Customer	**:**	**ESRIN**	**Document Ref**	**:**	**SST_CCI-ASR-UOE-001**
WP No	**:**		**Issue Date**	**:**	**30 July 2012**
			Issue	**:**	**1**

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

Title : **Algorithm Selection Report (ASR)**

Abstract : **This document contains the Algorithm Selection Report (ASR) for the SST_cci project.**

Author :

Chris Merchant &
Stuart MacCallum,
University of Edinburgh

Checked by :

Paul Spinks
Project Manager
Space ConneXions

Accepted by ESA:

Craig Donlon
ESA Technical Officer

Acceptance date: 30 July 2012

Distribution : SST_cci team members

ESA (Craig Donlon)

EUROPEAN SPACE AGENCY
CONTRACT REPORT

The work described in this report was done under ESA contract. Responsibility for the contents resides in the author or organisation that prepared it.

sst cci

AMENDMENT RECORD

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

AMENDMENT RECORD SHEET

ISSUE	DATE	REASON FOR CHANGE
A	4 July 2012	Initial Draft
1	30 July 2012	Revised following comments from Craig Donlon

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY 5

2. INTRODUCTION 7
2.1 Purpose and Scope 8
2.2 Structure of the Document 8
2.3 Referenced Documents 8
2.4 Acronyms and abbreviations 9

3. DEFINITIONS 12

4. PROTOCOLS FOR ALGORITHM SELECTION (ROUND ROBIN) 13
4.1 Role and Scope of Algorithm Selection in SST CCI 13
4.1.1 Role and Definition of Potential Scope 13
4.1.1.1 SST estimation 13
4.1.1.2 SST uncertainty estimation 13
4.2 Selection criteria and process 14
4.2.1 Over-arching principles 14
4.2.2 Definition of selection metrics 14
4.2.2.1 Bias 15
4.2.2.2 Non-systematic uncertainty (precision) 16
4.2.2.3 Stability 16
4.2.2.4 Independence from in situ SST 17
4.2.2.5 SST sensitivity 17
4.2.2.6 Generality 17
4.2.2.7 Improvability 17
4.2.2.8 Implementation considerations 18
4.2.3 Selection process 18
4.3 Tools 19

5. INFORMATION FOR ALGORITHM SELECTION 20
5.1 Guide to interpretation of comparison tables 20
5.2 Side-by-side comparison of results relating to AATSR 28
5.2.1 Presentation of metrics related to SST estimation for AATSR 28
5.2.2 Presentation of metrics related to SST uncertainty for AATSR 38
5.2.3 Comments on results for AATSR 39
5.3 Side-by-side comparison of results relating to ATSR-2 41
5.3.1 Presentation of metrics related to SST estimation for ATSR-2 41
5.3.2 Presentation of metrics related to SST uncertainty for ATSR-2 51
5.3.3 Comments on results for ATSR-2 52
5.4 Side-by-side comparison of results relating to Metop 53
5.4.1 Presentation of metrics related to SST estimation for Metop 54
5.4.2 Presentation of metrics related to SST uncertainty for Metop 69
5.4.3 Comments on results for Metop 70
5.5 Side-by-side comparison of results relating to NOAA-19 72
5.5.1 Presentation of metrics related to SST estimation for NOAA-19 72
5.5.2 Presentation of metrics related to SST uncertainty for NOAA-19 87
5.5.3 Comments on results for NOAA-19 88
5.6 Side-by-side comparison of results relating to NOAA-18 92
5.6.1 Presentation of metrics related to SST estimation for NOAA-18 92
5.6.2 Presentation of metrics related to SST uncertainty for NOAA-18 107
5.6.3 Comments on results for NOAA-18 108
5.7 Side-by-side comparison of results relating to NOAA-17 110
5.7.1 Presentation of metrics related to SST estimation for NOAA-17 110
5.7.2 Presentation of metrics related to SST uncertainty for NOAA-17 125
5.7.3 Comments on results for NOAA-17 126

6. ALGORITHM SELECTION 128
6.1 Summary of highlights from metrics 128
6.2 Analysis of strengths and weakness of algorithms 131
6.3 Conclusions on selection of algorithms 136

APPENDIX OF FULL RESULTS 137

A1. ALONG TRACK SCANNING RADIOMETER RESULTS 137
A1.1 Brief account of relevant instrument characteristics 137
A1.2 List of algorithms 137
A1.3 Metrics for ARC SST coefficient retrievals 138
A1.3.1 Bias 138
A1.3.2 Precision 142
A1.3.3 Sensitivity 146
A1.4 Metrics for CCI SST Optimal Estimation v1 148
A1.4.1 Bias 148
A1.4.2 Precision 152
A1.4.3 Sensitivity 156
A1.5 Metrics for CCI SST Optimal Estimation v2 158
A1.5.1 Bias 158
A1.5.2 Precision 162
A1.5.3 Sensitivity 166
A1.6 Metrics for Oxford-RAL SST Estimation 168
A1.6.1 Bias 168
A1.6.2 Precision 170

A2. ADVANCED VERY HIGH RESOLUTION RADIOMETER RESULTS 171
A2.1 Brief account of relevant instrument characteristics 171
A2.2 List of algorithms 172
A2.3 Metrics for CCI SST Optimal Estimation v1 172
A2.3.1 Bias 172
A2.3.2 Precision 180
A2.3.3 Sensitivity 188
A2.4 Metrics for CCI SST Optimal Estimation v2 192
A2.4.1 Bias 192
A2.4.2 Precision 200
A2.4.3 Sensitivity 208
A2.5 Metrics for Incremental Regression Estimation 212
A2.5.1 Bias 212
A2.5.2 Precision 228
A2.5.3 Sensitivity 244

A3. AVHRR RESULTS WITH SOLAR ZENITH ANGLE FILTERING APPLIED 252
A3.1 Motivation for additional analysis 252
A3.2 Summary Tables 252
A3.2.1 Metop 253
A3.2.2 NOAA-19 257
A3.2.3 NOAA-18 261
A3.2.4 NOAA-17 265
A3.3 Full Resolution Images 269
A3.3.1 Metop 269
A3.3.2 NOAA-19 278
A3.3.3 NOAA-18 287
A3.3.4 NOAA17 296

1. EXECUTIVE SUMMARY

Within the Sea Surface Temperature Climate Change Initiative (SST CCI) project, an algorithm selection exercise has been undertaken, to select the SST retrieval and uncertainty estimation algorithms to be used in SST CCI products. The conclusion is that Optimal Estimation (OE) will be the primary algorithm for the sensors included in the 20 year SST time series that SST CCI will generate. The formulation of OE chosen is one in which SSTs from Along Track Scanning Radiometers are used as a reference in order to preserve independence of the satellite time series from in situ observations, while delivering excellent bias characteristics (generally within 0.1 K in comparison with drifting buoys) and high levels of stability and sensitivity.

The algorithm selection approach has been extensive and rigorous. The data package for the exercise covered six selected sensors (ATSR2, AATSR, Metop AVHRR, NOAA 17 AVHRR, NOAA 18 AVHRR and NOAA 19 AVHRR). The data included all necessary satellite data, auxiliary data such as numerical weather prediction fields and validation data. All algorithm developers including those in the SST CCI team were given access to a common subset of data designated for training and testing of algorithms. Algorithm developers were given another subset of data designated the "selection dataset", with all information necessary to retrieve SST, but without validation data. SSTs for selection were therefore submitted "blind" to the validation results for the selection subset, giving maximum possible objectivity to the selection process. Metrics for comparing algorithm performance were defined in the SST CCI Product Validation Plan, prior to the algorithm selection procedure. The metrics define measures of important aspects of an SST climate data record, which are (in roughly decreasing order of weight): bias (global, regional, high latitude and coastal), stability (long-term, seasonal and diurnal), sensitivity to SST variability, precision (noisiness of SST), degree of independence of in situ observations, validity of associated uncertainty estimates, improvability, generality, and implementation issues. Metrics were calculated by SST CCI team members not involved in algorithm development, which, together with the predefinition of the metrics, again maximises objectivity in the selection process.

The selection process has been open and transparent. Ten external teams expressed interest in participation, of which two were able in practice to submit algorithm selection results in time for consideration. The submitted external algorithms were cutting edge algorithms of significant interest. Relevant to ATSRs was the Oxford-RAL Aerosol and Cloud retrieval (ORAC, submitted by Caroline Cox of RAL), an advanced optimal estimator recently extended to include SST, although only applicable to day time scenes. Relevant to AVHRRs was Incremental Regression (IR, submitted by Boris Petrenko of NOAA), which is a powerful fusion of model-based and empirical regression approaches. The internal algorithms included existing coefficient based retrievals for ATSRs, and a day-and-night (infra-red only) optimal estimator tuned (for both ATSRs and AVHRRs) to ATSR SSTs. ORAC as currently formulated is not sufficiently general (doesn't apply to night-time scenes) and gave out-of-target biases. IR was comparable to OEv2 for night-time AVHRR retrieval, but OEv2 performed a little better for day-time AVHRR retrieval on several quantitative metrics; moreover, OEv2 retains independence from in situ observations, whereas IR does not. OEv2 gives climate-quality results for ATSR-2 and AATSR, and therefore OEv2 is the best available, most consistent and independent algorithm for use by SST CCI for ATSR and AVHRR sensors.

The sequence of main conclusions from metric-based assessments, and the final conclusion, were as follows:

For AATSR and ATSR-2, ORAC and OEv1 showed some out-of-target biases, whereas OEv2 and ARC coefficients were comparable and within target. OEv2 gave better or comparable performance to ARC coefficients on measures of precision, stability and realism of uncertainty estimates. Both ARC and OEv2 are independent and fully sensitive

to SST variability. OEv2 allows in addition the calculation of extra quality information, which, although not a formal selection metric, is a relevant advantage. The interim conclusion was that either ARC coefficients or OEv2 are suitable for use for SST CCI, with marginally more evidence in favour of OEv2. OEv2 is not immediately applicable to ATSR-1, which tends to favour ARC coefficients for the later sensors. However, OEv2 and ARC are very consistent for ATSR-2 and AATSR (since OEv2 is tuned to ARC SSTs), so this is not a conclusive point. The performance of OEv2 for the AVHRR sensors needs to be considered before making the final conclusion for the ATSRs, and is summarized in the next paragraphs.

For night-time (3 channel) AVHRR retrievals, OEv2 and IR overall performed comparably on measures of bias, precision and SST sensitivity. OEv2 generally was better than IR for stability measures. OEv2 performed better than IR for high-latitude / marginal ice zones, whereas the converse was true in coastal zones. OEv2 is independent of in situ observations; IR is wholly dependent. The uncertainty estimates attached to OEv2 are more sophisticated than for IR. On balance, OEv2 had a slight advantage on the quantitative metrics, but the most significant difference was that it is an independent algorithm, rather than tuned to in situ SSTs.

For day-time (2 channel) AVHRR retrievals, both OEv2 and IR were within bias targets for Metop and NOAA-17. For NOAA-19 and NOAA-18, out of target biases were found for some northern mid-latitude regions for both OEv2 and IR, although they were considerably less marked in the OEv2 results. (It appears that restricting some conditions under which SST is retrieved may improve the situation regarding bias.) IR gave better precision than OEv2 when using two channels, but had low SST sensitivity, whereas for OEv2 SST sensitivity was close to ideal (100%). In fact, there is a trade-off between precision and sensitivity in this case, and parameters in either IR or OEv2 could be adjusted to maximize either metric at the expense of the other. For this selection, the metric of SST sensitivity had a greater weight than that for precision. OEv2 gave comparable or better performance on measures of stability. OEv2 and IR showed comparable performance in high latitudes, while OEv2 was markedly better in coastal zones. OEv2 is independent of in situ observations; IR is wholly dependent. The uncertainty estimates attached to OEv2 are more sophisticated than for IR. Because of the high weight attached to bias and stability considerations, the quantitative metrics in this case favour OEv2 somewhat. The other advantages of OEv2 (independence, ideal SST sensitivity and available of extra quality information, i.e., cost functions) add weight to this conclusion.

Taking together the night and day results for AVHRRs, the balance is in favour of selecting OEv2. Maximum possible algorithmic consistency across sensors is attractive, and this adds to the argument for using OEv2 for ATSR-2 and AATSR. The final selection was therefore of the OEv2 algorithm for SST.

2. INTRODUCTION

The SST_CCI project is part of the ESA Climate Change Initiative, which aims to produce and validate sea surface temperature (SST) essential climate variable (ECV) data products.

In order to identify the best performing algorithm or combination of algorithms, the SST_CCI project held an open round-robin (RR) algorithm intercomparison and product validation exercise following the protocol defined in this document and using the selection criteria defined in the Product Validation Plan (PVP, RD.216, Section 4). By maximising the number of users participating in the Round Robin exercise, ESA expects to identify the best algorithms for a future operational system.

The chosen algorithm(s) will then be implemented in an end-to-end system to generate the first SST_CCI data records. It is expected that future algorithm selection exercises will be carried out for each subsequent reprocessing to ensure the best performing algorithm is always implemented.

Although participation in the ESA SST_CCI RR was open to all, participants did have to follow a protocol (PVP, RD.216, Appendix C), which defined what was expected of each participant, how the RR would be run, how results should be submitted and what happens next.

All participants (including those internal to the ESA SST_CCI project team) were provided with a set of multi-sensor match-up dataset (MMD) files containing all of the necessary information to run their retrievals against. Each MMD match-up contained multiple satellite image extracts (roughly 11 km by 11 km) matched to a temperature history from an in situ platform.

Only match-ups to drifting buoys were used for the RR and the drifting buoy match-ups were randomly split into four categories: (1) training, (2) test, (3) selection and (4) validation. The training and test match-ups were provided at the start of the RR where the ‘training’ data could be used to tune a retrieval algorithm and the ‘test’ data were for the participant to evaluate their algorithms on an independent subset.

Towards the end of the RR period the participants were then supplied with the ‘selection’ match-ups, but this time the in situ measurements were withheld so the final choice of algorithms could be run on a blind sub-set. Participants then submitted SST estimates with appropriate uncertainties, which were then combined with the selection in situ data into the final RR data package for algorithm selection. The data package was then passed to the Science Team, who will carry out the final algorithm selection against the criteria defined in the PVP (RD.216).

In total, 10 research groups expressed interest in participating in the ESA SST_CCI RR, and contacts were made with Carol Anne Clayson (FSU now WHIO), Jim Cummings (US Navy), George Kruger (BoM), Haiyan Huang (Oxford), Bob Evans (Miami), Ajoy Kumar (Millersville), Igor Tomazic (ZIMO), Rene Preusker (FuB), Boris Petrenko (NOAA NESDIS) and Caroline Cox (RAL).

Of these potential participants only Petrenko and Cox submitted results to the final selection process (Cummings submission, 13 weeks after the public deadline, was too late for consideration). Unfortunately all other groups were unable to participate in the end due to other project commitments.

2.1 Purpose and Scope

This document presents the SST_CCI Algorithm Selection exercise, the submitted results, and the selection of SST retrieval algorithms.

2.2 Structure of the Document

After this introduction, the document is divided into a number of major sections that are briefly described below:

3 DEFINITIONS

This section defines key terms used within this document.

4 PROTOCOLS FOR ALGORITHM SELECTION (ROUND ROBIN)

Describes the scope, criteria, process, principles, and selection metrics used for algorithm selection.

5 IN**FORMATION FOR ALGORITHM** SELECTION

Presents the results (selection metrics) for different sensors on which algorithm selection is based.

6 ALGORITHM SELECTION

Interprets the results in Section 1, and states, discusses and justifies the selection of algorithms made.

APPENDIX OF FULL RESULTS

Full resolution graphics of selection metrics.

For convenience in navigating this large document on screen, hyperlinks are available as follows:

(i) The Table of Contents comprises active links to the corresponding section start.

(ii) Thumbnail images of selection metrics in Section 1 link to the corresponding full resolution graphic in the APPENDIX OF FULL RESULTS.

(iii) The caption below each full resolution graphic in the APPENDIX OF FULL RESULTS links to the selection table in which the corresponding thumbnail image appears in Section 1.

2.3 Referenced Documents

The following is a list of documents with a direct bearing on the content of this report. Where referenced in the text, these are identified as RD.n, where 'n' is the number in the list below:

RD.21 Mittaz, J. P. D., et al. (2009), A Physical Method for the Calibration of the AVHRR/3 Thermal IR Channels 1: The Prelaunch Calibration Data, Journal of Atmospheric and Oceanic Technology, 26(5), 996-1019.

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

RD.171 SST_CCI User Requirements Document, SST_CCI-URD-UKMO-001-Issue_2, http://www.esa-sst-cci.org/?q=webfm_send/46

RD.184 Embury, O., C. J. Merchant and G. K. Corlett (2012), A Reprocessing for Climate of Sea Surface Temperature from the Along-Track Scanning Radiometers: Initial validation, accounting for skin and diurnal variability, Rem. Sens. Env., pp62 - 78. DOI:10.1016/j.rse.2011.02.028

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

RD.211 Kennedy, J. J., et al. (2011), Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: biases and homogenisation., In press, Journal of Geophysical Research. 116, D14104, doi:10.1029/2010JD015220

RD.216 Casey, K.S., T.B. Brandon, P. Cornillon, and R. Evans (2010). "The Past, Present and Future of the AVHRR Pathfinder SST Program", in Oceanography from Space: Revisited, eds. V. Barale, J.F.R. Gower, and L. Alberotanza, Springer. DOI: 10.1007/978-90-481-8681-5_16

RD.221 Merchant C J, P Le Borgne, A Marsouin and H Roquet (2008), Optimal estimation of sea surface temperature from split-window observations, Rem. Sens. Env.,112(5), 2469-2484. doi:10.1016/j.rse.2007.11.011

RD.225 Merchant, C. J., (2011), SST CCI Algorithm Theoretical Basis Document v0, http://www.esa-sst-cci.org/?q=webfm_send/47

RD.246 O'Carroll, A.G., J.R. Eyre and R.W. Saunders, 2008: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations, J. Atmos. Ocean. Tech., 25, 1197-1207, doi: 10.1175/2007JTECHO542.1

RD.248 Merchant, C. J., A. R. Harris, H. Roquet, and P. Le Borgne (2009), Retrieval characteristics of non-linear sea surface temperature from the Advanced Very High Resolution Radiometer, Geophys. Res. Lett., 36, L17604, doi:10.1029/2009GL039843.

RD.272 Corlett, G., C.J. Merchant and N. Rayner, 2012: Product Validation Plan (PVP) – Issue 1, http://www.esa-sst-cci.org/?q=webfm_send/43

The current version of each SST_CCI project document is available via the SST CCI web pages at http://www.esa-sst-cci.org/?q=documents#.

2.4 Acronyms and abbreviations

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

Term	Definition
AATSR	Advanced ATSR
ASR	Algorithm Selection Report
AMSR-E	Advanced Microwave Scanning Radiometer – EOS
ARC	ATSR Reprocessing for Climate

ATSR	Along Track Scanning Radiometer
AVHRR	Advanced Very high Resolution Radiometer
BoM	Bureau of Meteorology
CASSTA	Composite Arctic Sea Surface Temperature Algorithm
CCI	Climate Change Initiative
CDR	Climate Data Record
D2	Dual-view 2-channel
D3	Dual-view 3-channel
DMI	Danmarks Meteorologiske Institut
ECV	Essential Climate Variable
EO	Earth Observation
ESA	European Space Agency
FSU	Florida State University
FuB	Freie Universität Berlin
GAC	Global Area Coverage
GCOS	Global Climate Observing System
GHRSST	Group for High Resolution SST
GUM	Guide to the expression of Uncertainty in Measurement
IR	Incremental Regression
METOP	Meteorological Operational Satellite
MMD	Multi-sensor Match-up Dataset
NESDIS	National Environmental Satellite, Data, and Information Service
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
OE	Optimal Estimation
OEv1	Optimal Estimation version 1
OEv2	Optimal Estimation version 2
ORAC	Oxford/RAL Retrieval of Aerosol and Cloud

PVP	Product Validation Plan
RAL	Rutherford Appleton Laboratory
RR	Round Robin
RRDP	Round Robin Data Package
RSD	Robust Standard Deviation
SD	Standard Deviation
SST	Sea Surface Temperature
SST_CCI	ESA Climate Change Initiative on SST
SZA	Solar Zenith Angle
TCWV	Total Column Water Vapour
UoE	University of Edinburgh
UoL	University of Leicester
URD	User Requirements Document
WHIO	Woods Hole Oceanographic Institution
ZIMO	Zavod za Istraživanje Mora i Okoliša

3. DEFINITIONS

The following definitions are used throughout this document:

Accuracy: For the term “accuracy” there seems to be two definitions in common circulation. In RD.150, GCOS considers accuracy to be measured by “the bias or systematic error of the data, i.e., the difference between the short-term average measured value of a variable and the truth” where the average referred to has been sufficient to render the random uncertainty in the measured value negligible. In contrast, the definition from the GUM [RD.191] is also used, whereby accuracy is “the closeness of agreement between the result of a measurement and a true value of a measurand” and therefore a measurement can be inaccurate either by virtue of a large systematic error or because it has a large random uncertainty. We find it useful to have a term available that distinguishes systematic and random uncertainty, and therefore in SST_CCI documents accuracy refers to the estimated magnitude of the systematic error (bias).

Precision: The difference between one result and the mean of several results obtained by the same method, i.e. reproducibility (includes random errors only).

Uncertainty: Is a parameter, associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand (given the measurement, in the light of our understanding of the sources of error in the measurement). Here, the parameter is the standard deviation of the dispersion, which is a confidence of 68% or (k=1).

Discrepancy: The difference between the result and the validation value.

Skin Sea Surface Temperature (SST-skin): The temperature measured by an infrared radiometer typically operating at wavelengths 3.7-12 µm (chosen for consistency with the majority of infrared satellite measurements) that represents the temperature within the conductive diffusion-dominated sub-layer at a depth of ~10-20 µm.

Sub-Skin Sea Surface Temperature (SST-subskin): The subskin temperature represents the temperature at the base of the conductive laminar sub-layer of the ocean surface.

Depth Sea Surface Temperature (SST-depth): Measurements of water temperature beneath the SSTsubskin, measured using a wide variety of platforms and sensors such as drifting buoys, vertical profiling floats, or deep thermistor chains at depths ranging from 10^{-2} - 10^{3}m. Here, the depth will usually be that associated with a drifting buoy (of order 20 cm) or a moored buoy (of order 1 m).

4. PROTOCOLS FOR ALGORITHM SELECTION (ROUND ROBIN)

4.1 Role and Scope of Algorithm Selection in SST CCI

The role, scope and process of Algorithm Selection (Round Robin) within SST CCI are fully defined in the Product Validation Plan (PVP) [RD.272 Section 6]. The statement of work for SST CCI requires that the present document also describe the Algorithm Selection process, with regards to the protocols, methods, reference data sets and tools used. Therefore, in this section, a summary of this information is given; however, the principal statement remains the PVP.

4.1.1 Role and Definition of Potential Scope

The role of algorithm selection in the SST CCI is to ensure that the most suitable algorithms are selected for creation of the long-term and short-term SST CDRs and are specified in the SST CCI processor. The algorithm selection process is open in the sense that algorithms for the SST CCI processor were not predefined, but will be determined by the algorithm selection process leading to this Algorithm Selection Report (ASR).

Algorithms considered via the algorithm selection process include, on an equal basis, those entered into the process by parties other than the SST CCI project team.

A consultation with potential external algorithm “competitors” was undertaken in April 2011 advertised via GHRSST, the Science Leader’s blog and by direct e-mails. The content for the call for interest is recorded in the blog (http://sst-cci.blogspot.com/2011/04/preparing-for-round-robin.html). There was external interest in participating in a formal Algorithm Selection for two aspects of SST CCI: “Estimation” and “Uncertainty” algorithms.

All algorithms in the categories covered within the Algorithm Selection process have been documented in the SST CCI “Algorithm Theoretical Basis Document v0” [RD.225].

4.1.1.1 SST estimation

Also known as retrieval or inversion, SST estimation is the process of inferring a value for SST from radiances (usually expressed as brightness temperatures).

4.1.1.2 SST uncertainty estimation

SST uncertainty estimation is the reasoned attribution of uncertainty information to an estimate of SST.

The total uncertainty in a single SST estimate reflects:

- the propagation of radiometric noise in the observed radiances through the estimation algorithm.
- the effect of algorithmic limitations, such as prior error and non-linearity error.
- the propagation of uncertainty in any ancillary information exploited in the estimation algorithm.
- the effect of classification errors and/or of undetected sub-pixel variations in the state.

- the uncertainty in radiance calibration propagated through the estimation algorithm.

- the uncertainty in true spatial location of the field(s) of view relative to the nominal geolocation.

An SST uncertainty estimate should quantify and combine at least the dominant sources of uncertainty.

4.2 Selection criteria and process

The selection of algorithms to be implemented in the SST CCI processor is made in this ASR by assessing algorithms against several criteria. To make comparisons fair, algorithm results need to be compared on common sets of matches within a common data set, called the Round Robin Data Package (RRDP).

4.2.1 Over-arching principles

Algorithms are compared on a fair basis by standardisation of the approach:

- Competing algorithms:
    - have been developed using identified ***training data*** within the RRDP (if necessary);
    - could be objectively assessed by developers internally using identified ***test data***; and
    - are here compared on the basis of results when applied to identified ***selection data*** , comprising cases for which all algorithm developers had no access to in situ validation data.
- All algorithm developers including the project EO team had access to the same data in the training and test categories (including in situ validation data), and to the same selection data (in situ data withheld).
- Common metrics describing the results (detailed below) are used for each type of algorithm to facilitate comparison of performance.

Algorithm selection requires joint assessment of a range of metrics and wider considerations. Not all properties of interest are quantifiable as metrics. Among measures that are quantifiable in principle, it may not always be feasible to undertake proper quantification within the scope of the project, and thus a qualitative approach may still be necessary.

4.2.2 Definition of selection metrics

Statements defining the selection metrics used later in this report are presented in this section. These are discussed in full in the PVP [RD.272].

4.2.2.1 Bias

Bias metrics

Metric for SST estimation:

For each sensor's selection dataset:

Calculate the mean and median SST discrepancy with respect to each class type of validation data. Here and below, the discrepancy is the difference between the validation data and the *individual satellite observation ("single pixel")* identified as the best match in the RRDP.

Map the mean and median SST discrepancy against drifting buoys, at a grid cell resolution that is specific to each sensor. The resolution will be chosen such that a mean 0.1 K discrepancy will be statistically significant for around 90% of grid cells. The same resolution will be used for all algorithms applied to a given sensor. Statistical significance will be assessed assuming drifting buoys with different buoy IDs have biases drawn from a Gaussian distribution of standard deviation 0.2 K and no random uncertainty. This approach is an approximation that doesn't account for drift in buoy calibration over time, and may be less valid for small numbers of matches.

Where sensors use different channel combinations and/or different algorithms in different situations, the above should be repeated for each situation.

"Less biased" algorithms will give mean and/or median values globally closer to the expected geophysical differences, and will have a narrower distribution of grid-cell mean and/or median values.

Metric for uncertainty estimation

For each sensor's selection dataset:

Calculate the chi-squared statistic, which measures the goodness of fit between the actual and estimated uncertainties of measurement and validation values, defined by:

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

where n is the number of discrepancies, i is an index, x is the SST estimate, v means validation value, δx is the SST uncertainty estimate, and δv is the uncertainty attributed to the validation value.

The expected value for χ^2 is unity. A value lower than this indicates the uncertainties attributed to the measurements or the validation values or both are too high. A value greater than unity indicates the uncertainties attributed to the measurements or the validation values or both are too low. Since the uncertainty to attribute to the validation values is itself uncertainty, the result for χ2 should not be over-interpreted, and confidence intervals in the result will be estimated in order to assess the significance of differences in χ2.

Where sensors use different channel combinations and/or different algorithms in different situations, the above should be repeated for each situation. Where the uncertainty estimate is known to be a function of some independent parameter (latitude, NWP[†] TCWV, etc.), the statistic should also be calculated for bins of this parameter to assess the validity of that dependence.

[†] in situ data is assimilated into NWP.

Table 4.2.1: Bias metrics for categories of algorithm.

4.2.2.2 Non-systematic uncertainty (precision)

Metrics of non-systematic uncertainty
Metric for SST estimation: For each sensor's selection dataset: At the same grid cell resolution used above to map bias, map the standard deviation of discrepancy in each cell. Also map the RSD of discrepancy in each cell. Calculate the mean of the cell standard deviations for the statistically significant cells, and likewise the median of the cell RSDs. "More precise" algorithms will have maps with smaller values of the above metrics than "less precise" algorithms. Where sensors use different channel combinations and/or different algorithms in different situations, the above should be repeated for each situation.

Table 4.2.2: Metric of non-systematic uncertainty for categories of algorithm.

4.2.2.3 Stability

Metrics of non-systematic uncertainty
Metric for SST estimation: The following need to be done for each channel-set/sensor combination relevant. Stability with respect to long-term trends: Put the N discrepancies for the stability subset in time order. From the standard deviation of discrepancy, calculate the number of points necessary for an average to have standard error ~0.01 K: $n \sim (SD / [0.01\ K])^2$. Divide the series into N/n periods, and calculate the average discrepancy and average time for each. Fit a linear (or low order polynomial, if appropriate) to these points, reporting the slope(s) of the fit as a stability estimate. Stability with respect to seasonality: Divide the N discrepancies into latitude bands (south of 15 S, 15 S to 15 N, north of 15 N). For each band, bin all discrepancies by month. Calculate mean and standard error for each latitude-month bin, and inspect the means for evidence of any annual cycle that is significant compared to the standard error. Stability with respect to diurnal cycle: calculate the mean discrepancy and standard error for day and night subsets, and check for significant differences. (Since some algorithms apply only at night, check for significant differences between any proposed pairs of algorithms for day and night use, also.) Repeat the trend procedure for day and night separately, to check for the long-term stability of any diurnal bias. Day and night periods are defined by solar zenith angle: day is when solar zenith angle is less than or equal to 85° and night is when solar zenith angle is greater than or equal to 95°. N.B. These definitions are modified from those given in the PVP (RD.272).

Table 4.2.3: Metric of stability for categories of algorithm.

4.2.2.4 Independence from in situ SST

	Metric of independence
SST estimation	Describe any usage of in situ observations in *defining* (not validating) algorithm.
SST uncertainty estimation	Describe any usage of in situ observations in *defining* (not validating) algorithm.

Table 4.2.4: Metric of independence from in situ SST

4.2.2.5 SST sensitivity

Metric of SST sensitivity
Calculate SST sensitivity for each match as per Merchant et al (2009) [RD.248]. Map the mean SST sensitivity for drifting buoy matches on the same grid cells as used for mapping SST bias. Where sensors use different channel combinations and/or different algorithms in different situations, the above should be repeated for each situation.

Table 4.2.5: Metric of SST sensitivity

4.2.2.6 Generality

The approach to consider generality is to create a list for each algorithm that states:

- to what sensors and/or channel-combinations and/or situations it applies.
- the degree and nature of adaptation required to apply the algorithm to a new sensor/channel-combination/situation.

On the basis of the comments, each algorithm's generality will be categorized as high, medium or low.

4.2.2.7 Improvability

For each algorithm, a statement is requested/provided specifying potential methodological improvements and the degree to which they are likely to improve accuracy, uncertainty or stability. On the basis of the information available, improvability will be categorized as high, medium or low.

4.2.2.8 Implementation considerations

For each algorithm, the following will be listed:

- The size and nature of any static auxiliary files required for the algorithm (e.g., look up tables).
- The size and nature of any dynamic auxiliary files required for the algorithm (e.g., NWP).
- The main steps of the algorithm, identifying where significant computation, use of external models, etc, is involved.

On the basis of the information available, implementation considerations will be categorized as high, medium or low.

4.2.3 Selection process

For each algorithm, all relevant metrics are generated and recorded in this report. The relevance of different metrics to categories of algorithm is summarized in Table 4.2.6.

	Esti-mation	**Uncer-tainty**
Bias	✓	✓
Uncertainty	✓	
Stability	✓	
Independence	✓	✓
Sensitivity	✓	
Generality	✓	✓
Improvability	✓	✓
Difficulty	✓	✓

Table 4.2.6: Relevance of different metrics to categories of algorithm for systematic consideration in the algorithm selection, all relevant metrics need to be provided and complied on a common basis.

Ultimately, we can expect competing algorithms to have strengths in different areas, and the decision regarding algorithm selection will consist of weighing the relative strengths and weaknesses. This selection process maps directly onto the trade-off analysis outlined in the statement of work [SST-TR-34]. Table 4.2.7 shows how each of the trade-off analysis criteria are met by the proposed metrics. The reasoned analysis leading to the algorithm selection is documented carefully for each selection, in the algorithm selection document.

Trade-off Analysis Criteria	Algorithm Selection Metric
Global retrieval accuracy	Bias
Degree of residual cloud contamination.	None applicable (no classification in RR)
Degree of residual aerosol contamination	None applicable (no classification in RR)
Performance of products in the MIZ.	Bias and non-systematic uncertainty assessed on high-latitude subset of test data
Performance of products in the coastal zone.	Bias and non-systematic uncertainty assessed on coastal subset of test data
Performance of products with respect to diurnal variability.	Stability
Ability to meet user and GCOS requirements.	Selection Process
Potential for further algorithm improvement to achieve ECV accuracy.	Improvability

Table 4.2.7: Match-up between trade-off selection criteria [SST-TR-34] and algorithm selection metrics.

4.3 Tools

The tools used in this algorithm selection exercise comprise routines in Interactive Data Language (IDL) which generate the metric data and figures.

5. INFORMATION FOR ALGORITHM SELECTION

5.1 Guide to interpretation of comparison tables

The selection of the algorithms for the SST CCI products is based on the comparison of the results presented in this present Section 1. The results are presented as comparison tables of the pre-defined algorithm selection metrics. This sub-section 5.1 describes these tables and offers a guide to their interpretation. For this purpose, refer to Table 5.2.1, the first comparison table presented.

The caption to the table defines the applicability of the results presented. In this case, the algorithm is for SST estimation from AATSR. The algorithms compared are global and this table shows their application to night-time observations. As some tables are spread over multiple pages, captions are repeated above and below each table for convenience. A hyperlink to the caption for each table is also provided in the first box of the table header (repeated at the start of each page the table covers).

Table 5.2.1: Comparison of metrics related to SST estimation for night-time AATSR observations, globally.

Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**

The top row of the table (above) identifies the algorithms by name. The first column of results relates to the ATSR Reprocessing for Climate (ARC) retrieval method based on coefficients. The next two columns refer to two versions of Optimal Estimation tested within the SST CCI project by the project team. The next column identifies an alternative optimal estimator submitted by the Rutherford Appleton Laboratory (RAL). Descriptions and references to each of these algorithms, and their equivalents for the AVHRR sensors later in the report, are given in the SST CCI project document "Algorithm Theoretical Basis Document v0" (RD.225). A feature of the RAL algorithm is its use of reflectance channels, making it only applicable to day time observations.

The first results row of the table (below) is labelled "Bias (mean discrepancy)". Bias (systematic error) is an important aspect of the SST CCI products to be generated for users (RD.171). The metric (measure) of bias presented in this row for each algorithm is the mean discrepancy of the satellite retrieved SST and the validation data (which are drifting buoys not used in the algorithm development -- see RD.272). The same adjustments for the sea surface skin effect (of order 0.2 K) and sub-skin to depth stratification (negligible at night) have been applied to all retrievals, and the ideal bias found would be zero. The "best" algorithm by this metric has the smallest magnitude of bias in this row. The SST CCI target for bias is to be within 0.1 K. Reading along the row, the ARC coefficients here have lowest bias (0.087 K) by this metric. The RAL column has N/A since the algorithm applies only to day time observations.

The final column along the row is titled Weight. This is a qualitative indicator of the relative importance of the different metrics to the algorithm selection decision. Since bias is a key user requirement, the Weight for this row is Very High.

Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Bias (mean discrepancy)	0.087 K	0.104 K	0.091 K	N/A	Very High

Mean discrepancy is not a perfect measure of bias, therefore the next three rows (below) have alternative metrics of bias for consideration. The next is median discrepancy, a metric which is less influenced by occasional outliers of discrepancy. Outliers can arise because of cloud contamination in the SST retrieval and because of poor drifting buoy values. Therefore this metric can be interpreted as more representative of the bias when everything is "right".

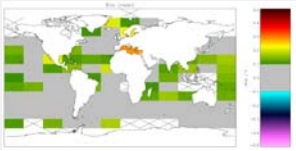
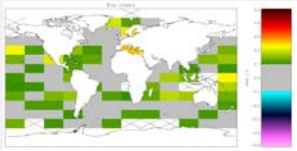
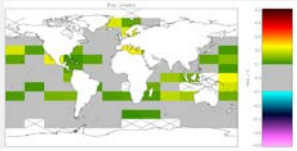
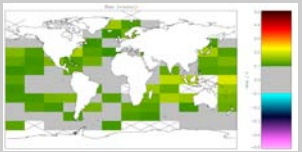
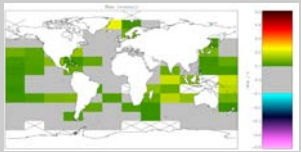
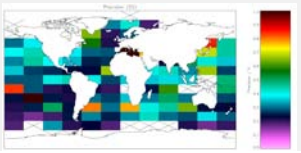
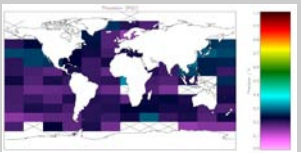
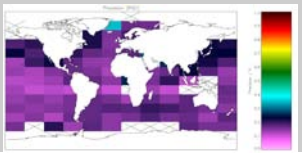
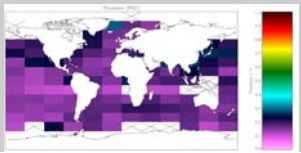
Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Bias (median discrepancy)	0.095 K	0.109 K	0.097 K	N/A	Very High
Bias (mean discrepancy map)				N/A	Very High

Table 5.2.1 AATSR Night	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Bias (median discrepancy map)				N/A	Very High

The next row (above) maps the mean discrepancy across the global oceans. Areas within target are coloured grey, and the ideal map would be completely grey. The cell size for this mapping has been chosen such that a magnitude of discrepancy greater than 0.1 K will arise by statistical chance only 5% of the time, based on an assumption that drifting buoy calibration uncertainty is 0.2 K (consistent with RD.246). Cells with discrepancy greater than 0.1 K appear coloured, and since, in each of these maps, more than 5% of the cells are coloured (generally green) rather than grey, we can infer there is a tendency in some regions for positive biases just above the 0.1 K target level. This is supported by the fact that the green cells have some geographical coherence. Cells where there are too few validation data to meet the criterion for statistical significance are marked by crosses on white.

Looking across the row, the image thumbnails allow easy comparison of the extent of bias outside the 0.1 K target for different algorithms. If the reader wants to scrutinize an image in detail, the document is hyperlinked such that clicking on the image thumbnail will jump to the full size version in the Appendix of this document. The reader can then navigate back to the comparison table from the caption to the full size image.

The fifth row of results (below) is labelled "Precision map (SD of discrepancy)". As indicated by this label, the mapped quantity here is the standard deviation (SD) of the discrepancies in each cell. The same cells are used as were valid for the bias maps higher in the table. Precision is ideally as small as possible, thus purple colours are good, blues and greens less so. Again, the thumbnails are hyperlinked in order to allow the reader to see the map in detail including the colour bar scale. Precision is of some importance to users, but the component of random errors that contribute to the precision is less important to climate users who often consider time and space scales over which there is a great deal of averaging satellite observations. Thus, the Weight for the precision metrics is Medium. Note also that the metric here is not a true reflection of the underlying satellite SST uncertainty, since drifting buoy errors contribute significantly to the observed values.

Table 5.2.1 AATSR Night	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Precision map (SD of discrepancy)				N/A	Medium
Precision map (RSD of discrepancy)				N/A	Medium
Precision (mean of cell SDs)	0.442 K	0.385 K	0.392 K	N/A	Medium
Precision (median of cell RSDs)	0.212 K	0.165 K	0.171 K	N/A	Medium

The next row (above) presents an alternative precision metric that is more robust to outliers -- robust standard deviation (RSD). The RSD is calculated as a scaled mean absolution deviation from the median, with the scaling such that it would give an identical answer to the conventional SD when applied to a Gaussian distribution. Here, the RSD precisions are generally smaller than the SD precisions, showing the outliers contribute significantly to the latter metric. The RSD in this example is useful in highlighting the better (tighter) precision (neglecting outliers) obtained by the optimal estimation techniques compared to ARC coefficient-based retrieval. This is also apparent in the global statistics on the next two rows (above).

Next there are statistics for "stability with respect to trend" (below). This is of Very High weight to climate users. The quoted metrics are statistical trend fits to the time-ordered discrepancies. The least squares fit trend +/- the 1 sigma fitting uncertainty on the trend are quoted. Points to note about this metric: (1) the fitting uncertainty quoted is not the full uncertainty, since the stability of the validation data (drifting buoys) as an ensemble is essentially unknown, and merely assumed to be small or comparable to these observed relative trends; (2) we can be statistically confident that an observed trend is detectably different from zero if the best estimate (the first number) is more different in magnitude from zero than (roughly speaking) twice the +/- number. Thus, in this example, no algorithm has given a statistically significant trend, nor are the algorithms

convincingly distinct from each other by this metric. The target stability for SST CCI is to be within 5 mK/yr of zero, and there is no evidence in this case of any of the algorithms causing the retrieved SSTs to have a trend outside that target.

Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Stability with respect to trend	*0.004* ± *0.003* K/yr	*-0.001* ± *0.003* K/yr	*-0.001* ± *0.003* K/yr	N/A	Very High
Stability with respect to season (amplitude of cycle)	North: *0.111* ± *0.585* K Equator: *0.106* ± *0.647* K South: *0.083* ± *0.687* K	North: *0.044* ± *0.685* K Equator: *0.051* ± *0.374* K South: *0.104* ± *0.893* K	North: *0.065* ± *0.526* K Equator: *0.063* ± *0.369* K South: *0.112* ± *0.878* K	N/A	Medium
Stability between day and night	Day-Night: *-0.005* ± *0.011* K. Trend: *0.007* ± *0.004* K/yr	Day-Night: 0.127 ± 0.01 K. Trend: *0.000* ± *0.004* K/yr	Day-Night: *-0.004* ± *0.010* K. Trend: *0.001* ± *0.004* K/yr.	N/A	Medium

"Stability with respect to season (amplitude of cycle)" (above) is a metric of how much the "bias" (mean discrepancy of satellite relative to validation data) changes with time of year. The discrepancies are binned by month of year and by latitude zone (North = north of 15° N, South = south of 15° S, Equator = remainder). For each zone, the maximum difference in mean discrepancy between any two months is found, and the magnitude reported along with the estimate of standard error.

"Stability between day and night" (above) is a metric of how much the bias changes between day and night. It is only applicable to algorithms that work both day and night.

For all stability estimates, cases which are not statistically significant within 95% confidence limits are highlighted in italics.

"Independence from in situ" (below) records the degree to which the SST retrieval is achieved without tuning or regression to any in situ observations. Independence was rated as essential by a significant minority of climate users in the SST CCI user requirements survey. An independent time series allows more objective assessment of in situ data sets and analyses, whereas a satellite data set that is tuned or regressed to in situ observations gives, inevitably, an over-optimistic picture of the degree of agreement between in situ and satellite records -- or, put another way, tends towards an under estimation of uncertainties, both in accuracy and stability.

Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Independence from in situ	SST retrieval fully independent.	Indirect dependence via use of NWP for forward model, but negligible in practice.	Indirect dependence via use of NWP for forward model, but negligible in practice.	N/A	High

"Map SST sensitivity" (below) is a map of $d\hat{x}/dx$, where $\hat{x}$ is retrieved SST and x is true SST (RD.248). For a perfect retrieval, a change of 1 K in the true SST will always lead to a 1 K change in retrieved SST, and the sensitivity is then 1. Where sensitivity differs from (generally is less than) 1, there are two implications. First is that SST variability on scales shorter than atmospheric correlation scales will be under-estimated in proportion to the degree to which the sensitivity is less than 1. Examples of situations were underestimation would be expected are fronts (short space scale) and rapid diurnal cycle variability (short time scale), although these expectations, it is fair to say, have not yet been comprehensively validated (RD.248). The second is a subtle point, but one that is very important for a climate data set. The sensitivity also measures the information content for the SST in the observations. Where sensitivity is not close to 1, the algorithm is providing some sort of prior SST information that significantly influences the SST retrieved. In methods such as optimal estimation via maximum a-posteriori probability, this prior information is explicit. It is not always understood that coefficient based retrievals also provide prior climatological SST information wherever the sensitivity is not 1. This tends to happen in regions where the atmospheric transmission is relatively low and a limited number of channels are used (e.g., 2); in such cases the coefficient based algorithms to some degree infer SST via climatological correlations between SST and lower-tropospheric humidity and temperature. Since in reality different geographical areas have different SST-lower-troposphere relationships, the result is often regional SST biases in the tropical zone (RD.248) -- even when coefficients are defined by regression to a match up data set that includes these same regions. In the maps, the idea colour is grey, which represents sensitivity between 0.99 and 1.01.

Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Map SST sensitivity				N/A	High

"Generality" (below) refers to the range of circumstances to which the algorithm is applicable. The approach to consider generality is to create a list for each algorithm that states: to what sensors and/or channel-combinations and/or situations it applies; and the degree and nature of adaptation required to apply the algorithm to a new sensor/channel-combination/situation. On the basis of the comments, each algorithm's generality is categorized as high, medium or low. The table contents brief summarize the range of applicability and then give the categorisation.

Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Generality	Coefficient design is specific to ATSRs.	General approach applicable to many sensors.	General approach applicable to many sensors, assuming matched high-quality reference SSTs are available.	N/A	Medium
Improvability	Reason to expect a full radiative transfer model upgrade may be beneficial. Otherwise, no obvious method to improve SST coefficients improvement.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	N/A	Medium

Table 5.2.1 AATSR Night	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Difficulty	Already implemented with modules and auxiliary coefficient files available.	Feasible in context of Bayesian cloud detection, since same simulations are required.	Feasible in context of Bayesian cloud detection, since same simulations are required. Requires new auxiliary information on simulation biases when applied to a new sensor.	N/A	Low

"Improvability" (above) specifies potential methodological improvements and the degree to which they are likely to improve accuracy, uncertainty or stability. On the basis of the information available, improvability will be categorized as high, medium or low.

"Difficulty" lists the nature of static and dynamic auxiliary information and external models used. The category is "High" is implementation is "easy" (i.e., limited volume of auxiliary information and modelling required). Although a relevant factor, implementation considerations should not determine an algorithm selection, and the weight on this aspect is Low.

In addition to tables presenting metrics of the SST retrieval, information to assess the associated uncertainty estimates is also provided. These uncertainty tables have a subset of the considerations for the main retrieval. The principal quantitative metric for the uncertainty estimates is the metric for bias (of the uncertainty). This is calculated using the equation in Table 4.2.1., assuming that drifting buoy observations have a certain precision (0.2 K). The 2 sigma (~95% confidence) range for the bias uncertainty is also provided as a + and - term to the central estimate. (The upper and lower range need to be specified separately since the distribution is asymmetric. They are calculated to account for our uncertainty in the appropriate drifting buoy precision, which is taken to be +/-1 sigma ~ +/- 0.02 K.) If the range does not include 1.0, then the uncertainty estimates associated with the SST retrieval are biased. Where the metric significantly exceeds 1.0, this means the uncertainty in the SST retrieval is under-estimated. Under-estimation of retrieval uncertainty is the more common case, since uncertainty models may not capture the effects of all the processes that contribute to the retrieval error distribution.

5.2 Side-by-side comparison of results relating to AATSR

5.2.1 Presentation of metrics related to SST estimation for AATSR

Table 5.2.1: Comparison of metrics related to SST estimation for night-time AATSR observations, globally.

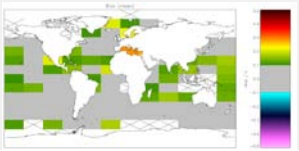
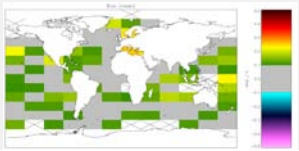
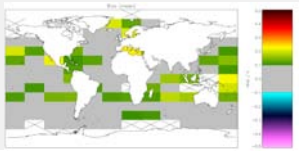
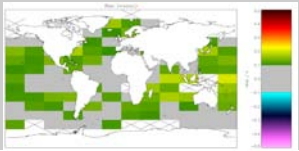
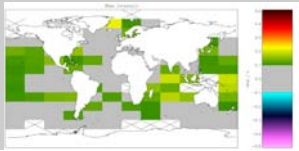
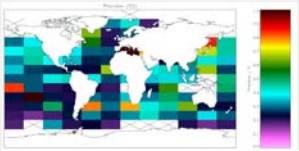
Table 5.2.1 AATSR Night	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Bias (mean discrepancy)	0.087 K	0.104 K	0.091 K	N/A	Very High
Bias (median discrepancy)	0.095 K	0.109 K	0.097 K	N/A	Very High
Bias (mean discrepancy map)				N/A	Very High
Bias (median discrepancy map)				N/A	Very High
Precision map (SD of discrepancy)				N/A	Medium

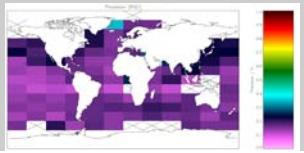
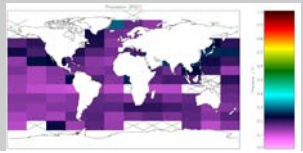
Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Precision map (RSD of discrepancy)				N/A	Medium
Precision (mean of cell SDs)	0.442 K	0.385 K	0.392 K	N/A	Medium
Precision (median of cell RSDs)	0.212 K	0.165 K	0.171 K	N/A	Medium
Stability with respect to trend	*0.004* ± *0.003* K/yr	*-0.001* ± *0.003* K/yr	*-0.001* ± *0.003* K/yr	N/A	Very High
Stability with respect to season (amplitude of cycle)	North: *0.111* ± *0.585* K Equator: *0.106* ± *0.647* K South: *0.083* ± *0.687* K	North: *0.044* ± *0.685* K Equator: *0.051* ± *0.374* K South: *0.104* ± *0.893* K	North: *0.065* ± *0.526* K Equator: *0.063* ± *0.369* K South: *0.112* ± *0.878* K	N/A	Medium
Stability between day and night	Day-Night: *-0.005* ± *0.011* K. Trend: *0.007* ± *0.004* K/yr	Day-Night: 0.127 ± 0.01 K. Trend: *0.000* ± *0.004* K/yr	Day-Night: *-0.004* ± *0.010* K. Trend: *0.001* ± *0.004* K/yr.	N/A	Medium
Independence from in situ	SST retrieval fully independent.	Indirect dependence via use of NWP for forward	Indirect dependence via use of NWP for forward	N/A	High

Table 5.2.1 **AATSR Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
		model, but negligible in practice.	model, but negligible in practice.		
Map SST sensitivity				N/A	High
Generality	Coefficient design is specific to ATSRs.	General approach applicable to many sensors.	General approach applicable to many sensors, assuming matched high-quality reference SSTs are available.	N/A	Medium
Improvability	Reason to expect a full radiative transfer model upgrade may be beneficial. Otherwise, no obvious method to improve SST coefficients improvement.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	N/A	Medium
Difficulty	Already implemented with modules and auxiliary coefficient files available.	Feasible in context of Bayesian cloud detection, since same simulations are required.	Feasible in context of Bayesian cloud detection, since same simulations are required. Requires new auxiliary information on simulation biases when applied to a new sensor.	N/A	Low

Table 5.2.1 AATSR Night	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Number of matches	21938	19563	20194	N/A	

Table 5.2.1: Comparison of metrics related to SST estimation for night-time AATSR observations, globally.

Table 5.2.2: Comparison of metrics related to SST estimation for day-time AATSR observations, globally.

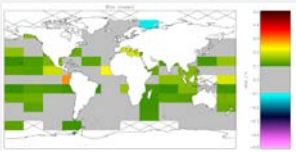
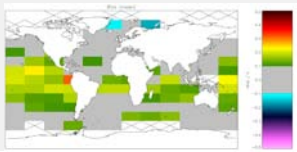
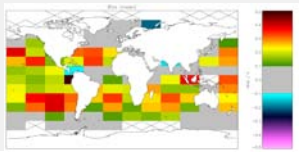
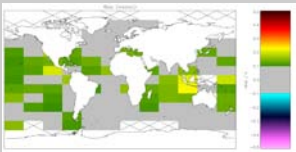
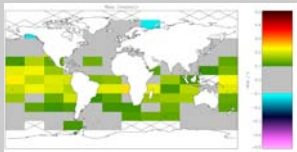
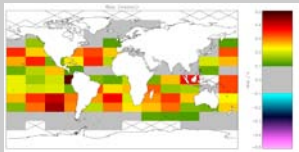
Table 5.2.2 AATSR Day	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Bias (mean discrepancy)	0.082 K	-0.023 K	0.087 K	0.206 K	Very High
Bias (median discrepancy)	0.094 K	0.003 K	0.101 K	0.221 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High

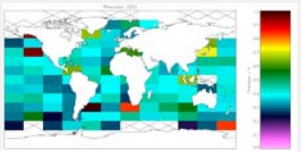
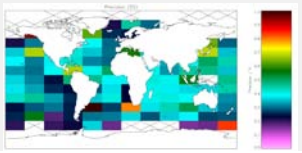
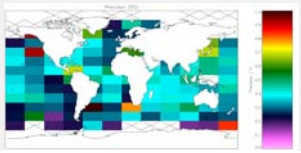
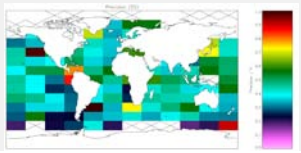
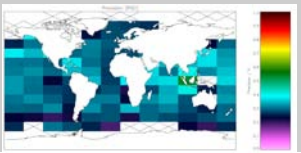
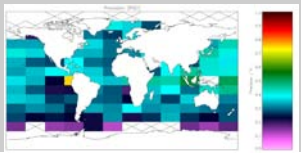
Table 5.2.2 **AATSR Day**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Precision map (SD of discrepancy)					Medium
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.463 K	0.432 K	0.434 K	0.497 K	Medium
Precision (median of cell RSDs)	0.306 K	0.269 K	0.264 K	0.349 K	Medium
Stability with respect to trend	*-0.003* ± *0.002* K/yr	*-0.001* ± *0.003* K/yr	*0.000* ± *0.003* K/yr	-0.008 ± 0.003 K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.139* ± *0.600* K Equator: *0.076* ± *0.543* K South: *0.067* ± *0.775* K	North: *0.211* ± *0.615* K Equator: *0.110* ± *0.526* K South: *0.067* ± *0.770* K	North: *0.171* ± *0.612* K Equator: *0.087* ± *0.508* K South: *0.063* ± *0.698* K	North: *0.140* ± *0.570* K Equator: *0.112* ± *0.737* K South: *0.084* ± *0.754* K	Medium
Stability between day and night	Day-Night: *-0.005* ± *0.011* K. Trend: *0.007* ± *0.004*	Day-Night: 0.127 ± 0.01 K. Trend: *0.000* ± *0.004*	Day-Night: *-0.004* ± *0.010* K. Trend: *0.001* ± *0.004*	N/A	Medium

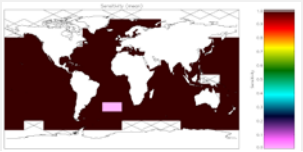
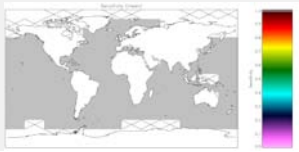
Table 5.2.2 **AATSR Day**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
	K/yr	K/yr	K/yr.		
Independence from in situ	SST retrieval fully independent.	Indirect dependence via use of NWP for forward model, but negligible in practice.	Indirect dependence via use of NWP for forward model, but negligible in practice.	Indirect dependence via use of NWP for forward model, but negligible in practice.	High
Map SST sensitivity				Not supplied but likely to be high (close to 1.0)	High
Generality	Coefficient design is specific to ATSRs.	General approach applicable to many sensors.	General approach applicable to many sensors, assuming matched high-quality reference SSTs are available.	General approach applicable to many sensors, but only envisaged for day time scenes at current time.	Medium
Improvability	Reason to expect a full radiative transfer model upgrade may be beneficial. Otherwise, no obvious method to improve SST coefficients improvement.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP, sensors (visible channels in particular), and particularly aerosol modes.	Medium
Difficulty	Already implemented	Feasible in context of	Feasible in context of	Some adaptation and	Low

Table 5.2.2 AATSR Day	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
	with modules and auxiliary coefficient files available.	Bayesian cloud detection, since same simulations are required.	Bayesian cloud detection, since same simulations are required. Requires new auxiliary information on simulation biases when applied to a new sensor.	recoding required.	
Number of matches	31659	30848	31053	27992	

Table 5.2.2: Comparison of metrics related to SST estimation for day-time AATSR observations, globally.

Table 5.2.3: Comparison of metrics related to SST estimation for night-time AATSR, for coastal regions.

Table 5.2.3 AATSR Night Coastal	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Bias (mean discrepancy)	0.093 K	0.144 K	0.127 K	N/A	Very High
Bias (median discrepancy)	0.097 K	0.115 K	0.111 K	N/A	Very High
Precision (mean of cell SDs)	0.612 K	0.532 K	0.570 K	N/A	Medium
Precision (median	0.267 K	0.205 K	0.216 K	N/A	Medium

Table 5.2.3 **AATSR Night Coastal**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
of cell RSDs)					
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	N/A	Very High
Stability with respect to season (amplitude of cycle)	North: *0.408 ± 1.139* K Equator: *0.706 ± 0.862* K South: *1.122 ± 1.628* K	North: *0.313 ± 1.094* K Equator: *0.556 ± 0.970* K South: *0.951 ± 1.326* K	North: *0.340 ± 1.049* K Equator: *0.552 ± 0.597* K South: *1.101 ± 1.729* K	N/A	Medium
Stability between day and night	Day-Night: *-0.012 ± 0.031* K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.209 ± 0.029 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.070 ± 0.030 K Trend: Not enough match-ups available to estimate trends.	N/A	Medium
Number of matches	922	760	800	N/A	

Table 5.2.3: Comparison of metrics related to SST estimation for night-time AATSR, for coastal regions.

Table 5.2.4: Comparison of metrics related to SST estimation for day-time AATSR, for coastal regions.

Table 5.2.4 **AATSR Day Coastal**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Bias (mean discrepancy)	0.081 K	-0.065 K	0.057 K	0.135 K	Very High
Bias (median discrepancy)	0.109 K	-0.033 K	0.072 K	0.154 K	Very High
Precision (mean of cell SDs)	0.552 K	0.534 K	0.531 K	0.575 K	Medium
Precision (median of cell RSDs)	0.337 K	0.316 K	0.300 K	0.403 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.408 ± 1.038* K Equator: *0.489 ± 0.693* K South: *0.641 ± 1.112* K	North: *0.420 ± 1.053* K Equator: *0.433 ± 0.748* K South: *0.674 ± 1.196* K	North: *0.437 ± 1.042* K Equator: *0.289 ± 0.544* K South: *0.698 ± 1.141* K	North: *0.374 ± 0.962* K Equator: *0.548 ± 1.091* K South: *0.552 ± 0.783* K	Medium
Stability between day and night	Day-Night: *-0.012 ± 0.031* K Trend: Not enough match-ups available to	Day-Night: -0.209 ± 0.029 K Trend: Not enough match-ups available to	Day-Night: -0.070 ± 0.030 K Trend: Not enough match-ups available to	N/A	Medium

Table 5.2.4 AATSR Day Coastal	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
	estimate trends.	estimate trends.	estimate trends.		
Number of matches	1202	1164	1169	990	

Table 5.2.4: Comparison of metrics related to SST estimation for day-time AATSR, for coastal regions.

5.2.2 Presentation of metrics related to SST uncertainty for AATSR

Table 5.2.5: Comparison of metrics related to SST uncertainty estimation for night-time AATSR observations, globally.

Table 5.2.5 AATSR Night	**ARC**		**Optimal Estimation v1**		**Optimal Estimation v2**		**ORAC**		**Weight**
Bias (chi-squared) (Estimate ± 2σ range)	1.7	+0.3 -0.2	2.3	+0.6 -0.5	2.3	+0.6 -0.5	N/A		High
Independence from in situ	Independent		Independent		Independent		N/A		Medium
Generality	Specific to ARC coefficient formulation		General to OE framework		General to OE framework		N/A		Low
Improvability	Not clear		Fuller assessment of forward model error covariance.		Fuller assessment of forward model error covariance.		N/A		Low
Difficulty	Moderate. New auxiliary analysis required for each sensor applied to.		No difficulty given OE framework.		No difficulty given OE framework.		N/A		Low

Table 5.2.5: Comparison of metrics related to SST uncertainty estimation for night-time AATSR observations, globally.

Table 5.2.6: Comparison of metrics related to SST uncertainty estimation for day-time AATSR observations, globally.

Table 5.2.6 AATSR Day	**ARC**		**Optimal Estimation v1**		**Optimal Estimation v2**		**ORAC**		**Weight**
Bias (chi-squared) (Estimate ± 2σ range)	2.5	+0.5 -0.5	1.1	+0.2 -0.1	1.1	+0.1 -0.1	1.3	+0.1 -0.2	High

Table 5.2.6 AATSR Day	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Independence from in situ	Independent	Independent	Independent	Independent	Medium
Generality	Specific to ARC coefficient formulation	General to OE framework	General to OE framework	General to OE framework	Low
Improvability	Not clear	Fuller assessment of forward model error covariance.	Fuller assessment of forward model error covariance.	Fuller assessment of forward model error covariance.	Low
Difficulty	Moderate. New auxiliary analysis required for each sensor applied to.	No difficulty given OE framework.	No difficulty given OE framework.	No difficulty given OE framework.	Low

Table 5.2.6: Comparison of metrics related to SST uncertainty estimation for day-time AATSR observations, globally.

5.2.3 Comments on results for AATSR

Considering night-time SST retrieval results (i.e., using dual-view and three-channels for retrieval), the AATSR SSTs based on ARC-coefficients applied to AATSR yield bias metrics about 0.05 K warmer than previously reported in RD.184. In terms of bias, ARC is comparable to OE v2. Since OE v2 is by design referenced to ARC via correction of brightness temperatures simulated using ARC SSTs, this is to be expected. OE v1 has a somewhat larger bias with more out-of-target areas on the bias maps. For both ARC and OE v2, most of the global oceans are grey in the bias maps, indicating the regional bias is within the 0.1 K target. The OE methods give lower (better) precision values than ARC. None of the algorithms show statistically significant metrics of stability with respect to trend, nor are differences between the results for different algorithms here significant. Other stability measures are not statistically clear cut except that OE v1 shows a significant relative bias of 0.127 K between day and night. Overall, for the night-time selection data, the two OE algorithms look to be the best and are quite comparable, although OE v2 is the better of the two for bias and day-night stability. Note that some retrievals are rejected as invalid based on limits on the retrieval cost, and that OE v1 rejects more than OE v2, which may advantage the OE v1 statistics somewhat.

Considering day time results, OE v1 has the lowest bias, although both ARC and OE v2 are within the target of mean bias <0.1 K. OE v1 is within the 0.1 K accuracy target for more of the global oceans than the other algorithms. The algorithm from RAL shows zones of relatively large bias (~0.4 K) and has the poorest precision of the four algorithms

considered. The ORAC algorithm is an optimal estimation that includes cloud/aerosol retrieval in its state vector. It currently applied only to day-time. The ORAC biases and standard deviations are somewhat larger than with OE v2, although in principle ORAC can cope with aerosol or cloud contamination better than OE v2. Areas of difference include the IR forward model (RTTOV v9 for ORAC, v10 for OE v2). Furthermore, the visible channel prior error covariances for the ORAC algorithm are recognised to require further development. Only the RAL results show a significant linear trend, with OE v2 giving a trend of zero K/year to 3 decimal places. Seasonal stability results are mixed, with none statistically significant. Again, OE v1 rejects more results based on retrieval cost than OE v2.

The uncertainty metrics for night-time AATSR SSTs are significantly above 1.0 for ARC, OE v1 and OE v2 – this means that the estimated uncertainties are smaller than the comparisons with drifting buoys suggest to be realistic. For day-time SSTs, the ARC uncertainty is also under-estimated, whereas for OE v1 and OE v2, it is not significantly biased.

Regarding the non-quantitative metrics, all of the retrieval algorithms here have complete or very high independence from in situ observations. Optimal estimation algorithms are more general (more readily widely applied) than the specific ATSR-series coefficients developed in ARC. The exception is the RAL algorithm that applies only to day time observations. There is in general a clearer path to improvement for optimal estimation, which in general involves either (i) reducing the uncertainty in the fast simulation of observations, or (ii) characterising more subtly the prior error covariances. The only algorithm that would present a challenge of implementation to the SST CCI is the RAL algorithm, since the Oxford/RAL Retrieval of Aerosol and Cloud (ORAC) is not implemented within SST CCI.

Overall, OE v1 and OE v2 are more often the best algorithms than ARC or RAL, with OE v1 having the edge over OE v2 for the night time results. The bias performance of OE v1 may be somewhat fortuitous (no effort having been made to bias correct the forward model), which can be assessed by looking at results for OE v1 for other sensors.

Interim conclusions considering AATSR metrics: OEv1, OEv2 and ARC coefficients are all appropriate for use for AATSR in SST CCI. OEv1 biases are generally good: this unexpected result may be fortuitous and may not apply for other cases. OEv2 is comparable to ARC regarding bias and stability, and better with regards to precision. Of the four candidates, only ORAC is excluded on the basis of the AATSR metrics, because of out-of-target biases and its present applicability only to day time cases.

5.3 Side-by-side comparison of results relating to ATSR-2

5.3.1 Presentation of metrics related to SST estimation for ATSR-2

Table 5.3.1: Comparison of metrics related to SST estimation for night-time ATSR-2 observations, globally.

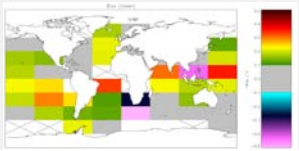
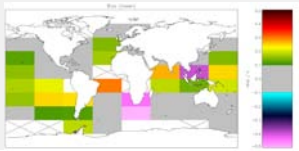
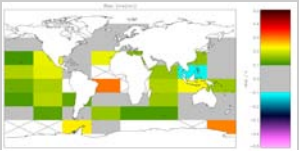
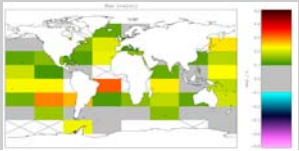
Table 5.3.1 ATSR-2 Night	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Bias (mean discrepancy)	0.062 K	0.130 K	0.081 K	N/A	Very High
Bias (median discrepancy)	0.107 K	0.161 K	0.113 K	N/A	Very High
Bias (mean discrepancy map)				N/A	Very High
Bias (median discrepancy map)				N/A	Very High
Precision map (SD of discrepancy)				N/A	Medium

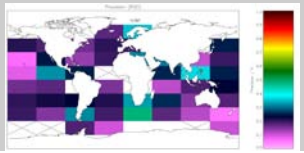
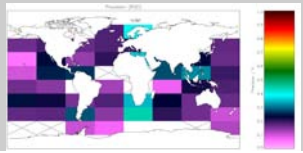
Table 5.3.1 **ATSR-2 Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Precision map (RSD of discrepancy)				N/A	Medium
Precision (mean of cell SDs)	0.485 K	0.436 K	0.444 K	N/A	Medium
Precision (median of cell RSDs)	0.250 K	0.202 K	0.207 K	N/A	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	N/A	Very High
Stability with respect to season (amplitude of cycle)	North: *0.347* ± *0.462* K Equator: *0.234* ± *0.444* K South: *0.763* ± *1.381* K	North: *0.152* ± *0.392* K Equator: *0.277* ± *0.166* K South: *0.867* ± *1.555* K	North: *0.213* ± *0.300* K Equator: *0.239* ± *0.153* K South: *0.920* ± *1.568* K	North: N/A Equator: N/A South: N/A	Medium
Stability between day and night	Day-Night: -0.077 ± 0.027 K. Trend: N/A	Day-Night: -0.232 ± 0.025 K Trend: N/A	Day-Night: 0.085 ± 0.025 K. Trend: N/A	Day-Night: N/A Trend: N/A	Medium
Independence from in situ	SST retrieval fully independent.	Indirect dependence via use of NWP for forward	Indirect dependence via use of NWP for forward	N/A	High

Table 5.3.1 **ATSR-2 Night**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
		model, but negligible in practice.	model, but negligible in practice.		
Map SST sensitivity				N/A	High
Generality	Coefficient design is specific to ATSRs.	General approach applicable to many sensors.	General approach applicable to many sensors, assuming matched high-quality reference SSTs are available.	N/A	Medium
Improvability	Reason to expect a full radiative transfer model upgrade may be beneficial. Otherwise, no obvious method to improve SST coefficients improvement.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	N/A	Medium
Difficulty	Already implemented with modules and auxiliary coefficient files available.	Feasible in context of Bayesian cloud detection, since same simulations are required.	Feasible in context of Bayesian cloud detection, since same simulations are required. Requires new auxiliary information on simulation biases when applied to a new sensor.	N/A	Low

Table 5.3.1 ATSR-2 Night	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Number of Matches	2075	1717	1809		

Table 5.3.1: Comparison of metrics related to SST estimation for night-time ATSR-2 observations, globally.

Table 5.3.2: Comparison of metrics related to SST estimation for day-time ATSR-2 observations, globally.

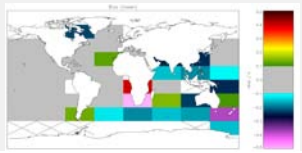
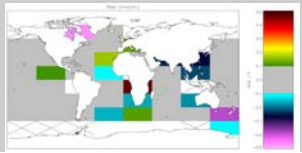
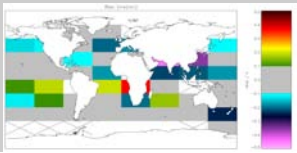
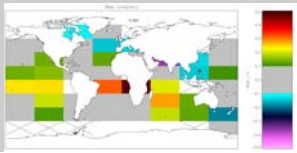
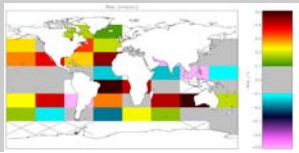
Table 5.3.2 ATSR-2 Day	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Bias (mean discrepancy)	-0.015 K	-0.102 K	-0.004 K	0.139 K	Very High
Bias (median discrepancy)	0.025 K	-0.053 K	0.039 K	0.174 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High

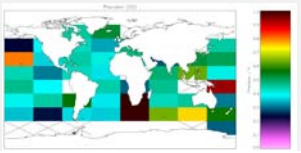
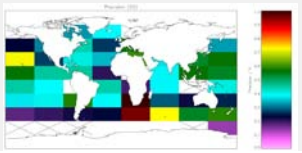
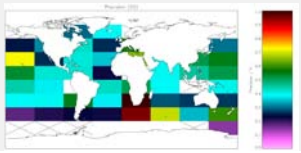
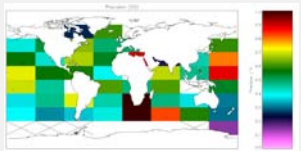
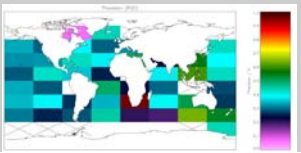
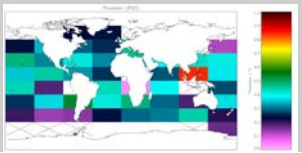
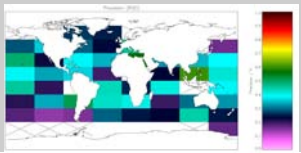
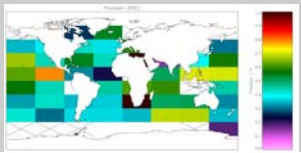
Table 5.3.2 **ATSR-2 Day**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Precision map (SD of discrepancy)					Medium
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.520 K	0.491 K	0.486 K	0.645 K	Medium
Precision (median of cell RSDs)	0.373 K	0.346 K	0.345 K	0.499 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.557 ± 1.173* K Equator: *0.526 ± 0.964* K South: *0.571 ± 1.169* K	North: *0.479 ± 1.133* K Equator: *0.259 ± 0.995* K South: *0.567 ± 1.035* K	North: *0.521 ± 1.128* K Equator: *0.148 ± 0.964* K South: *0.560 ± 1.031* K	North: *0.579 ± 1.125* K Equator: *0.206 ± 0.926* K South: *0.564 ± 1.152* K	Medium
Stability between day and night	Day-Night: -0.077 ± 0.027 K.	Day-Night: -0.232 ± 0.025 K	Day-Night: 0.085 ± 0.025 K.	Day-Night: N/A	Medium

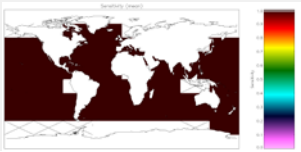
Table 5.3.2 **ATSR-2 Day**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
	Trend: N/A	Trend: N/A	Trend: N/A	Trend: N/A	
Independence from in situ	SST retrieval fully independent.	Indirect dependence via use of NWP for forward model, but negligible in practice.	Indirect dependence via use of NWP for forward model, but negligible in practice.	Indirect dependence via use of NWP for forward model, but negligible in practice.	High
Map SST sensitivity				N/A	High
Generality	Coefficient design is specific to ATSRs.	General approach applicable to many sensors.	General approach applicable to many sensors, assuming matched high-quality reference SSTs are available.	General approach applicable to many sensors, but only envisaged to day time scenes at current time.	Medium
Improvability	Reason to expect a full radiative transfer model upgrade may be beneficial. Otherwise, no obvious method to improve SST coefficients improvement.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP, sensors, and particularly aerosol modes.	Medium
Difficulty	Already implemented with modules and	Feasible in context of Bayesian cloud	Feasible in context of Bayesian cloud	For short delay operation, obtaining	Low

Table 5.3.2 **ATSR-2 Day**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
	auxiliary coefficient files available.	detection, since same simulations are required.	detection, since same simulations are required. Requires new auxiliary information on simulation biases when applied to a new sensor.	aerosol prior information may be difficult.	
Number of Matches	1996	1927	1954	1578	

Table 5.3.2: Comparison of metrics related to SST estimation for day-time ATSR-2 observations, globally.

Table 5.3.3: Comparison of metrics related to SST estimation for night-time ATSR-2, for coastal regions.

Table 5.3.3 **ATSR-2 Night Coastal**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Bias (mean discrepancy)	0.018 K	0.074 K	-0.002 K	N/A	Very High
Bias (median discrepancy)	0.051 K	0.095 K	0.047 K	N/A	Very High
Precision (mean of cell SDs)	0.590 K	0.522 K	0.622 K	N/A	Medium
Precision (median of cell RSDs)	0.324 K	0.285 K	0.299 K	N/A	Medium

Table 5.3.3 **ATSR-2 Night Coastal**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	N/A	Very High
Stability with respect to season (amplitude of cycle)	North: *1.126* ± *1.991* K Equator: *0.086* ± *0.506* K South: *0.471* ± *0.534* K	North: *1.069* ± *2.122* K Equator: *0.110* ± *0.528* K South: *0.346* ± *0.362* K	North: *1.111* ± *2.168* K Equator: *0.464* ± *0.501* K South: *0.288* ± *0.365* K	N/A	Medium
Stability between day and night	Day-Night: *-0.096* ± *0.106* K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.294 ± 0.102 K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.122* ± *0.110* K Trend: Not enough match-ups available to estimate trends.	N/A	Medium
Number of matches	135	98	105	N/A	

Table 5.3.3: Comparison of metrics related to SST estimation for night-time ATSR-2, for coastal regions.

Table 5.3.4: Comparison of metrics related to SST estimation for day-time ATSR-2, for coastal regions.

Table 5.3.4 **ATSR-2 Day Coastal**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
Bias (mean discrepancy)	-0.078 K	-0.220 K	-0.124 K	0.004 K	Very High
Bias (median discrepancy)	-0.014 K	-0.105 K	-0.022 K	0.055 K	Very High
Precision (mean of cell SDs)	0.687 K	0.664 K	0.671 K	0.582 K	Medium
Precision (median of cell RSDs)	0.394 K	0.288 K	0.324 K	0.405 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *1.160 ± 3.073* K Equator: *1.081 ± 1.328* K South: *0.652 ± 0.410* K	North: *1.006 ± 3.054* K Equator: *0.674 ± 1.160* K South: *0.253 ± 0.212* K	North: *1.059 ± 3.063* K Equator: *0.936 ± 1.173* K South: *0.153 ± 0.225* K	North: *0.690 ± 0.493* K Equator: Not enough match-ups available to estimate stability. South: *0.443 ± 0.871* K	Medium
Stability between day and night	Day-Night: *-0.096 ± 0.106* K Trend: Not enough	Day-Night: -0.294 ± 0.102 K Trend: Not enough	Day-Night: *-0.122 ± 0.110* K Trend: Not enough	N/A	Medium

Table 5.3.4 **ATSR-2 Day Coastal**	**ARC**	**Optimal Estimation v1**	**Optimal Estimation v2**	**ORAC**	**Weight**
	match-ups available to estimate trends.	match-ups available to estimate trends.	match-ups available to estimate trends.		
Number of matches	128	125	124	92	

Table 5.3.4: Comparison of metrics related to SST estimation for day-time ATSR-2, for coastal regions.

5.3.2 Presentation of metrics related to SST uncertainty for ATSR-2

Table 5.3.5: Comparison of metrics related to SST uncertainty estimation for night-time ATSR-2 observations, globally.

Table 5.3.5 **AATSR Night**	**ARC**		**Optimal Estimation v1**		**Optimal Estimation v2**		**ORAC**		**Weight**
Bias (chi-squared) (Estimate ± 2σ range)	0.9	+0.1 -0.1	2.8	+0.8 -0.5	2.8	+0.7 -0.5	N/A		High
Independence from in situ	Independent		Independent		Independent		N/A		Medium
Generality	Specific to ARC coefficient formulation		General to OE framework		General to OE framework		N/A		Low
Improvability	Not clear		Fuller assessment of forward model error covariance.		Fuller assessment of forward model error covariance.		N/A		Low
Difficulty	Moderate. New auxiliary analysis required for each sensor applied to.		No difficulty given OE framework.		No difficulty given OE framework.		N/A		Low

Table 5.3.5: Comparison of metrics related to SST uncertainty estimation for night-time ATSR-2 observations, globally.

Table 5.3.6: Comparison of metrics related to SST uncertainty estimation for day-time ATSR-2 observations, globally.

Table 5.3.6 **AATSR Day**	**ARC**		**Optimal Estimation v1**		**Optimal Estimation v2**		**ORAC**		**Weight**
Bias (chi-squared) (Estimate ± 2σ range)	2.6	+0.6 -0.4	1.6	+0.2 -0.3	1.4	+0.3 -0.2	2.0	+0.3 -0.2	High

Table 5.3.6 AATSR Day	ARC	Optimal Estimation v1	Optimal Estimation v2	ORAC	Weight
Independence from in situ	Independent	Independent	Independent	Independent	Medium
Generality	Specific to ARC coefficient formulation	General to OE framework	General to OE framework	General to OE framework	Low
Improvability	Not clear	Fuller assessment of forward model error covariance.	Fuller assessment of forward model error covariance.	Fuller assessment of forward model error covariance.	Low
Difficulty	Moderate. New auxiliary analysis required for each sensor applied to.	No difficulty given OE framework.	No difficulty given OE framework.	No difficulty given OE framework.	Low

Table 5.3.6: Comparison of metrics related to SST uncertainty estimation for day-time ATSR-2 observations, globally.

5.3.3 Comments on results for ATSR-2

Considering night-time SST retrieval results both ARC-coefficients and OE v2 retrievals yield biases within the target when considering mean discrepancy. However, none of the algorithms achieve the 0.1 K target when the median discrepancy is used, although both ARC and OE v2 are within ~0.01 K. Both OE methods give lower (better) precision values than ARC. Due to the much smaller number of match-ups for ATSR-2 it was not possible to estimate the stability with respect to trend, using the methods specified. Estimates of stability with respect to season are non-significant and broadly similar for all algorithms. Day-night biases are larger than for AATSR and significant for all algorithms, with the largest bias of -0.232 K observed for OE v1.

Day-time ARC and OE v2 biases are small using both mean and median forms of the metric. OE v1 is within the target bias of 0.01 K using the median and marginally above for the mean. The ORAC algorithm returns significantly higher biases, well above the target. This is highlighted in the bias maps, where ORAC returns fewer grey (target bias) cells and larger individual cell biases than the other algorithms. Precision values are again lower for the OE algorithms, with ORAC retrievals exhibiting more noise. As for night-time retrievals, too few match-ups were available to estimate stability with respect to trend. Seasonal stability metrics are again non-significant and broadly similar across algorithms.

Comments on the qualitative metrics are as for AATSR (section 5.2.3).

The uncertainty metrics for night-time AATSR SSTs are significantly above 1.0 for ARC, OE v1 and OE v2 – this means that the estimated uncertainties are smaller than the comparisons with drifting buoys suggest to be realistic. Day-time uncertainty metrics are significantly lower but still imply that uncertainties are underestimated in all algorithms.

Overall, OE v2 provides the best performance for most metrics, for both day and night. The bias performance of OE v1 is significantly worse than that for ARC and OE v2, supporting the earlier suggestion that the low OE v1 biases observed for AATSR may be somewhat fortuitous.

Interim conclusions considering ATSR-2 metrics: OEv1 is excluded on the basis of these results, because of poor bias performance, confirming the hypothesis that the good biases for OEv1 for AATSR were fortuitous. OEv2 and ARC are relatively comparable, but OEv2 has better precision and looks preferable overall for ATSR-2. The selection is finely balanced between ARC and OEv2 for AATSR and ATSR-2. Using ARC would maximize algorithmic consistency with ATSR-1, which will use ARC coefficients. However, if OEv2 performs well for the AVHRRs, it is probably more advantageous to maximize algorithmic consistency with the AVHRRs.

5.4 Side-by-side comparison of results relating to Metop

Section 1.1 and 1.1 addressed algorithm selection for AATSR and ATSR-2 respectively. This section addresses the Metop AVHRR, and following sections address other AVHRRs. It is worth noting the contrasts between the ATSRs and AVHRRs, indicated in Table 5.4.1.

ATSR typical characteristics	**AVHRR (GAC data) typical characteristics**
Dual view	Single view
Narrow swath (~500 km)	Wide swath (~3000 km)
1 km resolution in nadir, ~3 km resolution in forward view	4 km resolution at centre swath, ~25 km resolution towards edge of swath
Two-point calibration spanning SST-relevant range of radiance	Single calibration target and cold space
Actively cooled detectors, low noise	Detectors not actively cooled
Controlled ascending node crossing time (2200h or 2230h)	Orbit times drift (except for Metop, controlled to 0930h). Usually a “morning” and “afternoon” AVHRR is available.
Maximum view zenith angle: ~22° in nadir, ~56° in forward	Maximum view zenith angle (used): ~68°

Table 5.4.1: Comparison of ATSR and AVHRR instrument characteristics.

For the latest work on AVHRR calibration and a discussion of the instruments, see Mittaz et al. (2009) [RD.21]. Single view retrievals cannot be as robust to atmospheric aerosol as dual-view retrievals. On average the atmospheric path for AVHRR retrievals is longer than for the ATSR nadir view, which also implies SST retrieval is more challenging. For both ATSR and AVHRR, all portions of the swath width are representatively and adequately sampled in the selection data set.

5.4.1 Presentation of metrics related to SST estimation for Metop

Table 5.4.2: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop, globally.

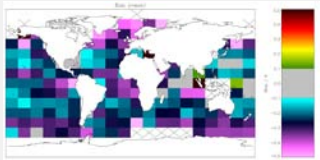
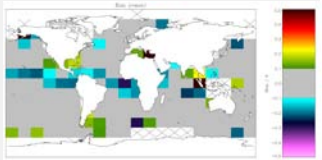
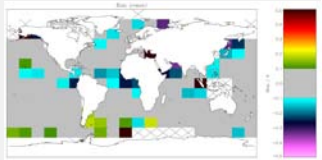
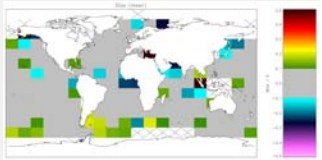
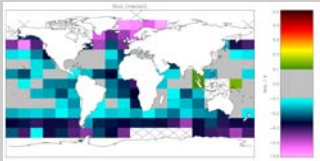
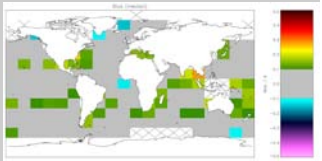
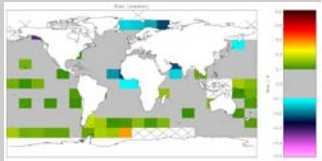
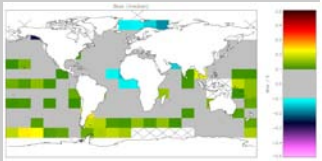
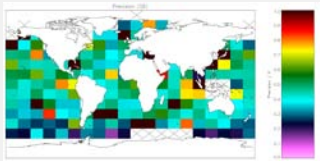
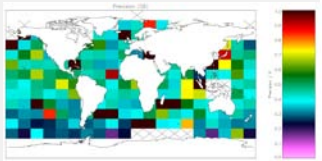
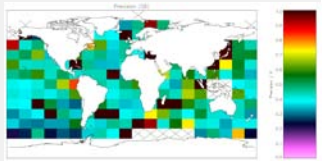
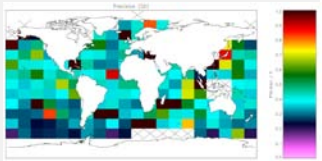
Table 5.4.2 **Metop Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Bias (mean discrepancy)	-0.213 K	-0.001 K	-0.011 K	0.013 K	Very High
Bias (median discrepancy)	-0.139 K	0.069 K	0.060 K	0.070 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

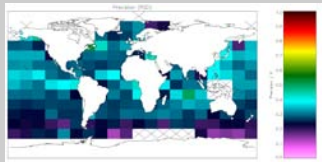
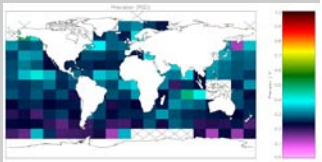
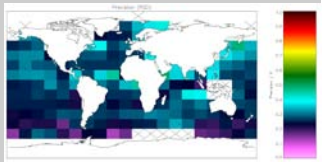
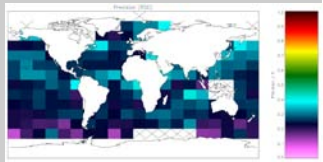
Table 5.4.2 **Metop Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.674 K	0.653 K	0.648 K	0.631 K	Medium
Precision (median of cell RSDs)	0.313 K	0.272 K	0.280 K	0.264 K	Medium
Stability with respect to trend	-0.018 ± 0.005 K/yr	-0.011 ± 0.003 K/yr	-0.019 ± 0.004 K/yr	-0.022 ± 0.003 K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.206 ± 1.236* K Equator: *0.127 ± 0.715* K South: *0.137 ± 1.078* K	North: *0.119 ± 1.211* K Equator: *0.108 ± 0.881* K South: *0.094 ± 1.293* K	North: *0.142 ± 1.171* K Equator: *0.082 ± 0.854* K South: 0*.045 ± 0.673* K	North: *0.138 ± 1.188* K Equator: *0.098 ± 0.804* K South: *0.044 ± 0.537* K	Medium
Stability between day and night	Day-Night: -0.155 ± 0.017 K Trend: *0.012 ± 0.011* K	Day-Night: -0.049 ± 0.016 K Trend: *0.008 ± 0.011* K	Day-Night: -0.068 ± 0.016 K Trend: *0.004 ± 0.011* K	Day-Night: -0.068 ± 0.015 K Trend: *0.007 ± 0.010* K	Medium
Independence from in situ	Fully independent retrieval based on radiative transfer.	Fully independent retrieval based on radiative transfer.	Fully dependent, based on empirical regression to drifting buoys.	Fully dependent, based on empirical regression to drifting buoys.	High

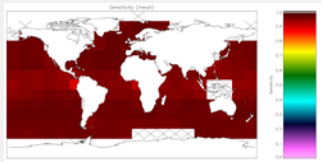
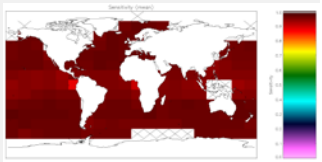
Table 5.4.2 **Metop Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Map SST sensitivity					High
Generality	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Medium
Improvability	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Investigated different algorithm forms. Improvement in prior forward modelling.	Investigated different algorithm forms. Improvement in prior forward modelling.	Medium
Difficulty	No implementation concerns.	No implementation concerns.	No implementation concerns.	No implementation concerns.	Low
Number of Matches	44732	56267	69332	56283	

Table 5.4.2: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop, globally.

Table 5.4.3: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop, globally.

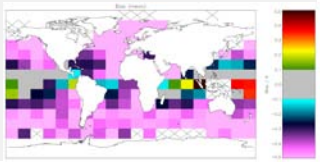
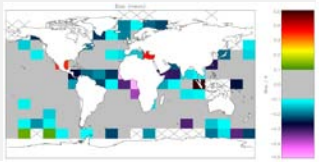
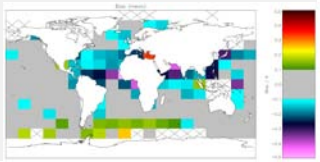
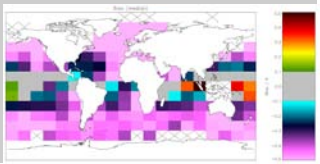
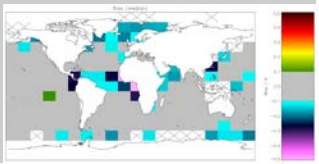
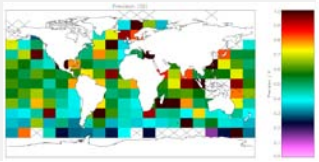
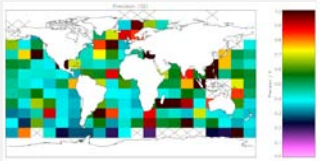
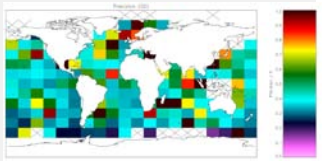
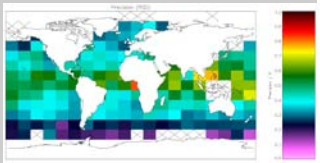
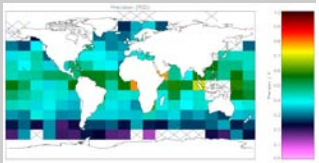
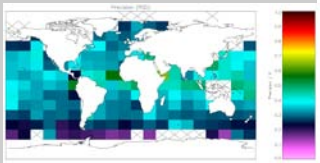
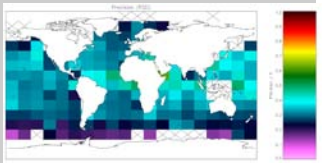
Table 5.4.3 **Metop Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Bias (mean discrepancy)	-0.368 K	-0.050 K	-0.079 K	-0.055 K	Very High
Bias (median discrepancy)	-0.375 K	-0.021 K	-0.036 K	-0.021 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium
Precision map (RSD of discrepancy)					Medium

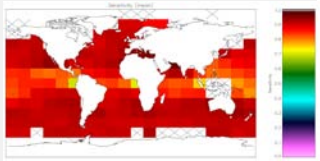
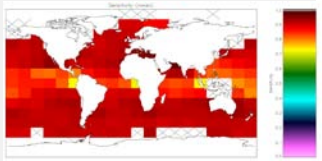
Table 5.4.3 **Metop Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Precision (mean of cell SDs)	0.734 K	0.687 K	0.647 K	0.619 K	Medium
Precision (median of cell RSDs)	0.433 K	0.400 K	0.345 K	0.328 K	Medium
Stability with respect to trend	*-0.006 ± 0.009* K/yr	*-0.003 ± 0.011* K/yr	*-0.015 ± 0.010* K/yr	*-0.015 ± 0.010* K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.286 ± 1.180* K Equator: *0.117 ± 1.058* K South: *0.192 ± 0.835* K	North: *0.230 ± 0.785* K Equator: *0.134 ± 0.832* K South: *0.146 ± 1.245* K	North: *0.139 ± 0.858* K Equator: *0.113 ± 1.021* K South: *0.086 ± 0.887* K	North: *0.121 ± 1.006* K Equator: *0.092 ± 1.097* K South: *0.087 ± 0.834* K	Medium
Stability between day and night	Day-Night: -0.155 ± 0.017 K Trend: *0.012 ± 0.011* K	Day-Night: -0.049 ± 0.016 K Trend: *0.008 ± 0.011* K	Day-Night: -0.068 ± 0.016 K Trend: *0.004 ± 0.011* K	Day-Night: -0.068 ± 0.015 K Trend: *0.007 ± 0.010* K	Medium
Independence from in situ	Fully independent retrieval based on radiative transfer.	Fully independent retrieval based on radiative transfer.	Fully dependent, based on empirical regression to drifting buoys.	Fully dependent, based on empirical regression to drifting buoys.	High
Map SST sensitivity					High

Table 5.4.3 Metop Day	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Generality	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Medium
Improvability	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Investigated different algorithm forms. Improvement in prior forward modelling.	Investigated different algorithm forms. Improvement in prior forward modelling.	Medium
Difficulty	No implementation concerns.	No implementation concerns.	No implementation concerns.	No implementation concerns.	Low
Number of Matches	50657	58363	74162	58391	

Table 5.4.3: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop, globally.

Table 5.4.4: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop, at latitudes above 60° N.

Table 5.4.4 **Metop Night MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.400 K	-0.038 K	-0.122 K	-0.411 K	-0.132 K	Very High
Bias (median discrepancy)	-0.434 K	-0.085 K	-0.135 K	-0.324 K	-0.143 K	Very High
Precision (mean of cell SDs)	0.641 K	0.696 K	0.690 K	0.979 K	0.724 K	Medium
Precision (median of cell RSDs)	0.271 K	0.316 K	0.331 K	0.729 K	0.387 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.499* ± *1.027* K	North: *0.434* ± *1.059* K	North: *0.347* ± *1.046* K	North: *1.087* ± *1.515* K	North: *0.240* ± *1.013* K	Medium
Stability between day and night	Day-Night: -0.191 ± 0.074 K Trend: Not enough match-ups available to	Day-Night: *-0.122* ± *0.073* K Trend: Not enough match-ups available to	Day-Night: *0.025* ± *0.071* K Trend: Not enough match-ups available to	Day-Night: 0.280 ± 0.100 K Trend: Not enough match-ups available to	Day-Night: 0.145 ± 0.076 K Trend: Not enough match-ups available to	Medium

Table 5.4.4 Metop Night MIZ North	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of matches	401	584	762	762	762	

Table 5.4.4: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop, at latitudes above 60° N.

Table 5.4.5: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop, at latitudes above 60° N.

Table 5.4.5 Metop Day MIZ North	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
Bias (mean discrepancy)	-0.591 K	-0.160 K	-0.097 K	-0.131 K	0.013 K	Very High
Bias (median discrepancy)	-0.604 K	-0.167 K	-0.108 K	-0.113 K	-0.002 K	Very High
Precision (mean of cell SDs)	0.726 K	0.709 K	0.701 K	0.998 K	0.767 K	Medium
Precision (median of cell RSDs)	0.339 K	0.309 K	0.305 K	0.718 K	0.381 K	Medium
Stability with respect to trend	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Very High

Table 5.4.5 **Metop Day MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trend.	estimate trend.	estimate trend.	estimate trend.	estimate trend.	
Stability with respect to season (amplitude of cycle)	North: *0.465* ± *0.489* K	North: *0.401* ± *0.480* K	North: *0.286* ± *1.220* K	North: *0.472* ± *0.904* K	North: *0.317* ± *1.355* K	Medium
Stability between day and night	Day-Night: -0.191 ± 0.074 K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.122* ± *0.073* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.025* ± *0.071* K Trend: Not enough match-ups available to estimate trends.	Day-Night: 0.280 ± 0.100 K Trend: Not enough match-ups available to estimate trends.	Day-Night: 0.145 ± 0.076 K Trend: Not enough match-ups available to estimate trends.	Medium
Number of matches	1251	1325	1486	1492	1492	

Table 5.4.5: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop, at latitudes above 60° N.

Table 5.4.6: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop, at latitudes below 50° S.

Table 5.4.6 **Metop Night MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.349 K	-0.022 K	0.046 K	-0.244 K	-0.020 K	Very High
Bias (median discrepancy)	-0.303 K	0.025 K	0.111 K	-0.066 K	0.031 K	Very High
Precision (mean of cell SDs)	1.230 K	1.084 K	0.998 K	1.148 K	1.027 K	Medium
Precision (median of cell RSDs)	0.214 K	0.210 K	0.222 K	0.480 K	0.304 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	South: *0.401* ± *3.320* K	South: *0.328* ± *2.982* K	South: *0.271* ± *2.773* K	South: *0.279* ± *2.872* K	South: *0.372* ± *0.747* K	Medium
Stability between day and night	Day-Night: *-0.140* ± *0.072* K Trend: Not enough match-ups available to	Day-Night: *-0.074* ± *0.063* K Trend: Not enough match-ups available to	Day-Night: *0.048* ± *0.056* K Trend: Not enough match-ups available to	Day-Night: *0.109* ± *0.068* K Trend: Not enough match-ups available to	Day-Night: *0.042* ± *0.059* K Trend: Not enough match-ups available to	Medium

Table 5.4.6 **Metop Night MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of matches	893	1182	1486	1486	1486	

Table 5.4.6: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop, at latitudes below 50° S.

Table 5.4.7: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop, at latitudes below 50° S.

Table 5.4.7 **Metop Day MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.489 K	-0.096 K	0.094 K	-0.135 K	0.022 K	Very High
Bias (median discrepancy)	-0.467 K	-0.081 K	0.113 K	-0.054 K	0.029 K	Very High
Precision (mean of cell SDs)	0.375 K	0.395 K	0.401 K	0.641 K	0.461 K	Medium
Precision (median of cell RSDs)	0.219 K	0.251 K	0.225 K	0.438 K	0.292 K	Medium
Stability with respect to trend	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Very High

Table 5.4.7 **Metop Day MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trend.	estimate trend.	estimate trend.	estimate trend.	estimate trend.	
Stability with respect to season (amplitude of cycle)	South: *0.262* ± *0.606* K	South: *0.268* ± *0.685* K	South: *0.310* ± *0.768* K	South: *0.438* ± *0.776* K	South: *0.351* ± *0.620* K	Medium
Stability between day and night	Day-Night: *-0.140* ± *0.072* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.074* ± *0.063* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.048* ± *0.056* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.109* ± *0.068* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.042* ± *0.059* K Trend: Not enough match-ups available to estimate trends.	Medium
Number of matches	1155	1197	1394	1394	1394	

Table 5.4.7: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop, at latitudes below 50° S.

Table 5.4.8: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop, for coastal regions.

Table 5.4.8 **Metop Night Coastal**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Bias (mean discrepancy)	-0.158 K	0.057 K	-0.006 K	0.033 K	Very High
Bias (median discrepancy)	-0.113 K	0.097 K	0.048 K	0.064 K	Very High
Precision (mean of cell SDs)	0.692 K	0.782 K	0.763 K	0.764 K	Medium
Precision (median of cell RSDs)	0.327 K	0.282 K	0.311 K	0.288 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.335* ± *1.644* K Equator: *0.520* ± *1.224* K South: *0.512* ± *0.598* K	North: *0.296* ± *1.957* K Equator: *0.332* ± *1.234* K South: *0.316* ± *0.531* K	North: *0.263* ± *1.861* K Equator: *0.222* ± *0.676* K South: *0.261* ± *0.635* K	North: *0.327* ± *1.988* K Equator: *0.258* ± *1.144* K South: *0.256* ± *1.107* K	Medium
Stability between day and night	Day-Night: -0.205 ± 0.038 K Trend: Not enough match-ups available to	Day-Night: -0.146 ± 0.037 K Trend: Not enough match-ups available to	Day-Night: -0.164 ± 0.034 K Trend: Not enough match-ups available to	Day-Night: -0.158 ± 0.035 K Trend: Not enough match-ups available to	Medium

Table 5.4.8 Metop Night Coastal	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of Matches	1415	1821	2365	1825	

Table 5.4.8: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop, for coastal regions.

Table 5.4.9: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop, for coastal regions.

Table 5.4.9 Metop Day Coastal	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	-0.363 K	-0.089 K	-0.170 K	-0.125 K	Very High
Bias (median discrepancy)	-0.382 K	-0.037 K	-0.104 K	-0.083 K	Very High
Precision (mean of cell SDs)	0.904 K	0.857 K	0.868 K	0.789 K	Medium
Precision (median of cell RSDs)	0.480 K	0.431 K	0.393 K	0.361 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Very High

Table 5.4.9 **Metop Day Coastal**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
	trend.	trend.	trend.	trend.	
Stability with respect to season (amplitude of cycle)	North: *0.438 ± 1.657* K Equator: *0.909 ± 2.114* K South: *0.797 ± 1.704* K	North: *0.448 ± 1.075* K Equator: *0.669 ± 1.864* K South: *0.662 ± 1.317* K	North: *0.371 ± 1.384* K Equator: *0.700 ± 2.446* K South: *0.496 ± 1.235* K	North: *0.393 ± 1.474* K Equator: *0.503 ± 1.593* K South: *0.619 ± 1.325* K	Medium
Stability between day and night	Day-Night: -0.205 ± 0.038 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.146 ± 0.037 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.164 ± 0.034 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.158 ± 0.035 K Trend: Not enough match-ups available to estimate trends.	Medium
Number of Matches	1750	1956	2561	1958	

Table 5.4.9: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop, for coastal regions.

5.4.2 Presentation of metrics related to SST uncertainty for Metop

Table 5.4.10: Comparison of metrics related to SST uncertainty estimation for night-time AVHRR onboard Metop, globally.

Table 5.4.10 **Metop Night**	**Optimal Estimation v1**		**Optimal Estimation v2**		**Incremental Regression**		**Incremental Regression (with OE2 QC)**		**Weight**
Bias (chi-squared) (Estimate ± 2σ range)	4.4	+0.7 -0.7	3.9	+0.8 -0.6	2.3	+0.2 -0.2	2.2	+0.2 -0.2	High
Independence from in situ	Independent		Independent		Empirical		Empirical		Medium
Generality	General to OE framework.		General to OE framework.		Specific to each match-up dataset used.		Specific to each match-up dataset used.		Low
Improvability	Fuller assessment of forward model error covariance.		Fuller assessment of forward model error covariance.		Investigated different algorithm forms. Improve-ment in prior forward modelling.		Investigated different algorithm forms. Improve-ment in prior forward modelling.		Low
Difficulty	No difficulty given OE framework.		No difficulty given OE framework.		Uncertainty statistics empirically defined outside of processing chain.		Uncertainty statistics empirically defined outside of processing chain.		Low

Table 5.4.10: Comparison of metrics related to SST uncertainty estimation for night-time AVHRR onboard Metop, globally.

Table 5.4.11: Comparison of metrics related to SST uncertainty estimation for day-time AVHRR onboard Metop, globally.

Table 5.4.11 **Metop Day**	**Optimal Estimation v1**		**Optimal Estimation v2**		**Incremental Regression**		**Incremental Regression (with OE2 QC)**		**Weight**
Bias	2.7	+0.4	1.8	+0.3	2.0	+0.2	1.9	+0.1	High

Table 5.4.11 **Metop Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
(chi-squared) (Estimate ± 2σ range)	-0.3	-0.2	-0.1	-0.2	
Independence from in situ	Independent	Independent	Empirical	Empirical	Medium
Generality	General to OE framework.	General to OE framework.	Specific to each match-up dataset used.	Specific to each match-up dataset used.	Low
Improvability	Fuller assessment of forward model error covariance.	Fuller assessment of forward model error covariance.	Investigated different algorithm forms. Improve-ment in prior forward modelling.	Investigated different algorithm forms. Improve-ment in prior forward modelling.	Low
Difficulty	No difficulty given OE framework.	No difficulty given OE framework.	Uncertainty statistics empirically defined outside of processing chain.	Uncertainty statistics empirically defined outside of processing chain.	Low

Table 5.4.11: Comparison of metrics related to SST uncertainty estimation for day-time AVHRR onboard Metop, globally.

5.4.3 Comments on results for Metop

Considering night-time results (using single-view and three channels), it is striking how comparable in performance OE v2 and Incremental Regression (IR) are for Metop. Both give biases well within target overall, with a coherent pattern of negative bias near the equator, whereas OE v1 (untuned to ARC) has widespread negative bias. All algorithms have similar precision, with OE v2 or Inc. Reg. coming out best for robust or standard statistics respectively. There is a statistically detectable trend in each of the results, being largest for IR. Results for stability with respect to season are inconclusive. The best day-night stability is with OE v2. The OE formulations are fully sensitive to SST, while the IR method is almost so. Noting that the OE v2 results are based on fewer matches than the IR results, because of quality level filtering using the OE cost, an additional column of IR results are shown with the same filtering applied (although not available to the IR method). This doesn't materially affect the comparison of the two algorithms.

Turning to the day time results (traditional split window channels used), again the untuned OE v1 is seen to be too biased for SST CCI use. Both OE v2 and IR are within target,

with OE v2 closer to zero. Again, the patterns of bias are similar, although IR has more negative bias in the tropics compensated by some positive bias in the Southern Ocean, not seen in OE v2. Patterns of precision are similar, with IR being significantly less noisy in subtropical regions than OE v2 (and over all). As with the night time results, OE v2 gives apparently better trend stability, although the differences are marginally significant statistically. IR may be better with respect to seasonal stability, although the statistical significance is too low to be sure. OE v2 is fully sensitive for mid and high latitudes and near-fully sensitive in the equator. Day time (two channel) IR is imperfectly sensitive to true SST, typically in range 0.8 to 0.9. While this is a disadvantage to the algorithm on this metric, the decreased sensitivity is probably also the reason that the precision results for IR are better than for OE v2. There is a trade-off. The OE v2 algorithm uses a very loose prior SST (assumed error covariance of $(10\ K)^2$) in order to ensure sensitivity close to 1.0, but this does mean the results are noisier than if a smaller prior were used.

Regarding the qualitative metrics, the main difference is in independence from in situ observations. OE v1 is based on untuned radiative transfer. For OE v2, there is tuning of the radiative transfer, using multi-sensor matches with AATSR observations. This is done without tying the Metop retrieval to in situ, so that OE v2 is also a fully independent retrieval. IR is fully dependent, since the regression coefficients for the increment applied to the prior SST are developed by empirical regression using a traditional match up data set.

Overall, OE v2 seems to be best as regards bias and trend, even when assessed against IR using retrieval-cost-based filtering. IR gives a very low noise retrieval, with precision markedly better than OE v2. Partly, this may reflect the very weak pre-conditioning placed on OE v2, the applied uncertainty in prior SST being 10 K. This is much less certain that even climatological knowledge of SST, and was chosen to ensure high levels of SST sensitivity. So, the selection, if based on Metop on isolation, would be between OE v2 and IR, and would trade off the benefits of best precision (IR) against slightly better bias, sensitivity and stability (OE v2), while factoring in the fact that IR is a dependent approach, while OE v2 preserves independence from in situ observations.

Considering the performance of the algorithms near the marginal ice zone, this has been addressed by extracting results for north and southern high latitude subsets. For these regions, the global algorithms can additionally be considered in the light of high latitude algorithms (CASSTA and a regression-based regional retrieval optimised by DMI). Across the four cases (north/south, day/night) OE v1 and CASSTA perform most poorly. With one exception (north, day) OE v2 is within the bias target. IR and DMI-regional are within the bias target except in the case of night-time observations in the northern high latitudes. OE v2 and IR consistently give similar, relatively good values for precision in high latitudes.

However, for coherence across the AVHRR GAC-based products of SST CCI, the selection on Metop needs to be made in conjunction with the results for the other AVHRRs in the algorithm selection process.

Interim conclusions considering Metop AVHRR: OEv1 is not suitable because of biases. OEv2 and IR are similar in terms of performance on quantitative metrics, although OEv2 is somewhat superior regarding trend stability of night-time observations and day-time biases. IR gives better precision; OEv2 gives better sensitivity; there is a direct trade-off between these qualities. OEv2 is independent of in situ observations, but IR is not. Assuming OEv2 always provides comparable or better results than IR for bias and stability, OEv2 will therefore be preferable. However, algorithmic consistency between AVHRRs is preferable, so the selection depends on the results for the other AVHRRs.

5.5 Side-by-side comparison of results relating to NOAA-19

5.5.1 Presentation of metrics related to SST estimation for NOAA-19

Table 5.5.1: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19, globally.

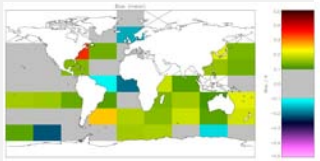
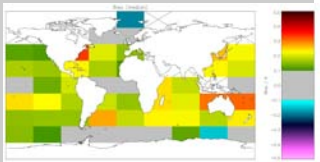
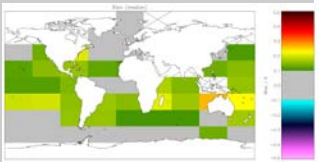
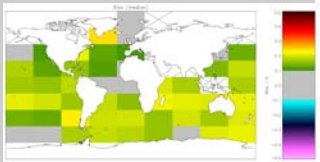
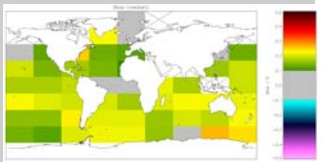
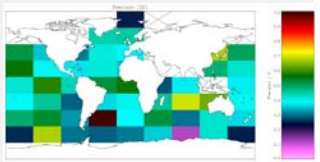
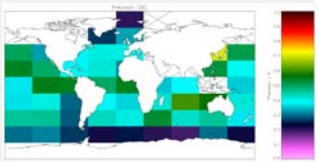
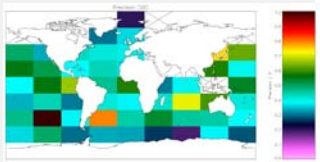
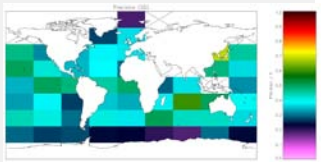
Table 5.5.1 NOAA-19 Night	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	0.109 K	0.052 K	0.072 K	0.104 K	Very High
Bias (median discrepancy)	0.184 K	0.126 K	0.155 K	0.170 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

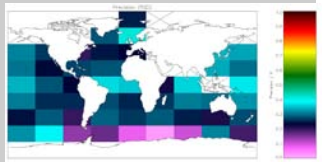
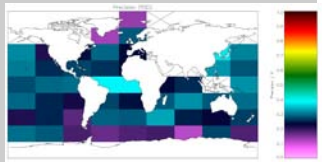
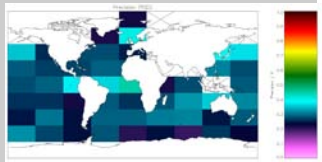
Table 5.5.1 **NOAA-19 Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.487 K	0.441 K	0.502 K	0.410 K	Medium
Precision (median of cell RSDs)	0.290 K	0.280 K	0.292 K	0.271 K	Medium
Stability with respect to trend	*0.011 ± 0.027* K/yr	*0.012 ± 0.012* K/yr	*0.008 ± 0.024* K/yr	*0.005 ± 0.014* K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.190 ± 0.635* K Equator: *0.213 ± 0.603* K South: *0.156 ± 1.403* K	North: *0.158 ± 0.527* K Equator: *0.156 ± 0.686* K South: *0.090 ± 0.767* K	North: *0.121 ± 0.793* K Equator: *0.125 ± 0.796* K South: *0.144 ± 0.848* K	North: *0.110 ± 0.557* K Equator: *0.166 ± 0.621* K South: *0.138 ± 0.737* K	Medium
Stability between day and night	Day-Night: 0.188 ± 0.021 K Trend: Not enough match-ups available to estimate day-time trend.	Day-Night: -0.213 ± 0.017 K Trend: *-0.017 ± 0.022* K	Day-Night: -0.367 ± 0.016 K Trend: *-0.025 ± 0.056* K	Day-Night: -0.336 ± 0.014 K Trend: *-0.004 ± 0.049* K	Medium
Independence	Fully independent	Fully independent	Fully dependent, based	Fully dependent, based	High

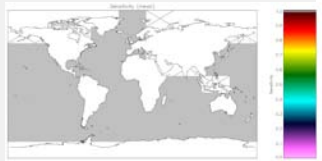
Table 5.5.1 **NOAA-19 Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
from in situ	retrieval based on radiative transfer.	retrieval based on radiative transfer.	on empirical regression to drifting buoys.	on empirical regression to drifting buoys.	
Map SST sensitivity					High
Generality	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Medium
Improvability	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Investigated different algorithm forms. Improvement in prior forward modelling.	Investigated different algorithm forms. Improvement in prior forward modelling.	Medium
Difficulty	No implementation concerns.	No implementation concerns.	No implementation concerns.	No implementation concerns.	Low
Number of Matches	8028	13693	17202	13696	

Table 5.5.1: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19, globally.

Table 5.5.2: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19, globally.

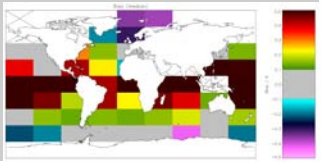
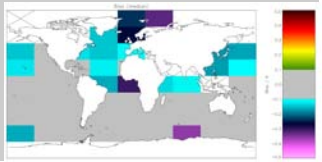
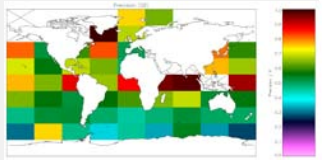
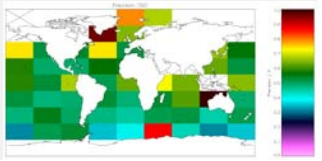
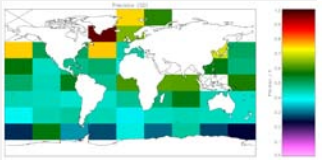
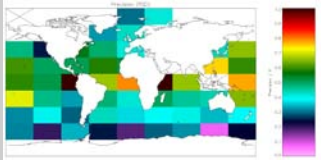
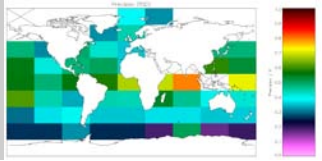
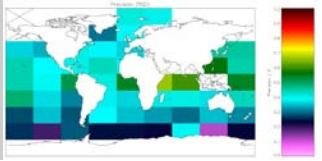
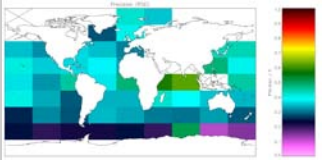
Table 5.5.2 NOAA-19 Day	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	0.297 K	-0.161 K	-0.295 K	-0.232 K	Very High
Bias (median discrepancy)	0.275 K	-0.076 K	-0.206 K	-0.169 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium
Precision map (RSD of discrepancy)					Medium

Table 5.5.2 **NOAA-19 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Precision (mean of cell SDs)	0.779 K	0.658 K	0.588 K	0.523 K	Medium
Precision (median of cell RSDs)	0.583 K	0.462 K	0.415 K	0.372 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	*-0.005 ± 0.018* K/yr	*-0.018 ± 0.051* K/yr	*0.001 ± 0.047* K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.471 ± 1.279* K Equator: *0.193 ± 1.316* K South: *0.368 ± 0.841* K	North: *0.346 ± 0.897* K Equator: *0.143 ± 1.032* K South: *0.128 ± 0.823* K	North: *0.335 ± 0.795* K Equator: *0.143 ± 0.804* K South: *0.101 ± 0.734* K	North: *0.300 ± 0.737* K Equator: *0.121 ± 0.710* K South: *0.116 ± 0.651* K	Medium
Stability between day and night	Day-Night: 0.188 ± 0.021 K Trend: Not enough match-ups available to estimate day-time trend.	Day-Night: -0.213 ± 0.017 K Trend: *-0.017 ± 0.022* K	Day-Night: -0.367 ± 0.016 K Trend: *-0.025 ± 0.056* K	Day-Night: -0.336 ± 0.014 K Trend: *-0.004 ± 0.049* K	Medium
Independence from in situ	Fully independent retrieval based on radiative transfer.	Fully independent retrieval based on radiative transfer.	Fully dependent, based on empirical regression to drifting buoys.	Fully dependent, based on empirical regression to drifting buoys.	High

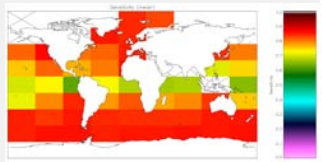
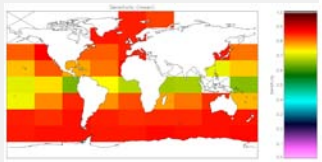
Table 5.5.2 **NOAA-19 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Map SST sensitivity					High
Generality	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Medium
Improvability	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Investigated different algorithm forms. Improvement in prior forward modelling.	Investigated different algorithm forms. Improvement in prior forward modelling.	Medium
Difficulty	No implementation concerns.	No implementation concerns.	No implementation concerns.	No implementation concerns.	Low
Number of Matches	11566	18381	24948	18383	

Table 5.5.2: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19, globally.

Table 5.5.3: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19, at latitudes above 60° N.

Table 5.5.3 **NOAA-19 Night MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.140 K	0.192 K	0.253 K	-0.298 K	0.191 K	Very High
Bias (median discrepancy)	-0.147 K	0.061 K	0.057 K	-0.198 K	0.066 K	Very High
Precision (mean of cell SDs)	0.297 K	0.691 K	0.825 K	1.045 K	0.817 K	Medium
Precision (median of cell RSDs)	0.370 K	0.205 K	0.237 K	0.650 K	0.360 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.436* ± *0.277* K	North: *0.557* ± *1.075* K	North: *0.811* ± *1.480* K	North: *1.608* ± *1.800* K	North: *0.928* ± *1.174* K	Medium
Stability between day and night	Day-Night: *-0.198* ± *0.113* K Trend: Not enough match-ups available to	Day-Night: -0.437 ± 0.140 K Trend: Not enough match-ups available to	Day-Night: -0.552 ± 0.148 K Trend: Not enough match-ups available to	Day-Night: *-0.167* ± *0.201* K Trend: Not enough match-ups available to	Day-Night: *-0.202* ± *0.151* K Trend: Not enough match-ups available to	Medium

Table 5.5.3 **NOAA-19 Night MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of matches	14	58	78	78	78	

Table 5.5.3: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19, at latitudes above 60° N.

Table 5.5.4: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19, at latitudes above 60° N.

Table 5.5.4 **NOAA-19 Day MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.338 K	-0.245 K	-0.299 K	-0.465 K	-0.011 K	Very High
Bias (median discrepancy)	-0.358 K	-0.268 K	-0.325 K	-0.343 K	-0.084 K	Very High
Precision (mean of cell SDs)	0.560 K	0.751 K	0.741 K	1.221 K	0.824 K	Medium
Precision (median of cell RSDs)	0.344 K	0.345 K	0.357 K	0.929 K	0.408 K	Medium
Stability with respect to trend	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Very High

Table 5.5.4 **NOAA-19 Day MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trend.	estimate trend.	estimate trend.	estimate trend.	estimate trend.	
Stability with respect to season (amplitude of cycle)	North: *0.719* ± *1.301* K	North: *0.463* ± *1.221* K	North: *0.623* ± *1.201* K	North: *1.158* ± *1.193* K	North: *0.695* ± *1.066* K	Medium
Stability between day and night	Day-Night: *-0.198* ± *0.113* K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.437 ± 0.140 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.552 ± 0.148 K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.167* ± *0.201* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.202* ± *0.151* K Trend: Not enough match-ups available to estimate trends.	Medium
Number of matches	256	561	643	643	643	

Table 5.5.4: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19, at latitudes above 60° N.

Table 5.5.5: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19, at latitudes below 50° S.

Table 5.5.5 **NOAA-19 Night MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	0.015 K	0.042 K	0.137 K	-0.270 K	-0.054 K	Very High
Bias (median discrepancy)	0.040 K	0.052 K	0.182 K	-0.102 K	-0.004 K	Very High
Precision (mean of cell SDs)	0.338 K	0.249 K	0.337 K	0.736 K	0.443 K	Medium
Precision (median of cell RSDs)	0.204 K	0.195 K	0.249 K	0.596 K	0.369 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	South: *0.306* ± *0.469* K	South: *0.096* ± *0.421* K	South: *0.334* ± *0.616* K	South: *0.627* ± *0.954* K	South: *0.411* ± *0.607* K	Medium
Stability between day and night	Day-Night: -0.115 ± 0.045 K Trend: Not enough match-ups available to	Day-Night: -0.143 ± 0.038 K Trend: Not enough match-ups available to	Day-Night: -0.160 ± 0.043 K Trend: Not enough match-ups available to	Day-Night: *-0.033* ± *0.083* K Trend: Not enough match-ups available to	Day-Night: *-0.078* ± *0.054* K Trend: Not enough match-ups available to	Medium

Table 5.5.5 **NOAA-19 Night MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of matches	169	419	547	547	547	

Table 5.5.5: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19, at latitudes below 50° S.

Table 5.5.6: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19, at latitudes below 50° S.

Table 5.5.6 **NOAA-19 Day MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.100 K	-0.101 K	-0.023 K	-0.303 K	-0.132 K	Very High
Bias (median discrepancy)	-0.076 K	-0.051 K	0.037 K	-0.144 K	-0.073 K	Very High
Precision (mean of cell SDs)	0.317 K	0.395 K	0.440 K	0.805 K	0.541 K	Medium
Precision (median of cell RSDs)	0.250 K	0.253 K	0.281 K	0.610 K	0.398 K	Medium
Stability with respect to trend	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Very High

Table 5.5.6 **NOAA-19 Day MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trend.	estimate trend.	estimate trend.	estimate trend.	estimate trend.	
Stability with respect to season (amplitude of cycle)	South: *0.325* ± *0.262* K	South: *0.382* ± *0.581* K	South: *0.250* ± *1.013* K	South: *0.985* ± *1.550* K	South: *0.460* ± *1.277* K	Medium
Stability between day and night	Day-Night: -0.115 ± 0.045 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.143 ± 0.038 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.160 ± 0.043 K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.033* ± *0.083* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.078* ± *0.054* K Trend: Not enough match-ups available to estimate trends.	Medium
Number of matches	145	294	346	346	346	

Table 5.5.6: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19, at latitudes below 50° S.

Table 5.5.7: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19, for coastal regions.

Table 5.5.7 **NOAA-19 Night Coastal**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Bias (mean discrepancy)	0.158 K	0.114 K	0.055 K	0.130 K	Very High
Bias (median discrepancy)	0.204 K	0.166 K	0.131 K	0.159 K	Very High
Precision (mean of cell SDs)	0.424 K	0.508 K	0.600 K	0.472 K	Medium
Precision (median of cell RSDs)	0.279 K	0.252 K	0.282 K	0.246 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.278* ± *0.673* K Equator: *0.887* ± *0.547* K South: *0.571* ± *0.692* K	North: *0.718* ± *1.179* K Equator: *0.871* ± *0.906* K South: *0.541* ± *0.399* K	North: *0.829* ± *1.675* K Equator: *0.492* ± *0.718* K South: *1.372* ± *1.861* K	North: *0.480* ± *1.131* K Equator: *0.682* ± *0.834* K South: *0.445* ± *0.477* K	Medium
Stability between day and night	Day-Night: 0.165 ± 0.060 K Trend: Not enough match-ups available to	Day-Night: -0.252 ± 0.048 K Trend: Not enough match-ups available to	Day-Night: -0.380 ± 0.043 K Trend: Not enough match-ups available to	Day-Night: -0.384 ± 0.042 K Trend: Not enough match-ups available to	Medium

Table 5.5.7 NOAA-19 Night Coastal	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of Matches	248	436	592	437	

Table 5.5.7: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19, for coastal regions.

Table 5.5.8: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19, for coastal regions.

Table 5.5.8 NOAA-19 Day Coastal	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	0.323 K	-0.138 K	-0.325 K	-0.254 K	Very High
Bias (median discrepancy)	0.238 K	-0.080 K	-0.248 K	-0.198 K	Very High
Precision (mean of cell SDs)	0.878 K	0.752 K	0.650 K	0.623 K	Medium
Precision (median of cell RSDs)	0.580 K	0.525 K	0.454 K	0.420 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Very High

Table 5.5.8 **NOAA-19 Day Coastal**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
	trend.	trend.	trend.	trend.	
Stability with respect to season (amplitude of cycle)	North: *0.457* ± *1.668* K Equator*:* *1.367* ± *2.312* K South: *1.187* ± *1.016* K	North: *0.593* ± *1.525* K Equator: *2.142* ± *1.177* K South: *0.518* ± *0.962* K	North: *0.519* ± *0.716* K Equator: *0.852* ± *1.297* K South: *0.515* ± *0.803* K	North: *0.572* ± *1.418* K Equator: *1.334* ± *1.021* K South: *0.549* ± *0.845* K	Medium
Stability between day and night	Day-Night: 0.165 ± 0.060 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.252 ± 0.048 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.380 ± 0.043 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.384 ± 0.042 K Trend: Not enough match-ups available to estimate trends.	Medium
Number of Matches	435	657	949	661	

Table 5.5.8: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19, for coastal regions.

5.5.2 Presentation of metrics related to SST uncertainty for NOAA-19

Table 5.5.9: Comparison of metrics related to SST uncertainty estimation for night-time AVHRR onboard NOAA-19, globally.

Table 5.5.9 **NOAA-19 Night**	**Optimal Estimation v1**		**Optimal Estimation v2**		**Incremental Regression**		**Incremental Regression (with OE2 QC)**		**Weight**
Bias (chi-squared) (Estimate ± 2σ range)	1.9	+0.2 -0.3	1.6	+0.2 -0.3	1.3	+0.1 -0.1	0.9	+0.1 -0.1	High
Independence from in situ	Independent		Independent		Empirical		Empirical		Medium
Generality	General to OE framework.		General to OE framework.		Specific to each matchup dataset used.		Specific to each matchup dataset used.		Low
Improvability	Fuller assessment of forward model error covariance.		Fuller assessment of forward model error covariance.		Investigated different algorithm forms. Improve-ment in prior forward modelling.		Investigated different algorithm forms. Improve-ment in prior forward modelling.		Low
Difficulty	No difficulty given OE framework.		No difficulty given OE framework.		Uncertainty statistics empirically defined outside of processing chain.		Uncertainty statistics empirically defined outside of processing chain.		Low

Table 5.5.9: Comparison of metrics related to SST uncertainty estimation for night-time AVHRR onboard NOAA-19, globally.

Table 5.5.10: Comparison of metrics related to SST uncertainty estimation for day-time AVHRR onboard NOAA-19, globally.

Table 5.5.10 **NOAA-19 Day**	**Optimal Estimation v1**		**Optimal Estimation v2**		**Incremental Regression**		**Incremental Regression (with OE2 QC)**		**Weight**
Bias	1.1	+0.1	1.2	+0.1	1.2	+0.0	0.9	+0.0	High

Table 5.5.10 **NOAA-19 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
(chi-squared) (Estimate ± 2σ range)	-0.0	-0.1	-0.1	-0.1	
Independence from in situ	Independent	Independent	Empirical	Empirical	Medium
Generality	General to OE framework.	General to OE framework.	Specific to each matchup dataset used.	Specific to each matchup dataset used.	Low
Improvability	Fuller assessment of forward model error covariance.	Fuller assessment of forward model error covariance.	Investigated different algorithm forms. Improve-ment in prior forward modelling.	Investigated different algorithm forms. Improve-ment in prior forward modelling.	Low
Difficulty	No difficulty given OE framework.	No difficulty given OE framework.	Uncertainty statistics empirically defined outside of processing chain.	Uncertainty statistics empirically defined outside of processing chain.	Low

Table 5.5.10: Comparison of metrics related to SST uncertainty estimation for day-time AVHRR onboard NOAA-19, globally.

5.5.3 Comments on results for NOAA-19

Considering the night-time results (single-view and three-channels used for retrieval), the bias for OE v2 and IR are both within target in the mean, with OE v2 around 0.02 K closer to zero. The regional biases are in target for more of the ocean in the case of OE v2. The precision of retrieval is better for OE v2 than for IR, although only marginally so when assessed using robust statistics. None of the tested algorithms gives a statistically significant trend. Day-night stability is poor for all, reflecting the poor bias for the day-time results discussed in the next paragraph. The OE retrievals are fully sensitive to SST and the IR is not fully sensitive, with cell average sensitivity typically between 0.6 and 0.85. The OE v2 algorithm defines a quality level in terms of retrieval cost and filters a proportion of the results. These may be cases that are slightly cloud contaminated (but not enough to be screened) and so it is not surprising that when filtered in the same way, the IR results become more positive, with a positive bias slightly outside of the target (rightmost column). The filtering improves the IR precision (although of course the filtering is only available when using OE v2).

The day-time results show a markedly less satisfactory outcome. All the algorithms are outside of the 0.1 K target for bias, the best being OE v2 with a bias of -0.161 K in the mean (or -0.076 K in the median). The negative biases are centred on the northern mid latitudes,. Although not formal SST CCI selection metrics, we considered the results as a function of various factors other than latitude, to try to understand them. The OE v2 showed relatively little variation against most factors, but did reveal a marked dependence on solar zenith angle being smaller than about 20 deg (Figure 5.5.3.1). A similar effect is present in the IR result (Figure 5.5.3.2). The explanation for this surprising finding has not been established, and, indeed, cannot be determined for schedule reasons before algorithm selection has to take place. Therefore the selection has to be made on the results available.

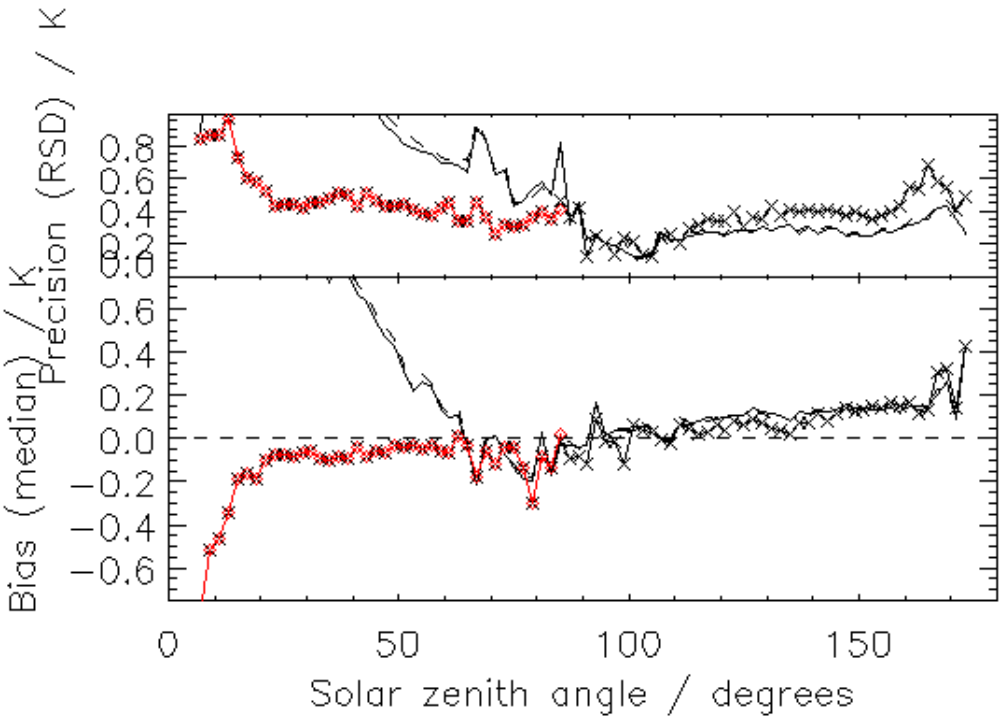


Figure 5.5.3.1: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for OE v2 retrievals for AVHRR onboard NOAA-19. The meaning of different lines/symbols is as follows: the line with symbols represents N2 retrievals (with red used to highlight daytime retrievals), the solid line (with no symbols) represents N3 retrievals, and the dashed line represents N2* (3.7 µm and 11 µm channel) retrievals.

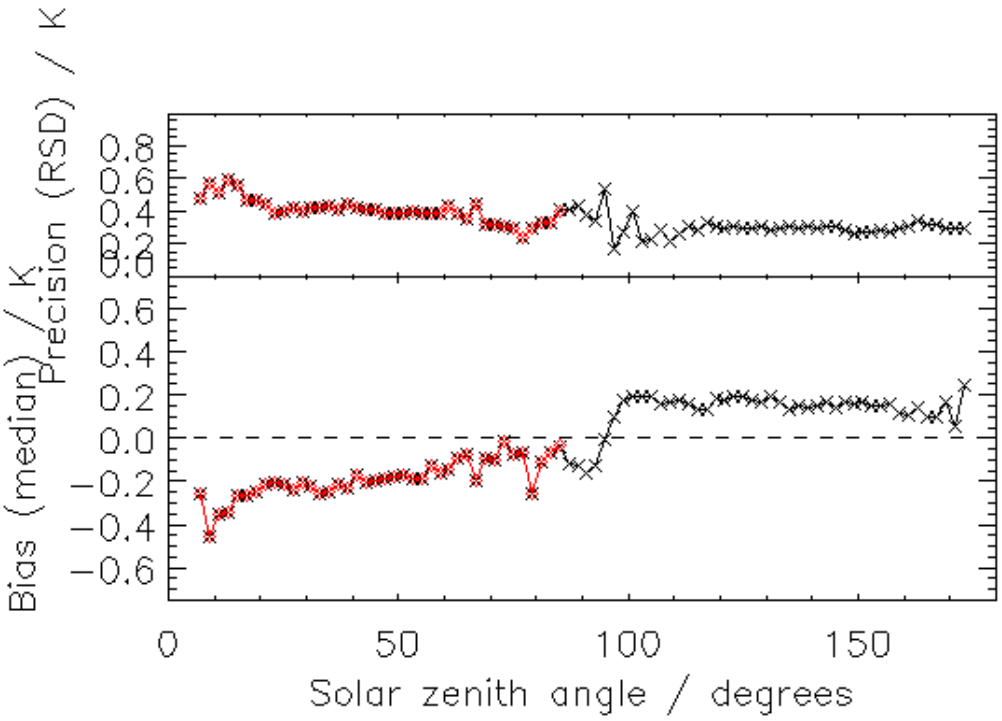


Figure 5.5.3.2: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for incremental regression retrievals for AVHRR onboard NOAA-19. The red and black lines with symbols represent retrievals using the daytime and night-time formulations of the incremental regression retrieval respectively.

Precision is better for IR than for OE v2, probably for similar reasons to those explained in the context of Metop GAC (section 5.4.3). No statistically significant temporal trend is found, while the results for stability with respect to season are statistically ambiguous.

Applying the OE v2 retrieval-cost-based filtering to IR improves the bias characteristics for IR, although OE v2 remains better. The precision of the IR results is also markedly improved and markedly better than for OE v2.

Comments on the qualitative metrics are as for Metop GAC (section 5.4.3).

Table 5.5.3 to Table 5.5.6 are provided of statistics of high-latitude matches, since the high latitudes and marginal ice zone (MIZ) are of interest. In these cases, the DMI CASSTA and regional algorithms are also provided. The CASSTA approach is not suited to wide-area application, as shown by the relatively high biases and precision it displays. The DMI regional algorithm does a good job in some cases (north, night-time; south, day-time), but generally has large (poor) precision compared to OE v2. With the exception of day-time southern hemisphere results, IR SST is more biased than from OE v2. The precision statistics are generally better for OE v2 than other methods. The stability performance in high latitudes / MIZ is difficult to compare because of the relatively few matches.

The uncertainty estimates associated with OE v2 are generally optimistic for night-time cases, while for day-time SSTs they are close to being unbiased. The uncertainty estimate for IR SSTs is a general value (same estimate applies to all cases), unlike for optimal estimation; this uncertainty on average is about right (the metric is close to 1.0 day and night).

Interim conclusions considering NOAA-19 AVHRR: here, OEv2 is better for bias and precision in night-time metrics, although both OEv2 and IR look appropriate for use for night-time NOAA-19 AVHRR. Both OEv2 and IR show disappointingly large biases for day-time observations, which may partly arise as a function of solar zenith angle. Applying some restriction on valid ranges may help keep biases within the SST CCI target. Overall, OEv2 is preferable to IR on the basis of NOAA-19 AVHRR.

5.6 Side-by-side comparison of results relating to NOAA-18

5.6.1 Presentation of metrics related to SST estimation for NOAA-18

Table 5.6.1: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18, globally.

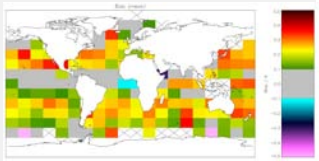
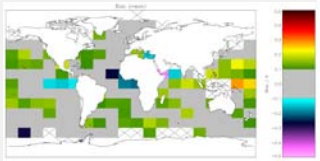
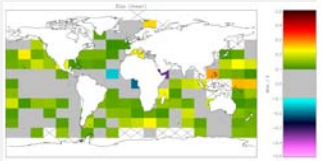
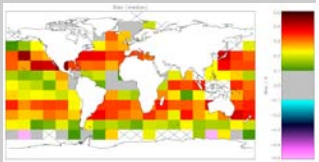
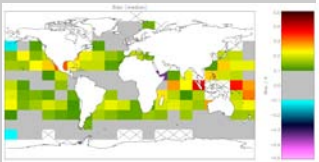
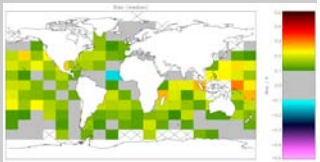
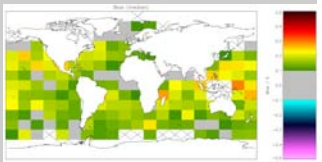
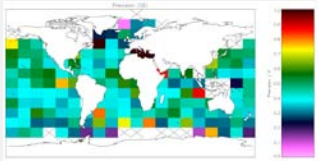
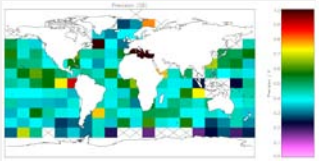
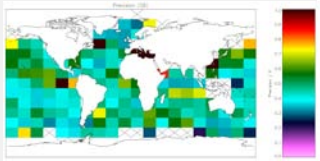
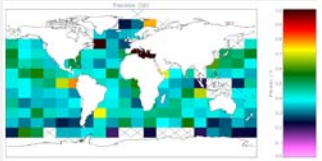
Table 5.6.1 **NOAA-18 Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Bias (mean discrepancy)	0.206 K	0.074 K	0.071 K	0.107 K	Very High
Bias (median discrepancy)	0.282 K	0.147 K	0.148 K	0.163 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

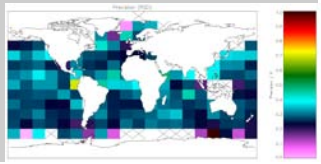
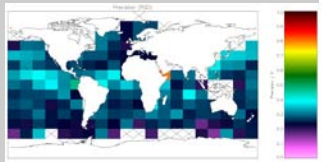
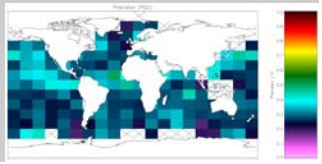
Table 5.6.1 **NOAA-18 Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.466 K	0.477 K	0.519 K	0.448 K	Medium
Precision (median of cell RSDs)	0.289 K	0.289 K	0.298 K	0.277 K	Medium
Stability with respect to trend	0.017 ± 0.004 K/yr	*0.002* ± *0.004* K/yr	0.013 ± 0.004 K/yr	0.013 ± 0.004 K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.140* ± *0.654* K Equator: *0.089* ± *0.630* K South: *0.082* ± *0.582* K	North: *0.136* ± *0.940* K Equator: *0.118* ± *0.664* K South: *0.098* ± *0.576* K	North: *0.139* ± *1.062* K Equator: *0.104* ± *0.782* K South: *0.093* ± *0.637* K	North: *0.142* ± *0.921* K Equator: *0.087* ± *0.587* K South: *0.096* ± *0.548* K	Medium
Stability between day and night	Day-Night: -0.147 ± 0.014 K Trend: 0.025 ± 0.010 K	Day-Night: -0.235 ± 0.014 K Trend: *-0.000* ± *0.009* K	Day-Night: -0.362 ± 0.014 K Trend: *0.003* ± *0.008* K	Day-Night: -0.348 ± 0.013 K Trend: *0.002* ± *0.008* K	Medium
Independence from in situ	Fully independent retrieval based on radiative transfer.	Fully independent retrieval based on radiative transfer.	Fully dependent, based on empirical regression to drifting buoys.	Fully dependent, based on empirical regression to drifting buoys.	High

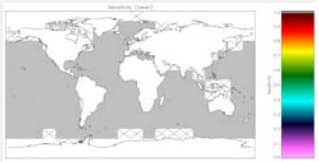
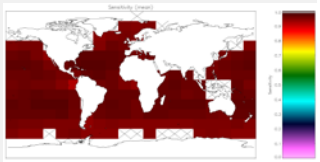
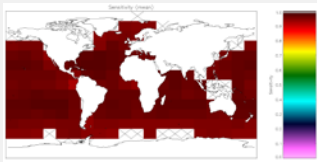
Table 5.6.1 **NOAA-18 Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Map SST sensitivity					High
Generality	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Medium
Improvability	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Investigated different algorithm forms. Improvement in prior forward modelling.	Investigated different algorithm forms. Improvement in prior forward modelling.	Medium
Difficulty	No implementation concerns.	No implementation concerns.	No implementation concerns.	No implementation concerns.	Low
Number of Matches	37785	49505	61250	48810	

Table 5.6.1: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18, globally.

Table 5.6.2: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18, globally.

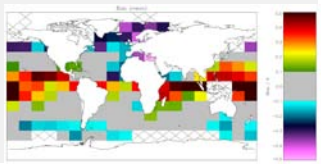
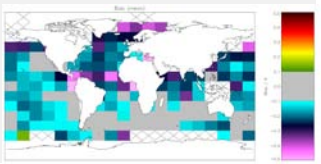
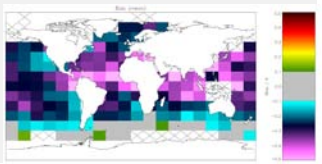
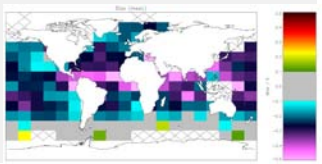
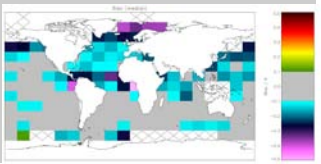
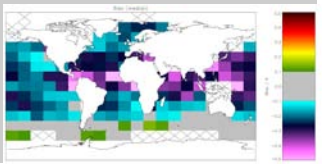
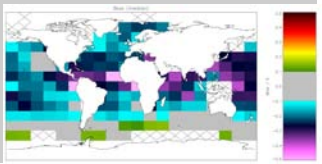
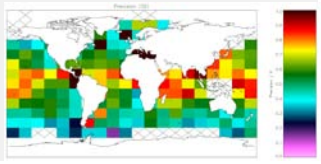
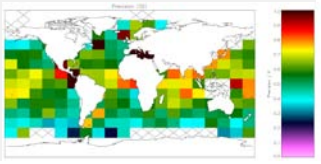
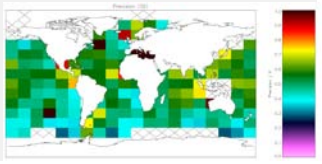
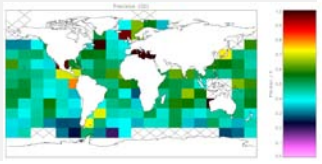
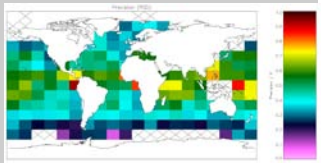
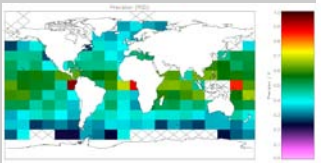
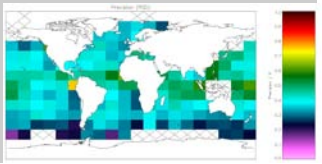
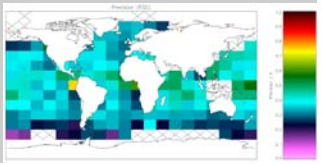
Table 5.6.2 NOAA-18 Day	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	0.059 K	-0.161 K	-0.291 K	-0.241 K	Very High
Bias (median discrepancy)	0.077 K	-0.090 K	-0.214 K	-0.184 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium
Precision map (RSD of discrepancy)					Medium

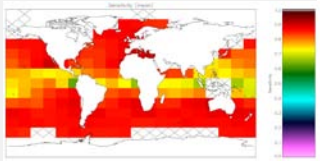
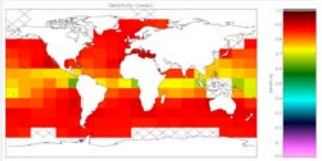
Table 5.6.2 **NOAA-18 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Precision (mean of cell SDs)	0.723 K	0.687 K	0.641 K	0.601 K	Medium
Precision (median of cell RSDs)	0.491 K	0.465 K	0.409 K	0.381 K	Medium
Stability with respect to trend	0.042 ± 0.010 K/yr	*0.002 ± 0.008* K/yr	*0.016 ± 0.008* K/yr	0.015 ± 0.007 K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.408 ± 1.631* K Equator: *0.301 ± 1.050* K South: *0.203 ± 0.818* K	North: *0.219 ± 1.512* K Equator: *0.184 ± 0.960* K South: *0.103 ± 0.776* K	North: *0.174 ± 0.904* K Equator: *0.164 ± 0.791* K South: *0.115 ± 0.718* K	North: *0.150 ± 0.852* K Equator: *0.175 ± 0.705* K South: *0.134 ± 0.669* K	Medium
Stability between day and night	Day-Night: -0.147 ± 0.014 K Trend: 0.025 ± 0.010 K	Day-Night: -0.235 ± 0.014 K Trend: *-0.000 ± 0.009* K	Day-Night: -0.362 ± 0.014 K Trend: *0.003 ± 0.008* K	Day-Night: -0.348 ± 0.013 K Trend: *0.002 ± 0.008* K	Medium
Independence from in situ	Fully independent retrieval based on radiative transfer.	Fully independent retrieval based on radiative transfer.	Fully dependent, based on empirical regression to drifting buoys.	Fully dependent, based on empirical regression to drifting buoys.	High
Map SST sensitivity					High

Table 5.6.2 **NOAA-18 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Generality	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Medium
Improvability	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Investigated different algorithm forms. Improvement in prior forward modelling.	Investigated different algorithm forms. Improvement in prior forward modelling.	Medium
Difficulty	No implementation concerns.	No implementation concerns.	No implementation concerns.	No implementation concerns.	Low
Number of Matches	57165	69579	91370	68726	

Table 5.6.2: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18, globally.

Table 5.6.3: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18, at latitudes above 60° N.

Table 5.6.3 **NOAA-18 Night MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	0.010 K	0.024 K	-0.016 K	-0.497 K	-0.006 K	Very High
Bias (median discrepancy)	0.009 K	0.043 K	0.032 K	-0.335 K	0.042 K	Very High
Precision (mean of cell SDs)	0.374 K	0.499 K	0.498 K	0.874 K	0.571 K	Medium
Precision (median of cell RSDs)	0.301 K	0.285 K	0.300 K	0.657 K	0.408 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.455* ± *0.584* K	North: 0.*353* ± *0.814* K	North: *0.287* ± *0.774* K	North: *1.093* ± *1.498* K	North: *0.556* ± *0.769* K	Medium
Stability between day and night	Day-Night: -0.342 ± 0.067 K Trend: Not enough match-ups available to	Day-Night: -0.367 ± 0.062 K Trend: Not enough match-ups available to	Day-Night: -0.238 ± 0.061 K Trend: Not enough match-ups available to	Day-Night: *-0.051* ± *0.109* K Trend: Not enough match-ups available to	Day-Night: *-0.007* ± *0.068* K Trend: Not enough match-ups available to	Medium

Table 5.6.3 **NOAA-18 Night MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of matches	69	258	351	358	358	

Table 5.6.3: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18, at latitudes above 60° N.

Table 5.6.4: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18, at latitudes above 60° N.

Table 5.6.4 **NOAA-18 Day MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.332 K	-0.343 K	-0.254 K	-0.548 K	-0.013 K	Very High
Bias (median discrepancy)	-0.323 K	-0.313 K	-0.218 K	-0.419 K	-0.005 K	Very High
Precision (mean of cell SDs)	0.620 K	0.613 K	0.612 K	1.129 K	0.670 K	Medium
Precision (median of cell RSDs)	0.339 K	0.347 K	0.357 K	1.043 K	0.425 K	Medium
Stability with respect to trend	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Very High

Table 5.6.4 **NOAA-18 Day MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trend.	estimate trend.	estimate trend.	estimate trend.	estimate trend.	
Stability with respect to season (amplitude of cycle)	North: *0.554* ± *0.560* K	North: *0.540* ± *0.465* K	North: *0.338* ± *0.518* K	North: *1.093* ± *2.163* K	North: *0.530* ± *0.598* K	Medium
Stability between day and night	Day-Night: -0.342 ± 0.067 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.367 ± 0.062 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.238 ± 0.061 K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.051* ± *0.109* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.007* ± *0.068* K Trend: Not enough match-ups available to estimate trends.	Medium
Number of matches	1214	1671	1855	1874	1874	

Table 5.6.4: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18, at latitudes above 60° N.

Table 5.6.5: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18, at latitudes below 50° S.

Table 5.6.5 **NOAA-18 Night MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	0.107 K	-0.028 K	0.058 K	-0.277 K	0.065 K	Very High
Bias (median discrepancy)	0.117 K	-0.020 K	0.102 K	-0.094 K	0.097 K	Very High
Precision (mean of cell SDs)	0.355 K	0.324 K	0.405 K	0.769 K	0.531 K	Medium
Precision (median of cell RSDs)	0.237 K	0.231 K	0.261 K	0.570 K	0.348 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	South: *0.516* ± *0.676* K	South: *0.224* ± *0.552* K	South: *0.370* ± *0.731* K	South: *0.503* ± *1.034* K	South: *0.537* ± *0.868* K	Medium
Stability between day and night	Day-Night: -0.185 ± 0.032 K Trend: Not enough match-ups available to	Day-Night: -0.085 ± 0.028 K Trend: Not enough match-ups available to	Day-Night: *-0.037* ± *0.030* K Trend: Not enough match-ups available to	Day-Night: *-0.090* ± *0.058* K Trend: Not enough match-ups available to	Day-Night: -0.098 ± 0.041 K Trend: Not enough match-ups available to	Medium

Table 5.6.5 NOAA-18 Night MIZ South	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of matches	513	1117	1441	1491	1491	

Table 5.6.5: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18, at latitudes below 50° S.

Table 5.6.6: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18, at latitudes below 50° S.

Table 5.6.6 NOAA-18 Day MIZ South	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
Bias (mean discrepancy)	-0.078 K	-0.113 K	0.021 K	-0.367 K	-0.033 K	Very High
Bias (median discrepancy)	-0.063 K	-0.092 K	0.065 K	-0.193 K	0.009 K	Very High
Precision (mean of cell SDs)	0.375 K	0.410 K	0.412 K	0.787 K	0.549 K	Medium
Precision (median of cell RSDs)	0.246 K	0.304 K	0.261 K	0.646 K	0.361 K	Medium
Stability with respect to trend	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Very High

Table 5.6.6 NOAA-18 Day MIZ South	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
	estimate trend.	estimate trend.	estimate trend.	estimate trend.	estimate trend.	
Stability with respect to season (amplitude of cycle)	South: *0.203* ± *0.445* K	South: *0.211* ± *0.523* K	South: *0.127* ± *0.590* K	South: *0.470* ± *1.172* K	South: *0.222* ± *0.646* K	Medium
Stability between day and night	Day-Night: -0.185 ± 0.032 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.085 ± 0.028 K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.037* ± *0.030* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.090* ± *0.058* K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.098 ± 0.041 K Trend: Not enough match-ups available to estimate trends.	Medium
Number of matches	963	1377	1545	1640	1640	

Table 5.6.6: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18, at latitudes below 50° S.

Table 5.6.7: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18, for coastal regions.

Table 5.6.7 **NOAA-18 Night Coastal**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Bias (mean discrepancy)	0.215 K	0.088 K	0.043 K	0.105 K	Very High
Bias (median discrepancy)	0.300 K	0.160 K	0.139 K	0.162 K	Very High
Precision (mean of cell SDs)	0.601 K	0.721 K	0.826 K	0.714 K	Medium
Precision (median of cell RSDs)	0.314 K	0.304 K	0.310 K	0.279 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.517* ± *1.600* K Equator: *0.430* ± *0.741* K South: *0.322* ± *0.274* K	North: *0.381* ± *1.361* K Equator: *0.519* ± *0.818* K South: *0.232* ± *0.426* K	North: *0.504* ± *2.598* K Equator: *0.594* ± *0.799* K South: *0.303* ± *0.701* K	North: *0.399* ± *2.257* K Equator: *0.469* ± *0.775* K South: *0.203* ± *0.412* K	Medium
Stability between day and night	Day-Night: -0.104 ± 0.043 K Trend: Not enough match-ups available to	Day-Night: -0.228 ± 0.043 K Trend: Not enough match-ups available to	Day-Night: -0.367 ± 0.039 K Trend: Not enough match-ups available to	Day-Night: -0.354 ± 0.041 K Trend: Not enough match-ups available to	Medium

Table 5.6.7 NOAA-18 Night Coastal	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of Matches	999	1437	1913	1420	

Table 5.6.7: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18, for coastal regions.

Table 5.6.8: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18, for coastal regions.

Table 5.6.8 NOAA-18 Day Coastal	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	0.111 K	-0.140 K	-0.324 K	-0.249 K	Very High
Bias (median discrepancy)	0.087 K	-0.106 K	-0.263 K	-0.224 K	Very High
Precision (mean of cell SDs)	1.106 K	1.117 K	1.006 K	1.057 K	Medium
Precision (median of cell RSDs)	0.566 K	0.491 K	0.452 K	0.403 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Very High

Table 5.6.8 **NOAA-18 Day Coastal**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
	trend.	trend.	trend.	trend.	
Stability with respect to season (amplitude of cycle)	North: *0.818 ± 2.754* K Equator: *0.455 ± 1.372* K South: *0.529 ± 0.805* K	North: *0.728 ± 3.021* K Equator*: 0.425 ± 1.482* K South: *0.354 ± 0.658* K	North: *0.521 ± 2.053* K Equator: *0.291 ± 1.035* K South: *0.341 ± 0.805* K	North: *0.660 ± 2.238* K Equator*: 0.382 ± 0.843* K South: *0.259 ± 0.540* K	Medium
Stability between day and night	Day-Night: -0.104 ± 0.043 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.228 ± 0.043 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.367 ± 0.039 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.354 ± 0.041 K Trend: Not enough match-ups available to estimate trends.	Medium
Number of Matches	1761	2070	2901	2060	

Table 5.6.8: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18, for coastal regions.

5.6.2 Presentation of metrics related to SST uncertainty for NOAA-18

Table 5.6.9: Comparison of metrics related to SST uncertainty estimation for night-time AVHRR onboard NOAA-18, globally.

Table 5.6.9 **NOAA-18 Night**	**Optimal Estimation v1**		**Optimal Estimation v2**		**Incremental Regression**		**Incremental Regression (with OE2 QC)**		**Weight**
Bias (chi-squared) (Estimate ± 2σ range)	2.2	+0.4 -0.3	2.0	+0.3 -0.3	1.1	+0.1 -0.1	0.9	+0.0 -0.1	High
Independence from in situ	Independent		Independent		Empirical		Empirical		Medium
Generality	General to OE framework.		General to OE framework.		Specific to each matchup dataset used.		Specific to each matchup dataset used.		Low
Improvability	Fuller assessment of forward model error covariance.		Fuller assessment of forward model error covariance.		Investigated different algorithm forms. Improve-ment in prior forward modelling.		Investigated different algorithm forms. Improve-ment in prior forward modelling.		Low
Difficulty	No difficulty given OE framework.		No difficulty given OE framework.		Uncertainty statistics empirically defined outside of processing chain.		Uncertainty statistics empirically defined outside of processing chain.		Low

Table 5.6.9: Comparison of metrics related to SST uncertainty estimation for night-time AVHRR onboard NOAA-18, globally.

Table 5.6.10: Comparison of metrics related to SST uncertainty estimation for day-time AVHRR onboard NOAA-18, globally.

Table 5.6.10 **NOAA-18 Day**	**Optimal Estimation v1**		**Optimal Estimation v2**		**Incremental Regression**		**Incremental Regression (with OE2 QC)**		**Weight**
Bias	1.9	+0.3	2.0	+0.3	1.2	+0.1	1.0	+0.1	High

Table 5.6.10 **NOAA-18 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
(chi-squared) (Estimate ± 2σ range)	-0.2	-0.2	-0.0	-0.0	
Independence from in situ	Independent	Independent	Empirical	Empirical	Medium
Generality	General to OE framework.	General to OE framework.	Specific to each matchup dataset used.	Specific to each matchup dataset used.	Low
Improvability	Fuller assessment of forward model error covariance.	Fuller assessment of forward model error covariance.	Investigated different algorithm forms. Improve-ment in prior forward modelling.	Investigated different algorithm forms. Improve-ment in prior forward modelling.	Low
Difficulty	No difficulty given OE framework.	No difficulty given OE framework.	Uncertainty statistics empirically defined outside of processing chain.	Uncertainty statistics empirically defined outside of processing chain.	Low

Table 5.6.10: Comparison of metrics related to SST uncertainty estimation for day-time AVHRR onboard NOAA-18, globally.

5.6.3 Comments on results for NOAA-18

Considering the night-time results (single-view and three-channels used for retrieval), the bias for OE v2 and IR are both within target in the mean and nearly identical (+0.07 K). The regional biases patterns are quite similar. The precision of retrieval is better for OE v2 than for IR, although only marginally so when assessed using robust statistics. When IR is filtered using OE v2 quality levels (based on the OE v2 retrieval cost) the IR precision is a little better than for OE v2. The OE v2 algorithm gives no statistically significant temporal trend, while there is a 0013 +/- 0.004 K/yr trend in the IR results that is unaffected by quality level filtering. Day-night stability is poor for all, reflecting the poor bias for the day-time results discussed in the next paragraph. The OE retrievals are fully sensitive to SST while for IR, sensitivity is typically 0.6 to 0.85.

The day-time results show a markedly less satisfactory outcome. All the algorithms are outside of the 0.1 K target for bias, the best being OE v2 with a bias of -0.161 K in the mean (or -0.090 K in the median). The negative biases are somewhat different in their

regional distribution between OE v2 and IR, with negative biases of ~0.4 K over a wide tropical and subtropical area in the case of IR. After removal of this regional bias variability, however, the precision of the IR retrieval is markedly better than that of OE v2. (As commented for Metop and NOAA-19, this may reflect the choice made about the prior SST error variance in the OE v2 being unnecessarily broad.) There is no temporal trend in the OE v2 results, whereas there is a marginally significant trend of 0.016 K/yr in the IR results. This trend is not removed when the IR results are filtered with the OE v2 retrieval cost (rightmost column). Stability with respect to season is inconclusive, and day-night stability is dominated by the different bias between the day (two channel) and night (three channel) versions of the algorithms. The cause of the relatively poor day time performance in terms of bias is not established.

In the high latitude tables, the DMI regional results are competitive in terms of bias. OE v2 tends to return the best precision results in high latitudes, which has been seen with other AVHRRs. For OE v2, the night time biases are within target, whereas the day time biases are too negative. IR SSTs are within the 0.1 K target except for day time northern hemisphere.

As with other AVHRRs, the uncertainty estimates associated with OE v2 SSTs are rather optimistic, whereas the generic uncertainty estimate for IR SST is about correct on average.

Comments on the qualitative metrics are as for Metop GAC (section 5.4.3).

Interim conclusions considering NOAA-18 AVHRR: night-time IR SSTs show a significant trend artefact, while OEv2 does not. Otherwise the performance is comparable. As with NOAA-18, the day-time results are disappointing for both OEv2 and IR, although IR biases are larger. Overall, the picture is similar to NOAA-19 AVHRR, and OEv2 is emerging as a preferable candidate for AVHRRs overall.

5.7 Side-by-side comparison of results relating to NOAA-17

5.7.1 Presentation of metrics related to SST estimation for NOAA-17

Table 5.7.1: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17, globally.

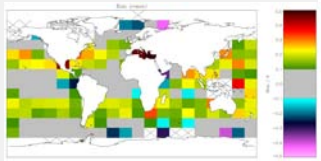
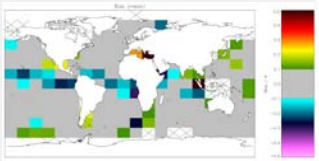
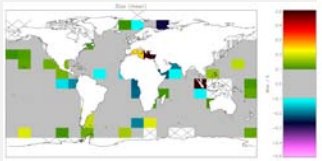
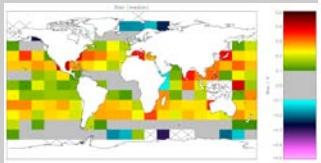
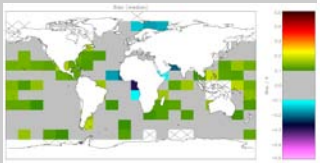
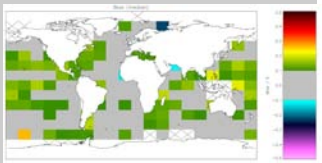
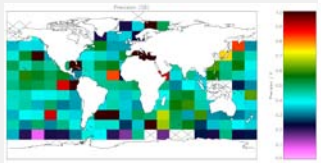
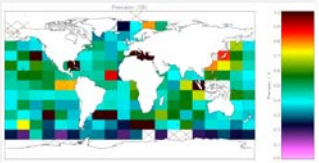
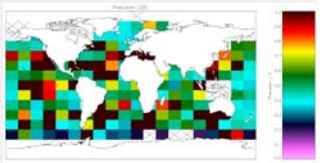
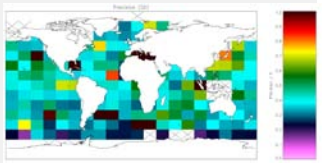
Table 5.7.1 NOAA-17 Night	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	0.147 K	0.005 K	0.007 K	0.035 K	Very High
Bias (median discrepancy)	0.217 K	0.077 K	0.080 K	0.092 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

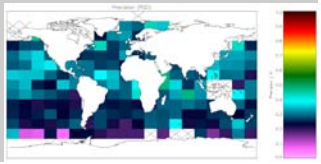
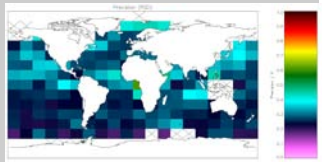
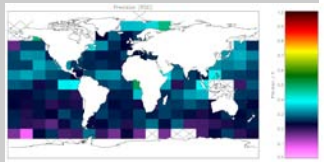
Table 5.7.1 **NOAA-17 Night**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.583 K	0.599 K	0.910 K	0.572 K	Medium
Precision (median of cell RSDs)	0.287 K	0.276 K	0.282 K	0.262 K	Medium
Stability with respect to trend	-0.019 ± 0.005 K/yr	*-0.006 ± 0.005* K/yr	*-0.006 ± 0.004* K/yr	-0.014 ± 0.005 K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.132 ± 1.153* K Equator: *0.111 ± 1.293* K South: *0.128 ± 0.898* K	North: *0.112 ± 0.687* K Equator: *0.097 ± 0.656* K South: *0.120 ± 1.071* K	North: *0.098 ± 1.091* K Equator: *0.139 ± 0.996* K South: *0.076 ± 1.356* K	North: *0.122 ± 1.078* K Equator: *0.078 ± 0.591* K South: *0.106 ± 1.038* K	Medium
Stability between day and night	Day-Night: 0.278 ± 0.018 K Trend: 0.035 ± 0.010 K	Day-Night: -0.055 ± 0.015 K Trend: *0.001 ± 0.010* K	Day-Night: -0.098 ± 0.019 K Trend: *0.009 ± 0.011* K	Day-Night: -0.094 ± 0.014 K Trend: *0.011 ± 0.011* K	Medium
Independence from in situ	Fully independent retrieval based on radiative transfer.	Fully independent retrieval based on radiative transfer.	Fully dependent, based on empirical regression to drifting buoys.	Fully dependent, based on empirical regression to drifting buoys.	High

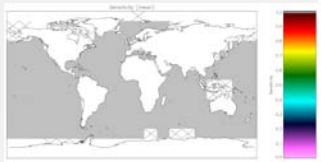
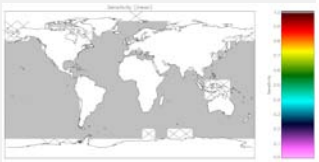
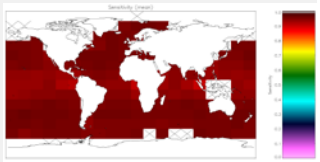
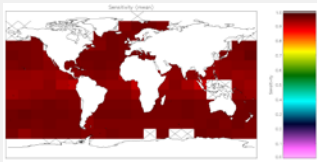
Table 5.7.1 NOAA-17 Night	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Map SST sensitivity					High
Generality	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Medium
Improvability	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Investigated different algorithm forms. Improvement in prior forward modelling.	Investigated different algorithm forms. Improvement in prior forward modelling.	Medium
Difficulty	No implementation concerns.	No implementation concerns.	No implementation concerns.	No implementation concerns.	Low
Number of Matches	38505	63108	80203	63751	

Table 5.7.1: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17, globally.

Table 5.7.2: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17, globally.

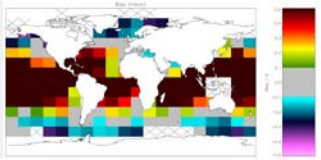
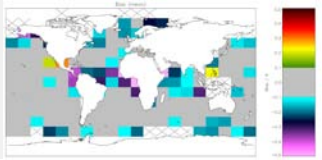
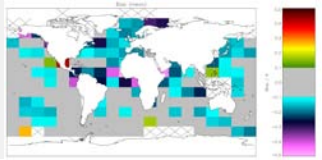
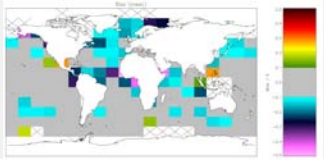
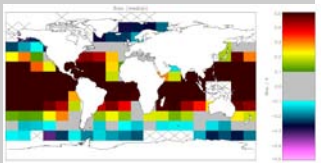
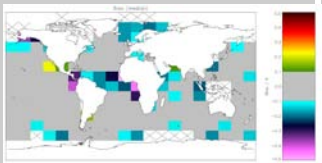
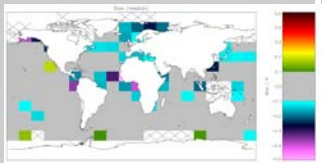
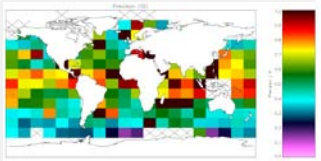
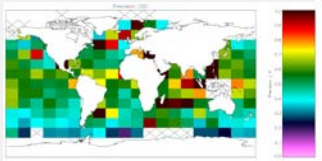
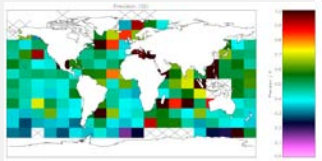
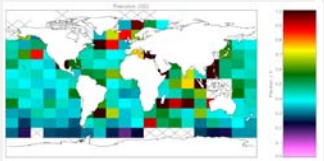
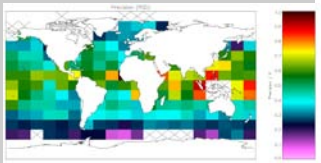
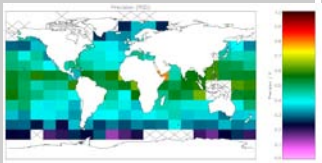
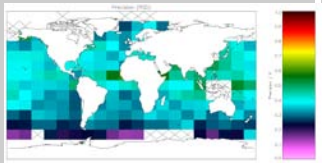
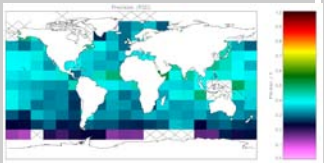
Table 5.7.2 AVHRR-17 Day	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	0.425 K	-0.050 K	-0.091 K	-0.059 K	Very High
Bias (median discrepancy)	0.357 K	-0.021 K	-0.048 K	-0.030 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium
Precision map (RSD of discrepancy)					Medium

Table 5.7.2 **AVHRR-17 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Precision (mean of cell SDs)	0.855 K	0.682 K	0.641 K	0.607 K	Medium
Precision (median of cell RSDs)	0.618 K	0.421 K	0.366 K	0.345 K	Medium
Stability with respect to trend	*0.016 ± 0.009* K/yr	*-0.005 ± 0.008* K/yr	*0.004 ± 0.010* K/yr	*-0.003 ± 0.010* K/yr	Very High
Stability with respect to season (amplitude of cycle)	North: *0.578 ± 1.409* K Equator: *0.297 ± 1.571* K South: *0.359 ± 1.128* K	North: *0.291 ± 1.132* K Equator: *0.166 ± 1.180* K South: *0.093 ± 1.075* K	North: *0.276 ± 0.890* K Equator: *0.164 ± 1.079* K South: *0.069 ± 0.805* K	North: *0.260 ± 0.827* K Equator: *0.185 ± 1.023* K South: *0.089 ± 0.636* K	Medium
Stability between day and night	Day-Night: 0.278 ± 0.018 K Trend: 0.035 ± 0.010 K	Day-Night: -0.055 ± 0.015 K Trend: *0.001 ± 0.010* K	Day-Night: -0.098 ± 0.019 K Trend: *0.009 ± 0.011* K	Day-Night: -0.094 ± 0.014 K Trend: *0.011 ± 0.011* K	Medium
Independence from in situ	Fully independent retrieval based on radiative transfer.	Fully independent retrieval based on radiative transfer.	Fully dependent, based on empirical regression to drifting buoys.	Fully dependent, based on empirical regression to drifting buoys.	High

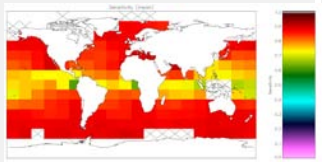
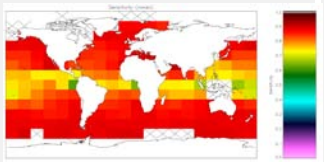
Table 5.7.2 **AVHRR-17 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Map SST sensitivity					High
Generality	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Applicable to other sensors.	Medium
Improvability	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Improvements will derive from improving fast radiative transfer and knowledge of error covariance characteristics of NWP and sensors.	Investigated different algorithm forms. Improvement in prior forward modelling.	Investigated different algorithm forms. Improvement in prior forward modelling.	Medium
Difficulty	No implementation concerns.	No implementation concerns.	No implementation concerns.	No implementation concerns.	Low
Number of Matches	48579	71778	92276	72541	

Table 5.7.2: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17, globally.

Table 5.7.3: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17, at latitudes above 60° N.

Table 5.7.3 **NOAA-17 Night MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.133 K	0.006 K	-0.086 K	-0.310 K	-0.071 K	Very High
Bias (median discrepancy)	-0.102 K	0.007 K	-0.084 K	-0.235 K	-0.042 K	Very High
Precision (mean of cell SDs)	0.535 K	0.534 K	0.665 K	0.907 K	0.696 K	Medium
Precision (median of cell RSDs)	0.316 K	0.328 K	0.397 K	0.657 K	0.389 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.297* ± *0.494* K	North: *0.272* ± *0.669* K	North: *0.256* ± *0.760* K	North: *0.995* ± *1.409* K	North: *0.196* ± *1.179* K	Medium
Stability between day and night	Day-Night: -0.100 ± 0.068 K Trend: Not enough match-ups available to	Day-Night: -0.167 ± 0.062 K Trend: Not enough match-ups available to	Day-Night: -0.126 ± 0.066 K Trend: Not enough match-ups available to	Day-Night: *0.164* ± *0.094* K Trend: Not enough match-ups available to	Day-Night: *0.082* ± *0.071* K Trend: Not enough match-ups available to	Medium

Table 5.7.3 NOAA-17 Night MIZ North	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of matches	267	551	846	849	849	

Table 5.7.3: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17, at latitudes above 60° N.

Table 5.7.4: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17, at latitudes above 60° N.

Table 5.7.4 NOAA-17 Day MIZ North	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
Bias (mean discrepancy)	-0.233 K	-0.161 K	-0.212 K	-0.146 K	0.011 K	Very High
Bias (median discrepancy)	-0.248 K	-0.147 K	-0.185 K	-0.090 K	0.013 K	Very High
Precision (mean of cell SDs)	0.687 K	0.686 K	0.665 K	0.971 K	0.707 K	Medium
Precision (median of cell RSDs)	0.346 K	0.321 K	0.348 K	0.682 K	0.381 K	Medium
Stability with respect to trend	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Very High

Table 5.7.4 **NOAA-17 Day MIZ North**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
	estimate trend.	estimate trend.	estimate trend.	estimate trend.	estimate trend.	
Stability with respect to season (amplitude of cycle)	North: *0.583* ± *0.602* K	North: *0.534* ± *0.431* K	North: *0.528* ± *0.557* K	North: *0.523* ± *1.285* K	North: *0.570* ± *1.147* K	Medium
Stability between day and night	Day-Night: *-0.100* ± *0.068* K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.167 ± 0.062 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.126 ± 0.066 K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.164* ± *0.094* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.082* ± *0.071* K Trend: Not enough match-ups available to estimate trends.	Medium
Number of matches	1148	1674	1929	1933	1933	

Table 5.7.4: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17, at latitudes above 60° N.

Table 5.7.5: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17, at latitudes below 50° S.

Table 5.7.5 **NOAA-17 Night MIZ South**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**DMI (CASSTA)**	**DMI (Regional)**	**Weight**
Bias (mean discrepancy)	-0.040 K	0.007 K	-0.043 K	-0.198 K	-0.002 K	Very High
Bias (median discrepancy)	-0.015 K	0.037 K	0.071 K	-0.037 K	0.049 K	Very High
Precision (mean of cell SDs)	0.366 K	1.016 K	1.490 K	1.106 K	0.989 K	Medium
Precision (median of cell RSDs)	0.252 K	0.215 K	0.234 K	0.494 K	0.305 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	South: *0.319* ± *0.508* K	South: *0.418* ± *3.189* K	South: *0.404* ± *3.057* K	South: *0.477* ± *3.022* K	South: *0.361* ± *2.929* K	Medium
Stability between day and night	Day-Night: -0.127 ± 0.029 K Trend: Not enough match-ups available to	Day-Night: *-0.091* ± *0.055* K Trend: Not enough match-ups available to	Day-Night: *0.048* ± *0.077* K Trend: Not enough match-ups available to	Day-Night: *0.102* ± *0.064* K Trend: Not enough match-ups available to	Day-Night: *0.019* ± *0.055* K Trend: Not enough match-ups available to	Medium

Table 5.7.5 NOAA-17 Night MIZ South	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of matches	574	1498	2020	2039	2039	

Table 5.7.5: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17, at latitudes below 50° S.

Table 5.7.6: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17, at latitudes below 50° S.

Table 5.7.6 NOAA-17 Day MIZ South	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
Bias (mean discrepancy)	-0.167 K	-0.084 K	0.005 K	-0.096 K	0.017 K	Very High
Bias (median discrepancy)	-0.171 K	-0.072 K	0.047 K	0.011 K	0.056 K	Very High
Precision (mean of cell SDs)	0.338 K	0.358 K	0.384 K	0.633 K	0.470 K	Medium
Precision (median of cell RSDs)	0.226 K	0.253 K	0.230 K	0.444 K	0.293 K	Medium
Stability with respect to trend	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Not enough match-ups available to	Very High

Table 5.7.6 NOAA-17 Day MIZ South	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	DMI (CASSTA)	DMI (Regional)	Weight
	estimate trend.	estimate trend.	estimate trend.	estimate trend.	estimate trend.	
Stability with respect to season (amplitude of cycle)	South: *0.202* ± *0.560* K	South: *0.207* ± *0.486* K	South: *0.141* ± *0.687* K	South: *0.311* ± *0.898* K	South: *0.166* ± *0.693* K	Medium
Stability between day and night	Day-Night: -0.127 ± 0.029 K Trend: Not enough match-ups available to estimate trends.	Day-Night: *-0.091* ± *0.055* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.048* ± *0.077* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.102* ± *0.064* K Trend: Not enough match-ups available to estimate trends.	Day-Night: *0.019* ± *0.055* K Trend: Not enough match-ups available to estimate trends.	Medium
Number of matches	1076	1381	1589	1589	1589	

Table 5.7.6: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17, at latitudes below 50° S.

Table 5.7.7: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17, for coastal regions.

Table 5.7.7 **NOAA-17 Night Coastal**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
Bias (mean discrepancy)	0.178 K	0.042 K	0.015 K	0.046 K	Very High
Bias (median discrepancy)	0.221 K	0.107 K	0.088 K	0.095 K	Very High
Precision (mean of cell SDs)	0.545 K	0.571 K	0.942 K	0.514 K	Medium
Precision (median of cell RSDs)	0.325 K	0.294 K	0.305 K	0.282 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Not enough match-ups available to estimate trend.	Very High
Stability with respect to season (amplitude of cycle)	North: *0.242* ± *0.579* K Equator: *0.651* ± *0.781* K South: *0.488* ± *0.909* K	North: *0.163* ± *1.256* K Equator: *0.403* ± *1.352* K South: *0.394* ± *0.990* K	North: *0.273* ± *1.632* K Equator: *0.351* ± *1.189* K South: *0.379* ± *1.088* K	North: *0.124* ± *0.719* K Equator: *0.374* ± *1.303* K South: *0.302* ± *0.961* K	Medium
Stability between day and night	Day-Night: 0.303 ± 0.038 K Trend: Not enough match-ups available to	Day-Night: -0.117 ± 0.031 K Trend: Not enough match-ups available to	Day-Night: -0.154 ± 0.035 K Trend: Not enough match-ups available to	Day-Night: -0.147 ± 0.028 K Trend: Not enough match-ups available to	Medium

Table 5.7.7 NOAA-17 Night Coastal	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
	estimate trends.	estimate trends.	estimate trends.	estimate trends.	
Number of Matches	1126	1901	2593	1917	

Table 5.7.7: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17, for coastal regions.

Table 5.7.8: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17, for coastal regions.

Table 5.7.8 NOAA-17 Day Coastal	Optimal Estimation v1	Optimal Estimation v2	Incremental Regression	Incremental Regression (with OE2 QC)	Weight
Bias (mean discrepancy)	0.481 K	-0.075 K	-0.139 K	-0.101 K	Very High
Bias (median discrepancy)	0.369 K	-0.042 K	-0.086 K	-0.068 K	Very High
Precision (mean of cell SDs)	0.967 K	0.836 K	0.792 K	0.760 K	Medium
Precision (median of cell RSDs)	0.769 K	0.466 K	0.408 K	0.381 K	Medium
Stability with respect to trend	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Not enough match-ups available to estimate	Very High

Table 5.7.8 **NOAA-17 Day Coastal**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
	trend.	trend.	trend.	trend.	
Stability with respect to season (amplitude of cycle)	North: *0.588* ± *1.828* K Equator: *0.643* ± *1.304* K South: *0.500* ± *0.969* K	North: *0.547* ± *1.152* K Equator: *0.534* ± *1.626* K South: *0.348* ± *0.843* K	North: *0.496* ± *1.050* K Equator: *0.395* ± *2.625* K South: *0.215* ± *0.785* K	North: *0.502* ± *0.999* K Equator: *0.560* ± *3.157* K South: *0.350* ± *0.780* K	Medium
Stability between day and night	Day-Night: 0.303 ± 0.038 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.117 ± 0.031 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.154 ± 0.035 K Trend: Not enough match-ups available to estimate trends.	Day-Night: -0.147 ± 0.028 K Trend: Not enough match-ups available to estimate trends.	Medium
Number of Matches	1586	2322	3090	2346	

Table 5.7.8: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17, for coastal regions.

5.7.2 Presentation of metrics related to SST uncertainty for NOAA-17

Table 5.7.9: Comparison of metrics related to SST uncertainty estimation for night-time AVHRR onboard NOAA-17, globally.

Table 5.7.9 **NOAA-17 Night**	**Optimal Estimation v1**		**Optimal Estimation v2**		**Incremental Regression**		**Incremental Regression (with OE2 QC)**		**Weight**
Bias (chi-squared) (Estimate ± 2σ range)	3.3	+0.5 -0.5	3.5	+0.6 -0.6	2.1	+0.1 -0.2	1.7	+0.2 -0.1	High
Independence from in situ	Independent		Independent		Empirical		Empirical		Medium
Generality	General to OE framework.		General to OE framework.		Specific to each matchup dataset used.		Specific to each matchup dataset used.		Low
Improvability	Fuller assessment of forward model error covariance.		Fuller assessment of forward model error covariance.		Investigated different algorithm forms. Improve-ment in prior forward modelling.		Investigated different algorithm forms. Improve-ment in prior forward modelling.		Low
Difficulty	No difficulty given OE framework.		No difficulty given OE framework.		Uncertainty statistics empirically defined outside of processing chain.		Uncertainty statistics empirically defined outside of processing chain.		Low

Table 5.7.9: Comparison of metrics related to SST uncertainty estimation for night-time AVHRR onboard NOAA-17, globally.

Table 5.7.10: Comparison of metrics related to SST uncertainty estimation for day-time AVHRR onboard NOAA-17, globally.

Table 5.7.10 **NOAA-17 Day**	**Optimal Estimation v1**		**Optimal Estimation v2**		**Incremental Regression**		**Incremental Regression (with OE2 QC)**		**Weight**
Bias	2.1	+0.3	1.7	+0.3	1.8	+0.2	1.6	+0.1	High

Table 5.7.10 **NOAA-17 Day**	**Optimal Estimation v1**	**Optimal Estimation v2**	**Incremental Regression**	**Incremental Regression (with OE2 QC)**	**Weight**
(chi-squared) (Estimate ± 2σ range)	-0.2	-0.2	-0.1	-0.1	
Independence from in situ	Independent	Independent	Empirical	Empirical	Medium
Generality	General to OE framework.	General to OE framework.	Specific to each matchup dataset used.	Specific to each matchup dataset used.	Low
Improvability	Fuller assessment of forward model error covariance.	Fuller assessment of forward model error covariance.	Investigated different algorithm forms. Improve-ment in prior forward modelling.	Investigated different algorithm forms. Improve-ment in prior forward modelling.	Low
Difficulty	No difficulty given OE framework.	No difficulty given OE framework.	Uncertainty statistics empirically defined outside of processing chain.	Uncertainty statistics empirically defined outside of processing chain.	Low

Table 5.7.10: Comparison of metrics related to SST uncertainty estimation for day-time AVHRR onboard NOAA-17, globally.

5.7.3 Comments on results for NOAA-17

Considering the night-time results (single-view and three-channels used for retrieval), the bias for OE v2 and IR are both within target in the mean and nearly identical (+0.006 K). The regional biases patterns are quite similar. The precision of retrieval is better for OE v2 than for IR, although only marginally so when assessed using robust statistics. When IR is filtered using OE v2 quality levels (based on the OE v2 retrieval cost) the IR precision is a little better than for OE v2. There is no statistically significant trend with either the OE v2 or IR algorithms, while the OE v1 algorithm shows a -0.0019 ± 0.005 K/yr trend. A trend of -0.014 ± 0.005 K/yr is also observed in the IR results with the OE v2 quality filtering applied. The best day-night stability is with OE v2. Seasonal stability results are mixed, with none statistically significant. The OE retrievals are fully sensitive to SST and the IR method has sensitivity in the same range as for other AVHRRs -- 0.6 to 0.85.

For day time results the untuned OE v1 is seen to be too biased for SST CCI use. Both OE v2 and IR are within target, with OE v2 closer to zero. Again, the patterns of bias are similar, particularly when considering the median discrepancy. Patterns of precision are similar, with IR being significantly less noisy in subtropical regions than OE v2 (and over all). This is a result of the IR retrievals being significantly less sensitive to SST than the OE v2 retrievals in this region. There is no statistically significant trend with any of the algorithms, nor are there any statistically significant seasonal stability results.

Comments on the qualitative metrics are as for Metop GAC (section 5.4.3).

The uncertainty metrics for night-time NOAA-17 SSTs are significantly above 1.0 for all retrievals, particularly both OE retrievals – this means that the estimated uncertainties are smaller than the comparisons with drifting buoys suggest to be realistic. Results for day-time retrievals suggest uncertainty estimates are better for all algorithms (particularly OE v2) but are still underestimated.

High latitude results show similar patterns to the AVHRRs previously discussed. OE v2 and IR do comparably well in terms of bias with OE v2 typically having better precision. Both have out-of-target negative bias in the case of day-time northern hemisphere matches, but are within 0.1 K otherwise. DMI regional tends to be good in terms of bias but inferior in terms of precision.

Interim conclusions considering NOAA-17 AVHRR: OEv2 and IR results are similar for this sensor in terms of bias and stability. As before, precision is better for IR, while sensitivity is better for OEv2. On its own, the quantitative metrics would not strongly discriminate OEv2 and IR. However, the better performance of OEv2 for some other AVHRRs together with the independence of OEv2 favours selection of OEv2.

6. ALGORITHM SELECTION

6.1 Summary of highlights from metrics

Table 6.1.1: Summary of metrics for ATSR-2 and AATSR.

Table 6.1.1	AATSR		ATSR-2	
	Day (D2)	**Night (D3)**	**Day (D2)**	**Night (D3)**
Bias	OEv1 is least biased. OEv2 and ARC are comparable and within 0.1 K for most locations. ORAC shows positive biases in mid latitudes.	ARC and OEv2 are similar, with slightly positive bias, although within 0.1 K target. OEv1 biases are larger.	OEv2 followed by ARC give the lowest biases, and are within 0.05 K of zero globally. OEv1 is negatively biased in tropics, while ORAC shows positive midlatitude biases.	ARC gives lowest bias, followed closely by OEv2. Both are within 0.1 K for most of the ocean. OEv1 biases are larger.
Precision	Best precision from OEv2 and OEv1.	Best precision from OEv1 and OEv2.	Best precision from OEv2 and OEv1.	Best precision from OEv1 and OEv2.
Stability measures	OEv2 is best on trend and day/night stability. ARC shows best seasonal stability (though not significantly so).	No significant trend for any algorithm. Seasonal and diurnal stability best for OEv2.	Not assessable with respect to trend stability. ARC and OEv2 have good day/night stability.	Not assessable with respect to trend stability. ARC and OEv2 have good day/night stability.
Independence	All algorithms considered are independent of in situ observations.			
SST sensitivity	All algorithms considered have sensitivity acceptably close to ideal (1.0).			
Generality	OEv1 is a general algorithm. ARC relies an on ATSR-specific coefficient formulation, that could be adapted. OEv2 is a general formulation, and relies on existence of ARC retrievals for brightness temperature tuning. ORAC is a general algorithm, although applicable only to day time (at present).			
Improvability	ARC coefficients for ATSR-2 and AATSR are close to ideal: may be improved with newer radiative transfer; ATSR-1 needs more development. For OEv1, OEv2 and ORAC, approach to improvement is to improve accuracy of forward modelling and to improve understanding of error covariances.			
Implementation issues	ARC, OEv1 and OEv2 are already available within ARC processing chain and readily prototyped for SST CCI. ORAC would require effort to			

Table 6.1.1	AATSR		ATSR-2	
	Day (D2)	**Night (D3)**	**Day (D2)**	**Night (D3)**
	implement within SST CCI, but no fundamental difficulties.			
Uncertainty estimates	Unbiased, best results for OEv2.	Least biased for OEv2, but nevertheless, under-estimated.	Good (least biased) results for OEv2.	Least biased for OEv2, but nevertheless, under-estimated.
Coastal Zone	Best results for OEv2 in most cases.	Only ARC meets target bias but better precision for OE retrievals.	ARC and ORAC meet target bias for both mean and median metrics.	Lowest biases for OE v2. Best precision for OEv1.

Table 6.1.1: Summary of metrics for ATSR-2 and AATSR.

Table 6.1.2: Summary of metrics for all night-time AVHRR.

Table 6.1.2	Night-time AVHRR results			
	Metop	**NOAA-19**	**NOAA-18**	**NOAA-17**
Bias	Biases well within target and comparable for OEv2 and IR.	Better results for OEv2 than IR. Both show slightly positive biases when using median-based metric.	Comparable results for OEv2 and IR. Both show slightly positive biases when using median-based metric.	Comparable, within-target results for OEv2 and IR.
Precision	Comparable for OEv2 and IR.	Comparable for OEv2 and IR.	Comparable for OEv2 and IR.	Comparable for OEv2 and IR.
Stability	OEv2 gives best results on all stability measures.	No significant trend in any case. Day-night stability best in OEv2.	OEv2 gives best overall results for stability. IR SSTs show significant trend.	No significant trends for OEv2 or IR. Day-night stability best in OEv2.
Independence	OEv2 is independent. IR is wholly dependent on in situ.			
SST sensitivity	Best in OEv2 (essentially ideal, ~1.0)	Best in OEv2 (essentially ideal, ~1.0)	Best in OEv2 (essentially ideal, ~1.0)	Best in OEv2 (essentially ideal, ~1.0)
Generality	All approaches applicable to all AVHRRs			
Improvability	OEv2 can be improved by refining method for cross-referencing to ATSRs, by improving forward modelling, and improving parameterisation			

	of error covariance measurements. IR can be improved through different algorithm forms and by improving prior forward modelling.			
Implementation issues	Matches become more scarce prior to 1990s – presents a challenge for all methods considered.			
Marginal ice zone	Best results from OEv2	Best results from OEv2	Variable results for bias, except DMI regional is consistently good. OEv2 best for precision.	In most cases, best results from OEv2.
Coastal zone	Best results from IR.	Only IR mean-bias metric meets target.	Large bias in OEv1. IR and OEv2 within target using mean metric.	Best results from IR in most cases.
Uncertainty estimates	OEv2 gives uncertainty case-by-case, and tends to underestimate uncertainty. IR provides a single generic estimate per sensor, and is close to unbiased.			

Table 6.1.2: Summary of metrics for all night-time AVHRR.

Table 6.1.3: Summary of metrics for all day-time AVHRR.

Table 6.1.3	**Day-time AVHRR results**			
	Metop	**NOAA-19**	**NOAA-18**	**NOAA-17**
Bias	Comparable, within-target results from OEv2 and IR.	OEv2 is least biased, but shows some out-of-target biases in northern midlatitudes.	OEv2 is least biased, but shows some out-of-target biases in northern midlatitudes.	Comparable, within-target results from OEv2 and IR.
Precision	Best precision from IR.	Best precision from IR.	Best precision from IR.	Best precision from IR.
Stability	Trend stability best for OEv2. Other measures comparable between OEv2 and IR.	No significant trend in any case. OEv2 better for day-night stability. No significance to seasonal stability comparison.	No trend in OEv2; marginally significant trend in IR. Other aspects comparable.	No significant trend in any case. OEv2 better for day-night stability. No significance to seasonal stability comparison.
Independence	OEv2 is independent. IR is wholly dependent on in situ.			

SST sensitivity	Close to ideal from OEv2. IR well below ideal.	Close to ideal from OEv2. IR well below ideal.	Close to ideal from OEv2. IR well below ideal.	Close to ideal from OEv2. IR well below ideal.
Generality	All approaches applicable to all AVHRRs			
Improvability	OEv2 can be improved by refining method for cross-referencing to ATSRs, by improving forward modelling, and improving parameterisation of error covariance measurements. IR can be improved through different algorithm forms and by improving prior forward modelling.			
Implementation issues	Matches become more scarce prior to 1990s – presents a challenge for all methods considered.			
Marginal ice zone	DMI regional best for biases. OEv2 and IR comparable on other aspects.	Bias results are mixed – inconsistent between hemispheres.	DMI regional best for biases. OEv2 and IR comparable on other measures.	DMI regional best for biases. OEv2 and IR comparable on other measures.
Coastal zone	Only OEv2 meets target bias but with slightly lower precision than IR.	Only OEv2 median-bias metric meets target. Large biases from other algorithms. Precision better using IR.	Large biases in IR. OE biases close to target. Precision better using IR.	Lowest biases in OEv2. IR also within target using median metric. Precision better using IR.
Uncertainty estimates	OEv2 is case-by-case uncertainty, while IR is a general value per algorithm-sensor. Comparable biases in uncertainty between OEv2 and IR, although OEv2 more biased in case of NOAA-18.			

Table 6.1.3: Summary of metrics for all day-time AVHRR.

6.2 Analysis of strengths and weakness of algorithms

First, we consider the algorithms under consideration for use with ATSRs, in the light of the results presented for ATSR-2 and AATSR.

The ARC algorithm for retrieving SST using coefficients has the following strengths. SSTs retrieved using the ARC algorithm in general meet the target bias of 0.1 K for most of the global ocean, as assessed by mean and median discrepancy against drifting buoys for both sensors. The SSTs are relatively stable with respect to long-term trends, seasons and (where assessable) day-night differences; in general stability is comparable to OEv2 (discussed below) and either ARC or OEv2 performs best on any given measure. Since OEv2 is cross-referenced to ARC, this is the expected behaviour. ARC SSTs are defined independently from in situ observations. The results presented here assess ARC SSTs individually for ATSR-2 and AATSR, but it is worth noting that ARC coefficients have been homogenised to maximise consistency between sensors. This is a particular strength for ARC SSTs, and is only indirectly achieved for OEv2 because OEv2 is crossed-referenced to ARC SSTs. The sensitivity for ARC SSTs is close to, but not identical to, 1.0; it is

nonetheless suitable for a climate data set, and therefore counts as a strength. Lastly, note that ATSR-1 coefficients exist for ARC, while other algorithms have not been tested for ATSR-1 within this algorithm selection procedure (since no "competitor" algorithms were proposed).

The ARC algorithm for retrieving SST using coefficients has the following relative weaknesses. The ARC coefficient approach is adaptable to other sensors, but this has not been done and would involve some effort. The precision of coefficient based retrieval compared to OE is generally poorer (larger SDs and RSDs) and that is true in the results shown here. Nonetheless, the precision of ARC SSTs is very good in absolute terms (in comparison to earlier coefficient designs not considered here, for example). It is not clear how to improve the ARC retrievals further, since they already benefit from of order 3 person-years of development within the ARC project. In principle, an update of the underlying spectroscopy used in the line-by-line radiative transfer code should improve the coefficients, but with forward modelling errors already apparently being <<0.1 K, this is not guaranteed. Probably, an update to spectroscopic information would be justified only in the context of other reasons to update the radiative transfer basis.

The optimal estimation v1 algorithm has the following strengths. OEv1 is a relatively low-noise algorithm: either OEv1 or OEv2 returned the best precision for the ATSRs in each situation. It is driven by radiative transfer modelling, and is therefore independent of in situ observations. It returns a sensitivity of extremely close to 1.0. It is a general algorithm that can readily be applied to other sensors, provided the relevant auxiliary files for the fast simulation model (RTTOV) are available. As with all optimal estimators, the routes to improvement are to improve the simulation capability and/or better represent prior error covariance information in the retrieval.

OEv1 has the following relative weaknesses. The fast simulation model does not automatically guarantee within target biases. In this exercise, OEv1 biases were acceptable in the case of AATSR day-time, but not in the other situations. This is because the simulation capability is not at the <<0.1 K level of accuracy that supports coefficient-based retrieval using full line-by-line radiative transfer: fast models require approximations that introduce errors. This in itself precludes use of optimal estimation without bias-correction of simulated vs. observed brightness temperatures (BTs). This weakness rules out OEv1, as demonstrated by the bias metrics presented in this document. OEv1 will not be considered further for either ATSRs or AVHRRs.

The OEv2 algorithm includes bias-correction of BTs, preserving independence from in situ by using ARC SSTs as the true skin SST in the bias correction procedure (although see comments below under weaknesses). OEv2 is within the global 0.1 K target for day and night, ATSR-2 and AATSR observations, and gives regional biases within target over most of the global oceans. Broadly, its bias performance approaches that of ARC SSTs. The OEv2 algorithm is low noise, giving very good precision, better than other algorithms. OEv2 shows excellent stability of discrepancy relative to drifting buoys. The sensitivity is very close to the ideal 1.0, globally. There is scope to improve forward modelling, bias adjustment of BTs and specification of prior error covariances, but already OEv2 is appropriate for use in a climate context with respect to the various metrics reported here. OEv2 also generates as a ready by-product of the retrieval process, a retrieval "cost" which will be able to be exploited further within SST CCI.

The OEv2 algorithm is dependent on the existence of the ARC retrieval, for BT bias correction, which is arguably a weakness. Since the ARC SSTs perform well, this is acceptable, although, as commented above, the practical improvability of ARC SSTs is not fully known. The OE approach is very general, and, presents no implementation difficulty in a system with access to an appropriate feed of numerical weather prediction (NWP) fields.

The Oxford-RAL Retrieval of Aerosol and Cloud (ORAC) is another optimal estimation algorithm, this time including use of visible channels and the option to retrieval aerosol amount. In principle, ORAC should have various strengths compared to OEv2, but these seem not to be realised at the current state of development of ORAC for SST. It is an independent algorithm.

The relative weakness of the ORAC results are as follows. The results for ORAC tend to be positively biased by more than 0.1 K in midlatitudes. For the ATSRs, the precision and stability measures tend to be less good than for OEv2. Lastly, a significant restriction on the generality of the algorithm in its current development is that it is not applicable at night. Therefore, we see the good (but not best available) results from ORAC in this exercise as demonstrating future potential, rather than pointing to immediate implementation within SST CCI.

For the ATSRs, the choice therefore narrows to ARC or OEv2. Reviewing Table 6.1.1 and the detailed information in Section 1, it is clear that there is comparable performance on most metrics – depending on the situation (day/night, ATSR2/AATSR) one or other may be marginally ahead. The metrics that should be considered paramount are those with the highest weighting (Very High), namely, Bias, and Stability with respect to Trend. ARC and OEv2 biases are very similar (within 2 cK typically). Regarding Stability measures, OEv2 arguably has the edge (see, for example, the trend stability results for AATSR), but the differences are statistically marginal or non-significant. ARC and OEv2 are also equivalent on the Highly weighted considerations (Sensitivity, and Independence from In Situ). The first clear advantage, working down the hierarchy of weightings, is apparent for the Precision metric, where OEv2 is clearly favoured. It is an interesting scientific point that the optimal retrieval can be tuned to the coefficient based retrieval as regards bias, and then, because of better noise propagation characteristics, emerge as better than the coefficients in terms of precision. Other considerations relate to the uncertainty and coastal-zone performance of the two algorithms. The uncertainty estimates arising from the OEv2 appear to be more realistic (Table 6.1.1). Against this, ARC retrievals give better biases for matches in coastal zones, for reasons not clear.

We want to select one of ARC and OEv2 to apply to day and night and to ATSR2 and AATSR, for reasons of consistency. Regarding the most highly weighted metrics (Bias, Stability, Independence), there is no clear winner. OEv2 is markedly better with regards to precision and uncertainty estimation, which needs to be balanced against the better performance of ARC retrievals in the coastal zone. Since the former are important aspects of global behaviour, OEv2 is favoured.

One caveat is that OEv2 has not been formulated for ATSR-1. While possible in principle, it is complicated by issues around the Pinatubo aerosol event and the trend in ATSR-1 detector temperature, unique to that sensor. Therefore, there is no selection to be made for ATSR-1: ARC coefficients will be used in this phase of SST CCI. Thus, the trade-off favours OEv2 for the ATSR-2 and AATSR individually, but a more consistent selection across all three ATSRs would be ARC retrievals. Having said that, OEv2 is tuned to ARC for ATSR-2 and AATSR, and should in that sense be compatible with ARC applied to ATSR-1.

OEv2 is not a stratospheric aerosol robust algorithm, unlike the ARC coefficients. OEv2 can be applied to ATSR2 and AATSR, because we know that no further stratospheric aerosol events occurred during their lifetimes. But ATSR-1 was affected by post-Pinatubo aerosol. The priority in SST CCI was to develop the new OEv2 in detail for the normal situation, and future work will be able to extend the technique to include Pinatubo aerosol. This will likely require having such aerosol in RTTOV10 (used for fast forward modelling in OEv2), which is not presently the case.

Since the ATSR selection should also be considered in light of the results for the other sensors in the long-term time-series, the discussion now turns to the AVHRR results.

As discussed in the context of the ATSRs, OEv1 is limited in its bias performance. Since bias is a metric with very high weight, OEv1 need not be considered further for the selection of an AVHRR algorithm.

Night-time AVHRR results are summarised in Table 6.1.2. OEv2 and incremental regression both perform comparably well regarding bias for night-time AVHRR SSTs across the four sensors. In one case (NOAA-18) OEv2 is notably less biased than IR; in the other cases, results are comparable. Across the metrics of stability, it is often the case that no statistically significant instability is found for either algorithm. However, in one case, NOAA-18 again, IR results in a statistically significant trend in discrepancy, while OEv2 does not. The precision for night-time SSTs is also comparable between OEv2 and IR. The similarity in these measures is interesting, and probably arises from some aspects of the algorithms which are intrinsically similar. Both work by taking a simulation of BTs using NWP as a linearization point. In this controlled selection process, those simulations are in common, which means that non-algorithmic differences (to do with the simulation) are excluded and a clean comparison is possible. Both algorithms then calculate an increment to the prior SST. For OEv2, this increment depends on simulated jacobians. For IR, the jacobians are, in effect, estimated by empirical regression. The similarity of results in this case points to both methods be similarly efficient in this step. Thus, there is no clear cut case to make for either algorithm over the other on the basis of the Bias and Stability (both with Very High weight) or Precision (Medium weight).

The Highly weighted metrics are the SST Sensitivity and Independence. The SST Sensitivity is essentially ideal for OEv2 and is very close to ideal for IR. It is in regards to Independence that the algorithms sharply diverge, since OEv2 (cross-referenced to ATSRs) is wholly independent of in situ observations, whereas the empirical determination of the coefficients for IR makes it wholly dependent. This qualitative metric is therefore the one on which the selection of algorithm must hinge, favouring OEv2. (It is worth noting in retrospect that a hybrid approach could be interesting, in which the coefficients for IR are defined by regressing to matched ATSR SSTs, preserving independence. However, this is not available for algorithm selection in the present exercise.)

Although the Independence criterion effectively determines the selection for AVHRR night-time, it is worth reviewing the trade-off for the remaining metrics. There is arguably more room for improvement with OEv2, since the forward modelling and error covariance assumptions are unlikely to be fully optimised. OEv2 is a new algorithm developed within SST CCI. The parameters in the algorithm have been optimized so far using only the designated matches for development and testing within the Round Robin exercise, so that the SST CCI team has no advantage. This restriction on the SST CCI algorithm development was a result of our high priority to ensure parity for all algorithms being considered. It is likely that more data will allow better optimization in future, as experience with the algorithm is accumulated. Both approaches will face challenges when applied in the early 1990s, when drifting buoy matches are scarcer. The high-latitude / marginal ice zone results clearly favour OEv2, whereas the subset of results in global coastal zones clearly favours IR. Probably, IR fares worse in high latitudes because matches are few, while the cross-referencing method for OEv2 extrapolates better to high latitude conditions. On the associated uncertainty estimates for the two approaches, OEv2 uncertainties do discriminate “good” and “fair” SSTs with some skill, but over all are underestimates of the true uncertainty. In contrast, the IR uncertainty is unbiased on average, but is a single value for all SSTs, and therefore not in reality as informative.

So, in brief for night-time AVHRR cases, OEv2 and IR are similar in their quantitative performance. OEv2 has the advantage of being independent of in situ observations, which is important to a substantial minority of users.

The case of day-time AVHRR results is more complex. Regarding the metrics of Bias, OEv2 and IR are comparable and perform within target (globally) for Metop and NOAA-

17. In the cases of NOAA-19 and 18, both show negative biases outside of the 0.1 K target, particularly in northern midlatitudes. As was discussed in Section 5.5.3., low solar zenith angles correlate to more negatively biased SSTs for NOAA-19. In fact, this is true to varying degrees for all four sensors. Of the two algorithms, OEv2 shows this effect less. This strongly favours selection of OEv2 over IR.

Considering the Stability measures, OEv2 seems to come out marginally favoured, since there is a marginally significant trend in the IR SST discrepancies in the case of NOAA-18. Across the other stability metrics, the two algorithms are comparable.

Regarding the Highly weighted aspects of Independence and SST Sensitivity, OEv2 is favoured on both counts, across all four sensors. OEv2 is independent of in situ observations, while IR is wholly dependent. OEv2 has close to ideal sensitivity, whereas IR returns sensitivities well below ideal. This is different to the finding for night-time SSTs. The reason is the reduced information content with two split window channels (as here) compared to also having a 3.7 μm channel available (as in the night-time case). The OEv2 by design maintains close to ideal SST sensitivity – at the cost of increased noise levels (see next paragraph). IR coefficients are chosen to optimise for low noise, and in essence, sacrifice SST sensitivity to this end.

For this reason, the Precision metric favours IR over OEv2 very clearly. But this is a trade-off against decreased SST sensitivity. In this exercise, SST sensitivity is a more highly weighted metric than Precision, and therefore OEv2 selects the “right” trade-off between these aspects of retrieval. Why is this weighting appropriate? The reason is that non-zero SST sensitivity is an indicator of the presence of significant prior SST information being present in the retrieval – meaning that true SST variability is underestimated. This is undesirable in a climate data set. On the other hand, unbiased noise in SST is diminished by spatio-temporal averaging in many climate applications. For example, climate model inter-comparisons generally use 1°x1°, monthly data. In situ-based datasets are gridded to even lower resolutions: e.g., HadSST3 is widely used and is 5°x5° monthly data (Kennedy et al., 2011; RD. 211). At such resolutions in satellite data, it is the correlated/systematic errors (including those due to insensitivity to SST) that are dominant. Thus, in the SST CCI user requirements survey (RD.171) 75% of respondents were concerned about precision only for spatial scales of 100 km or greater or had no applicable precision requirement. User requirement UR-QUF-49 is that 0.1 °C precision is required over a 100 km space scale. None of the differences in precision at pixel level between algorithms the algorithms considered compromise this target.

Considering other metrics in summary, OEv2 and IR do comparably for marginal ice zone / high latitude cases, with results often mixed for a given sensor. However, for day-time retrievals, the DMI regional algorithm is better over all (not found in the night-time case). It would be locally advantageous in high latitudes to use the DMI regional algorithm, but the loss of consistency is an argument against selecting regional algorithms. However, the DMI regional results do suggest that there is scope to improve OEv2 in the MIZ in future. In the coastal cases, OEv2 gives markedly better performance on bias than IR. As before, the IR uncertainty estimate is appropriate on average, but doesn’t give discrimination between retrievals as the OEv2 uncertainty estimate does.

Taken together for day-time SST retrieval from the AVHRRs, it is clear that OEv2 is preferable. IR SSTs are less noisy, but either comparable or less good on other metrics.

Although selection of OEv2 for day-time AVHRRs is clear relative to the option of incremental regression, the results include some negative biases that are out of target. We therefore need some additional strategies to improve these biases. The strategies and improvements attainable are presented and discussed in Appendix A3.

The findings of this section are synthesised into algorithm selection conclusions in the next section (6.3).

6.3 Conclusions on selection of algorithms

The findings of Section 6.2 can be summarized as follows.

There is very similar performance on the most highly weighted metrics (bias and stability) between the ARC SST coefficient algorithm and the optimal estimator, OEv2, which uses brightness temperatures adjusted so as to be consistent with ARC SSTs. This is true both day and night. OEv2 SSTs are less noisy (better precision) and are associated with less biased uncertainty estimates, whereas ARC SSTs gave better coastal zone results. OEv2 is available only for ATSR-2 and AATSR, and is by design consistent with ARC. Use of OEv2 for the later sensors and ARC for ATSR-1 is a plausible selection outcome, and is favoured by the selection outcome for AVHHR discussed in the next two paragraphs.

Two algorithms, OEv2 and incremental regression, IR, are difficult to separate for the case of night-time AVHRRs in terms of the quantitative metrics such as bias, stability, precision and sensitivity. The main distinction is that OEv2 is cross-referenced to ARC SSTs and therefore independent of in situ observations, whereas IR is empirically tuned to drifting buoys.

There is a clearer separation between OEv2 and IR for day-time AVHRR retrievals. OEv2 performs either comparably to or better than IR (depending on the sensor) in terms of bias and stability. OEv2 by design delivers near-ideal SST sensitivity, at the expense being noisier (poorer precision), whereas IR coefficients are defined by a least squares fit that minimises noise but is typically only 60% to 85% sensitive to SST variations. SST sensitivity has a higher weight in our selection criteria than precision. Marginal ice zone biases are comparable; coastal zone biases are better with OEv2. OEv2 is independent, IR wholly dependent. For day-time AVHRR SSTs, therefore, OEv2 comes out better across the metrics as a whole. We note in addition that the optimal estimation framework includes the possibility to calculate various “cost” parameters. These measure the degree to which the solution fits the observations and prior assumptions, relative to their uncertainties. The cost is a useful quality indicator (Merchant et al., 2008; RD.221) and therefore an important asset within the processing system. The IR approach does not offer this possibility.

There is an advantage of coherence and consistency to SST CCI in selecting the same algorithm for different sensors, where justifiable. This will make the products more consistent over the SST CCI outputs as a whole, and has a lower overhead in implementing and maintaining algorithms. OEv2 and IR do comparably (other than on independence, which favours OEv2) for night-time AVHRR, while OEv2 is clearly preferable for day-time AVHRR. Therefore, a justifiable and maximally consistent outcome is to select OEv2 for all AVHRR observations.

The outcome of the SST CCI algorithm selection is therefore as follows: OEv2 is the selected algorithm, based on metrics comparing algorithm performance for ATSR-2, AATSR, NOAA 17, 18 and 19, and Metop AVHRR. OEv2 will be developed for remaining AVHRRs to be used within SST CCI (but not considered in the selection exercise), while ARC coefficients will be used for ATSR-1.

APPENDIX OF FULL RESULTS

A1. Along Track Scanning Radiometer results

A1.1 Brief account of relevant instrument characteristics

The algorithms assessed (A1.2) in the Algorithm Selection process for ATSR-2 and AATSR all utilise the dual-view capability of these instruments. The selection of available channels used varies between day and night and across algorithms. ARC, OE v1, and OE v2 use a 3-channel (3.7 µm, 11 µm and 12 µm) retrieval during the night and a 2-channel (11 µm and 12 µm) during the day. The 3.7 µm channel is not used during the day due to solar contamination effects. The ORAC retrieval uses all visible and infrared channels and can therefore only be applied during the day.

The spatial resolution of the ATSRs is 1 km^2 at the centre of the nadir swath and 1.5 x 2 km at the centre of the forward swath. For ATSR-2, limitations on data rates can result in a reduced swath width (180 km compared to 512 km) during the day, therefore reducing the spatial coverage of the instrument.

ATSR-2 orbits with equator crossing time of 10:30 while AATSR orbits with equator crossing time of 10:00 (both descending).

A1.2 List of algorithms

The following algorithms have been assessed for AVHRR in the Algorithm Selection process:

- ARC SST coefficient retrievals (A1.3)
- CCI SST Optimal Estimation v1 (A1.4)
- CCI SST Optimal Estimation v2 (A1.5)
- Oxford-RAL SST Estimation (A1.6)

All algorithms listed above have been documented in the SST CCI “Algorithm Theoretical Basis Document v0” [RD.225].

A1.3 Metrics for ARC SST coefficient retrievals

A1.3.1 Bias

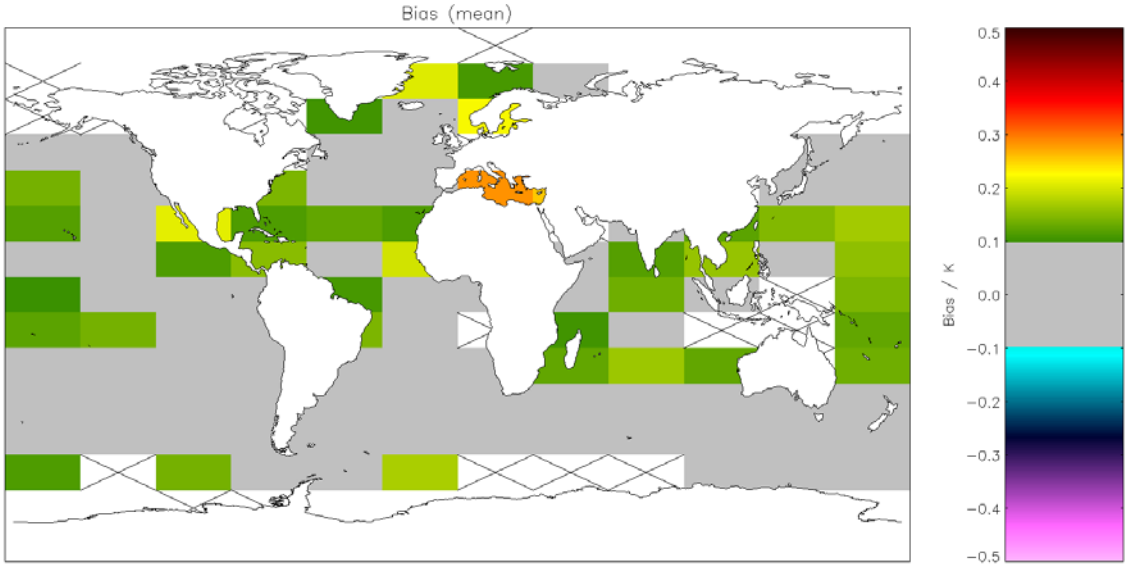


Figure A1.3.1-1: Global map of mean bias (satellite-buoy) using ARC SST coefficient retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.087 K (**Table 5.2.1**).

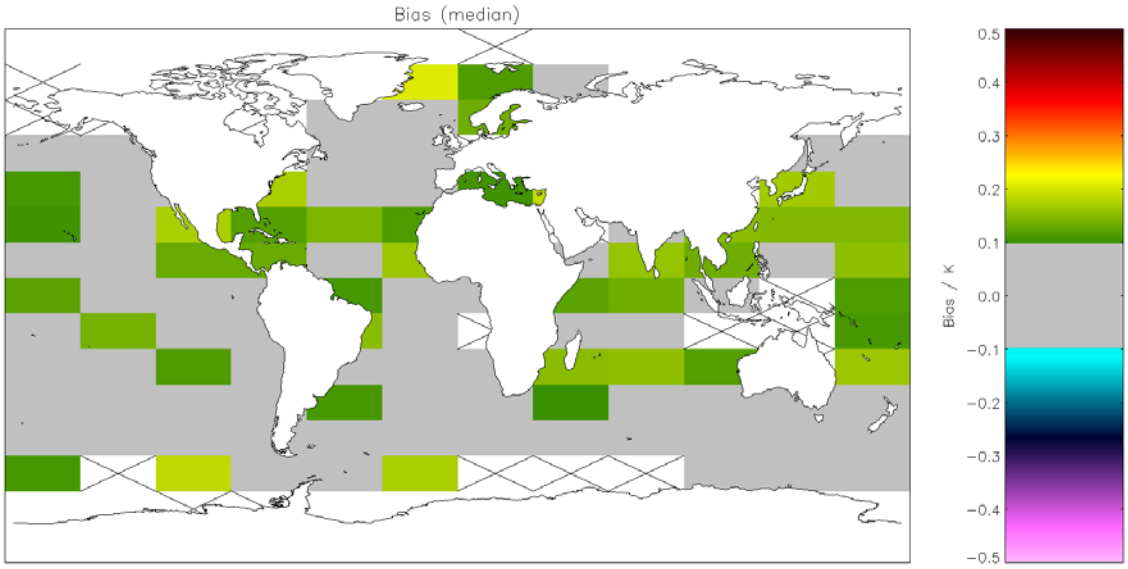


Figure A1.3.1-2: Global map of median bias (satellite-buoy) using ARC SST coefficient retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.095 K (**Table 5.2.1**).

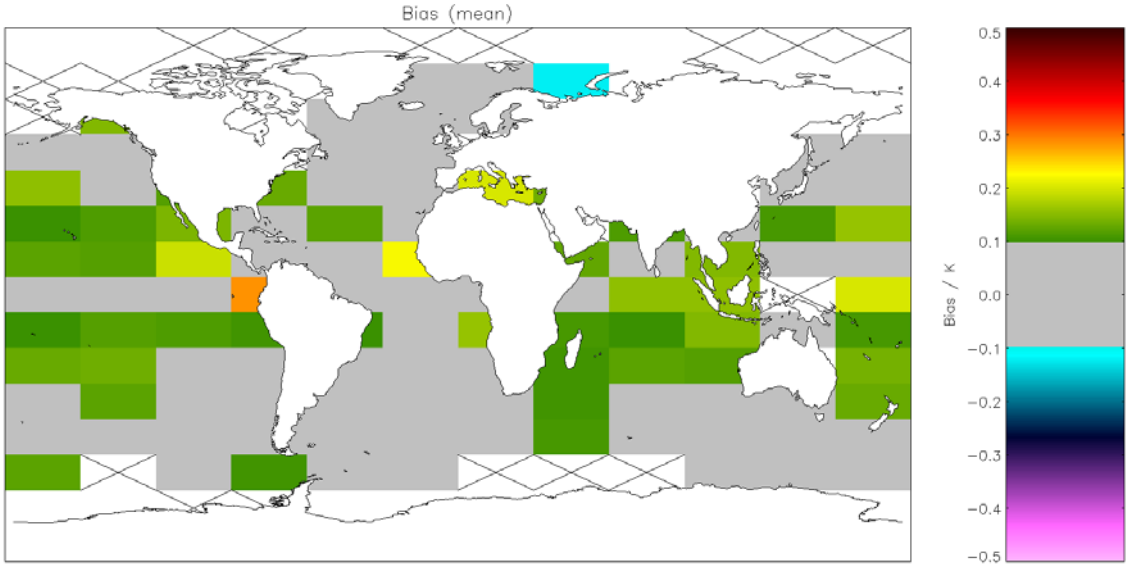


Figure A1.3.1-3: Global map of mean bias (satellite-buoy) using ARC SST coefficient retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.082 K (**Table 5.2.2**).

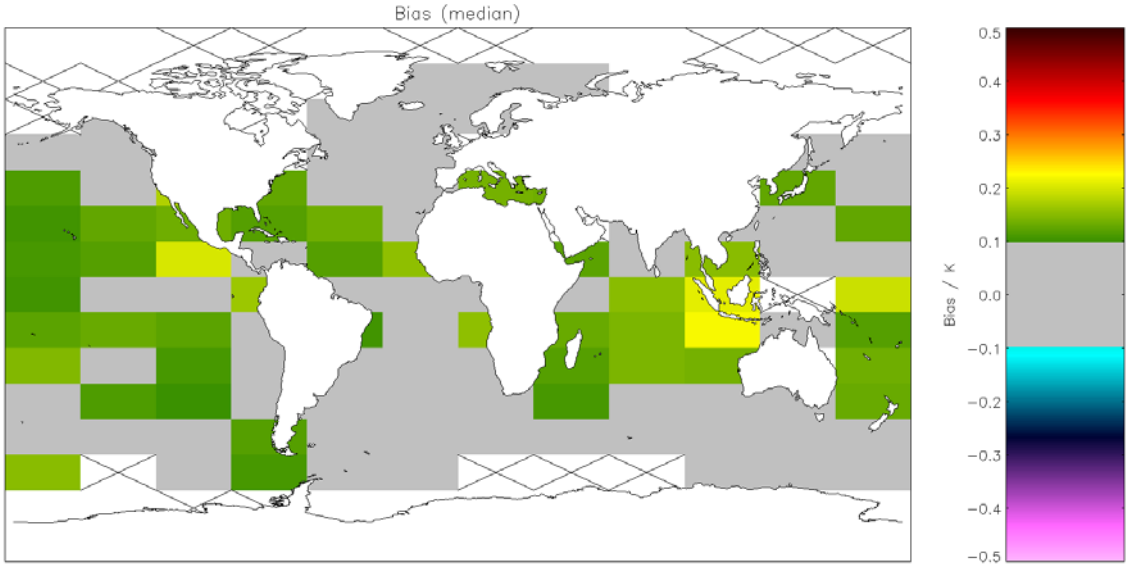


Figure A1.3.1-4: Global map of median bias (satellite-buoy) using ARC SST coefficient retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.094 K (**Table 5.2.2**).

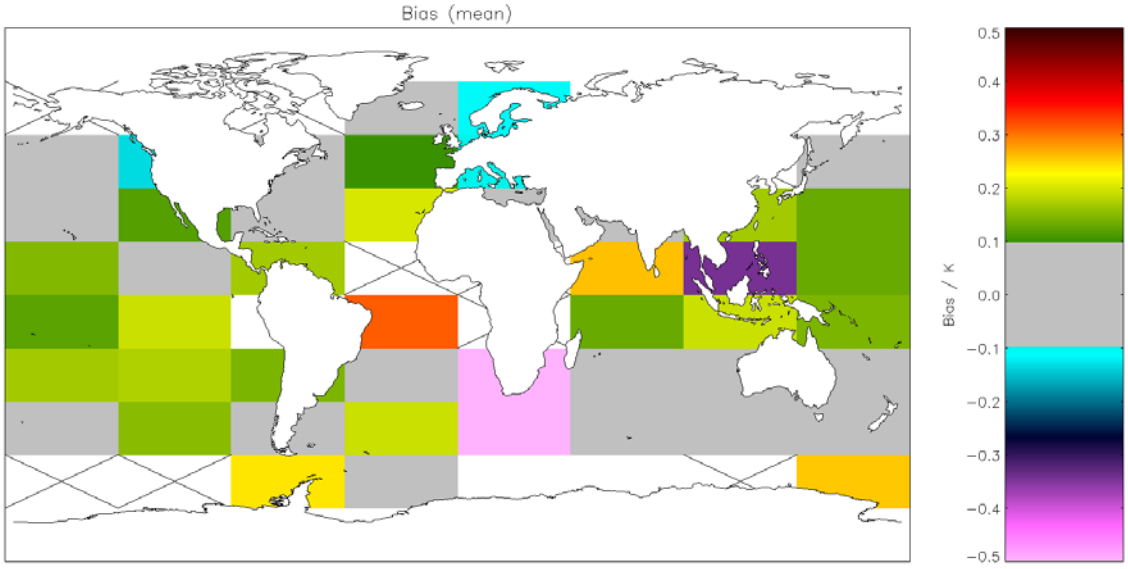


Figure A1.3.1-5: Global map of mean bias (satellite-buoy) using ARC SST coefficient retrieval for night-time ATSR2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global mean bias = 0.062 K (**Table 5.3.1**).

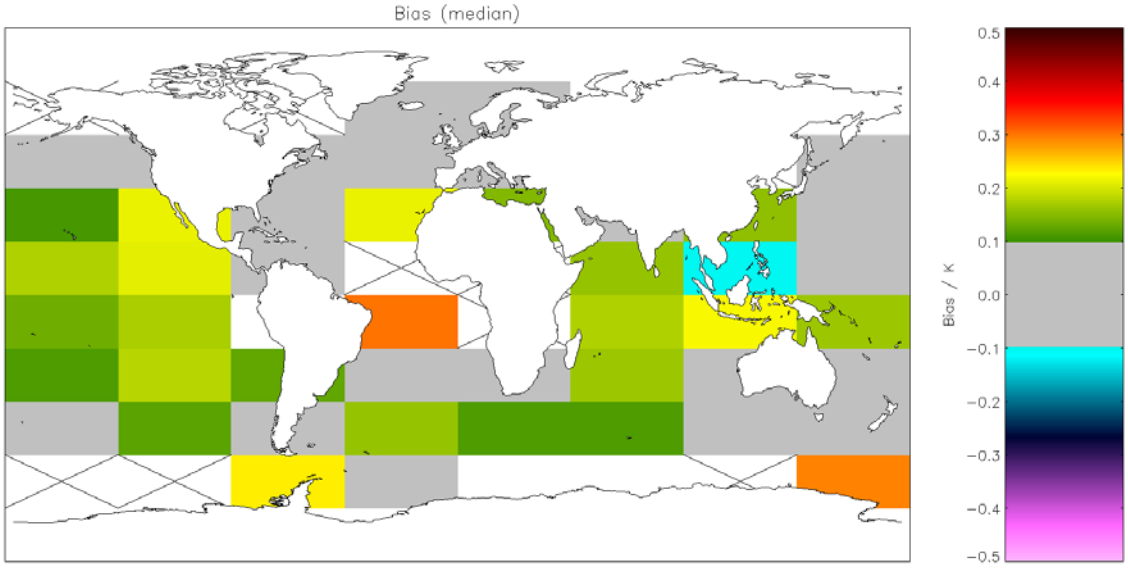


Figure A1.3.1-6: Global map of median bias (satellite-buoy) using ARC SST coefficient retrieval for night-time ATSR2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global median bias = 0.107 K (**Table 5.3.1**).

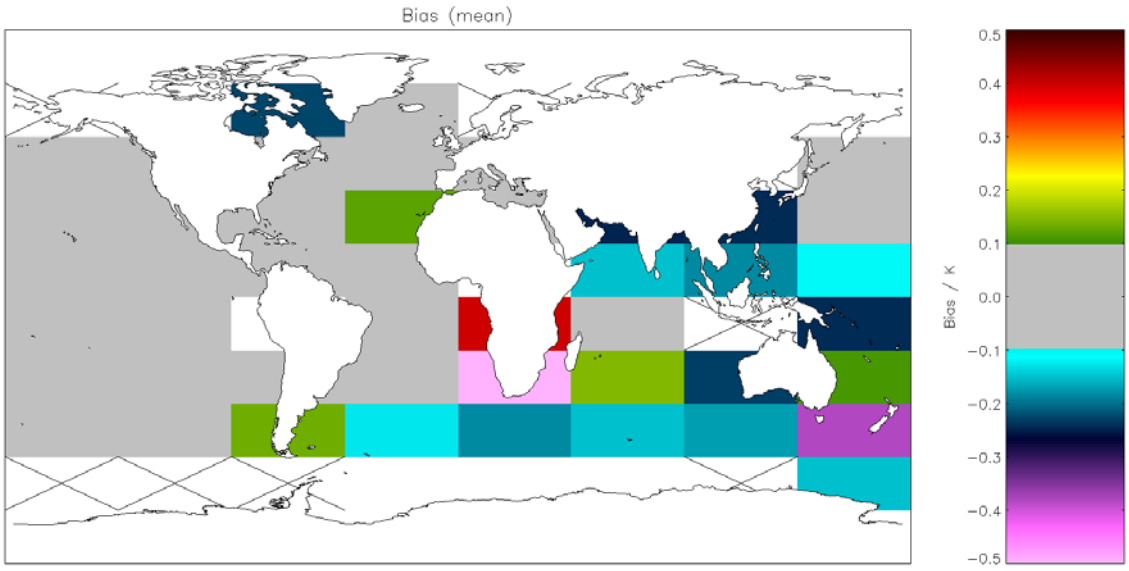


Figure A1.3.1-7: Global map of mean bias (satellite-buoy) using ARC SST coefficient retrieval for day-time ATSR2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global mean bias = -0.015 K (**Table 5.3.2**).

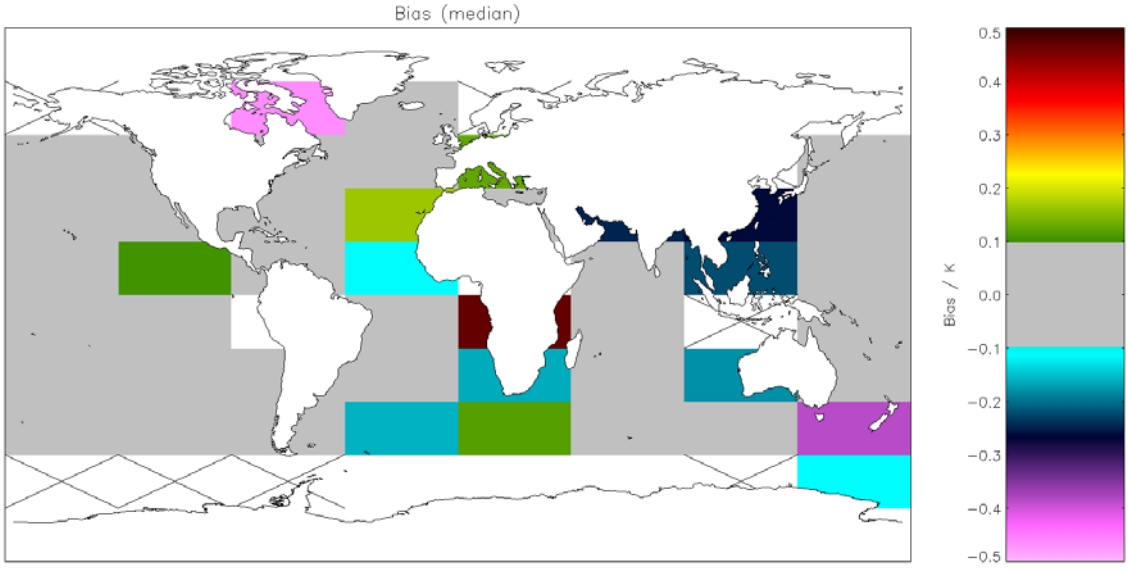


Figure A1.3.1-8: Global map of median bias (satellite-buoy) using ARC SST coefficient retrieval for day-time ATSR2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global median bias = 0.025 K (**Table 5.3.2**).

A1.3.2 Precision

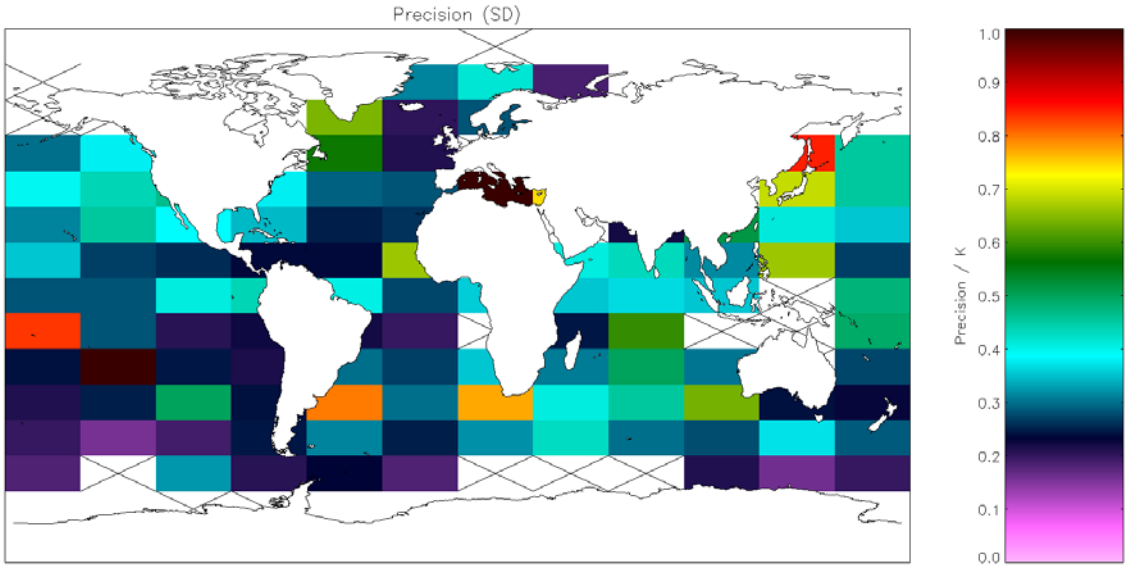


Figure A1.3.2-1: Global map of precision (mean of cell SDs) using ARC SST coefficient retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.442 K (**Table 5.2.1**).

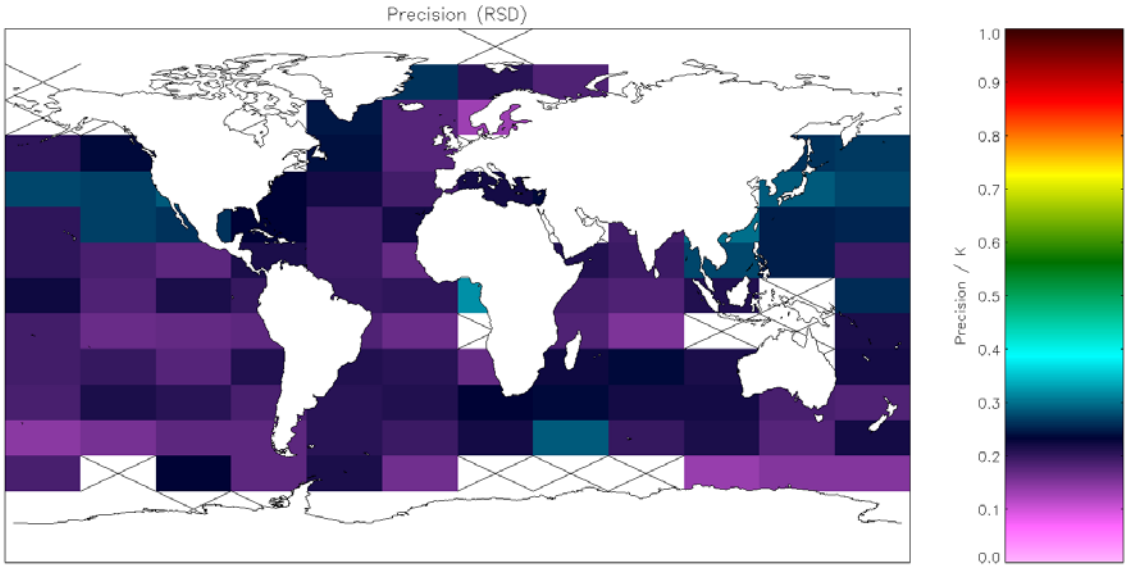


Figure A1.3.2-2: Global map of precision (median of cell RSDs) using ARC SST coefficient retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.212 K (**Table 5.2.1**).

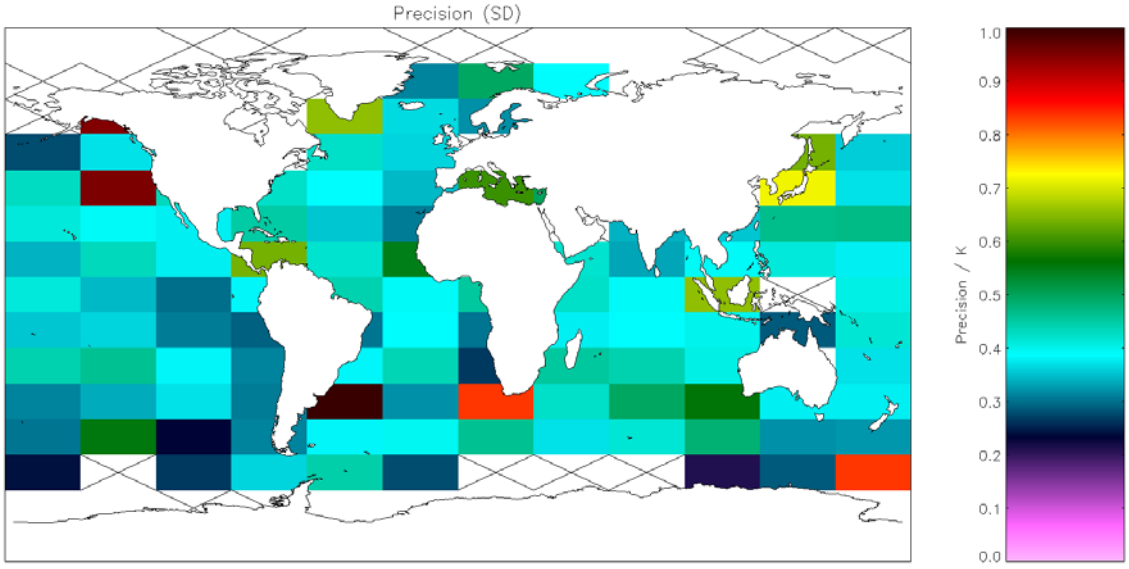


Figure A1.3.2-3: Global map of precision (mean of cell SDs) using ARC SST coefficient retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.463 K (**Table 5.2.2**).

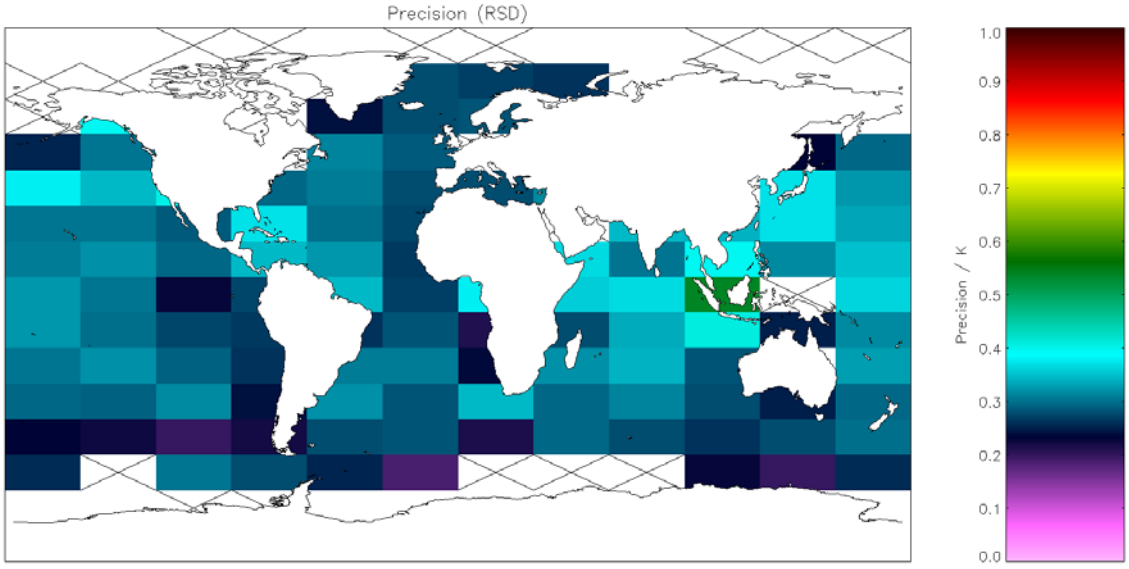


Figure A1.3.2-4: Global map of precision (median of cell RSDs) using ARC SST coefficient retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.306 K (**Table 5.2.2**).

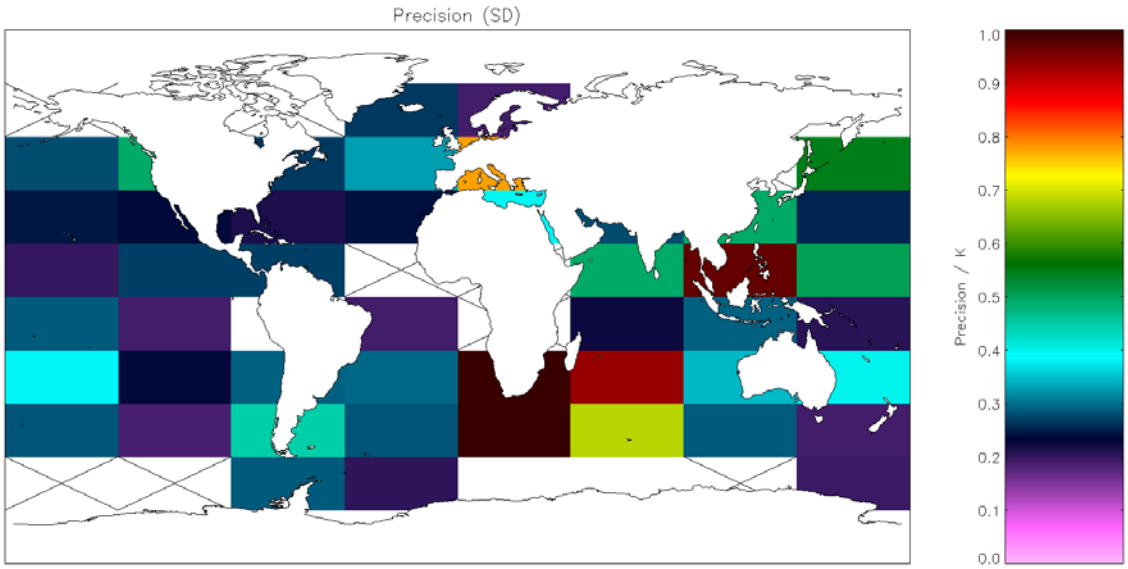


Figure A1.3.2-5: Global map of precision (mean of cell SDs) using ARC SST coefficient retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global SD = 0.485 K (**Table 5.3.1**).

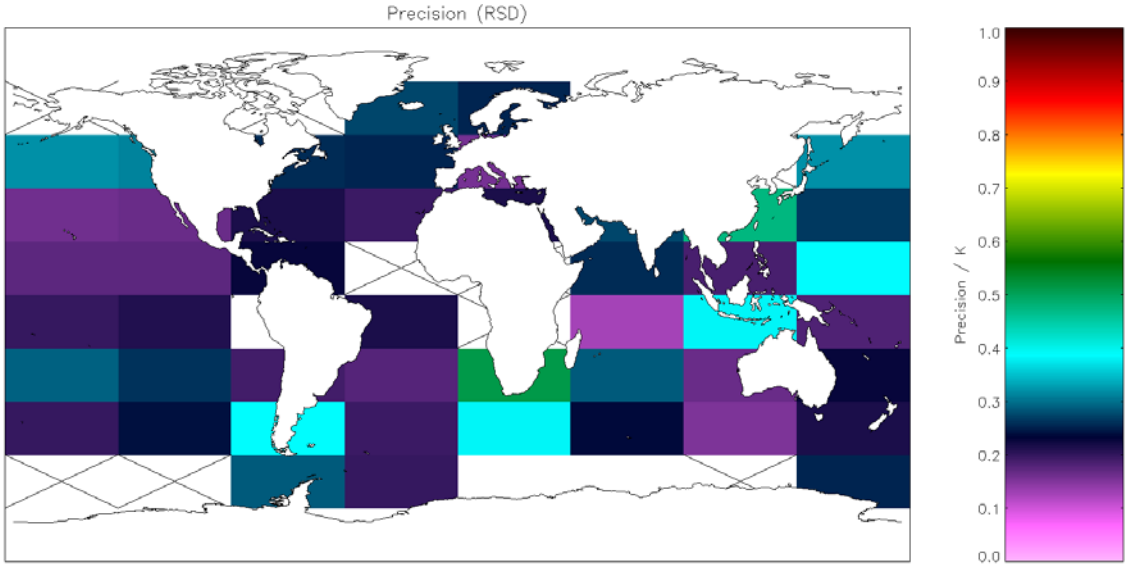


Figure A1.3.2-6: Global map of precision (median of cell RSDs) using ARC SST coefficient retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global RSD = 0.250 K (**Table 5.3.1**).

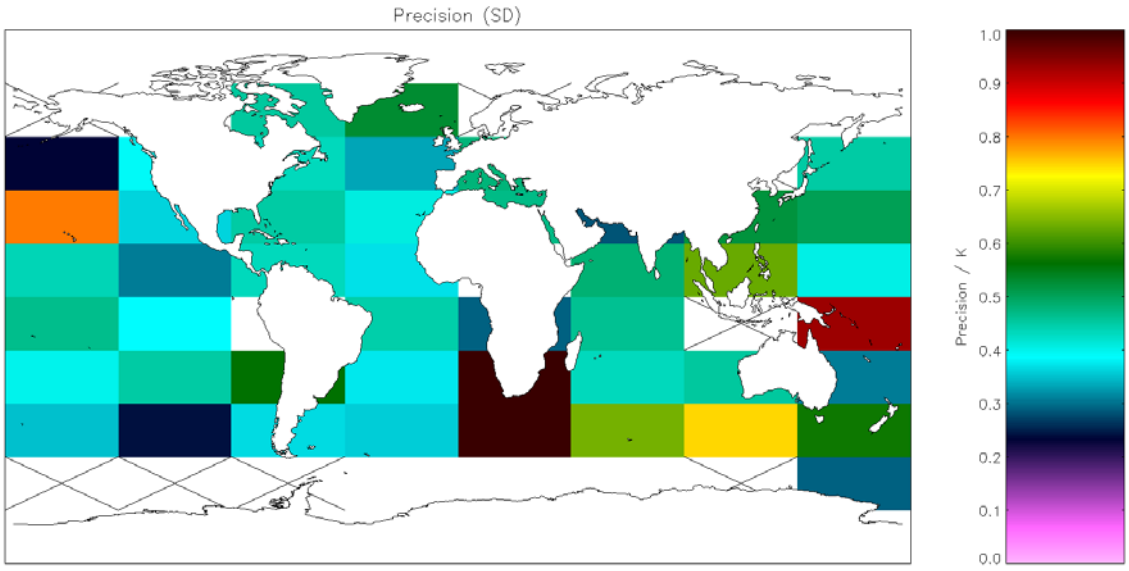


Figure A1.3.2-7: Global map of precision (mean of cell SDs) using ARC SST coefficient retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global SD = 0.520 K (**Table 5.3.2**).

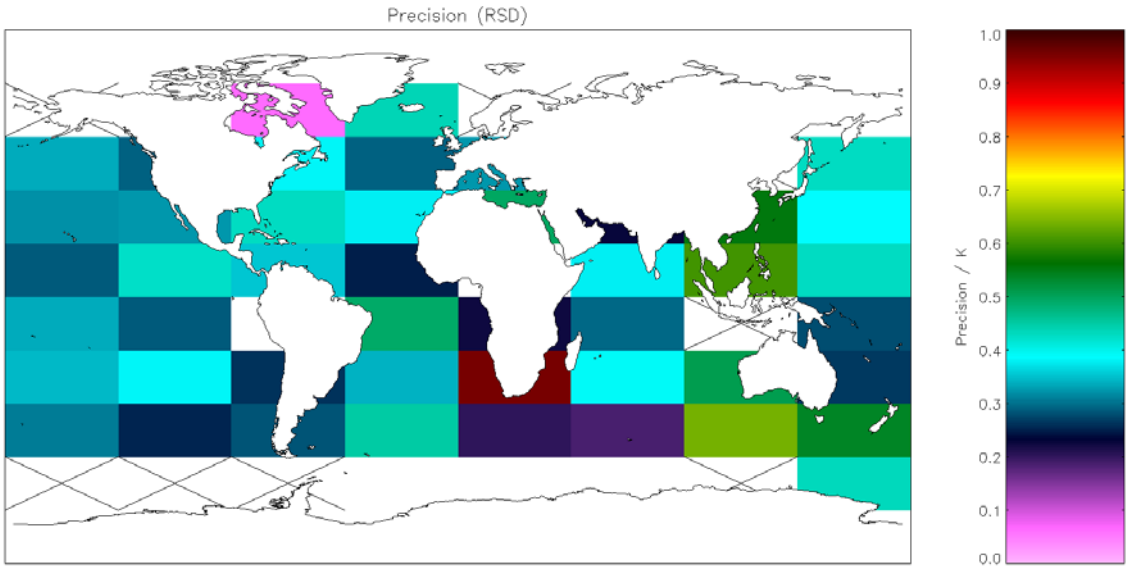


Figure A1.3.2-8: Global map of precision (median of cell RSDs) using ARC SST coefficient retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global RSD = 0.520 K (**Table 5.3.2**).

A1.3.3 Sensitivity

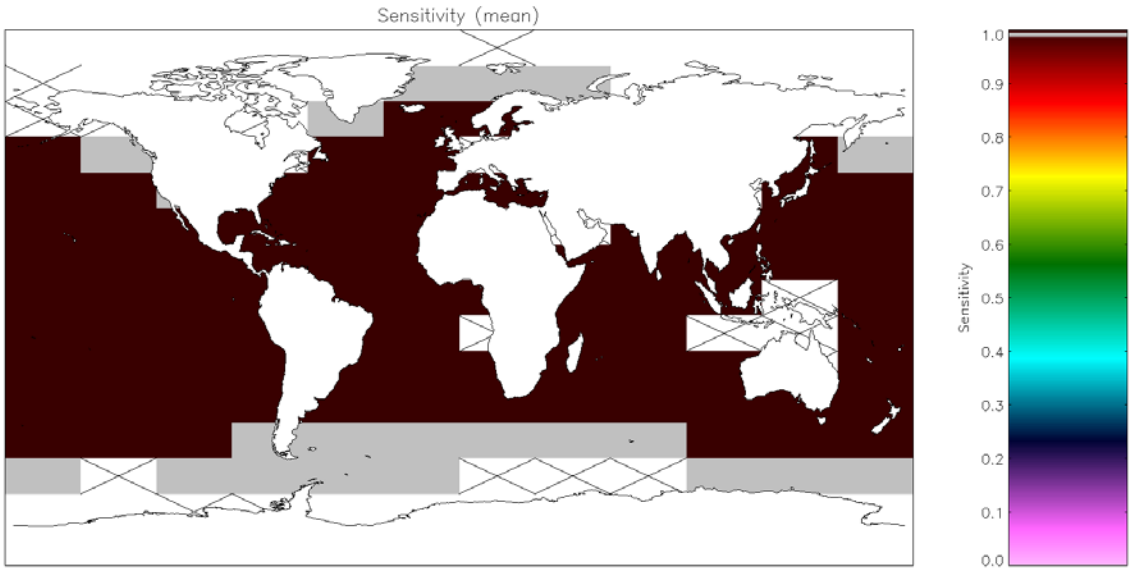


Figure A1.3.3-1: Global map of SST sensitivity using ARC SST coefficient retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.2.1**).

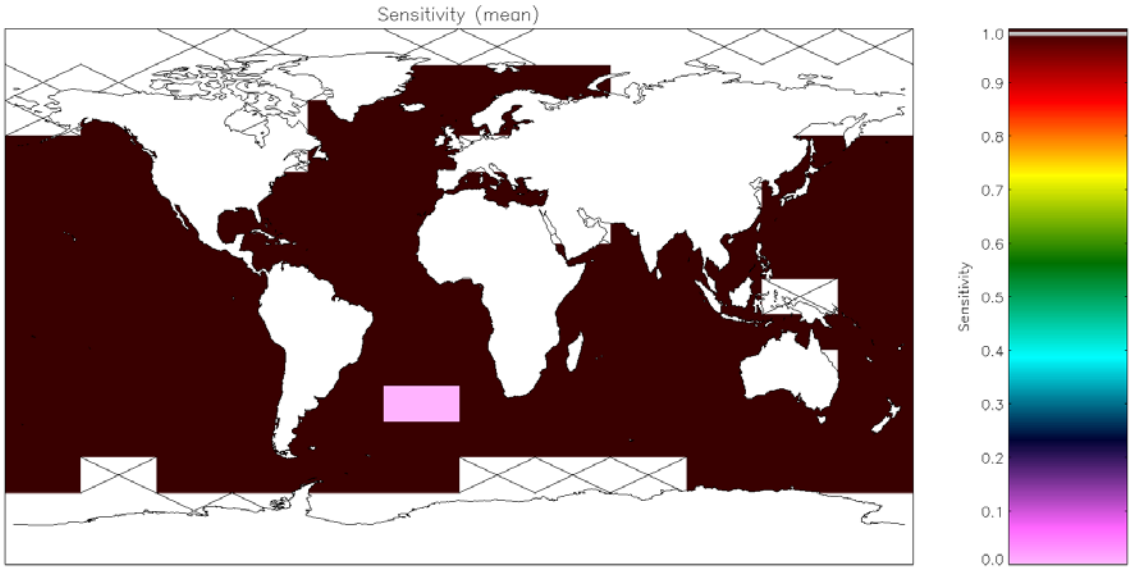


Figure A1.3.3-2: Global map of SST sensitivity using ARC SST coefficient retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.2.2**).

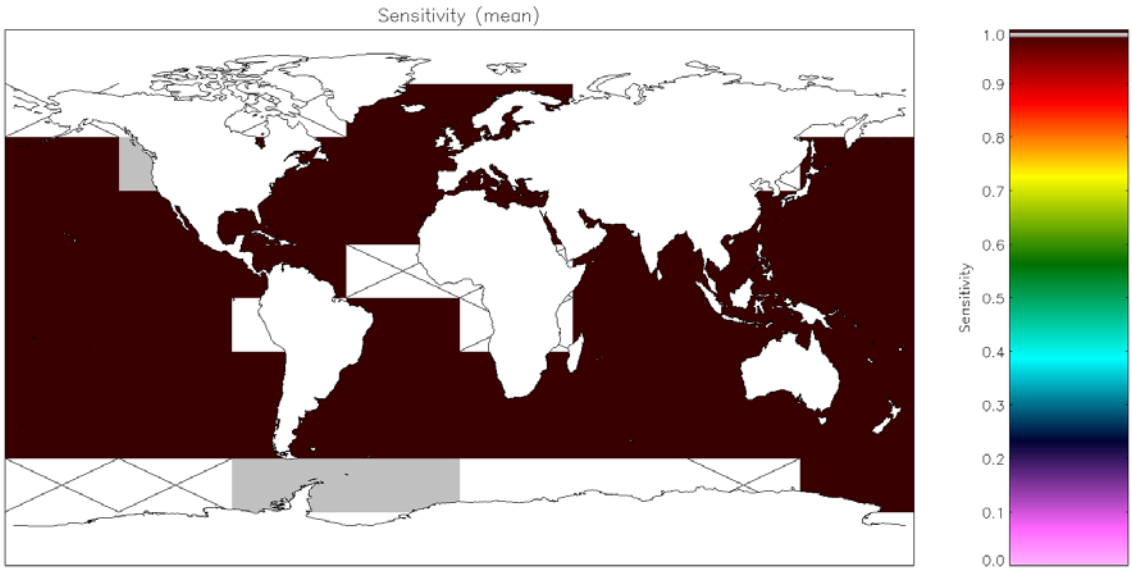


Figure A1.3.3-3: Global map of SST sensitivity using ARC SST coefficient retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.3.1**).

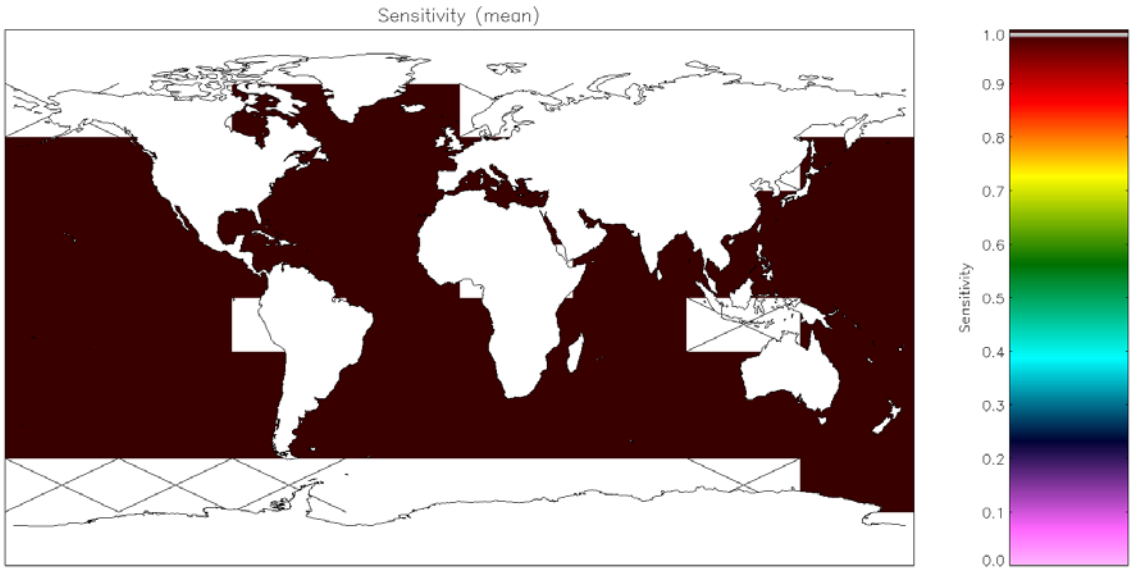


Figure A1.3.3-4: Global map of SST sensitivity using ARC SST coefficient retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.3.2**).

A1.4 Metrics for CCI SST Optimal Estimation v1

A1.4.1 Bias

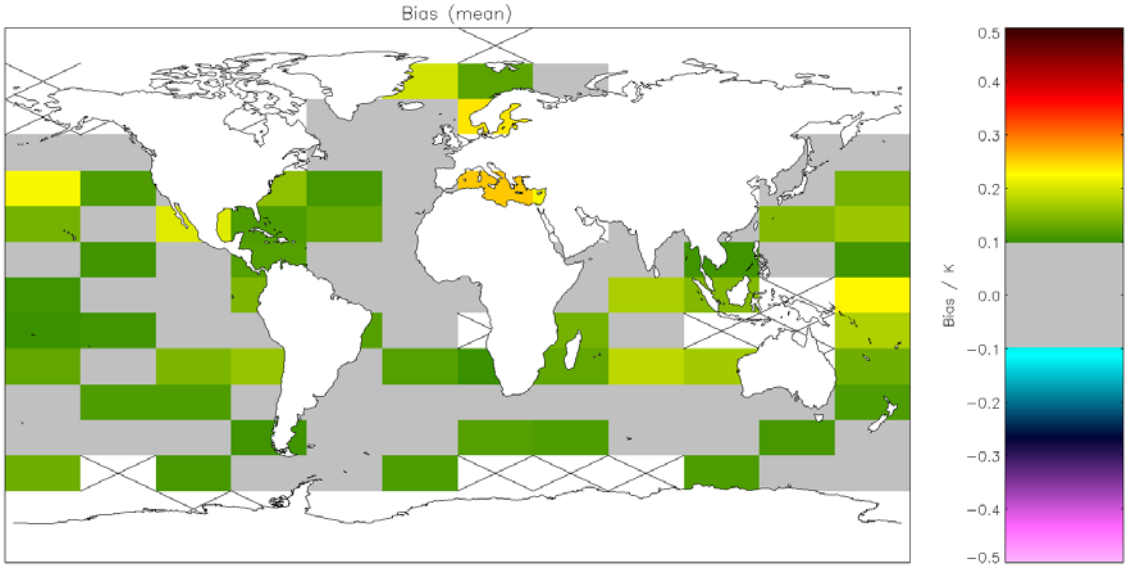


Figure A1.4.1-1: Global map of mean bias (satellite-buoy) using OE v1 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.104 K (**Table 5.2.1**).

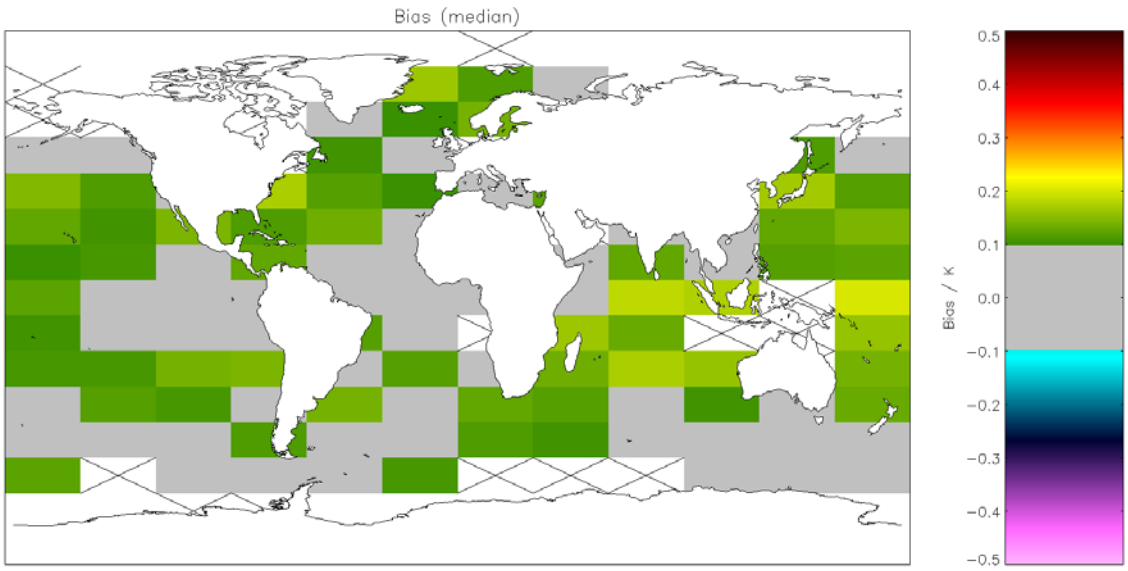


Figure A1.4.1-2: Global map of median bias (satellite-buoy) using OE v1 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.109 K (**Table 5.2.1**).

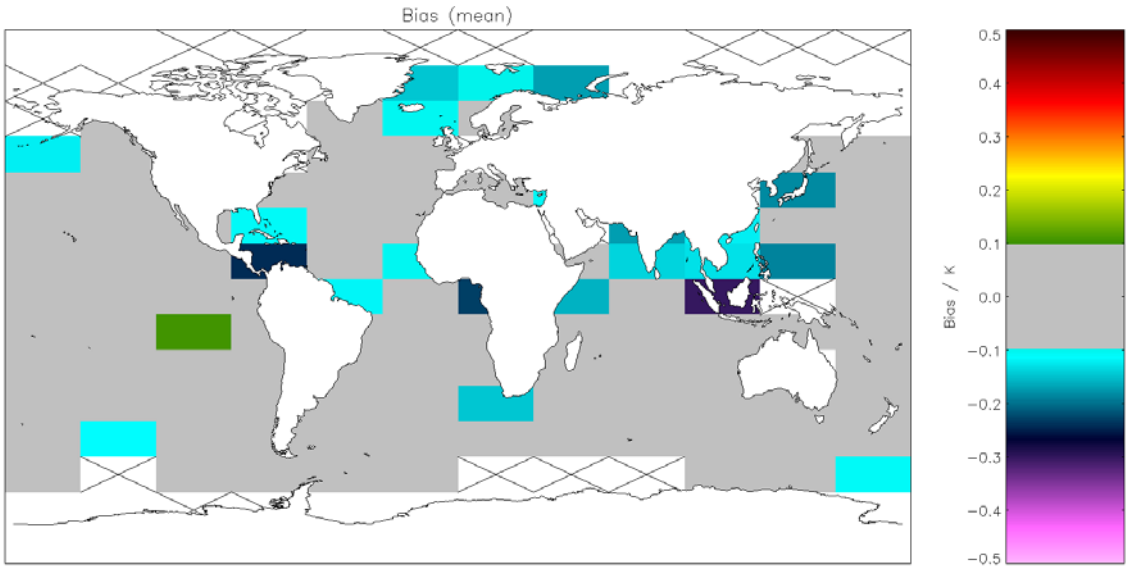


Figure A1.4.1-3: Global map of mean bias (satellite-buoy) using OE v1 retrieval for daytime AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.023 K (**Table 5.2.2**).

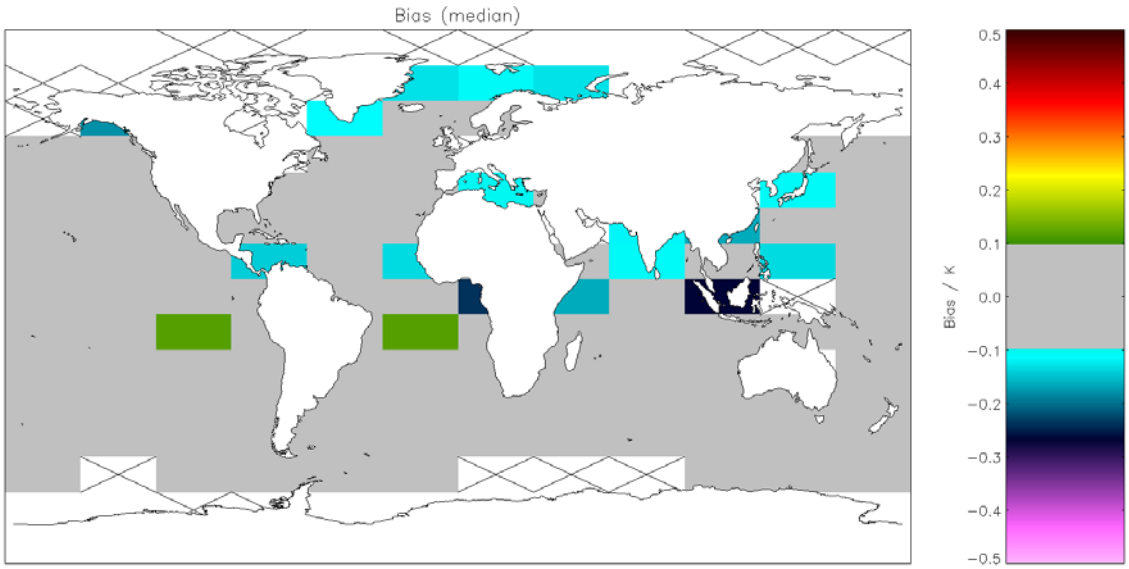


Figure A1.4.1-4: Global map of median bias (satellite-buoy) using OE v1 retrieval for daytime AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.003 K (**Table 5.2.2**).

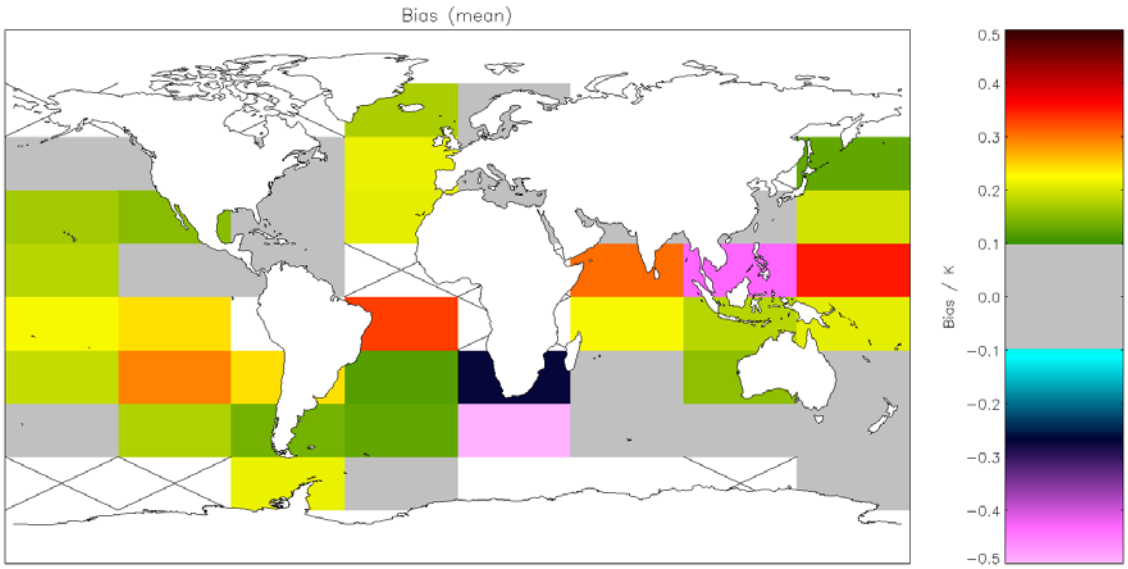


Figure A1.4.1-5: Global map of mean bias (satellite-buoy) using OE v1 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell. Global mean bias = 0.130 K (**Table 5.3.1**).

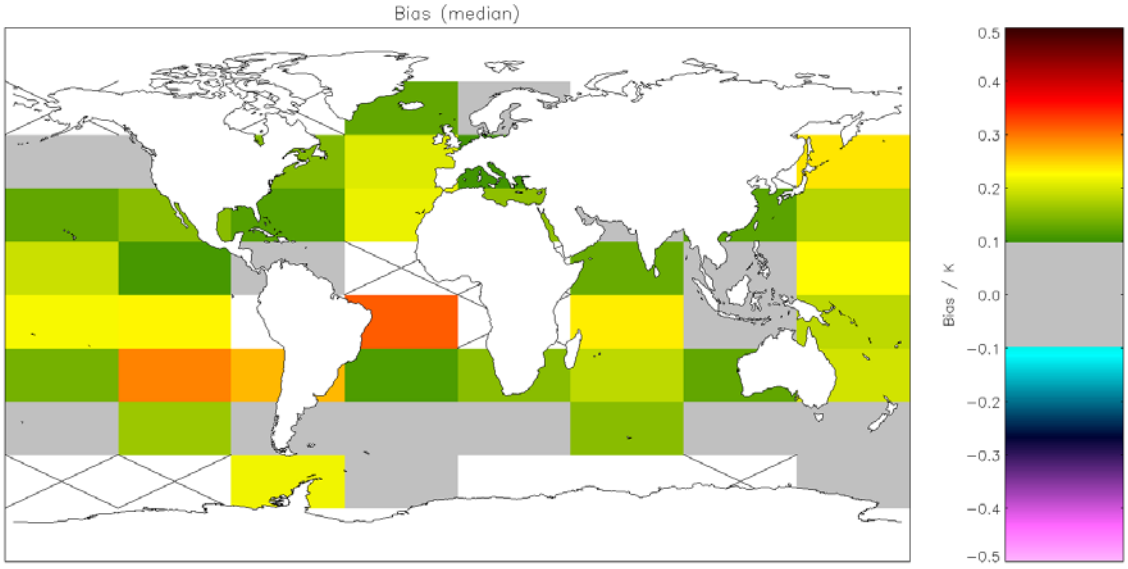


Figure A1.4.1-6: Global map of median bias (satellite-buoy) using OE v1 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell. Global median bias = 0.161 K (**Table 5.3.1**).

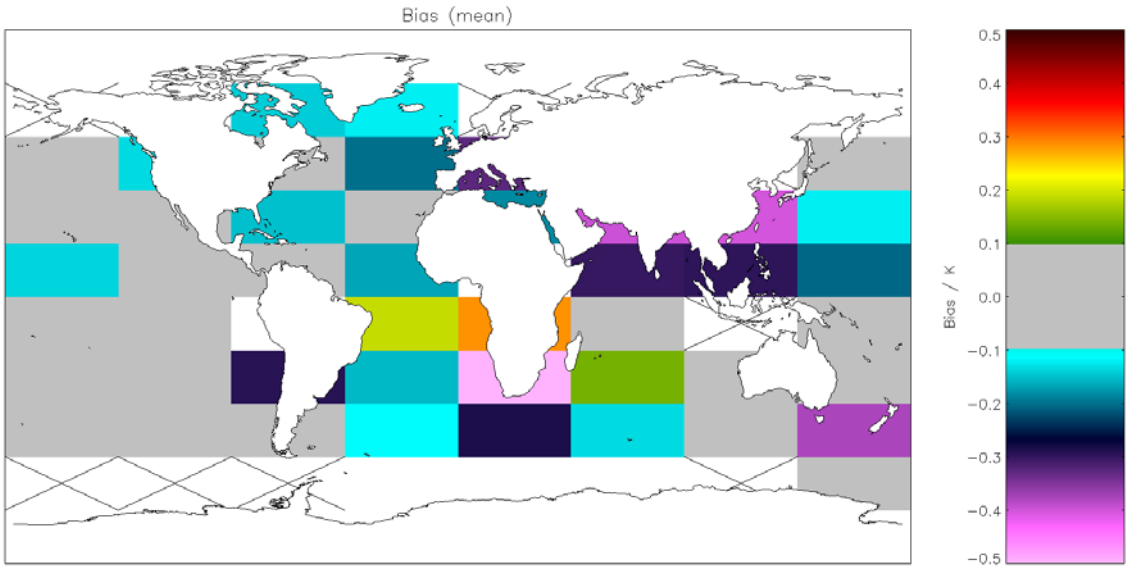


Figure A1.4.1-7: Global map of mean bias (satellite-buoy) using OE v1 retrieval for daytime ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell. Global mean bias = -0.102 K (**Table 5.3.2**).

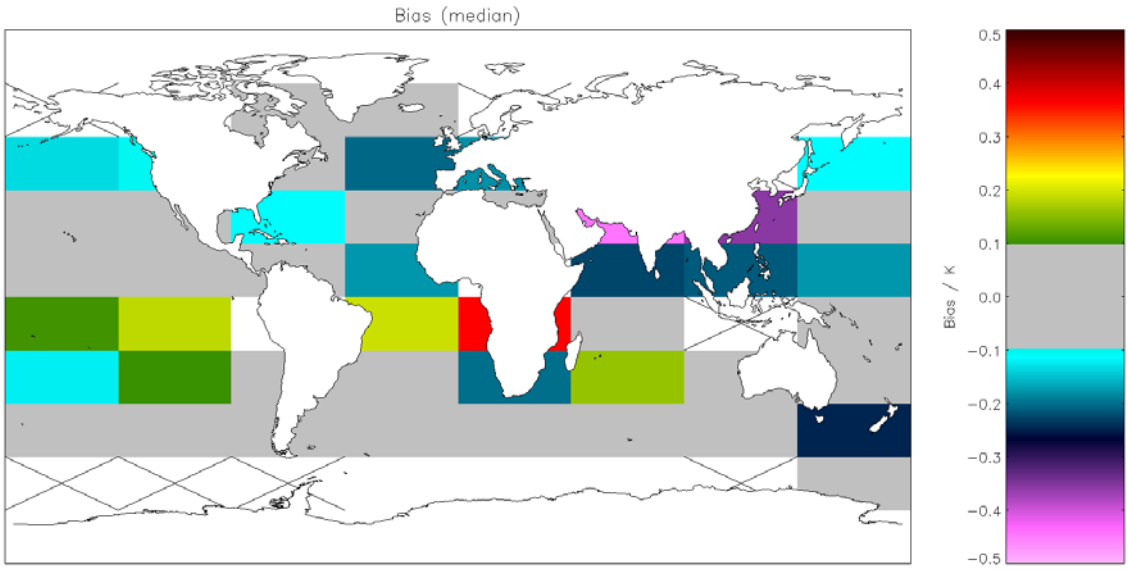


Figure A1.4.1-8: Global map of median bias (satellite-buoy) using OE v1 retrieval for daytime ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell. Global median bias = -0.053 K (**Table 5.3.2**).

A1.4.2 Precision

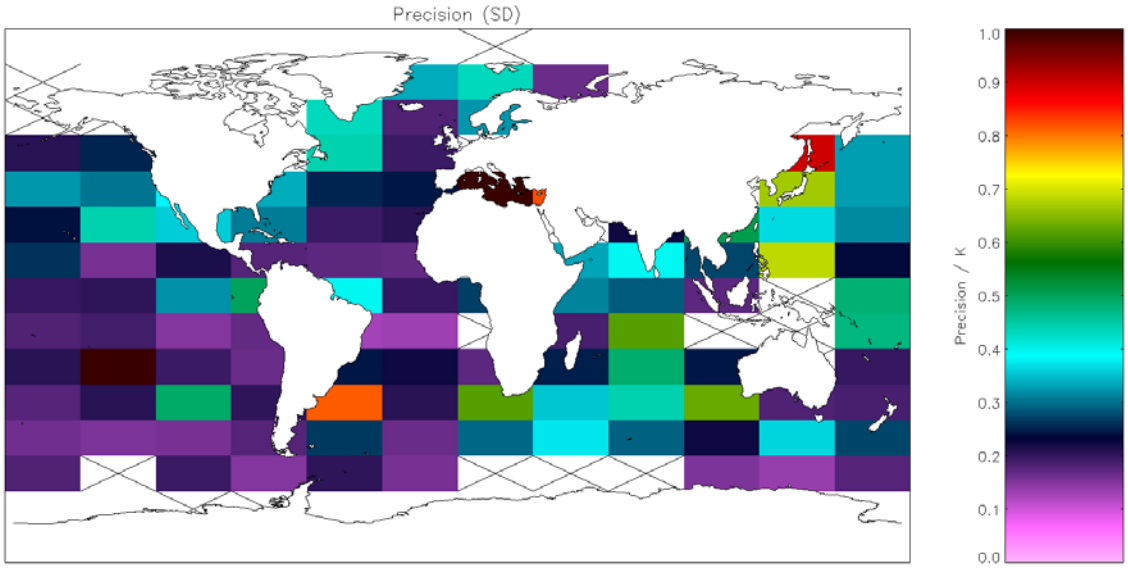


Figure A1.4.2-1: Global map of precision (mean of cell SDs) using OE v1 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.385 K (**Table 5.2.1**).

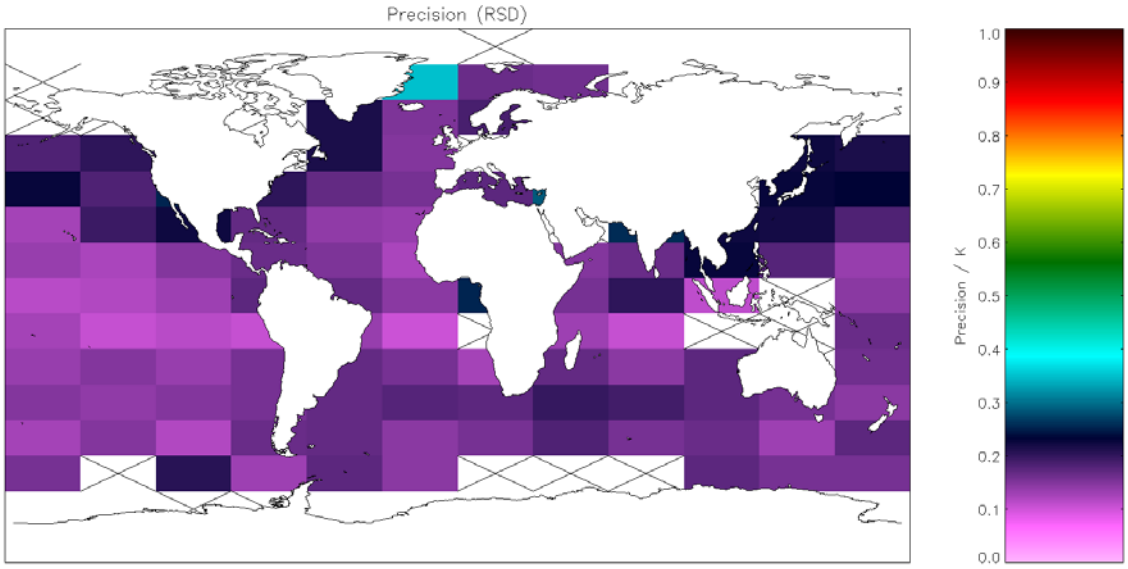


Figure A1.4.2-2: Global map of precision (median of cell RSDs) using OE v1 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.165 K (**Table 5.2.1**).

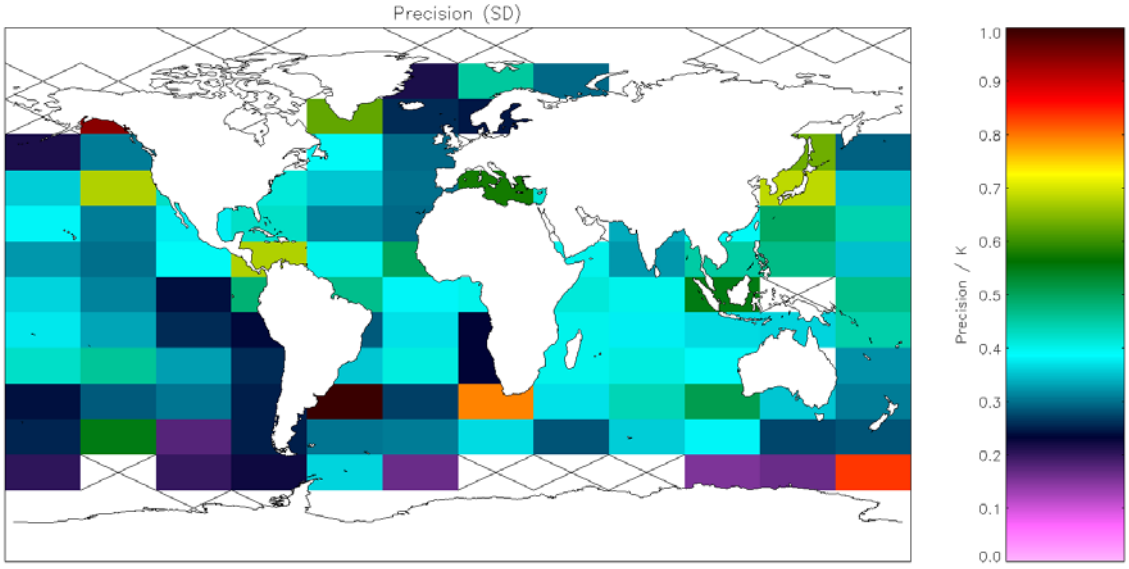


Figure A1.4.2-3: Global map of precision (mean of cell SDs) using OE v1 retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.432 K (**Table 5.2.2**).

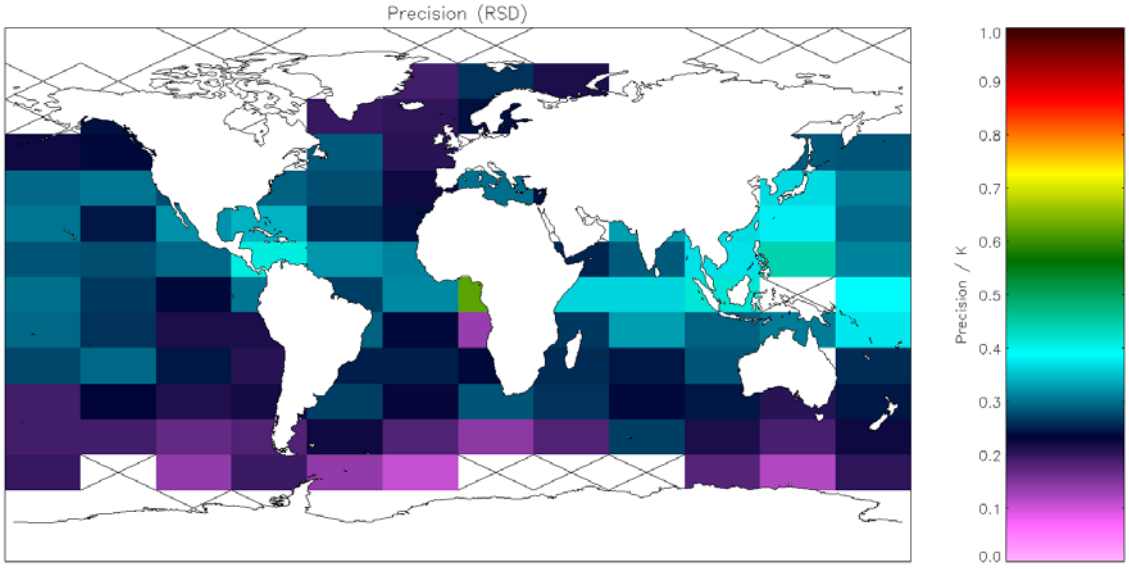


Figure A1.4.2-4: Global map of precision (median of cell RSDs) using OE v1 retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.269 K (**Table 5.2.2**).

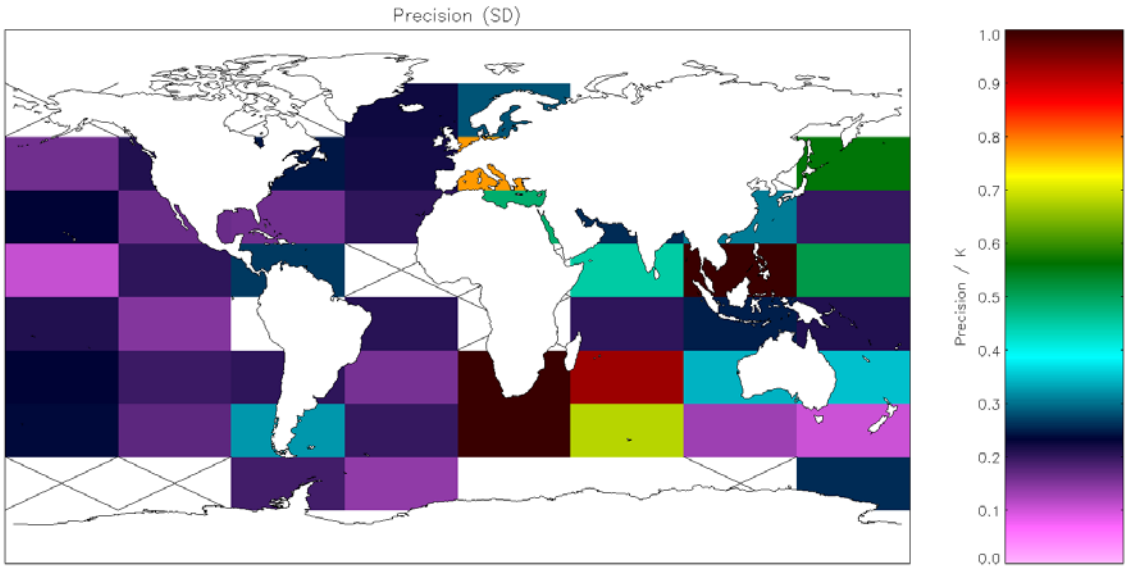


Figure A1.4.2-5: Global map of precision (mean of cell SDs) using OE v1 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell. Global SD = 0.436 K (**Table 5.3.1**).

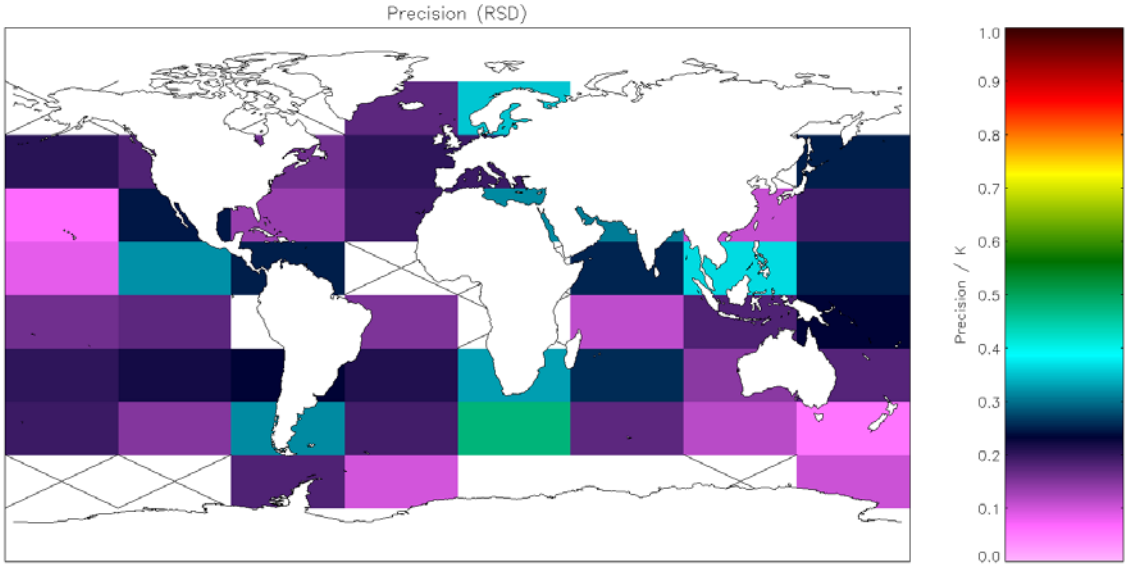


Figure A1.4.2-6: Global map of precision (median of cell RSDs) using OE v1 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell. Global RSD = 0.202 K (**Table 5.3.1**).

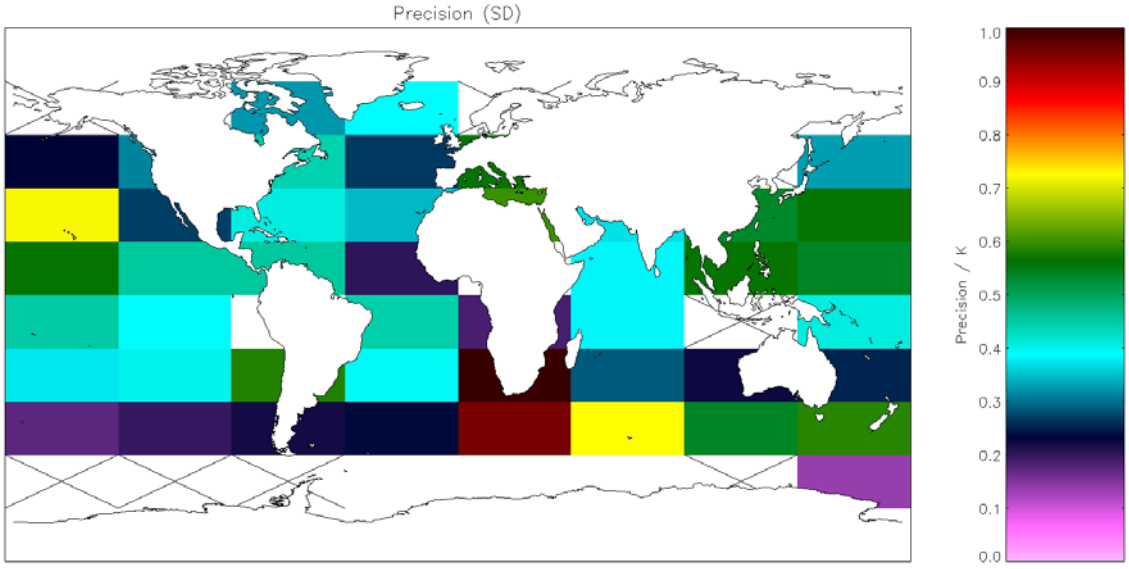


Figure A1.4.2-7: Global map of precision (mean of cell SDs) using OE v1 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell. Global SD = 0.491 K (**Table 5.3.2**).

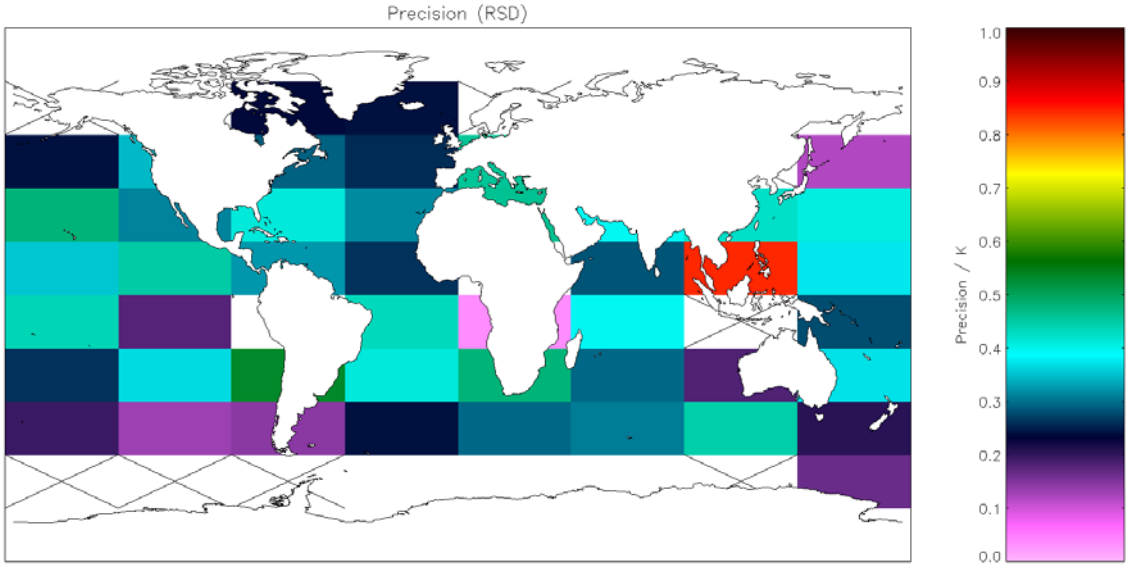


Figure A1.4.2-8: Global map of precision (median of cell RSDs) using OE v1 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell. Global RSD = 0.346 K (**Table 5.3.2**).

A1.4.3 Sensitivity

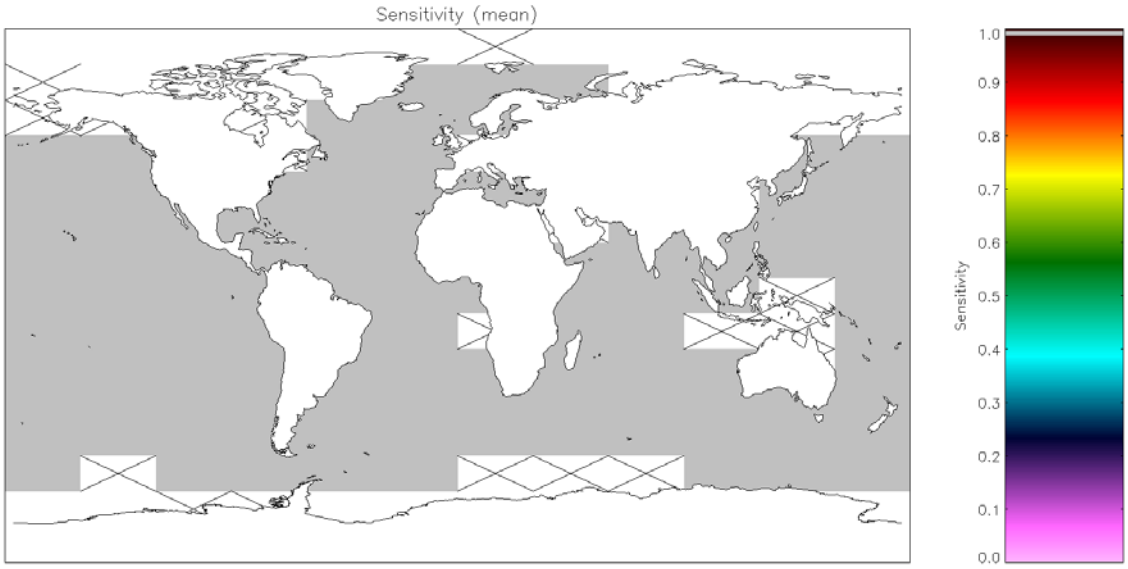


Figure A1.4.3-1: Global map of SST sensitivity using OE v1 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.2.1**).

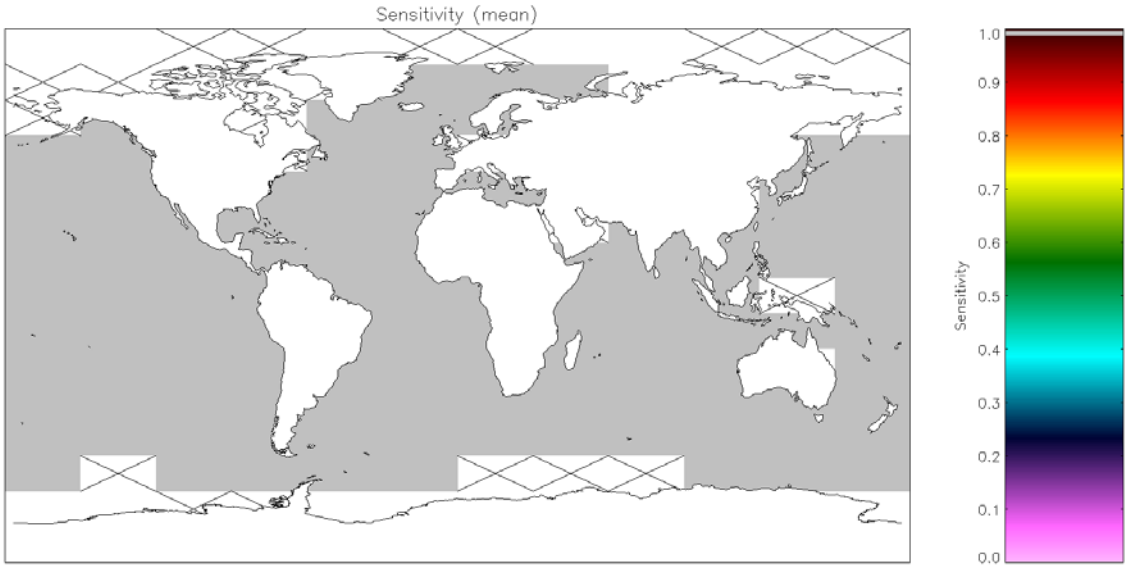


Figure A1.4.3-2: Global map of SST sensitivity using OE v1 retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.2.2**).

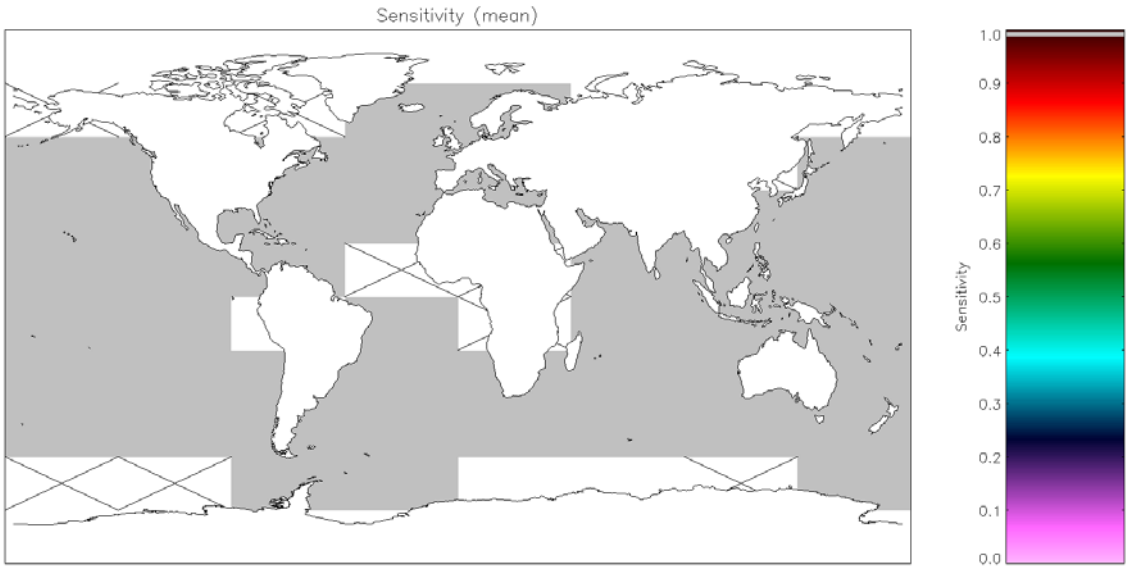


Figure A1.4.3-3: Global map of SST sensitivity using OE v1 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell (**Table 5.3.1**).

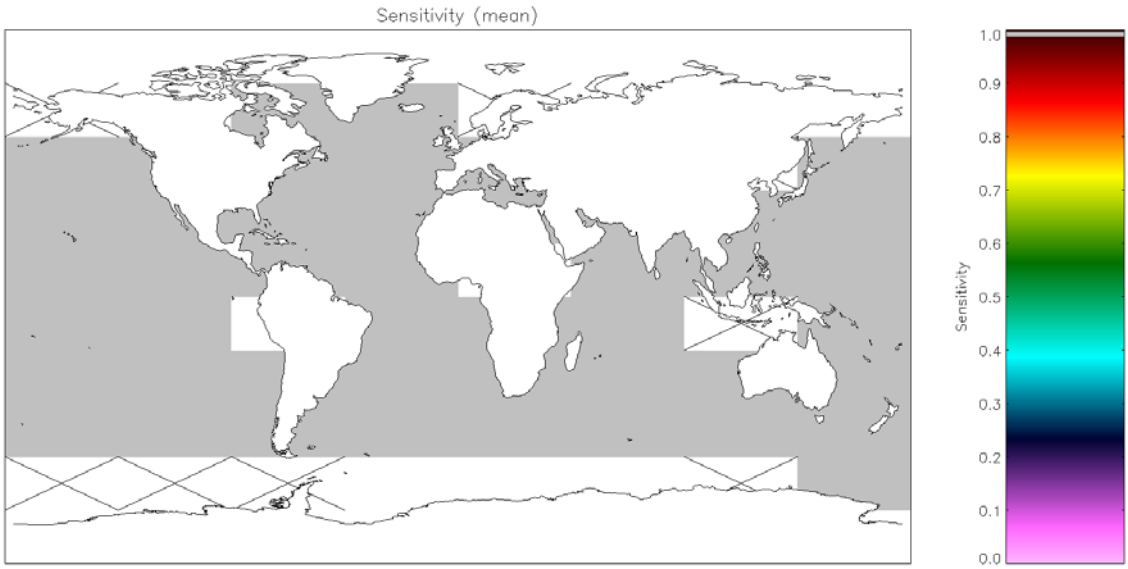


Figure A1.4.3-4: Global map of SST sensitivity using OE v1 retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.2 K being insignificant for the cell (**Table 5.3.2**).

A1.5 Metrics for CCI SST Optimal Estimation v2

A1.5.1 Bias

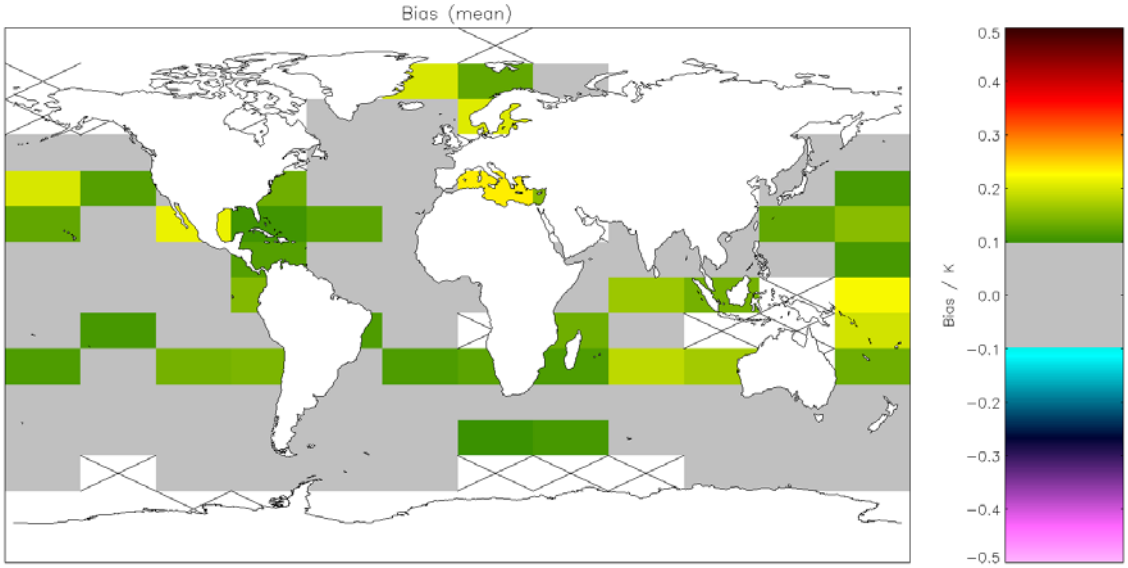


Figure A1.5.1-1: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.091 K (**Table 5.2.1**).

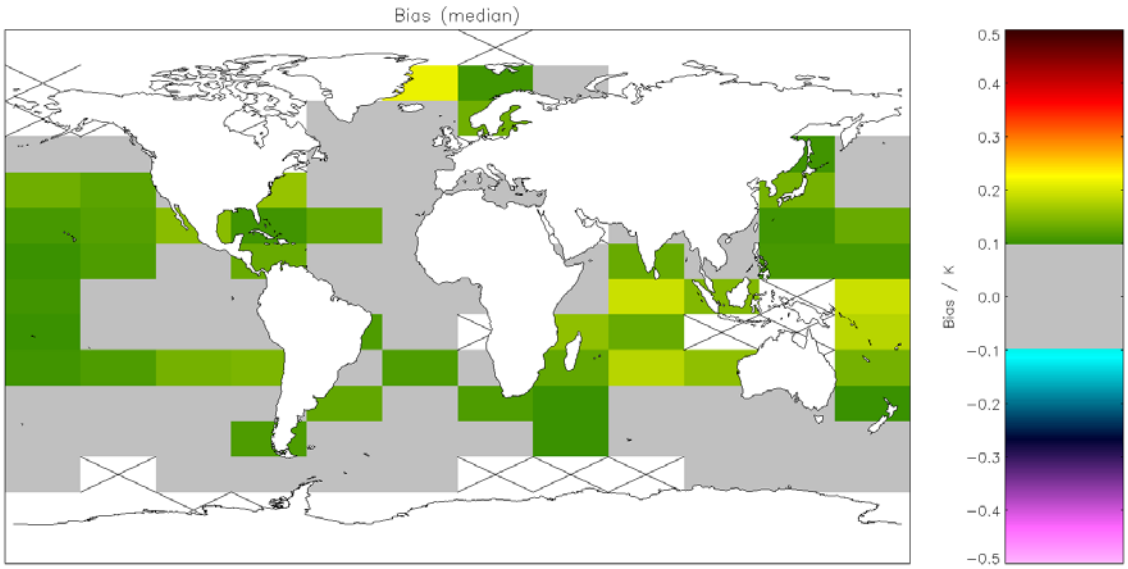


Figure A1.5.1-2: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.097 K (**Table 5.2.1**).

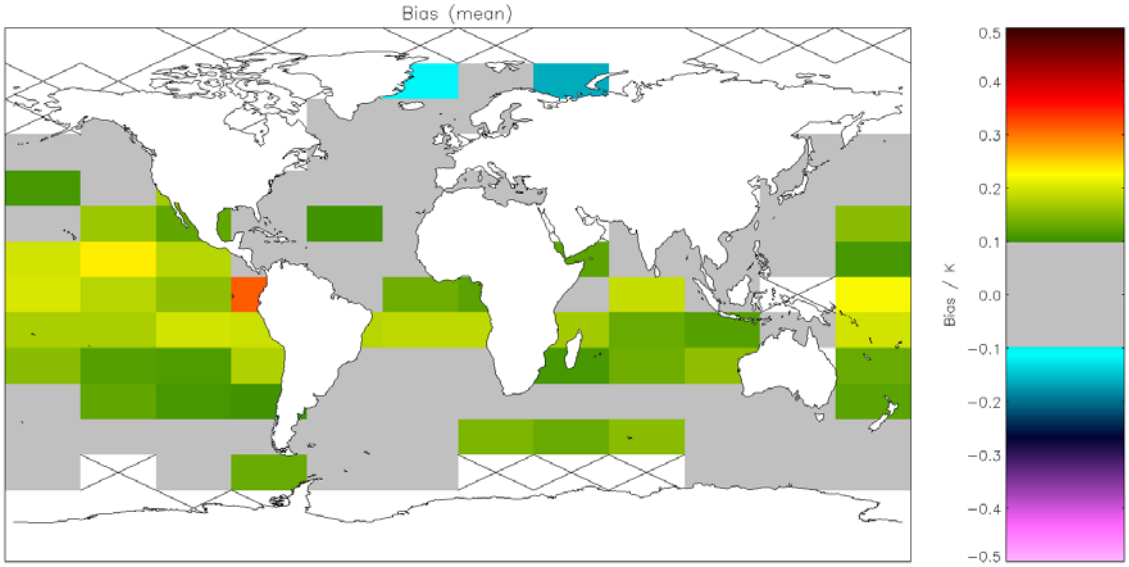


Figure A1.5.1-3: Global map of mean bias (satellite-buoy) using OE v2 retrieval for daytime AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.087 K (**Table 5.2.2**).

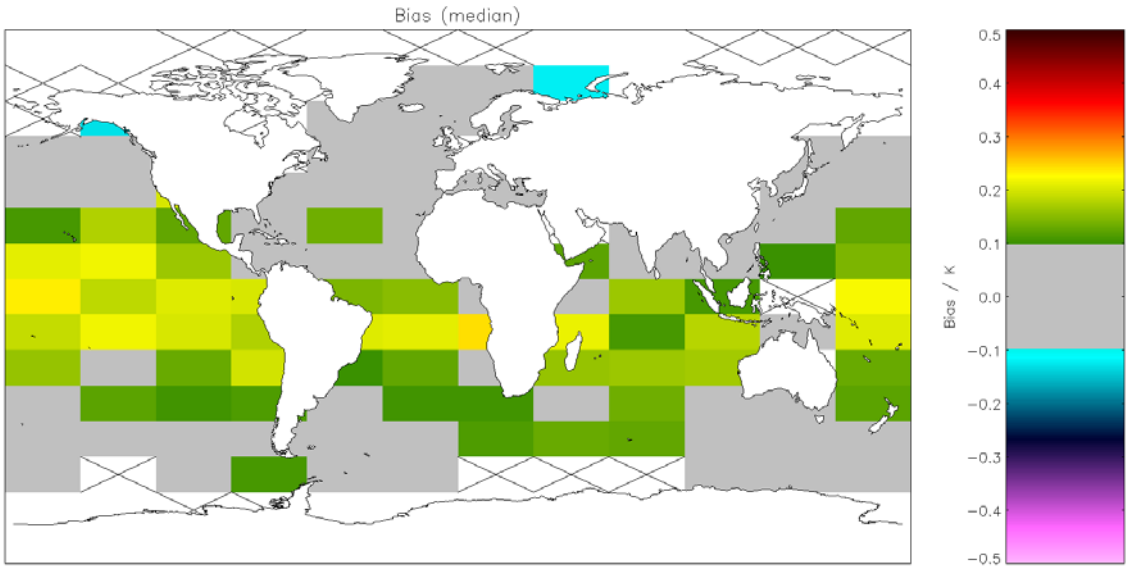


Figure A1.5.1-4: Global map of median bias (satellite-buoy) using OE v2 retrieval for daytime AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.101 K (**Table 5.2.2**).

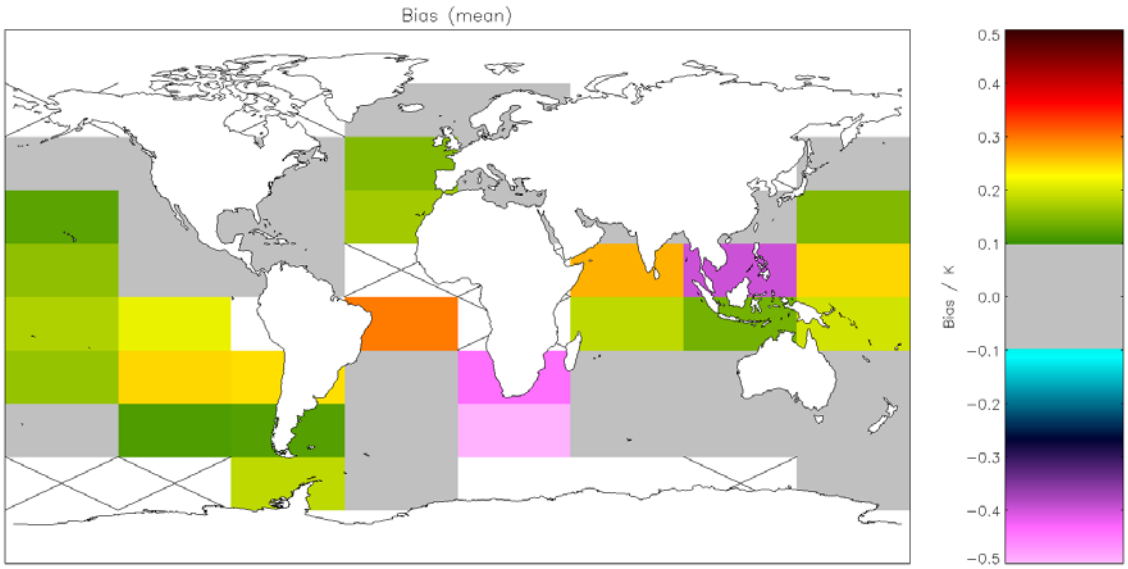


Figure A1.5.1-5: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global mean bias = 0.081 K (**Table 5.3.1**).

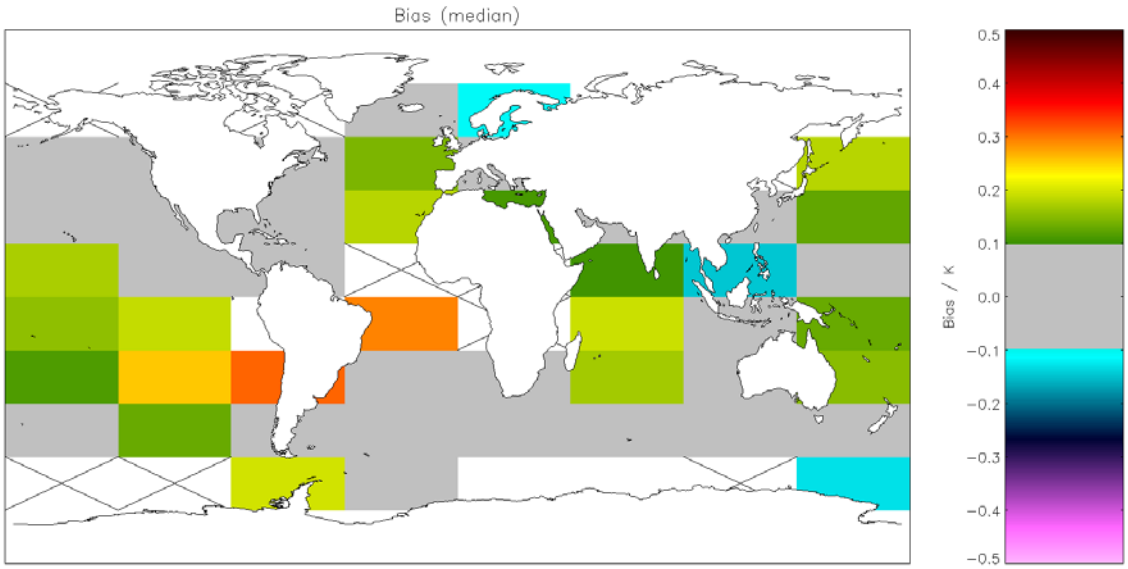


Figure A1.5.1-6: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global median bias = 0.113 K (**Table 5.3.1**).

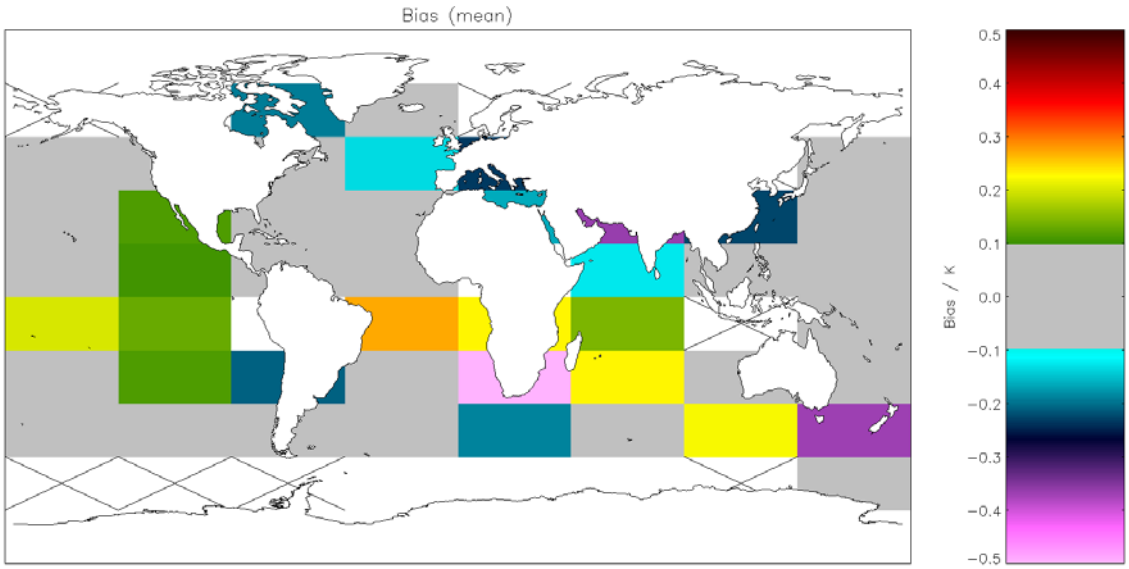


Figure A1.5.1-7: Global map of mean bias (satellite-buoy) using OE v2 retrieval for daytime ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global mean bias = -0.004 K (**Table 5.3.2**).

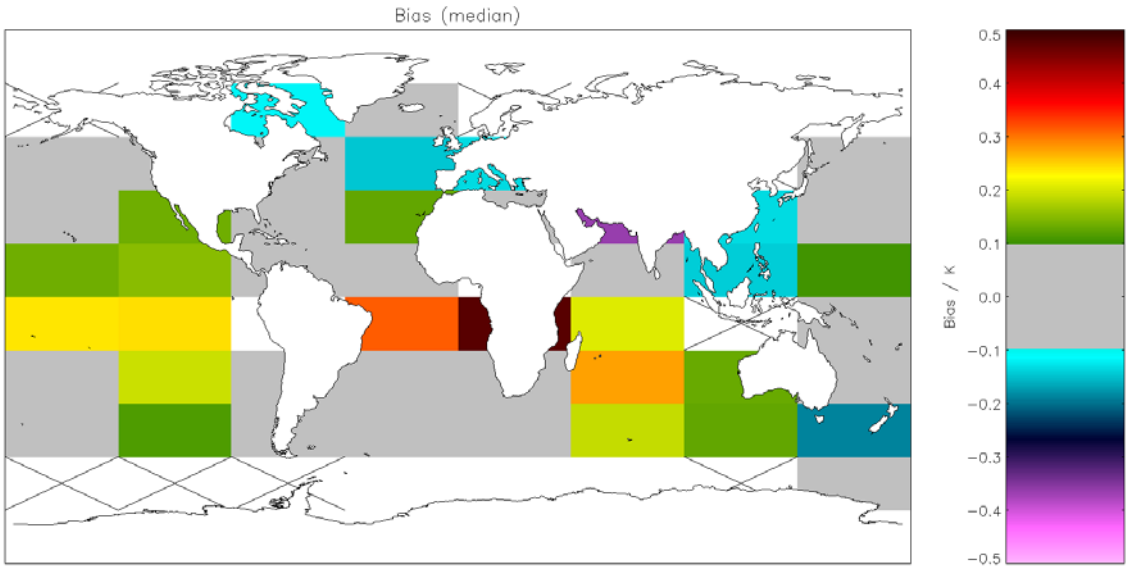


Figure A1.5.1-8: Global map of median bias (satellite-buoy) using OE v2 retrieval for daytime ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global median bias = 0.039 K (**Table 5.3.2**).

A1.5.2 Precision

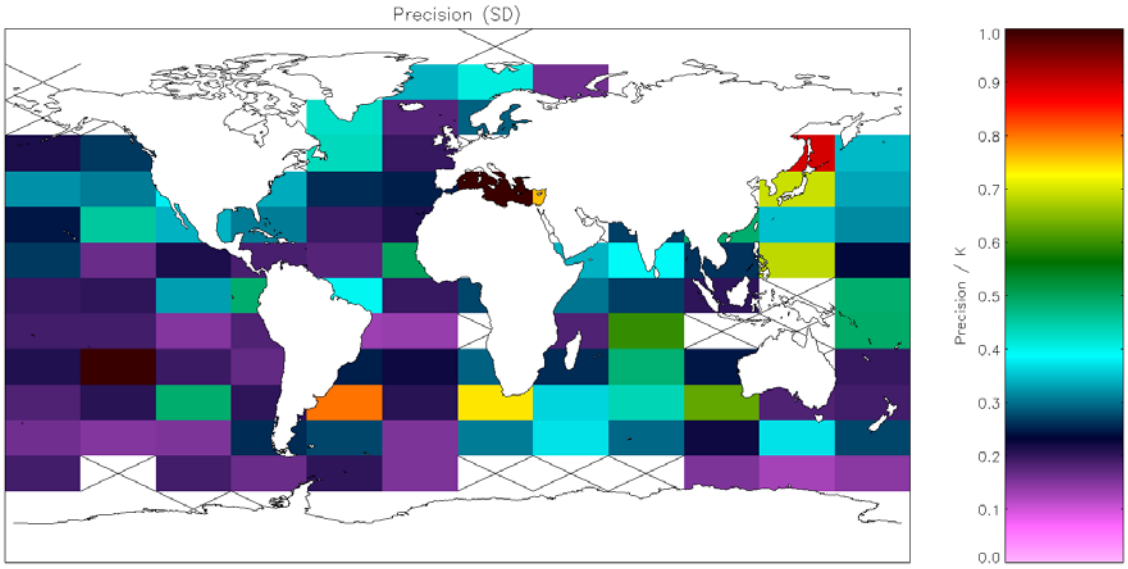


Figure A1.5.2-1: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.392 K (**Table 5.2.1**).

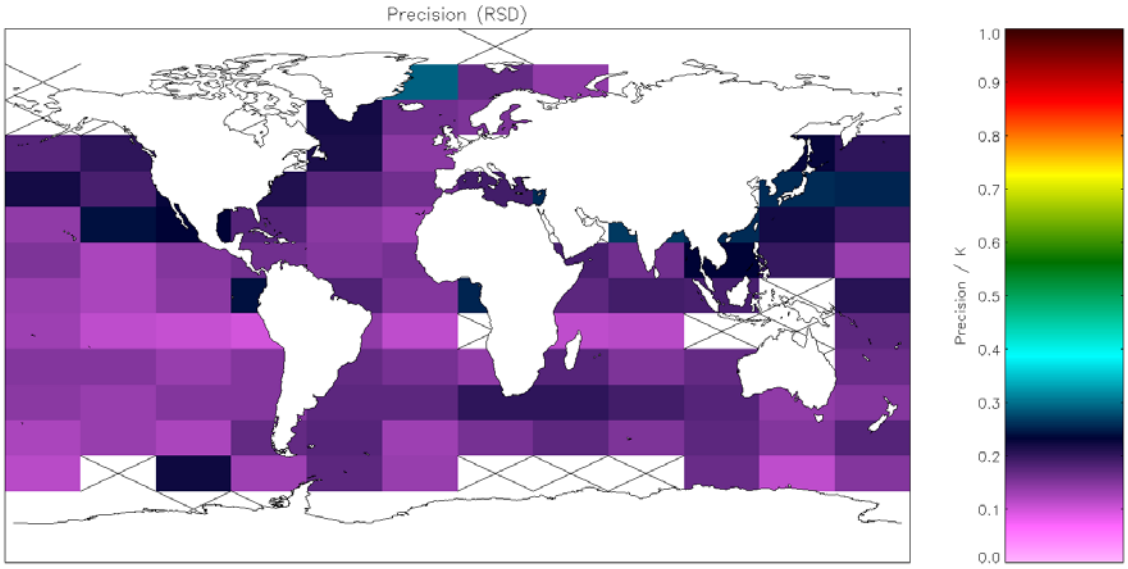


Figure A1.5.2-2: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.171 K (**Table 5.2.1**).

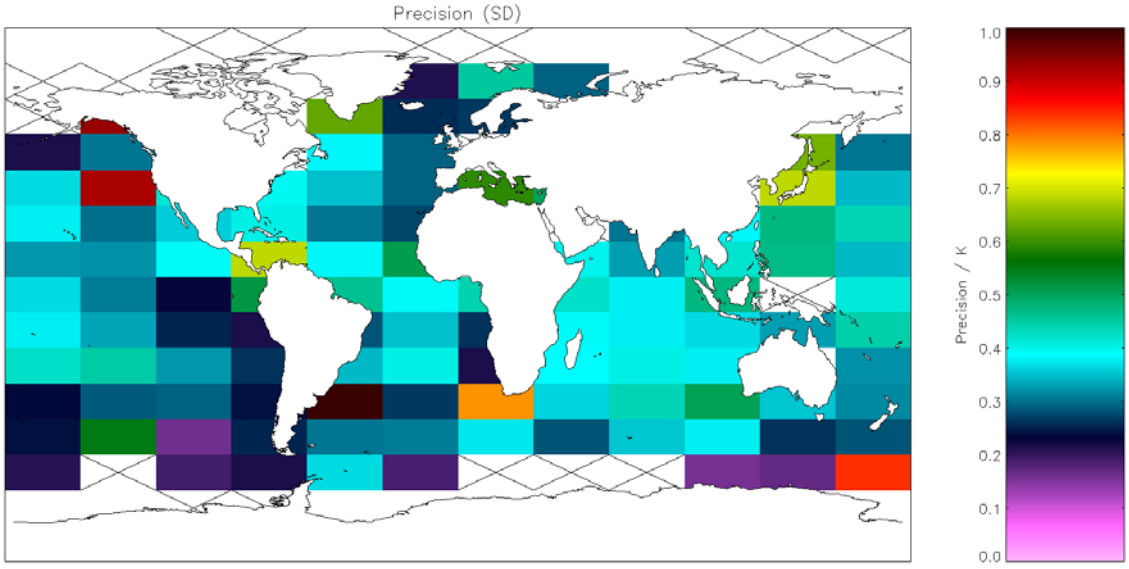


Figure A1.5.2-3: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.434 K (**Table 5.2.2**).

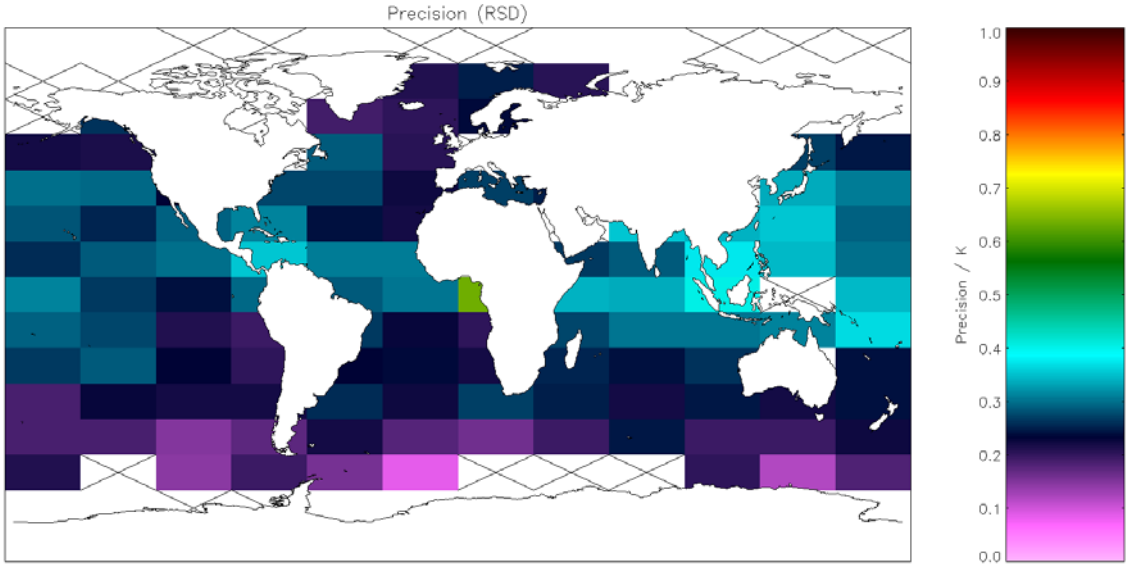


Figure A1.5.2-4: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.264 K (**Table 5.2.2**).

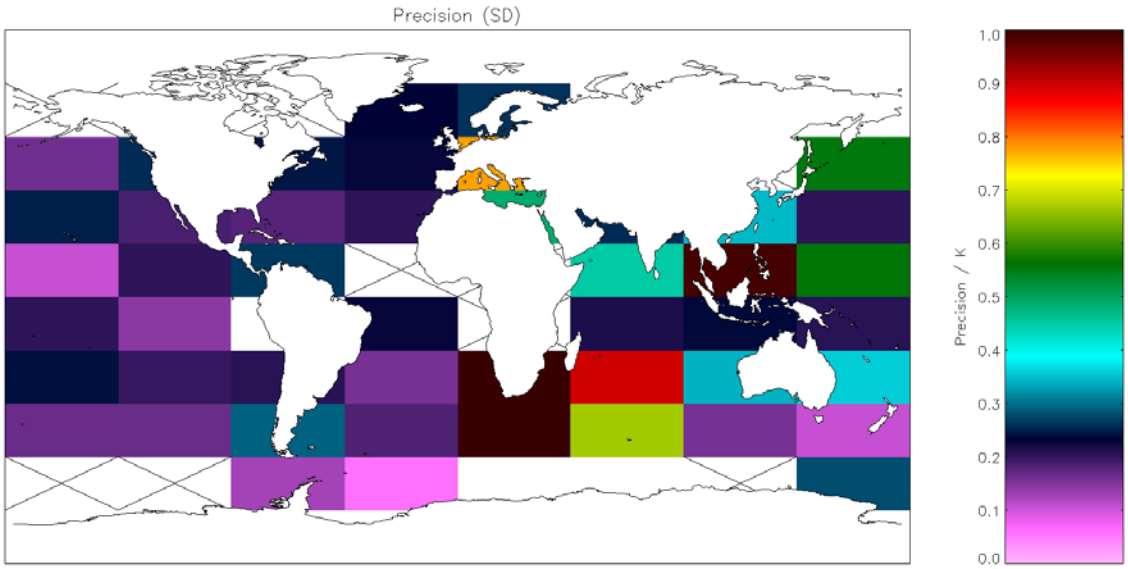


Figure A1.5.2-5: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.444 K (**Table 5.3.1**).

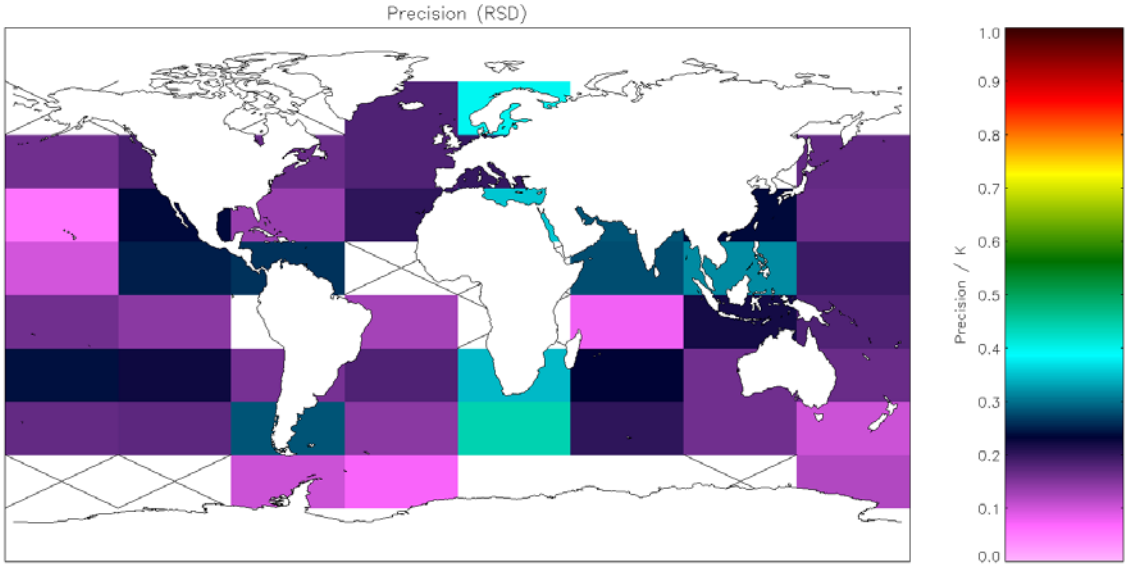


Figure A1.5.2-6: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.207 K (**Table 5.3.1**).

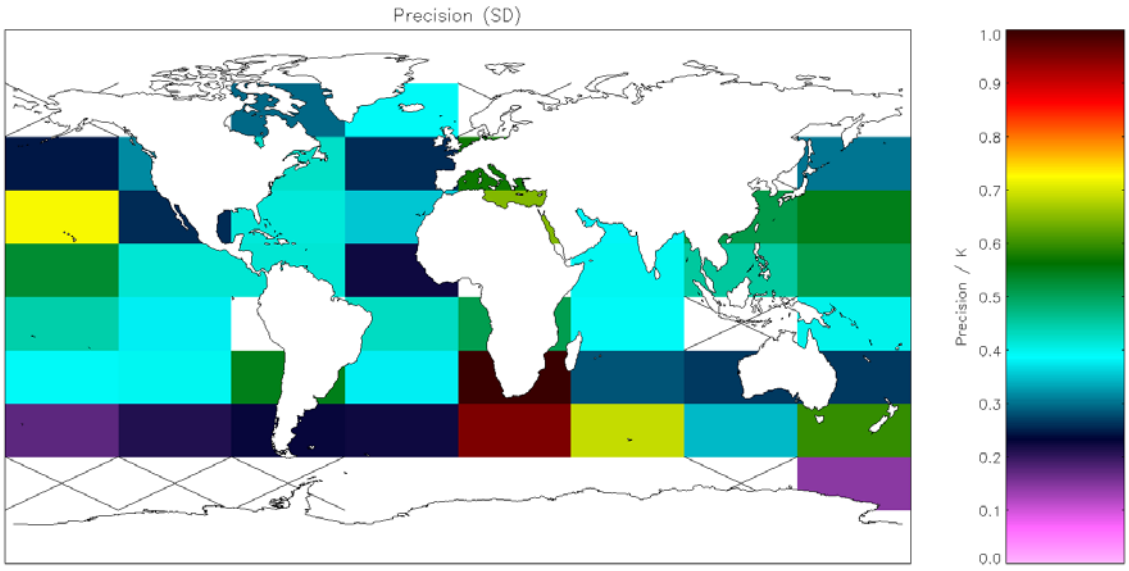


Figure A1.5.2-7: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.486 K (**Table 5.3.2**).

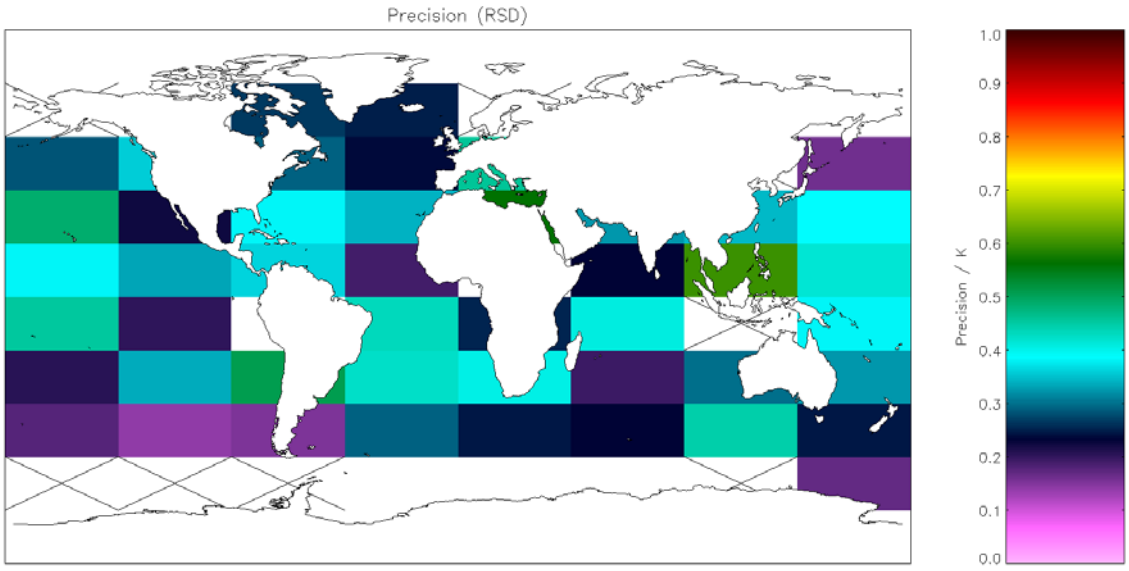


Figure A1.5.2-8: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.345 K (**Table 5.3.2**).

A1.5.3 Sensitivity

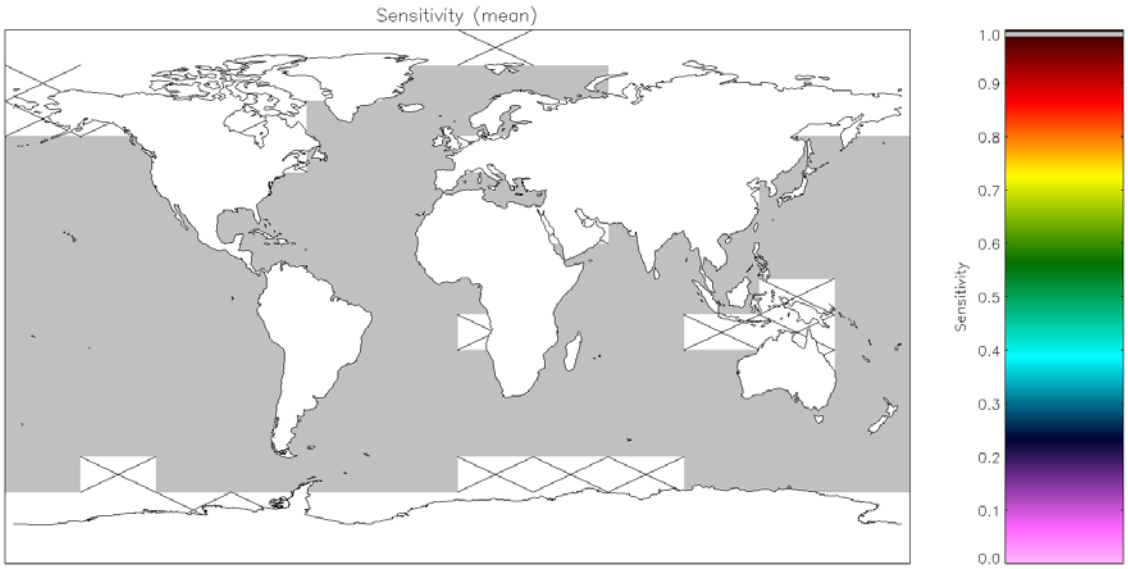


Figure A1.5.3-1: Global map of SST sensitivity using OE v2 retrieval for night-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.2.1**).

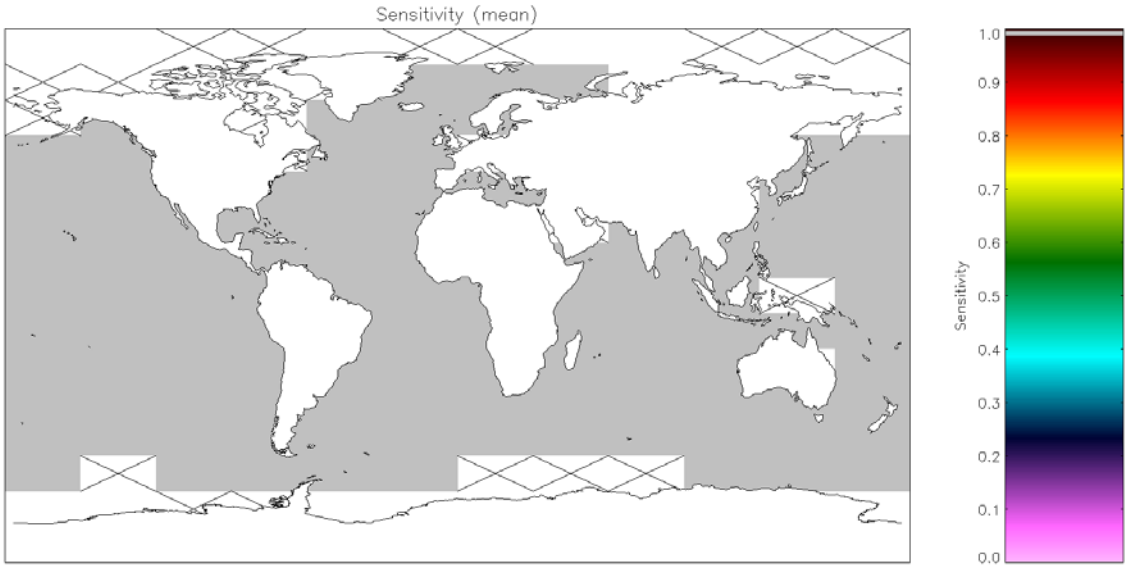


Figure A1.5.3-2: Global map of SST sensitivity using OE v2 retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.2.2**).

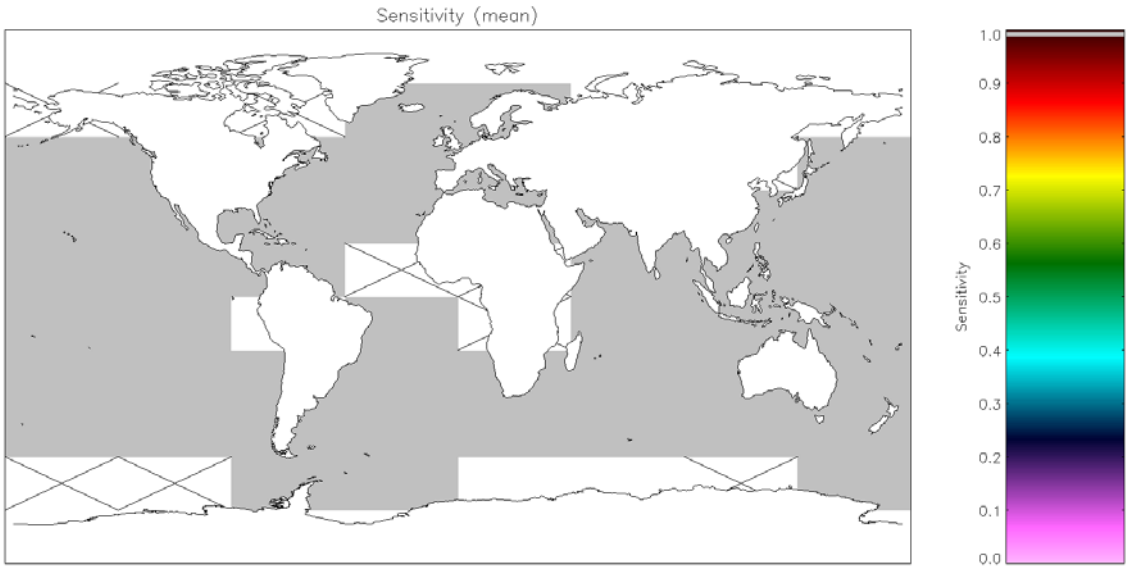


Figure A1.5.3-3: Global map of SST sensitivity using OE v2 retrieval for night-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell (**Table 5.3.1**).

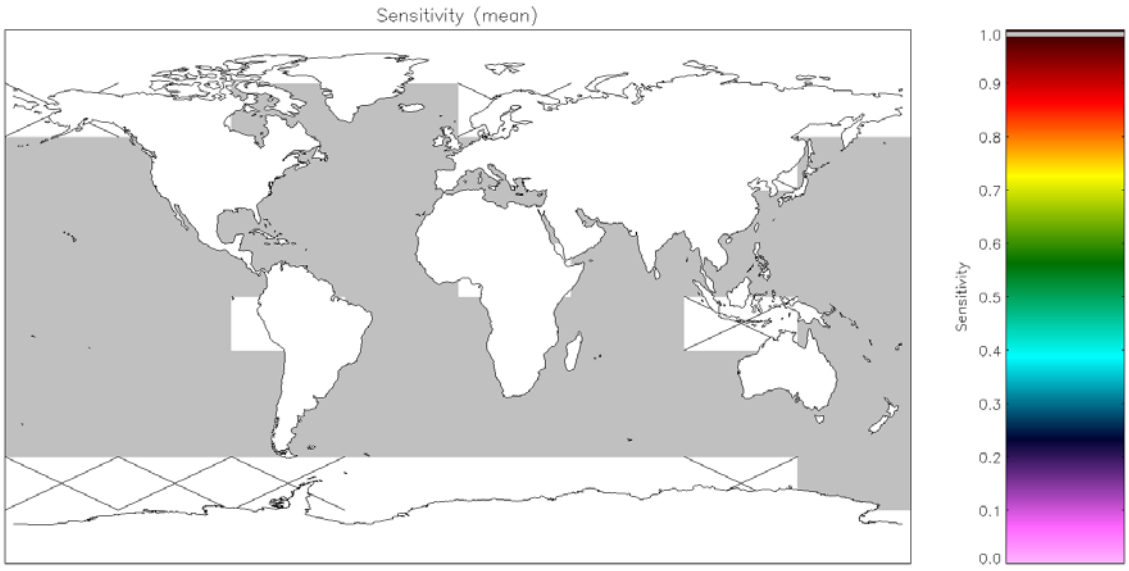


Figure A1.5.3-4: Global map of SST sensitivity using OE v2 retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell (**Table 5.3.2**).

A1.6 Metrics for Oxford-RAL SST Estimation

A1.6.1 Bias

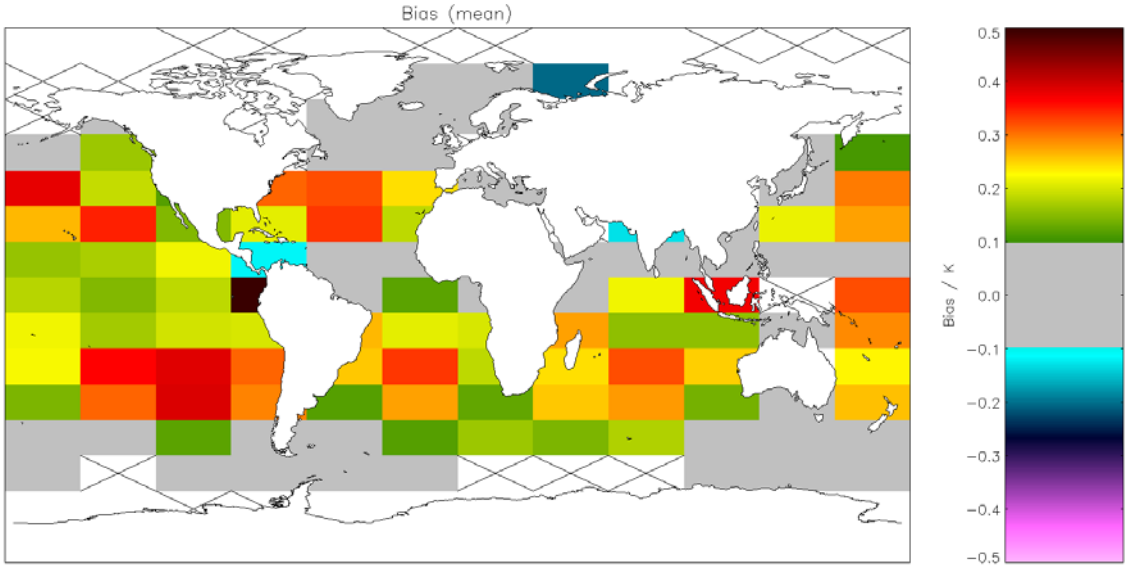


Figure A1.6.1-1: Global map of mean bias (satellite-buoy) using Oxford-RAL retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.206 K (**Table 5.2.2**).

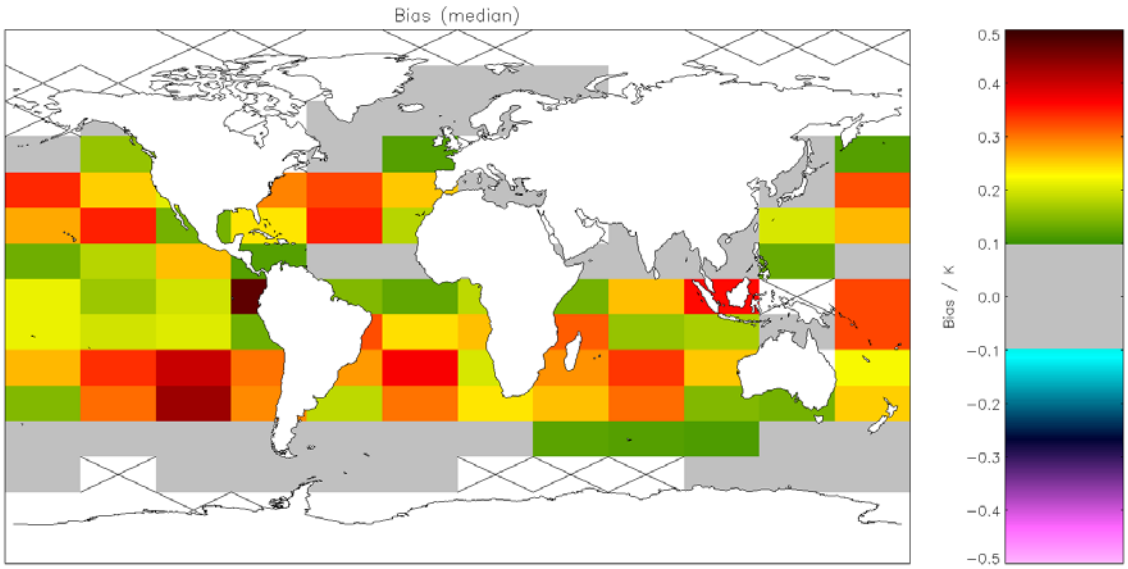


Figure A1.6.1-2: Global map of median bias (satellite-buoy) using Oxford-RAL retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.221 K (**Table 5.2.2**).

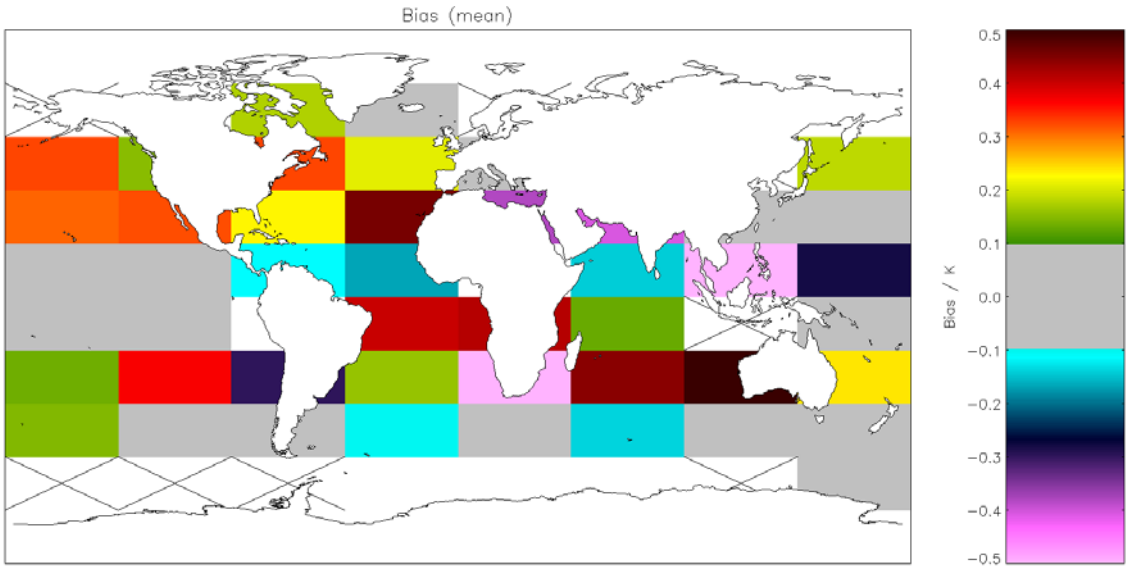


Figure A1.6.1-3: Global map of mean bias (satellite-buoy) using Oxford-RAL retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global mean bias = 0.139 K (**Table 5.3.2**).

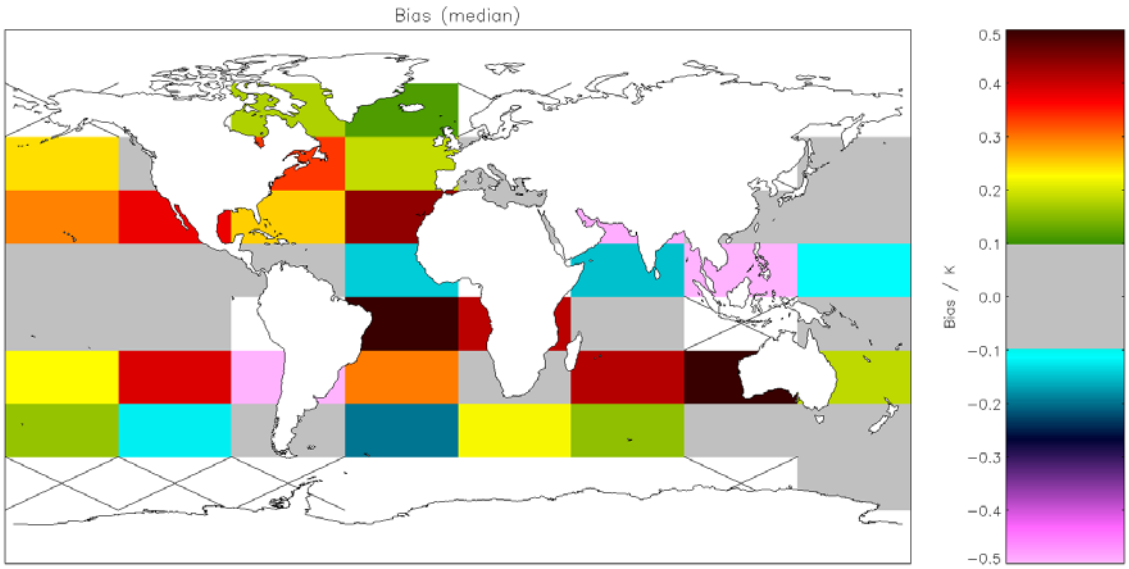


Figure A1.6.1-4: Global map of median bias (satellite-buoy) using Oxford-RAL retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global median bias = 0.174 K (**Table 5.3.2**).

A1.6.2 Precision

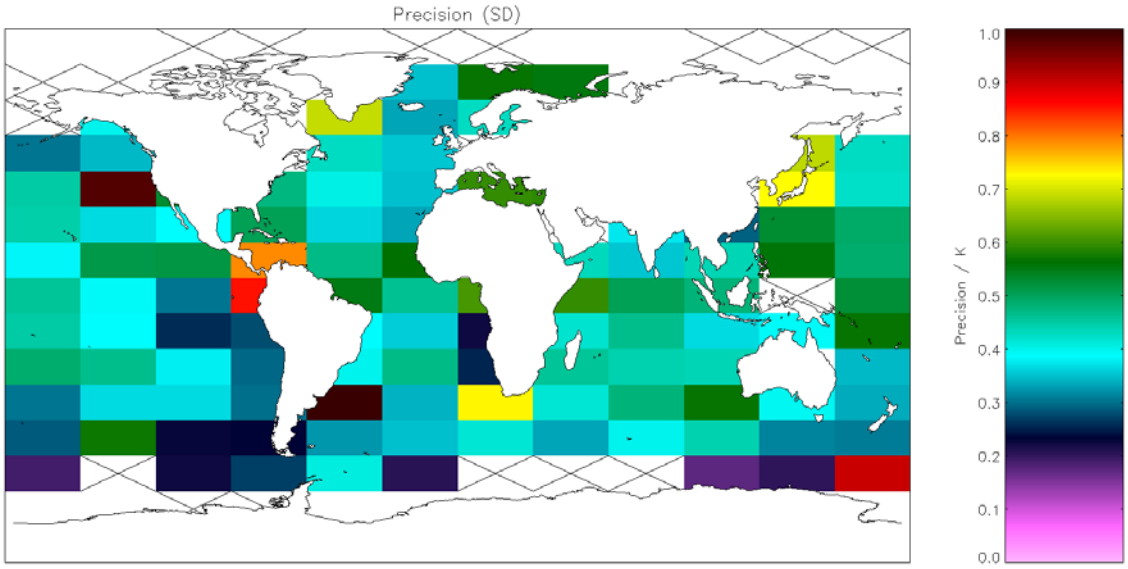


Figure A1.6.2-1: Global map of precision (mean of cell SDs) using Oxford-RAL retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.497 K (**Table 5.2.2**).

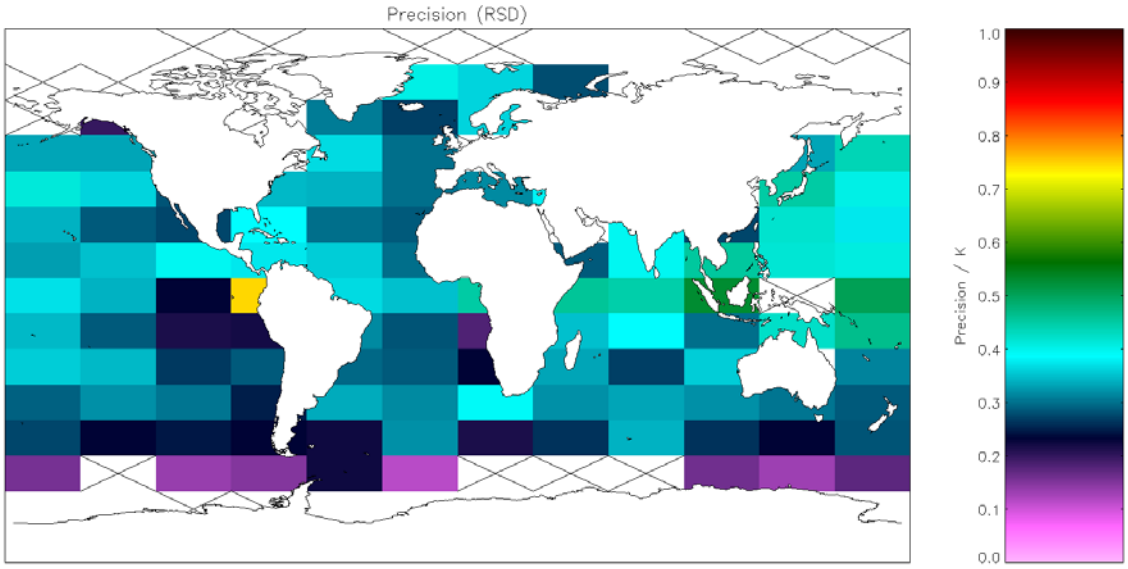


Figure A1.6.2-2: Global map of precision (median of cell RSDs) using Oxford-RAL retrieval for day-time AATSR observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.349 K (**Table 5.2.2**).

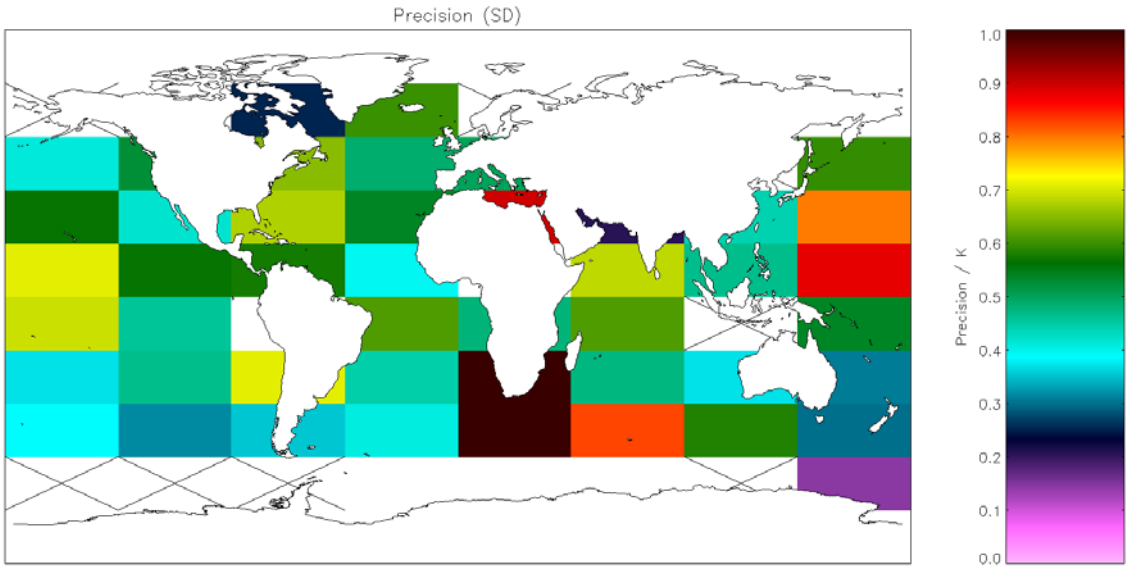


Figure A1.6.2-3: Global map of precision (mean of cell SDs) using Oxford-RAL retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global SD = 0.139 K (**Table 5.3.2**).

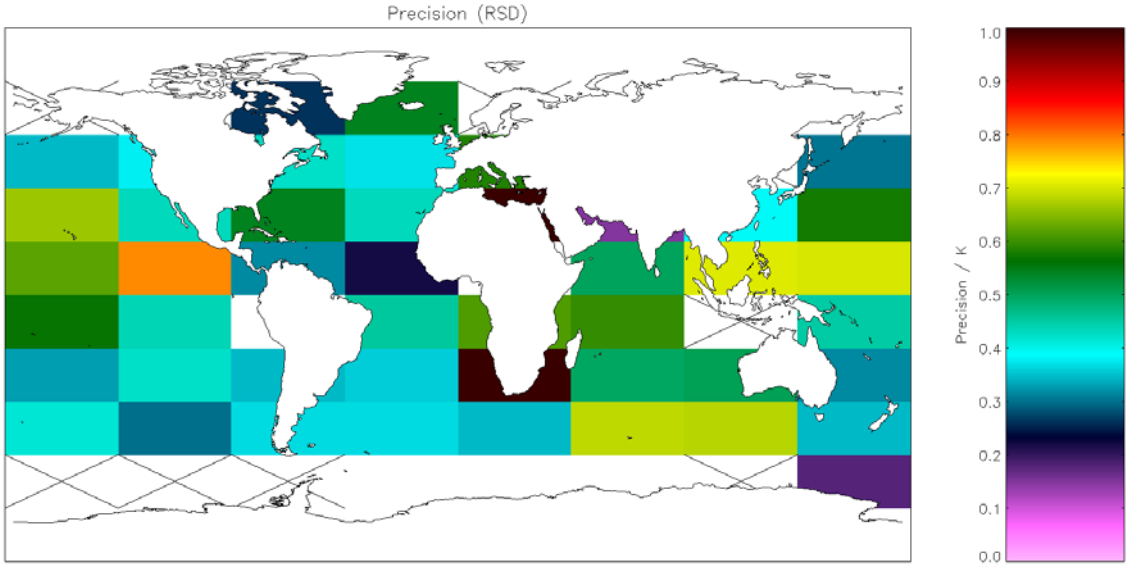


Figure A1.6.2-4: Global map of precision (median of cell RSDs) using Oxford-RAL retrieval for day-time ATSR-2 observations. Crosses indicate grid cells excluded due to discrepancy of 0.15 K being insignificant for the cell. Global RSD = 0.174 K (**Table 5.3.2**).

A2. Advanced Very High Resolution Radiometer results

A2.1 Brief account of relevant instrument characteristics

The algorithms assessed (A2.2) in the Algorithm Selection process for AVHRRs onboard NOAA-17, NOAA-18, NOAA-19 and Metop utilise a different selection of the available channels for day and night-time retrievals. During the night, all three infrared channels (3.7 µm, 11 µm and 12 µm) are used. The 3.7 µm channel is not used during the day due

to solar contamination effects, so day-time retrievals are performed using only the 11 µm and 12 µm channels.

AVHRR GAC data are derived from a sample averaging of the full resolution (1.1 km^2 at swath centre) AVHRR data, yielding a spatial resolution 4 km^2, covering a swath width of 2600 km at the equator.

NOAA-17 and Metop orbit with equator crossing times of ~10:00 (descending) while NOAA-18 and NOAA-19 orbit with equator crossing times of ~14:00 (ascending).

A2.2 List of algorithms

The following algorithms have been assessed for AVHRR in the Algorithm Selection process:

- CCI SST Optimal Estimation v1 (A2.3)
- CCI SST Optimal Estimation v2 (A2.4)
- Incremental Regression Estimation (A2.5)

All algorithms listed above have been documented in the SST CCI "Algorithm Theoretical Basis Document v0" [RD.225].

A2.3 Metrics for CCI SST Optimal Estimation v1

A2.3.1 Bias

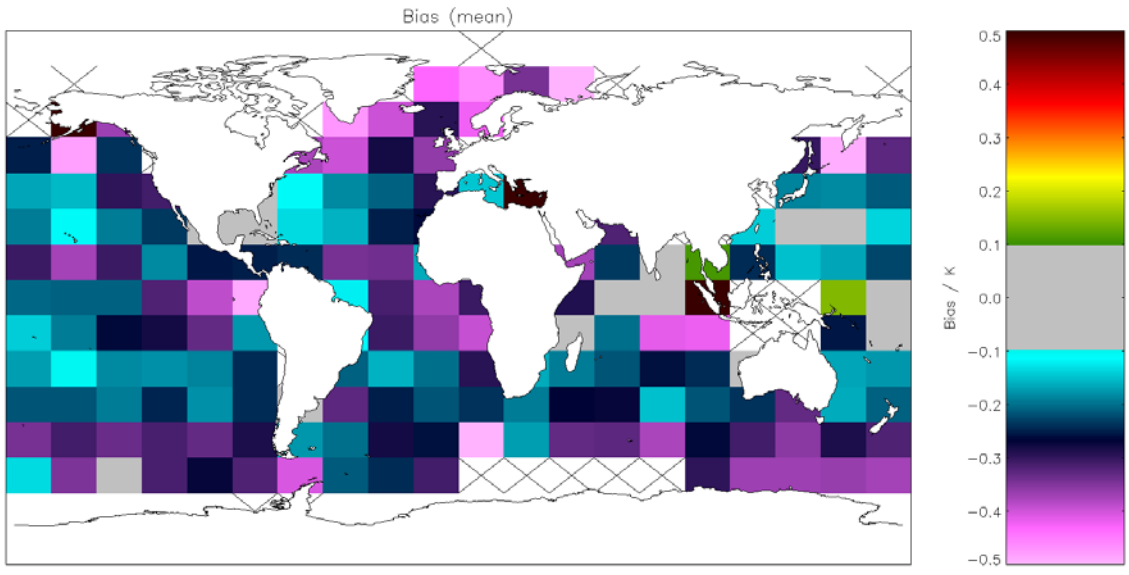


Figure A2.3.1-1: Global map of mean bias (satellite-buoy) using OE v1 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.213 K (**Table 5.4.2**).

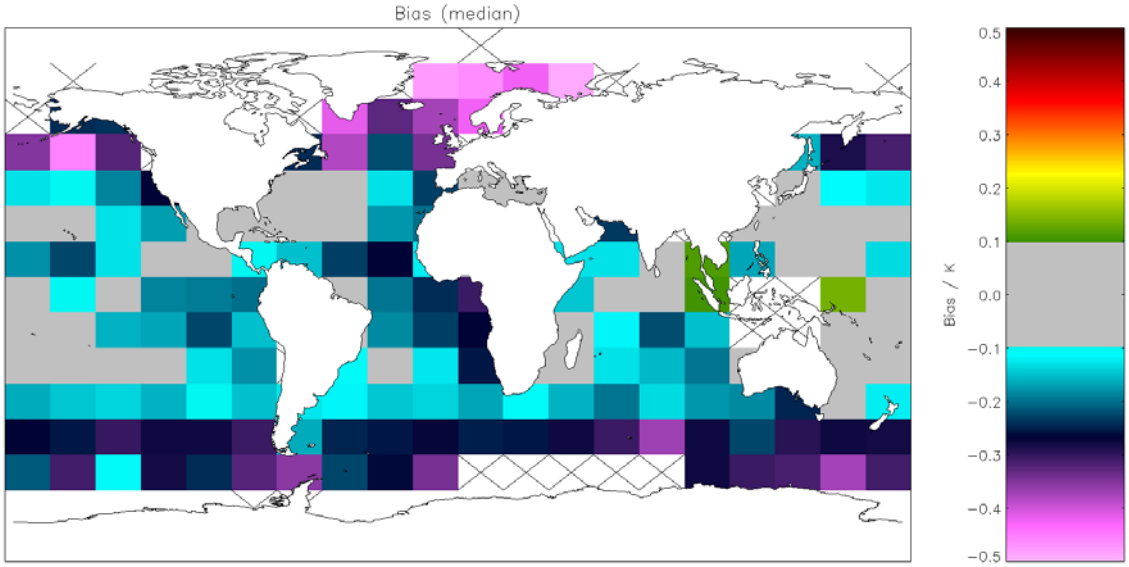


Figure A2.3.1-2: Global map of median bias (satellite-buoy) using OE v1 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.139 K (**Table 5.4.2**).

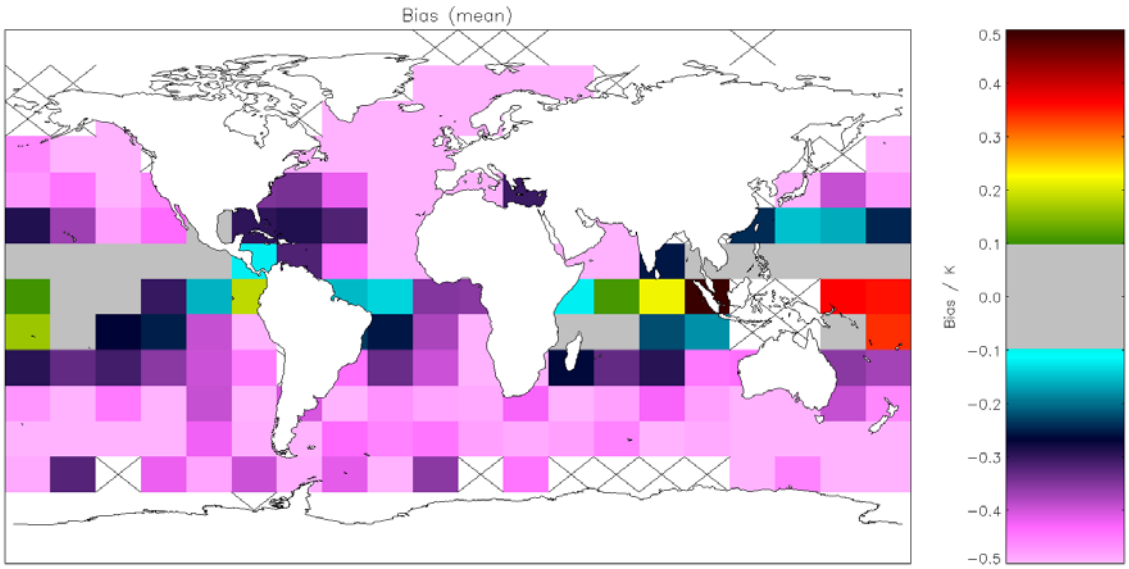


Figure A2.3.1-3: Global map of mean bias (satellite-buoy) using OE v1 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.368 K (**Table 5.4.3**).

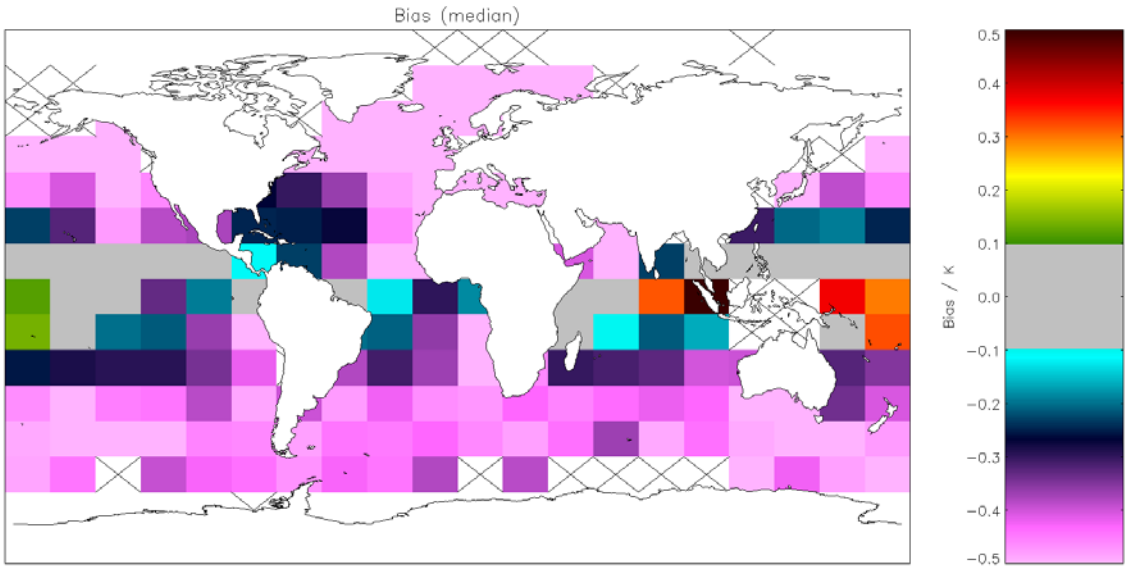


Figure A2.3.1-4: Global map of median bias (satellite-buoy) using OE v1 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.375 K (**Table 5.4.3**).

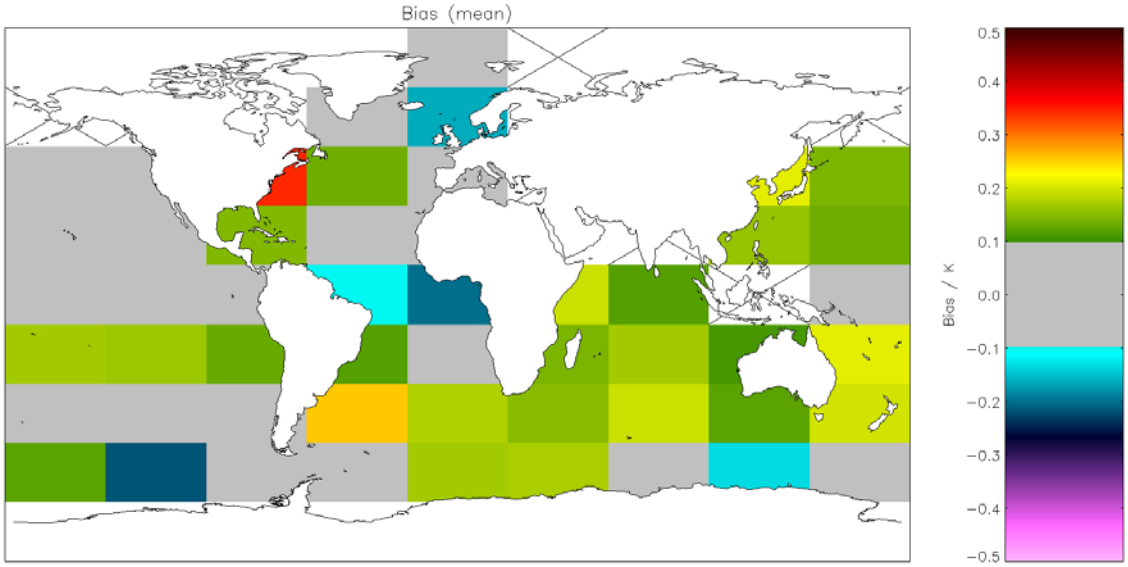


Figure A2.3.1-5: Global map of mean bias (satellite-buoy) using OE v1 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.109 K (**Table 5.5.1**).

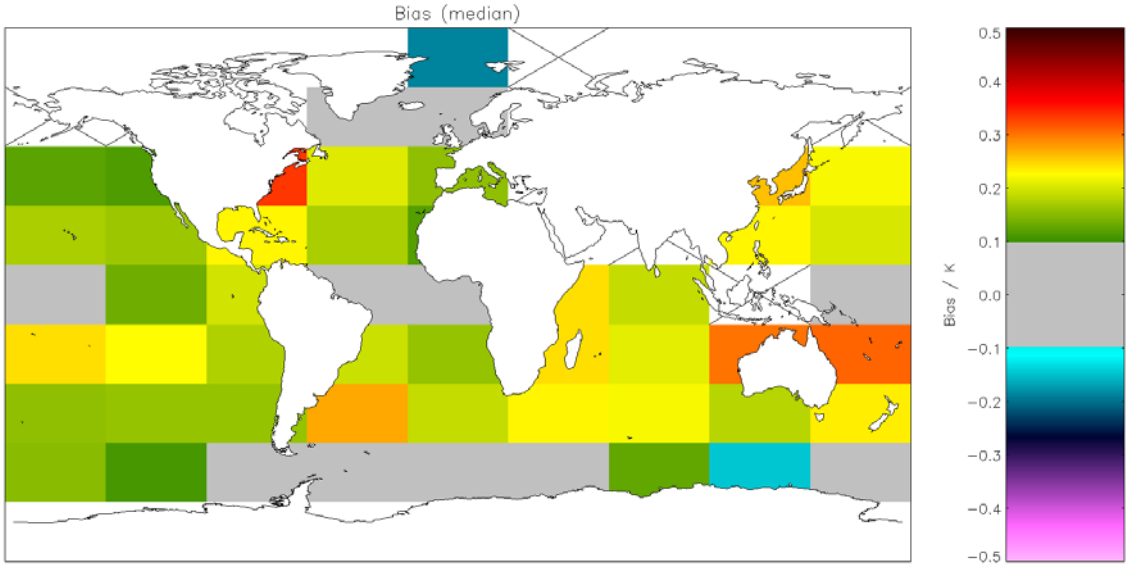


Figure A2.3.1-6: Global map of median bias (satellite-buoy) using OE v1 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.184 K (**Table 5.5.1**).

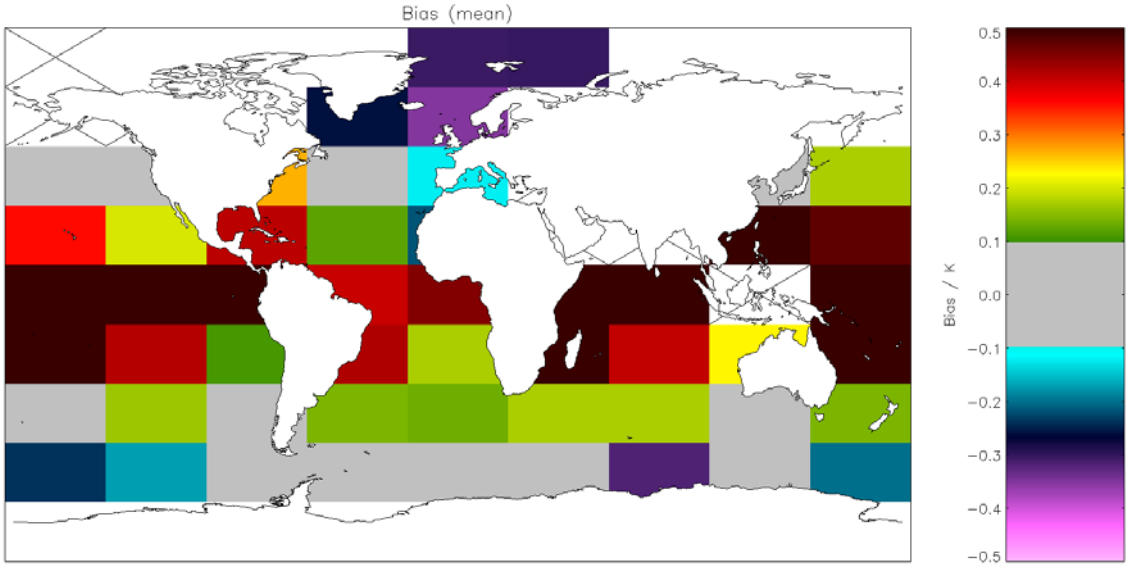


Figure A2.3.1-7: Global map of mean bias (satellite-buoy) using OE v1 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.297 K (**Table 5.5.2**).

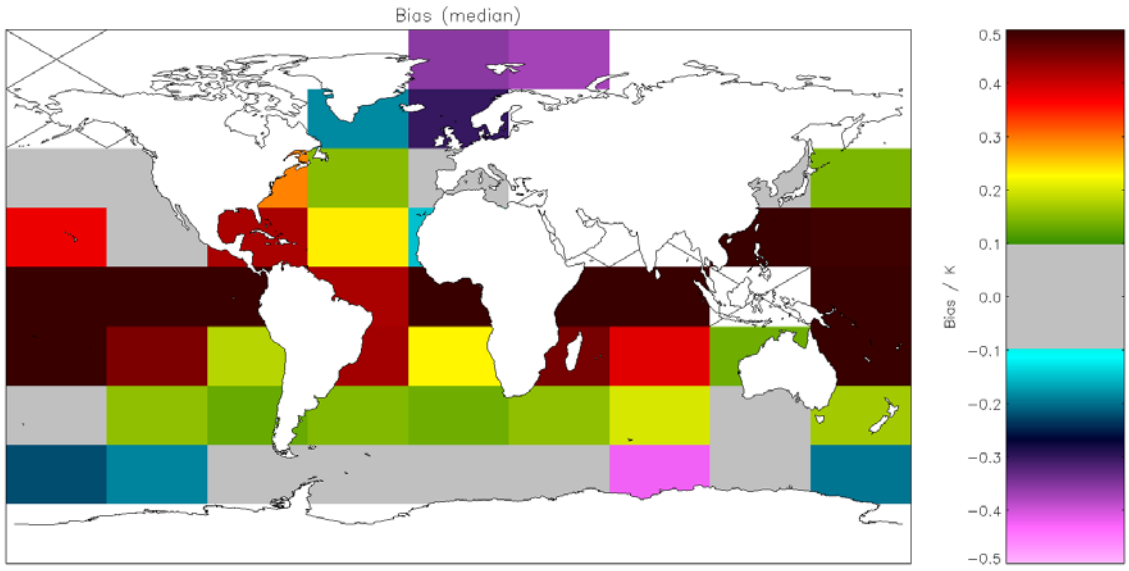


Figure A2.3.1-8: Global map of median bias (satellite-buoy) using OE v1 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.275 K (**Table 5.5.2**).

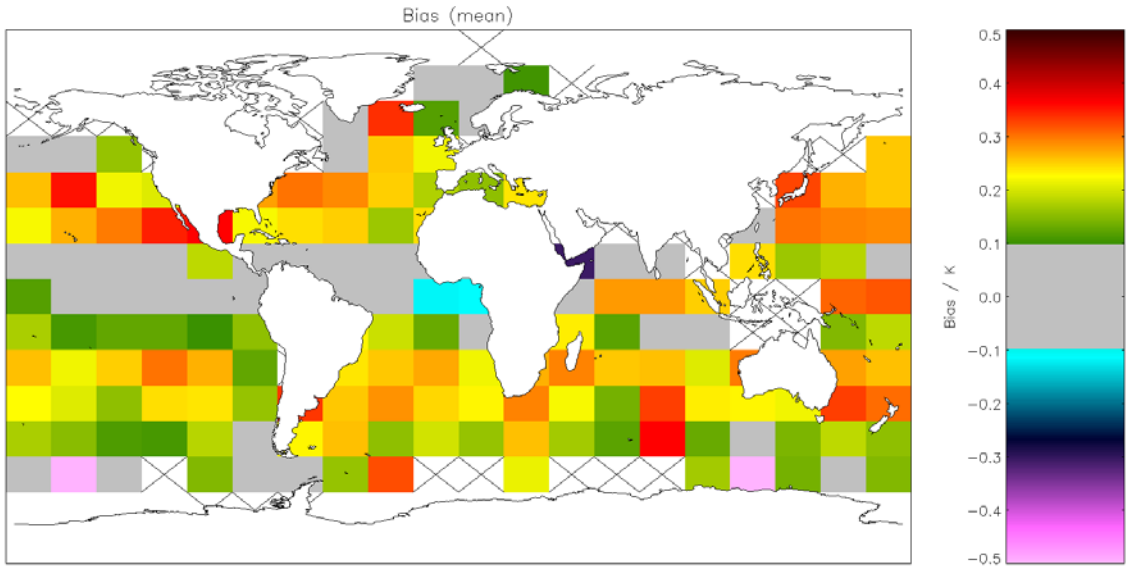


Figure A2.3.1-9: Global map of mean bias (satellite-buoy) using OE v1 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.206 K (**Table 5.6.1**).

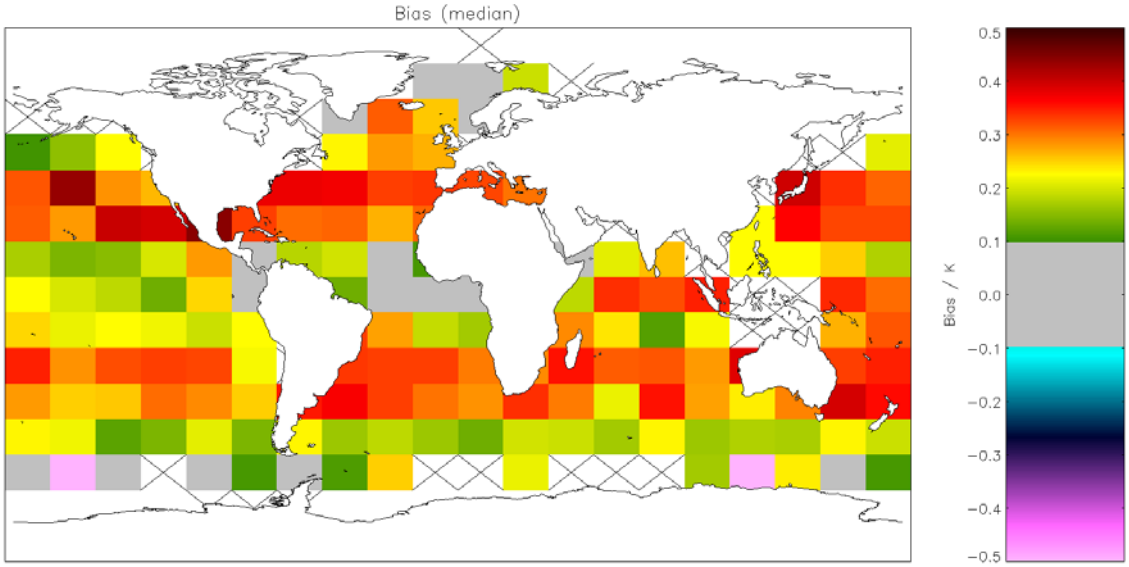


Figure A2.3.1-10: Global map of median bias (satellite-buoy) using OE v1 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.282 K (**Table 5.6.1**).

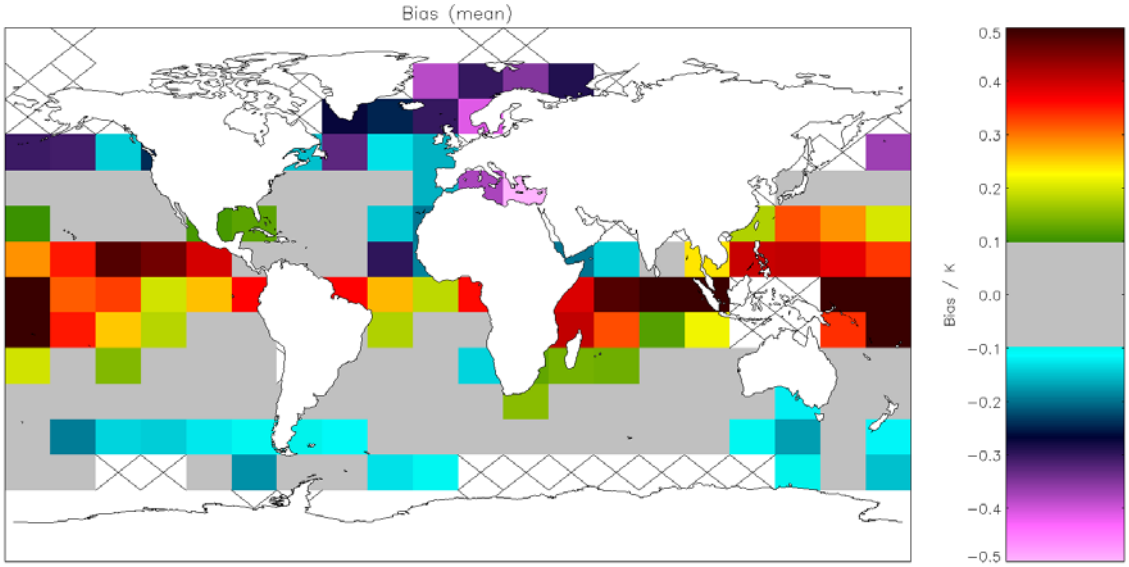


Figure A2.3.1-11: Global map of mean bias (satellite-buoy) using OE v1 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.059 K (**Table 5.6.2**).

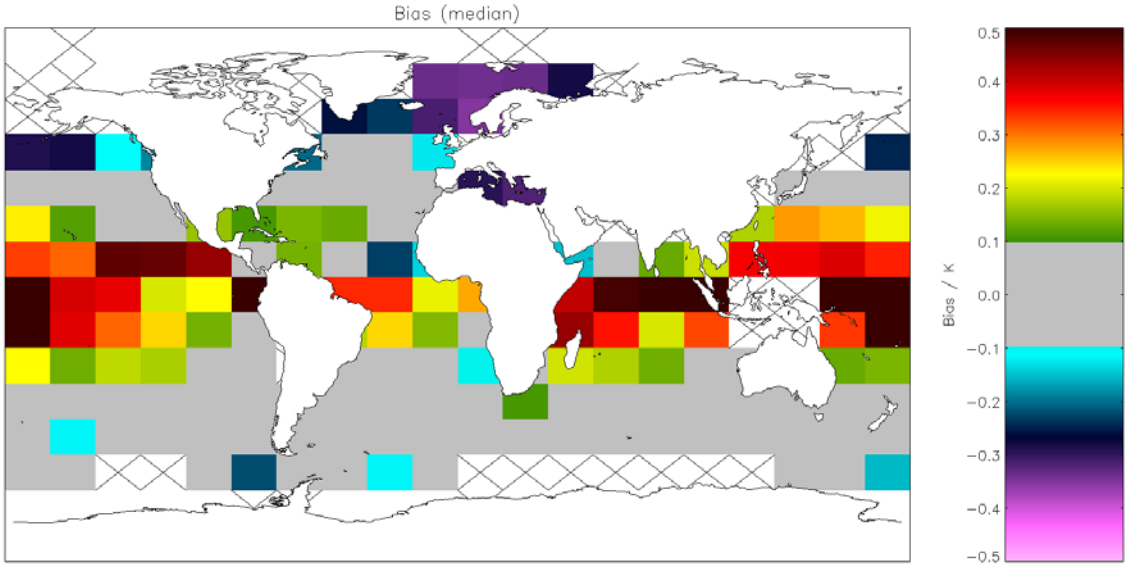


Figure A2.3.1-12: Global map of median bias (satellite-buoy) using OE v1 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.077 K (**Table 5.6.2**).

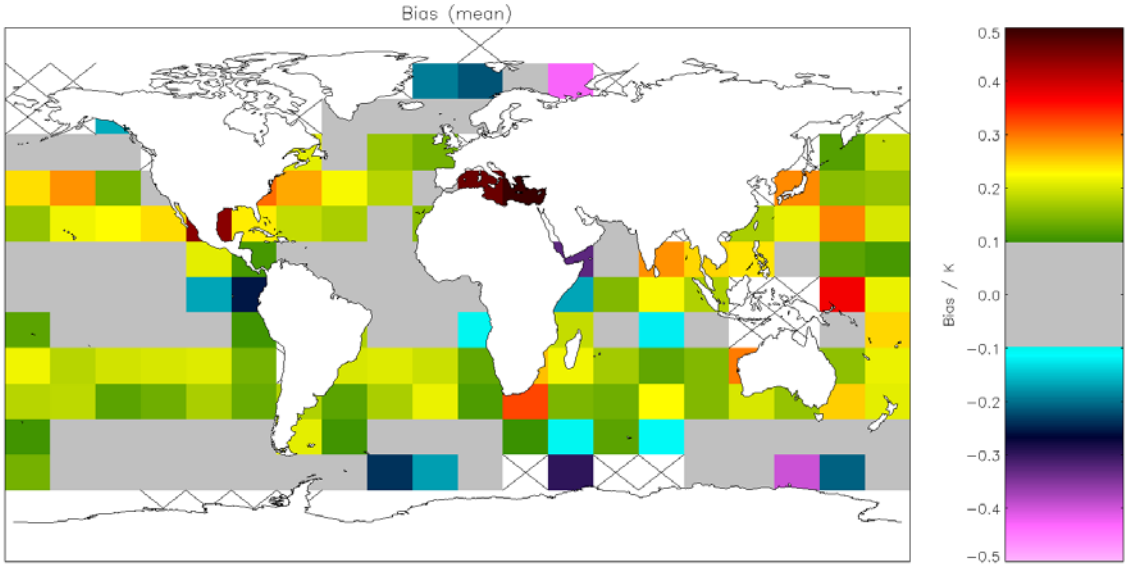


Figure A2.3.1-13: Global map of mean bias (satellite-buoy) using OE v1 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.147 K (**Table 5.7.1**).

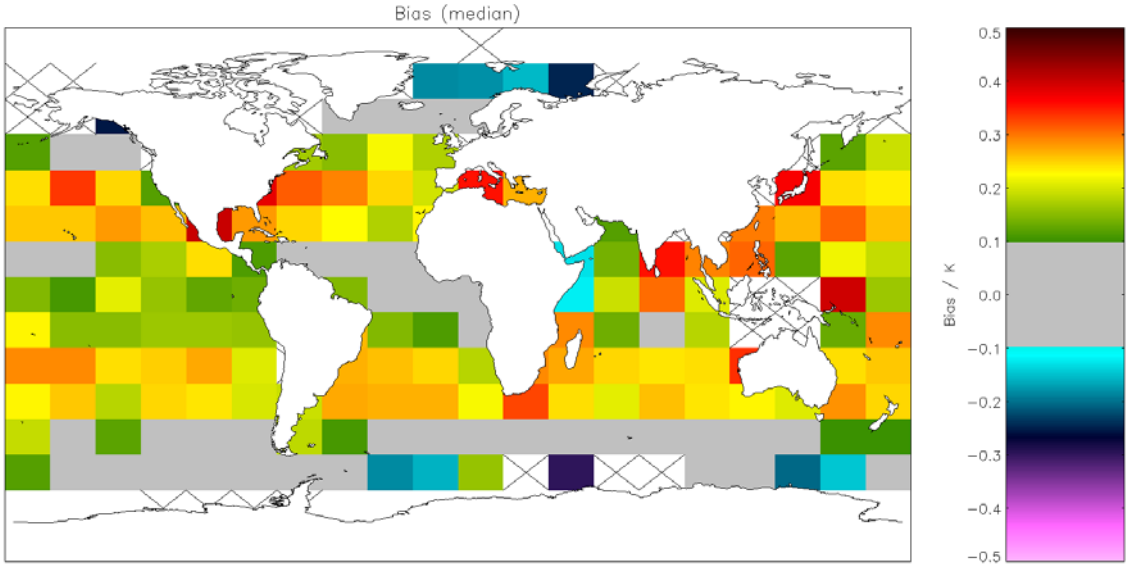


Figure A2.3.1-14: Global map of median bias (satellite-buoy) using OE v1 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.217 K (**Table 5.7.1**).

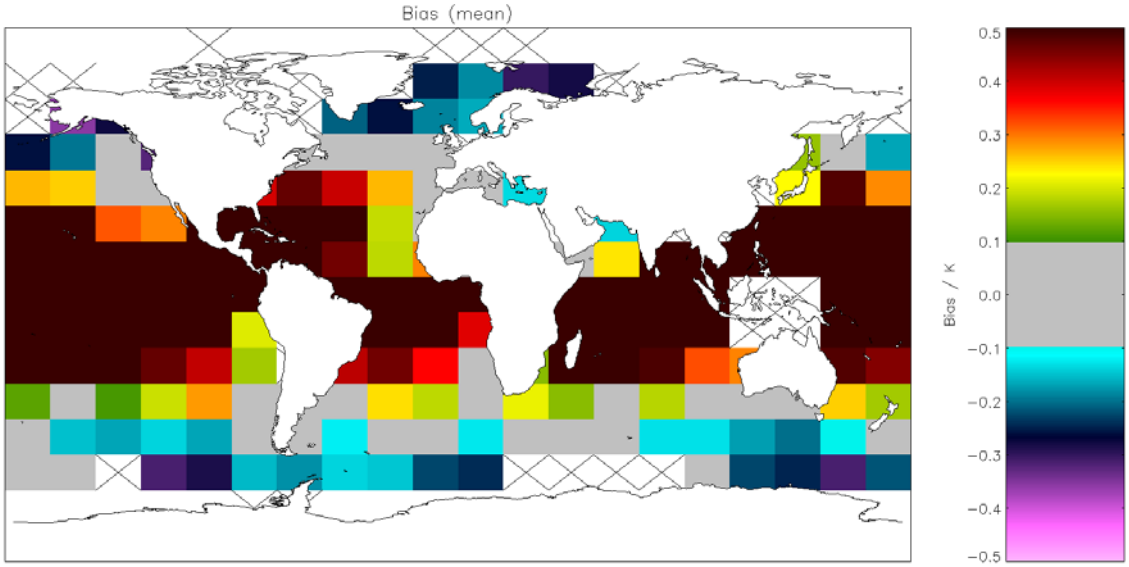


Figure A2.3.1-15: Global map of mean bias (satellite-buoy) using OE v1 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.425 K (**Table 5.7.2**).

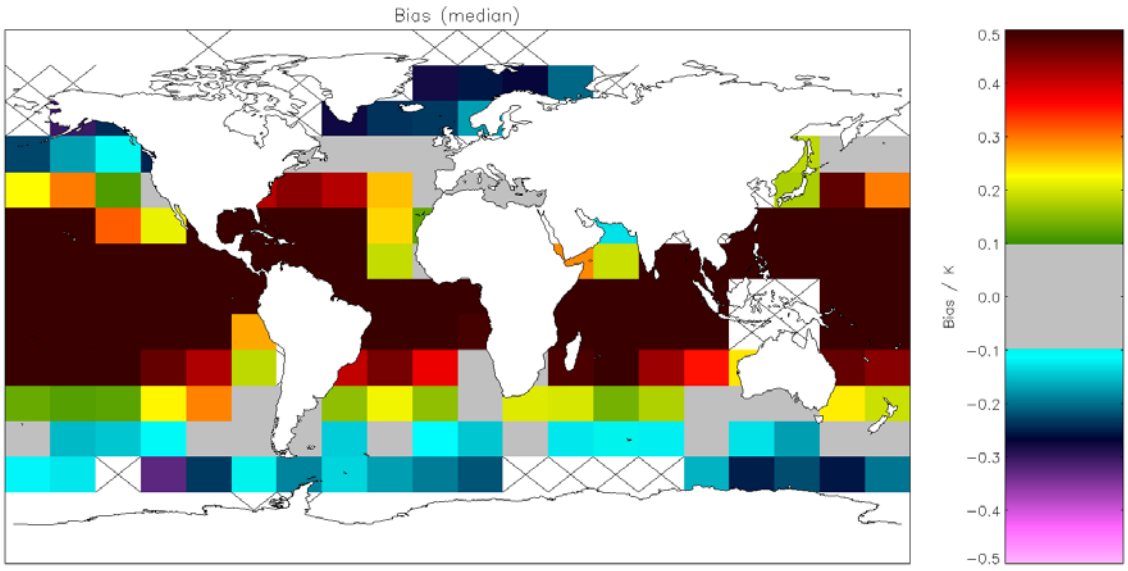


Figure A2.3.1-16: Global map of median bias (satellite-buoy) using OE v1 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.357 K (**Table 5.7.2**).

A2.3.2 Precision

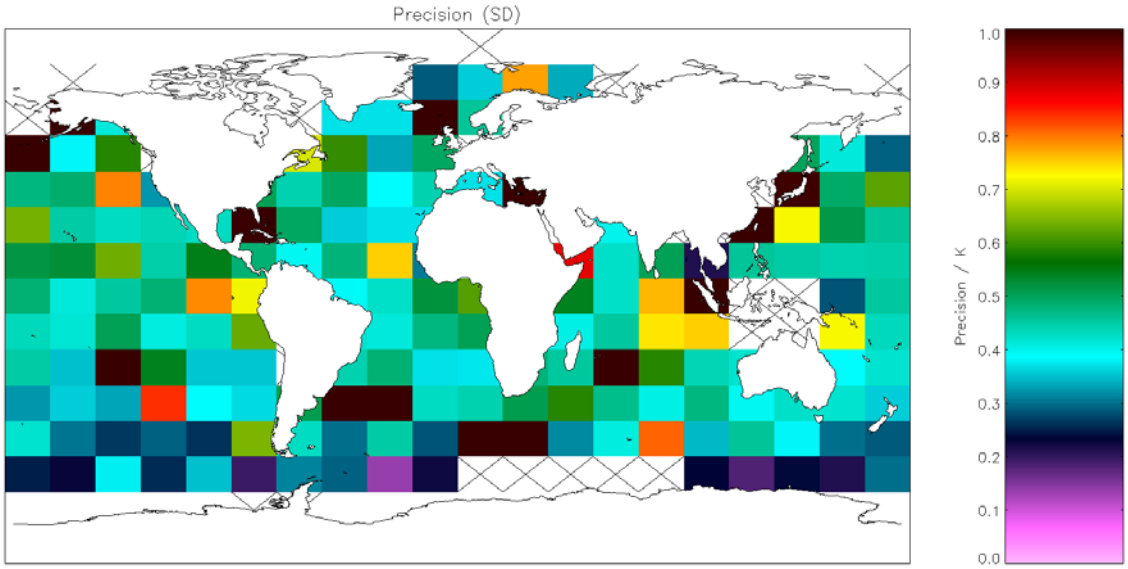


Figure A2.3.2-1: Global map of precision (mean of cell SDs) using OE v1 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.674 K (**Table 5.4.2**).

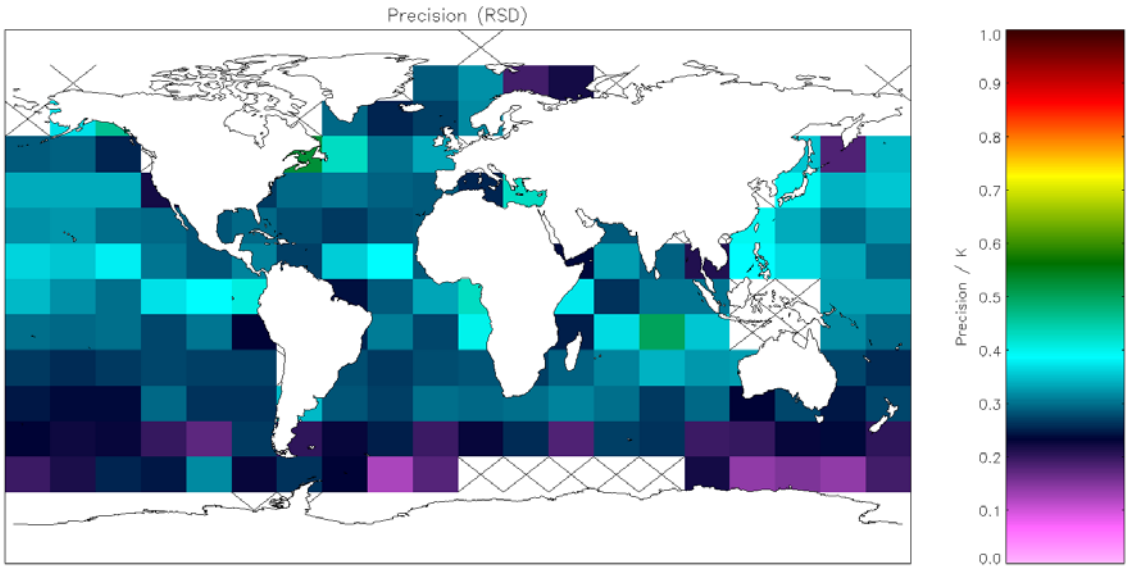


Figure A2.3.2-2: Global map of precision (median of cell RSDs) using OE v1 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.313 K (**Table 5.4.2**).

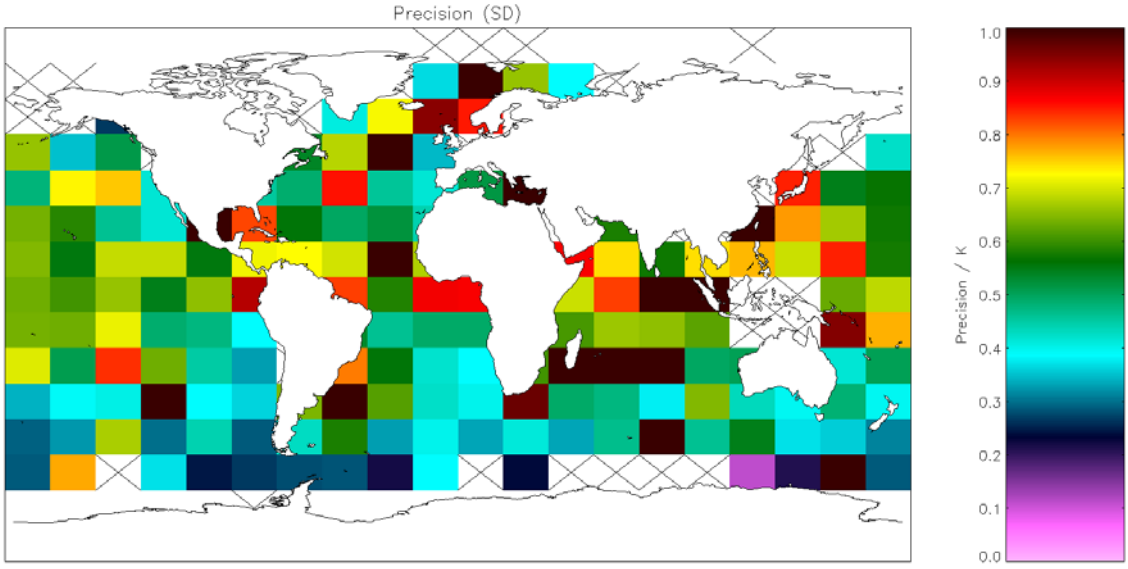


Figure A2.3.2-3: Global map of precision (mean of cell SDs) using OE v1 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.734 K (**Table 5.4.3**).

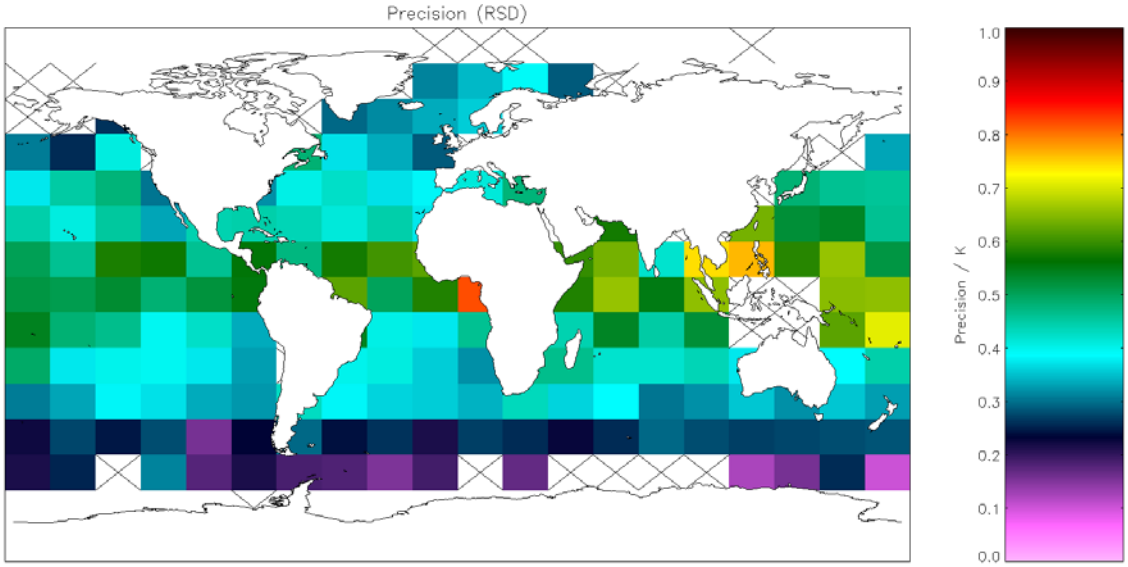


Figure A2.3.2-4: Global map of precision (median of cell RSDs) using OE v1 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.433 K (**Table 5.4.3**).

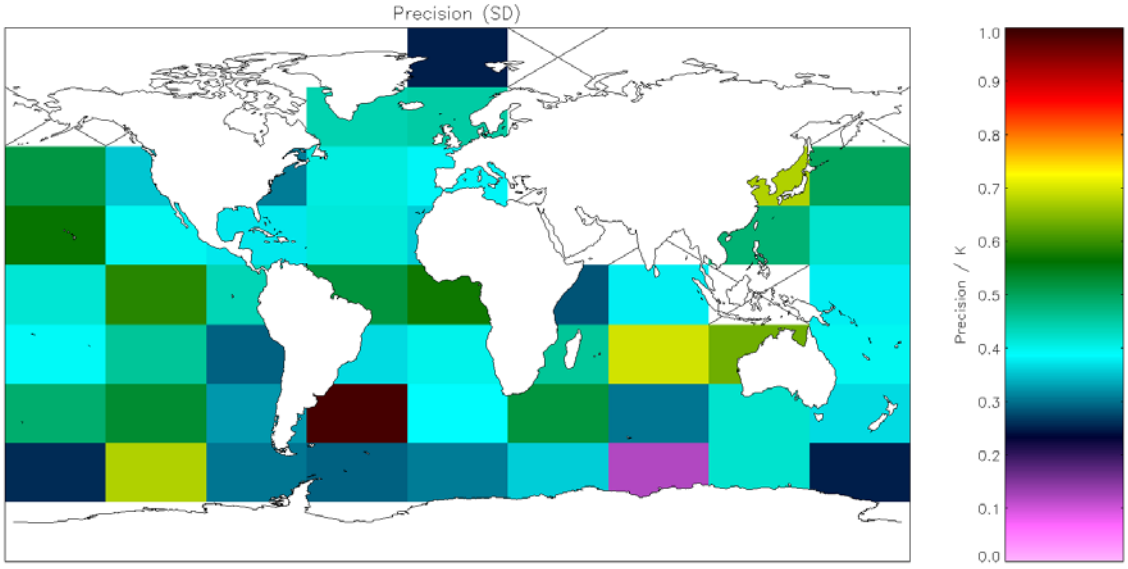


Figure A2.3.2-5: Global map of precision (mean of cell SDs) using OE v1 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.487 K (**Table 5.5.1**).

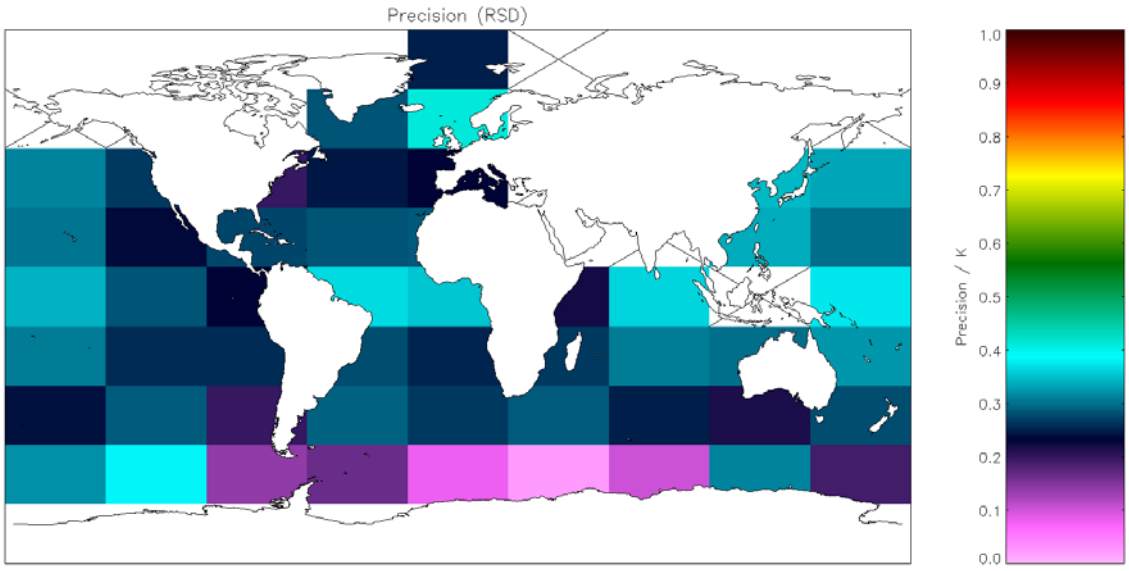


Figure A2.3.2-6: Global map of precision (median of cell RSDs) using OE v1 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.290 K (**Table 5.5.1**).

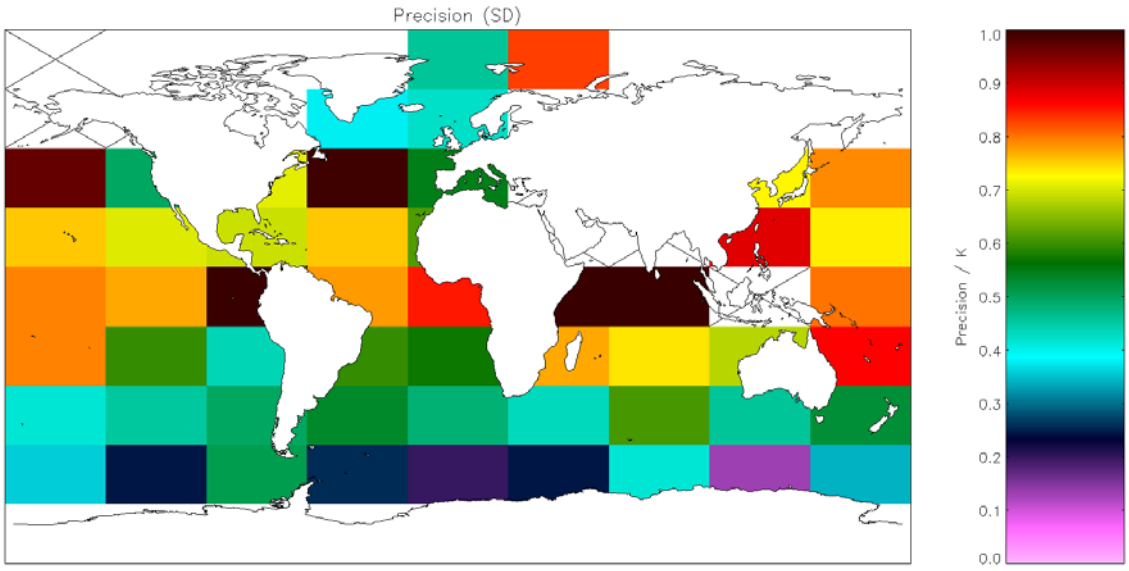


Figure A2.3.2-7: Global map of precision (mean of cell SDs) using OE v1 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.779 K (**Table 5.5.2**).

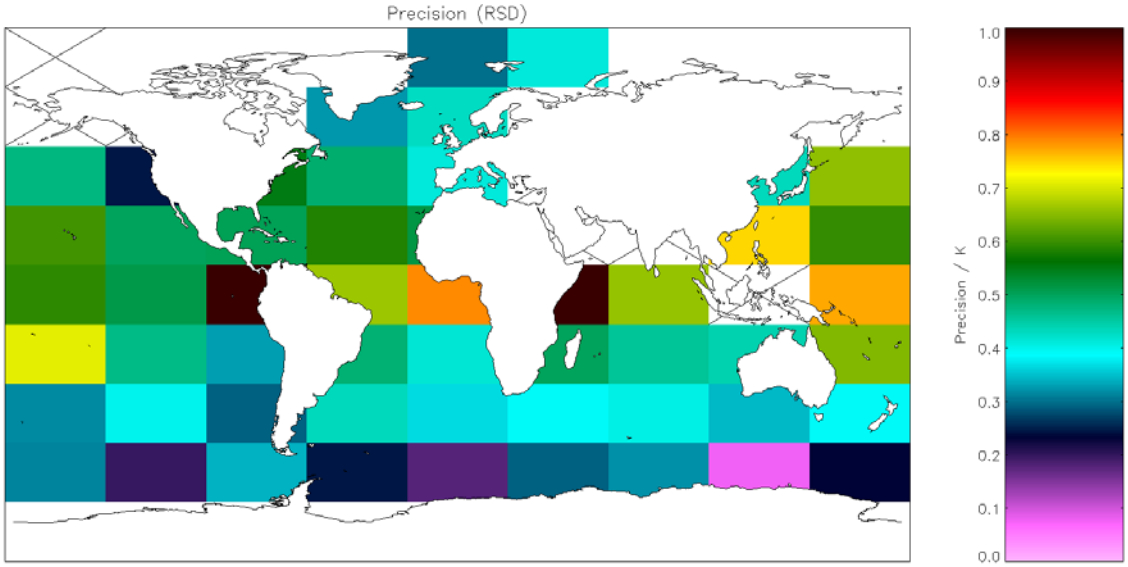


Figure A2.3.2-8: Global map of precision (median of cell RSDs) using OE v1 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.583 K (**Table 5.5.2**).

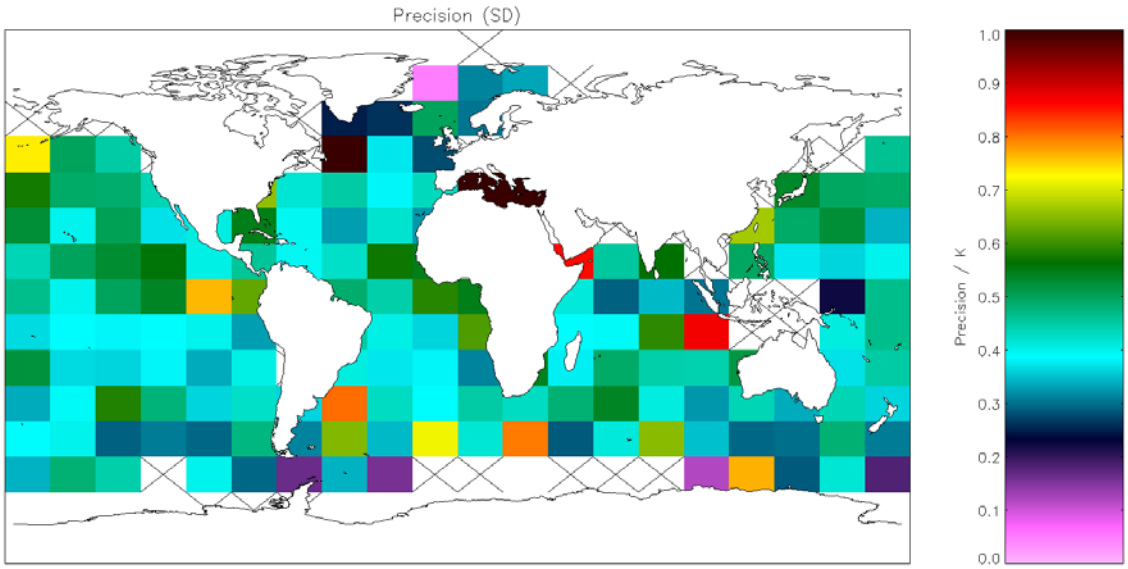


Figure A2.3.2-9: Global map of precision (mean of cell SDs) using OE v1 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.466 K (**Table 5.6.1**).

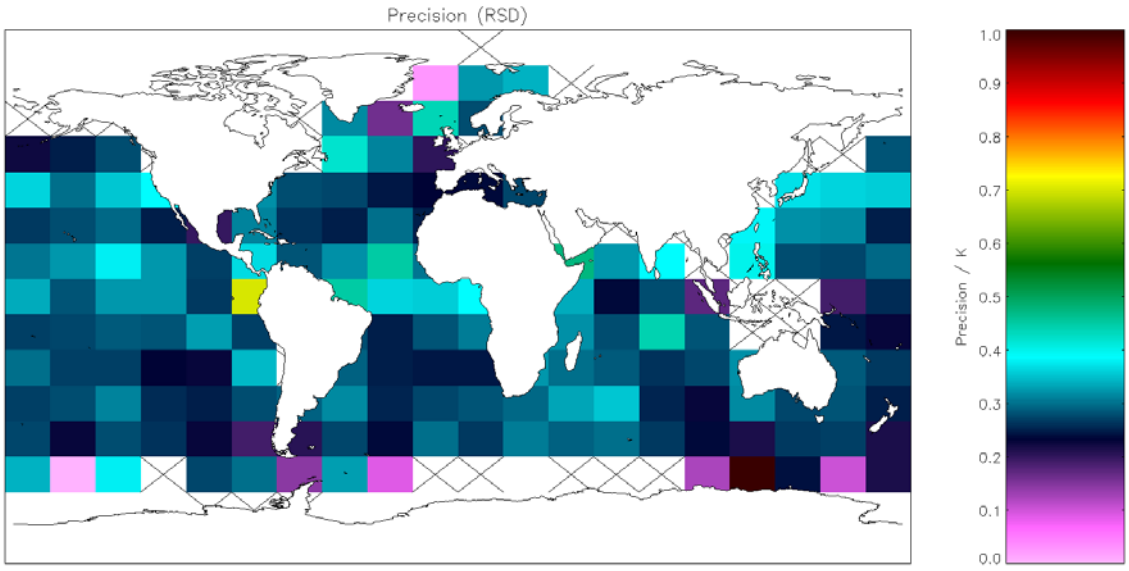


Figure A2.3.2-10: Global map of precision (median of cell RSDs) using OE v1 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.289 K (**Table 5.6.1**).

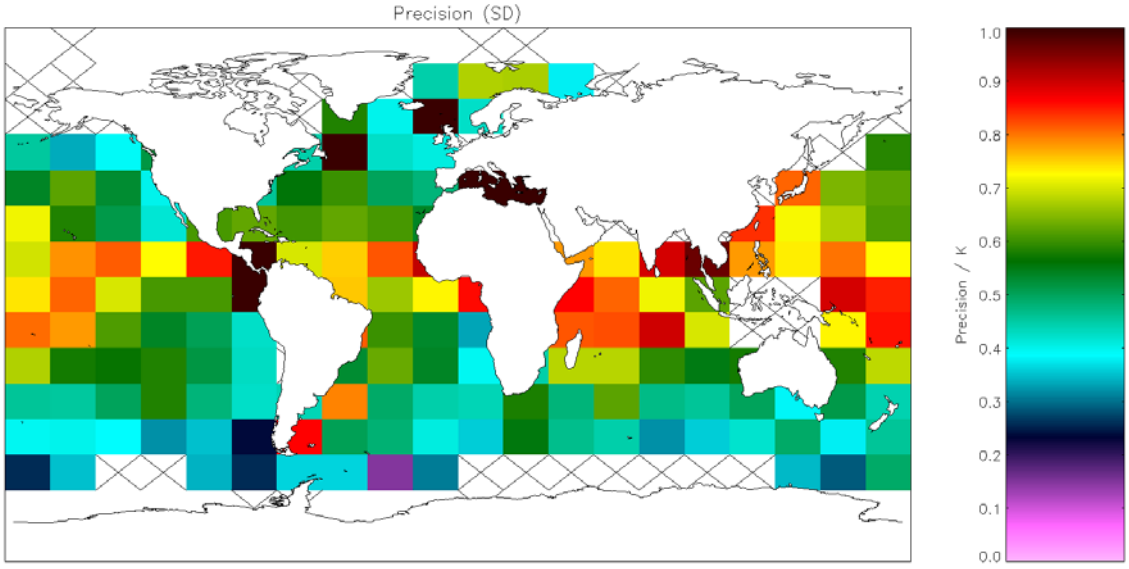


Figure A2.3.2-11: Global map of precision (mean of cell SDs) using OE v1 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.723 K (**Table 5.6.2**).

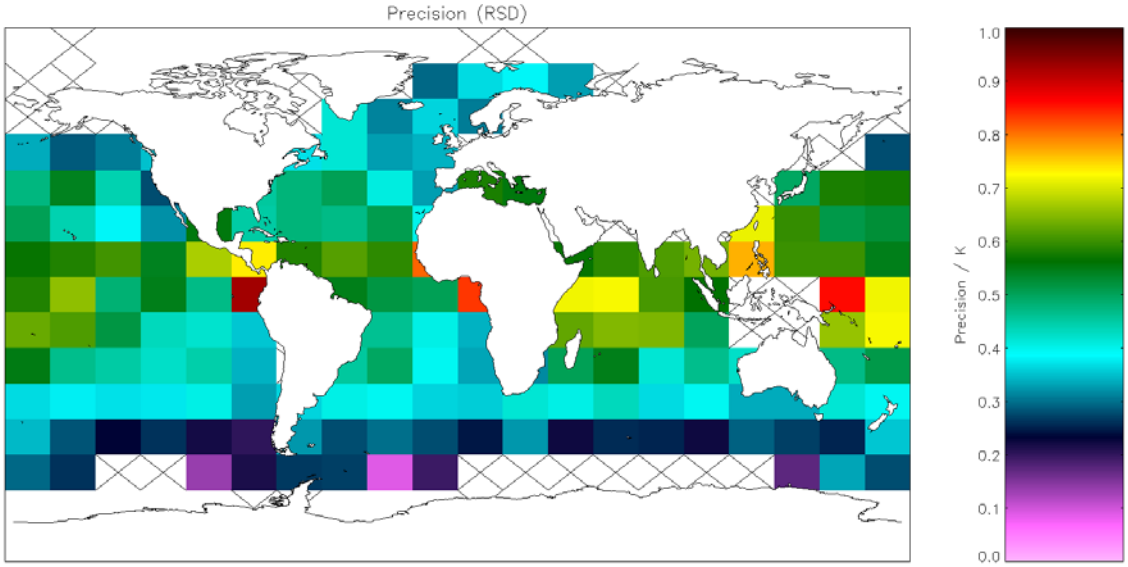


Figure A2.3.2-12: Global map of precision (median of cell RSDs) using OE v1 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.491 K (**Table 5.6.2**).

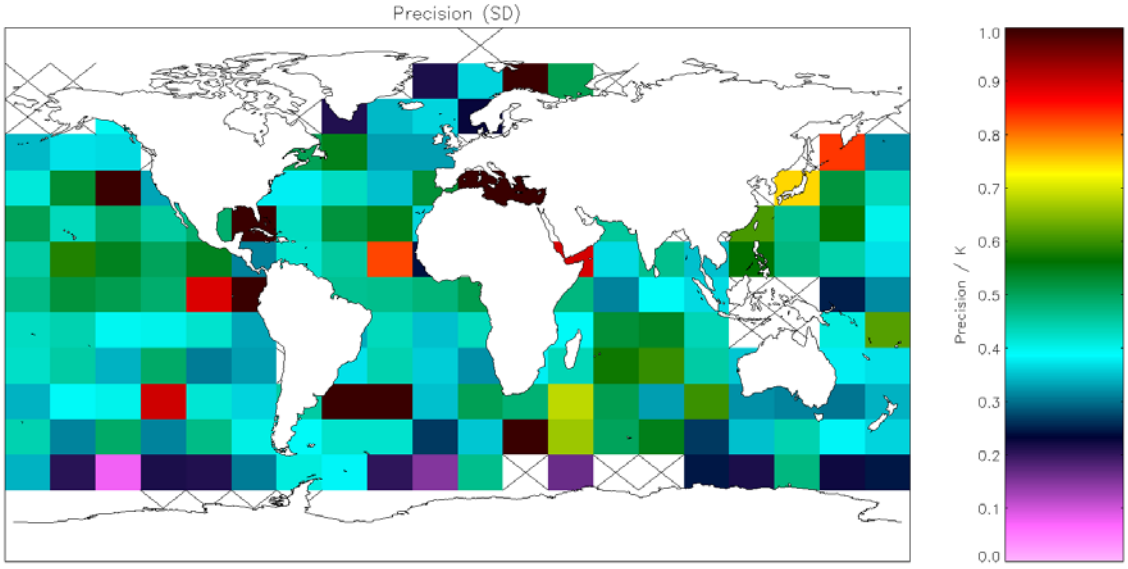


Figure A2.3.2-13: Global map of precision (mean of cell SDs) using OE v1 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.583 K (**Table 5.7.1**).

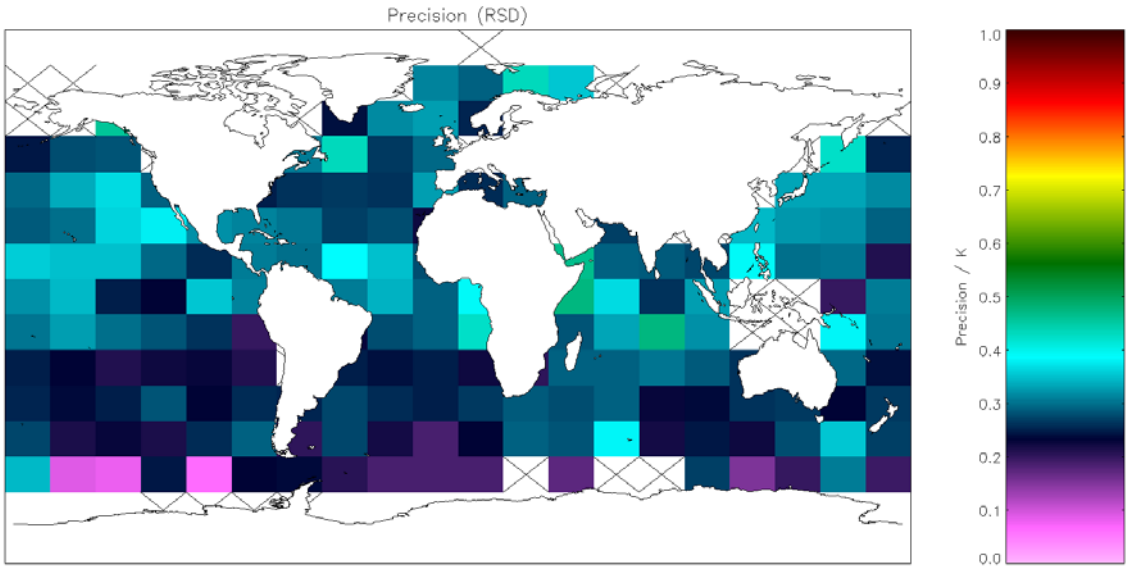


Figure A2.3.2-14: Global map of precision (median of cell RSDs) using OE v1 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.287 K (**Table 5.7.1**).

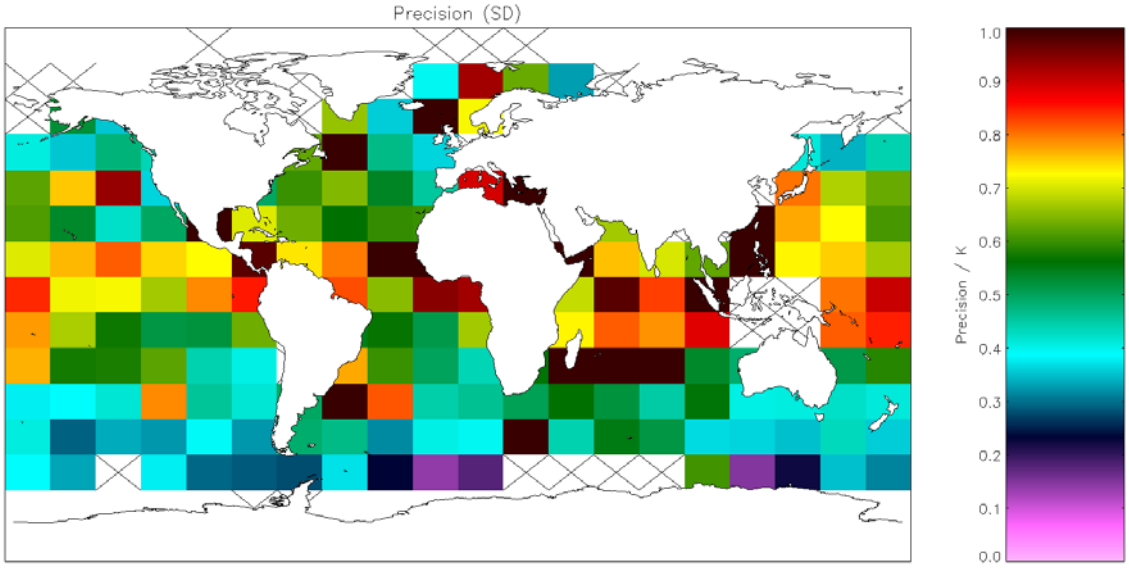


Figure A2.3.2-15: Global map of precision (mean of cell SDs) using OE v1 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.855 K (**Table 5.7.2**).

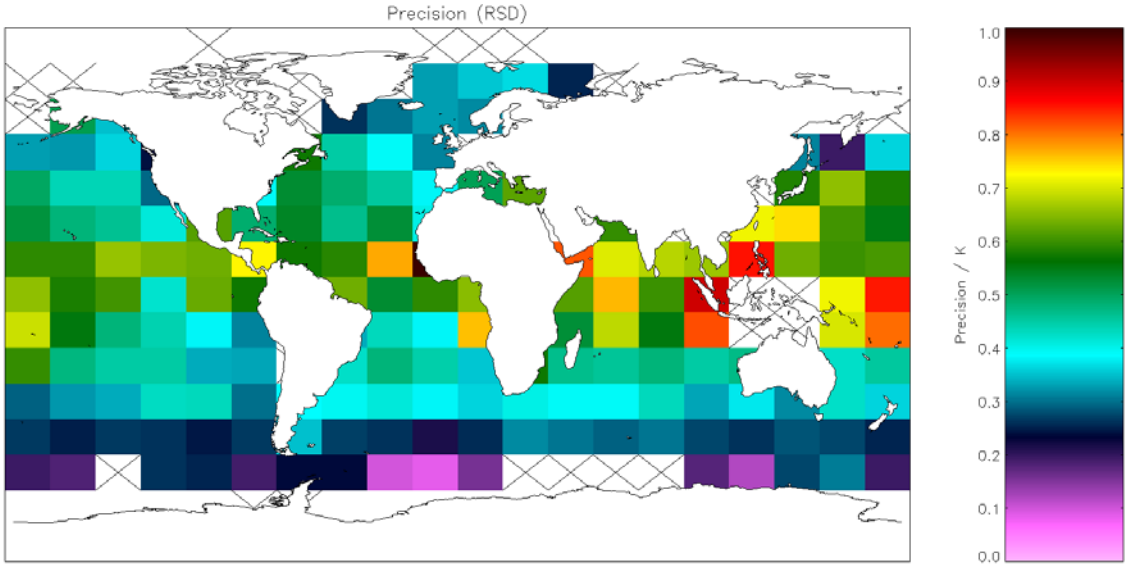


Figure A2.3.2-16: Global map of precision (median of cell RSDs) using OE v1 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.618 K (**Table 5.7.2**).

A2.3.3 Sensitivity

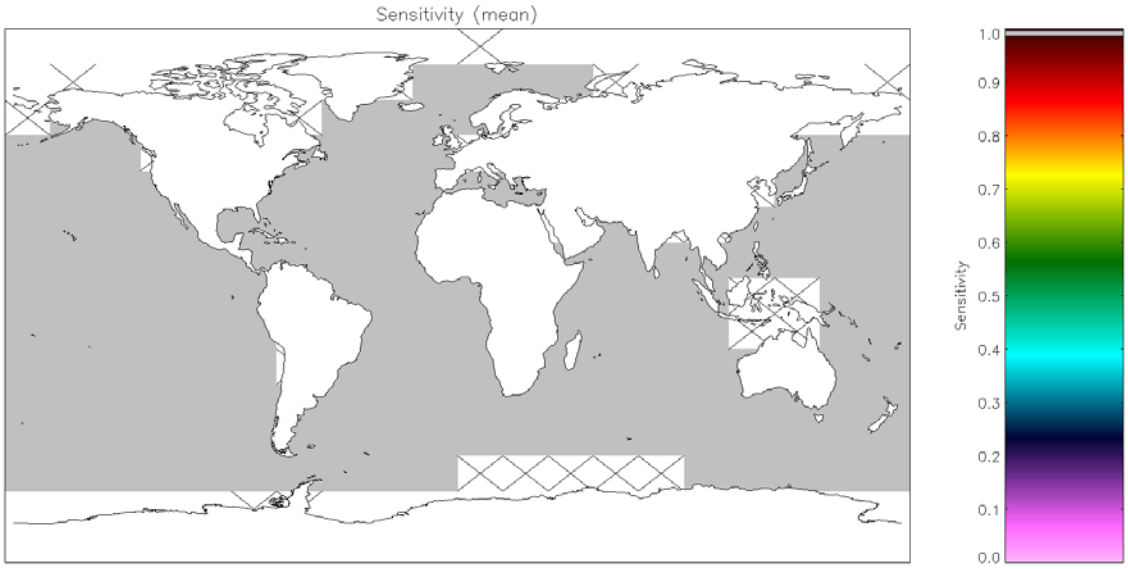


Figure A2.3.3-1: Global map of SST sensitivity using OE v1 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.4.2**).

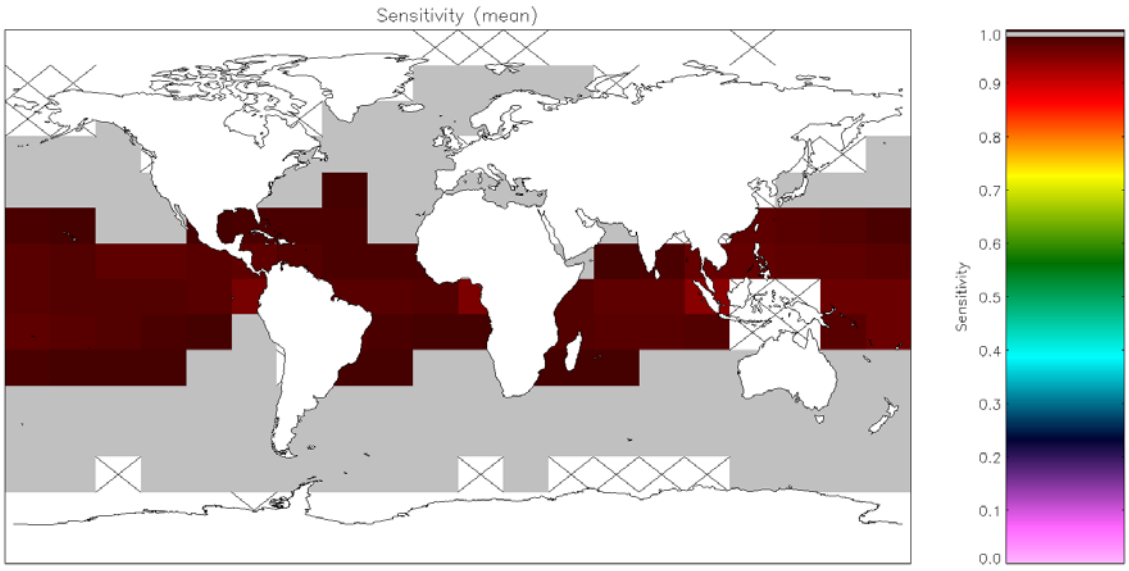


Figure A2.3.3-2: Global map of SST sensitivity using OE v1 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.4.3**).

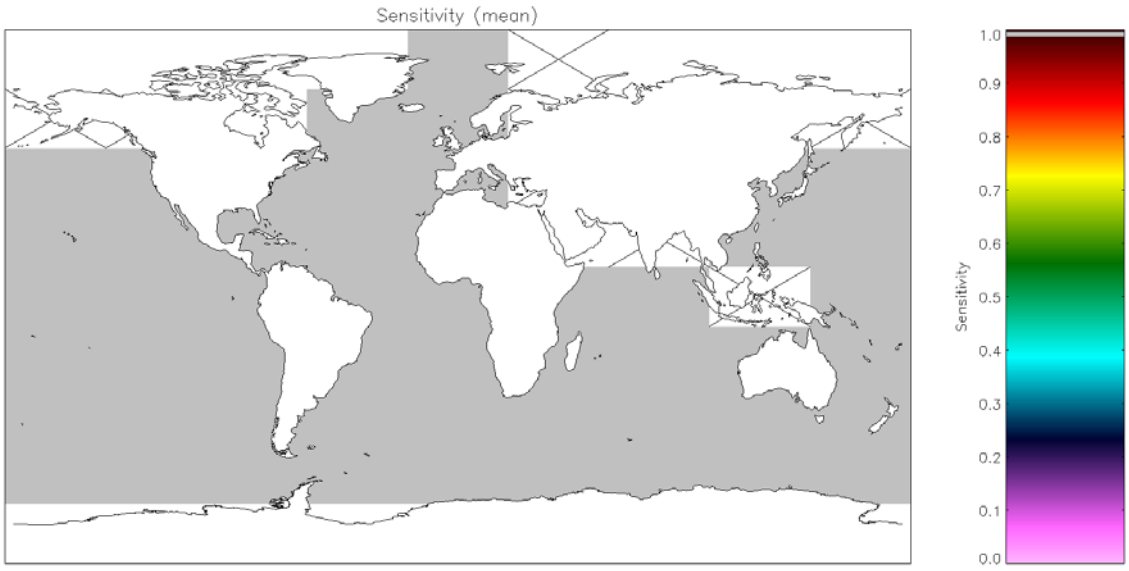


Figure A2.3.3-3: Global map of SST sensitivity using OE v1 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.5.1**).

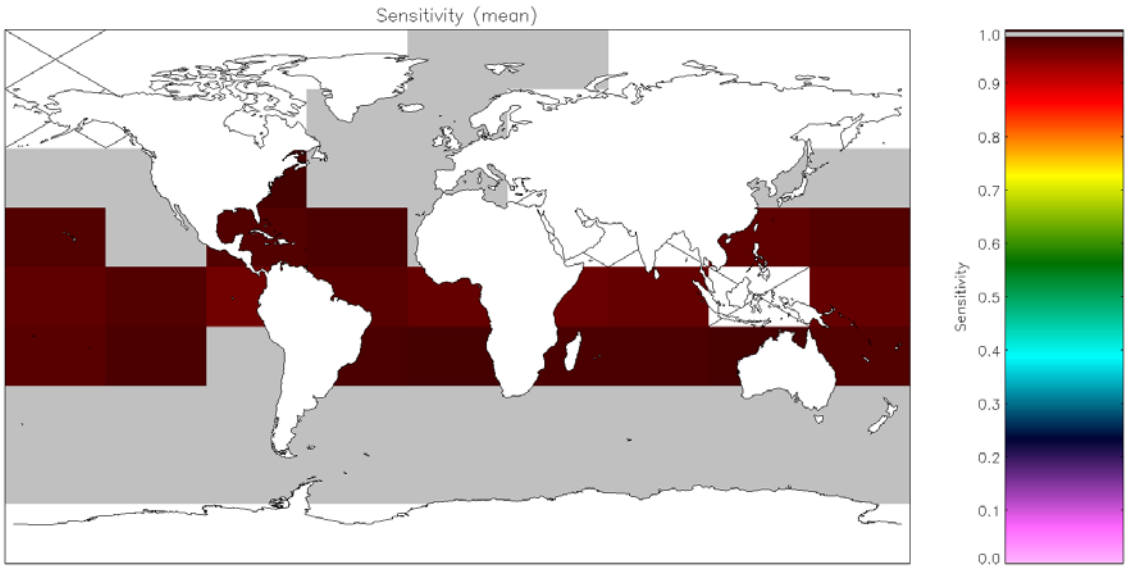


Figure A2.3.3-4: Global map of SST sensitivity using OE v1 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.5.2**).

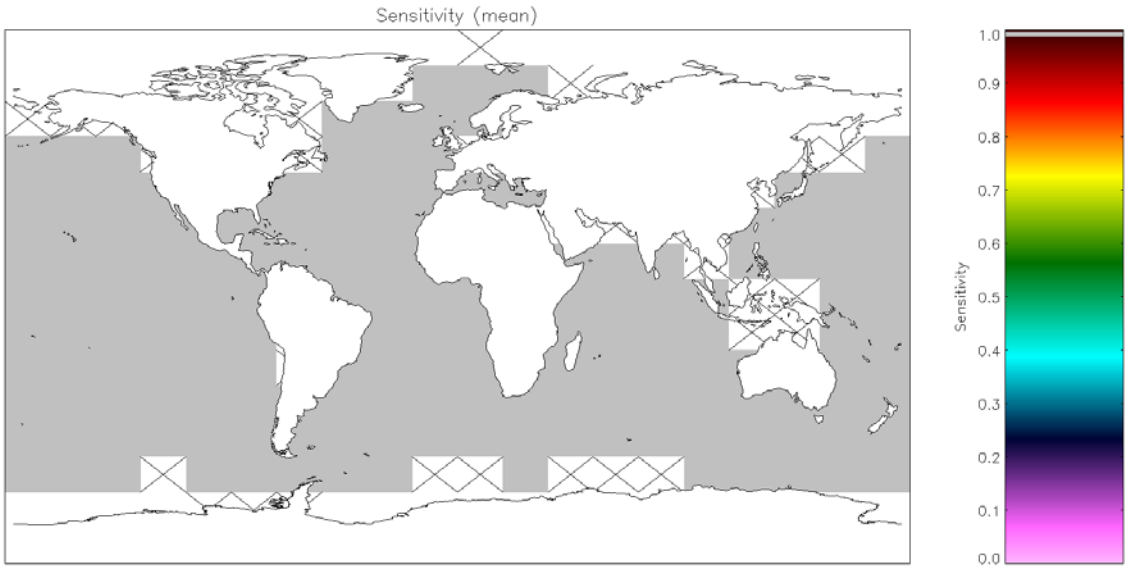


Figure A2.3.3-5: Global map of SST sensitivity using OE v1 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.6.1**).

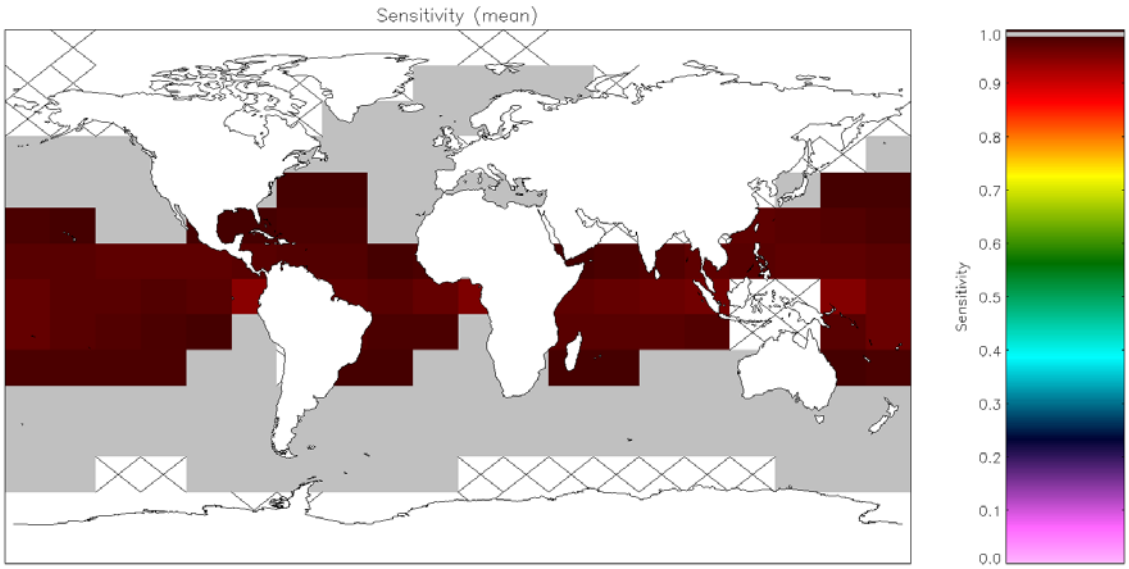


Figure A2.3.3-6: Global map of SST sensitivity using OE v1 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.6.2**).

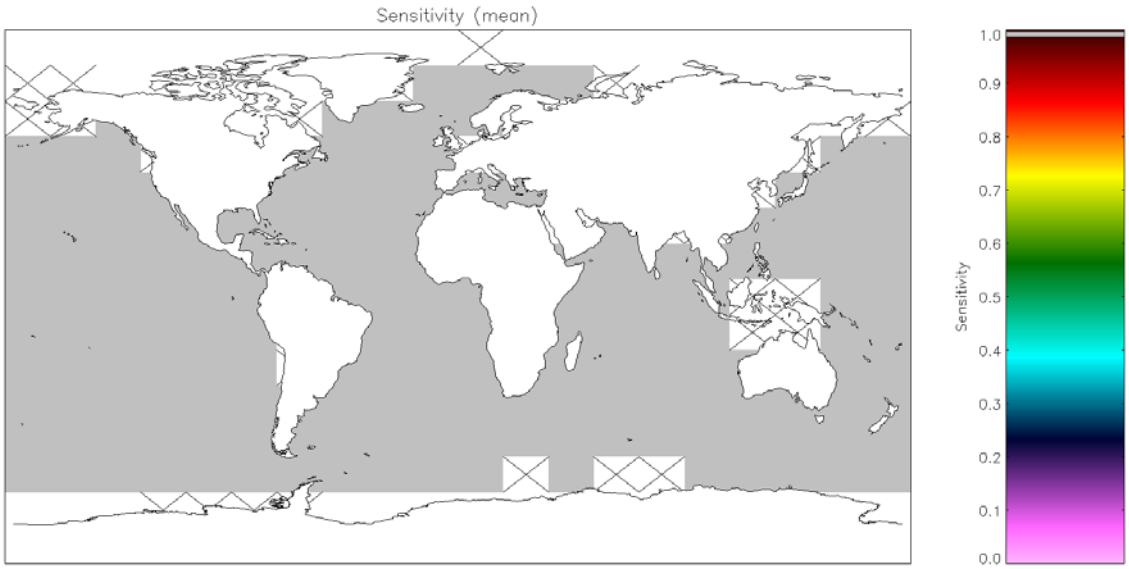


Figure A2.3.3-7: Global map of SST sensitivity using OE v1 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.7.1**).

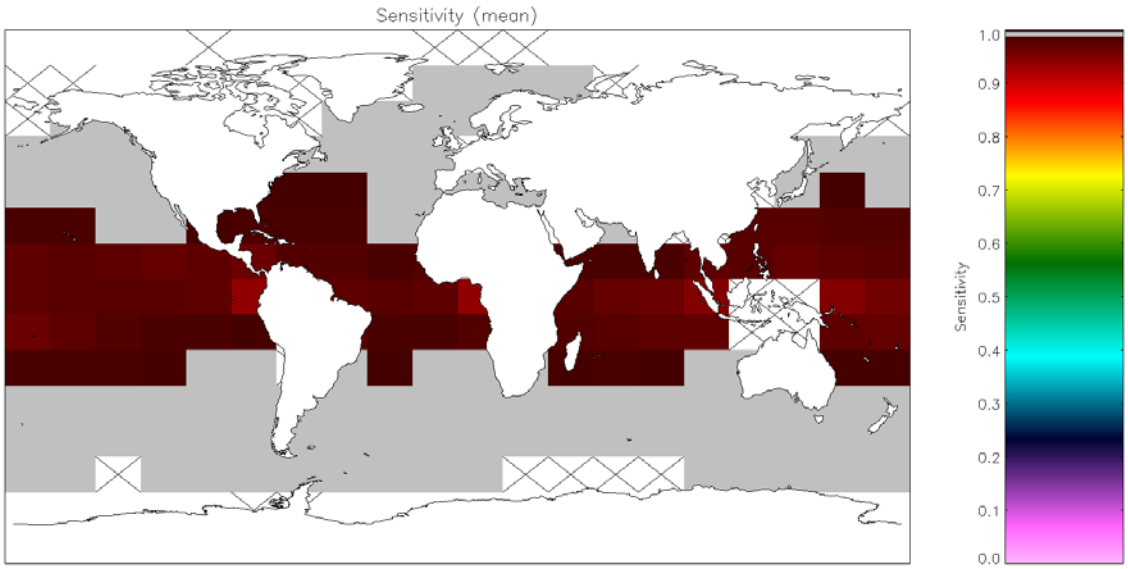


Figure A2.3.3-8: Global map of SST sensitivity using OE v1 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.7.2**).

A2.4 Metrics for CCI SST Optimal Estimation v2

A2.4.1 Bias

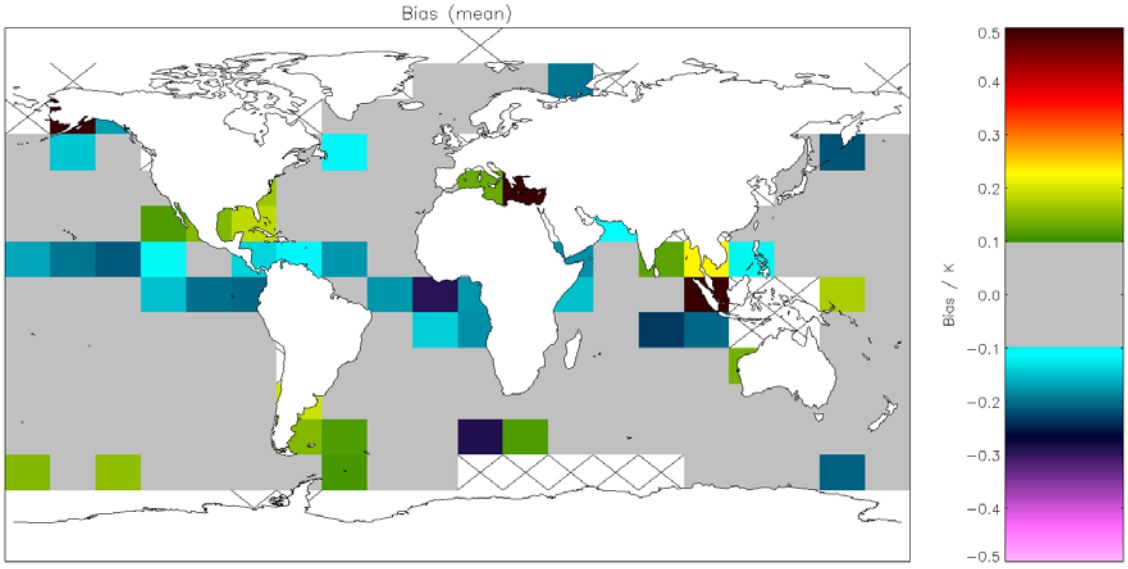


Figure A2.4.1-1: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.001 K (**Table 5.4.2**).

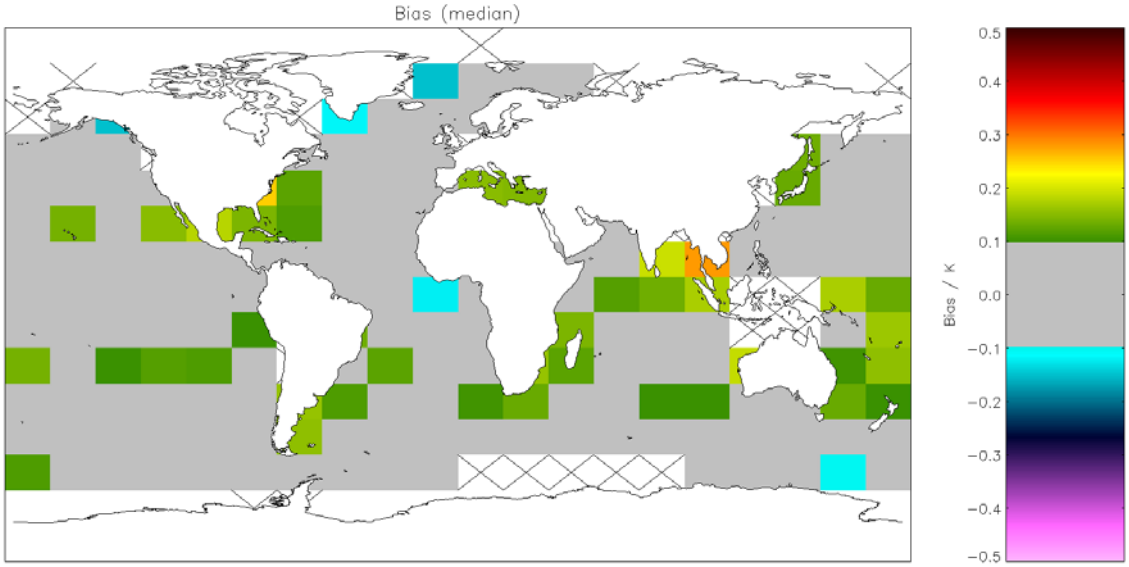


Figure A2.4.1-2: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.069 K (**Table 5.4.2**).

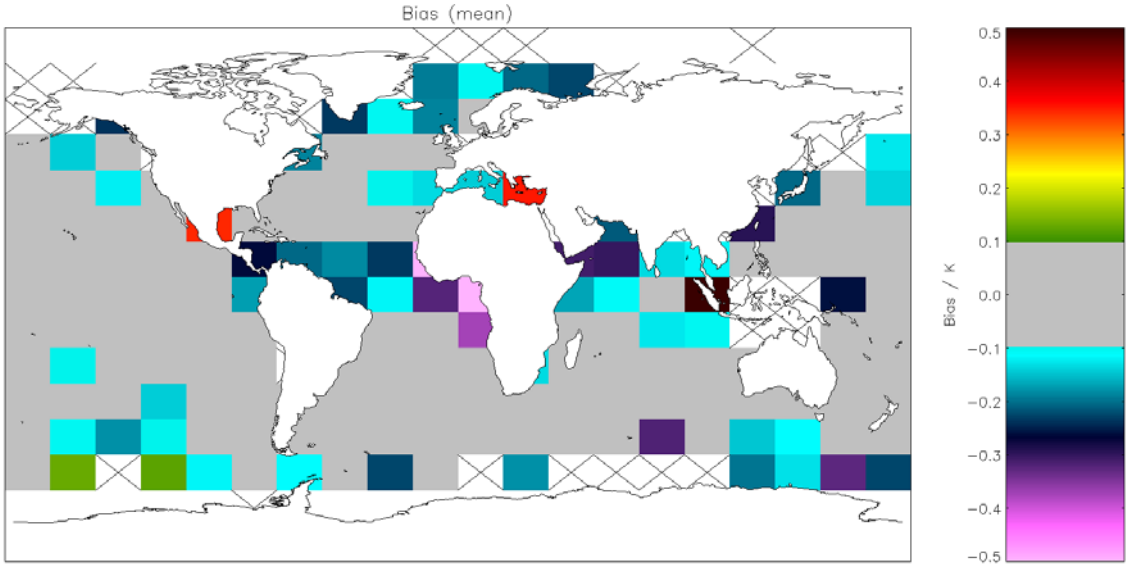


Figure A2.4.1-3: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.050 K (**Table 5.4.3**).

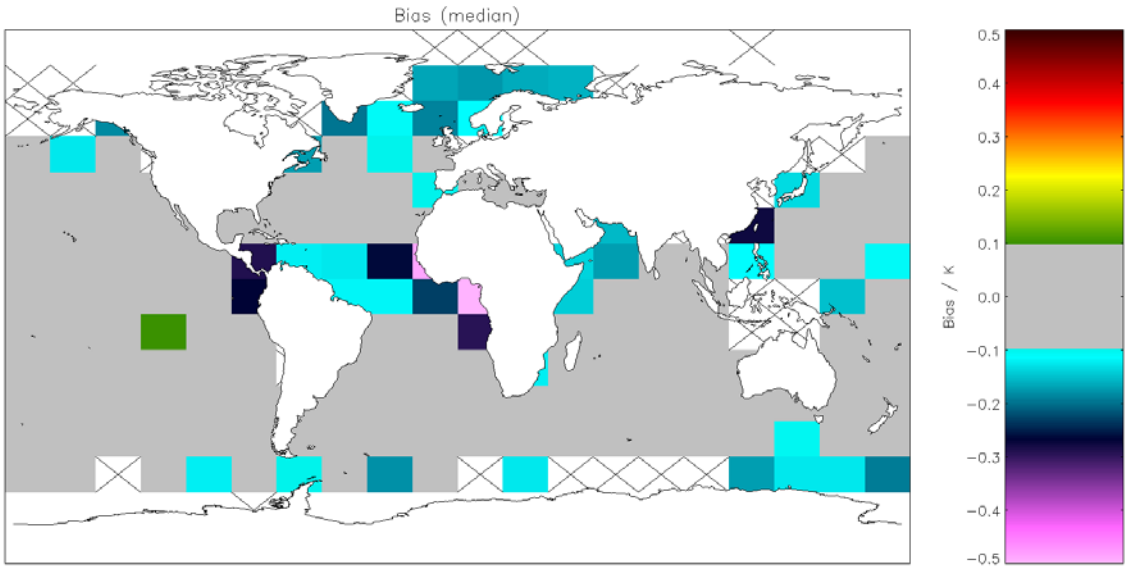


Figure A2.4.1-4: Global map of median bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.021 K (**Table 5.4.3**).

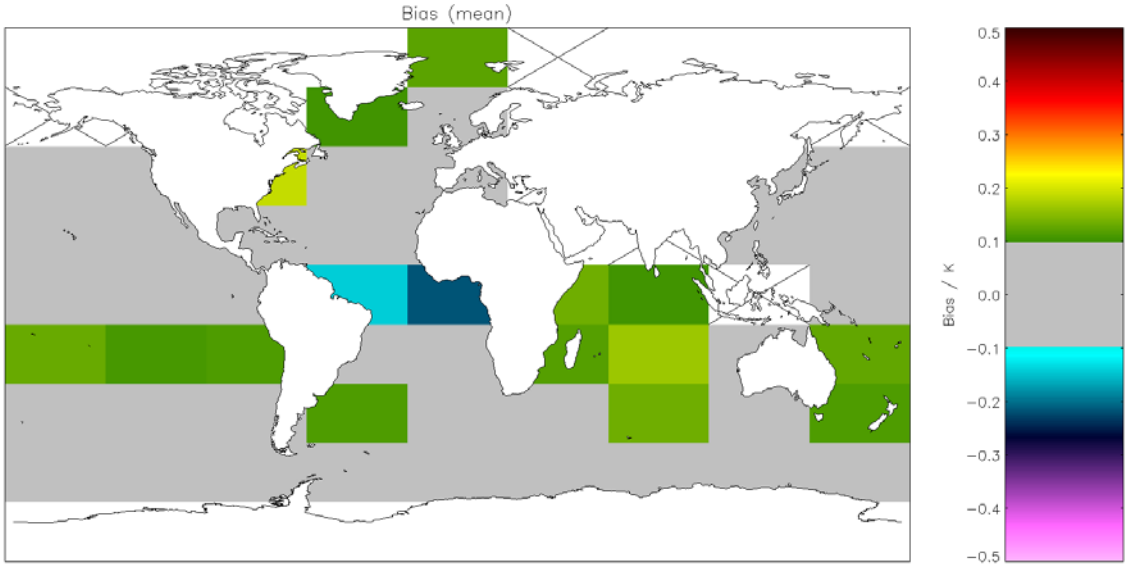


Figure A2.4.1-5: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.052 K (**Table 5.5.1**).

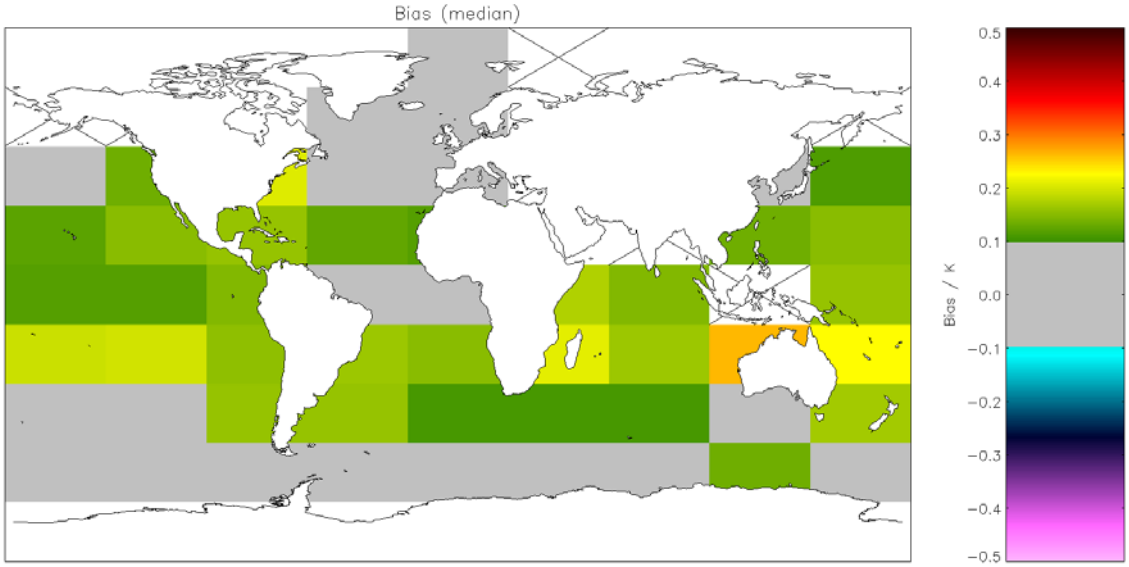


Figure A2.4.1-6: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.126 K (**Table 5.5.1**).

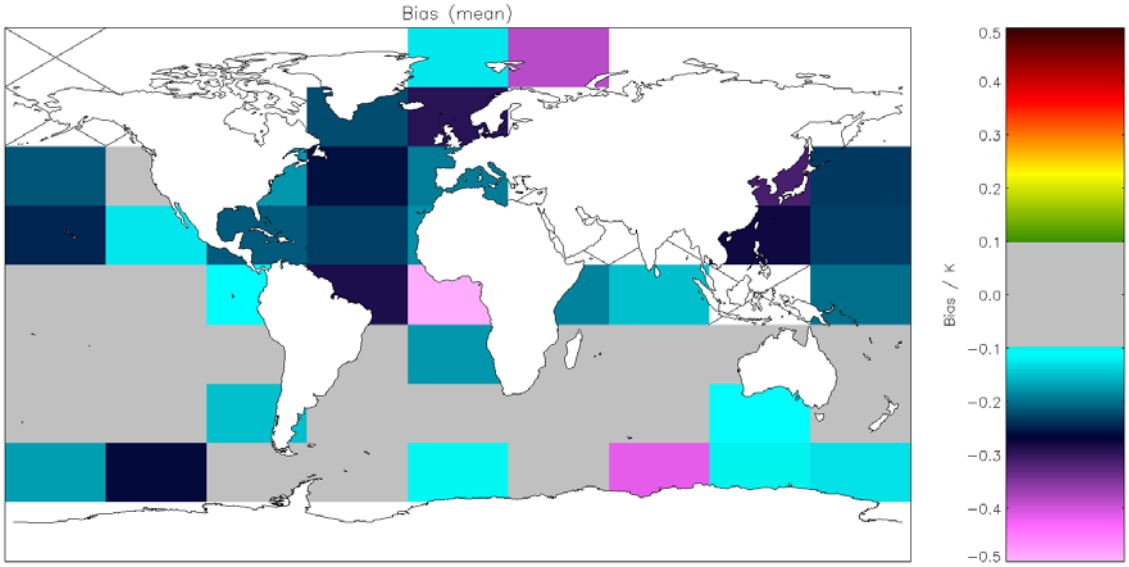


Figure A2.4.1-7: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.161 K (**Table 5.5.2**).

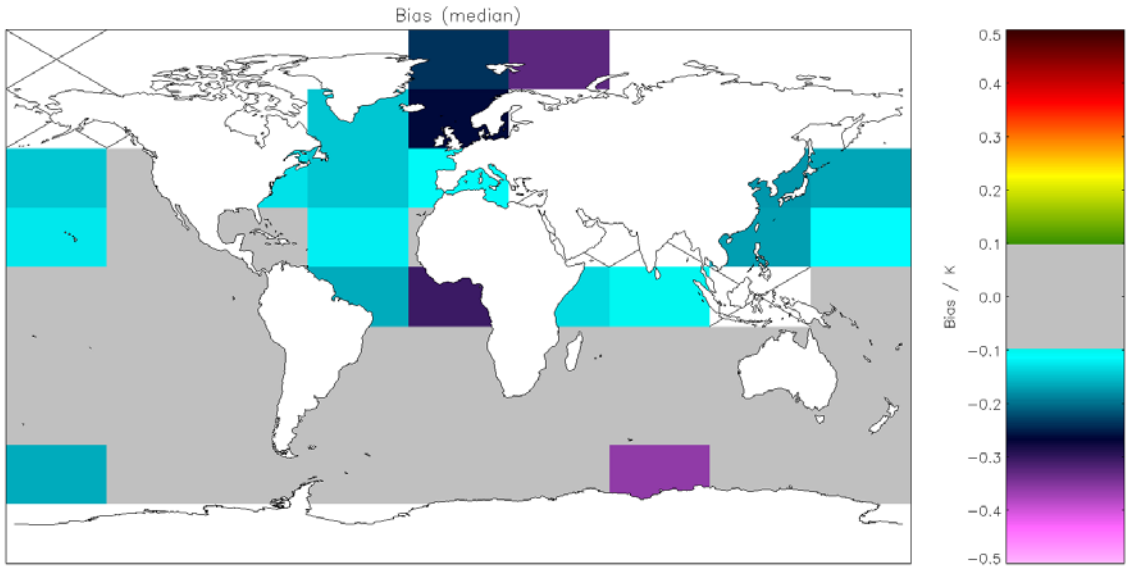


Figure A2.4.1-8: Global map of median bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.076 K (**Table 5.5.2**).

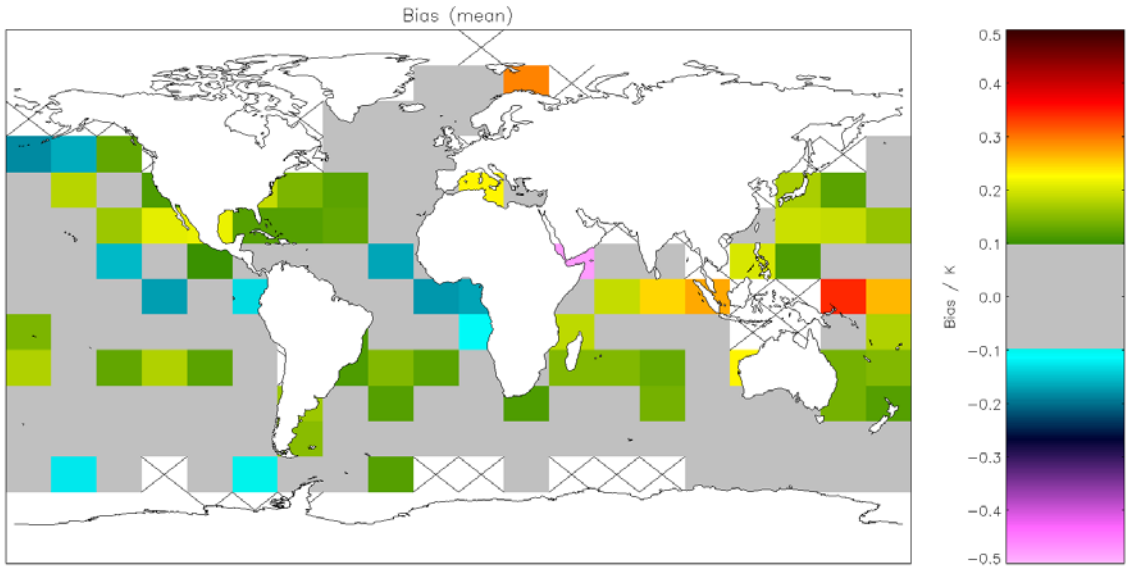


Figure A2.4.1-9: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.074 K (**Table 5.6.1**).

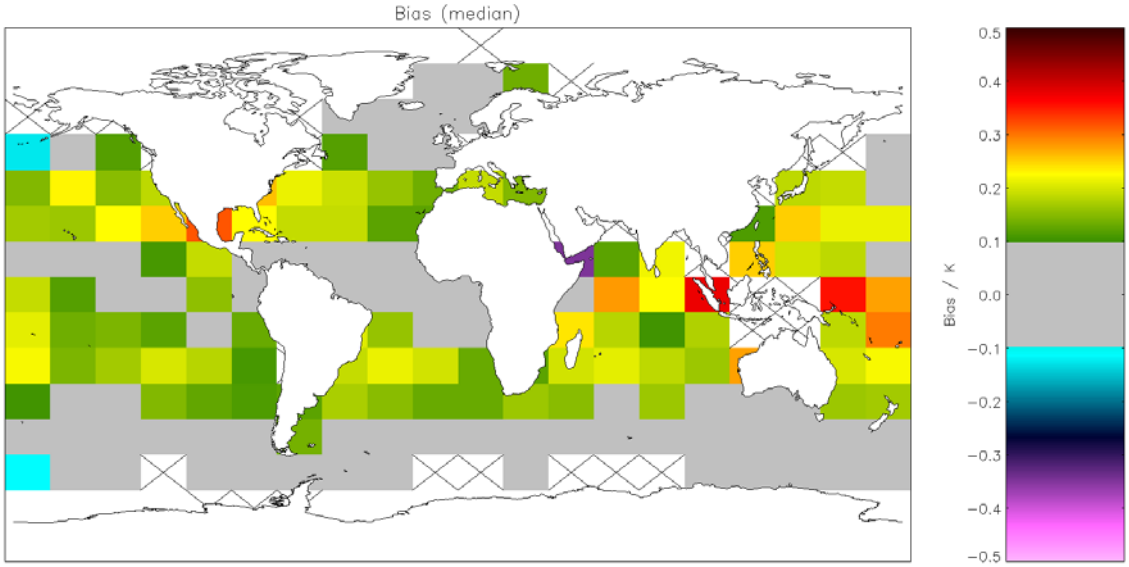


Figure A2.4.1-10: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.147 K (**Table 5.6.1**).

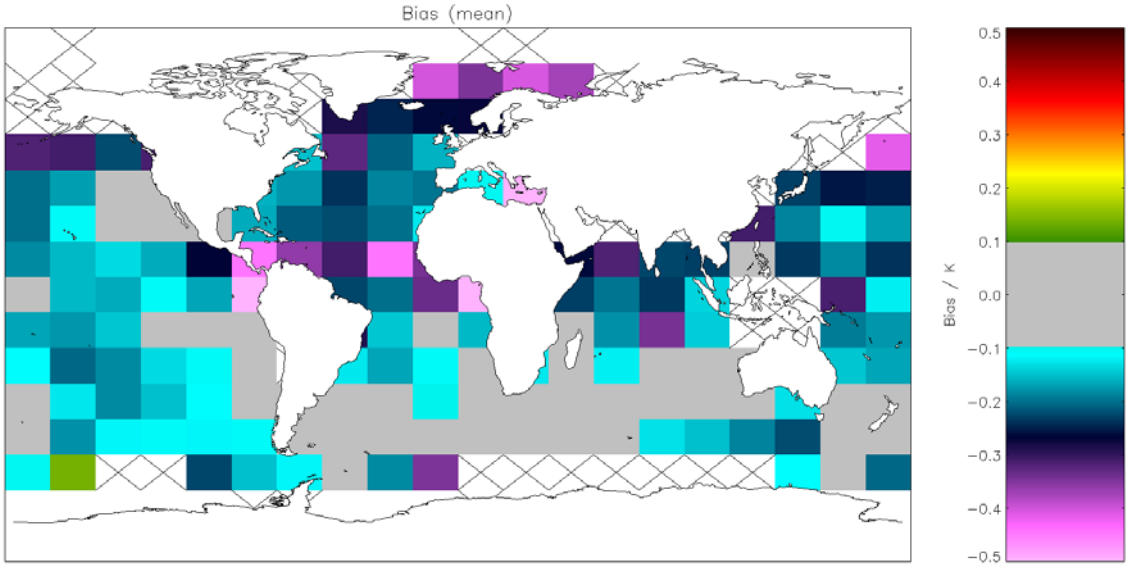


Figure A2.4.1-11: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.161 K (**Table 5.6.2**).

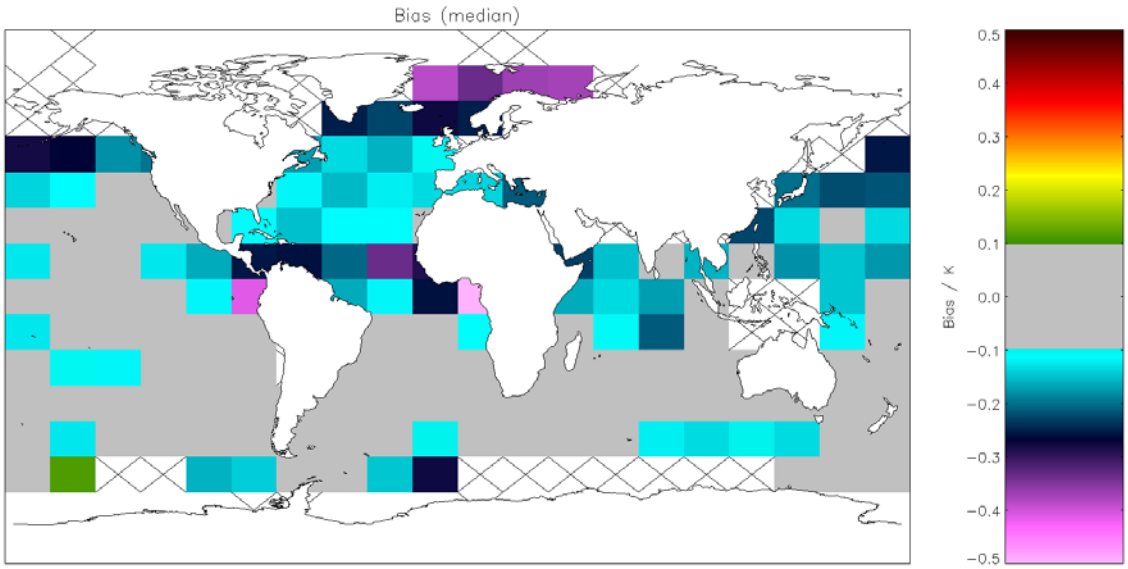


Figure A2.4.1-12: Global map of median bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.090 K (**Table 5.6.2**).

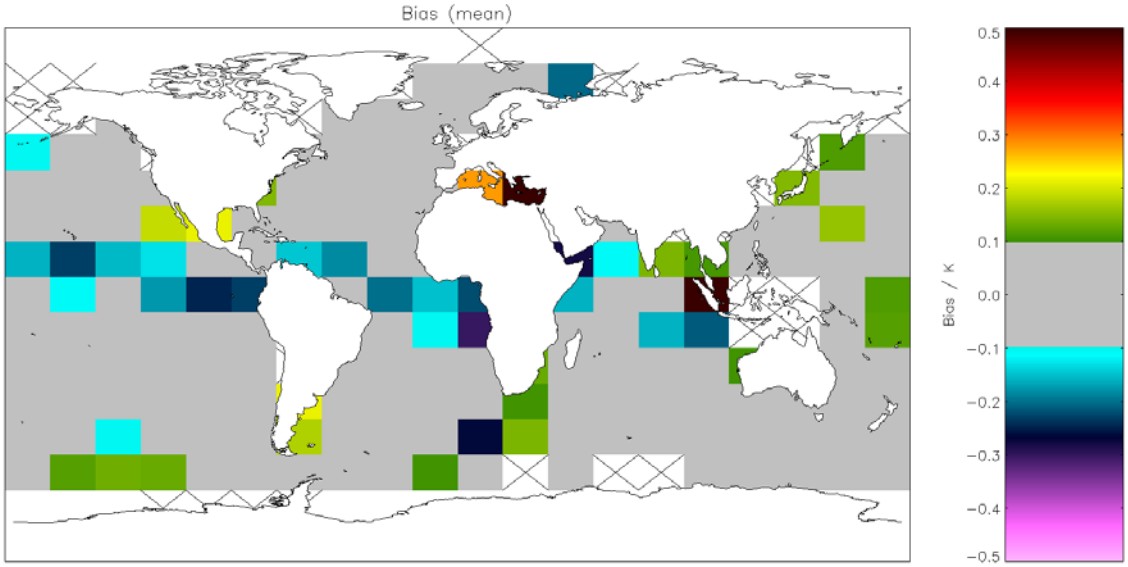


Figure A2.4.1-13: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.005 K (**Table 5.7.1**).

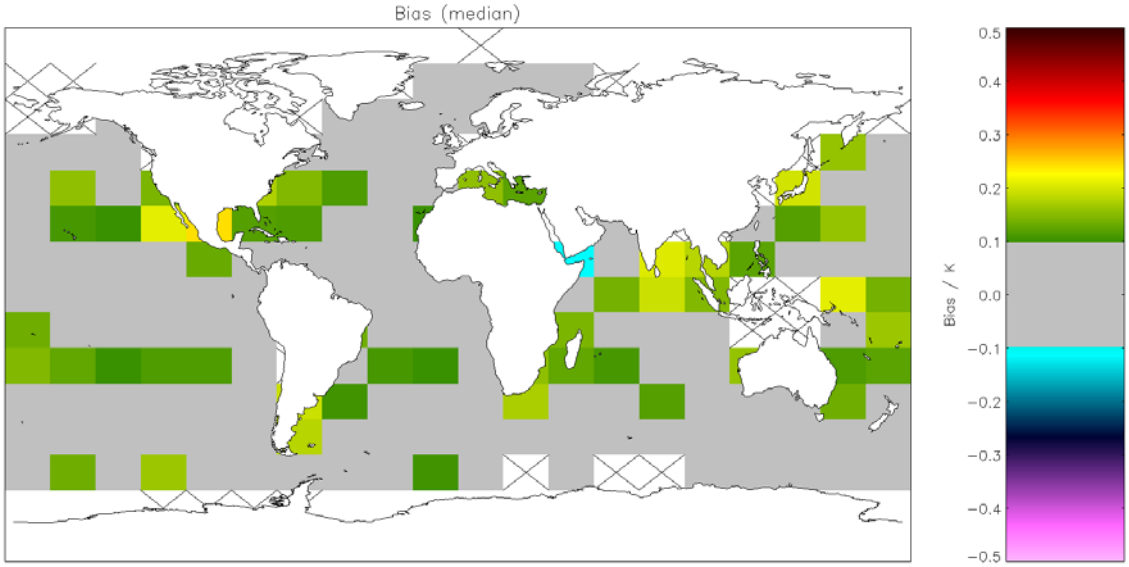


Figure A2.4.1-14: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.077 K (**Table 5.7.1**).

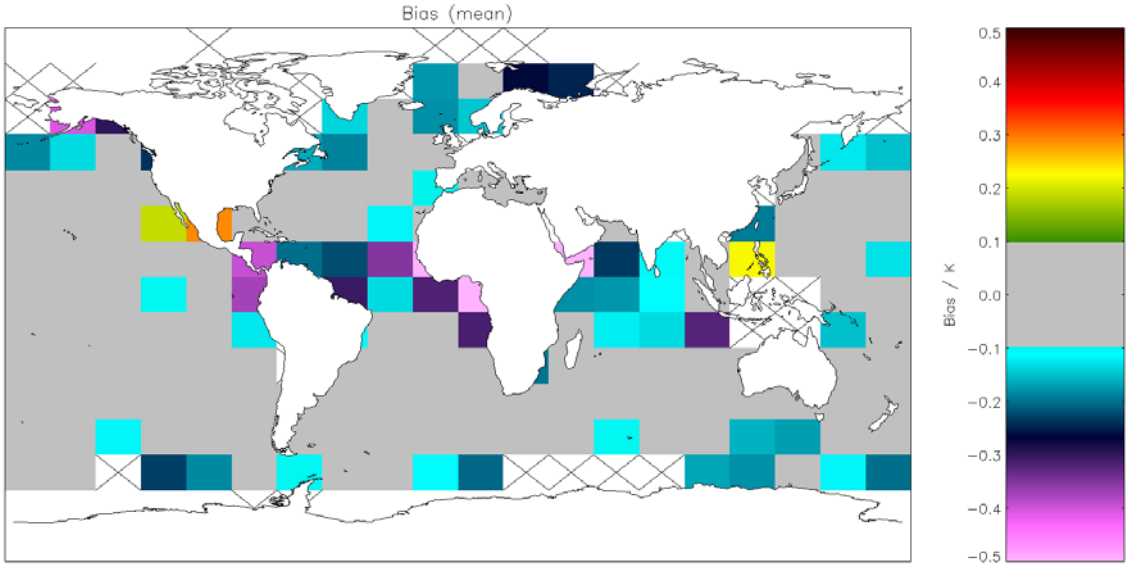


Figure A2.4.1-15: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.050 K (**Table 5.7.2**).

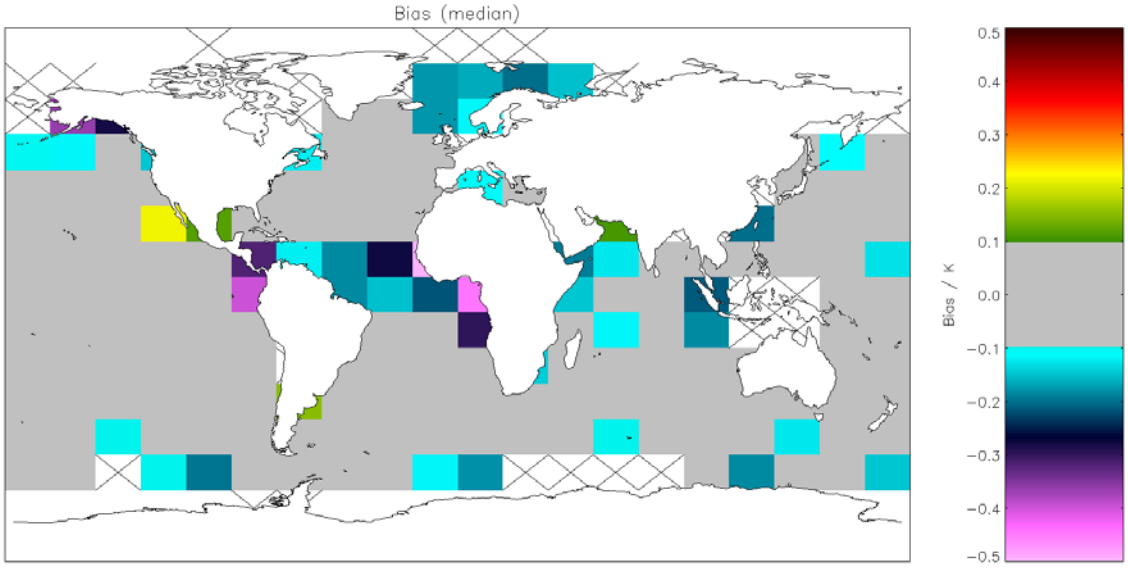


Figure A2.4.1-16: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.021 K (**Table 5.7.2**).

A2.4.2 Precision

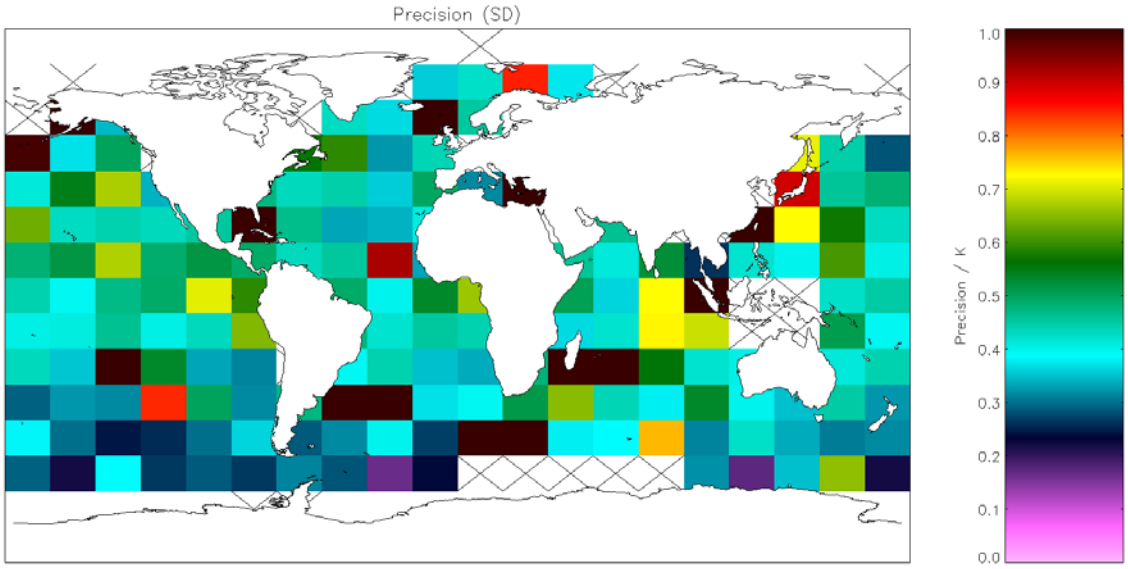


Figure A2.4.2-1: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.653 K (**Table 5.4.2**).

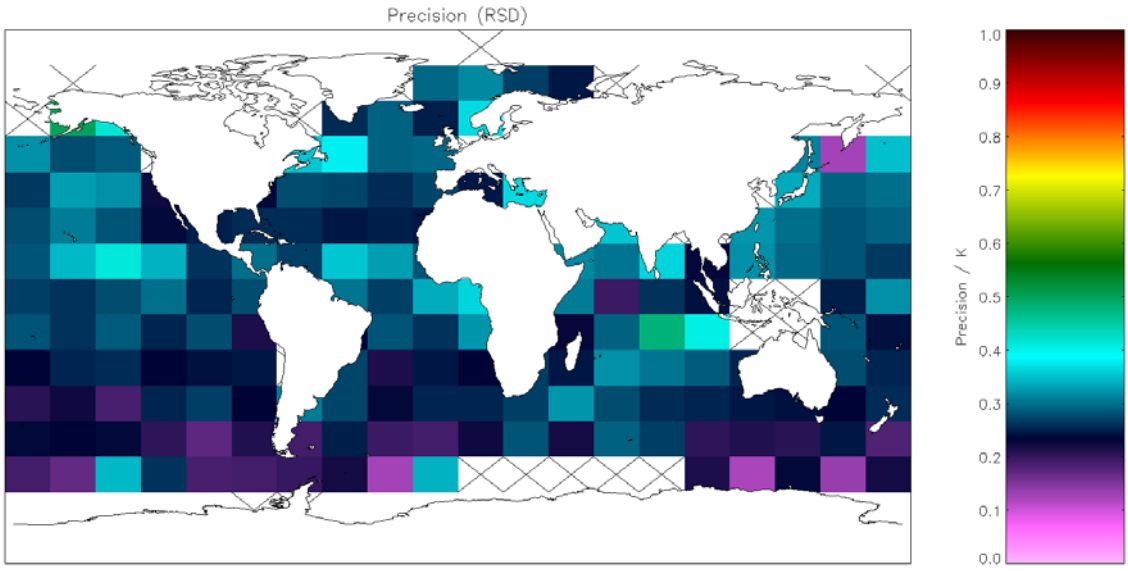


Figure A2.4.2-2: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.272 K (**Table 5.4.2**).

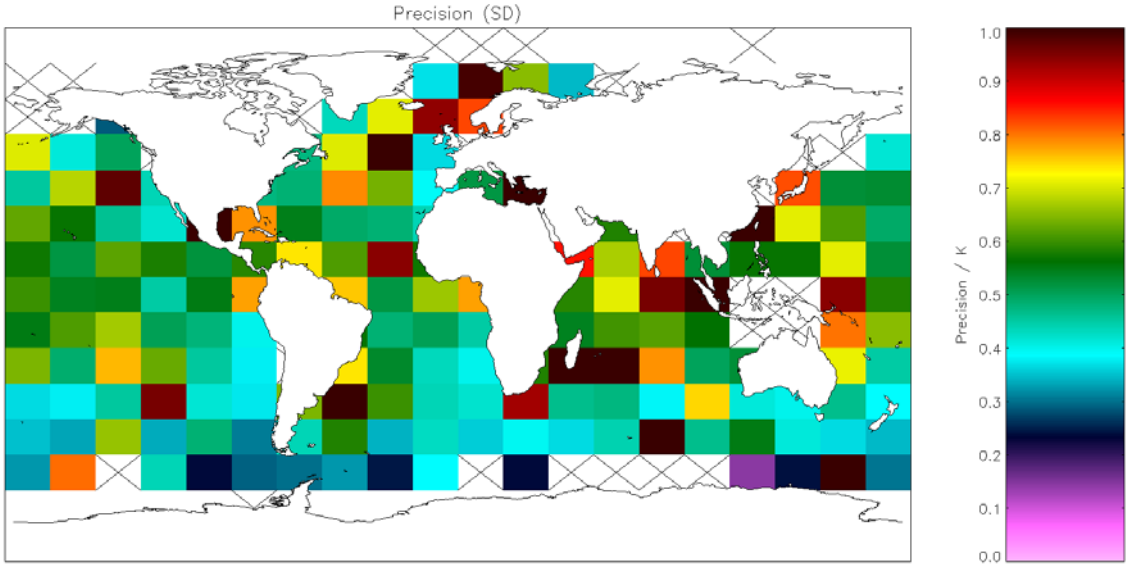


Figure A2.4.2-3: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.687 K (**Table 5.4.3**).

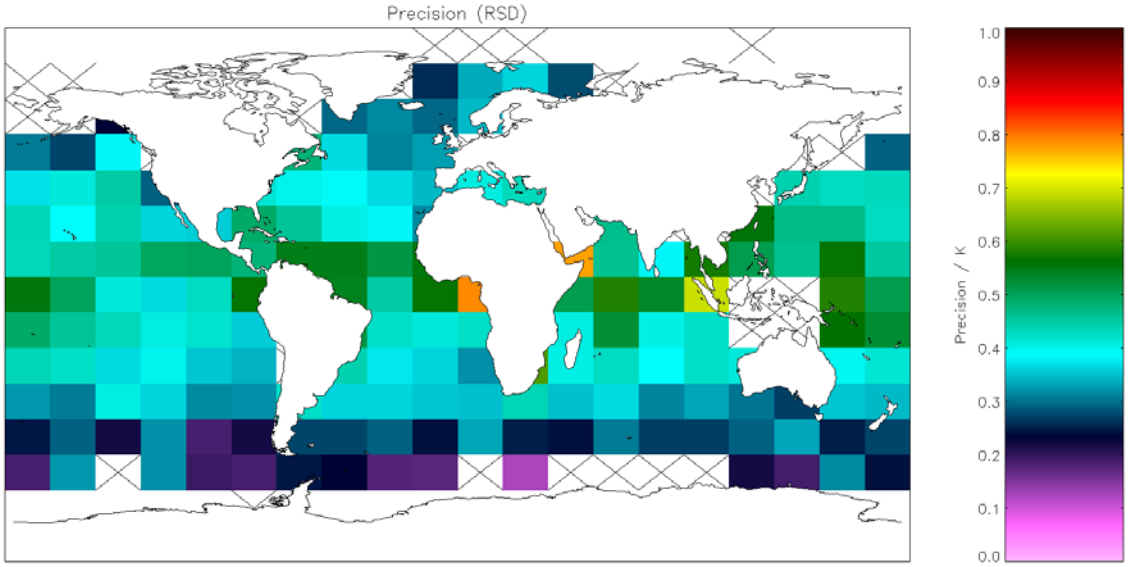


Figure A2.4.2-4: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.400 K (**Table 5.4.3**).

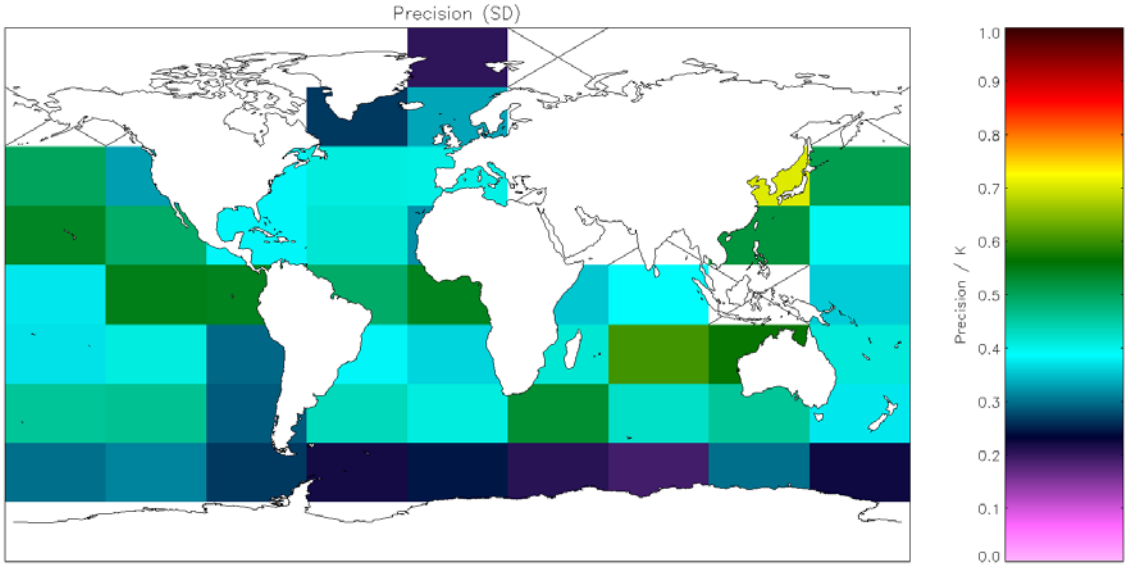


Figure A2.4.2-5: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.441 K (**Table 5.5.1**).

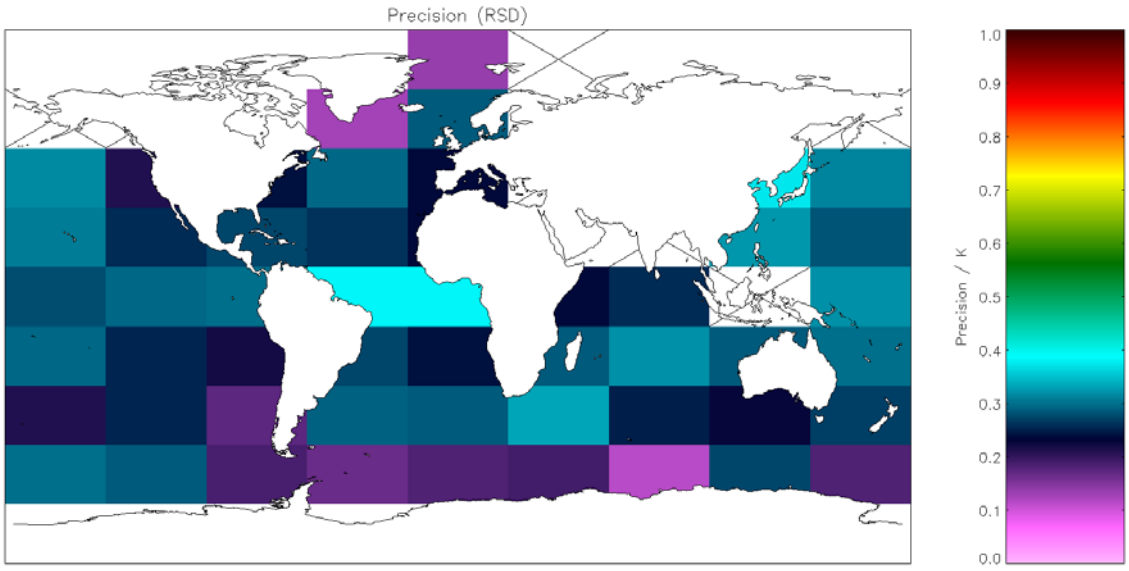


Figure A2.4.2-6: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.280 K (**Table 5.5.1**).

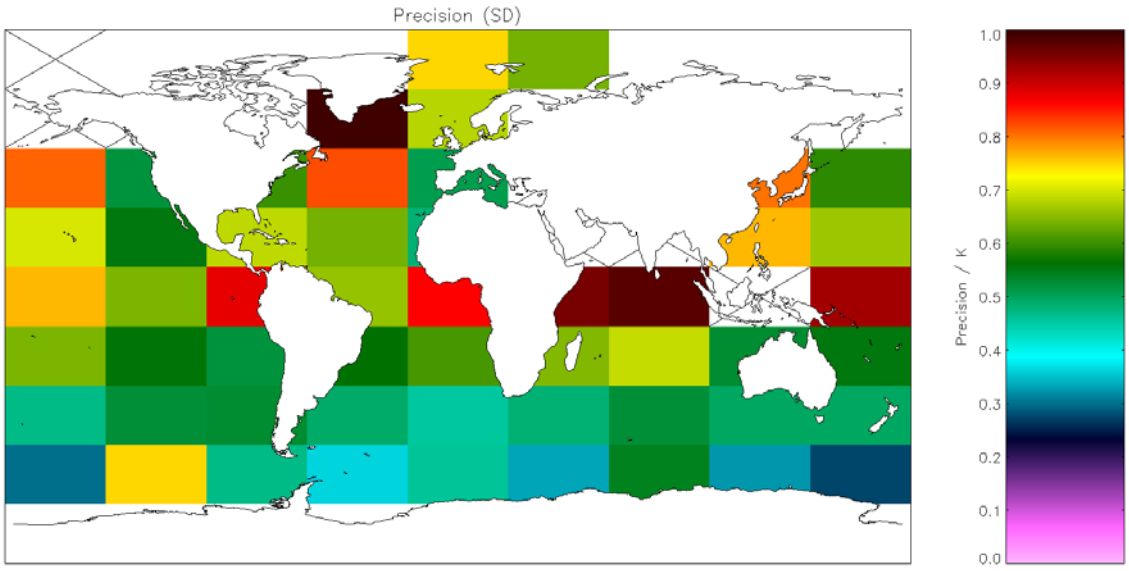


Figure A2.4.2-7: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.658 K (**Table 5.5.2**).

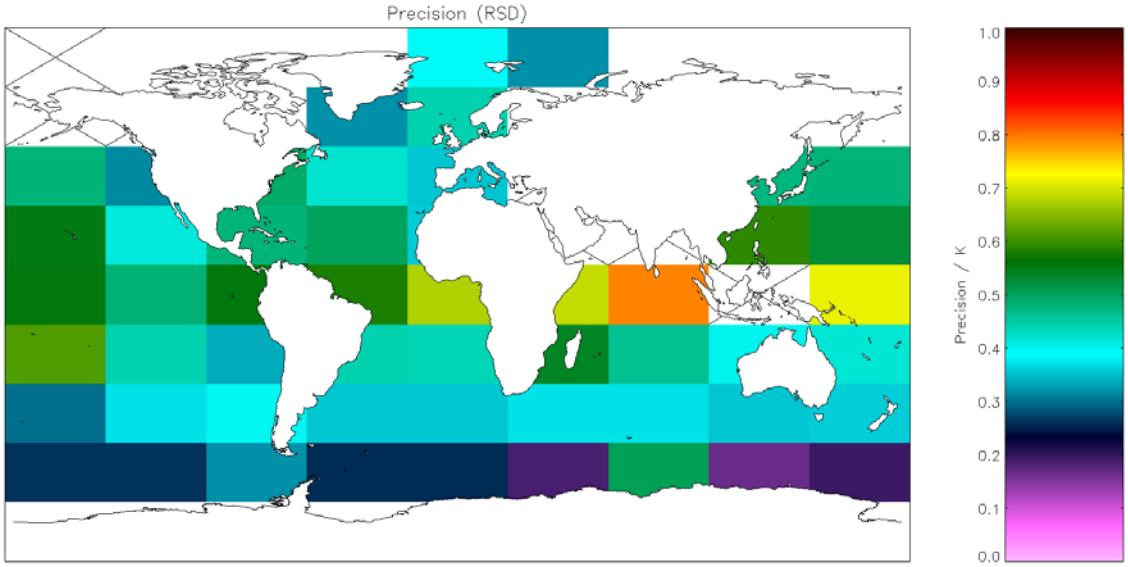


Figure A2.4.2-8: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.462 K (**Table 5.5.2**).

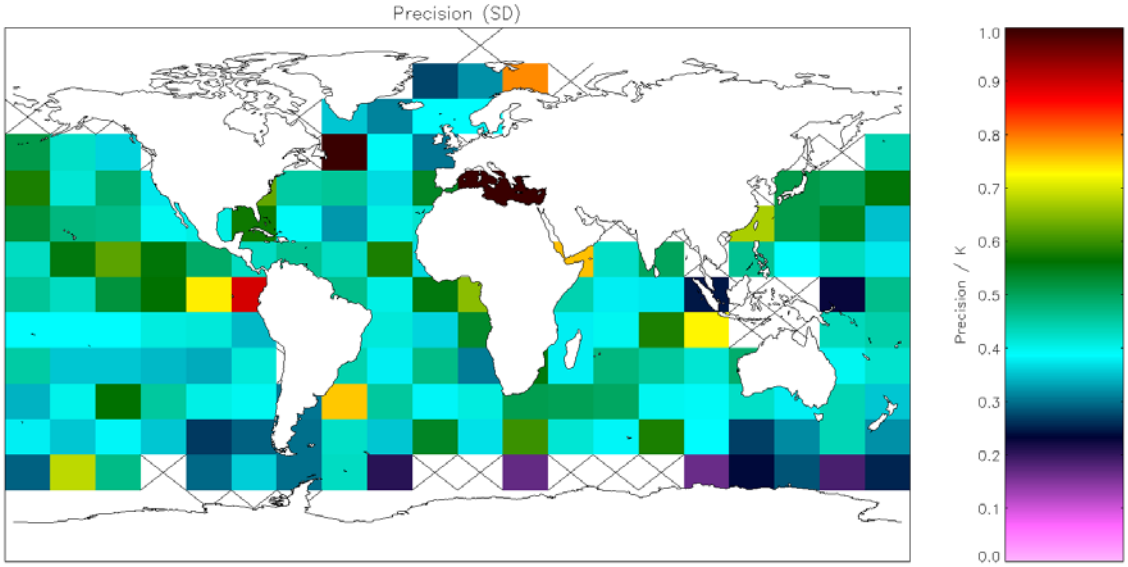


Figure A2.4.2-9: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.477 K (**Table 5.6.1**).

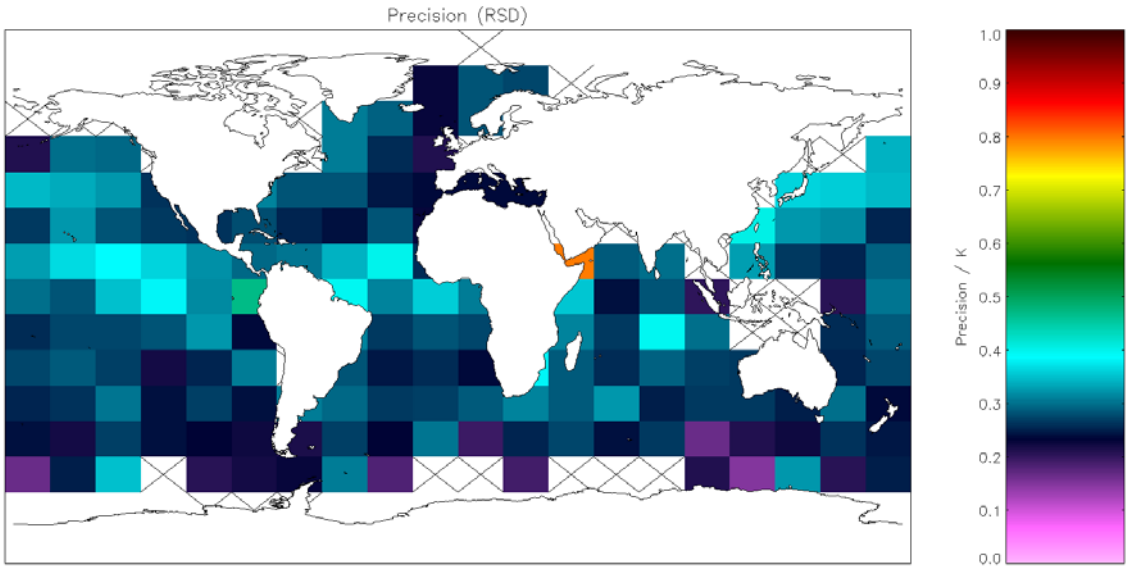


Figure A2.4.2-10: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.289 K (**Table 5.6.1**).

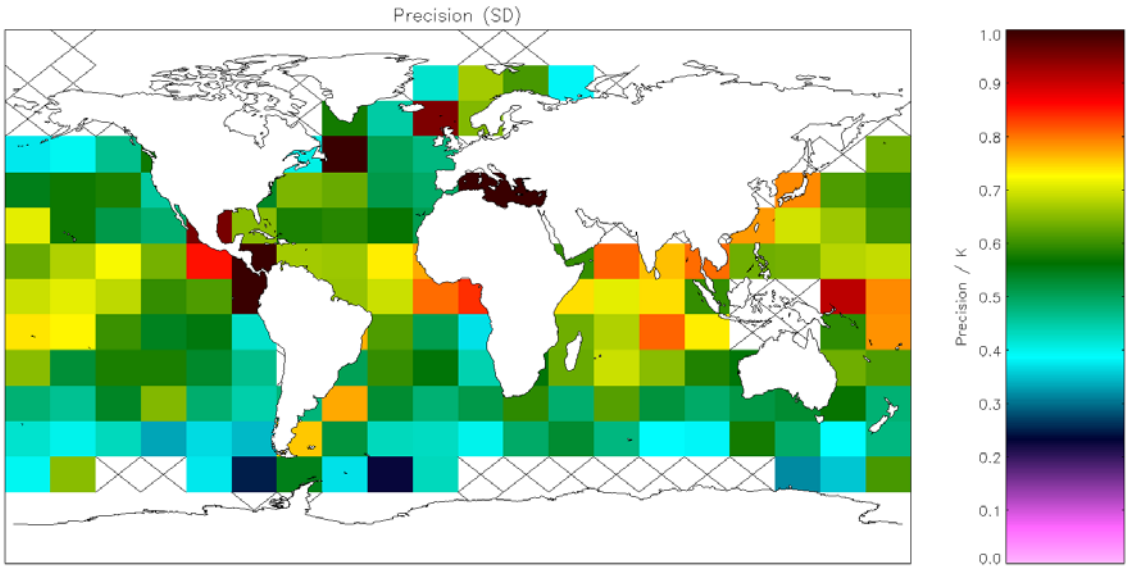


Figure A2.4.2-11: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.687 K (**Table 5.6.2**).

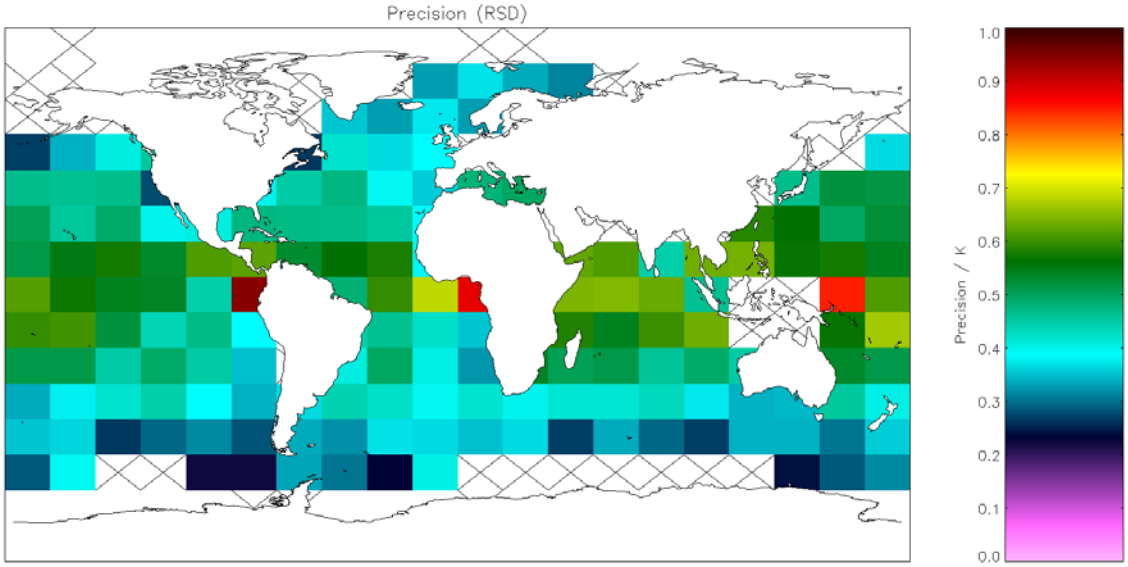


Figure A2.4.2-12: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.465 K (**Table 5.6.2**).

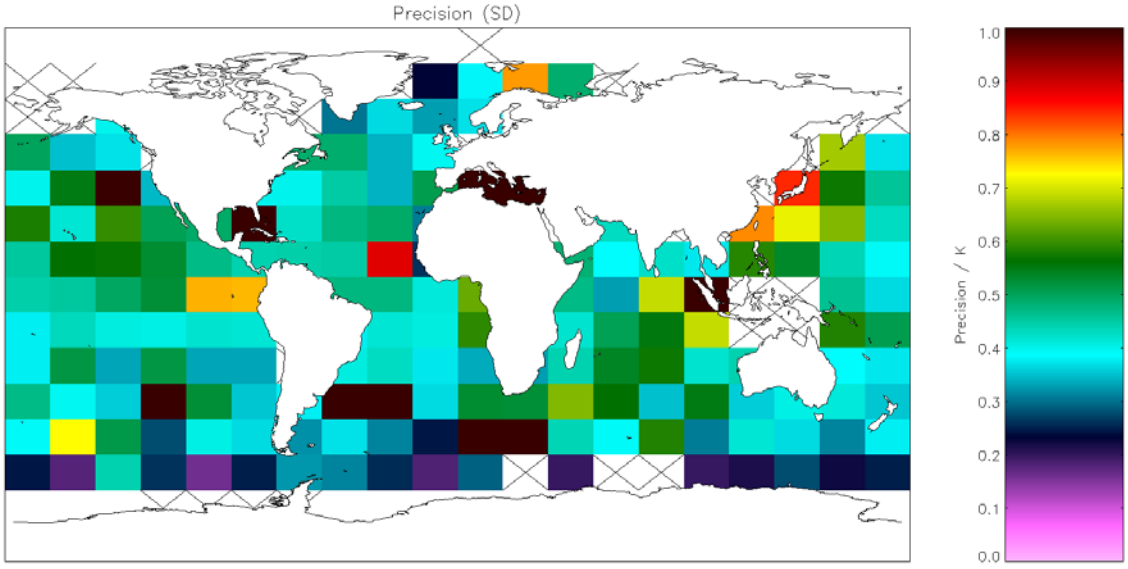


Figure A2.4.2-13: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.599 K (**Table 5.7.1**).

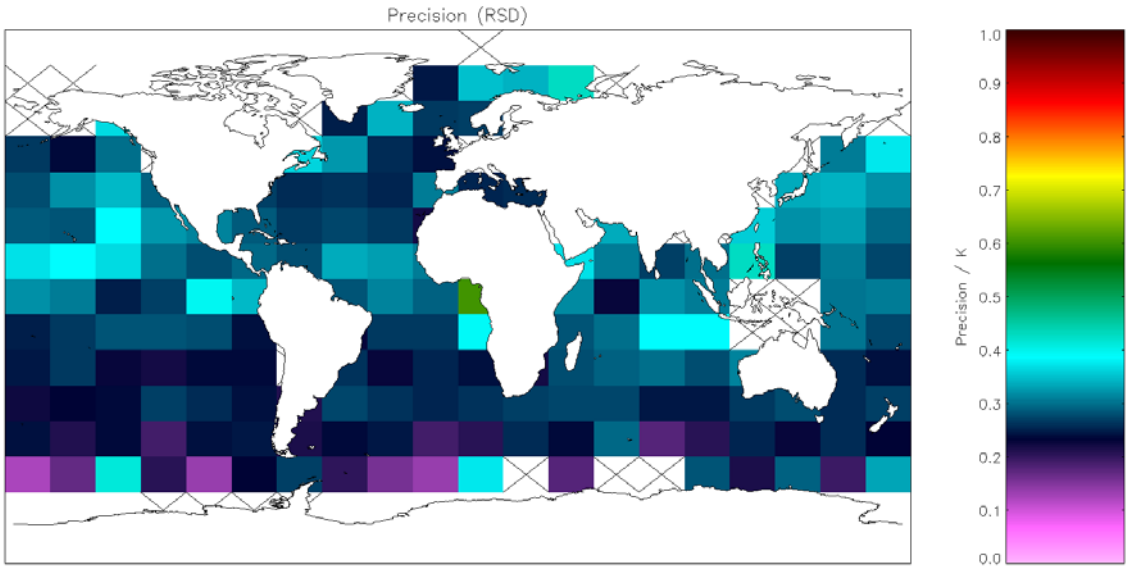


Figure A2.4.2-14: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.276 K (**Table 5.7.1**).

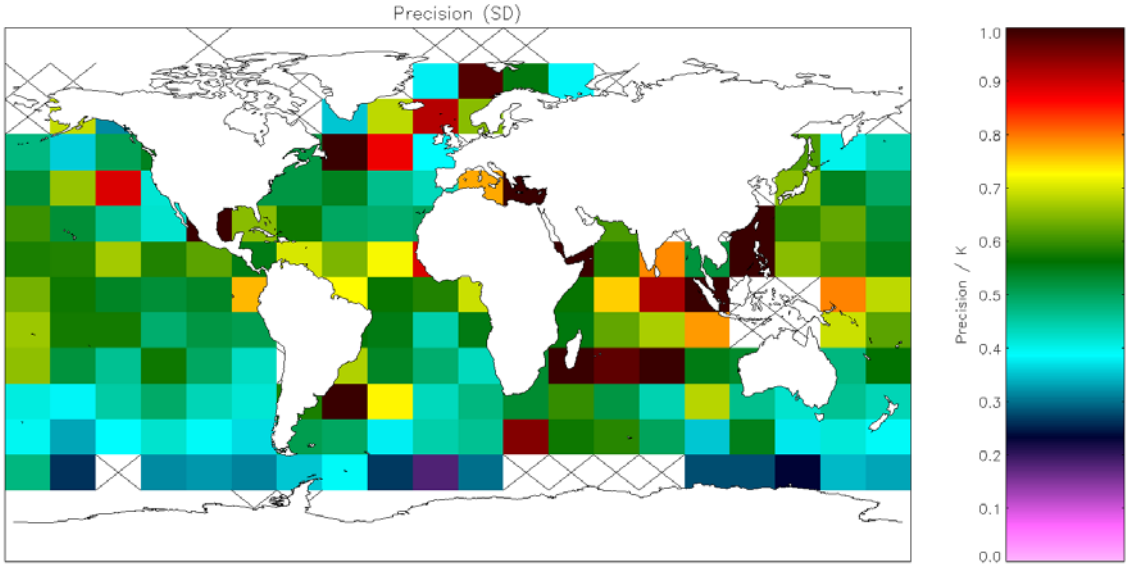


Figure A2.4.2-15: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.682 K (**Table 5.7.2**).

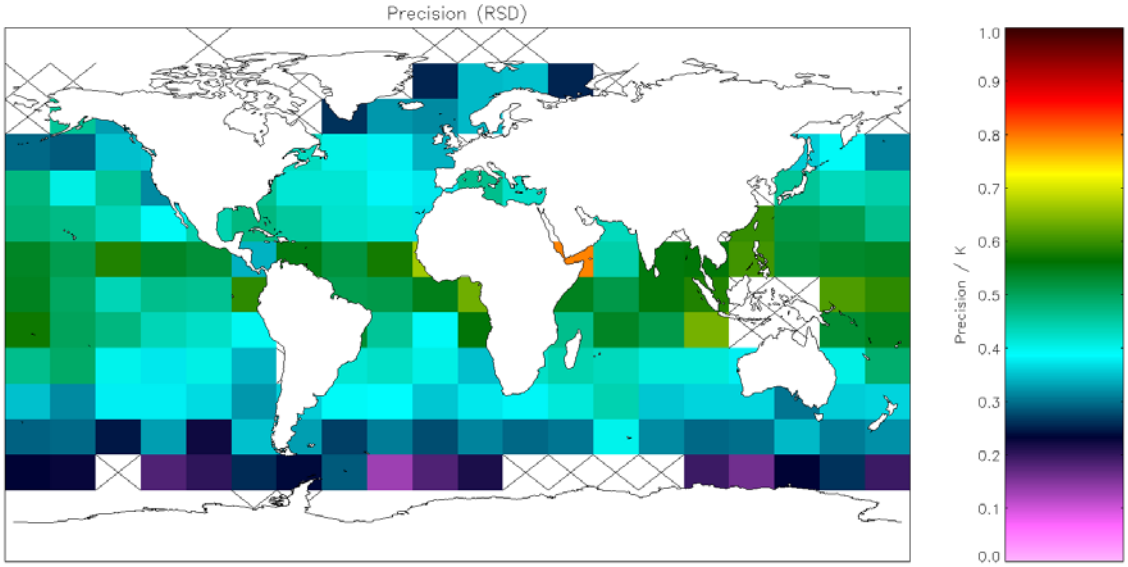


Figure A2.4.2-16: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.421 K (**Table 5.7.2**).

A2.4.3 Sensitivity

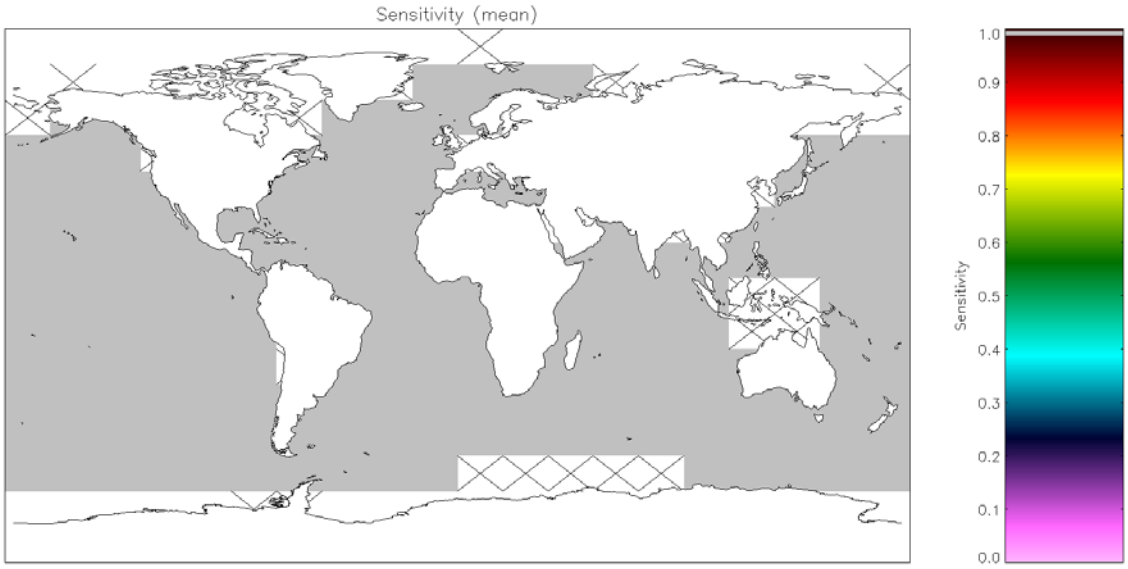


Figure A2.4.3-1: Global map of SST sensitivity using OE v2 retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.4.2**).

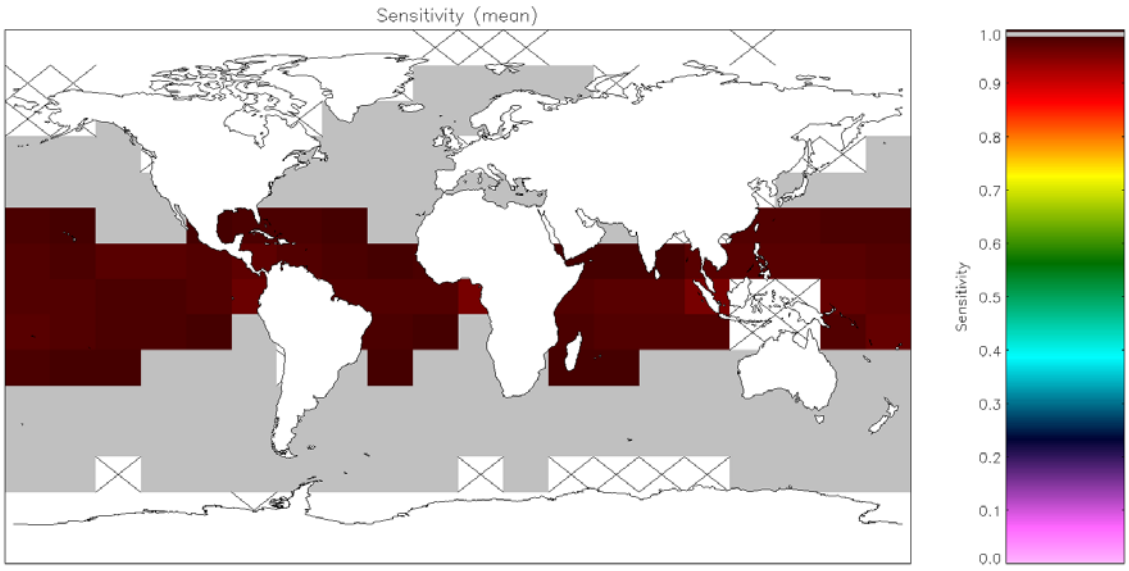


Figure A2.4.3-2: Global map of SST sensitivity using OE v2 retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.4.3**).

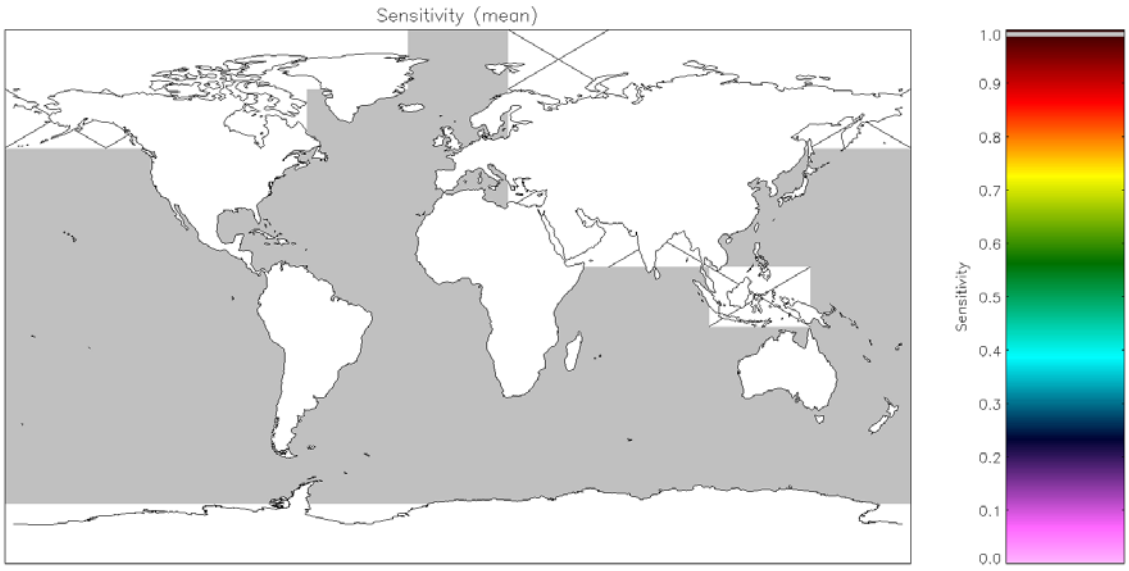


Figure A2.4.3-3: Global map of SST sensitivity using OE v2 retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.5.1**).

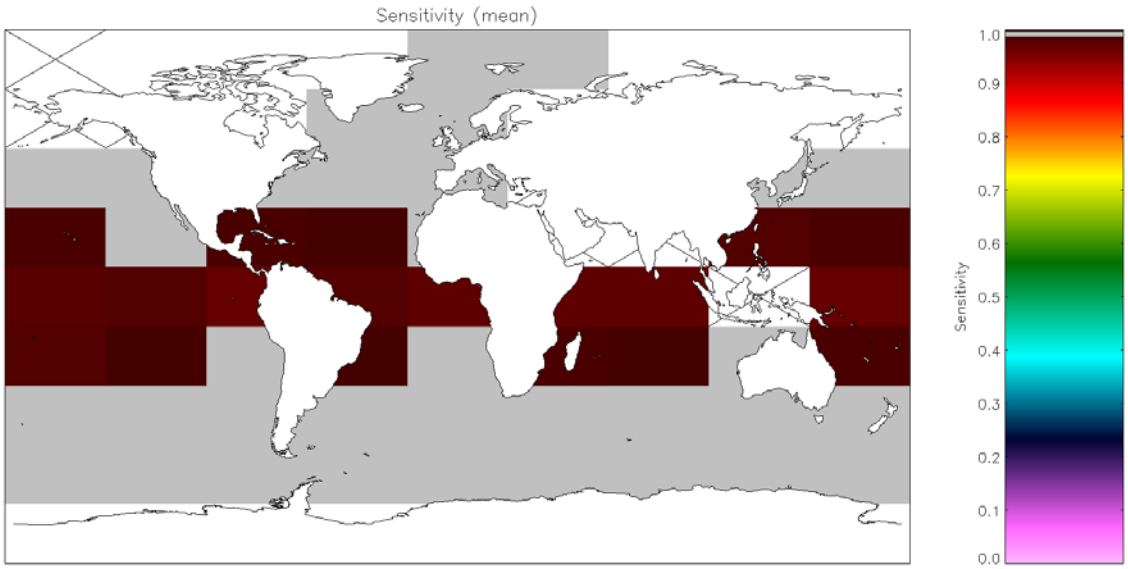


Figure A2.4.3-4: Global map of SST sensitivity using OE v2 retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.5.2**).

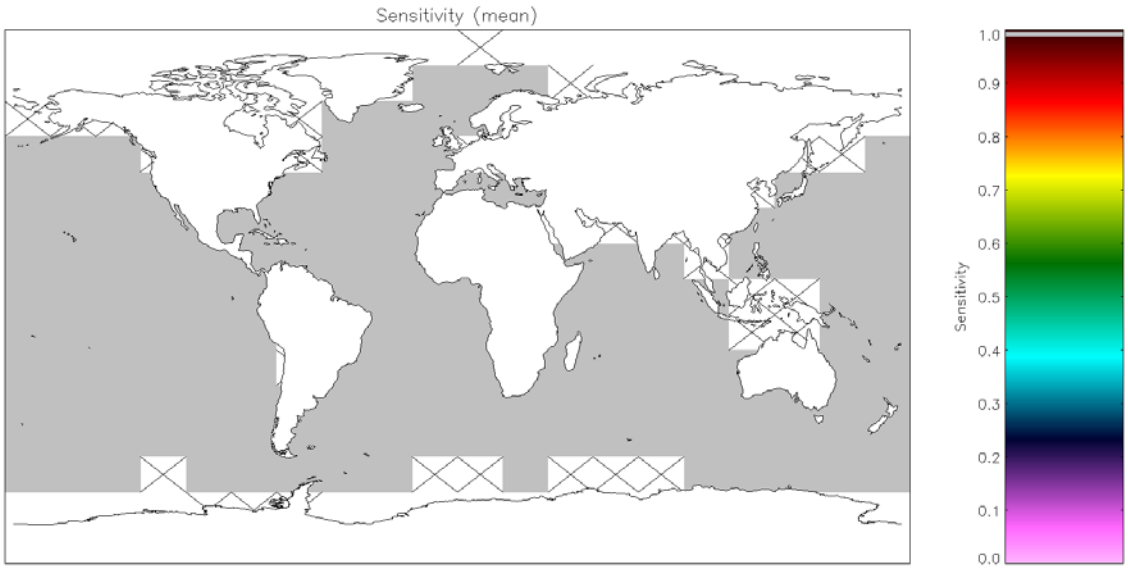


Figure A2.4.3-5: Global map of SST sensitivity using OE v2 retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.6.1**).

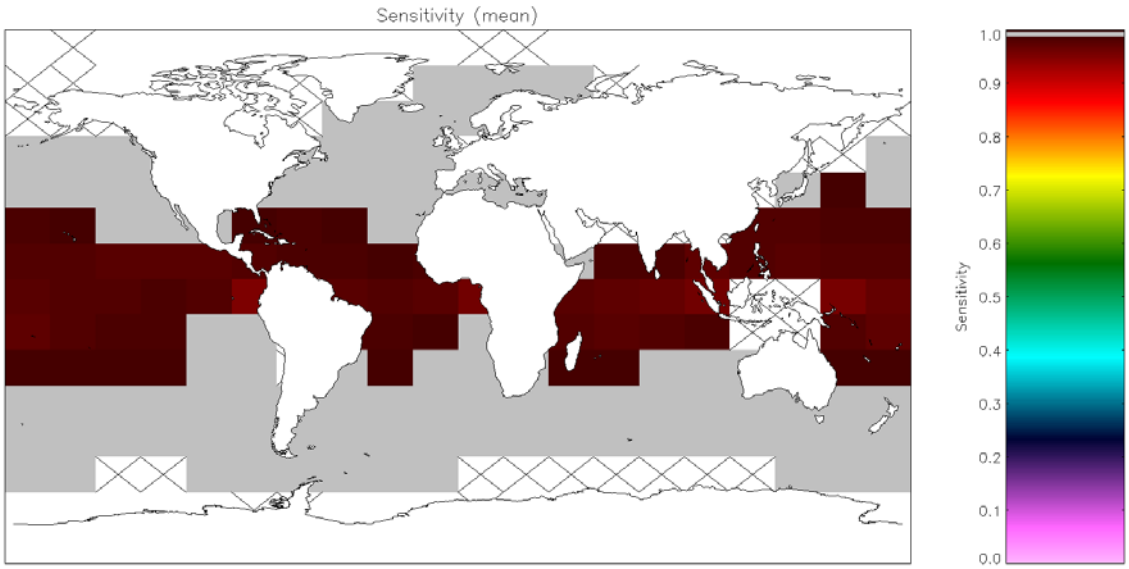


Figure A2.4.3-6: Global map of SST sensitivity using OE v2 retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.6.2**).

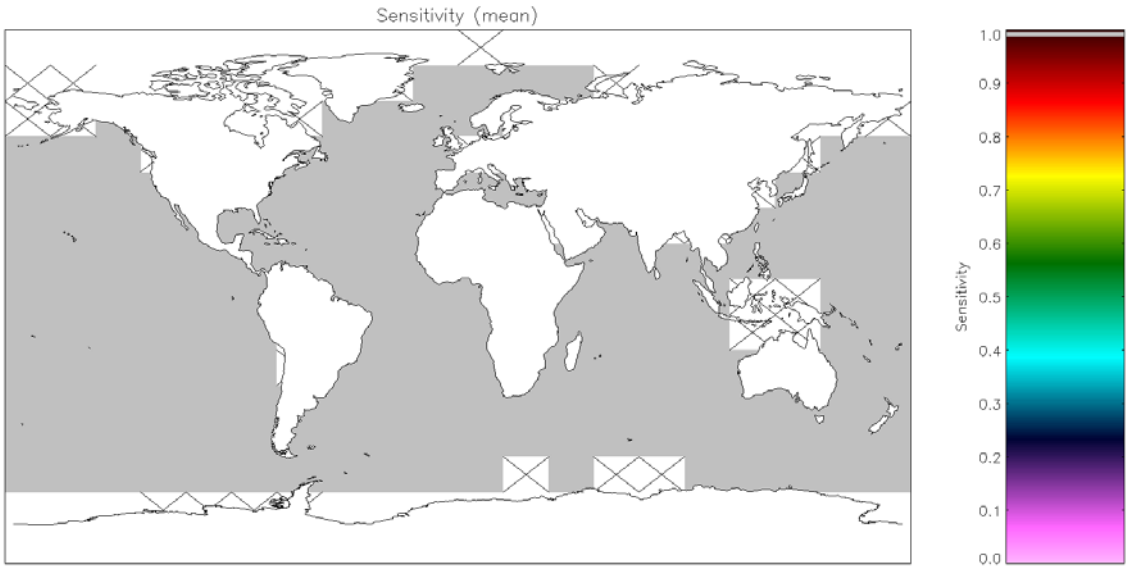


Figure A2.4.3-7: Global map of SST sensitivity using OE v2 retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.7.1**).

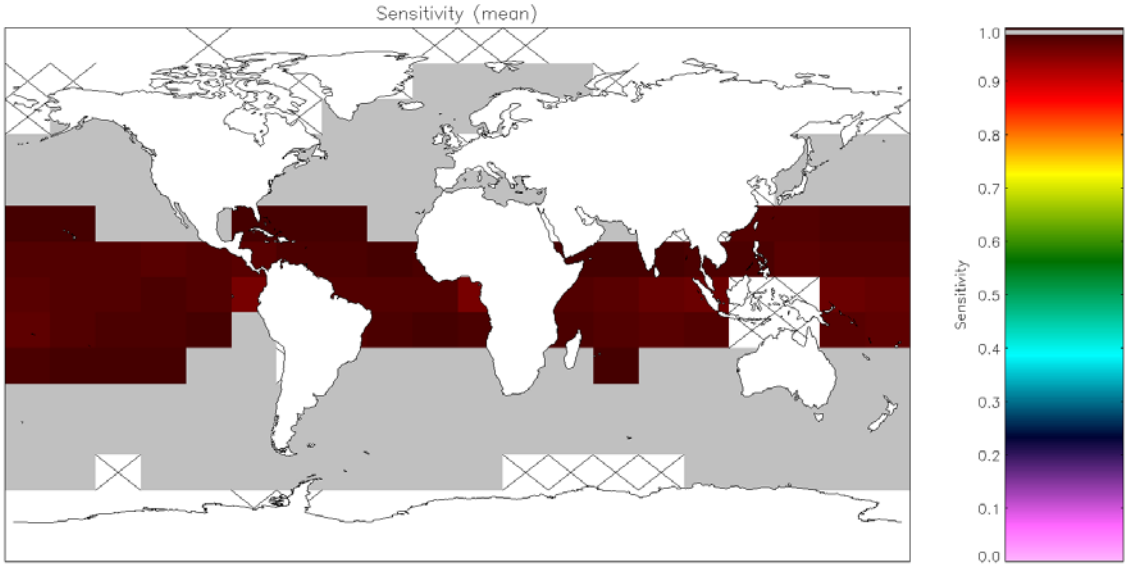


Figure A2.4.3-8: Global map of SST sensitivity using OE v2 retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.7.2**).

A2.5 Metrics for Incremental Regression Estimation

A2.5.1 Bias

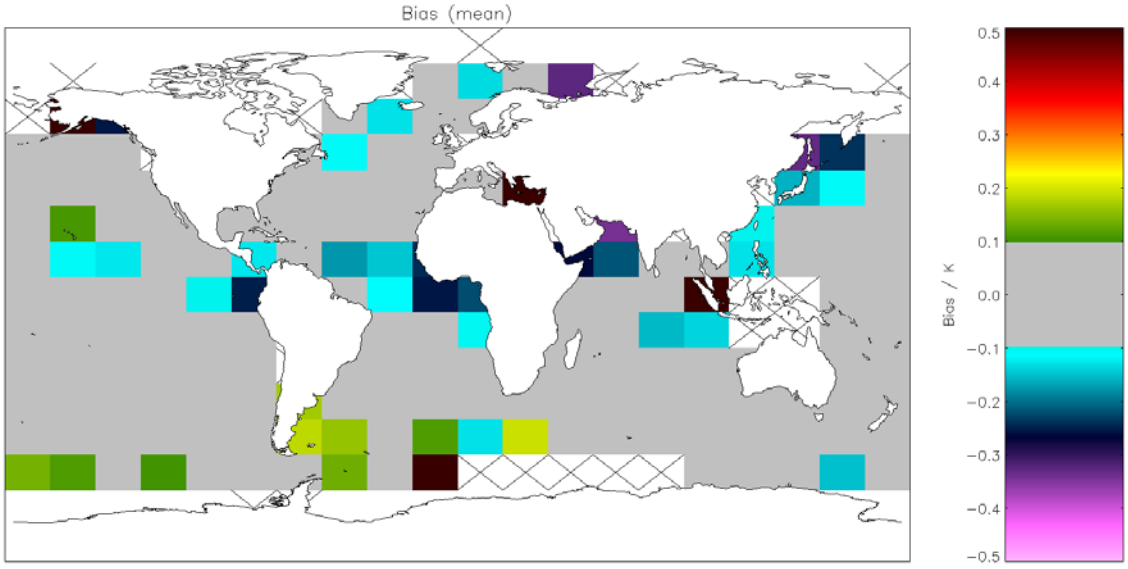


Figure A2.5.1-1: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.011 K (**Table 5.4.2**). See **Figure A2.5.1-17** for equivalent result with OE v2 QC filtering applied.

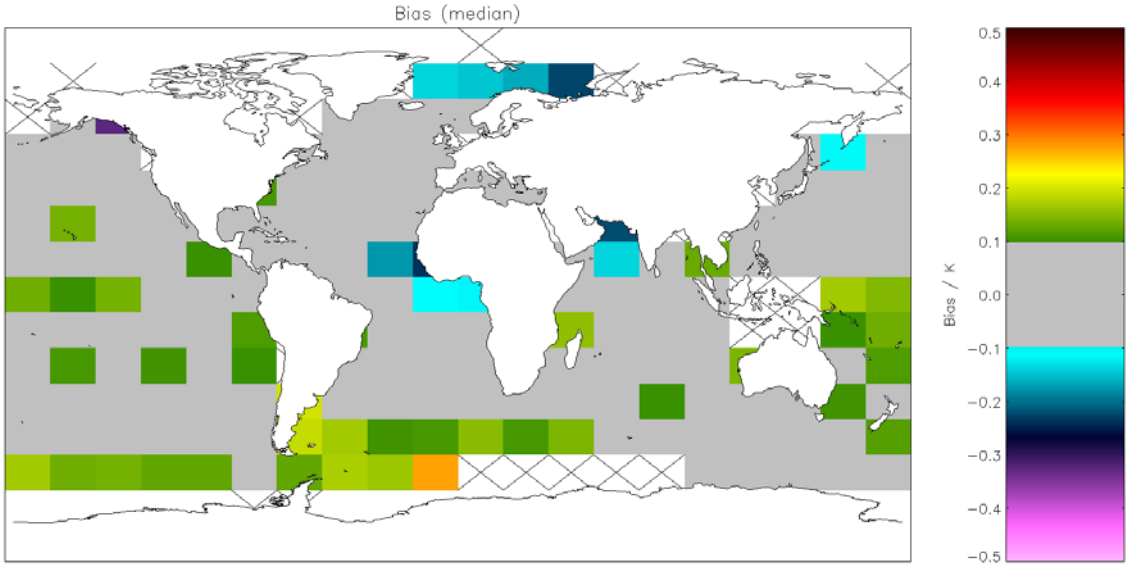


Figure A2.5.1-2: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.060 K (**Table 5.4.2**). See **Figure A2.5.1-18** for equivalent result with OE v2 QC filtering applied.

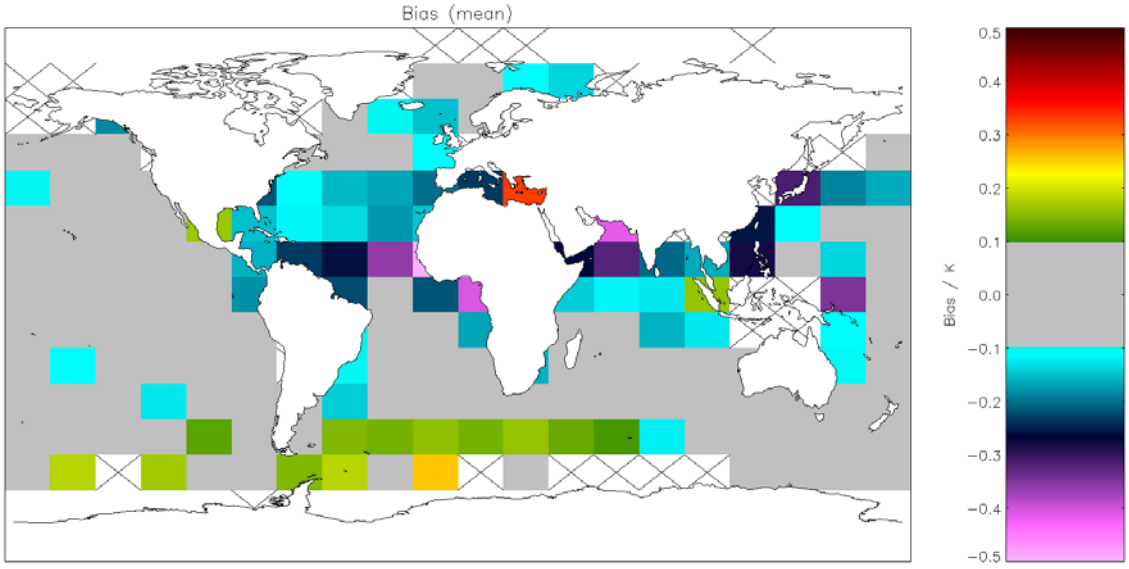


Figure A2.5.1-3: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.079 K (**Table 5.4.3**).

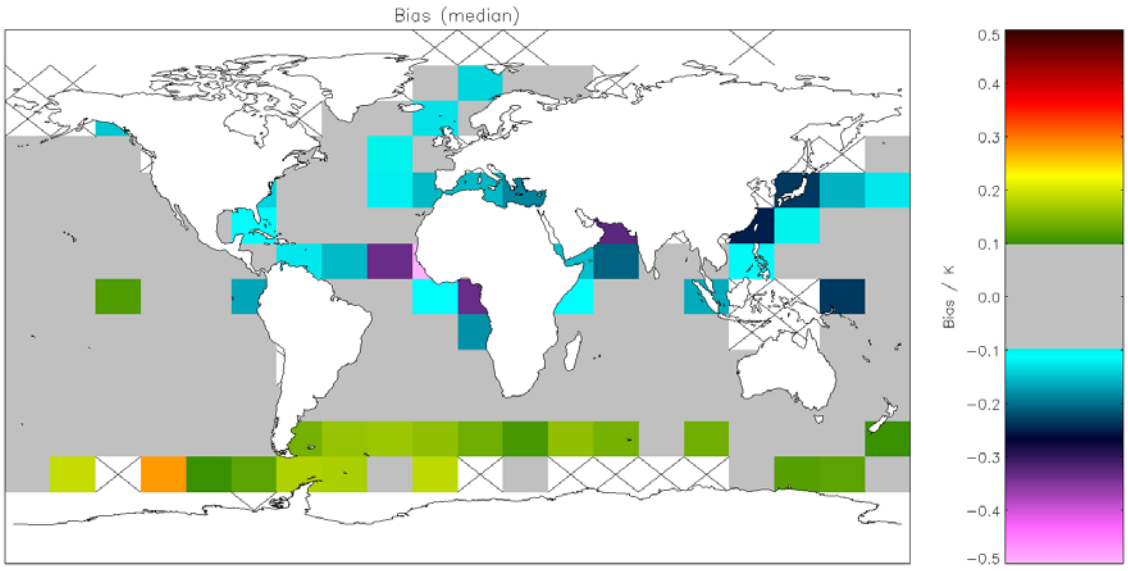


Figure A2.5.1-4: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.036 K (**Table 5.4.3**).

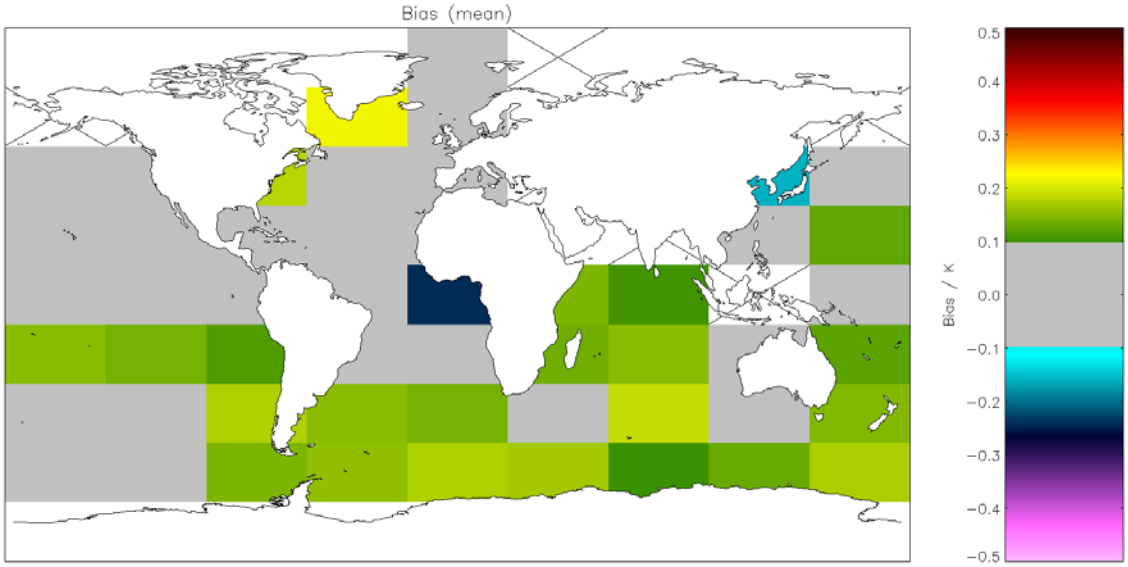


Figure A2.5.1-5: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.072 K (**Table 5.5.1**).

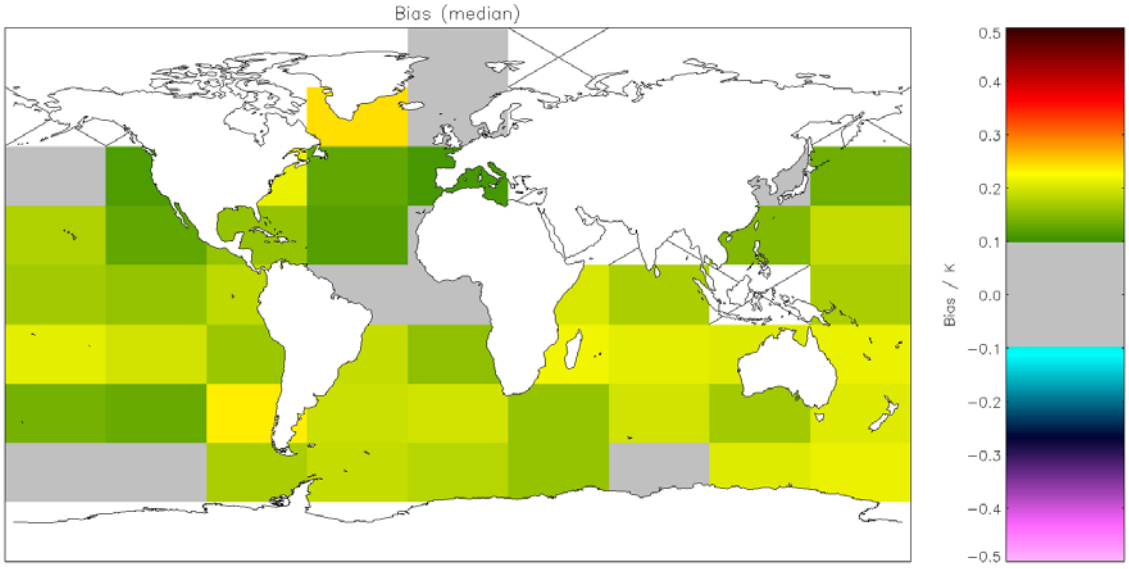


Figure A2.5.1-6: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.155 K (**Table 5.5.1**).

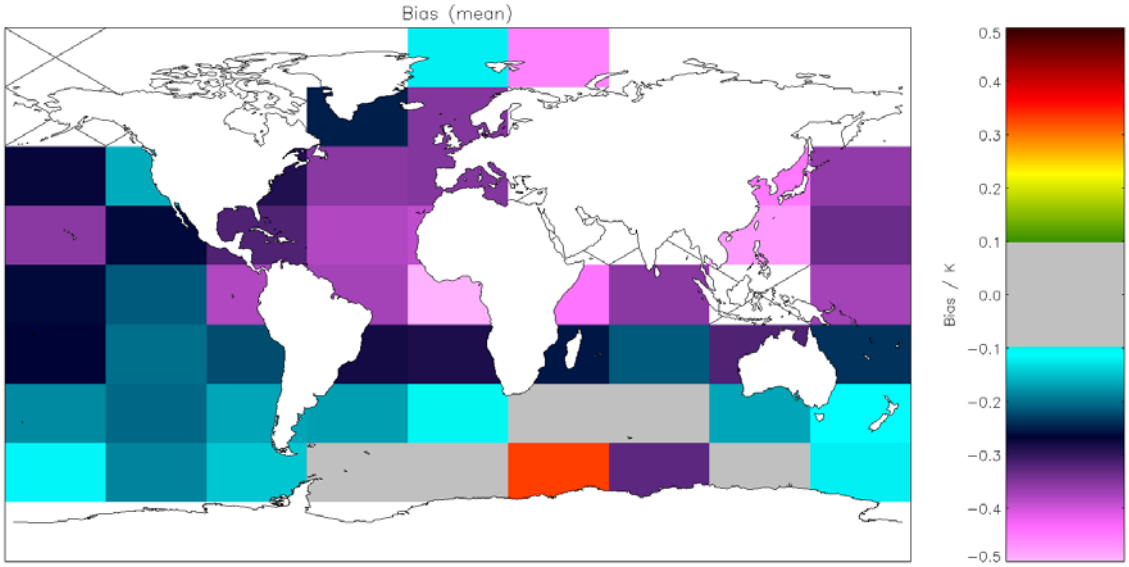


Figure A2.5.1-7: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.295 K (**Table 5.5.2**).

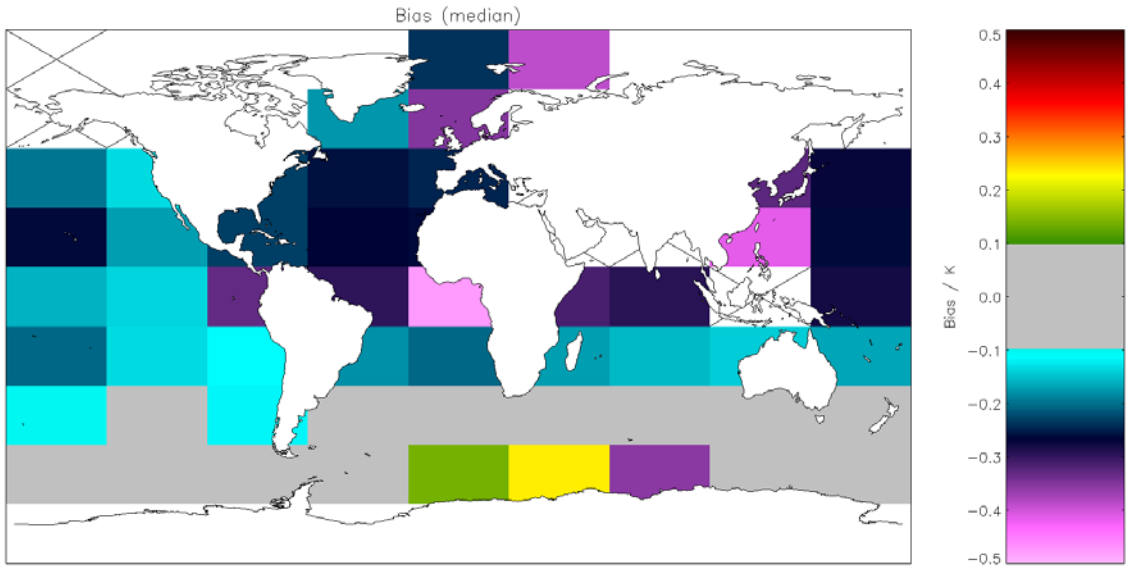


Figure A2.5.1-8: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.206 K (**Table 5.5.2**).

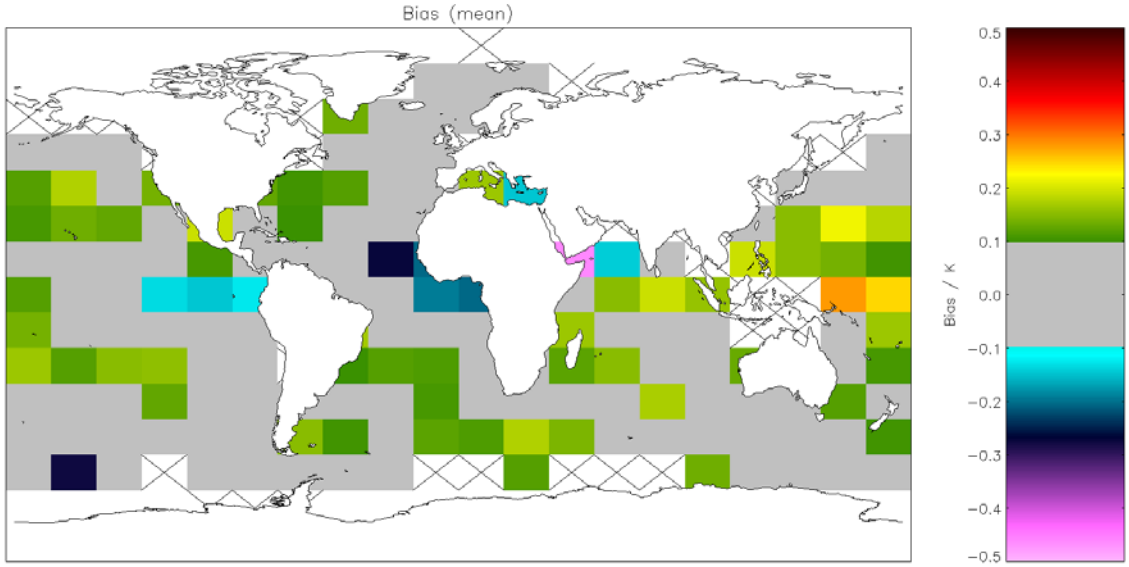


Figure A2.5.1-9: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.071 K (**Table 5.6.1**).

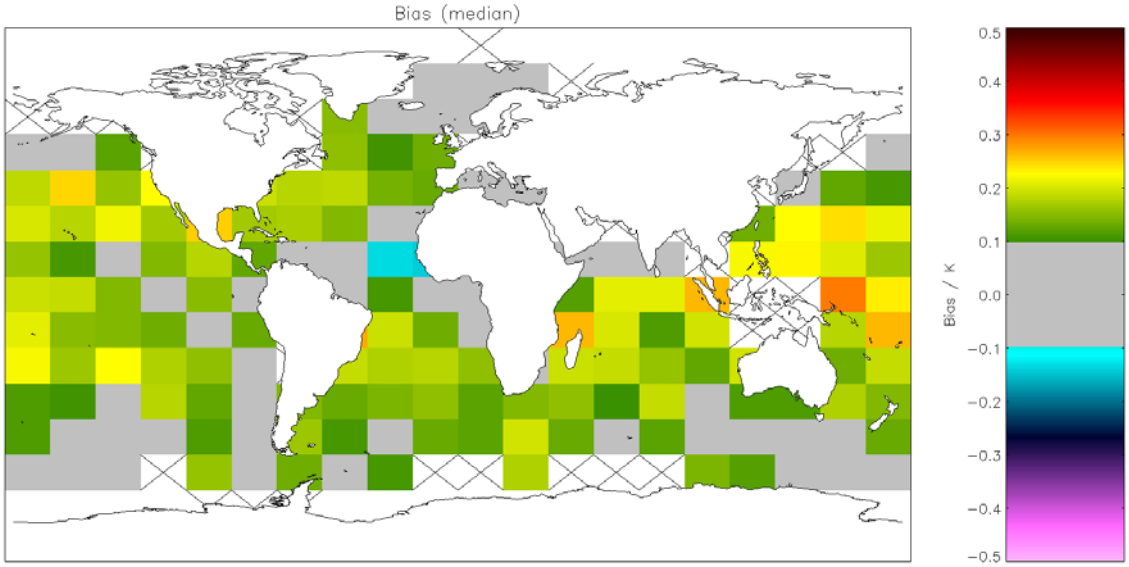


Figure A2.5.1-10: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.148 K (**Table 5.6.1**).

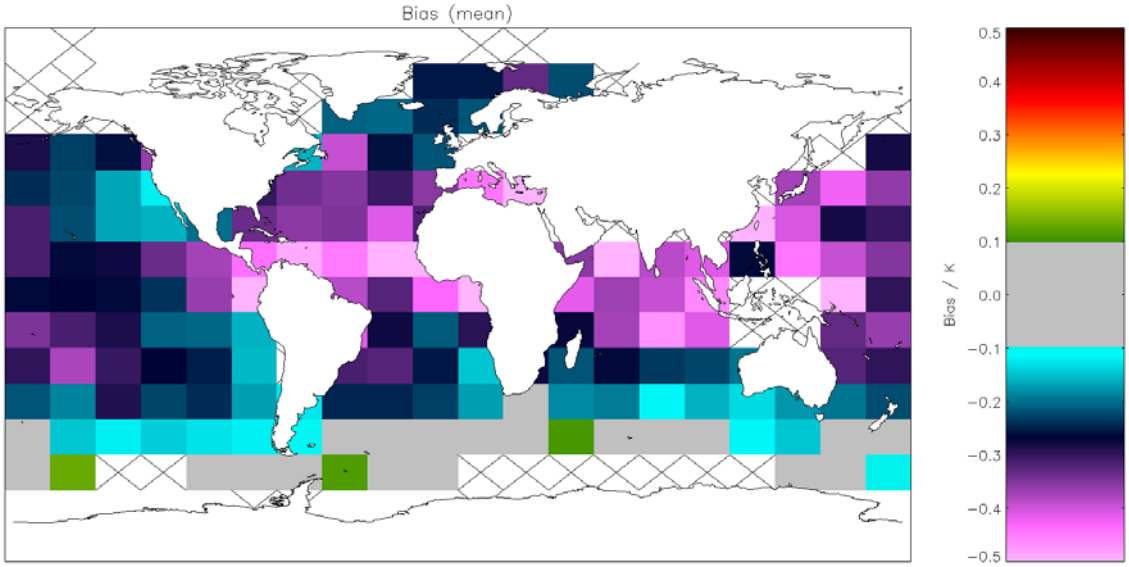


Figure A2.5.1-11: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.291 K (**Table 5.6.2**).

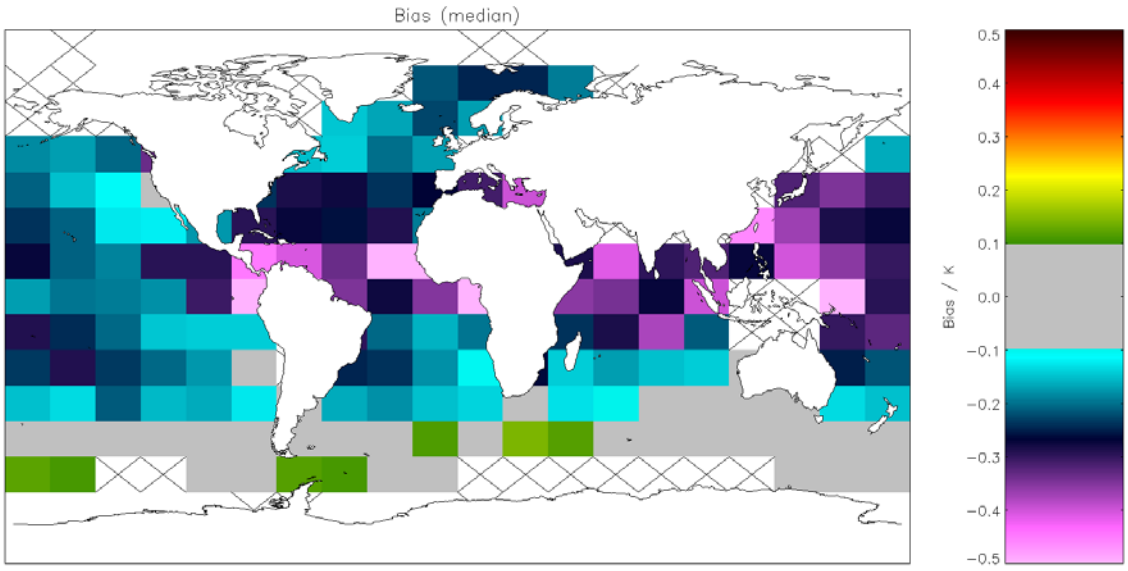


Figure A2.5.1-12: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.214 K (**Table 5.6.2**).

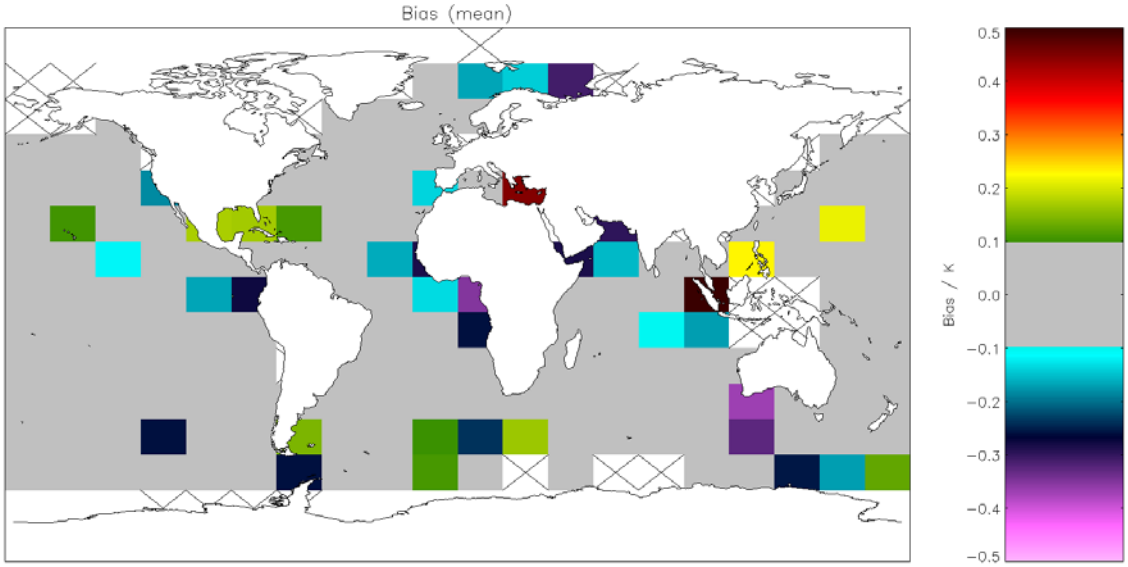


Figure A2.5.1-13: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.007 K (**Table 5.7.1**).

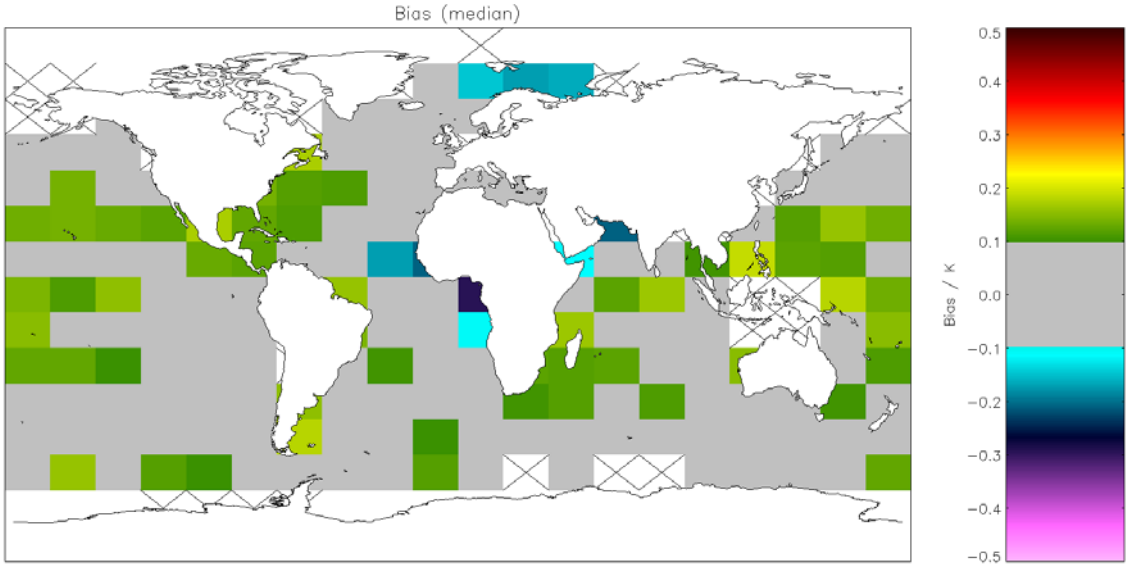


Figure A2.5.1-14: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.080 K (**Table 5.7.1**).

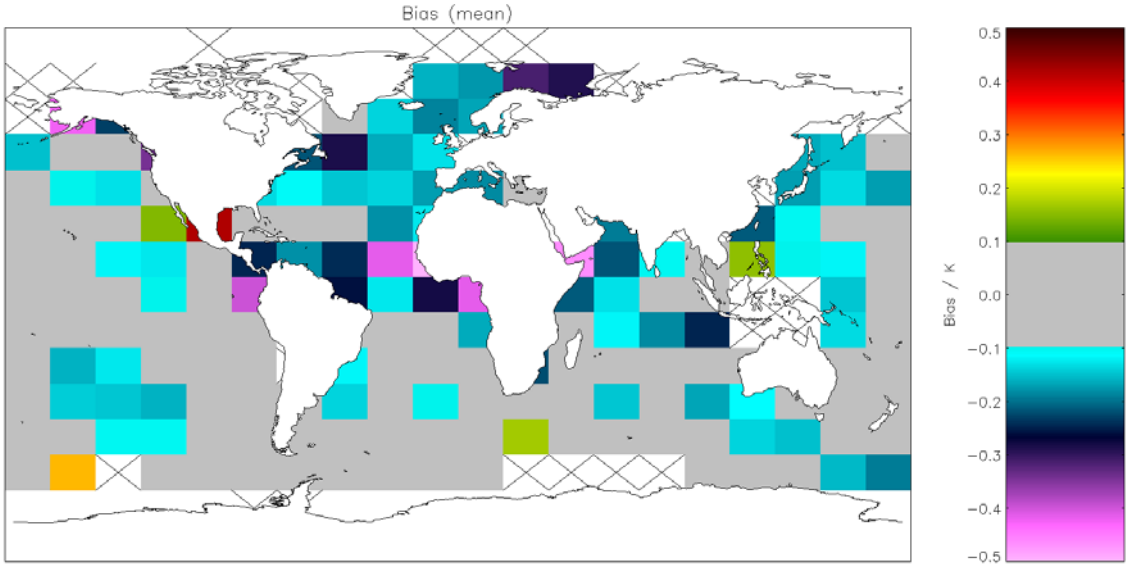


Figure A2.5.1-15: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.091 K (**Table 5.7.2**).

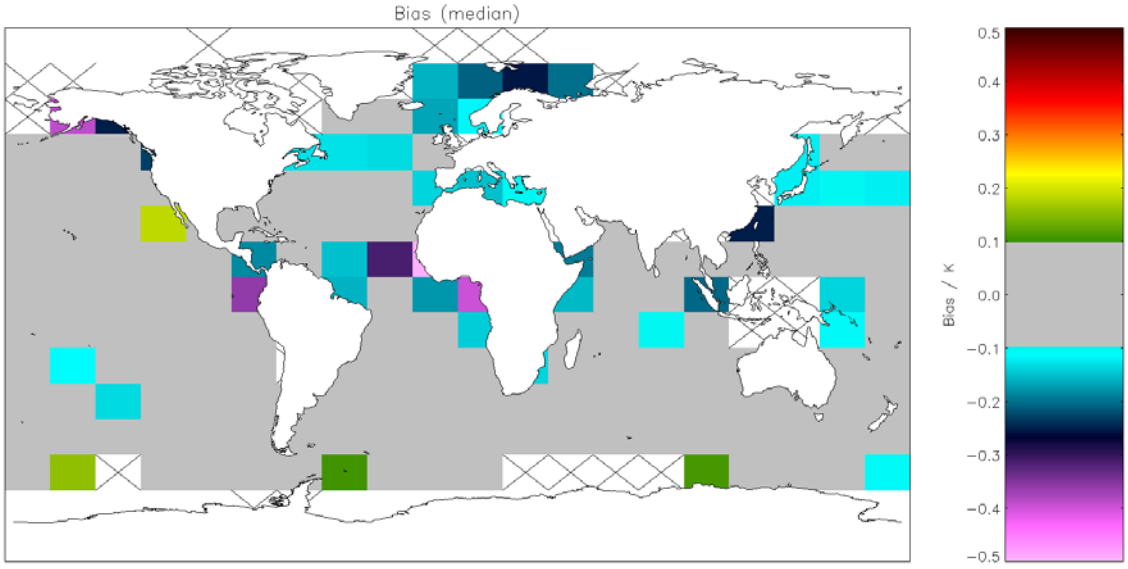


Figure A2.5.1-16: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.048 K (**Table 5.7.2**).

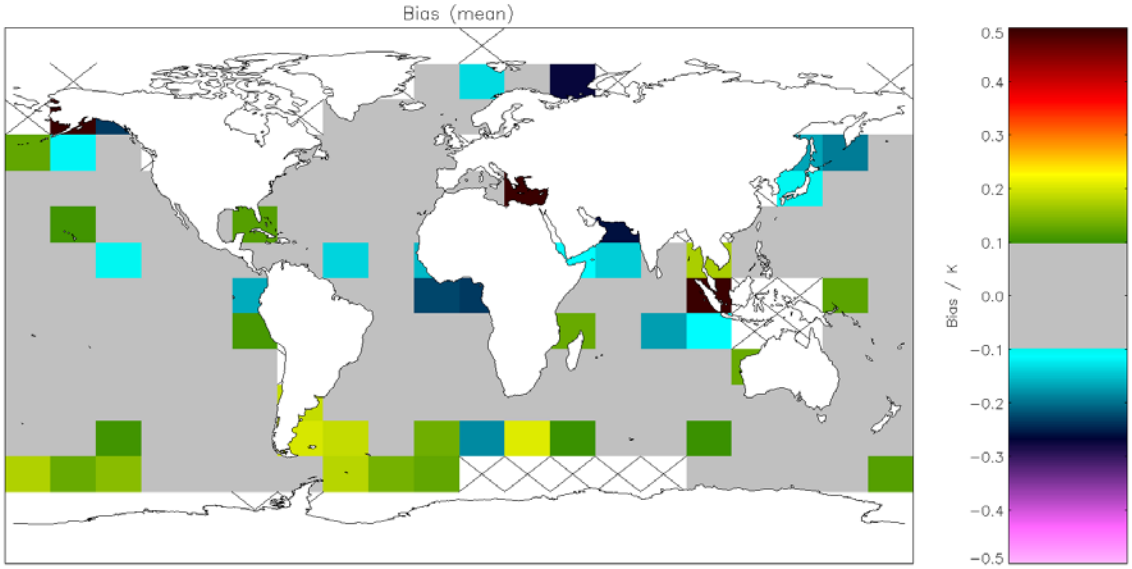


Figure A2.5.1-17: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.013 K (**Table 5.4.2**). As **Figure A2.5.1-1** but with OE v2 QC filtering applied.

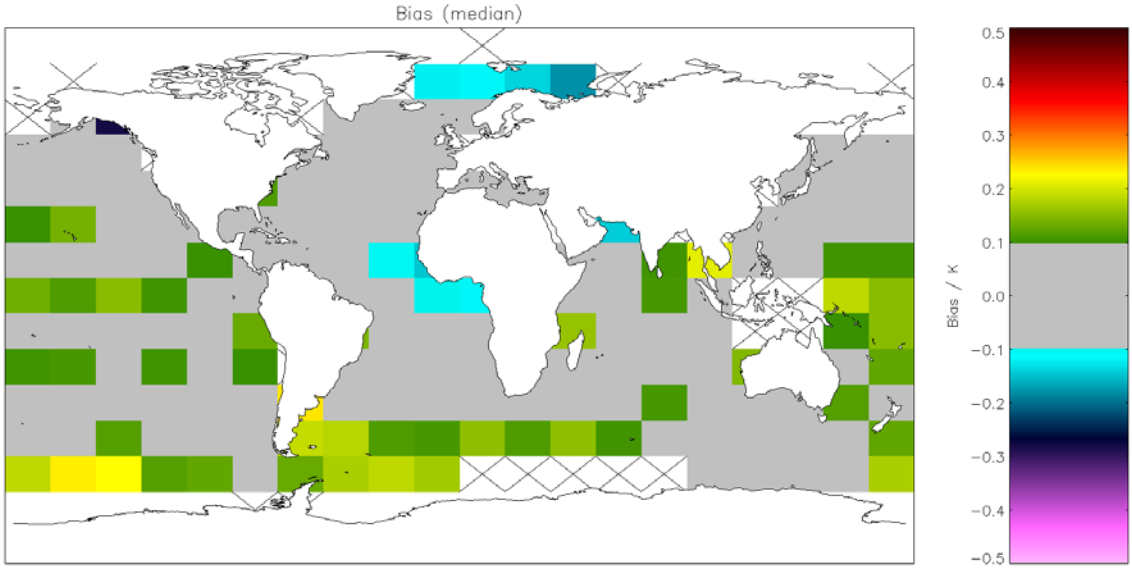


Figure A2.5.1-18: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.070 K (**Table 5.4.2**). As **Figure A2.5.1-2** but with OE v2 QC filtering applied.

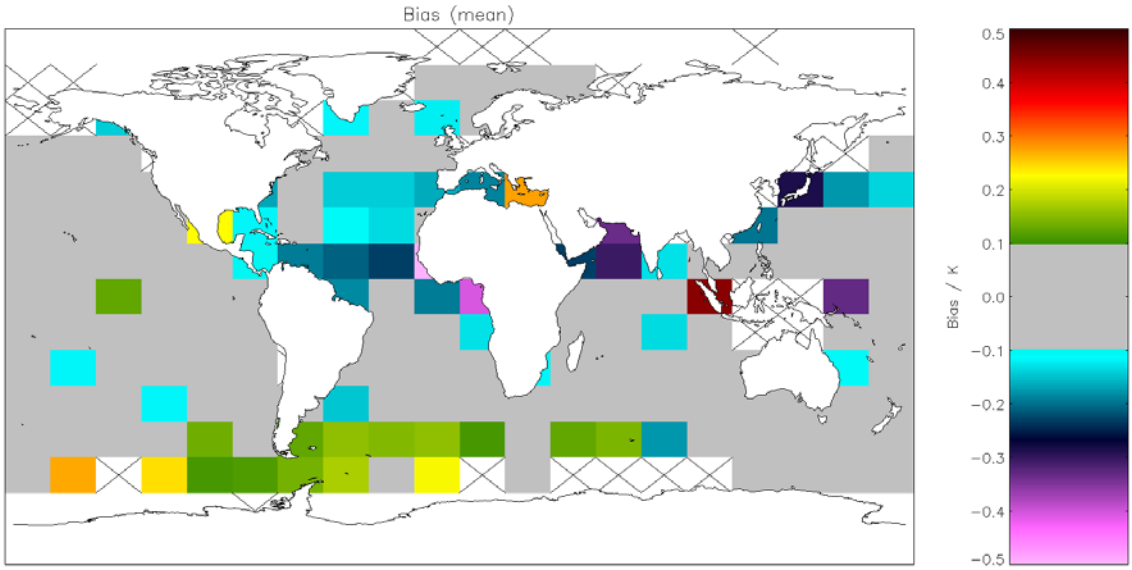


Figure A2.5.1-19: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.055 K (**Table 5.4.3**). As **Figure A2.5.1-3** but with OE v2 QC filtering applied.

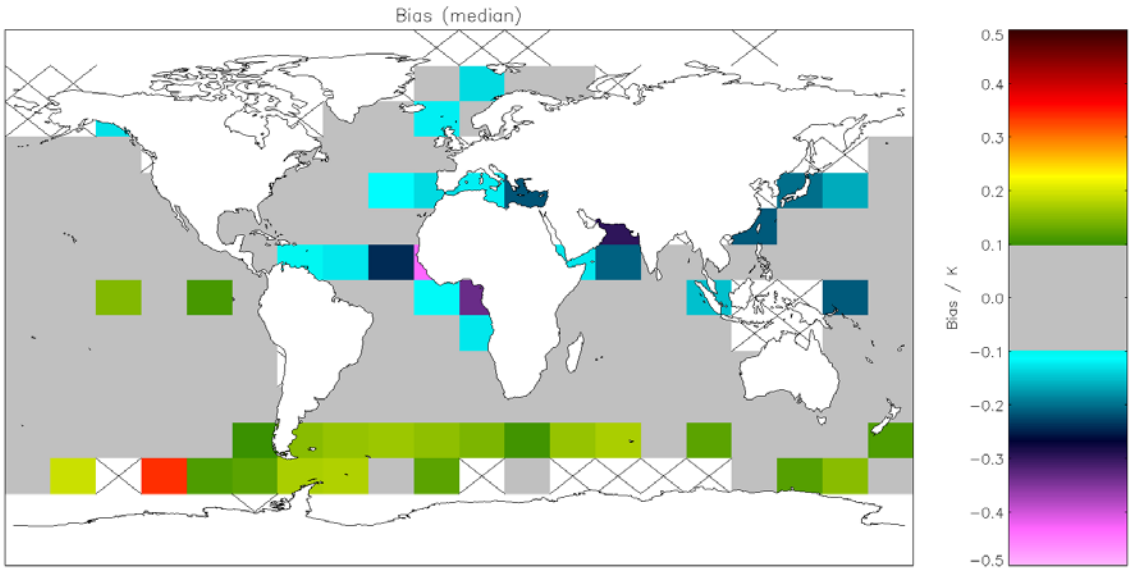


Figure A2.5.1-20: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.021 K (**Table 5.4.3**). As **Figure A2.5.1-4** but with OE v2 QC filtering applied.

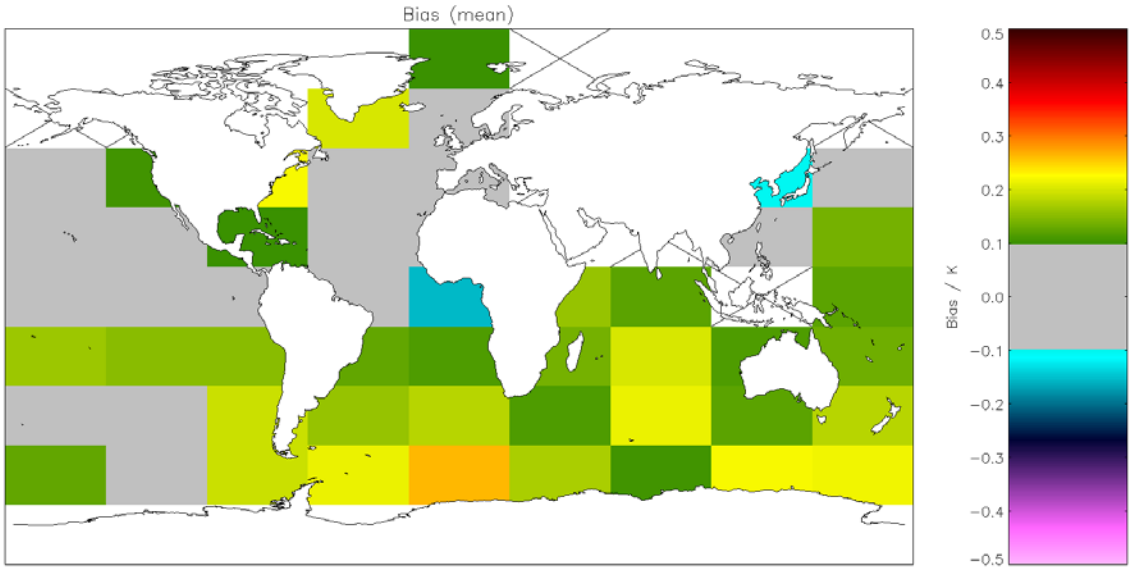


Figure A2.5.1-21: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.104 K (**Table 5.5.1**). As **Figure A2.5.1-5** but with OE v2 QC filtering applied.

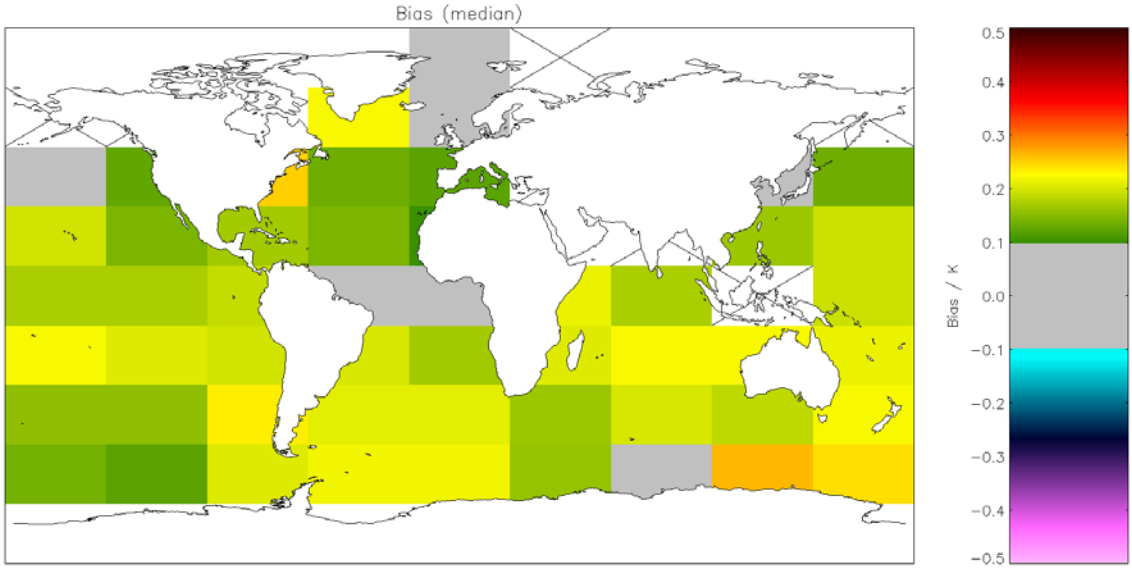


Figure A2.5.1-22: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.170 K (**Table 5.5.1**). As **Figure A2.5.1-6** but with OE v2 QC filtering applied.

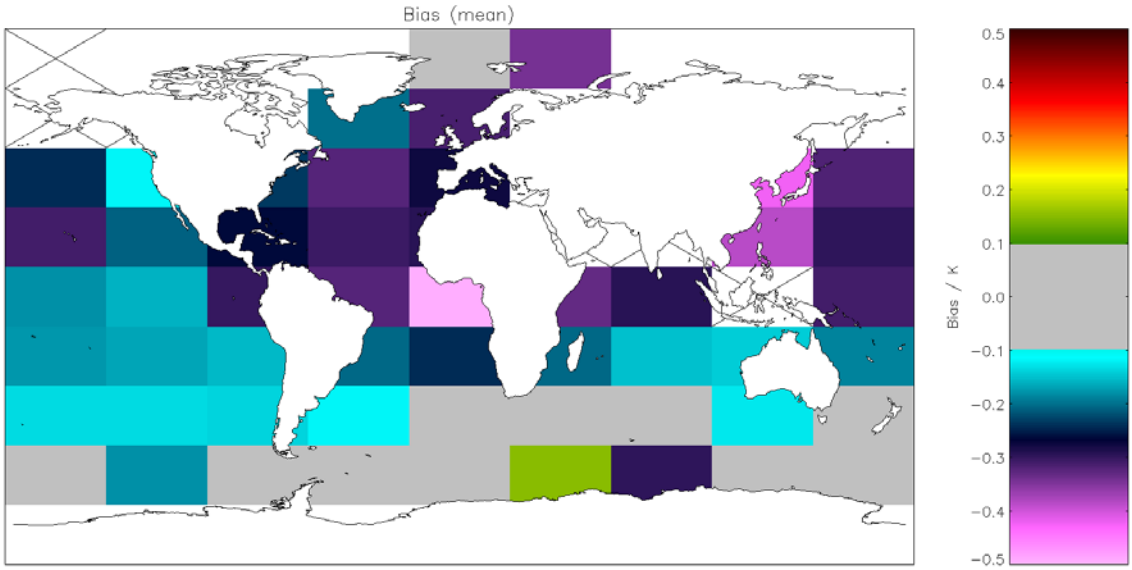


Figure A2.5.1-23: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.232 K (**Table 5.5.2**). As **Figure A2.5.1-7** but with OE v2 QC filtering applied.

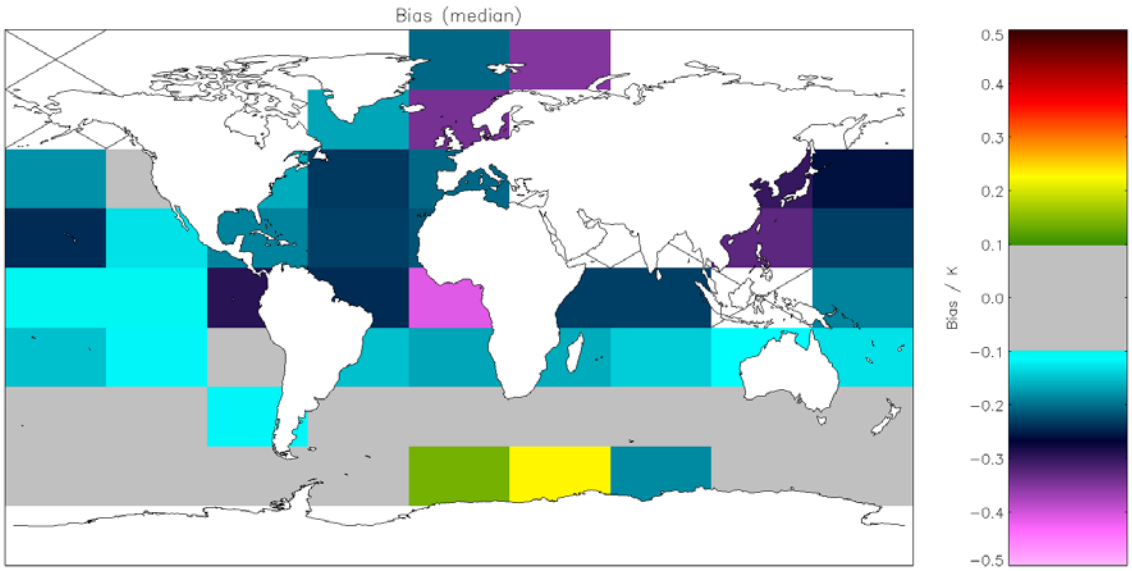


Figure A2.5.1-24: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.169 K (**Table 5.5.2**). As **Figure A2.5.1-8** but with OE v2 QC filtering applied.

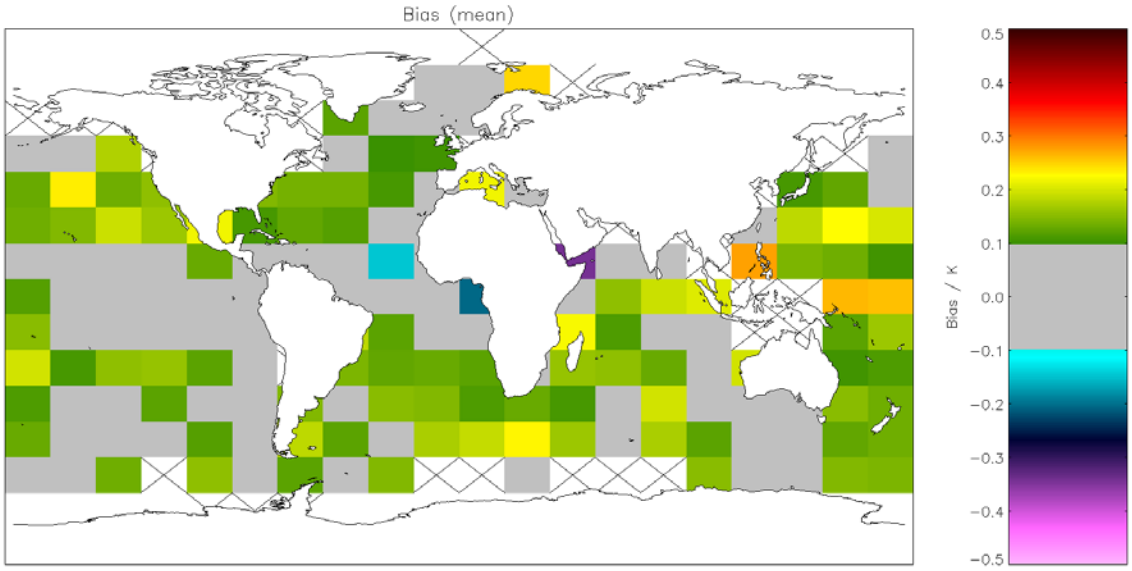


Figure A2.5.1-25: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.107 K (**Table 5.6.1**). As **Figure A2.5.1-9** but with OE v2 QC filtering applied.

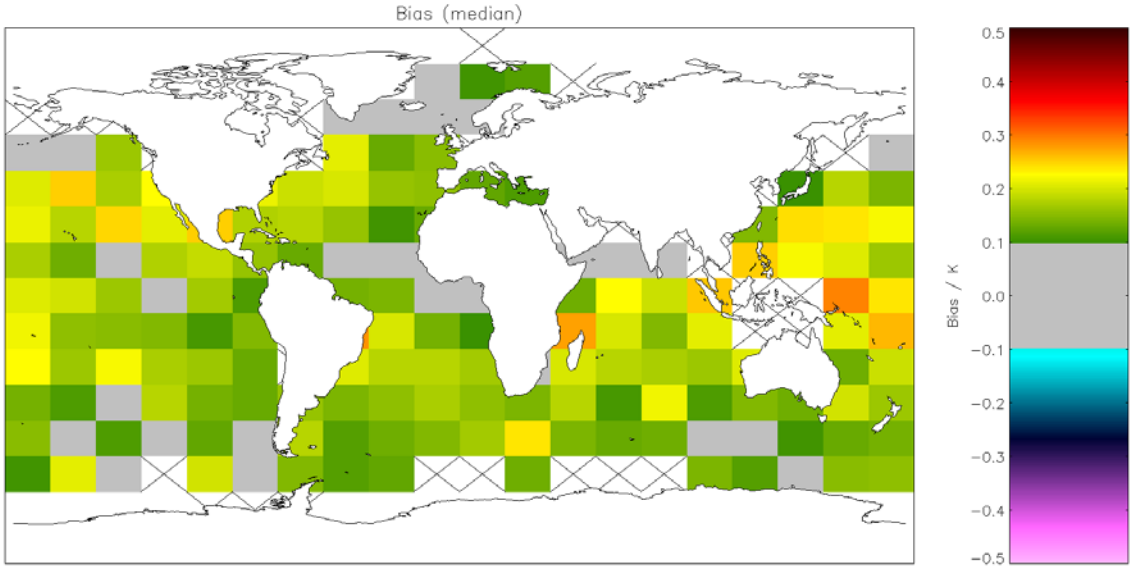


Figure A2.5.1-26: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.163 K (**Table 5.6.1**). As **Figure A2.5.1-10** but with OE v2 QC filtering applied.

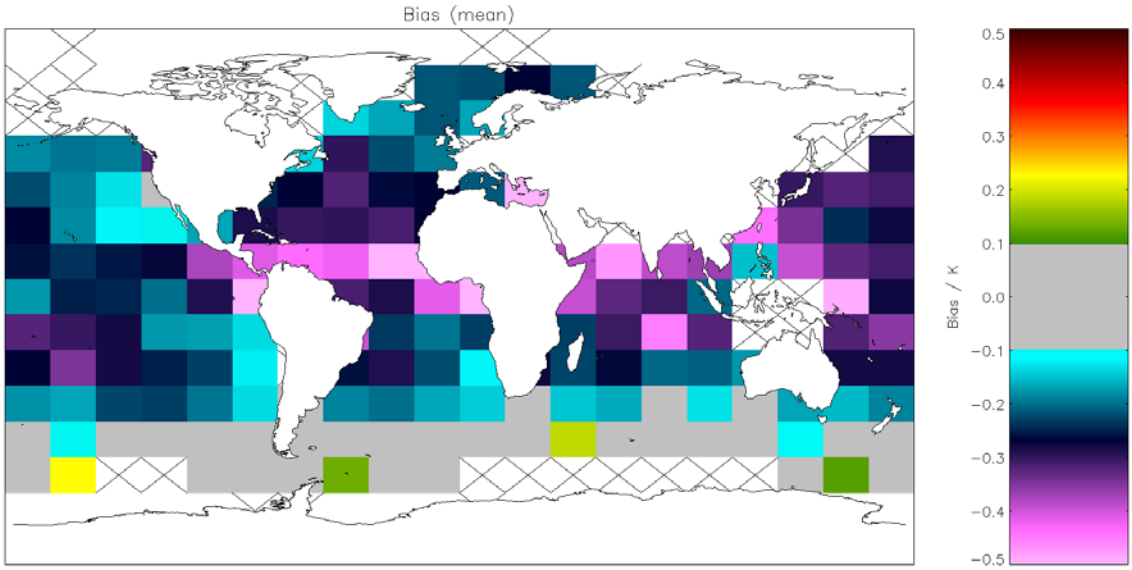


Figure A2.5.1-27: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.241 K (**Table 5.6.2**). As **Figure A2.5.1-11** but with OE v2 QC filtering applied.

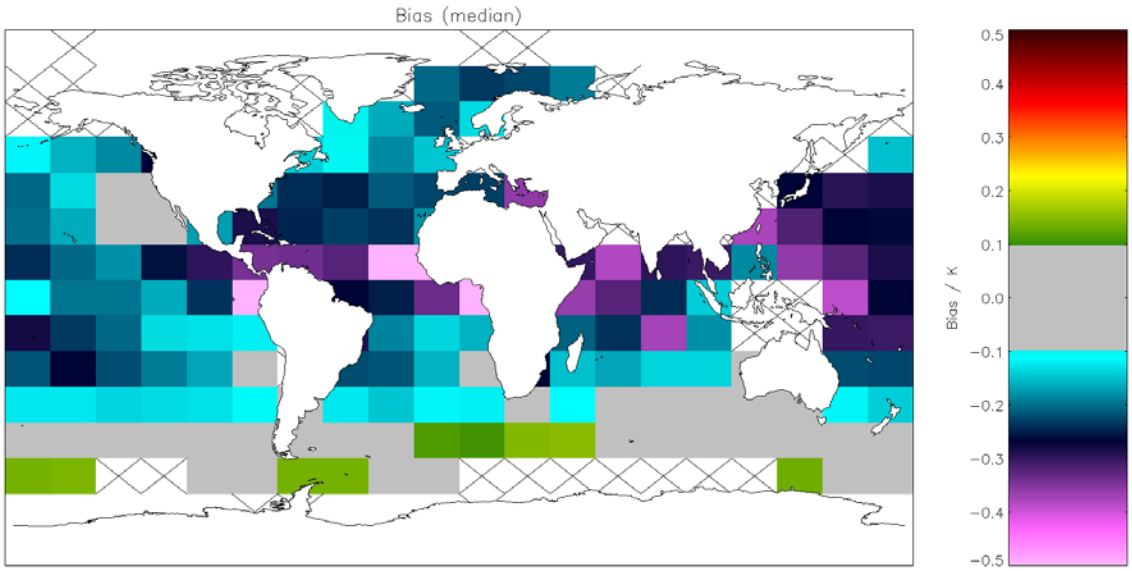


Figure A2.5.1-28: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.184 K (**Table 5.6.2**). As **Figure A2.5.1-12** but with OE v2 QC filtering applied.

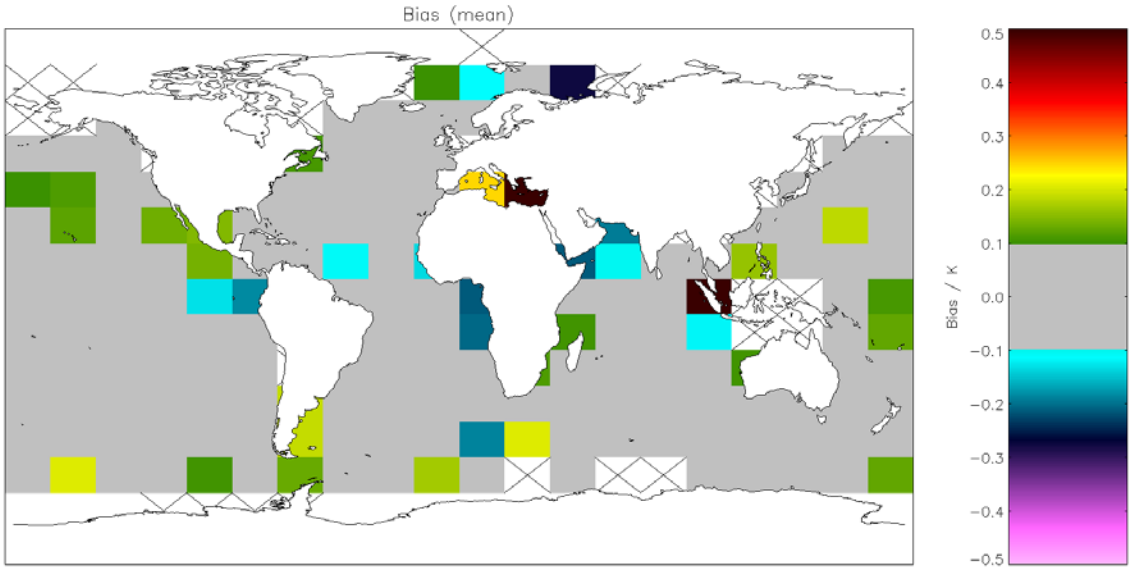


Figure A2.5.1-29: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.035 K (**Table 5.7.1**). As **Figure A2.5.1-13** but with OE v2 QC filtering applied.

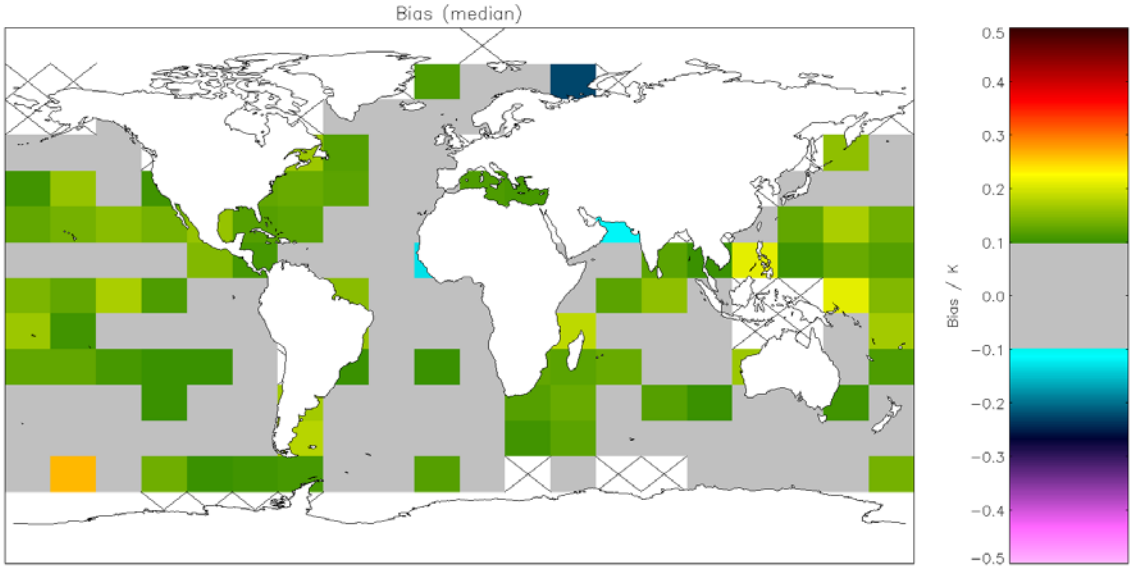


Figure A2.5.1-30: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.092 K (**Table 5.7.1**). As **Figure A2.5.1-14** but with OE v2 QC filtering applied.

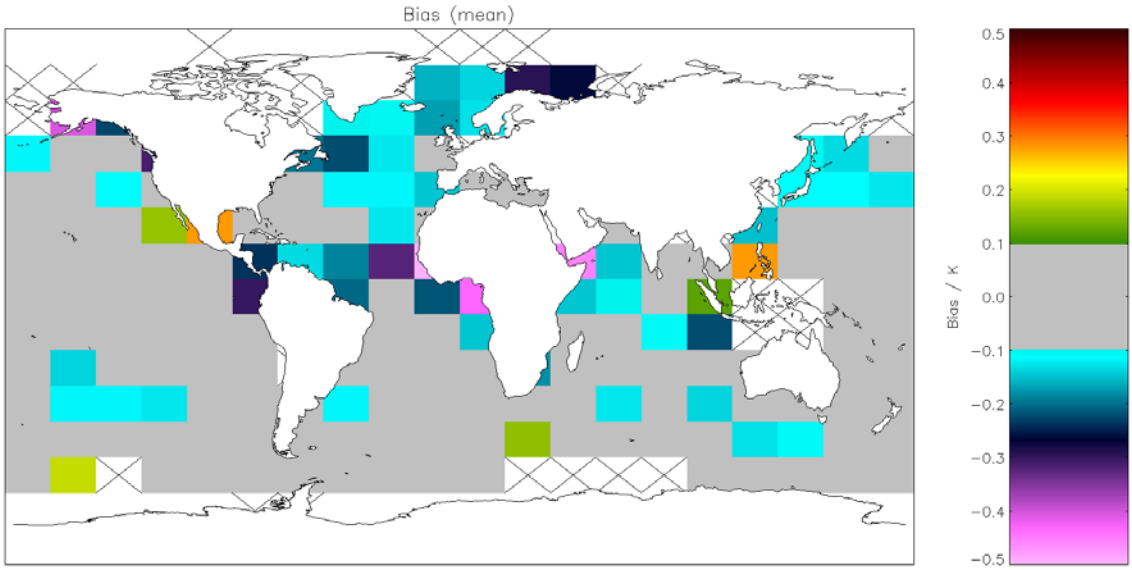


Figure A2.5.1-31: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.059 K (**Table 5.7.2**). As **Figure A2.5.1-15** but with OE v2 QC filtering applied.

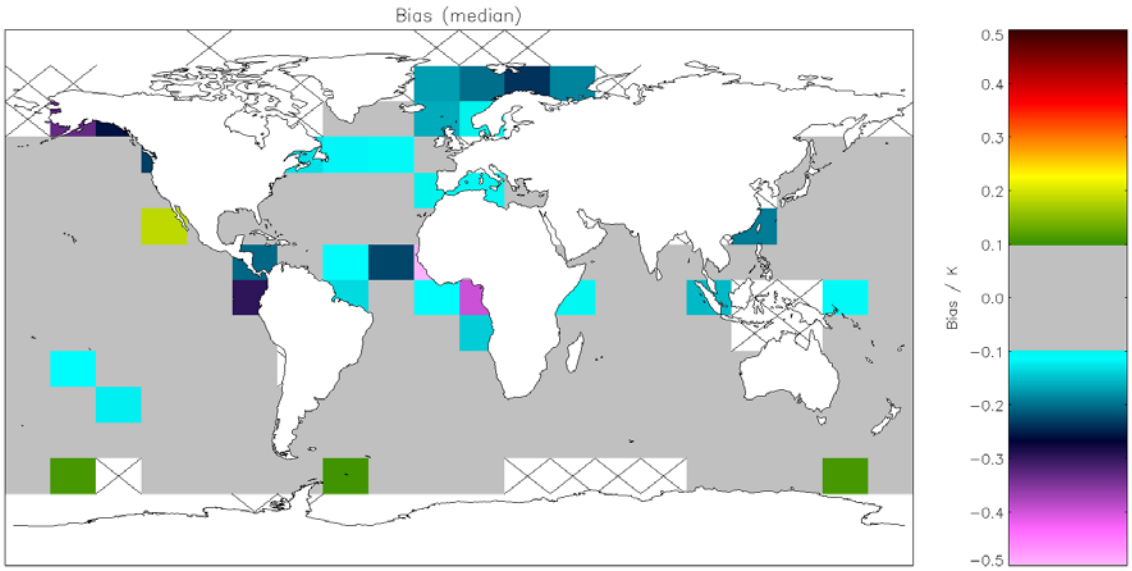


Figure A2.5.1-32: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.030 K (**Table 5.7.2**). As **Figure A2.5.1-16** but with OE v2 QC filtering applied.

A2.5.2 Precision

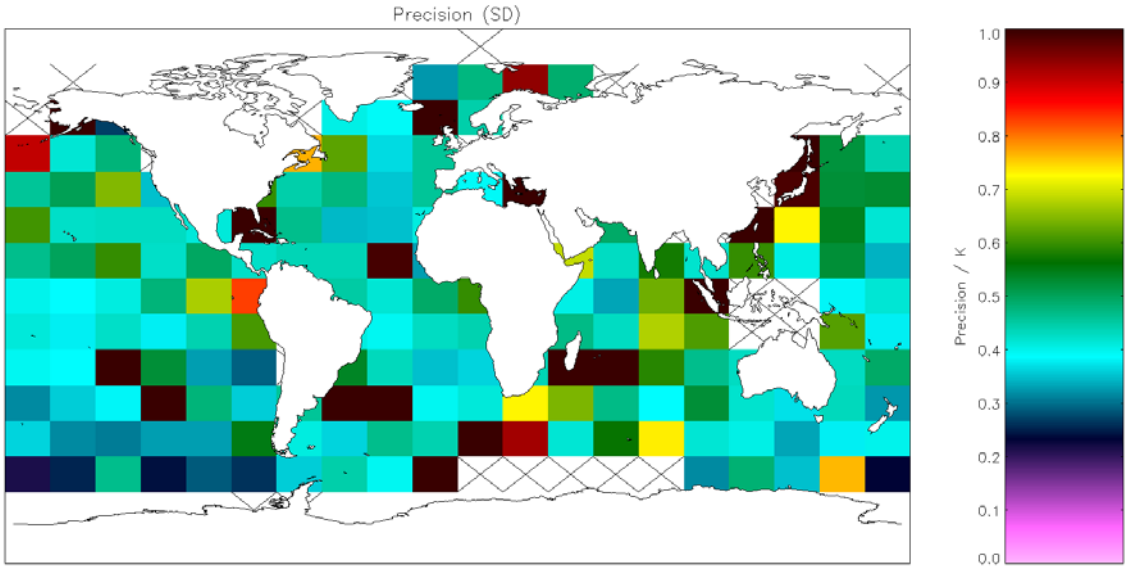


Figure A2.5.2-1: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.648 K (**Table 5.4.2**).

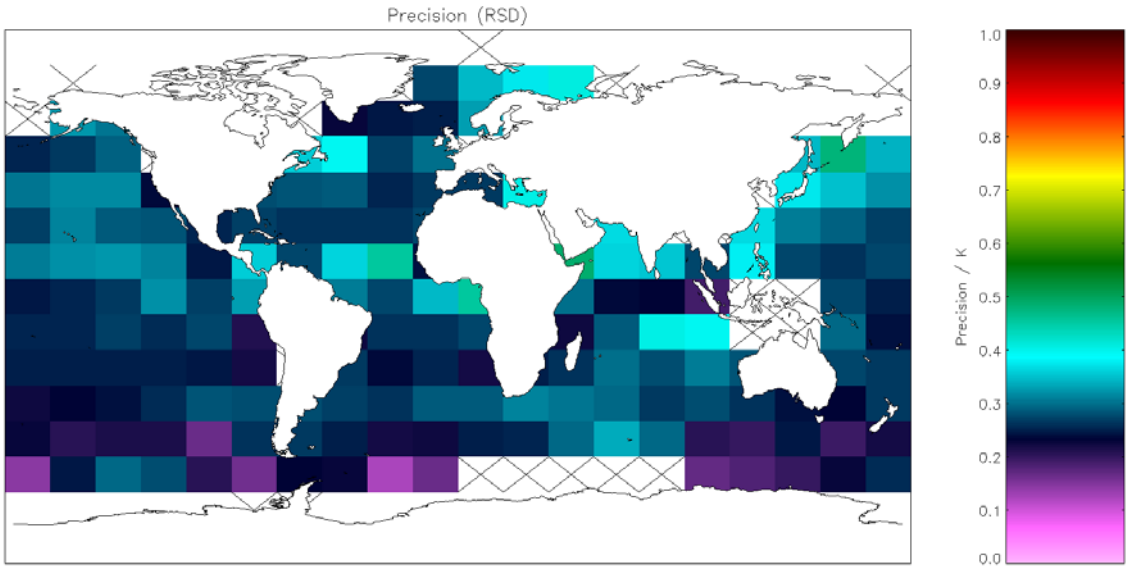


Figure A2.5.2-2: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.280 K (**Table 5.4.2**).

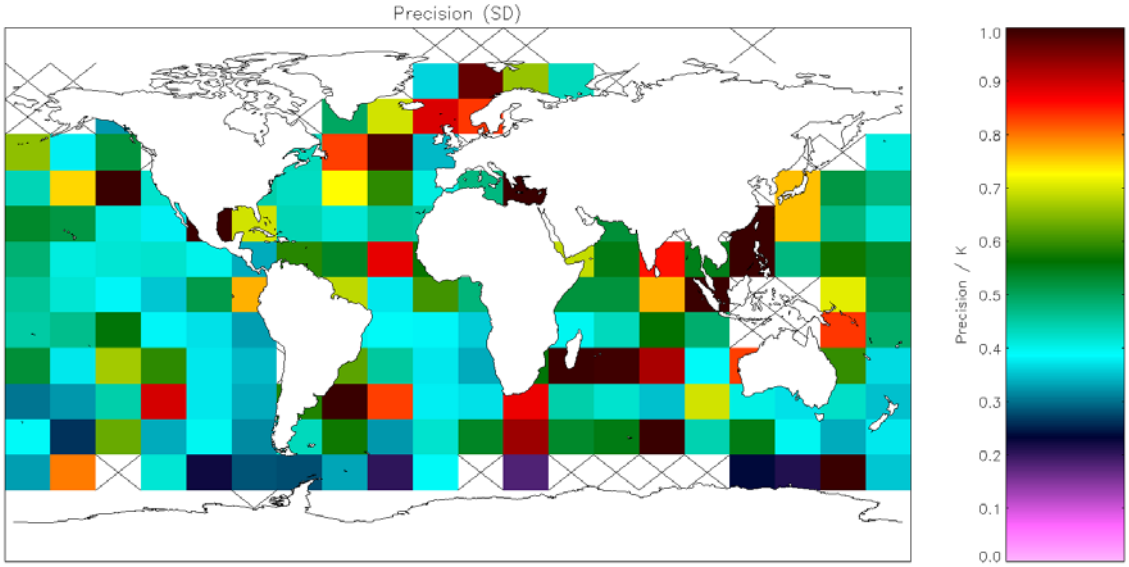


Figure A2.5.2-3: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.647 K (**Table 5.4.3**).

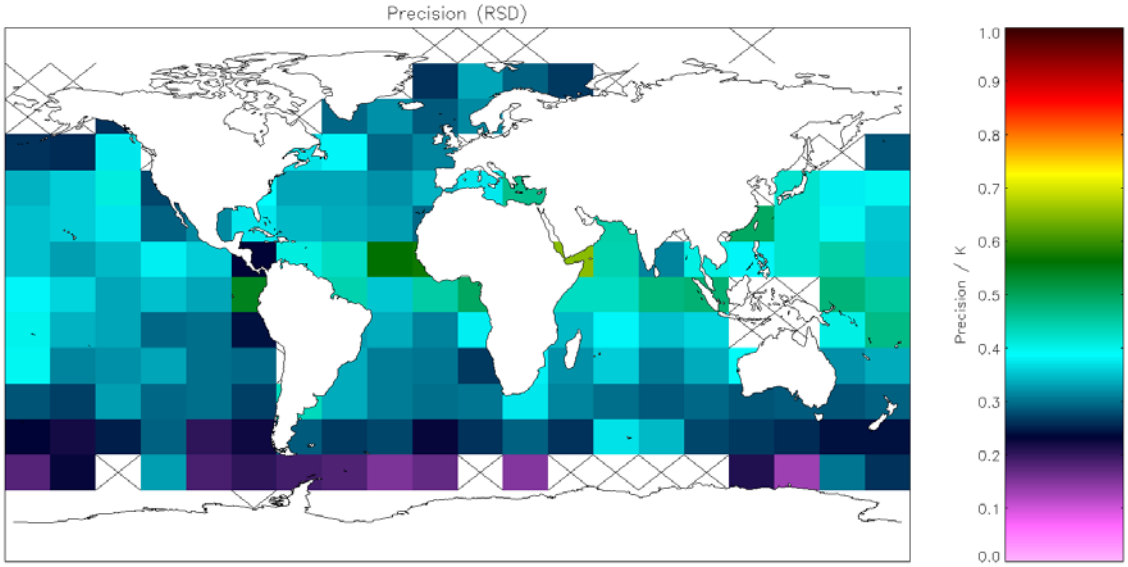


Figure A2.5.2-4: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.345 K (**Table 5.4.3**).

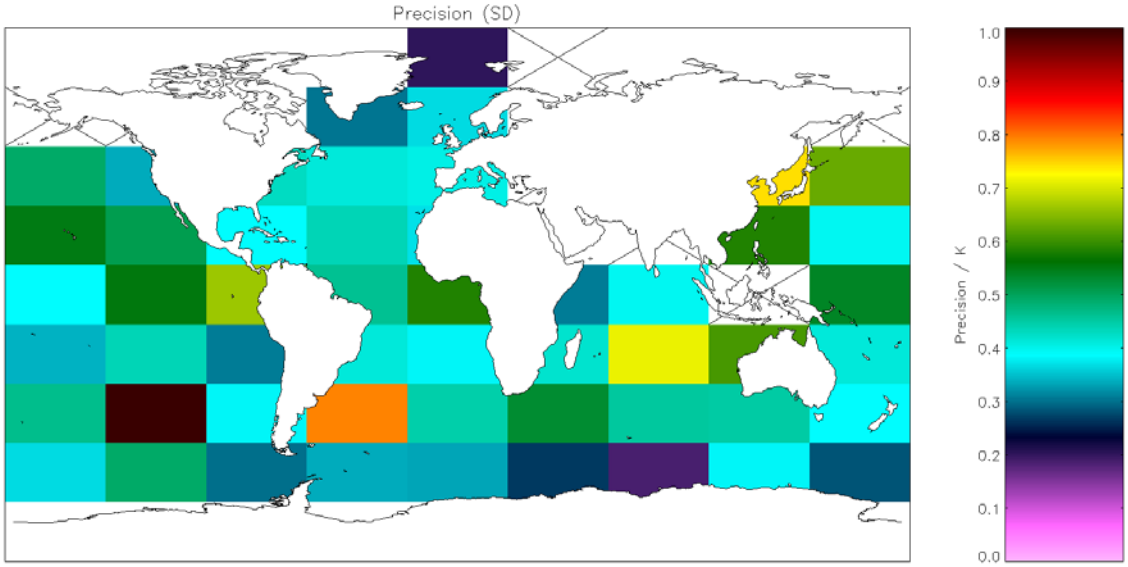


Figure A2.5.2-5: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.502 K (**Table 5.5.1**).

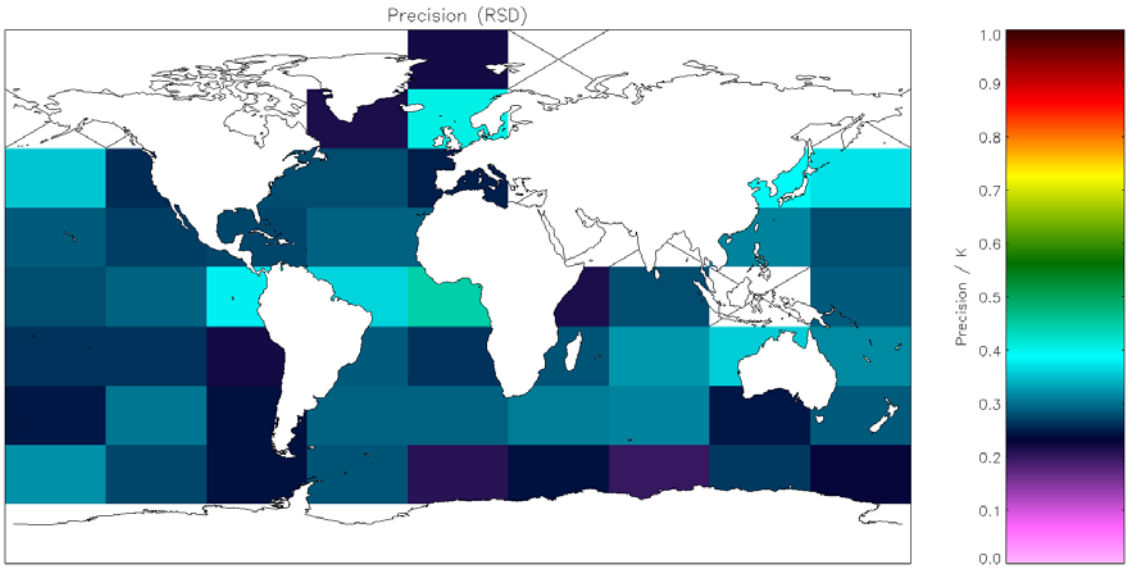


Figure A2.5.2-6: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.292 K (**Table 5.5.1**).

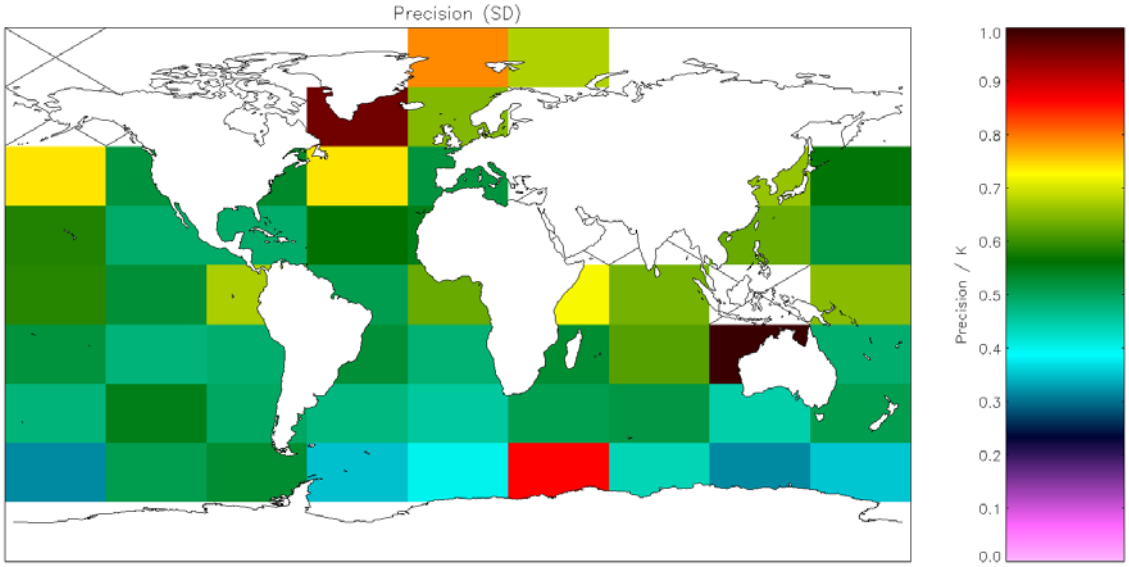


Figure A2.5.2-7: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.588 K (**Table 5.5.2**).

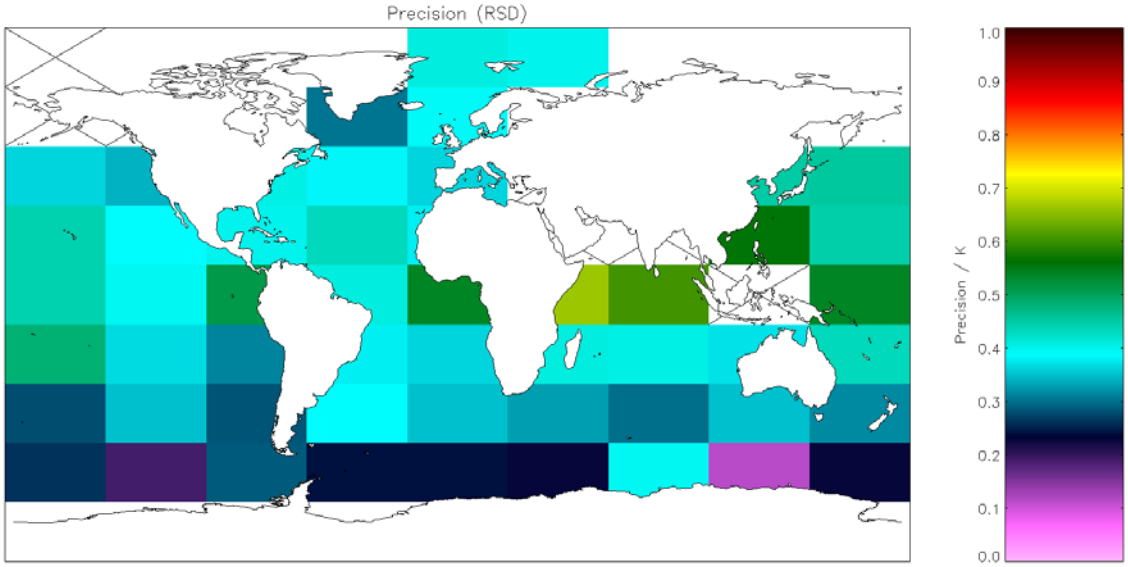


Figure A2.5.2-8: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.415 K (**Table 5.5.2**).

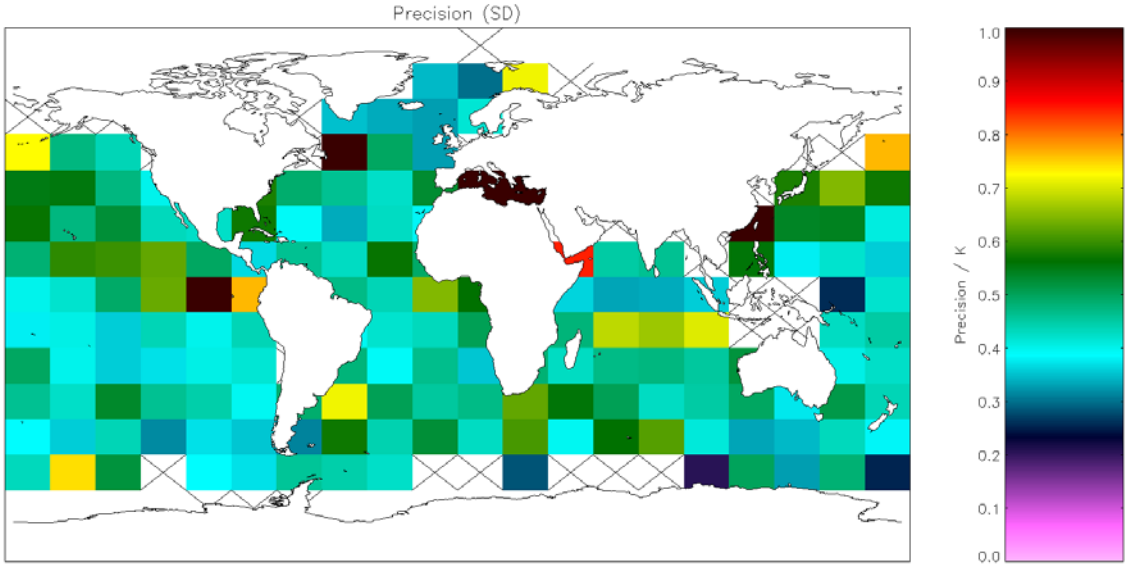


Figure A2.5.2-9: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.519 K (**Table 5.6.1**).

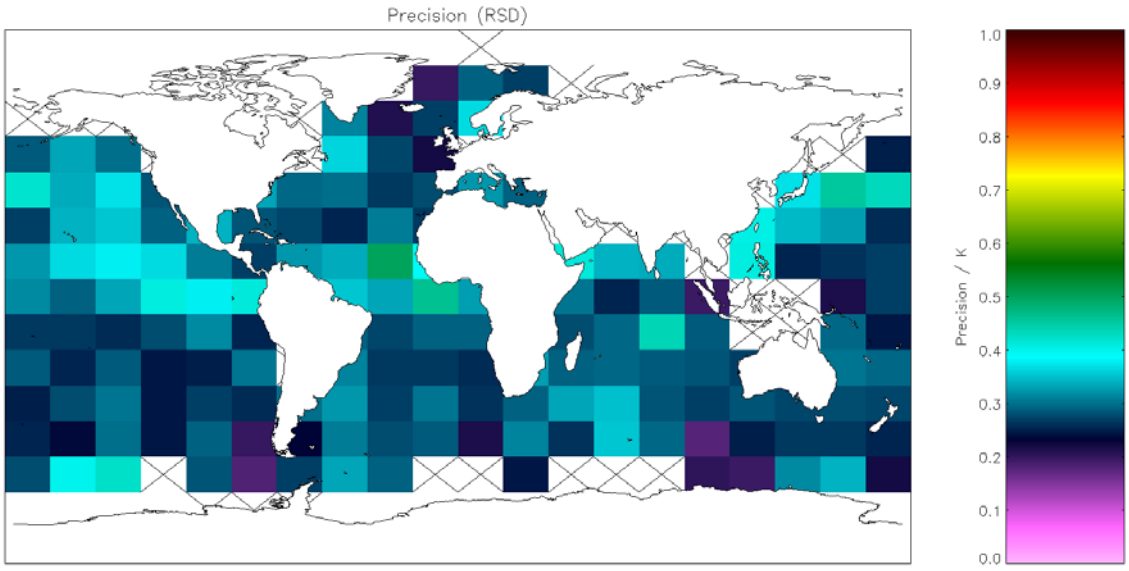


Figure A2.5.2-10: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.298 K (**Table 5.6.1**).

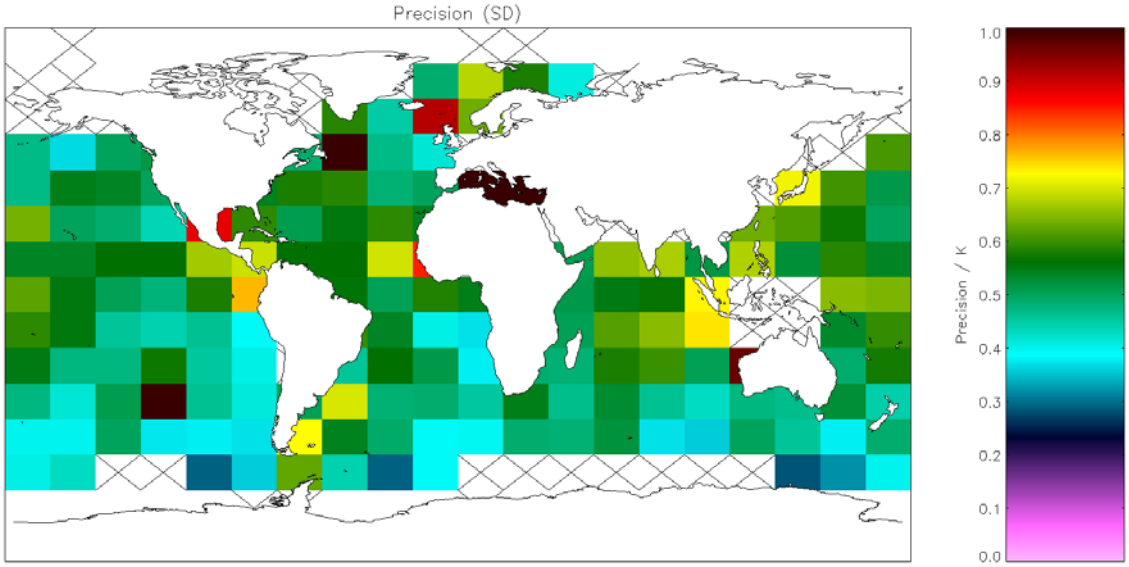


Figure A2.5.2-11: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.641 K (**Table 5.6.2**).

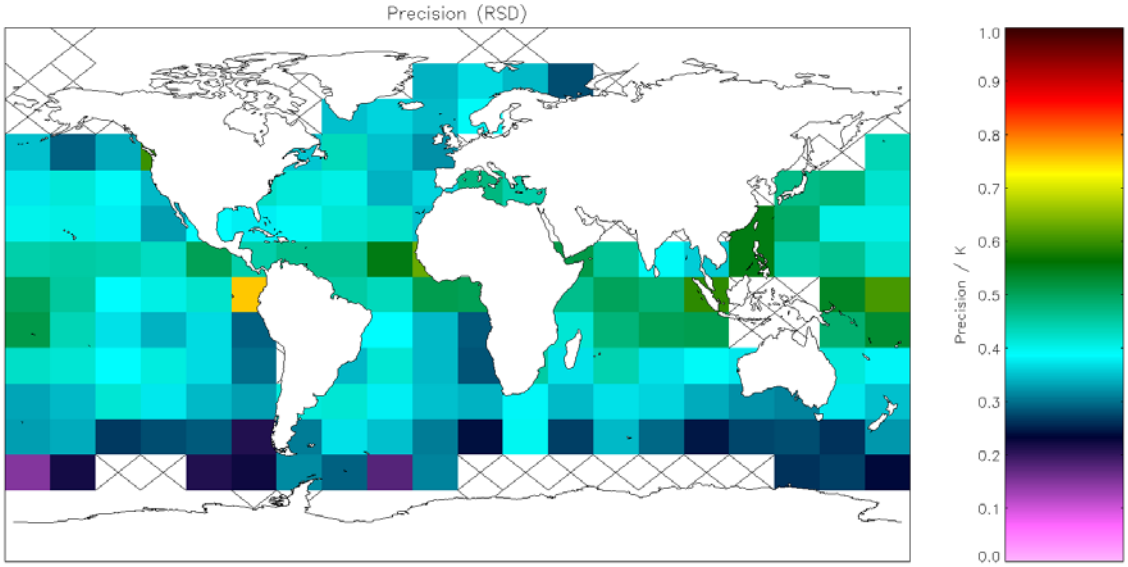


Figure A2.5.2-12: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.409 K (**Table 5.6.2**).

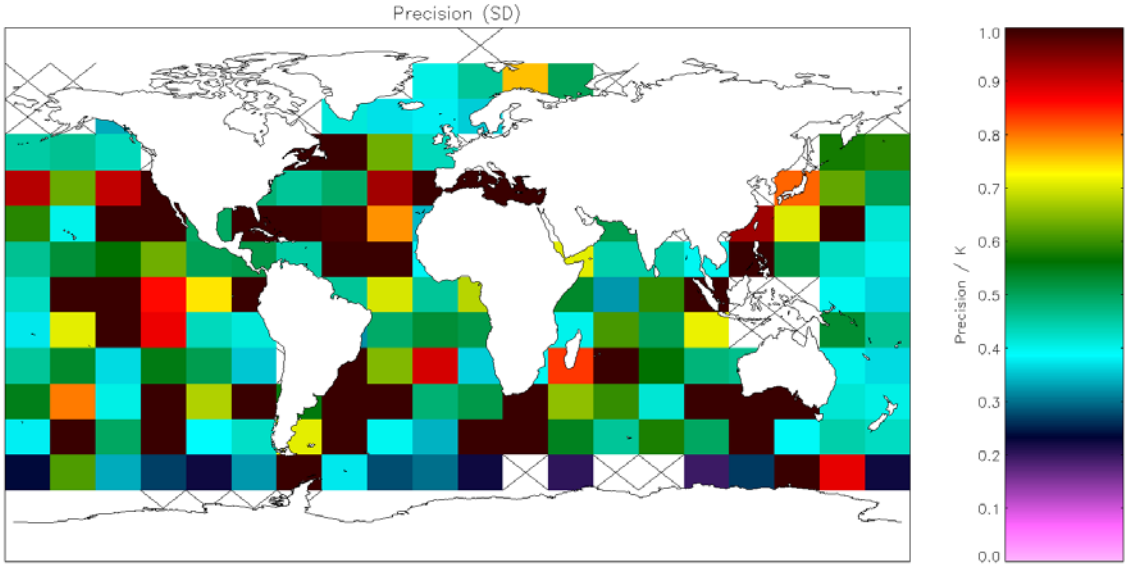


Figure A2.5.2-13: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.910 K (**Table 5.7.1**).

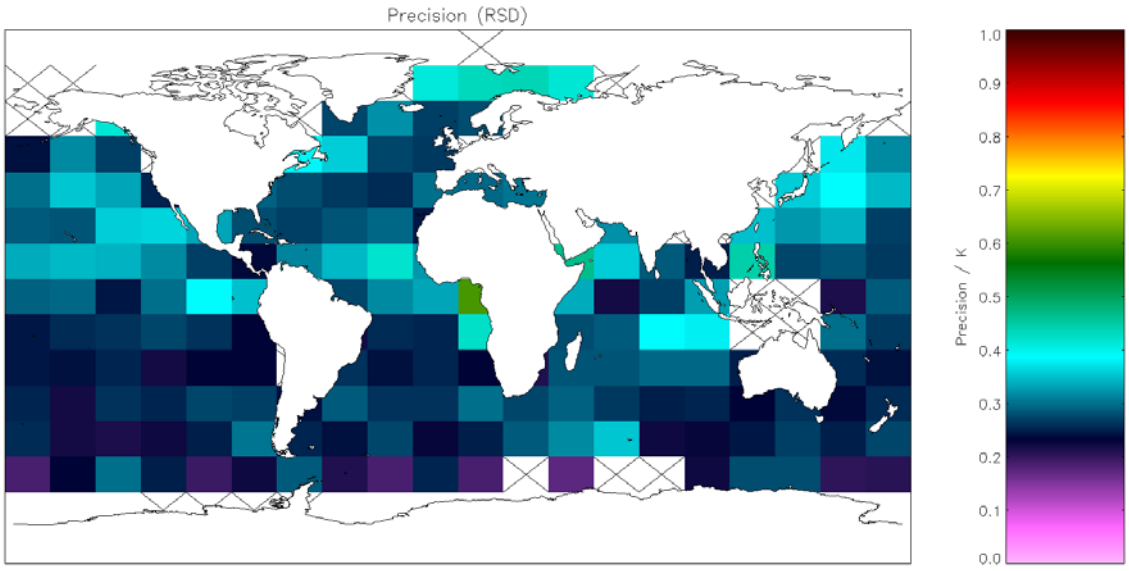


Figure A2.5.2-14: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.282 K (**Table 5.7.1**).

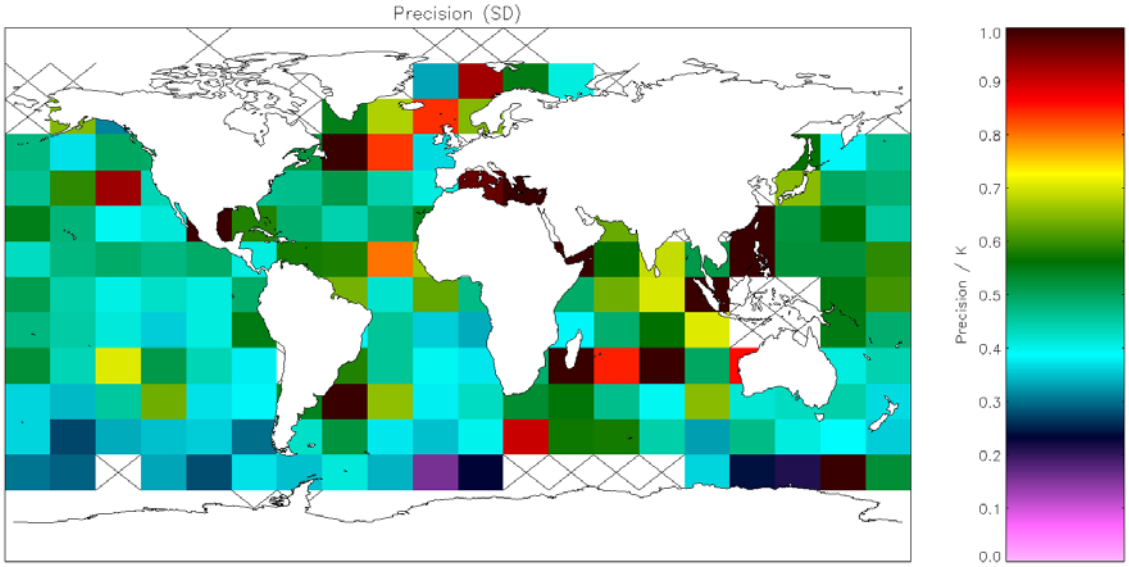


Figure A2.5.2-15: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.641 K (**Table 5.7.2**).

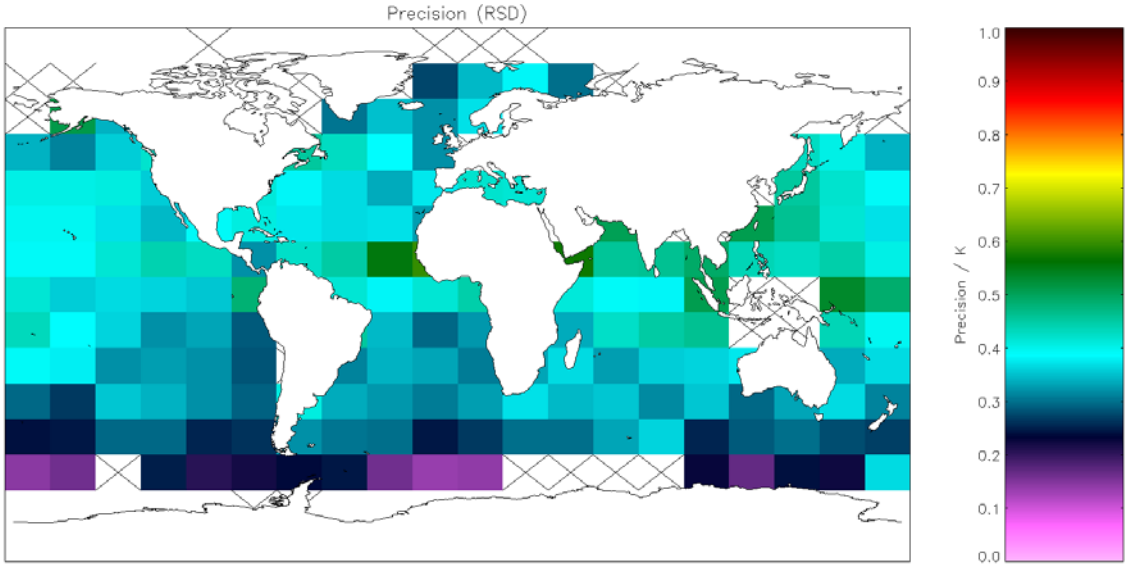


Figure A2.5.2-16: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.366 K (**Table 5.7.2**).

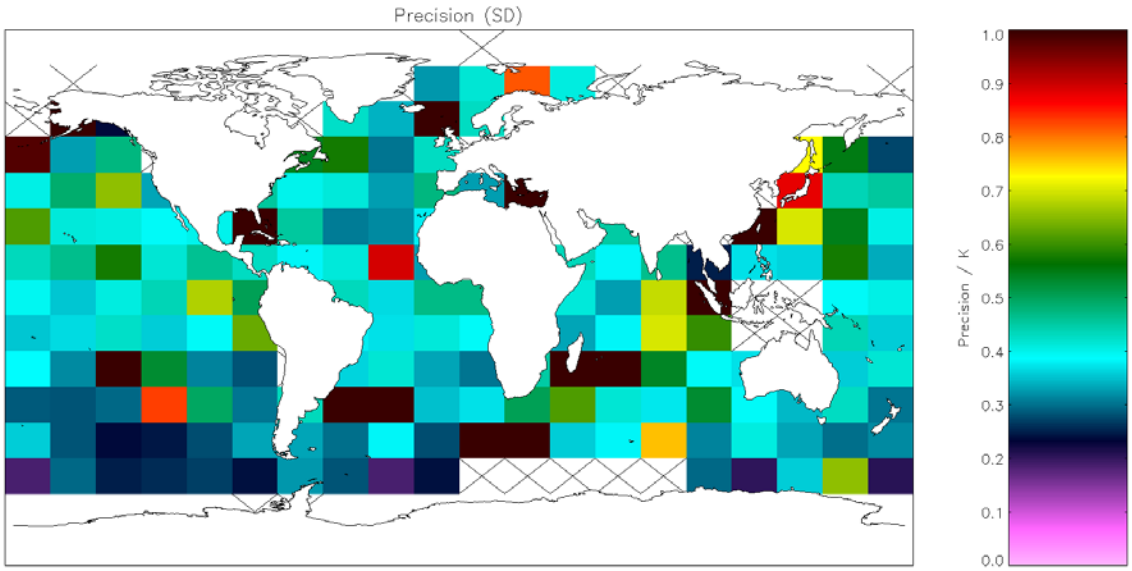


Figure A2.5.2-17: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.631 K (**Table 5.4.2**). As **Figure A2.5.2-1** but with OE v2 QC filtering applied.

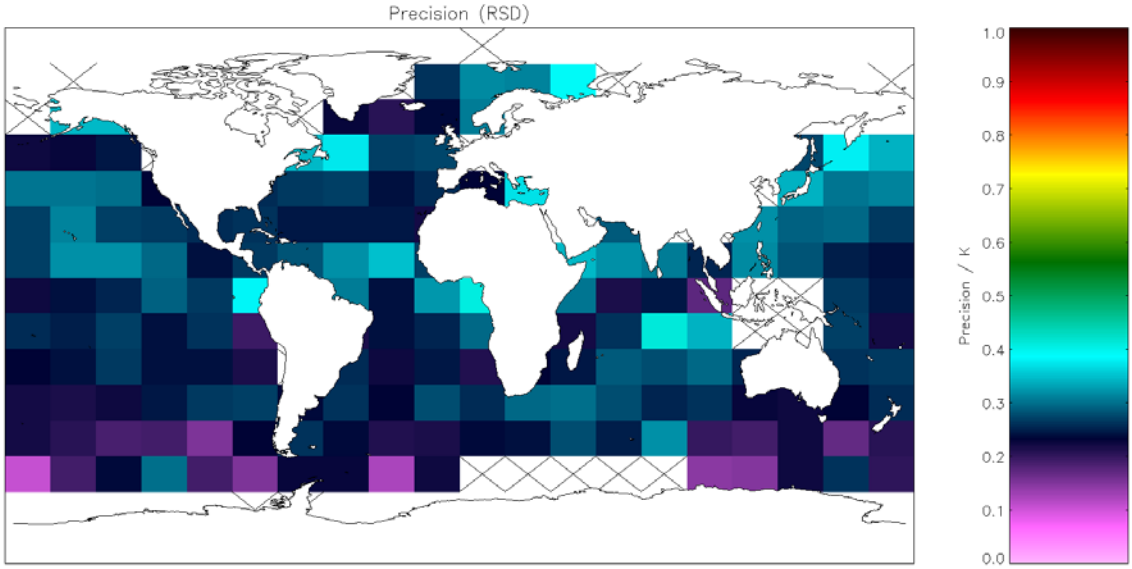


Figure A2.5.2-18: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.264 K (**Table 5.4.2**). As **Figure A2.5.2-2** but with OE v2 QC filtering applied.

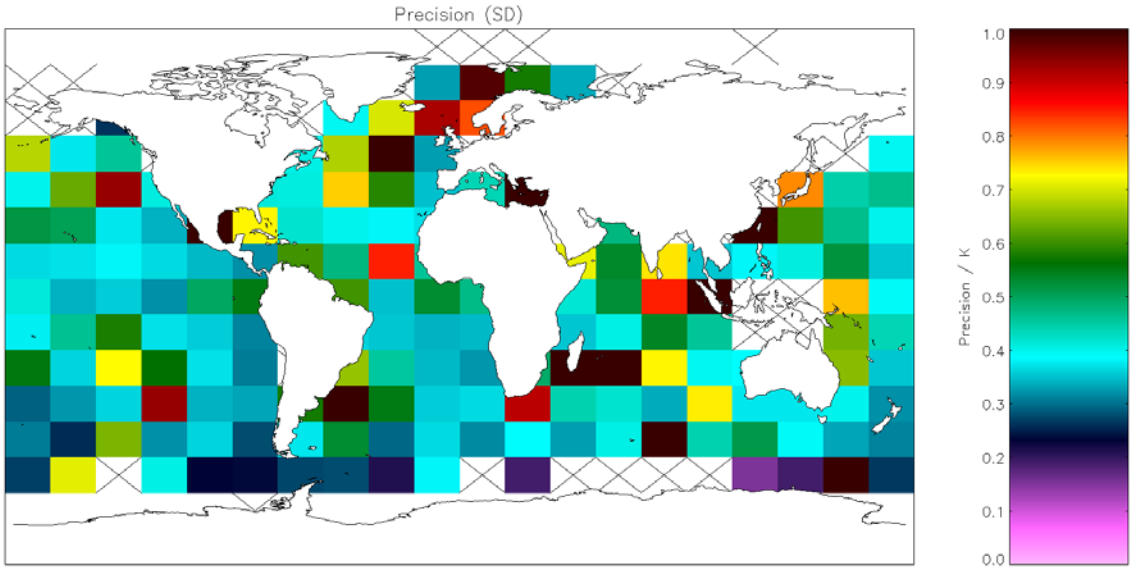


Figure A2.5.2-19: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.619 K (**Table 5.4.3**). As **Figure A2.5.2-3** but with OE v2 QC filtering applied.

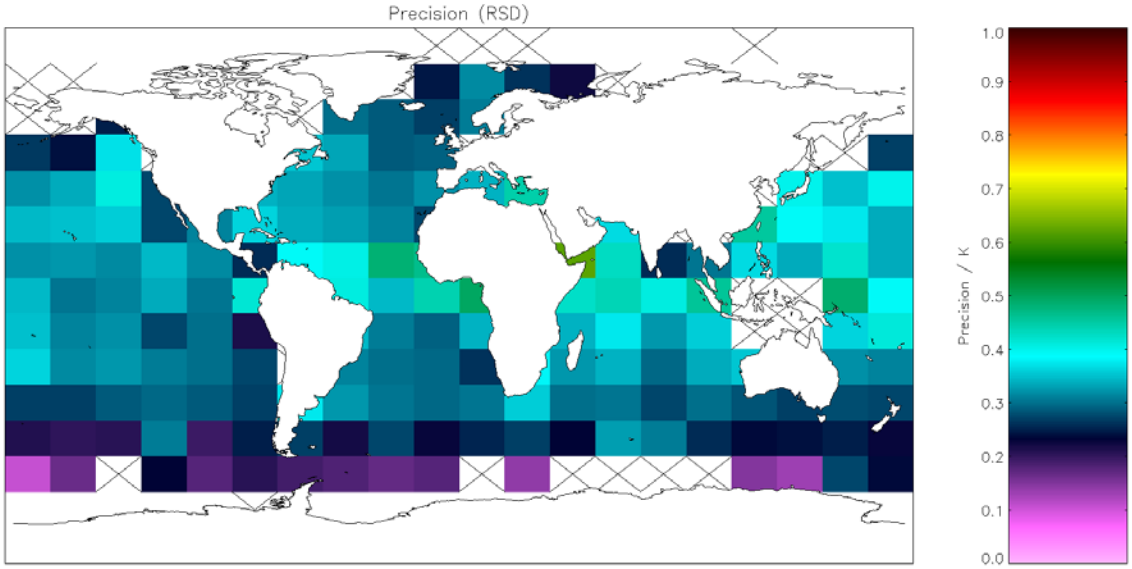


Figure A2.5.2-20: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.328 K (**Table 5.4.3**). As **Figure A2.5.2-4** but with OE v2 QC filtering applied.

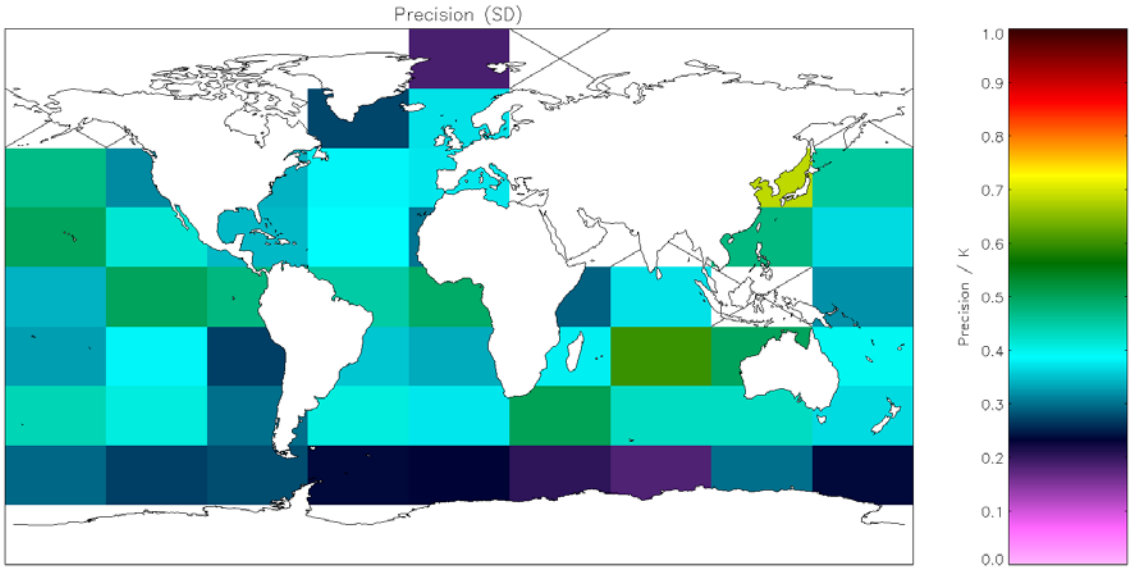


Figure A2.5.2-21: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.410 K (**Table 5.5.1**). As **Figure A2.5.2-5** but with OE v2 QC filtering applied.

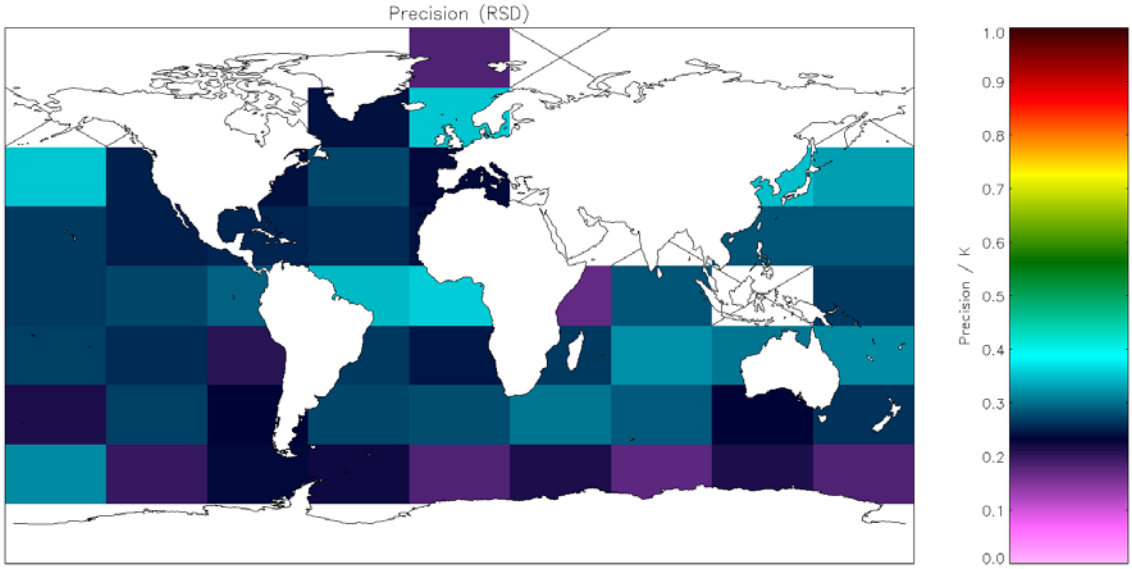


Figure A2.5.2-22: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.271 K (**Table 5.5.1**). As **Figure A2.5.2-6** but with OE v2 QC filtering applied.

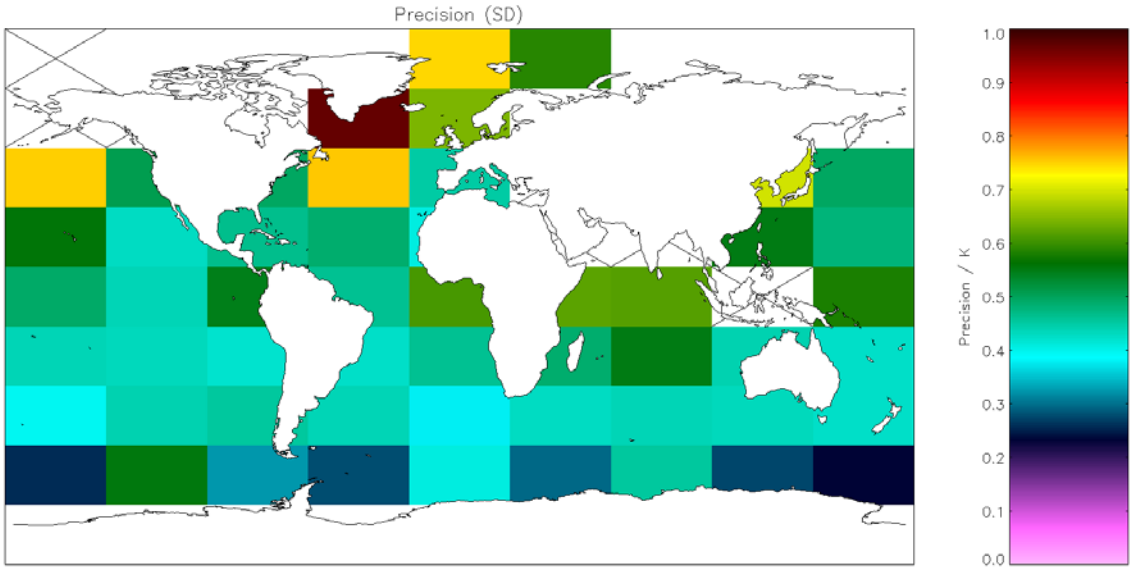


Figure A2.5.2-23: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.523 K (**Table 5.5.2**). As **Figure A2.5.2-7** but with OE v2 QC filtering applied.

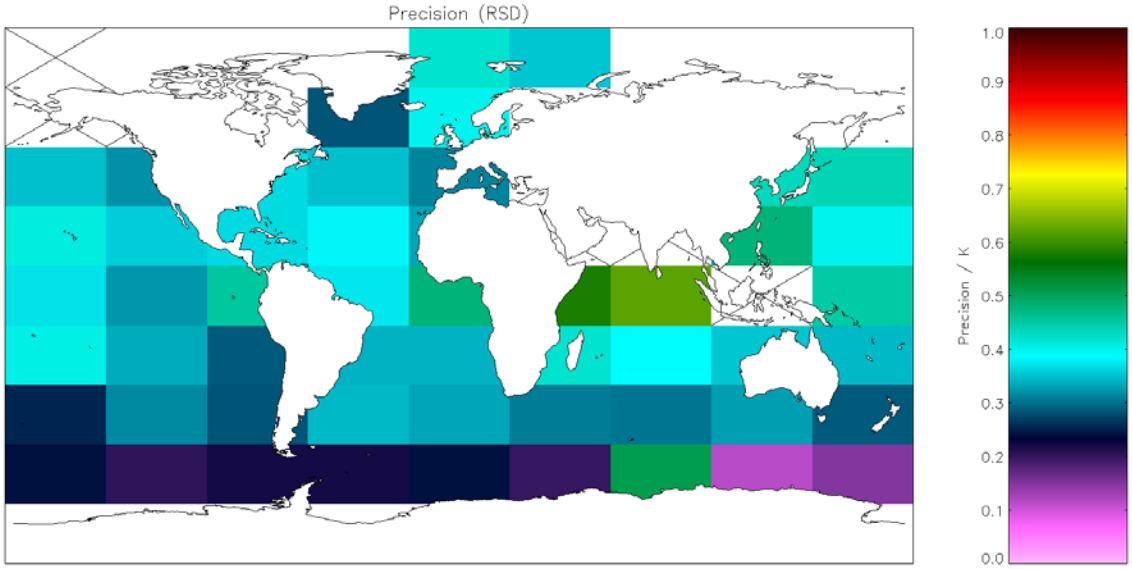


Figure A2.5.2-24: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.372 K (**Table 5.5.2**). As **Figure A2.5.2-8** but with OE v2 QC filtering applied.

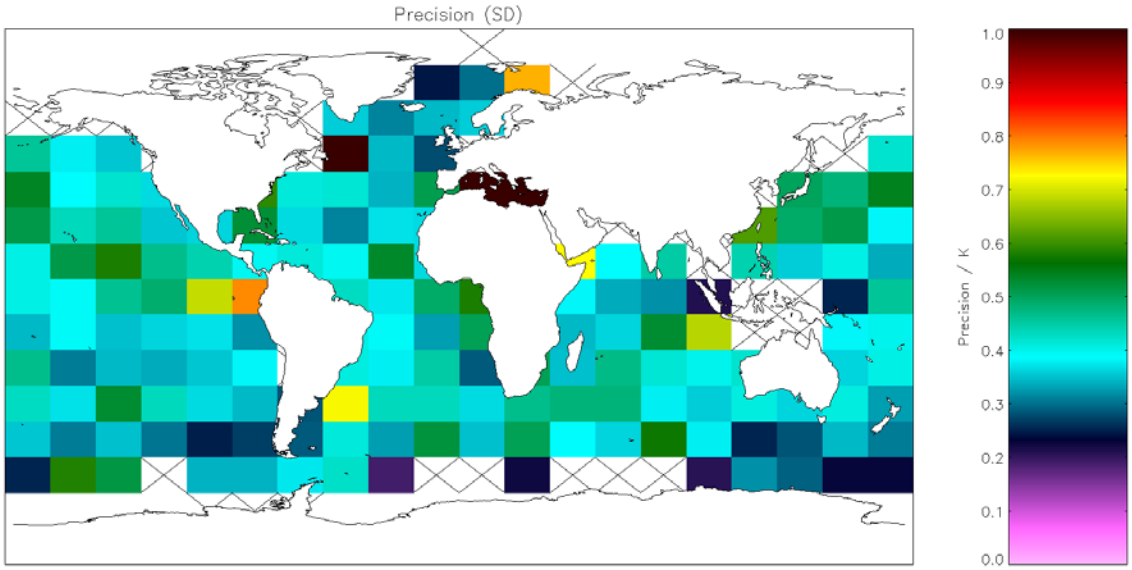


Figure A2.5.2-25: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.448 K (**Table 5.6.1**). As **Figure A2.5.2-9** but with OE v2 QC filtering applied.

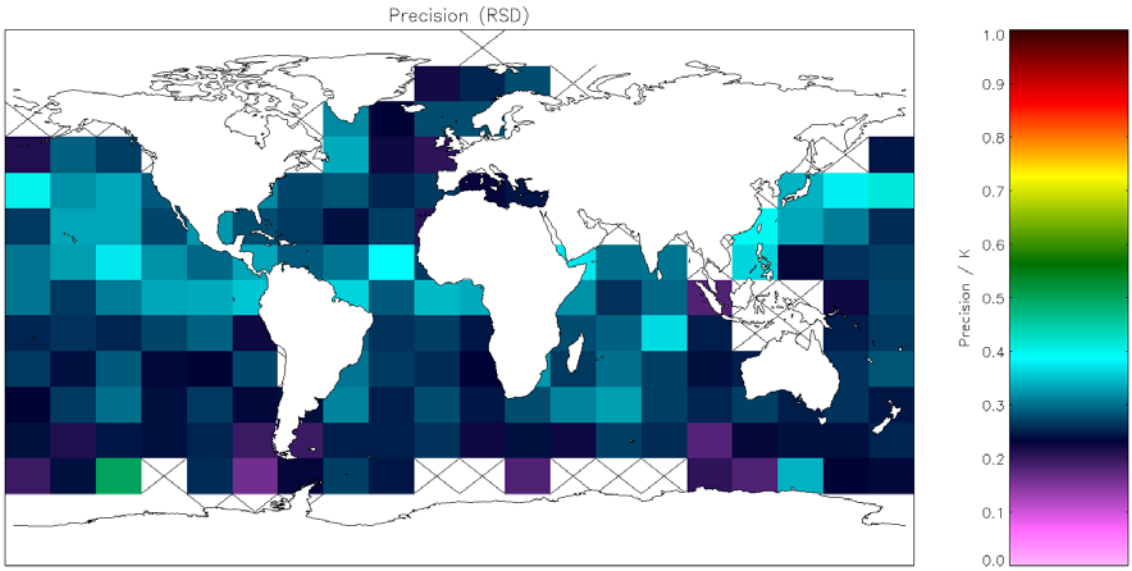


Figure A2.5.2-26: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.277 K (**Table 5.6.1**). As **Figure A2.5.2-10** but with OE v2 QC filtering applied.

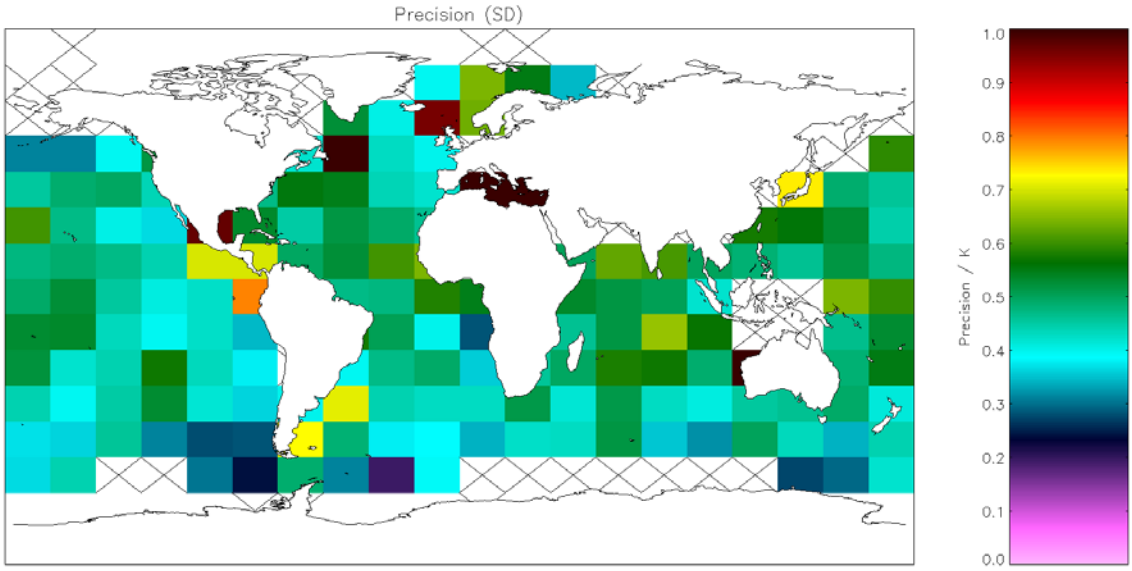


Figure A2.5.2-27: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.601 K (**Table 5.6.2**). As **Figure A2.5.2-11** but with OE v2 QC filtering applied.

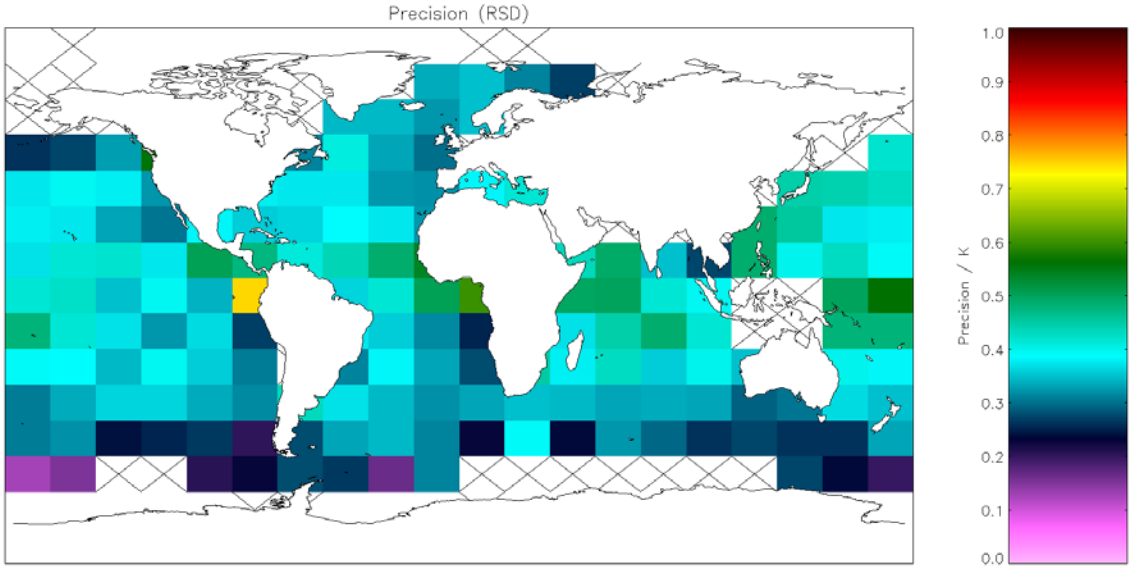


Figure A2.5.2-28: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.381 K (**Table 5.6.2**). As **Figure A2.5.2-12** but with OE v2 QC filtering applied.

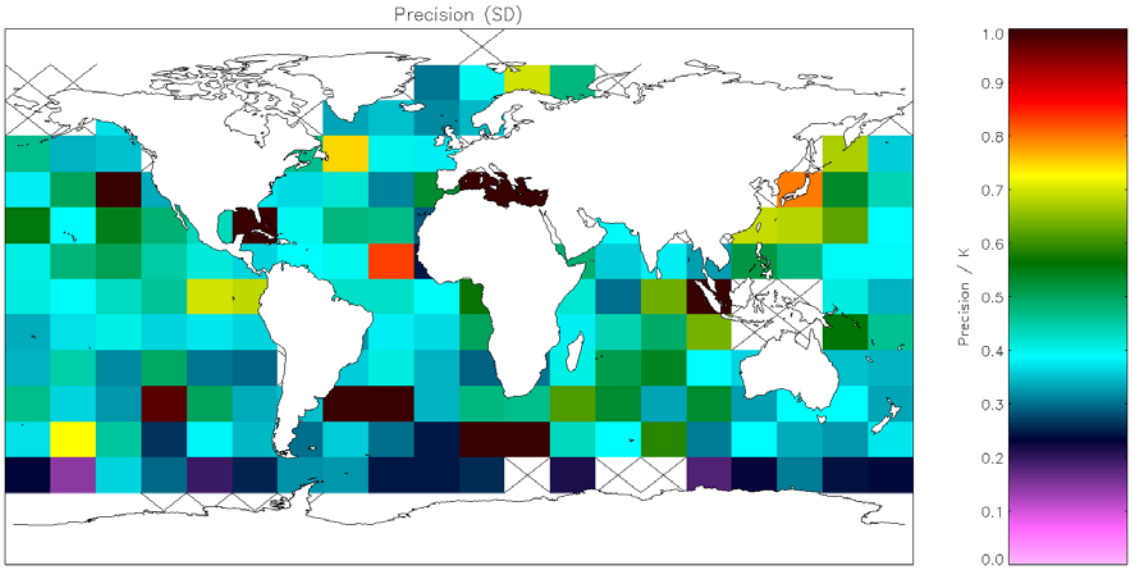


Figure A2.5.2-29: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.572 K (**Table 5.7.1**). As **Figure A2.5.2-13** but with OE v2 QC filtering applied.

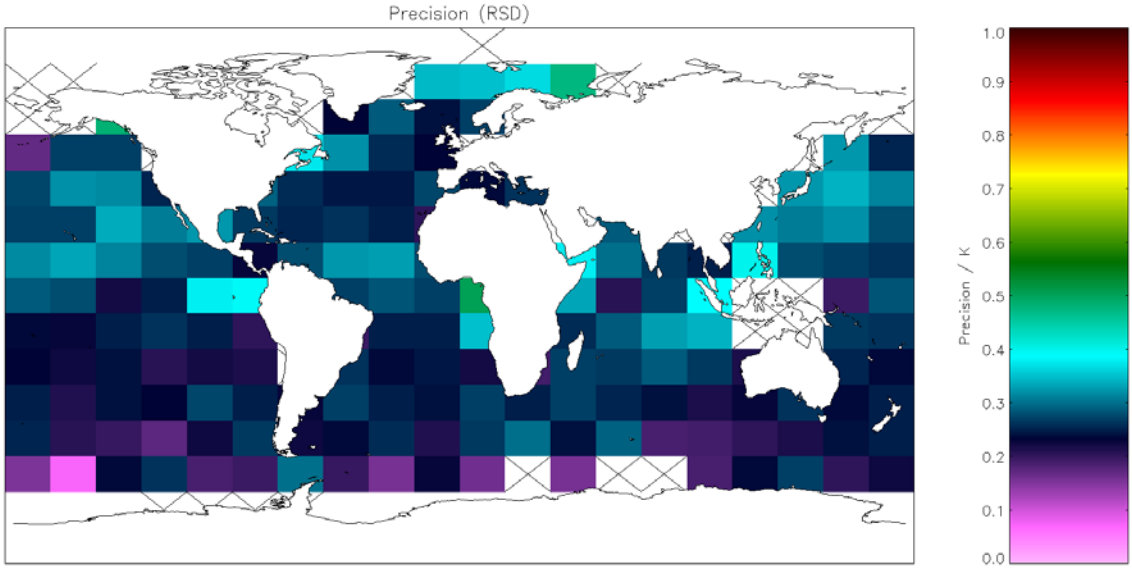


Figure A2.5.2-30: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.262 K (**Table 5.7.1**). As **Figure A2.5.2-14** but with OE v2 QC filtering applied.

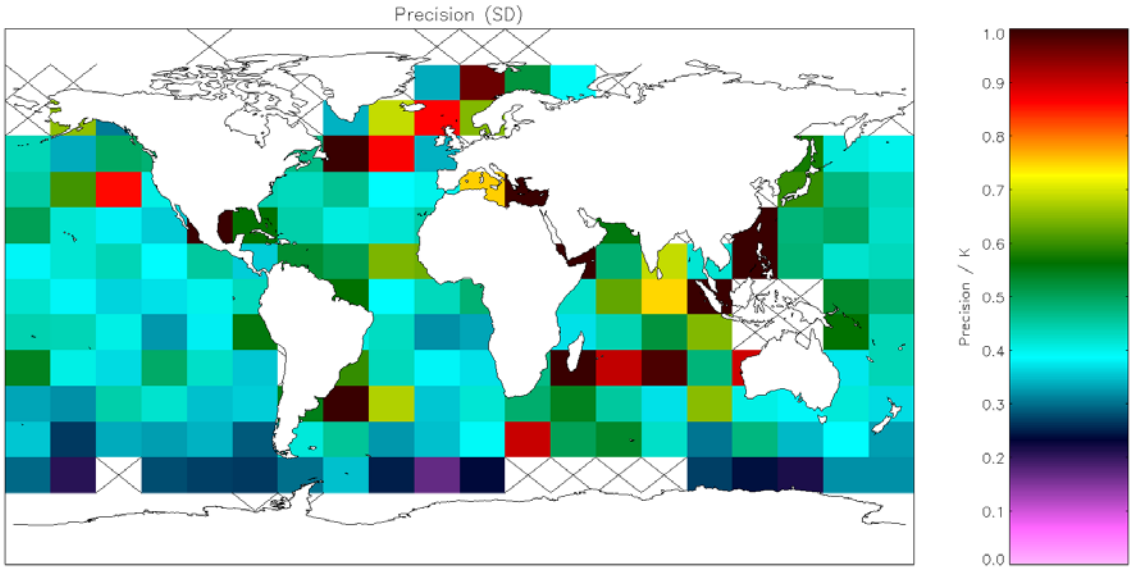


Figure A2.5.2-31: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.607 K (**Table 5.7.2**). As **Figure A2.5.2-15** but with OE v2 QC filtering applied.

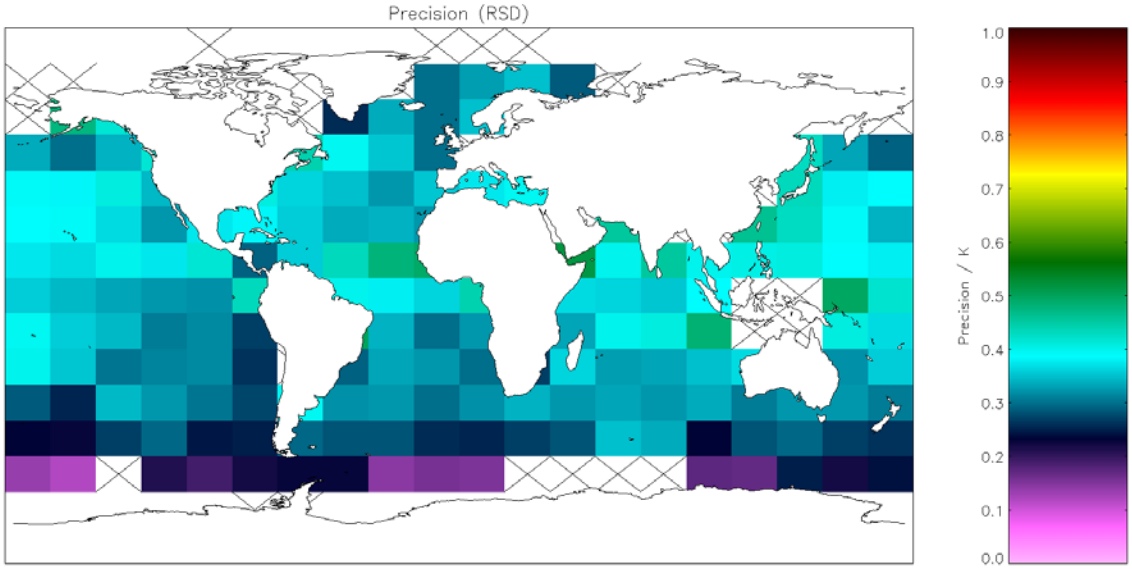


Figure A2.5.2-32: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.345 K (**Table 5.7.2**). As **Figure A2.5.2-16** but with OE v2 QC filtering applied.

A2.5.3 Sensitivity

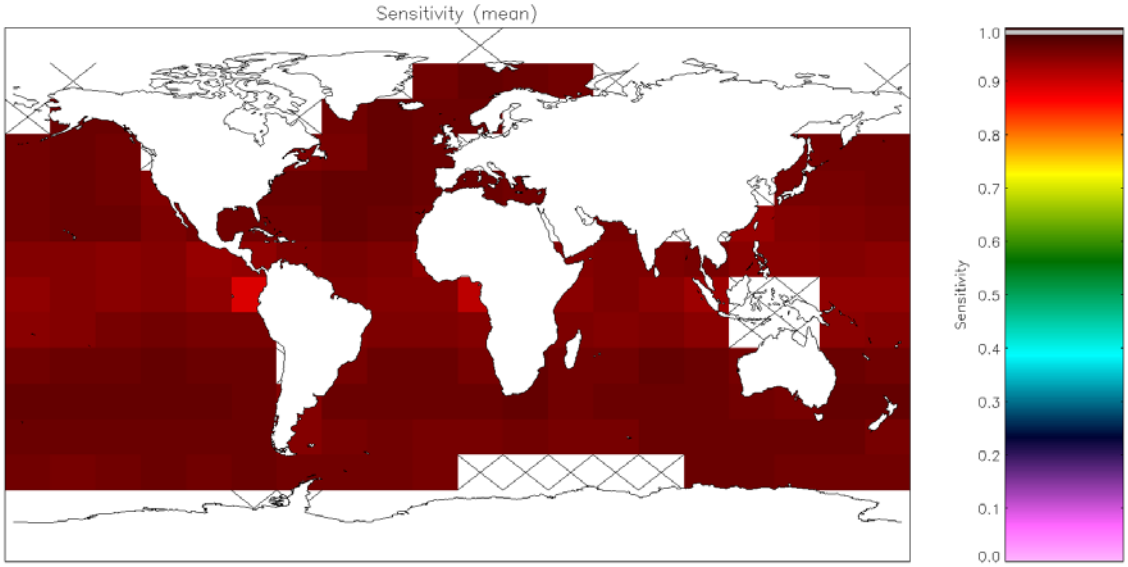


Figure A2.5.3-1: Global map of SST sensitivity using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.4.2**).

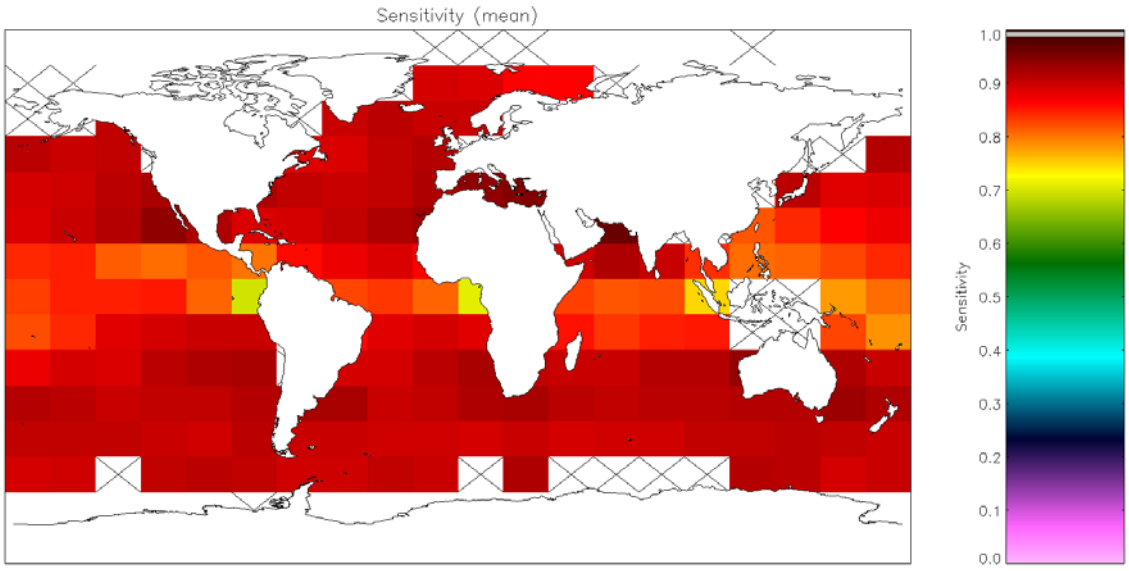


Figure A2.5.3-2: Global map of SST sensitivity using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.4.3**).

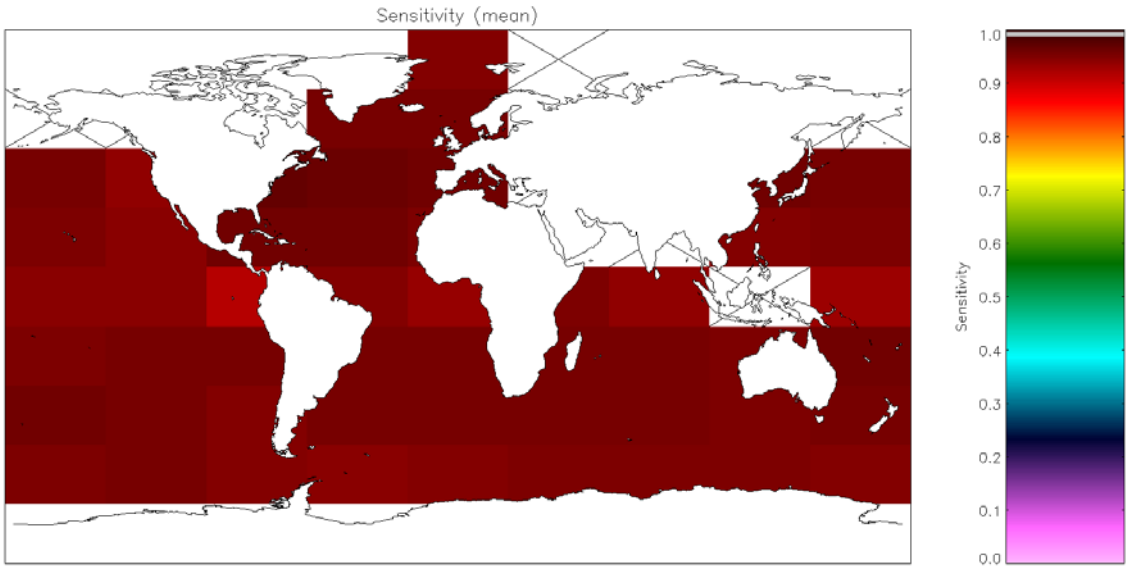


Figure A2.5.3-3: Global map of SST sensitivity using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.5.1**).

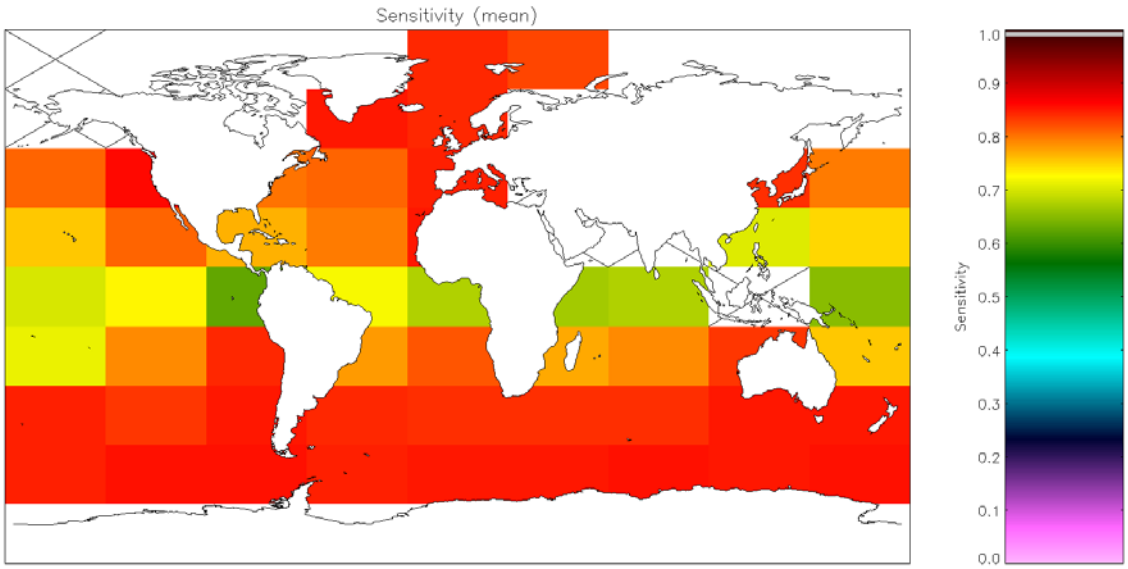


Figure A2.5.3-4: Global map of SST sensitivity using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.5.2**).

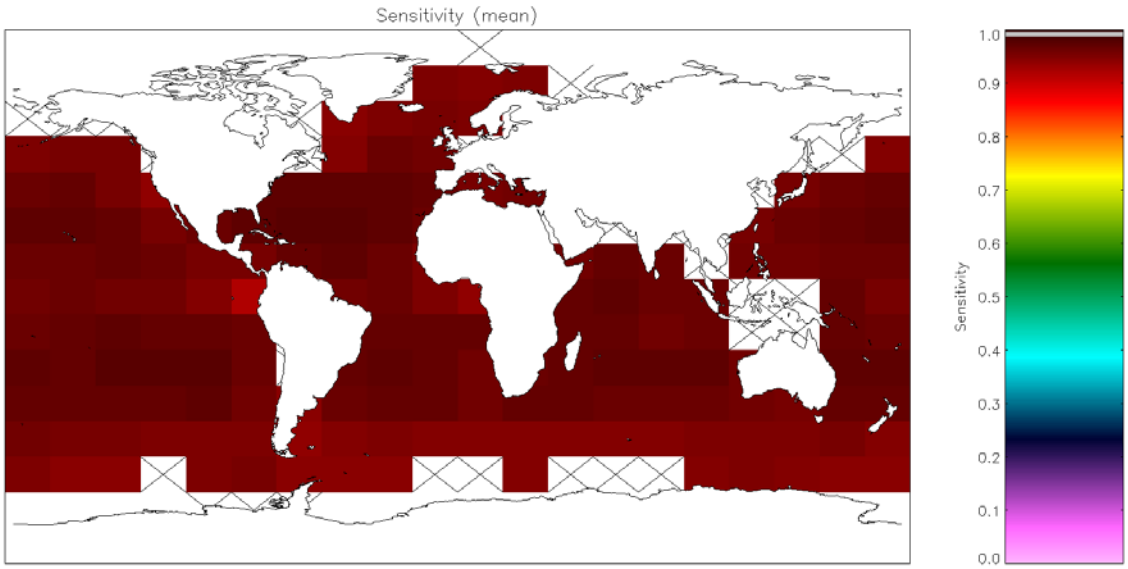


Figure A2.5.3-5: Global map of SST sensitivity using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.6.1**).

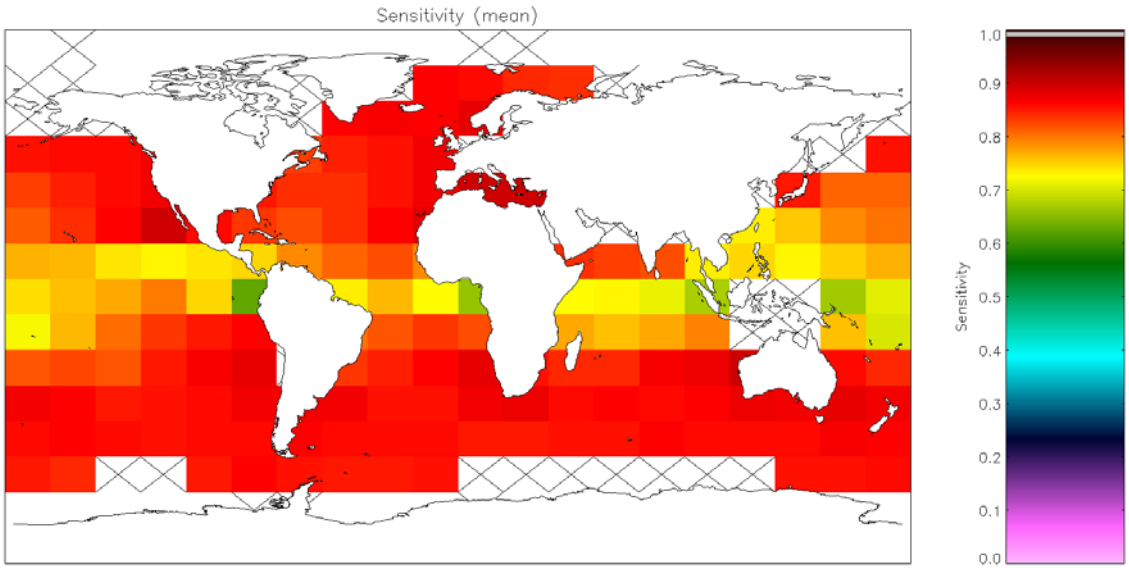


Figure A2.5.3-6: Global map of SST sensitivity using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.6.2**).

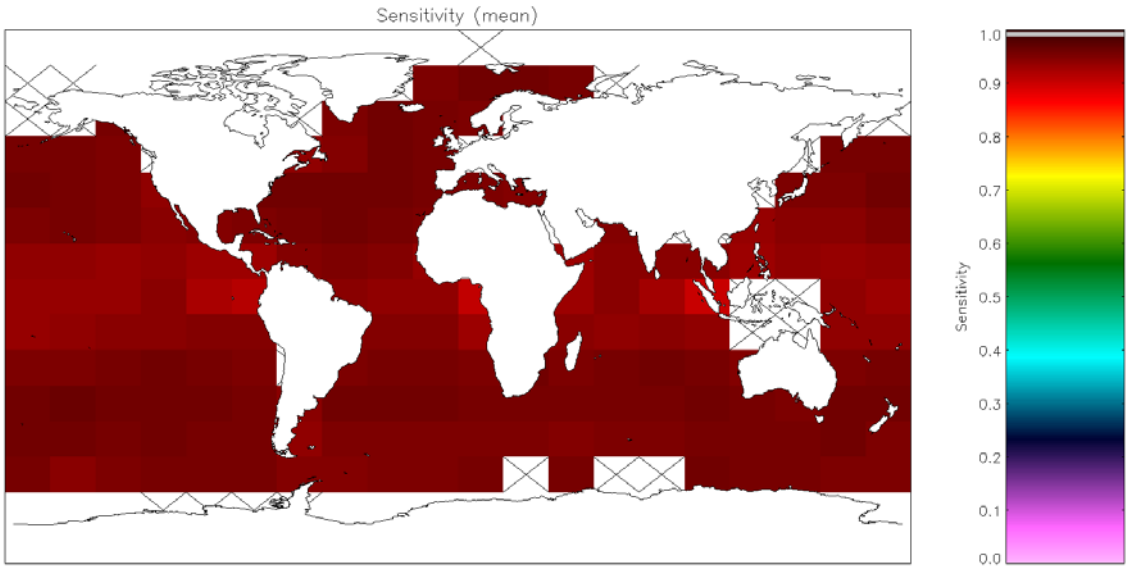


Figure A2.5.3-7: Global map of SST sensitivity using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.7.1**).

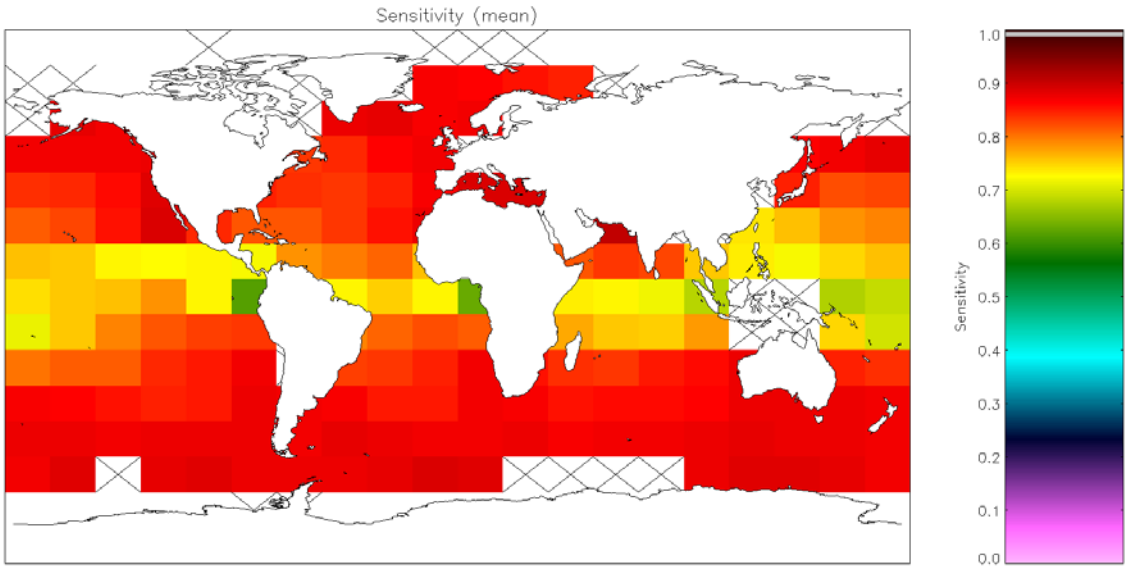


Figure A2.5.3-8: Global map of SST sensitivity using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.7.2**).

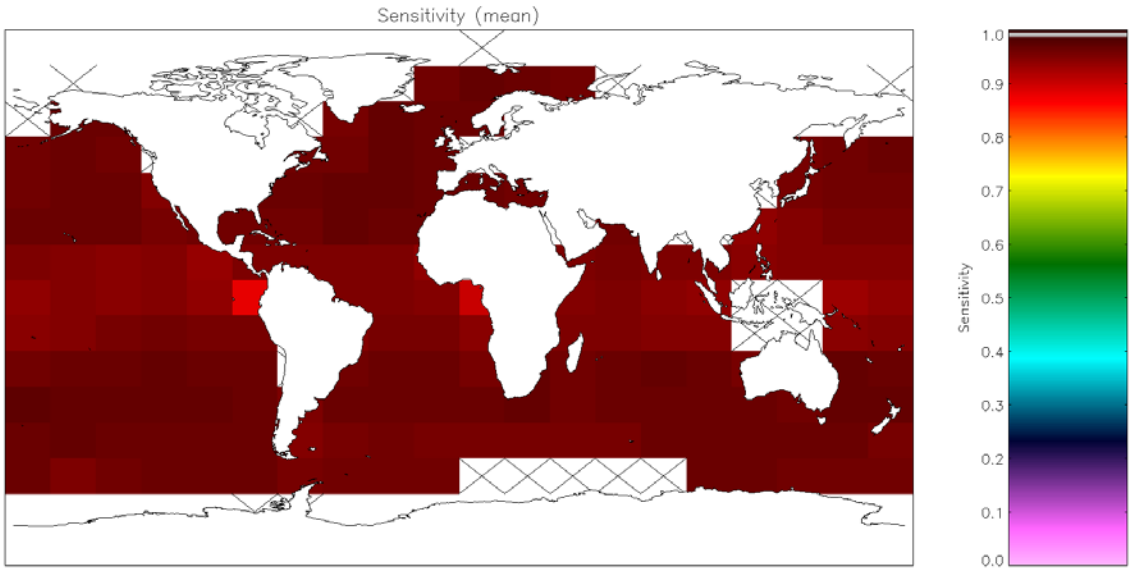


Figure A2.5.3-9: Global map of SST sensitivity using incremental regression retrieval for night-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.4.2**). As **Figure A2.5.3-1** but with OE v2 QC filtering applied.

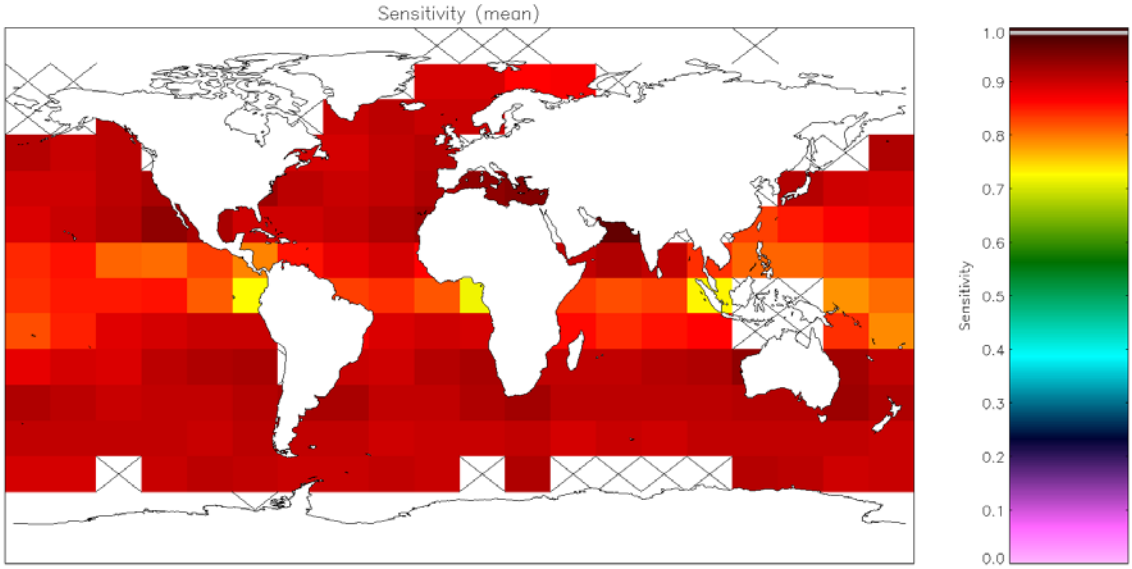


Figure A2.5.3-10: Global map of SST sensitivity using incremental regression retrieval for day-time AVHRR-Metop observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.4.3**). As **Figure A2.5.3-2** but with OE v2 QC filtering applied.

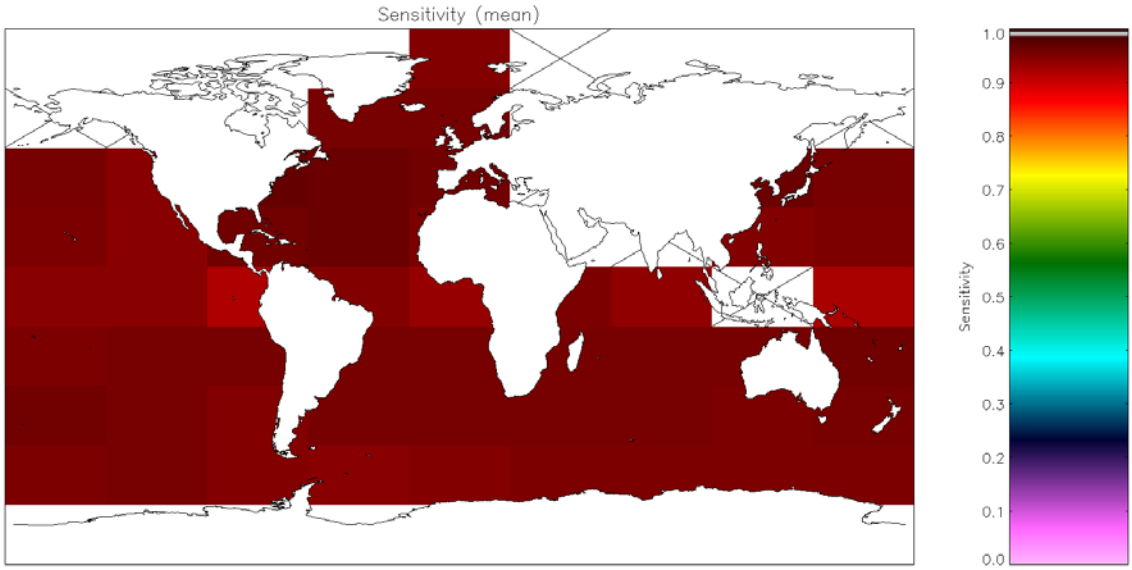


Figure A2.5.3-11: Global map of SST sensitivity using incremental regression retrieval for night-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.5.1**). As **Figure A2.5.3-3** but with OE v2 QC filtering applied.

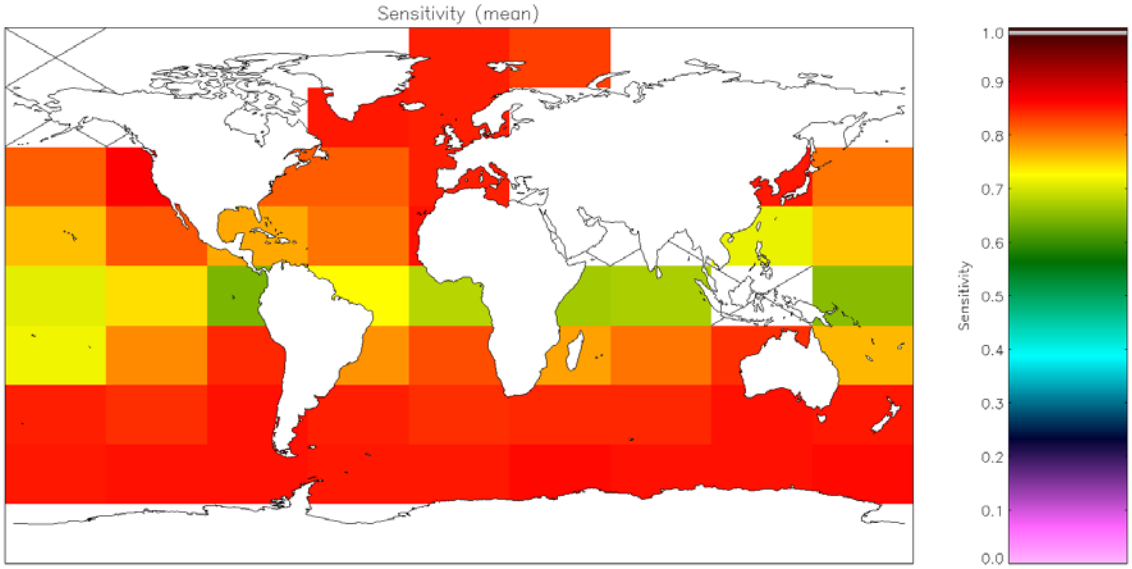


Figure A2.5.3-12: Global map of SST sensitivity using incremental regression retrieval for day-time AVHRR-19 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.5.2**). As **Figure A2.5.3-4** but with OE v2 QC filtering applied.

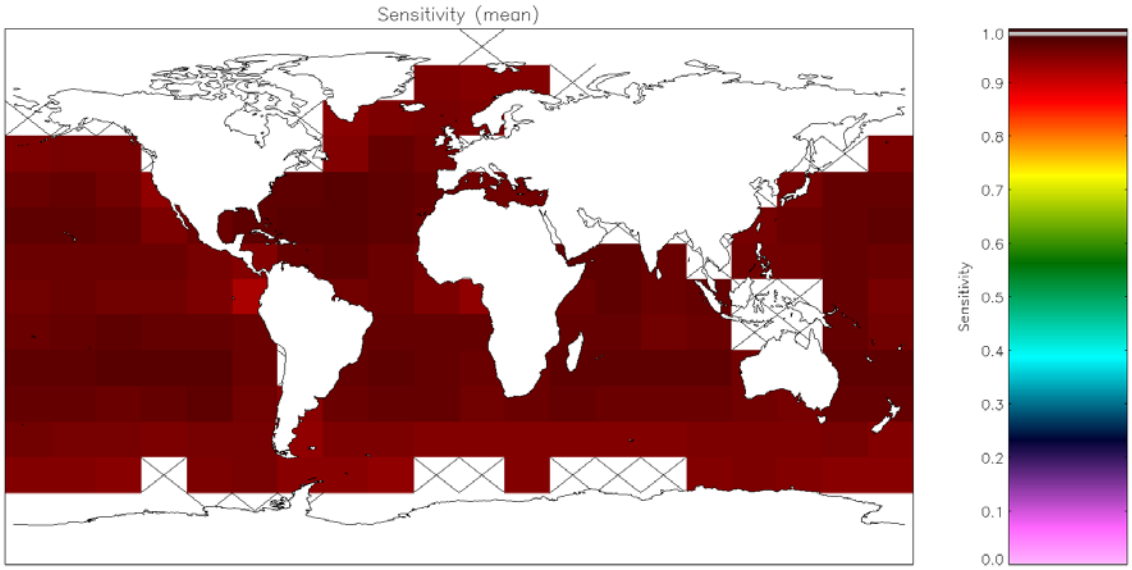


Figure A2.5.3-13: Global map of SST sensitivity using incremental regression retrieval for night-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.6.1**). As **Figure A2.5.3-5** but with OE v2 QC filtering applied.

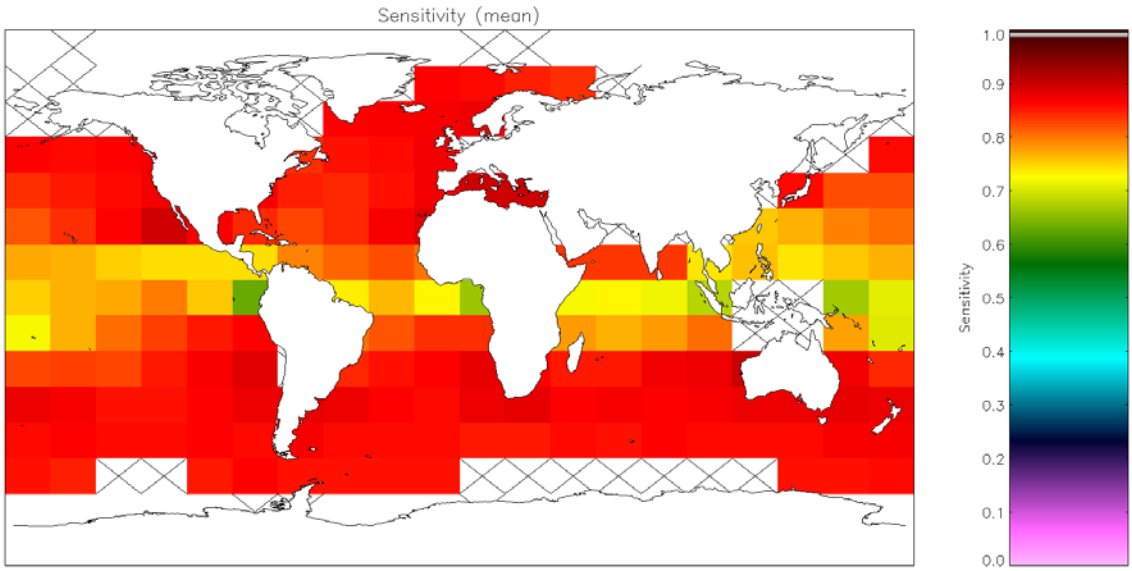


Figure A2.5.3-14: Global map of SST sensitivity using incremental regression retrieval for day-time AVHRR-18 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.6.2**). As **Figure A2.5.3-6** but with OE v2 QC filtering applied.

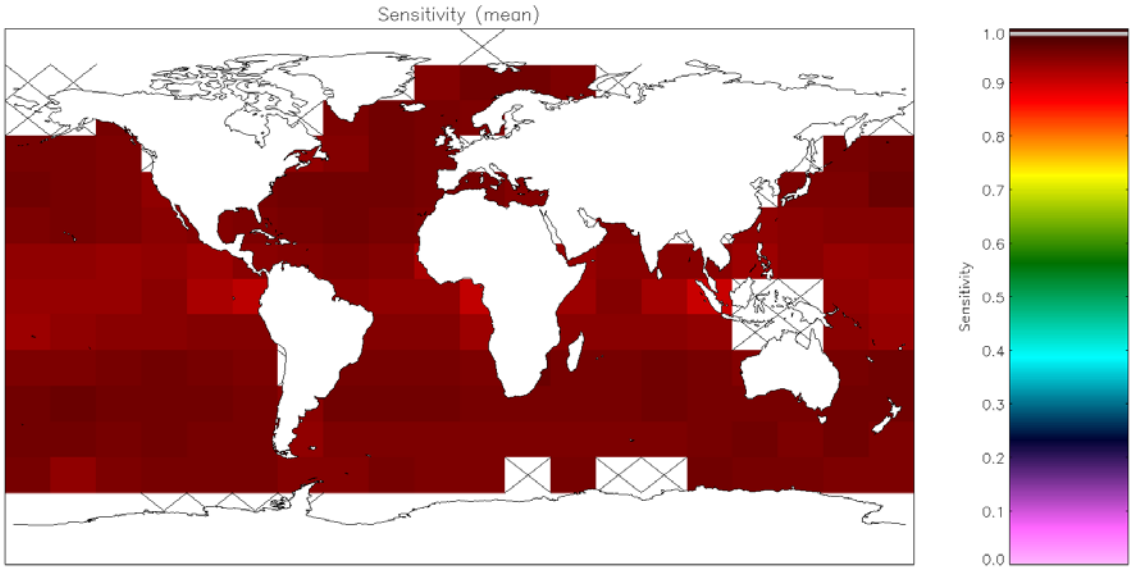


Figure A2.5.3-15: Global map of SST sensitivity using incremental regression retrieval for night-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.7.1**). As **Figure A2.5.3-7** but with OE v2 QC filtering applied.

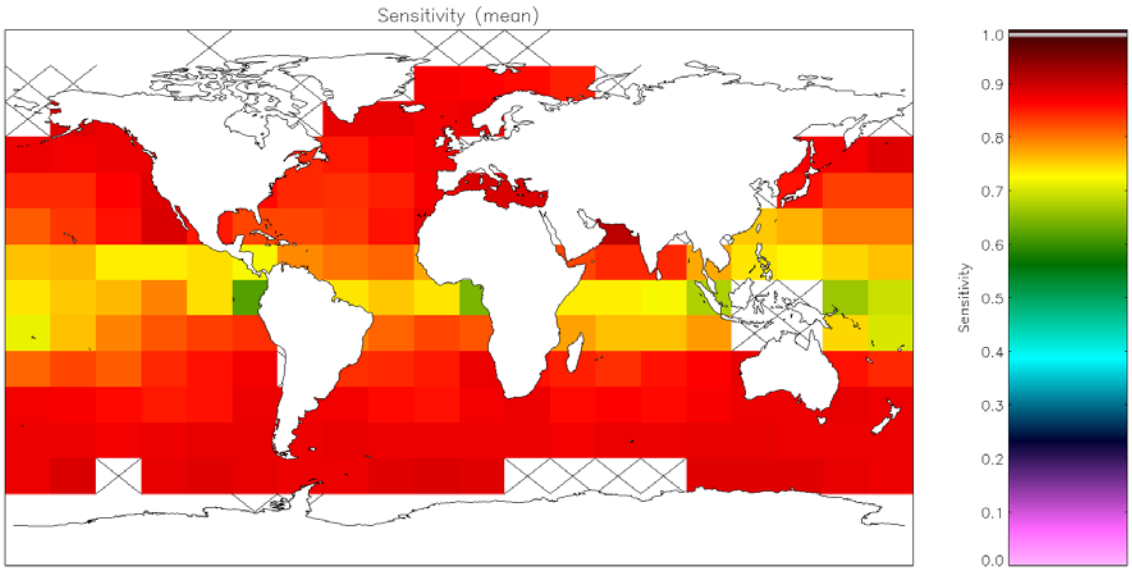


Figure A2.5.3-16: Global map of SST sensitivity using incremental regression retrieval for day-time AVHRR-17 observations. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell (**Table 5.7.2**). As **Figure A2.5.3-8** but with OE v2 QC filtering applied.

A3. AVHRR results with solar zenith angle filtering applied

A3.1 Motivation for additional analysis

The results in this appendix A3 are metrics for the AVHRR algorithms within the SST CCI algorithm selection, when limited to a restricted range of solar zenith angle. The motivation is, as noted in Section 5.5.3, that (i) both OEv2 and IR show out-of-target biases (more negative than 0.1 K) and higher noise for some geographical regions, and (ii) when plotted against solar zenith angle, there is a clear dependence of noise and bias on solar zenith angle, particularly for NOAA19 and NOAA18 AVHRR. This means that neither OEv2 or IR meet SST CCI retrieval quality aspirations for the full range of solar zenith angles. Therefore, SST CCI products will output SST with the best quality flag only for solar zenith angles between 20° and 160°. The tables and figures in this Appendix show the degree of improvement achieved by this means. They also verify that this step does not change the algorithm selection conclusions made on the basis of the full solar zenith angle range.

A3.2 Summary Tables

A3.2.1 Metop

Table A3.2-1: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle > 160° are excluded).

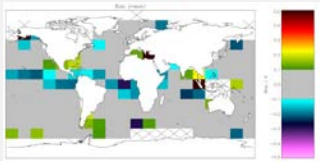
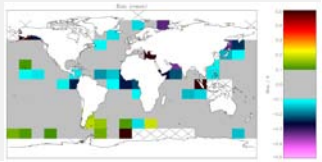
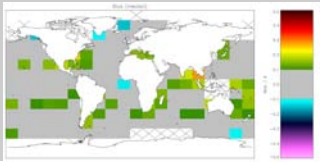
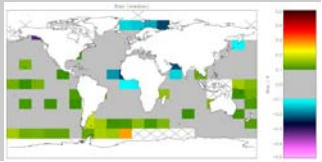
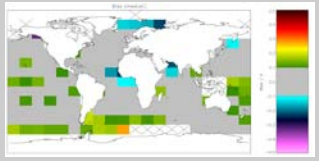
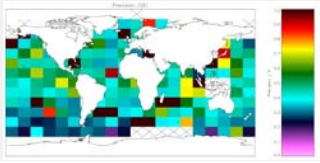
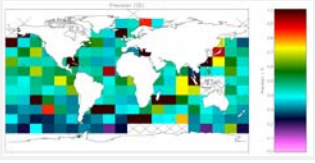
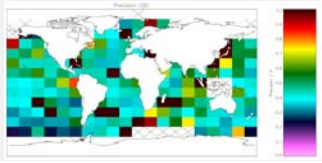
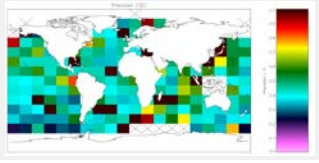
Table A3.2-1 **Metop Night** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Bias (mean discrepancy)	-0.001 K	-0.001 K	-0.011 K	-0.011 K	Very High
Bias (median discrepancy)	0.069 K	0.069 K	0.060 K	0.060 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

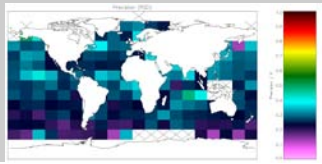
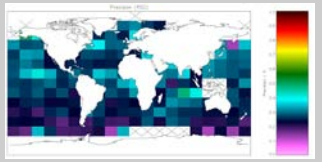
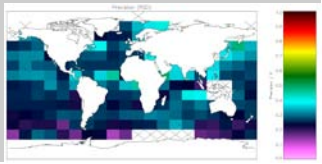
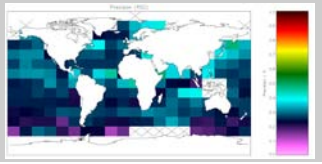
Table A3.2-1 **Metop Night** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.653 K	0.654 K	0.648 K	0.648 K	Medium
Precision (median of cell RSDs)	0.272 K	0.272 K	0.280 K	0.279 K	Medium
Number of Matches	56267	56175	69332	69220	

Table A3.2-1: Comparison of metrics related to SST estimation for night-time AVHRR onboard Metop with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle > 160° are excluded).

Table A3.2-2: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle < 20° are excluded).

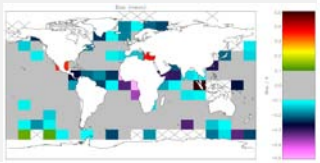
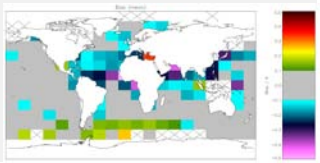
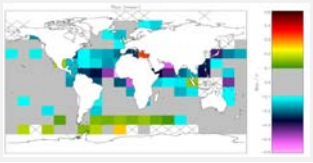
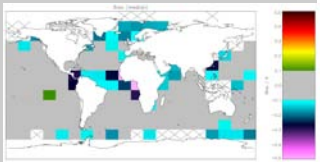
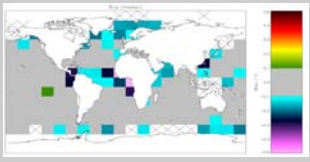
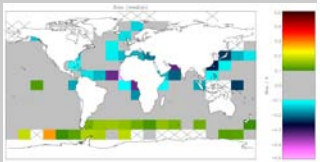
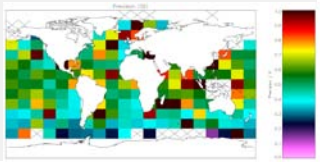
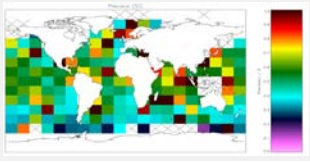
Table A3.2-2 **Metop Day** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Bias (mean discrepancy)	-0.050 K	-0.050 K	-0.079 K	-0.079 K	Very High
Bias (median discrepancy)	-0.021 K	-0.021 K	-0.036 K	-0.036 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

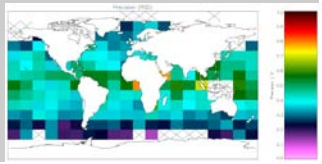
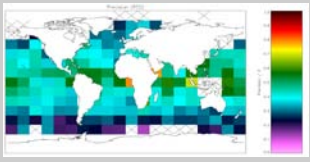
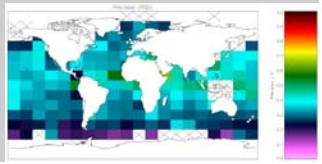
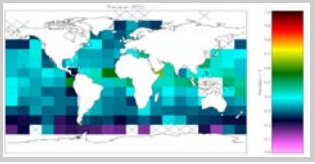
Table A3.2-2 **Metop Day** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.687 K	0.687 K	0.647 K	0.648 K	Medium
Precision (median of cell RSDs)	0.400 K	0.400 K	0.345 K	0.345 K	Medium
Number of Matches	58363	58275	74162	74048	

Table A3.2-2: Comparison of metrics related to SST estimation for day-time AVHRR onboard Metop with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle < 20° are excluded).

A3.2.2 NOAA-19

Table A3.2-3: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle > 160° are excluded).

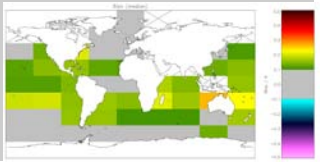
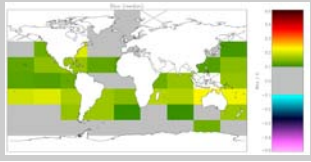
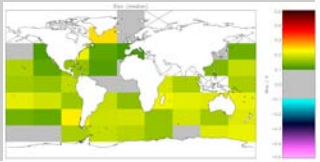
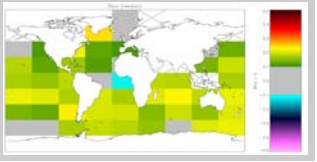
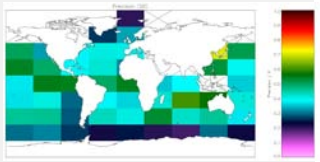
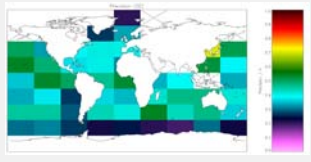
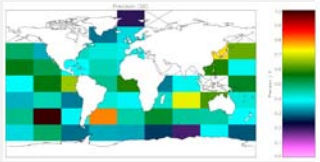
Table A3.2-3 NOAA-19 Night (SZA filtering)	Optimal Estimation v2	Optimal Estimation v2 (with SZA filter)	Incremental Regression	Incremental Regression (with SZA filter)	Weight
Bias (mean discrepancy)	0.052 K	0.052 K	0.072 K	0.076 K	Very High
Bias (median discrepancy)	0.126 K	0.124 K	0.155 K	0.159 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

Table A3.2-3 **NOAA-19 Night** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.441 K	0.429 K	0.502 K	0.401 K	Medium
Precision (median of cell RSDs)	0.280 K	0.275 K	0.292 K	0.268 K	Medium
Number of Matches	13693	12679	17202	15963	

Table A3.2-3: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-19 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle > 160° are excluded).

Table A3.2-4: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle < 20° are excluded).

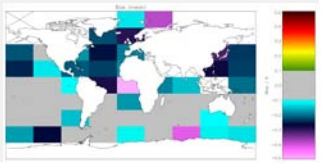
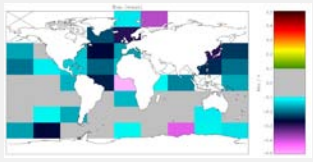
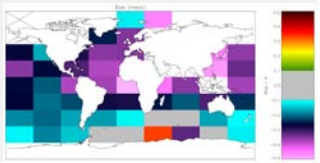
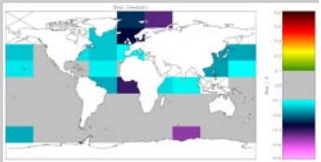
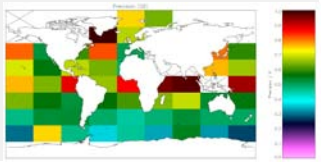
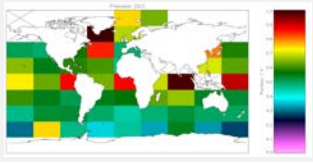
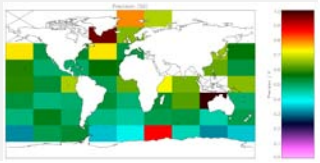
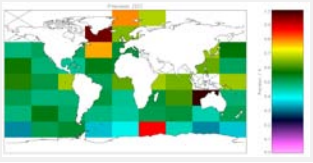
Table A3.2-4 **NOAA-19 Day** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Bias (mean discrepancy)	-0.161 K	-0.144 K	-0.295 K	-0.288 K	Very High
Bias (median discrepancy)	-0.076 K	-0.067 K	-0.206 K	-0.201 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

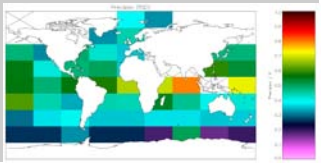
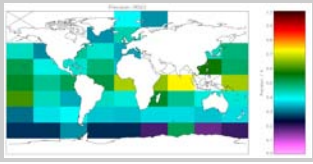
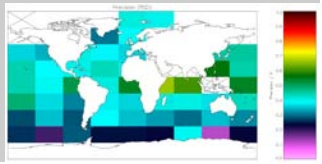
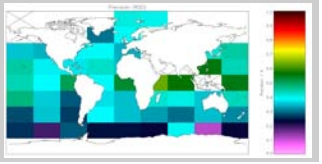
Table A3.2-4 **NOAA-19 Day** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.658 K	0.629 K	0.588 K	0.579 K	Medium
Precision (median of cell RSDs)	0.462 K	0.446 K	0.415 K	0.409 K	Medium
Number of Matches	18381	16947	24948	23056	

Table A3.2-4: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-19 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle < 20° are excluded).

A3.2.3 NOAA-18

Table A3.2-5: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle > 160° are excluded).

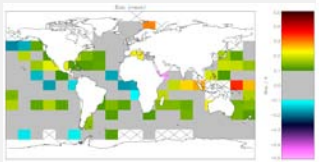
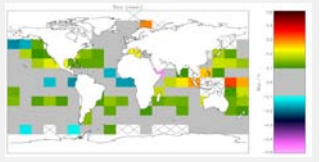
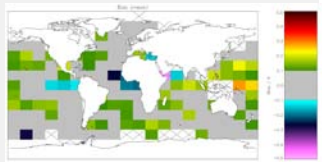
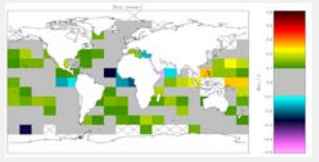
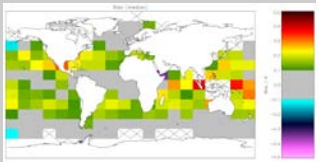
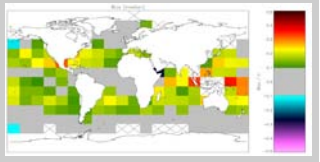
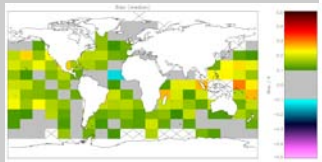
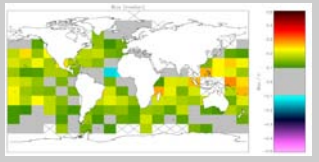
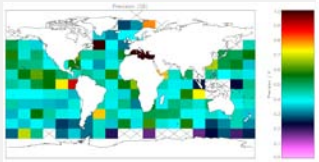
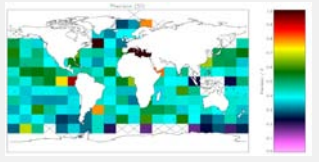
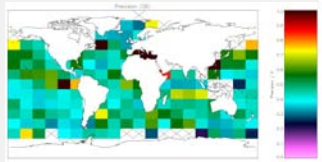
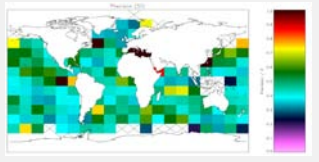
Table A3.2-5 **NOAA-18 Night** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Bias (mean discrepancy)	0.074 K	0.074 K	0.071 K	0.076 K	Very High
Bias (median discrepancy)	0.147 K	0.145 K	0.148 K	0.152 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

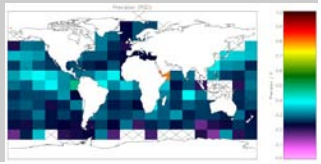
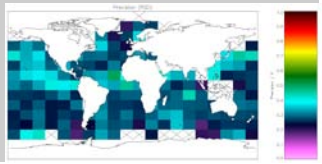
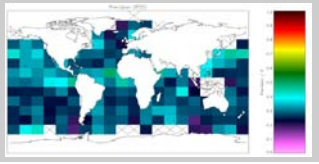
Table A3.2-5 **NOAA-18 Night** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.477 K	0.474 K	0.519 K	0.520 K	Medium
Precision (median of cell RSDs)	0.289 K	0.286 K	0.298 K	0.295 K	Medium
Number of Matches	49505	44721	61250	55535	

Table A3.2-5: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-18 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle > 160° are excluded).

Table A3.2-6: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle < 20° are excluded).

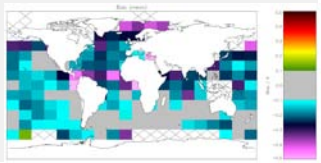
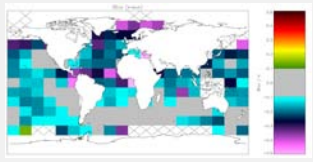
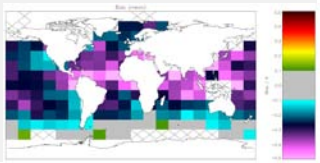
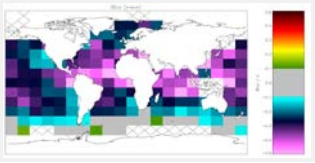
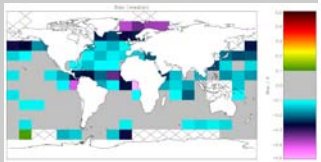
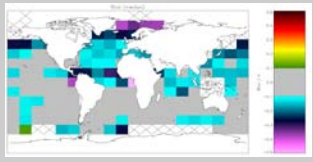
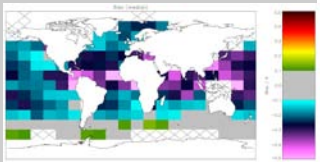
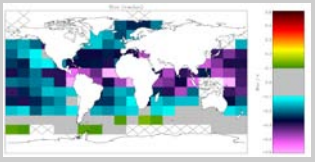
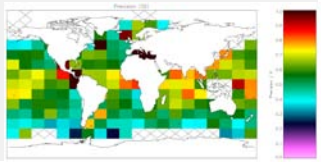
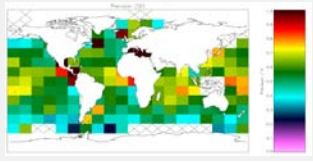
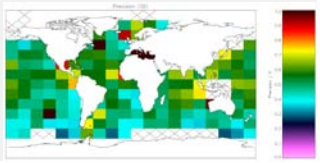
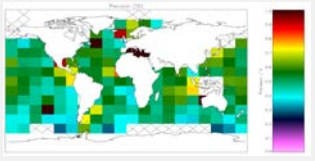
Table A3.2-6 **NOAA-18 Day** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Bias (mean discrepancy)	-0.161 K	-0.152 K	-0.291 K	-0.288 K	Very High
Bias (median discrepancy)	-0.090 K	-0.086 K	-0.214 K	-0.211 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

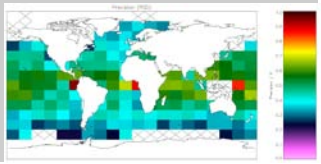
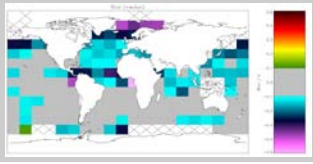
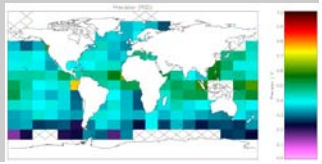
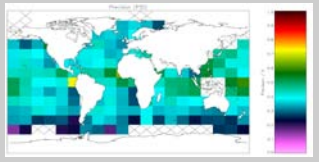
Table A3.2-6 **NOAA-18 Day** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.687 K	0.667 K	0.641 K	0.631 K	Medium
Precision (median of cell RSDs)	0.465 K	0.453 K	0.409 K	0.403 K	Medium
Number of Matches	69579	64257	91370	84140	

Table A3.2-6: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-18 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle < 20° are excluded).

A3.2.4 NOAA-17

Table A3.2-7: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle > 160° are excluded).

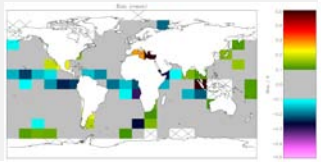
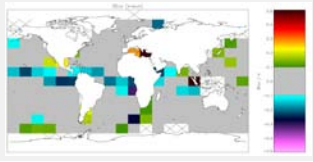
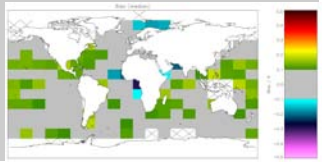
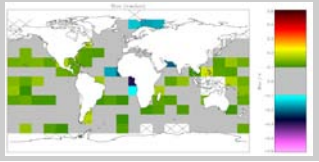
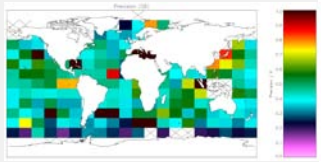
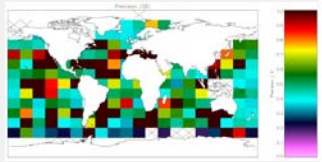
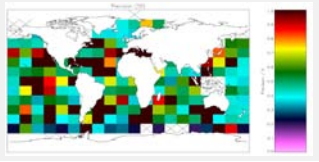
Table A3.2-7 **NOAA-17 Night** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Bias (mean discrepancy)	0.005 K	0.006 K	0.007 K	0.009 K	Very High
Bias (median discrepancy)	0.077 K	0.077 K	0.080 K	0.082 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

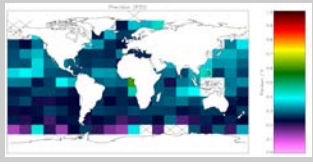
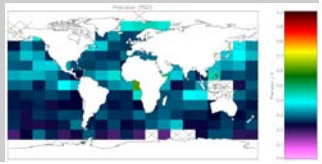
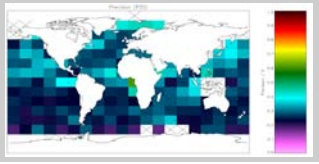
Table A3.2-7 **NOAA-17 Night** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.599 K	0.603 K	0.910 K	0.923 K	Medium
Precision (median of cell RSDs)	0.276 K	0.275 K	0.282 K	0.282 K	Medium
Number of Matches	63108	60437	80203	76974	

Table A3.2-7: Comparison of metrics related to SST estimation for night-time AVHRR onboard NOAA-17 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle > 160° are excluded).

Table A3.2-8: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle < 20° are excluded).

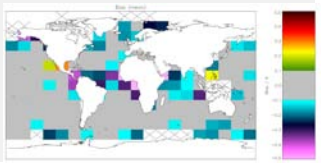
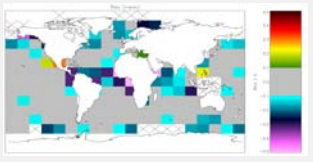
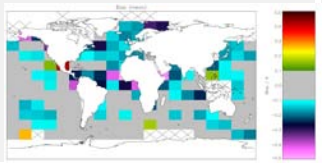
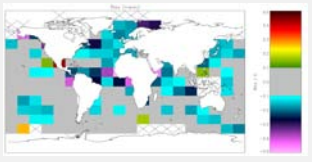
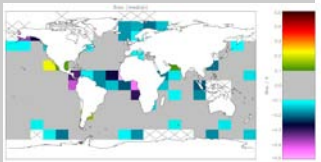
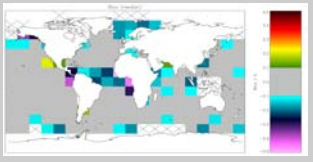
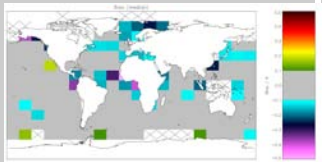
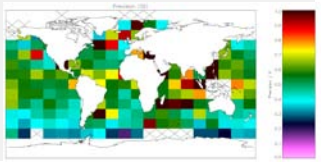
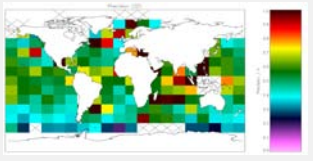
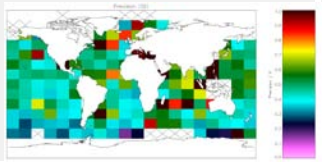
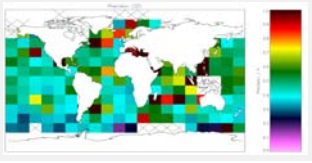
Table A3.2-8 **AVHRR-17 Day** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Bias (mean discrepancy)	-0.050 K	-0.048 K	-0.091 K	-0.089 K	Very High
Bias (median discrepancy)	-0.021 K	-0.020 K	-0.048 K	-0.047 K	Very High
Bias (mean discrepancy map)					Very High
Bias (median discrepancy map)					Very High
Precision map (SD of discrepancy)					Medium

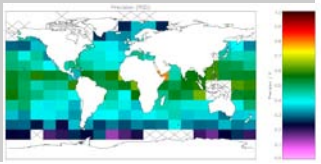
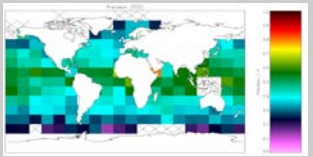
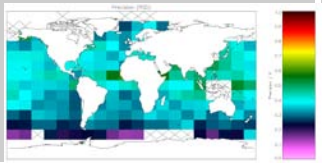
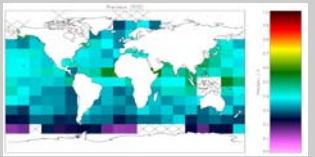
Table A3.2-8 **AVHRR-17 Day** **(SZA filtering)**	**Optimal Estimation v2**	**Optimal Estimation v2 (with SZA filter)**	**Incremental Regression**	**Incremental Regression (with SZA filter)**	**Weight**
Precision map (RSD of discrepancy)					Medium
Precision (mean of cell SDs)	0.682 K	0.679 K	0.641 K	0.638 K	Medium
Precision (median of cell RSDs)	0.421 K	0.417 K	0.366 K	0.363 K	Medium
Number of Matches	71778	69490	92276	89219	

Table A3.2-8: Comparison of metrics related to SST estimation for day-time AVHRR onboard NOAA-17 with and without solar zenith angle filtering applied ("with SZA filter" indicates match-ups with solar zenith angle < 20° are excluded).

A3.3 Full Resolution Images

A3.3.1 Metop

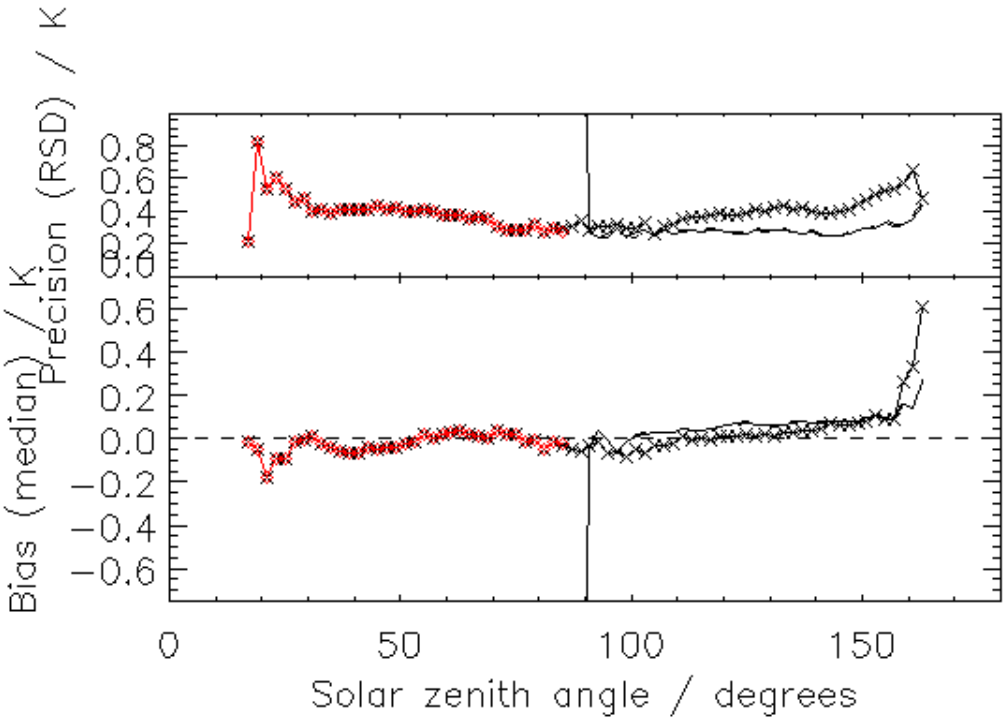


Figure A3.3.1-1: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for OE v2 retrievals for AVHRR onboard Metop. The meaning of different lines/symbols is as follows: the line with symbols represents N2 retrievals (with red used to highlight daytime retrievals), the solid line (with no symbols) represents N3 retrievals, and the dashed line represents N2* (3.7 µm and 11 µm channel) retrievals.

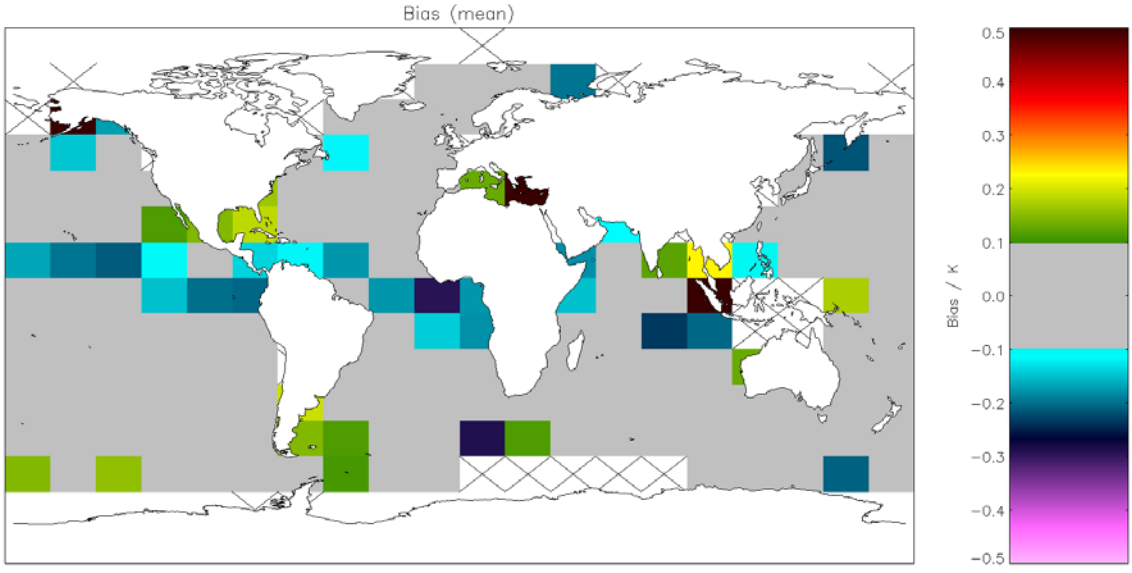


Figure A3.3.1-2: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses

indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.001 K (**Table A3.2-1**).

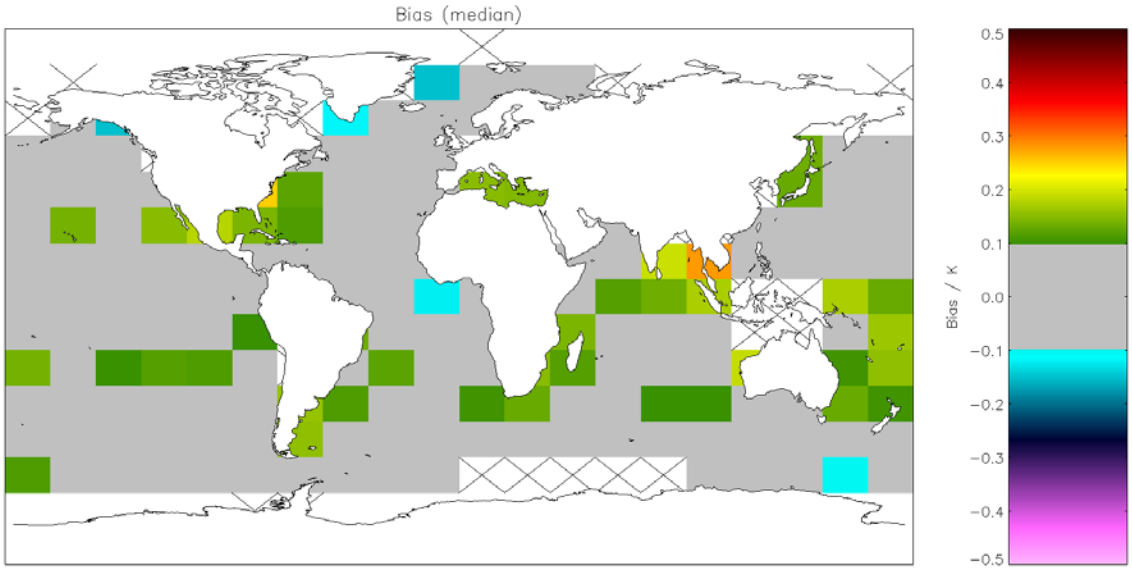


Figure A3.3.1-3: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.069 K (**Table A3.2-1**).

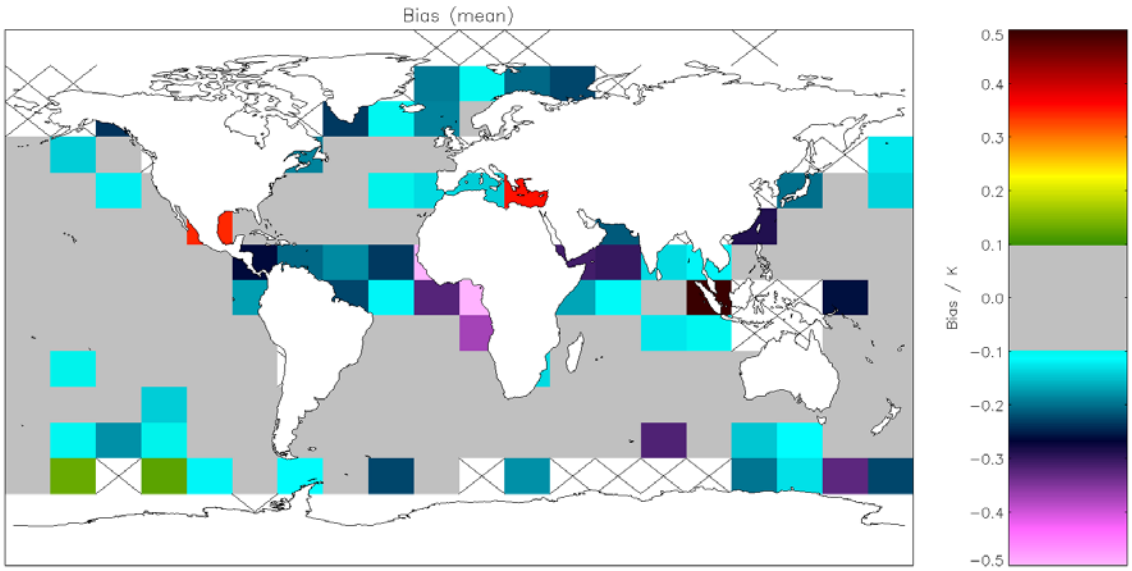


Figure A3.3.1-4: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-Metop observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.050 K (**Table A3.2-2**).

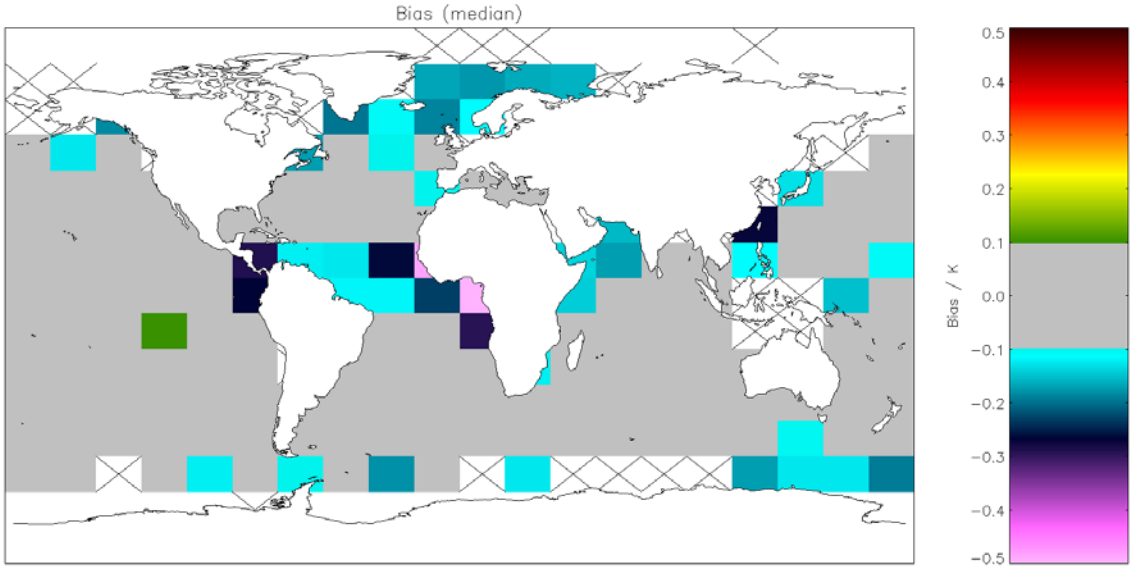


Figure A3.3.1-5: Global map of median bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-Metop observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.050 K (**Table A3.2-2**).

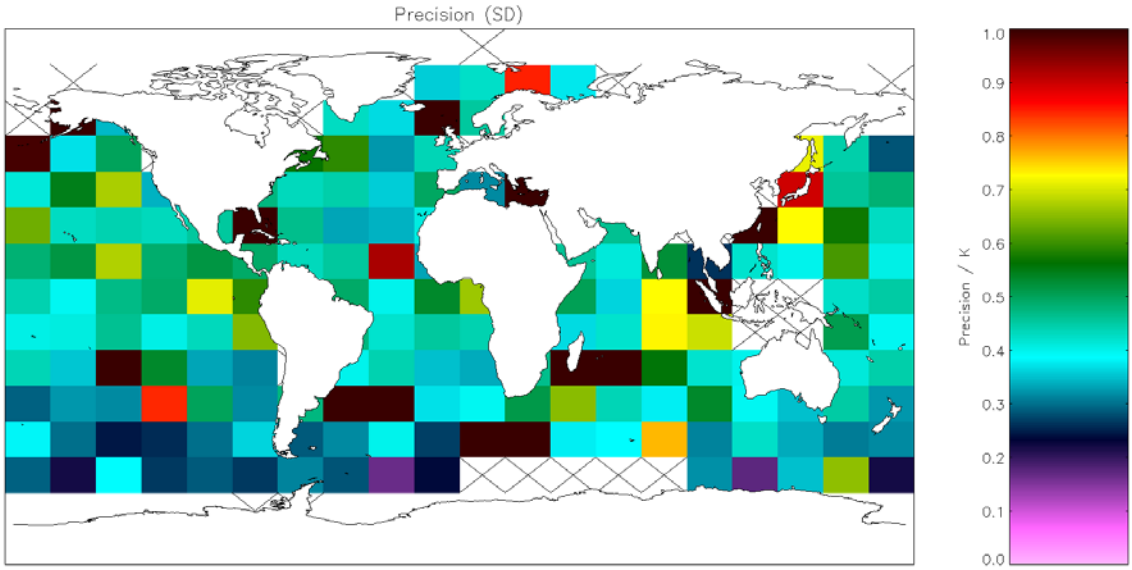


Figure A3.3.1-6: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.654 K (**Table A3.2-1**).

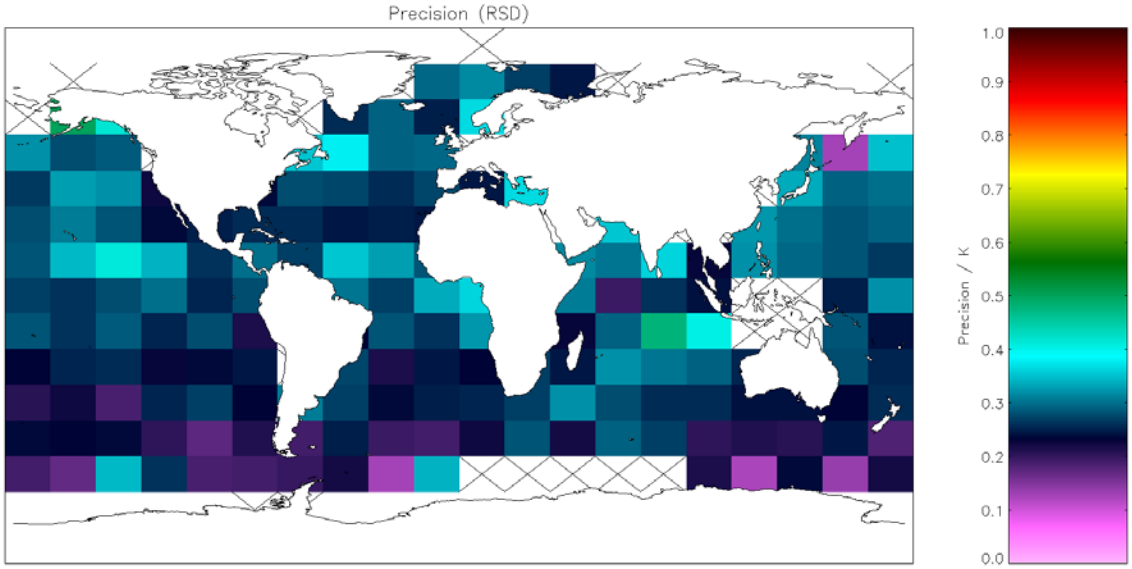


Figure A3.3.1-7: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.687 K (**Table A3.2-1**).

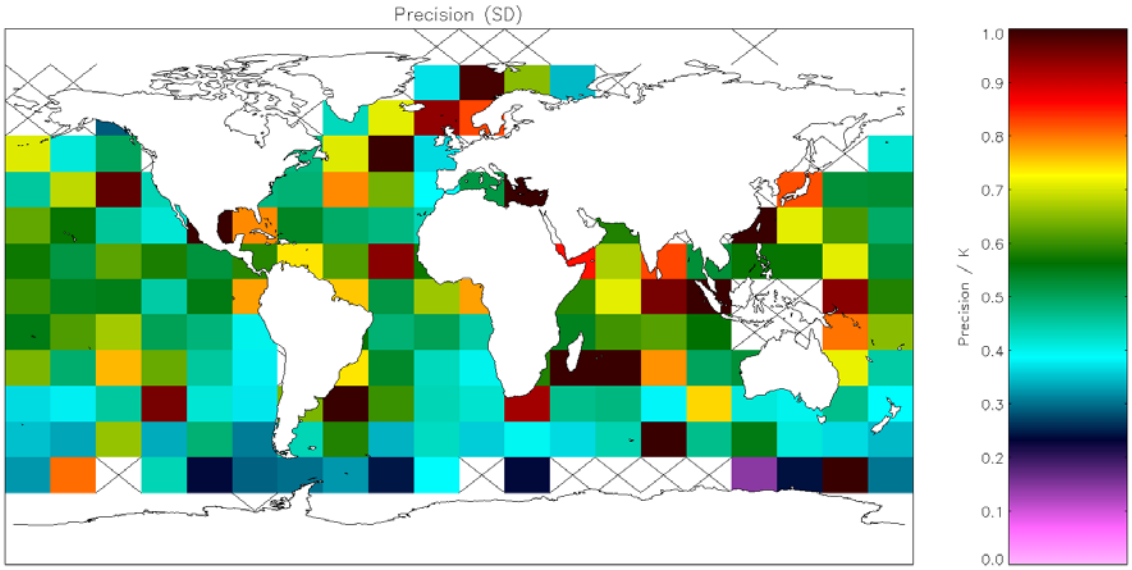


Figure A3.3.1-8: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.687 K (**Table A3.2-2**).

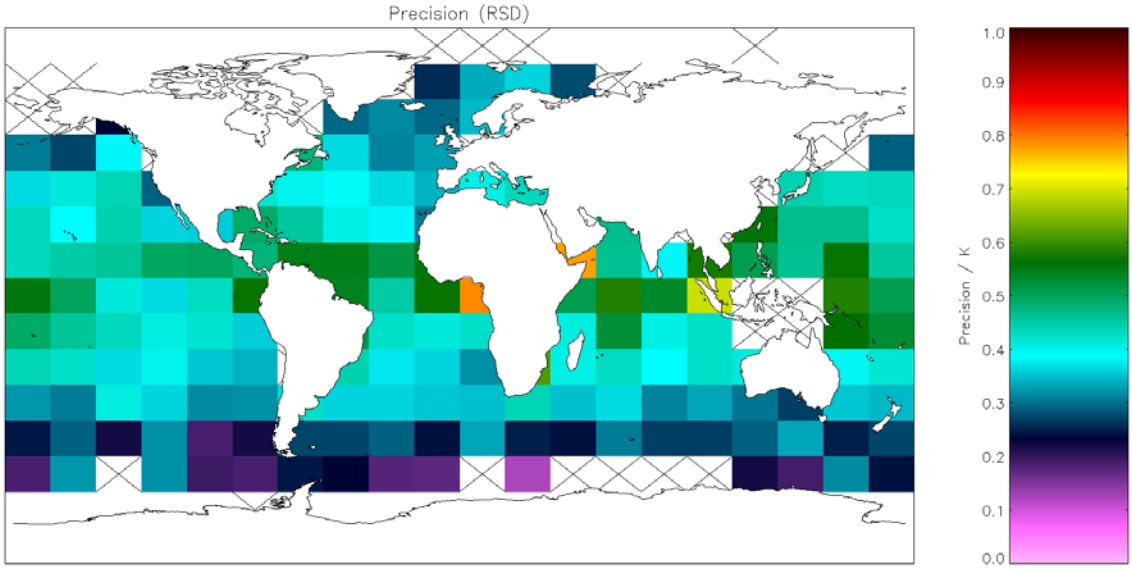


Figure A3.3.1-9: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.400 K (**Table A3.2-2**).

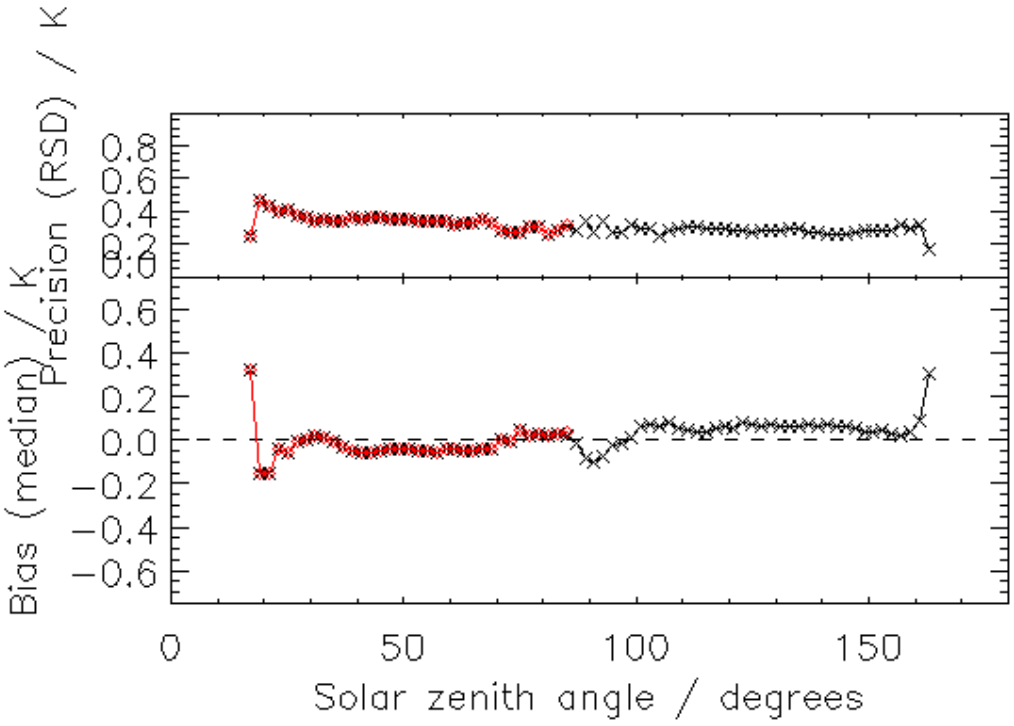


Figure A3.3.1-10: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for incremental regression retrievals for AVHRR onboard Metop. The red and black lines with symbols represent retrievals using the daytime and night-time formulations of the incremental regression retrieval respectively.

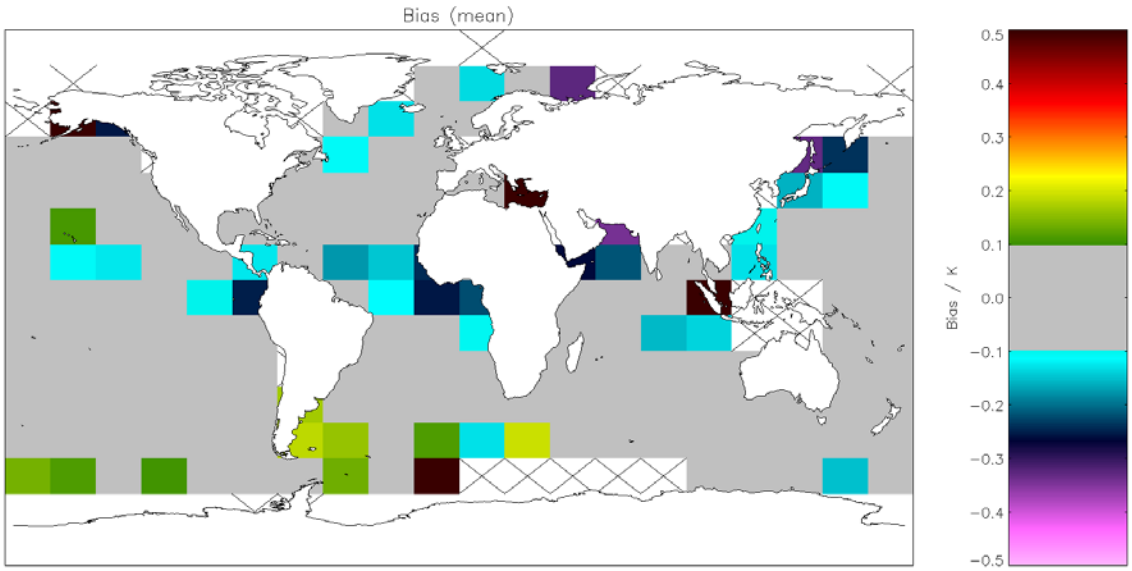


Figure A3.3.1-11: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.011 K (**Table A3.2-1**).

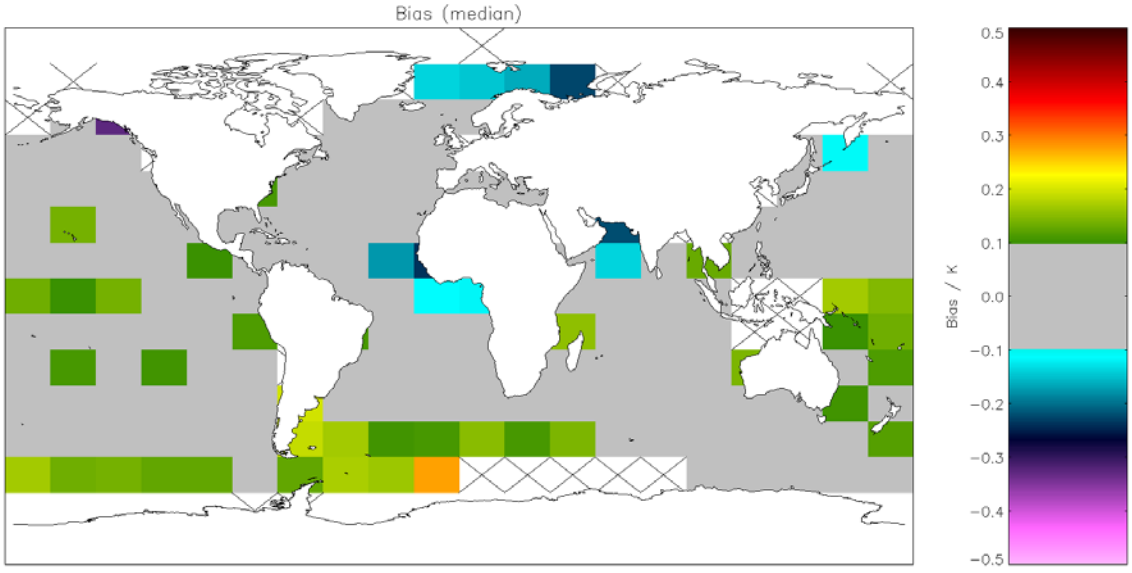


Figure A3.3.1-12: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.060 K (**Table A3.2-1**).

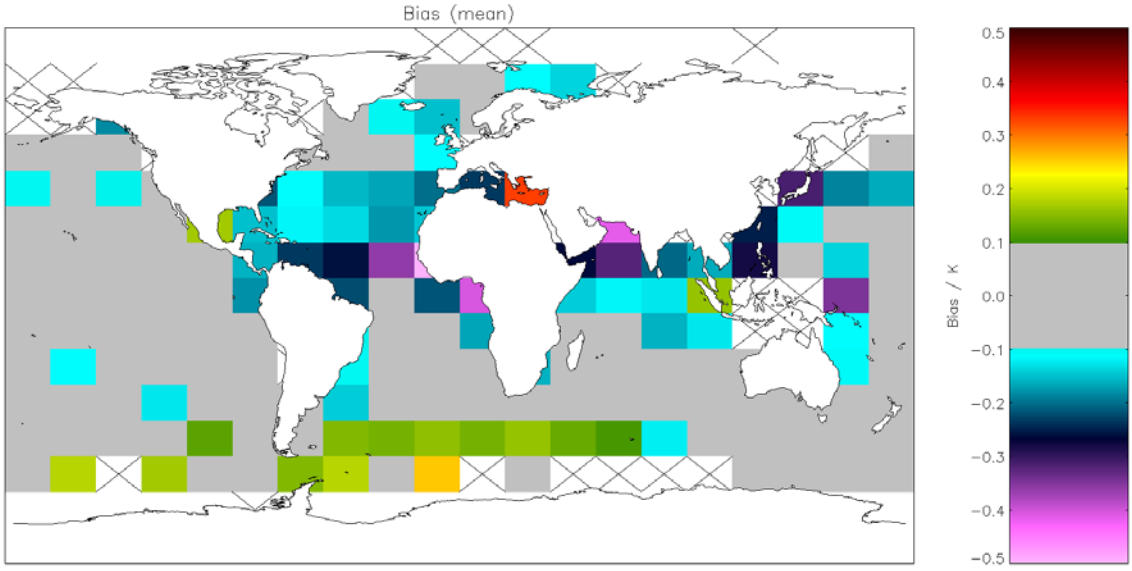


Figure A3.3.1-13: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-Metop observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.079 K (**Table A3.2-2**).

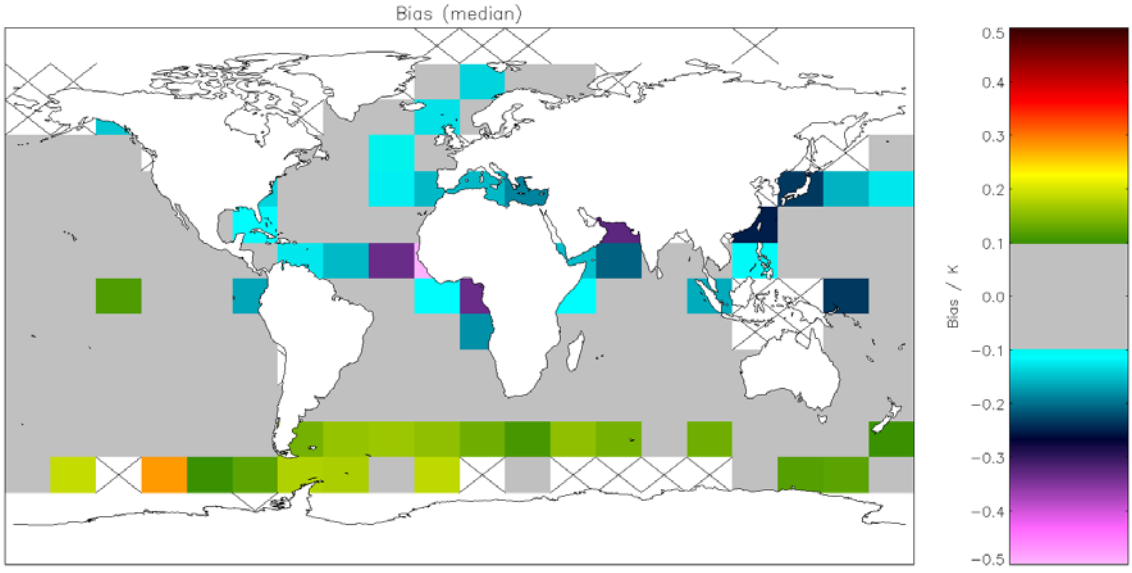


Figure A3.3.1-14: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-Metop observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.036 K (**Table A3.2-2**).

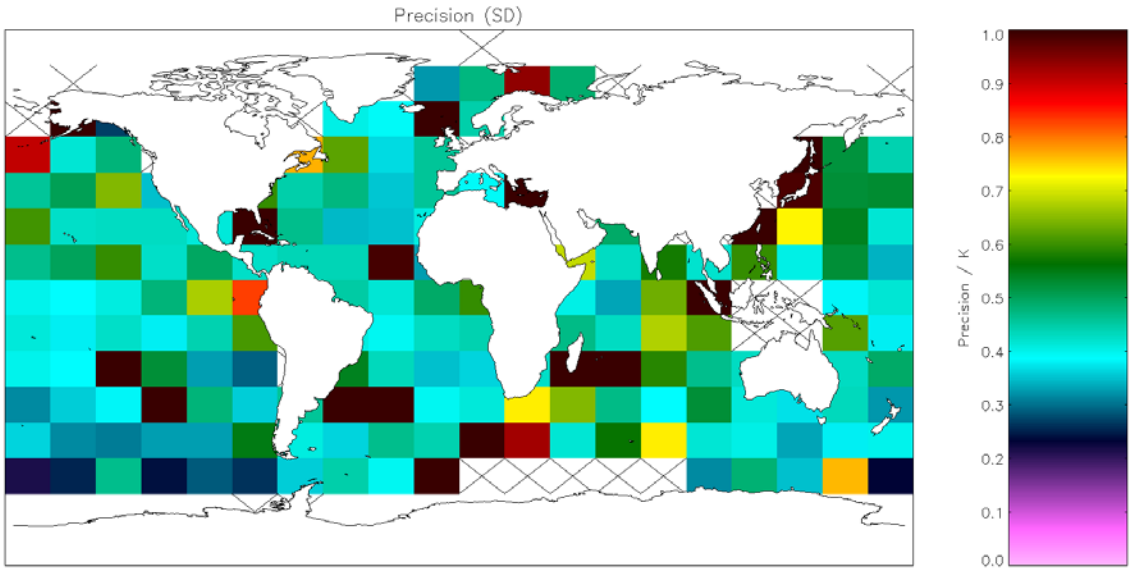


Figure A3.3.1-15: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.648 K (**Table A3.2-1**).

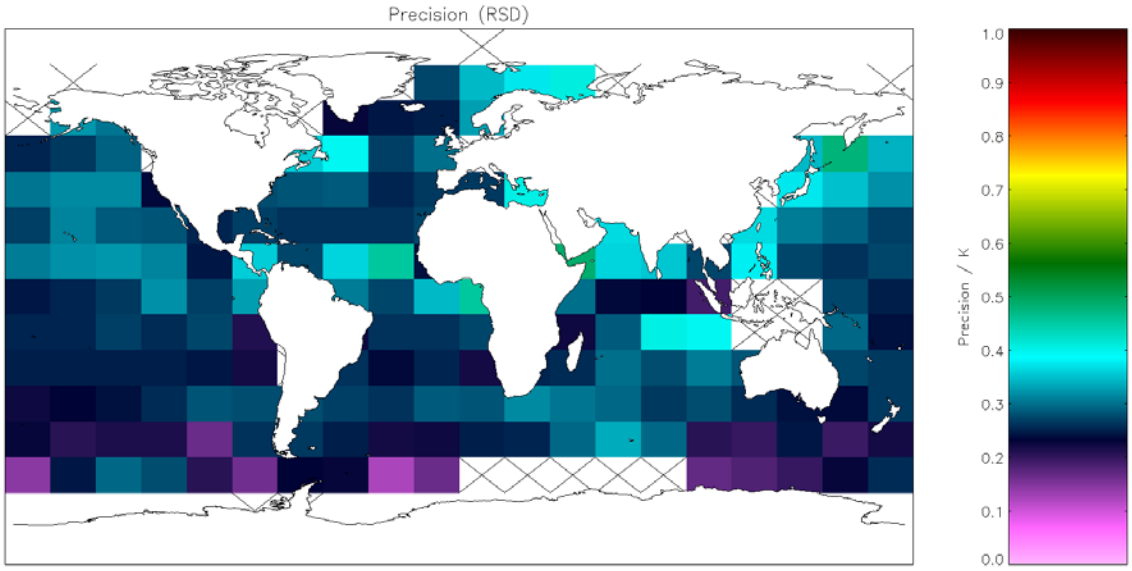


Figure A3.3.1-16: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.279 K (**Table A3.2-1**).

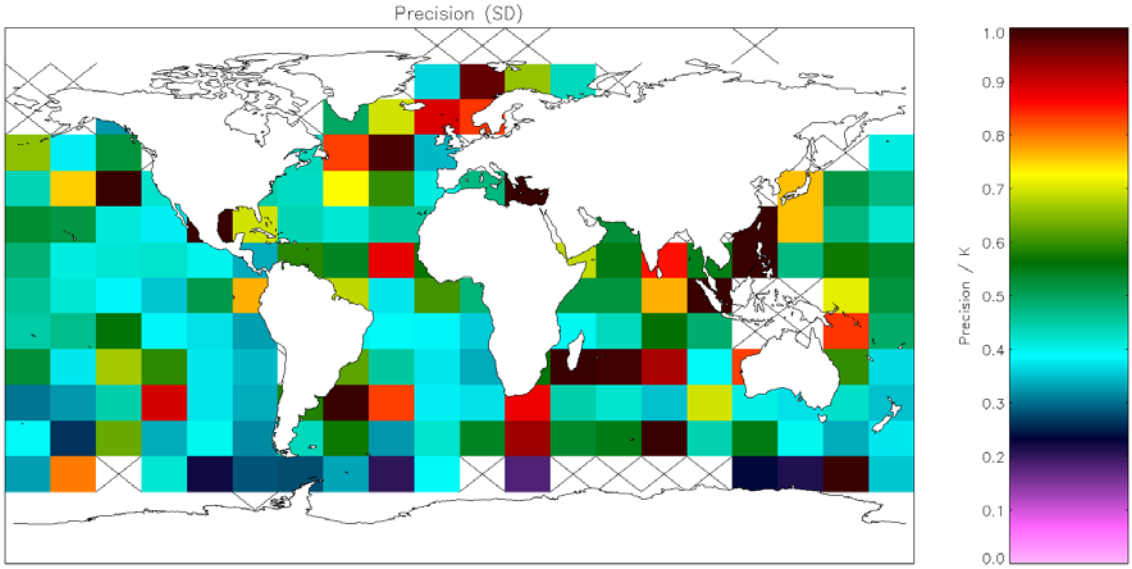


Figure A3.3.1-17: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.648 K (**Table A3.2-2**).

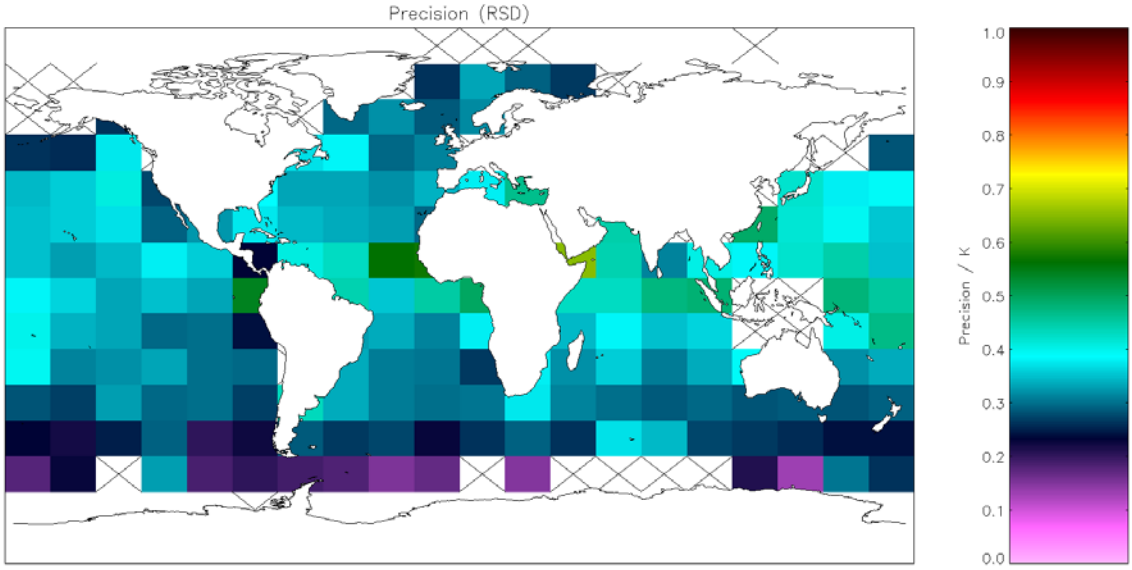


Figure A3.3.1-18: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-Metop observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.345 K (**Table A3.2-2**).

A3.3.2 NOAA-19

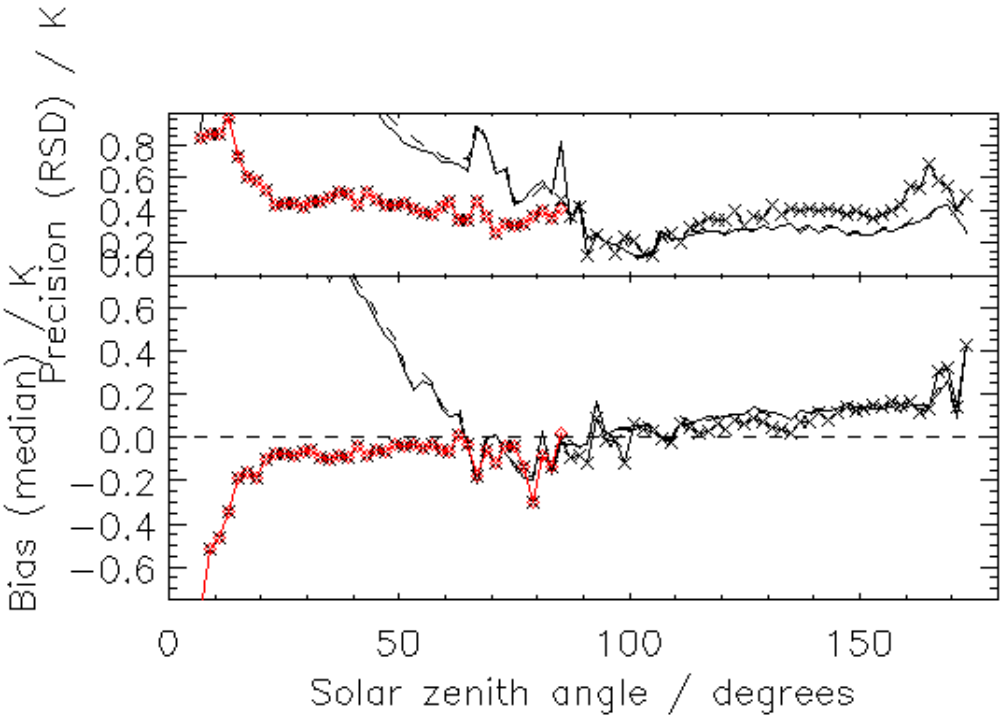


Figure A3.3.2-1: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for OE v2 retrievals for AVHRR onboard NOAA-19. The meaning of different lines/symbols is as follows: the line with symbols represents N2 retrievals (with red used to highlight daytime retrievals), the solid line (with no symbols) represents N3 retrievals, and the dashed line represents N2* (3.7 µm and 11 µm channel) retrievals.

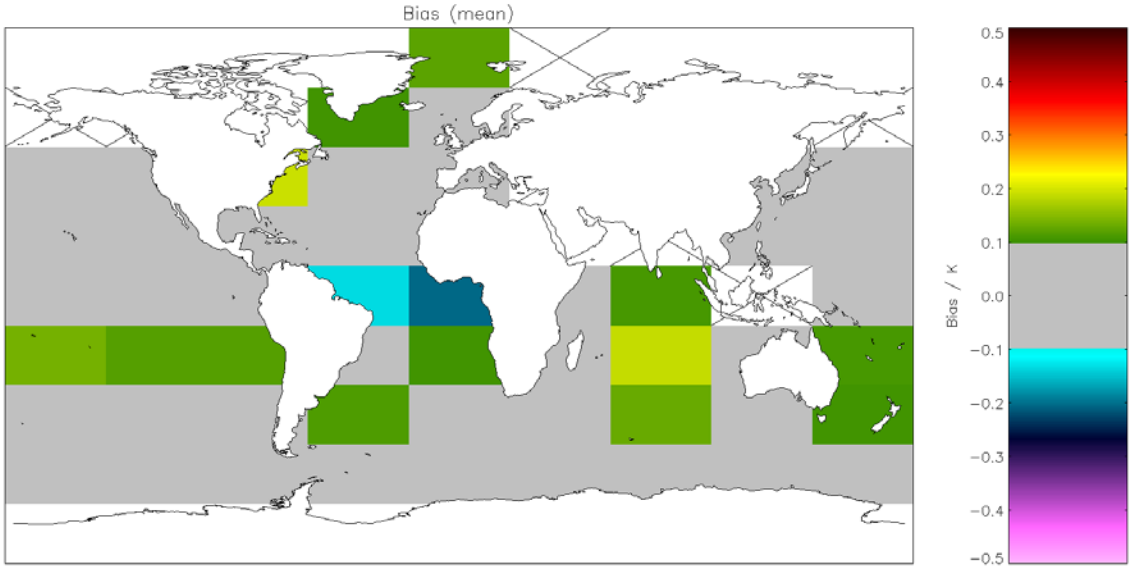


Figure A3.3.2-2: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-19 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.052 K (**Table A3.2-3**).

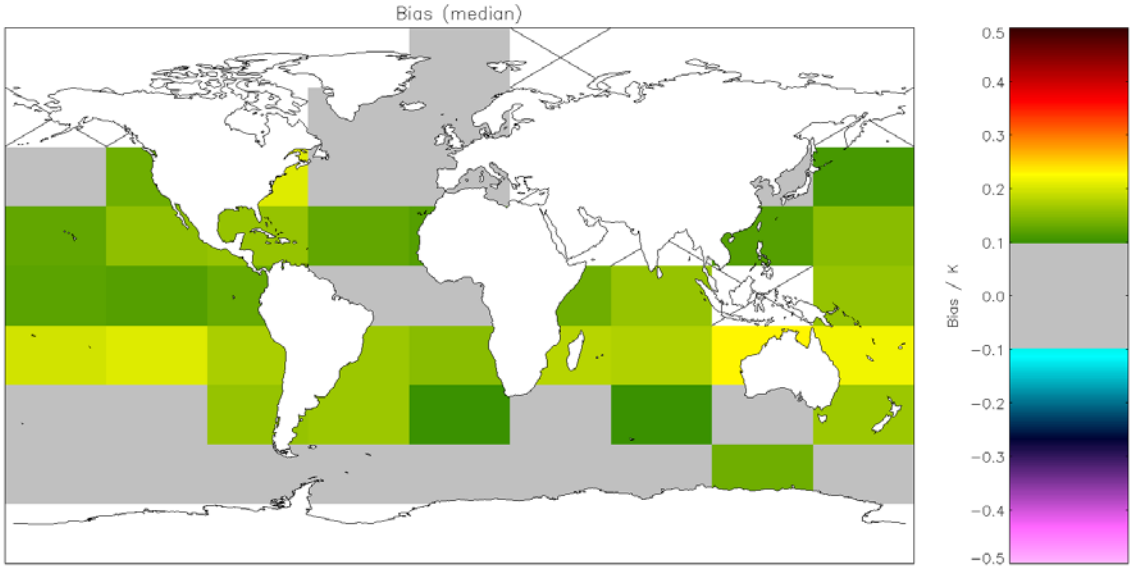


Figure A3.3.2-3: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-19 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.124 K (**Table A3.2-3**).

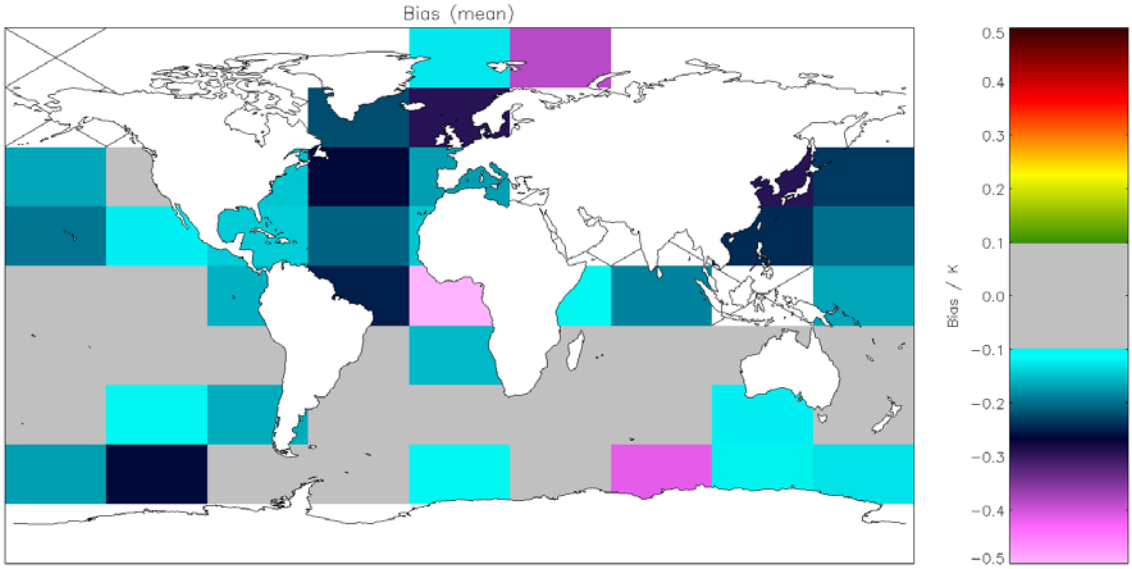


Figure A3.3.2-4: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-19 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.144 K (**Table A3.2-4**).

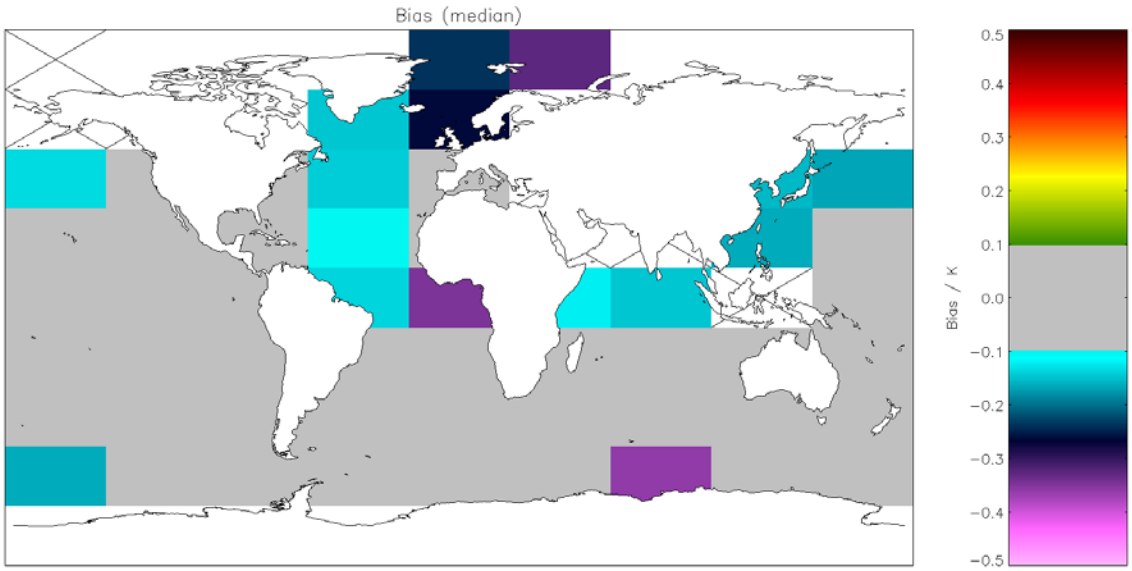


Figure A3.3.2-5: Global map of median bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-19 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.067 K (**Table A3.2-4**).

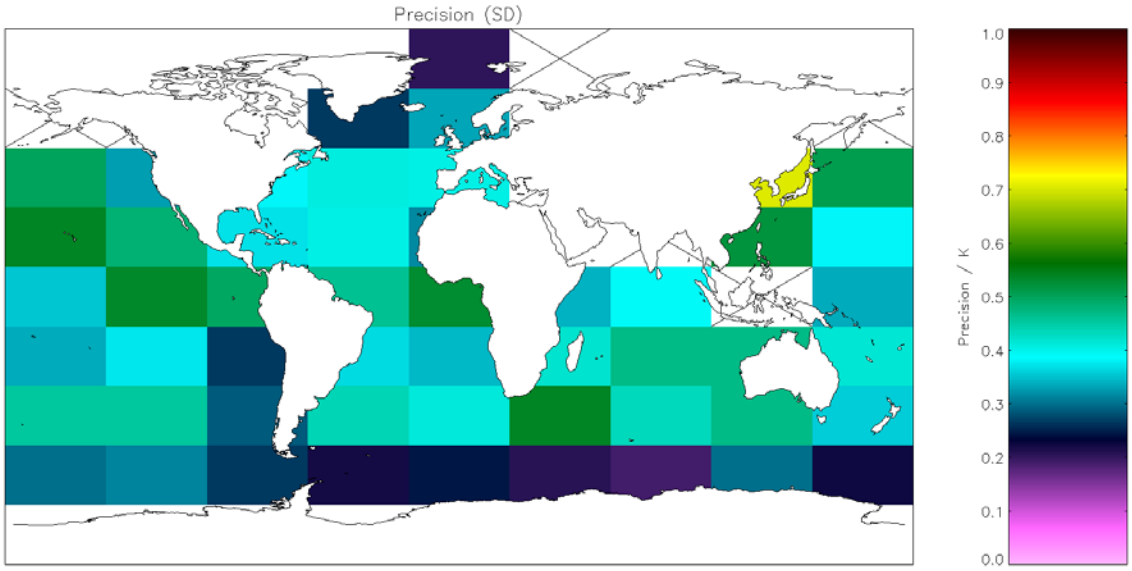


Figure A3.3.2-6: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-19 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.429 K (**Table A3.2-3**).

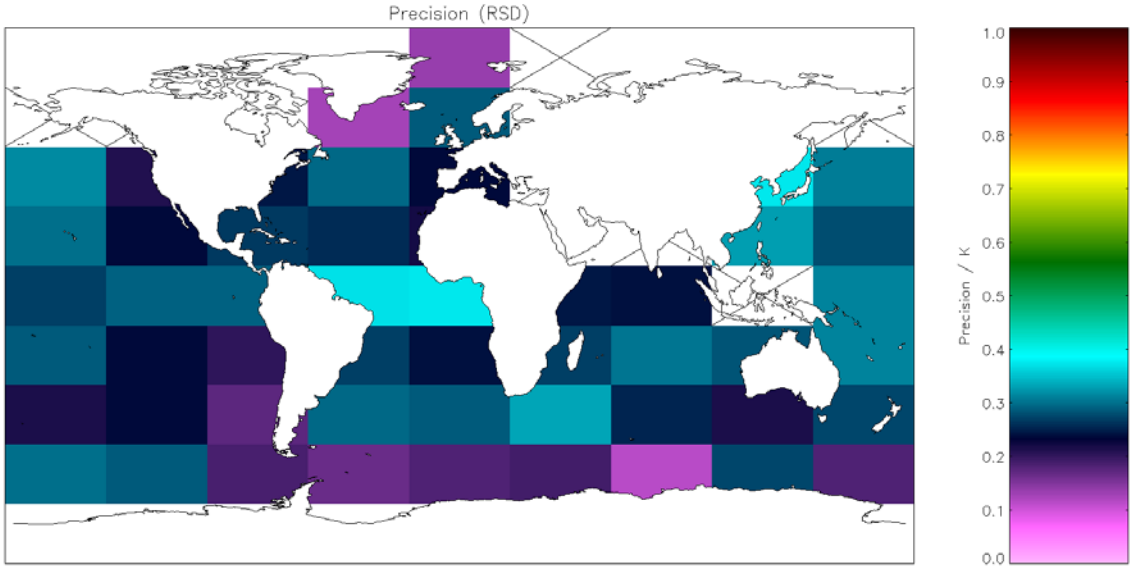


Figure A3.3.2-7: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AVHRR-19 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.275 K (**Table A3.2-3**).

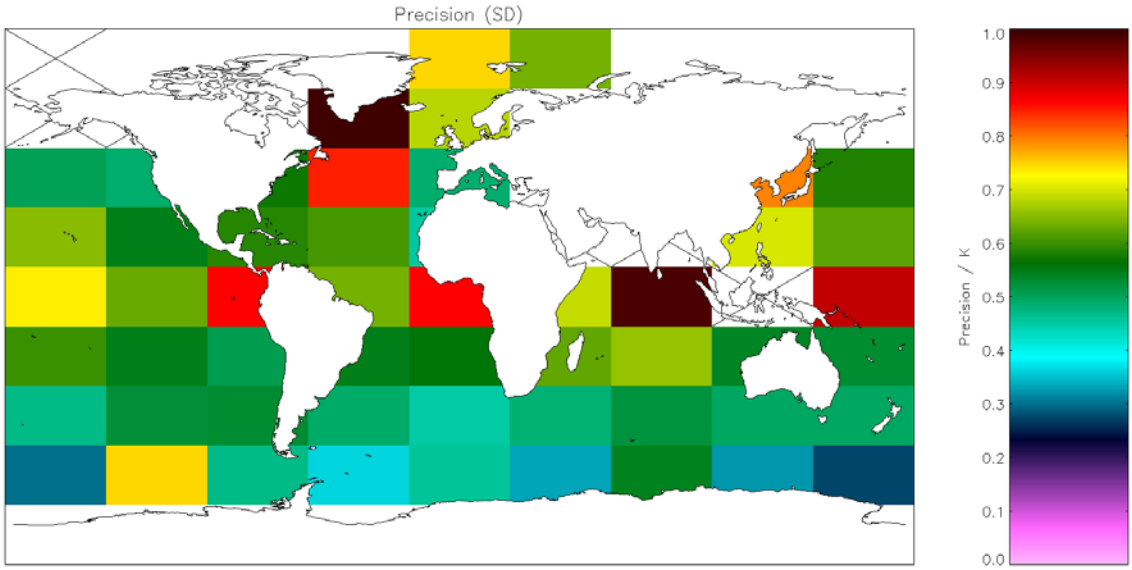


Figure A3.3.2-8: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time AVHRR-19 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.629 K (**Table A3.2-4**).

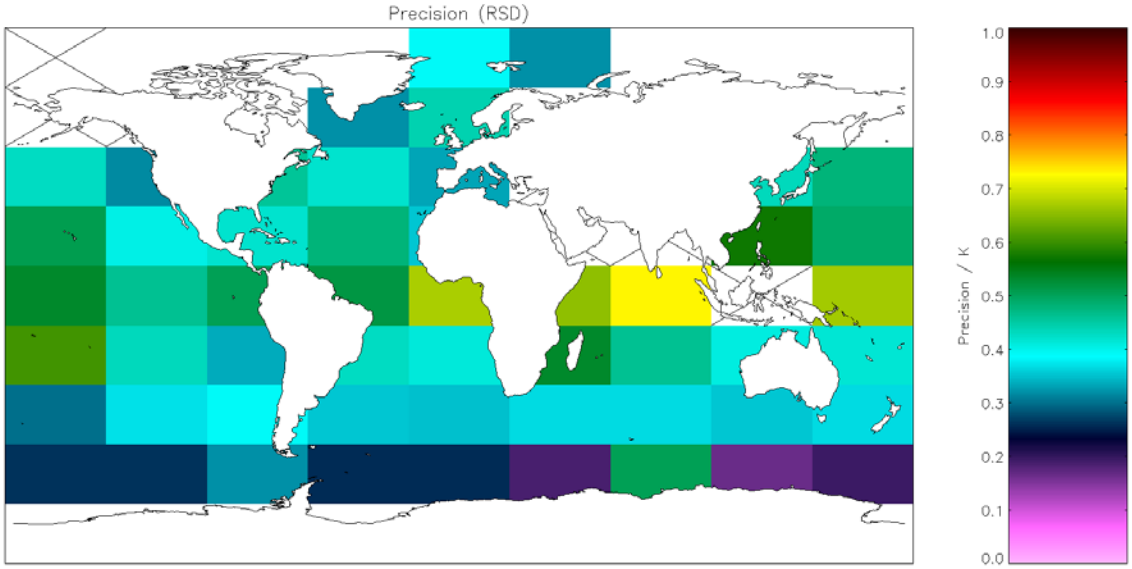


Figure A3.3.2-9: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time AVHRR-19 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.446 K (**Table A3.2-4**).

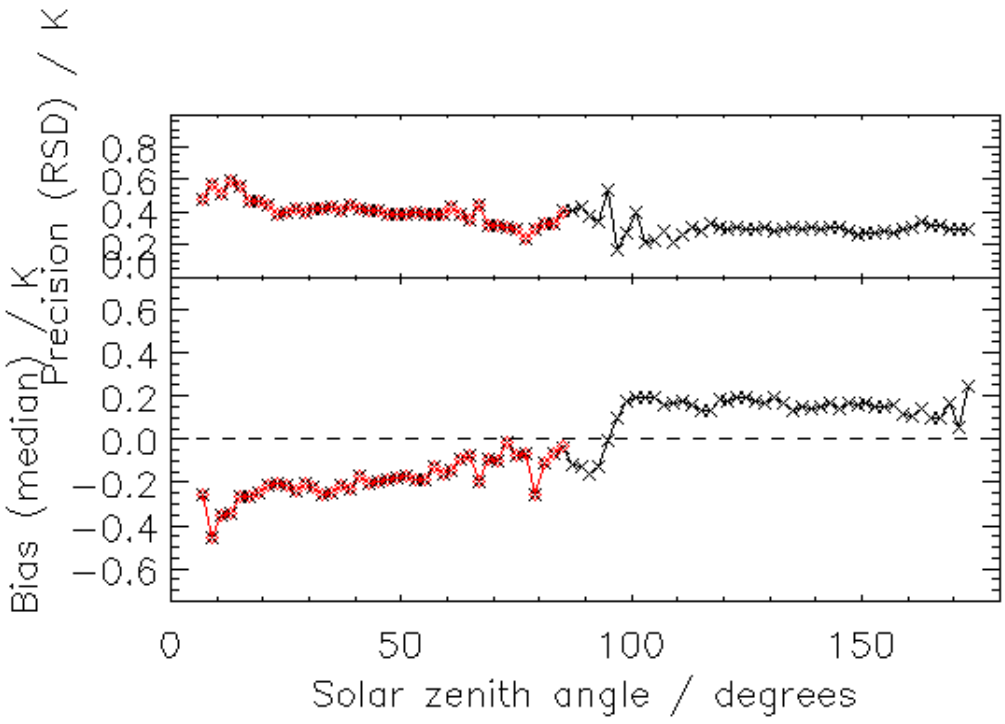


Figure A3.3.2-10: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for incremental regression retrievals for AVHRR onboard NOAA-19. The red and black lines with symbols represent retrievals using the daytime and night-time formulations of the incremental regression retrieval respectively.

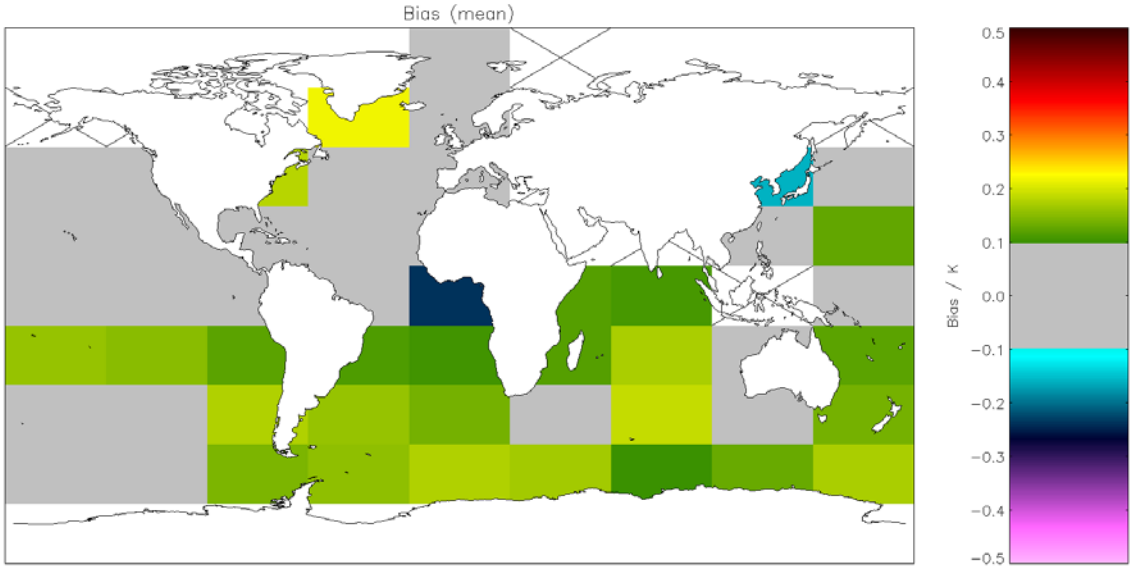


Figure A3.3.2-11: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-19 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.076 K (**Table A3.2-3**).

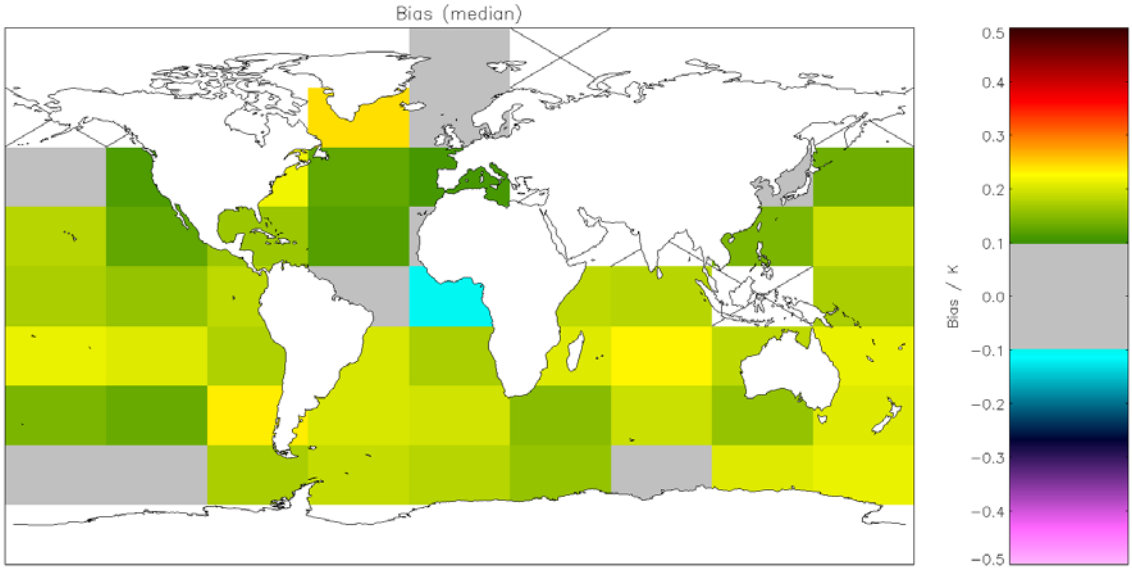


Figure A3.3.2-12: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-19 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.159 K (**Table A3.2-3**).

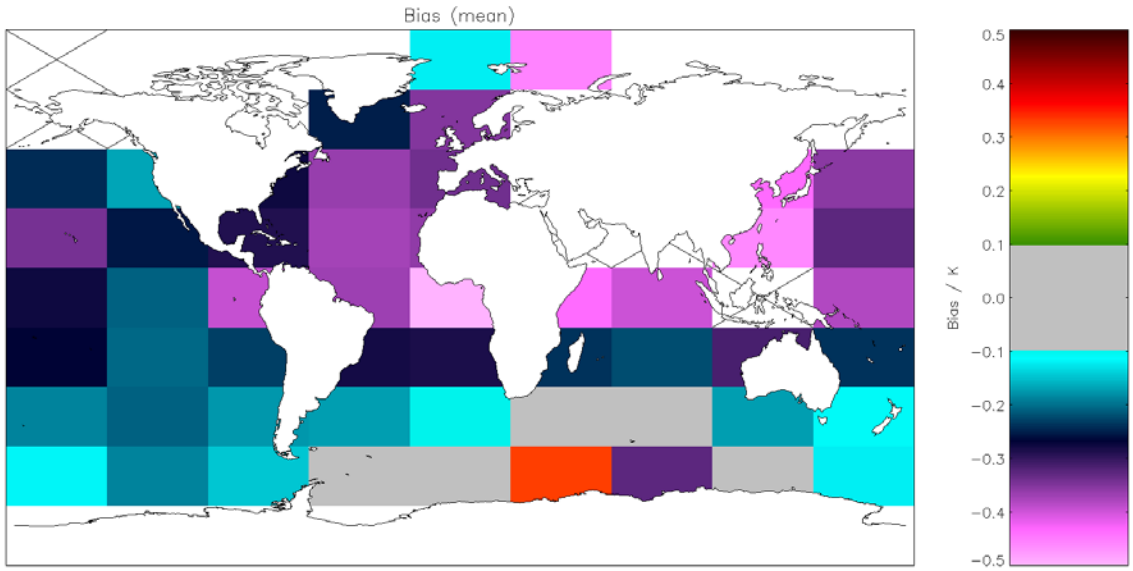


Figure A3.3.2-13: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-19 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.288 K (**Table A3.2-4**).

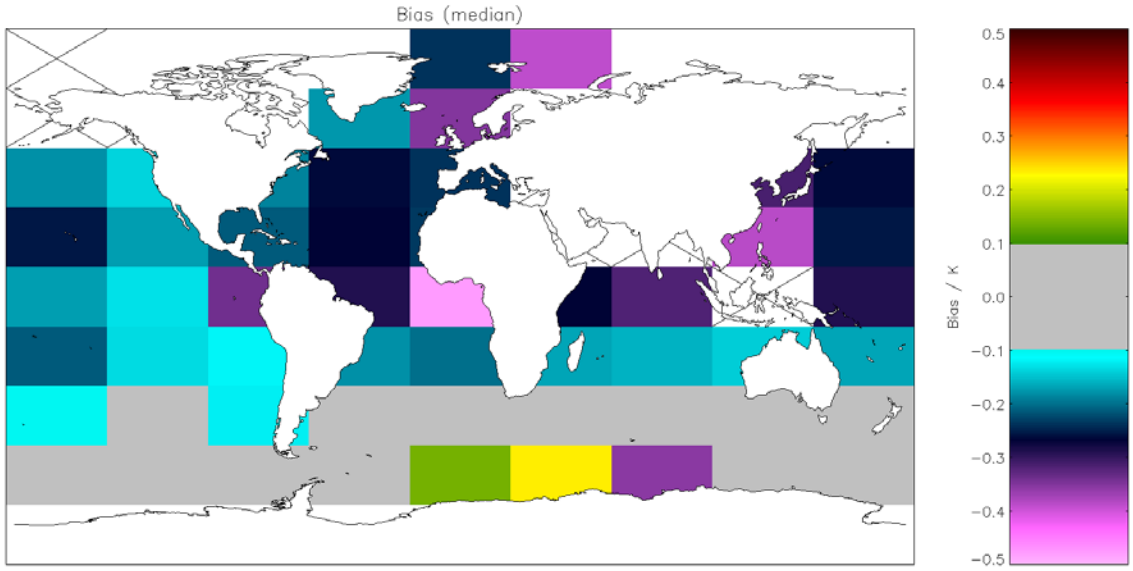


Figure A3.3.2-14: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-19 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.201 K (**Table A3.2-4**).

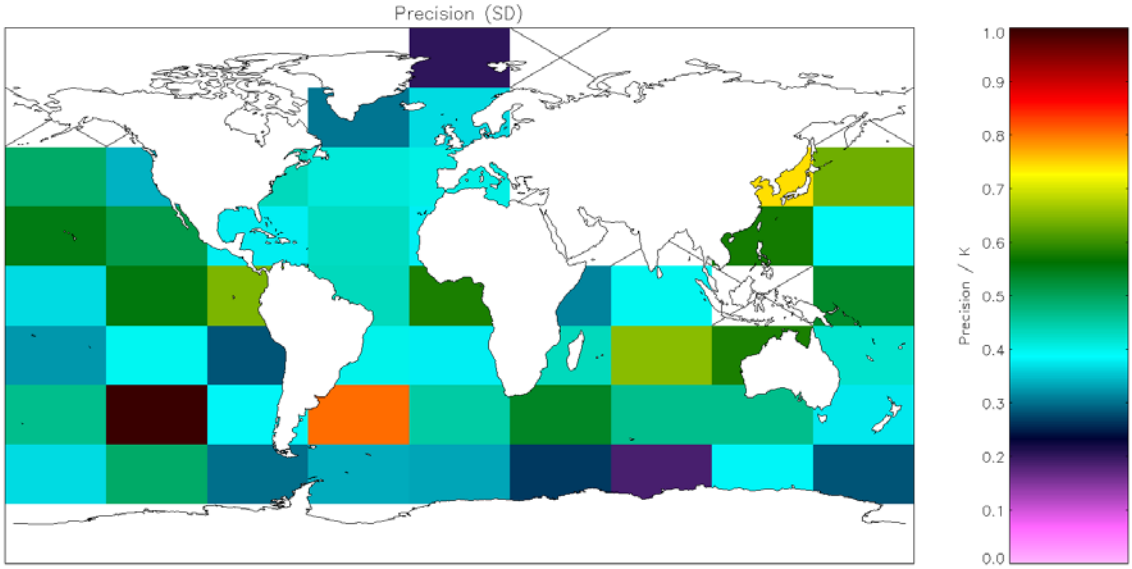


Figure A3.3.2-15: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-19 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.520 K (**Table A3.2-3**).

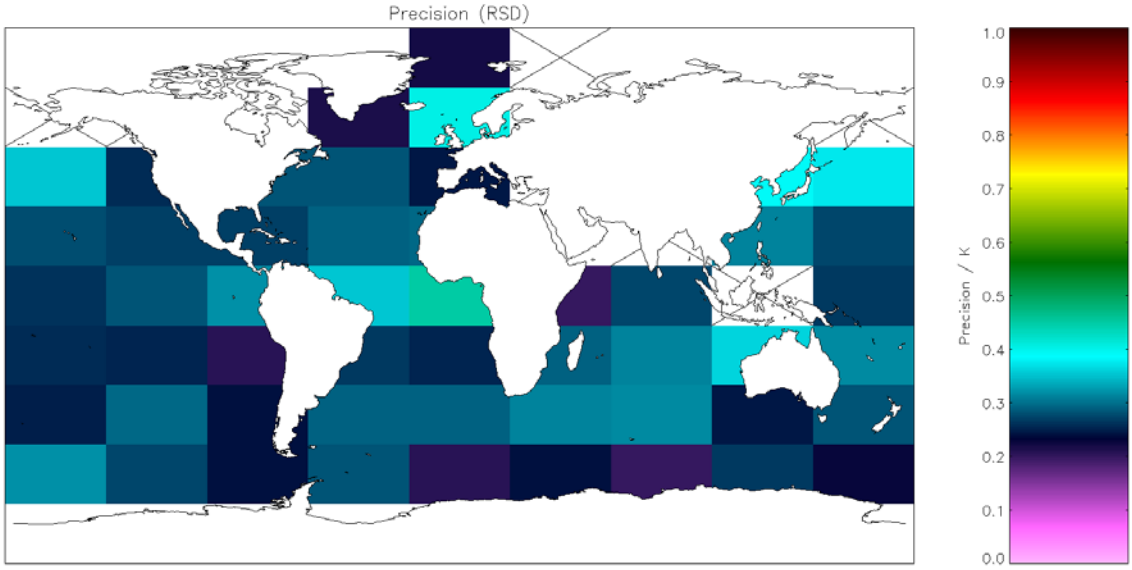


Figure A3.3.2-16: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-19 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.295 K (**Table A3.2-3**).

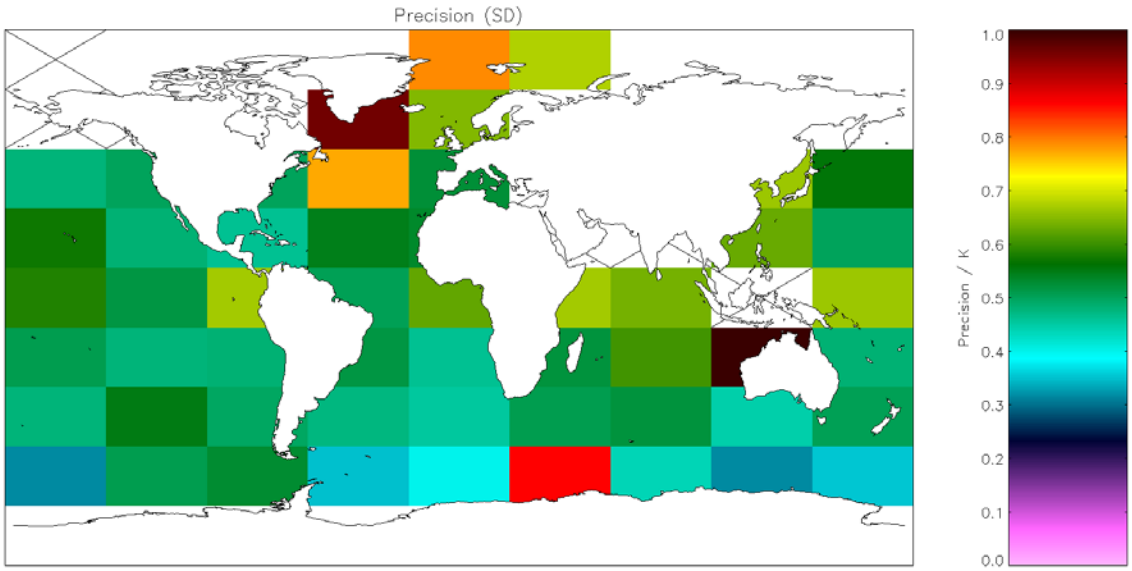


Figure A3.3.2-17: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-19 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.631 K (**Table A3.2-4**).

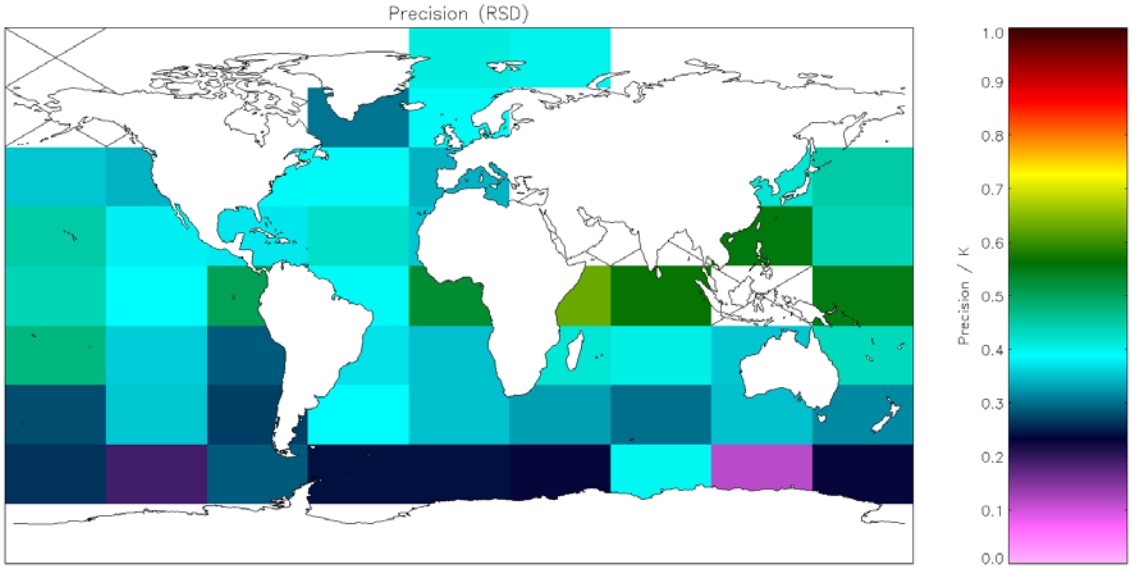


Figure A3.3.2-18: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-19 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.403 K (**Table A3.2-4**).

A3.3.3 NOAA-18

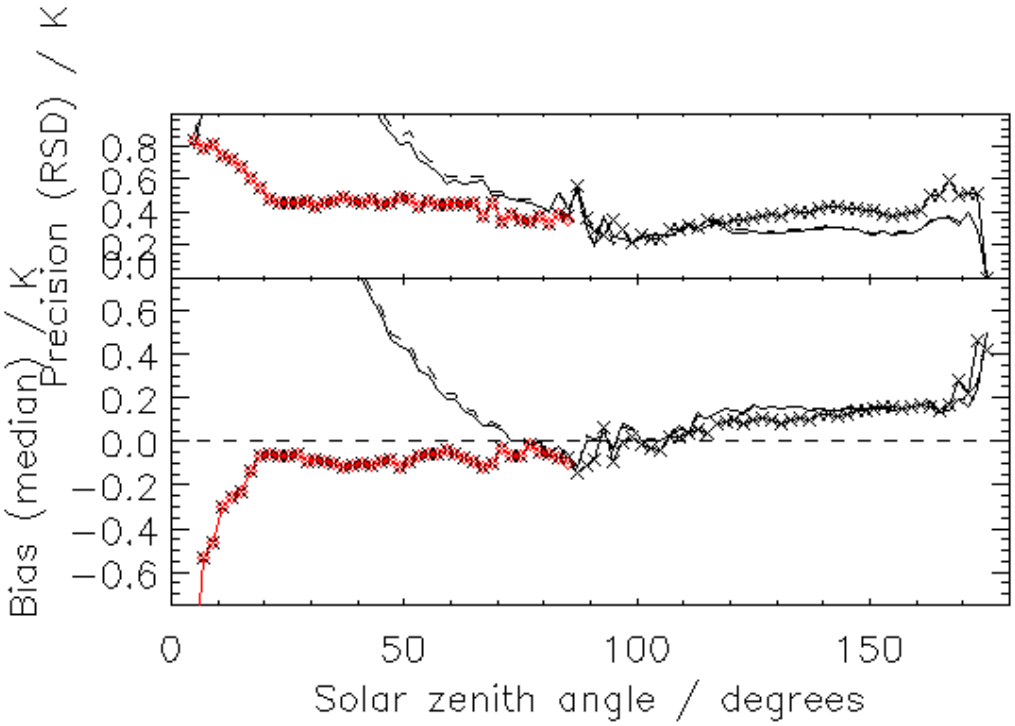


Figure A3.3.3-1: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for OE v2 retrievals for AVHRR onboard NOAA-18. The meaning of different lines/symbols is as follows: the line with symbols represents N2 retrievals (with red used to highlight daytime retrievals), the solid line (with no symbols) represents N3 retrievals, and the dashed line represents N2* (3.7 µm and 11 µm channel) retrievals.

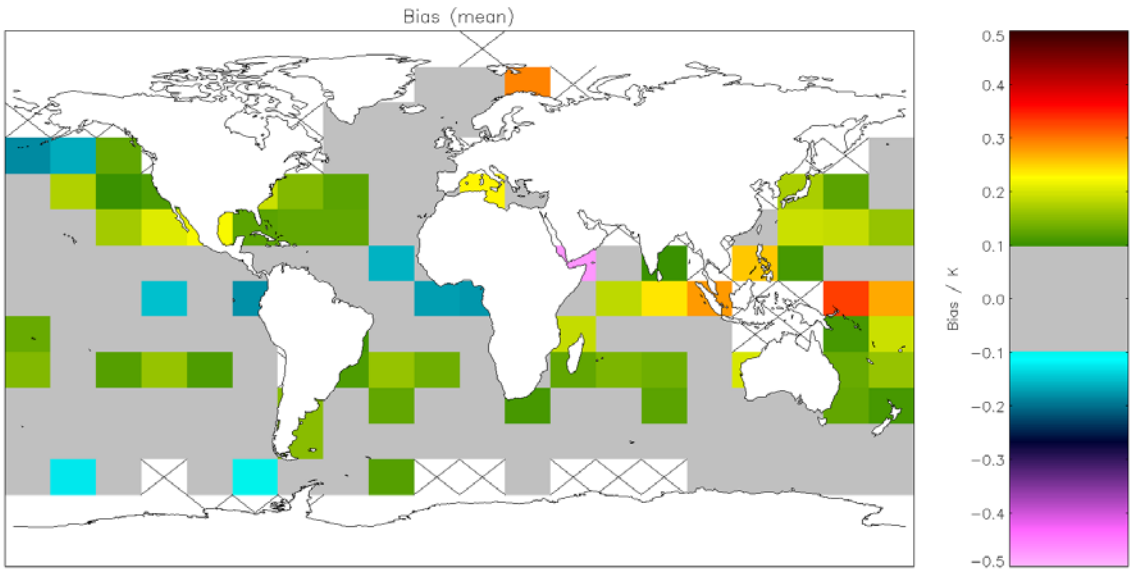


Figure A3.3.3-2: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-18 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.074 K (**Table A3.2-5**).

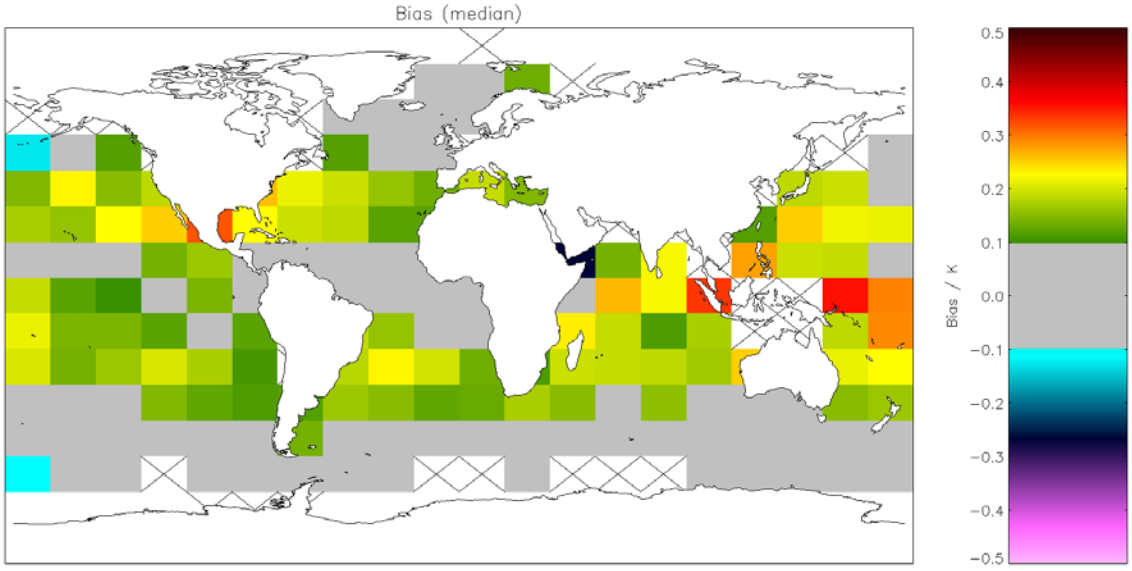


Figure A3.3.3-3: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-18 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.145 K (**Table A3.2-5**).

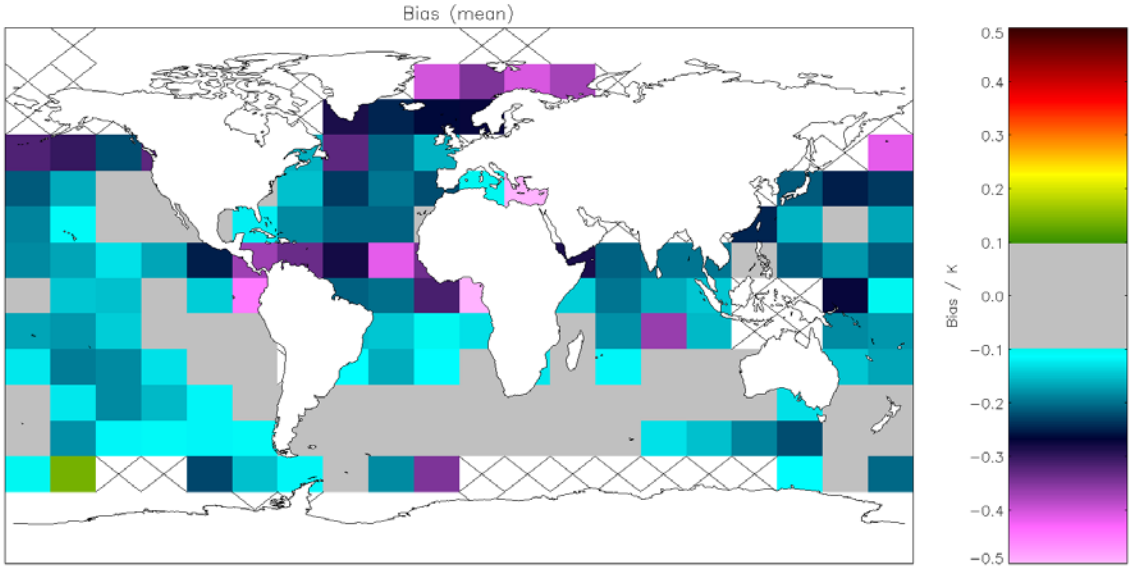


Figure A3.3.3-4: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-18 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.152 K (**Table A3.2-6**).

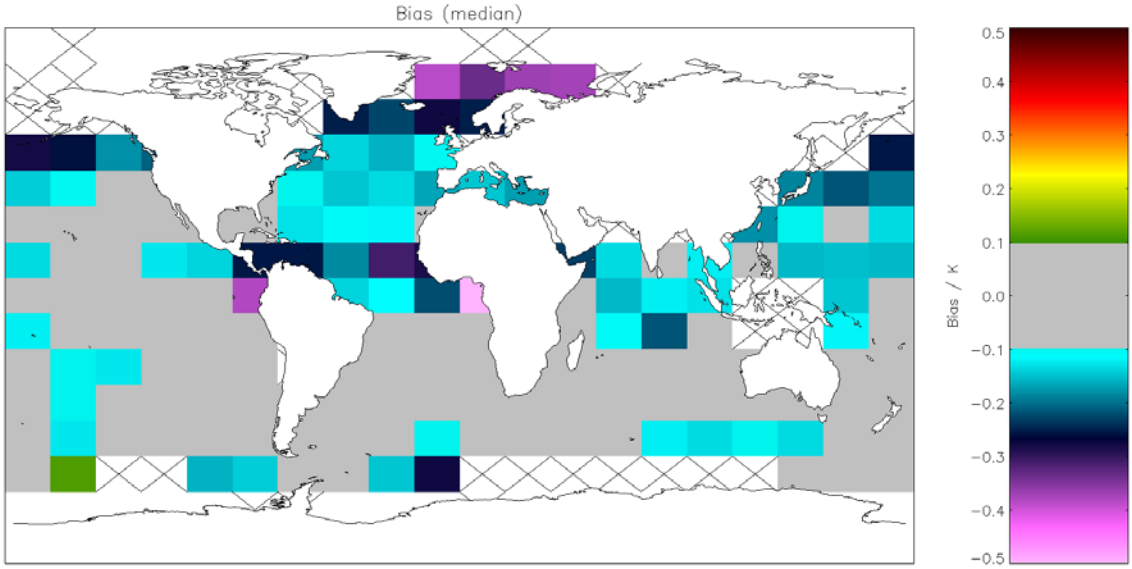


Figure A3.3.3-5: Global map of median bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-18 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.086 K (**Table A3.2-6**).

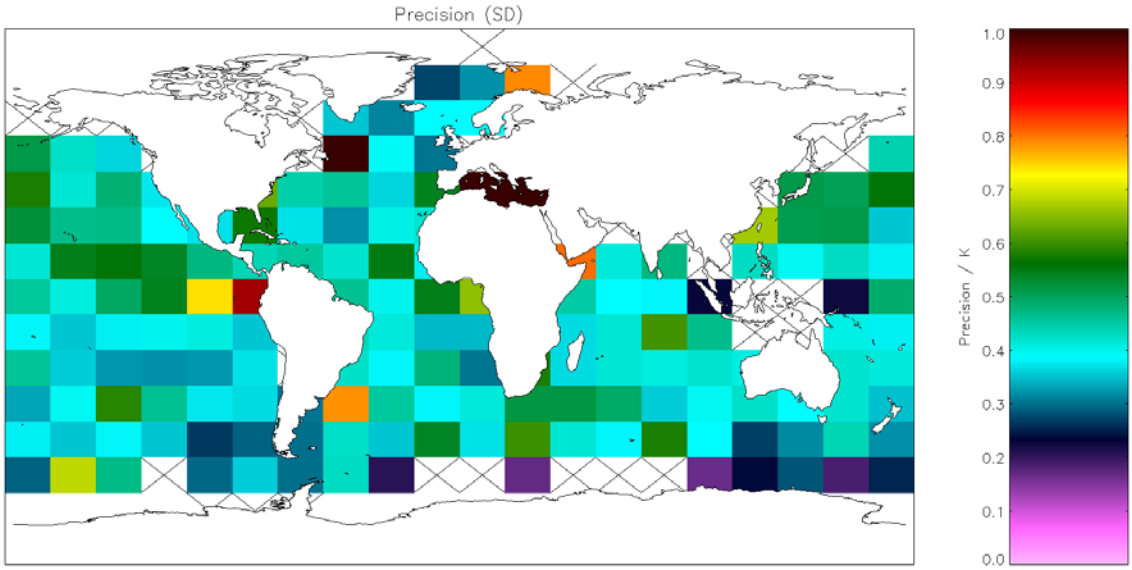


Figure A3.3.3-6: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-18 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.474 K (**Table A3.2-5**).

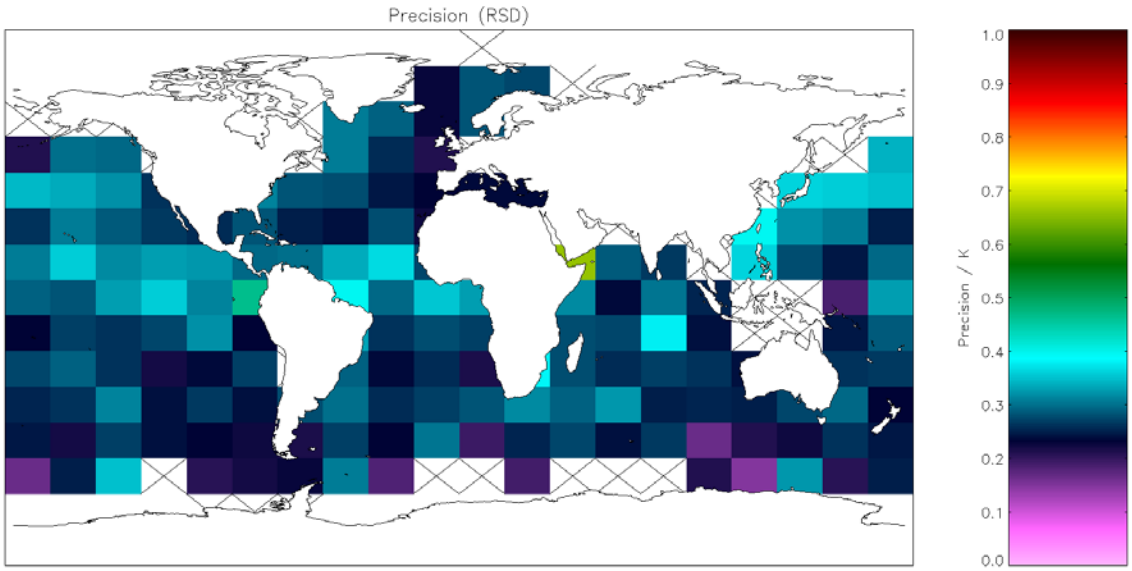


Figure A3.3.3-7: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AVHRR-18 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.145 K (**Table A3.2-5**).

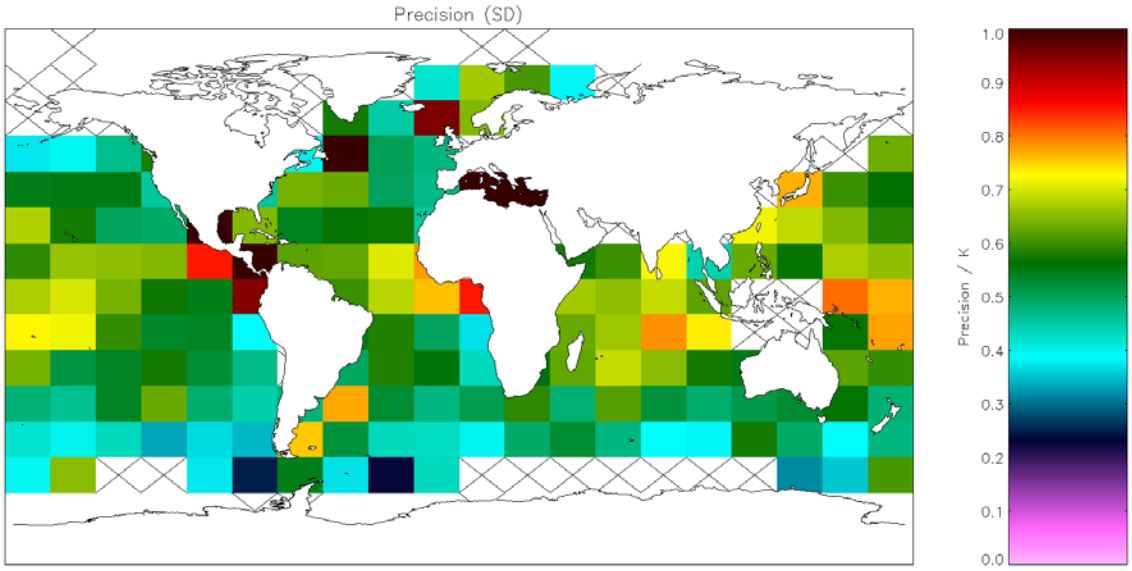


Figure A3.3.3-8: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time AVHRR-18 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.667 K (**Table A3.2-6**).

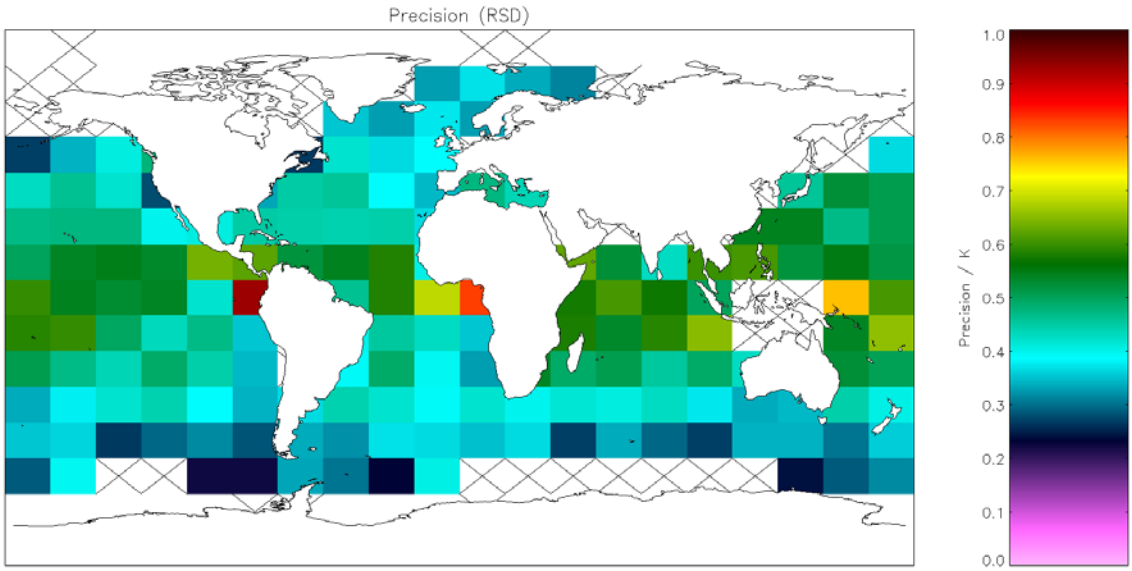


Figure A3.3.3-9: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time AVHRR-18 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.453 K (**Table A3.2-6**).

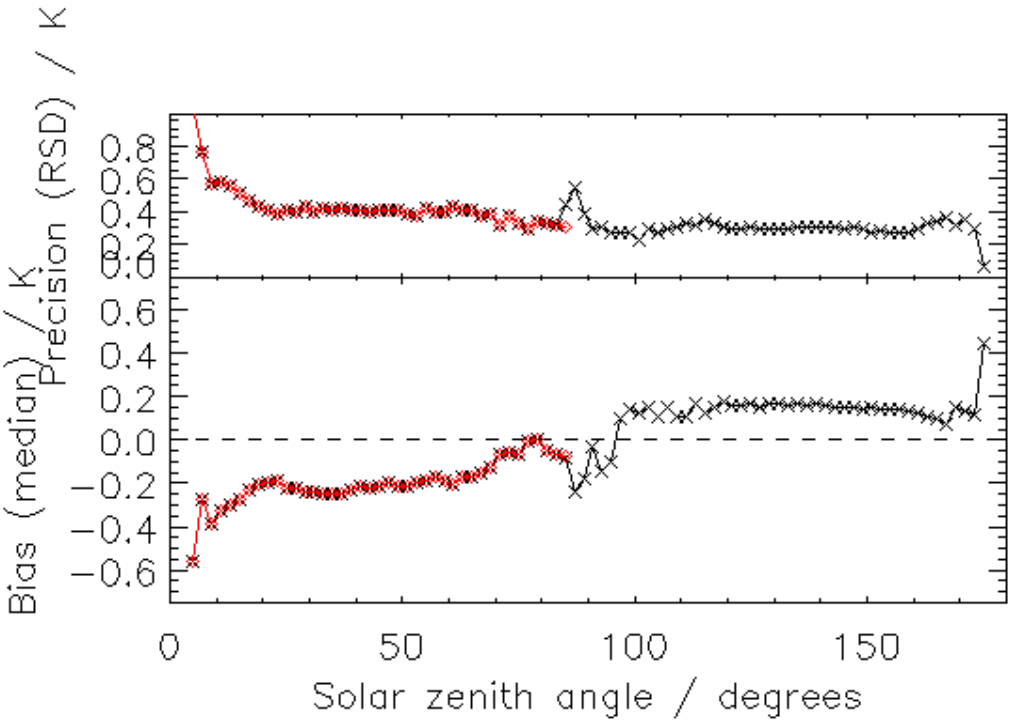


Figure A3.3.3-10: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for incremental regression retrievals for AVHRR onboard NOAA-18. The red and black lines with symbols represent retrievals using the daytime and night-time formulations of the incremental regression retrieval respectively.

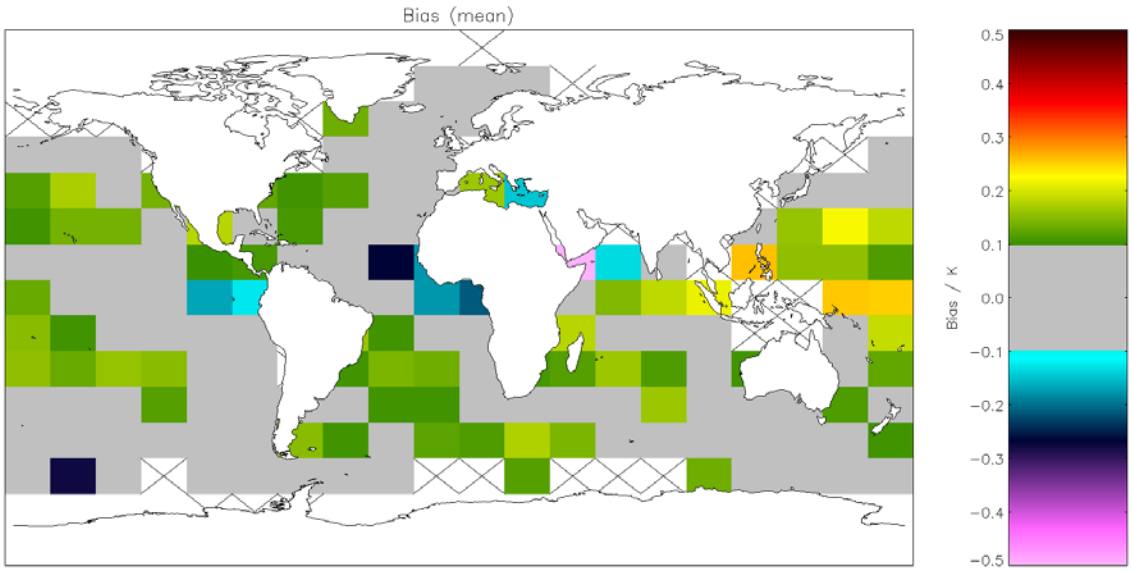


Figure A3.3.3-11: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-18 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.076 K (**Table A3.2-5**).

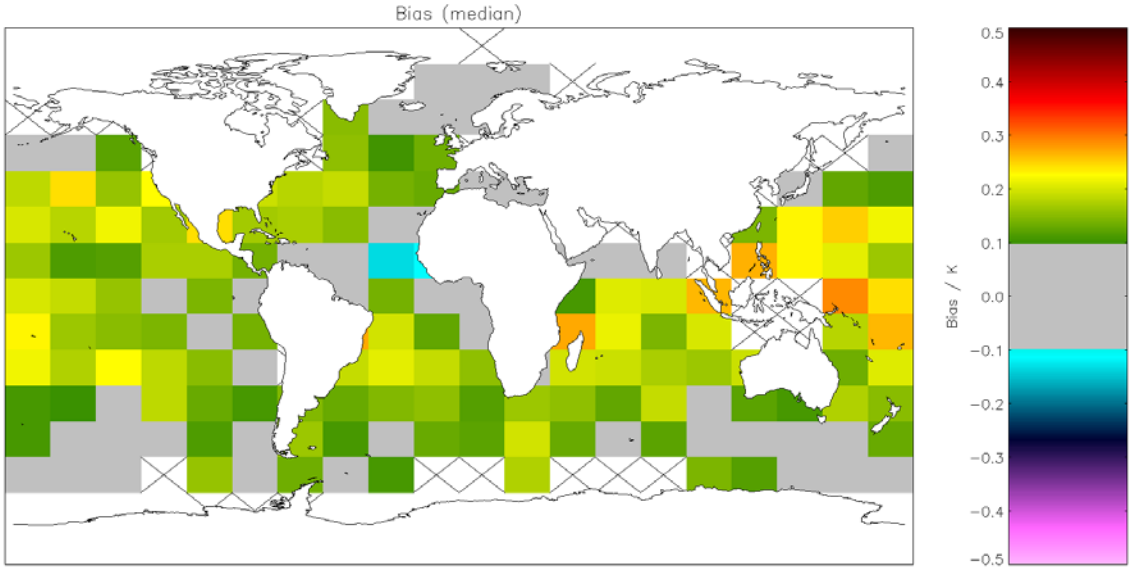


Figure A3.3.3-12: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-18 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.152 K (**Table A3.2-5**).

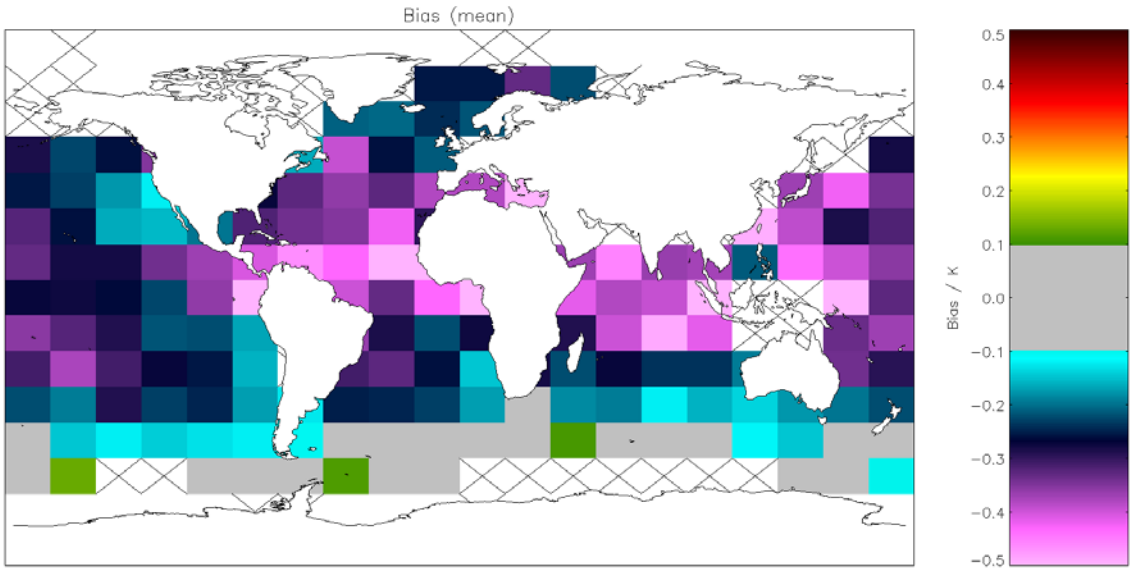


Figure A3.3.3-13: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-18 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.228 K (**Table A3.2-6**).

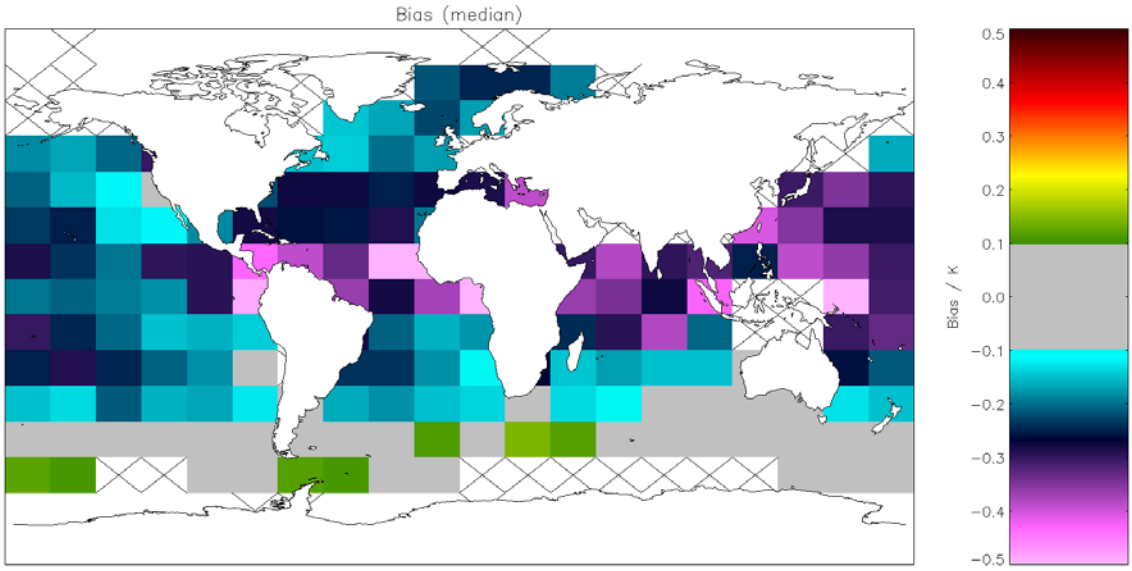


Figure A3.3.3-14: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-18 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.211 K (**Table A3.2-6**).

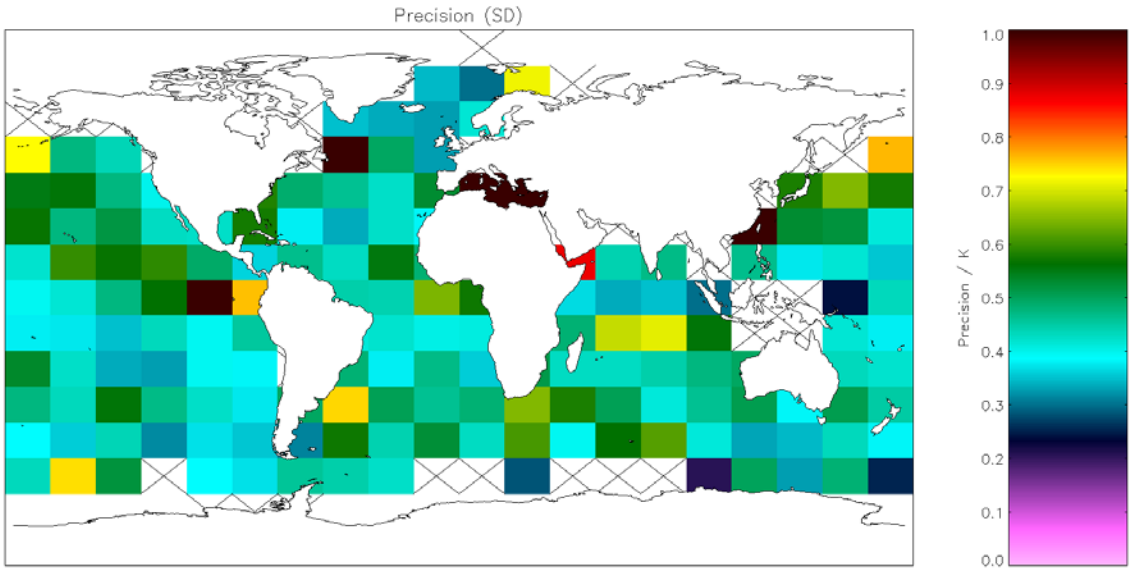


Figure A3.3.3-15: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-18 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.520 K (**Table A3.2-5**).

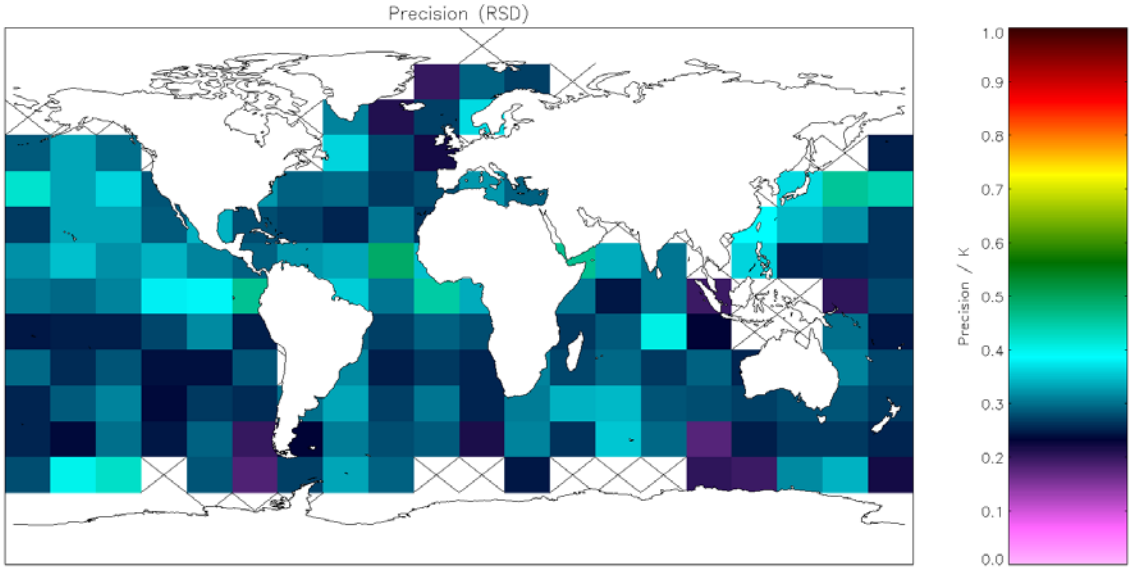


Figure A3.3.3-16: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-18 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.295 K (**Table A3.2-5**).

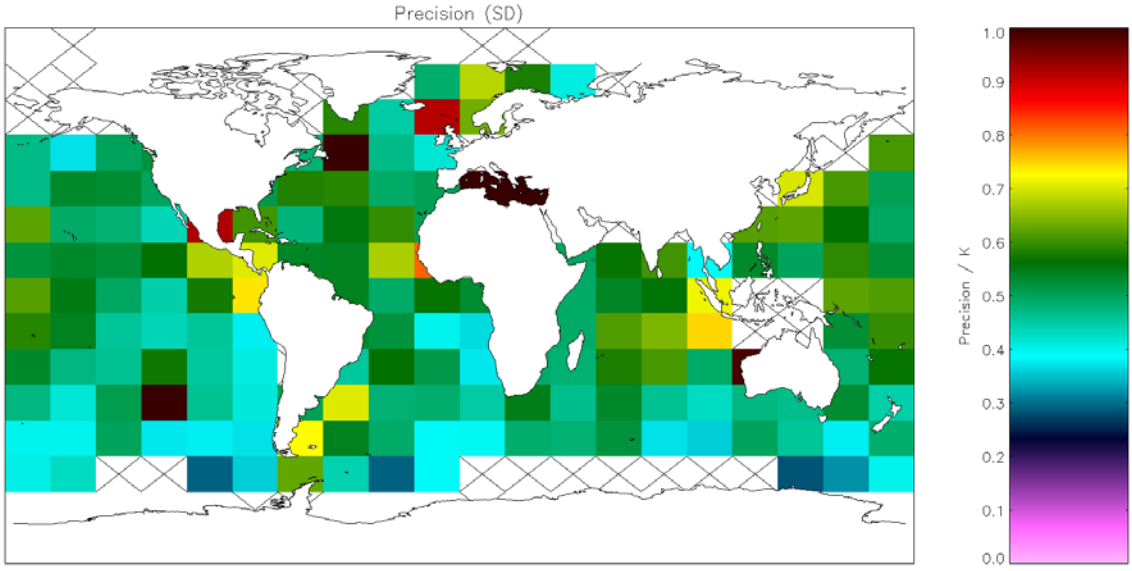


Figure A3.3.3-17: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-18 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.631 K (**Table A3.2-6**).

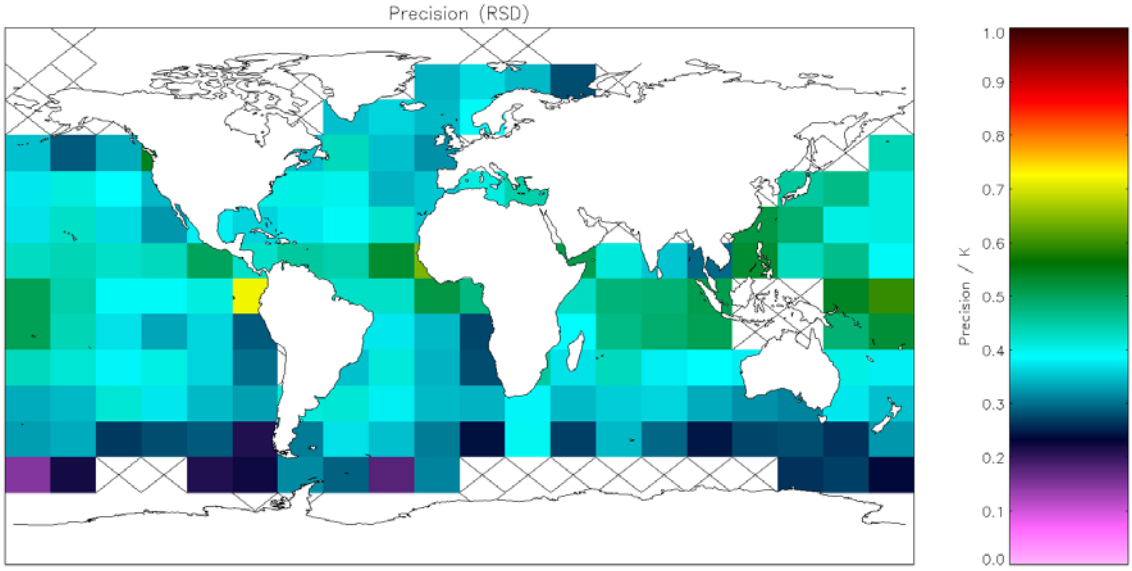


Figure A3.3.3-18: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-18 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.403 K (**Table A3.2-6**).

A3.3.4 NOAA17

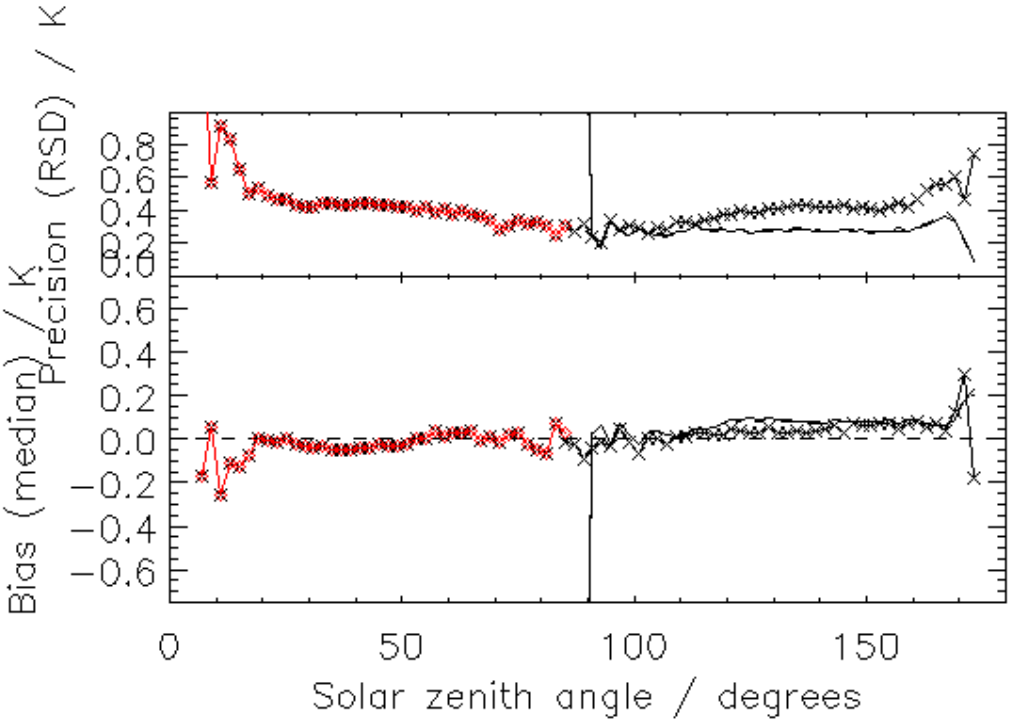


Figure A3.3.4-1: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for OE v2 retrievals for AVHRR onboard NOAA-17. The meaning of different lines/symbols is as follows: the line with symbols represents N2 retrievals (with red used to highlight daytime retrievals), the solid line (with no symbols) represents N3 retrievals, and the dashed line represents N2* (3.7 µm and 11 µm channel) retrievals.

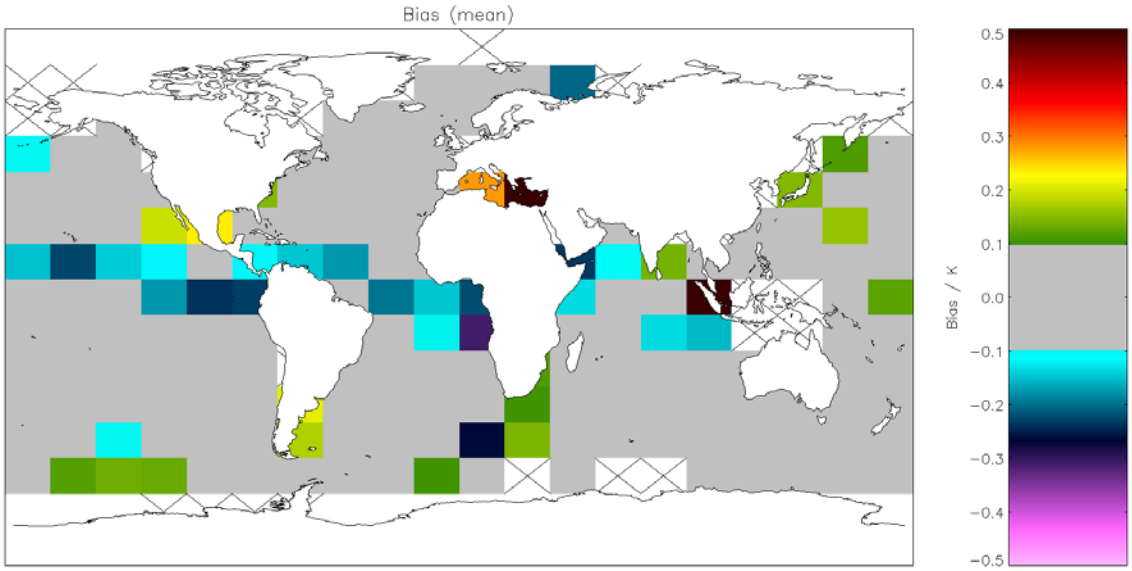


Figure A3.3.4-2: Global map of mean bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-17 observations with solar zenith angles > 160° excluded. Crosses indicate

grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.006 K (**Table A3.2-7**).

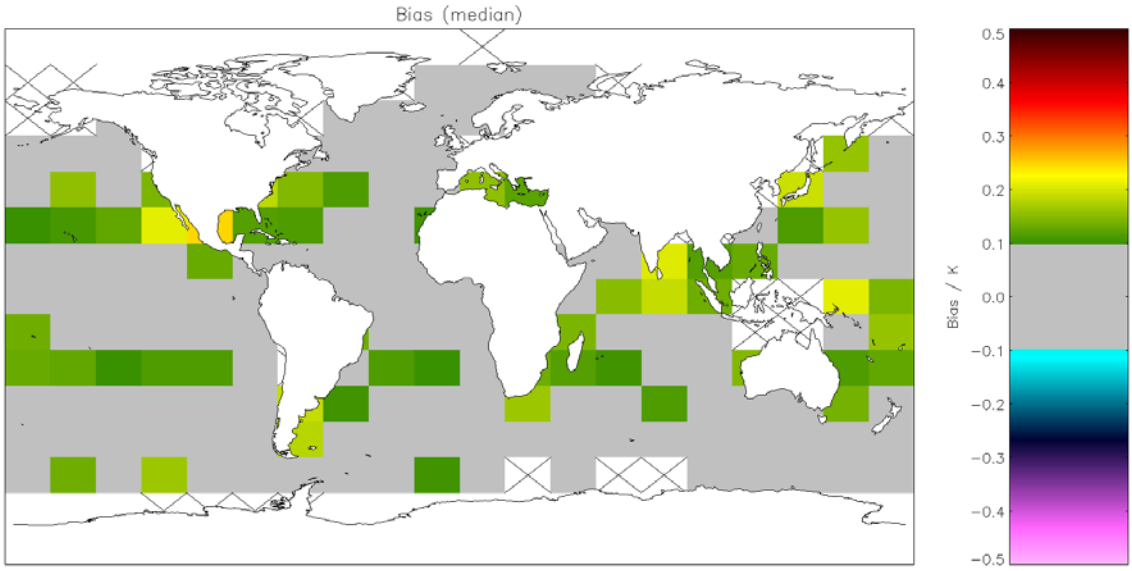


Figure A3.3.4-3: Global map of median bias (satellite-buoy) using OE v2 retrieval for night-time AVHRR-17 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.077 K (**Table A3.2-7**).

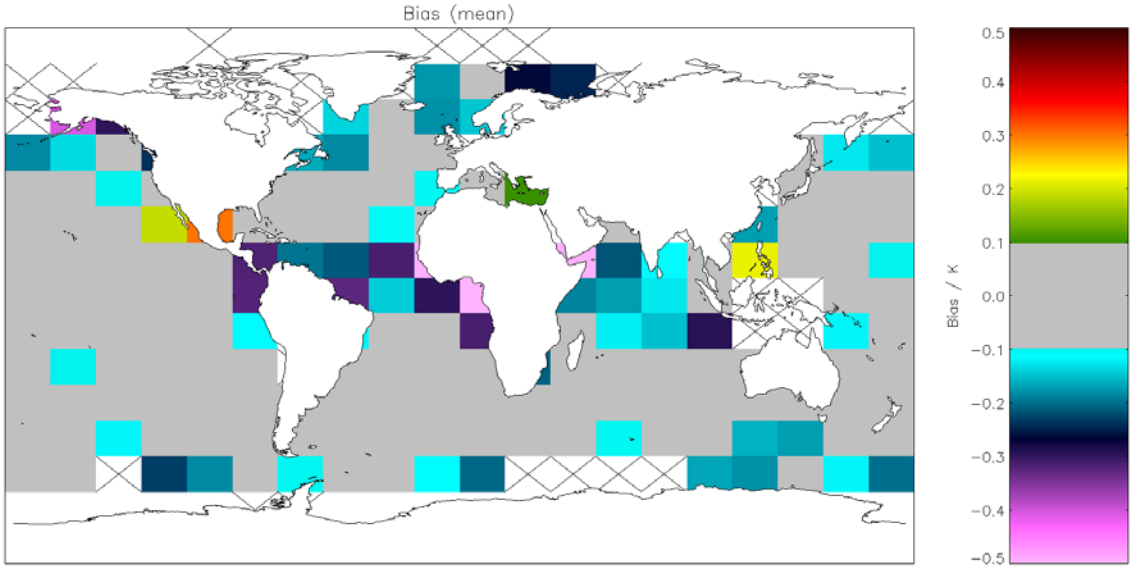


Figure A3.3.4-4: Global map of mean bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-17 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.048 K (**Table A3.2-8**).

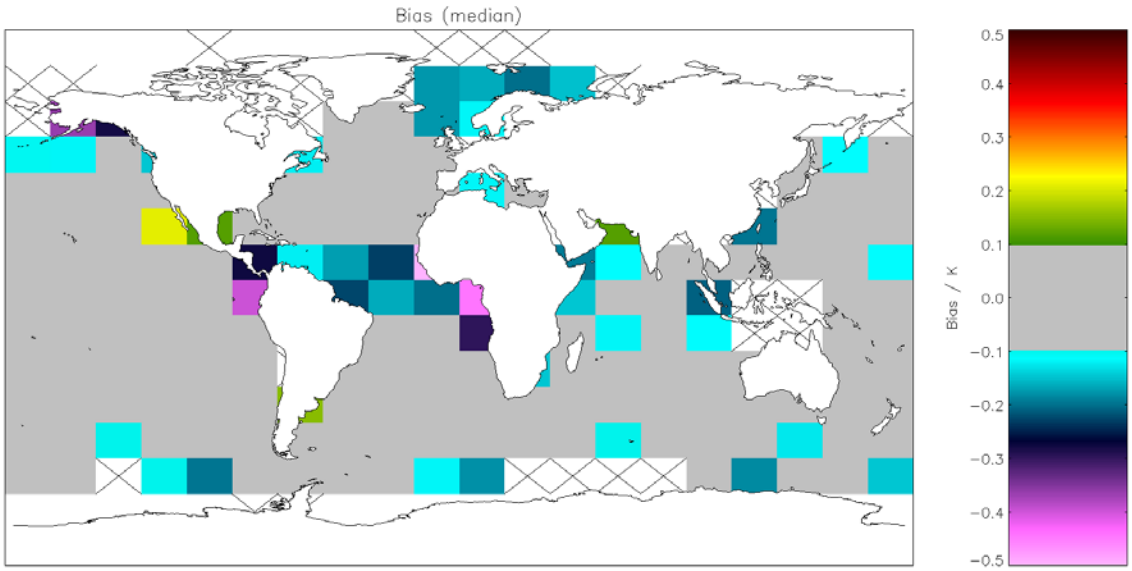


Figure A3.3.4-5: Global map of median bias (satellite-buoy) using OE v2 retrieval for day-time AVHRR-17 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.020 K (**Table A3.2-8**).

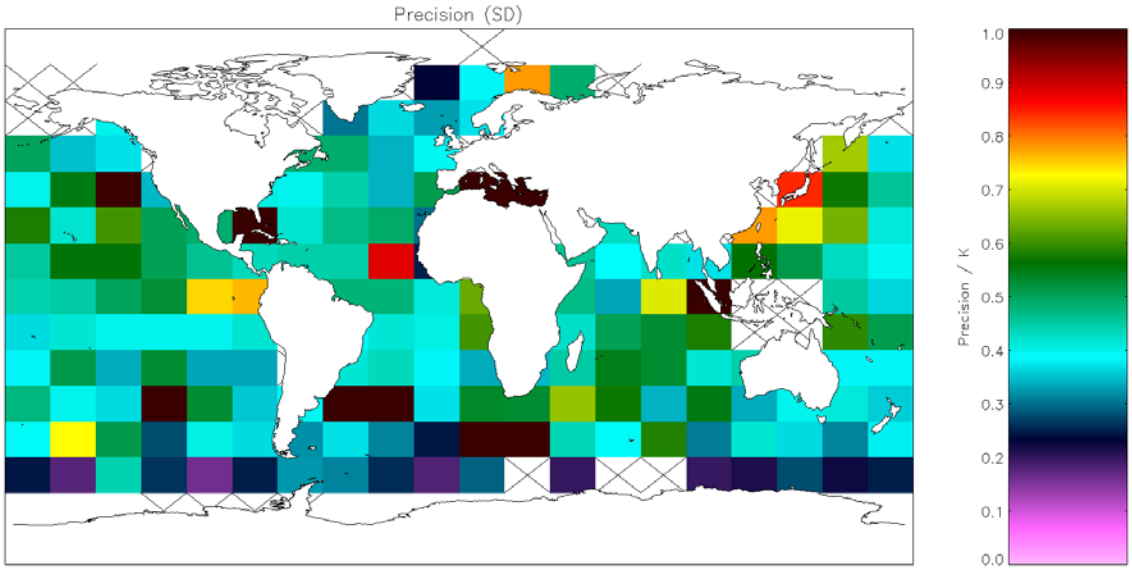


Figure A3.3.4-6: Global map of precision (mean of cell SDs) using OE v2 retrieval for night-time AVHRR-17 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.603 K (**Table A3.2-7**).

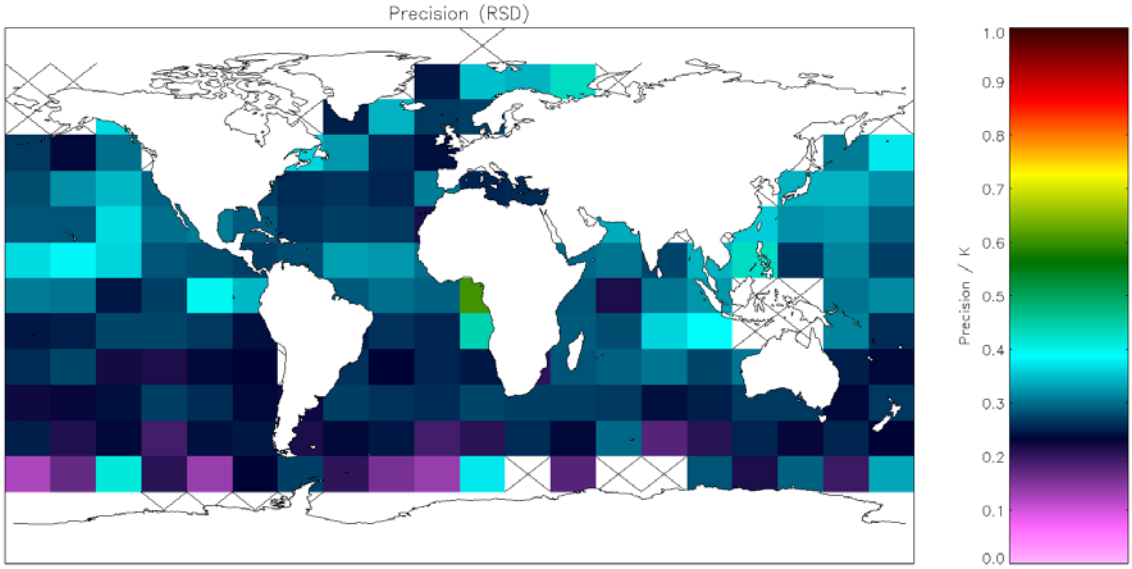


Figure A3.3.4-7: Global map of precision (median of cell RSDs) using OE v2 retrieval for night-time AVHRR-17 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.275 K (**Table A3.2-7**).

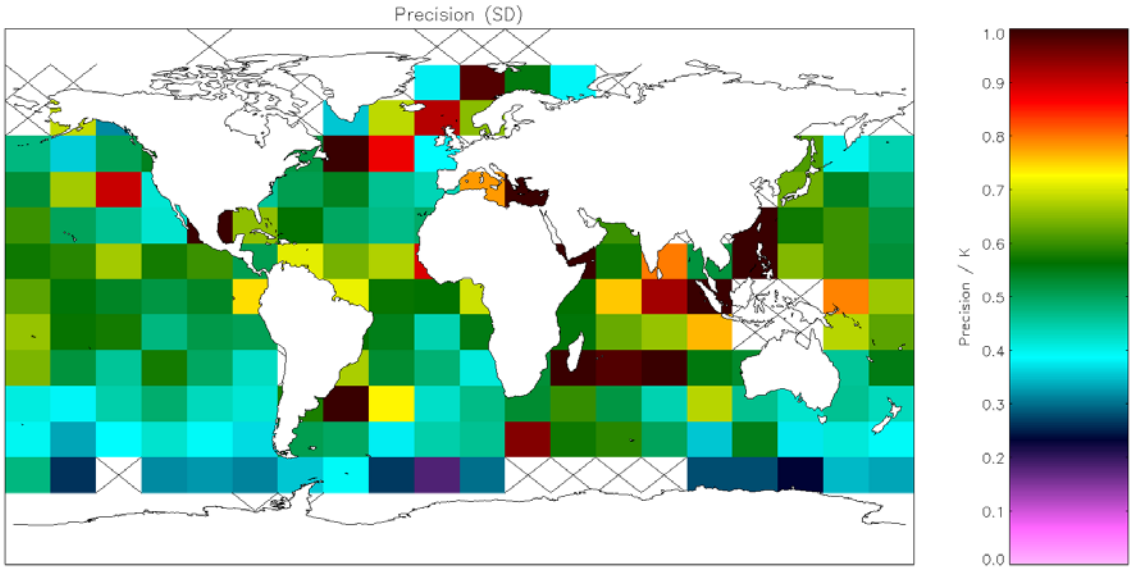


Figure A3.3.4-8: Global map of precision (mean of cell SDs) using OE v2 retrieval for day-time AVHRR-17 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.679 K (**Table A3.2-8**).

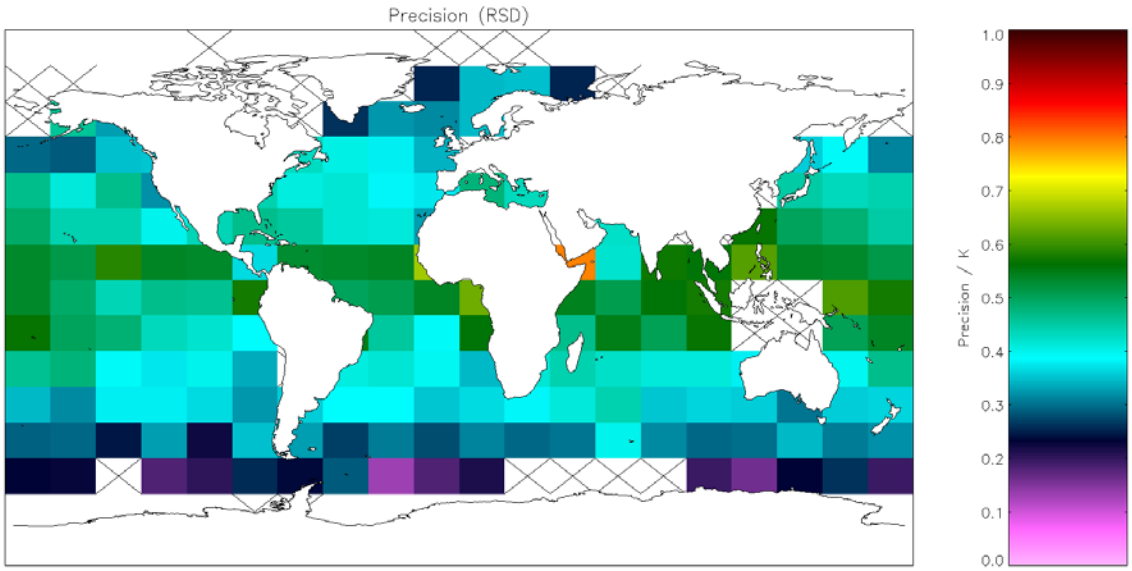


Figure A3.3.4-9: Global map of precision (median of cell RSDs) using OE v2 retrieval for day-time AVHRR-17 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.417 K (**Table A3.2-8**).

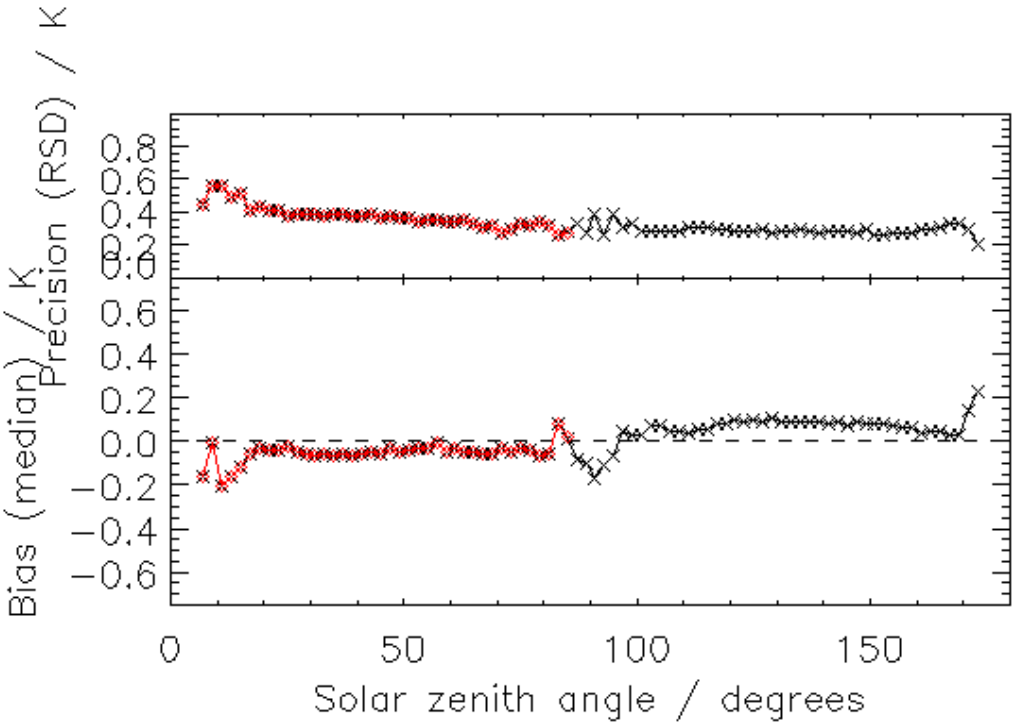


Figure A3.3.4-10: Bias (median discrepancy) and precision (RSD of discrepancies) as a function of solar zenith angle for incremental regression retrievals for AVHRR onboard NOAA-17. The red and black lines with symbols represent retrievals using the daytime and night-time formulations of the incremental regression retrieval respectively.

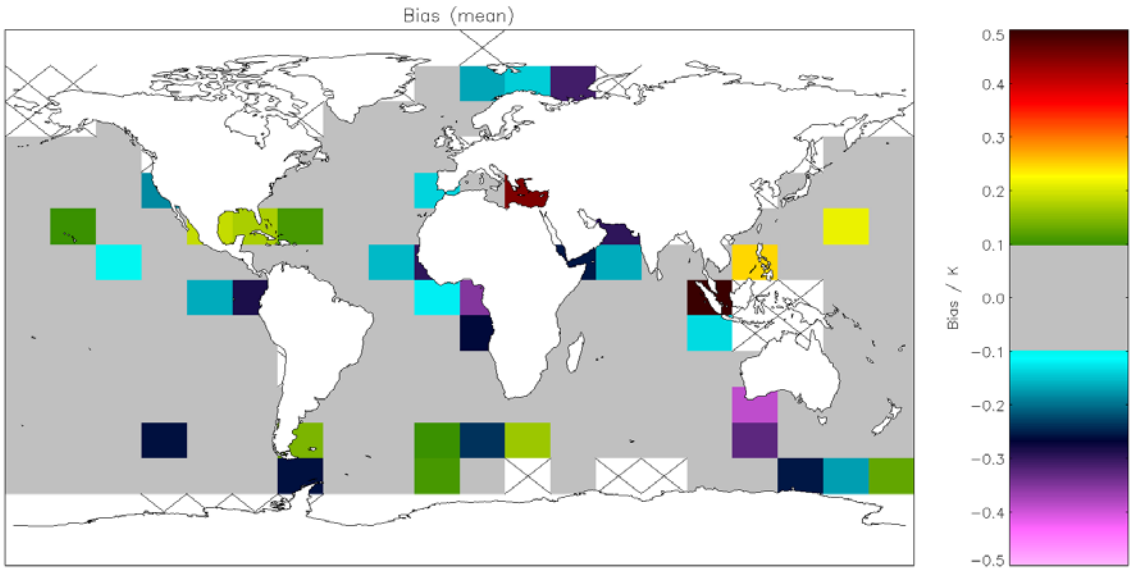


Figure A3.3.4-11: Global map of mean bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-17 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = 0.009 K (**Table A3.2-7**).

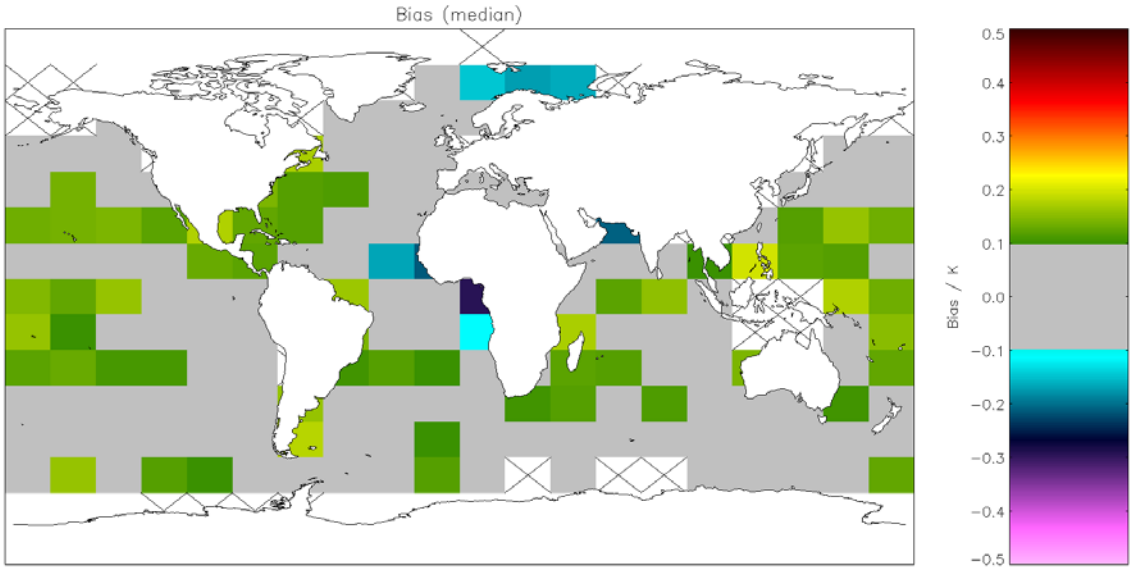


Figure A3.3.4-12: Global map of median bias (satellite-buoy) using incremental regression retrieval for night-time AVHRR-17 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = 0.082 K (**Table A3.2-7**).

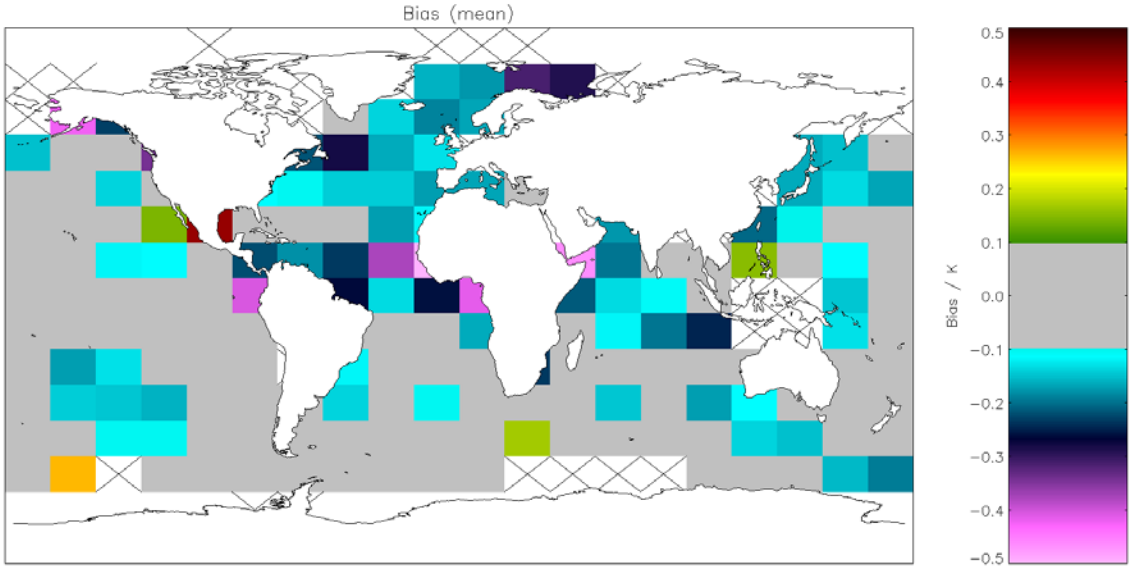


Figure A3.3.4-13: Global map of mean bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-17 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global mean bias = -0.089 K (**Table A3.2-8**).

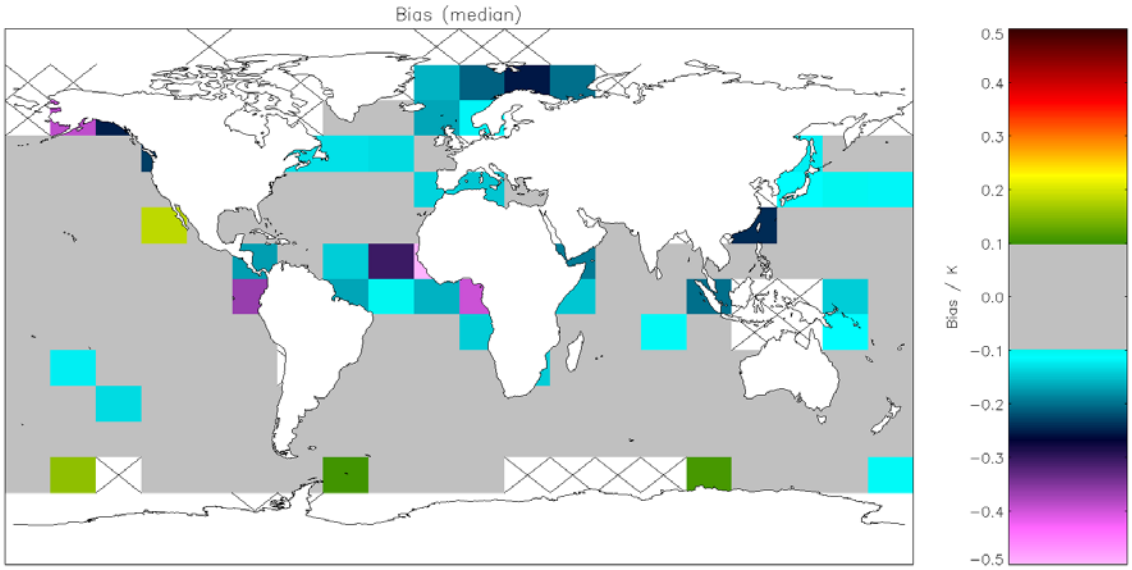


Figure A3.3.4-14: Global map of median bias (satellite-buoy) using incremental regression retrieval for day-time AVHRR-17 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global median bias = -0.047 K (**Table A3.2-8**).

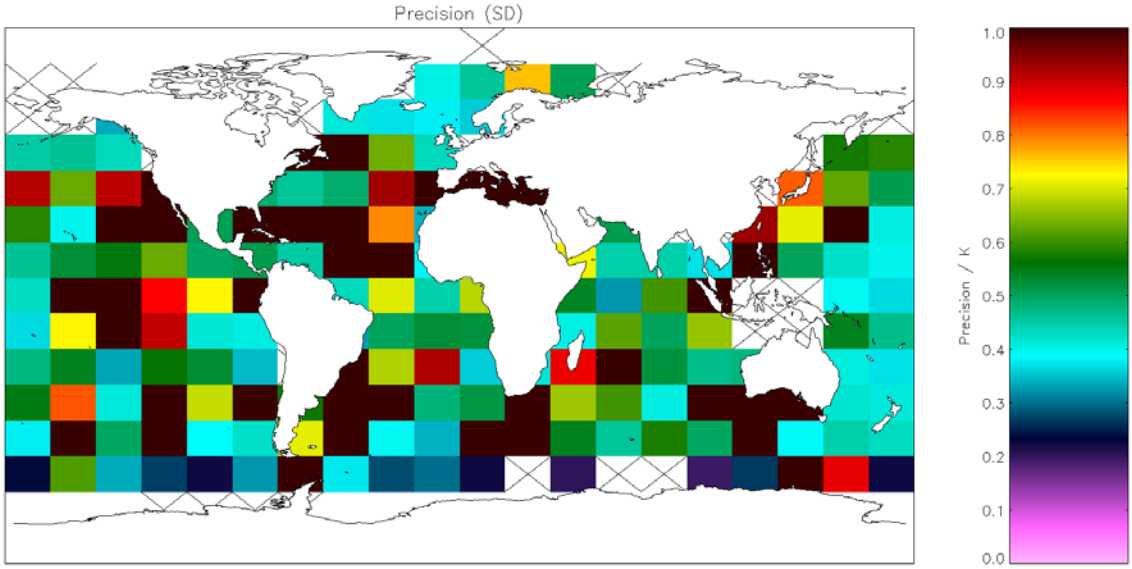


Figure A3.3.4-15: Global map of precision (mean of cell SDs) using incremental regression retrieval for night-time AVHRR-17 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.923 K (**Table A3.2-7**).

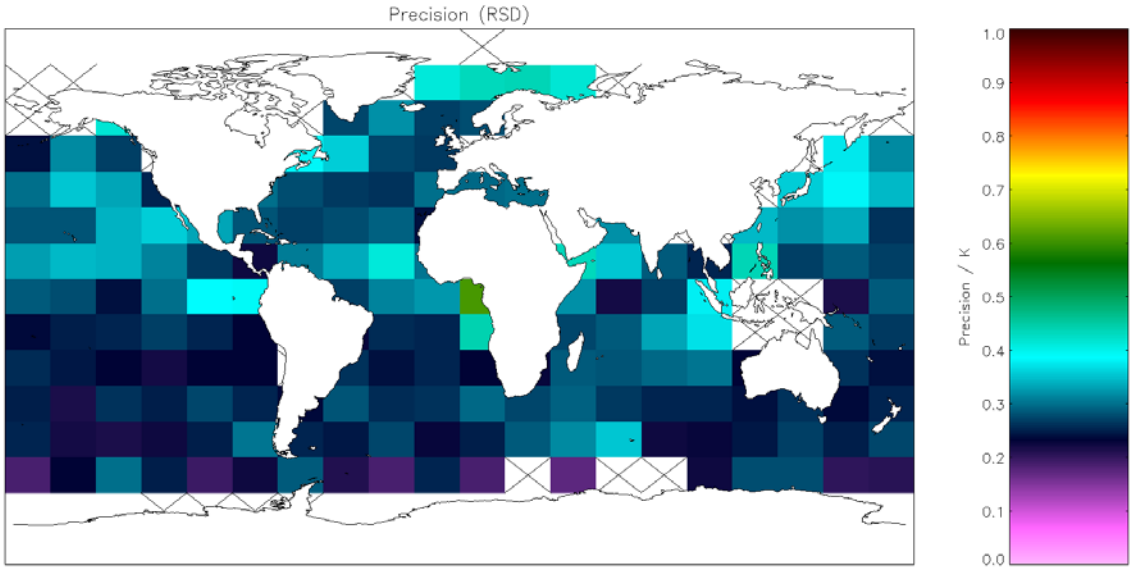


Figure A3.3.4-16: Global map of precision (median of cell RSDs) using incremental regression retrieval for night-time AVHRR-17 observations with solar zenith angles > 160° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.282 K (**Table A3.2-7**).

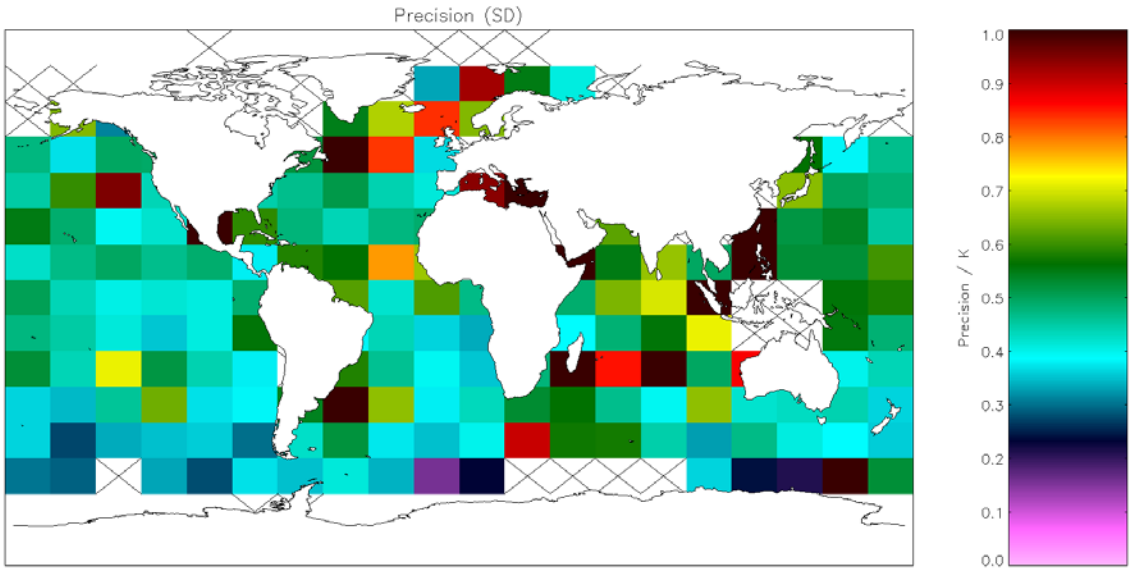


Figure A3.3.4-17: Global map of precision (mean of cell SDs) using incremental regression retrieval for day-time AVHRR-17 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global SD = 0.638 K (**Table A3.2-8**).

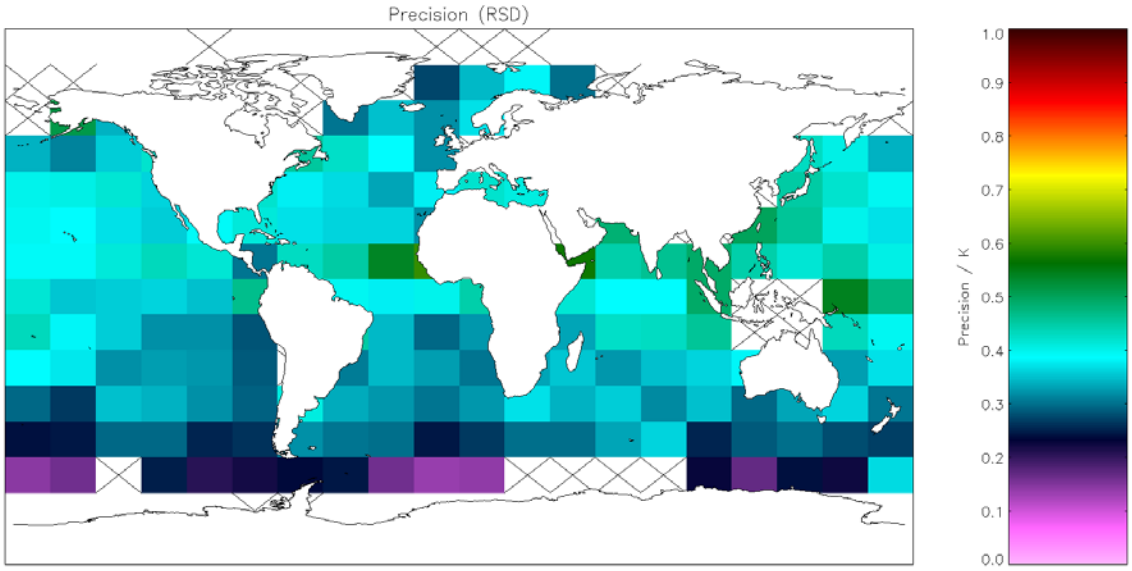


Figure A3.3.4-18: Global map of precision (median of cell RSDs) using incremental regression retrieval for day-time AVHRR-17 observations with solar zenith angles < 20° excluded. Crosses indicate grid cells excluded due to discrepancy of 0.1 K being insignificant for the cell. Global RSD = 0.363 K (**Table A3.2-8**).