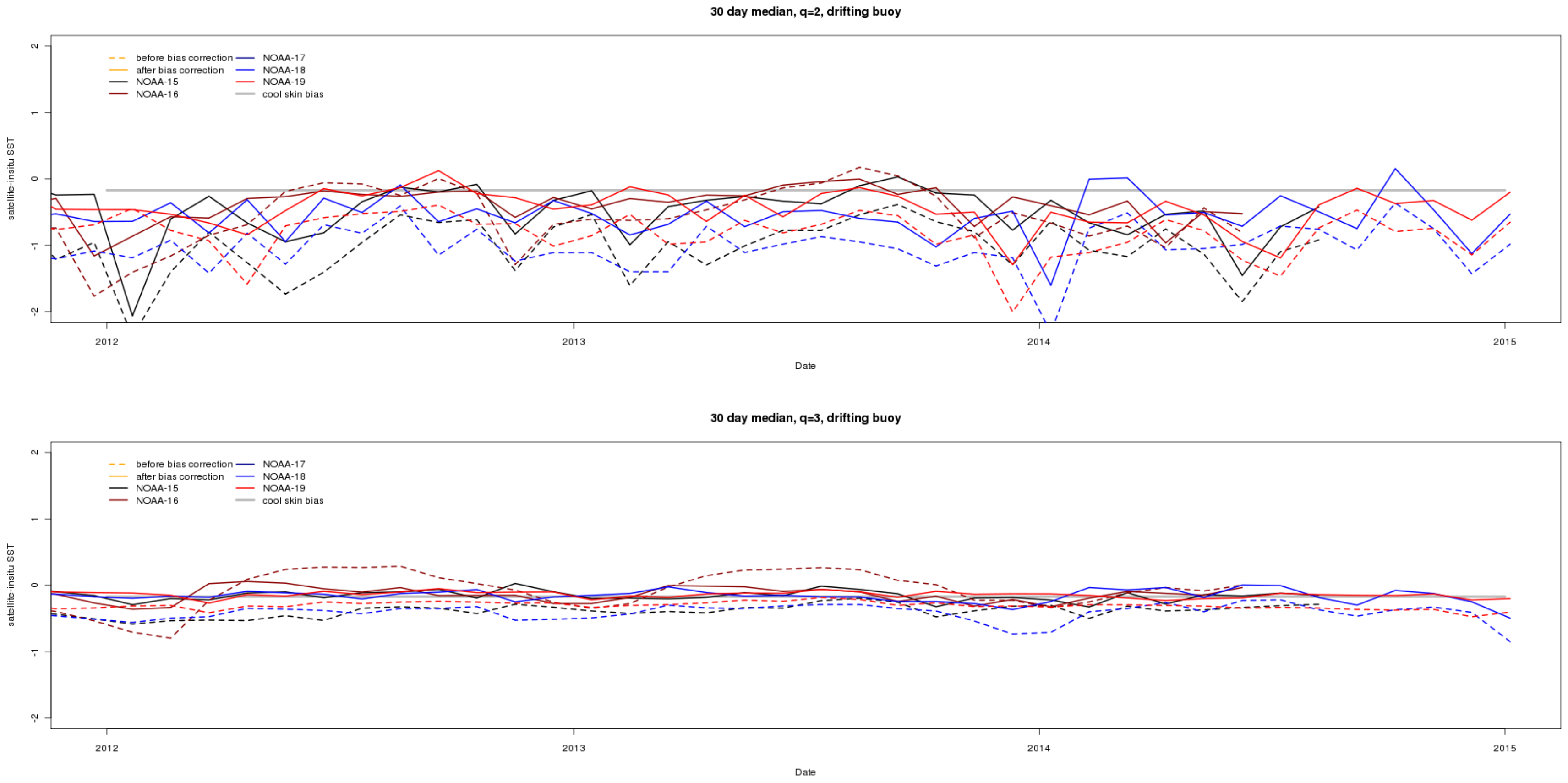


Figure 24: Impact of applying SSES bias correction (fv02 - empirical model approach) to the 30 day median of $T_{i,\text{satellite}} - T_{i,\text{insitu}}$, $q = 2, 3$. Dotted lines are before correction, solid lines are after correction.

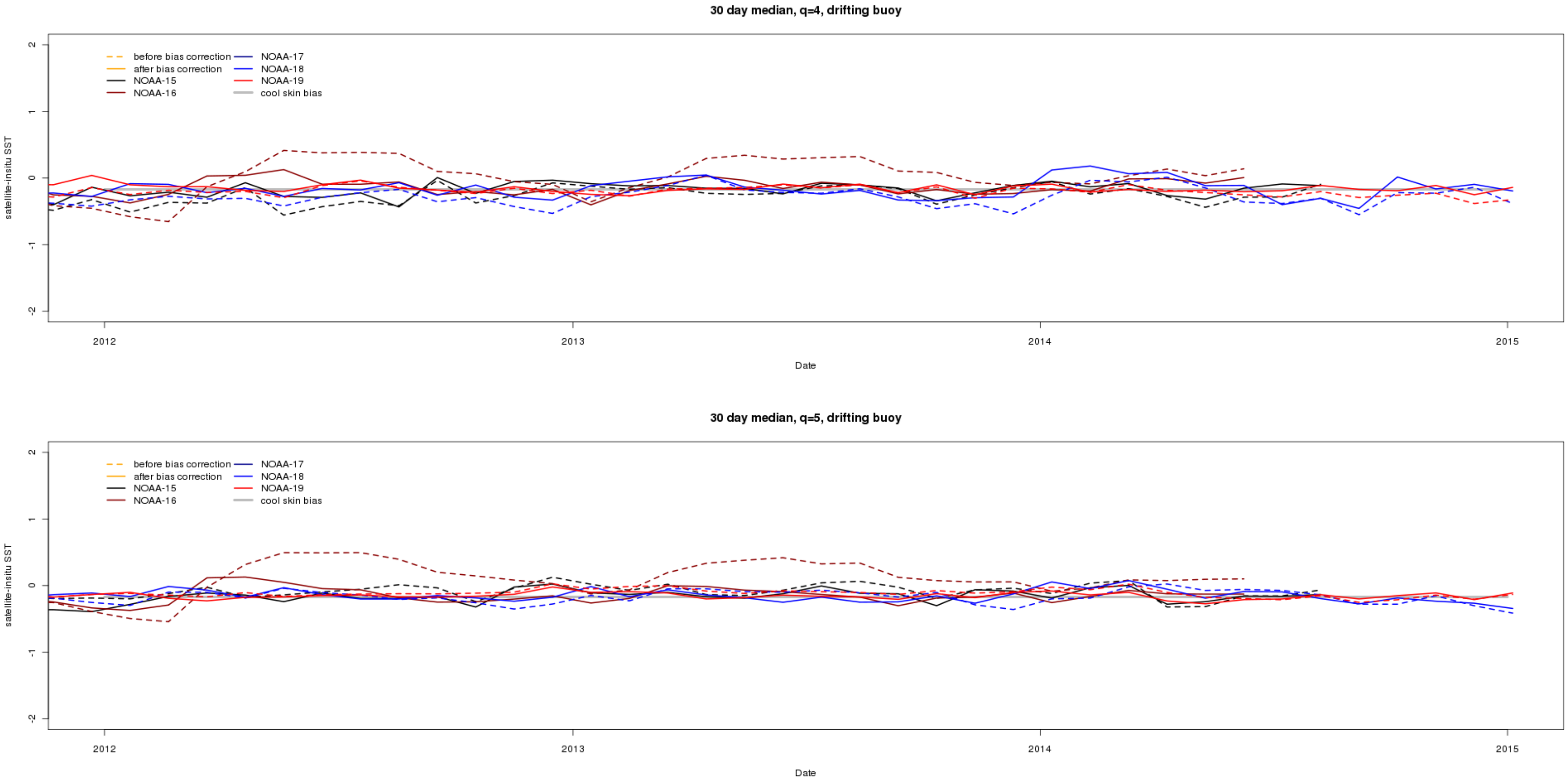


Figure 25: Impact of applying SSES bias correction (fv02 - empirical model approach) to the 30 day median of $T_{i,\text{satellite}} - T_{i,\text{insitu}}$, $q = 4, 5$. Dotted lines are before correction, solid lines are after correction.

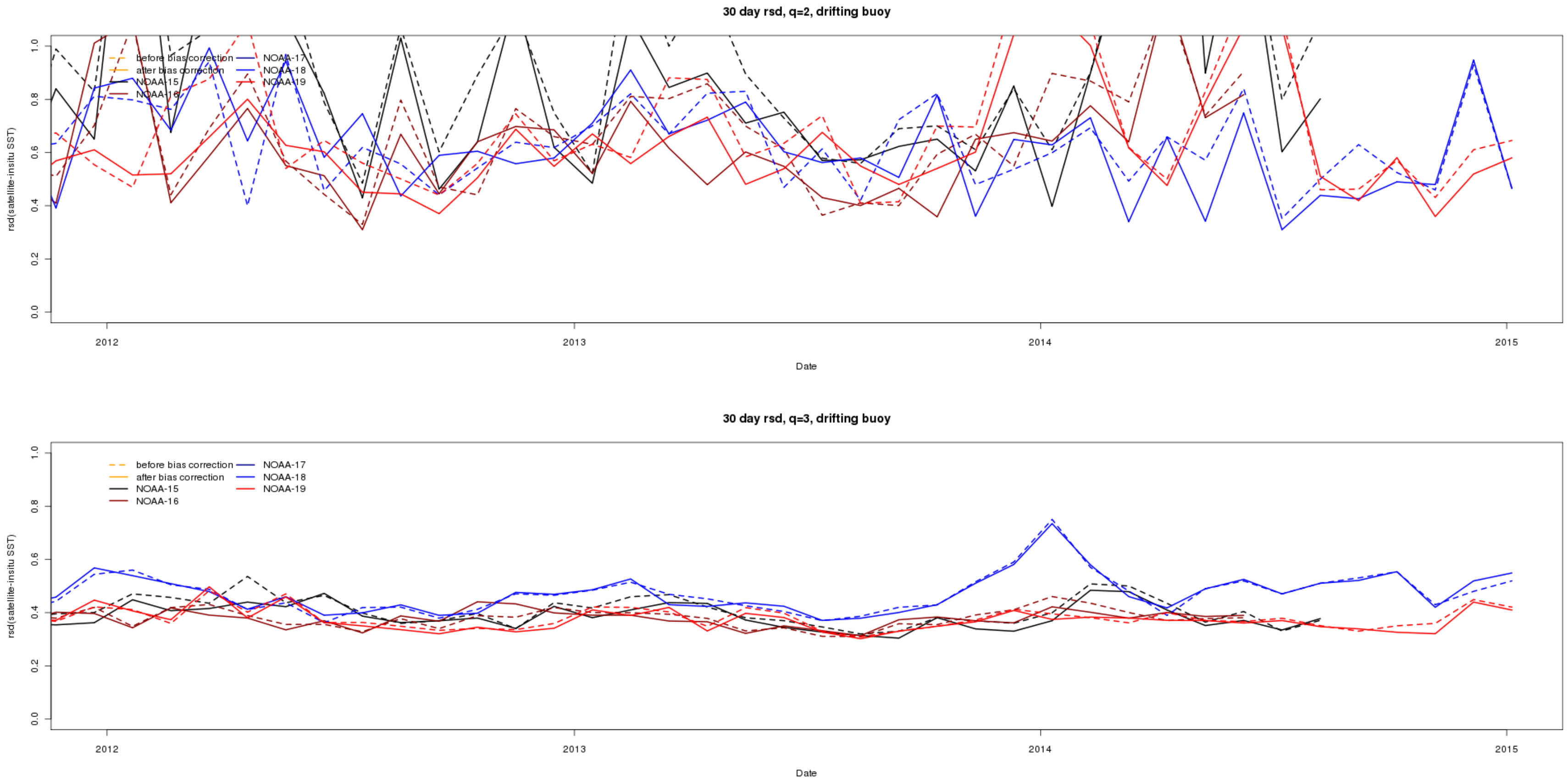


Figure 26: Impact of applying SSES bias correction (fv02 - empirical model approach) to the 30 day rsd of $T_{i,\mathrm{satellite}} - T_{i,\mathrm{insitu}}$, $q = 2, 3$. Dotted lines are before correction, solid lines are after correction.

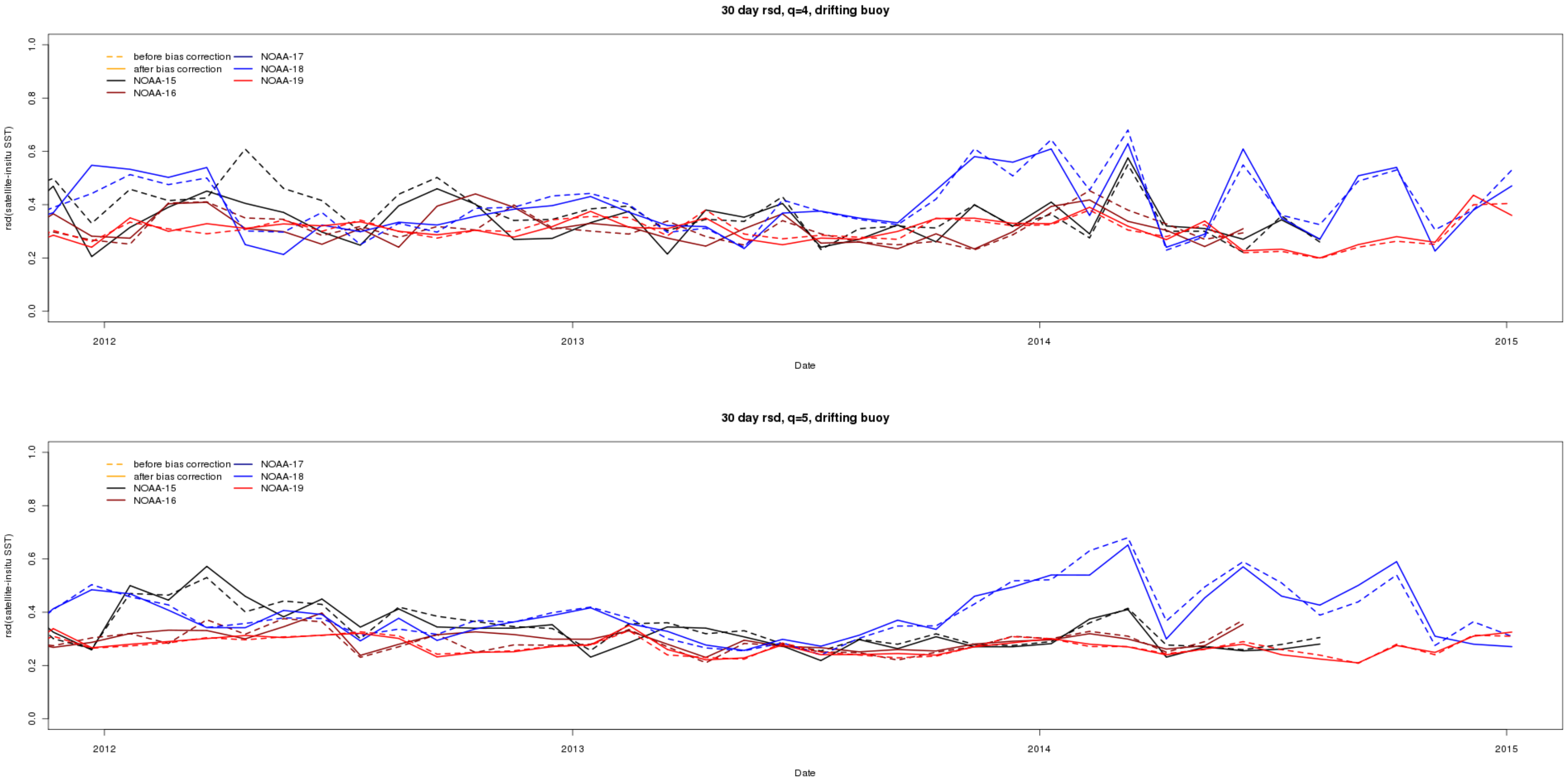


Figure 27: Impact of applying SSES bias correction (fv02 - empirical model approach) to the 30 day rsd of $T_{i,\text{satellite}} - T_{i,\text{insitu}}$, $q = 4, 5$. Dotted lines are before correction, solid lines are after correction.

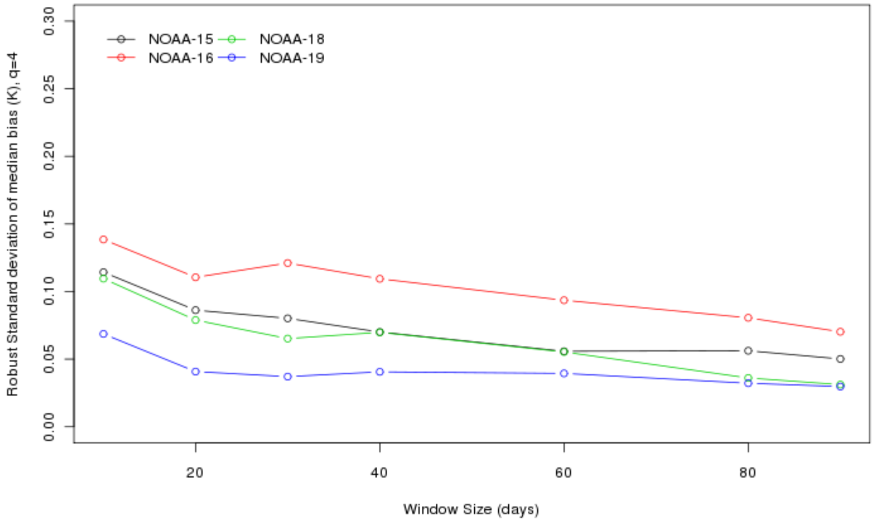


Figure 28: Impact of the window size can be assessed by computing the robust standard deviation of the difference between *in situ* and bias corrected SST, $T_{i,\mathrm{satellite}} - T_{i,\mathrm{bias}} - T_{i,\mathrm{insitu}}$ as a function of the window size, for empirical model based SSES assessment on dynamic SST retrieval (fv02). The chart above corresponds to quality level 4 measurements, over the time period from 1st January 2012 to 31st December 2014.

3 Quality control of matches between *in situ* and satellite SST

Our standard approaches to SST retrieval and SSES estimation require matching satellite with *in situ* measurements. This process involves first choosing an *in situ* data set, and then deciding which subset of *in situ* measurements are of sufficient quality and usefulness for the task at hand. The process involves removing measurements that have a higher chance of negatively influencing the task at end, either because of clear errors in reporting or measuring, or because they are too atypical.

3.1 Selection criteria for favourable conditions for *in situ* to satellite SST matches

Lookup table based SSES determination requires a sample of best *in situ* to satellite matches which at the same time provides representative variation, such that standard deviations can be estimated. In order to achieve this, it is desirable to filter measurements such that outliers are removed. The following criteria are used to quality control such matches.

- The *in situ* measurement must be in the field of view, and from drifting buoys only (not moored buoys, not ship or animal based measurements, and not from Argo).
- The *in situ* buoy is not on the Météo-France buoy blacklist.[8] (The list is updated as issued by Météo-France).

- δ_i should be in the range $-3K$ to $3K$, to ensure that anomalous measurements do not inappropriately bias error estimates. Measurements which show a large deviation are biased to be either cloud affected $\delta_i < 0$, from areas of large thermal gradients near the surface due to very calm seas (for example), or high diurnal variability areas $\delta_i > 0$. Cloud affected variation has the potential of being much larger than diurnal variation (which most of the time will fall within $\pm 3K$. Uncertainties arising from these physical effects are in many cases not a direct result of measurement error, and can be comparatively large. To remove possible sources of bias that may arise, these measurements are removed.[**paltaglou**],

- The wind speed should be in the range 6ms^{-1} to 20ms^{-1} during the day or 2ms^{-1} to 20ms^{-1} at night, to ensure that there is adequate mixing between the ocean surface and the location where the *in situ* measurements are taken, after Donlan *et al*[6, 10].

- The difference between SST and analysis SST from the previous day should be in the range $-3K$ to $3K$, to ensure that high diurnal warming events do not inappropriately bias error estimates.

- The distance between the nominal SST measurement location and the buoy location (based on latitude and longitude supplied by each measurement device) should be in the range 0 to 2km, to ensure that the *in situ* measurement is geographically appropriate.

- The absolute time difference between the satellite and buoy measurements should be in the range 0 to 60min, to ensure that the *in situ* measurement is temporally appropriate.

3.2 Selection of *in situ* data sources

The *in situ* data sets used for retrieval and SSES model calibration are chosen such that a respectable number of *in situ* measurements are available.After a review of the number of measurements available over time, and a desire to have a dataset that was able to be regressed against SST every month against a running window of *in situ* measurements, the following criteria were developed.

After pre-screening for quality (we choose matches between satellite and *in situ* measurements where the satellite based proximity to cloud quality level is greater than or equal to 5km, corresponding to assigned quality level $q = 5$), the following set of rules is applied sequentially to determine the measurement set to be used,

- If there are less than 100 drifting buoy measurements over the time period of interest, use the ship and moored buoy measurements in their entirety, with no drifting buoy measurements.

- If there are greater than 5 drifting buoy measurements per day and the time coverage for drifting buoy measurements exceeds 65% of the time period of interest, use the drifting buoy measurements entirely.

- If there are fewer than 5 drifting buoy measurements per day or the time coverage for drifting buoy measurements is less than 65% of the time period of interest, and the drifting buoy time coverage is less than 80% of the moored buoy time coverage, use both drifting buoy and moored buoy measurements.

- If there are fewer than 5 drifting buoy measurements per day or the time coverage for drifting buoy measurements is less than 65% of the time period of interest, and the drifting buoy

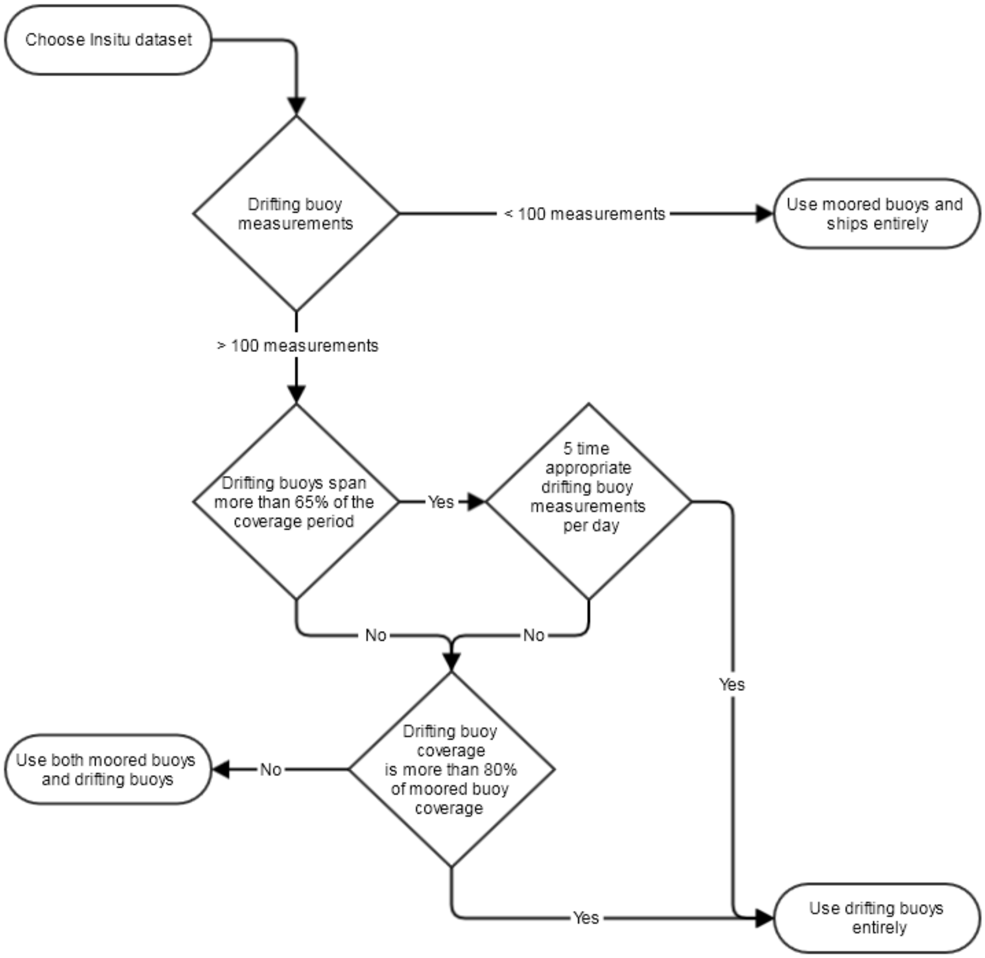

Choose Insitu dataset
Drifting buoy measurements
< 100 measurements
Use moored buoys and ships entirely
> 100 measurements
Drifting buoys span more than 65% of the coverage period
Yes
5 time appropriate drifting buoy measurements per day
No
No
Yes
Use both moored buoys and drifting buoys
No
Drifting buoy coverage is more than 80% of moored buoy coverage
Yes
Use drifting buoys entirely

Figure 29: Choice of data set for model regression follows this flow chart applied to matches between satellite and *in situ* that have $q \geq 2$.

time coverage is greater than 80% of the moored buoy time coverage, use the drifting buoy measurements.

This procedure is summarized on a flowchart in figure 29. The proportions 65% and 80% were investigated and the selection algorithm was determined not to be sensitive to small changes in these parameters.

The choices of a minimum 5 measurements a day was chosen as the smallest data quantity which could yield useful information about the distribution of the data on a daily basis.

Where fewer than 100 measurements of drifting buoys were available, the drifting buoy measurements that are available tend to be more highly temporally and spatially biased than the available moored buoys over the same time period. It thus was deemed safer to mix the moored with drifting buoy data in such cases.

3.3 LatQC: Quality control of *in situ* Temperature measurements

Suitable matches between satellite and *in situ* are determined by matching spatially and temporally related measurements of *in situ* SST and the corresponding satellite SST. The data sets are chosen based on the rules outlined in section 3.2. It is possible that *in situ* SST measurements are inaccurate due to the way that they are measured, or come from defective measurement devices, and these will provide spurious matches, which contribute outliers to the matches between satellite and *in situ* measurements. Since our retrieval and SSES models rely on a clean data set, we provide a simple quality control to these potential outliers by looking at the *in situ* SST deviation from a typical value for that *in situ* at the given location and time.

The typical *in situ* SST is modelled by a simple harmonic latitude dependency, plus a harmonic bias with annual period, $t = \{0 \dots 1\}$,

$$T_{insitu,\mathrm{model}} = \eta_0 + \eta_1 \sin 2\pi t + \eta_2 \cos 2\pi t + \eta_3 \sin^2\left(\frac{\pi}{180}\,\theta_{\mathrm{lat}}\right) \tag{31}$$

This model was determined to set the base scale for the measurement, rather than provide a precise geophysical model. The model embodies the crude observation that the sea is warmer near the equator and on average may have an annual seasonal cycle over the southern hemisphere (since most of our observations cover the southern hemisphere and the northern tropics), which is approximately fixed in terms of phase and amplitude.

When the *in situ* measurement deviates from $T_{insitu,\mathrm{model}}$ by more than a fixed amount, (typically of the order of $\sim$ 5K, but in practise based on the distribution of the actual *in situ* measurements, as is outlined below), we can classify the *in situ* measurements are being acceptable (small deviation) or abnormal (large deviation), without any further input other than the *in situ* dataset.

Figure 30 shows a typical fit for *in situ* SST which have matches to Australian reception NOAA-12 data, over two years prior to December 1st 1995. The blue area represents the result of the application of the model, while black circles correspond to *in situ* measurements that are well modelled. Red circles near the periphery are flagged as abnormal due to an extreme measurement, and are thus not used in the SST process. Notice that cold measurements at lower latitudes are effectively removed, as are very warm mid-latitude measurements.

Regressing this model with a two year rolling window on a monthly basis over a longer period of time is demonstrated in figure 31, which shows the equatorial baseline η_0, annual variation η_1, hemispherical variation η_3 and seasonal phase $\arctan\left(\frac{\eta_3}{\eta_2}\right)$, of *in situ* matched to NOAA-12 in a rolling two year window over the period from 1993 to 2007 (the period from 2005-2007 was produced on a single model construction due to a greater than 2 year break in reception of NOAA-12).

Regressing monthly over the entire historical data set for matches to all platforms, with a two year running window, is shown in figure 33. Although there is broad general agreement, the magnitude of the seasonal fluctuation shows some discrepancy from platform to platform which could be accounted by the differences and evolution of equator crossing times for platforms considered. This information may form a basis for determining diurnal or other biases in the *in situ* data sets used on each platform, although the details of this is beyond the scope of this work.

It is worth emphasizing one more time that the data presented herein is completely *in situ* in nature. The satellite dependency was purely on the basis that it had a clear view to the ocean at the same time (within 60 minutes) and place (within 2km), and so this purely reflects any biases that may exist in the use of *in situ* dataset matches each platform.

The deviation between *in situ* measurements and the model is determined,

$$\Delta T_{i,\mathrm{I,m}} = T_{\mathrm{insitu}} - T_{\mathrm{insitu,model}} \tag{32}$$

We accept measurements based on $\Delta T_{i,\mathrm{I,m}}$ falling within 4.5 standard deviations, $\sigma_{\mathrm{I,m}}$, of the median $\Delta T_{i,\mathrm{I,m}}$,

$$S_{\theta_{\mathrm{lat}},\mathrm{ok}} = \{i : |\Delta T_{i,\mathrm{I,m}} - \mathrm{median_j}(\Delta T_{i,\mathrm{I,m}})| < 4.5\,\sigma_{\mathrm{I,m}}\} \tag{33}$$

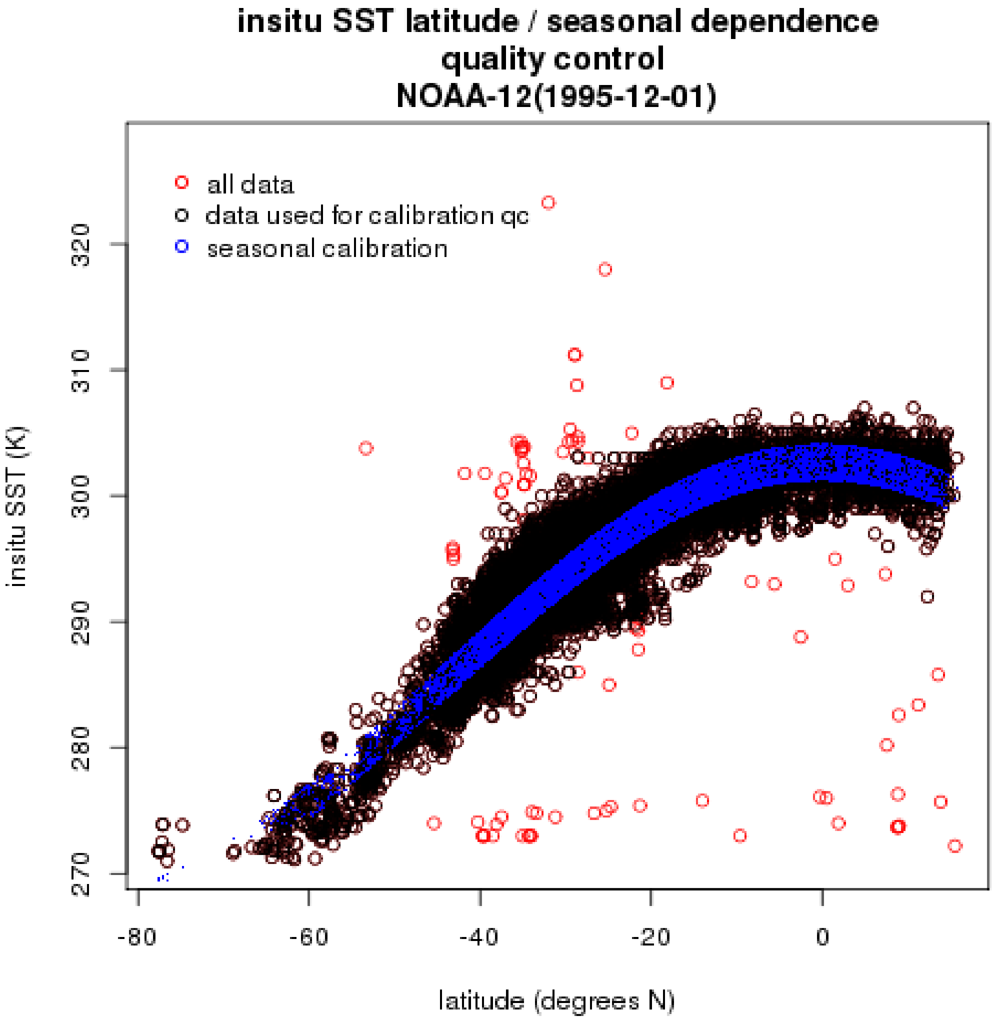
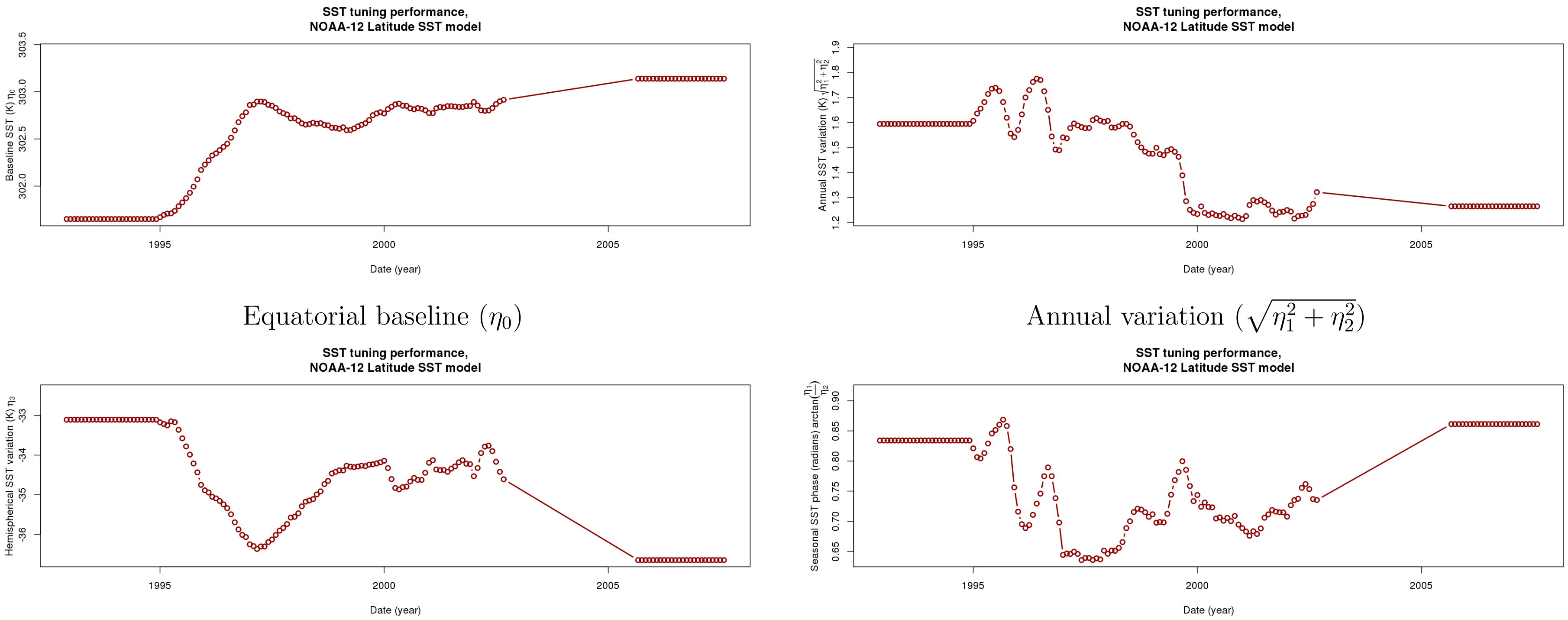


Figure 30: *in situ* measurement latitude model, *in situ* SST matched to NOAA-12 observations for 2 years prior to December 1st 1995. Actual measurements are shown in black and red, whereas model values are blue. The red data are removed due to excessive deviation from the model.



Equatorial baseline (η_0)

Annual variation ($\sqrt{\eta_1^2 + \eta_2^2}$)

Hemispherical variation (η_3)

Seasonal phase ($\arctan\left(\frac{\eta_1}{\eta_2}\right)$)

Figure 31: *in situ* measurement latitude model. Evolution of parameters, for *in situ* within the NOAA-12 field of view, 1993 to 2007. Prior to 1995 and after 2003 data is unavailable or missing for large periods of time, and the estimates derived are considered constant throughout these periods.

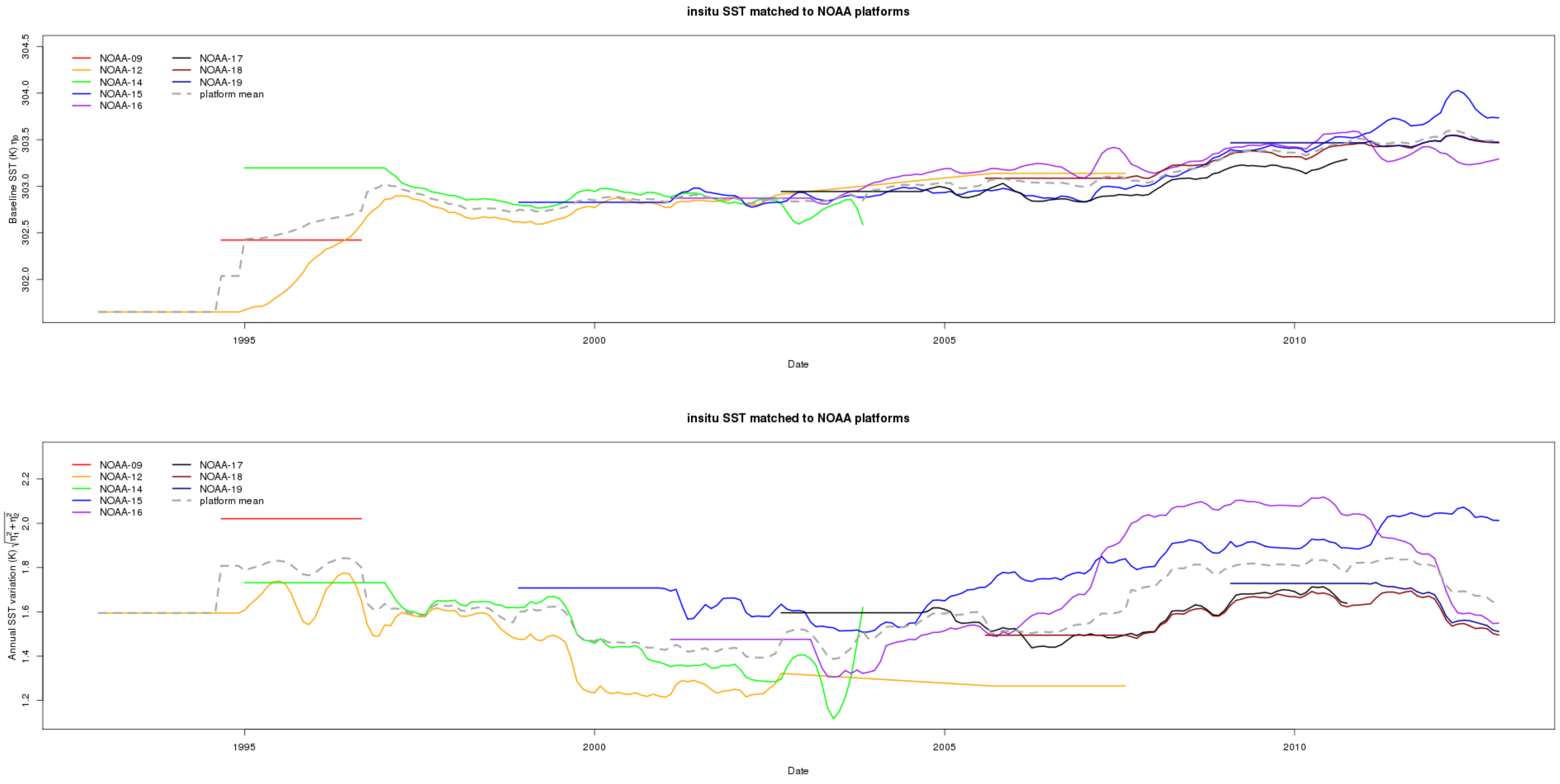


Figure 32: *In situ* measurement latitude model. Evolution of parameters, all observations, 1994 to 2013

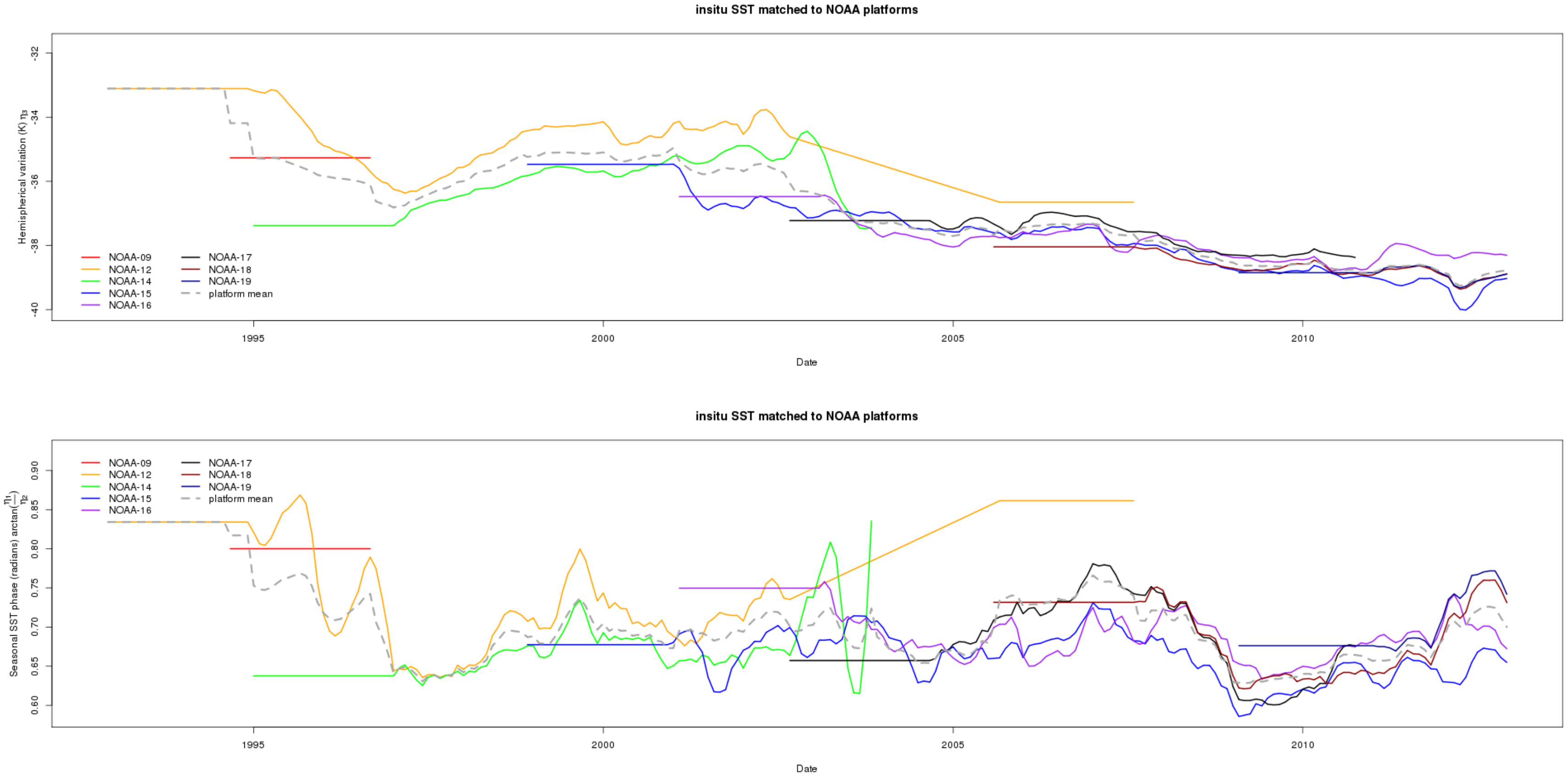


Figure 33: *In situ* measurement latitude model. Evolution of parameters, all observations, 1994 to 2013

which by Chebyshev's theorem[2], considering $\sigma_{\mathrm{I,m}}$ estimated from the dataset itself, will result in the exclusion of no more than approximately 5% of the available data, irrespective of the distribution, and negligible exclusion if $\Delta T_{i,\mathrm{I,m}}$ is distributed normally.

$\sigma_{\mathrm{I,m}}$ is computed based on a range of $\Delta T_{\mathrm{I,m}}$ defined by the 1st and 99th percentile, highest quality satellite observation ($4 \leq q_i \leq 5$) limits.

$$S_{\theta_{\mathrm{lat}}} = \{i : P_1(\{\Delta T_{j,\mathrm{I,m}} : 4 \leq q_j \leq 5\}) < \Delta T_{i,\mathrm{I,m}} < P_{99}(\{\Delta T_{j,\mathrm{I,m}} : 4 \leq q_j \leq 5\})\} \tag{34}$$

$$\sigma^2_{\mathrm{I,m}} = \mathrm{var}_{i \in S_{\theta_{\mathrm{lat}}}}(\Delta T_{i,\mathrm{I,m}}) \tag{35}$$

where $P_i(A)$ is the i^{th} percentile of the set A, and $\mathrm{var}_{i\in A}(B_i)$ is the population variance,

$$\mathrm{var}_{i\in A}(B_i) = \frac{\sum_{i\in A}\left(B_i - \mathrm{mean}_{j\in A}(B_j)\right)^2}{\mathrm{count}(A)} \tag{36}$$

This quality control method is self referential and does not require input from external sources or other temperature measurements that could be themselves derived from the *in situ* measurements. It thus serves as an open-loop filtering process, which results in an outlier filtered collection of measurements.

This model works reasonably well over a single hemisphere, however if the data spans both hemispheres, a latitudinal model based on a polynominal of quartic or higher order in θ_{lat} provides a more robust model for *in situ* selection which can accommodate differences in the measurement density in the north and south without sacrificing accuracy of fit using ordinary least squares,

$$T_{insitu,\mathrm{model}} = P_{\{4,5,6,\ldots\}}\left(\frac{\theta_{\mathrm{lat}}}{90}\right) \tag{37}$$

3.4 L4QC: Quality control of *in situ* measurements using level 4 analysis SST

The goal of quality control of *in situ* measurements using analysis SST is to reduce the list of available measurements into a set of reasonable measurements $S_{4,\mathrm{ok}}$, which have a higher degree of certainty of being geophysically accurate, under the assumption that the level 4 reanalysis (typically foundation SST) provides a better estimate of the expected SST than the crude model considered in section 3.2, and it does so without applying potentially adverse biases to the data set. Outliers are likely to include measurements that are subject to large amounts of diurnal variation, varying ocean conditions and problems with cloud clearing, and although the measurements may be good, for the purpose of model determination, an inaccurate result which may be detrimental to the model should be avoided. We are thus less concerned about measurements that may be of scientific interest, and caused by *bona fide* geophysical phenomena, and more concerned about the probability that observations are stable and well correlated between satellite and *in situ* measurement.

Our priority in this context is to remove measurements that represent outliers, and preserve all best measurements without bias even if this means preserving some (approximately) symmetrically distributed outliers[2].

Since we want to ensure that these algorithms can be applied in real time and without retrospective bias, and thus cannot be used to produce data that advise the analysis, the analysis field is chosen to be a daily foundation for the previous day.

[2] The resulting data selection will be used in ordinary least squares fitting, which assumes Gaussian (and thus symmetric) residuals

Apart from this obvious difference in timeliness and the associated uncertainty this creates, analysis estimates are further subject to other spacial and temporal limitations. For example, temporally, since we use one day foundation measurements, there will be differences due to the diurnal cycle. Spatial differences in grid resolution and assumptions made in smoothing, interpolation or models used to gap fill, also affect the analysis, and typically result in poor agreement in coastal areas, and in shallow ocean. Thus we do not expect the analysis to be able to faithfully estimate SST on the same scale or with the same detail as the satellite or *in situ* observations.

Hence there are potentially two problems, to determine which analysis SST measurements to use, and then to determine if the analysis SST can guide the choice of *in situ* measurements, by looking for agreement. In what follows we provide a crude but effective solution to these problems.

Given $S_{4,\mathrm{qc}}$ is the set of measurements based on reasonable analysis SST and $S_{4.\mathrm{insitu}}$ is the set of measurements with reasonable *in situ* observations given the set of measurements with reasonable analysis SST, we assume separability and define the measurements of interest, $S_{4,\mathrm{ok}}$ by the intersection,

$$S_{4,\mathrm{ok}} = S_{4,\mathrm{qc}} \cap S_{4,\mathrm{insitu}}(S_{4,\mathrm{qc}}) \tag{38}$$

To find $S_{4,\mathrm{qc}}$, we consider the deviation between analysis SST and satellite SST,

$$\Delta T_{i,\mathrm{a,s}} = T_{i,\mathrm{analysis}} - T_{i,\mathrm{satellite}} \tag{39}$$

We accept measurements based on $\Delta T_{i,\mathrm{a,s}}$ falling within 4.5 standard deviations, $\sigma_{\mathrm{a,s}}$, of the median $\Delta T_{i,\mathrm{a,s}}$,

$$S_{4,\mathrm{qc}} = \{i : |\Delta T_{i,\mathrm{a,s}} - \mathrm{median_j}(\Delta T_{j,\mathrm{a,s}})| < 4.5\,\sigma_{\mathrm{a,s}}\} \tag{40}$$

which by Chebyshevs theorem[2], assuming $\sigma_{\mathrm{a,s}}$ is calculated from the data set itself, will result in the exclusion of no more than approximately 5% of the available data, and negligible exclusion if the data set is normally distributed.

The upper chart in figure 36 shows that the relationship between the number of measurements accepted using percentiles and standard deviation are linearly related over a large range of rates of exclusion (from close to zero to 60%).

$\sigma_{\mathrm{a,s}}$ is computed based on a range of $\Delta T_{\mathrm{a,s}}$ defined by the 1st and 99th percentile, highest quality satellite observation ($4 \leq q_i \leq 5$) limits.

$$S_{4,\sigma} = \{i : P_1(\{\Delta T_{j,\mathrm{a,s}} : 4 \leq q_j \leq 5\}) < \Delta T_{i,\mathrm{a,s}} < P_{99}(\{\Delta T_{j,\mathrm{a,s}} : 4 \leq q_j \leq 5\})\} \tag{41}$$

$$\sigma^2_{\mathrm{a,s}} = \mathrm{var}_{i \in S_{4,\sigma}}(\Delta T_{i,\mathrm{a,s}}) \tag{42}$$

In the estimation of $\sigma_{\mathrm{a,s}}$, choosing a subset $S_{4,\sigma}$ devoid of the most extreme 2% measurements removes outliers in the high quality dataset which are most likely to represent discrepancies between the satellite and *in situ* measurement due to misjudgement of the level of cloud, abnormalities in the SST retrieval from brightness temperatures, or localized discrepancies of temperature on a temporal or spacial scale that is small compared to the scale of the SST analysis model, all of which are generally unsuitable for use in model building. Thus the range determined quantifies a reasonable expected range of variation of $\Delta T_{i,\mathrm{a,s}}$ under conditions where these undesirable effects are minimized.

This method of quality control does not require fixed expectations of the deviation from analysis and is thus self scaling, and can be applied across time domains where the analysis model and satellite platform vary somewhat, with minimal impact.

If the number of matches between *in situ* and satellite is reduced, which is characterised by our use of moored buoy measurements in the data set, we use a more lenient quality condition,

$3 \leq q_j \leq 5$ rather than $4 \leq q_j \leq 5$, and more relaxed percentile limits, P_5 rather than P_1 and P_{95} rather than P_{99}.

In the event that the retrieval scheme for the satellite SST $T_{i,\text{satellite}}$ has not been determined, a simple linear model (which we call the T_4T_5 model) based on two AVHRR brightness temperature channels is assumed,

$$T_{i,\text{satellite}} = t_0 + t_1 T_{i,4} + t_2 T_{i,5} \tag{43}$$

The model parameters $\{t_i\}$ are chosen based on all of the measurements satisfying the appropriate quality condition, $q \geq 4$, if no moored measurements are used, and $q \geq 3$ if moored measurements are used.

We desire that the SST will be estimated to within a few degrees K, approximately less than half the deviation than we expect would be characteristic of an outlier. We exclude non-linear terms and satellite field of view terms because of the asymmetric biases that might be introduced.

Details about the rate of exclusion based on monthly application of this algorithm to a NOAA-16 observed *in situ* data set is shown in figure 34 for reference. The boundaries based on P_1 and P_99 percentiles (red) show a clear asymmetry due to cold bias (primarily from misclassified cloud), which is not evident in the 4.5σ based assessment (green), which by construction assumes symmetry.

As a rule of thumb, an order $5K$ discrepancy seems like a reasonable choice for the removal of outliers in the data set, over a long period of time. The wider ranges in earlier years correspond to poorer quality *in situ* data.

A monthly parameter fit of the T_4T_5 model is shown in figure 35. Parameters show somewhat smoothly varying performance over time. Additionally, there is a rather large bias (t_0) in the model which is possibly due to geometric and sampling biases (which were not included in the model structure). The Model sensitivity however appears relatively stable over time and is close to 1. Anti-correlation between bias and sensitivity appears to be relatively large because there were no constraints placed on the amount the model coefficients were allowed to change from month to month.

The number of accepted matches compared to the total of $q \geq 3$ measurements for each of the two exclusion methods is shown in figure 36. At least in quantity, we at most we reject 2% of our measurements over most of the period considered. Using the 4.5σ approach actually results in us being slightly more lenient than accepting all $q \geq 3$ measurements, meaning that $q < 3$ measurements will also be included in the assessment. The applied filtering is thus, not particularly restrictive on the data set.

Changing attention to $S_{4,\text{insitu}}$, we consider the deviation between satellite SST and *in situ* SST, $\Delta T_{i,\text{s,I}}$, and the deviation between analysis SST and *in situ* SST, $\Delta T_{i,\text{a,I}}$, and the relationship between them. The analysis SST and *in situ* SST should be very highly correlated, as should the satellite SST and *in situ* SST. ΔT are therefore expected to be nominally zero, with some spread in distribution commensurate with the degree of correlation. Furthermore, if the *in situ* measurement is in error, both ΔT will be shifted in the same direction, along the line $\Delta T_{i,\text{s,I}} = \Delta T_{i,\text{a,I}}$. Although it may be difficult to discern if a pair of ΔT are influenced by an inaccurate *in situ* measurements, over a population of measurements, those that are close to the line of equality, $\Delta T_{i,\text{s,I}} = \Delta T_{i,\text{a,I}}$, can be removed, yielding $S_{4,\text{insitu}}$, the set of reasonable *in situ* measurements,

$$N_0 = \left\{ i : \frac{|\Delta T_{i,\text{s,I}} - \Delta T_{i,\text{a,I}}|}{\max\left(|\Delta T_{i,\text{s,I}}|, |\Delta T_{i,\text{a,I}}|\right)} > \gamma \right\} \tag{44}$$

$$\sigma^2_{s,I} = \text{var}_{i \in N_0 \cap S_{4,\text{qc}}}(\Delta T_{i,\text{s,I}}) \tag{45}$$

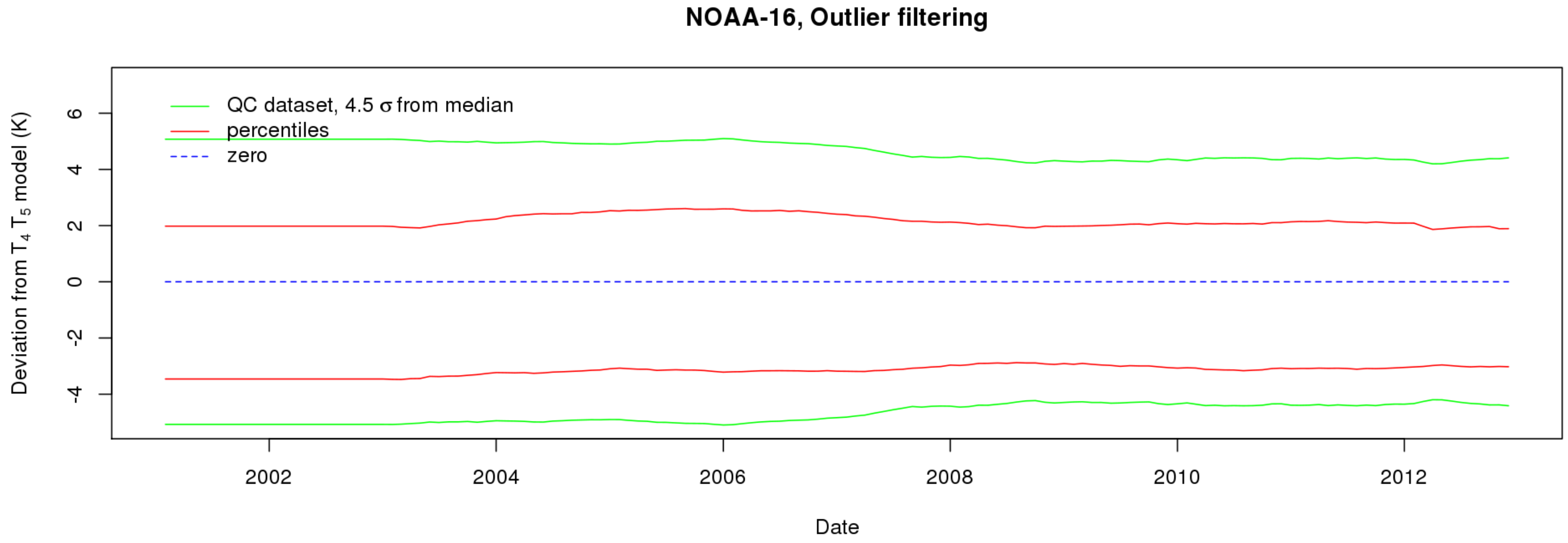


Figure 34: Filtering process and the determination of $S_{4,\sigma}$ (range given by red lines) and $S_{4,\mathrm{qc}}$ (range given by green lines) for NOAA-16 *in situ* matchups from 2001 to end of 2012, with the T_4T_5 model $T_{i,\mathrm{satellite}} = t_0 + t_1T_{i,4} + t_2T_{i,5}$, used for SST retrieval.

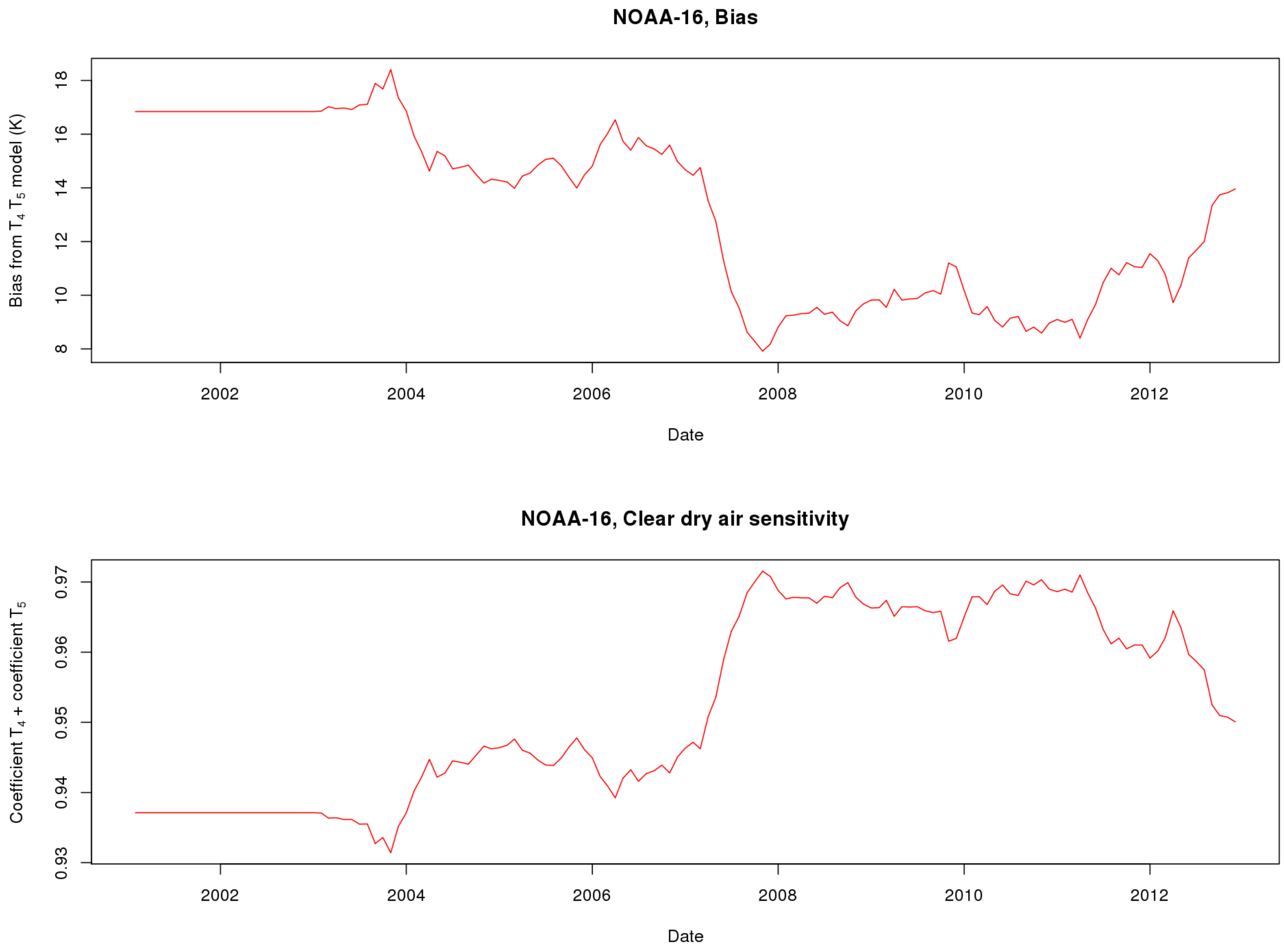


Figure 35: Filtering process and the determination of $S_{4,\sigma}$ and $S_{4,\mathrm{qc}}$ from NOAA-16 *in situ* matchups from 2001 to end of 2012. The upper chart shows t_0, the bias from the T_4T_5 model, which the lower chart shows $t_1 + t_2$, representing the sensitivity to sea temperature changes, and should be close to 1 in clear dry air.

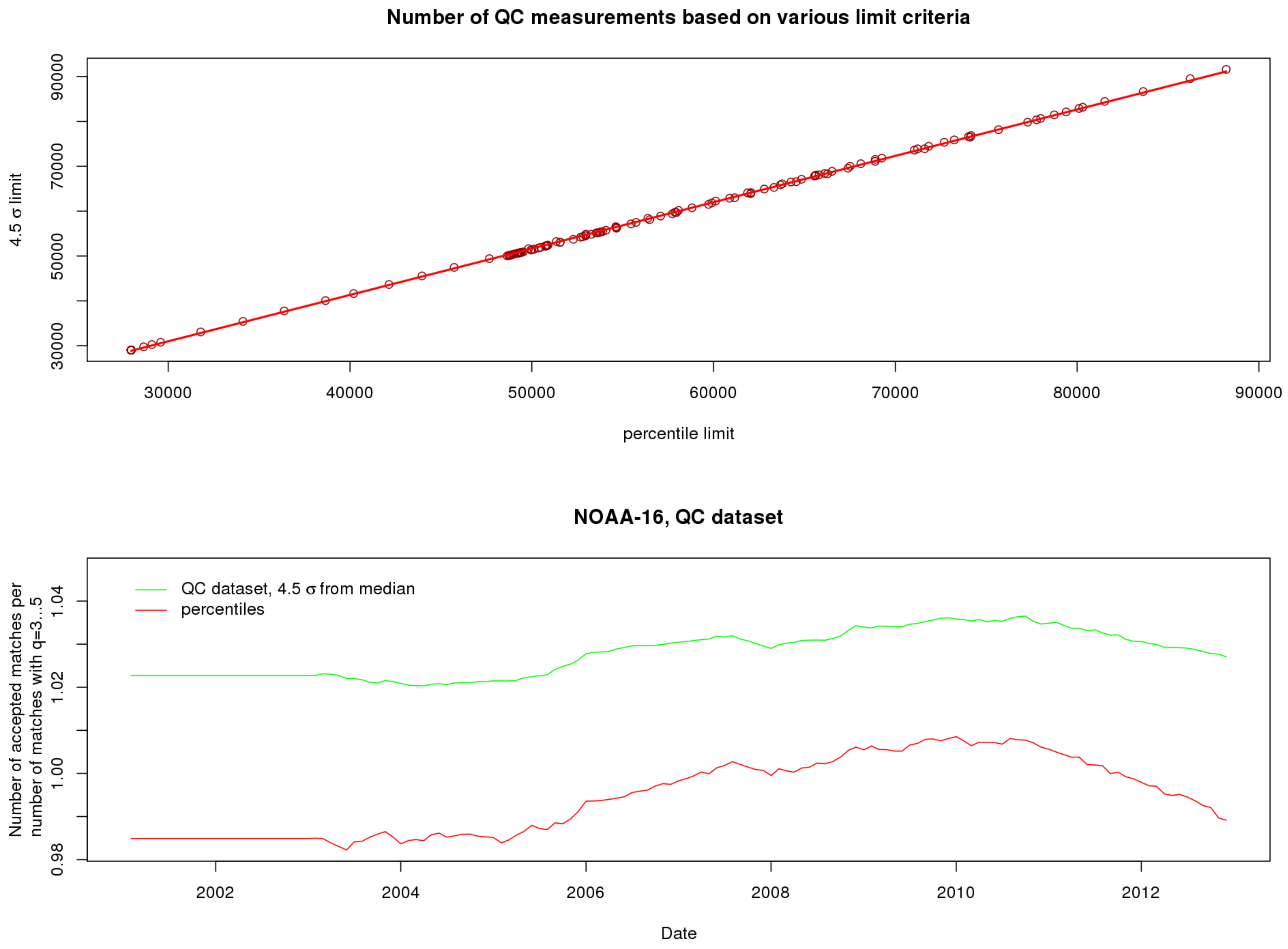


Figure 36: Filtering process and the determination of $S_{4,\sigma}$ and $S_{4,\mathrm{qc}}$ from NOAA-16 *in situ* matchups from 2001 to end of 2012. The upper chart explores the relationship between how many *in situ* measurements are accepted for both the percentile and σ based limits, as the percentile limit amount is changed. The lower chart shows how many are rejected based on a P_1 to $P_9 9$ acceptance region.

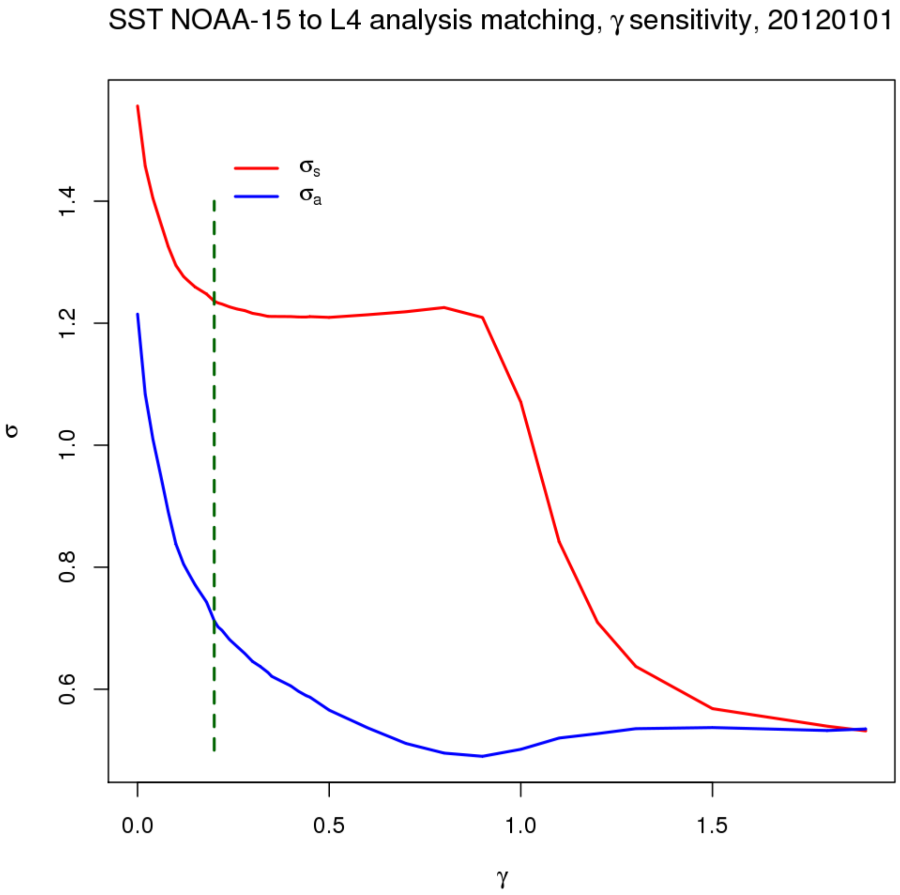


Figure 37: How choices in γ in the selection process for N_0 affect $\sigma_{s,I}$ and $\sigma_{a,I}$. We require γ as small as possible, so that the determination of both variances are stable. Choosing $\gamma = 0.2$ (dashed line) provides a value near the start of the plateau in both σ.

$$\sigma^2_{a,I} = \mathrm{var}_{i \in N_0 \cap S_{4,\mathrm{qc}}}(\Delta T_{i,\mathrm{a,I}}) \tag{46}$$

$$N = \left\{ i : \frac{|\Delta T_{i,\mathrm{s,I}} - \Delta T_{i,\mathrm{a,I}}|}{\sqrt{\sigma^2_{s,I} + \sigma^2_{a,I}}} > \min\left(\frac{1}{\sigma_{s,I}} |\Delta T_{i,\mathrm{s,I}}|, \frac{1}{\sigma_{a,I}} |\Delta T_{i,\mathrm{a,I}}| \right) \right\} \tag{47}$$

$$N_1 = \left\{ i : \max\left(\frac{1}{\sigma_{s,I}} |\Delta T_{i,\mathrm{s,I}}|, \frac{1}{\sigma_{a,I}} |\Delta T_{i,\mathrm{a,I}}| \right) < 4.5 \right\} \tag{48}$$

$$S_{4,\mathrm{insitu}} = S_{4,\mathrm{qc}} \cap (N \cup N_1) \tag{49}$$

In equations 44 to 49, outliers are removed to form measurement set N_0, which is used to estimate $\sigma_{s,I}$ and $\sigma_{a,I}$, the measurement uncertainties associated with the *in situ* and analysis differences to SST. The choice of $\gamma = 0.2$ in the filtering condition to form N_0 is largely arbitrary, and $\sigma_{s,I}$ and $\sigma_{a,I}$ are not particularly sensitive to this choice as long as it is neither too large or too small. See figure 37 for further information.

The set N is formed by excluding the measurements a closer number of standard deviations to $\Delta T_{i,\mathrm{s,I}} = \Delta T_{i,\mathrm{a,I}}$, than to $\Delta T_{i,\mathrm{s,I}} = 0$ or $\Delta T_{i,\mathrm{a,I}} = 0$. The set N_1 includes all of the measurement differences close to zero, using the 4.5σ limit we have used before, and $S_{4,\text{insitu}}$ is the union of both of these sets intersected with $S_{4,\text{qc}}$. The process is illustrated on a sample dataset in figure 38.

3.5 2 channel SST regression quality control.

The two channel regression model **BT45SZ** is used to do further quality control the suitability of data for SST regression. **BT45SZ** is the simplest *Day and Night* regression model, of reasonable accuracy that considers view angle[7], and is often used in studies of diurnal warming where the same retrieval scheme is required in both day and night,

$$f_{\mathbf{BT45SZ}} = a_0 + a_1 T_4 + a_2(T_4 - T_5) + a_3(T_4 - T_5)(\sec\theta_z - 1) \tag{50}$$

Given the residual of the model fit $R_{\mathbf{BT45SZ}}$, based on the data set, $D = \{d_i\}$, we determine the cleaned data set, D_{clean}, by retaining data such that the deviation of the residual from the median residual is less than γ standard deviations, where the standard deviation is calculated after 2% outliers are removed from $R_{\mathbf{BT45SZ}}(D)$,

$$\begin{aligned} D_{98} &= \{i : P_1(R_{\mathbf{BT45SZ}}(D)) < R_{\mathbf{BT45SZ}}(d_i) < P_{99}(R_{\mathbf{BT45SZ}}(D))\} && (51)\\ \sigma_D^2 &= \text{var}(D_{98}) && (52)\\ D_{\text{clean}} &= \{i : |R_{\mathbf{BT45SZ}}(d_i) - \text{median}(D_{98})| \le \gamma\,\sigma_D\} && (53) \end{aligned}$$

Since the distribution is expected to be skewed, generally due to poorly discriminated cloud which results in erroneous assignment of quality level, γ is chosen to be the minimum of the greatest positive deviation, the greatest negative deviation, and 4.5,

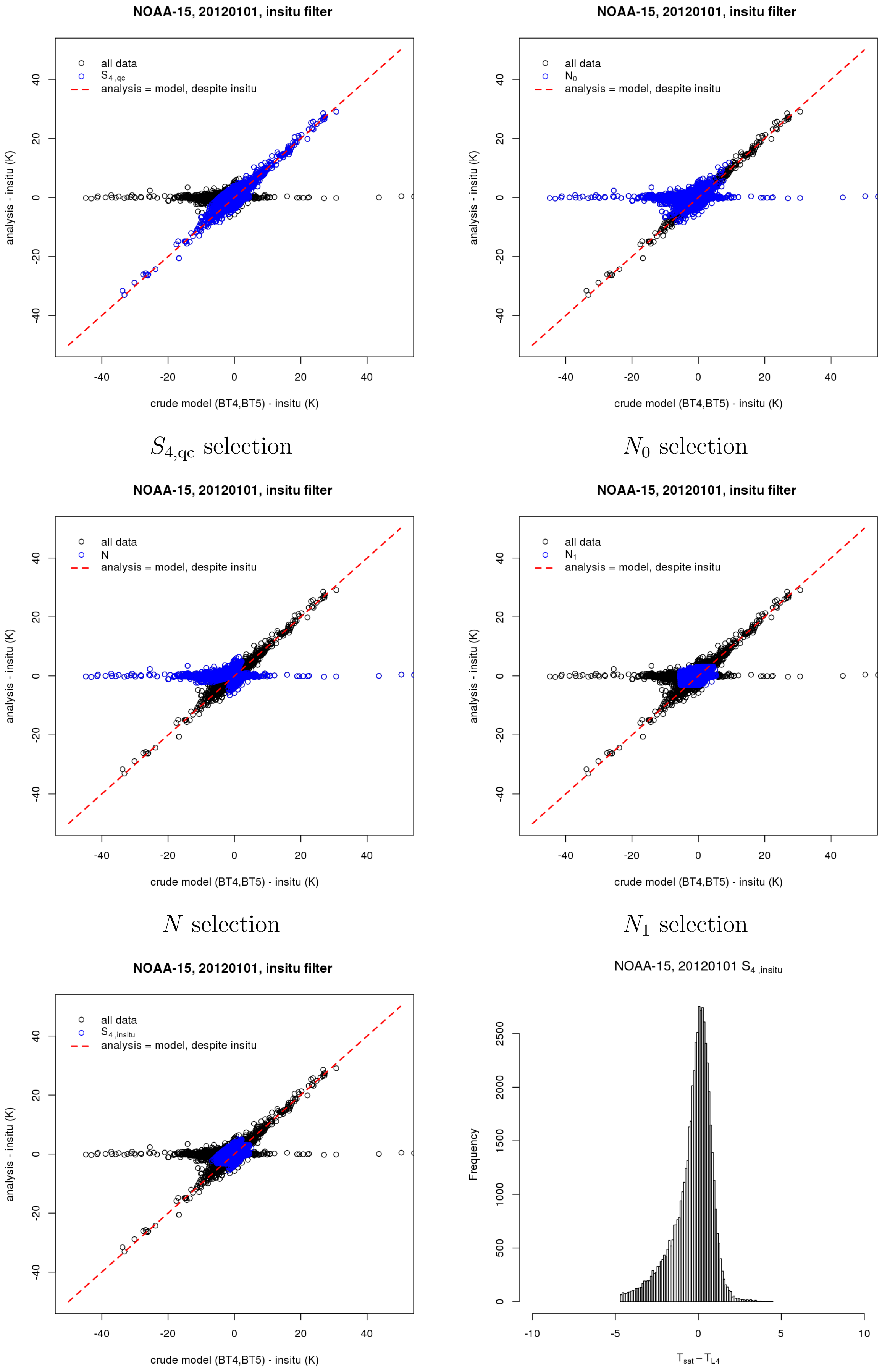
$$\begin{aligned} \gamma = \min\Bigg(& \frac{1}{\sigma_D} \max_{i\in D_{98}}(R_{\mathbf{BT45SZ}}(d_i) - \text{median}(D_{98}))\,, \\ & \frac{1}{\sigma_D} \max_{i\in D_{98}}(\text{median}(D_{98}) - R_{\mathbf{BT45SZ}}(d_i))\,, 4.5\Bigg) \end{aligned} \tag{54}$$

Since the distributions of measurements are unevenly distributed for day and night, and we wish to avoid a day / night bias, we perform a weighted ordinary least squares fit, with weight w_i,

$$\begin{aligned} w_i &= \frac{1}{n_{\text{day}}},\ \text{for day data} \\ & \frac{1}{n_{\text{night}}},\ \text{for night data} \end{aligned} \tag{55}$$

Where n_p are the number of measurements at the different periods of the day. If the residuals are normally distributed, the measurement sample sizes are sufficiently small, such that the $\gamma = 4.5$ limit would not result in any of data being removed from the data set. The outlier criteria results in the standard deviation being underestimated by about 7%, and we would be removing outliers more extreme than 4.5 standard deviations, which is less than 2 per 100000.

With a skewed distribution, only the outliers on one side of the distribution, which would generally bias least squares regression (which assumes the residuals are normally distributed), will be removed. Regression over the reduced data set is thus expected to be better centered.



$S_{4,\mathrm{qc}}$ selection

N_0 selection

N selection

N_1 selection

$S_{4,\mathrm{insitu}}$ selection

$S_{4,\mathrm{insitu}}$ distribution

Figure 38: Set selection in *in situ* and analysis filtering.

3.6 Data sets for fitting and quality control

Data sets for fitting, verification and validation, are chosen from the same source satellite to *in situ* data. Fitting data sets represent a subset of the data set that corresponds to the best set of measurements, that likely have good correspondence between satellite and *in situ* . Verification data are less stringently controlled and suitable for the verification of a modelling algorithm that may have been fitted with the Fitting data. Validation data sets provide some geophysical quality control to derive satellite to *in situ* matched events where there is a good chance that a coincident measurements is possible.

Table 13 summarizes the differences and similarities between these data sets and what they were used for.

Data set	Wind speed filter (m/s)	Lat QC	L4 QC	2Ch SST	Location	Quality	Use
Validation for both fv01 and fv02	$6 \leq v \leq 20$(d) $2 < v < 20$(n)	yes	no	no	within 2km 1hour	no filter	For independant validation of SST retrievals of the overall system.
SSES Lookup table fv01	$6 \leq v \leq 20$(d) $2 < v < 20$(n) Additional criteria are applied per section 3.1	no	no	no	within 2km 1hour	no filter	For independant validation of SST retrievals of the overall system.
Verification fv02	$6 \leq v \leq 20$(d) $2 \leq v \leq 20$(n)	yes	no	no	within 2km 1hour	$4 \leq q \leq 5$, or $3 \leq q \leq 5$ if moored buoy data is used	For verification of SST and SSES models. Summarizing the performance of model building activities.
Fitting fv02	$6 \leq v \leq 20$(d) $2 < v < 20$(n)	yes	yes	yes	same pixel 1hour	$4 \leq q \leq 5$, or $3 \leq q \leq 5$ if moored buoy data is used.	Model fitting for SST retrieval, and SSES determination.

Table 13: Matched data sets between satellite and *in situ* SST measurements used for fitting, verification and validation. Data sets are originally chosen as described in section 3.2, then the constraints applied as tabulated. Wind filters apply separately for day (d) and night (n). LatQC is applied according to section 3.3. L4QC is applied according to section 3.4. 2Ch SST filtering is applied according to section 3.5. Co-location requires the *in situ* measurement to be in the same pixel on the satellite image as the satellite measurement. Dropping the co-location requirement requires the matched data to be located within a specified number of kilometres of the *in situ* data. Quality filtering depends on the availability of data. If moored buoy data is required to be included due to inadequate coverage for drifting buoys, the quality requirements are less stringent, because the data is sparse.

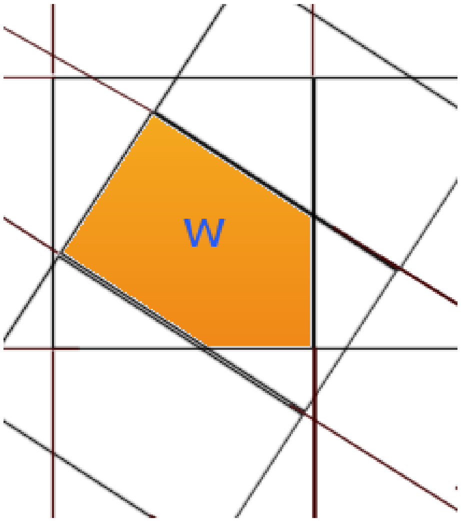


Figure 39: Identification of grid overlap weight.

3.7 L3U class product and the computation of L3U from L2P product

Rectangularly gridded L3U products are produced by remapping each single swath of SST product in native coordinates onto a fixed grid. The purpose of this is to provide a consistent coordinate system for products that allows for easier comparison from swath to swath, and the opportunity to provide the product to a grid resolution that is commensurate with the requirements of other downstream applications.

The process of gridding SST consists of projecting an ungridded swath, pixel by pixel, onto a regular fixed grid, weighting each pixel contribution by the area of overlap between the source and target pixels, w_i. In the computation of w_i, we assume that both the source and target pixel arrays consist of many parallelograms with side lengths given by the centre to centre distance of the pixels. The source image is assumed to be rotated, and the size and centres of the pixels vary over the field of view in both the source and target pixels. For the particular supported grids, cylindrical coordinates ensure that the target locations are regularly spaced over the chosen rectilinear grid extent.

The weight corresponds to the area of overlap, as demonstrated in figure 67.

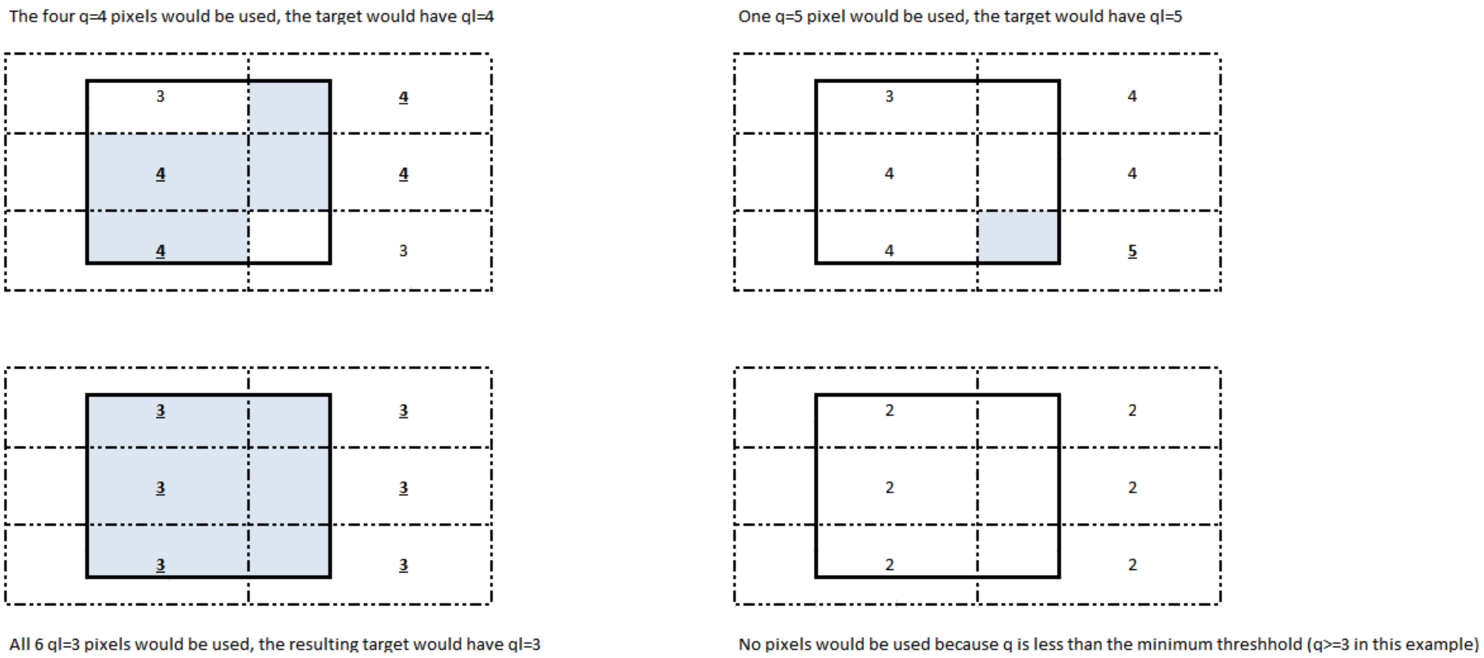
For a particular target pixel, the overlapping source pixels are sorted based on quality level, and those with the highest quality are merged by applying weighted sums. Figure 68 shows an example of how pixels are chosen. The SST and other similar parameters are mapped using the standard weighted average method, with weights $w_{i,j}$ representing the overlap area of source pixel i into target pixel j,

$$T_{\text{satellite},U,j} = \frac{\sum_{i \in j} w_{i,j} \, T_{\text{satellite},i}}{\sum_{i \in j} w_{i,j}}, \tag{56}$$

wherein $\sum_{i \in j}$ represents the sum over all suitable pixels that contribute to a given target pixel j, determined based on the best quality pixels available at the given target.

$T_{\text{satellite},U,j}$ defined in this way corresponds to an area weighted overage of best quality SST measurements at the pixel of interest. The same averaging technique is applied to other ancillary fields (such as wind speed, aerosol dynamic indicator, analysis SST, observation time).

Pixel by pixel SSES, the gridded degrees of freedom $n_{U,j}$, bias $\mu_{U,j}$ and standard error $\sigma_{U,j}$, are



The four q=4 pixels would be used, the target would have ql=4

One q=5 pixel would be used, the target would have ql=5

All 6 ql=3 pixels would be used, the resulting target would have ql=3

No pixels would be used because q is less than the minimum threshhold (q>=3 in this example)

Figure 40: Identification of contributing pixels based on source candidate quality ensures only pixels of highest quality contribute to the target L3 pixel.

determined using slightly different algorithms. The bias, which is expected to be an estimated offset to SST, considers area based weighting in exactly the same manner as the ancillary fields,

$$\mu_{U,j} = \frac{\sum_{i\in j} w_{i,j}\, \mu_{P,i}}{\sum_{i\in j} w_{i,j}} \tag{57}$$

The number of gridded degrees of freedom reflects the number of pixels that went into the average, by considering the sum of each pixels weighted contribution, normalized by the largest contribution,

$$n_{U,j} = \frac{\sum_{i\in j} w_{i,j}}{\max_{i\in j} w_{i\in j}}, \tag{58}$$

where as before, the understanding of $\{i \in j\}$ is all of the best quality source pixels i that overlap with the target pixel j. The resulting count value $n_{U,j}$ is a non integer value representative of the number of pixels that went into the computation. The use of the maximum weight as normalisation allows the pixel with the largest contribution to count as one, and those that contribute relatively less to be counted as such according to the weight relative to the largest contribution. A per pixel normalization of this kind does not affect the interpretation of the weight in the computation of other weighted averages, but allows the interpretation as number of significant measurements to be applied over the field of view irrespective of the relative overlap area in different regions of the remapping. The standard deviation estimate, which is derived from a population of $n_{P,i}$ *in situ* measurements, is also weighted by pixel overlap, $w_{i,j}$

$$\sigma_{U,j} = \sqrt{\frac{\sum_{i\in j} w_{i,j}\,(\sigma_i^2 + \mu_{P,i}^2)}{\sum_{i\in j} w_{i,j}} - \left(\frac{\sum_{i\in j} w_{i,j}\, \mu_{P,i}}{\sum_{i\in j} w_{i,j}}\right)^2} \tag{59}$$

None of the SSES depend on $n_{P,i}$, and their formation is thus not affected by missing degrees of freedom data in L2P files. This is appropriate, since the SST measurement itself has no relation to the number of degrees of freedom used to generate the bias and standard error estimates, and we are forced to use the same method of averaging for all three of these parameters to maintain consistency in the application of the bias as a correction.

In addition to ancillary fields such as wind and aerosol, The GHRSST specification also includes time based fields, which are also remapped from L2P to L3U as the weighted average time since epoch, under the assumption the the linear variation of measured value is best described by a linear variation in time.

The L2P f_{L2p} parameter, which describes possible exceptions or causes that may influence pixel quality and interpretation (generally negatively) is combined using the local OR of all of the source pixels, respecting the desire to record any possible influence that may contribute to the interpretation of the target pixel behaviour.

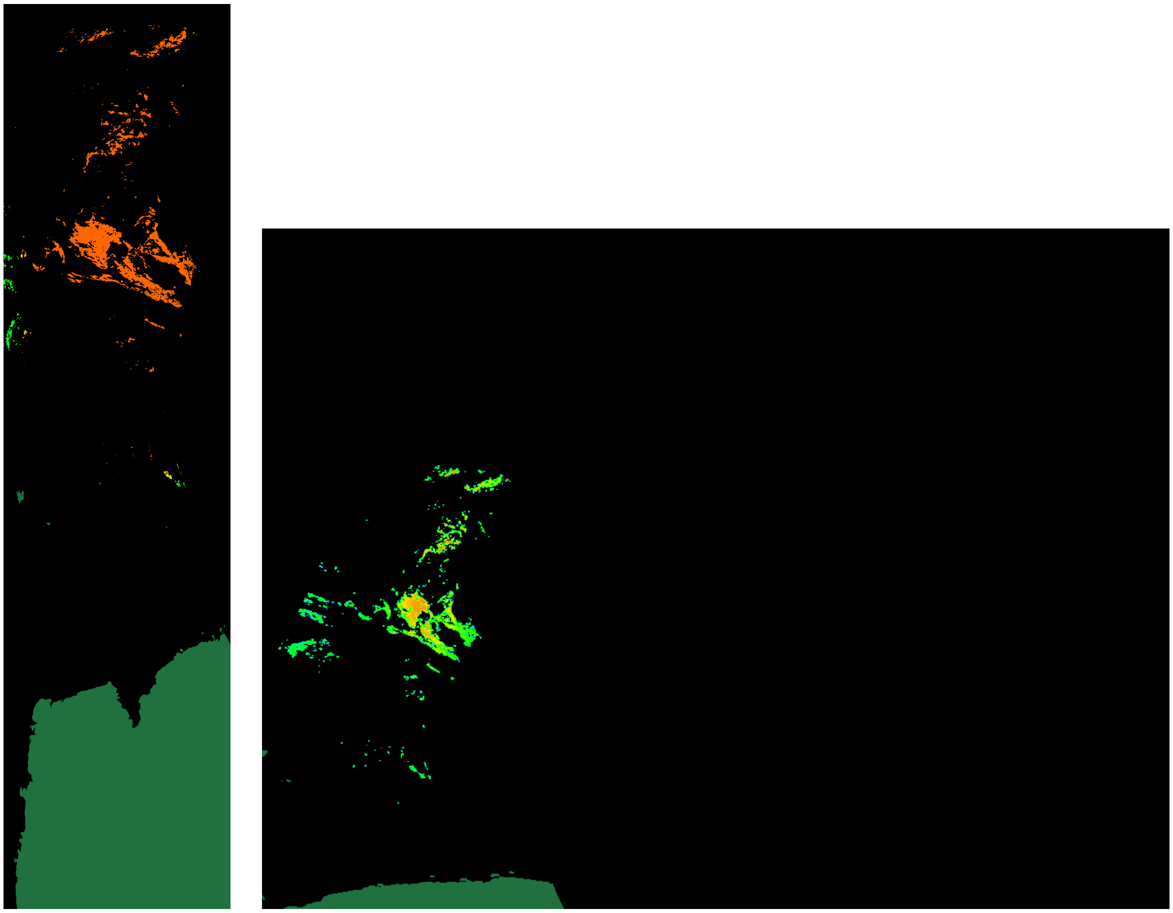
In this manner, all of the points on the L2P swath are mapped to an L3U set,

$$\left\{T_{\mathrm{satellite},U,j}, t_{U,j}, q_{U,j}, \mu_{U,j}, \sigma_{U,j}, n_{U,j}, \mathrm{ancillary}_j, f_{\mathrm{L2p},U,j}\right\} \tag{60}$$

and this information is stored in the SSES fields for L3U files with the same indicative names as those used for L2P files, as outlined in table 19.

Parameter name	Symbol	L2P		L3U	
`sses_count`	n	n_P	Indicative of the number of *in situ* measurements made under similar viewing conditions	n_U	Indicative number of best quality L2P pixels merged when converted to a fixed grid.
`sses_bias`	μ	μ_P	Indicative median bias for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing conditions.	μ_U	Indicative gridded median bias for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing and merging conditions.
`sses_standard_deviation`	σ	σ_P	Indicative standard deviation for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing conditions.	σ_U	Indicative gridded standard deviation for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing and merging conditions.
`l2p_flags`	f_{L2p}	$f_{\text{L2p},P}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions.	$f_{\text{L2p},U}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location.
`quality_level`	q	q_P	Quality level as a measure of proximity to detected cloud in kilometres. $q = 0$ is also used to indicate invalid data for other reasons.	q_U	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.
`sea_surface_temperature`	$T_{\text{satellite}}$	$T_{\text{satellite},P}$	Retrieved sea surface temperature	$T_{\text{satellite},U}$	An average of the retrieved sea surface temperature based on all of the measurements contributing to the gridded location.

Table 14: Association between field names in GHRSST compliant files and symbols used in this text, with a short description of the intent of the parameter and symbol, for L2P and L3U files.



L2P n_P

L3U n_U

Figure 41: Comparison between `sses_count` field for L2P compared to L3U, for NOAA-16 at Jan 1st 2011, 00:36UTC. The L2P count reflects the fact that there are more *in situ* measurements on a long term average at the center of the swath (red) compared to the edges (green), and the number of measurements diminishes as we go further south (green). The L3U count reflects the fact that there is a higher overlap of L2P pixels in the middle of cloud clear regions near the center of swath (orange), compared to the edges of both the cloud and the swath (green).

It should be noted that the most significant difference between L2P and L3U SSES just described is the use of the `sses_count` field, which corresponds to an indicative number of *in situ* measurements that contribute to SSES estimates, n_P, in the L2P file (defaulting to 1 if not present), and an indicative number of incumbent best quality L2P pixels, n_U, in the L3U files, with highest n_U in places where larger numbers of low σ_P pixels contribute to the average. This is illustrated in figure 69.

3.8 L3C class product and the computation of L3C from L3U product

L3C products consist of merges of multiple swaths from the same instrument and platform over a period of time that is small on the scale of significant changes in the underlying SST. Since a single swath of a polar orbiting satellite provides only a small snapshot of the ocean temperature, merging multiple swaths allows greater regional coverage to be delivered at the expense of reduced temporal resolution. Thus we consider these as a common grid merge of multiple L3U data sources. Each L3C product has a measurement window - a time period and time domain - of interest, in our case we consider day-time one and three day products as well as night-time one and three day products, four different measurement windows.

The process of merging gridded SST estimates consists of taking the highest quality gridded estimates from multiple sources on the same grid, over the measurement window, and providing a merged value of these measurements, by averaging. The resulting average thus represents a characteristic measurement for the platform, over the measurement window in question. This process is complicated by the reality that there is expected to be some time dependent variation on the measured $\mathrm{SST_U}$, which will be averaged out in the averaging process, but should be considered when we think about estimates of the standard error.

This time dependent variation does not correspond to a typical measurement error, rather qualifies the variation in the estimate of SST due to natural variation over the measurement window. As the measurement window is enlarged, and the interpretation of the SST is maintained as characteristic measurement commensurate with the time taken in by the measurement window, the uncertainty due to this variation is expected to scale out as $\frac{1}{\sqrt{N}}$, where N is the number of measurements, swaths, or the amount of time involved, in keeping with standard error estimates, whereas the *in situ* errors will not, since they represent uncertainties associated with the reported SST value in the context of the measurement apparatus.

Furthermore, since it is likely at the edge of swath that there may be overlap in the measurements, and that the sensor specific error statistics reflect uncertainties based on satellite zenith angle, or that there are different error estimates associated with different times of the day, it is desirable to weight measurements by their significance measured by the number of degrees of freedom, n_U and the estimate of the measurement error, σ_U, under the assumption that measurements with a larger n_U and a smaller σ_U are more certain to be representative of the pixel in question over the period over which the merge is considered, and each measurement made can be considered somewhat independent.

Following this rationalisation, we choose the representative value for $T_{\mathrm{satellite},C}$, the gridded instrument specific merged SST over a fixed time period and other merged parameters to be computed with a simple quantity over variance $(\frac{n}{\sigma^2})$ weighting as if the combined measurements are from uncorrelated sources,

$$T_{\mathrm{satellite},C,j} = \frac{\sum_{i\in j} \frac{n_{U,i}}{\sigma^2_{U,i}} T_{\mathrm{satellite},U,i}}{\sum_{i\in j} \frac{n_{U,i}}{\sigma^2_{U,i}}} \tag{61}$$

where the sum in the above expression over $i \in j$, is assumed to be over all of the best quality source L3U pixels from all of the swath files over the time window, i at the common target location, j.

SSES $\{n_C, \mu_C, \sigma_C\}$, are determined by a similarly weighted average for the number of degrees of freedom and the bias,

$$\begin{aligned} n_{C,j} &= \frac{\sum_{i\in j} \frac{n_{U,i}}{\sigma^2_{U,i}}}{\sum_{i\in j} \frac{1}{\sigma^2_{U,i}}} && (62) \\ \mu_{C,j} &= \frac{\sum_{i\in j} \frac{n_{U,i}}{\sigma^2_{U,i}} \mu_{U,i}}{\sum_{i\in j} \frac{n_{U,i}}{\sigma^2_{U,i}}} && (63) \end{aligned}$$

The sensor specific standard deviation is similarly computed,

$$\sigma_{Cs,j}^2 = \frac{\sum_{i\in j}\frac{n_{U,i}}{\sigma_{U,i}^2}(\sigma_{U,i}^2+\mu_{U,i}^2)}{\sum_{i\in j}\frac{n_{U,i}}{\sigma_{U,i}^2}} - \mu_{C,j}^2 \qquad (64)$$

but is corrected for the time window variation of the SST which adds an additional uncertainty to the time window characteristic SST computed, resulting in the SSES estimate, σ_C,

$$\sigma_{C,j} = \sqrt{\sigma_{Cs,j}^2 + \frac{\sigma_{w,C,j}^2}{n_{C,j}}} \qquad (65)$$

which scales the environmental component, $\sigma_{w,C}$ out as $\sim \frac{1}{\sqrt{n_C}}$, as any estimate of the standard error of the mean, by the central limit theorem.

$\sigma_{w,C}$ is the standard deviation of the environmental component of the SST over the time window, which is estimated by making use of the time window variability parameters, $\{n_{w,C}, T_{w,C}, \sigma_{w,C}\}$, determined by equally weighting all of the SST measurements over the period of interest, irrespective of n_U and σ_U.

$$n_{w,C,j} = \text{count}\,\{i\in j\} \qquad (66)$$

$$T_{w,C,j} = \frac{\sum_{i\in j} T_{\text{satellite},U,i}}{n_{w,j}} \qquad (67)$$

$$\sigma_{w,C,j}^2 = \frac{\sum_{i\in j} T_{\text{satellite},U,i}^2}{n_{w,j}} - T_{w,C,j}^2 \qquad (68)$$

where, as before, the notation $\{i\in j\}$ refers to the set of all of the valid pixels of the best quality from the multiple L3U sources covering the time window at the position j, and the count function counts them. This serves as an indication of the amount of variation possible under the assumption that all of the measurements made have no error and are of good quality, thus any variation seen is an estimate of the environmental variation over the time window rather than instrument variation. In the event that the instrument and environment are uncorrelated, this will be an overestimate or conservative estimate of the possible environmental variation. The variation over the time window is added in quadrature (with the assumption of normality) to the instrument contribution to σ_C, because it is expected the environment will be uncorrelated with the instrument variation as a first approximation.

In this manner, all of the points of the L3U source data are mapped to an L3C data set,

$$\left\{T_{\text{satellite},C,j}, t_{C,j}, q_{C,j}, \mu_{C,j}, \sigma_{C,j}, n_{C,j}, T_{w,j}, \sigma_{w,C,j}, n_{w,C,j}, \texttt{ancillary}_j, f_{\text{L2p},C,j}\right\} \qquad (69)$$

and this information is stored in the SSES fields for L3C files with the same indicative names as those used for L3U files, as outlined in table 20. Ancillary fields are treated the same way as SSTfields, and f_{L2p} are bitwise or-ed, as before.

Note that there are four additional fields to those recommended in the GDS version 2.0r5. In addition to `sses_count`, n_C, we have added the three time window variation fields corresponding to $\{T_{w,C,j}, \sigma_{w,C,j}, n_{w,C,j}\}$, `sst_mean`, `sst_standard_deviation` and `sst_count`, representing the equally weighted SST, standard deviation, and count. Having these stored in L3C files allows combinations of L3C files to be considered and compared, and the observed environmental parameters combined. This aids in the merging of L3C to L3S, where differences in the bias due to different platforms are considered. See section A.3 for further details.

3.9 L3S class product and the computation of L3S from L3C product

L3S class product provides a typical characteristic SST over a (possibly) extended time window, and multiple instruments, by combining single day, single platform L3C files. In order to consider the SST provided indicative of the time period, we assume that all best quality measurements from all platforms and days contribute equally. To remove the impact of unstable or end of life platforms, we only include the missions that we consider production quality on the day in question. See figure 70 for details about which missions are included over the full period covered by the archive.

The equal weighting simplifies the composition process of L3S files, and allows multiple L3S files to be composed and generated progressively if the coverage period is very long,

$$T_{\text{satellite},S,j} = \frac{\sum_{i\in j} n_{C,i}\, T_{\text{satellite},C,i}}{\sum_{i\in j} n_{C,i}} \tag{70}$$

As before, the sum is over all of the best quality pixels at the same target location j over the time window and range of platforms.

The number of degrees of freedom, combined bias and standard deviation with respect to *in situ* are estimated based on equal weighting after first removing the time window variation from the L3C SSES,

$$\begin{aligned}
n_{S,j} &= \sum_{i\in j} n_{C,i} && (71)\\
\mu_{S,j} &= \frac{\sum_{i\in j} n_{C,i}\mu_{C,i}}{n_{S,j}} && (72)\\
\sigma^2_{Cs,i} &= \sigma^2_{C,i} - \frac{\sigma^2_{w,C,i}}{n_{C,i}} && (73)\\
\sigma^2_{Sb,j} &= \frac{\sum_{i\in j} n_{C,i}(\sigma^2_{Cs,i} + \mu^2_{Cb,i})}{n_{S,j}} - \mu^2_{S,j} && (74)
\end{aligned}$$

In estimating SSES, some care is required in treating time window variation and *in situ* based variation separately. We use the same six factor representation of SSES introduced in section A.2, being careful to apply platform biases to the measured SST in the composition of the mean time window SST, and the variance of the time window SST,

$$\begin{aligned}
n_{w,S,j} &= \sum_{i\in j} n_{w,C,i} && (75)\\
T_{w,S,j} &= \frac{\sum_{i\in j} n_{w,C,i}\,(T_{w,C,i} - \mu_{C,i})}{n_{w,S,j}} + \mu_{S,j} && (76)\\
\sigma^2_{w,S,j} &= \frac{\sum_{i\in j} n_{w,C,i}\left(\sigma^2_{w,C,i} + \mu^2_{w,C,i}\right)}{n_{w,S,j}} - (T_{w,S,j} - \mu_{S,j})^2\\
&\quad + \frac{\sum_{i\in j} n_{w,C,i}\mu_{C,i}\,(\mu_{C,i} - 2T_{w,C,i})}{n_{w,S,j}} && (77)
\end{aligned}$$

Additional terms in the above remove contributions due to the platform biases in $T_{w,C}$, and the correlation between $T_{w,C}$ and μ_C, the measured temperature and the bias in the measurement equipment.

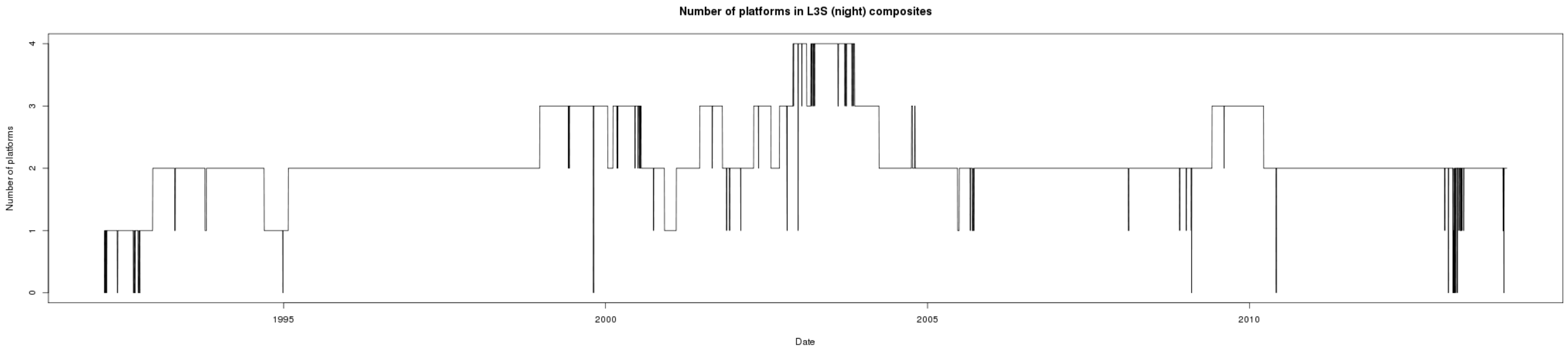


Figure 42: Number of NOAA-AVHRR missions available for inclusion in ABOM L3S night composites since 1995. Satellites are excluded or included based on NOAA mission status, reception quality, and the success of accurate navigation correction. For the greater part of the period, more than 2 platforms provide coverage for L3S files over continental Australia.

As before, the standard deviation is a composite of the sensor related component added in quadrature to the environmental component, where the environmental component is treated as a standard error of mean, and scaled out with the number of degrees of freedom,

$$\sigma_{S,j} \;=\; \sqrt{\sigma^2_{Sb,j} + \frac{\sigma^2_{w,S,j}}{n_{S,j}}} \tag{78}$$

Thus we form the L3S data set,

$$\left\{T_{\mathrm{satellite},S,j}, t_{S,j}, q_{S,j}, \mu_{S,j}, \sigma_{S,j}, n_{S,j}, T_{w,S,j}, \sigma_{w,S,j}, n_{w,S,j}, \mathtt{ancillary}_j, f_{\mathrm{L2p},S,j}\right\} \tag{79}$$

and this information is stored in the resulting L3S fields with the same indicative names as those used for L3C files, as outlined in table 20. Ancillary fields are treated the same way as SST fields, and f_{L2p} are bitwise or-ed, as before.

This treatment allows L3S and L3C files to be combined hierarchically, producing L3S files at an intermediate step that can be further combined. Longer time period product with many individual data sources can thus be produced recursively with the resulting SSES independent of the exact order in which the files were combined. For example, annual L3S SST could be generated by combining four quarterly L3S SST products which are in turn derived from three monthly L3S SST product, each of which are composed of daily L3S product, which are in turn composed of the L3C product from various source instruments on their respective days.

The resulting L3S product contains estimates of the time window SST variation $\sigma_{w,S}$, the *in situ* error, σ_S, the number of measurements $n_{w,S}$ and the number of high quality measurements n_S, with biases corresponding to mean bias over all platforms, μ_S.

4 Application of SSES

4.1 A reassessment of quality using SSES

Once the SSES are estimated, the quality of an SST can be estimated based on the size of the standard deviation normalized against the minimum acceptable standard deviation on the field of view, and the measured bias as a number of standard deviations. The method and rationale is outlined in the general discussion of appendix A.

Weighting both of these contributions equally, and assuming the argument of the square root is always positive, equation 163 in section A.4.3 can be written as,

$$q_s = \frac{1}{\sqrt{2}}\sqrt{\left(\frac{\sigma}{\sigma_0}\right)^2 + \left(\frac{\mu}{\sigma}\right)^2 - 1} \tag{80}$$

q_s defined in this way is effectively the uncorrelated mean z score derived from the standard deviation spread from minimum and bias shift, under the assumption of normally distributed error estimates.

Because out estimates of σ and μ are derived directly from q, there is a strong correlation between q_s and q, however under conditions where σ and μ vary greatly at fixed q, the boundaries between levels of q_s will become less distinct, and q_s will better the reflect the quality of the measurement, provided we have a reasonable estimate of the SSES. Moreover, whereas the assessment of q is based on independent assessments of every field of view, (for example, $q = 5$ in a day field of view has the same quality as $q = 5$ in a night field of view, even though cloud clearing and retrieval

Parameter name	Symbol	L3C		L3S	
`sses_count`	n	n_C	Indicative number of good quality L3U measurements merged to L3C	n_S	Indicative number of good quality L3U measurements merged to L3S
`sses_bias`	μ	μ_C	An estimate of the median bias of the platform and sensor over the time window of the L3C file.	μ_S	An estimate of the median bias over all measurements over the time window of consideration.
`sses_standard_deviation`	σ	σ_C	Indicative uncertainty over the time window, including contributions from natural variation as they affect the estimate of the mean SST.	σ_S	Indicative uncertainty over the time window, including contributions from natural variation as they affect the estimate of the mean SST.
`sst_count`	n_w	$n_{w,C}$	Number of measurements merged to L3C.	$n_{w,S}$	Number of measurements merged to L3S.
`sst_mean`	T_w	$T_{w,C}$	Unweighted mean measured sea surface temperature.	$T_{w,S}$	Unweighted mean measured sea surface temperature.
`sst_standard_deviation`	σ_w	$\sigma_{w,C}$	Unweighted standard deviation of measured sea surface temperature.	$\sigma_{w,S}$	Unweighted standard deviation of measured sea surface temperature.
`l2p_flags`	f_{L2p}	$f_{\mathrm{L2p},C}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location over the time window.	$f_{\mathrm{L2p},S}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location over the time window.
`quality_level`	q	q_C	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.	q_S	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.
`sea_surface_temperature`	$T_{\mathrm{satellite}}$	$T_{\mathrm{satellite},C}$	Estimate of the sea surface temperature characteristic of the time window.	$T_{\mathrm{satellite},S}$	Estimate of the sea surface temperature characteristic of the time window.

Table 15: Association between field names in GHRSST compliant files and symbols used in this text, with a short description of the intent of the parameter and symbol, for L3C and L3S.

algorithms are distinctly different), the determination of q_s is not. q_s thus should serve as a better basis for comparing pixel quality between different scenes, at different times of the day, and over longer periods of time. Due to its experimental nature, q_s is only available in L2P files, where it is called `sses_quality` and as of fv01 and fv02 is not used in the formation of L3U files or gridded composites. This possibility is under investigation for future product improvements. Please see appendix A for further information concerning the derivation, rationalization and further use of q_s as a quality assessment in the context of generating L3S composites from external suppliers.

5 Computation of retrieval sensitivity

Retrieval sensitivity computations are the GHRSST recommended way for evaluating the accuracy and stability of an SST retrieval. The sensitivity gives a measure of how sensitive retrieved SST is to changes in physical SST, and is most usually evaluated using radiative transfer models based on reanalysis of the state of the atmosphere over the time of the retrieval. We choose a simpler method that relies on a numerical estimate derived from the same set of *in situ* measurements from which the retrieval is derived, which has less dependency on the details surrounding the construction of the radiative transfer and atmospheric modelling.

5.1 Computation of retrieval sensitivity from *in situ* measurements

The sensitivity of the retrieval, S can be defined as the change of satellite retrieved SST, $T_{\text{satellite}}$ with respect to a change in the physical SST as follows,

$$S = \frac{\partial T_{\text{satellite}}}{\partial T_{\text{physical}}} \tag{81}$$

Over many measurements, assuming that the *in situ* SST with cool skin correction is representative of the physical skin SST, at least on median[3] we can approximate S as follows,

$$S \approx \text{median}_i \left(\frac{\partial T_{\text{satellite}}}{\partial T_{i,\text{insitu}}} \right) \tag{82}$$

Given an SST retrieval at pixel i is a simple analytic function in terms of a set of 3 AVHRR brightness temperatures, as summarized in table 4, $\{T_{i,3}, T_{i,4}, T_{i,5}\}$, we can write the satellite measurement as follows,

$$T_{i,\text{satellite}} = f(T_{i,3}, T_{i,4}, T_{i,5} \ldots) \tag{83}$$

where we acknowledge that there may be able factors in the dependency, but ignore these for the moment. The sensitivity can be approximated thus,

$$\begin{aligned} S \ &\approx\ \text{median}_i \Bigg[\left(\frac{\partial T_{\text{satellite}}}{\partial T_3} \right) \left(\frac{\partial T_3}{\partial T_{i,\text{insitu}}} \right) \\ &+\ \left(\frac{\partial T_{\text{satellite}}}{\partial T_4} \right) \left(\frac{\partial T_4}{\partial T_{i,\text{insitu}}} \right) \\ &+\ \left(\frac{\partial T_{\text{satellite}}}{\partial T_5} \right) \left(\frac{\partial T_5}{\partial T_{i,\text{insitu}}} \right) \Bigg] \end{aligned} \tag{84}$$

[3] We use the median to improve the robustness of the approach. Since we expect the distribution of retrievals be asymmetric if cloud identification is poorly performed, and close to symmetric if performed accurately, choosing the median provides an assessment that is more immune to cloud misclassification.

Since, in clear sky, $\{T_{i,3}, T_{i,4}, T_{i,5}\}$ are all surface temperature estimates (subject to corrections due to absorption at different rates by atmospheric components) they are by construction highly correlated. Terms involving the partial derivative of $T_{\text{satellite}}$, can be computed analytically since the retrieval is an algebraic function, which leaves the *in situ* sensitivities S_j

$$S_{i,j} = \frac{\partial T_{i,\text{insitu}}}{\partial T_j} \tag{85}$$

to be estimated.

If we assume that the variability of the sensitivity is expected to be primarily influenced by field of view and seasonal fluctuations which we limit to satellite zenith angle, latitude, and time of year, the *in situ* temperature can be written $T_{i,\text{insitu}}(\theta_z, \theta_{\text{lat}}, t)$. If a linear approximation is also reasonable, and furthermore we assume that the median of the products is approximately the product of median, we can consider $\text{median}_{\text{i}}(S_{i,j})$, which can be estimated from *in situ* measurements by binning the data accordingly, then approximating the partial derivatives by forming a linear regression of the *in situ* measurements against the individual brightness temperatures, with regression coefficients $\{a_i\}$,

$$\begin{aligned} T_{i,\text{insitu}}(\theta_z, \theta_{\text{lat}}, t) &= a_0(\theta_z, \theta_{\text{lat}}, t) + a_3(\theta_z, \theta_{\text{lat}}, t)\, T_{i,3}(\theta_z, \theta_{\text{lat}}, t) \\ &= a_1(\theta_z, \theta_{\text{lat}}, t) + a_4(\theta_z, \theta_{\text{lat}}, t)\, T_{i,4}(\theta_z, \theta_{\text{lat}}, t) \\ &= a_2(\theta_z, \theta_{\text{lat}}, t) + a_5(\theta_z, \theta_{\text{lat}}, t)\, T_{i,5}(\theta_z, \theta_{\text{lat}}, t) \end{aligned} \tag{86}$$

from which the partial derivatives of the retrieval algorithm can be approximated where we the fitting algorithm defines an appropriate context in which the mean is taken,

$$\begin{aligned} \text{median}_i(S_{i,3}(\theta_z, \theta_{\text{lat}}, t)) &= a_3(\theta_z, \theta_{\text{lat}}, t) \\ \text{median}_i(S_{i,4}(\theta_z, \theta_{\text{lat}}, t)) &= a_4(\theta_z, \theta_{\text{lat}}, t) \\ \text{median}_i(S_{i,5}(\theta_z, \theta_{\text{lat}}, t)) &= a_5(\theta_z, \theta_{\text{lat}}, t) \end{aligned} \tag{87}$$

This results in an expression for the sensitivity from equation 84,

$$S \approx \text{median}_i \left(\frac{1}{a_3} \left(\frac{\partial f}{\partial T_{i,3}} \right) + \frac{1}{a_4} \left(\frac{\partial f}{\partial T_{i,4}} \right) + \frac{1}{a_5} \left(\frac{\partial f}{\partial T_{i,5}} \right) \right) \tag{88}$$

We propose this method for evaluating algorithm sensitivity based on SST retrievals that are algebraic (rather than based on the physical retrieval).

6 Interpretation and use of l2p_flags

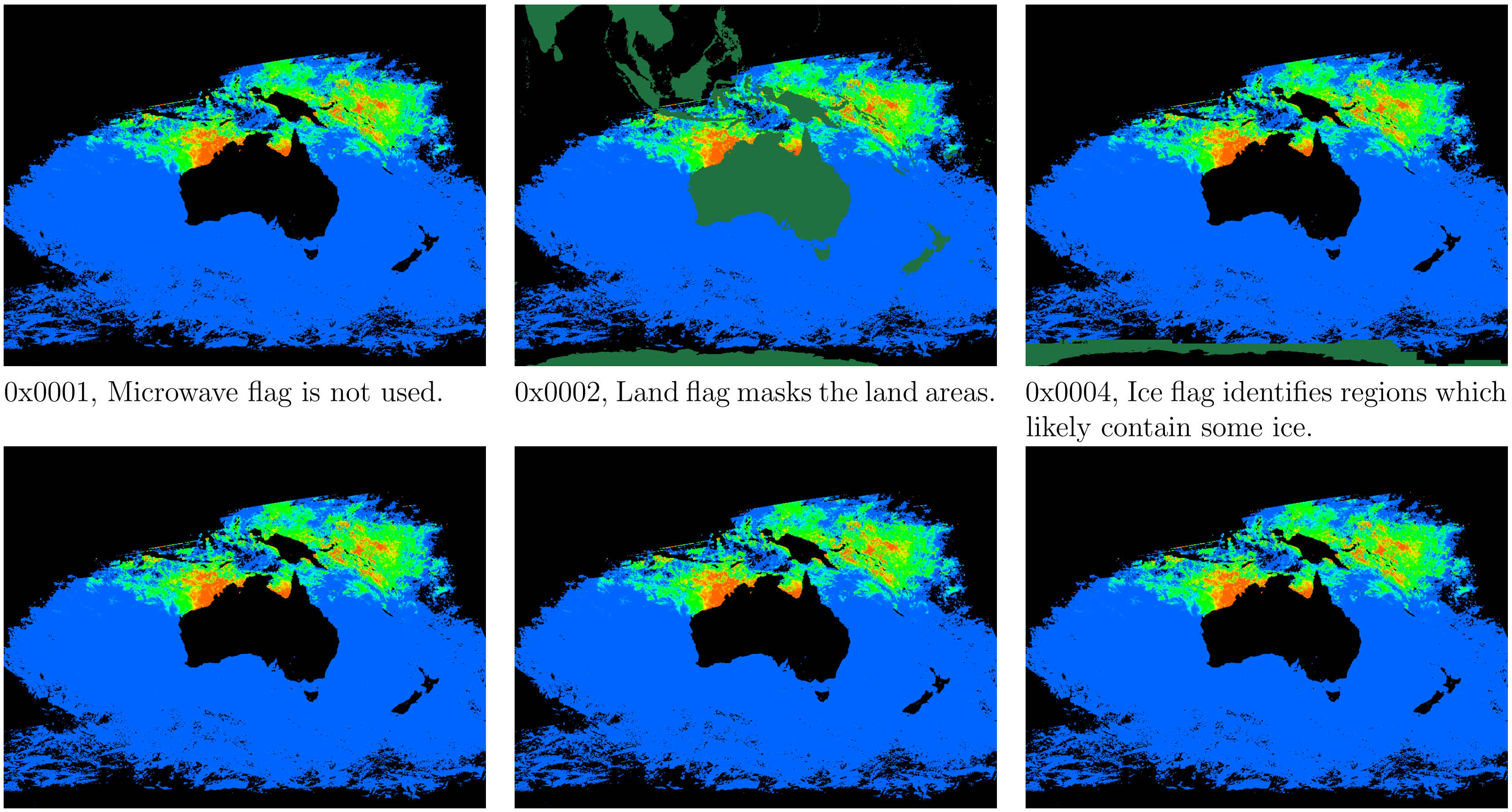
A list of flags used to qualify SST measurements in shown in table 16, Sample appearance of the flags in L3U files are shown in figure 43. Through the merging process, the flags of merged pixels combined using the local OR, a flag is set in the merged data set if it is set in any of the constituent data sets, and cleared if cleared in all of the constituent data sets.

7 SSES model fitting

This section deals with a first attempt to estimate biases and uncertainties between a retrieved SST, and a quality controlled set of *in situ* measurements, which allows said biases and uncertainties to

0x0001	microwave (not set for AVHRR)
0x0002	land - land
0x0004	ice - this region of the sea is covered by ice
0x0008	lake - reserved but not currently set
0x0010	river - reserved but not currently set
0x0020	spare - not currently used
0x0040	aerosol - aerosol value is too large
0x0080	analysis - difference between level 4 analysis and measured SST is large
0x0100	lowwind - NWP wind is low
0x0200	highwind - NWP wind is high
0x0400	edge - pixel is at swath edge (pixel spread is large compared to spread at the center of the swath)
0x0800	terminator - pixel is in solar termination region
0x1000	reflector - pixel is a high amplitude reflection if the surface of the earth was smooth
0x2000	swath - the pixel is likely to be visible by the swath
0x4000	deltadn - day and night sst differs greatly from standard SST, in fileversion 02.0 files, this corresponds to a deviation of 3σ or greater.

Table 16: Flags qualifying Satellite SST measurements. Note: a current deficiency is that the current product data set does not include a flag indicating day / night status.



0x0001, Microwave flag is not used.

0x0002, Land flag masks the land areas.

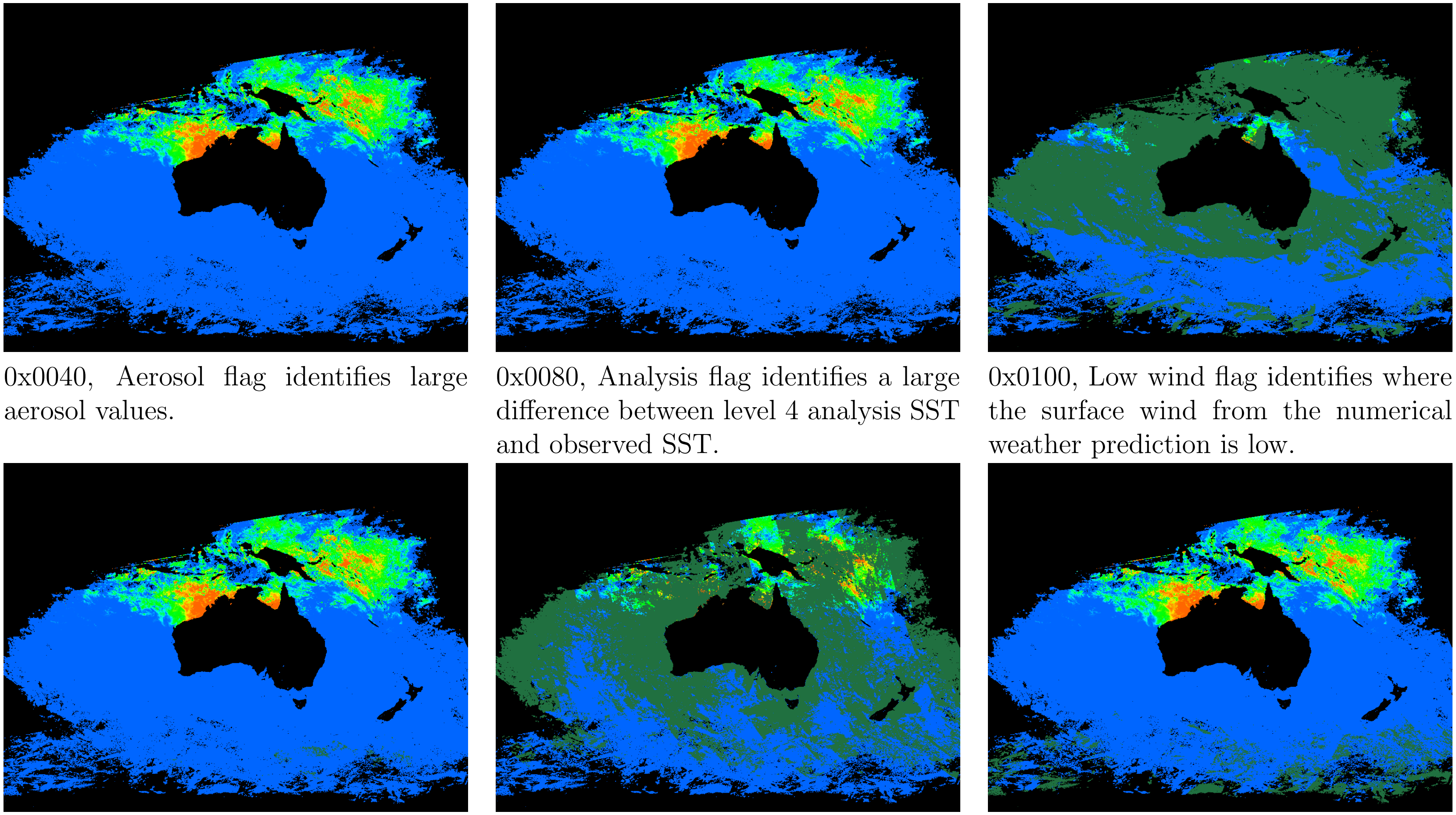
0x0004, Ice flag identifies regions which likely contain some ice.

0x0008, Lake flag is defined to identify lakes, but is not used.

0x0010, River flag is defined to identify rivers, but is not used.

0x0020

Figure 43: Distribution of flags qualifying Satellite SST measurements for a typical L3 file, bits 0 to 4, illustrated on a 14 day multi-platform composite (which due to availability of data consists of NOAA-11 alone), with characteristic time 6th April 1992, 15:20 UTC. Pixels with the appropriate bit field are colored dark green.



0x0040, Aerosol flag identifies large aerosol values.

0x0080, Analysis flag identifies a large difference between level 4 analysis SST and observed SST.

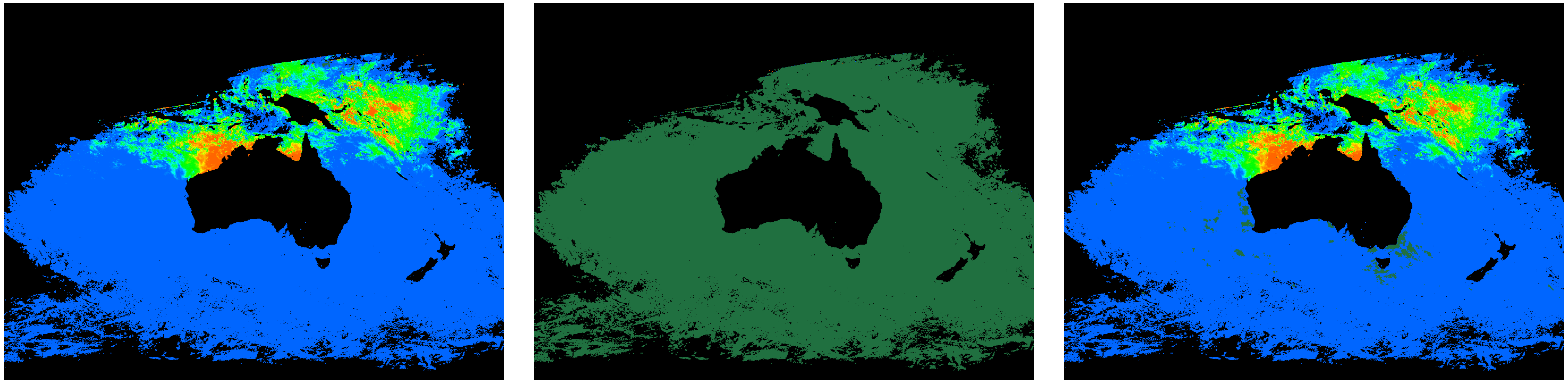
0x0100, Low wind flag identifies where the surface wind from the numerical weather prediction is low.

0x0200, High wind flag identifies where the surface wind from the numerical model prediction is high.

0x0400, Edge flag identifies pixels at the edge of swath (pixel density is low compared to density at the center of the swath).

0x0800, Terminator flag indicates pixels in the solar termination region (at sunrise or sunset).

Figure 44: Distribution of flags qualifying Satellite SST measurements for a typical L3 file, bits 5 to 9, illustrated on a 14 day multi-platform composite (which due to availability of data consists of NOAA-11 alone), with characteristic time 6th April 1992, 15:20 UTC. Pixels with the appropriate bit field are colored dark green.



0x1000, Reflector flag indicates if the pixel would demonstrate a high intensity of solar reflection if the surface was smooth and reflective.

0x2000, Swath flag indicates that the pixel is on a swath. All SST measurements will have this flag set.

0x4000, Retrieval discrepency flag indicates that the day/night SST varies greatly from the standard SST.

Figure 45: Distribution of flags qualifying Satellite SST measurements for a typical L3 file, bits 10 to 14, illustrated on a 14 day multi-platform composite (which due to availability of data consists of NOAA-11 alone), with characteristic time 6th April 1992, 15:20 UTC. Pixels with the appropriate bit field are colored dark green.

be assigned smoothly over the field of view on a pixel by pixel basis. Furthermore, a best estimate of the errors should include both swath dependent anomalies and geographical anomalies, essentially independently, and allow for slow time variation, which is more appropriate on the seasonal scale for geographical anomalies or long term drift of the retrievals.

The discussion begins with a duplicate of section 2.3.

We consider an empirical model for the number of degrees of freedom, n, median bias, μ, and standard deviation, σ, which is seperable in swath $\{n_{\text{swath}}, \mu_{\text{swath}}, \sigma_{\text{swath}}\}$, and geographical components $\{g_n, g_\mu, g_\sigma\}$, as follows,

$$n = n_{\text{swath}} g_n \tag{89}$$

$$\mu = \mu_{\text{swath}} + g_\mu \tag{90}$$

$$\sigma = \sigma_{\text{swath}} g_\sigma \tag{91}$$

We choose the median bias as the basis for our model, because the distribution of the difference between *in situ* and satellite measurements is asymmetric with a larger tail towards cool satellite measurements which becomes more pronounced at lower quality level, see figure 46.

The primary cause of this is atmospheric contamination, such as cloud, where the temperature observed suffers from atmospheric interference rather than being representative of the sea surface. Targeting the median thus provides a more robust representative value that will be close to the mean for high quality pixels and less sensitive to contaminated pixels at lower quality level.

In our basic determination of $\{n_{\text{swath}}, \mu_{\text{swath}}, \sigma_{\text{swath}}\}$, we consider functional dependencies which depend on the first and second harmonics of the day night cycle, an interaction between the day / night cycle and the quality level, and variation over the satellite field of view, modelling systematic biases that relate to the field of view, using the three dimensions,

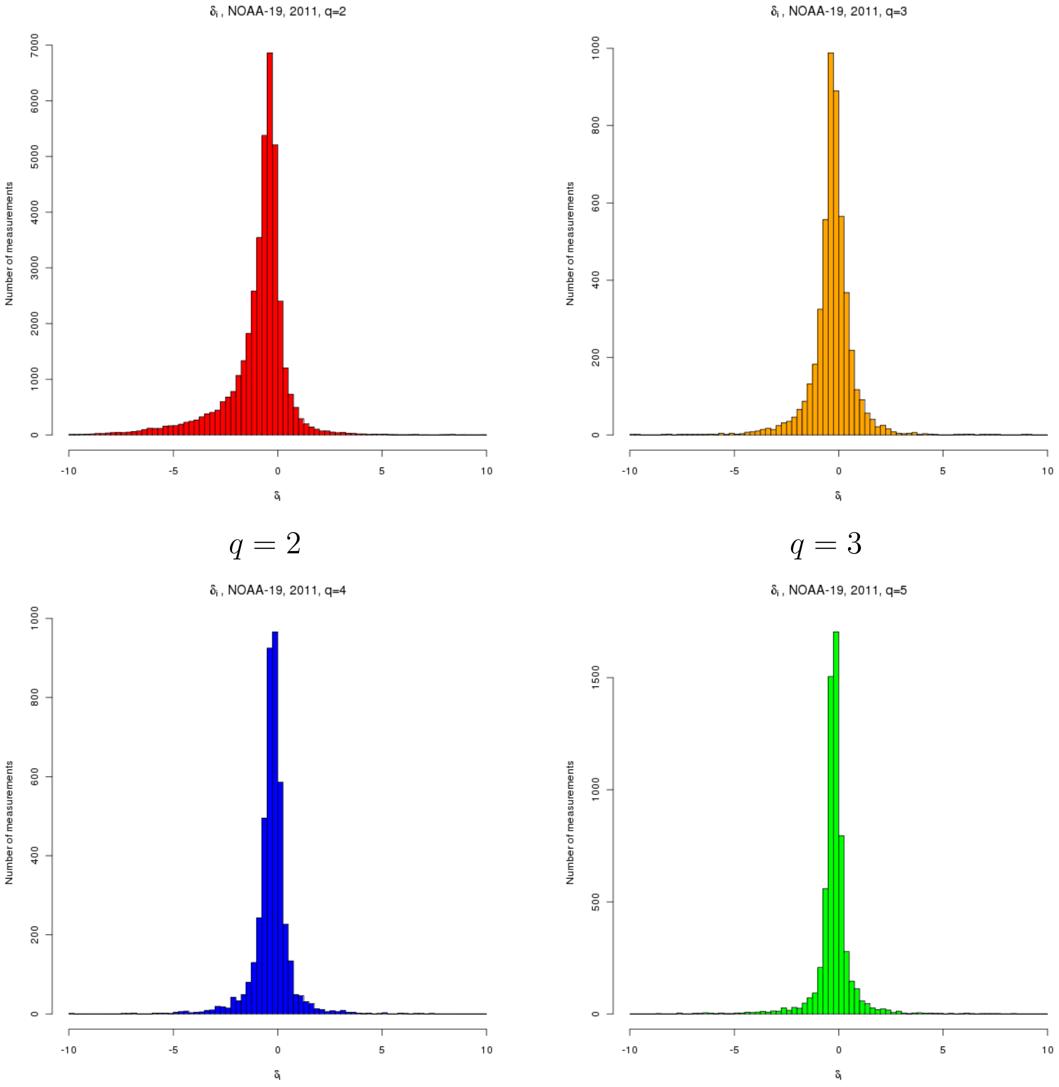
θ_z The satellite zenith angle at the point of observation. Angular dependence on the amount of atmosphere between the sensor and the sea depends on $(\sec\theta_z - 1)$, which is the standard form used to introduce this dimension.

θ_s The sun zenith angle at the point of observation, (corrected so that angles prior to midday are negative). The first two harmonics of the diurnal cycle are represented by the four terms $\cos\theta_s$, $\sin\theta_s$, $\cos^2\theta_s$ and $\cos\theta_s\sin\theta_s$

q The quality level, defined as the distance to detected cloud in kilometres, with a maximum of 5 corresponding to cloud free. Since the result of proximity to cloud is typically to reduce the SST, q is introduced as a number between -1 (lowest quality) and 0 (best quality), using $\left(\frac{q}{5} - 1\right)$.

These pose a natural generalization to the binned approach described in the previous section, where the parameters have assumed dependencies on time of day, view angle and quality level. The advantage in this approach however is the possibility of smooth variation over the time of day and field of view.

This swath determination is corrected by $\{g_n, g_\mu, g_\sigma\}$, which represent interactions with latitude, longitude, quality and time referenced to the current time t_0 at which the model is considered optimal, and in the process introducing the following three dimensions in addition to the quality level,



$q = 2$ $q = 3$

$q = 4$ $q = 5$

Figure 46: Typical Distribution of the difference between *in situ* and satellite measurements for different quality levels. NOAA-19, 2011.

$t - t_0$ The Julian date, in days, offset from the time at which the model is considered optimal which is usually the time at which the last *in situ* measurement was recorded. We assume that this applies linearly, and represents a very low frequency drift.

θ_{lat} The latitude. We choose a polynomial representation of this dimension, since it allows the coupling between this and other dimensions to be introduced more simply. Coupling between latitude and longitude can be modelled with a single $\theta_{\text{lat}}\phi_{\text{lon}}$ term, rather than worrying about two terms that consider the shift in the amplitude and phase of a harmonic function in lat, for example.

ϕ_{lon} The longitude. Since we are dealing with only a small section of the globe, and we wish to introduce dimensional coupling more simply, we use a polynomial representation of this dimension.

When time dependence is modelled, a relatively large set of historical data can be used in the analysis, permitting better statistical estimates, while weighting more recent measurements to allow sensitivity to recent trending behaviour. The initial period of the platform is regressed with a time independent model to ensure that there artifacts due to a reduced data set are minimized.

In its entirety, the model can be represented as shown in equations 92 to 97,

$$\begin{aligned}\log\, n_{\mathrm{swath}} &= a_0\,(b_{0,q} + b_1\cos\theta_s + b_2\sin\theta_s + b_3\cos\theta_s\sin\theta_s \\ &\quad + b_4\Big(\frac{q}{5}-1\Big)\cos\theta_s + b_5\Big(\frac{q}{5}-1\Big)\sin\theta_s + b_6\cos^2\theta_s \\ &\quad - c_0(1-e^{-(\sec\theta_z - 1)})\big) + a_1\end{aligned}\tag{92}$$

$$\begin{aligned}\mu_{\mathrm{swath}} &= \Big(f_0 + f_1\Big(\frac{q}{5}-1\Big) + f_2(\sec\theta_z - 1) + f_3\cos\theta_s + f_4\sin\theta_s \\ &\quad + f_5\Big(\frac{q}{5}-1\Big)(\sec\theta_z - 1) + f_6\Big(\frac{q}{5}-1\Big)\cos\theta_s + f_7\Big(\frac{q}{5}-1\Big)\sin\theta_s \\ &\quad + f_8\cos\theta_s\sin\theta_s + f_9(\sec\theta_z - 1)\cos\theta_s + f_{10}(\sec\theta_z - 1)\sin\theta_s \\ &\quad + f_{11}\Big(\frac{q}{5}-1\Big)^2 + f_{12}(\sec\theta_z - 1)^2 + f_{13}\cos^2\theta_s \\ &\quad + \big(g_0\sigma^2_{\mathrm{swath}} + g_1\big)\big)\,h_1 + h_0\end{aligned}\tag{93}$$

$$\begin{aligned}\sigma^2_{\mathrm{swath}} &= d_0 + d_1\Big(\frac{q}{5}-1\Big) + d_2(\sec\theta_z - 1) + d_3\cos\theta_s + d_4\sin\theta_s \\ &\quad + d_5\Big(\frac{q}{5}-1\Big)(\sec\theta_z - 1) + d_6\Big(\frac{q}{5}-1\Big)\cos\theta_s + d_7\Big(\frac{q}{5}-1\Big)\sin\theta_s \\ &\quad + d_8\cos\theta_s\sin\theta_s + d_9(\sec\theta_z - 1)\cos\theta_s + d_{10}(\sec\theta_z - 1)\sin\theta_s \\ &\quad + d_{11}\Big(\frac{q}{5}-1\Big)^2 + d_{12}(\sec\theta_z - 1)^2 + d_{13}\cos^2\theta_s + (e_1\Big(\frac{q}{5}-1\Big) + e_2)^2\end{aligned}\tag{94}$$

$$\begin{aligned}\log g_n &= \alpha_0 + \alpha_1\theta_{\mathrm{lat}} + \alpha_2\phi_{\mathrm{lon}} \\ &\quad + \alpha_3\theta^2_{\mathrm{lat}} + \alpha_4\phi^2_{\mathrm{lon}} + \alpha_5\theta_{\mathrm{lat}}\phi_{\mathrm{lon}} \\ &\quad + \alpha_6\theta_{\mathrm{lat}}\phi^2_{\mathrm{lon}} + \alpha_7\phi^3_{\mathrm{lon}} + \alpha_8\theta_{\mathrm{lat}}\phi^3_{\mathrm{lon}} + \alpha_9\theta^2_{\mathrm{lat}}\phi^2_{\mathrm{lon}}\end{aligned}\tag{95}$$

$$\begin{aligned}g_\mu &= \beta_0 + \beta_1\theta_{\mathrm{lat}} + \beta_2\Big(\frac{q}{5}-1\Big) + \beta_3(t_0 - t) + \beta_4(t_0 - t)\Big(\frac{q}{5}-1\Big) \\ &\quad + \beta_5\theta_{\mathrm{lat}}\Big(\frac{q}{5}-1\Big) + \beta_6\theta_{\mathrm{lat}}(t_0 - t) \\ &\quad + \beta_7\theta^2_{\mathrm{lat}} + \beta_8\theta^2_{\mathrm{lat}}\Big(\frac{q}{5}-1\Big) + \beta_9\theta^2_{\mathrm{lat}}(t_0 - t)\end{aligned}\tag{96}$$

$$\begin{aligned}\log g_\sigma &= \gamma_0 + \gamma_1\theta_{\mathrm{lat}} + \gamma_2\Big(\frac{q}{5}-1\Big) + \gamma_3(t_0 - t) + \gamma_4(t_0 - t)\Big(\frac{q}{5}-1\Big) \\ &\quad + \gamma_5\theta_{\mathrm{lat}}\Big(\frac{q}{5}-1\Big) + \gamma_6\theta_{\mathrm{lat}}(t_0 - t) \\ &\quad + \gamma_7\theta^2_{\mathrm{lat}} + \gamma_8\theta^2_{\mathrm{lat}}\Big(\frac{q}{5}-1\Big) + \gamma_9\theta^2_{\mathrm{lat}}(t_0 - t)\end{aligned}\tag{97}$$

Greek symbols represent free parameters in the geographical model components of the fit, and roman symbols correspond to the swath or view model components.

The *in situ* data used to fit the empirical model is not necessarily optimally distributed for the application of the model over the full range of view and geographical conditions, with *in situ* measurement devices sparsely represented in the tropics and southern ocean and well represented on the mid latitudes. In an attempt to manage this artifact of measurement, the modelling procedure is progressive, and based on preserving desirable functional forms.

The following details how the coefficients of these model equations are determined.

7.1 SSES determination - fv02 - detailed procedure

The following step by step procedure outlines in detail how model coefficients are determined for the model based SSES estimates used in fv02 processing.

in situ **measurement selection**

Since we are going to use the data set to estimate a standard deviation, and this estimate should be reasonably robust, we wish to remove outliers in the data set that will strongly influence the computation of the standard deviation and are more likely due to technical rather than physical reasons.

We consider unfiltered data from all non zero quality levels, $q \geq 1$, we consider triplets of *in situ* measurements $\{T_{i,\mathrm{analysis}}, T_{i,\mathrm{satellite}}, T_{i,\mathrm{insitu}}\}$, and the three pairwise differences,

$$\begin{aligned} \Delta T_{i,\mathrm{a,s}} &= T_{i,\mathrm{analysis}} - T_{i,\mathrm{satellite}} \\ \Delta T_{i,\mathrm{I,s}} &= T_{i,\mathrm{insitu}} - T_{i,\mathrm{satellite}} \\ \Delta T_{i,\mathrm{I,a}} &= T_{i,\mathrm{insitu}} - T_{i,\mathrm{analysis}} \end{aligned} \tag{98}$$

We retain the data within the middle 99% from the non-parametric distributions of each of these pairwise differences, throwing out no more than 3% of the data set, forming S_{mdb},

$$\begin{aligned} S_{\mathrm{mdb,a,s}} &= \{i : P_{0.5}(\Delta T_{\mathrm{a,s}}) < \Delta T_{i,\mathrm{a,s}} < P_{99.5}(\Delta T_{\mathrm{a,s}})\} \\ S_{\mathrm{mdb,I,s}} &= \{i : P_{0.5}(\Delta T_{\mathrm{I,s}}) < \Delta T_{i,\mathrm{I,s}} < P_{99.5}(\Delta T_{\mathrm{I,s}})\} \\ S_{\mathrm{mdb,I,a}} &= \{i : P_{0.5}(\Delta T_{\mathrm{I,a}}) < \Delta T_{i,\mathrm{I,a}} < P_{99.5}(\Delta T_{\mathrm{I,a}})\} \end{aligned} \tag{99}$$

$$S_{\mathrm{mdb}} = S_{\mathrm{mdb,a,s}} \cap S_{\mathrm{mdb,I,s}} \cap S_{\mathrm{mdb,I,a}} \tag{100}$$

Choosing measurements where the deviation between L4 analysis and satellite is greater than 20% of the maximum of the deviation between *in situ* and satellite and the deviation between *in situ* and L4 analysis

$$S_{\mathrm{mdb,diag}} = \{i : |\Delta T_{i,\mathrm{a,s}}| > 0.2 \max\left(|\Delta T_{i,\mathrm{I,s}}|, |\Delta T_{i,\mathrm{I,a}}|\right)\} \tag{101}$$

in situ measurements are selected for suitability beginning with the basic quality controlled fitted data set for fv02, defined in section 3.6.

Aggregate by $\left(\frac{q}{5} - 1\right)$**,** $(\sec\theta_z - 1)$**,** $\cos\theta_s$**,** $\sin\theta_s$

Aggregation proceeds by binning for each quality level, for $(\sec\theta_z - 1)$ in steps of 0.1, $\cos\theta_s$ in steps of 0.5, $\sin\theta_s$ in steps of 0.5, determining the median, sample size, and standard deviation of $\Delta T_{i,\mathrm{I,s}}$. Additionally standard deviations of $\Delta T_{i,\mathrm{a,s}}$ and $\Delta T_{i,\mathrm{I,a}}$ are also estimated. To improve robustness, only bins with at least 3 measurements, with non zero standard deviation of $\Delta T_{i,\mathrm{I,s}}$, and where the median of $\Delta T_{i,\mathrm{I,s}}$ is less than 10 standard deviations from zero are used for fitting.

Normalize the characteristic *in situ* count over quality levels

The sample size of each quality level is determined, and counts normalized by dividing out the quality level sample size. Thus counts determined in the binning process are replaced with relative normalized counts, which are effectively normalized over different quality levels.

Determine view angle compensation to $\log n$

The normalized counts are characterized in terms of $(\sec\theta_z - 1)$, using a weighted linear model,

$$\log n = A_0 + A_1 e^{-(\sec\theta_z - 1)} \tag{102}$$

The weights chosen, $\frac{\log n}{\sigma^2}$, attempt to trade off those measurements sets with large n with those that are well known (σ^2 is small)[4]

Model the normalized $\log n$ based on $\left(\frac{q}{5} - 1\right)$, $\cos\theta_s$, $\sin\theta_s$

Renormalize $\log n$ to correct for offset and scale discrepancies

Determine the minimum σ based on $\left(\frac{q}{5} - 1\right)$

Fit the residual to $\left(\frac{q}{5} - 1\right)$, $(\sec\theta_z - 1)$, $\cos\theta_s$, $\sin\theta_s$

Model median bias based on σ^2

Fit the median bias residual to $\left(\frac{q}{5} - 1\right)$, $(\sec\theta_z - 1)$, $\cos\theta_s$, $\sin\theta_s$

Aggregate by θ_{lat}, ϕ_{lon}, $(t_0 - t)$

Model the residual $\log n$ geographically

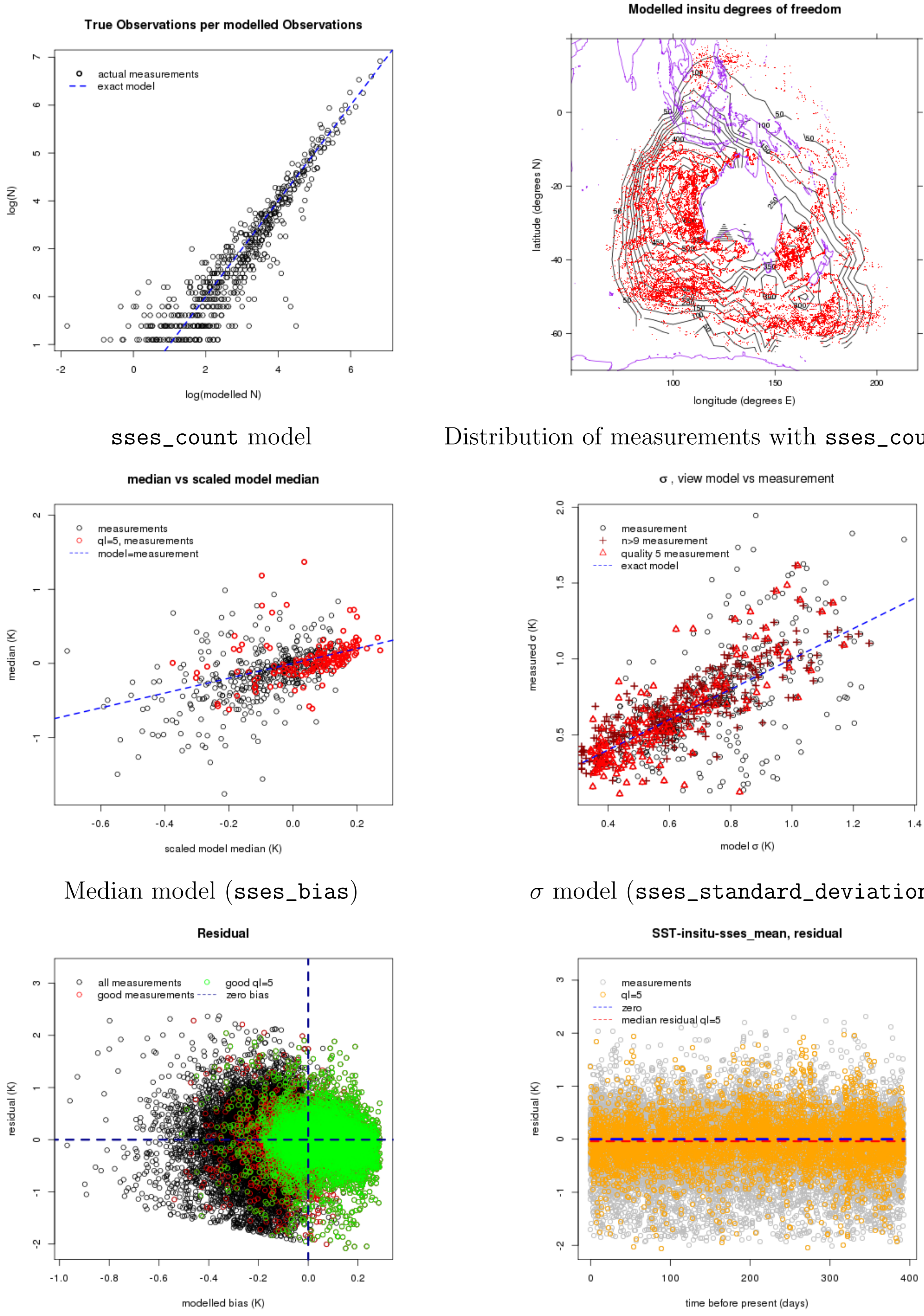
Model the residual median bias geographically

Model the residual σ geographically

7.2 Performance of the SSES model

Figure 47 summarizes the typical model performance. In practise the SSES model is updated every five days so that it can adapt to rapidly changing conditions.

[4] σ is the standard deviation of the residual over the aggregated space.



`sses_count` model

Distribution of measurements with `sses_count` model

Median model (`sses_bias`)

σ model (`sses_standard_deviation`)

Overall residual, $SST - insitu$

Overall residual over time

Figure 47: SSES model performance. Typical behaviour for NOAA-19, from reference date 28th September 2014

8 SST model fitting

SST model calibration requires two *in situ* measurement data sets, the fit data set for determining the optimal SST model, and the verification data set, which is a superset of the fit data set, for verifying it. *In situ* data sets are determined based on the rules outlined in section 3.2.

8.1 Verification data set

The data set used for verification is fv02 Verification data set from table 13 and is determined from the set of *in situ* data chosen by the following rules,

- *in situ* measurement must pass the latitude based quality control, outlined in section 3.3.
- Data must be observed under suitable wind conditions,
- Observations in the verification data set need not be co-located with *in situ* (they can be up to 10km away from each other).
- Satellite observation can be of any quality level $q \geq 2$.

8.2 Fit data set

The data set used for fitting is the fv02 Fit data set from table 13 and is determined from the set of *in situ* data by the following rules (the first two rules are the same as for the verification data set, which assures that the fit data set is always a subset of the verification data set),

- *in situ* measurement must pass the latitude based quality control, outlined in section 3.3.
- Data must be observed under suitable wind conditions,
- *in situ* measurement must pass the analysis based quality control, outlined in section 3.4.
- *In situ* and satellite observations must be co-located.
- Satellite observation must be of quality level $q \geq 4$.
- *In situ* and satellite observations must pass two channel regression quality control, outlined in section 3.5.

8.3 Model types

SST models are determined by regressing polynomial expressions involving the brightness temperatures, $\{T_3, T_4, T_5\}$, and satellite zenith angle, $(\sec\theta_z - 1)$ (and in one case [**paltaglou**], the latitude, $(1 + \cos^2\theta_{\text{lat}})\sec\theta_z)$ to *in situ* measurements on the fit data set. There are broad classes of models employed based on the availability and applicability of brightness temperatures to the retrieval of SST, and the algebraic structure of the underlying models,

Day and Night model

Day and night models are a single model that can be applied to both day and night scenarios. These models are generally based on a single regression between the *in situ* SST and the brightness temperature channels that are applicable to both day and night, channels 4 and 5. SST retrieval is based on regression of a single equation, f,

$$T_{\text{insitu}} = f(T_4, T_5, (\sec\theta_z - 1), \ldots) \tag{103}$$

The advantage that these models have in general is that they can be applied to day and night data and have the same functional form during the day and night by construction. This is desirable if we wish to study diurnal variation, where it is desirable to have both day and night treated the same way. This model (with or without $(\sec\theta_z - 1)$ components) is used when a first guess of SST is required, for quality control and is also considered as a basis for comparison.

Day model, Night model

Day model, Night model are pairs of models that are separately applied to day and night data, making best use of the available information; channels 4 and 5 during the day, and channels 3, 4, and 5 at night. SST retrieval is based on regression of a pair of equations, f_{day} and f_{night}, which could have algebraically different structure,

$$\begin{aligned} T_{\text{insitu,day}} &= f_{\text{day}}(T_4, T_5, (\sec\theta_z - 1), \ldots) \\ T_{\text{insitu,night}} &= f_{\text{night}}(T_3, T_4, T_5, (\sec\theta_z - 1), \ldots) \end{aligned} \tag{104}$$

Maintaining consistency in the interpretation of how the brightness temperatures determine the SST across the day/night boundary is non trivial and cannot in general be easily assured, because the functional forms of f_{day} and f_{night} are different, and fitted coefficients are optimized to each of these scenarios. The advantage that these models have in general is that they use all of the available information at hand, which generally results in considerably more accurate retrievals at night, where the additional T_3 information is available.

Hybrid Day and Night model

Hybrid Day and Night models attempt to blend the advantages of both the previously discussed models, by having a single retrieval equation, f_H that allows consistent regression over both day and night, and allowing T_3 to be used at night to retain the accuracy that this affords for night observations. This is done by considering the nominal dependencies between brightness temperature channel 3 and brightness temperature channels 4 and 5, and then using these nominal dependencies to derive a proxy for channel 3 during the periods where it is not available, using a regression of a functional form f_{T_3},

$$\begin{aligned} T_{3,\text{night}} &= T_3 \\ T_{3,\text{day}} &= f_{T_3}(T_4, T_5, \ldots) \end{aligned} \tag{105}$$

$$\begin{aligned} T_{\text{insitu,day}} &= f_H(T_{3,\text{day}}, T_4, T_5, (\sec\theta_z - 1), \ldots) \\ T_{\text{insitu,night}} &= f_H(T_{3,\text{night}}, T_4, T_5, (\sec\theta_z - 1), \ldots) \end{aligned} \tag{106}$$

The resulting model represents an attempt to harmonize the fits for day and night, utilizing the same functional form while retaining the better performance at night which comes from the additional measured T_3

In our determination, we provide SST computed based on a *Day model, Night model*, and a *Hybrid Day and Night model*, the earlier providing a best SST with the information available, and the later providing an optimal diurnal SST. Differences between these two models are likely to be most pronounced when day time conditions are such that there is a large deviation from the typical night behaviour, which may additionally provide a useful indication that the retrieval may not be reliable due to abnormal atmospheric conditions.

8.4 Model Regression

In an attempt to evaluate model performance, we consider the following *Day and Night models* (with $\{a_i, b_i, c_i\}$ representing the regression coefficients)[7]

$$\begin{aligned} f_{\mathbf{BT45}} &= a_0 + a_1 T_4 + a_2(T_4 - T_5) && (107)\\ f_{\mathbf{BT45SZ}} &= a_0 + a_1 T_4 + a_2(T_4 - T_5) + a_3(T_4 - T_5)(\sec\theta_z - 1) && (108)\\ f_{\mathbf{BT34L4}} &= a_0 + a_1 T_4 + a_2 T_{\text{analysis}}(T_4 - T_5) + a_3(T_4 - T_5)(\sec\theta_z - 1) && (109)\end{aligned}$$

Additionally, we evaluate the *Day model, Night models*, with separate equations for day and night retrieval,[4]

$$\begin{aligned} f_{\mathbf{NLSST},\text{night}} &= a_0 + a_1 T_4 + a_2 T_3(T_3 - T_5) + a_3(\sec\theta_z - 1)\\ f_{\mathbf{NLSST},\text{day}} &= b_0 + b_1 T_4 + b_2 T_4(T_4 - T_5) + b_3(T_4 - T_5)(\sec\theta_z - 1) && (110)\end{aligned}$$

$$\begin{aligned} f_{\mathbf{BT345SZ},\text{night}} &= a_0 + a_1 T_3 + a_2 T_4 + a_3 T_5 + a_4(T_4 - T_5)(\sec\theta_z - 1)\\ f_{\mathbf{BT345SZ},\text{day}} &= b_0 + b_1 T_4 + b_2 T_5 + b_3(T_4 - T_5)(\sec\theta_z - 1) && (111)\end{aligned}$$

$$\begin{aligned} f_{\mathbf{P2011},\text{night}} &= a_0 + a_1 T_3 + a_2 T_4 + a_3 T_5 + a_4(\sec\theta_z - 1) + a_5(\sec\theta_z - 1)(T_3 - T_5)\\ &\quad + a_6(\sec\theta_z - 1)^2 + a_7(\sec\theta_z - 1)^3 && (112)\\ f_{\mathbf{P2011},\text{day}} &= b_0 + b_1 T_4 + b_2 T_4(T_4 - T_5) + b_3(T_4 - T_5)^2 + b_4(1 + \cos^2\theta_{\text{lat}})\sec\theta_z\\ &\quad + b_5(1 + \cos^2\theta_{\text{lat}})\sec\theta_z(T_4 - T_5) + b_6(1 + \cos^2\theta_{\text{lat}})\sec\theta_z T_4\\ &\quad + b_7((1 + \cos^2\theta_{\text{lat}})\sec\theta_z)^2 && (113)\end{aligned}$$

and the following *Hybrid Day and Night models*, using view angle corrected brightness temperatures (which will be explained in further detail in section 8.6),

$$\begin{aligned} f_{\mathbf{VBT345},H} &= a_0 + a_1 T_{3,v} + a_2 T_{4,v} + a_3 T_{5,v} && (114)\\ f_{\mathbf{VBT3SZ},H} &= a_0 + a_1 T_{3,v} + a_2(T_{5,v} - T_{4,v}) + a_3(T_{5,v} - T_{4,v})(\sec\theta_z - 1) && (115)\\ f_{\mathbf{VBT345SZ},H} &= a_0 + a_1 T_{3,v} + a_2 T_{4,v} + a_3 T_{5,v}\\ &\quad + a_4(T_{5,v} - T_{4,v})(\sec\theta_z - 1) + a_5(T_{5,v} - T_{3,v})(\sec\theta_z - 1) && (116)\end{aligned}$$

sharing a common f_{T_3},

$$f_{T_{3,v}} = c_0 + c_1 T_{4,v} + c_2 T_{5,v} + c_3 (T_{4,v} - T_{5,v})^2 + c_4 (T_{4,v} - T_{5,v})^3 + c_5 (T_{4,v} - T_{5,v})^4 \tag{117}$$

`fv01` SST use a platform specific configuration of 110 and 113 to determine SST `sea_surface_temperature`

For fv02, we use equation 110 for our `sea_surface_temperature` estimate, and equation 116 for `sea_surface_temperature_day_night`. See [11] for an example of the use of the hybrid model in diurnal studies using the fv02 dataset over the tropical warm pool.

8.5 Model performance

We consider model performance based on a monthly fit of a two year rolling window of insitu measurements.

8.6 Hybrid Day and Night models.

We have already introduced *Hybrid Day and Night models*, as an attempt to provide the benefits of day / night SST retrievals where additional information is available in brightness temperature channel T_3, without the drawbacks of separate retrieval equations for day and night, with the view of providing a data stream that is optimal for the investigation and measurement of diurnal warming and cooling events. One such retrieval scheme is provided in our GHRSST compliant data set, and is constructed as follows,

- Correct brightness temperatures for satellite view angle dependencies, without resorting to calibration with *in situ* measurements. This should be theoretically possible, because the dependence on satellite view angle is a purely geometrical effect. The resulting brightness temperatures are view corrected, $\{T_{3,v}, T_{4,v}, T_{5,v}\}$.
- Determine the relationship between $T_{3,v}$ and $\{T_{4,v}, T_{5,v}\}$ that characterizes nominal measurement conditions, based on night measurements, when all three channels are expected to characterize SST the best.
- Use the relationship determined to construct a virtual $T_{3,v}$ for day time measurements.
- Regress both day and night measurements, day measurements with the virtual $T_{3,v}$ and night measurements with the measured $T_{3,v}$ against a common regression model.

8.6.1 Correcting brightness temperatures for satellite view angle

In order to correct for the dependency of measured brightness temperature on the satellite view angle, we consider the deviation of the measured brightness temperature from *in situ* measurements at night. To a first order approximation, the optical path length of the measurement lengthens by a factor of magnitude $(\sec\theta_z - 1)$, which corresponds to a first order brightness temperature dependency,

$$T \sim T_0 - \alpha T_0 (\sec\theta_z - 1) \tag{118}$$

where T_0 is the brightness temperature which would have been observed if the same measurement were viewed at nadir. With respect to the global average brightness temperature, T_g,

$$T - T_0 \sim -\alpha T_g \left(1 - \frac{T_0 - T_g}{T_g}\right) (\sec\theta_z - 1) \tag{119}$$

and the second term on the right provides a typically less than 20% correction to the very much simplified dependency,

$$T - T_{\text{insitu}} \approx -\alpha_g(\sec\theta_z - 1) \tag{120}$$

where we have substituted the *in situ* value as the zero order estimate for the nadir measurement. It is possible that the *in situ* measurements have a systematic bias that also depends on $(\sec\theta_z - 1)$, by virtue of the common equator crossing times and the location of the *in situ* measurement devices relative to the satellite motion, and this bias can be easily removed if the compensation is a single global constant independent of T, which is assumed in the above approximation.

Figure 48 demonstrates this relationship, and the regression of the three brightness temperature channels at night. The determination of a global set of corrections, $\{\alpha_{g,3}, \alpha_{g,4}, \alpha_{g,5}\}$, allows an initial compensation of the view angle to be made on all of the measured brightness temperatures, irrespective of their source or value. The relationship between channel differences

$$\begin{aligned}
\Delta T_{54,v} &= T_{5,v} - T_{4,v} && (121)\\
\Delta T_{34,v} &= T_{3,v} - T_{4,v} && (122)\\
\Delta T_{53,v} &= T_{5,v} - T_{3,v} && (123)
\end{aligned}$$

are determined by regressing to polynomials of degrees from 1 to 4 and using the Bayesian Information Criterion [24] to determine the simplest optimal fit, on night observations, assuming $\Delta T_{54,v}$, which is available during the day and night, is the independent variable,

$$\begin{aligned}
\Delta T_{34,v} = &= \kappa_{34,0} + \kappa_{34,1}\Delta T_{54,v} + \kappa_{34,2}\Delta T_{54,v}^2 + \kappa_{34,3}\Delta T_{54,v}^3 + \kappa_{34,4}\Delta T_{54,v}^4 && (124)\\
\Delta T_{53,v} = &= \kappa_{53,0} + \kappa_{53,1}\Delta T_{54,v} + \kappa_{53,2}\Delta T_{54,v}^2 + \kappa_{53,3}\Delta T_{54,v}^3 + \kappa_{53,4}\Delta T_{54,v}^4 && (125)
\end{aligned}$$

Two model estimates of $T_{3,v}$ can be determined, and the average $T_{3,v,\text{est}}$ can be calculated during either the day or night,

$$\begin{aligned}
T_{3,v,A} &= \Delta T_{34,v}(\Delta T_{54,v}) + T_{4,v} && (126)\\
T_{3,v,B} &= \Delta T_{54,v} - \Delta T_{53,v}\,(\Delta T_{54,v}) + T_{4,v} && (127)\\
T_{3,v,\text{est}} &= \frac{1}{2}(T_{3,v,A} + T_{3,v,B}) && (128)
\end{aligned}$$

During the night, the residual of the estimate computed in this way against $T_{3,v}$ shows some $T_{3,v,\text{est}}$ dependency, due to the variability of the number and magnitude of *in situ* measurements over the range of T_3, involved in the characterization, and the possible difference in scale between the brightness temperature and *in situ* SST which is used as the zero order approximate. This is removed simply by application of a regressed self normalization $\{s_0, s_1\}$,

$$T_{3,v,\text{est}} - T_{3,v} = s_0 + s_1 T_{3,v,\text{est}} \tag{129}$$

The required correction is typically a few percent in the scale and a half a degree in the offset at 290K. Figure 49 shows further details. The resulting estimate for $T_{3,v,\text{est}}$, is used to supplement $T_{4,v}$ and $T_{5,v}$ during the day, with the observed $T_{3,v}$ still in use for night observations.

8.7 Fixed retrieval regression - fv01

A summary of the retrievals performed on the satellites from fv01 can be found in [**paltaglou**].

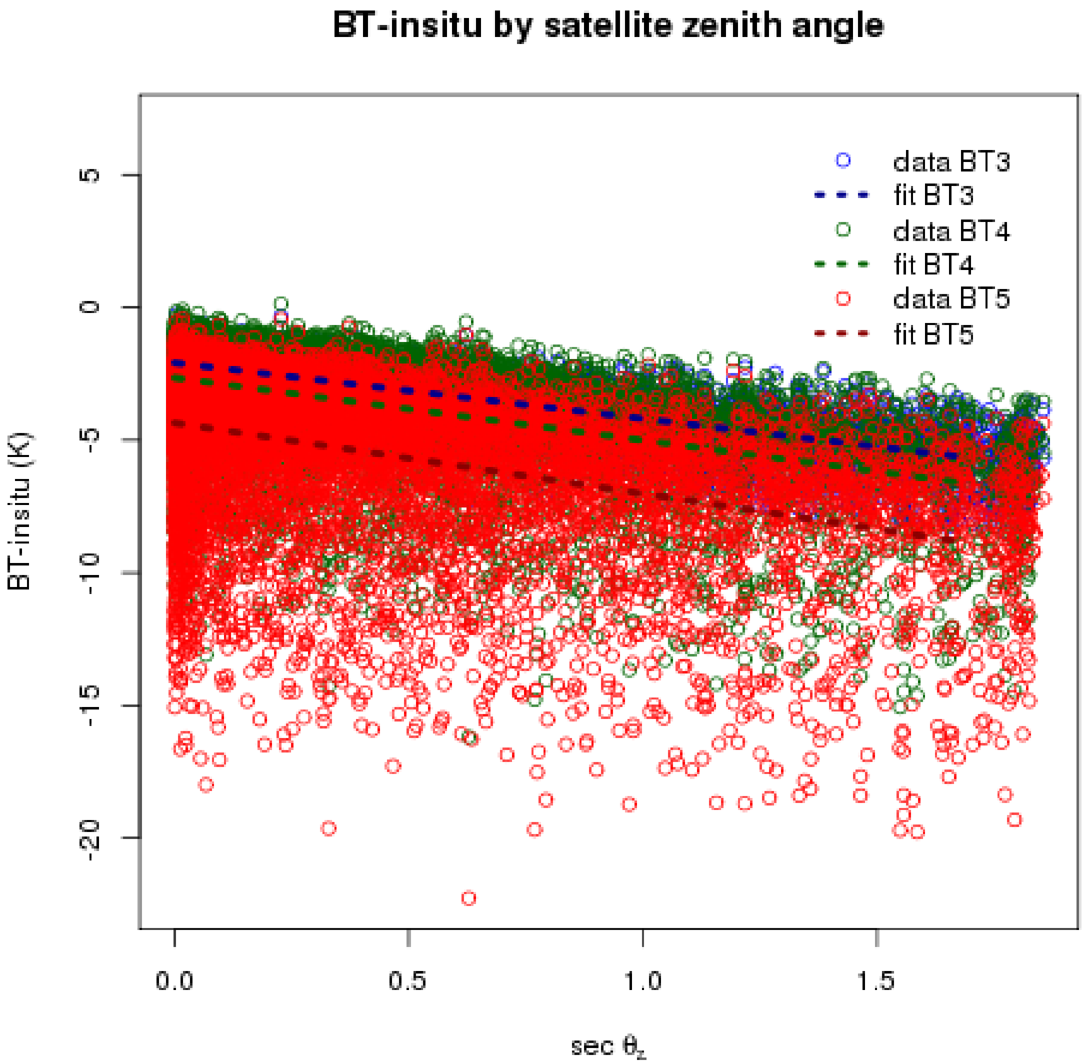
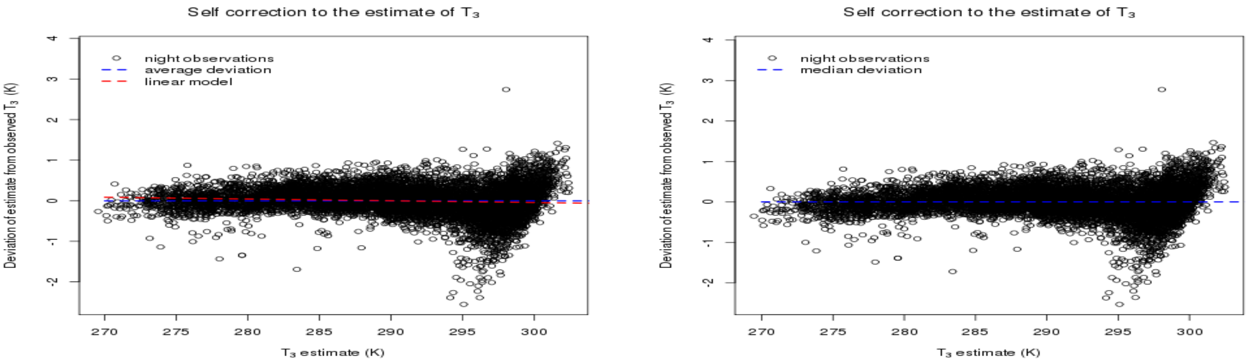


Figure 48: θ_z dependency for NOAA-19 brightness temperature measurements. The difference between *in situ* and brightness temperature for the best quality night matches over 2 years prior to 1^{st} October 2014. This is indicative of the systematic correction that can be made to the 3 AVHRR brightness temperature channels.



before correction

after correction

Figure 49: Linear self correction to $T_{3,v,\text{est}}$, to compensate for uniform coverage in *in situ* data.The figures show the role the compensation has in rebalancing *in situ* biases, for NOAA-19 over the two years up to and including 1st October 2014.

8.8 Progressive retrieval regression - fv02

It is expected that satellite retrievals potentially show some slow time dependence, which is generally measured on the scale of years, due to instrument degradation as well as slow variation in the operating environment (one obvious source for earlier platforms is the change of equator crossing time). This information can accommodated for in a time continuous regression scenario, where the satellite is regressed periodically against *in situ* , to ensure that the impact of this dependence is compensated for over time. In this section, we consider such scenarios, by regressing the models introduced earlier over a fixed window.

We evaluate the performance by regressing *in situ* SST to brightness temperatures over a running window, which is typically on the scale of years, evaluating the bias to *in situ* over a much smaller current window, which is typically on the scale of months, as well as the ability for the same regressed relationship to be used over the same duration current future window. The performance on the current past compared to the current future are directly compared and this is used as a measure of how well algorithms can be dynamically tuned.

Assessment of the Bayesian Information Criterion [24] for each model over time allows comparison between models that respects the trade-off between accuracy and complexity. →*TODO:***A lot more comments based on the results here. May need to justify postfacto why BOM12010 is no longer in vogue. Note that we cannot use the sum of the BIC for day and night as a comparison. We need to consider day and night separately.**← The following outlines the detailed procedure used to determine retrieval algorithms for progressive retrieval regression.

***in situ* measurement selection**

In situ measurement selection follows the procedure for the fitted data selection outlined in section 3.2.

Apply 2 channel regression quality control

The day cycle is divided into day and night measurement sets. Satellite brightness temperatures are regressed against *in situ* SST, weighted inversely with the number of measurements

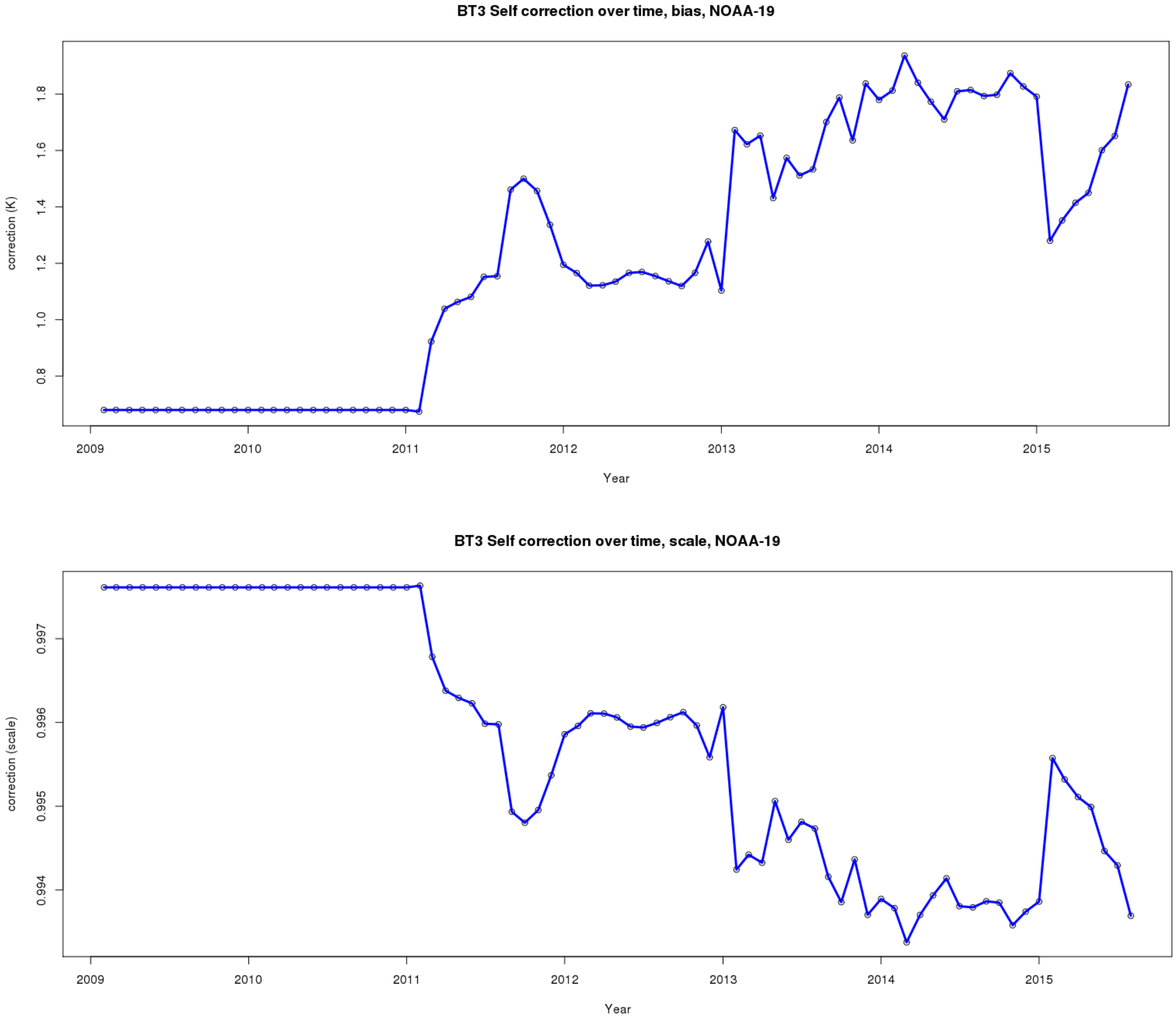


Figure 50: Self correction to $T_{3,v,\mathrm{est}}$, to compensate for uniform coverage in *in situ* data for NOAA-19.The figures show the compensation bias and scale required over an extended period of time.

in the same day or night. Quality control is performed on the residuals, as outlined in section 3.5.

Fit BT45SZ

The optimal **BT45SZ** model is determined based on the quality controlled data, inversely weighted by the number of samples for day and night, per equation 55. The fitting coefficients are stored although this algorithm is not used in producing SST. This represents a cross check on the distribution of residuals and expected variation of the quality controlled measurements. Performance is reported on the Verification dataset constructed as outlined in section 3.6.

Fit BT4BT5

The optimal **BT4BT5** model is determined based on the quality controlled data, inversely weighted by the number of samples for day and night, per equation 55. The fitting coefficients are stored although this algorithm is not used to produce SST. Comparing with **BT45SZ** allows some assessment of the impact of the sun zenith angle. Performance is reported on the Verification dataset constructed as outlined in section 3.6.

Fit PF1998

The optimal **PF1998** model is determined based on the quality controlled data, inversely weighted by the number of samples for day and night, per equation 55. The fitting coefficients are stored although this algorithm is not used to produce SST. Performance is reported on the Verification dataset constructed as outlined in section 3.6.

Fit NLSST

The optimal non linear SST (**NLSST**) model is determined based on the quality controlled data. Different Algorithms are used for day and night, per equation 110. Because different algorithms are used for day and night, the weights for measurements are equal. This algorithm is used to provide `sea_surface_temperature` estimates. Performance is reported on the Verification dataset constructed as outlined in section 3.6.

Fit P2011

The legacy fv01 model **P2011** is determined based on the quality controlled data. Different Algorithms are used for day and night, per equation 113. Because different algorithms are used for day and night, the weights for measurements are equal. The fitting coefficients are stored although this algorithm is not used to produce SST. Performance is reported on the Verification dataset constructed as outlined in section 3.6.

Fit BT345SZ

The linear model **BT345SZ** is determined based on the quality controlled data. Different Algorithms are used for day and night, per equation 113. Because different algorithms are used for day and night, the weights for measurements are equal. The fitting coefficients are stored although this algorithm is not used to produce SST. Performance is reported on the Verification dataset constructed as outlined in section 3.6.

Fit Hybrid models VBT345, VBT3SZ, VBT345SZ

The optimal hybrid 3 channel models are determined based on the quality controlled data. The same algorithm is used for day and night, based on a three channel retrieval. Since there is contamination in one of the channels during the day, the third day channel is simulated based on historical night performance. The algorithm to determine the simulated channel is discussed in detail in section 8.6. The fit for **VBT345SZ** is used to provide `sea_surface_temperature_day_night` estimates. Performance is reported on the Verification dataset constructed as outlined in section 3.6.

8.9 NOAA-11 performance

Figures 51, and 52 contains the details about algorithm tuning for NOAA-11. Although we had just almost years navigated data, the number of day and night *in situ* matches was more balanced, and larger than NOAA-09. None the less, lower means and medians for the 3 channel non-linear day / night model, with better sensitivity performance, and lower BIC, show a similar conclusion.

8.10 NOAA-12 performance

Figures 53, and 54 contains the details about algorithm tuning for NOAA-12.

8.11 NOAA-14 performance

Figures 55, and 56 contains the details about algorithm tuning for NOAA-14.

8.12 NOAA-15 performance

Figures 57, and 58 contains the details about algorithm tuning for NOAA-15.

8.13 NOAA-16 performance

Figures 59, and 60 contains the details about algorithm tuning for NOAA-16.

8.14 NOAA-17 performance

Figures 61, and 62 contains the details about algorithm tuning for NOAA-17.

8.15 NOAA-18 performance

Figures 63, and 64 contains the details about algorithm tuning for NOAA-18.

8.16 NOAA-19 performance

Figures 65, and 66 contains the details about algorithm tuning for NOAA-19.

9 GHRSST Compliant file format.

The files produced are fully GHRSST 2.0r5 compliant. The relevant specification, [29] provides a reference to common features of these files, and should be use as a primary reference.

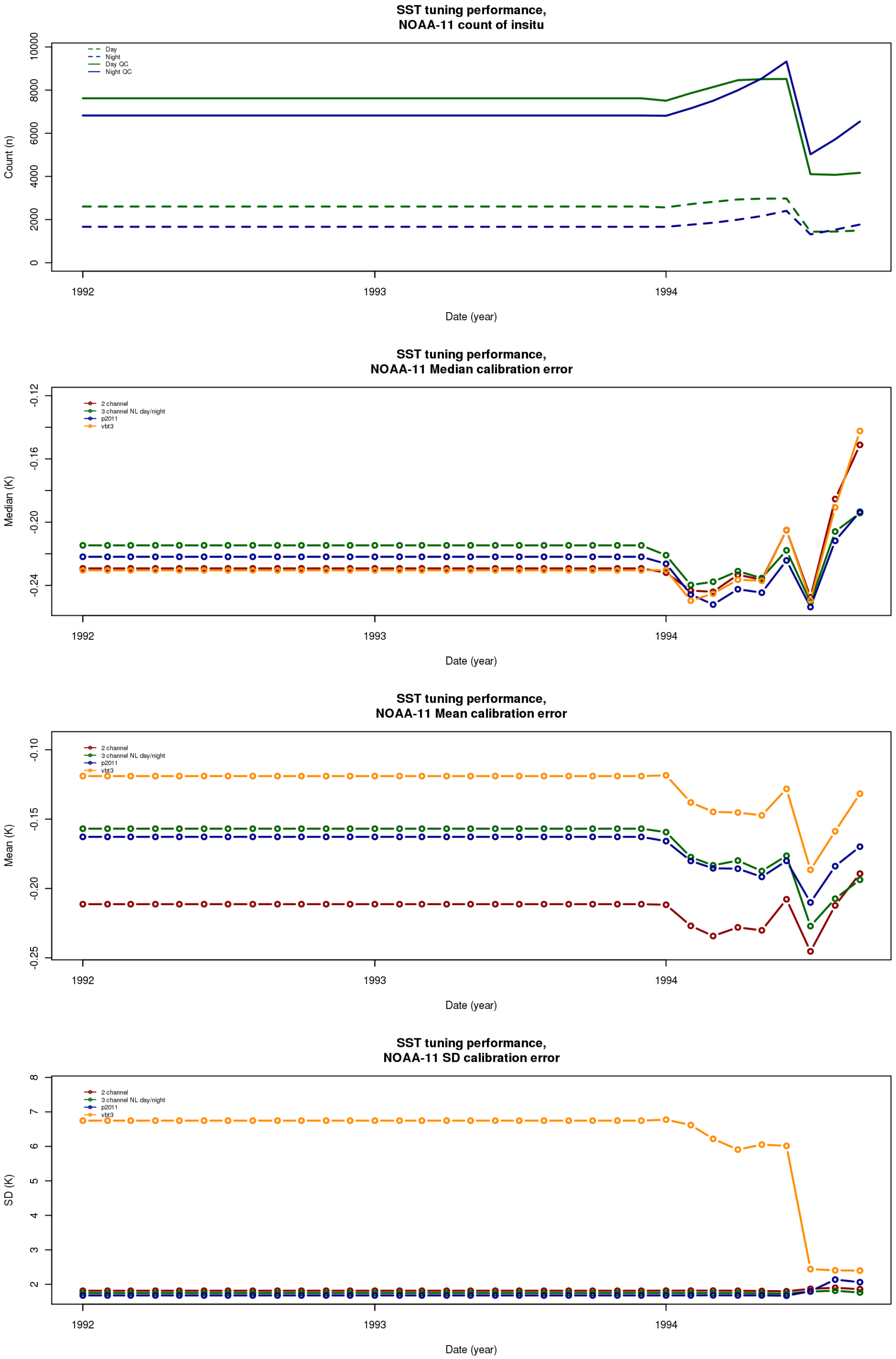
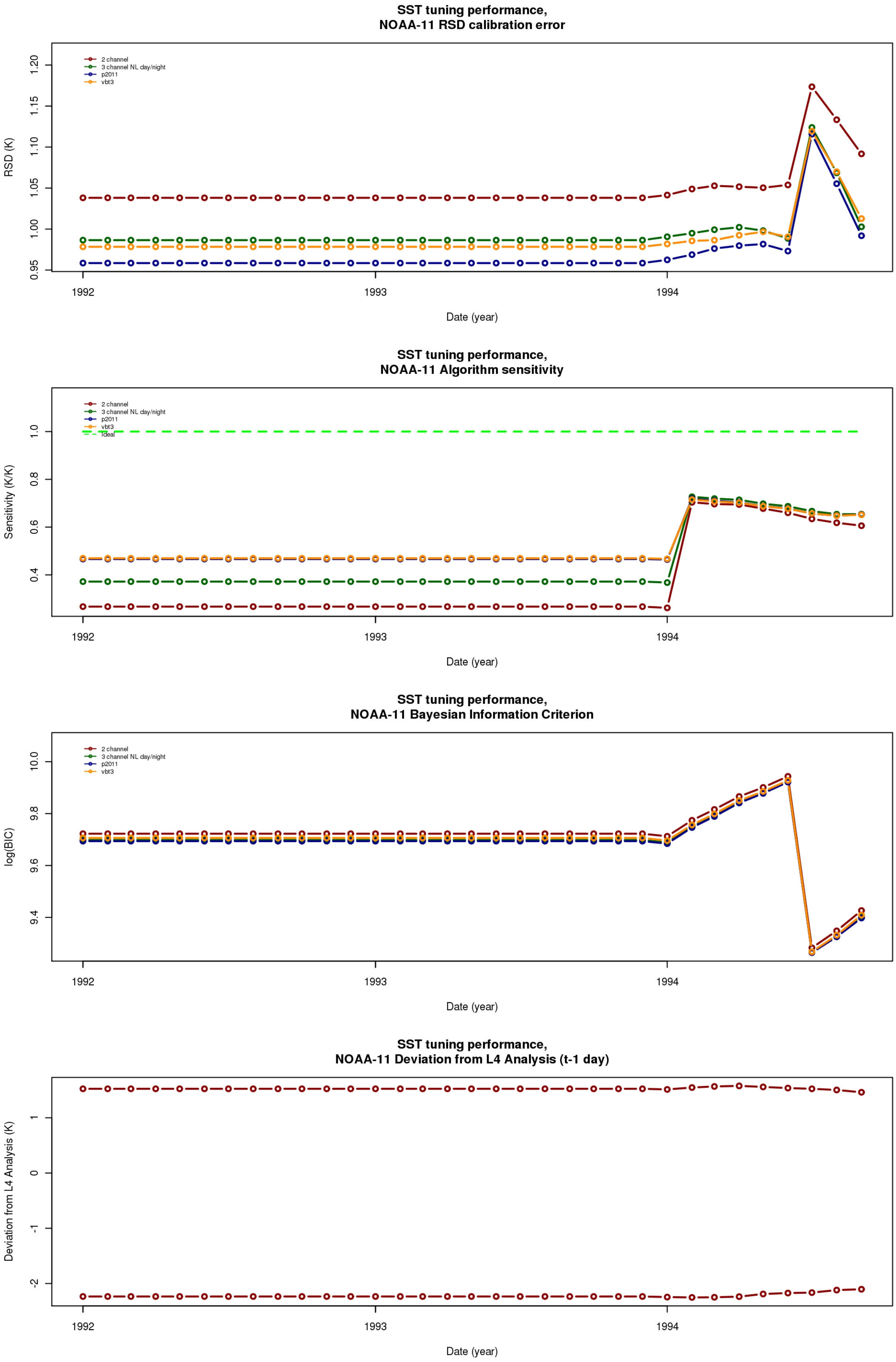


Figure 51: NOAA-11 Regression results over baseline dataset, comparative algorithm details

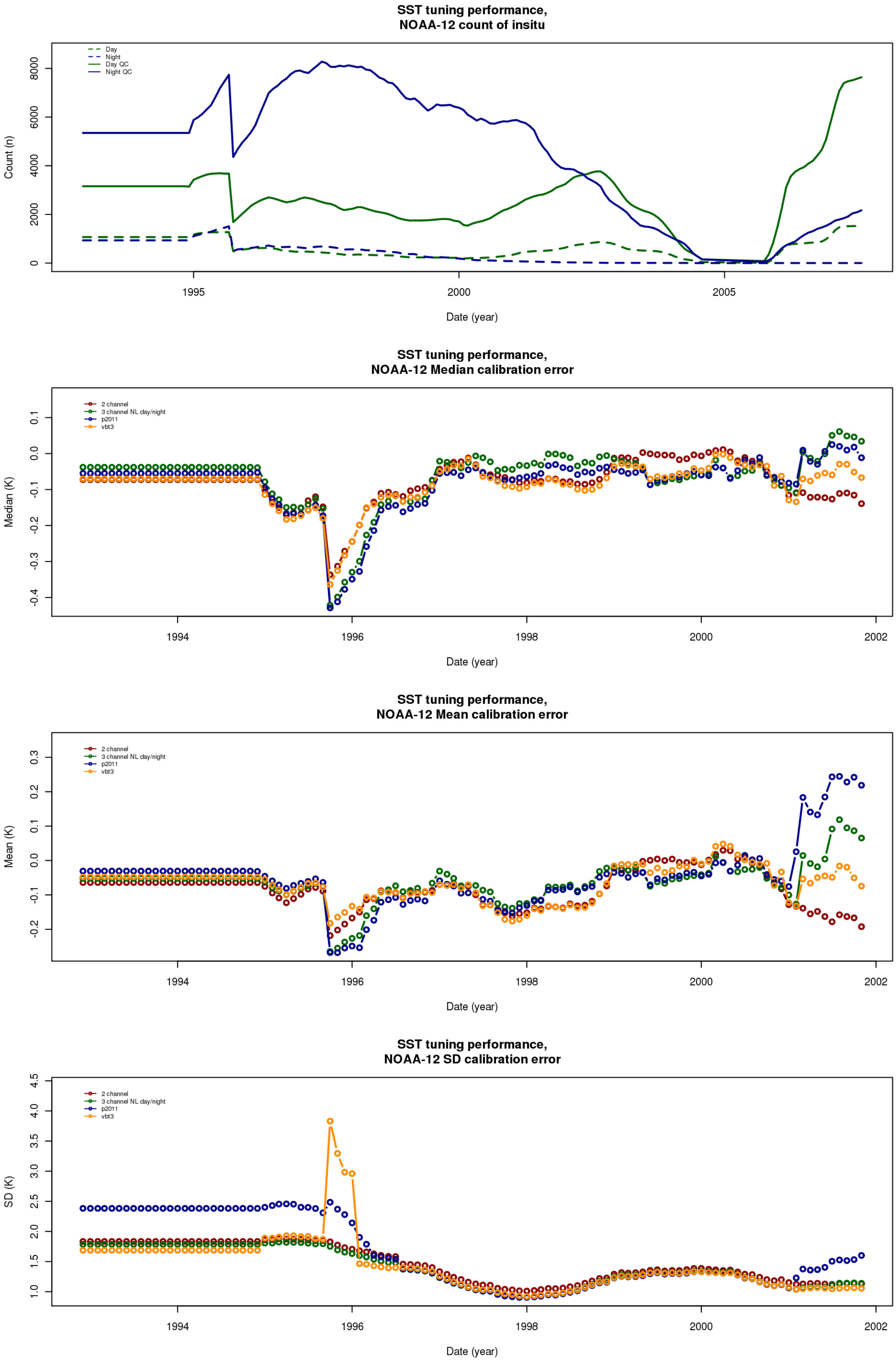

SST tuning performance,
NOAA-11 RSD calibration error

SST tuning performance,
NOAA-11 Algorithm sensitivity

SST tuning performance,
NOAA-11 Bayesian Information Criterion

SST tuning performance,
NOAA-11 Deviation from L4 Analysis (t-1 day)

Figure 52: NOAA-11 Regression results over baseline dataset, comparative algorithm performance metrics



SST tuning performance,
NOAA-12 count of insitu

SST tuning performance,
NOAA-12 Median calibration error

SST tuning performance,
NOAA-12 Mean calibration error

SST tuning performance,
NOAA-12 SD calibration error

Figure 53: NOAA-12 Regression results over baseline dataset, comparative algorithm details

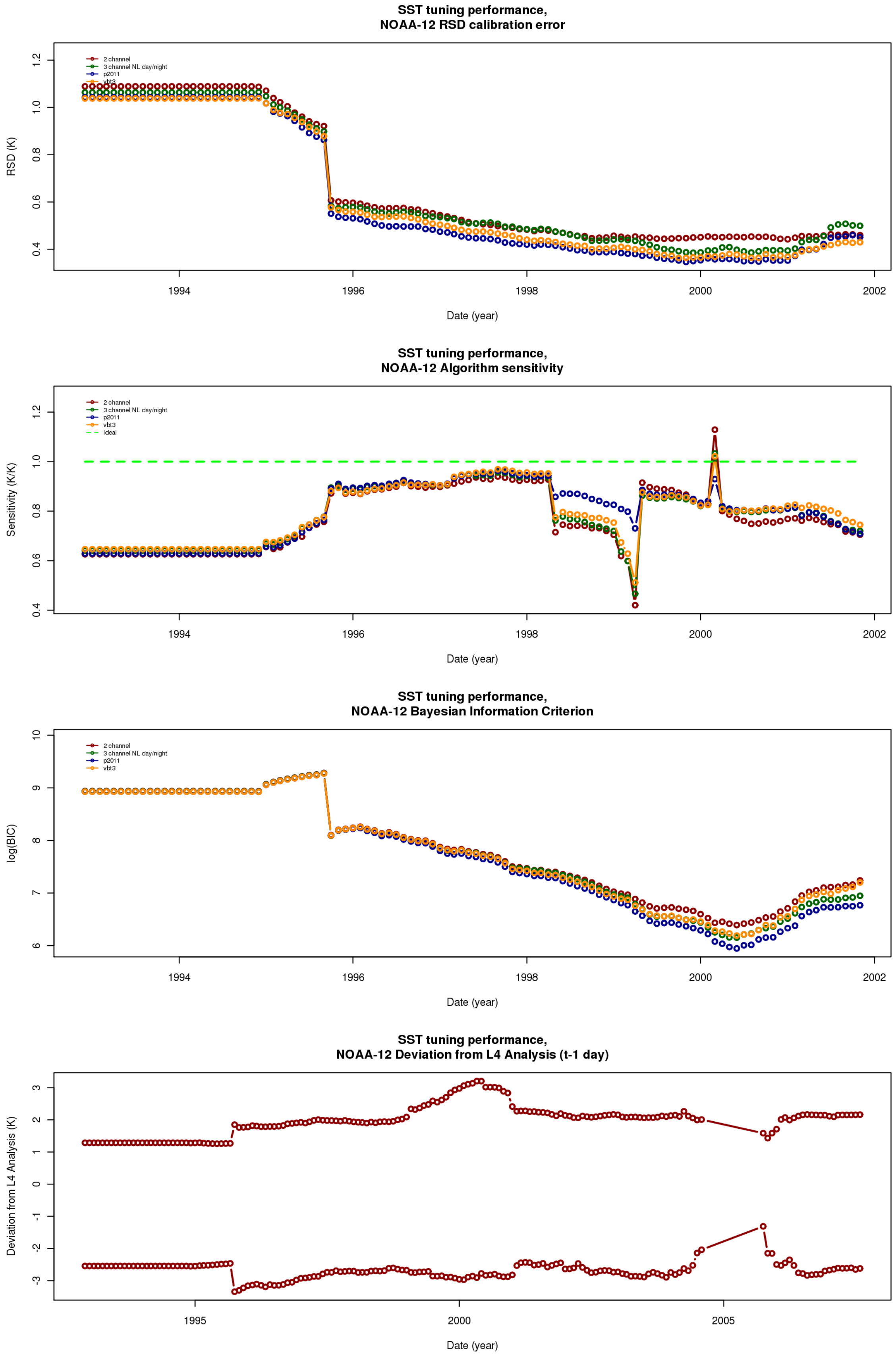
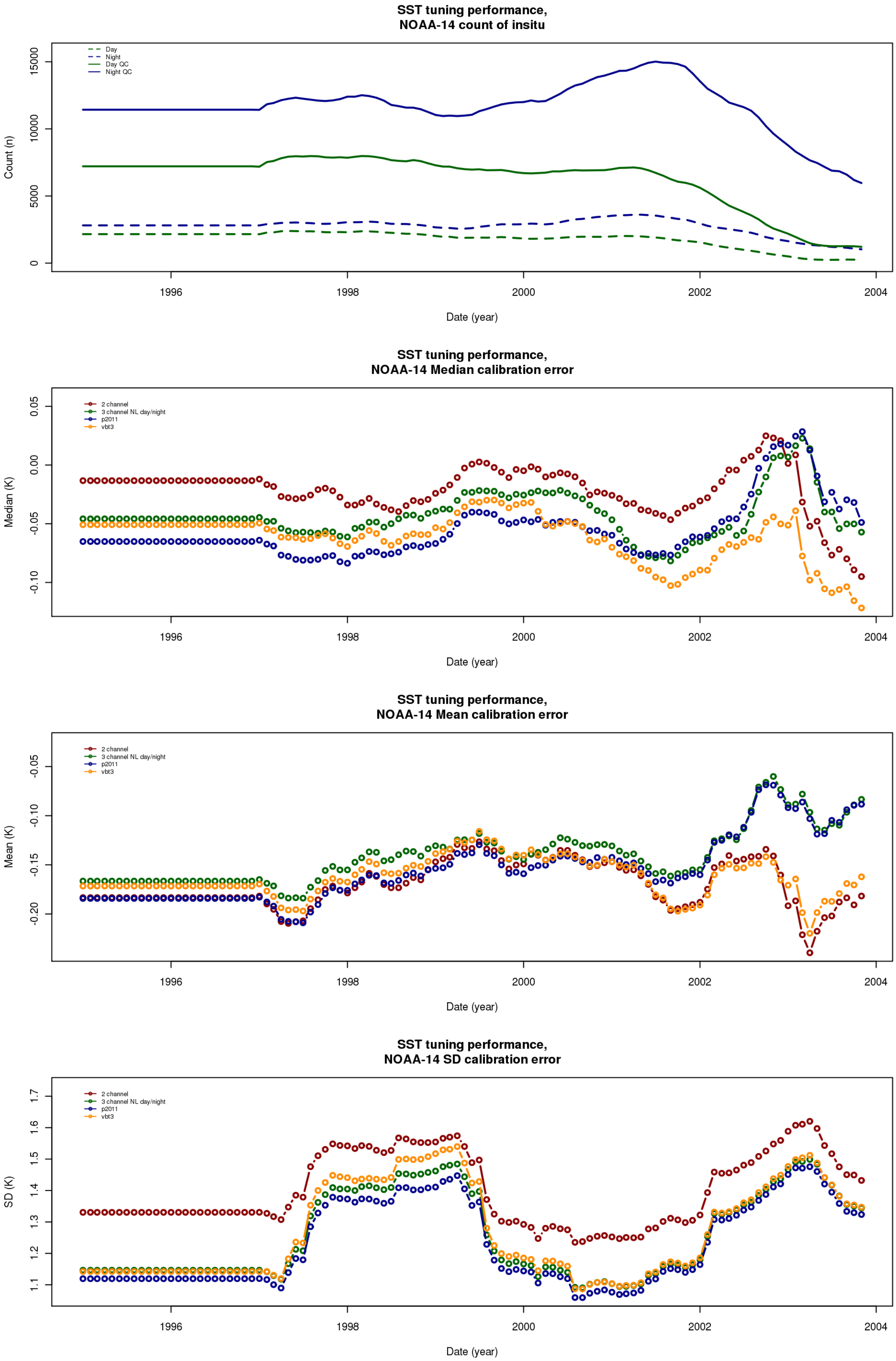


Figure 54: NOAA-12 Regression results over baseline dataset, comparative algorithm performance metrics



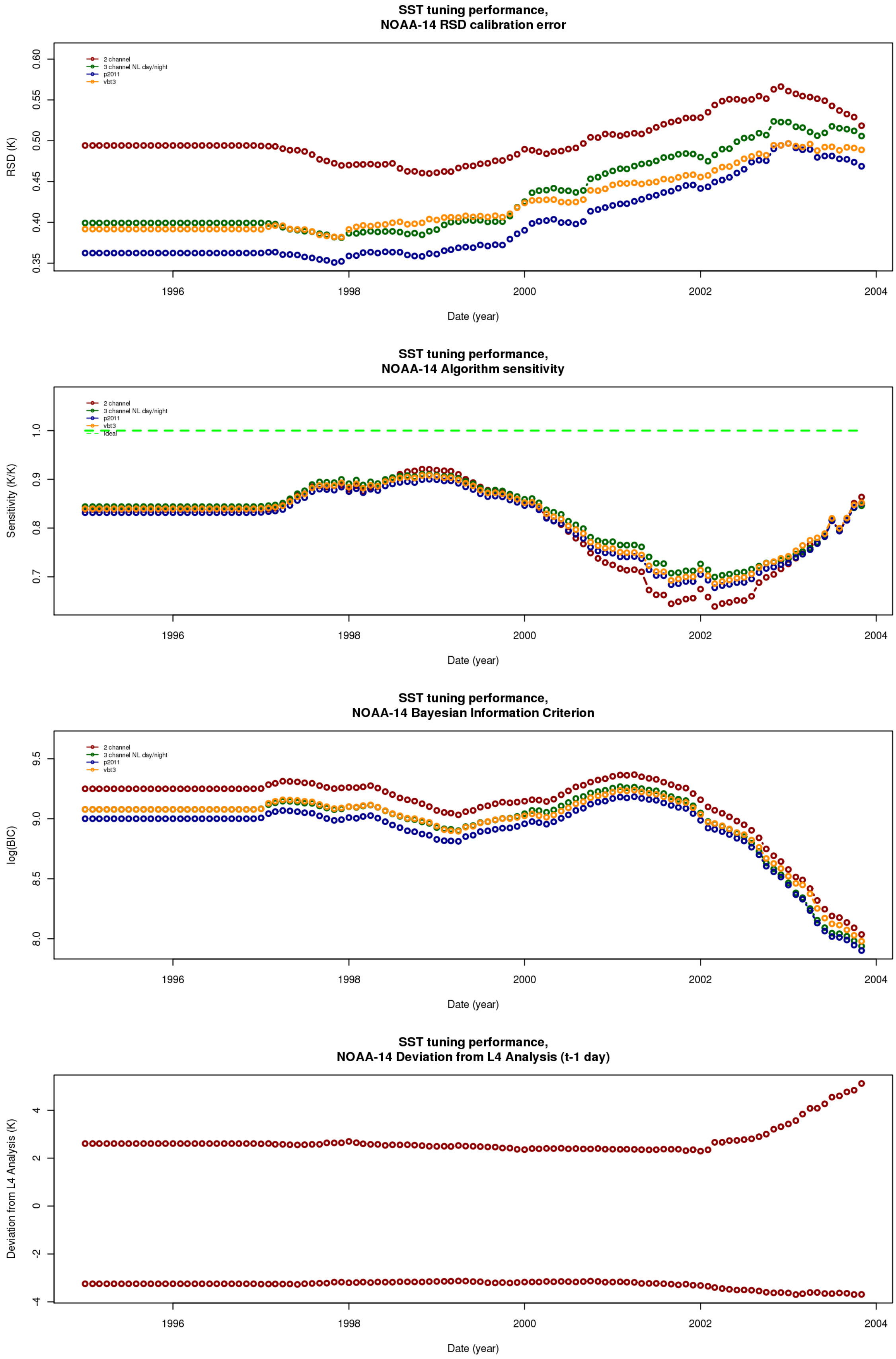
SST tuning performance,
NOAA-14 count of insitu

SST tuning performance,
NOAA-14 Median calibration error

SST tuning performance,
NOAA-14 Mean calibration error

SST tuning performance,
NOAA-14 SD calibration error

Figure 55: NOAA-14 Regression results over baseline dataset, comparative algorithm details



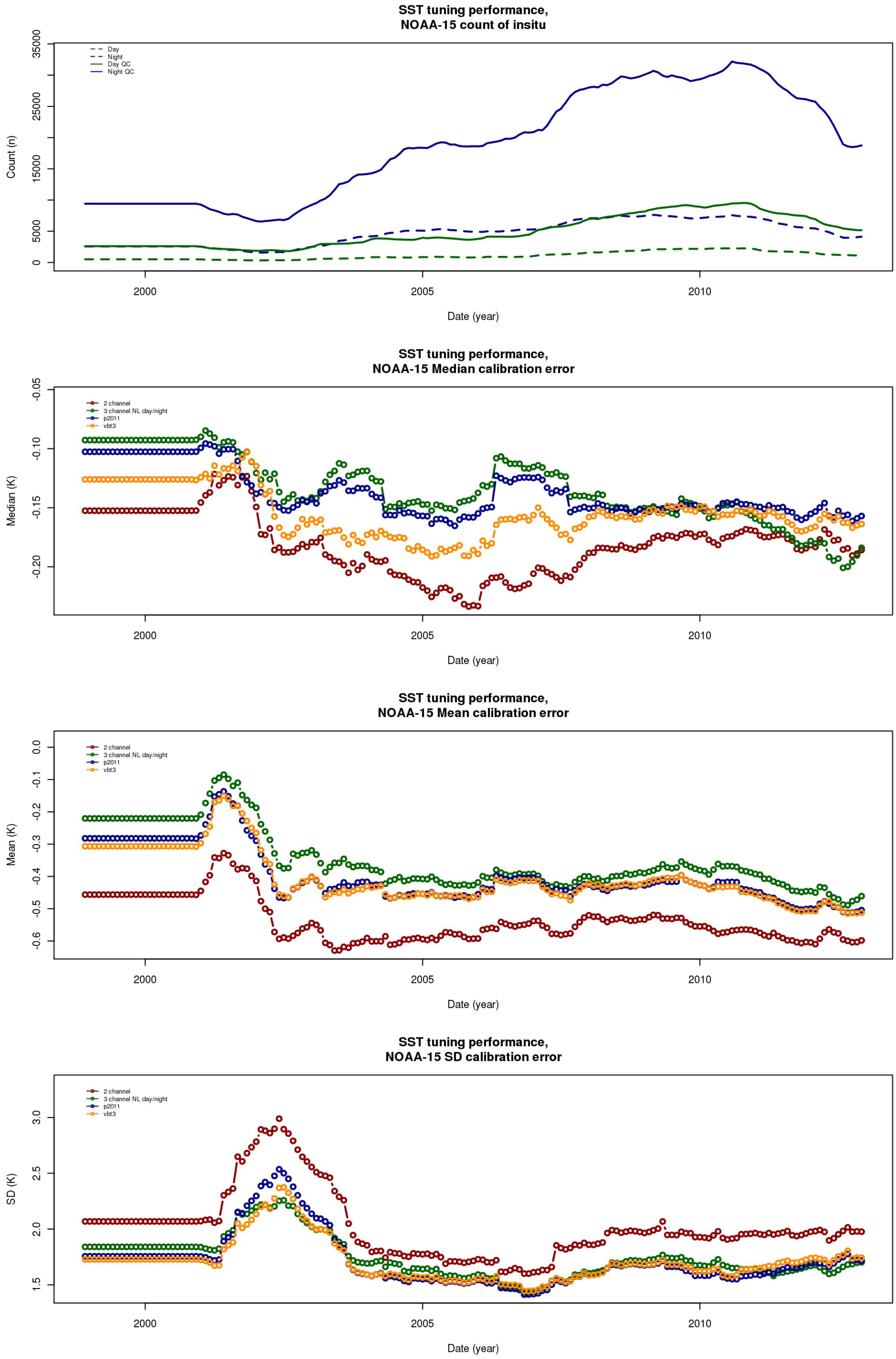
SST tuning performance,
NOAA-14 RSD calibration error

SST tuning performance,
NOAA-14 Algorithm sensitivity

SST tuning performance,
NOAA-14 Bayesian Information Criterion

SST tuning performance,
NOAA-14 Deviation from L4 Analysis (t-1 day)

Figure 56: NOAA-14 Regression results over baseline dataset, comparative algorithm performance metrics


SST tuning performance,
NOAA-15 count of insitu

SST tuning performance,
NOAA-15 Median calibration error

SST tuning performance,
NOAA-15 Mean calibration error

SST tuning performance,
NOAA-15 SD calibration error

Figure 57: NOAA-15 Regression results over baseline dataset, comparative algorithm details

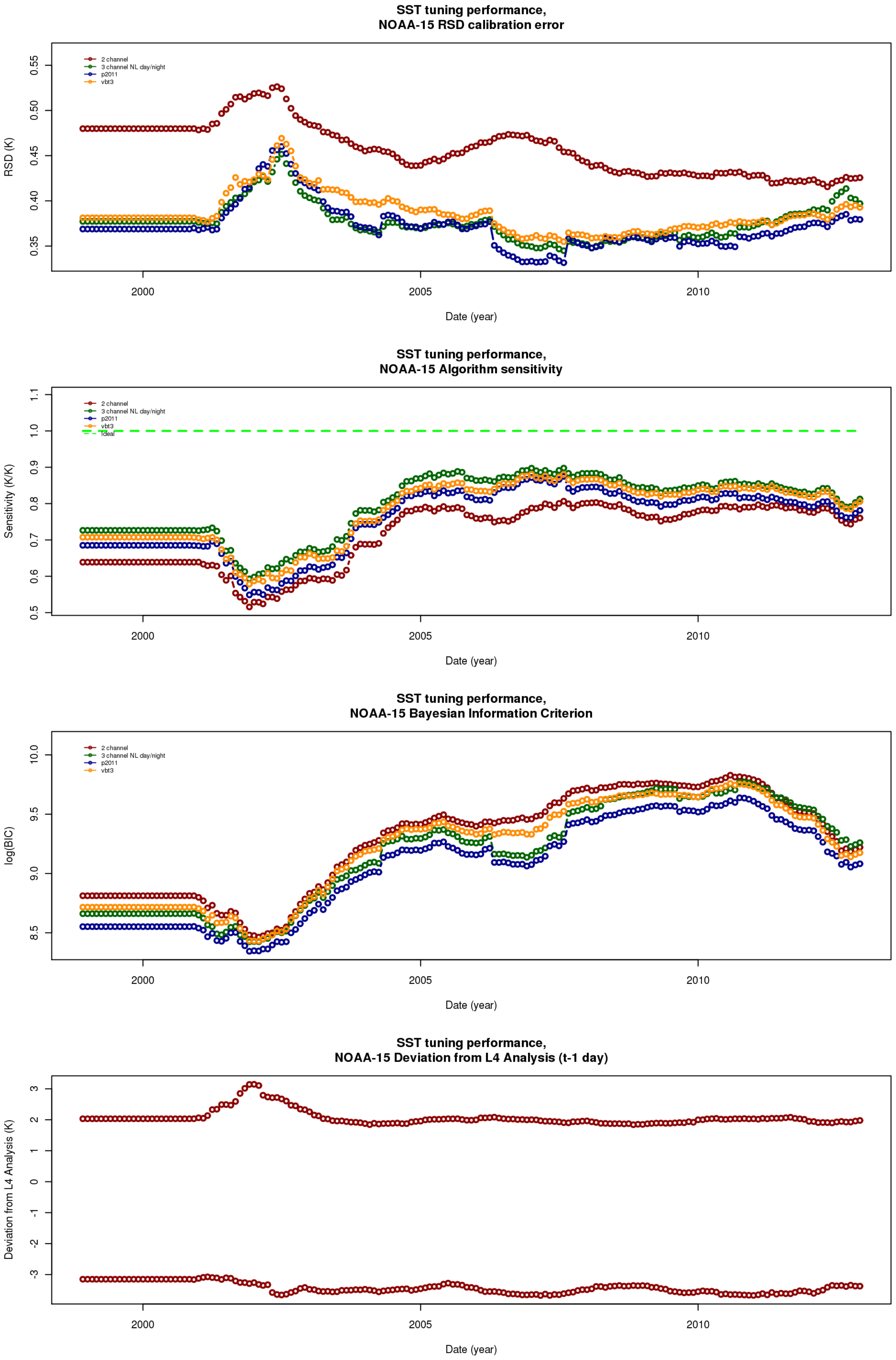


Figure 58: NOAA-15 Regression results over baseline dataset, comparative algorithm performance metrics

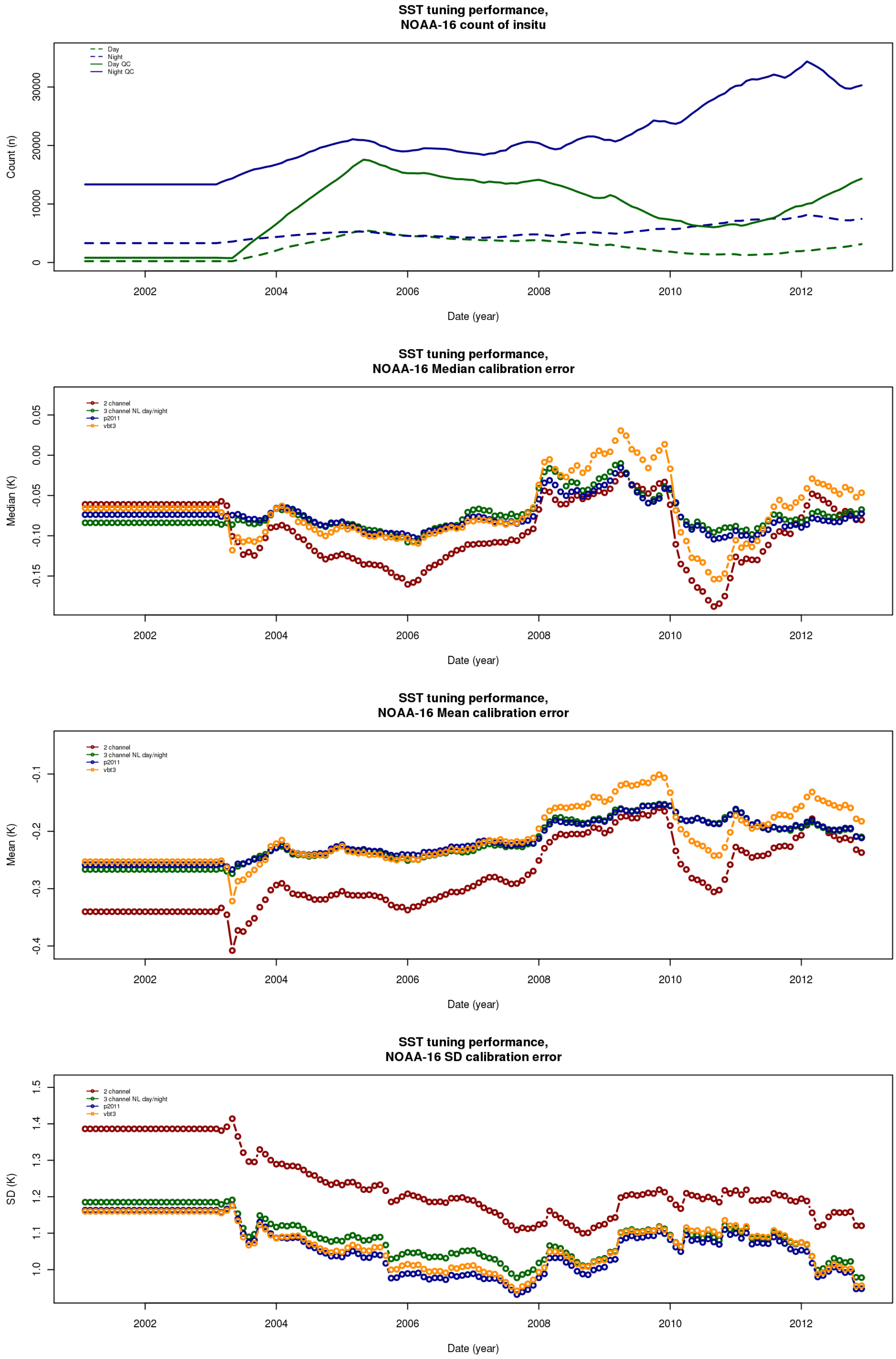


Figure 59: NOAA-16 Regression results over baseline dataset, comparative algorithm details

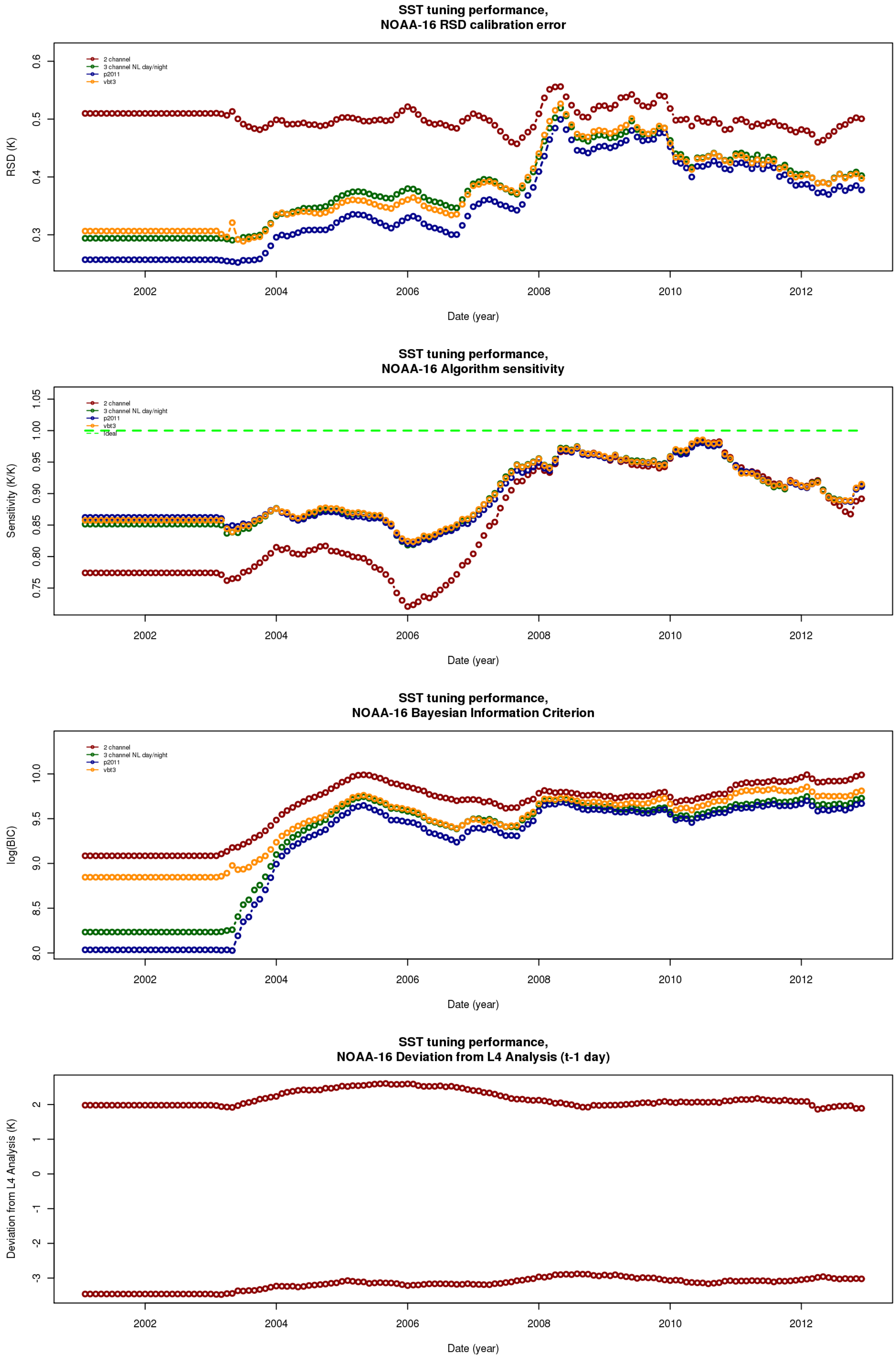


Figure 60: NOAA-16 Regression results over baseline dataset, comparative algorithm performance metrics

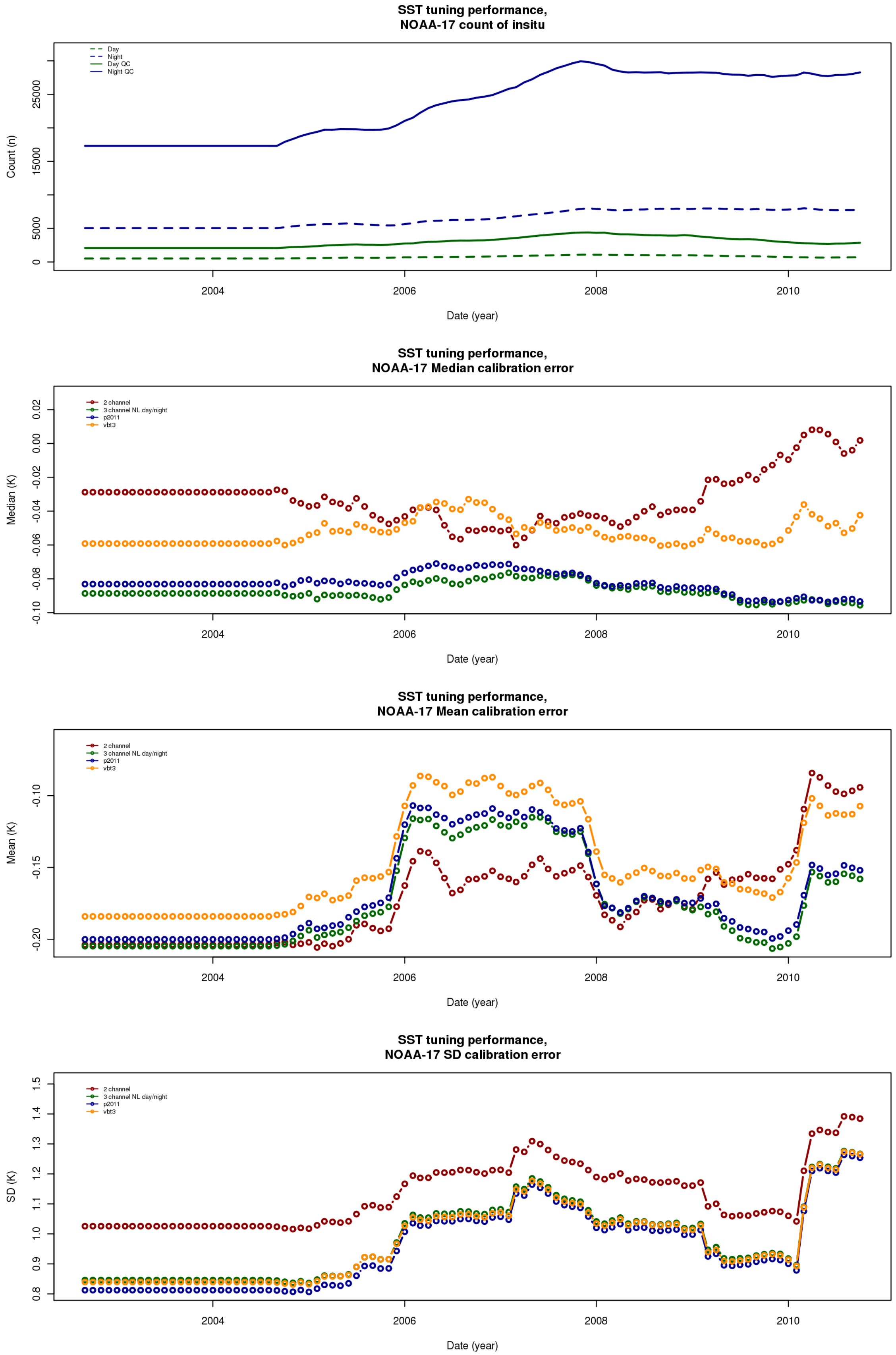


Figure 61: NOAA-17 Regression results over baseline dataset, comparative algorithm details

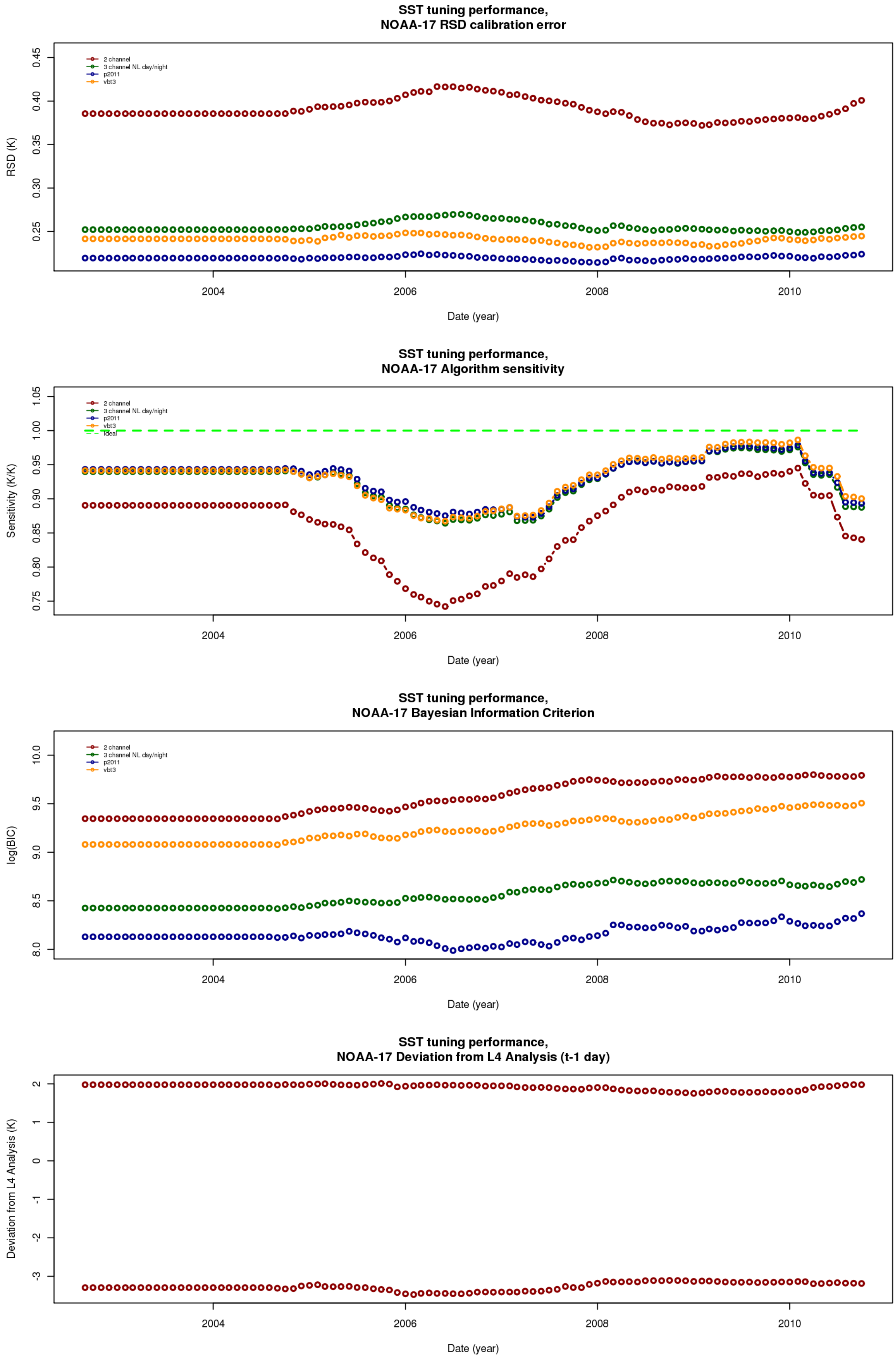


Figure 62: NOAA-17 Regression results over baseline dataset, comparative algorithm performance metrics

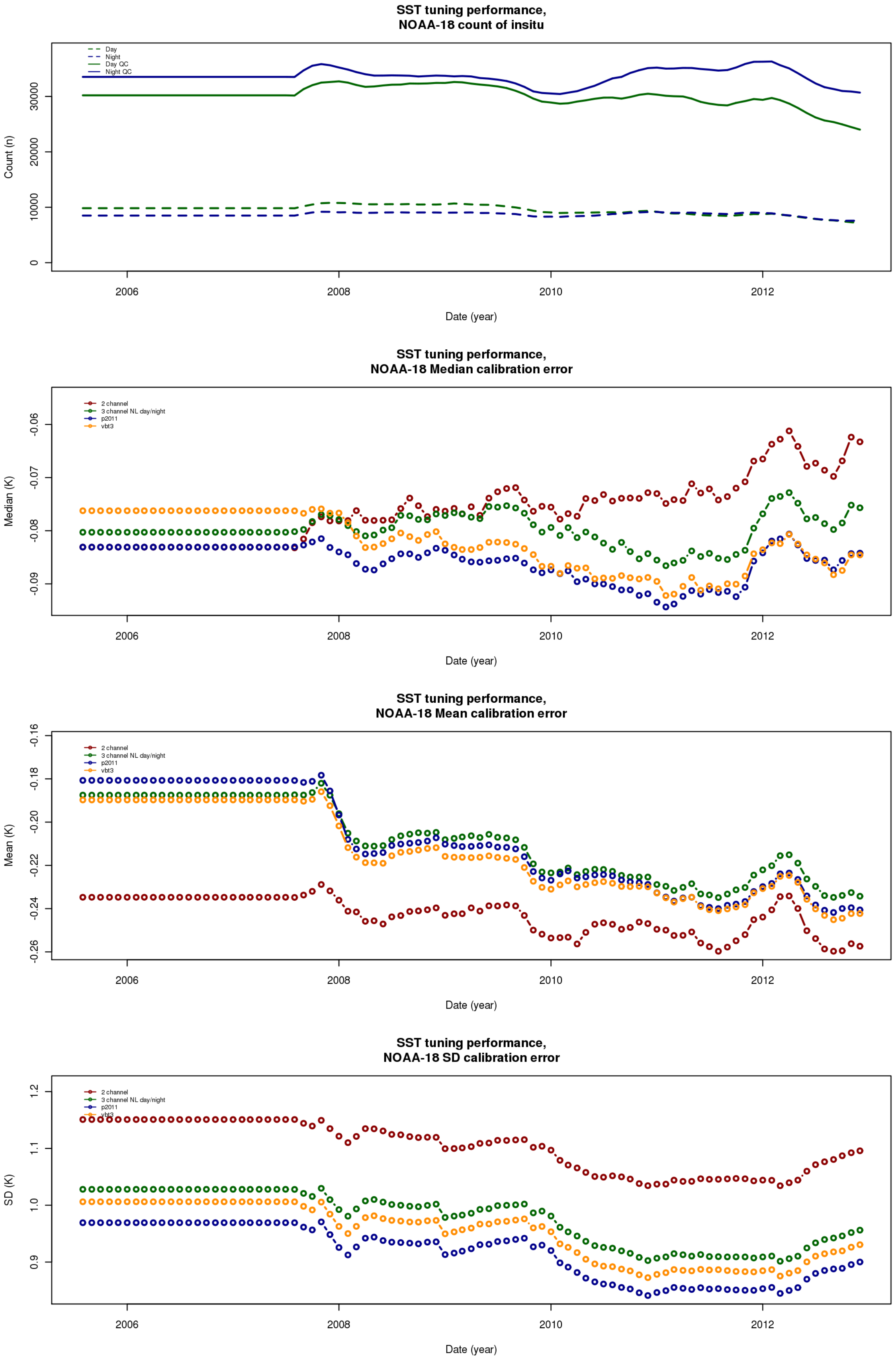
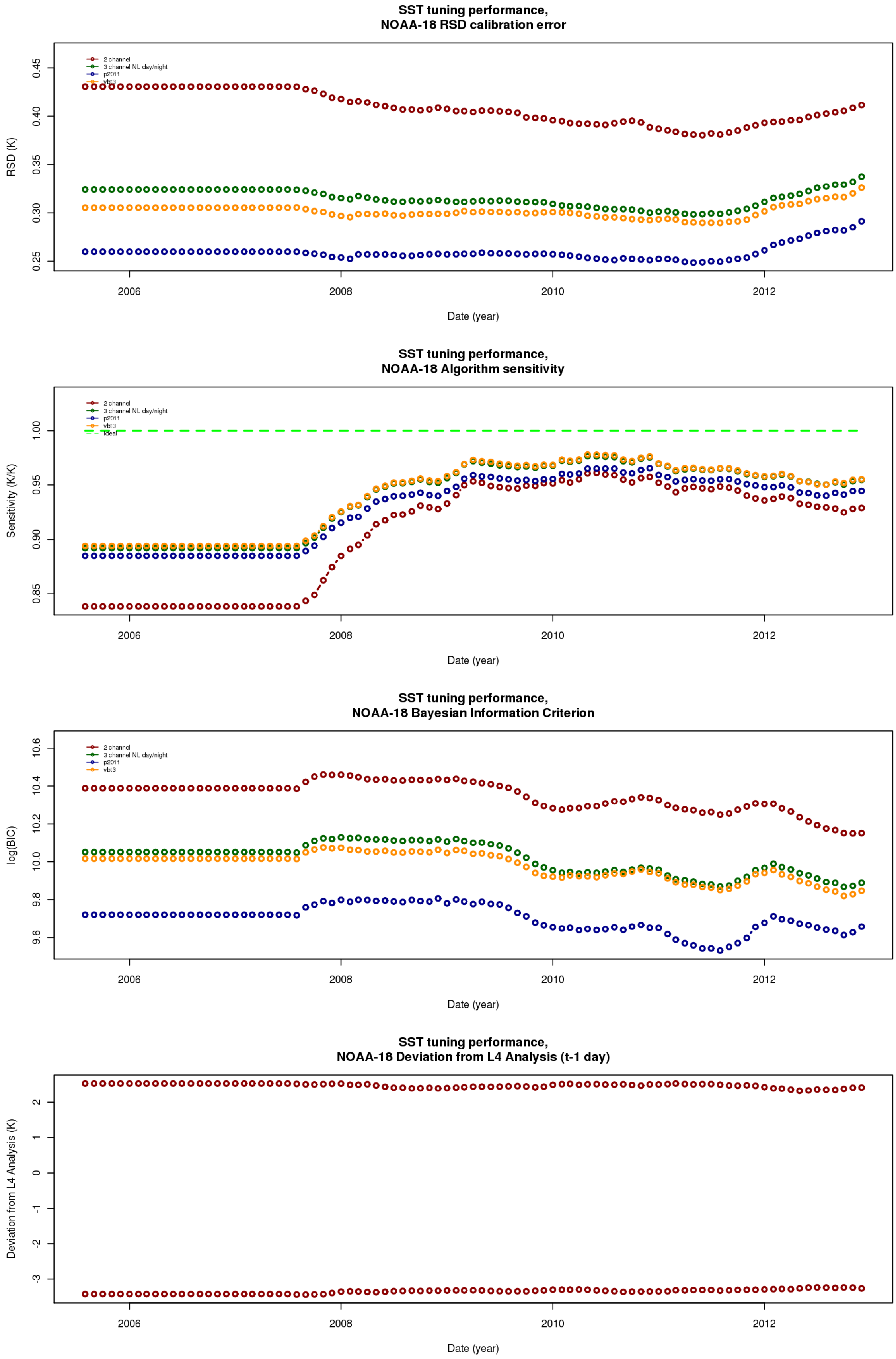


Figure 63: NOAA-18 Regression results over baseline dataset, comparative algorithm details


SST tuning performance,
NOAA-18 RSD calibration error

SST tuning performance,
NOAA-18 Algorithm sensitivity

SST tuning performance,
NOAA-18 Bayesian Information Criterion

SST tuning performance,
NOAA-18 Deviation from L4 Analysis (t-1 day)

Figure 64: NOAA-18 Regression results over baseline dataset, comparative algorithm performance metrics

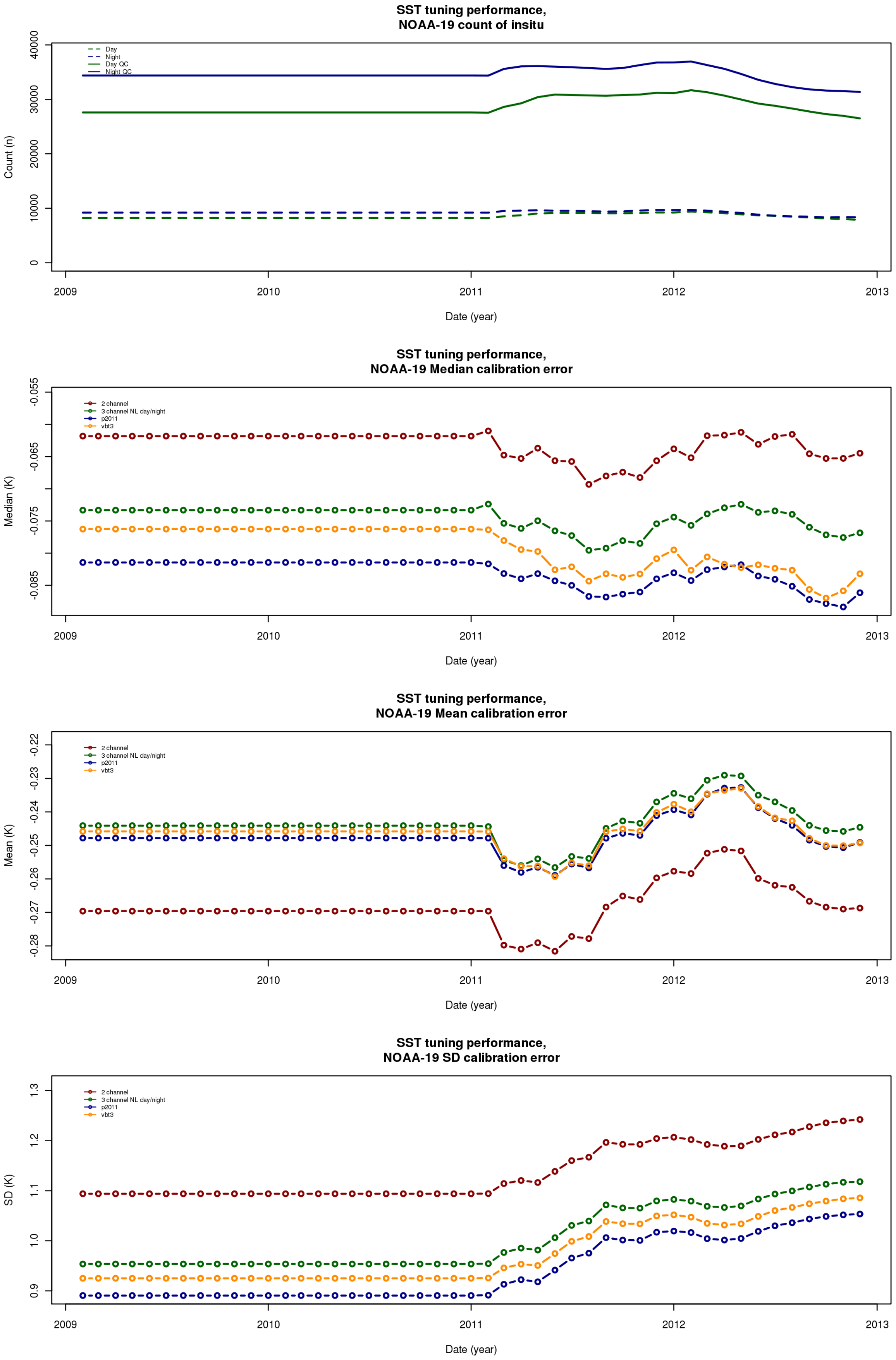


Figure 65: NOAA-19 Regression results over baseline dataset, comparative algorithm details

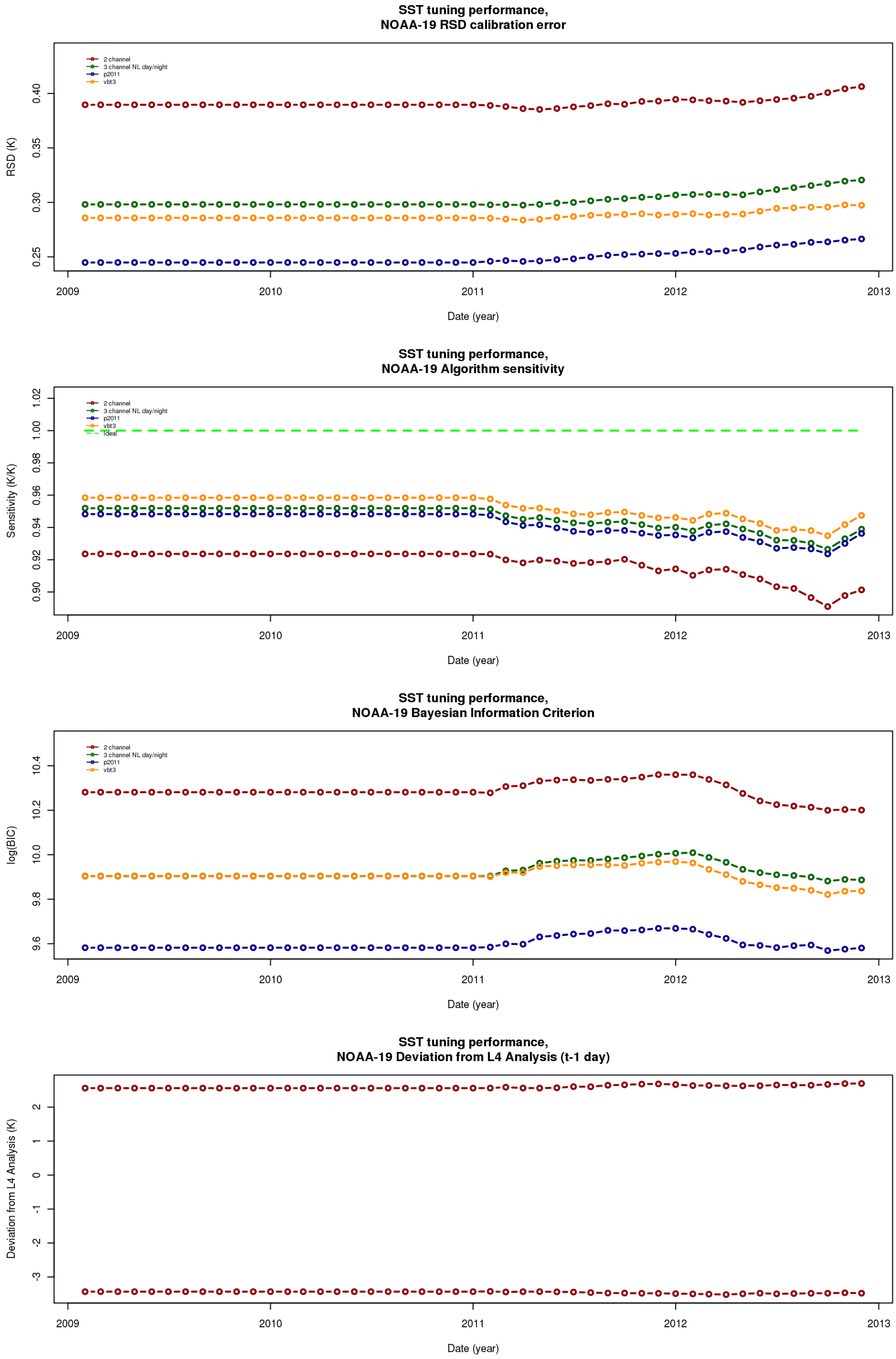


Figure 66: NOAA-19 Regression results over baseline dataset, comparative algorithm performance metrics

9.1 File naming convention

The file naming convention employed follows that suggested in the GHRSST standard[29], with the following specific structure specific to our product, →*TODO:***Put the description here**←

9.2 File field description

In addition, it is worth making some comments on some of the elements of the files that relate to site specific interpretations or information,

history The `history` metadata contains much information about the history of the file during processing as well as the tools that processed the file and other status information that relates to the processing. `history` is a comma delimited `parameter=value` list. The exact list of parameters present in the list is subject to change and is largely contextual, however some standard parameters are as documented in table 17. One of the most significant parameters is the `issue` parameter from which an overall assessment of file quality is determined.

quality_level The pixel quality based on a measure of the distance to the nearest cloud in kilometers, or 5 if the distance is greater than 5km. Pixel quality is assumed to be primarily driven by the degree of cloud contamination, so this forms the best first indication of the quality of a given pixel. Further indications of pixel quality can be garnered from `l2p_flags`, ancillary information, and the `sses_bias` and `sses_standard_deviation` which provide estimates of bias and error based on *in situ* measurements under similar view, quality and geographical conditions.

l2p_flags Additional `l2p_flags` have been added to reflect various ancillary conditions. This allows standard decisions based on ancillary fields to be made by reading a bit array rather than processing the ancillary fields. See table 16 for a more detailed description of the flags available.

sses_count A measure of the indicative number of degrees of freedom or measurements which went into validating or generating the SST measurement. For L2P files, this is an indication of the number of *in situ* measurements made under similar view, quality and geographical conditions, based on an empirical model. For L3U files, this is an indication of the number and proportion of L2P observed pixels that went into the composition of the gridded cell. For L3C and L3S files, `sses_count` is an indication of the weighted number of degrees of freedom that contributed to the SST. Empirical degrees of freedom models (L2P files) are updated frequently - currently every five days - and represent relatively high frequency changes in the condition of the SST retrieval.

sses_bias An estimate of the bias of the measurement with respect to *in situ* measurements made under similar view, quality and geographical conditions. This may be interpolated based on an empirical model. The `sses_bias` is empirically determined, and not based on regression with physically inspired SST retrieval models. However, addition of the bias does result in an overall more accurate match between retrieved SST and *in situ* measurement. `sses_bias` can be subtracted from the SST if this empirical compensation is required to be included in the assessment of the SST. `sses_bias` strictly is derived from `sea_surface_temperature`, and should not be blindly applied to `sea_surface_temperature_day_night`. Empirical bias

Parameter	Description
source_file	The primary source file (a stitched ASDA / HRIT file) that was processed to produce the data.
orbit	The satellite orbit number, as recorded in the HRIT file.
adi_source	The source for the aerosol dynamic indicator.
wind_source	The source for the NWP wind speed.
analysis_source	The source for the level 4 reanalysis.
ice_source	The source of the ice data.
satname	The name of the satellite, somewhat similar to the platform global parameter.
ad	The orbit type (a = ascending - from the south to the north, d = descending - from the north to the south).
isSouthern	A flag (=0 or =1) indicating that the file contains information with Southern Ocean coverage, and may (depending on the cloud cover) contain useful Southern ocean SST.
isPolar	A flag (=0 or =1) indicating that the file was produced using information from polar receiving stations.
isTropic	A flag (=0 or =1) indicating that the file contains information with Tropical Warm Pool coverage, and may (depending on the cloud cover) contain useful Tropical Warm Pool SST.
quality	realtime, archive, or fault, designates if the data internal to the file is considered finalized or not, and in this way indicates the processing quality of the file. A value of archive indicates that the file was finished and satisfactory information was available for all fields. A value of realtime indicates that the file was not finished, and that the additional information required to finish the file is possible to be provided at a later time. A value of fault indicates that the file was not finished due to some kind of processing error.
file_quality_level	the same as the file_quality_level global parameter, indicates the overall quality of the file.
issue	A semicolon delimited parameter value list of issues that result in the file having a non-archive processing quality. The list assigns values to the reasons for the quality degradation. See sectio 9.3 for more information about the names and hanndling of the issue parameter.

Table 17: A short list of history parameters that may be of use.

models, if used, are updated frequently - currently every five days - and represent relatively high frequency changes in the condition of the SST retrieval.

sses_standard_deviation An indication of the uncertainty associated with the SST estimate, compared to *in situ* measurements made under similar view, quality and geographical conditions, possibly interpolated by an empirical model. The `sses_standard_deviation` is empirically determined, and not based on regression with physically inspired SST retrieval models. However, `sses_standard_deviation` can be compared to the `sses_bias` to give some idea about the confidence in using this as a correction to the SST. `sses_standard_deviation` strictly is derived from `sea_surface_temperature`, and thus will not directly apply to `sea_surface_temperature_` however, it may be used to determine the scale on which the difference between the two sea surface temperatures may show significant difference. Empirical standard deviation models are updated frequently - currently every five days - and represent relatively high frequency changes in the condition of the SST retrieval.

sst_count The unbiased number of SST observations that was merged into the pixel at this location. This field is not present in L2P files, and will be absent if all valid values are 1. This field is typically used in multiday *L3S* over long periods of time, to correct SSES estimates.

sst_standard_deviation The unbiased standard deviation of the SST observations that were merged into the pixel. This field is not present in L2P files, and will be absent if all valid values are ill-defined because `sst_count` is less than 2. This field is typically used in multiday *L3S* over long periods of time, to correct SSES estimates.

sst_mean The unbiased mean of the SST observations that were merged into the pixel. This field is not present in L2P files, and will be absent if `sst_count` is less than 2. This field is typically used in multiday *L3S* over long periods of time, to correct SSES estimates.

sea_surface_temperature The SST based on a non-linear retrieval scheme with seperate day and night algorithms, similar to that presented in [4]. The retrieval scheme is calculated monthly on a 2 year rolling window regression of observations against *in situ* measurements.

sea_surface_temperature_day_night An optional SST based on a retrieval scheme which uses a single algorithm for both day and night processing, resulting in a consistent retrieval for both day and night. The SST thus retrieved is therefore expected to be more suited to studies of the diurnal cycle. It is expected that when there is a large deviation between `sea_surface_temperature` and `sea_surface_temperature_day_night` during the day, that this corresponds to a transitory atmospheric condition. Large differences thus detected, of 3 times `sses_standard_deviation` or greater, result in a flag to be set in the `l2p_flags` field. The retrieval scheme is calculated monthly on a 2 year rolling window regression of observations against *in situ* measurements, commensurate with the longer term stability expected of the AVHRR sensor.

dt_analysis The deviation between the SST `sea_surface_temperature`, and a level 4 re-analysis of the previous day's foundation temperature. For skin SST, the mean value in a wind environment that encourages top level mixing, will be 0.17K. Prior to July 23rd, 2008, level 4 re-analysis makes use of the NCDC AVHRR SST analysis (”Reynolds”)[23]. After July 23rd, 2008, GAMSSA foundation SST is used as the level 4 re-analysis.[30]

wind_speed The 10 meter wind estimated from re-analysis numerical weather prediction models. Prior to 1st September 2009, ECMWF historical re-analysis is used.[21] After 1st September 2009, the Australian Bureau of Meteorology ACCESS-G winds are used.[14]

sea_ice_fraction A number between zero and one, representing the sea-ice fraction, is taken from the NCEP re-analysis of the previous day.[9]

aerosol_dynamic_indicator Aerosol Dynamic Indicator (ADI) is added based on the NOAA AERO100 data set, after 27th November, 1998[3], re-analysis based on the previous day. OSDPD daily files are used after 11th January 2011. Early SST retrievals prior to 27th November, 1998, may not contain ADI fields.

sst_difference_1 An optional field which represents the difference between the SST reported in the file and another reference SST. The field `comment` attribute provides further details about the source of the reference field.

sst_difference_2 An optional field which represents the difference between the SST reported in the file and another reference SST. The field `comment` attribute provides further details about the source of the reference field.

sst_difference_3 An optional field which represents the difference between the SST reported in the file and another reference SST. The field `comment` attribute provides further details about the source of the reference field.

9.3 Issue description

`issue` is an optional parameter value pair which appears in the `history` global parameter of a GHRSST compliant netCDF file. `issue` contains a semicolon delimited list of parameter value pairs which describe an issue (as the parameter), and its severity (as the value). Severity values follow the following conventions,

- 0 — the quality degradation is trivial or non-existent, and serves as an observation, and did not affect quality significantly. Files produced with issues of severity 0 do not suffer any degradation in quality due to these issues.

- 1 — the quality degradation was caused by something that could be considered a processing fault, and is thus considered a serious quality compromise. Files produced with any issues of severity 1 are designated as faulty using the `quality=fault` descriptor in the `history` global parameter.

- 2 — the quality degradation is significant but can most likely be recovered at a later time or when a full complement of processing information, ancillaries and data files are available. It is neither serious or trial at the present time and most likely will become trivial. Files produced with any issues of severity 2 that have no severity 1 issues are designated as faulty using the `quality=realtime` descriptor in the `history` global parameter.

- 3 — there may or may not be a significant quality degradation (its possible, but not possible to give a definitive answer right now) typically because there is sufficient information for good quality, however the information is lower in quantity than would be expected. The

Severity 0	Severity 1	Severity 2	Severity 3	Assigned Quality
Yes	No	No	No	`quality=archive`
Yes or No	Yes	Yes or No	Yes or No	`quality=fault`
Yes or No	No	Yes	Yes or No	`quality=realtime`
Yes or No	No	Yes or No	Yes	`quality=realtime`

Table 18: Mapping between multiple issues and severities and overall quality

legitimacy of this issue is typically determined in time or the context in which the data are used downstream. Files produced with any issues of severity 3 that have no severity 1 issues are designated as faulty using the `quality=realtime` descriptor in the `history` global parameter.

When multiple issues and severity values are present, the rules for assigning overall quality are applied according to table 18.

The `issue` parameters are inherited by downstream products of the processing chain, when the context of the data use is appropriate for this to occur, and also other supporting data sets such as the *in situ* match-up data base (MDB), as appropriate, which can in turn influence algorithm calibration, and thus the SST retrieved at a later time. Thus resolving issues is important if their potential consequences are required to be minimized.

A Use of SSES for quality assessments

There is a great many GHRSST compatible products provided by many different authorities. The GHRSST format ensures that each of them must provide an assessment of quality, and Sensor Specific Error Statistics (SSES). Sensor specific bias and sensor specific standard deviation are determined on a pixel by pixel basis. The quality assessment is as an indication of degree of cloudiness, or water content, but the precise definition is left to the data provider, and is generally understood to be relative to the scene in view. Some data providers do not apply graduated quality assessment at all.

Generally speaking sensor specific bias is an attempt to express the SST retrieval against an *in situ* standard, and sensor specific standard deviation is an attempt to estimate the possible error in the retrieval, which is largely based on the method employed to compute the retrieval and if *in situ* measurements are also involved in such a system. Whereas the quality assessment contains other ancillary information, which expresses the relative probability that the retrieval was performed to adequate accuracy.

Although the use of the datasets is somewhat standardized by the GHRSST interface, comparison between these data sets is hampered by the ambiguity in quality assessment, since there is a range of ways in which it is determined, and this information does not generally relate to other uncertainty estimates or to data from different scenes. Furthermore the quality assessment as implemented is effectively a non-parametric assessment of quality. On the other hand, consideration using bias and standard deviation assessments may more useful quantitatively, provided the estimates can be assumed to include factors that correlate with the determination of the quality assessment. However it is desirable to have a quality assessment that includes information from both sources, allowing the non-parametric nature to be fully exploited in the comparison process.

A simple proposal is provided as a way of harmonizing the quality assessment and the other uncertainty estimates without being concerned about the underlying details of how the various qualities and SSES are estimated. We show how this works against ABOM NOAA-POES AVHRR data sets, as well as those from ACSPO VIIRS data, and consider the process of using this information to aggregate data from these various sources in a way that allows the quality assessment in its non-parametric sense to be used as a robust filter that preserves the best estimates in the aggregation process.

The merge process has two basic types, the first is a pixel by pixel merge which represents a change of coordinate frame. This process allows multiple pixels in the source dataset to be merged into a single target, as well as multiple target pixels to be associated with a single source pixel. This represents the primary transformation between GHRSST L2P, native coordinate data, to GHRSST L3U, rectangularly gridded data. The second type of merging process considers data multiple views of the same grid that needs to be aggregated. This represents the primary transformation between GHRSST L3U and L3C datasets, or between L3C and L3S datasets. The later aggregation can be considered in some sense as a special case of the earlier, which we discuss in what follows. For more information on both methods, see [**GHRSSTdoc**].

A.1 L3U class product and the computation of L3U from L2P product

Rectangularly gridded L3U products are produced by remapping each single swath of SST product in native coordinates onto a fixed grid. The purpose of this is to provide a consistent coordinate system for products that allows for easier comparison from swath to swath, and the opportunity

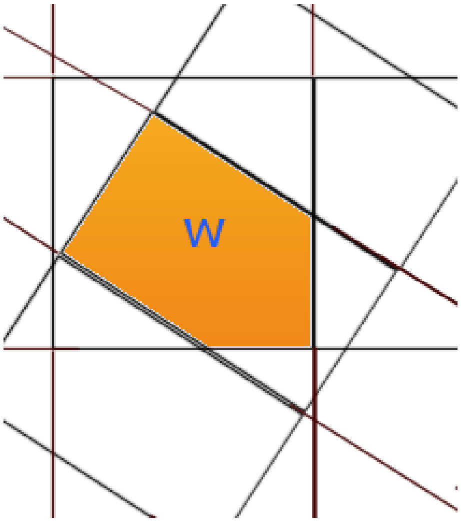


Figure 67: Identification of grid overlap weight.

to provide the product to a grid resolution that is commensurate with the requirements of other downstream applications.

The process of gridding SST consists of projecting an ungridded swath, pixel by pixel, onto a regular fixed grid, weighting each pixel contribution by the area of overlap between the source and target pixels, w_i. In the computation of w_i, we assume that both the source and target pixel arrays consist of many parallelograms with side lengths given by the centre to centre distance of the pixels. The source image is assumed to be rotated, and the size and centres of the pixels vary over the field of view in both the source and target pixels. For the particular supported grids, cylindrical coordinates ensure that the target locations are regularly spaced over the chosen rectilinear grid extent.

The weight corresponds to the area of overlap, as demonstrated in figure 67.

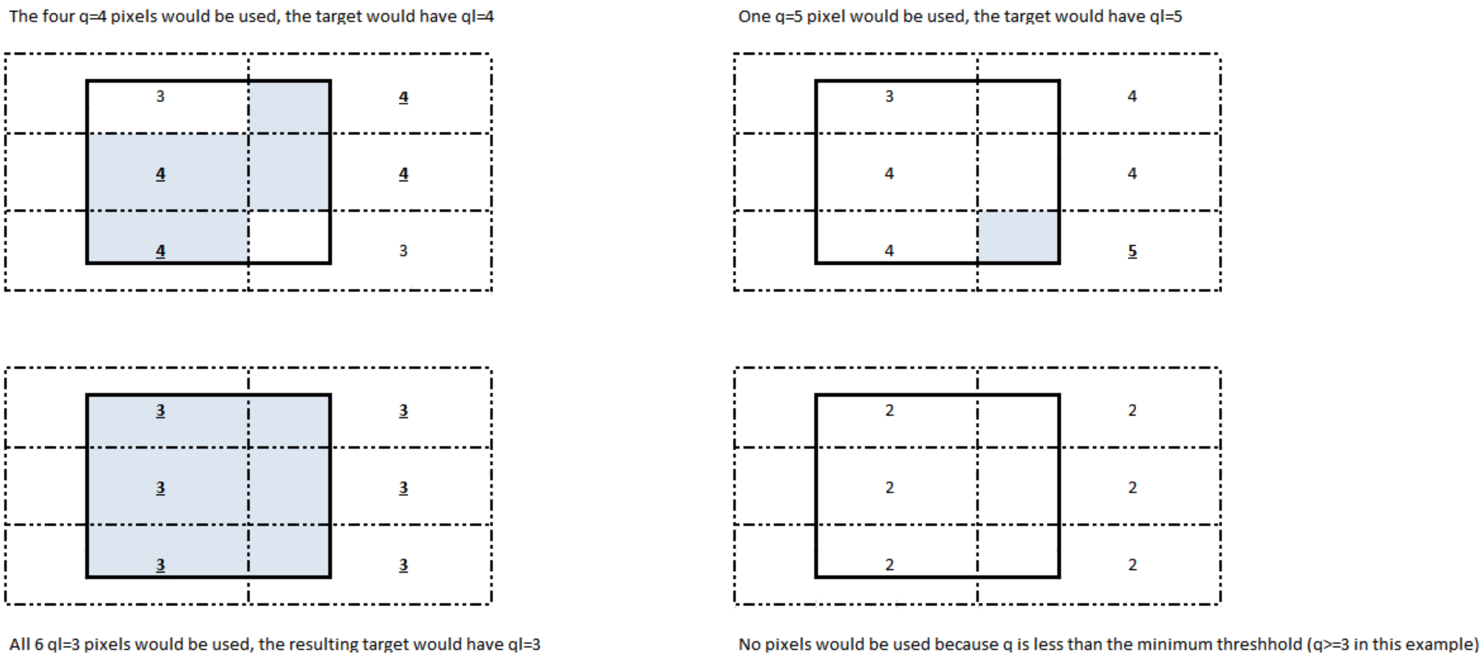
For a particular target pixel, the overlapping source pixels are sorted based on quality level, and those with the highest quality are merged by applying weighted sums. Figure 68 shows an example of how pixels are chosen. The SST and other similar parameters are mapped using the standard weighted average method, with weights $w_{i,j}$ representing the overlap area of source pixel i into target pixel j,

$$T_{\text{satellite},U,j} = \frac{\sum_{i \in j} w_{i,j} \, T_{\text{satellite},i}}{\sum_{i \in j} w_{i,j}}, \tag{130}$$

wherein $\sum_{i \in j}$ represents the sum over all suitable pixels that contribute to a given target pixel j, determined based on the best quality pixels available at the given target.

$T_{\text{satellite},U,j}$ defined in this way corresponds to an area weighted overage of best quality SST measurements at the pixel of interest. The same averaging technique is applied to other ancillary fields (such as wind speed, aerosol dynamic indicator, analysis SST, observation time).

Pixel by pixel SSES, the gridded degrees of freedom $n_{U,j}$, bias $\mu_{U,j}$ and standard error $\sigma_{U,j}$, are determined using slightly different algorithms. The bias, which is expected to be an estimated offset to SST, considers area based weighting in exactly the same manner as the ancillary fields,



The four q=4 pixels would be used, the target would have ql=4

One q=5 pixel would be used, the target would have ql=5

All 6 ql=3 pixels would be used, the resulting target would have ql=3

No pixels would be used because q is less than the minimum threshhold (q>=3 in this example)

Figure 68: Identification of contributing pixels based on source candidate quality ensures only pixels of highest quality contribute to the target L3 pixel.

$$\mu_{U,j} = \frac{\sum_{i\in j} w_{i,j}\,\mu_{P,i}}{\sum_{i\in j} w_{i,j}} \tag{131}$$

The number of gridded degrees of freedom reflects the number of pixels that went into the average, by considering the sum of each pixels weighted contribution, normalized by the largest contribution,

$$n_{U,j} = \frac{\sum_{i\in j} w_{i,j}}{\max_{i\in j} w_{i\in j}}, \tag{132}$$

where as before, the understanding of $\{i \in j\}$ is all of the best quality source pixels i that overlap with the target pixel j. The resulting count value $n_{U,j}$ is a non integer value representative of the number of pixels that went into the computation. The use of the maximum weight as normalisation allows the pixel with the largest contribution to count as one, and those that contribute relatively less to be counted as such according to the weight relative to the largest contribution. A per pixel normalization of this kind does not affect the interpretation of the weight in the computation of other weighted averages, but allows the interpretation as number of significant measurements to be applied over the field of view irrespective of the relative overlap area in different regions of the remapping. The standard deviation estimate, which is derived from a population of $n_{P,i}$ *in situ* measurements, is also weighted by pixel overlap, $w_{i,j}$

$$\sigma_{U,j} = \sqrt{\frac{\sum_{i\in j} w_{i,j}\,(\sigma_i^2 + \mu_{P,i}^2)}{\sum_{i\in j} w_{i,j}} - \left(\frac{\sum_{i\in j} w_{i,j}\,\mu_{P,i}}{\sum_{i\in j} w_{i,j}}\right)^2} \tag{133}$$

None of the SSES depend on $n_{P,i}$, and their formation is thus not affected by missing degrees of freedom data in L2P files. This is appropriate, since the SST measurement itself has no relation to the number of degrees of freedom used to generate the bias and standard error estimates, and we are forced to use the same method of averaging for all three of these parameters to maintain consistency in the application of the bias as a correction.

In addition to ancillary fields such as wind and aerosol, The GHRSST specification also includes time based fields, which are also remapped from L2P to L3U as the weighted average time since epoch, under the assumption the the linear variation of measured value is best described by a linear variation in time.

The L2P f_{L2p} parameter, which describes possible exceptions or causes that may influence pixel quality and interpretation (generally negatively) is combined using the local OR of all of the source pixels, respecting the desire to record any possible influence that may contribute to the interpretation of the target pixel behaviour.

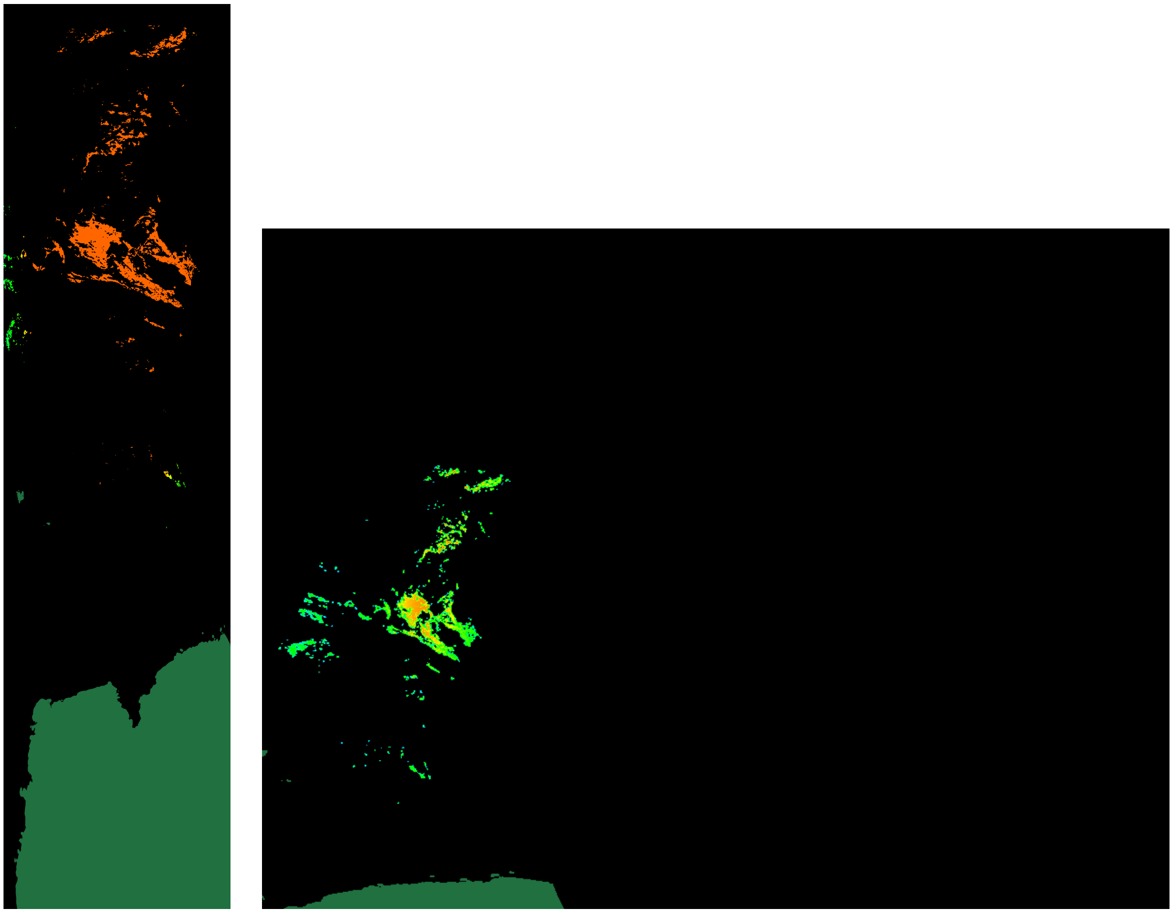
In this manner, all of the points on the L2P swath are mapped to an L3U set,

$$\left\{T_{\mathrm{satellite},U,j}, t_{U,j}, q_{U,j}, \mu_{U,j}, \sigma_{U,j}, n_{U,j}, \mathrm{ancillary}_j, f_{\mathrm{L2p},U,j}\right\} \tag{134}$$

and this information is stored in the SSES fields for L3U files with the same indicative names as those used for L2P files, as outlined in table 19.

Parameter name	Symbol	L2P		L3U	
`sses_count`	n	n_P	Indicative of the number of *in situ* measurements made under similar viewing conditions	n_U	Indicative number of best quality L2P pixels merged when converted to a fixed grid.
`sses_bias`	μ	μ_P	Indicative median bias for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing conditions.	μ_U	Indicative gridded median bias for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing and merging conditions.
`sses_standard_deviation`	σ	σ_P	Indicative standard deviation for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing conditions.	σ_U	Indicative gridded standard deviation for $T_{\text{satellite}}$ compared to *in situ* measurements made under similar viewing and merging conditions.
`l2p_flags`	f_{L2p}	$f_{\text{L2p},P}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions.	$f_{\text{L2p},U}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location.
`quality_level`	q	q_P	Quality level as a measure of proximity to detected cloud in kilometres. $q = 0$ is also used to indicate invalid data for other reasons.	q_U	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.
`sea_surface_temperature`	$T_{\text{satellite}}$	$T_{\text{satellite},P}$	Retrieved sea surface temperature	$T_{\text{satellite},U}$	An average of the retrieved sea surface temperature based on all of the measurements contributing to the gridded location.

Table 19: Association between field names in GHRSST compliant files and symbols used in this text, with a short description of the intent of the parameter and symbol, for L2P and L3U files.



L2P n_P

L3U n_U

Figure 69: Comparison between `sses_count` field for L2P compared to L3U, for NOAA-16 at Jan 1st 2011, 00:36UTC. The L2P count reflects the fact that there are more *in situ* measurements on a long term average at the center of the swath (red) compared to the edges (green), and the number of measurements diminishes as we go further south (green). The L3U count reflects the fact that there is a higher overlap of L2P pixels in the middle of cloud clear regions near the center of swath (orange), compared to the edges of both the cloud and the swath (green).

It should be noted that the most significant difference between L2P and L3U SSES just described is the use of the `sses_count` field, which corresponds to an indicative number of *in situ* measurements that contribute to SSES estimates, n_P, in the L2P file (defaulting to 1 if not present), and an indicative number of incumbent best quality L2P pixels, n_U, in the L3U files, with highest n_U in places where larger numbers of low σ_P pixels contribute to the average. This is illustrated in figure 69.

A.2 L3C class product and the computation of L3C from L3U product

L3C products consist of merges of multiple swaths from the same instrument and platform over a period of time that is small on the scale of significant changes in the underlying SST. Since a single swath of a polar orbiting satellite provides only a small snapshot of the ocean temperature, merging multiple swaths allows greater regional coverage to be delivered at the expense of reduced temporal resolution. Thus we consider these as a common grid merge of multiple L3U data sources. Each L3C product has a measurement window - a time period and time domain - of interest, in our case we consider day-time one and three day products as well as night-time one and three day products, four different measurement windows.

The process of merging gridded SST estimates consists of taking the highest quality gridded estimates from multiple sources on the same grid, over the measurement window, and providing a merged value of these measurements, by averaging. The resulting average thus represents a characteristic measurement for the platform, over the measurement window in question. This process is complicated by the reality that there is expected to be some time dependent variation on the measured $\mathrm{SST_U}$, which will be averaged out in the averaging process, but should be considered when we think about estimates of the standard error.

This time dependent variation does not correspond to a typical measurement error, rather qualifies the variation in the estimate of SST due to natural variation over the measurement window. As the measurement window is enlarged, and the interpretation of the SST is maintained as characteristic measurement commensurate with the time taken in by the measurement window, the uncertainty due to this variation is expected to scale out as $\frac{1}{\sqrt{N}}$, where N is the number of measurements, swaths, or the amount of time involved, in keeping with standard error estimates, whereas the *in situ* errors will not, since they represent uncertainties associated with the reported SST value in the context of the measurement apparatus.

Furthermore, since it is likely at the edge of swath that there may be overlap in the measurements, and that the sensor specific error statistics reflect uncertainties based on satellite zenith angle, or that there are different error estimates associated with different times of the day, it is desirable to weight measurements by their significance measured by the number of degrees of freedom, n_U and the estimate of the measurement error, σ_U, under the assumption that measurements with a larger n_U and a smaller σ_U are more certain to be representative of the pixel in question over the period over which the merge is considered, and each measurement made can be considered somewhat independent.

Following this rationalisation, we choose the representative value for $T_{\text{satellite},C}$, the gridded instrument specific merged SST over a fixed time period and other merged parameters to be computed with a simple quantity over variance $(\frac{n}{\sigma^2})$ weighting as if the combined measurements are from uncorrelated sources,

$$T_{\text{satellite},C,j} = \frac{\sum_{i\in j}\frac{n_{U,i}}{\sigma^2_{U,i}}T_{\text{satellite},U,i}}{\sum_{i\in j}\frac{n_{U,i}}{\sigma^2_{U,i}}} \tag{135}$$

where the sum in the above expression over $i \in j$, is assumed to be over all of the best quality source L3U pixels from all of the swath files over the time window, i at the common target location, j.

SSES $\{n_C, \mu_C, \sigma_C\}$, are determined by a similarly weighted average for the number of degrees of freedom and the bias,

$$n_{C,j} = \frac{\sum_{i\in j}\frac{n_{U,i}}{\sigma^2_{U,i}}}{\sum_{i\in j}\frac{1}{\sigma^2_{U,i}}} \tag{136}$$

$$\mu_{C,j} = \frac{\sum_{i\in j}\frac{n_{U,i}}{\sigma^2_{U,i}}\mu_{U,i}}{\sum_{i\in j}\frac{n_{U,i}}{\sigma^2_{U,i}}} \tag{137}$$

The sensor specific standard deviation is similarly computed,

$$\sigma^2_{Cs,j} \quad = \quad \frac{\sum_{i\in j} \frac{n_{U,i}}{\sigma^2_{U,i}}(\sigma^2_{U,i} + \mu^2_{U,i})}{\sum_{i\in j} \frac{n_{U,i}}{\sigma^2_{U,i}}} - \mu^2_{C,j} \tag{138}$$

but is corrected for the time window variation of the SST which adds an additional uncertainty to the time window characteristic SST computed, resulting in the SSES estimate, σ_C,

$$\sigma_{C,j} \quad = \quad \sqrt{\sigma^2_{Cs,j} + \frac{\sigma^2_{w,C,j}}{n_{C,j}}} \tag{139}$$

which scales the environmental component, $\sigma_{w,C}$ out as $\sim \frac{1}{\sqrt{n_C}}$, as any estimate of the standard error of the mean, by the central limit theorem.

$\sigma_{w,C}$ is the standard deviation of the environmental component of the SST over the time window, which is estimated by making use of the time window variability parameters, $\{n_{w,C}, T_{w,C}, \sigma_{w,C}\}$, determined by equally weighting all of the SST measurements over the period of interest, irrespective of n_U and σ_U.

$$n_{w,C,j} \quad = \quad \text{count}\,\{i \in j\} \tag{140}$$

$$T_{w,C,j} \quad = \quad \frac{\sum_{i\in j} T_{\text{satellite},U,i}}{n_{w,j}} \tag{141}$$

$$\sigma^2_{w,C,j} \quad = \quad \frac{\sum_{i\in j} T^2_{\text{satellite},U,i}}{n_{w,j}} - T^2_{w,C,j} \tag{142}$$

where, as before, the notation $\{i \in j\}$ refers to the set of all of the valid pixels of the best quality from the multiple L3U sources covering the time window at the position j, and the count function counts them. This serves as an indication of the amount of variation possible under the assumption that all of the measurements made have no error and are of good quality, thus any variation seen is an estimate of the environmental variation over the time window rather than instrument variation. In the event that the instrument and environment are uncorrelated, this will be an overestimate or conservative estimate of the possible environmental variation. The variation over the time window is added in quadrature (with the assumption of normality) to the instrument contribution to σ_C, because it is expected the environment will be uncorrelated with the instrument variation as a first approximation.

In this manner, all of the points of the L3U source data are mapped to an L3C data set,

$$\left\{T_{\text{satellite},C,j}, t_{C,j}, q_{C,j}, \mu_{C,j}, \sigma_{C,j}, n_{C,j}, T_{w,j}, \sigma_{w,C,j}, n_{w,C,j}, \texttt{ancillary}_j, f_{\text{L2p},C,j}\right\} \tag{143}$$

and this information is stored in the SSES fields for L3C files with the same indicative names as those used for L3U files, as outlined in table 20. Ancillary fields are treated the same way as SSTfields, and f_{L2p} are bitwise or-ed, as before.

Note that there are four additional fields to those recommended in the GDS version 2.0r5. In addition to `sses_count`, n_C, we have added the three time window variation fields corresponding to $\{T_{w,C,j}, \sigma_{w,C,j}, n_{w,C,j}\}$, `sst_mean`, `sst_standard_deviation` and `sst_count`, representing the equally weighted SST, standard deviation, and count. Having these stored in L3C files allows combinations of L3C files to be considered and compared, and the observed environmental parameters combined. This aids in the merging of L3C to L3S, where differences in the bias due to different platforms are considered. See section A.3 for further details.

A.3 L3S class product and the computation of L3S from L3C product

L3S class product provides a typical characteristic SST over a (possibly) extended time window, and multiple instruments, by combining single day, single platform L3C files. In order to consider the SST provided indicative of the time period, we assume that all best quality measurements from all platforms and days contribute equally. To remove the impact of unstable or end of life platforms, we only include the missions that we consider production quality on the day in question. See figure 70 for details about which missions are included over the full period covered by the archive.

The equal weighting simplifies the composition process of L3S files, and allows multiple L3S files to be composed and generated progressively if the coverage period is very long,

$$T_{\mathrm{satellite},S,j} = \frac{\sum_{i\in j} n_{C,i}\, T_{\mathrm{satellite},C,i}}{\sum_{i\in j} n_{C,i}} \tag{144}$$

As before, the sum is over all of the best quality pixels at the same target location j over the time window and range of platforms.

The number of degrees of freedom, combined bias and standard deviation with respect to *in situ* are estimated based on equal weighting after first removing the time window variation from the L3C SSES,

$$\begin{aligned}
n_{S,j} &= \sum_{i\in j} n_{C,i} && (145)\\
\mu_{S,j} &= \frac{\sum_{i\in j} n_{C,i}\mu_{C,i}}{n_{S,j}} && (146)\\
\sigma^2_{Cs,i} &= \sigma^2_{C,i} - \frac{\sigma^2_{w,C,i}}{n_{C,i}} && (147)\\
\sigma^2_{Sb,j} &= \frac{\sum_{i\in j} n_{C,i}(\sigma^2_{Cs,i} + \mu^2_{Cb,i})}{n_{S,j}} - \mu^2_{S,j} && (148)
\end{aligned}$$

In estimating SSES, some care is required in treating time window variation and *in situ* based variation separately. We use the same six factor representation of SSES introduced in section A.2, being careful to apply platform biases to the measured SST in the composition of the mean time window SST, and the variance of the time window SST,

$$\begin{aligned}
n_{w,S,j} &= \sum_{i\in j} n_{w,C,i} && (149)\\
T_{w,S,j} &= \frac{\sum_{i\in j} n_{w,C,i}\,(T_{w,C,i} - \mu_{C,i})}{n_{w,S,j}} + \mu_{S,j} && (150)\\
\sigma^2_{w,S,j} &= \frac{\sum_{i\in j} n_{w,C,i}\left(\sigma^2_{w,C,i} + \mu^2_{w,C,i}\right)}{n_{w,S,j}} - \left(T_{w,S,j} - \mu_{S,j}\right)^2\\
&\quad + \frac{\sum_{i\in j} n_{w,C,i}\mu_{C,i}\,(\mu_{C,i} - 2T_{w,C,i})}{n_{w,S,j}} && (151)
\end{aligned}$$

Additional terms in the above remove contributions due to the platform biases in $T_{w,C}$, and the correlation between $T_{w,C}$ and μ_C, the measured temperature and the bias in the measurement equipment.

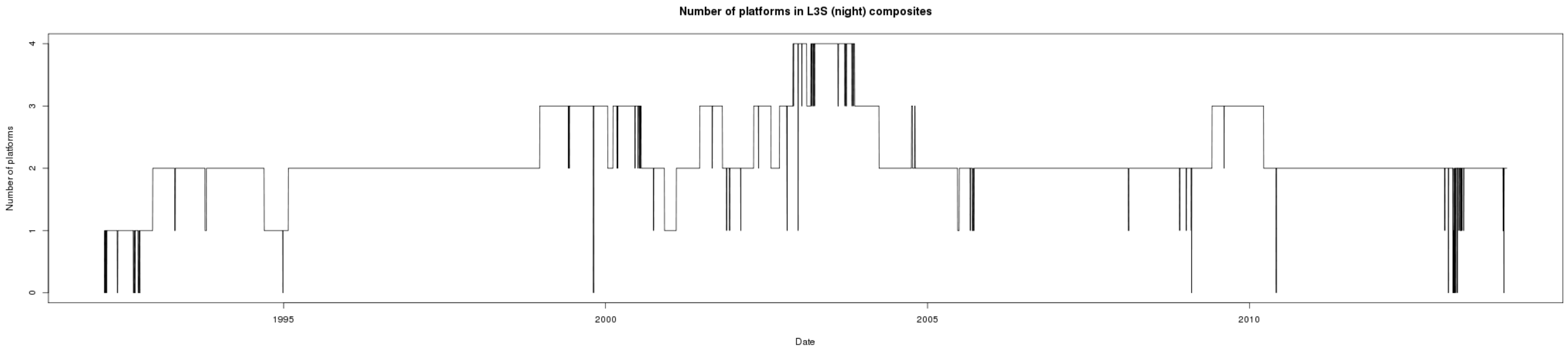


Figure 70: Number of NOAA-AVHRR missions available for inclusion in ABOM L3S night composites since 1995. Satellites are excluded or included based on NOAA mission status, reception quality, and the success of accurate navigation correction. For the greater part of the period, more than 2 platforms provide coverage for L3S files over continental Australia.

As before, the standard deviation is a composite of the sensor related component added in quadrature to the environmental component, where the environmental component is treated as a standard error of mean, and scaled out with the number of degrees of freedom,

$$\sigma_{S,j} = \sqrt{\sigma_{Sb,j}^2 + \frac{\sigma_{w,S,j}^2}{n_{S,j}}} \tag{152}$$

Thus we form the L3S data set,

$$\left\{T_{\text{satellite},S,j}, t_{S,j}, q_{S,j}, \mu_{S,j}, \sigma_{S,j}, n_{S,j}, T_{w,S,j}, \sigma_{w,S,j}, n_{w,S,j}, \texttt{ancillary}_j, f_{\text{L2p},S,j}\right\} \tag{153}$$

and this information is stored in the resulting L3S fields with the same indicative names as those used for L3C files, as outlined in table 20. Ancillary fields are treated the same way as SST fields, and f_{L2p} are bitwise or-ed, as before.

This treatment allows L3S and L3C files to be combined hierarchically, producing L3S files at an intermediate step that can be further combined. Longer time period product with many individual data sources can thus be produced recursively with the resulting SSES independent of the exact order in which the files were combined. For example, annual L3S SST could be generated by combining four quarterly L3S SST products which are in turn derived from three monthly L3S SST product, each of which are composed of daily L3S product, which are in turn composed of the L3C product from various source instruments on their respective days.

The resulting L3S product contains estimates of the time window SST variation $\sigma_{w,S}$, the *in situ* error, σ_S, the number of measurements $n_{w,S}$ and the number of high quality measurements n_S, with biases corresponding to mean bias over all platforms, μ_S.

A.4 A new definition of quality level from SSES

In order to approach a more unified approach to the treatment of quality, we consider the three basic contributions to quality provided for GHRSST compatible SST and define them as follows.

Bias A continuous variable that describes how well the retrieved SST matches the skin SST inferred from *in situ* sources. Since the most extensive sets of *in situ* measurements are not skin measurements, the inferencing process may involve models of varying degrees of complexity. In any case, if the difference is large in magnitude, then the quality would be considered worse. The degree in which the difference increases would in some sense be related to the degree in which the quality diminishes. In GHRSST compliant data sets, Bias is determined on a pixel by pixel basis under the field name `sses_bias`. We will use the symbol μ_{sses} to represent the bias.

Standard Deviation A continuous variable that describes how accurately the retrieved SST is determined. There are many possible methods for determining this parameter. Some may relate to comparisons with *in situ* measurements, others may include binning, while others may relate to the accuracy of the retrieval method (derived from covariances of retrievals from radiative transfer, for example). In any case, there will be a minimum value of standard deviation which represents the capability of the equipment. Furthermore, increases in standard deviation will result in decreased quality, and in some sense this should be proportionate in much the same way as the bias is. In GHRSST compliant data sets, Standard Deviation is determined on a pixel by pixel basis under the field name `sses_standard_deviation`. We will use the symbol σ_{sses} to represent the Standard Deviation.

Parameter name	Symbol	L3C		L3S	
sses_count	n	n_C	Indicative number of good quality L3U measurements merged to L3C	n_S	Indicative number of good quality L3U measurements merged to L3S
sses_bias	μ	μ_C	An estimate of the median bias of the platform and sensor over the time window of the L3C file.	μ_S	An estimate of the median bias over all measurements over the time window of consideration.
sses_standard_deviation	σ	σ_C	Indicative uncertainty over the time window, including contributions from natural variation as they affect the estimate of the mean SST.	σ_S	Indicative uncertainty over the time window, including contributions from natural variation as they affect the estimate of the mean SST.
sst_count	n_w	$n_{w,C}$	Number of measurements merged to L3C.	$n_{w,S}$	Number of measurements merged to L3S.
sst_mean	T_w	$T_{w,C}$	Unweighted mean measured sea surface temperature.	$T_{w,S}$	Unweighted mean measured sea surface temperature.
sst_standard_deviation	σ_w	$\sigma_{w,C}$	Unweighted standard deviation of measured sea surface temperature.	$\sigma_{w,S}$	Unweighted standard deviation of measured sea surface temperature.
l2p_flags	f_{L2p}	$f_{\mathrm{L2p},C}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location over the time window.	$f_{\mathrm{L2p},S}$	Bit flags indicating pixel state, based on relationships to land, ancillary fields, and other measurement conditions of all measurements contributing to the gridded location over the time window.
quality_level	q	q_C	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.	q_S	Quality level as a measure of cloud proximity for all of the measurements contributing to the gridded location.
sea_surface_temperature	$T_{\mathrm{satellite}}$	$T_{\mathrm{satellite},C}$	Estimate of the sea surface temperature characteristic of the time window.	$T_{\mathrm{satellite},S}$	Estimate of the sea surface temperature characteristic of the time window.

Table 20: Association between field names in GHRSST compliant files and symbols used in this text, with a short description of the intent of the parameter and symbol, for L3C and L3S.

Quality level Quality level assessments are most often made by assessments on cloud cover alone, although it is possible that other factors are also included. Quality level is assumed to be an ordinal only assessment. Lower quality indicates a lower likelihood of a measurement being good, with the degree of likelihood remaining inconsistent. In GHRSST compliant data sets, quality level is determined on a pixel by pixel basis under the field name `quality_level`.

Due to the possibility of different methods of determination, it is not immediately clear that the Bias and Standard Deviation assessments are correlated in any way, and we will make an assumption that in general they are not. We also consider the sources of quality determination from Bias and Standard Deviation as potentially distinct. Ignoring what is possibly a positive correlation between μ and σ (empirically this is evident in ABOM data sets, see [**GHRSSTdoc**] for more details), will result in an over estimation of uncertainty, but we consider this tolerable on the basis that uncertainty is better to overestimate than to underestimate.

On the other hand, it is expected that the `quality_level` assessment, if it really is to reflect quality, should be at worst quantitatively related to either greater absolute bias or greater standard deviation, or both.

Thus, we consider a complimentary quality indicator, q_s, computed based on the SSES parameters, which displays a similar scaling behaviour to `quality_level`, and choose to supplement the `quality_level` with this new indicator, by choosing the minimum,

$$\texttt{quality\_level} \rightarrow \min(\texttt{quality\_level}, q_s) \tag{154}$$

q_s is defined in such a way that it varies over the history of the data set and from scene to scene, and is computed on a pixel by pixel basis, allowing the overall pixel quality to be compared between scenes, and between different data sets over different time periods.

A.4.1 Standard Deviation component to quality

The SSES standard deviation assessment is composed of two components. The first component represents the absolute sensor in context capability, σ_{sensor}, the best accuracy we could expect from the sensor at the particular time of life in the general context of making SST measurements. The second component is from other sources, which include geophysical as well as algorithmic, that relate specifically to the direct application of the retrieval technique to the individual measurements, σ_{other}. We assume that in general these other sources are uncorrelated with the sensor in context sources, and with the assumption of Gaussian measurements which is implicitly assumed throughout, we can write,

$$\sigma^2_{\text{sses}} = \sigma^2_{\text{sensor}} + \sigma^2_{\text{other}} \tag{155}$$

Furthermore, we assume that the sensor component is always present in the SSES estimate, and bounded below by the ideal best possible sensor performance σ_0, then the extent to which the measurement is not ideal can be expressed by the variance, $\sigma_{\sigma,q}$,

$$\sigma^2_{\sigma,q} = \sigma^2_{\text{sses}} - \sigma^2_0 \tag{156}$$

As $\sigma_{\sigma,q}$ increases we expect the overall quality of measurement to decrease, and in the limit that other sources of quality assessment are negligible, it is natural to expect a qualitative relationship such as,

$$q_{\text{sses}} \sim \frac{\sigma_{\sigma,q}}{\sigma_0} \tag{157}$$

The natural scale of the variance being the sensor performance σ_0.

A.4.2 Bias component to quality

The bias assessment, μ_{sses} allows us to determine how far from the typical expected bias to *in situ* measurements, μ_0, each measurement falls. In order to decouple this assessment from standard deviation related assessments, we could standardize the distribution to one that has mean 0 and standard deviation σ_0, under the assumption that the best estimate of the standard deviation of the distribution of μ_{sses} is given by the estimate of the variation above the natural sensor performance, σ_{sses},

$$\mu_q = \frac{\mu_{\text{sses}} - \mu_0}{\sigma_{\text{sses}}}\sigma_0 \tag{158}$$

As μ_q increases in magnitude, we naturally expect the quality to decrease (or the risk that the retrieval may be suspect increases, because the measurement is further placed from a standard), thus we consider the contribution to the quality assessment from the mean, $\sigma_{\mu,q}$,

$$\sigma^2_{\mu,q} = \mu_q^2 \tag{159}$$

$$= \left(\frac{\mu_{\text{sses}} - \mu_0}{\sigma_{\text{sses}}}\right)^2 \sigma_0^2 \tag{160}$$

and, as before, it is natural to expect a relationship such as,

$$q_{\text{sses}} \sim \frac{\sigma_{\mu,q}}{\sigma_0} \tag{161}$$

Further, standardization of the contribution results in a natural scale of $\sigma_{\mu,q}$ comparable to the natural scale of $\sigma_{\sigma,q}$.

A.4.3 Overall quality assessment from Bias and Standard Deviation

We can combine both the standard deviation and bias assessment of quality, since the natural scale of each parameter is the same, by assuming that both contribute equally to the assessment. Averaging the variances, and normalizing each component by σ_0^2 so that the quality assessment, $q_s ses$ is dimensionless,

$$q_{\text{sses}} = \frac{1}{\sigma_0}\sqrt{\frac{1}{2}\left(\sigma^2_{\sigma,q} + \sigma^2_{\mu,q}\right)} \tag{162}$$

Substituting equations 156 and 161,

$$q_{\text{sses}} = \frac{1}{\sqrt{2}}\sqrt{\max\left(\left(\frac{\sigma_{\text{sses}}}{\sigma_0}\right)^2 + \left(\frac{\mu_{\text{sses}} - \mu_0}{\sigma_{\text{sses}}}\right)^2 - 1, 0\right)} \tag{163}$$

When evaluating q_{sses}, imposing the condition that the argument of the square root is greater than zero compensates for errors in estimating κ, λ, σ_0 and μ_0 without significantly changing the overall assessment as long as these errors are small.

q_{sses} defined in this way is the uncorrelated mean absolute z score derived from the standard deviation spread from minimum in the large sample limit, and bias shift (under that assumption the σ_{sses} characterizes the standard distribution of the bias shift), for Gaussian measurements. q_{sses} is furthermore unbounded.

The GHRSST standard demands that the quality level be represented as a number between 0 and 5 inclusive, which is the range that we wish to apply to q_s. The standard dictates that pixel of higher quality should have a higher probability of being good quality, but does not impose any further interpretation. We propose translating q_{sses} into a variable that meets these general requirements, by exponential scaling, as follows,

$$q_s = \lfloor 5 \exp^{\eta q_{\text{sses}}} \rceil \tag{164}$$

Where the nearest integer function is represented by $\lfloor x \rceil$. The η parameter sets the scale for q_s, and is chosen such that the degradation in quality determined by SSES measurements is similar to the observed degradation in `quality_level` over a period of time where the sensor is known to perform well.

The exponential function represents a cumulative distribution function for the maximum entropy distribution of a positive unbounded random variable - a natural choice of distribution given. Thus, our definition of q_s, in addition to providing the basic requirements, ensures that there is an intuitive relation between the quality level and the cumulative likelihood that the pixel is good. Moreover, the definition of q_s is entirely fixed by the constants $\{\sigma_0, \mu_0, \eta\}$, which are fixed over the life of the sensor retrievals, ensuring that the quality designation is maintained consistently and comparatively over time. This means for example that $q_s = 5$ retrievals have equivalent quality during times when the sensor performs well, and during times when the sensor does not perform well. During periods of degraded performance, the number of retrievals of high q_s will decrease, as it will during times of the day when performance may also be questionable due to uncertainties associated with the retrieval method.

Moreover, different data sources can also be compared using the same quality assessment, so long as η and σ_0 for each data source are chosen to be relatively constant,

$$\frac{\eta}{\sigma_0} = \text{constant} \tag{165}$$

A.4.4 Determination of characteristic parameters

In order to determine the appropriate values of the characteristic parameters which apply for a given GHRSST product, the following procedure was considered,

- Identify σ_0 from studies concerning the sensor used to measure brightness temperatures in the context of the application at hand. Alternatively, if an independent estimate is not available, the minimum value of standard deviation assessed against an *in situ* standard over a range of retrievals over a sufficiently long period of time (at least one season) when the sensor was considered well performing, could be determined, and this value assumed as σ_0. If data is being compared across instruments of the same type, and the quality indication is required to reflect the better quality of one instrument compared to the other, σ_0 should be kept constant.

For all NOAA AVHRR platforms, for example we use a single value of σ_0 so that all platforms can reference the same standard.

- Identify μ_0 by assessing the algorithm used to determine `sses_bias`. If this is not available, validate the provided data set against *in situ* measurements over a sufficiently long period of time when the sensor is well performing, and assume μ_0 is the mean bias against this source, corrected for systematic biases such as skin measurements.

- Identify η by comparing `quality_level` with q_{sses} over a sufficiently long period of time, when a sensor is performing well. For sensors of the same type, a constant η is maintained, so that performances can be assessed relatively. When sensors of different types are compared, the use of equation 165 allows the scale to be maintained consistently.

For ABOM GHRSST L2P NOAA AVHRR data sets, the `sses_standard_deviation` computation is performed by binning against *in situ* measurements. σ_0 is determined from information provided by direct measurements of brightness temperature. In the review of Minnett[15] for example, several validation studies involving the AVHRR sensor based SST are discussed, indicating a typical standard deviation of $\sigma \sim 0.23K$ for buoy measurements of SST[20] and a minimum standard deviation of $\sigma \sim 0.24K$ for AVHRR to M-AERI radiometer SST validations[17]. Since ABOM SST retrievals are based on regression to buoys, and it has a similar standard deviation to validations with other radiometers, we choose $\sigma_0 = 0.23K$ as a typical value representative of this uncertainty and refer to this as the sensor in context uncertainty. It should be noted that this is considerably higher than the quoted instrument noise or Noise Equivalent Temperature Difference ($NE\Delta T$) of the infra-red channels of the AVHRR sensor, which is $0.12K$[5], but better reflects the working uncertainties that we expect from an SST retrieval in application. The `sses_bias` estimation algorithm has $0K$ skin temperature offset, thus $\mu_0 = 0$. What remains is a determination of η to set the scale of q_{sses} against `quality_level`.

To this end, we consider the cumulative distribution of the quality of retrievals, on a view by view basis,

$$q_{\mathrm{cum}}(q,t) = \frac{\alpha(t)}{\sum_{\mathtt{quality\_level}} n(\mathtt{quality\_level}, t)} \sum_{\mathtt{quality\_level}=q}^{\mathtt{quality\_level}=5} n(\mathtt{quality\_level}, t) \tag{166}$$

The cumulative distribution is scaled by an arbitrary normalization $\alpha(t)$, which is dependant on the geophysical as well as algorithmic aspects of the retrieval which we characterize by the time of observation, t. Determination of α can be done away with by considering the relative cumulative distribution, which we expect to scale with q_{sses}, at least statistically, according to equation 164,

$$\log\left(\frac{q_{\mathrm{cum}}(5,t)}{q_{\mathrm{cum}}(2,t)}\right) \sim \eta\left(q_{\mathrm{sses}}\big|_{\mathtt{quality\_level}=5} - q_{\mathrm{sses}}\big|_{\mathtt{quality\_level}=2}\right) \tag{167}$$

Since the retrievals use different algorithms for day and night, we consider these as statistically distinct populations. For robust statistics, we consider the median q_{sses} for every quality level of each observation, and construct a 14 day rolling median of these medians for day and night populations. (The choice of 14 days was chosen as a time scale which is consistent with the persistence of the SST and the degree of coverage. The result is not sensitive to this choice.) We consider both the ABOM real time (fv01) and archival (fv02) L2P data sets, over different periods of time, and the

Platform	real time (fv01) η	archival (fv02) η
NOAA-15 AVHRR L2P	-0.1007	-0.2161
NOAA-16 AVHRR L2P	n.a.	-0.3982
NOAA-17 AVHRR L2P	n.a.	-0.5175
NOAA-18 AVHRR L2P	-0.0497	-0.3339
NOAA-19 AVHRR L2P	-0.0894	-0.2614

Table 21: Determination of the quality scaling parameter η for both ABOM L2P data sets and a range of NOAA POES platforms. The smaller coefficients in fv01 data sources indicate weaker discriminatory power in the SSES estimation towards quality level assessment, compared to fv02 data sources.

NOAA POES platforms available over those time periods. Figure 71 shows the 14 day median of median q_{sses} at different `quality_level` over time, for different data sources, demonstrating the evolution of the quality of observation over time. The data for NOAA-19 AVHRR is shown for brevity, and because we wish to standardize our quality assessments to this platform. The dark lines represent night data, whereas bright lines represent day data. Generally speaking, higher quality data corresponds to lower q_{sses}, however in the real time processing the distinction is not so clear. In the real time case, the estimate of uncertainties is based on recent *in situ* measurements, whereas the historical estimates are based on a longer term model, which has less variability.

For each data source, η be estimated by linear regression, as shown in figure 72.

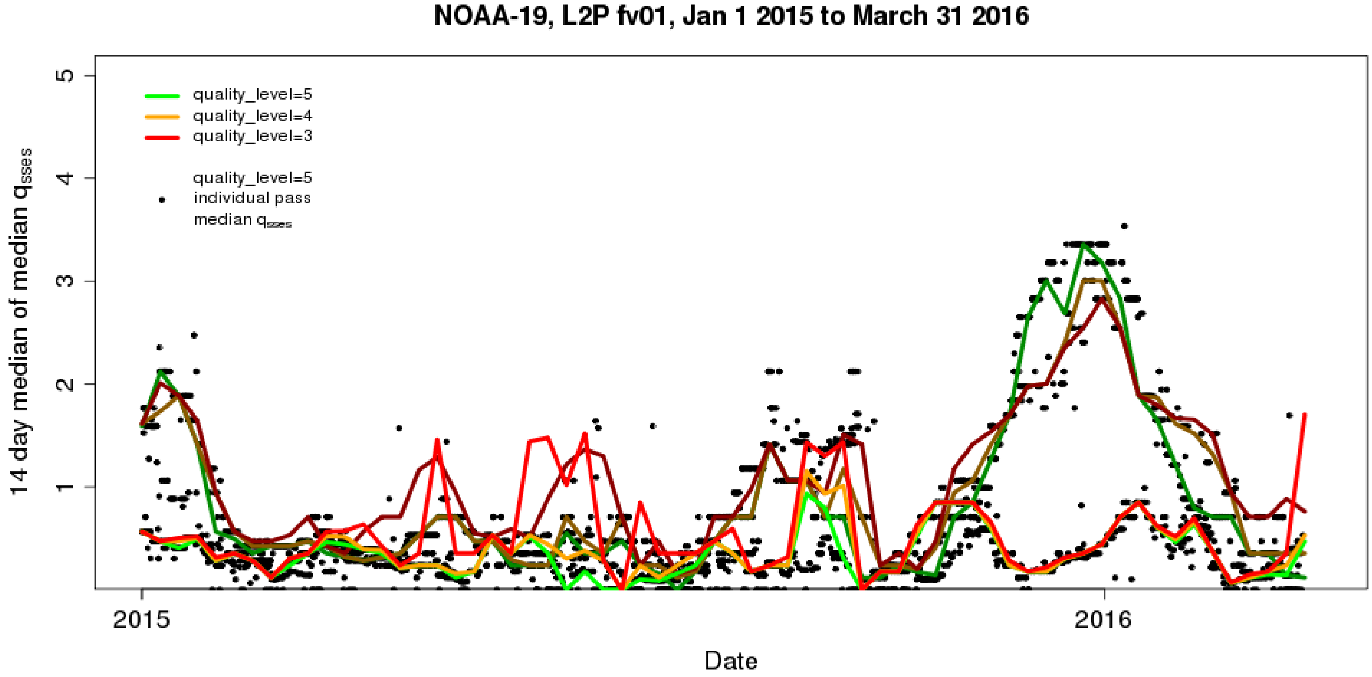
The values of η determined from each platform and data set are tabulated in table 21. For a standardized basis of comparison we choose NOAA-19 on the archival data set as the reference, since this represents the best performing platform in our AVHRR data set, thus $\eta = -0.2614$

Using $\eta = -0.2614$ and $\sigma_0 = 0.23K$, we can remap the 14 day median of median q_s real time data to compare real time quality. Figure 73 shows q_s prior to the application of the nearest integer function. The q_s values on the time series can be directly compared, for example, near the start of 2016 there is a period when both NOAA-19 night and NOAA-18 retrievals suffer a distinct drop in quality - the NOAA-19 drop is more severe than NOAA-18. NOAA-18 retrievals are generally of lower quality than NOAA-19 throughout, and NOAA-15 retrievals are typically the worst of the three, although over the start of 2016, night retrievals from NOAA-15 appear to be of better quality than NOAA-19. This information allows decisions to be made comparatively on the value of three platforms over the time period, and how the best quality information from each source could be used.

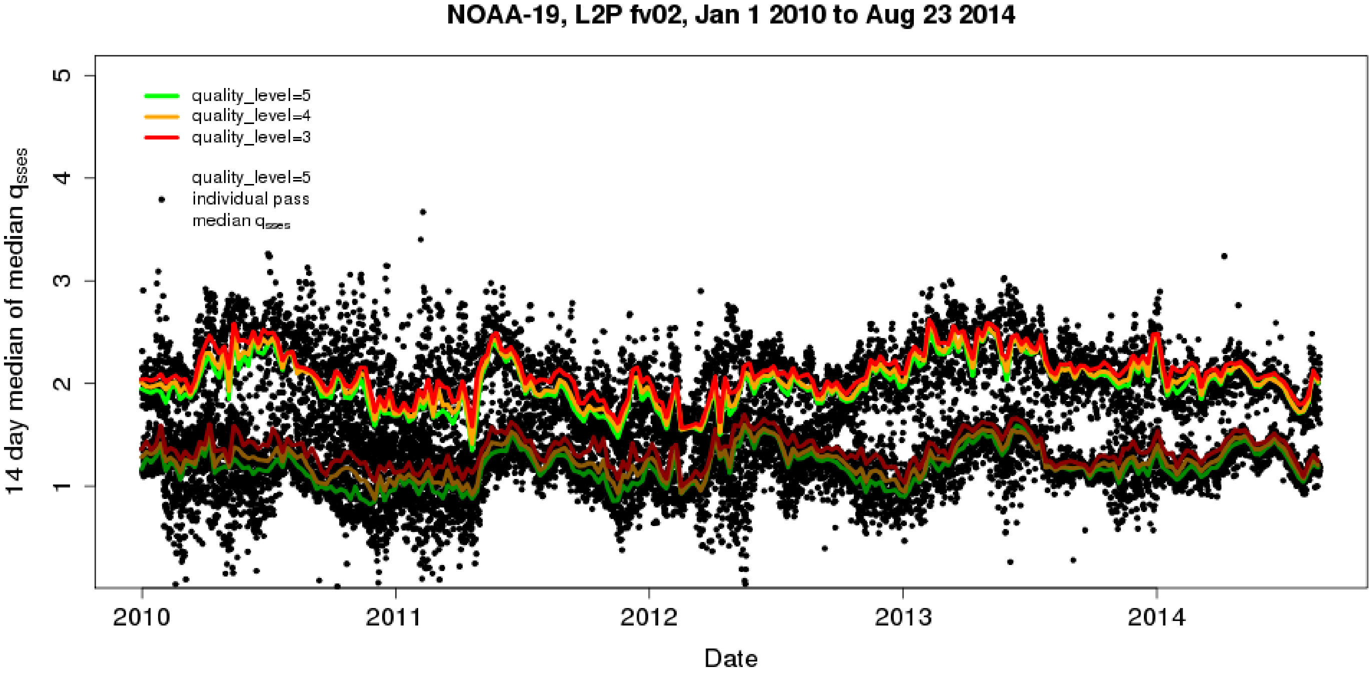
On the NPP VIIRS L3U platform from ACSPO [1], σ_0 can be determined by considering the differences in $NE\Delta T$ for VIIRS against AVHRR, and adjusting in quadrature,

$$\sigma^2_{0,\text{VIIRS}} = \sigma^2_{0,\text{AVHRR}} - NE\Delta T^2_{\text{AVHRR}} + NE\Delta T^2_{\text{VIIRS}} \tag{168}$$

For VIIRS infra red channels, $NE\Delta T$ is in the region of $0.037K$[16], about one third of the value of AVHRR. However, the ACSPO method uses a retrieval that references *in situ* measurements, which in the context of AVHRR retrievals have been given an uncertainty of $0.23K$, which far dominates the $NE\Delta T$ estimate. Making use of equation 168, we derive an NPP VIIRS estimate of $\sigma_0 = 0.227$. For η, we make use of equation 165, leading to $\eta = -0.17$. The `sses_bias` estimation algorithm has $0K$ skin temperature offset, thus $\mu_0 = 0K$, however, since a global sub-skin to skin bias not been removed, and `sea_surface_temperature` needs to be corrected for this systematic bias if direct comparison with the NOAA AVHRR data is required.



NOAA-19, fv01, real time data stream.



NOAA-19, fv02, delayed mode historical data stream.

Figure 71: 14 day median of median q_{sses} for various `quality_level` for day (bright colors) and night (dark colors) over several years for historical and real time ABOM data streams. There is a distinction between day and night processing, where different algorithms are employed.

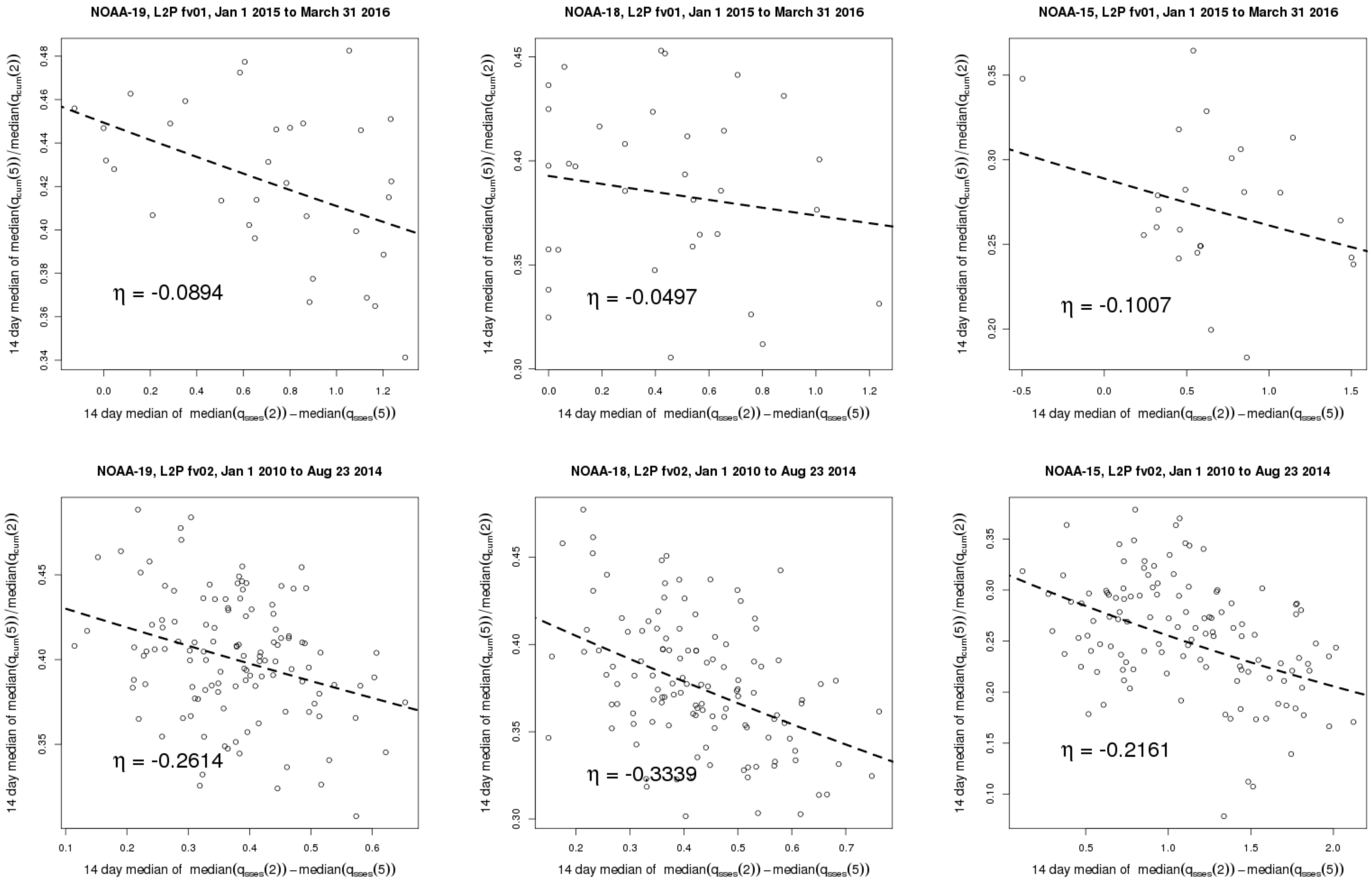
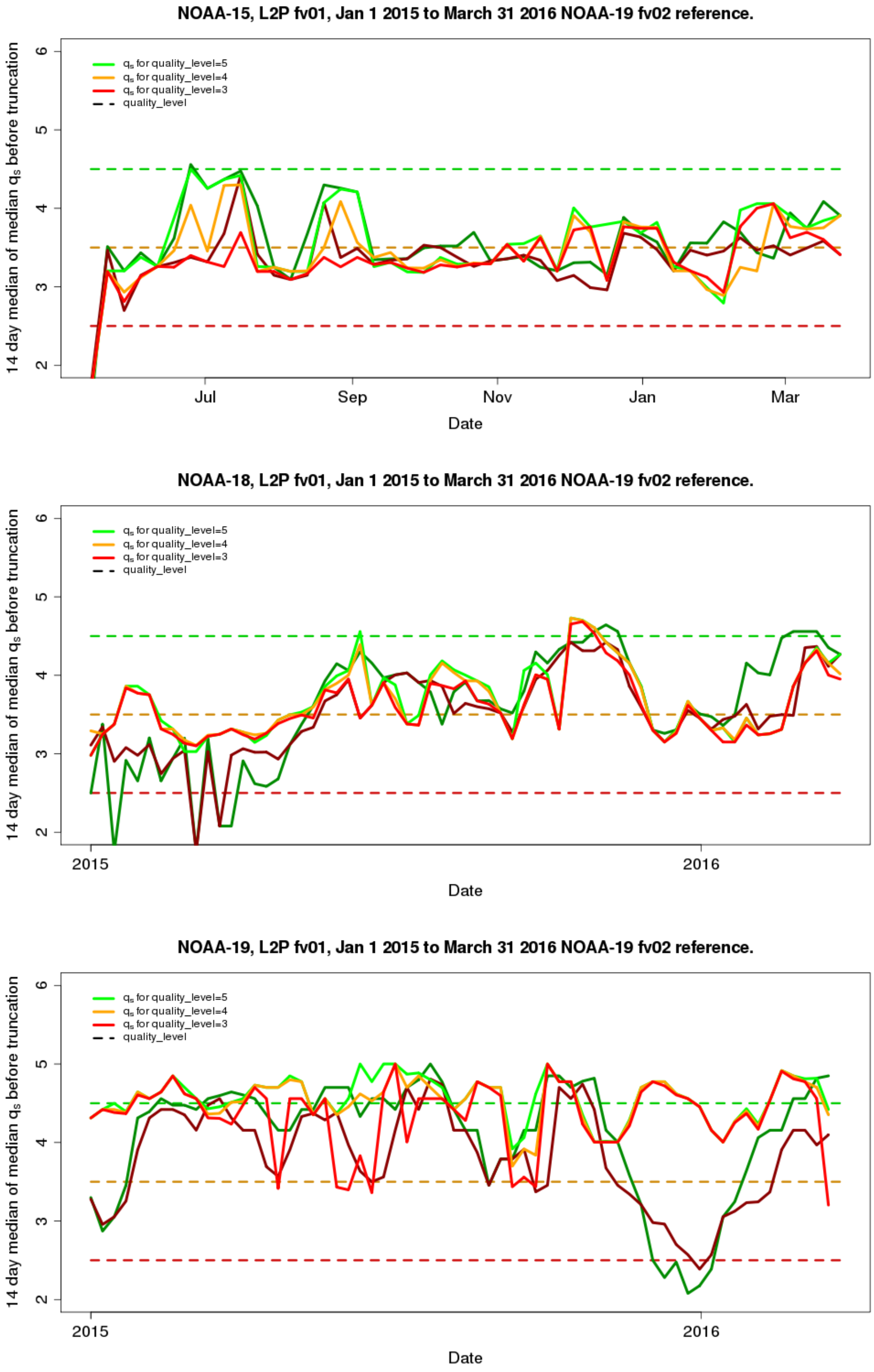


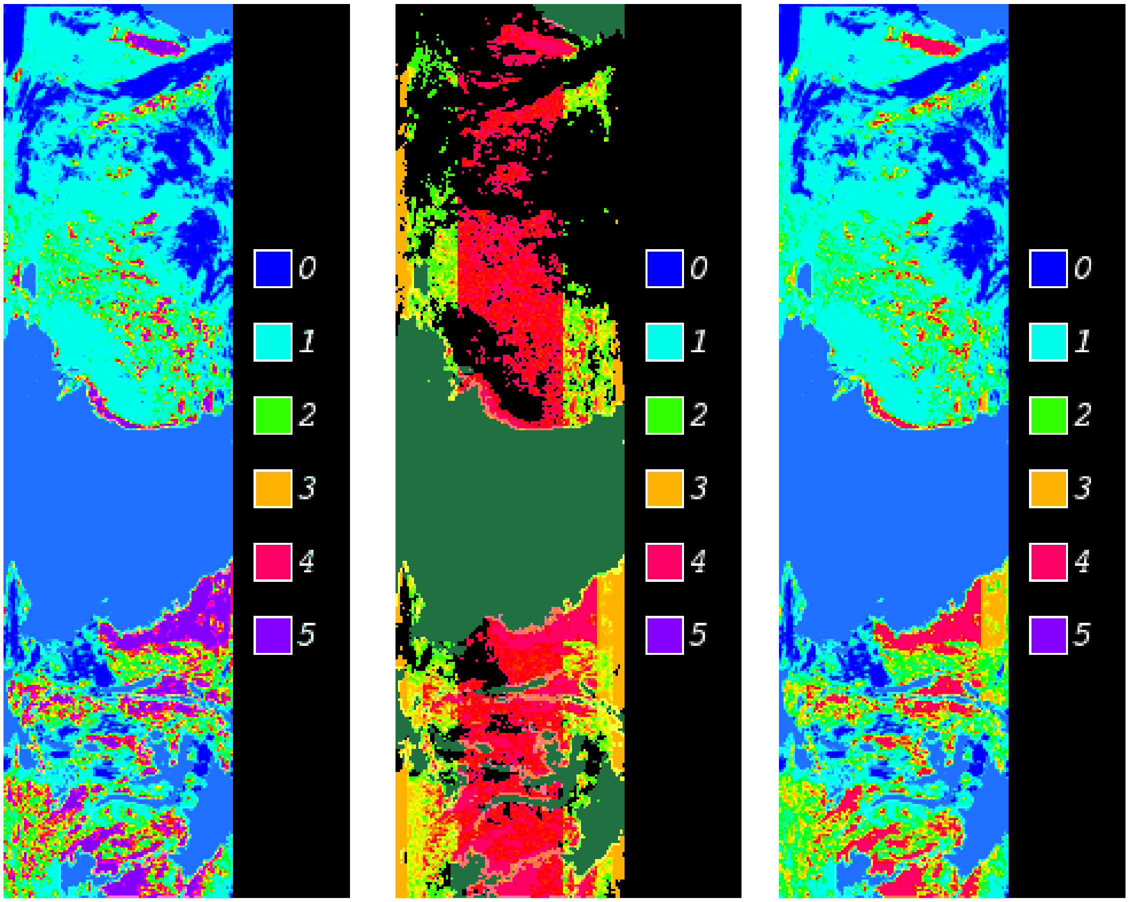
Figure 72: Estimates of η for different NOAA POES platforms. η indicates the scaling of q_{sses} against the assignment of `quality_level`. Data shown for the NOAA POES platforms that


NOAA-15, L2P fv01, Jan 1 2015 to March 31 2016 NOAA-19 fv02 reference.

NOAA-18, L2P fv01, Jan 1 2015 to March 31 2016 NOAA-19 fv02 reference.

NOAA-19, L2P fv01, Jan 1 2015 to March 31 2016 NOAA-19 fv02 reference.

Figure 73: Time series of median q_s at various `quality_level`, for real time data on NOAA-15, NOAA-18 and NOAA-19 platforms. Day data is represented by bright colors, whereas night data is darker. Reception of NOAA-15 before mid April 2015 had some difficulties.



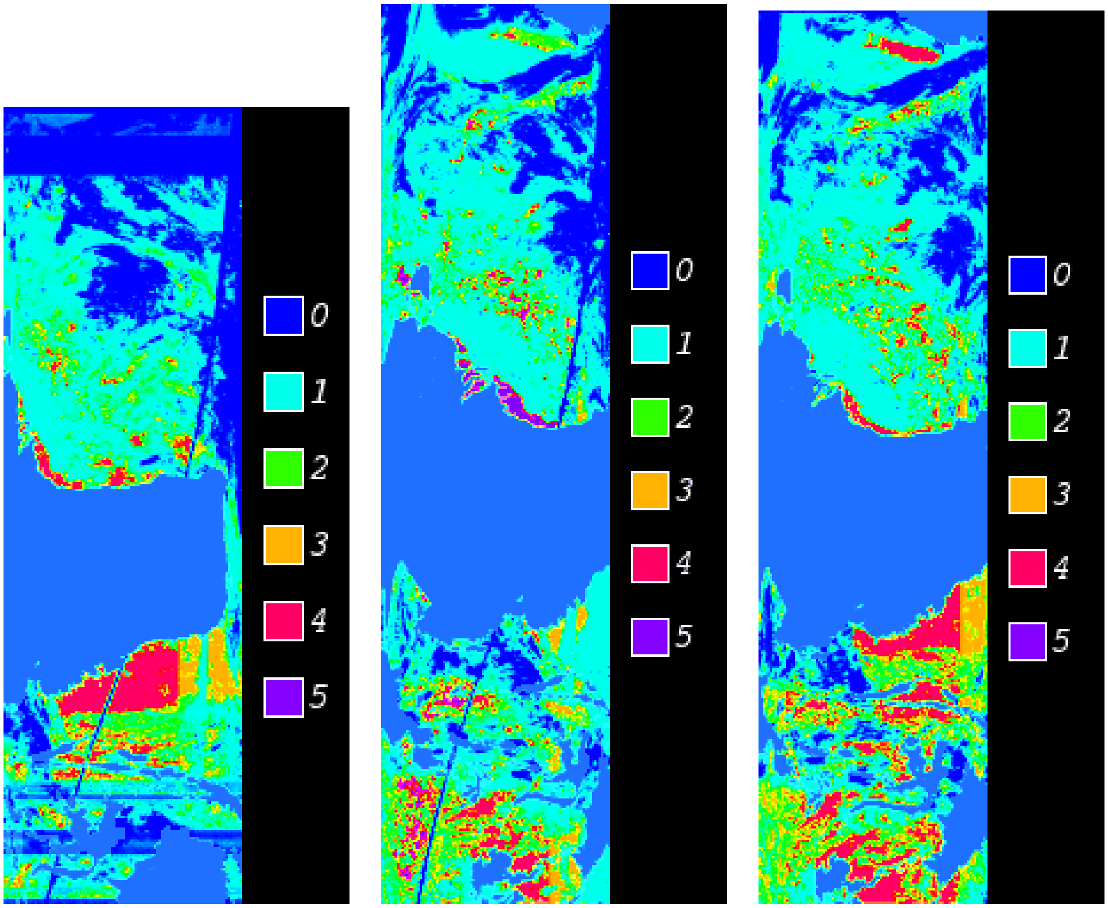
`quality_level` assignment q_s assignment remapped `quality_level`

Figure 74: Comparison between `quality_level` and q_s assignments for a single swath of real time (fv01) AVHRR NOAA-19 SST. In the swath shown above, in addition to distance to cloud, the view angle on the swath provides a significant contribution to the quality reassignment assignment. q_s assignments can only be made where SSES are provided, and they are not provided for low quality observations in the ABOM fv01 data set.

A.4.5 Using quality reassessments to provide best quality merges from multiple sources

In order to see how this can be used in practice, we consider the task of aggregating data from three NOAA AVHRR platforms and one NPP VIIRS platform as a single pass, each with approximately the same acquisition time for the ABOM real time (fv01) processing stream. The aggregation follows the general method outlined in section A.1, and involves a statistical variance weighted mean of the best quality pixels, with appropriate weightings for overlap (in the event that grid squares are not colocated). To determine the best quality, reassign the `quality_level` based on equation 154, tuned to the archival NOAA19 AVHRR assessment standard.

The reassignment of `quality_level` for every observation is illustrated visually in figure 74. q_s determination places an upper bound of quality four over the central region of the swath, with degraded quality towards the edges. When this is combined with the original `quality_level` assignment, the resulting reassignment downgrades the quality at higher zenith angle. Having a lower quality provides a non-parametric filter against this data, which will tend to naturally favour data near the center of the field of view when aggregating.

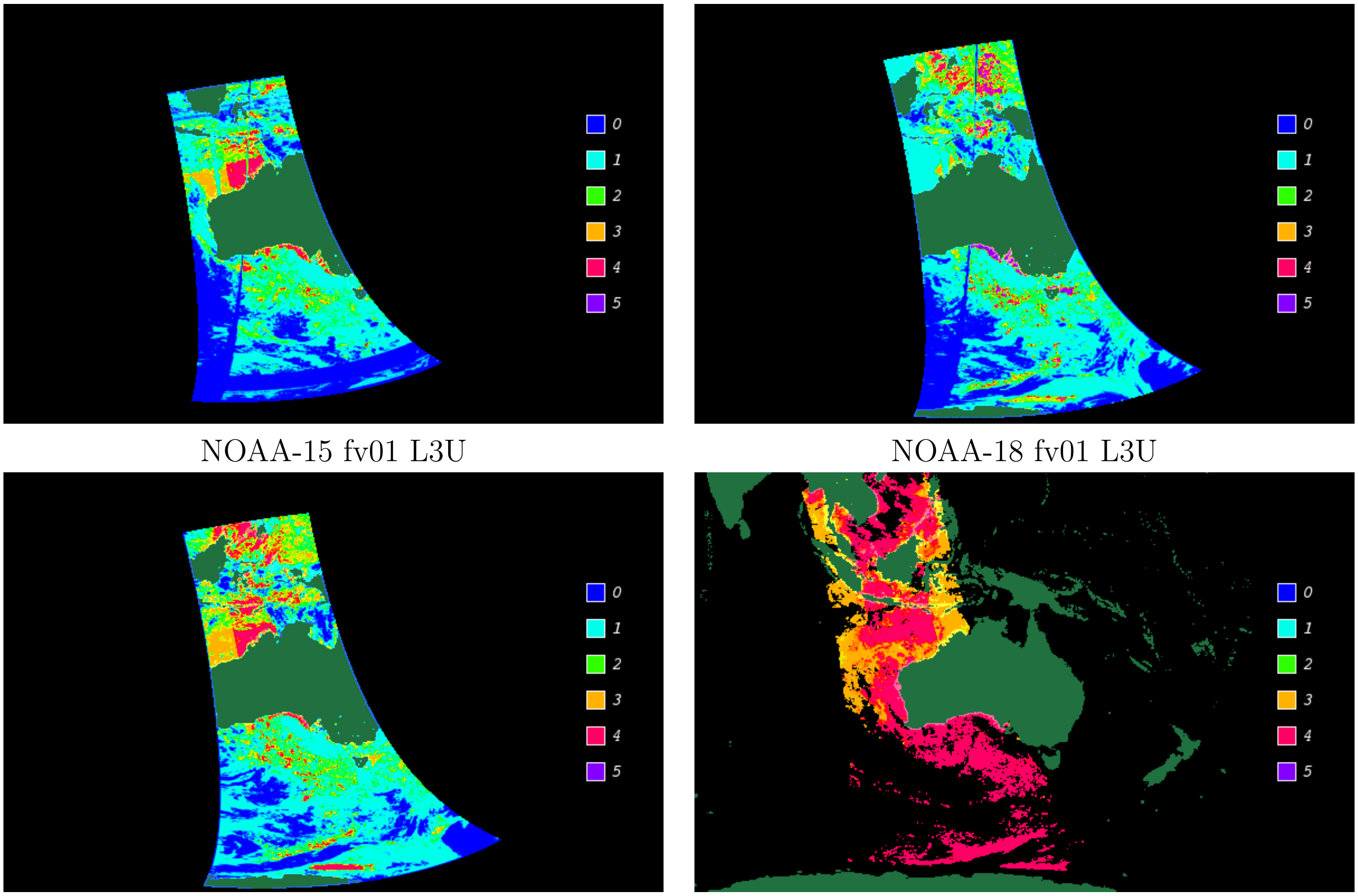


NOAA-15 fv01 NOAA-18 fv01 NOAA-19 fv01

Figure 75: Remapped `quality_level` assignments applied to single passes from three different platforms. NOAA-15 is considerably degraded in terms of coverage and number of good quality measurements compared to NOAA-19, however there are still some good quality measurements that may be worth considering.

Considering three NOAA AVHRR platforms, swaths from similar time periods after quality reassignment can be similarly compared. See figure 75. It is clear from the reassignment that NOAA-15 data is of considerably degraded in quality compared to NOAA-19, in terms of the number of good quality observations and in previous real time merges, we would have excluded NOAA-15 completely from the process. However, in this particular example, it is also clear from the images that there is information from the NOAA-15 swath that may have value since it is either missing or on the edges of the NOAA-18 and NOAA-19 field of view, such as good quality in the center of the NOAA-15 swath, indicating value including this data. The reassigned `quality_level` allows the determination of which NOAA-15 observations to be included to be made, without polluting the observations with a possibly degraded measurement.

To make a comparison with NPP VIIRS SST, we remap the platforms onto a common L3U grid, following the methods of section A.1. The ACSPO NPP VIIRS L3U SST data set `quality_level` is typically not graduated, so all of the graduation of the quality is derived from q_s. Applying the same procedure to this data set, with the adjusted σ_0 and η to bring the NPP VIIRS to the NOAA-19 AVHRR baseline, over a single swath near the same time as the AVHRR data, yields



NOAA-15 fv01 L3U

NOAA-18 fv01 L3U

NOAA-19 fv01 L3U

L3U NPP VIIRS

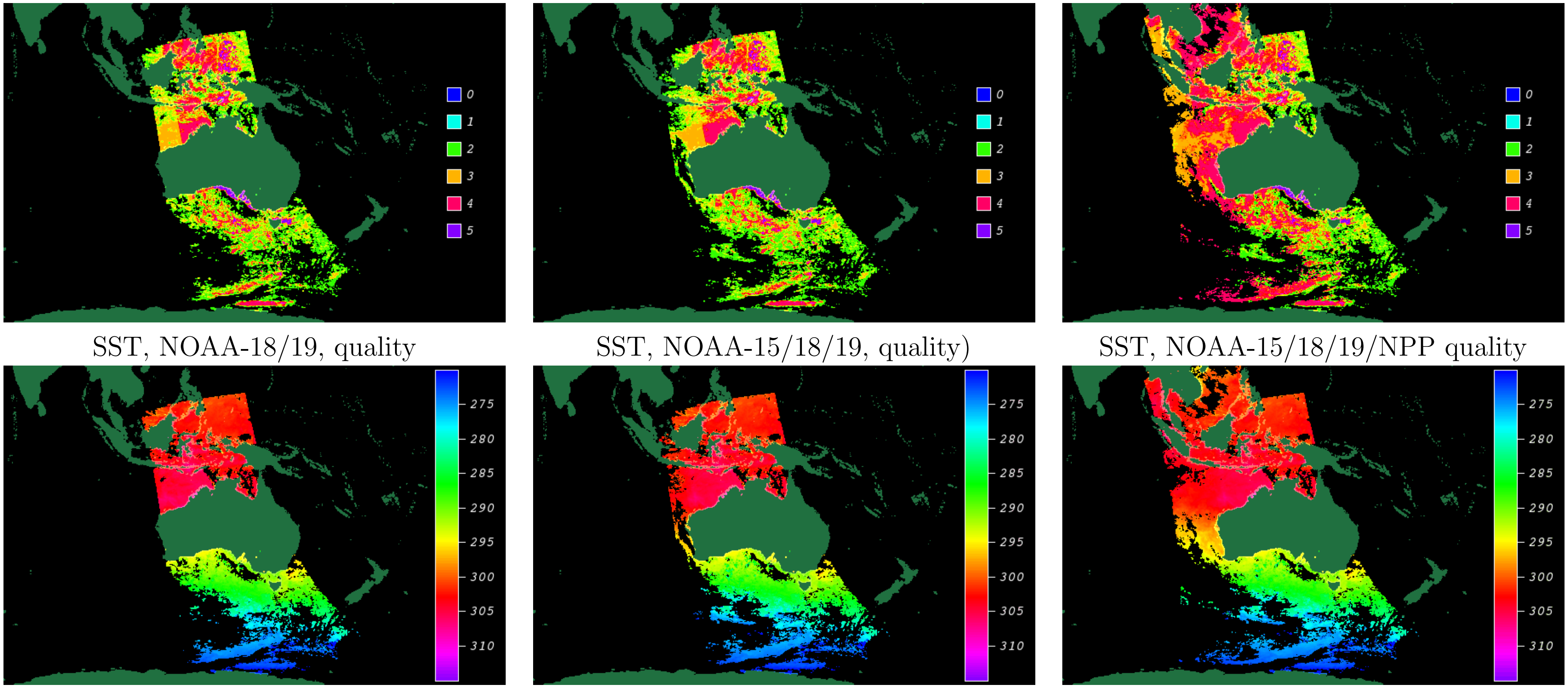
Figure 76: remapped `quality_level` applied to single passes from three different NOAA AVHRR platforms and NPP VIIRS, on a common rectangular coordinate system. The NPP VIIRS SST retrieval has superior overall coverage, with quality degradation over the tropics near the edges of the field of view, however there are still measurements from the AVHRR platforms that could be used to increase the coverage further.

a quality comparison that is illustrated in figure 76. A similar situation is observed, although the NPP VIIRS L3U product has considerably better coverage, there are regions where the quality and coverage of NOAA AVHRR could extend it, and the remapped `quality_level` allows this to be done selectively.

Figure 77 shows the result of the aggregation. The rightmost images show two platform composites of remapped `quality_level` and SST, using the rules on which the current data sets are defined. The center image shows the result using data from all three available AVHRR platforms, the coverage is slightly improved by the selective addition of the degraded NOAA=15 platform. The right image includes NPP VIIRS, resulting in a significant improvement in both coverage and extent of quality, particularly in the northern latitudes.

A.4.6 Generalizing to include other quality factors in the assessment

GHRSST compliant data sets include ancillary fields which estimate wind speed, aerosols, and ice. These are included in the files because of the possible impact they may have on quality of retrieval.



SST, NOAA-18/19, quality

SST, NOAA-15/18/19, quality)

SST, NOAA-15/18/19/NPP quality

SST, NOAA-18/19, SST

SST, NOAA-15/18/19, SST

SST, NOAA-15/18/19/NPP SST

Figure 77: Aggregated composites based on various combinations of platforms. Best reassigned `quality_level` is used in the aggregation process.

Although these are geophysical parameters and are expected to be fixed in time irrespective of the observing instrument, it might be of interest including uncertainties due to these parameters for comparisons over time. This could be simply done by adding variance rations in equation 163, that include the additional information.

A.4.7 Generalizing for measurement distributions that are expected to be asymmetric and/or fat tailed

The approach outlined previously can be generalized to measurement distributions that are non-Gaussian and/or asymmetric. This is not necessarily appropriate for SST retrievals, but may be required for using similar methods on other geophysical parameters. The generalization is relatively straight forward and can be expressed analytically if we restrict ourselves to the stable distributions, which are the distributions which are outcomes of the most general form of the central limit theorem. The generalization requires that the σ and μ parameters which are formally divergent or undefined if considered to be mean and standard deviation parameters in many of these distributions, are considered as proxies for the shape and location parameters. This assumption is reasonable provided sample size variations of σ and μ are fixed or small over comparative determinations of σ and μ and the typical sample sizes involved. (The formal divergences of these parameters appear as large sample limits in positive powers of the sample size, $n^{|p|} \to \infty$ as $n \to \infty$, thus, the sample size can be thought of as a regularizing parameter.)

Making use of the stability parameter α, and a skewness parameter β, the properties of the stable distributions suggest that q_{sses} can be generalized as follows,[22]

$$\begin{align}
q_{\mathrm{sses}} &= 2^{-\frac{1}{\alpha}} \left\{ \left| \frac{\sigma_{\mathrm{sses}}}{\sigma_0} \right|^{\alpha} + \left| \frac{\mu_{\mathrm{sses}} - \mu_{0,\alpha}}{\sigma_{\mathrm{sses}}} \right|^{\alpha} - 1 \right\}^{\alpha^{-1}} \tag{169} \\
\mu_{0,\alpha} &= \mu_0 + \mathrm{sgn}\left(\mu_{\mathrm{sses}} - \mu_0\right) \beta \left|\sigma_{\mathrm{sses}}\right|^{\alpha} \tan\left(\frac{\pi\alpha}{2}\right), \ (\alpha \neq 1) \notag \\
&= \mu_0 + \mathrm{sgn}\left(\mu_{\mathrm{sses}} - \mu_0\right) \frac{\beta}{\pi} \left|\sigma_{\mathrm{sses}}\right| \log\left(\left|\frac{\sigma_{\mathrm{sses}}}{\sqrt{2}}\right|\right), \ (\alpha = 1) \tag{170}
\end{align}$$

The applicability and use of α and β in the assessments of q_{sses} is a focus of ongoing research.

References

[1] Ignatov A. et al. *GHRSST v2 Level 3U Global Skin Sea Surface Temperature from the Visible Infrared Imaging Radiometer Suite (VIIRS) on the Suomi NPP satellite created by the NOAA Advanced Clear-Sky Processor for Ocean (ACSPO) (GDS version 2). NOAA National Centers for Environmental Information. Dataset.* 2015. URL: `https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:GHRSST-VIIRS_NPP-OSPO-L3U`.

[2] Markov A. "On certain applications of algebraic continued fractions". PhD thesis. St. Petersburg, 1884.

[3] *Aerosol Optical Thickness 100km.* URL: `https://www.class.ncdc.noaa.gov/saa/products/search?sub_id=0&datatype_family=AERO100`.

[4] Walton C. C. et al. "The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites". In: *J. Geophys. Res.* C12.103 (1998), pp. 279–280.

[5] A. P. Cracknell. *The Advanbced Very High Resolution Radiometer.* Tailor and Francis, 1997. ISBN: 0-7484-0209-8.

[6] C.J. Donlon et al. "Towards improved validation of satellite sea surface skin temperature measurements for Climate Research." In: *J. Climate* 15 (2002), pp. 353–369.

[7] "Earth Surface Temperatures". In: A. P Crackell. *The Advanced Very High Resolution Radiometer.* Taylor and Francis, 1998. Chap. 4.

[8] Meteo France. *Marine Observation Monitoring, Quality Control Tools.* URL: `http://www.meteo.shom.fr/qctools/`.

[9] R.W Grumbine. "Automated Passive Microwave Sea Ice Concentration Analysis at NCEP". In: *Tech. Note, NOAA/NCEP* (1996), 13pp. URL: `http://polar.ncep.noaa.gov/seaice/Analyses.shtml`.

[10] Beggs H. et al. "Enhancing ship of opportunity sea surface temperature observations in the Australian region." In: *Journal of Operational Oceanography (ISSN: 1755-8778)* (5 2012), pp. 59–73. URL: `http://www.ingentaconnect.com/content/imarest/joo/2012/00000005/00000001/art00006`.

[11] Zhang H. et al. "Seasonal patterns of SST diurnal variation over the Tropical Warm Pool region". In: *J. Geophys. Res. Oceans* 121 (2016). DOI: `10.1002/2016JC012210`. URL: `http://onlinelibrary.wiley.com/doi/10.1002/2016JC012210/epdf`.

[12] *IMOS Satellite Remote Sensing Sea Surface Temperature (SST) Products.* URL: `http://imos.org.au/sstproducts.html`.

[13] Carlos M. Jarque and Anil K. Bera. "A test for normality of observations and regression residuals". In: *International Statistical Review* 55.2 (1987), pp. 163–172.

[14] Puri K. et al. "Implementation of the initial ACCESS numerical weather prediction system". In: *Aust. Meteorol. Oceanogr. J.* 63.0 (2013), pp. 265–284.

[15] P. J. Minnett. "The Validation of Sea Surface Temperature Retrievals". In: *"Oceanography from Space Revisited"*. Ed. by V. Barale, J.F.R Gower, and L Alberotanza. Springer, 2010. Chap. 14, pp. 229–247. ISBN: 978-90-481-8680-8.

[16] P. J. Minnett and R. H. Evans. *SST from VIIRS on NPP: prelaunch preparations and post-launch validation.* NASA SST Science Team Meeting. Seattle, 2010. URL: http://depts.washington.edu/papers/sst2010/plenary_calibration_etc/Minnett_SST_from_VIIRS_on_NPP.pptx.

[17] P. J. Minnett et al. "The marine atmospheric emitted radiance interferometer (M-AERI), a high-accuracy, sea-going infrared spectrometer." In: *J. Atmos. Oceanic Tech.* 18 (2001), pp. 994–1013.

[18] *NOAA POES Operational Status.* URL: http://www.ospo.noaa.gov/Operations/POES/status.html.

[19] *NOAA Satellite Information System, Advanced Very High Resolution Radiometer - AVHRR.* URL: http://noaasis.noaa.gov/NOAASIS/ml/avhrr.html.

[20] A. G. O'Carroll, J. R. Eyre, and Saunders R. W. "Three way error analysis between AATSR, AMSR-E and in situ sea surface temperature observations." In: *J. Atmos. Oceanic Tech.* 25 (2008), pp. 1197–1207.

[21] Lee D. P. et al. "The ERA-Interim reanalysis: configuration and performance of the data assimilation system". In: *Quarterly Journal of the Royal Meteorological Society* (2011). DOI: 10.1002/qj.828. URL: http://onlinelibrary.wiley.com/doi/10.1002/qj.828/abstract.

[22] Nolan J. P. *Stable Distributions. Models for Heavy Tailed Data.* 2009, 34pp. URL: http://www.math.ucla.edu/~biskup/275b.1.13w/PDFs/Nolan.pdf.

[23] Reynolds R.W. et al. "Daily high-resolution blended analyses for sea surface temperature". In: *J. Climate* 20.0 (2007), pp. 5473–5496.

[24] G. Schwartz. "Estimating the dimension of a model." In: *Annals of Statistics* 6 (1978), pp. 461–464.

[25] McClain E. P. Stowe L. L. Davis P. A. "Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the Advanced Very High Resolution Radiometer". In: *Journal of Atmospheric and Oceanic Technology* 16.6 (1999), pp. 656–681.

[26] P. Sykes et al. *Diurnal Variability in Sea Surface Temperature: Observation and model assessment.* 556. Met Office, Exeter, UK., 2011, 45pp. URL: http://www.metoffice.gov.uk/media/pdf/b/s/FRTR556.pdf.

[27] The GHRSST Science Team. *SST common priciples for generating SSES.* Web page no longer exists on www.ghrsst.org.

[28] The GHRSST Science Team. *SST definitions.* URL: https://www.ghrsst.org/ghrsst-data-services/products/.

[29] The GHRSST Science Team. *The Recommended GHRSST Data Specification (GDS) 2.0 revision 5.* GHRSST Int. Proj. Off. Reading, U. K. 2010. URL: https://www.ghrsst.org/governance-documents/ghrsst-data-processing-specification-2-0-revision-5/.

[30] Aihong Zhong and Helen Beggs. "Operational Implementation of Global Australian Multi-Sensor Sea Surface Temperature Analysis". In: *Analysis and Prediction Operations Bulletin* 00.77 (2008). URL: http://www.bom.gov.au/australia/charts/bulletins/apob77.pdf.