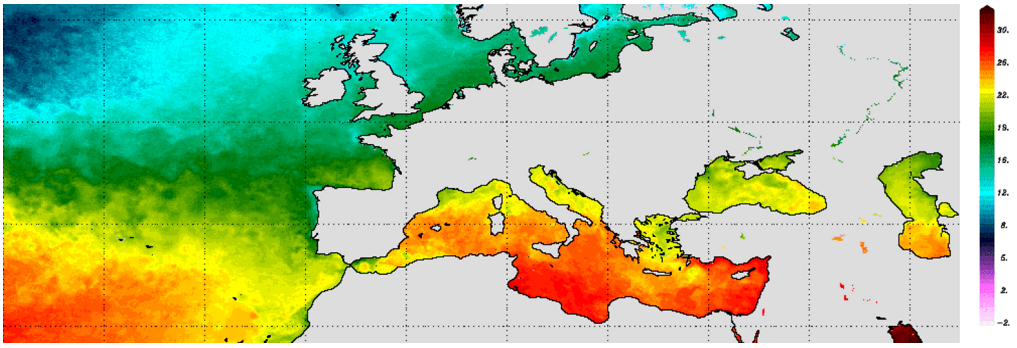
ESA CCI Phase 3 Sea Surface Temperature (SST)



Product Validation and Intercomparison Report D4.1 v2.1

Issue Date: **20 January 2023**

Document Ref: SST_CCI_D4.1_PVIR_v2.1

Contract: ESA/AO/1-9322/18/I-NB

SIGNATURES AND COPYRIGHT

Title : **ESA CCI Phase 3 Sea Surface Temperature (SST)**

Volume : **Product Validation and Intercomparison Report D4.1 v2.1**

Issued : **20 January 2023**

Authored :

Digitally signed by Craig James Donlon
Date: 2023.06.12 16:14:17 +02'00'

Owen Embury (UoR)

Address : University of Reading,
Whiteknights,
Reading,
Berkshire,
RG6 6AH,
United Kingdom

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Purpose and Scope

This document is the SST_CCI Product Validation and Intercomparison Report (PVIR). This version covers the Level 3 products of the version 3 Climate Data Record (CDR).

1.2 Executive summary

The SST CCI CDR v3 Level 3 products and their uncertainties have been validated against the in situ reference dataset. This is primarily based on drifting buoys supplemented by moored buoys, bottle, CTD, MBT, and XBT measurements prior to 1995 to improve coverage. Further comparison against ship SST measurements are used to verify bias spatial patterns in the 1980s.

A brief summary of the results per product is:

- ATSR / SLSTR
  - Absolute median difference to reference data: mostly $\lesssim$ 0.02 K
  - ATSR1 robust standard deviation: ~0.45 K
  - Other dual-view sensors have robust standard deviation: ~0.20 K night-time and 0.21-0.28 K daytime
  - Uncertainties are generally well estimated
  - Data do not show large scale systematic biases related to desert dust
  - Data do not show significant biases related to stratospheric aerosol from Mount Pinatubo eruption
- AVHRR
  - Absolute median difference to reference data: mostly $\lesssim$ 0.05 K
  - Robust standard deviations vary from ~0.55 K at start of record to ~0.25 K at end.
  - Early data (1980s and into 1990s) still show fluctuations in monthly global bias of few tenths of kelvin.
  - Night-time uncertainties are well estimated, but daytime are over estimated.
  - Desert-dust related biases are greatly reduced compared to previous CDR versions. Night-time data still show some residual cold biases to west of Africa and the Arabian sea at reduced magnitude 0.1 K (compared to 1 K previously).
  - Global timeseries is not affected by stratospheric aerosol related biases. However, a small ($\lesssim$ 0.4 K) regional bias is evident in night-time data immediately after the Mount Pinatubo eruption. A smaller regional bias may be present after the 1982 El Chichón.
- Analysis
  - Absolute median difference to reference data: mostly $\lesssim$ 0.05 K on decadal scales.
  - Robust standard deviations vary from ~0.50 K at start of record to ~0.22 K at end.
  - Desert-dust related biases no longer visible in L4 data.

- Analysis shows increased biases during periods based on just one or two AVHRRs in earlier decades. These are reduced in magnitude compared to CDRv2 and are mostly comparable to uncertainty in reference in situ.
- 1-2 months in 2001 remains an outlier (two poor AVHRRs).

Assessment of the uncertainties provided with the SST data suggests that the daytime AVHRR uncertainties are overestimated (i.e. the SST data are more accurate than the uncertainty estimates would indicate). The night-time AVHRR and ATSR (both day and night) uncertainties are well estimated and are a good representation of the spread of errors in the data across the range of uncertainties that occur.

1.3 Acronyms

The following acronyms and abbreviations have been used in this report with the meanings shown:

Acronym	Definition
AMSR	Advanced Microwave Scanning Radiometers
AVHRR	Advanced Very High Resolution Radiometers
ATSR	Along-Track Scanning Radiometers
C3S	Copernicus Climate Change Service
CCI	Climate Change Initiative
CDR	Climate Data Record
CMEMS	Copernicus Marine Environment Monitoring Service
DV	Diurnal Variability
ECMWF	European Centre for Medium-Range Weather Forecasts
EPS	EUMETSAT Polar System
ERS	European Remote Sensing
ESA	European Space Agency
GAC	Global Area Coverage
GTMBA	Global Tropical Moored Buoy Array
GTS	Global Telecommunication System
HadIOD	Hadley Centre Integrated Ocean Dataset
ICOADS	International Comprehensive Ocean-Atmosphere Dataset
L2	Level 2
L3	Level 3
L3C	Level 3 Collated
L3U	Level 3 Uncollated
L4	Level 4
MD	Matchup Dataset
MMS	Multi-sensor Matchup System
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
POES	Polar Operational Environmental Satellites
RSD	Robust Standard Deviation
SIRDS	SST CCI Independent Reference Data Set

SLSTR	Sea and Land Surface Temperature Radiometers
SST	Sea Surface Temperature
SST-CCI	ESA Climate Change Initiative on SST
VOS	Voluntary Observing Ship
WMO	World Meteorological Organization

2. PRODUCTS VALIDATED

This document describes the validation of products in the version 3.0 SST Climate Data Record, undertaken on Level 3 (that are gridded from Level 2) and Level 4 products.

The SST-CCI CDR provides global SST in the period 1980 though 2021 derived from three series of thermal infra-red sensors: the Advanced Very High Resolution Radiometers (AVHRRs), the Along-Track Scanning Radiometers (ATSRs), and the Sea and Land Surface Temperature Radiometers (SLSTRs); and two microwave sensors: the Advanced Microwave Scanning Radiometers (AMSR). The temporal coverage for each sensor type is shown in Figure 1 and sensor resolution in Table 1. Data are provided in four levels: L2P – source resolution in the satellite swath projection; L3U – remapped into files of grid-cell-mean SST at 0.05° latitude-longitude resolution; L3C – collated to daily (single-sensor) files; L4 – blended multi-sensor analysis to provide a daily gap-filled product. Figure 2 shows the different data levels.

Compared to the previous v2.1 CDR the major changes are:

- Longer time series: 1980 to 2021 (previous CDR was Sept 1981 to 2016)
- Improved retrieval to reduce systematic biases using bias-aware optimal methods (for single view sensors)
- Improved retrieval with respect to desert-dust aerosols
- Addition of dual-view SLSTR data from 2016 onwards
- Addition of early AVHRR/1 data in 1980s, and improved AVHRR processing to reduce data gaps in 1980s.
- Use of full-resolution MetOp AVHRR data (previously used ‘global area coverage’ Level 1 data)
- Inclusion of L2P passive microwave AMSR data

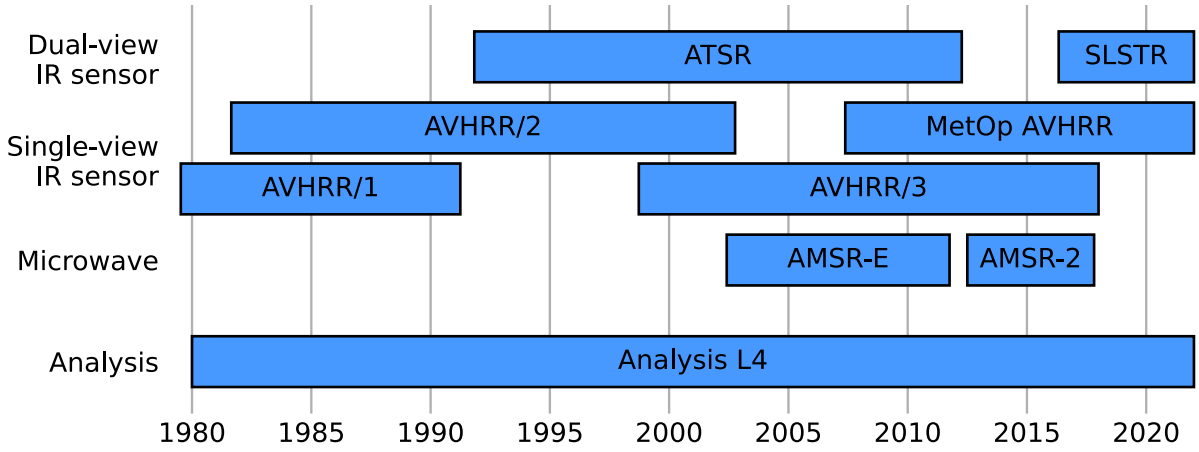


Figure 1: Overview of products included in CDR3.0

Sensor	Orbit	Level 2	Level 3	Notes
NOAA AVHRR	Drifting AM/PM	4-20 km	Y	
MetOp AVHRR	9:30 am	1.1-5 km	Y	MTA drifting from 2017
AMSR	1:30 pm	~50 km		PMW resolution is lower than SST-CCI level 3 grid so only provided as level 2 products
ATSR	10:00 / 10:30 am	1 km	Y	ATSR products unchanged from CDR v2.1
SLSTR	10:00 am	1 km	Y	

Table 1: Summary of sensors used in CDRv3

The ATSR and SLSTR instruments are both well calibrated, dual-view radiometers designed to produce long-term consistent SST observations. Three ATSRs were flown on board ESA's European Remote Sensing (ERS) satellite and Envisat between 1991 and 2012; the first SLSTR instrument was carried onboard the Sentinel-3A satellite launched in 2016, and there are currently two SLSTRs in operation. All ATSR and SLSTR sensors have been in stable sun-synchronous orbits with near-constant equatorial crossing times (10:30 for the two ERS satellites, and 10:00 for Envisat and the Sentinel-3 platforms).

The AVHRRs are a series of multipurpose imaging instruments carried onboard the National Oceanic and Atmospheric Administration (NOAA) Polar Operational Environmental Satellites (POES) and EUMETSAT Polar System (EPS) MetOp satellites. The first AVHRR instrument was carried onboard the TIROS-N satellite launched in October 1978 and the final onboard MetOp-C in November 2018. The equator crossing times of the various satellites are shown in Figure 3. The NOAA satellites are all in drifting orbits, meaning that the equator crossing times are slowly changing. The EUMETSAT MetOp satellites are in controlled orbits with equator crossing times of 9:30. Global full resolution data are available for MetOp AVHRRs, whereas for NOAA AVHRRs, global data are available only at the reduced resolution referred to as "global area coverage" (GAC).

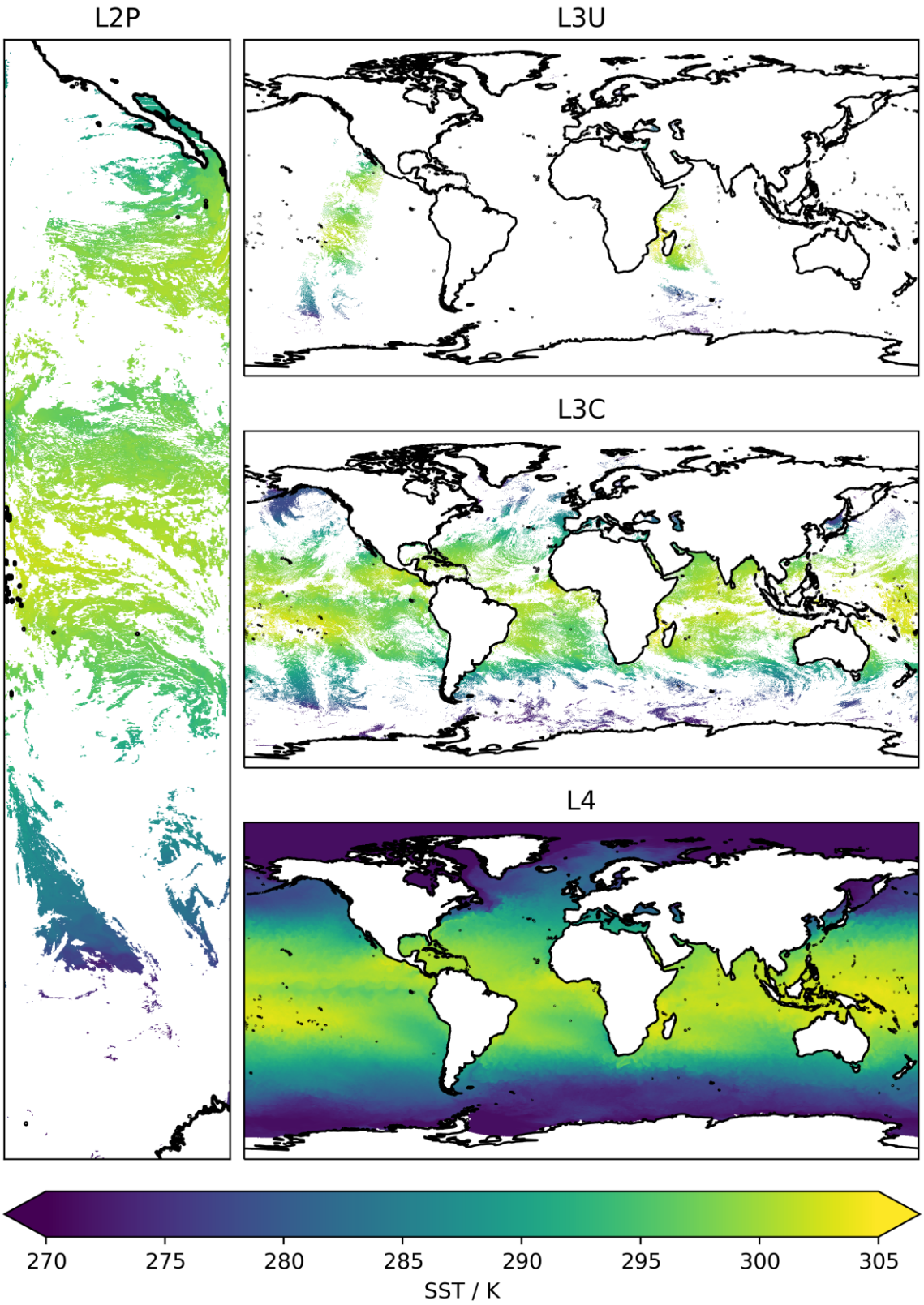


Figure 2: Illustration of how data are stored according to the "level" of the product. White areas correspond to locations with no SSTs. These occur, for example, due to cloud preventing the SST retrieval or because the location corresponds to land or ice.

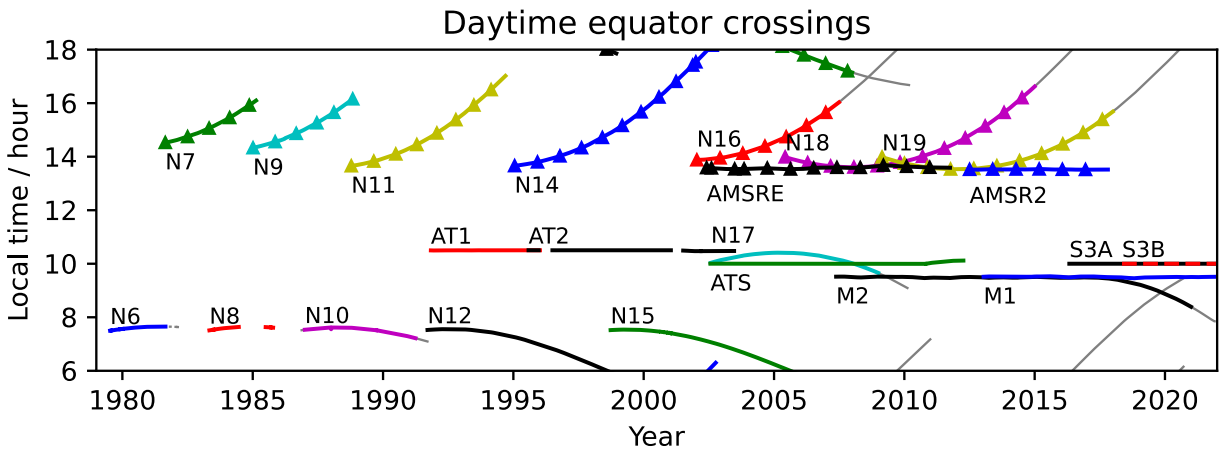


Figure 3: Satellite equator crossing times. Solid lines indicate descending node crossings; lines with triangles indicate ascending node crossings; thin grey lines indicate data were not used in CDR v3.

3. PRODUCT VALIDATION

3.1 Reference Dataset

The SST CCI Independent Reference Data Set (SIRDS) comprises in situ SST observations extracted from the Met Office Hadley Centre Integrated Ocean Dataset (HadIOD) v1.2.0.0 (Atkinson et al. 2014). The dataset was originally created during CCI Phase 2 with later updates (2017 onwards) for the European Union Copernicus Climate Change Service (C3S, https://climate.copernicus.eu/, C3S_312a_Lot_3). See Atkinson et al. 2014 for details on HadIOD (which includes both SST and salinity observations from 1900 onwards), Rayner et al. (2006) and Atkinson et al. (2013) for details of the additional quality control checks. Data are included for the following platform types:

1. Drifting buoys
2. Global Tropical Moored Buoy Array (GTMBA)
3. Moored buoys (excluding GTMBA)
4. Voluntary observing ships
5. Argo floats
6. Animals
7. Bottles
8. Conductivity-Temperature-Depth casts (CTDs)
9. Mechanical BathyThermographs (MBTs)
10. eXpendable BathyThermographs (XBTs)

The various data types have a range of characteristics with different instrument types, measurement uncertainties, sampling frequency, spatial location, and length of record which are discussed in the following sections. Figure 4 shows the temporal evolution of the datasets. In the present report, the primary in situ types used are: drifter for global validation since mid-1990s; all non-ship observations for global validation in 1980s and early 1990s; ship-based observations for regional effects in the 1980s; and GTMBA for estimating the precision in tropical areas (particularly for the dual-view ATSR and SLSTRs).

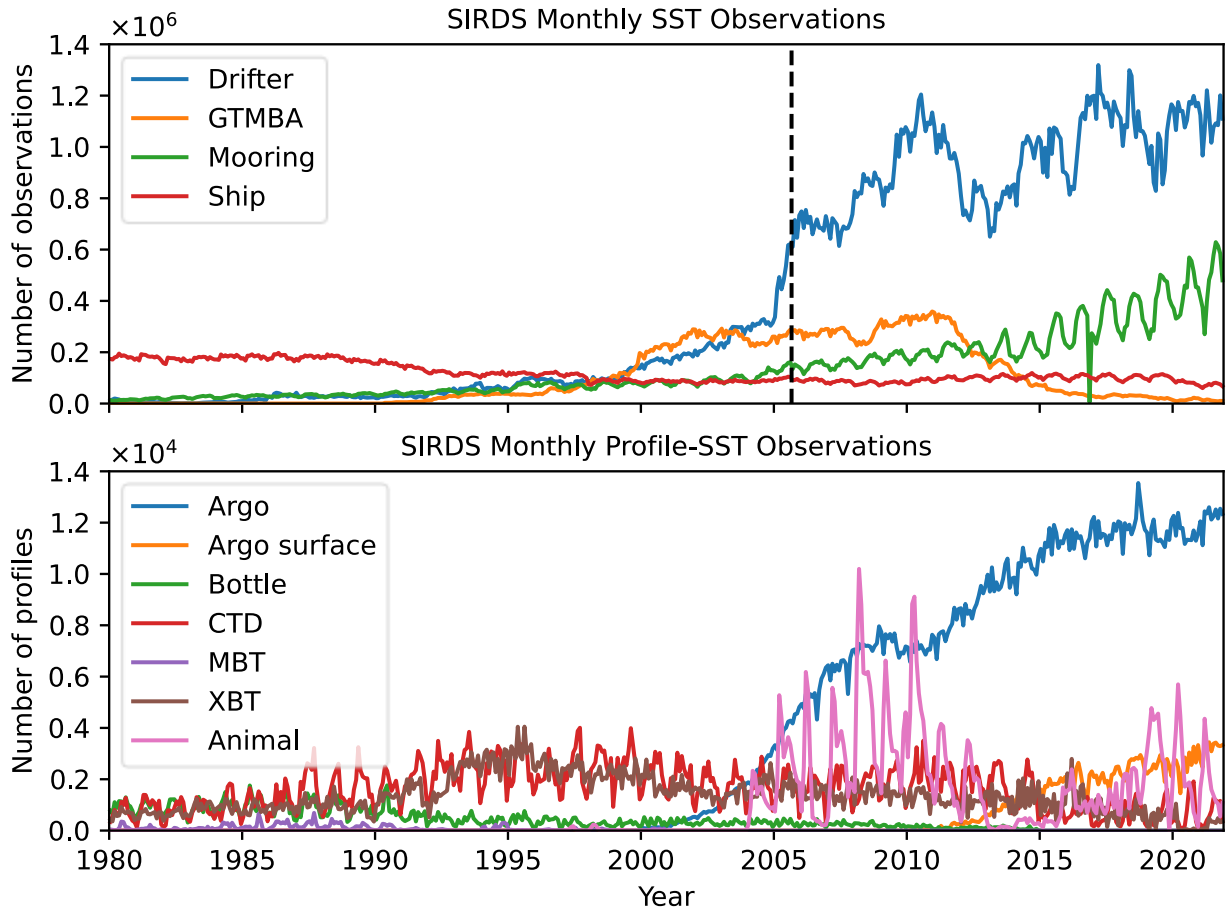


Figure 4: Monthly counts of in situ SST observations in reference dataset. Upper panel shows surface-based platforms; lower panel shows sub-surface profiling platforms. The vertical dashed line shows when the drifter network reached "full" coverage in September 2005.

3.1.1 Drifters

Drifting buoys comprise a surface float, of approximately 30 cm diameter, tethered to a sub-surface drogue (or sea-anchor) which ensures the drifter will closely follow the ocean currents (Lumpkin and Pazos, 2007). If the drogue is lost, then the drifter movement will be more affected by wind and waves and more prone to air-exposure of its thermistor. All drifters are equipped with an SST sensor which sits approximately 15 cm below the ocean surface in nominal conditions, and with equipment to transmit the collected data via satellite.

Initially the design of drifters was highly variable but work towards a standardized design began in 1982 as part of the World Climate Research Program. The first modern Surface Velocity Program (SVP) drifters were deployed in 1993 and the global drifter program array grew rapidly through the 1990s and was completed in September 2005 (observations in each 5 x 5 degree grid of open ocean) as shown in Figure 4 and the Hovmöller distribution in Figure 5. Drifters now provide the most complete in situ coverage of the world; however, this coverage is not uniform as the drifters follow ocean currents and collect in ocean convergence zones.

Drifting buoy observations were taken from the International Comprehensive Ocean-Atmosphere Dataset (ICOADS) dataset (Woodruff et al., 2011) and from CMEMS (marine.copernicus.eu). The number of drifter observations present in the ICOADS dataset starts to decrease from mid-2016 onwards due to a change in the way the buoy IDs were encoded, therefore the CMEMS drifters are used for 2016 onwards.

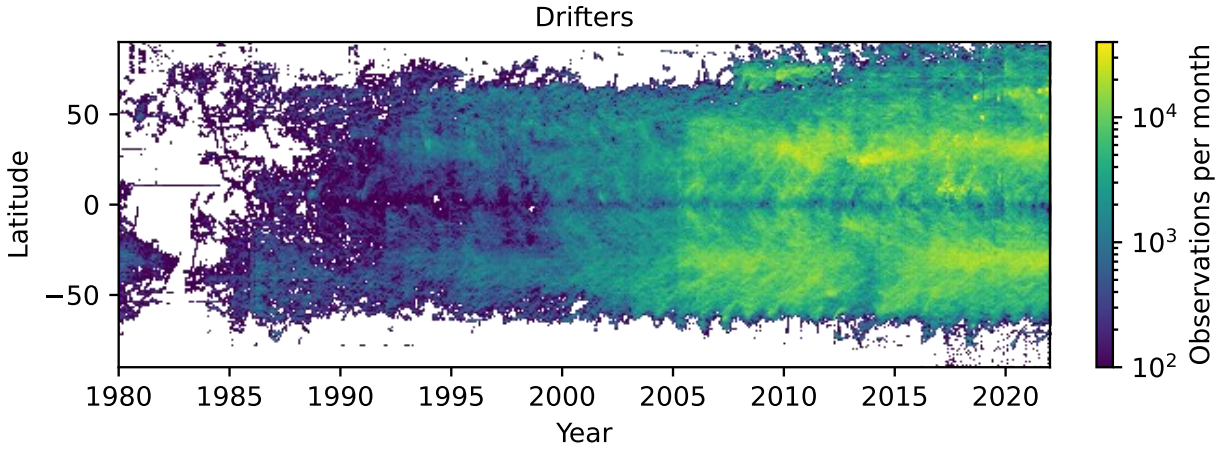


Figure 5: Hovmöller distribution of drifter observations. Data to end-2016 are from ICOADS, while data for 2016 onwards are taken from CMEMS.

3.1.2 GTMBA

The GTMBA includes tropical buoys managed by three different programmes: the Tropical Atmosphere Ocean/Triangle Trans Ocean Buoy Network (TAO/TRITON) in the Pacific, the Prediction and Research Moored Array (PIRATA) in the Atlantic, and the Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction (RAMA) in the Indian Ocean. All three components of the GTMBA use Autonomous Temperature Line Acquisition System (ATLAS) moorings which measure a wide range of meteorological and sub surface parameters, including SST at a depth of 1 m. The instruments measure SST at a resolution of 0.001 K and accuracy of 0.02 K every 10 minutes. The buoys transmit daily average and hourly average SSTs while a communication satellite is overhead (hence hourly measurements are not transmitted for all 24 hours in the day). However, the full resolution data is recorded on the buoys and recovered when the buoys are serviced. The temporal evolution of the GTMBA is shown in Figure 4 and the locations in Figure 6.

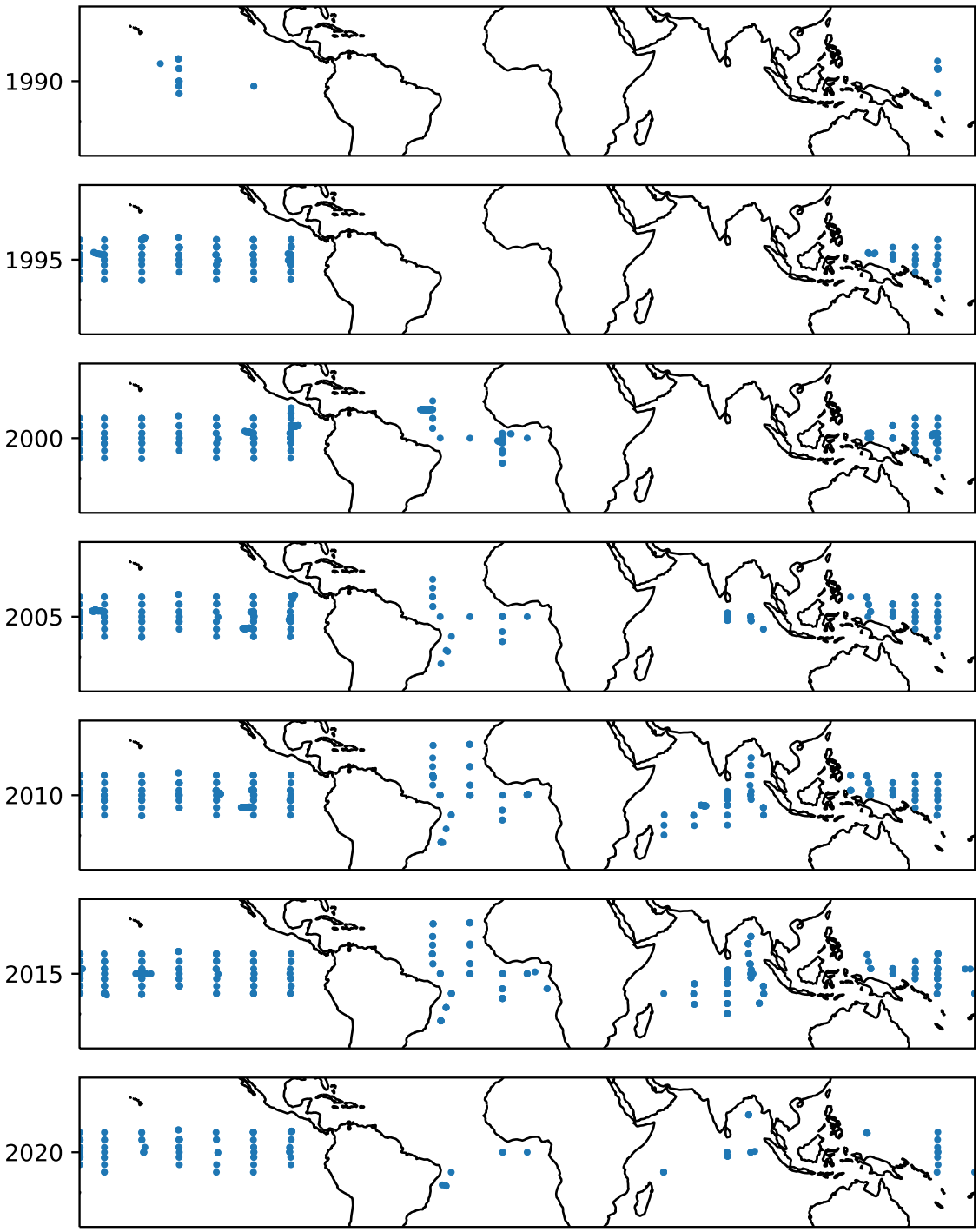


Figure 6: Location of active GTMBA stations from 1990 to 2020.

3.1.3 Voluntary Observing Ships

Ship-based measurements of SST provide the longest running record of SST available with usable observations dating back to the 1850s when the then sea-faring nations agreed on standardised measurement techniques at the Brussels Maritime Conference of 1853 (Rayner et al. 2006; Maury, 1858, 1859). Today the Voluntary Observing Ship (VOS) scheme is in

international programme run by the WMO where national weather services recruit ships to take and record weather observations. The size of the VOS fleet peaked in the mid-1980s with over 7500 active ships, the number of participating ships has declined since then; however, ships still provided the majority of in situ SST observations into the mid-1990s as seen in Figure 4. Due to their nature VOS data are concentrated along the major shipping lanes which will change over the years.

Various methods of measuring SST from ships have been employed over the years. Historically water was collected with a bucket to be measured with a thermometer; while modern ships may report the temperature of engine intake water used to cool the engine or use a dedicated hull contact sensor.

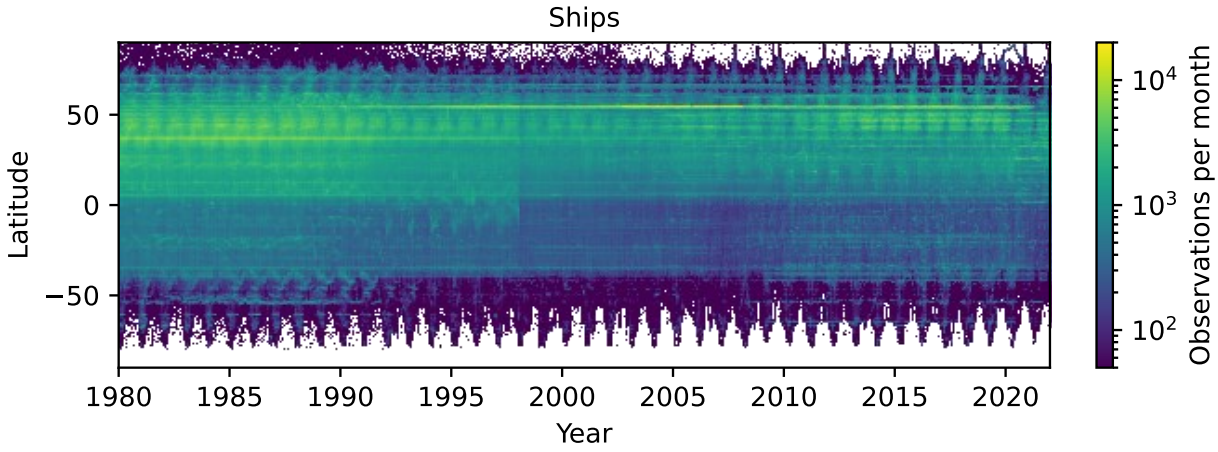


Figure 7: Hovmöller distribution of ship-based observations.

3.2 Validation Methodology

3.2.1 Matchup strategy

A matchup dataset (MD) of coincident satellite and reference in situ observations is needed in order to validate the CDR. Due to the changing data density and resolution we use two different matchup strategies: a simple “direct” matching for recent sensors, and a more complex approach using the Multi-sensor Matchup System (MMS; Block et al. 2018). In recent years we can obtain sufficient numbers of matches with a simple “direct” matching between in situ and Level 3 (or 4) satellite data. In this scheme the matchup algorithm is:

- For each Level 3 product find all in situ within +/- 2 hours of product extent
- For each unique in situ platform:
- For each in situ observation extract corresponding Level 3 cell
    - Reject any pairings where level 3 cell is missing data
    - Reject any pairing where time difference is > 2 hours
    - Use pairing with smallest time difference as the matchup

This approach is appropriate for the gridded level 3 and 4 products and provides a direct comparison of the satellite and in situ measurements. However, it is only effective with recent high-density data. The direct matching of in situ to level 3 cell is equivalent to a spatial criterion of 0.025° separation (i.e. half the grid cell size; approximately 2.5 km at the equator). While this works well for modern sensors it is too strict for the low-resolution NOAA AVHRR (GAC) which has coarser sampling than the level 3 grid except close to satellite nadir view. Furthermore, prior to the increased in situ data density of recent decades we need to allow a much larger time window in order to obtain sufficient matches which means we must account for diurnal changes in the SST between the satellite and in situ observations (see section 3.3).

When the direct matching of in situ to level 3 approach is insufficient, we use the MMS (Block et al. 2018) which performs a complete matchup search between the in situ and Level 1 satellite radiances allowing for the variable pixel size across swath. Once the pixel-level match has been found, the corresponding Level 3 SST is also added, and this value is used here for validation for consistency. In addition, the MMS will also add collocated Numerical Weather Prediction (NWP) fields which can be used to calculate the relevant diurnal adjustment for larger temporal separations.

Sensor	Matchup Method	Spatial Criterion	Temporal Criterion	Primary Reference
NOAA-06 to -12	MMS	12 km	12 hours	All non-ship observations
NOAA-14 to -19	MMS	12 km	4 hours	Drifters
ATSR-1	MMS	1 km	4 hours	All non-ship observations
ATSR-2 / AATSR	MMS	1 km	4 hours	Drifters
MetOp AVHRR	Direct	0.025°	2 hours	Drifters
SLSTR	Direct	0.025°	2 hours	Drifters
Level 4 (to 1996)[1]	Direct	0.025°	12 hours	CTD, Drifter, GTMBA, Moorings, XBT[2]
Level 4 (1996+)	Direct	0.025°	12 hours	Drifters

Table 2: Matchup criteria for different sensors. (1) Used for full record for coastal comparison in section 4.3.3. (2) Initially pre-1996 Level 4 was compared against all non-ship observations; however, bottle and MBTs showed unexplained outliers and were removed (see section 4.3)

3.2.2 Validation of SST

The differences between the satellite SSTs and reference data were analysed using robust statistics, which means statistics that are less influenced by outliers in the distribution of differences. Outliers arise in both satellite and validation data, and robust statistics better describe the majority of data that are more normally distributed. The measures used here are the median and the robust standard deviation (RSD) given by 1.4826 times the median absolute deviation. For a normal distribution the RSD and standard deviation are equal.

Satellite and reference datasets are compared using global statistics, time-series, spatial maps, and Hovmöller plots.

3.2.3 Validation of uncertainties

The approach used was to compare robust standard deviation of differences between the analyses and the reference data to the combination of the in situ data uncertainty and the uncertainties provided with the SST products. Statistics were generated for different levels of uncertainty ascribed to the analyses, in order to determine if the uncertainties were valid across the full range of possible uncertainties.

Figure 8 shows a typical uncertainty validation plot. Along the x-axis is the estimated uncertainty which is calculated as part of the retrieval (i.e. it is estimated without the use of any in situ data). The y-axis represents the discrepancy between the satellite retrieval and in situ measurement, which will be affected by errors in both the satellite and in situ observations. Given an estimate of the uncertainty in the in situ data (in this example we assume 0.2 K for drifter observations), we can estimate the expected spread in the satellite to in situ comparison as $\sigma = \sqrt{\sigma_{ins} + \sigma_{sat}}$ which is shown in the plot with solid blue lines.

The shaded grey area shows the RSD difference as a function of the estimated uncertainty (x-axis). If the uncertainties have been correctly estimated, then the shaded area should match the expected RSD envelope shown in the solid blue lines. In the example shown in Figure 8 (left panel) the uncertainties are over estimated (grey area fall short of the envelope), while in Figure 8 (right panel) the uncertainties are well estimated (grey area matches the envelope). Additionally, we show the bias as a function of the estimated uncertainty with the orange error bars, and the distribution of estimated uncertainties in the green violin plot. In the example plot we see that daytime retrievals have expected uncertainty 0.15 K and 0.95 K, though the majority of data are found between 0.25 and 0.45; while the night-time data range between 0.15 K and 0.65 K, with the majority under 0.25 K.

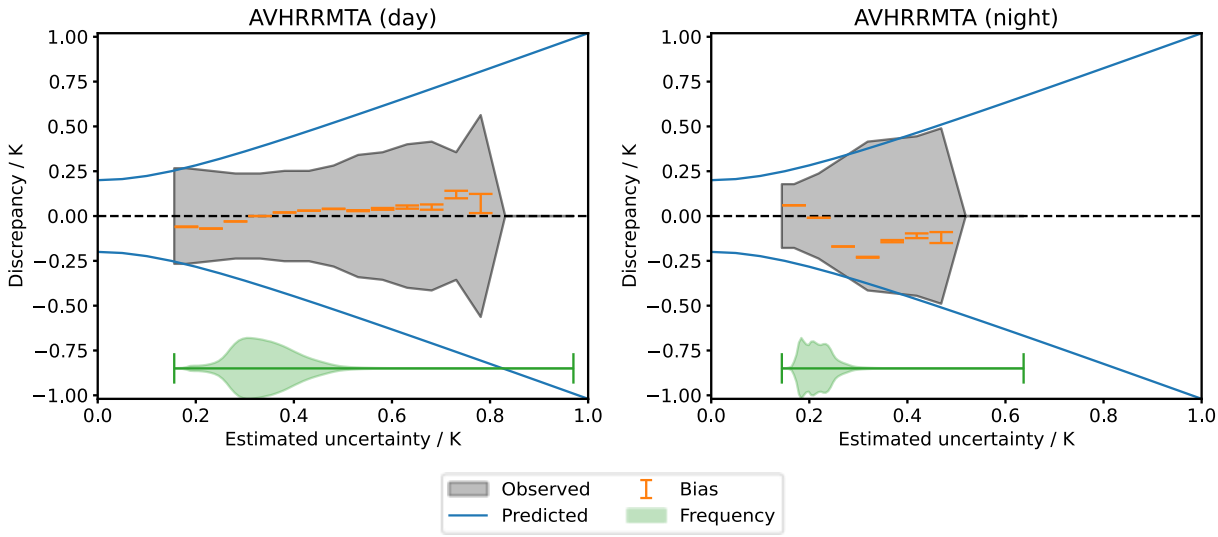


Figure 8: Example of uncertainty validation plot against in situ drifters for daytime (left) and nighttime (right) data. Shaded grey area shows RSD discrepancy as a function of estimated uncertainty Solid blue line shows expected relationship based on assumed in situ uncertainty of 0.2 K. Orange error bars show median discrepancy in each bin. Green violin plot shows distribution of data.

3.3 Diurnal corrections

The near-surface ocean goes through a diurnal cycle as it is heated by the sun during the day and cools at night. Typically, this results in a diurnal cycle in SST of a few tenths of kelvin, but in cases of sustained low wind speed and high incident solar radiation can exceed 5 K (Gentemann et al. 2008). Figure 9 shows a climatological diurnal cycle for SST under clear-sky conditions between 40N and 50N derived from drifting buoy measurements by Morak-Bozzo et al. (2016). Diurnal warming is higher when there is more input solar heating (hence the larger magnitude in summer) and that heat is trapped in the near-surface layer (higher wind speeds increase mixing so the temperature cycle is lower). The diurnal SST anomaly is close to zero around 1030 h or 2230 h which is why the SST-CCI depth SSTs are standardized to a 10:30 am/pm local time.

Differences in the observation time between satellite and in situ measurements can therefore affect the validation results depending on the size of the temporal separation and satellite overpass time. Figure 10 shows the distribution of satellite – in situ observation time for NOAA-14 to drifters for both the SST_{skin} and $SST_{0.2m}$ estimates. During the late 1990s, when the NOAA-14 satellite was operational, in situ drifters were transmitting SST observations every two hours when a receiving Global Telecommunication System (GTS) satellite was in range. As the NOAA satellites are part of the GTS we get very close matches between SST_{skin} and drifters with most less than 30 minutes apart (Figure 10 left). However, there is reduced in situ coverage around 10:30 local time, so for $SST_{0.2m}$ we get three much broader peaks centred on -2, 0, and +2 hours separation (Figure 10 right). In the modern era all drifters are reporting hourly or better throughout the whole day as can be seen in Figure 11 which shows drifter matches to 10:30 local time for MetOp-A AVHRR.

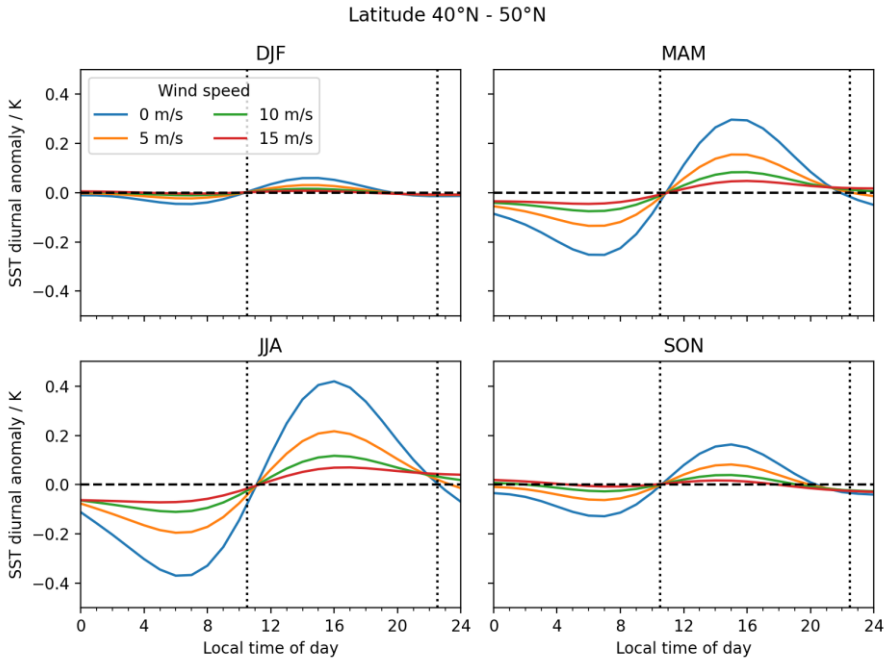


Figure 9: Diurnal cycle in daily SST anomaly for 40°N to 50°N from Morak-Bozzo et al. 2016. Panels show northern hemisphere winter (DJF), spring (MAM), summer (JJA), and autumn (SON). Vertical dotted lines show standard 10:30 local time used in SST CCI products.

In order to minimise these uncertainties in the historic matches (NOAA AVHRR and ATSR), the satellite – in situ differences are adjusted for using a skin effect (Fairall et al., 1996) and warm layer model (Kantha and Clayson, 1994) driven by the NWP fluxes in the MMS-based matches. This is the same model as used in the processing chain to calculate the satellite $SST_{0.2m}$ @ 10:30 estimate [see ATBD]. For the newer MetOp AVHRR and SLSTR sensors the temporal separation between satellite SST and drifter is typically less than 30 minutes so the diurnal correction is not required.

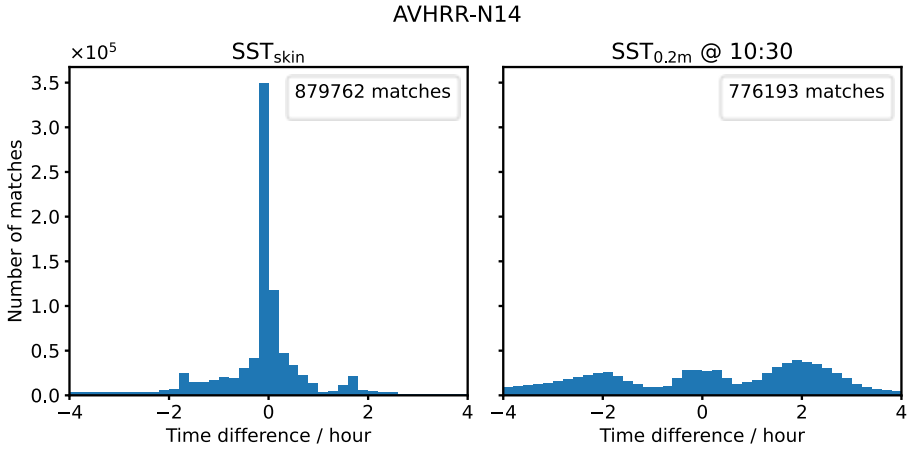


Figure 10: Distribution of temporal separation for drifter matches to NOAA-14 AVHRR. Left: matching to observation at satellite overpass. Right: matching to SST at 10:30 local time.

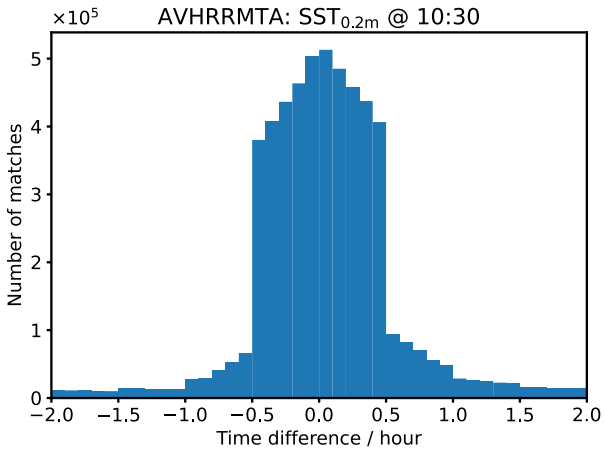


Figure 11: Distribution of temporal separation for drifter matches to MetOp-A AVHRR.

4. SUMMARY OF VALIDATION RESULTS

4.1 Validation of L3 SST

Global validation statistics comparing the Level 3 SST product against reference in situ are shown in Table 3. The reference in situ used are drifters for all sensors, supplemented by other non-ship-based observations for NOAA AVHRR 06–12 and ATSR-1. All sensors are performing within target with global biases < 0.1 K. The dual-view ATSR and SLSTR sensors have the most consistently low biases which are generally only a couple of cK, while several AVHRR sensors show biases 0.05 K or greater.

For sensors in a late morning orbit (ATSRs and NOAA-17) the SST_{skin} and SST_{20cm} statistics are virtually identical. This is expected as the diurnal correction will be minimal in these cases and indicates that the same will be true for the directly matched MetOp and SLSTR sensors where only SST_{20cm} validation was performed. However, for the afternoon sensors there is a slight divergence with the time and depth-adjusted SST_{20cm} being 0.03 – 0.05 K warmer at day and 0.02 – 0.03 K colder at night relative to in situ. This is most obvious for the three most recent afternoon AVHRRs (NOAA-16, -18, and -19) which benefit from both being newer, better calibrated AVHRRs, and from better in situ coverage. This may suggest scope to improve the diurnal variability adjustment model, although the degree of divergence is within expected accuracy.

Tropical validation statistics comparing the Level 3 data against GTMBA are shown in Table 4 beginning with AVHRR-11 at the early stage of TAO deployment. Here the discrepancy RSDs are lower than the corresponding drifter cases – particularly for night-time ATSR and SLSTR cases – which is due to the in situ uncertainty contributing to the reported RSDs. There is more variation in the AVHRR median biases as the GTMBA results are for a limited region (see Figure 6) rather than global statistics. However, with the exception of AVHRR-12 night-time all other biases other are less than 0.1 K.

	SST skin				SST 0.2m @ 10:30			
	Day		Night		Day		Night	
	Median	RSD	Median	RSD	Median	RSD	Median	RSD
AVHHR-06			-0.00	0.56			+0.02	0.55
AVHRR-07	-0.00	0.52	+0.09	0.52	+0.00	0.53	+0.07	0.53
AVHRR-08			+0.02	0.55			+0.02	0.57
AVHRR-09	+0.00	0.48	+0.04	0.51	+0.02	0.49	+0.02	0.51
AVHRR-10			-0.05	0.51			-0.04	0.52
AVHRR-11	+0.05	0.42	+0.08	0.40	+0.07	0.43	+0.05	0.41
AVHRR-12	+0.04	0.39	-0.00	0.41	+0.02	0.40	-0.00	0.41
AVHRR-14	+0.02	0.36	+0.05	0.38	+0.04	0.37	+0.02	0.38
AVHRR-15	+0.05	0.32	+0.04	0.33	+0.03	0.32	+0.03	0.34
AVHRR-16	-0.01	0.29	-0.00	0.28	+0.05	0.30	-0.03	0.29
AVHRR-17	+0.07	0.25	+0.05	0.26	+0.07	0.25	+0.06	0.26
AVHRR-18	-0.02	0.27	+0.01	0.26	+0.03	0.28	-0.02	0.27
AVHRR-19	+0.00	0.27	-0.00	0.25	+0.05	0.28	-0.03	0.25
AVHRRMTA					-0.01	0.25	-0.01	0.24
AVHRRMTB					+0.01	0.25	+0.02	0.24
ATSR-1	+0.04	0.45	-0.00	0.45	+0.04	0.45	+0.01	0.45
ATSR-1 (d3)			-0.01	0.26			+0.00	0.26
ATSR-2	-0.00	0.28	+0.01	0.21	-0.00	0.28	+0.02	0.21
AATSR	+0.01	0.21	-0.00	0.18	0.01	0.21	+0.01	0.18
SLSTR-A					+0.02	0.25	+0.00	0.19
SLSTR-B					-0.03	0.24	-0.01	0.19

Table 3: Summary of validation against reference in situ (recent sensors use drifters, with other non-ship data added for sensors up to NOAA-12 and ATSR-1; see Table 2 for details). SST skin results compare the satellite retrieval to in situ adjusted for the skin effect and to the satellite observation time. SST 0.2m results compare the SST depth estimate to in situ adjusted to the 10:30 am/pm local time. NOTE – MetOp/SLSTR comparisons do not include additional DV adjustments as the in situ observations are within 30 minutes of the standardised times.

	SST skin				SST 0.2m @ 10:30			
	Day		Night		Day		Night	
	Median	RSD	Median	RSD	Median	RSD	Median	RSD
AVHRR-11	+0.00	0.32	+0.01	0.34	+0.05	0.33	-0.01	0.35
AVHRR-12	+0.03	0.29	-0.14	0.36	+0.04	0.29	-0.12	0.36
AVHRR-14	-0.04	0.31	-0.04	0.32	-0.01	0.32	-0.07	0.32
AVHRR-15	+0.04	0.26	-0.01	0.31	+0.01	0.26	-0.01	0.31
AVHRR-16	-0.01	0.26	-0.01	0.26	+0.03	0.26	-0.04	0.26
AVHRR-17	+0.07	0.21	+0.04	0.23	+0.07	0.21	+0.05	0.23
AVHRR-18	-0.02	0.25	-0.02	0.24	+0.02	0.25	-0.04	0.25
AVHRR-19	+0.02	0.25	-0.03	0.24	+0.06	0.25	-0.05	0.24
AVHRRMTA					+0.05	0.20	-0.01	0.21
AVHRRMTB					+0.06	0.21	+0.01	0.21
ATSR-1	+0.04	0.45	-0.04	0.44	+0.04	0.45	-0.03	0.44
ATSR-1 (d3)			-0.01	0.13			-0.00	0.14
ATSR-2	-0.01	0.22	-0.02	0.11	-0.01	0.22	-0.01	0.11
AATSR	+0.01	0.19	-0.02	0.13	-0.00	0.19	-0.01	0.13
SLSTR-A					-0.04	0.24	-0.03	0.15
SLSTR-B					-0.09	0.24	-0.03	0.16

Table 4: Summary of validation against GTMBA reference. NOTE – MetOp/SLSTR do not include any additional DV adjustments.

A timeseries of the median and RSD of $SST_{0.2m}$ – in situ discrepancy for each sensor series is shown in Figure 12. There are several regimes noticeable in the AVHRR data. First, from 1980 to late 1991 the AVHRR dataset is based on an AVHRR/1 in the morning orbit and an AVHRR/2 instrument in the afternoon orbit. Data in this period are relatively noisy (higher RSD) and have fluctuations in the relative bias of a few tenths of kelvin; however, part of what is observed is likely due to the limited in situ coverage in this period as there are few in situ measurements in this period (Figure 4) and drifters have very limited geographical coverage (Figure 5). For the next period from late 1991 to end-2000 we have an AVHRR/2 instrument (NOAA-12) and then an AVHRR/3 (NOAA-15) in the morning orbit along with a large increase in the number of in situ drifters. Data in this period show a noticeable improvement over the first decade with lower noise and smaller fluctuations in the relative bias. For the next period from 2001 to end-2009 there are up-to four AVHRR/3 instruments in operation at any one time, while the in situ drifter is completed in 2005 (see Section 3.1). Initially two of the AVHRRs (NOAA-15 and NOAA-17) are in the morning orbit, but by 2006 the NOAA-15 orbit has drifted to become an afternoon overpass (see Figure 3). During this period the RSDs decrease to ~0.25 K and the monthly fluctuations in relative bias largely disappear, while a seasonal cycle in the daytime SST becomes more apparent. Finally, from 2010 onwards the NOAA AVHRR is based on just afternoon sensors NOAA-18 and NOAA-19 (as the MetOp-AVHRRs are now available for the morning orbit). During this period there is a slight divergence (up to ~0.05 K) in the daytime and night-time SSTs, with daytime becoming slightly warm while the night-time cools relative to drifters. This appears to be an artefact of the diurnal adjustment from the satellite overpass time between 2 and 4pm local time to the

reference 10:30 local time and is not apparent in the equivalent SST_{skin} comparisons shown in Figure 13.

Figure 12 does not show any evidence of the eruptions of El Chichón (April 1982) and Mount Pinatubo (June 1991) in the global time series of relative bias; however, the Hovmöller distributions in Figure 14 (against reference in situ) and Figure 15 (against ship in situ) shows there is a brief regional signal with biases up to ~0.4 K in the night-time AVHRR SST between 10S and 10N from the Mount Pinatubo eruption. There is some sign of a weaker regional bias associated with El Chichón in the ship comparison (Figure 15), but it is no larger than other seasonal bias signals in the 1980s. For comparison, satellite SST products which do not account for stratospheric aerosol will likely see global SST biases over 0.5 K and regional biases over 2 K for AVHRR (Zhang et al. 2004, Blackmore et al. 2012) or 1.5 K for ATSR (Merchant et al. 1999).

Also shown in Figure 14 and Figure 15 are the spatial distribution of the mean $SST_{0.2m}$ @ 10:30 local time minus in situ difference for the AVHRR sensors. Here we see that the desert-dust related biases which affected the previous versions of the CDR have been largely eliminated. The remaining mission level plots are shown for MetOp AVHRR in Figure 16, ATSR in Figure 17, and SLSTR in Figure 18. The MetOp AVHRR results (Figure 16) are consistent with the recent NOAA AVHRRs, showing little seasonal dependence or desert-dust related biases. There is some year-to-year variability in the Arctic with 2012 through 2016 appearing cooler relative to drifting buoys than other years. There are two notable features in the ATSR results (Figure 17), firstly the failure of the ATSR-1 3.7 µm channel in May 1992 meant the night-time SST falls back to a two-channel retrieval. This combined with the increased operating detector temperature resulted in a increase in the instrument noise which can be seen as a band of noisier data between June 1992 and August 1995. Secondly there are two gaps in the ATSR-2 data record: Jan to Jun 1996 due to the scan mirror failure, and Jan to Jun 2001 due to the onboard gyro failure.

A full range of single-sensor plots are available in the appendix.

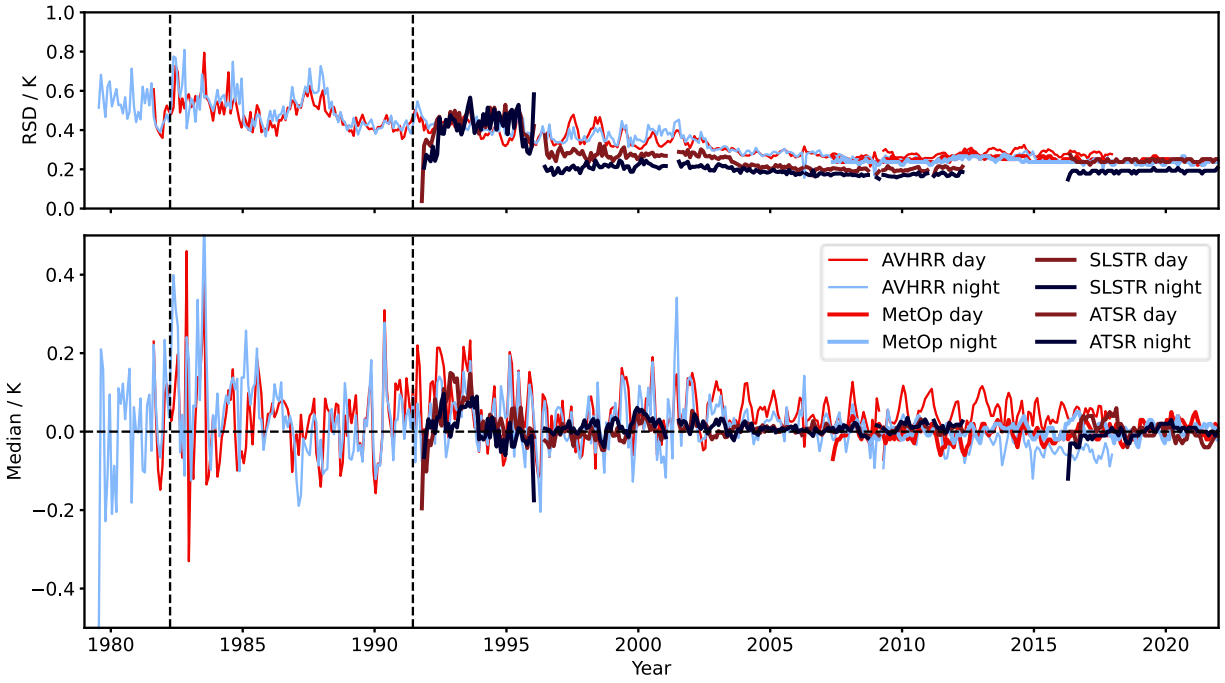


Figure 12: Timeseries of validation results. Monthly robust standard deviation (top panel) and median discrepancy (lower panel) for comparison of SST 0.2m @ 10:30 local time and reference in situ (see Table 2 for matchup criteria and in situ types used). Vertical dashed lines show time of El Chichón (April 1982) and Mount Pinatubo (June 1991) eruptions.

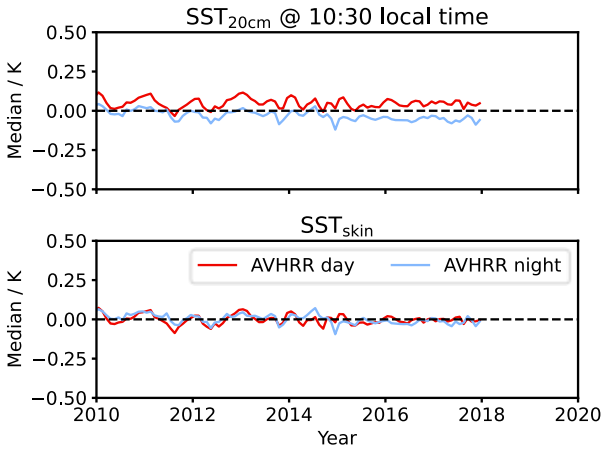


Figure 13: Timeseries of AVHRR validation results for SST 0.2m @ 10:30 local time (top panel) and SST skin (lower panel)

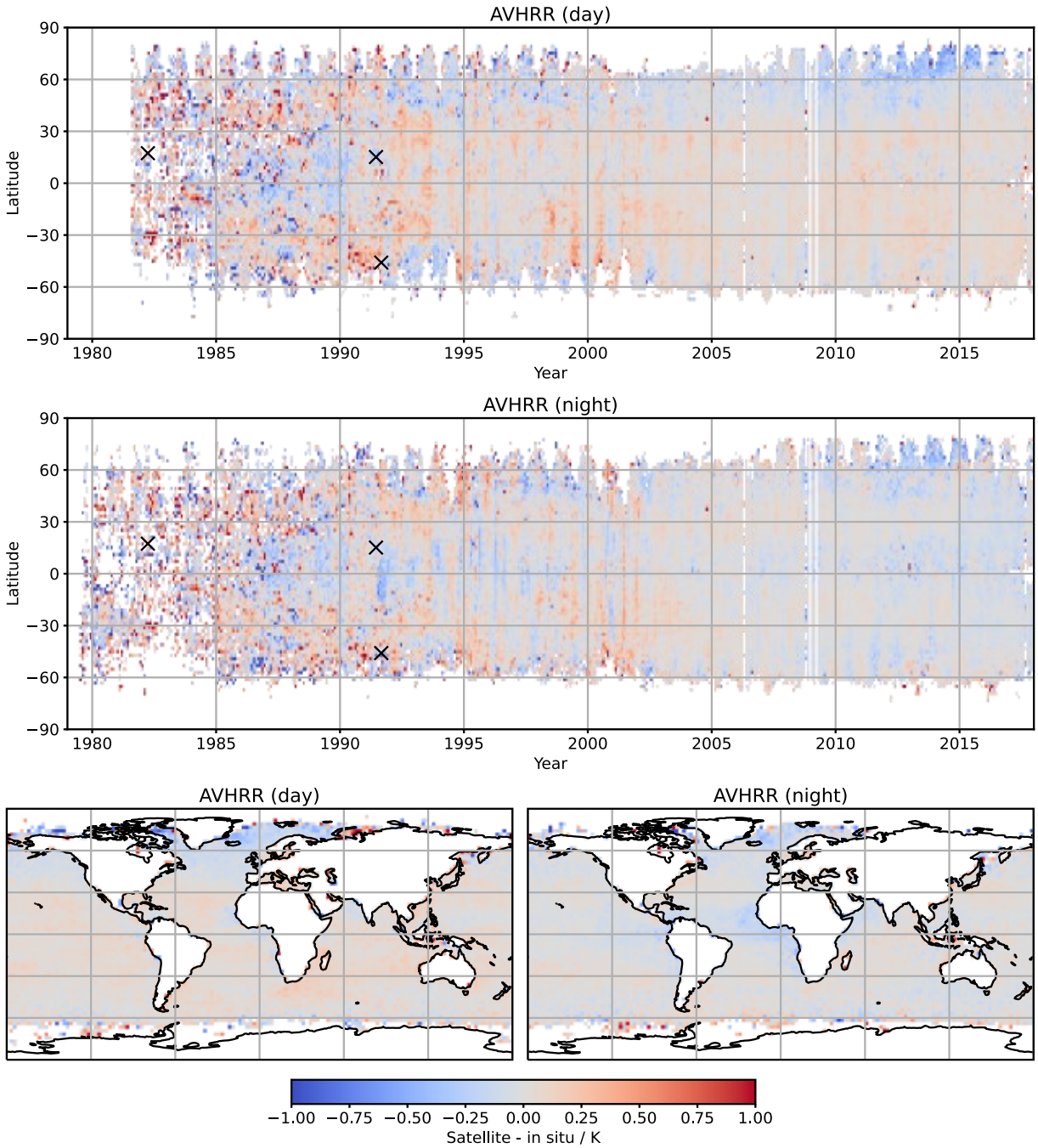


Figure 14: NOAA AVHRR $SST_{0.2m}$ @ 10:30 local time minus reference in situ difference (see Table 2). Top: day-time hovmöller distribution; centre; night-time hovmöller; lower-left: day-time spatial; lower-right: night-time spatial. X symbols mark major volcanic eruptions: El Chichón (April 1982), Mount Pinatubo (June 1991), and Mount Hudson (September 1991)

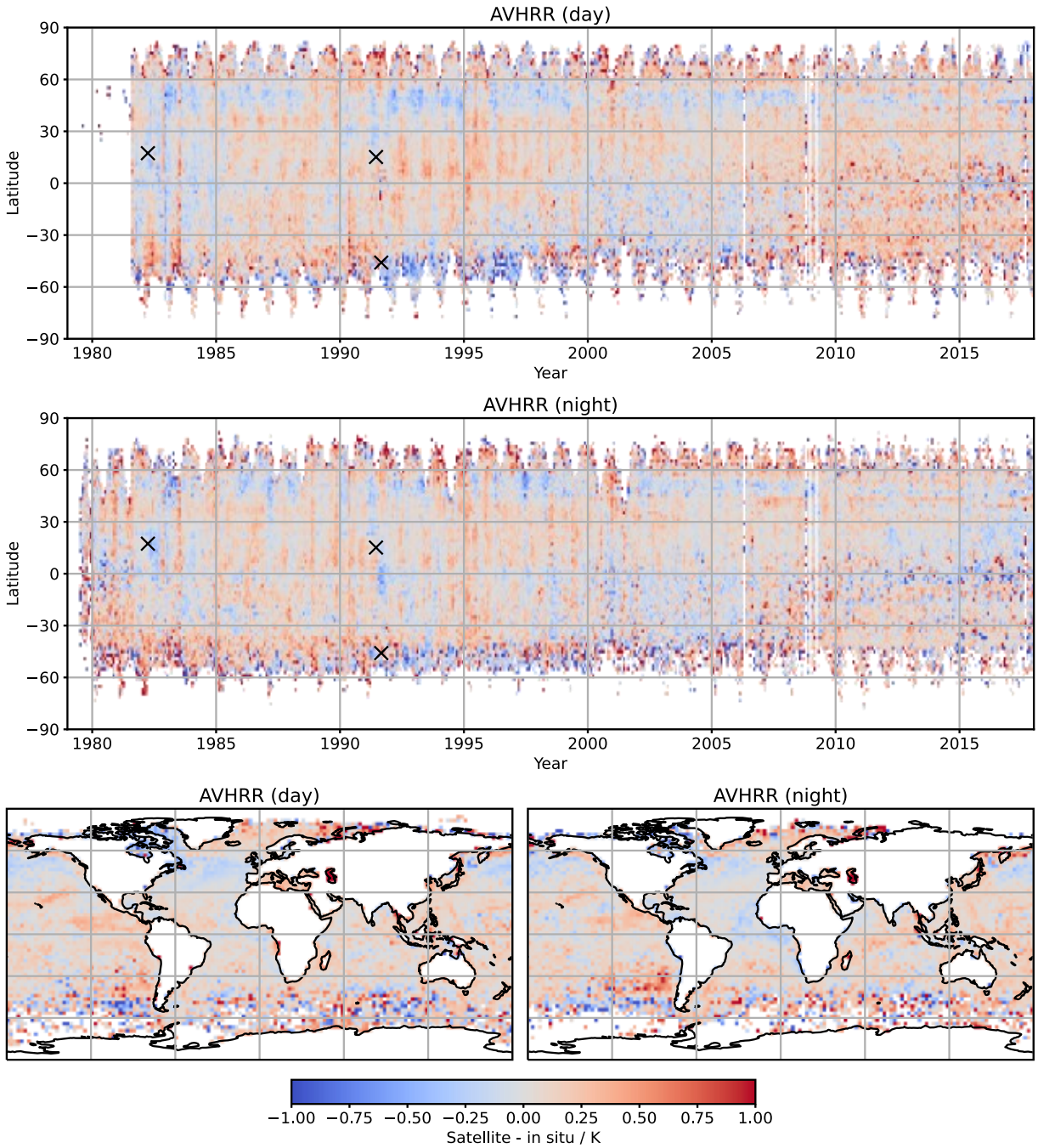


Figure 15: NOAA AVHRR $SST_{0.2m}$ @ 10:30 local time minus ship in situ difference. Top: day-time hovmöller distribution; centre; night-time hovmöller; lower-left: day-time spatial; lower-right: night-time spatial. X symbols mark major volcanic eruptions: El Chichón (April 1982), Mount Pinatubo (June 1991), and Mount Hudson (September 1991)

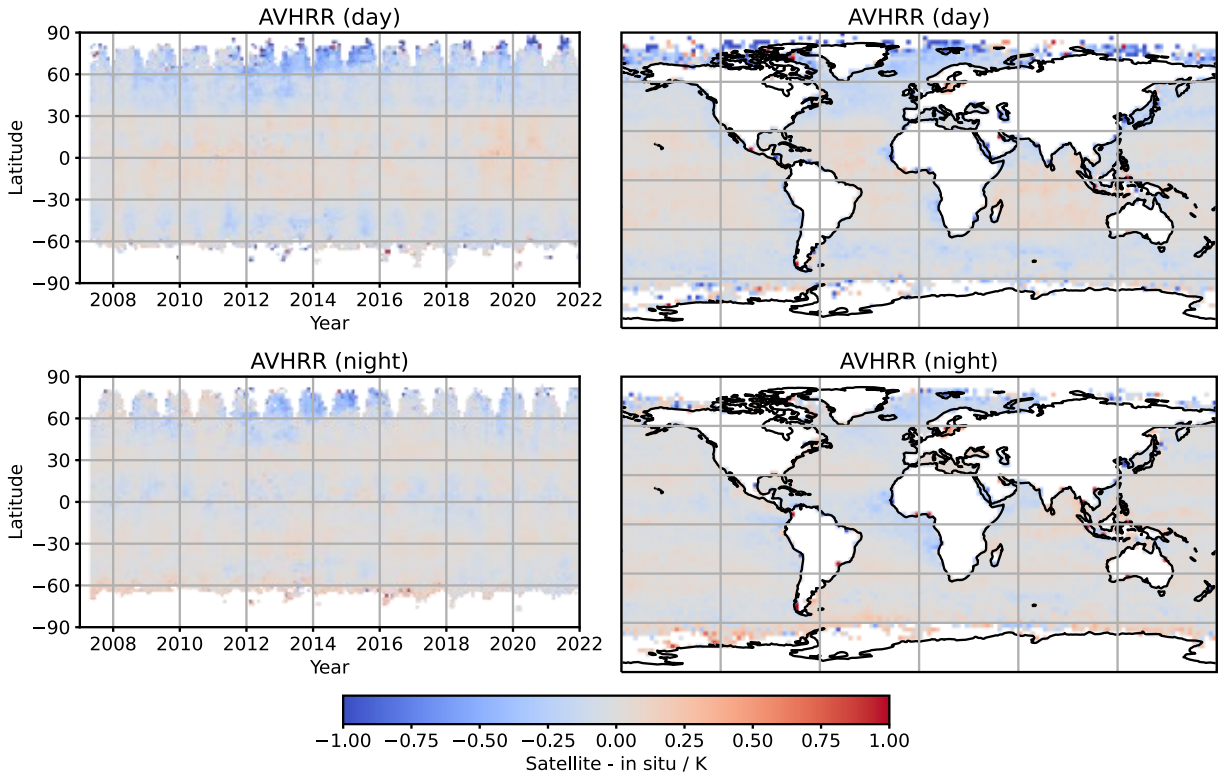


Figure 16: MetOp AVHRR $SST_{0.2m}$ @ 10:30 local time minus drifter in situ difference. Top-left: day-time hovmöller distribution; top-right; night-time hovmöller; lower-left: day-time spatial; lower-right: night-time spatial.

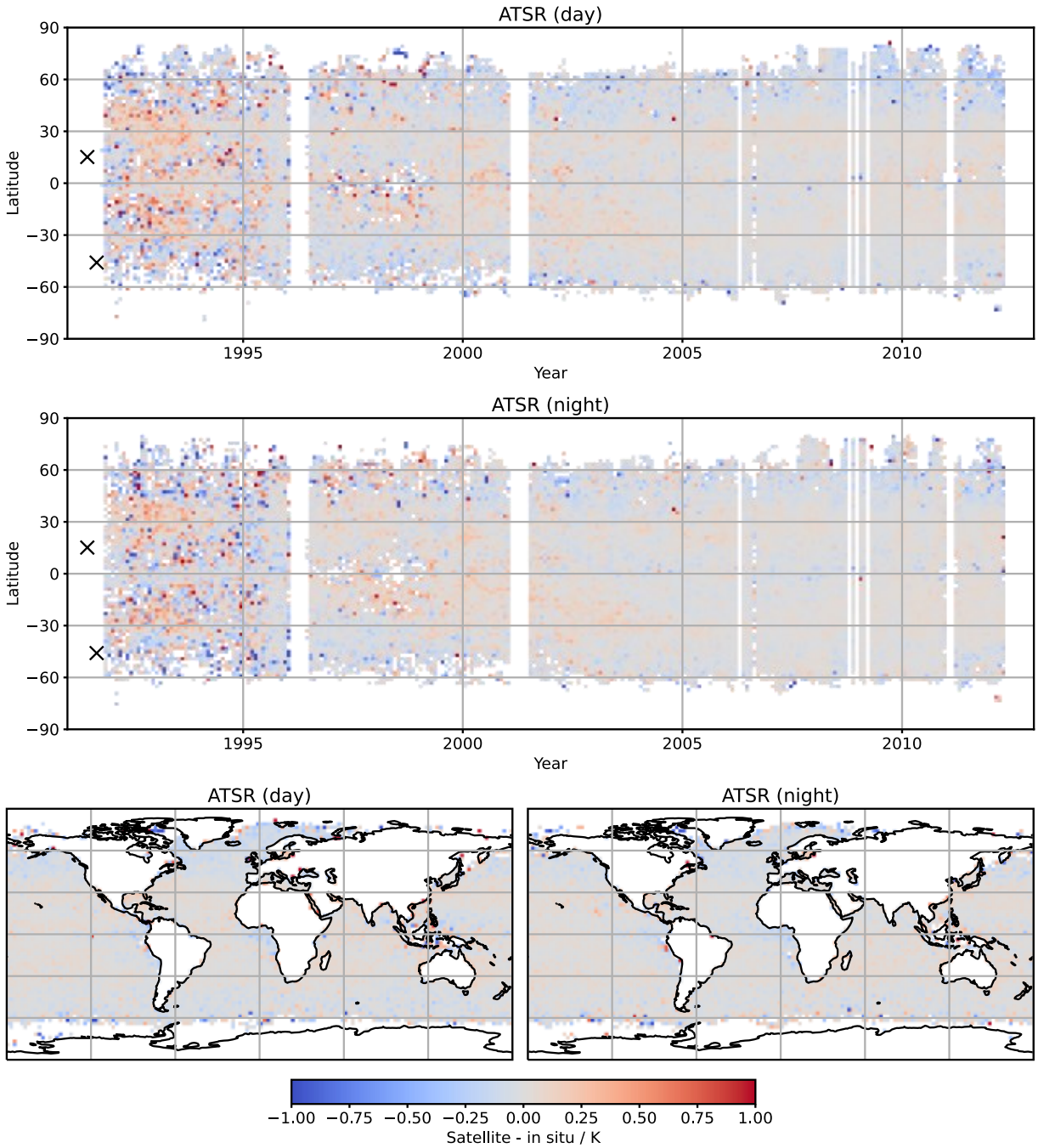


Figure 17: ATSR $SST_{0.2m}$ @ 10:30 local time minus reference in situ difference (see Table 2). Top: day-time hovmöller distribution; centre; night-time hovmöller; lower-left: day-time spatial; lower-right: night-time spatial. X symbols mark major volcanic eruptions: Mount Pinatubo (June 1991), and Mount Hudson (September 1991)

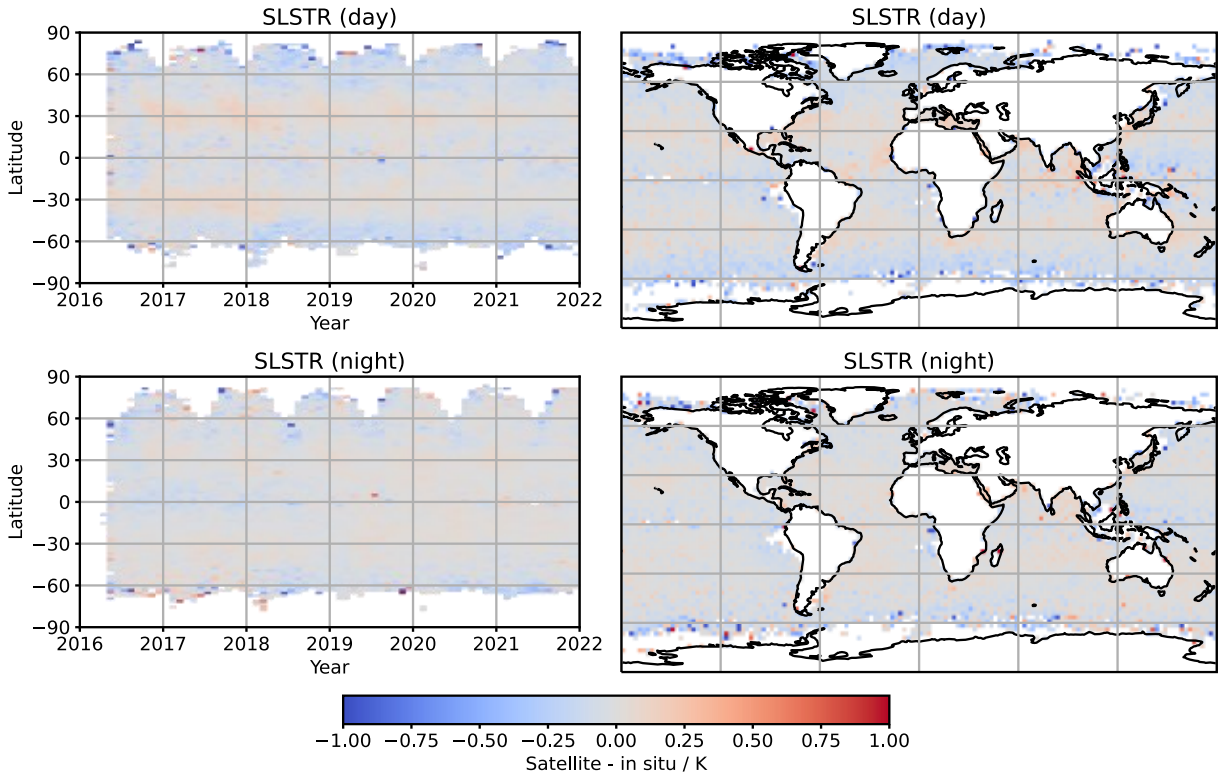


Figure 18: SLSTR $SST_{0.2m}$ @ 10:30 local time minus drifter in situ. Top: day-time hovmöller distribution; centre; night-time hovmöller; lower-left: day-time spatial; lower-right: night-time spatial.

4.2 Validation of L3 uncertainties

Uncertainty validation plots are shown for a selection of L3 products in Figure 19. Other sensors show the same behaviour, with dual-view ATSR/SLSTR uncertainties well estimated (the estimated uncertainties provide a good indication of the observed satellite – in situ RSD differences. For the single-view AVHRR sensors the night-time uncertainties are well estimated; however, the daytime uncertainties all tend to be over-estimated – i.e. actual satellite – in situ differences are smaller than predicted by the uncertainty field.

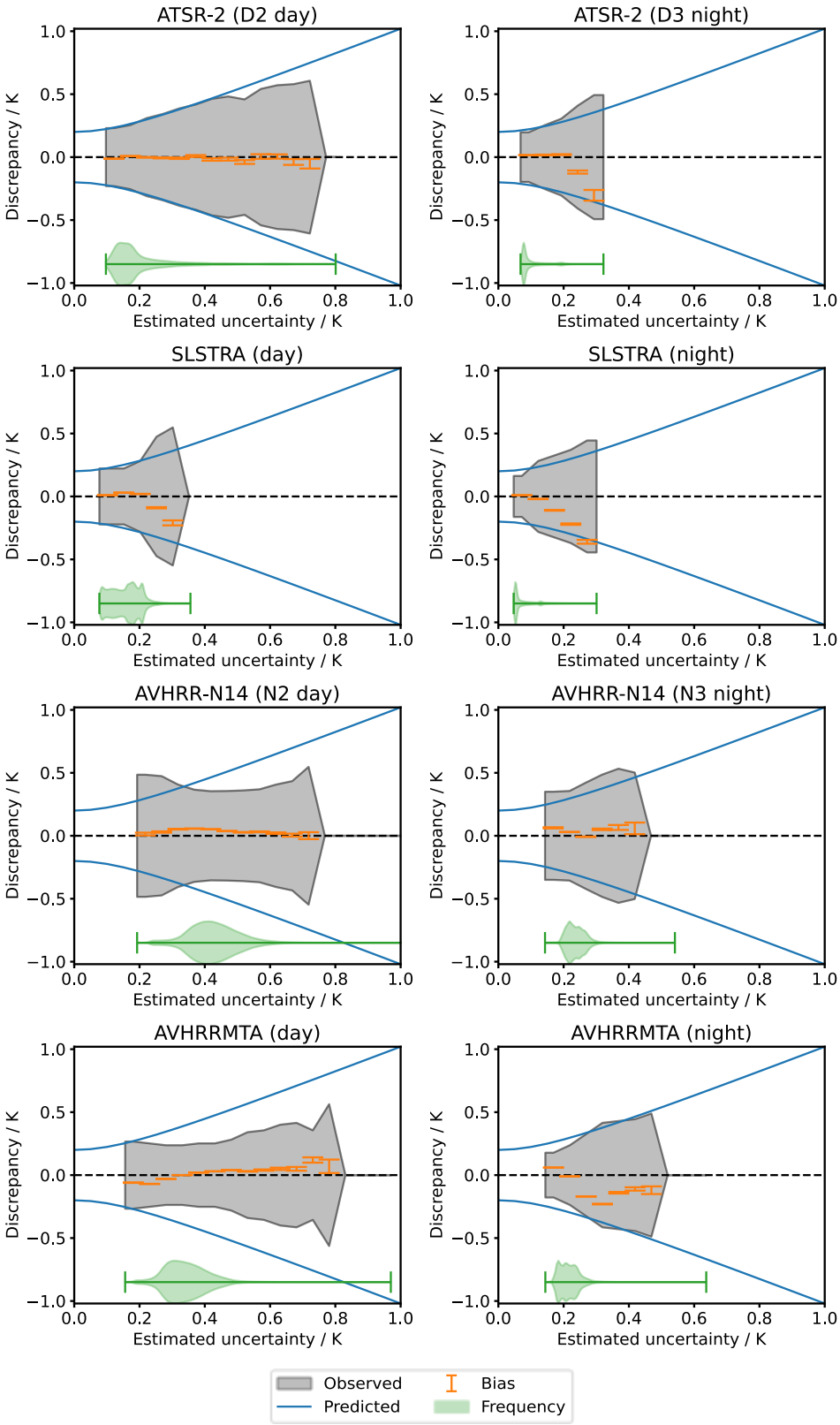


Figure 19: Dependence of satellite – in situ difference on estimated uncertainty. Shaded grey area shows RSD of difference. Solid blue line shows expected relationship based on assumed in situ drifter uncertainty of 0.2 K. Orange error bars show median difference in each bin. Green violin plot shows distribution of data.

4.3 Validation of L4 SST

4.3.1 L4 Matchup Criteria

Initial validation of the L4 SST product against all the non-ship in situ showed artefacts during the earlier part of the record which were not evident in the L3 results presented in section 4.1. These can be seen as horizontal artifacts in the hovmöller plot in Figure 20. In order to determine if these are related to the satellite data or the in situ we show the results for individual platform types in Figure 21 through Figure 24. The equivalent plots against in situ ship observations are shown in Figure 25. From the histogram plots in Figure 21 we see that the bottle and MBT types have a large number of outliers with a strong warm tail (L4 SST warmer than in situ measurements) and the highest standard deviations (over 1 K). The spatial distribution of the differences are shown in Figure 22, both bottle and MBT plots show large biases in the north-west Pacific around Japan which are not present in the CTD or XBT results (drifter and moorings platform types do not cover this region in the 1980s), or the ship in situ shown in Figure 25. Next the latitude-time hovmöller diagram in Figure 23 shows that some of the strongest artefacts seen in Figure 20 are associated with the bottle measurements around 35N. Finally, Figure 24 shows the bias and standard deviations in a distance-to-land versus time hovmöller diagram, where we see that the strongest biases and standard deviations in the bottle comparisons occur in coastal areas (distance to land < 10 km). Ships (Figure 25) do show increased biases and standard deviations in coastal areas, but not to the same extent as the bottle measurements. CTD and XBT comparisons both have increased bias near coasts, but without the increase in standard deviation.

Closer examination of the bottle measurements revealed that many of them were taken in the Seto Inland Sea. This is a narrow region between the three southern islands of Japan, joining the Sea of Japan to the Pacific Ocean, with a width varying between 13 and 50 km. Due to the resolution of the AVHRR data available in the 1980s, which is between approximately 4 and 20 km, most pixels over the Inland Sea will be rejected due to cloud or land contamination. As such the outliers could be caused by problems in the bottle SST measurements, the (lack of) satellite SST estimates, or both.

Table 5 summarises the SD and RSD statistics for each platform type for all matches (as shown in Figure 21) and “ocean” matches defined here as any match at least 10 km from land. Removing the near-coastal matches does reduce the SD and RSD for bottle, CTD, and MBT comparisons consistent with Figure 24. However, the SD and RSD of bottle and MBTs still remain higher than the other non-ship platforms.

For the global Level 4 SST validation will be limited to matchups at least 10 km from the coast; using CTD, Drifter, GTMBA, Moorings, and XBT observations up to end of 1995; and drifters only from 1996 onwards.

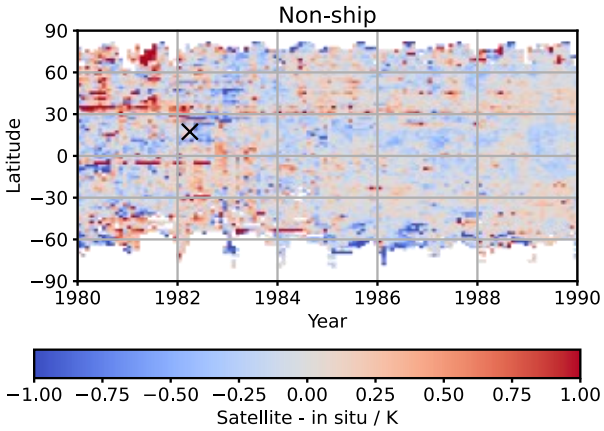


Figure 20: Hovmöller distribution of Level 4 SST0.2m minus reference in situ difference during 1980s. Several horizontal artefacts can be seen where satellite and in situ SST differ at fixed latitudes.

	Global			Ocean (over 10 km from land)		
In situ type	Number	SD / K	RSD / K	N	SD / K	RSD / K
Bottle	103005	1.36	0.74	67909	1.06	0.56
CTD	123814	0.98	0.54	104339	0.90	0.51
Drifter	257794	0.78	0.48	253766	0.75	0.47
MBT	15892	1.37	0.73	14199	1.29	0.69
Mooring	139118	0.80	0.56	125561	0.79	0.55
XBT	99045	0.83	0.49	96861	0.82	0.48
Ship	18055483	1.28	1.00	17508166	1.27	1.00

Table 5: Number, standard deviation, and robust standard deviation of satellite L4 minus in situ difference for each in situ type during 1980s. Left columns show all matches; right columns show open-ocean matches (cases where match is over 10 km from land)

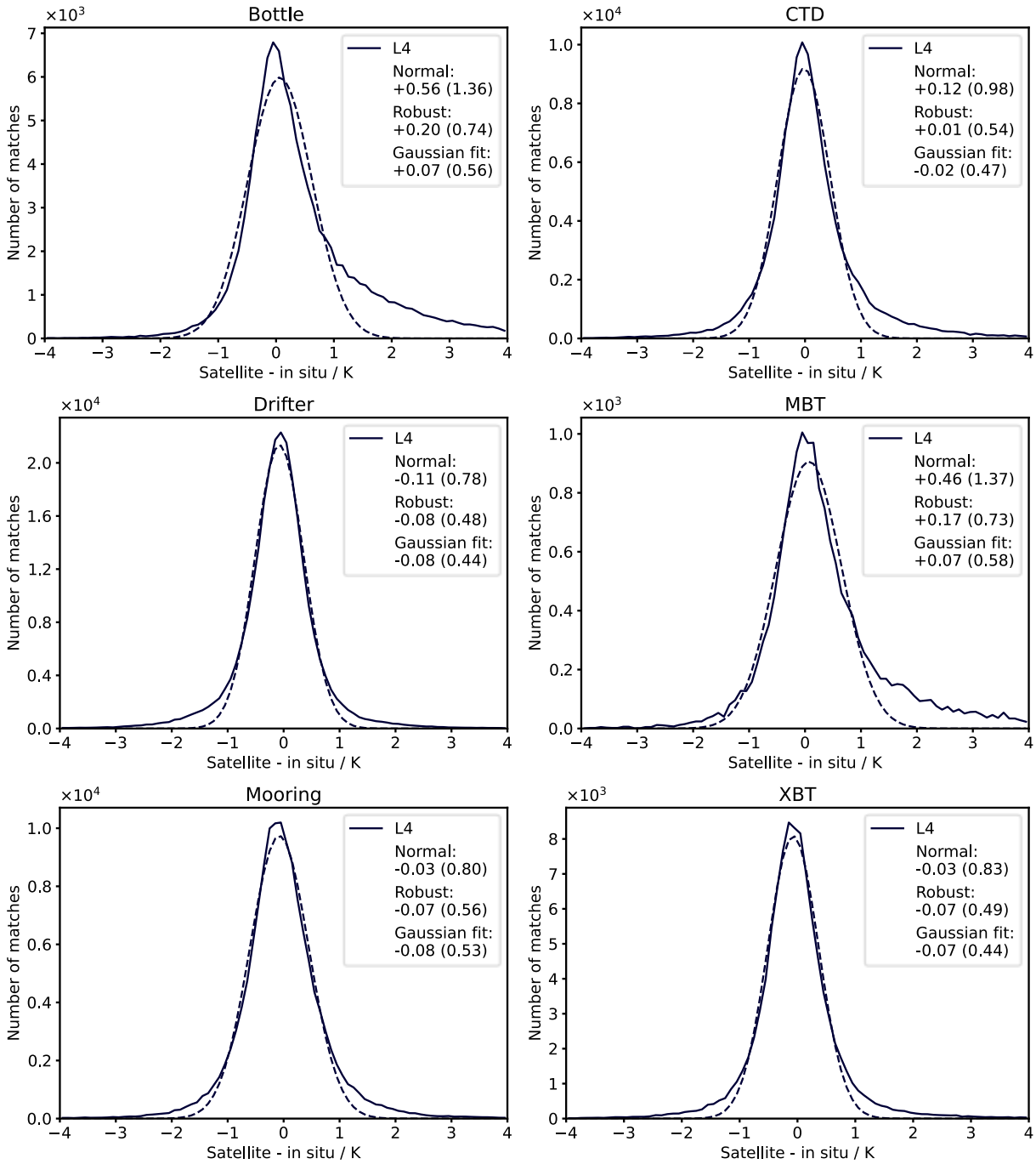


Figure 21: Histograms of L4 SST minus in situ by in situ type during 1980s.

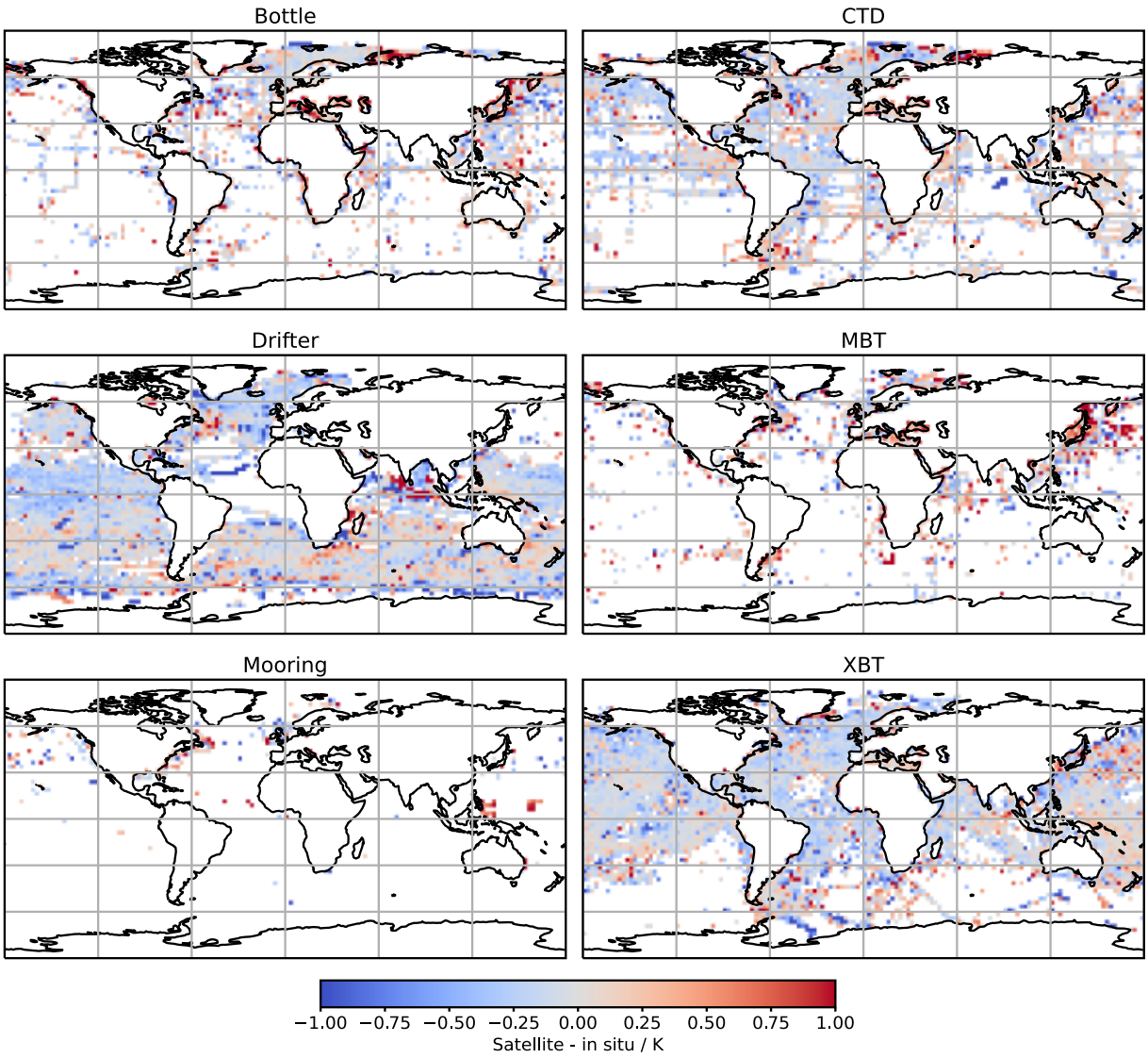


Figure 22: Spatial distribution of L4 SST minus in situ by in situ type during 1980s.

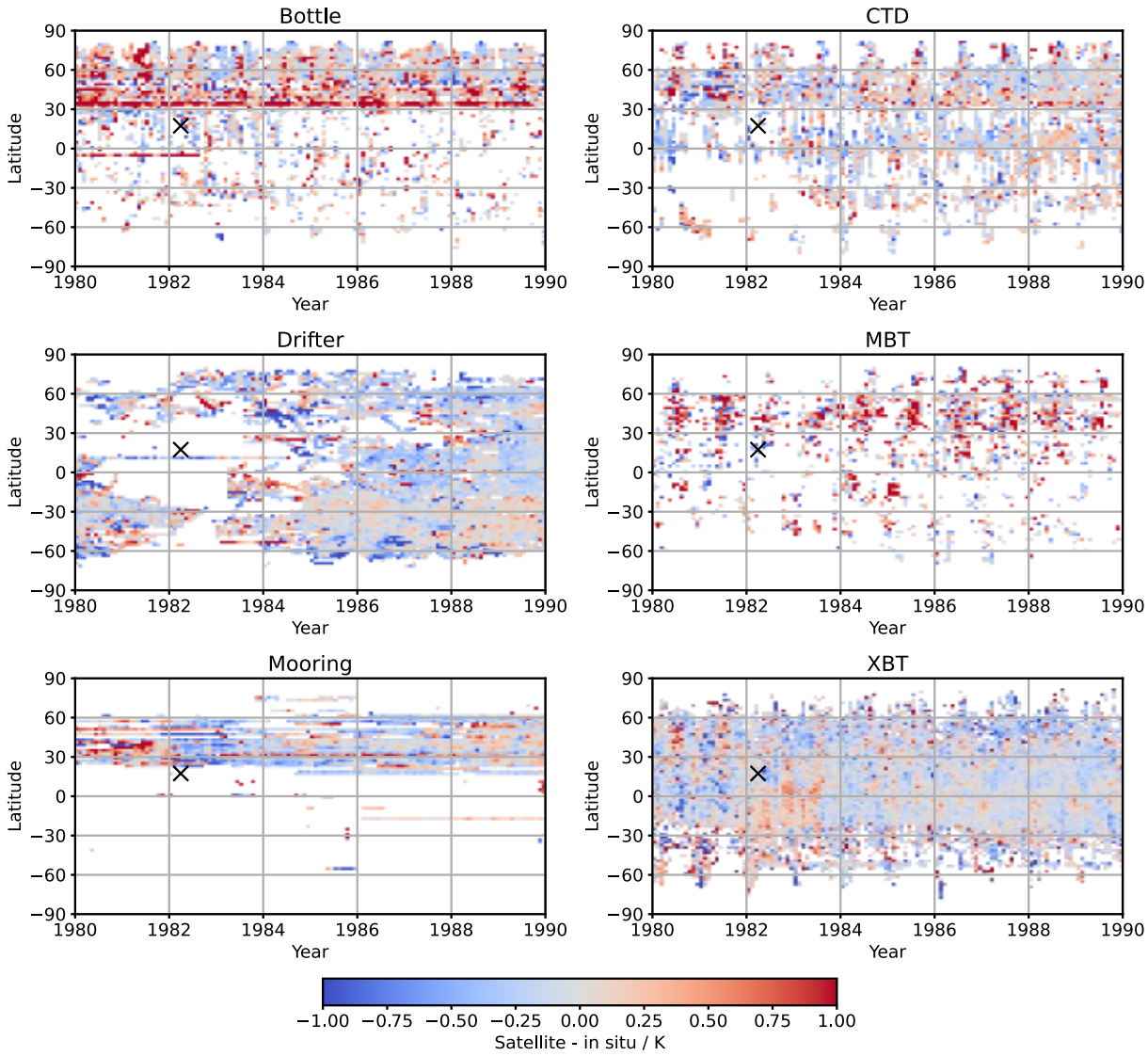


Figure 23: Latitude versus time Hovmöller diagram of L4 SST minus in situ by in situ type during 1980s.

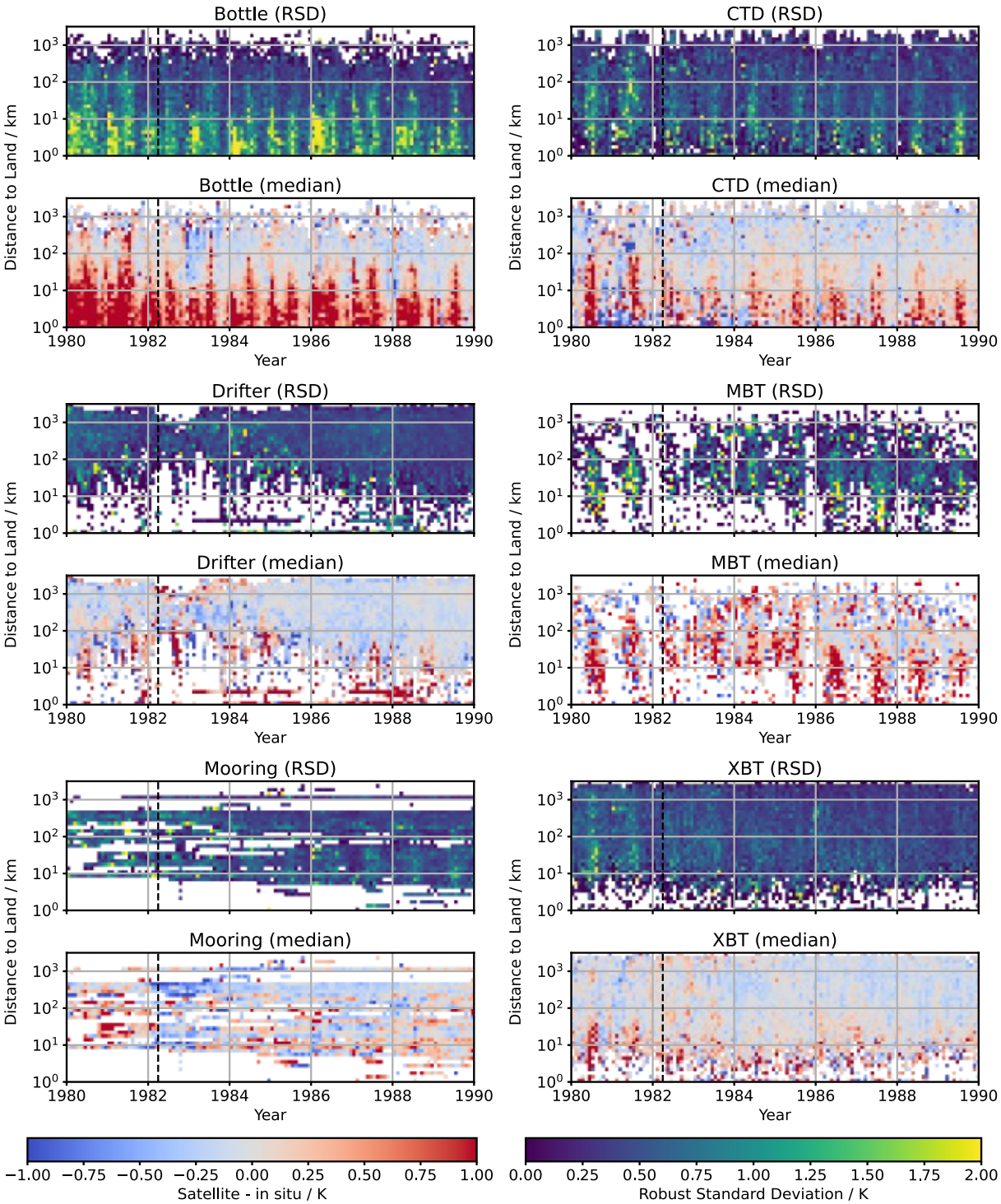


Figure 24: Distance-to-land versus time Hovmöller diagram of L4 SST minus in situ by in situ type during 1980s. Due to the resolution of the L4 grid (0.05°) and reporting of in situ locations (0.1° during this time period) distance estimates will become unreliable below ~10 km.

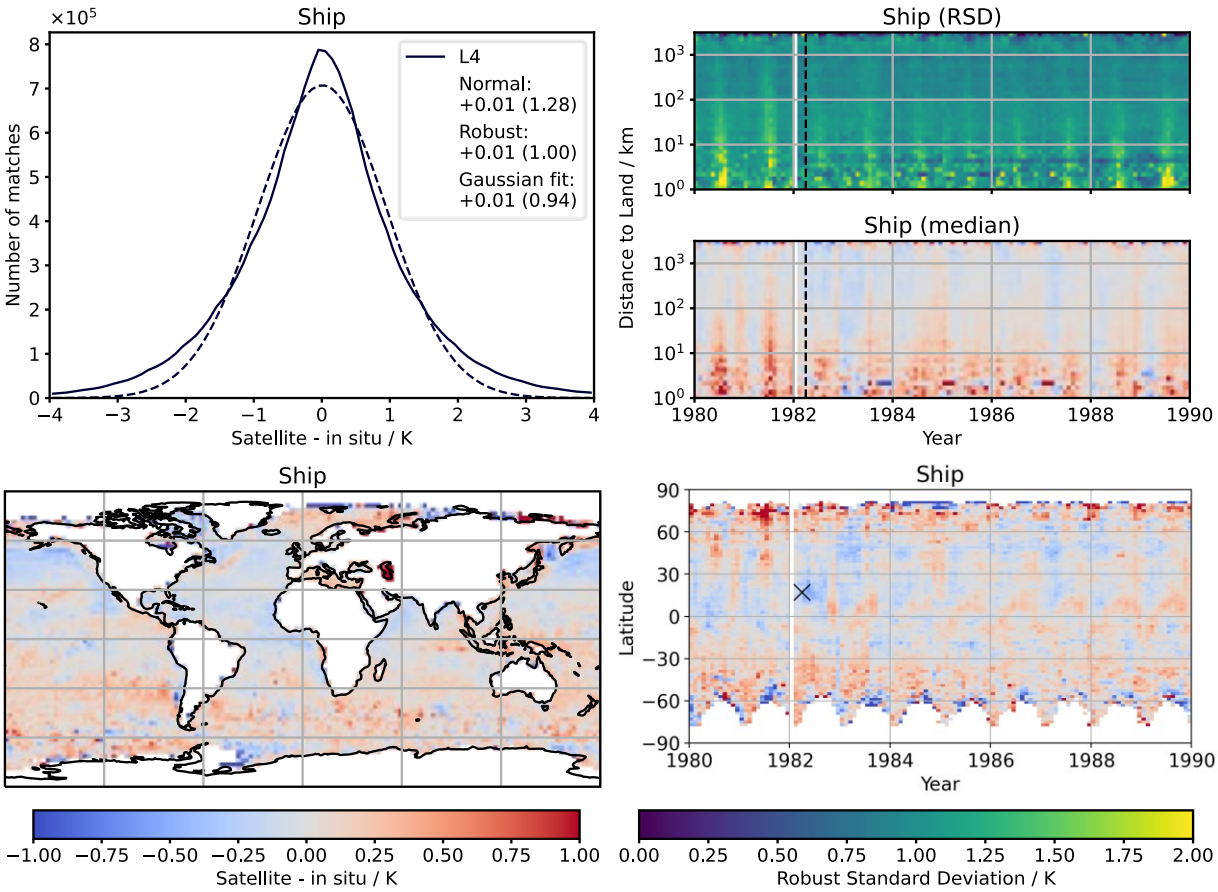


Figure 25: Plots of L4 SST minus ship differences during 1980s. Top-left: histogram; top-right: spatial distribution; lower-left: latitude versus time Hovmöller diagram; lower-right: distance-to-land versus time Hovmöller diagram.

4.3.2 Global L4 SST

Global validation statistics comparing the Level 4 SST product against in situ are shown in Table 6. All matches are at least 10 km from land to avoid issues with coastal data in the 1980s (see section 4.3.1); the reference in situ includes CTD, Drifter, GTMBA, Moorings, and XBT observations up to end of 1995 and drifters only from 1996 onwards. The decrease in RSD of Level 4 - reference in situ with time is consistent with that seen for the Level 3 results (see Table 3 and Figure 12), going from ~0.5 K in 1980s to 0.22 K for the last decade. In all cases the performance is within target with biases less than 0.1 K.

Period	**Reference in situ**		**Ship in situ**	
	Median	RSD	Median	RSD
1980	-0.07	0.50	+0.00	1.00
1990	-0.03	0.35	+0.01	0.92
2000	-0.03	0.23	+0.03	0.80
2010	-0.05	0.22	+0.04	0.74
All	-0.04	0.25	+0.02	0.89

Table 6: Summary of L4 validation against in situ for each decade and total time series. Reference in situ includes CTD, drifter, GTMBA, moorings, and XBT up to end-1995; and drifters-only from 1996 onwards.

The time-series of monthly median bias and RSD against both reference in situ and ships is show in Figure 26. The comparison against reference in situ shows a noticeable improvement in agreement around 2002 corresponding to the introduction of new sensors (also see Figure 3): AVHRR-16 in January, AMSR-E in June, AATSR in July, and AVHRR-16 in August. The drifter network was being rapidly expanded during this time; however, the largest change in the drifters occurs slightly later in 2005 (see Figure 4). There is a short 1–2 month spike in the bias during early 2001 when the L4 depends on just two AVHRR inputs (AVHRR-14 and AVHRR-15) both of which were affected by intermittent biases in this period, while ATSR-2 was unavailable the first half of 2001 due to the failure of the navigation gyros.

Plots of the bias as a function of latitude are shown in Figure 27. There is a noticeable cold bias relative to the reference in situ in the Arctic (Figure 27) – this is mostly due to the period without any dual-view sensors between end of AATSR in 2012 and start of SLSTR in 2016 as seen in the Hovmöller plot in Figure 28. The spatial distribution in Figure 29 shows that the Level 4 data are no longer affected by desert-dust biases.

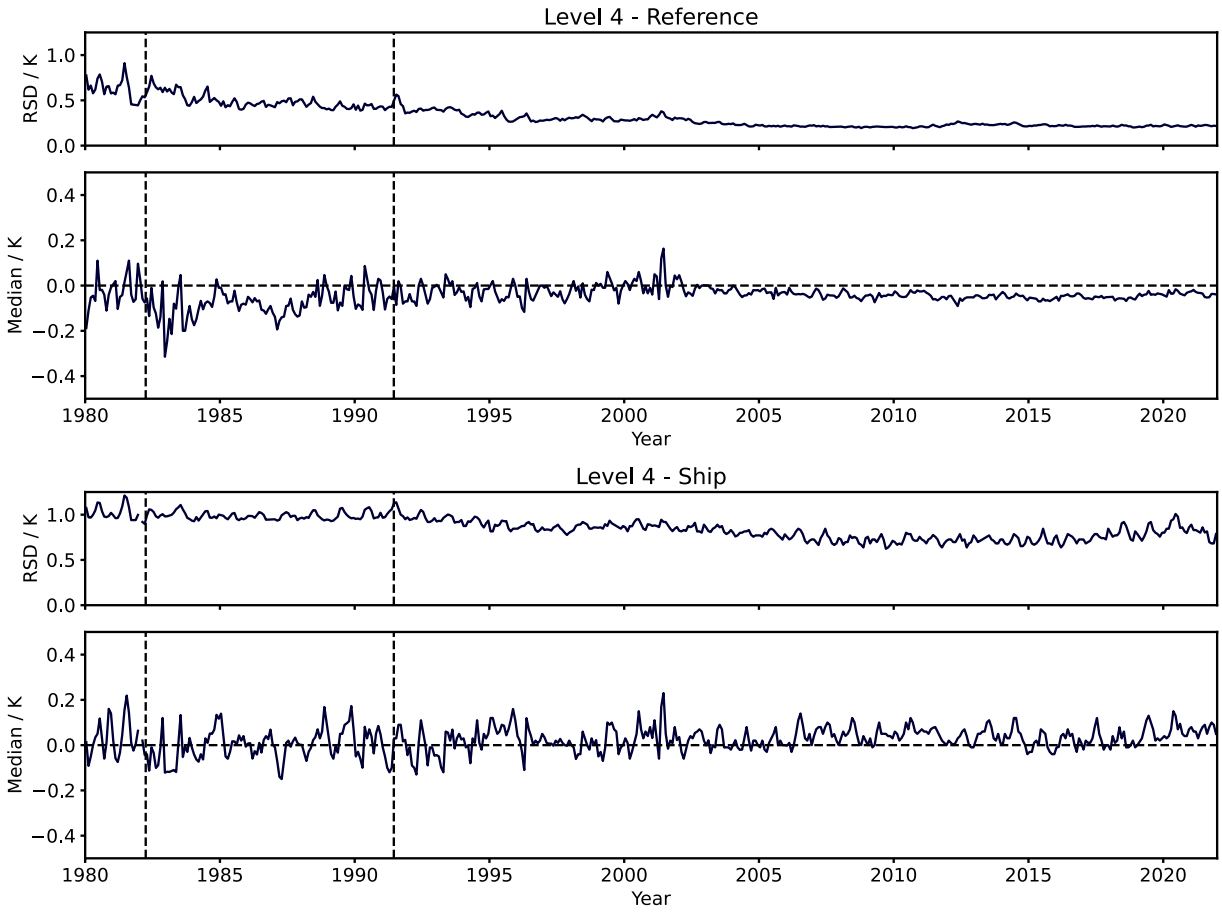


Figure 26: Time-series of Level 4 validation results. Top panel: reference in situ. Lower panel: ship in situ. Vertical dashed lines show time of El Chichón (April 1982) and Mount Pinatubo (June 1991) eruptions.

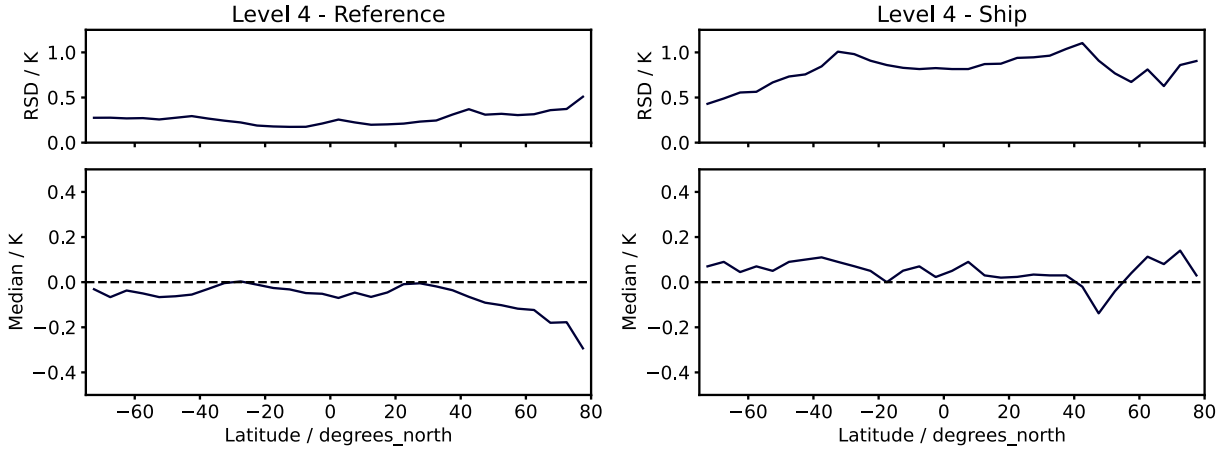


Figure 27: Level 4 validation as a function of latitude. Left panel: reference in situ. Right panel: ship in situ.

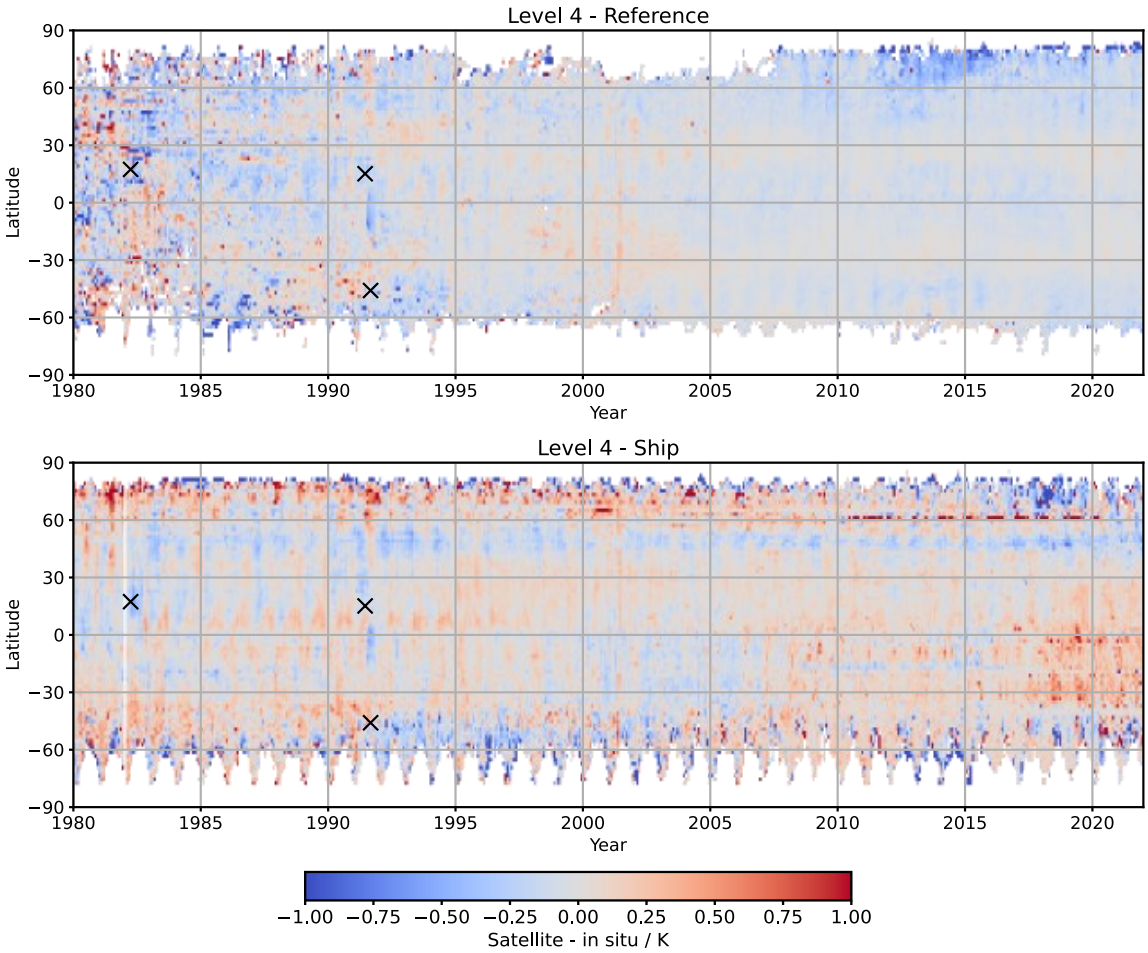


Figure 28: Hovmöller distribution of Level 4 minus in situ SST mean difference. Top panel: reference in situ. Lower panel: ship in situ. X symbols mark major volcanic eruptions: El Chichón (April 1982), Mount Pinatubo (June 1991), and Mount Hudson (September 1991).

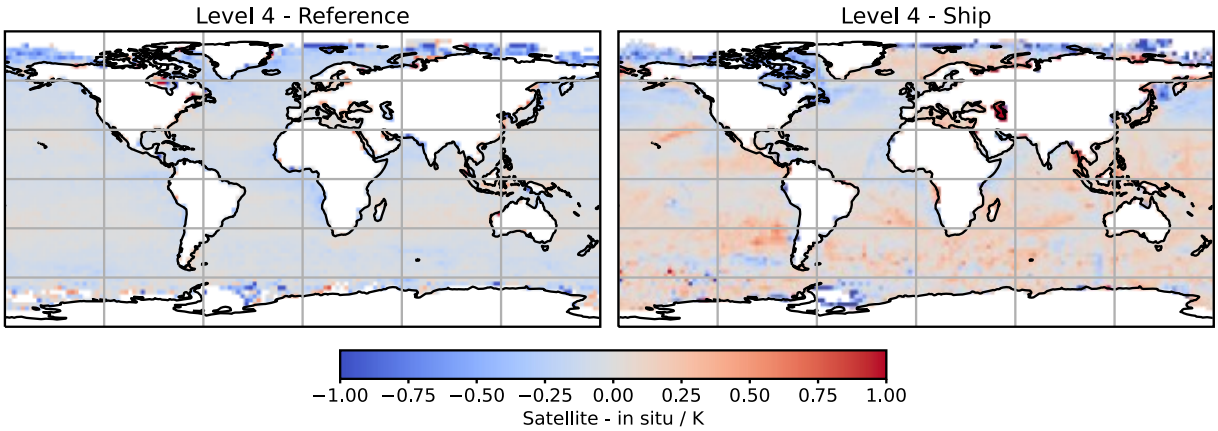


Figure 29: Spatial distribution of Level 4 minus in situ SST mean difference. Left panel: reference in situ. Right panel: ship in situ.

4.3.3 Coastal L4 SST

Validation statistics comparing coastal (within 10 km of land) and open-ocean (over 10 km from land) are shown in Table 7 and histograms of the satellite-in situ differences in Figure 30. Note – the coastal comparison includes the additional in situ types (ctd, mooring etc.) in the reference comparison for the full period of the data record. The RSD and SD against reference in situ is almost doubled in coastal regions compared to open-ocean – going from 0.25 K and 0.49 K in open-ocean to 0.46 K and 0.96 K close to land. From the histograms in Figure 30 we see a warm tail in the coastal L4-reference histogram (L4 SST warmer than in situ), such that the mean bias is 0.1 K while the median is 0.00 K.

A different behaviour is observed when comparing to ships – the RSD and SD do not change significantly between coastal and open-ocean (expected as these will be dominated by uncertainty in the ship measurements). There is no indication of skewness in the comparison against ships – all three bias estimates agree within 0.01 K – however, this maybe due to the much higher uncertainty in the ship comparisons (open-ocean RSD is 0.89 K compared to 0.25 K for reference in situ). The biases are different near the coast; however, the change is now in the opposite direction with L4-ship being slightly cooler by ~0.05 K near the coast compared to open-ocean.

The NOAA AVHRR sensor which provides the majority of the satellite observations until mid-2007 has a ground resolution of ~4 to 20 km depending on view angle. As such it cannot retrieve SST close to the coast and the L4 analysis must extrapolate from observations further out into the ocean. The dual-view ATSR instruments do have a higher ground resolution providing data from 1991 to 2012, but due to the narrow swath take approximately three days to acquire global coverage. It is not until June 2007 that we have global 1 km data from MetOp AVHRR.

	Ocean (over 10 km from land)			**Coastal (under 10 km from land)**		
In situ type	N / million	SD / K	RSD / K	N / million	SD / K	RSD / K
Reference	10.81	0.49	0.25	0.54	0.96	0.46
Ship	45.03	1.22	0.89	3.23	1.25	0.88

Table 7: Number, standard deviation, and robust standard deviation of satellite L4 minus in situ difference for reference and ship-only matches. Left columns shows open-ocean matches (cases where match is over 10 km from land); right columns shows coastal matches.

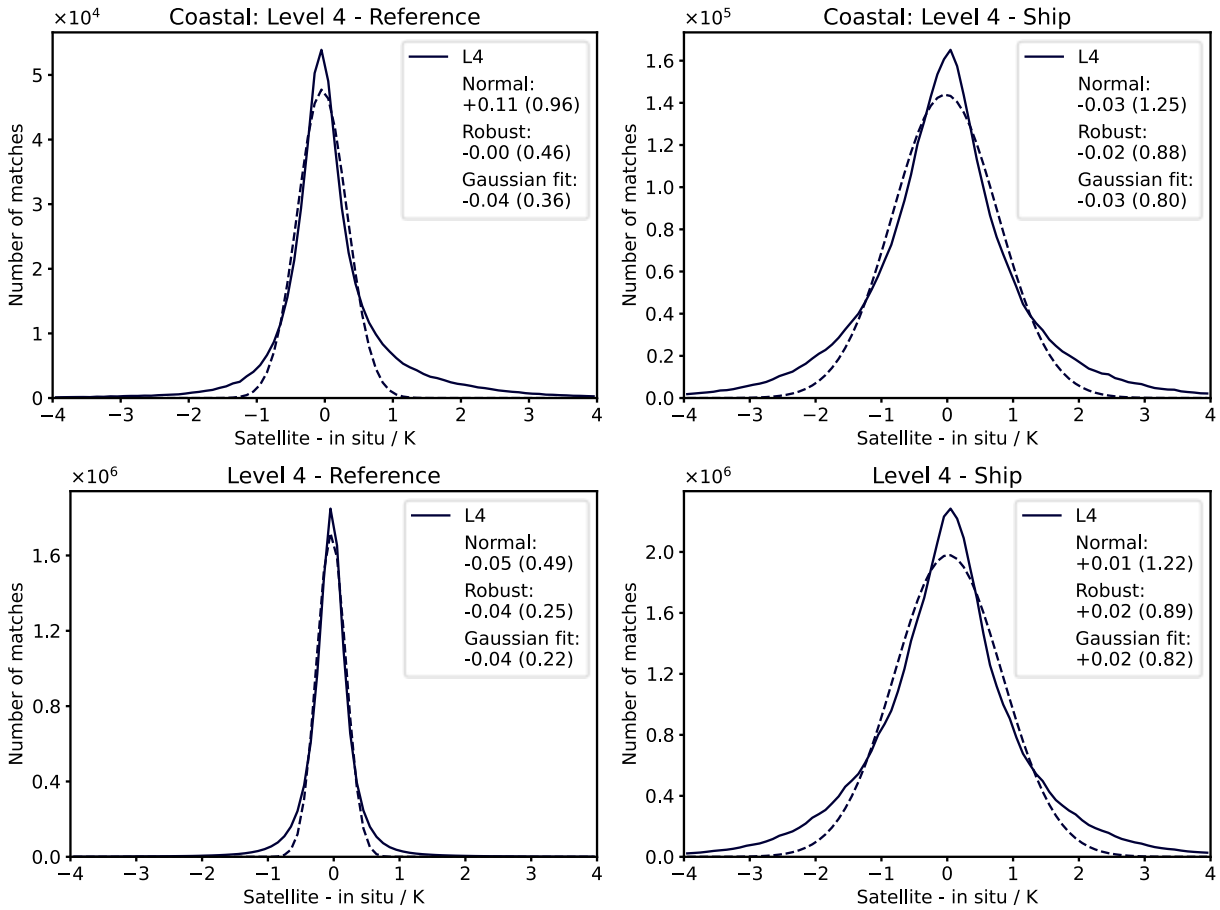


Figure 30: Histograms of L4 SST minus in situ differences. Top-row: coastal locations within 10 km of land. Lower-row: open water over 10 km from land. Left column: reference in situ (Note – coastal plot includes moorings and other non-drifter observations for complete record). Right column: ship in situ.

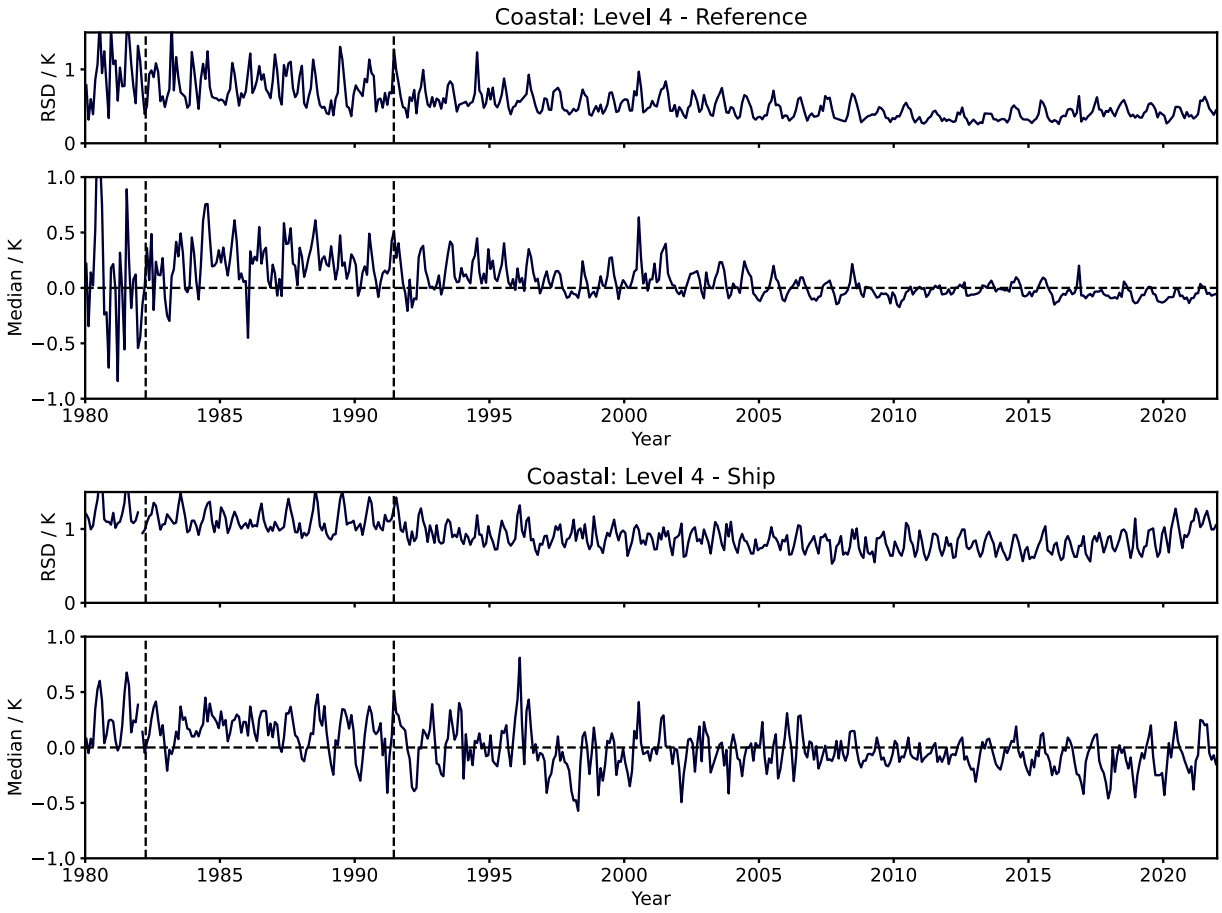


Figure 31: Time-series of L4 minus reference in situ SST.

4.4 Validation of L4 uncertainties

Uncertainty validation plots for the global and coastal Level 4 data against reference in situ and ship are show in Figure 32. In all cases the spread of RSDs is slightly narrower than the expected envelope – indicating that either the estimated Level 4 uncertainties are too high, or the assumed in situ uncertainty is too high. In this case we believe the assumed in situ uncertainty is too high: as the reference data includes a mixture of in situ types we use the median reported uncertainty from the SIRDS dataset which at 0.39 K is significantly higher than 0.2 K assumed for drifter-only datasets.

In the comparison of ocean data to reference in situ (Figure 32 top-left) we see that distribution of estimated uncertainty peaks between 0.2 and 0.3 K with the majority of all data under 0.5 K. With coastal data (Figure 32 lower-left) the estimated uncertainties are noticeably higher, with a peak in the distribution between 0.3 and 0.5 K and a noticeable fraction up to 1.0 – 1.5 K. In both regions, extreme cases with estimated uncertainties over 3 K do exist (mostly in the earlier record), there are insufficient to estimate the RSD or median biases above about 2 – 2.5 K. In the comparison against ships (Figure 32 right column) the distributions are shifted towards slightly higher estimated uncertainties due to the spatial and temporal sampling.

Figure 33 shows the uncertainty validation against reference in situ and ships for each of the four complete decades: 1980s, 1990s, 2000s, 2010s.

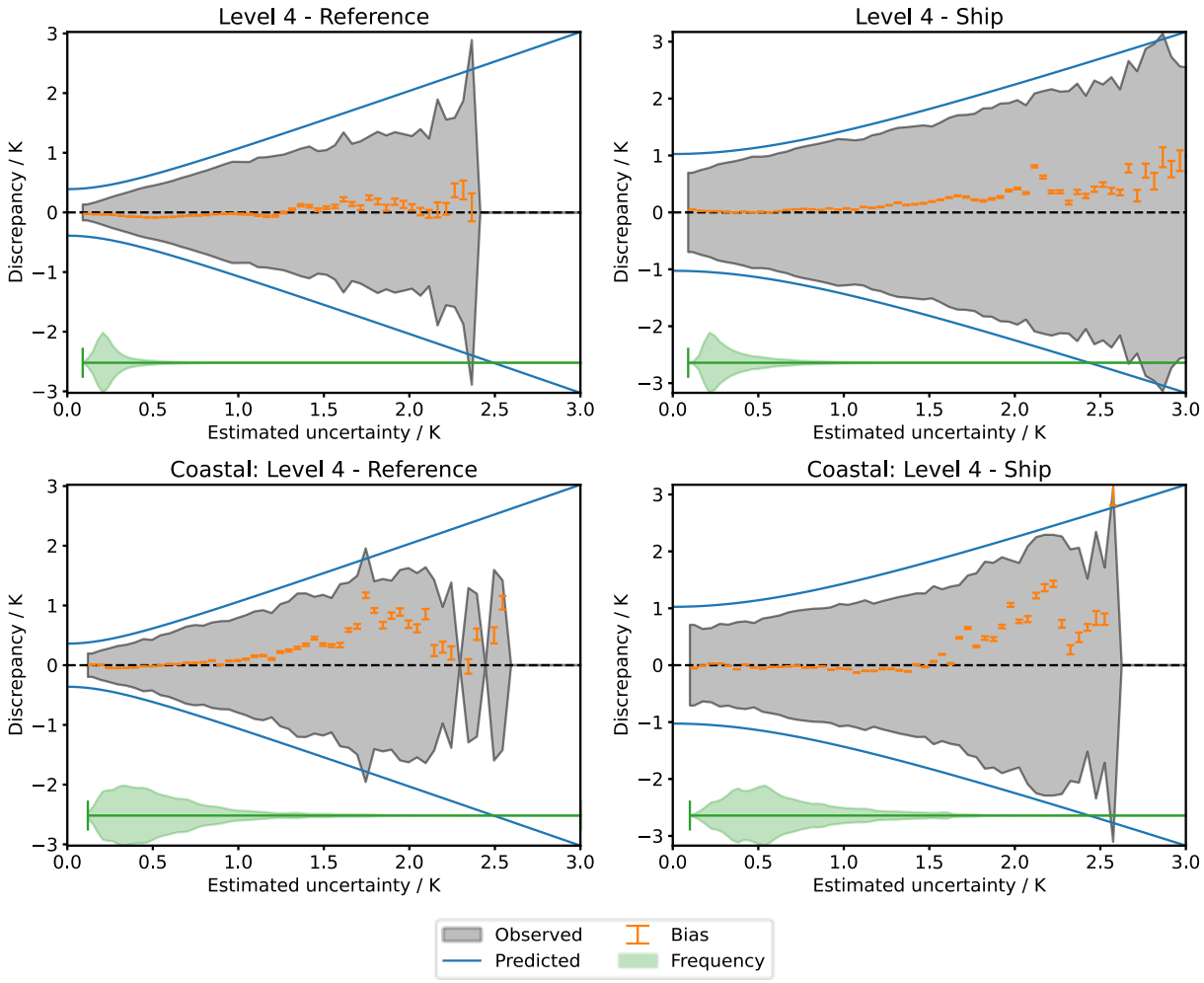


Figure 32: Dependence of the median (error bars) and robust standard deviation (grey shaded area) between L4 SST and in situ SST discrepancies as a function of estimate uncertainty. Blue line shows expected enveloped based on assumed in situ uncertainties. Green violin plots show distribution of data.

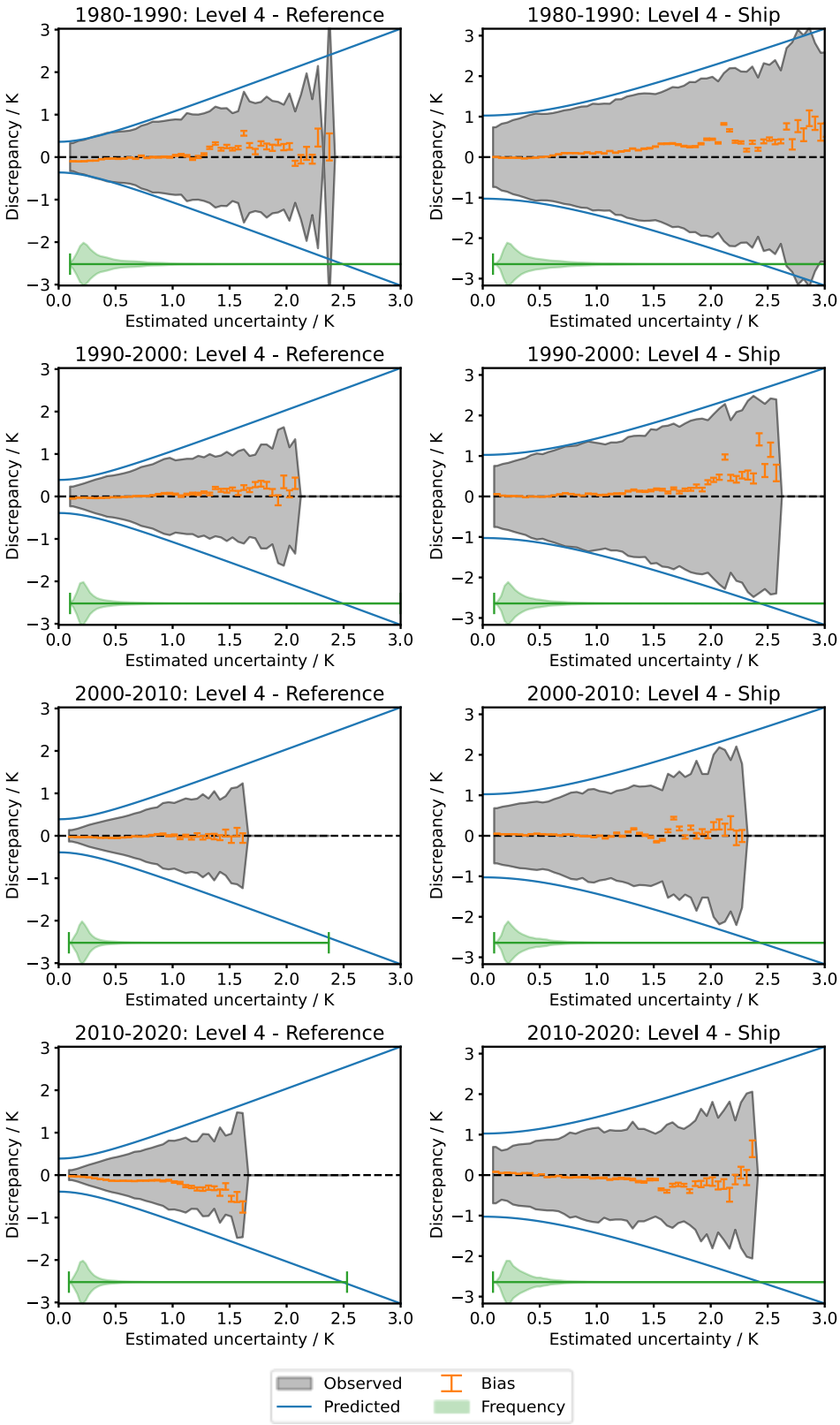


Figure 33: Dependence of the median (error bars) and robust standard deviation (grey shaded area) between L4 SST and in situ SST discrepancies as a function of estimate uncertainty by decade. Blue line shows expected enveloped based on assumed in situ uncertainties. Green violin plots show distribution of data.

5. REFERENCES

Atkinson, C.P., Rayner, N.A., Roberts-Jones, J. and Smith, R.O., (**2013**), Assessing the quality of sea surface temperature observations from drifting buoys and ships on a platform-by-platform basis, *J. Geophys. Res. Oceans*, 118, doi:10.1002/jgrc.20257

Atkinson, C. P., Rayner, N. A., Kennedy, J. J., and Good, S. A. (**2014**) An integrated database of ocean temperature and salinity observations, *J. Geophys. Res. Oceans*, 119, 7139-7163, doi:10.1002/2014JC010053

Blackmore, T., O'Carroll, A., Fennig, K., and Saunders, R. (**2012**) Correction of AVHRR Pathfinder SST data for volcanic aerosol effects using ATSR SSTs and TOMS aerosol optical depth, *Remote Sensing of Environment*, 116, 107-117, doi: 10.1016/j.rse.2011.04.040

Block, T., Embacher, S., Merchant, C. J. and Donlon, C. (**2018**) High performance software framework for the calculation of satellite-to-satellite data matchups (MMS version 1.2). *Geoscientific Model Development*, 11 (6). pp. 2419-2427. doi:10.5194/gmd-11-2419-2018

Fairall, C. W., Bradley, E. F., Godfrey, J. S., Wick, G. A., Edson, J. B., and Young, G. S. (**1996**), Cool-skin and warm-layer effects on sea surface temperature, *J. Geophys. Res.*, 101(C1), 1295-1308, doi:10.1029/95JC03190

Gentemann, C. L., Minnett, P. J., Le Borgne, P. and Merchant, C. J. (**2008**) Multi-satellite measurements of large diurnal warming events. *Geophysical Research Letters*, 35 (22). L22602. doi:10.1029/2008GL035730

Kantha, L. H., and Clayson, C. A. (**1994**), An improved mixed layer model for geophysical applications, *J. Geophys. Res.*, 99(C12), 25235-25266, doi:10.1029/94JC02257

Lumpkin, R., and Pazos, M. (**2007**). Measuring surface currents with Surface Velocity Program drifters: The instrument, its data, and some recent results. In A. Griffa, A. Kirwan, Jr., A. Mariano, T. Özgökmen, & H. Rossby (Eds.), *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics* (pp. 39-67). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511535901.003

Merchant, C. J., Harris, A. R., Murray, M. J., and Závody, A. M. (**1999**), Toward the elimination of bias in satellite retrievals of sea surface temperature: 1. Theory, modeling and interalgorithm comparison, *J. Geophys. Res.*, 104(C10), 23565-23578, doi:10.1029/1999JC900105

Morak-Bozzo, S., Merchant, C. J., Kent, E. C., Berry, D. I. and Carella, G. (**2016**) Climatological diurnal variability in sea surface temperature characterized from drifting buoy data. *Geosci. Data J.*, 3: 20-28. doi:10.1002/gdj3.35

Rayner, N.A., Brohan, P., Parker, D.E., Folland, C.K., Kennedy, J.J., Vanicek, M., Ansell, T.J. and Tett S.F.B., (**2006**), Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset, *J. Climate*, 19, 446-469, doi:10.1175/JCLI3637.1

Woodruff, S., Worley, S., Lubker, S., Ji, Z., Freeman, J., Berry, D., Brohan, P., Kent, E., Reynolds, R., Smith, S. and Wilkinson, C., (**2011**), ICOADS release 2.5: extensions and enhancements to the surface marine meteorological archive, *International Journal of Climatology*, 31, 951-967, doi:10.1002/joc.2103

Zhang, H.-M., Reynolds, R. W., and Smith, T. M. (**2004**), Bias characteristics in the AVHRR sea surface temperature, *Geophys. Res. Lett.*, 31, L01307, doi:10.1029/2003GL018804

APPENDIX A: DETAILED SINGLE-SENSOR RESULTS

The following section contains the detailed, per-sensor, validation results. For each sensor we provide:

- Histograms of the discrepancies for day and night separately
- Median satellite - in situ discrepancy as a function of latitude and year.
- Hovmöller plots of the mean satellite - in situ discrepancy.
- Spatial plots of the mean satellite - in situ discrepancy.

A.1. NOAA AVHRR

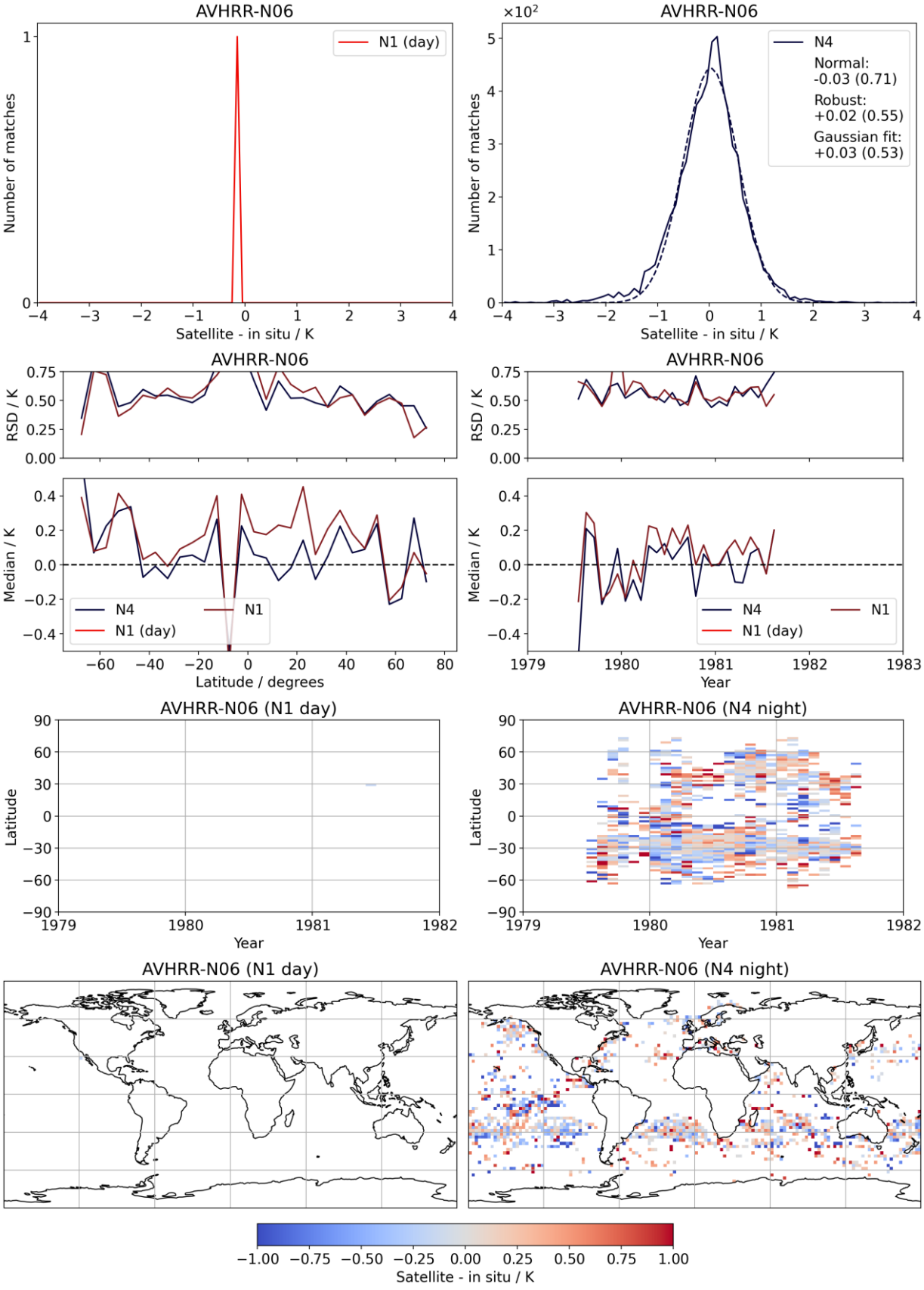


Figure 34: AVHRR-06 SST0.2m versus all non-ship in situ.

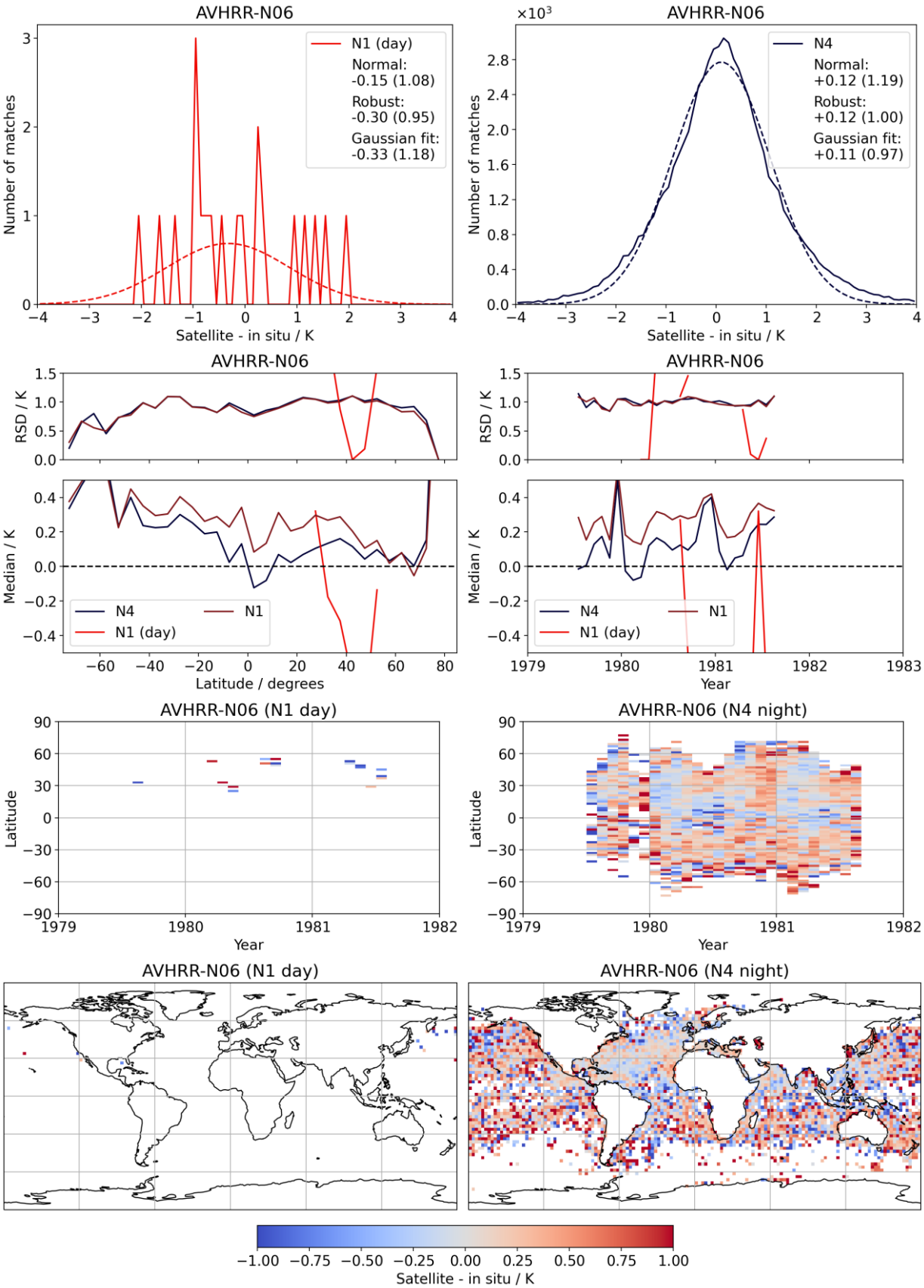


Figure 35: AVHRR-06 SST0.2m versus ships.

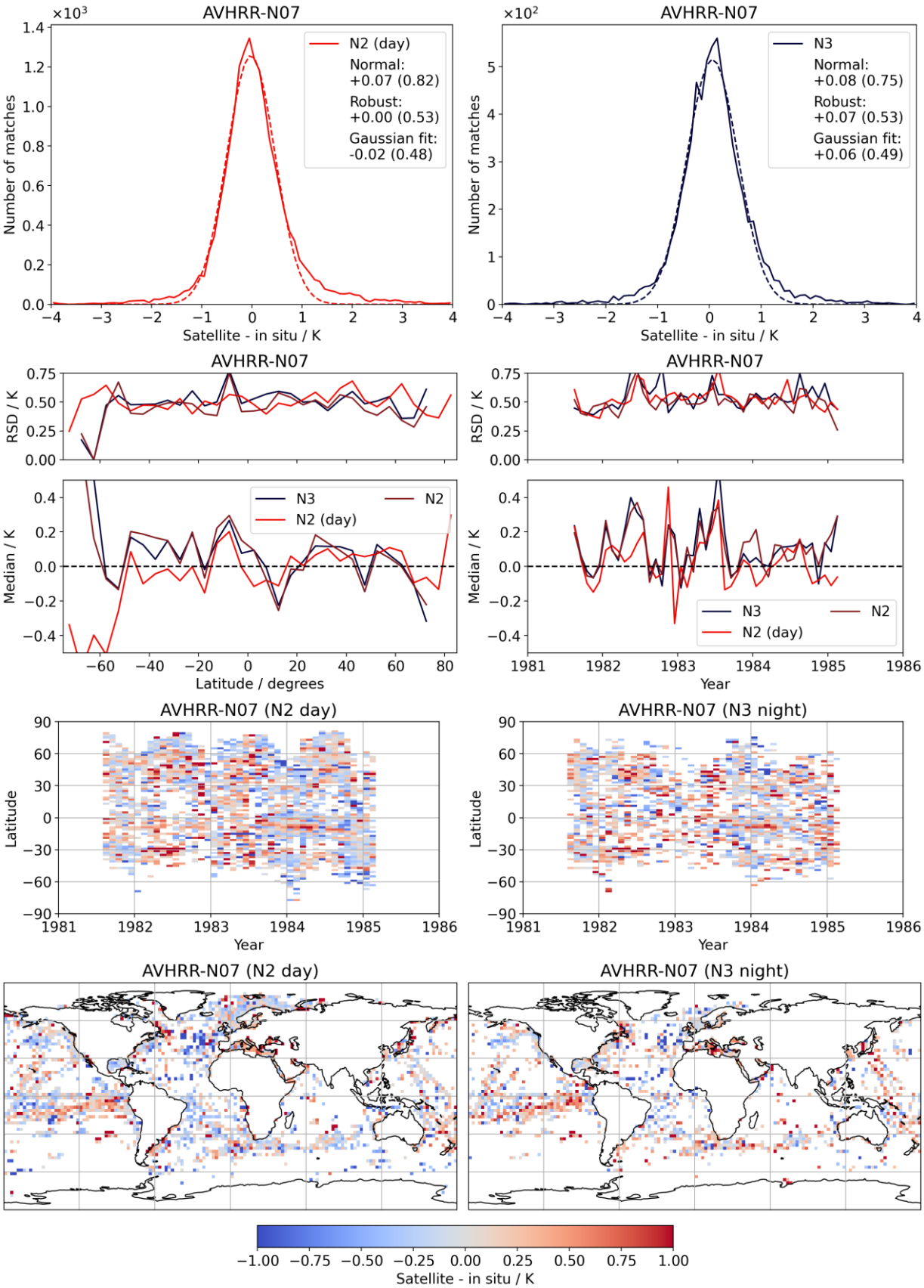


Figure 36: AVHRR-07 SST0.2m versus all non-ship in situ.

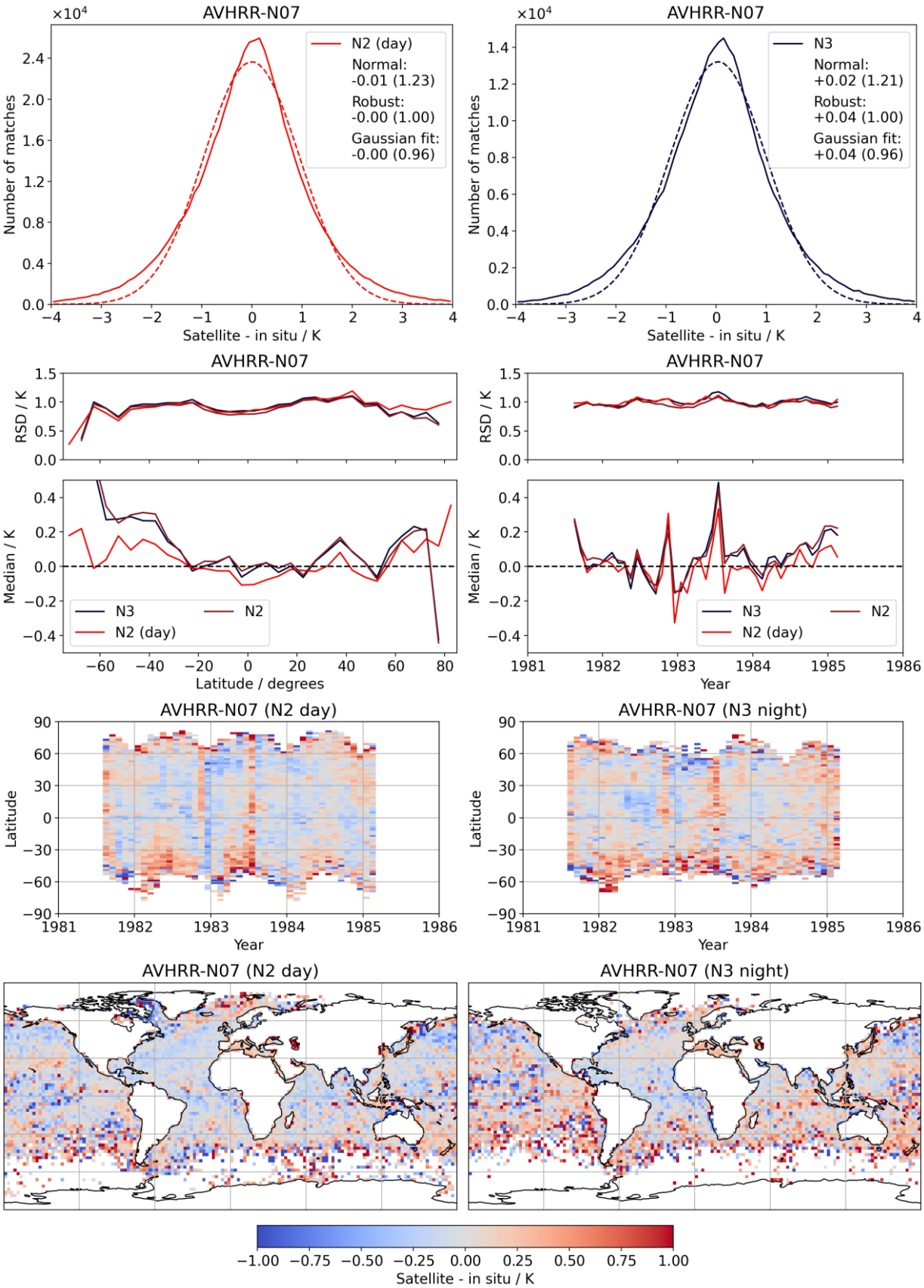


Figure 37: AVHRR-07 SST0.2m versus ships.

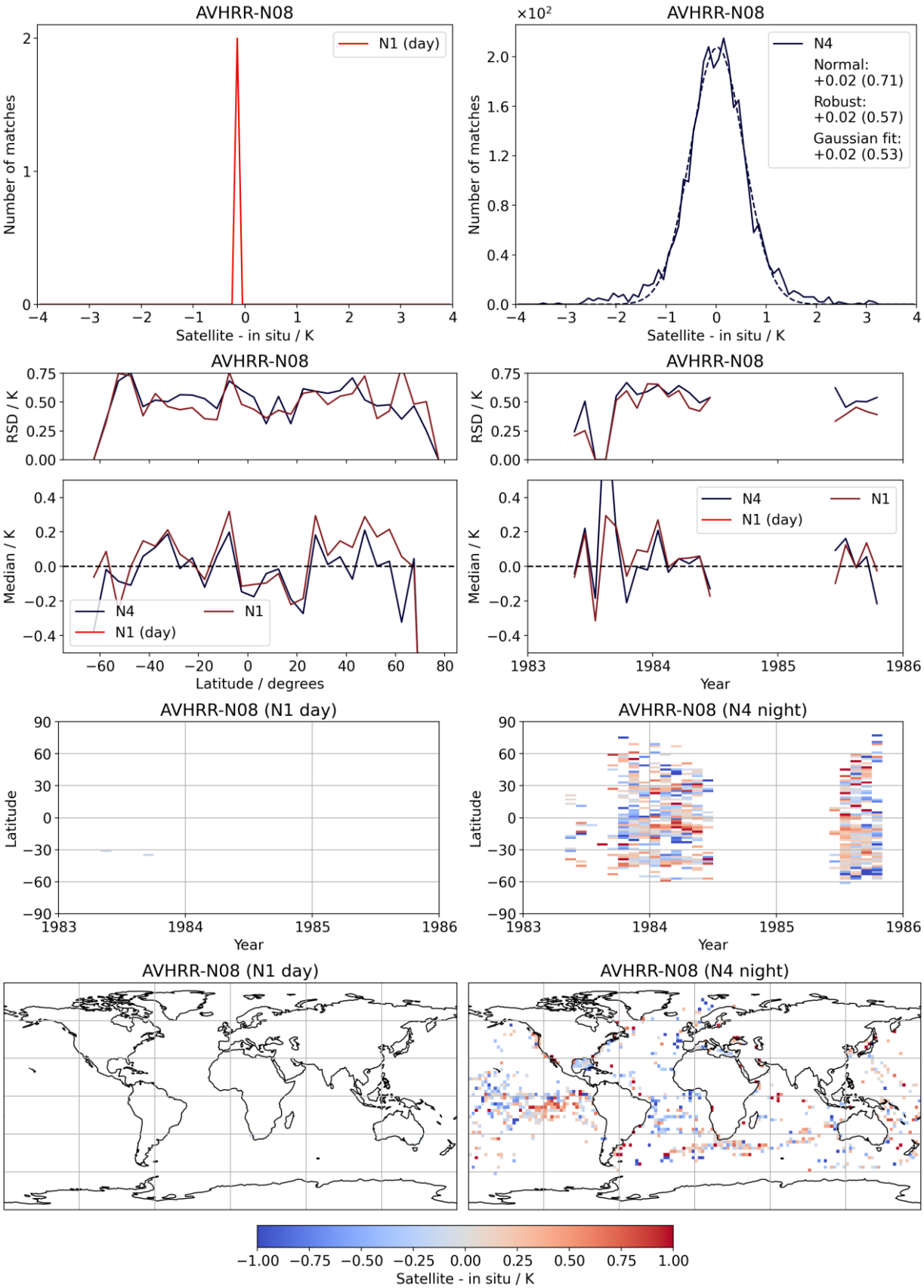


Figure 38: AVHRR-08 SST0.2m versus all non-ship in situ.

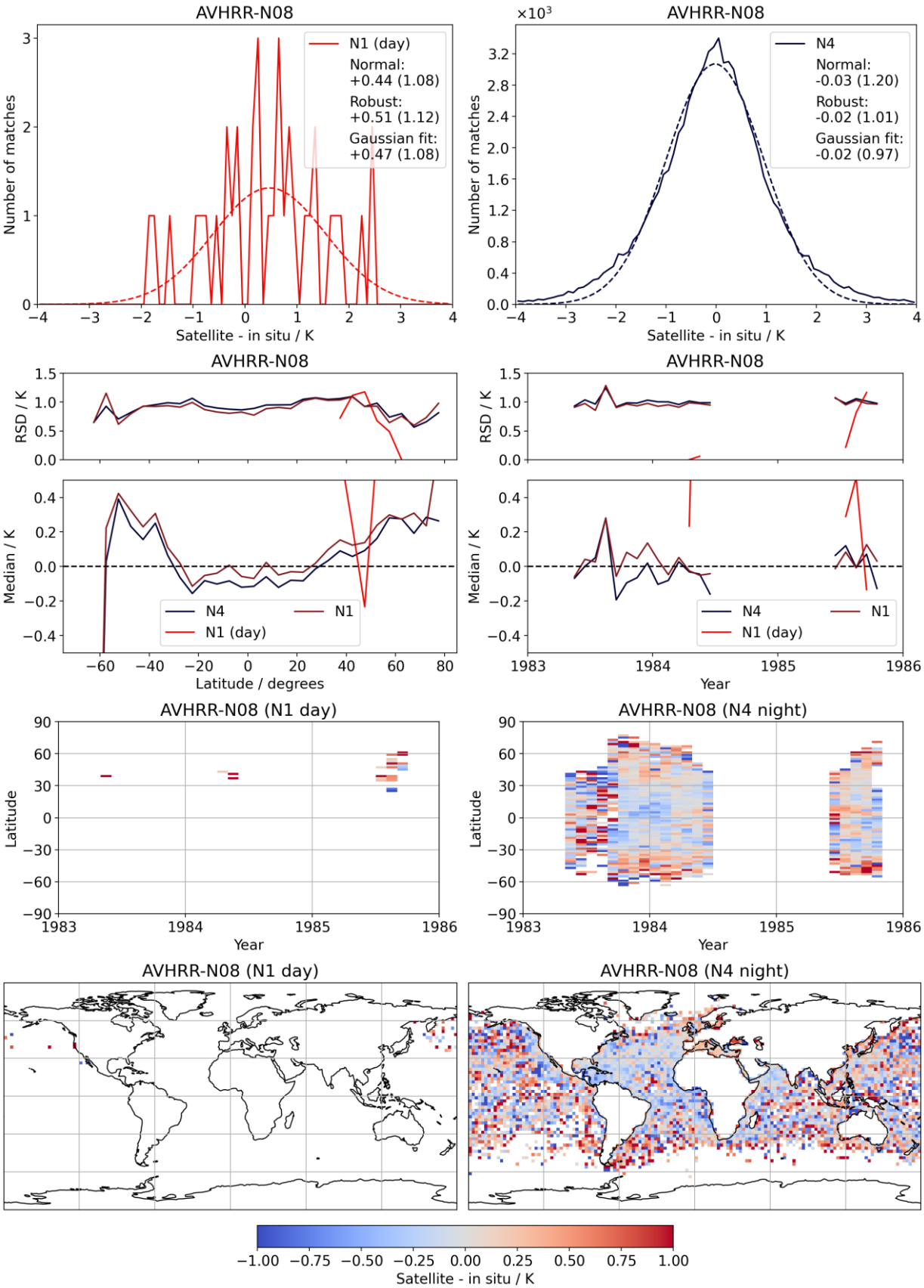


Figure 39: AVHRR-08 SST0.2m versus ships.

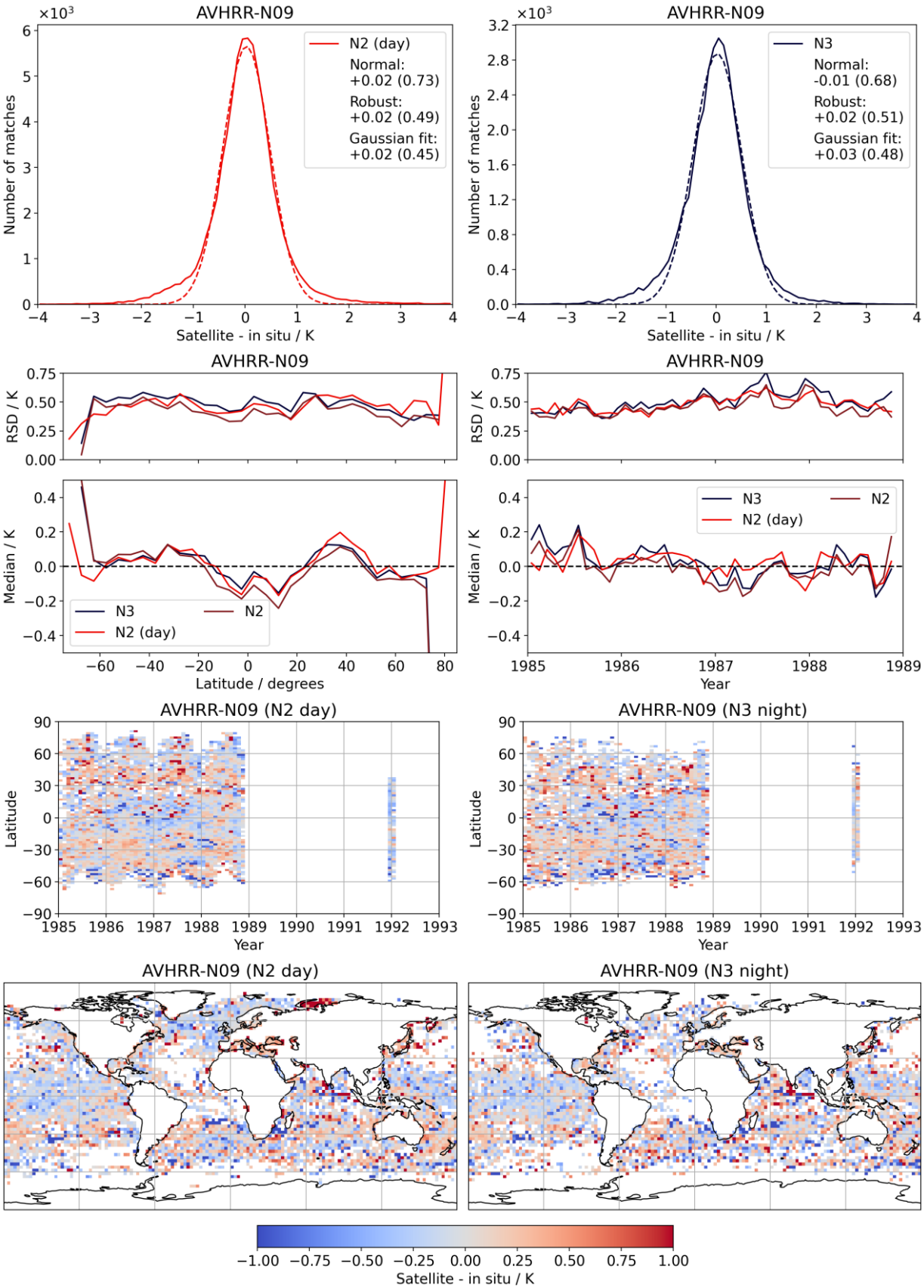


Figure 40: AVHRR-09 SST0.2m versus all non-ship in situ.

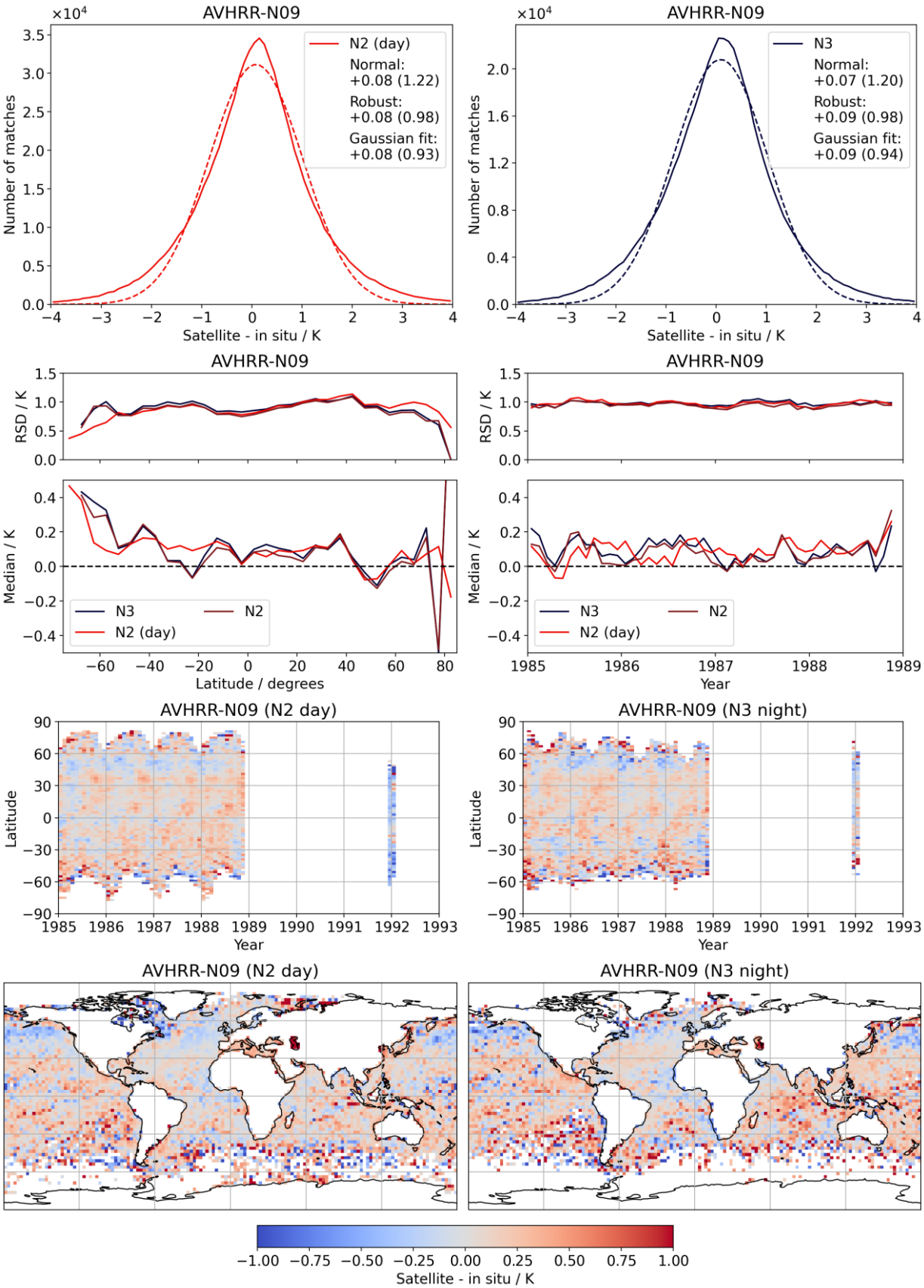


Figure 41: AVHRR-09 SST0.2m versus ships.

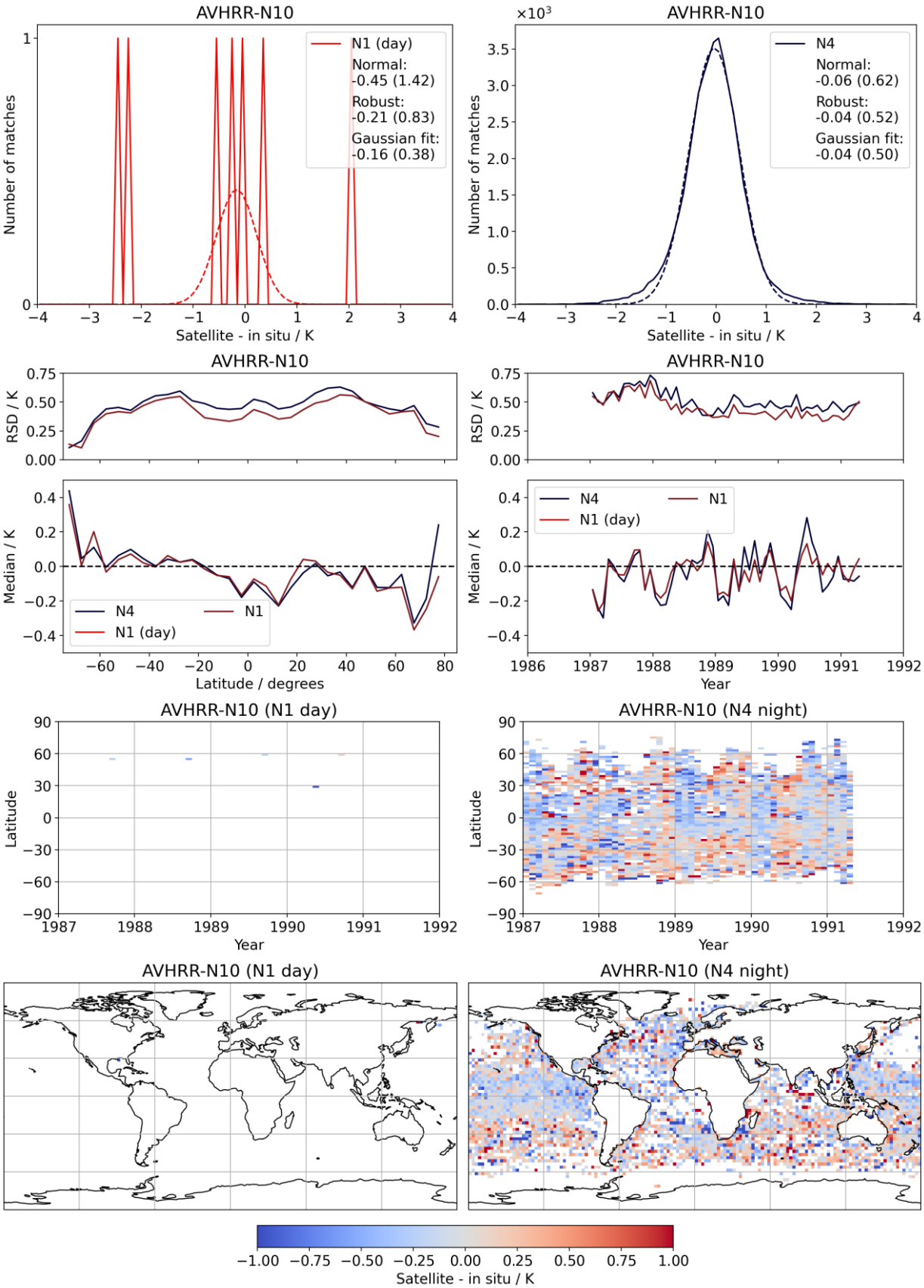


Figure 42: AVHRR-10 SST0.2m versus all non-ship in situ.

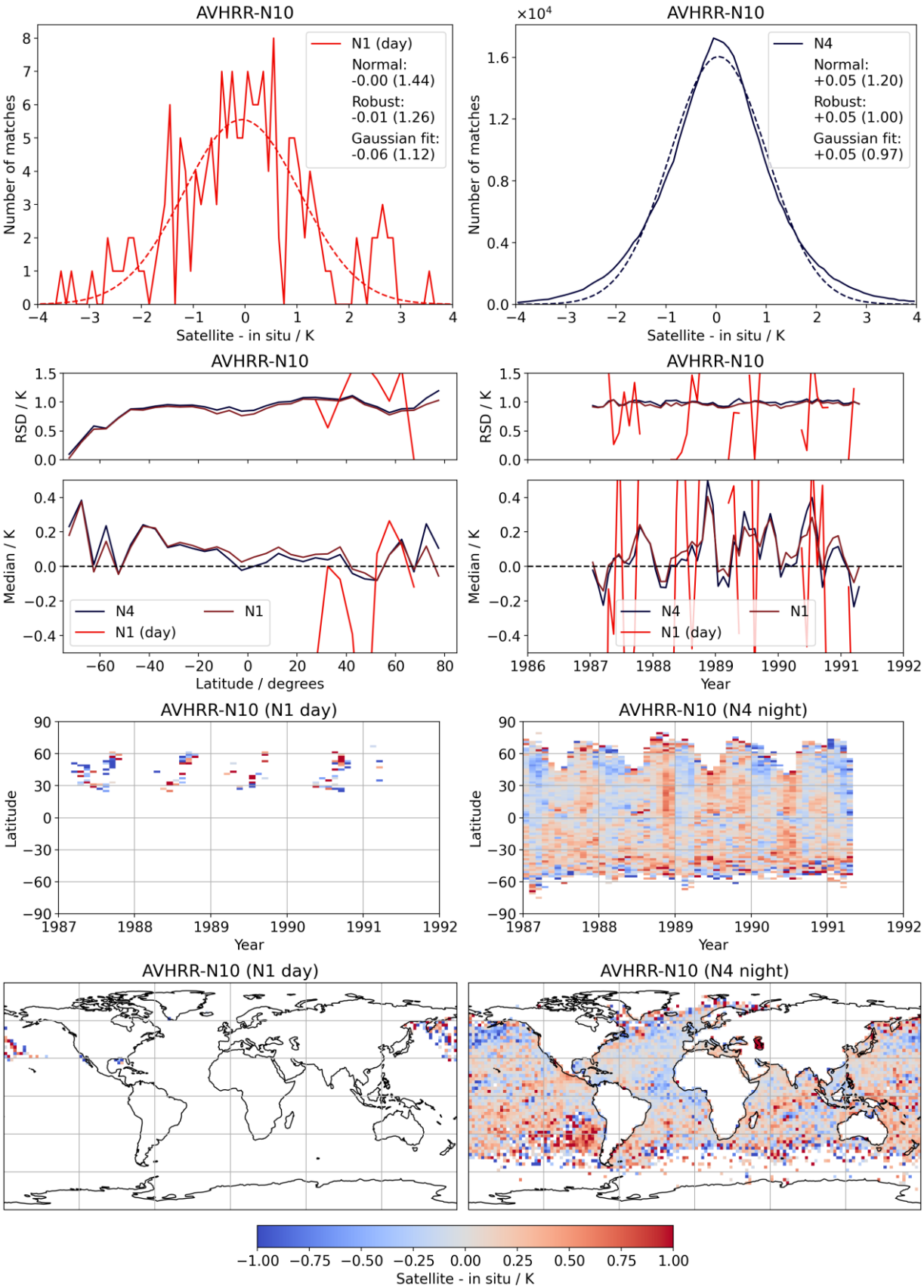


Figure 43: AVHRR-10 SST0.2m versus ships.

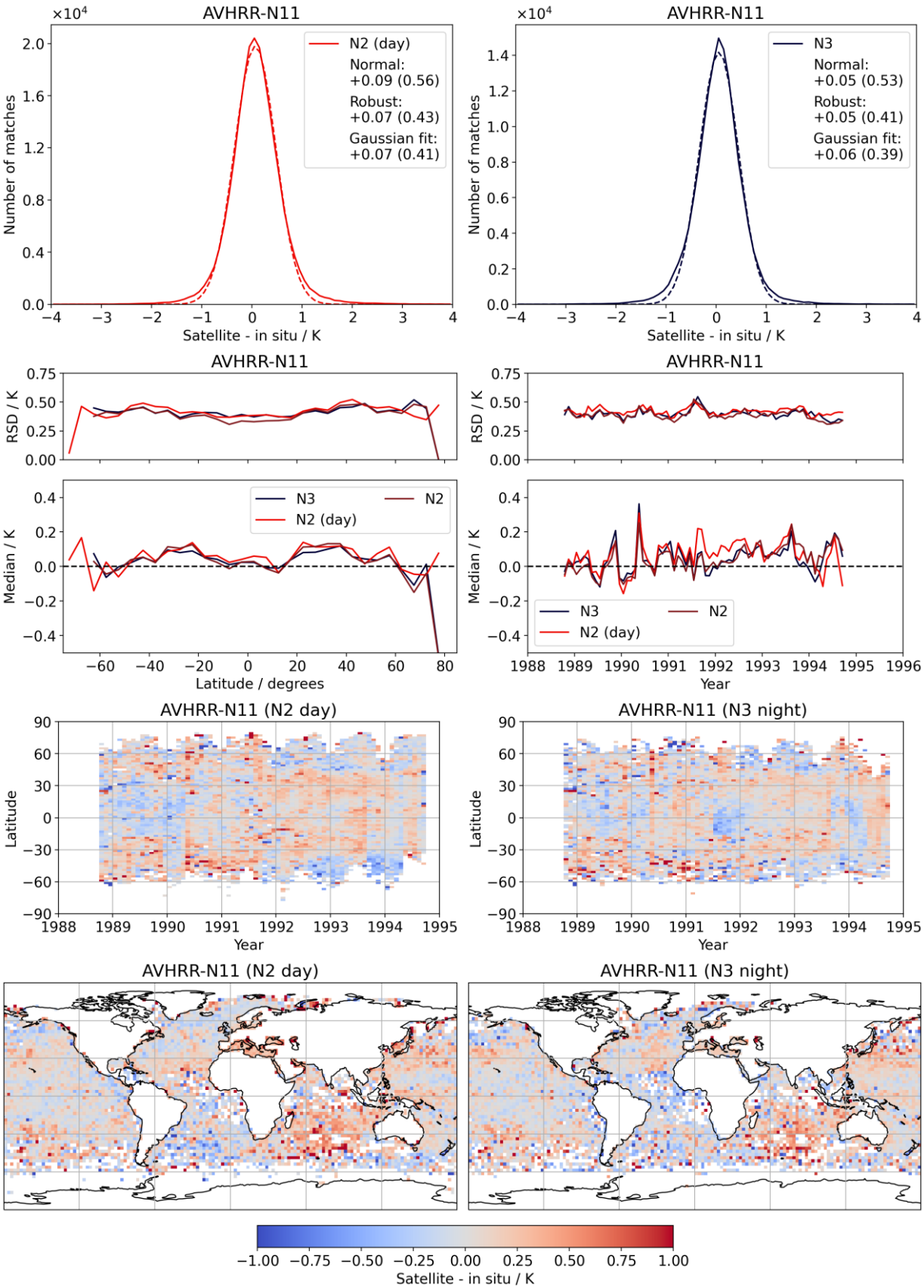


Figure 44: AVHRR-11 SST0.2m versus all non-ship in situ.

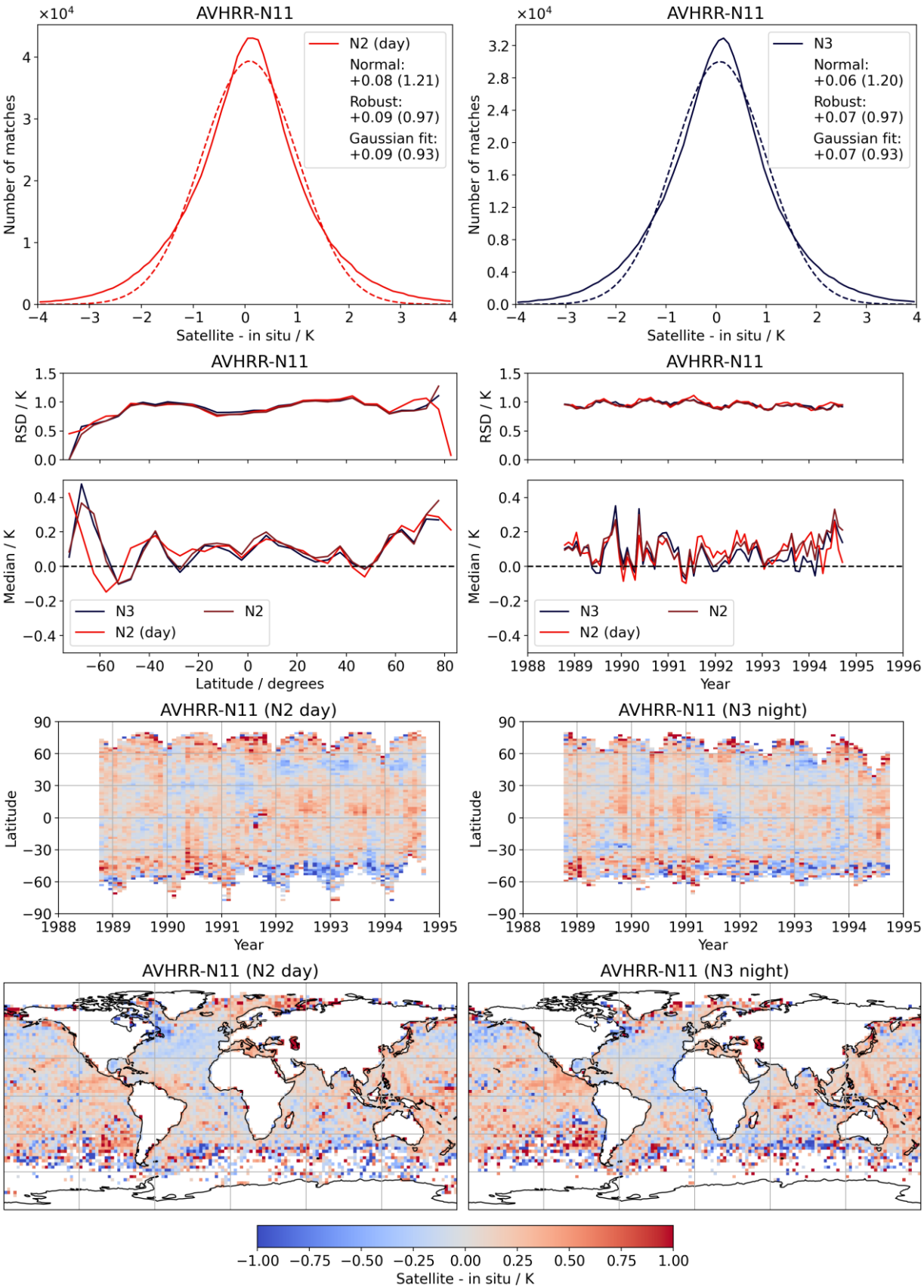


Figure 45: AVHRR-11 SST0.2m versus ships.

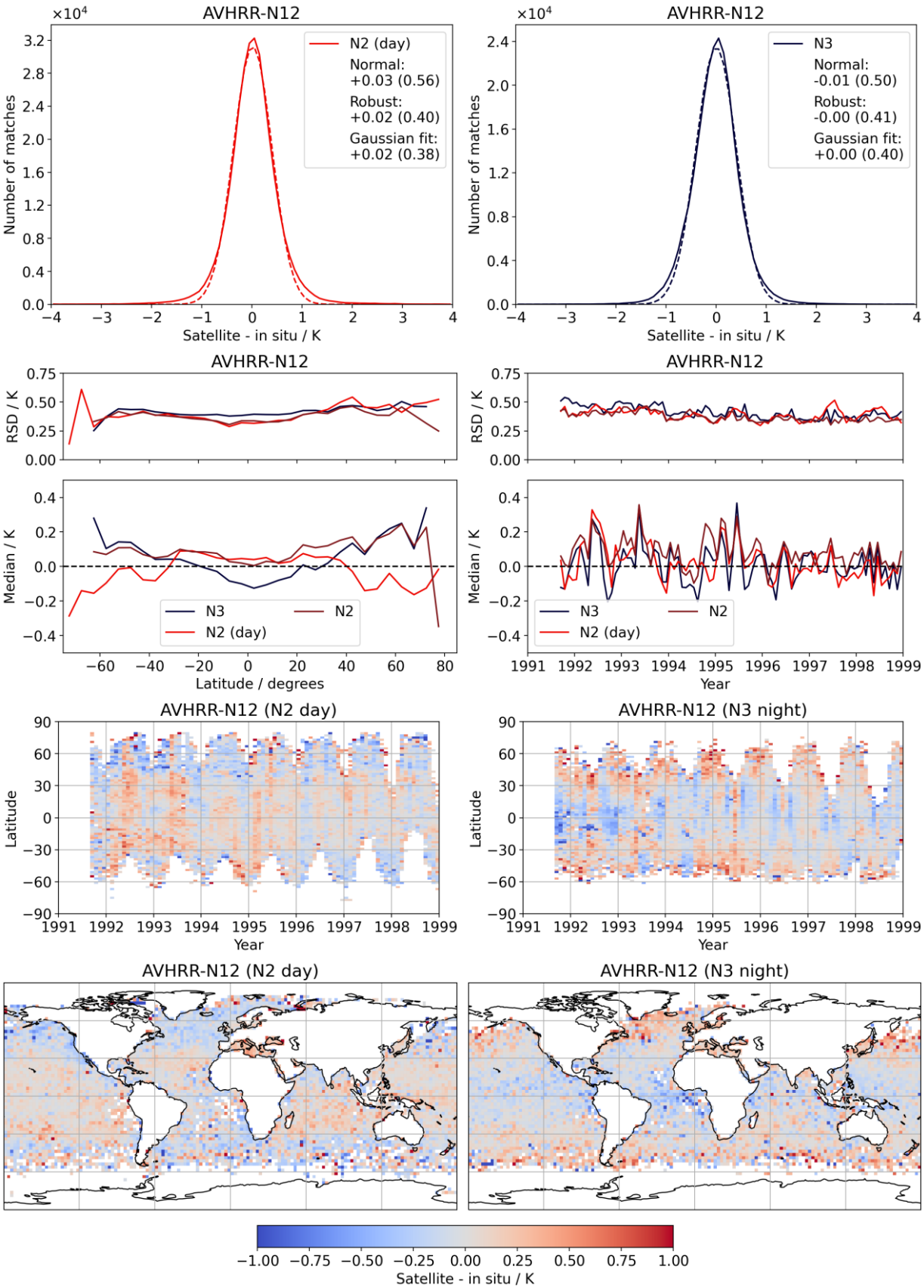


Figure 46: AVHRR-12 SST0.2m versus all non-ship in situ.

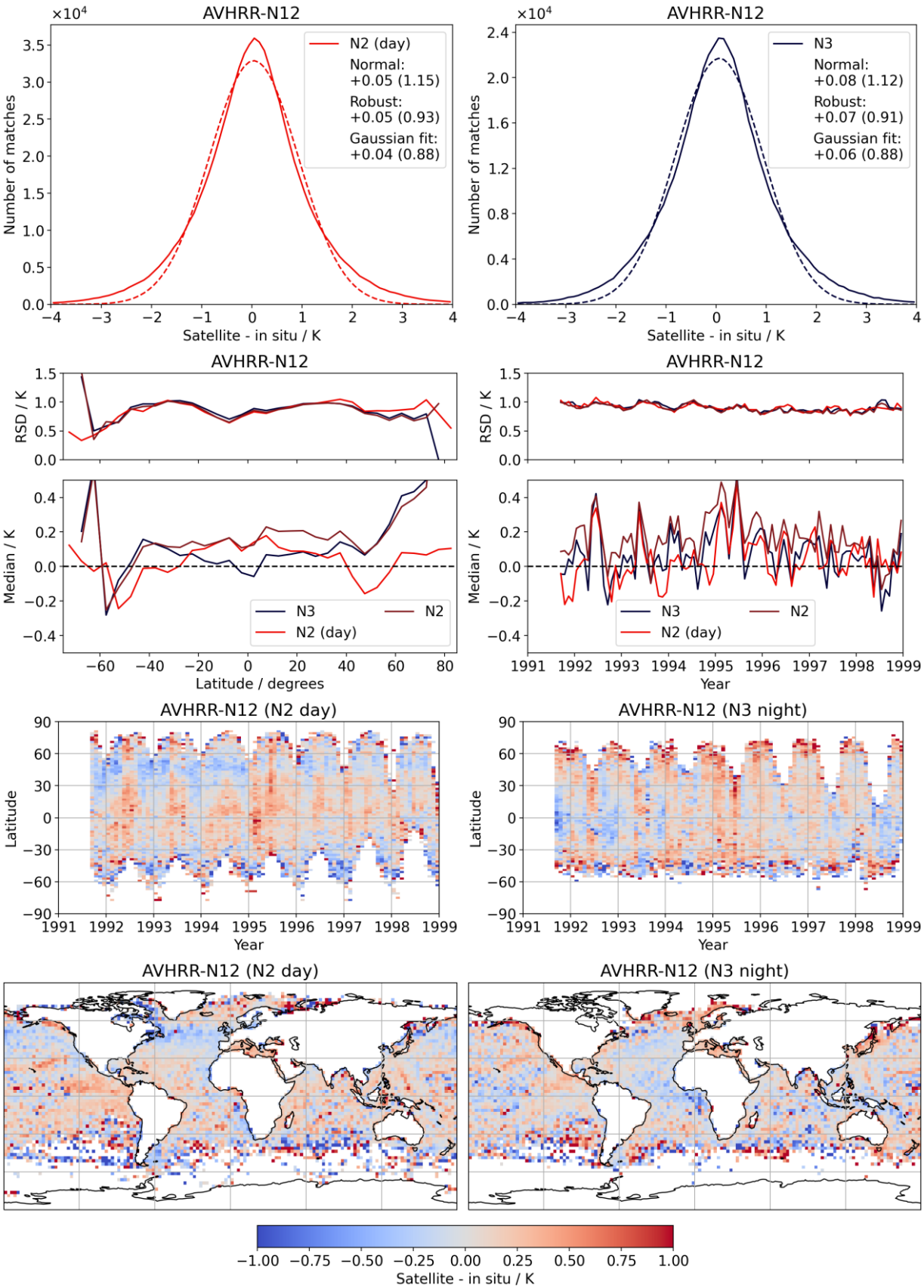


Figure 47: AVHRR-12 SST0.2m versus ships.

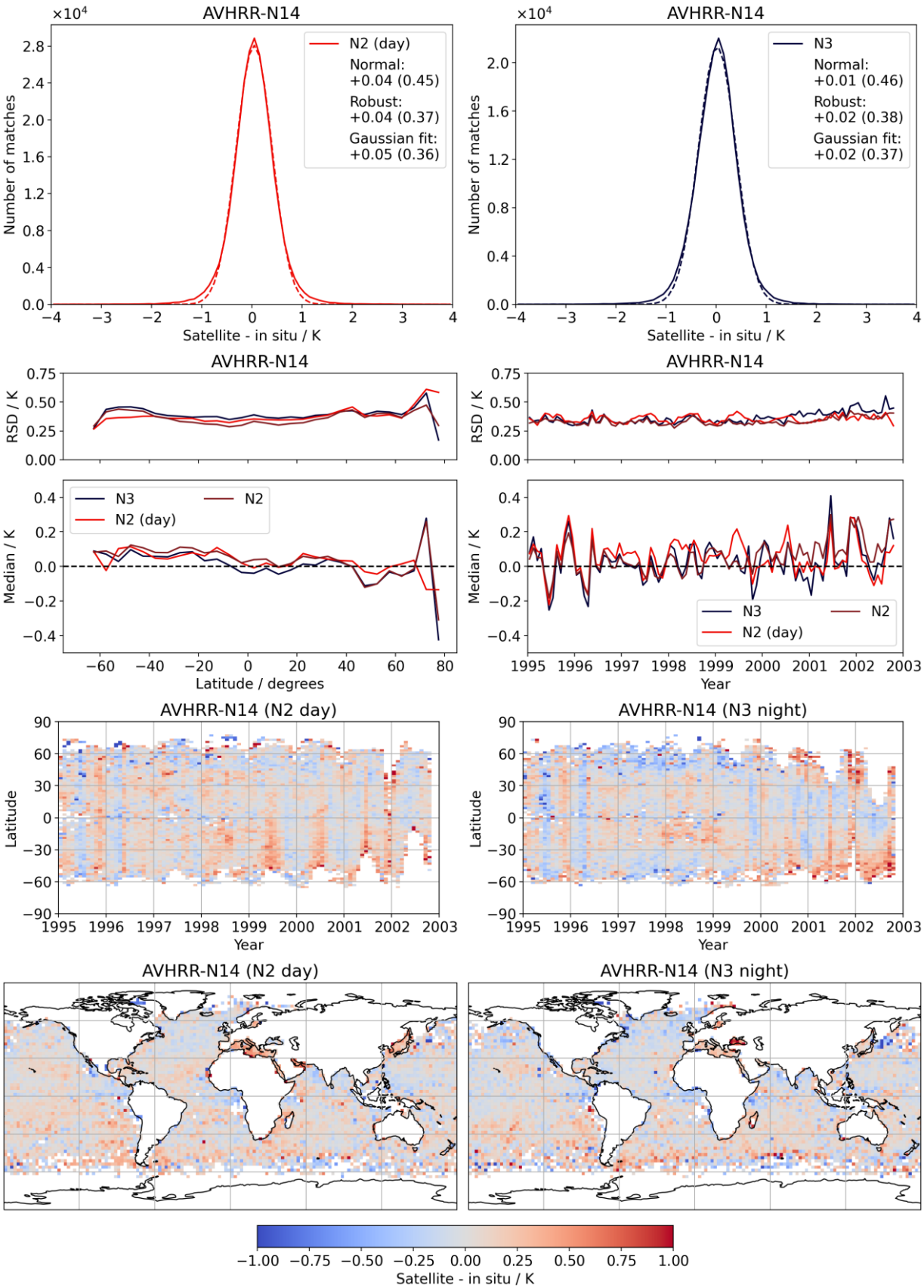


Figure 48: AVHRR-14 SST0.2m versus drifting buoys.

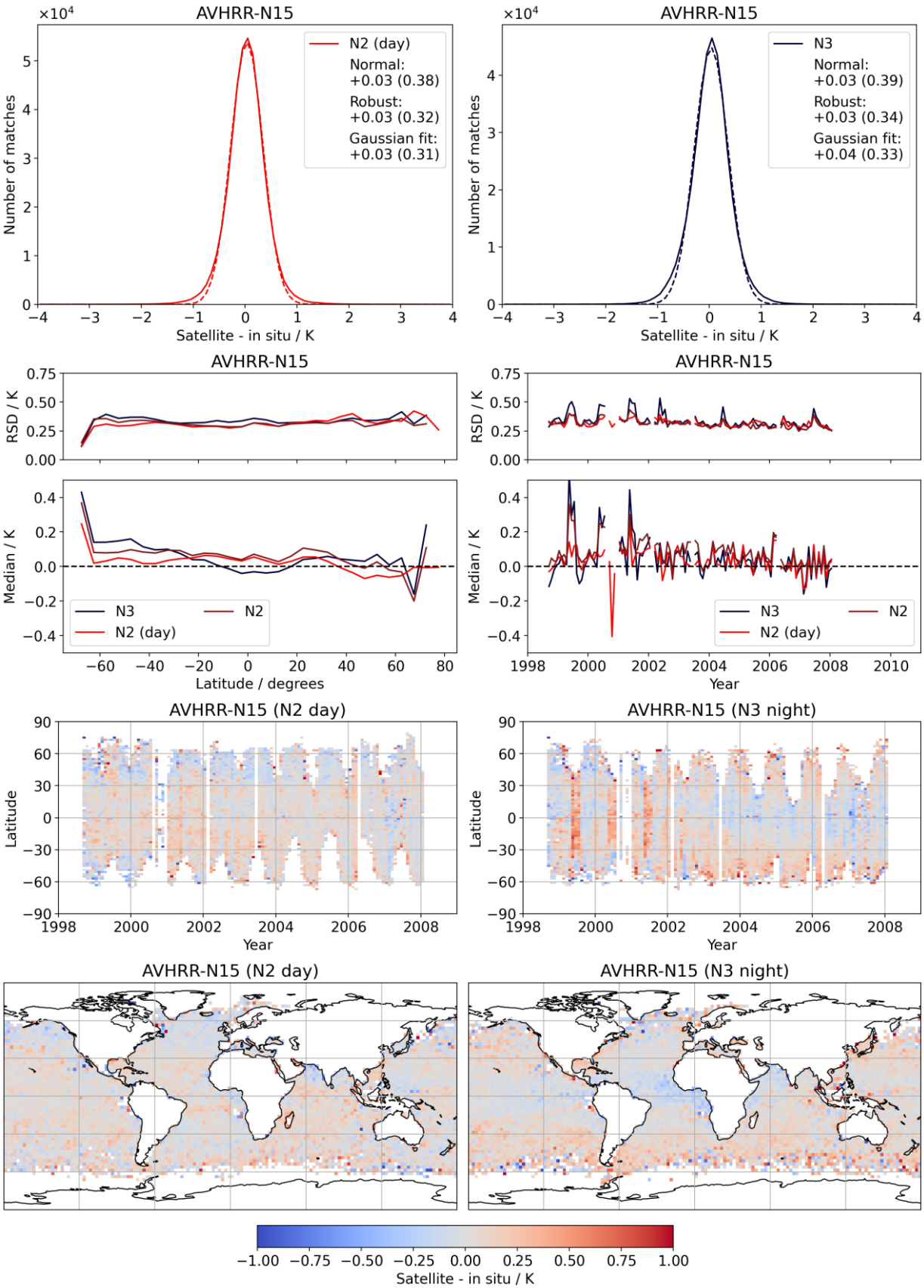


Figure 49: AVHRR-15 SST0.2m versus drifting buoys.

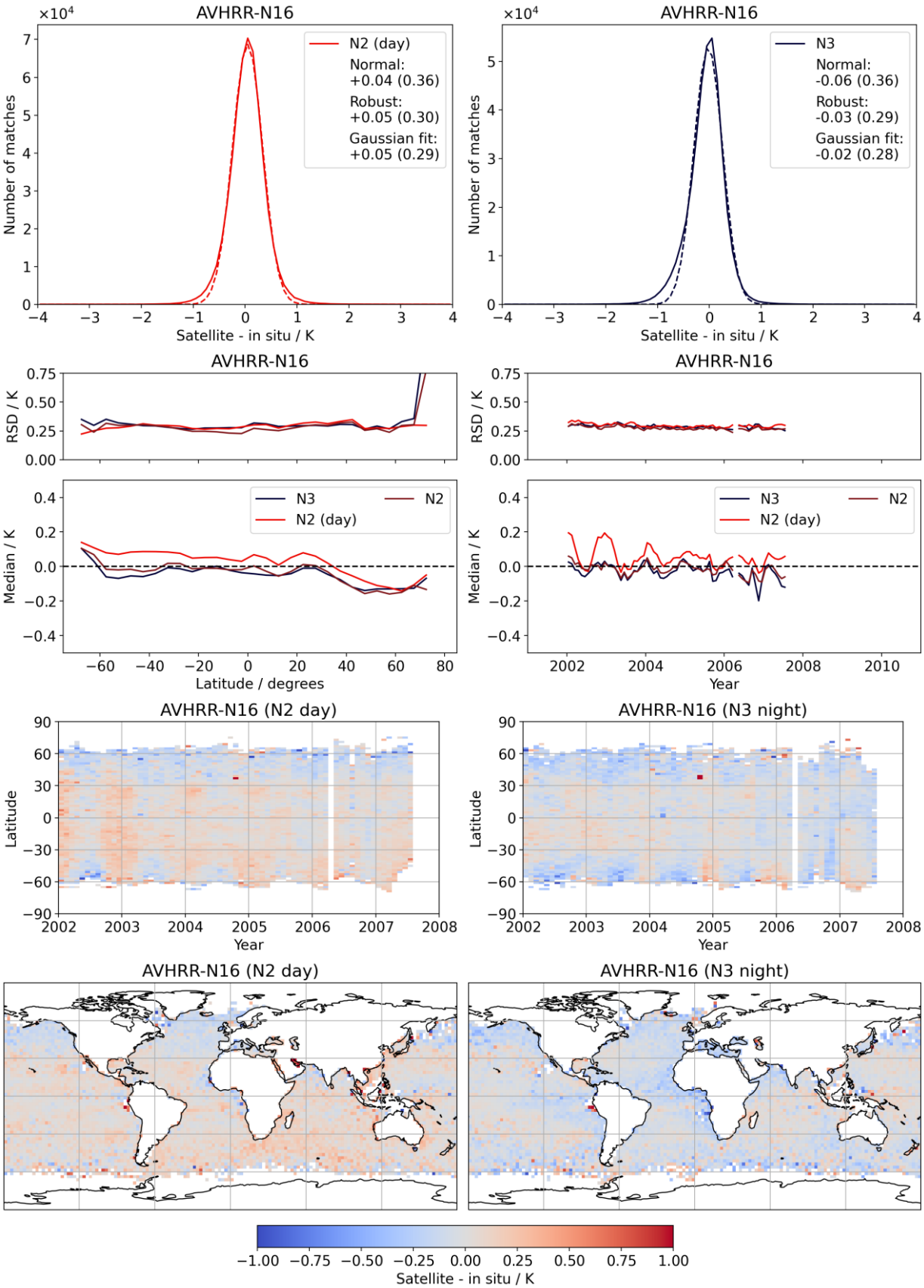


Figure 50: AVHRR-16 SST0.2m versus drifting buoys.

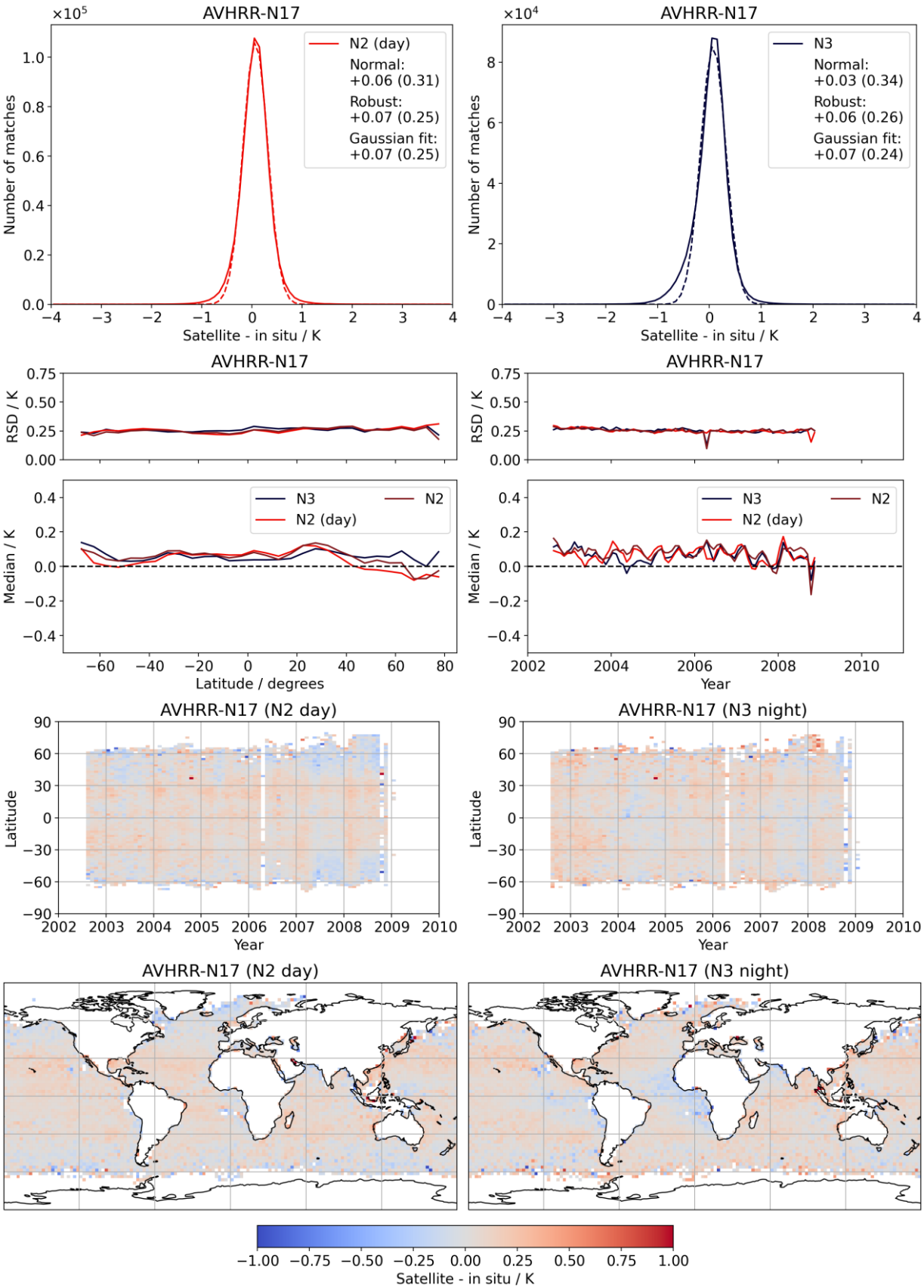


Figure 51: AVHRR-17 SST0.2m versus drifting buoys.

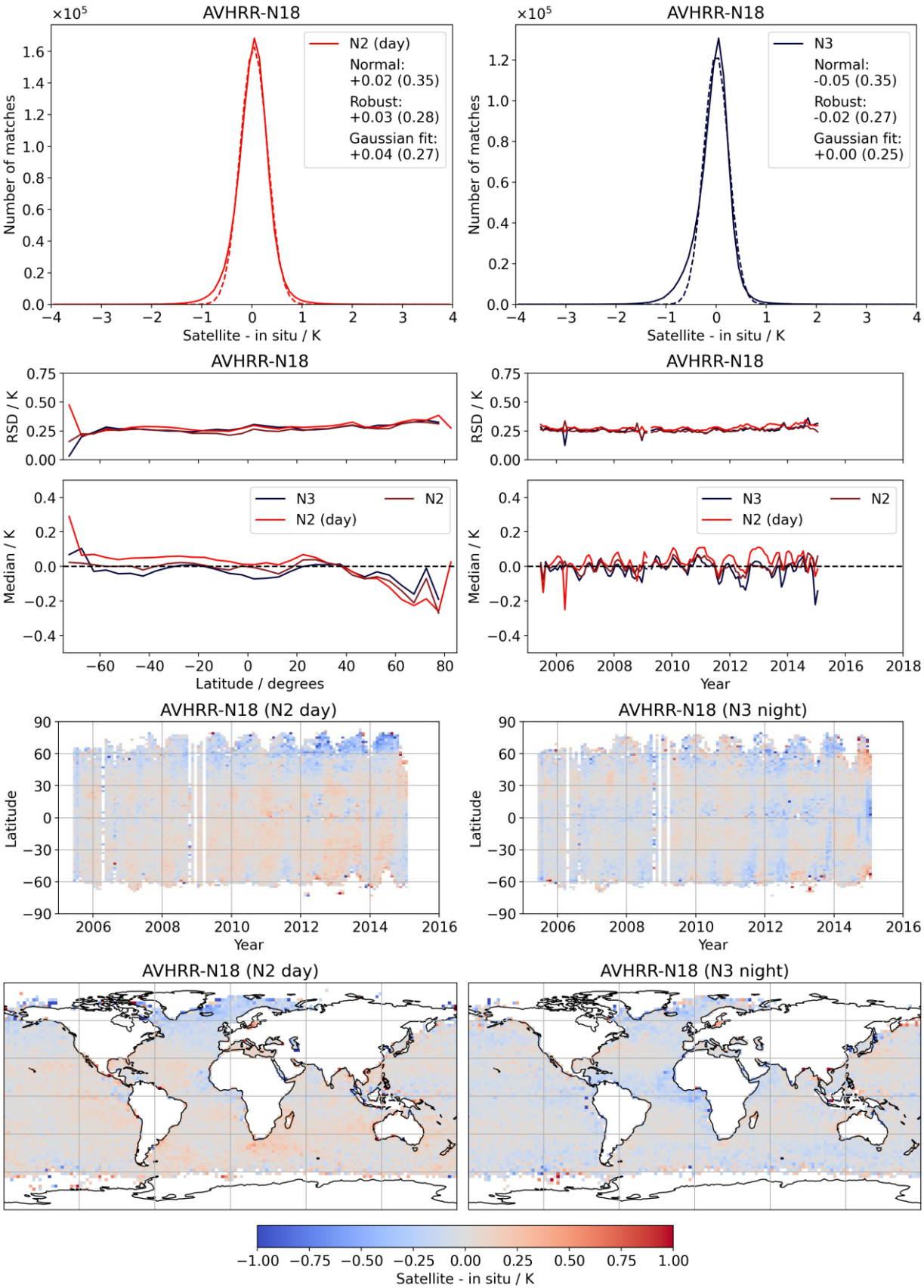


Figure 52: AVHRR-18 SST0.2m versus drifting buoys.

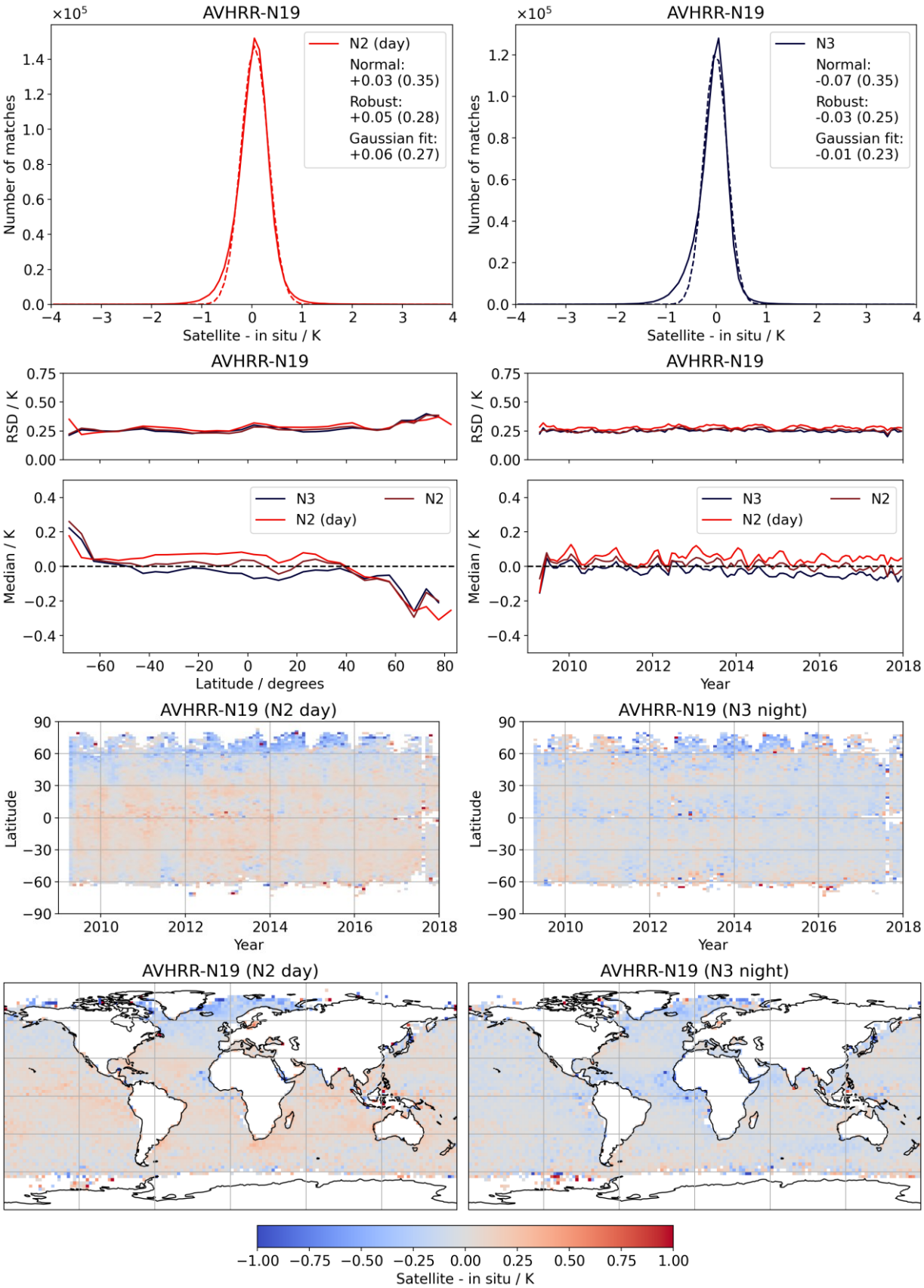


Figure 53: AVHRR-19 SST0.2m versus drifting buoys.

A.2. ATSR

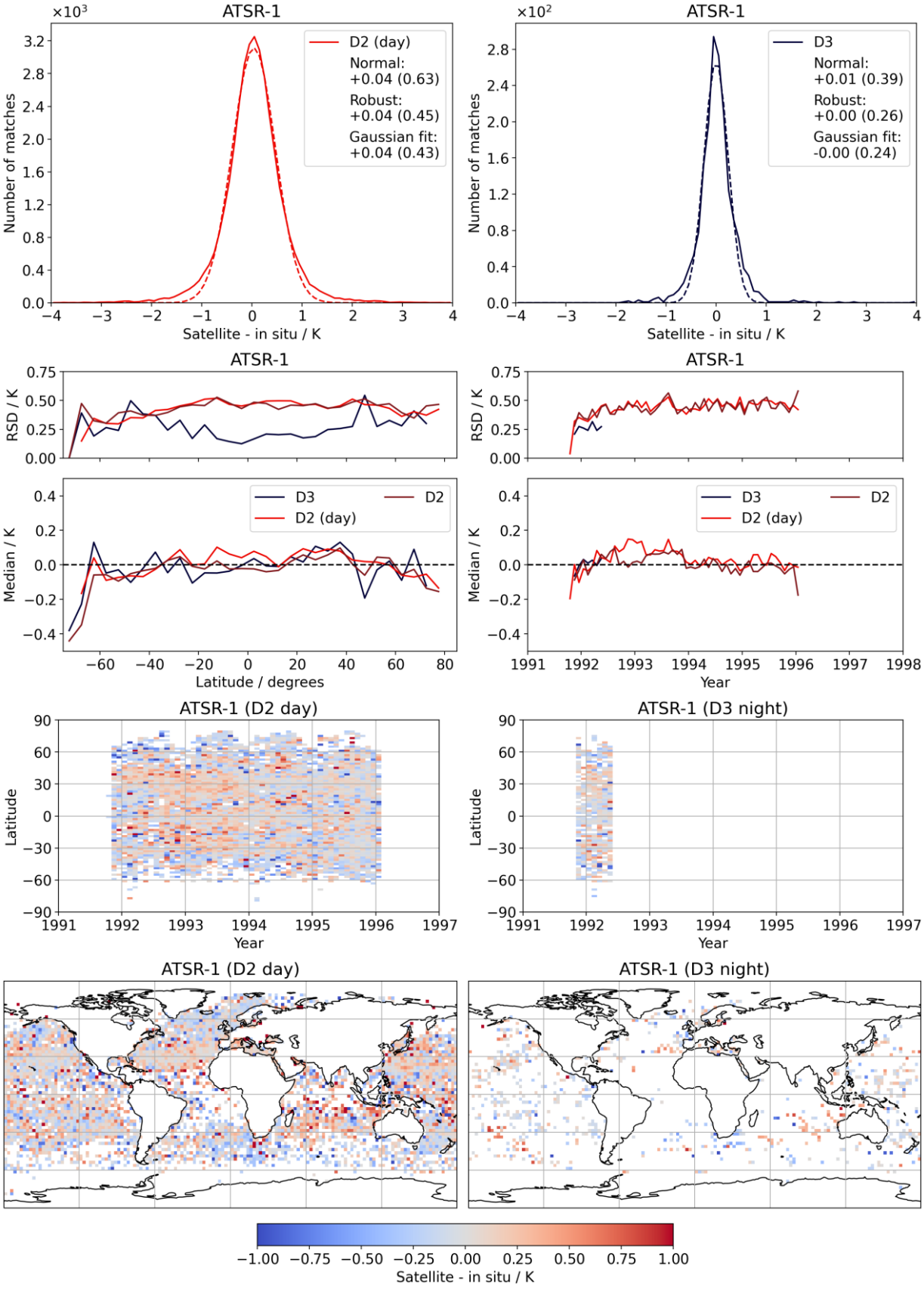

ATSR-1
×10³
Number of matches
Satellite - in situ / K
D2 (day)
Normal: +0.04 (0.63)
Robust: +0.04 (0.45)
Gaussian fit: +0.04 (0.43)
ATSR-1
×10²
Number of matches
Satellite - in situ / K
D3
Normal: +0.01 (0.39)
Robust: +0.00 (0.26)
Gaussian fit: -0.00 (0.24)
ATSR-1
RSD / K
ATSR-1
RSD / K
D3
D2
D2 (day)
Median / K
Latitude / degrees
D3
D2
D2 (day)
Median / K
Year
ATSR-1 (D2 day)
Latitude
Year
ATSR-1 (D3 night)
Latitude
Year
ATSR-1 (D2 day)
ATSR-1 (D3 night)
Satellite - in situ / K

Figure 54: ATSR-1 SST0.2m versus all non-ship in situ.

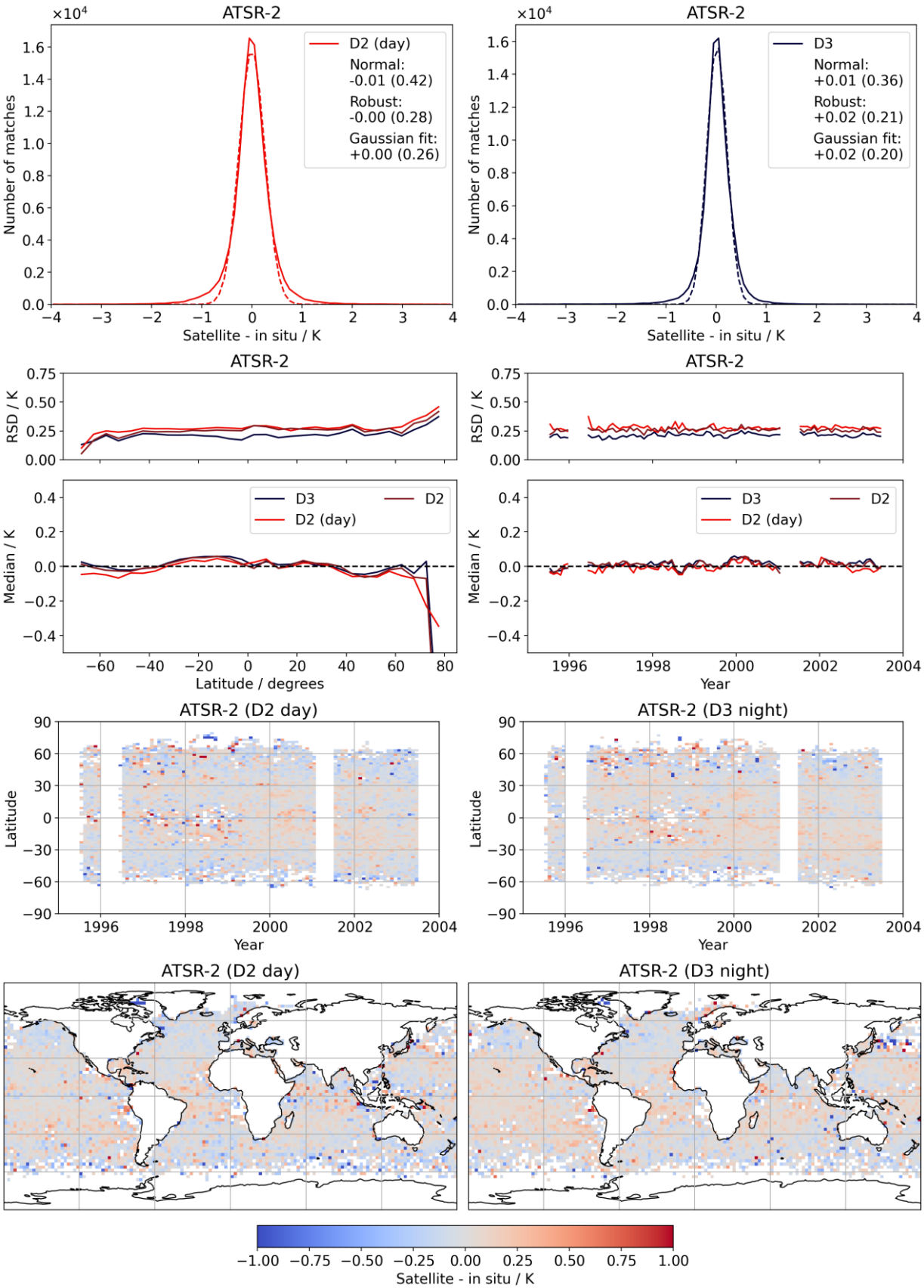


Figure 55: ATSR-2 SST0.2m versus drifting buoys.

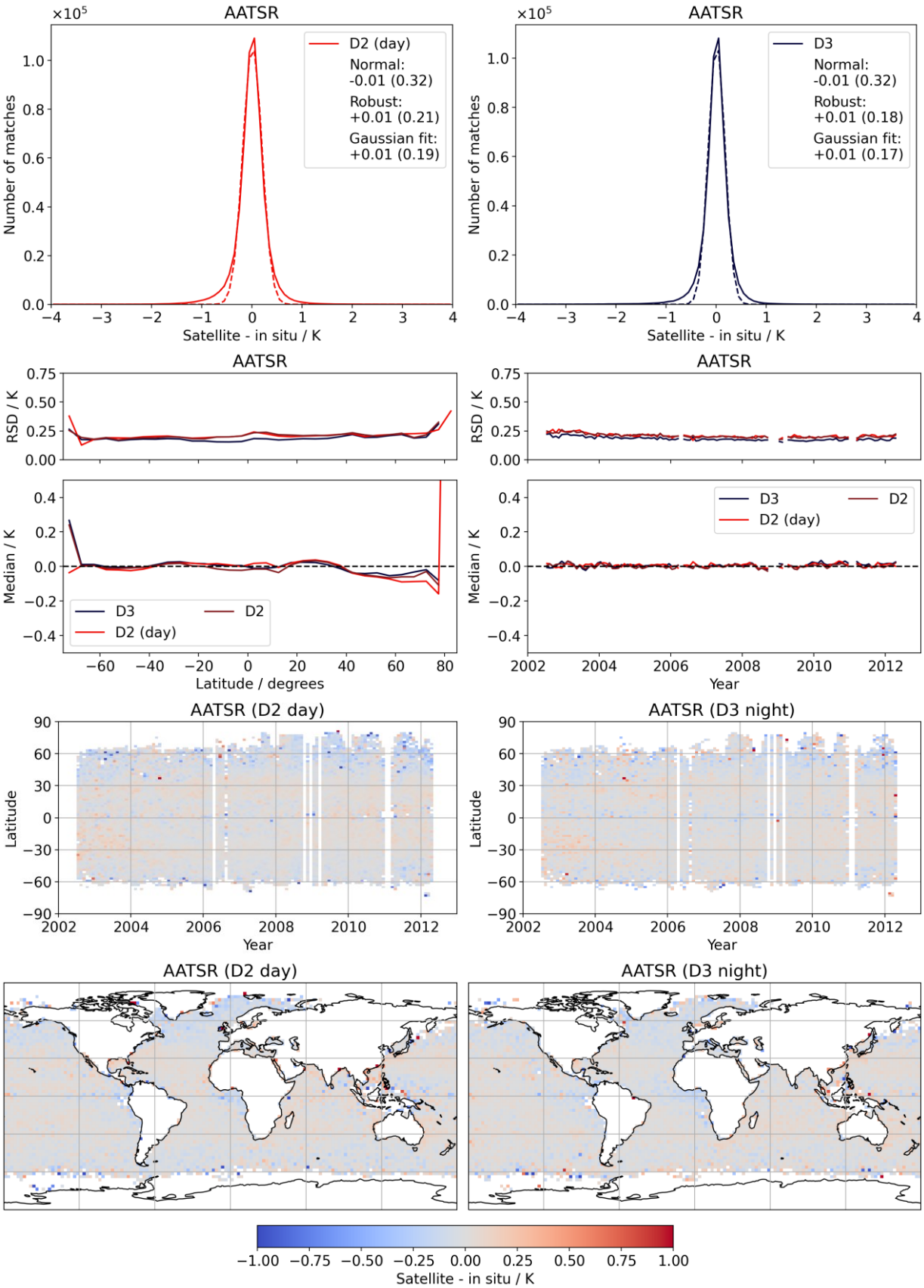


Figure 56: AATSR SST0.2m versus drifting buoys.

A.3. MetOp AVHRR

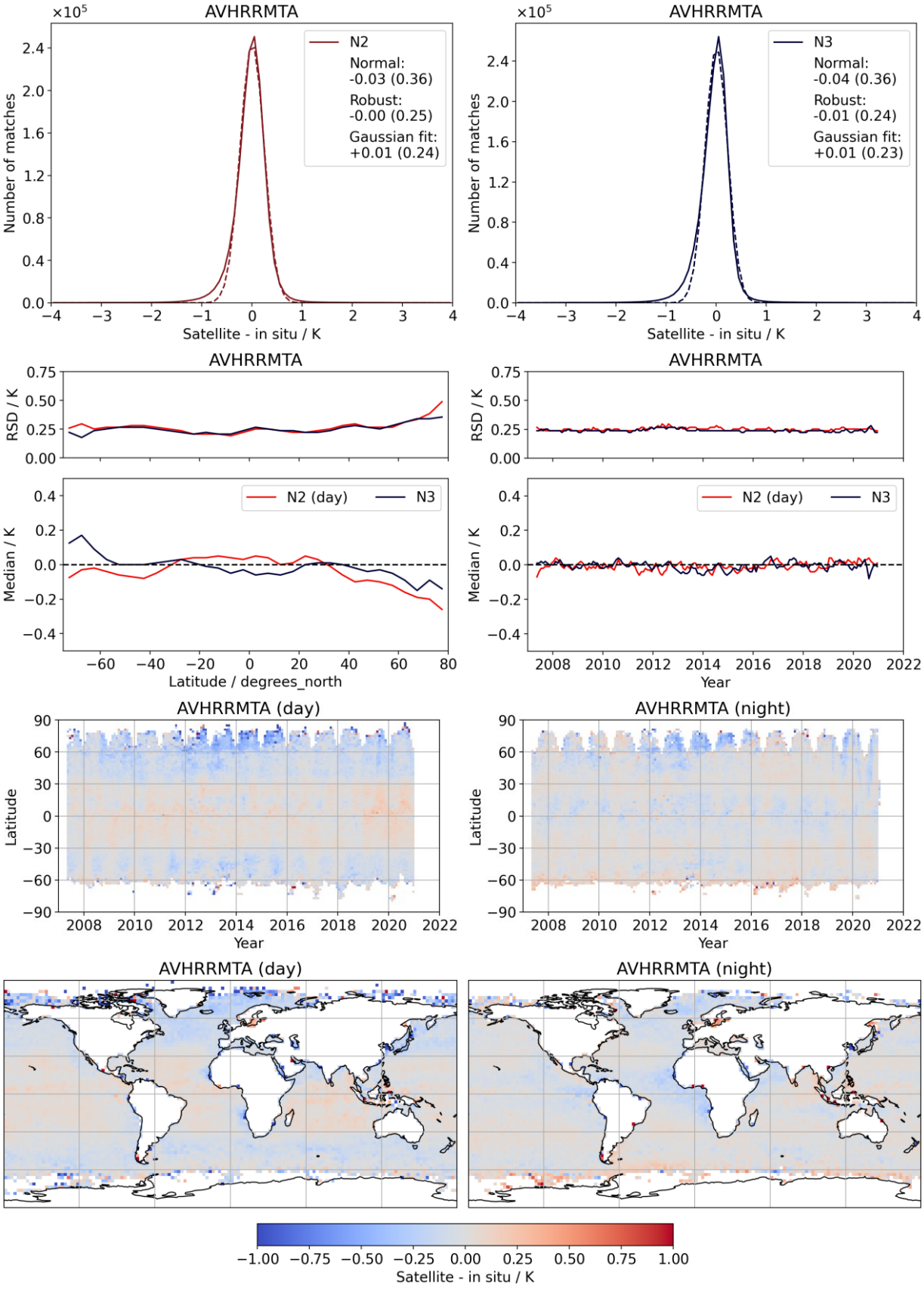


Figure 57: MetOp-A AVHRR SST0.2m versus drifting buoys.

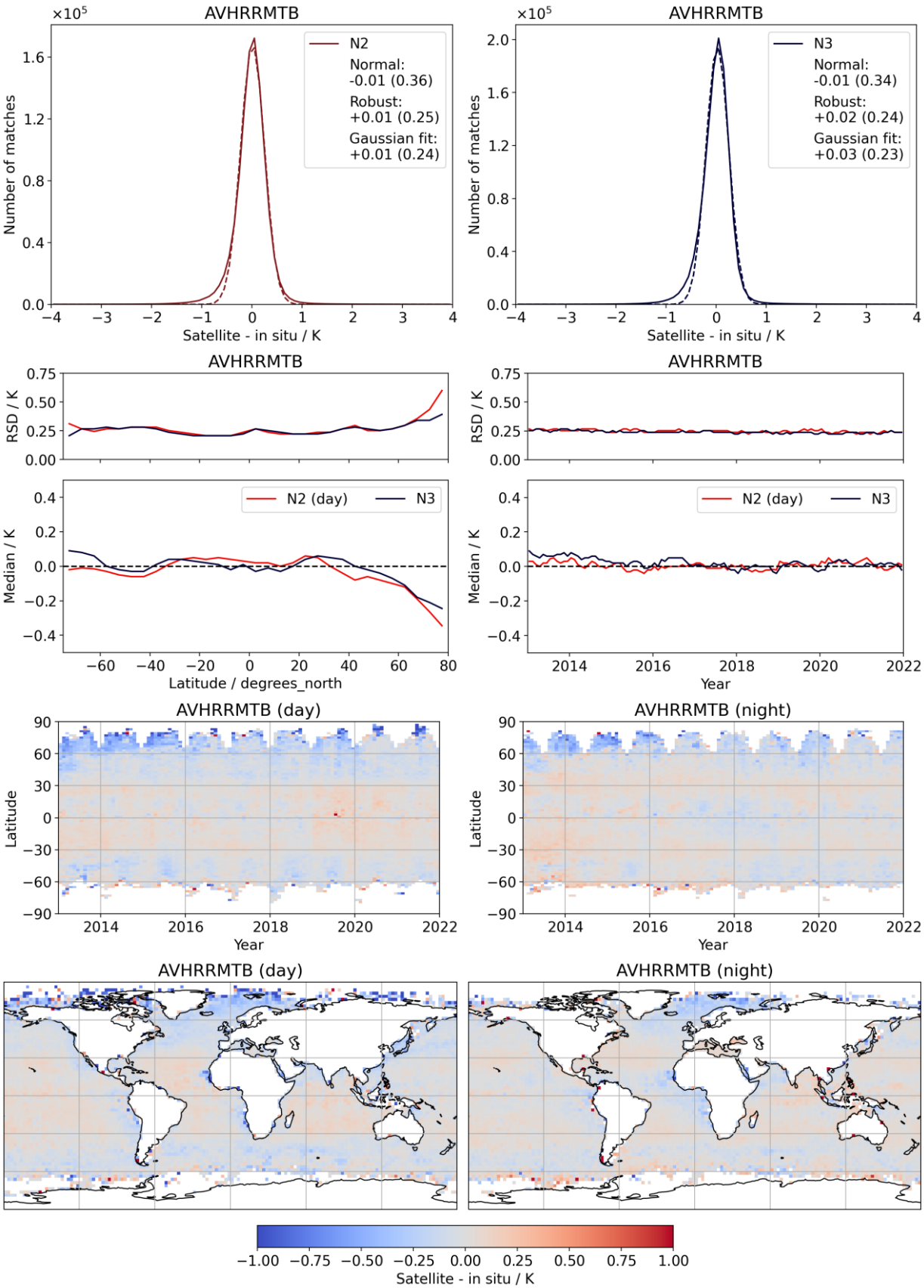


Figure 58: MetOp-B AVHRR SST0.2m versus drifting buoys.

A.4. SLSTR

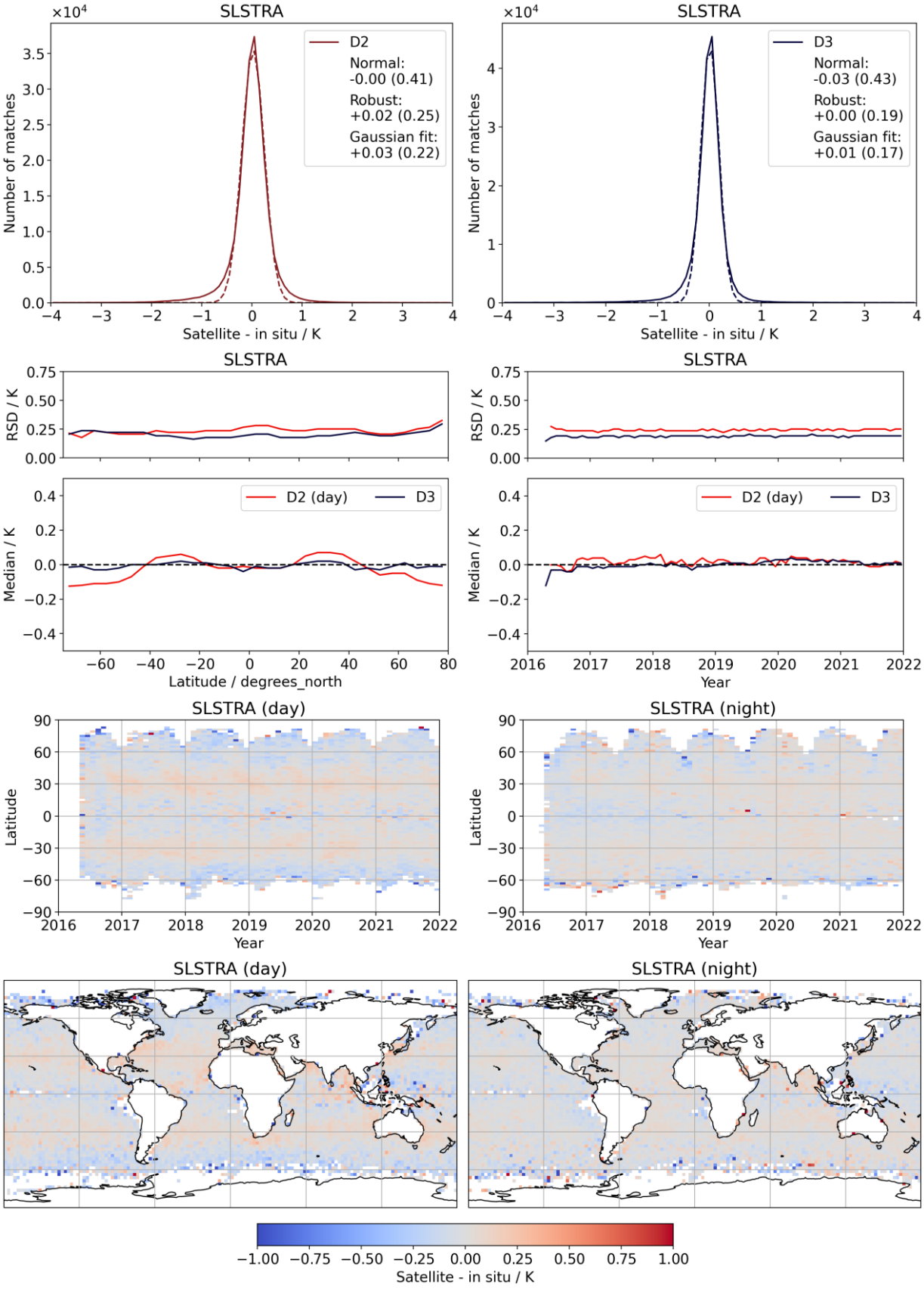

SLSTRA
D2
Normal: -0.00 (0.41)
Robust: +0.02 (0.25)
Gaussian fit: +0.03 (0.22)
Number of matches
Satellite - in situ / K
SLSTRA
D3
Normal: -0.03 (0.43)
Robust: +0.00 (0.19)
Gaussian fit: +0.01 (0.17)
Number of matches
Satellite - in situ / K
SLSTRA
RSD / K
SLSTRA
RSD / K
D2 (day)
D3
Median / K
Latitude / degrees_north
D2 (day)
D3
Median / K
Year
SLSTRA (day)
Latitude
Year
SLSTRA (night)
Latitude
Year
SLSTRA (day)
SLSTRA (night)
Satellite - in situ / K

Figure 59: SLSTR-A SST0.2m versus drifting buoys.

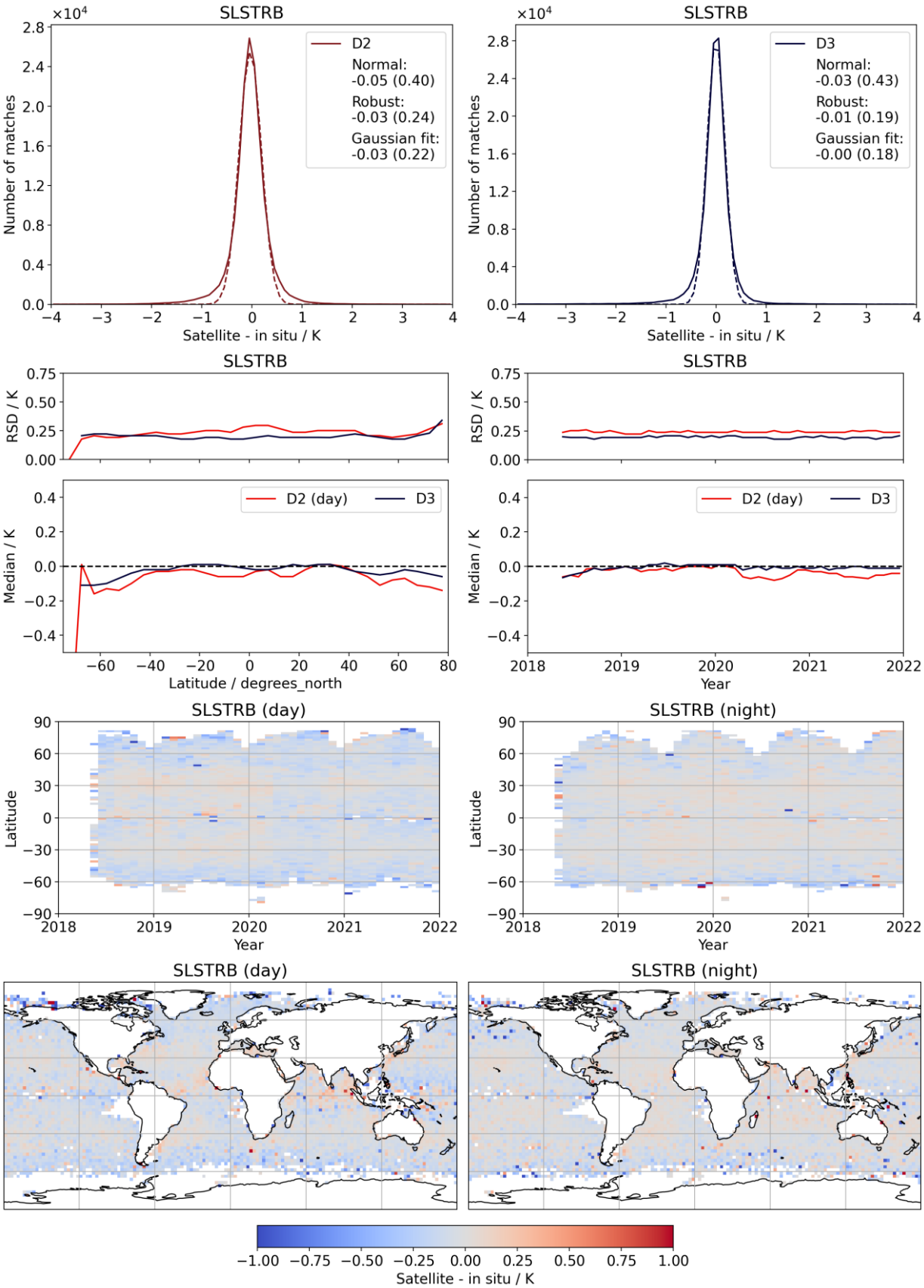


Figure 60: SLSTR-B SST0.2m versus drifting buoys.

A.5. Level 4

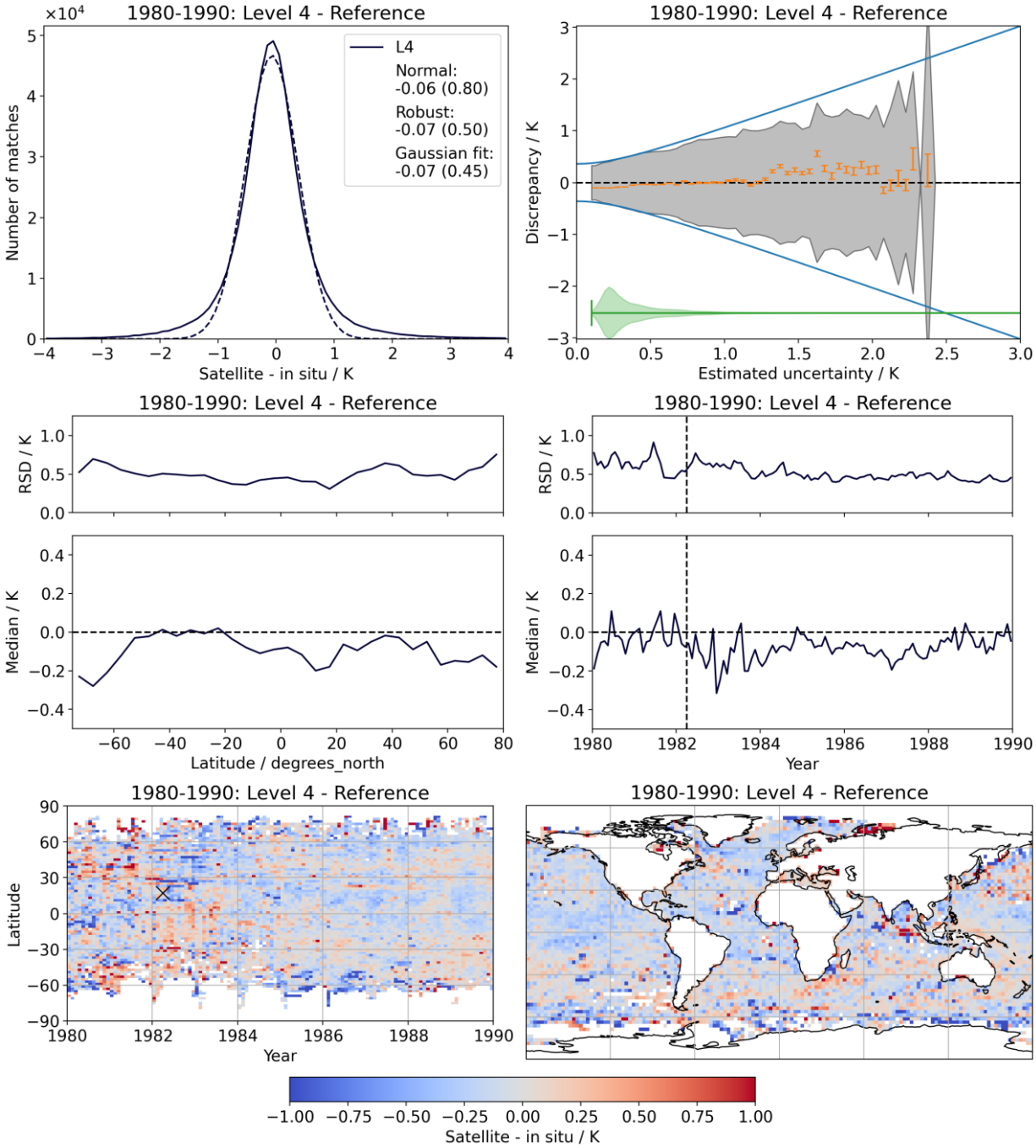


Figure 61: Level 4 versus reference in situ (1980 - 1990)

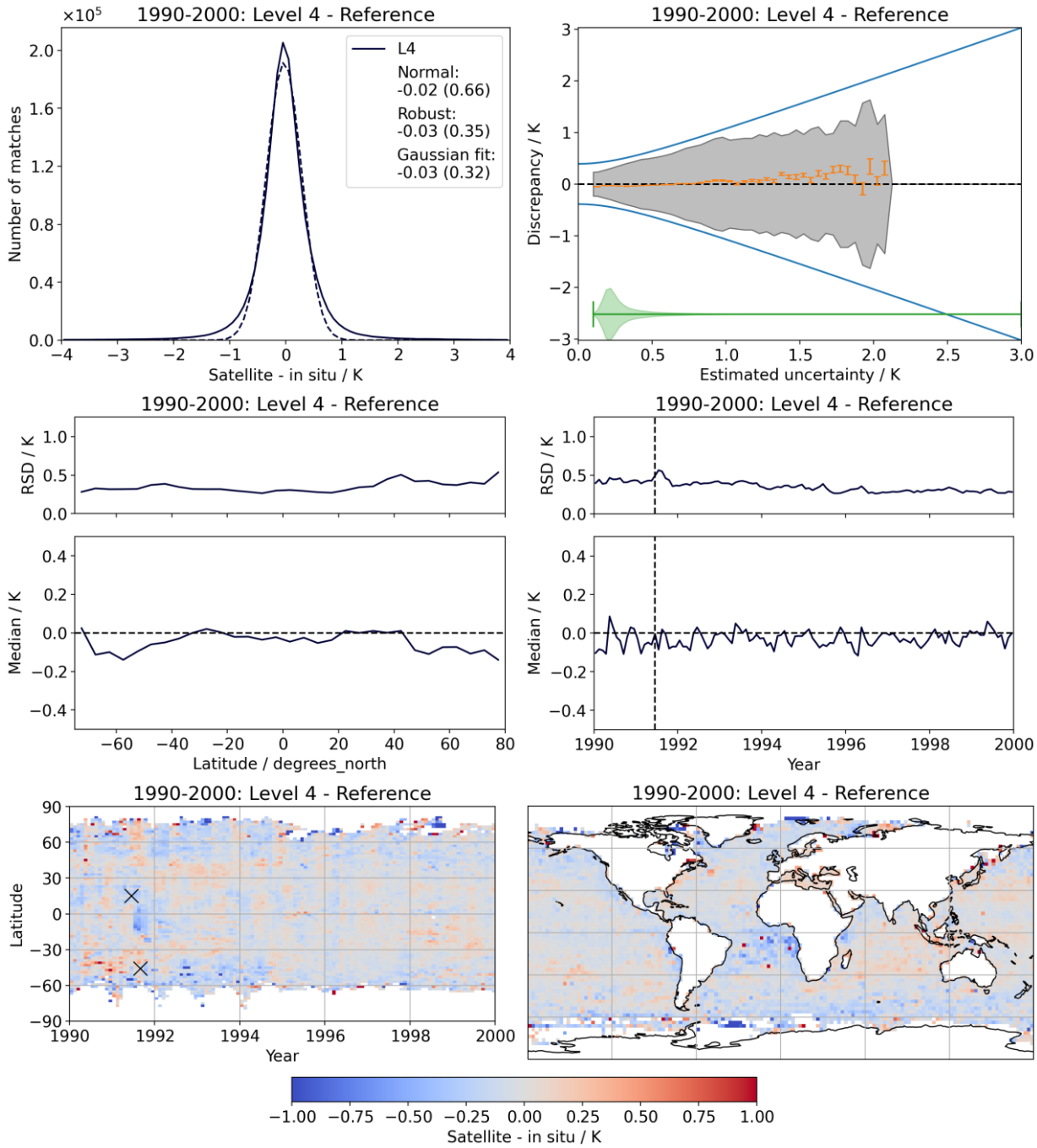


Figure 62: Level 4 versus reference in situ (1990 - 2000)

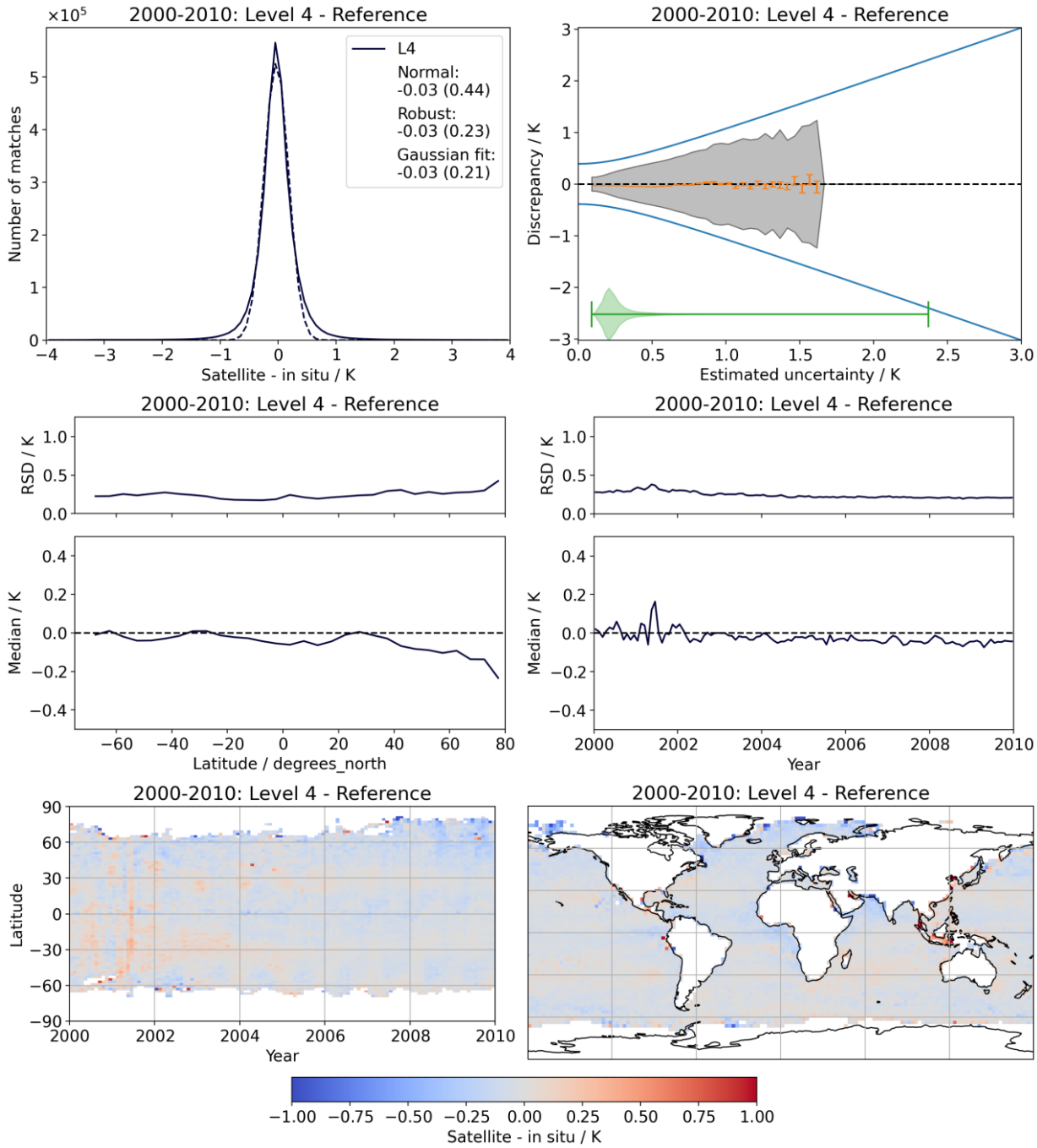


Figure 63: Level 4 versus reference in situ (2000 - 2010)

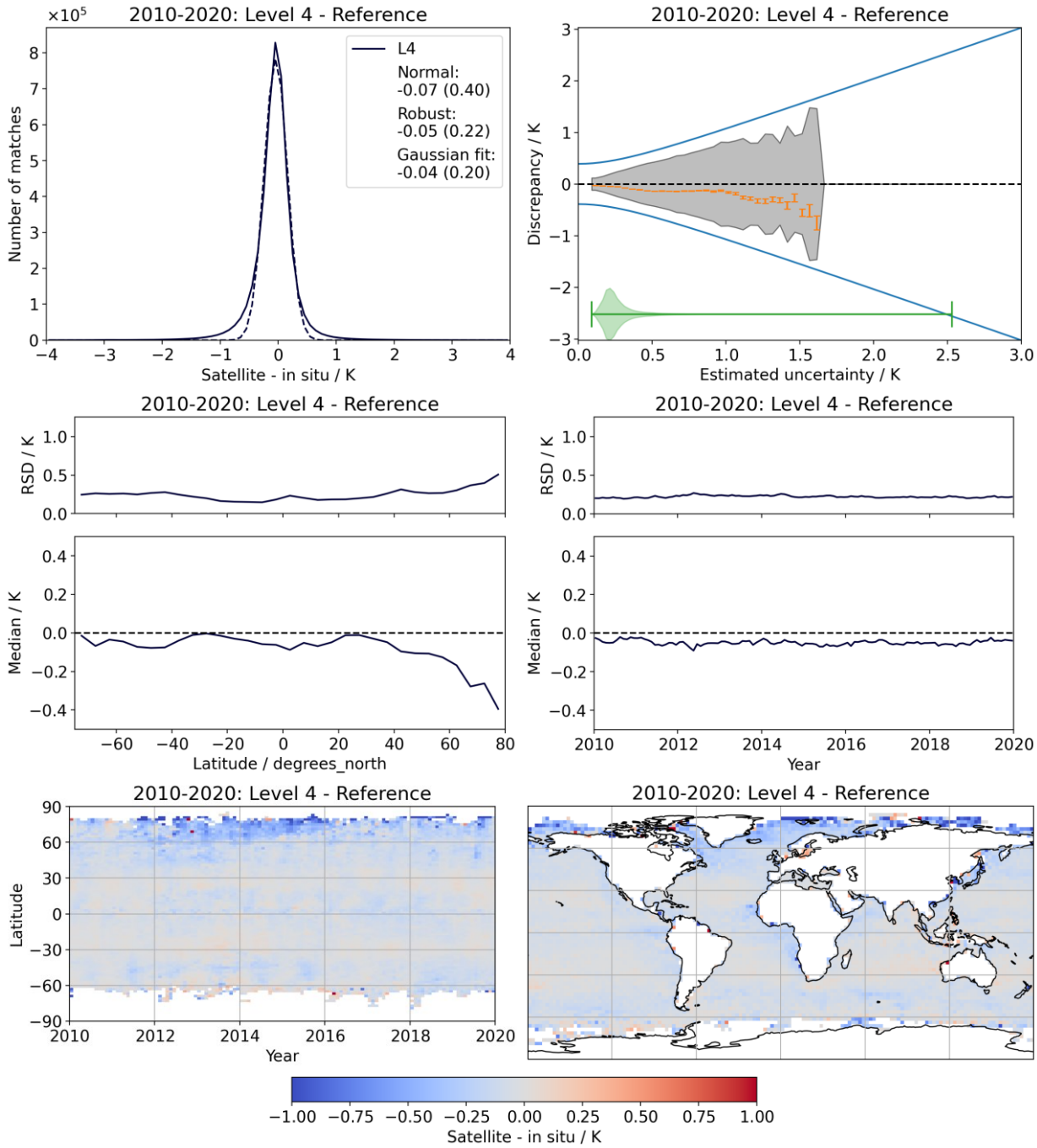


Figure 64: Level 4 versus reference in situ (2010 - 2020)