Funded by the EU Copernicus programme

Scientific service framework for Copernicus sea and sea-ice surface temperature product improvement and CAL/VAL tool development and evolution

Activity 1: Improvement to Sentinel-3 SLSTR sea and sea-ice Surface Temperature product quality

[D2-A1_1] Product Development Plan for the Sea and Land Surface Temperature Radiometer: Sea Surface Temperature

Issue. 1 – Rev. 1

153 rue du Lac – 31670 Labège – France
Tél. : +33 (0)562 88 11 11 – Fax : +33 (0)562 88 11 12 – E-mail : noveltis@noveltis.fr

Document status

Scientific service framework for Copernicus sea and sea-ice surface temperature product improvement and CAL/VAL tool development and evolution Activity 1: Improvement to Sentinel-3 SLSTR sea and sea-ice Surface Temperature product quality [D2-A1_1] Product Development Plan			
Issue	**Revision**	**Date**	**Reason for the revision**
1	0	28/10/21	Initial version
1	1	15/11/21	Updates according to EUMETSAT's feedback

Modification status				
Issue	Rev	Status *	Modified pages	Reason for the modification
1	1	M	§A2	Adding precision on the rating of "I-stripe Reflectance Products for use with IR" task

* *I = Inserted* *D = Deleted* *M = Modified*

Acknowledgements

This service is funded by the European Union's Copernicus programme.

Acronyms

AATSR	Advanced along track scanning radiometer
ADF	Auxiliary data file
ARC	AATSR Reprocessing for Climate
ASCII	American Standard Code for Information Interchange
ASTER	Advanced Spaceborne Thermal Emission and Reflection
ATSR	Along track scanning radiometer
AVHRR	Advanced very high resolution radiometer
BAOE	Bias-aware Optimal Estimation
BCDS	Bayesian cloud detection study
BRDF	Bi-directional Reflective Distribution Function
BT	Brightness temperature
CAL	Calibration
CAMS	Copernicus atmospheric monitoring service
CCI	Climate change initiative
netCDF	Network common data format
CGT	Coefficient Generation Tool
DINEOF	Dynamically Interpolating Empirical Orthogonal Functions
DV	Diurnal Variability
DW	Diurnal Warming
ERA	ECMWF re-analysis
ESL	Expert support laboratory
EUSTACE	European Surface Temperature for All Corners of Earth
FOV	Field of view
GEO	Geostationary Earth orbit
GHRSST	Group for High Resolution Sea Surface Temperature
GOES	Geostationary Operational Environmental Satellite
HRSST	High resolution SST
IR	InfraRed
IST	ice surface temperature
LSD	Local Standard Deviation
LUT	Look-up table
MDB	Matchup database

MET-TX	Meteorological fields
Metop SG	Metop Second Generation
MF	Meteo France
MM	Multisensor match
MMD	Multisensor matchup database
MODIS	Moderate resolution Imaging Spectrometer
NRT	Near real time
NTC	Non-time critical
NWP	Numerical weather prediction
OE	Optimal Estimation
OPAC	Optical Properties of Aerosols and Clouds
OSI-SAF	Ocean and sea ice satellite application facility
OSTIA	Operational Sea Surface Temperature and Ice Analysis (
PDF	Probability density function
PDP	Product Development Plan
PM	Person Months
POES	Polar orbiting environmental satellite
PP	Prototype Processor
QL	Quality Level
RT	Radiative transfer
RTM	Radiative transfer model
RTTOV	Name of a radiative transfer model
SD	Standard deviation
SEVIRI	Spinning enhanced visible and infrared radiometer
SLSTR	Sea and Land Surface Temperature Radiometer
SNAP	Sentinel applications
SRF	Spectral response function
SSES	Single sensor error statistics
SST	Sea Surface Temperature
ST	Skin Temperature
SWIR	Short wave infrared
TCWV	Total column water vapour
TOA	Top of atmosphere
TRISHNA	Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment

VAL	Validation
VIS	Visible
WMO	World Meteorological Organisation

[RD1]	Carrea, L., Embury, O. and Merchant, C. J. (2015) Datasets related to in-land water for limnology and remote sensing applications: distance-to-land, distance-to-water, water-body identifier and lake-centre co-ordinates. Geoscience Data Journal, 2 (2). pp. 83-97. ISSN 2049-6060 doi: https://doi.org/10.1002/gdj3.32
[RD2]	Bulgin, C. E., Eastwood, S., Embury, O., Merchant, C. J., Donlon, C. (2015). Sea surface temperature climate change initiative: alternative image classification algorithms for sea-ice affected oceans. Remote Sensing of Environment, 162. pp 396-407.
[RD3]	Bulgin, C. E., Merchant, C. J., Ghent, D., Klueser, L., Popp, T., Poulsen, C. and Sogacheva, L. (2018) Quantifying uncertainty in satellite-retrieved land surface temperature from cloud detection errors. Remote Sensing, 10 (4). 616. ISSN 2072-4292 doi: https://doi.org/10.3390/rs10040616
[RD4]	Bulgin, C. E. and Merchant, C. J. (2018). Validation and investigations into Bayesian cloud detection performance, Preparation and operations of the mission performance centre (MPC) for the Copernicus Sentinel-3 mission. S3MPC.RAL.VR.033.
[RD5]	Merchant, C., Saux-Picart, S. and Waller, J. (2020) Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references. Remote Sensing of Environment, 237. 111590. ISSN 0034-4257 doi: https://doi.org/10.1016/j.rse.2019.111590
[RD6]	Merchant, C. J., Block, T., Corlett, G. K., Embury, O., Mittaz, J. P. D. and Mollard, J. D. P. (2020) Harmonization of space-borne infrared sensors measuring sea surface temperature. Remote Sensing, 12 (6). 1048. ISSN 2072-4292 doi: https://doi.org/10.3390/rs12061048
[RD7]	Merchant, C.J. and Harris, A.R. (1999) Toward the elimination of bias in satellite retrievals of skin sea surface temperature. 2: Comparison with in situ measurements. Journal of Geophysical Research, 104 (C10). ISSN 0148-0227 doi: https://doi.org/10.1029/1999JC900106
[RD8]	Bingkun Luo, Peter J. Minnett, Nicholas R. Nalli (2021) Infrared satellite-derived sea surface skin temperature sensitivity to aerosol vertical distribution–Field data analysis and model simulations, Remote Sensing of Environment, Volume 252, https://doi.org/10.1016/j.rse.2020.112151.
[RD9]	Merchant, C. J., Harris, A. R., Maturi, E., Embury, O., MacCallum, S. N., Mittaz, J. and Old, C. P. (2009) Sea surface temperature estimation from the Geostationary Operational Environmental Satellite 12 (GOES-12). Journal of Atmospheric and Oceanic Technology, 26 (3). pp. 570-581. ISSN 1520-0426 doi: https://doi.org/10.1175/2008JTECHO596.1
[RD10]	Merchant, C. J., Harris, A. R., Roquet, H. and Le Borgne, P. (2009) Retrieval characteristics of non-linear sea surface temperature from the Advanced Very High Resolution Radiometer. Geophysical Research Letters, 36 (17). L17604. ISSN 0094-8276 doi: https://doi.org/10.1029/2009GL039843
[RD11]	Merchant, C. J., Le Borgne, P., Marsouin, A. and Roquet, H. (2008) Optimal estimation of sea surface temperature from split-window observations. Remote Sensing of Environment, 112 (5). pp. 2469-2484. ISSN 0034-4257 doi: https://doi.org/10.1016/j.rse.2007.11.011
[RD12]	Merchant, C. J., Le Borgne, P., Roquet, H. and Legendre, G. (2013) Extended optimal estimation techniques for sea surface temperature from the Spinning Enhanced Visible and Infrared Imager (SEVIRI). Remote Sensing of Environment, 131. pp. 287-297. ISSN 0034-4257 doi: https://doi.org/10.1016/j.rse.2012.12.019
[RD13]	Rayner, N. A., Auchmann, R., Bessembinder, J., Brönnimann, S., Brugnara, Y., Capponi, F., Carrea, L., Dodd, E. M. A., Ghent, D., Good, E., Høyer, J. L., Kennedy, J. J., Kent, E. C., Killick, R. E., van der Linden, P., Lindgren, F., Madsen, K. S., Merchant, C. J., Mitchelson, J. R., Morice, C. P., Nielsen-Englyst, P., Ortiz, P. F., Remedios, J. J., van der Schrier, G., Squintu, A. A., Stephens, A., Thorne, P.

	W., Tonboe, R. T., Trent, T., Veal, K. L., Waterfall, A. M., Winfield, K., Winn, J. and Woolway, R. I. (2020) The EUSTACE project: delivering global, daily information on surface air temperature. Bulletin of the American Meteorological Society, 101 (11). E1924-E1947. ISSN 1520-0477 doi: https://doi.org/10.1175/BAMS-D-19-0095.1
[RD14]	Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofeiva, V. F. and Wagner, W. (2017) Uncertainty information in climate data records from Earth observation. Earth System Science Data, 9 (2). pp. 511-527. ISSN 1866-3516 doi: https://doi.org/10.5194/essd-9-511-2017
[RD15]	Fairall, C.W., Bradley, E.F., Godfrey, J.S., Wick, G.A., Edson, J.B., Young, G.S., 1996, Cool-skin and warm-layer effects on sea surface temperature, JGR (Oceans), 101, 1295-1308, doi: https://doi.org/10.1029/95JC03190
[RD16]	Kantha, L., Wittmann, P., Sclavo, M., Carniel, S., 2009, A preliminary estimate of the Stokes dissipation of wave energy in the global ocean, GRL (Oceans), 36, L02605, doi: https://doi.org/10.1029/2008GL036193
[RD17]	Kantha, L.H. & Clayson, C.A., 1994, An improved mixed layer model for geophysical applications, JGR (Oceans), 99, 25235-25266, doi: https://doi.org/10.1029/94JC02257

Table of contents

1. Introduction

This is a Product Development Plan (PDP) for ocean products from the Sea and Land Surface Temperature Radiometer (SLSTR), focused on sea surface temperature (SST), but bearing in mind future plans for marine ice surface temperature developments. It was prepared under the project "Science for Marine Surface Temperature" (Sci4MaST).

The primary purpose of this plan is to propose to EUMETSAT for consideration a route for the development of improved "Day 2" SST products for SLSTR, given the experience of SLSTR operations to date and drawing from the state-of-the-art of the wider SST scientific community. This primary purpose is principally fulfilled in sections 2 to 3 following. System implications of the planned developments are outlined in section 4. These sections are intentionally succinct.

The secondary purpose of this document is to capture the scientific considerations and investigations that lie behind the proposed route for development, thinking for which was consolidated within Sci4Mast. This secondary purpose is fulfilled by the appendices of the document, A1 to A8, which are intentionally extensive.

2. Overview of Priority Developments for SLSTR SST

2.1. "Day 2" Developments

Quality of cloud detection, uncertainty of SST, validity of SST uncertainty estimates and utility of the quality indicator are all important aspects contributing to SLSTR SST products. These areas are addressed in this PDP, as is fundamental work on relating skin and depth SSTs by modelling.

Twenty tasks, all being beneficial advances on the state of the art, have been identified by review and, among them, nine more tractable tasks are identified as recommended priorities achievable on a two-year research timescale (in combination).

The full list of tasks identified as desirable is presented below. (In A1, the tasks are additionally labelled with categories of person-month effort.)

T1 SLSTR new ADFs for cloud likelihood

T2 Synthetic SLSTR scenes for algorithm developments

T3 I-stripe Reflectance Products for use with IR

T4 New algorithm for prior SST near land in MET TX given NWP

T5 Quantification of coastal spatial SST variability at <1 km resolution

T6 Joint dual-view cloud detection

T7 Improved dust/aerosol forward model simulations

T8 Forward model accuracy and uncertainty for sunglint in SLSTR

T9 Propose and evaluate new Bayesian measures

T10 Three-way classification distinguishing water and sea ice

T11 V2 Coefficient Generation Tool

T12 Future-proofing SST coefficients

T13 Desert Dust Aerosol SST Corrections

T14 Daytime use of 3.7 µm coefficients

T15 Baseline bias aware optimal estimation

T16 Extended vector optimal estimation

T17 2d SST field estimation

T18 Review quality level definitions and practice

T19 Exploitation of OE goodness of fit

T20 Uncertainty model review and improvement

The recommended nine priority developments towards "Day-2" SLSTR SST products are listed below, plus an integration task and the diurnal variability task (for off-line validation).

Task	Type	Level
T1 SLSTR new ADFs for cloud likelihood	Classifier	L2
T4 New algorithm for prior SST near land in MET TX given NWP	Class./Retr.	L1
T7 Improved dust/aerosol forward model simulations	Class./Retr.	L2
T10 Three-way classification distinguishing water and sea ice	Classifier	L2
T11 V2 Coefficient Generation Tool	Retrieval	Offline
T13 Desert Dust Aerosol SST Corrections	Retrieval	L2
T14 Daytime use of 3.7 µm coefficients	Retrieval	L2
T19 Exploitation of OE goodness of fit	Quality Ind.	L2
T20 Uncertainty model review and improvement	Uncertainty	L2
Combi. Assess combined impact of above and implement in PP.	All	Offline
DV. Diurnal variability model for validation etc.	All	Offline/L2

The recommended approach in terms of “Retrieval” developments (T11 and T14) is continuity, based on a choice to focus on coefficient-based techniques. This represents a modification of current approaches to retrieval of SST rather than replacement. However, we note that a more radical but equally coherent choice could be made to focus on optimal estimation retrieval developments, which have the breadth and maturity of application to be considered for operational SST. This alternative approach would prioritize tasks T15, T16 and T17. It is not possible to determine *a priori* which of these two lines of development would ultimately lead to the better SST products, hence the default recommendation of a lesser-risk approach, incremental improvement of the coefficient-based performance. In principle, however, the reward of developing optimal estimation for SLSTR should be superior.

As to resources to permit, the best approach is to pursue both strands of retrieval development, push the methods to their limits and then select the better – or mix the more successful elements of each (e.g., OE-based estimates of correction quantities may be used to improve coefficient-based retrievals).

In terms of “Classifier” developments, all have well-founded expectations of solid product improvement. While T4 is most logically an L1 implementation task, it can be addressed at L2 if need be. T10 is a critical task towards integrated marine surface temperature products which are consistent across water/sea-ice boundaries. A further recommendation is that a PDP exercise with respect to ice surface temperature is undertaken in the coming year. T3 is considered of fundamental importance, but the interim solution provided via the ESL and PP already brings a significant advance.

The “Quality” and “Uncertainty” tasks are broad and fundamental, since these areas have not been intensively addressed in recent years, and present practice carries forward approaches that have been little developed since their first formulation (at which point they were state-of-the-art).

In addition to the tasks in the table, we recommended a task looking at the combination of the developments, to assess improvements and guard against unwanted interactions between developments. We recommend the development of a maintainable diurnal variability model, as discussed in the appendix A7.

2.2. “Day 3” developments

We recommend bias-aware optimal estimation (BAOE) methods as the core focus development towards “Day 3” SLSTR products. Note that the developments in three-way classification and coefficient adjustments are underpinned by radiative transfer and other developments that will also efficiently underpin BAOE. As mentioned, the BAOE tasks (T15 to T17) can beneficially be pursued in parallel to the recommended “Day 2” tasks from a scientific perspective (although this is not accounted for in the plan of section 3 from the perspective of available effort and expertise); or, T15 etc. could follow on from the prioritised plan.

The motivations for BAOE development are:

- BAOE provides a tool for empirically exploring the poorly understood BT-instrument temperature dependencies that affect SLSTR stability (see G. Corlett, SLSTR Quality Working Group presentation, February 2021).
- For wide-swath, single-view retrievals during daytime, BAOE is very likely to give reduced SST uncertainties compared to coefficient-based retrieval (even after improvement to the latter).
- The maintainability of the processing chain is improved in the sense that the necessary RTTOV aspects for cloud detection are also used in retrieval.
- BAOE can be extended to SLSTR matches to other EUMETSAT SST instruments, giving the framework in which the SST constellation can be systematically harmonised with dual-view SLSTR and HRSST reference sensors.

T17, which is 2d-SST field estimation, is also recommended for Day 3. Part of the motivation for EUMETSAT is that the techniques developed here would be equally applicable to any retrieved product where there is a complicated relationship between instrument geometry and an image grid appropriate for L2 or L3 products. In terms of SST, a 2d approach offers uncertainty reduction (as well as a better use of all SLSTR pixels) because “smooth-atmosphere” constraints are accommodated.

3. Plan for Priority Development Tasks

An example plan for is shown below. A two-year programme of "Day 2" developments is assumed to commence on April 1st 2022, which aligns with EUMETSAT's Day-2 target time scale in 2024. Time for combined assessment of developments is included. Elapsed time (with reasonable assumptions about the availability of expertise) is estimated. The "Combi" task involves assessment of the combination of developments and systematic inclusion in an updated prototype processor (PP). The "DV" task refers to the improved diurnal variability modelling discussed in A7.

Year	2022			2023				2024
Quarter	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**
T1								
T7								
T4								
T10								
T11								
T14								
T13								
T18								
T20								
Combi								
DV								

4. Implications for System

T1 SLSTR new ADFs for cloud likelihood

Impact on system: minimal.

The expected outcome of ADF cloud-likelihood improvements are ADFs that act as like-for-like replacements for the operational ADFs.

T4 New algorithm for prior SST near land in MET TX given NWP

Impact on system: recommended L1 solution implies coding changes in the processor calculating met-tx components of SLSTR L1 from NWP flows; alternative solution in L2 processor involves either (1) operational access to NRT OSTIA SST, or (2) significant additional ADFs (SST climatological information).

The forecast NWP fields used in creating met-tx files take a particular approach to specifying the met-tx surface temperature field near land which requires revision to achieve improved (unbiased) values in L1 products. The modified method is likely to be equivalent in its processing overhead to that of the status quo.

The alternative solution at L2 will substitute the surface temperatures in met-tx files with new values based on alternative information sources and a different method of extrapolation of SST “into” land areas to ensure an unbiased prior exists for all SLSTR sea-viewing pixels at full resolution.

The best scientific quality solution is for the alternative information source to be an operational L4 analysis SST, specifically the OSTIA product: this is a significant additional system requirement and dependence. A workable solution is for the alternative information to consist of a substantial new ADF of SST climatology (from SST CCI) to be used with the existing met-tx information (details TBC, but involving new interpolation steps).

T7 Improved dust/aerosol forward model simulations

Impact on system: additional ADFs and slower radiative transfer modelling at L2.

The L2 processor will require significant additional ADFs containing aerosol climatological information. Radiative transfer simulations (RTTOV) will be slightly slower (small impact expected).

T10 Three-way classification distinguishing water and sea ice

Impact on system: L2 processing time increases, additional L2 ADFs and L2 product format changes to accommodate sea-ice class.

The processing overhead in 3-way compared to 2-way classification reflects additional computation and additional ADF look-ups, and is expected to be moderate. Significant additional ADFs are required, such as prior sea ice probability climatology and ice-class likelihood tables. To accommodate the sea-ice class on output, small changes to L2 format are implied.

T11 V2 Coefficient Generation Tool

Impact on system: none (directly).

Note that the CGT v2 will facilitate activity such as future-proofing SST coefficients, that could in turn imply an additional dimension (time) for the interpolation of the coefficient ADFs.

T13 Desert Dust Aerosol SST Corrections

Impact on system: L2 processor, slower radiative transfer.

Additional, no-aerosol simulations will be required, meaning additional runs of RTTOV. The processing overhead increase is expected to be moderate.

T14 Daytime use of 3.7 µm coefficients

Impact on system: L2, additional/re-organised ADFs; additional radiative transfer simulations.

The likely solution will combine two-channel and three-channel coefficients with weighting determined by an algorithm to be developed. RTTOV will need to be run for day-time scenes with the solar source and additional 3.7 µm channel, with an expected moderate impact on processing overhead.

T19 Exploitation of OE goodness of fit

Impact on system: likely minimal.

Modified algorithms are likely to be similar in processing overhead and ADF usage as is presently operational.

T20 Uncertainty model review and improvement

Impact on system: likely small.

Modified uncertainty algorithms are likely to be similar in processing overhead and ADF usage as is presently operational, but small additional processing overheads may arise.

A1. List of Potential Developments

The tasks are labelled as requiring low (L), moderate (M) or high (H) person-months (PM) of effort, as provisionally estimated at the level of task definition provided in this PDP.

Key:

- L = 1 to 4 PM
- M = 5 to 11 PM
- H = 12 to 20 PM

T1	SLSTR new ADFs for cloud likelihood	L
T2	Synthetic SLSTR scenes for algorithm developments	M
T3	I-stripe Reflectance Products for use with IR	M
T4	New algorithm for prior SST near land in MET TX given NWP	M
T5	Quantification of coastal spatial SST variability at <1 km resolution	H
T6	Joint dual-view cloud detection	M
T7	Improved dust/aerosol forward model simulations	L
T8	Forward model accuracy and uncertainty for sunglint in SLSTR	M
T9	Propose and evaluate new Bayesian measures	M
T10	Three-way classification distinguishing water and sea ice	H
T11	V2 Coefficient Generation Tool	L
T12	Future-proofing SST coefficients	L
T13	Desert Dust Aerosol SST Corrections	M
T14	Daytime use of 3.7 µm coefficients	M
T15	Baseline bias aware optimal estimation	M
T16	Extended vector optimal estimation	M
T17	2d SST field estimation	H
T18	Review quality level definitions and practice	L
T19	Exploitation of OE goodness of fit	L
T20	Uncertainty model review and improvement	M
Combi	Assessment and verification of combined effect of all tasks	L
DV	Diurnal variability model optimisation	M

A2. Prioritisation Methodology

The possible tasks T1 to T20 were scored on:

- anticipated benefits to SLSTR L2 products for SST (coverage, uncertainty, etc.),
- scientific risk (confidence of success in scientific development),
- implementation risk (complexity of implementation or processing impacts).

The scoring was 1 to 5 with: 5 for the largest benefits; 5 for the lowest scientific or implementation risk. Additionally, the set of "winners" (scores of 12 or more) were reviewed to ensure they constitute a coherent, mutually beneficial set of actions, which they do.

	Task	Benefit	Sci. Risk	Imp. Risk	Total
T1	SLSTR new ADFs for cloud likelihood	4	4	5	**13**
T2	Synthetic SLSTR scenes for algorithm developments	3	1	5	**9**
T3	I-stripe Reflectance Products for use with IR*	3	3	2	**8**
T4	New algorithm for prior SST near land in MET TX given NWP	5	4	3	**12**
T5	Quantification of coastal spatial SST variability at <1 km resolution	4	2	3	**9**
T6	Joint dual-view cloud detection	3	3	3	**9**
T7	Improved dust/aerosol forward model simulations	4	5	5	**14**
T8	Forward model accuracy and uncertainty for sunglint in SLSTR	3	3	3	**9**
T9	Propose and evaluate new Bayesian measures	3	2	4	**9**
T10	Three-way classification distinguishing water and sea ice	5	4	3	**12**
T11	V2 Coefficient Generation Tool	4	5	5	**14**
T12	Future-proofing SST coefficients	3	4	4	**11**
T13	Desert Dust Aerosol SST Corrections	5	3	4	**12**
T14	Daytime use of 3.7 µm coefficients	5	3	5	**13**
T15	Baseline bias aware optimal estimation	5	3	3	**11**
T16	Extended vector optimal estimation	4	3	3	**10**
T17	2d SST field estimation	5	3	2	**10**
T18	Review quality level definitions and practice	4	5	4	**13**
T19	Exploitation of OE goodness of fit	4	4	3	**11**
T20	Uncertainty model review and improvement	5	3	4	**12**

** This refers to work beyond the baseline on this topic already established within BCDS. The BCDS pre-processor is already a significant step for a useful "I stripe", and this is the reason this fundamental aspect gets a lower prioritisation here.*

A3. Development Tasks: Clear-sky Determination

A3.1. Introduction

This section details the development tasks aimed at improving the performance of clear-sky determination within the SLSTR SST production chain.

SST retrieval is undertaken under conditions of clear-sky and ice-free ocean. Most cloud cover is sufficiently opaque as to render the sensitivity of TOA BTs zero or small. Even where cloud is optically thin in the IR, it may change the relationship between window channels relative to clear sky conditions, and bias the retrieval. Additionally, atmospheric aerosol (typically desert dust) may sometimes be too optically thick for retrieved SSTs (or their attached uncertainty estimates) to be reliable, and complete or partial sea-ice in the field of view may have a skin temperature very different from valid SSTs, and thus also needs to be avoided for SST retrieval. This is often referred to as "cloud detection". Here it is more generally referred to as "clear-sky determination", addressing the need to identify image pixels for which the SST retrieval and uncertainty estimate are most likely to be valid.

The long-term framework for clear-sky determination is assumed to be Bayesian, for sound practical and theoretical reasons. The problem is clear sky determination for the SLSTR image given the observations and prior knowledge about factors affecting observations in both clear and cloudy circumstances. Thus, the problem is inherently Bayesian in nature, the Bayesian approach is natural. The implementation of Bayes' theorem can be extended and improved piecemeal as new insights are gained, by developing new "Bayesian measures", which are properties of the observations that give some separation between clear-sky and other conditions.

Essentially, any progress in clear-sky determination can be classified as:

- better using the available observations to provide measures with improved separation,
- better using the available background knowledge to constrain the observations expected for clear vs. other situations.

Clear-sky ice-free ocean may be distinguishable from cloudy and other conditions by virtue of a contrast in some or all of the following, depending on the case:

- reflectance (brightness – e.g., ocean is often darker than cloud),
- reflectance spatial variability (clouds often more variable than ocean),
- reflectance spectrum (clouds are often whiter than other surfaces across reflectance channels),
- temperature (clouds are often colder than sea surface),
- temperature differences (i.e. effectively the infrared spectrum, affected by spectral emissivity as well as by differential atmospheric absorption, is often distinct),
- temperature spatial variability (clouds often more spatially variable than ocean).

Note the recurring "often" in the above list: none of these contrasts between clear and other is found in all circumstances. The need to combine and 'trade-off' different contrasts is met by adopting a Bayesian approach. While obvious clouds are easily identified, no single measure above on its own is satisfactory to discriminate more marginal cases (thin or subpixel cloud fractions, low warm clouds under low illumination, etc.) with an optimal balance of detection and false alarm rates. The Bayesian approach gives an objective means to automatically combine the measures, with appropriate weights.

A3.2. SLSTR-specific cloud likelihood ADFs

Problem addressed: existing cloud likelihood look-up tables used for SLSTR are adapted from other sensors with different viewing geometry. Some extreme look-up situations give rise to artefacts in the probability field because of

statistical sampling limitations. After a few years of operation, it is now possible to generate statistically robust SLSTR-specific tables, greatly reducing the incidence of such artefacts.

Here is an example of a probability of clear field for a region of extreme SST, with an unusual diamond-shaped artefact of low clear-sky probability (blues and yellow in the lower left quadrant), whereas the coastal and cloud clear-sky determination for the remainder of the image is geophysically convincing.

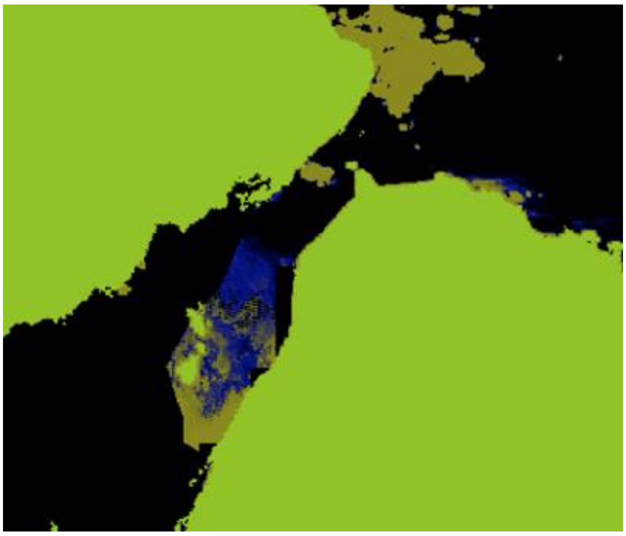

Figure 1 Black-blue-yellow: Field of probability of clear over ocean, blues and yellow being below the usual threshold applied to clear skies. Green: land. From image tile 23070739-BAYES-Pclear-SLSTR-v02.0-fv01.0.html in forward view, cloud-screened with the operational cloud mask, except with corrected interpolation of NWP fields near coasts, in order to highlight the look-up-table related artefact. The area is the Gulf of Aden

T1 SLSTR new ADFs for cloud likelihood

Objective	To generate SLSTR-specific cloudy look-up tables fully with more robust statistical sampling properties under extreme conditions.
Required Research	Review of LUT axes/bin-structure appropriate to SLSTR-specific geometry, for all foreseen cloud-detection configurations. Calculation of data volume required for LUT population. Decision on baseline cloud detection to be applied to bootstrap the LUT. Sample the required (large) volume of SLSTR-A segments, building the LUTs. Populate LUTs. Inspect and test. Building new LUTs would give the opportunity to include a PDF with the 3.7 µm observations during the day, providing a basis for testing the usefulness of this channel for daytime cloud detection purposes.
Required Algorithm Development	Calculate the (expected small) reflectance/BT transformations to apply to SLSTR-B etc. observations for use of LUT derived from A.
Comments on implementation	Implementation of new LUTs is at L2.
Benefits beyond state of the art	Improved performance of cloud detection under extreme look-up conditions, specifically configured to SLSTR geometry.

A3.3. Synthetic Scenes to Support Developments

Problem addressed: Validation of clear-sky determination is difficult because of the absence of truth.

Expert judgement has been used to develop complete reference cloud masks for 10 ocean scenes for SLSTR (Bulgin and Merchant, 2018), but this is surprisingly labour intensive. Expert judgement using a pixel selection approach is another option. A problem with both expert approaches is that the most ambiguous cases to the expert are precisely those cases which are of interest to test automated methods: in complete-cloud-mask expert references, these ambiguous cases are more likely to be misclassified in the reference; in pixel-selection references, ambiguous cases are likely to be under-sampled because of a selection bias towards pixels whose cloud / clear status is confidently assessable.

Therefore it is worthwhile to consider whether synthetic scenes could help the objective development and improvement of cloud detection. Having synthetic cases where the synthetic truth for all aspects of the surface and atmospheric state is known could support clear-sky determination algorithm development more objectively. Advantages of synthetic scenes are:

- objective quantifiable comparison of classified/retrieved status compared to the synthetic truth,
- ability to manipulate scenarios such as instrumental noise levels, calibration biases, RTM biases, prior biases, etc., for sensitivity analysis,
- via synthesis of scenes both with and without clouds, new means to objectively quantify the retrieval impacts of classification failures,
- additional applications beyond algorithm development, including use within unit tests for L2 processors.

T2 Synthetic SLSTR scenes for algorithm developments

Objective	To generate synthetic SLSTR observations for which the 'true' surface and atmospheric state is fully known, for use in objective algorithm testing (for clear-sky determination and other uses)
Required Research	Scoping of number and sampling of synthetic scenes. Definition of radiative transfer modelling approach. Prescription of all necessary geophysical variables at pixel-relevant resolution for radiative transfer (geophysically plausible downscaling from NWP resolution required). Radiative transfer simulation of synthetic observations (error free): both with and without clouds. Procedure for selecting error covariances and applying (and varying) associated synthetic errors.
Required Algorithm Development	Efficient radiative transfer simulation strategy for synthetic scenes, and control framework for applying synthetic errors.
Comments on implementation	Not part of operational implementation, except that these could provide a basis for some L2-processor unit tests.
Benefits beyond state of the art	Objective quantitative comparison of L2 processing options for clear-sky determination (new). Objective quantification of product errors arising from classification errors (new).

A3.4. Reflectance on the Infrared Image Grid

Problem addressed: inconsistency of fields of view when using operational reflectance and infrared image grids together.

For SLSTR, reflectance channels are available to aid clear-sky determination in daytime scenes. Some complexity arises because there is no co-registration between reflectance channels and the infrared channels. (This section draws on work done in early 2021 within the ESL and BCDS.)

SLSTR infrared imagery in current products is regridded to an image grid with a nominal 1 km resolution. The reflectance channels are regridded to a finer image grid, nominally 0.5 km resolution. Grid cell corners are conceptually aligned between the grids. The "natural" way of creating a reflectance image to match the infrared image grid is therefore to do 2x2 pixel averaging of the reflectance image (usually the A stripe by default).

Analyses within the BCDS confirm that 2x2 averaging can be improved upon, to reduce VIS-IR inconsistencies that arise from:

- duplicate pixels (used to fill the image grid when the instrument grid undersamples the image grid),
- independence between the mapping of instrument to image grid for the reflectance versus infrared channels: this inevitably leads to some incompatible choices.

Some examples of the estimated FOVs obtained by 2x2 averaging are shown in Figure 2. The degree of correspondence between the reflectance and IR FOVs is seen to be quite variable.

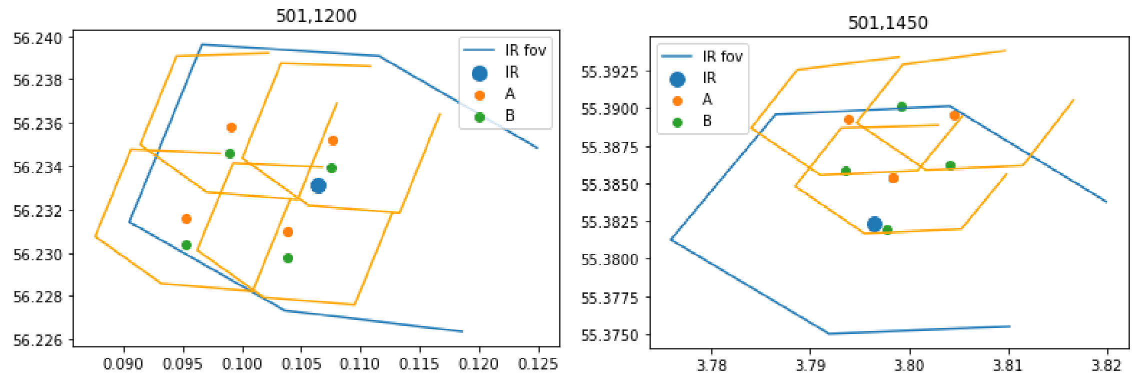

Figure 2 Comparison of estimated FOV between reflectance and infrared pixels. Cylindrical longitude-latitude projection. Blue: approximate IR FOV. Orange: A-stripe reflectance FOVs for the 2x2 pixels corresponding to the IR pixel in the image grids. Green: pixel centres for B-stripe. Left: for image coordinates (501,1200); in this case the reflectance pixels align well with the boundary of ~3/4 of the IR FOV; A and B stripe pixels are highly overlapping. Right: for image coordinates (501,1450); here there are three A-stripe FOVs because of cosmetic duplication in the image grid; the reflectance FOVs leave a large fraction of the IR FOV uncovered; A and B stripe pixels are better interleaved.

Within the ESL and through the BCDS study, a preprocessor module in Fortran that more optimally selects reflectance pixels to match an IR FOV has been developed. The basis is simply to select the N reflectance pixels with the shortest Euclidean distance between their centres and the IR FOV centres, using the x-y information in the SLSTR products. The following aspects are selectable:

- N, number of reflectance pixels
- Use A-stripe only or A- and B- stripe where available (in the latter case, collecting 2N reflectance pixels)
- Aggregate the selected pixels using arithmetic mean, maximum or standard deviation.

Note that orphan pixels are interrogated as well as pixels in the image grid, when identifying the nearest N. Experience suggests N = 5 is an effective choice, and favours using the A- and B-stripes where both are available (2N pixels).

An example of the improved consistency of the reflectance i-stripe (reflectance channels on the IR grid) thus obtained is shown in Figure 3.

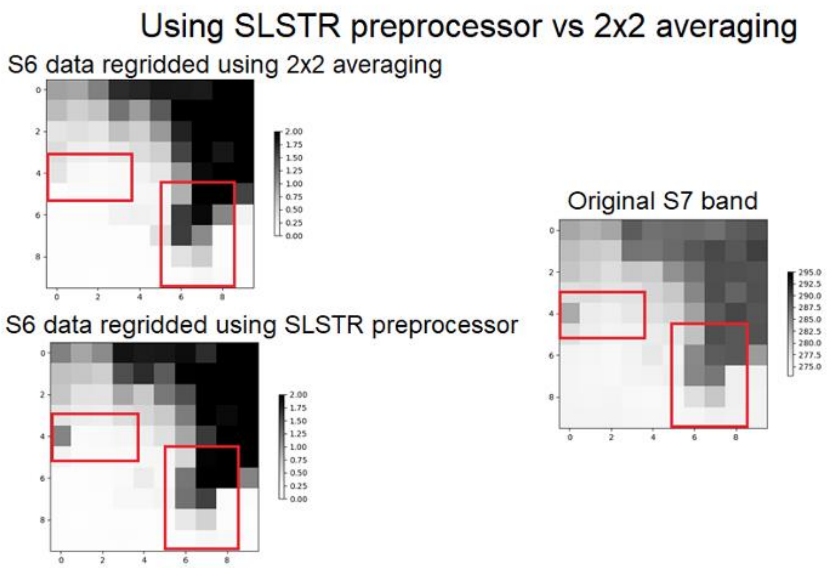

Figure 3 Left: SLSTR channel 6 extract regridded using (upper) 2x2 averaging of image-grid pixels or (lower) the SLSTR preprocessor with N = 5 (A-stripe only). Right: the corresponding IR extract in channel 7. This is a coastal zone with islands, and because the land is both brighter and warmer than the sea, there is strong correlation between the extracts. Red boxes indicate where improved reflectance-IR spatial consistency is most obvious.

Improved reflectance-IR consistency on the IR image grid benefits all cloud detection, including Bayesian clear-sky determination, by reducing the incidence of contradictory indications of cloudiness or clearness between the reflectance and IR. Particularly over fields of scattered cumulus, which are often on scales similar to the pixel FOV, it is important that the effective reflectance FOV overlaps with the whole IR FOV.

While the approach described above does bring a useful improvement (see outcomes of BCDS study), further developments should be considered. These are perhaps best addressed within the ground segment generating L1 products rather than at L2.

T3 I-stripe Reflectance Products for use with IR

Objective	Provide at L1 a reflectance product, in addition to the A and B stripes, which is optimised for use with the IR image.
Required Research	Case-study verification of useful impact. Define practical, effective method. Could amount to optimisation of the X% and Y% parameters mentioned below.
Required Algorithm Development	A practical, effective method needs to be defined. One attractive algorithm that could work is described below: FOV estimates (in the form of hexagonal shapes) can now be calculated for SLSTR imagery, although they are not standard products. Implement FOV estimates in the L1 processing chain as the basis for “reflectance i-stripe” estimation. Given available FOV estimates, develop an algorithm that optimally combines A and B stripe reflectance pixels to cover the IR FOV. The principles to be encoded in the algorithm for doing so could be: (A) the reflectance pixels that are combined cover between them >X% of the IR FOV, (B) none of the reflectance pixels combined has more than Y% of its area beyond the IR FOV.

	Starting point estimates for these parameters: X ~ 90%; Y ~ 20%. It is possible that cases arise that cannot simultaneously satisfy both parameters, and a refined decision process for such a case will need to be designed.
Comments on implementation	An important part of algorithm development would be to make the reflectance i-stripe calculation computationally efficient, as searches and calculations of overlapping areas in higher-level languages can be slow. For resilience to possible scenarios, the implementation should be configurable as to whether it searches the A stripe, the B stripe or both. The recommendation that this is a task upstream of L2 arises because this reflectance i-stripe would be valuable to many product streams, potentially. However, it may be that the algorithmic complexity is such that timeliness could not be achieved for NRT and that the reflectance i-strip is made available for NTC products.
Benefits beyond state of the art	It is not known how much practical benefit this much more intensive, rigorous decision process for C-stripe pixels would bring, in comparison to the nearest-N approach implemented in the L2 preprocessor that exists. For this reason, a precursor study is recommended to assess this (in which speed considerations are not prioritised).

A3.5. Prior SST from Numerical Weather Prediction

Problem addressed: severe cloud detection artefacts near some coastlines.

The 'MET TX' fields in the SLSTR product facilitate operational clear-sky determination by Bayesian (or any forward model-dependent) methods, and are also used in optimal estimation (such as in the Copernicus Global Land Service Lake Surface Water Temperature products). This is because the surface temperature, wind speed and atmospheric temperature and humidity profiles are used as inputs into radiative transfer modelling.

The NWP systems from which the MET TX fields are taken operate at coarser resolution than the SLSTR tie point or image grids. There is therefore a method to interpolate these fields spatially.

The surface temperature variables are particularly important priors that may vary on spatial scales finer than the atmospheric length scales. Two NWP temperatures are available: SST (defined only over water) and skin temperature (ST), which is the radiometric surface temperature over all surfaces.

The best use of these two NWP fields for prior clear-sky determination and SST estimation should be more thoroughly investigated in the light of the current NWP forecasting data-streams available and the definition and properties of the two surface temperature fields. (The current approach is a legacy from when ERA-40 re-analysis fields were applicable, which were at much coarser resolution than those available now.)

The nominal operational approach is simple bilinear interpolation of NWP surface temperature to SLSTR tie points, using SST where available and ST where SST is not available. (NB: In the BCDS, it has been found that this recommendation is not successfully implemented for NRT processing, with apparently 0 degrees Celsius entering the interpolation for NWP locations with no SST, instead of the ST. That is not the topic of this development task, however, which is more fundamental. The NRT issue is the topic of a Software Problem Report at time of writing.)

The limitations of the current approach (even after correction of NRT issue) are:

- In cases of extreme land-sea temperature contrast, use of ST may drive the coastal prior SST far from a realistic value, triggering false flagging of a coastal 'halo'.
- Coastal zones are expected to have relatively small-scale variability that are not expected to be captured within NWP SST/ST in current NWP systems. This could also lead to false flagging if the variability is large enough.

Concepts that could address the first of these limitations (the halo from land-sea contrast) include:

- prior to interpolation/extrapolation of the NWP SST/ST, expand the coverage of SST (by copying SST values into adjacent land locations, reducing the instances on which ST is used); or
- instead of ST, use a LUT of spatially expanded SST climatology as the value to interpolate towards where SST is not defined; or
- as above, but accounting for the nearest SST anomaly (rather than absolute) of the nearest NWP SST value to modify the expanded SST climatology; or
- context-dependent algorithms for extrapolation that use only NWP SST information (not additional ST or LUT values);
- use of an operational analysis L4 (OSTIA) expanded into land regions to replace the NWP-based surface temperature field.

The question of small-scale variability near coasts requires a more ambitious approach, since such variability will not by definition be addressable using NWP. To explore small-scale coastal variability in SST for the purpose of support SLSTR observations at 1 km scales, it would be ideal to quantify such variability with observations at finer resolution, on the scale of resolution (say) of the distance-to-land mask of [RD1] which informs SLSTR processing, i.e., 150 to 300 m.

An ambitious approach to this would be a comprehensive assessment of patterns and amplitudes of near-coastal variability using Landsat thermal observations (which have a real resolution in the thermal of <200 m) supplemented by ASTER, ECOSTRESS and (from 2024) TRISHNA observations. The resulting mean and modes of spatial variability would be quantified in appropriate modal patterns (e.g., using dynamically interpolation empirical orthogonal functions) and the weights of these modes parameterised in terms of information available to a processing chain (time, tidal phase and wind-speed strength and direction are the most likely informative parameters) could be used to modify the NWP SST near coasts more realistically than linear interpolation. Given predicted patterns of spatial SST anomaly on a ~300 m scale relative to the larger scale available from NWP tie points, estimation of a prior SST per SLSTR pixel becomes possible for clear-sky determination.

This is a major undertaking, and not likely justifiable for SLSTR product improvement alone; however, the results of such a project would be of widespread scientific interest and application to other SST processing chains, e.g., in the OSI-SAF, as well as for the coastal-zone information systems that increasingly being called for in the context of information to support climate resilience decision-making.

(The question of handling coastlines where tidal range is significant at the SLSTR observational scale should also be addressed where information is available, but is not considered further here.)

T4 New algorithm for prior SST near land in MET TX given NWP

Objective	**Develop a means to handle the NWP information on prior surface temperature that is less prone to creating a "halo" of false flagging near costs**
Required Research	Testing of alternative options, for example on the BCDS diagnostic data set
Required Algorithm Development	Depending on the solution: • robust routines for flexible extrapolation of NWP SST into land locations, • development of LUT of spatially-expanded SST climatology and module to utilise the climatology as required.
Comments on implementation	The uninterpolated NWP information is part of the processing chain to level 1, and the improved NWP SST prior will therefore need to be implemented in the pre-level-1 processing chain at the point where MET TX fields are populated.
Benefits beyond state of the art	False-flagging "halos" are currently relatively common at distances within ~10 km coasts and significantly diminish the coverage by SLSTR of near-coastal zones. Even after correction of

the NRT bug, "halos" will continue to arise when land-sea temperature contrasts are large, in both NRT and NTC products.

T5 Quantification of coastal spatial SST variability at <1 km resolution

Objective	Provide a means to estimate prior SST in coastal zones on a scale fine enough to characterise coastal zone spatial variability in SLSTR observations
Required Research	Collection of all coast-zone observations at ~300 m or better capable of SST patterns on to a common grid. Auxiliary data on tides and winds. Application of SST pattern extraction (e.g., DINEOF) and analysis of dependencies of patterns on tidal and wind-driven dynamics.
Required Algorithm Development	Develop an ADF of SST fine-scale coastal spatial variability relative to off-shore SST, mean, modes, and parameterisation of mode weights on dynamically relevant parameters (tidal phase and wind, the latter being available from the NWP) Determine options for a reliable data-stream of adequate tidal phase information. Algorithmic use of ADF and parameterisation given tidal phase information and wind state from NWP, to map into prior SST per pixel in SLSTR nadir and oblique views, and also how fine-scale variability modifies the expected local standard deviation of the 11 um channel used in clear-sky determination.
Comments on implementation	The implementation in the L2 chain may be relatively complex with significant auxiliary data files and may require an additional data stream of tidal phase information. The implementation therefore may not be feasible for NRT – this would need to be explored.
Benefits beyond state of the art	It is not known how much practical benefit to clear-sky determination in coastal zones such a development would bring, although in principle it could make the difference between a system that is able to track important events such as upwelling in the SST signature, and one that is intractably incapable of this because open-ocean assumptions in the clear-sky determination lead to permanent exclusion of the affected coastal area. For this reason, a precursor study is recommended to assess the benefit, which although unpredictable may be significant. The coastlines of France and the UK would be appropriate for such a precursor study.

A3.6. Joint Dual-View Masking

Problem addressed: joint dual-view constraint on plausible clear-sky radiance combinations is not used when each view is individually screened. Reduced SST errors at the expense of reduced coverage was observed when this was last studied for ATSR.

For the dual-view SST retrievals, the BTs in both the nadir and oblique views should be clear-sky over ice-free ocean (hereafter, simply "clear"). "Dual-view masking" refers to simultaneously assessing the probability of clear with reference to BTs from both nadir and oblique views together in a joint probability distribution. This has been done for ATSR-series sensors before, although the present operational scheme for SLSTR uses multiple channels in each view separately.

In practice, the fully joint solution is tractable for two channels in each view at most. This is because the cloudy probability density function is evaluated via an ADF, which becomes too large to populate with statistical confidence if there are too many dimensions (in practice, more than 4 or 5).

In principle, it is expected that the joint dual-view constraint should be more discriminating of cloud in any circumstance where the cloud temperature is more non-uniform spatially on the scale of nadir pixels than is the sea. The anecdotal experience from the ATSR work was that where the forward view was less likely to be confident about the presence of cloud (because of its lower resolution and longer atmospheric path), joint cloud detection reduced the rate of missed cloud in the forward view compared to forward-only calculations; the same could be expected for the SLSTR oblique view.

There is more than one option for screening with greater exploitation of the dual-view:

First option: treat the views as independent events but threshold the probabilities jointly. Presently, two threshold tests are applied: let the Bayesian probability of the nadir being clear be $P_n = P(c|\boldsymbol{x}, \boldsymbol{y}_n)$ and the corresponding probability for the oblique be $P_o = P(c|\boldsymbol{x}, \boldsymbol{y}_o)$. Currently, the truth values of $P_n > p_{th}$ and $P_o > p_{th}$ are evaluated and the dual-view SST is given a flag corresponding to confidently clear retrieval if both are true. The simplest joint approach is instead to evaluate $P_o P_n > p_{th}$. For example, if the threshold were $p_{th} = 0.9$ and P_o and P_n were respectively 0.91 and 0.98, the dual-view retrieval would presently be given the highest quality level. However, viewed as two independent events, the correct estimate of the probability of the two views both being clear is $P_o P_n$ which would be 0.89 and below the threshold for the highest quality level.

Second option: do a joint Bayesian calculation under the approximation that the two views are independent. Within the Bayesian calculation there is effectively a calculation of the ratio of likelihoods of the observations under clear and cloudy conditions. This option is a dual-view implementation where the approximation is made:

$$\frac{\boldsymbol{P}\left(\begin{bmatrix}\boldsymbol{y}_n\\ \boldsymbol{y}_o\end{bmatrix}\middle|\boldsymbol{x},\boldsymbol{c}\right)}{\boldsymbol{P}\left(\begin{bmatrix}\boldsymbol{y}_n\\ \boldsymbol{y}_o\end{bmatrix}\middle|\boldsymbol{x},\overline{\boldsymbol{c}}\right)} \approx \frac{\boldsymbol{P}(\boldsymbol{y}_n|\boldsymbol{x},\boldsymbol{c})}{\boldsymbol{P}(\boldsymbol{y}_n|\boldsymbol{x},\overline{\boldsymbol{c}})}\frac{\boldsymbol{P}(\boldsymbol{y}_o|\boldsymbol{x},\boldsymbol{c})}{\boldsymbol{P}(\boldsymbol{y}_o|\boldsymbol{x},\overline{\boldsymbol{c}})}$$

where $P(a|b)$ is the probability density of occurrence a conditional on b; $\boldsymbol{y_n}$ is a set of observations in the nadir view, with $\boldsymbol{y_o}$ being that in the oblique view; $\boldsymbol{x}$ is the prior state; c is the situation of being clear sky, and $\overline{\boldsymbol{c}}$ of being cloudy.

This option can be equivalent to the first option if the same sets of Bayesian measures are used as in the single-view tests, but allows for this not to be the case. For example, with this approach, it may be considered best not to use the local standard deviation (LSD) of the oblique view BT in the joint Bayesian calculation, because the oblique view has poorer real spatial resolution; however, in the oblique single-view evaluation, however, the LSD could still be used.

Third option: fully evaluate $P\left(\begin{bmatrix}\boldsymbol{y}_n\\ \boldsymbol{y}_o\end{bmatrix}\middle|\boldsymbol{x},\boldsymbol{c}\right)$ and its cloudy-sky counterpart. In theory, this is the most effective approach since the full multi-dimensional constraint is imposed. The evaluation of the clear-sky probability density is readily evaluated by the same means (based on radiative transfer simulation) as are used in each individual view. The evaluation of the cloudy-sky probability is done via a ADF and this brings some limitations. The ADF must be empirically populated, and this becomes difficult to achieve with statistical confidence when the number of dimensions becomes too large. Joint use of 4 BTs is feasible, but not ADFs exceeding five dimensions.

The three different options thus have different trade-offs which would benefit from systematic analysis.

In addition to these options, there are other aspects to joint dual-view cloud detection that ideally should be explored for SLSTR.

First aspect: Consideration of oblique/nadir relative pixel geometries. Across the oblique (i.e., shared nadir and oblique) swath width, there is no co-registration between the pixels of the two views, despite them being presented on a common "image grid" – see Figure 4.

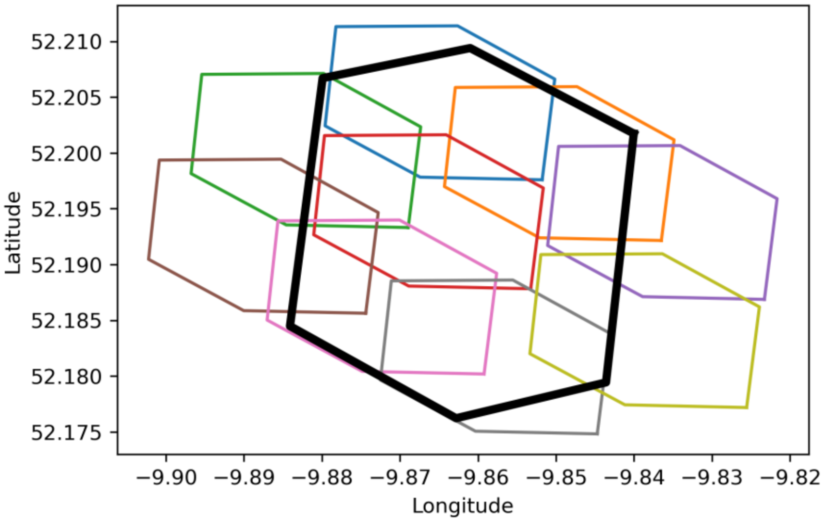

Figure 4 Field of view estimates (courtesy of Caroline Cox, pers. comm.) for an oblique IR pixel (thick black polygon) and the nine nadir pixels whose FOVs overlap with the oblique FOV by 30% or more (coloured polygons). Under dual-view joint cloud detection implementations, it may be beneficial to account for FOV-relationships between views, rather than relying on the pairings suggested by the SLSTR L1 nadir and oblique image grids

Oblique pixels are larger, and, for this reason, isolated sub-pixel clouds are more difficult to detect in the oblique view. It may be that by using not a single nadir BT in joint cloud detection, but an appropriate selection of overlapping nadir BTs, additional precision is introduced to the expected BT relationships (reduced noise in the nadir BT) that is beneficial to the joint determination. On the other hand, it is geometrically possible for the oblique FOV to be truly clear, while there is cloud present in some of the overlapping nadir pixels, so some subtlety may be required to achieve the right trade-offs.

Second aspect: Interaction with nadir-only cloud detection. Is joint dual-view detection best approached independently of nadir-only results? The nadir-only cloud detection results have an advantage of higher spatial resolution, which enables better detection of the presence of cloud within a pixel. How could the results of a nadir-only calculation of probability of clear sky inform the dual-view cloud detection? It may seem appropriate to impose the condition that a pixel flagged in nadir-only cloud detection is automatically flagged in the context of joint dual-view cloud detection. On the other hand,

- all cloud detection results have a non-zero rate of false flagging, and additional information from the oblique view doubtless would sometimes restore some flagged nadir pixels to being judged as clear,
- the oblique view pixel of relevance to a given nadir-view pixel is not a constant, but instead it depends on the cloud top and base heights because of parallax effects.

Systematic consideration of these aspects is required if fundamental progress is to be made.

T6 Joint dual-view cloud detection

Objective	Quantify any advantages that accrue for SLSTR SSTs from a joint dual rather than view-by-view approach. Select and implement a joint-view approach from various possibilities.
Required Research	Systematic testing of performance of alternative options, for example on the BCDS diagnostic data set or synthetic SLSTR scenes.
Required Algorithm Development	Adaptation of per-view to joint dual-view Bayesian calculation of probability of clear sky.

Comments on implementation	Implementation will be in the Bayesian module at the L2 processor. NB joint screening is only possible for part of the nadir swath, of course. For this reason, this step is in addition to per-view cloud detection and would be applied for dual-view SSTs.
Benefits beyond state of the art	Considered likely to decrease failure to detect cloud in the oblique view.

A3.7. Forward Model Simulations: Dust

Problem addressed: current forward model simulations assume that clear-sky is aerosol-free, whereas dust-affected scenes likely return negatively biased SSTs (where the dust is not thick enough to trigger cloud detection).

The current Bayesian clear-sky determination is a binary classifier which classifies the pixels based on their similarity to the forward model simulations (RTTOV clear-sky, aerosol free, conditions) or the cloudy PDF lookup table. Pixels which contain desert-dust (e.g. Saharan dust outbreaks over the Atlantic Ocean) are not well suited to this scheme:

- Optically thin aerosol can be difficult to distinguish from clear-sky and single-view classification as currently used will often classify dust affected pixels as clear-sky.
- However, with joint dual-view masking (see A3.6) it is likely that dust affected pixels will be classified as cloud.
- Some applications (e.g. SST retrieval) need to determine cloud-free and dust-free pixels, while aerosol-retrievals need to distinguish cloudy pixels from dusty pixels.
- Future retrieval work (see A4.5) may improve performance of SST retrievals in the presence of dust aerosol in which case we will want to classify dust-affected pixels as not-cloudy.

The current version of RTTOV includes support for radiative transfer simulations including the effects of aerosol using either OPAC or CAMS aerosol scattering properties. If combined with a suitable aerosol climatology or aerosol prior (e.g. CAMS analysis), this will lead to better control of the classifier in the presence of aerosol.

This could also be combined with an extension to the multi-way classifier work (A3.10) to classify pixels as clear, clear-with-dust-aerosol or cloud (away from sea-ice).

This work is also a necessary precursor to improve dust-robustness for SST retrieval, whether coefficient-based or OE-based.

T7 Improved dust/aerosol forward model simulations

Objective	Improve forward model simulations in the presence of dust and other aerosol
Required Research	Determine required result - should dust-affected areas be considered "clear" or "not-clear" Investigate available RTTOV aerosol options and available prior aerosol information
Required Algorithm Development	Adaptation of RTTOV inputs to include suitable aerosol prior – possibly from climatology or aerosol analysis. It is noted that parallel work is ongoing within the consortium (MF) in preparation for Metop SG with which this can be co-ordinated, as well as with SST CCI research.
Comments on implementation	Implementation will be in the Bayesian module at the L2 processor.
Benefits beyond state of the art	Better control if "aerosol" is classified as "clear" or "not-clear" If combined with A3.10 would allow classification as "clear", "dust", or "cloud"

A3.8. Forward Model Simulations: Sunglint and Uncertainty

Problem addressed: optimum cloud detection with areas of sunglint depends on an effective uncertainty model for glint-affected simulations, while the existing uncertainty model is generic (e.g., independent of fetch and wind speed) and based on earlier simulation capability (ignoring aerosol).

In Bayesian clear-sky determination, reflectance and infrared channels under clear-sky assumptions are simulated. RTTOV is used, and includes capability of simulation of sunglint into the satellite view from solar reflection in the wind-roughened sea. The sea surface in such a geometry has a bright reflectance whose value depends on the closeness of the geometry to specular reflection and the bi-directional reflectance distribution function (BRDF) of the sea surface, which is sea-state dependent. Typically, NWP winds are used to parameterise the sea state, since empirically wind speed and wave-facet slope distributions have been found to be relatively tightly coupled (although other factors also affect wave-facet slope distribution).

The limitations for clear-sky determination that arise are as follows:

- For any given reflectance channel, there is a possibility that by chance the reflectance from a cloud within the glint region closely matches the reflectance that would be obtained from glint with no cloud: in this circumstance, the reflectance measure offers no discrimination for geophysical reasons.
- Clear-sky sea areas that are shadowed by cloud receive less direct sunlight than assumed in the forward calculation and look unfeasibly low-reflectance compared to the RTTOV result.
- The NWP wind fields are imperfect, particularly around coasts and islands because of complex orographic interactions as well as near weather systems where spatial gradients in wind magnitude are large. The result is patterns of real smaller-scale BRDF variability whose disagreement with the RTTOV simulation is larger in magnitude (in either direction) than elsewhere (where the RTTOV BRDF and true BRDF agree more closely)
- Atmospheric aerosol modifies the sunglint reflectance and anecdotally extends the marginal area of elevated reflectance around the sunglint geometry.

The uncertainty attributed to the reflectance simulation is important to give the difference between observed and simulated reflectance appropriate weight in the calculation of Bayesian probability. For the current SLSTR Bayesian implementation, the following are true:

- The uncertainty assumption is based on legacy work undertaken on AVHRR on an out-dated version of RTTOV (with no aerosol capability), not updated to reflect improvements in NWP, nor for SLSTR geometries specifically
- Dependences of the sunglint-region uncertainty estimate on factors with which it may vary have not been evaluated (i.e. a context-independent assumption is implemented). Relevant factors to investigate include the proximity to coast (because of the BRDF influence), wind fetch (related to proximity to coast, but direction dependent), wind speed and the RTTOV-simulated reflectance itself.

T8 Forward model accuracy and uncertainty for sunglint in SLSTR

Objective	Explore discrepancies between simulated and SLSTR reflectance values for sunglint over ocean to determine: any useful bias adjustments to the forward model specific to the RTTOV version and SLSTR sensor; a model for the uncertainty of the RTTOV simulation in terms of any relevant factors
Required Research	Systematic data analysis of observations and simulations of SLSTR reflectances under sunglint conditions, to obtain any bias correction and an uncertainty model Pertinent factors for the analysis include: • coastal proximity, • fetch (given wind direction), • reflectance magnitude simulated,

	• aerosol, • geometry, including relative azimuth and tilt angles.
Required Algorithm Development	Parameterisation of any bias correction and of uncertainty model
Comments on implementation	Implementation will be in the Bayesian module at the L2 processor.
Benefits beyond state of the art	Decreased rate of missed cloud and false alarms in areas of sunglint.

A3.9. New Ideas for Bayesian Measures

Problem addressed: not all possible discriminating measures between clear and non-clear states have been explored or developed within a Bayesian context.

The “Bayesian measures” here refer to the quantities that offer some degree of separation between clear and cloudy sky distributions, and which are therefore used as observations in the Bayesian calculation. Examples (currently in use, or previously investigated) include: brightness temperature, brightness temperature local standard deviation, reflectance, and reflectance local standard deviation. (Within BCDS, sub-IR-pixel reflectance standard deviation was discussed as another measure, and this will not be repeated here.)

An additional Bayesian measure needs to satisfy the following:

- Clear and non-clear distributions of the measure need substantial separation.
- The separation should be distinct from separation in other measures: i.e. in some domains where existing measures are ambiguous, the new measure should be less ambiguous.
- A forward model (this conceptually includes a look-up of an ADF) for the measure is available for both clear and non-clear.
- The uncertainty distributions of the forward models are known and narrow compared to the distribution and separation.

Concepts for alternative / additional Bayesian measures are necessarily somewhat speculative until the above are evaluated. Here are some ideas:

- “blueness” or “non-whiteness”, i.e. using the spectral distribution across multiple reflectance channels in a combination that assesses the closeness of multi-channel reflectance ratios to what is expected of the scene under clear and cloud conditions,
- local BT combinations (to be defined) that may be less sensitive to the organised variability across ocean thermal fronts but as sensitive to the less regular edges and effects of clouds; may include multi-channel BT combinations; may need to be larger in scale than the 3x3 used for LSD (which could be a disadvantage),
- 3.7 µm BT during the day, which, for some cloud types gives distinct information compared to the split window channels, because of cloud-reflected solar irradiance only present in this channel.

T9 Propose and evaluate new Bayesian measures

Objective	Find and implement additional (or replacement) Bayesian measures that increase separation of clear and non-clear pixels
Required Research	Literature review for additional candidate measures, including image analysis literature on texture measures. Formulate each candidate measure in a form for use within Bayesian framework. For each candidate measure: • explore degree of separation of measure between clear and cloud cases, and understand behaviour in aerosol, sea-ice and partial land pixels, • assess degree to which separation is complementary to existing measures, • (if above steps suggest continuation with the candidate) develop a forward model or ADF for the measure, along with uncertainty estimates.
Required Algorithm Development	Implementation of additional measure within the Bayesian framework.
Comments on implementation	Implementation will be in the Bayesian module at the L2 processor.
Benefits beyond state of the art	Decreased rate of missed cloud and false alarms in areas of ambiguity given existing measures.

A3.10. Extension of number of Bayesian Classes

Problem addressed: for seamless surface temperature determination over water and sea-ice-covered ocean areas, e.g. marginal ice zones, multi-way classification is required.

SST and IST are desirable as integrated, consistent products, with retrieved ST values wherever observations are cloud free, requiring classification into:

- cloud-free over sea water,
- cloud-free over sea ice,
- cloudy over sea water or ice.

Multi-way classification in a single approach is desirable to ensure consistency of SST and IST products from SLSTR, i.e. that SSTs and ISTs are reported for mutually exclusive sets of cloud-free image pixels.

Significant challenges exist in classifying pixels in marginal ice zones. Newly formed ice is often dark, with a temperature close to the freezing point, making it difficult to distinguish from open water at both infrared and reflectance wavelengths. [RD2] demonstrates the benefits of three-way classification at high latitudes for AATSR data using additional information at reflectance wavelengths during the day. This included observations at wavelengths equivalent to S2 and S3, as well as the local standard deviation at 1.6 microns (equivalent to S5). The biggest benefit was to the classification of clear-sky-over-ocean observations, with an 89.9 % success rate in comparison to 65 % for the ARC algorithm, with a 9.9 % increase in overall classifier performance. Figure 5 shows a number of single channel and channel ratio PDFs for SLSTR observations, [RD3]. In channels S1-S3 we see good separability between water and sea-ice, and in the S2/S5, ratio a good separability between cloud and non-cloud pixels (water or sea ice). This provides a good indication that a PDF including S2, S3 and S5 has the potential to improve classification in marginal ice zones as demonstrated with AATSR data.

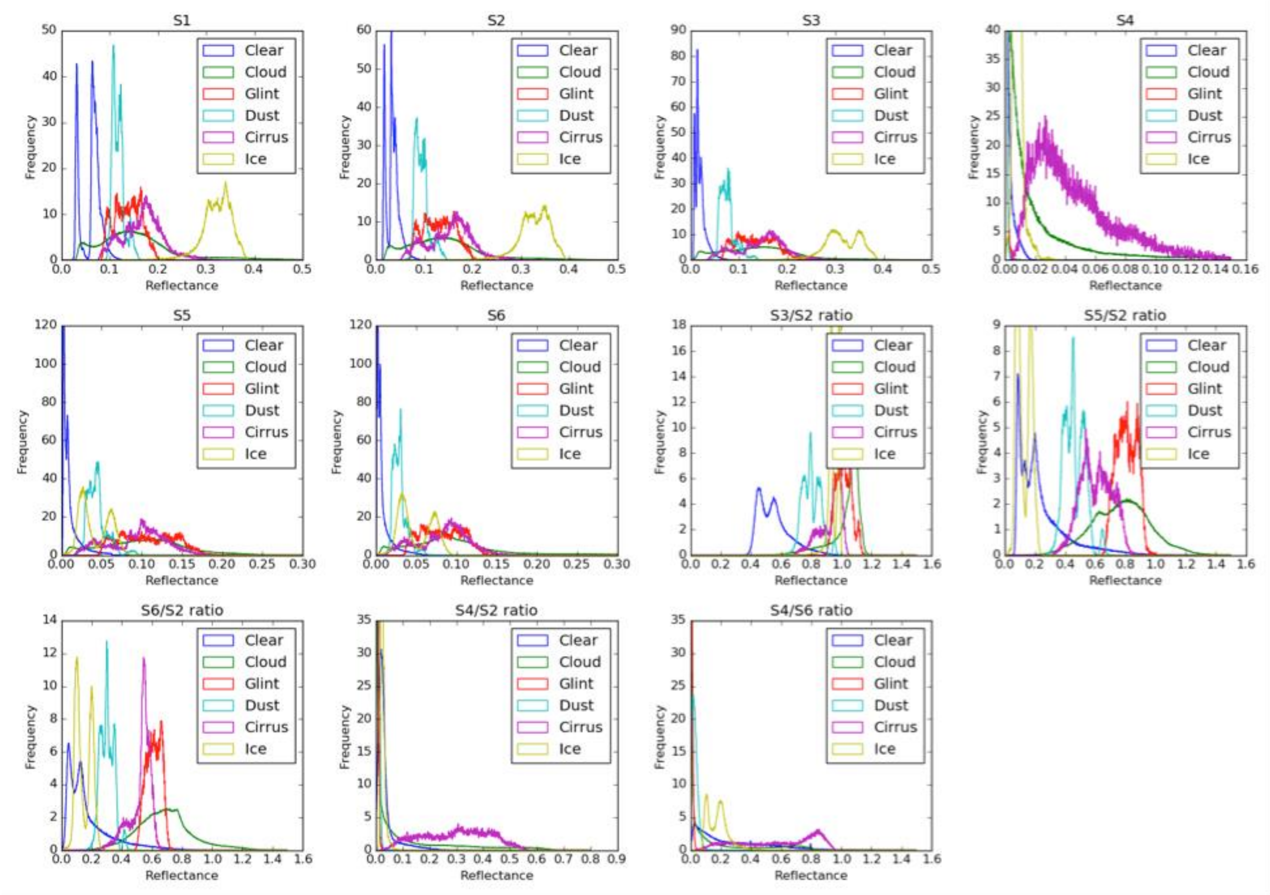

Figure 5 Single dimension PDF comparisons for data classified as clear (blue), cloud (green), sunglint (red), dust (cyan), cirrus (purple) and ice (yellow). Considering plots from left to right, then top to bottom, PDFs show individual channels (S1-S6), and then the following ratios: S3/S2, S5/S2, S6/S2, S4/S2 and S4/S6 [RD4].

Note that there is a distinction between having 3 classes in the Bayesian calculation and how this is reflected in SST and IST quality levels (which reflect the degree of trust in the validity of the retrieval context and associated uncertainty estimate). For example, some pixels in reality observe sub-pixel mixtures of water and sea ice, and can reasonably be expected to return intermediate probabilities for both cloud-free classes. It is not known whether an additional water/ice mixture class would be helpful to SST and IST products. Because of the different requirements for SST vs. IST, it may be appropriate to differently map classification probability into quality level for each surface type. Importantly, however, having a common classifier ensures that the SST and IST products can nonetheless be consistent.

Given the figure above, an important aspect of a 3-way classifier for ice will be use of S2 and S3 channels. These are also sensitive to ocean turbidity, both coastally generated and open-ocean blooms. The behaviour of an extended classifier in reflective-ocean conditions may need to be considered. The prior probability of sea-ice is strictly zero over many parts of the ocean where high ocean reflectance could mislead reflectance metrics towards a sea-ice class. Thermal metrics should work against this, and desirable behaviour should be verified.

NOTE – this task is potentially extensible to three-way classification of clear, clear-with-dust-aerosol and cloudy, if T7 is also undertaken prior to this task.

T10 Three-way classification distinguishing water and sea ice

Objective	To extend the SLSTR day-time Bayesian classifier to two cloud-free classes (for water and sea-ice surfaces), to facilitate consistency between SST and IST products
Required Research	Consolidate lessons from previous AATSR study in light of new RTTOV capabilities, SLSTR channels, and any subsequent research understanding. Establish test scenes for open ocean, full sea ice cover and marginal zones. Establish true classes for selected pixels of test scenes. Design, implement (likely “offline”) and assess performance of 3-way classifier on test scenes.
Required Algorithm Development	Addition of N-way classifier option within Bayesian prototype processor: complexity of algorithm development may be significant; this is to be assessed. N-way (including N = 2 and 3) is preferably built in at this stage in a generic way to future-proof against future developments.
Comments on implementation	Implementation will be in the Bayesian module at the L2 processor, at a point before L2 processor branches between SST and IST retrievals.
Benefits beyond state of the art	Fully consistent classification for SST and IST products from SLSTR

A4. Development Tasks: Sea Surface Temperature Retrieval

A4.1. Introduction

A4.1.1. Coefficients

This introduction starts with contextual information on SLSTR SST retrieval coefficients.

SLSTR SSTs are obtained using pre-defined coefficients, $\{a_0, \mathbf{a}\}$, that weight a set of observed brightness temperatures, $\mathbf{y}$, (the set being selected relative to the coefficient type) to give the SST estimate:

$$\hat{x} = a_0 + \mathbf{a}^{\mathrm{T}}\mathbf{y}$$

For SLSTR, the coefficient types are “N2”, N3”, “D2” and “D3” meaning that the observations respectively comprise:

- nadir 11 µm and nadir 12 µm,
- as above plus nadir 3.7 µm,
- nadir and oblique 11 µm and nadir and oblique 12 µm,
- as above plus nadir and oblique 3.7 µm.

The coefficients are pre-defined in strata against relevant factors, the stratifying variables being presently:

- zenith angle (in each view in the case of dual-view coefficients),
- TCWV (from NWP).

Time is a further potential stratifying variable, but this dimension is not operationally implemented for SLSTR. Having time as a stratifying variable is intended to address changing trace gas concentrations. In reprocessing activities, where the trajectory of trace gas concentrations is known retrospectively, coefficients are defined for the atmospheric conditions of every four years and interpolated in time. Operationally, the coefficients are held constant reflecting the concentration conditions of the epoch at which they are defined.

In previous work for ATSR-1, stratification by detector temperature was also applied, as the detector temperature had a strong temporal trend through the mission and was known to affect the spectral response function of particularly the 12 µm channel. (ATSR series coefficients were in general stratified by across track pixel number, which was more convenient for their geometry compared to SLSTR, rather than zenith angle.)

The coefficients are generated in two steps.

First, representative simulations of BTs given SST are inverted to give the coefficients that fit the SSTs with the least sum-of-squares residual. This may be done imposing an additional constraint (via Lagrange multipliers) that ensures the SSTs are unaffected by any BT variability along one or more axes in BT space: this enables dual-view retrievals to be resilient to potential major stratospheric aerosol events (although such an event has not occurred since 1991). The spread of residuals from the regression defines the uncertainty from the inversion process ambiguity, which depends on the atmospheric conditions observed and is taken to be associated with errors in SST that are correlated on synoptic scales of the atmosphere.

Second, the coefficients are applied to a set of observed BTs. The SSTs for different sets should theoretically be unbiased relative to each other (i.e., zero mean difference between SSTs derived from different sets), but in practice they are not. This is because the simulations of the BTs have some level of bias relative to satellite calibration. For the SST biases between algorithmic forms to be acceptably small (<0.1 K), the BT simulation and satellite calibration biases would need to be of order 0.01 K or smaller. They are not that small, in general. Among the various reasons why not, fundamentally 0.01 K is beyond the specification of the accuracy of the SLSTR instrument calibration by a significant margin.

Therefore, a reference algorithm is selected and within each stratum the offset coefficient of the other algorithms is adjusted to achieve the same mean SST as the reference algorithm on the set of observed BTs. This completes the process of definition. The reference algorithm is that which, on the basis of past experience and current validation, seems least biased at the SST level by the calibration/simulation biases. Usually a dual-view three-channel algorithm is the reference retrieval.

A4.1.2. Wider retrievals landscape

Next, consider the wider context of SST retrieval, which includes advances in the application of optimal estimation to SST.

Optimal estimation (OE) is an explicitly Bayesian retrieval formulation that adjusts prior expectations for the retrieved variable in the light of the observations relative to simulations of the observations based on the prior expectations. How the state affects the observations is represented through derivatives (Jacobians) that are also obtained from simulation. It can be an iterative process, but for a nearly linear case such as IR-based retrieval of SST, it has generally been done as a single linear step. In this case, it is equivalent to a retrieval based on linear coefficients except that the coefficients are dynamical determined for the given prior.

OE should in principle supersedes retrieval methods using fixed coefficients because:

- It is an optimal (e.g. minimum error variance) solution given prior knowledge of the retrieval context.
- The prior knowledge can supply understanding of the atmospheric state that helps disambiguate the retrieval in the presence of noise, which is particularly relevant for single-view two-channel ("N2") retrievals which are fundamentally deficient in retrieval degrees of freedom.
- OE automatically adapts to situations such as missing channels, changes in channel noise, etc. (so long as these situations are correctly passed to the algorithm, of course!).
- Standard OE equations provide uncertainty estimates, averaging kernel estimates (including "SST sensitivity") and quality-of-fit indicators that can be useful for quality level assignment.

However, for low-noise dual-view reference sensors like SLSTR, it has previously been found in practice to be more effective to use retrieval coefficients (except for N2 SSTs). This is because of the absence of procedures equivalent to the "offset adjustment" procedure for physics-based retrieval coefficients, to cope with calibration/simulation biases. OE theory simply assumes an unbiased prior state and unbiased simulations relative to observations.

The other question that has been raised about OE relates to the error covariance assumptions that are needed within the retrieval. These determine the degree to which the observations influence the answer and the intended optimal solution requires the prior and observation error covariances to have the right ratio. (Correct uncertainty estimation further requires that each has the right magnitude.) Previously, methods to assess the error covariance assumptions have been rather ad hoc

Recent progress [RD5], [RD6] has introduced "bias-aware optimal estimation" (BAOE) to remote sensing, and in particular to SST retrieval. BAOE adapts and combines two established techniques. A modification of concepts of parameter estimation by Kalman filtering is used to obtain parameters for bias adjustments of the prior state and the observations. Diagnostic ('Desroziers') equations from bias-aware data assimilation are adapted to provide estimates of the prior and observation error covariances.

BAOE techniques in principle offer a framework both to mature SLSTR as an SST reference sensor, and rigorously harmonise other sensors in the SST constellation to the SLSTR reference. The establishment of the SLSTR reference could be based on radiative transfer modelling (e.g., the D3 SST retrievals from coefficients acts as a reference) and/or in situ references (radiometers and reference buoys): exemplars for both approaches exist.

Beyond OE and BAOE, various approaches for progress in SST retrieval have been explored. These include alternatives to OE within the family of formal inverse theory, numerous complexifying formulations of coefficient equations and machine-learning-style extensions to regression. While it is like further progress could be made in those directions, none has yet provided a compelling reason to supersede coefficient-based retrievals and (BA)OE as a fruitful research direction. Moreover, implementing coefficient-based methods for SLSTR amounts to an adaptation of the current approach, and implementing OE-based methods is rendered relatively easy because the quantities to be simulated and calculated in OE are already available from the Bayesian clear-sky determination implementation. Therefore, this plan focuses on developments in coefficient retrieval and OE for SST.

A4.2. SST Coefficient Capability: Coefficient Generation Tool v2

Problem addressed: replicability and efficiency of full coefficient generation workflow.

The Coefficient Generation Tool v1 (deliverable SD2S-CGT) is illustrated in relation to the full coefficient generation process in Figure 6 below.

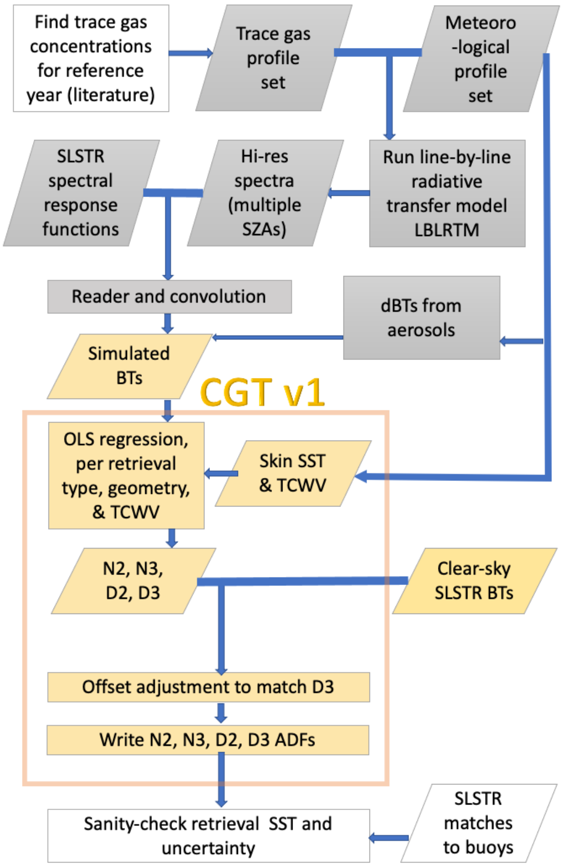

Figure 6 Flow chart and boundary of coefficient generation tool v1.

The CGT v1 delivers the core capability of coefficient and ADF generation given the supplied auxiliary brightness temperatures (in Figure 6, CGT v1 and auxiliary data are in gold). To expand the CGT capability to the full end-to-end process of generation, all the steps presently in grey can be included in the CGT as a complete workflow, constituting CGT v2. This step would help make efficient the further concepts for SST Coefficient Capability described below (A4.3 to A4.6).

T11 V2 Coefficient Generation Tool

Objective	Expand the scope of the CGT.
Required Research	None.
Required Algorithm Development	Formalise additional steps into CGT modules.
Comments on implementation	Offline tool.
Benefits beyond state of the art	Increased traceability and reproducibility for SLSTR coefficients.

A4.3. SST Coefficient Capability: Instrument Temperature

Problem addressed: the responses seen in routine SST (SSES) monitoring to SLSTR detector temperature changes are not well understood. An important contributing factor in the outcome is the spectral response function change as a function of temperature, which is constrained by laboratory characterisation only from the "warm" side of the operating temperature. This task researches whether non-linear spectral response function changes are adequate to explain the responses.

This work is to fit parameters for a solid-state detector model using the available laboratory data, and use this to give a physics-constrained extrapolation to operating temperature regime for the SLSTR-A SRFs. From these SRFs, detector-temperature dependent coefficients (using the CGT, but without inter-algorithm bias correction) will be derived. The SST perturbations for the different operating detector temperature combinations across channels can be derived and compared to the responses seen empirical by EUMETSAT in validation. We note that the current proposed detector temperature model is based on measured values which do not cover the lowest in-orbit operational temperature and so involve some level of extrapolation and while the use of a detector model is also an extrapolation it would be constrained by some level of physical understanding of how a HdCdTe detector works.

To fit the detector model would require access to both measurements of the SRF and different focal plane temperatures as well as corresponding pre-launch counts/temperatures etc. This is because to fitting a detector model will involve fitting certain parameters such as doping levels and photon flux scaling which together with the detector temperature will impact both the detector quantum efficiency (effectively changing the shape of the SRF) as well as changing the detectors non-linearity which can only be evaluated using the pre-launch data. One the detector model is optimised it can then be compared with the current detector temperature model and if as good would provide more confidence in extrapolating to the required lower detector temperatures.

More preliminary analysis of this potential task is required before a task table can be formulated; it is noted for completeness.

A4.4. SST Coefficient Capability: Prospective Trace Gas Trends

Problem addressed: operational coefficients are retrospective with respect to the changing trace gases in the atmosphere, but predictability of changes is sufficient to support coefficients that have future time evolution built in, to optimise SST stability operationally.

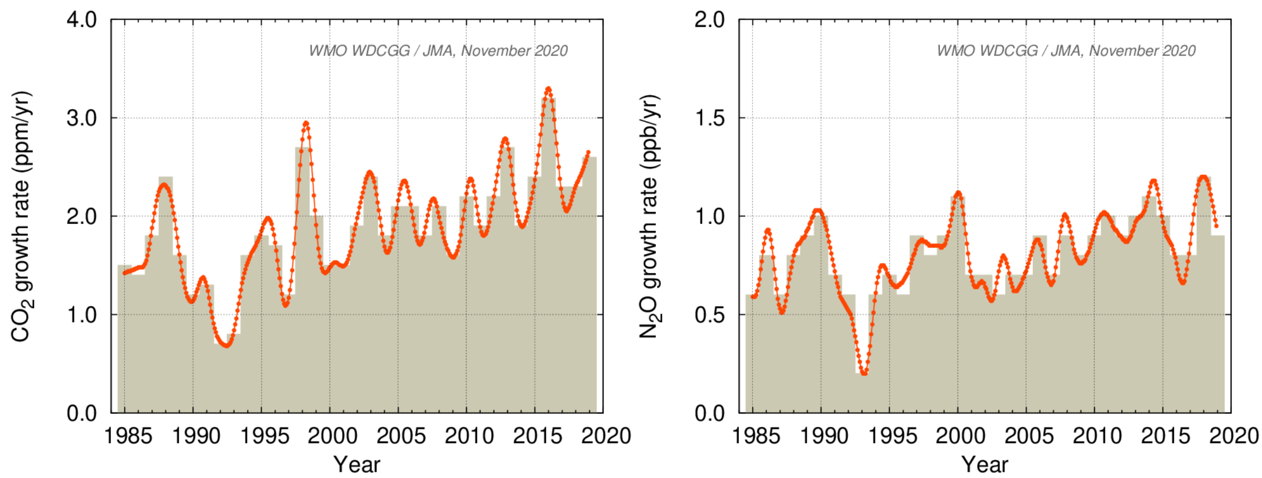

Figure 7 Trace gas growth rates as labelled, showing multi-annual predictability.

As these images from the WMO Greenhouse Gas Bulletin No.16, 2020 illustrate, growth rates of the main non-vapour radiatively active gases affecting SST retrieval can be relatively stable and therefore predictable on timescales of 5 - 10 years. Rather than fixing SLSTR SST coefficients with respect to recent trace gas concentrations, it is therefore feasible to make coefficients "future proof" on a mission timescale by projecting future trace gas concentrations using forecast growth rates. Trace gas increase at a fixed fractional growth rate is an exponential time increase, while trace gas perturbations of brightness temperatures tend to be logarithmic with respect to concentration. This means that linear interpolation of two sets of coefficients in time between "starting" and "future" conditions is adequate to achieve long-term SST stability compensation for changing trace gases.

T12 Future-proofing SST coefficients

Objective	Provide SST coefficients with a time-dimension accounting for extrapolated trace gas trends
Required Research	Extrapolate trace gas trends out to (say) 2030.
Required Algorithm Development	Run CGT v2 for trace gases at intervals (2018, 2022, 2026, 2030). Assume empirical offset adjustments persist. Define coefficient ADFs with time dimension added.
Comments on implementation	Implementation of time-interpolation in L2 SST.
Benefits beyond state of the art	Reduced trends from trace gas changes during mission life.

A4.5. SST Coefficient Capability: Desert Dust Aerosol

Problem addressed: in areas affected by significant desert dust aerosol, SLSTR SST retrievals are biased negative (to a degree that varies with retrieval type and interacts with the degree to which dust is screened in cloud detection).

Desert dust is the most significant aerosol for SST biases and this section will be written with respect to addressing dust specifically, although many comments could also be applied to other aerosol types in principle.

Dual-view retrieval embeds a significant amount of robustness from aerosol (i.e., relative bias insensitivity such that $\frac{\partial \hat{x}}{\partial M}$ is small, where M is column aerosol mass), because of the purely geometric (path-length) constraints provided by the dual-view observations. Nonetheless:

- Residual dust sensitivity is expected to be present in dual-view observations, at levels not presently quantified and plausibly significant for higher aerosol loadings.
- SLSTR also has a wider single-view swath whose SST properties hopefully match that of the dual-view swath as much as possible, despite single-view retrievals having no geometric-based robustness to aerosol.

One approach to reducing dust biases is to build aerosol robustness into the coefficients, as has been done with respect to stratospheric volcanic aerosols in dual-view coefficients, with highly effective outcomes notably for ATSR-1 [RD7]. To the degree that $\boldsymbol{k} = \partial \boldsymbol{y}/\partial M$ is known and constant, robustness can be achieved by constraining the coefficient regression with the condition $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{k} = 0$. The advantage of such an approach, if possible, is that it offers a prior-free, ADF-free solution, at the expense of increased error variance in the retrieval. However:

- For N2 SSTs this approach cannot work because the degrees of freedom for the coefficient regression are too small: loosely speaking, we need at least three BTs to deal with three geophysical variables (SST, atmospheric gas absorption and atmospheric aerosol impacts).
- For N3 SSTs this approach is highly dependent on the appropriateness and relative constancy of $\boldsymbol{k}$; but for desert dust aerosol, the pattern of BT impact across channels is not relatively invariant, but instead is sensitive to the aerosol height (and therefore temperature) as recently demonstrated by Luo et al. (2021) [RD8].

In summary:

- Dust robustness by $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{k} = 0$ with a generic $\boldsymbol{k}$ is not a convincing strategy for the wide, single-view swath of SLSTR SSTs.
- For dual view retrievals, SST sensitivity to dust is relatively low (good geometric robustness expected), but could be designed to approach zero more closely.

The proposed new approach is to calculate a dynamic estimate of $\boldsymbol{k} = \Delta\boldsymbol{y}/\Delta M$ using RTTOV (via some extra simulations compared to those used for clear-sky determination currently) and from this calculate an adjustment for each retrieval type (N2, N3, D2, D3). For a given pixel, the bias caused by aerosol is $\boldsymbol{a}^{\mathrm{T}}(\Delta\boldsymbol{y}/\Delta M)M$, where M is the mass loading of dust present in the simulation and assuming that $\Delta\boldsymbol{y}/\Delta M$ has been adequately simulated. (Note that the coefficients $\boldsymbol{a}$ differ from pixel to pixel because of changes in the look-up factors.) The concept is thus to adjust coefficient-retrieved SSTs by $-\boldsymbol{a}^{\mathrm{T}}(\Delta\boldsymbol{y}/\Delta M)M$.

An estimate for M could in principle be sourced from the Copernicus Atmospheric Monitoring Service forecasts (if the MET fields were extended to include relevant dust mass loads), or could be climatological (based on CAMS: an ADF applicable at L2 for this can be developed available based on work within SST CCI). As part of the investigation, the effectiveness of a climatological versus forecast/analysis estimate of dust can be examined.

Note that the uncertainty model, which currently neglects possible dust, will be expanded here to account for the correction uncertainty based on uncertainty in M.

T13 Desert Dust Aerosol SST Corrections

Objective	To reduce the sensitivity of SLSTR SSTs to the presence of desert dust aerosol.
Required Research	Off-line demonstration of correction approach, including relative benefits of forecast/analysis cf. climatological knowledge of M. Determine how to estimate the uncertainty of the corrections, such that the uncertainty model for desert-dust affected observations is also improved.
Required Algorithm Development	Addition of desert dust mass loading to processing. Additional RTTOV runs (with and without dust) for BT impact. Calculation of correction per retrieval-type and pixel, and uncertainty contribution.
Comments on implementation	Implementation with forecast desert dust distributions would require significant addition to the MET fields in L1. Implementation with climatological distributions can be done at L2.
Benefits beyond state of the art	Improved SLSTR performance as a reference sensor in the areas affected significantly by desert dust (Mediterranean, Red, and Arabian Seas, and north-east subtropical Atlantic). More realistic uncertainty estimation in desert-dust areas (presently the uncertainty impact of desert dust is neglected).

A4.6. SST Coefficient Capability: Day-night Three-Channel

Problem addressed: three-channel retrieval is better than two-channel, with markedly lower large-scale biases and smaller uncertainty; despite previous experience showing utility in using the 3.7 µm channel in day-time for SST, away from specular angles, this has never been investigated for SLSTR.

In situations of potential scattering/reflectance of solar irradiance at 3.7 µm, use of the channel for SST is not plausible, because the strength of scattering will be highly uncertainty in general. Thus, for quasi-specular angles (a large halo around the sunglint maximum), the 3.7 µm (of the affected view) should not be considered in the SST retrieval. The geometries and aerosol types for which 3.7 µm scattering by aerosol needs to be avoided in SST retrieval are not systematically quantified to our knowledge, and would be researched in this context.

However, for other situations, there is the possibility that the superior retrieval performance of night-time three-channel retrievals could be approached for low-scattering day-time scenes. This opinion is based on the retrieval of SST using 3.7 and 11 µm channels for the GOES-12+ missions, which had no 12 µm channel [RD9]. In that work, the solar-scattering in the (as on GOES) 3.9 µm channel was heavily parameterised.

Here, it is assumed that the new capabilities of RTTOV enable useful situation-specific estimates of the solar-scattering impact on the 3.7 µm channel, something that the research will explore and quantify, in terms of the uncertainty in correction achievable.

Another (or an additional) possibility is to constrain 3.7 µm scattering by the reflectance observed in SLSTR's SWIR channels (S4 to S6). Presently, this is a speculative possibility.

For consistency with existing retrieval coefficients, and ease of implementation (minimising additional coefficient sets), an appealing approach is to "blend" between two-channel and three-channel retrieval coefficients as a function of 3.7 µm solar-scattering (i.e., the estimated BT effect), and apply the coefficients to the BTs after adjustment of the 3.7 µm BTs to an equivalent night-time value. The appropriate blending function would depend on the uncertainty with which the 3.7 µm BTs could be corrected for solar scattering, and would be defined to minimise the retrieval uncertainty.

Approaching sunglint or other high-scattering situations, the SST retrievals would consist of the two-channel retrievals currently used for all day-time pixels. Under more favourable conditions, the retrieval uncertainty for day-time SSTs would be reduced towards that obtained for night-time scenes.

T14 Daytime use of 3.7 µm coefficients

Objective	To reduce the uncertainty of day-time SLSTR SSTs under low-solar-scattering conditions, by adjusting the 3.7 µm channel using the lower-error night-time coefficients.
Required Research	Investigate predictability of solar scattering impact on 3.7 µm BTs using: • additional RTTOV simulations, • SLSTR S4 to S6 reflectance, • a combination of the above. Develop adjustment-to-night-time-BT method for 3.7 µm channel, including uncertainty model Explore performance of night-time SST coefficients in day-time as a function of adjustment uncertainty Devise optimum blending of day and night SST coefficients for reducing day-time SST biases and uncertainty
Required Algorithm Development	Algorithms for all steps indicated in "Required Research".
Comments on implementation	Implementation at L2. Likely some additional RTTOV simulations and small ADFs required.
Benefits beyond state of the art	Reduce day-time SST errors towards those obtained for night-time scenes: a very significant SST development.

A4.7. Optimal Estimation: Bias Corrections and Covariances

Problem addressed: biases and uncertainty in SST for N2 retrievals are very likely to be reduced using optimal estimation with bias-aware methods (BAOE) that optimise consistency with other retrieval types. (It is not known *a priori* whether N3, D2 and/or D3 SST retrievals could also be improved compared to coefficient-based methods.)

The classic split-window channels in single view do not in reality contain enough information for high-sensitivity SST retrieval in tropical regions [RD10], particularly in off-nadir views. In these circumstances, all published retrievals (whether explicitly or implicitly) rely heavily on prior information, and are prone to bias whenever the prior SST departs from the observed SST.

Minimising this effect involves making the prior information as informative for a particular retrieval as possible, and optimal estimation [RD11] is the principal theory that systematically enables prior information to be exploited, both as prior and as a linearisation point for retrieval. The overhead in introducing OE in the SLSTR processor is small, given that the necessary NWP flows and simulations are used in Bayesian cloud detection. Therefore, OE should be explored for SLSTR operational SSTs, with a view to using it for N2 at least (where the biggest benefit will be clear), and of assessing the potential for other retrieval types (N3, D2 and D3).

The limitations of OE are: biases between necessarily fast radiative transfer (RTTOV) and the satellite calibration are small but large enough to cause out-of-specification SST biases (>0.1 K); error covariance matrices are required that were often constructed on an informed but ad hoc basis. Recently, bias-aware optimal estimation methods have been developed that address these limitations [RD5], [RD6].

This task is therefore to apply BAOE to SLSTR for all channel combinations, assessing performance relative to the corresponding coefficient-based retrievals. The OE formulation will be the current standard formulation of a reduced state vector of SST and total column water vapour (TCWV), and a one-pixel-at-a-time BT-only observation vector. {Aside: EUMETSAT may wish to consider TCWV over the dual-view swath as a potential delivered product.}

The reference SST for the bias adjustments will be the D3 SST (skin temperature), and the objective is to 'calibrate' OE SSTs against this reference. Adjustments for BTs in all channels will be inferred as an additive piecewise-linear function of plausible geophysical and instrument parameters, such as TCWV and zenith angle. Because the D3 SST doesn't cover all nadir-swath angles, a strategy for interpolating and extrapolating will need to be devised. Additionally, a prior-bias adjustment may be obtained, along with error covariance matrices (which may also have piecewise linear dependencies).

T15 Baseline bias aware optimal estimation

Objective	To reduce the biases and uncertainty of the N2 SST (and potentially of other retrieval types, TBD) by use of bias-aware optimal estimation, referred to D3 coefficient-based skin SSTs.
Required Research	Apply BAOE to SLSTR, and obtain bias corrections and covariance matrices. Research the set of geophysical and instrument predictors most successful in reducing retrieval biases. Look at temporal stability of the BAOE solution, and consider implications for update (if necessary) of bias corrections as instruments age in orbit, and with respect to events such as outgassing, changes in instrument temperature, etc.
Required Algorithm Development	Algorithm for optimal estimation using reduced (two-element) state vector and between 2 and 6 BTs as observation vector (exists: just adaptation to SLSTR). Extend to include the bias-aware aspect, in which: • models for instrument-RT biases are proposed in terms of factors related to SST errors, including instrument temperature, • BT bias-correction parameters and retrieval error covariance matrices are derived using matches to reference measurements,

	• these parameters are then used within the OE to generated reduced-bias SST retrievals.
Comments on implementation	Implementation of OE is at L2. ADFs for bias correction parameters and covariance matrices will be required, which are derived off-line (using matches). Note: in principle a real-time, updating implementation of BAOE is possible for NRT. This goes beyond what is described here, undertaken offline. An updating implementation would require organisational development of software and data streams.
Benefits beyond state of the art	Reduce day-time N2 SST errors, retaining consistency with coefficient-based SSTs. Depending on performance, OE may be considered for all channel combinations.

A4.8. Optimal Estimation: Extended Vector Formulations

Problem addressed: "standard" formulation of OE for SST doesn't account for desert dust aerosol.

The standard OE formulation has a retrieved state vector of $[x, w]^{\mathrm{T}}$ – i.e., SST and TCWV are retrieved. A full state vector (surface temperature and atmospheric profiles) is used for the RTTOV simulations within OE, and these simulations can account for SST-significant aerosol, specifically desert dust. Without dust in the retrieved state vector, the prior dust information is effectively assumed to be correct, whereas of course there is uncertainty in that prior (quite large uncertainty if the prior is the climatological dust loading, since dust is variable on short space and time scales).

Adding the column mass of desert dust as a retrieved state vector quantity, $[x, w, m]^{\mathrm{T}}$, should reduce dust-related SST biases if there is observational information to constrain the dust mass. For SLSTR, thermal channels in dual view (D2 and D3) may provide a useful level of constraint (TBC). Moreover, for daytime retrieval, channels S4 – S6 may provide additional constraint in addition to the thermal channels. For this reason, an extended observation vector to include reflectance channels is also considered in this task.

Note that using a column mass of desert dust involves simplifying assumptions similar to the use of only TCWV to represent atmospheric humidity and temperature profiles. The BT impact of desert dust does vary with the dust particle size distribution and the vertical dust mass distribution, neither of which is captured in m. As with TCWV, only the leading order dependence is retrieved, because of the limited information content in the SLSTR channels. The particle size and vertical dust distributions need to be specified for simulation purposes, and for SLSTR operations, a climatological ADF for these seems most plausible. However, the Copernicus Atmospheric Monitoring Service does forecast desert dust (see Figure 8 below.)

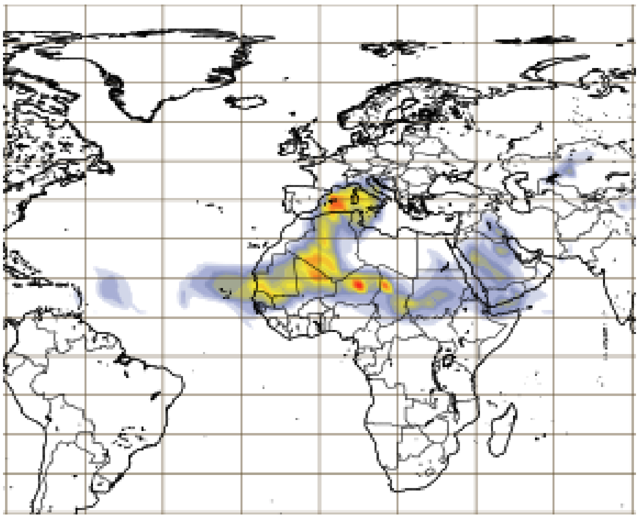

Figure 8 Forecast of desert dust mass from CAMS, for 00z on 16 September 2021 with 2-day lead time. Levels of dust mass significant for SST retrieval are seen off West Africa and in the western Mediterranean.

The forecast dust mass (although not all distributions) may be a feasible addition to MET-TX information for SLSTR, which would reduce the prior dust uncertainty relative to relying solely on climatology.

Two notes: (1) This task is complementary to T13, addressing the SLSTR desert dust bias problem for coefficient retrievals. It is unfortunately not predictable which of these two approaches will be more fruitful in addressing the problem. (2) Desert dust can be significantly screened by "cloud" detection, but it is preferable not to do so, but rather to have retrieval methods that cope with dust, in order to maximise SST coverage.

T16 Extended vector optimal estimation

Objectives	To reduce the biases and uncertainty in SST retrieved under desert dust conditions, by extending optimal estimation methods to estimate dust by inclusion of dust in the retrieved state vector and including dust-constraining observations where possible.
Required Research	Determine the SST reference for the development, most likely drifting buoys. Unlike non-dusty situations, D3 SST skin retrievals do not (presently) have a proven bias performance when desert-dust aerosol is present. Assess dusty-sky simulations using climatological and CAMS forecast dust masses, and climatological dust distributions. Apply BAOE to obtain prior or simulation corrections, in a variety of configurations (thermal only, thermal and reflectance, reflectance only; single and dual view). Design extended OE on basis of results. Test and validate retrieval using forecast and climatological dust mass to determine magnitude of benefit from adding dust to MET-TX information.
Required Algorithm Development	Algorithms for using dust prior information and ADFs at image resolution. Extended optimal estimator (with uncertainty estimate).
Comments on implementation	Purely climatological implementation of dust-mass prior is possible at L2, with addition of significant ADFs. Implementation with CAMS forecast dust mass operationally would be to add this variable to MET-TX information at L1.
Benefits beyond state of the art	Reference quality SSTs from SLSTR also for desert-dust conditions.

A4.9. 2D SST Field Optimal Estimation

Problem addressed: (1) the use of oblique and nadir BTs via the image grid for dual-view SST retrievals is simplistic, because of the complexity of the real pixel positions, leading to distortion of SST frontal intensities and front positions; (2) one-at-a-time SST retrieval neglects the fact that the TCWV (and dust) fields should vary smoothly on pixel scales, and omitting this constraint misses the opportunity for suppression of SST noise.

Figure 4 illustrates the disparity in area of an oblique versus forward pixel. Despite this, after mapping by nearest neighbour methods onto a common "image grid", oblique and nadir pixels are used in dual-view SST retrieval as if they represent a common area. Over many low-gradient regions, this is of little consequence, but in dynamic areas, this "blurs" dual-view retrievals of SST fronts such that errors are introduced into any derived gradient estimates and frontal locations may be distorted at the level of 1 pixel. Moreover, the image-grid step makes duplicate use of some BTs, while consigning other BTs to "orphan" pixels that are then "wasted" from the point of view of retrieving SST.

The proper and general solution for complex viewing geometry with differently sized fields of view should be to estimate the two-dimensional SST field (2DF) given the actual locations and spatial extent of all observations. There is no fundamental need to act as if oblique and nadir pixels have the same field of view, nor is there any need to over-use (duplicate) or exclude (orphan) some observations. These practices have merely emerged as relatively simple to conceive and implement.

Other than SLSTR, these comments apply to MODIS (bowtie effect region) and all microwave radiometers for SST (and indeed for other variables). So, a general solution this problem could be of wide applicability.

In a 2DF solution, the spatial coherence of different retrieval variables can also be appropriately constrained. For example, in the case of SST retrieval with two retrieved variables (SST, TCWV), short-scale (1 km) variations of SST are appropriate because of the possibility of SST fronts, whereas clear-sky TCWV varies smoothly on such scales. When one-at-a-time retrieval is performed, the TCWV is not smooth, but noisy, and there is complementary noise in the SST field. This is an indication that imposing smoothness on the retrieved TCWV field will reduce noise in SST, which has been confirmed by studies using extended optimal estimation techniques [RD12] using multiple pixels within the observation vector. The ability to constrain the smoothness of TCWV in a controlled way in a 2DF solution is a reason for applying it also when observation geometries are simpler (such as in single view retrieval).

The target grid for the 2DF solution can be chosen for convenience of the SST product to be obtained: essentially the swath geometry can be by-passed to estimate the 2D SST field on (in the case of SLSTR, say) a 0.01° grid, or an equal area grid of choice, as shown in Figure 9.

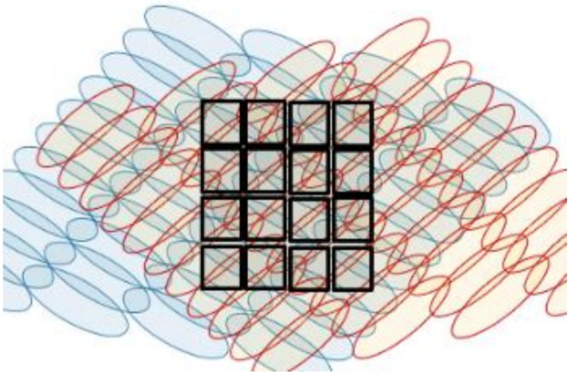

Figure 9 Concept of directly retrieving a field of (a) geophysical variable(s) (square black grid) directly from the observation values and positions (fields of view) of overlapping, non-coregistered observations (ellipses).

This is a more complex formulation of retrieval, but it has the following advantages:

- It avoids injection, in the intermediate step, of regridding errors that are then propagated to the retrieved product.
- It provides an optimal use of any over-sampling to maximise the achieved spatial resolution of retrieval.

In the case of retrieval of geophysical features known to vary with differing length scales, there are further advantages to a direct 2D approach which may be even more significant:

- The geophysical scales can be utilised in the retrieval to reduce the retrieval ambiguity (compared to 1D one-pixel-at-a-time) and drive lower noise in the result.
- Spatial geophysical correlations known to exist between variables can similarly be utilised to constrain retrieved answers, reducing errors.

Although implemented only crudely (not accounting for real observation geometries) this latter insight is already applied to sea surface temperature retrieval using optimal estimation, in which SST and total column water vapour, W, are jointly retrieved. SST can vary at fronts on distances of ~1 km while W varies spatially at least an order of magnitude more slowly. By retrieving SST over an area with a common W, noise amplification in the inverse is significantly reduced.

Geospatial statistical methods used in the EU project EUSTACE [RD13] for estimation of global surface air temperature from a mix of different observation types give a basis for 2D retrieval methods. These methods avoid brute-force inversion of large matrices, are computationally highly efficient, can be iterated, account for spatial correlation in the target variable(s), and account for a “footprint” of influence of each observation on the target field – which is how the FOV of each satellite observation could be accounted for.

The research and development required is to combine these geospatial estimation capabilities with the SLSTR multi-channel OE retrieval of SST and TCWV (and potentially dust mass). Note: this would necessarily involve collaboration with the developer of the EUSTACE techniques, Finn Lindgren at University of Edinburgh.

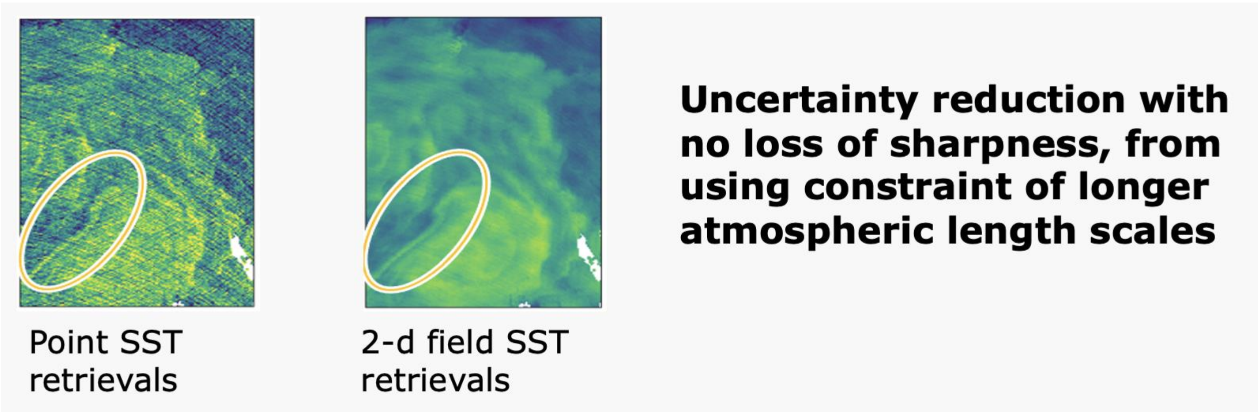

Figure 10 Effectiveness of smooth-atmosphere constraints in reducing uncertainty for ATSR-1 (from SST CCI).

T17 2d SST field estimation

Objective	To reduce geometry-induced errors and reduce retrieval noise in by solving the 2D SST field, rather than one-at-a-time retrieval.
Required Research	Formalise the combination of geospatial and spectral inversion into a 2D retrieval in a general mathematic formulation (“algorithm”), including accounting for geophysical, spatial and error correlation constraints Define the interfaces needed to present satellite, forward model and prior data to this algorithm (e.g., models for FOV, interface to radiative transfer simulator, formulation of constraints) Extend existing open-source software libraries to implement algorithm, testing on synthetic data with known truth (via interface formats) Add complexity to synthetic data to challenge algorithm, such as cloud coverage to break up 2D field, desert dust. Application to real SLSTR data, retrieving SST, TCWV and possibly dust mass fields. Compare results with 1D retrieval. NB: Formal approach above ensures adaptability to other retrieval situations and sensors.
Required Algorithm Development	See above. A combination of satellite retrieval and geostatistical expertise with strong theoretical and practical background will be required.
Comments on implementation	L2 implementation, but radically changing the concept of the “L2” product (not on SLSTR image geometry). Solution may be operationally slower than 1D retrieval (TBC), so processing speed needs to be considered.
Benefits beyond state of the art	Lowest-noise, most spatially faithful SSTs of which it is presently possible to conceive.

A5. Development Tasks: Quality Flagging

A5.1. Quality level definitions: review of practice

Problem addressed: need to review quality-level definitions with respect to scientific coherence and fitness-for-purpose in the eyes of users.

SLSTR quality indicators follow GHRSST practice, but GHRSST practice is broadly defined and does not enforce coherent and optimal practices. The rules for quality level indicators in SST products from SLSTR have developed incrementally over time, and there is merit in a reflective review of the QL practice from two points of view:

- Are the current QLs as scientifically coherent and effective as they could be, and if not, how could this be improved?
- How are users interpreting and using present QLs, and does that suggests improvements?

The relevance to other SSTs within the OSI-SAF could also be considered within this task.

T18 Review quality level definitions and practice

Objective	To systematically appraise the scientific coherence and user relevance of SLSTR quality indicators.
Required Research	User survey in relation to their interpretation, use and expectation of quality indicators in SLSTR SST products. Review of practice and identification of options for improvement.
Required Algorithm Development	Would follow after considering review outcomes.
Comments on implementation	Would follow after considering review outcomes.
Benefits beyond state of the art	Better alignment of quality indicators and user expectations, if possible.

A5.2. Optimal Estimation Goodness-of-Fit

Problem addressed: use of additional OE diagnostics in context of quality levels of products.

In optimal estimation, a number of goodness-of-fit indicators can be calculated. Previous work has shown such "χ^2" metrics are effective discriminators of bias and uncertainty against independent validation data [RD6]. Here is an example:

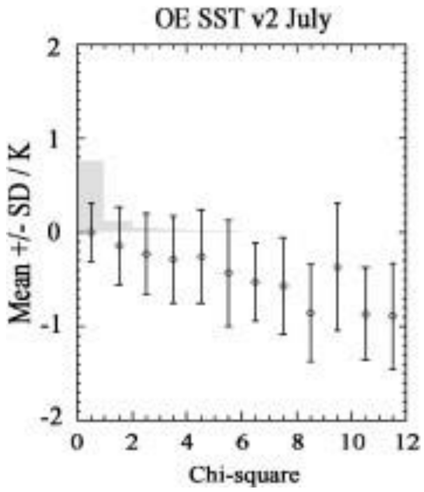

Figure 11 Validation statistics (mean and SD) against a χ^2 metric, showing the latter to be strongly predictive of bias and additional retrieval uncertainty.

Quality indicators have no precise objective and definition under GHRSST conventions, but it has been argued [RD14] that they should indicate a degree of confidence that the uncertainty of a retrieval has been well estimated (rather than merely presenting simplified uncertainty information). Uncertainty estimates are provided in SLSTR SST products, and are available from OE retrievals; these are internal to the OE and assume valid retrieval conditions. Discrepancies between the OE uncertainty and the error statistics one might impute given the χ^2-validation characterisation may indicate circumstances in which the OE uncertainty estimate is invalidated (because the circumstances of the retrieval are not as the OE assumes, e.g., there is some residual cloud contamination).

SLSTR SSTs may continue to be obtained by coefficient-based methods. However, the overhead for a simple OE is small, given that Bayesian clear-sky determination is implemented, so it could be explored whether high-χ^2 is also an indicator of larger-than-modelled error statistics when using coefficients for retrieval. OE could have a role in setting quality levels even if not the primary SST retrieval method.

T19 Exploitation of OE goodness of fit

Objective	**To determine whether χ^2 measures are useful indicators of quality level for either OE or coefficient-based SLSTR retrievals.**
Required Research	Explore validation statistics for SLSTR SSTs as a function of χ^2, in comparison to statistics expected given the uncertainty model. Devise method for exploiting χ^2, if shown useful, for quality level setting.
Required Algorithm Development	Run OE (irrespective of whether main SST retrieval is done by coefficients) and convert χ^2 into a factor in setting the quality level.
Comments on implementation	L2. Running OE is a small overhead, given Bayesian clear-sky determination has already run.
Benefits beyond state of the art	Improved quality information in SLSTR SST products.

A6. Development Tasks: Uncertainty Evaluation

A6.1. L1 Uncertainty Information and L2 Uncertainty Model

Problem addressed: uncertainty estimates for SLSTR SSTs are not validating as well as it is considered they should.

Note: the operational coefficient ADFs are known to include approximated specification of the retrieval uncertainty. This will be addressed via the CGT, and is not the problem addressed here.

This task is to undertake a systematic review of the SLSTR uncertainty model and validation method, using metrological approaches within the limits of a study undertaken at L1 and L2. This will involve: review of the intended definition and actual calculation of L1 uncertainty information (for thermal and reflectance channels, since reflectance channels may be used in some retrievals); review of the methods of propagation of L1 uncertainty; review of the uncertainty model for inversion of perfect data using coefficients (after correction of the problem noted above); review of the "cloud proximity" uncertainty contribution, which could use tandem-phase comparisons to better quantify this effect; identification and estimation of effects not included within the uncertainty model; review of the combination of uncertainty information and the allocation of uncertainty components to spatio-temporal scales; and review of the uncertainty attributed to matched in situ data in the process of uncertainty validation.

Having reviewed and modified the uncertainty model (and uncertainty validation model) in the light of the review outcomes, the updated uncertainty model will be validated, to quantify the hoped-for improvements in uncertainty validation.

T20 Uncertainty model review and improvement

Objective	**To understand why SLSTR SST uncertainties do not validate as expected, and rectify any detected problems.**
Required Research	Review definition of L1 uncertainty information and assess whether it appears to be quantitatively correct. Review L2 uncertainty model (L1 uncertainty propagation and combination with perfect-data inversion uncertainty). Review the assumptions made in uncertainty validation about uncertainty in the matched reference SSTs. Correct any uncovered problems and look for improvements in the uncertainty validation arising.
Required Algorithm Development	Potentially changes at L2 in uncertainty model.
Comments on implementation	L2 implementation of any changes in SST uncertainty model. If L1 uncertainty is questioned, this will raise an implementation issue at a lower level. Implementation of changes to validation comparison is "offline".
Benefits beyond state of the art	Improved uncertainty model and/or validation thereof.

A7. Skin-depth Difference Modelling ("DV")

SLSTR measures skin SST by physics-based retrievals, and these retrievals are validated against in situ measurements at various depths: radiometric (skin) measurements, drifting buoy SSTs and deeper moorings or Argo measurements.

To assess the skin measurements and their uncertainty in validation, the true geophysical variability between the point in situ measurement and the areal satellite measurement should be taken account of. While the point-to-area aspects contribute "noise", the depth-related aspects are systematic, and require knowledge of the stratification of the near-surface ocean, both from cool skin and diurnal warm-layer effects.

Within SST CCI, a legacy skin-depth model from the Met Office remains in use, but is unsupported, does not reflect more recent science, and is not considered maintainable. Therefore, this task to assess options for a DV model relevant to SLSTR SST product development, requiring that it be:

- maintainable,
- scientifically current,
- able to be run from NWP inputs (operational and re-analysis),
- able to be validated/tuned for consistency with various observations.

The best candidate identified is the NOAA Diurnal Model code base, which is a modified version of the code running operationally at NOAA to correct the POES-GOES blended Level 4. The core physical basis is the physics of Kantha & Clayson (1994) [RD16], which is the basis of the successfully applied "black-box" model in SST CCI, but more recent insights from (unpublished) work of Gary Wick has been incorporated. Moreover, work by Mittaz at Reading several years ago made progress on debugging some behaviours of the original NOAA version of the code so that it is advantageous to start from this updated version.

The worked reported here, therefore, looks at several observed diurnal warming events, the model's performance, and discuss possible options for its improvement by tuning the model for use with the SLSTR sea surface temperature (SST) products.

A7.1. The NOAA DW model

The NOAA Diurnal Warming (DW) model is a one dimensional second order turbulent closure model designed to model the water column under the influence of solar heating of the surface layers. The original version of the model is described in [RD16] and the current version of the code also includes the effect of Stokes dissipation of wave energy (see, for example, Kantha et al. 2009, [RD17]). The model also includes the Fairall skin model (Fairall et al. 1996) and can use different water absorption models. The code is based on source code provided by Gary Wick from the NOAA Physical Sciences Laboratory which was all rewritten in Fortran 95 using double precision arithmetic by Jon Mittaz (when at NOAA) and has been extensively code-reviewed and debugged as part of an earlier project at Reading.

A7.2. The parameter file

To run the model a number of parameters need to be defined, some of which were determined by laboratory and/or field campaigns and can be considered fixed and some of which can be changed. All the code settings plus the parameters themselves are currently defined in a single parameter file which has the form shown in Table 1. Note, some of the terms are for testing purposes.

Table 1 List of parameters found in the configuration file

Parameter Name		Default value
file_turb_cnts	=	turb_constants_kantha.txt *Input turbulent constants from a file*
use_ascii_input	=	n *There are multiple input formats for the model/flux data. The original code used an ascii input format (used if use_ascii_input=y) whereas the operational NOAA code used a netCDF format and other analysis can use a netCDF MMD format (use_ascii_input=n).*
grid_file_stem	=	_sst.dat *If ASCII input format is selected it needs two separate files. This file stem contains a single line per location with position/initial SST/grid location etc. information*
flux_file_stem	=	_flx.dat *If ASCII input format is selected this file stem is for the file containing the time series of fluxes etc. used to create the time series of SST for each location defined in the _sst.dat (grid_file_stem parameter) file.*
output_file	=	sens.000 *Output file name*
test_output	=	n *Gary Wicks reduced output format (ASCII only)*
output_file_netcdf_format	=	n *Output the model in the NOAA operational netCDF format*
output_file_cci_format	=	y *Output the model in a format related to the CCI MDB input format*
output_ascii_profiles	=	n *Output the model data in Gary Wicks ASCII full format*
output_file_profile	=	profile.dat *Output depth profile data for a range of parameters. Note, these files can be very large if care is not taken regarding how many times/locations are to be output as files are in ASCII format.*
ngrid	=	1 *Number of grid points to be processed. Can be updated internally dependent on which input file is read in as can the following grid parameters.*
file_grid_nx	=	720 *Output grid size (longitude)*
file_grid_ny	=	240 *Output grid size (latitude)*
file_grid_min_lon	=	0.25 *Minimum longitude of grid*
file_grid_max_lon	=	359.75 *Maximum longitude of grid*

file_grid_min_lat	=	-59.75 *Minimum latitude of grid*
file_grid_max_lat	=	59.75 *Maximum latitude of grid*
days	=	2.5 *Numbers of days to generate a DW time series. Note that this needs to include spin up time for the model.*
print_freq_hours	=	1 *Frequency in hours to print out certain information to stdout*
nlevels	=	140 *Number of depth levels*
dt1	=	300 *Time step between DW estimates in seconds*
smooth	=	0.1 *Smoothing parameter which smooths backward/current values to help with algorithmic stability.*
bottom_depth	=	50 *Depth of the bottom in meters*
reference_depth	=	0.2 1.0 1.5 10. *Depths for which the model will output temperatures in meters. Note that there is a possible maximum number of reference depths of 10 (N_REFERENCE_DEPTH is defined in Types.f90).*
bottom_salinity	=	35.2 *Value of the salinity at the bottom boundary layer.*
solid_bottom_temp	=	280 *Temperature at the bottom depth (assumed solid)*
KHB_init	=	1E-5 *Kinematic diffusivity for the average flow*
KMB_init	=	1E-5 *Mean eddy diffusivity*
FCT	=	0.75 *Scaling factor between previous and updated diffusivity components for updated values*
shear_enhancement_factor	=	1.0 *Enhancement factor used to scale the square of the velocity gradients used in calculating the Richardson number/shear production values*
langmuir_top_boundary_factor	=	1.0 *Scaling factor for adding the square of the langmuir velocity (Stokes drift term) to the VHP term used to update the kinetic energy term*
wave_enhancement		10.

		Scaling factor for extra term for the surface VHP term if wave breaking is turned on
surface_abs_enhancement	=	0.1 *Scaling of net solar to the total flux if surface absorption is allowed*
use_skin_temperature	=	y Use input skin temperature rather than the top of the temperature profile.
use_solid_bottom	=	n *If turned on uses a fixed temperature for the lowest profile temperature (see solid_bottom_temp parameter for the bottom temperature used)*
use_estimated_wavedata	=	y *Will force the use of the stokes drift parameterization. Does not currently force the use of estimated wavedata itself as that decision involves the presence of input wavedata only so parameter may be redundant.*
wave_breaking	=	y *Adds in wave breaking (see also wave_enhancement)*
use_langmuir	=	y *Add stoke drift components to model*
fairall_skin_effect	=	y *Add in the Fairall et al. (1996)* [RD15] *skin model*
surface_abs	=	n *Add in extra radiation at surface related to the net_solar parameter (see surface_abs_enhancement)*
shear_flag	=	y *Add shear to diffusivity terms*
modify_fluxes	=	y *Modifies the surface fluxes based on any diurnal warming from the original model state*
use_evaporation	=	n *Add in the effect of evaporation on the salinity flux term*
exponential_damping	=	n *Damping of the surface parameters*
original_solar_absorption	=	y Use the original water absorption coefficients rather than a version which includes satellite view angle/TCWV terms.
use_mmd_format	=	n Use the CCI MMD data as input to the model

As can be seen from Table 1 there are a range of different options that either change a value or switch an option, both of which will change the behaviour of the model. Changes to the values stored in the parameter file can be made either by changing the parameter file itself or by using command line options because every parameter listed in Table 1 can be changed using a command line option, for example:

./kantha_clayson_dwmodel.exe default.cfg $MODEL_DIR $DATE_STRING use_mmd_format=y

where default.cfg is the parameter file described above, $MODEL_DIR is the location of the NWP model data, $DATE_STRING is the start date in a form YYYY-MM-DD and use_mmd_format=y is modifying the value stored in the parameter file. Any of the parameters can be changed using the form param=value.

Note that many of the parameters in the parameter file are related to input/output data information so do not need to be changed. Also, constants stored in the **file_turb_cnts** file are based on laboratory and detailed ocean observations/field campaigns so will not be modified as part of this study. The parameters that will be studied here are listed in Table 2, though there are other further parameters that could also be modified such as those for the bottom boundary conditions which could be important for coastal regions. Here we consider the open ocean case.

Table 2 Parameters that will be modified to look for changes in the DW model output in this study

Modified Parameter	Description
KHB_init	Mean (background) kinematic diffusivity
KMB_init	Mean (background) eddy diffusivity
FCT	Scaling factor between previous and updated diffusivity components
shear_enhancement_factor	Enhancement factor in calculating the Richardson number/shear production values
langmuir_top_boundary_factor	Scaling factor for the Stokes drift term used to update the kinetic energy term
wave_enhancement	Scaling factor for extra term for the surface VHP term
surface_abs_enhancement	Scaling of the net solar contribution to the total flux if surface absorption is allowed

A7.3. Test datasets

To understand the impact of changing some of the parameters listed above on the diurnal warming estimates we have first looked at some of the matchup files created by Chris Old when he did a limited project looking at the possibility of tuning the DW model to an AATSR-to-in situ matchup dataset. The extant matchup datasets consist of ECA CCI formatted data with AATSR observations matched to in-situ measurements together with associated meteorological data (fluxes etc.) taken from the ERA-Interim reanalysis.

To select a small number of locations and times for further study we have looked at positions that satisfy the following criteria:

1. cases where the individual matchups in the AATSR MMD have a GHRSST quality flag of five (excellent data),
2. cases which are within ±30° of the SEVIRI subsatellite point (effectively 0° latitude, 0° longitude) to make sure the SEVIRI data are at relatively low viewing angles,
3. cases where over a three-day period the matchup location has clear skies based on a time series of the SEVIRI GHRSST quality level.
4. cases where the model runs indicate a minimum diurnal warming signal of at least 0.5 K.

This yields 11 locations and times (see Table 3) which, despite their small number, do provide enough information to begin to understand how to tune the model. In reality of course, there will be many more locations where significant diurnal warming can be seen: they can be found by looking more globally and/or refining the selection criteria. Therefore, the selection of locations presented here is only meant as a quick look at the behaviour of the NOAA diurnal model in a limited location rather than a complete analysis.

Table 3 List of locations where diurnal warming is present

Matchup Id	Start Time	Latitude	Longitude
469816829	2004-06-27T06:00:00	35.099907	19.701702
469873388	2005-06-24T06:00:00	36.599327	12.799703
469873405	2005-06-24T06:00:00	36.39959	13.100175
469873430	2005-06-24T06:00:00	36.69931	12.79940
471962715	2004-07-06T06:00:00	38.99979	18.30161
471976757	2004-07-28T18:00:00	38.799824	26.100056
471986505	2004-07-08T18:00:00	36.299644	16.600061
471986516	2004-07-08T18:00:00	36.40005	19.299421
471986522	2004-07-08T18:00:00	36.30017	19.49958
479386216	2011-07-07T06:00:00	40.319862	17.270628
479765952	2011-05-14T06:00:00	35.788223	25.890598

A7.4. Examples of SST variations from SEVIRI data

As discussed above, we have looked at locations that have both an AATSR-to-in situ matchup as well as SEVIRI SST data. The SEVIRI SST data used is that from the EUMETSAT OSI-SAF SST archive found at http://tds0.ifremer.fr/thredds/catalog/OSI-206-series/${yr}/${dy}/${yr}${mn}${day}${hr}0000-OSISAF-L3C_GHRSST-SSTsubskin-SEVIRI_SST-ssteqc_meteosat08_${yr}${mn}${day}_${hr}0000-v02.0-fv01.0.nc, as there is good coverage with respect to the AATSR-to-in situ MMD (2002-2012), and the data is available in its original L2P form which has a time resolution of 15 minutes. We do note that the SSTs retrieved by SEVIRI may have a reduced sensitivity to the real SST variations which needs to be borne in mind when comparing the data to the DW model. Figure 12 shows the SST time series for one of the locations where a diurnal variation of close to 3 K can be seen. Further examples are shown below.

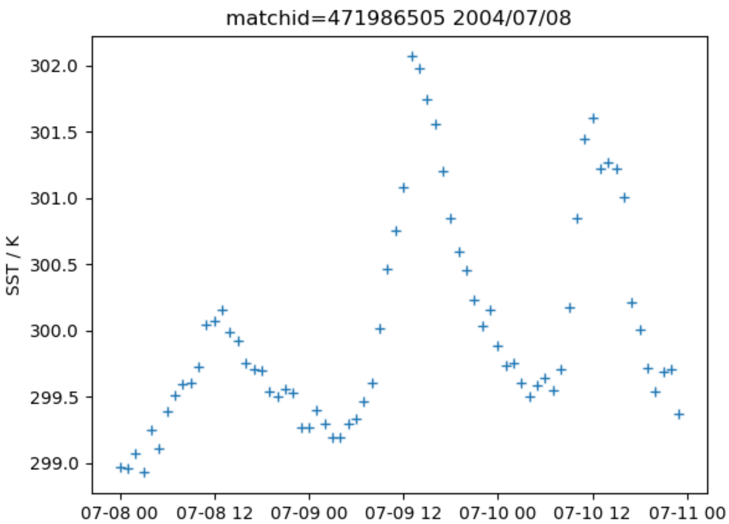

Figure 12 Time series of OSI-SAF SEVIRI SST data at the location of the AATSR-to-in situ matchup id=471986505 showing evidence of reasonably strong diurnal signals.

A7.5. Comparison with the default diurnal model

Figure 13 shows all the SST time series we have looked at together with the diurnal model (orange line) for each case. Note that the diurnal model has been shifted to approximately match the SEVIRI data and is using the default parameters shown in Table 1 only. Also shown in Figure 13 is the NWP wind speed as this parameter is one of the key parameters in determining the size of any diurnal warming event and is plotted for information only.

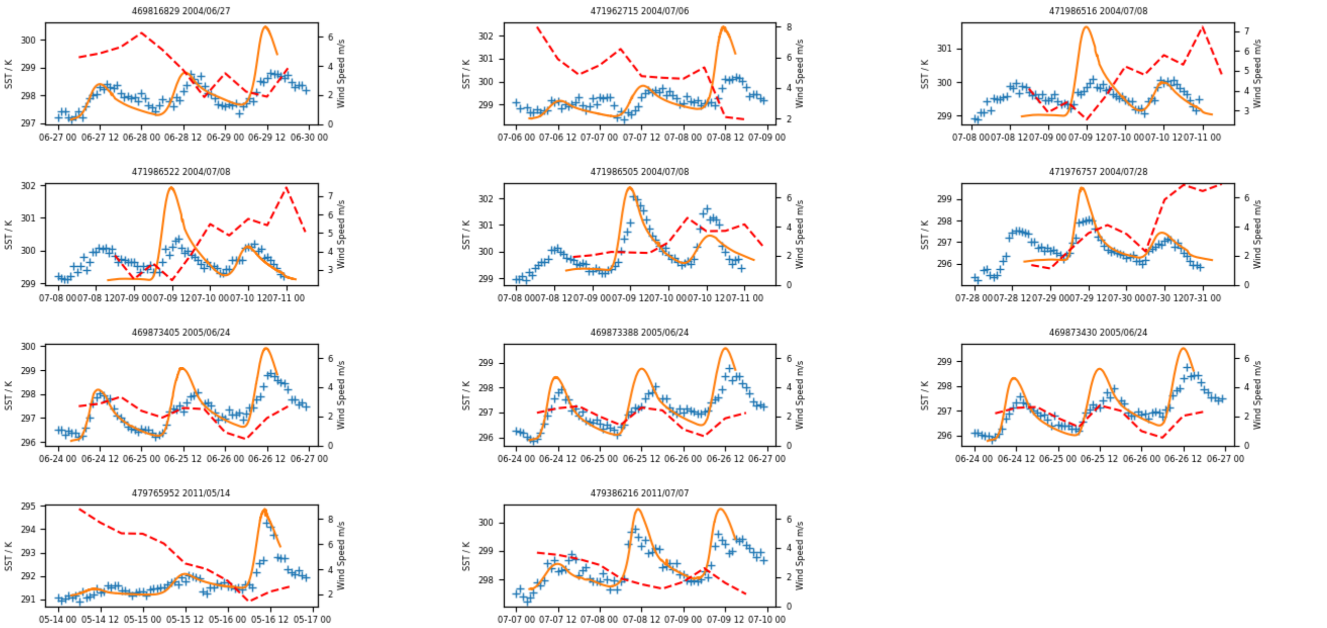

Figure 13 SEVIRI SST time series (blue crosses) with the diurnal model using default parameters (orange line) together with the NWP wind speed (dashed red line)

Even though the number of cases is small, Figure 13 shows a range of differences between the predicted warming and the observed SEVIRI SST signal. In general, the timing of the modelled peak warming is reasonably consistent with the observed signal and there are a couple of times/locations when the model agrees well with the SEVIR SST signal, at least of a single warming event (for example see matchup id=469873405, col:1 row:3). All three warming events seem to be reasonably modelled for matchup id=479765952 (col:1 row:4), but in general looking at all the data there do seem to be systematic differences where the observed warming is smaller than the predicted warming. Sometimes this may be due to an underestimate of the NWP wind speed, as may be indicated by matchup id=471986516 (col:3 row:1) where at the lowest wind speeds the diurnal model is significantly overestimating the diurnal warming signal. For the other cases, the origin of the differences between the modelled and observed diurnal signal is not particularly clear. Therefore, for all cases a tuning of some of model parameters may help improve the statistical match between the model and SST diurnal variations.

A7.6. Impact of parameter variation on modelled diurnal variation

To understand how to tune the diurnal model it is important to understand which of the parameters listed in Table 1 significantly impact the size and shape of the modelled signal. In this report we have concentrated on those parameters listed in Table 2. In terms of the impact on the model output, examples from two locations are shown in Figure 14 and Figure 15 which show the DW model using the default parameters together with the SEVIRI SST data (right-hand side) and shows the model only time series where the parameters have been varied (left-hand side.) For each parameter (Table 2) the variation shown corresponds to varying the given parameter between 50% and 150% of the default value, leaving all other parameters at the default setting. Further work will need to be done on ensuring that the range of variation of these parameters is reasonable so these variation are only intended as a guide to the impact of the changing parameters.

What is clear in Figure 14 and Figure 15 is that some parameters have a limited effect on the DW model (such as KMB_init, langmuir_top_boundary_factor and wave_enhancement). The FCT parameter seems to show a two-state condition which is probably because FCT is more related to the stability of the algorithm rather than the size of warming. Therefore, in terms of parameters which could be used to tune the model we are left with KHB_init, shear_enhancement_factor and surface_abs_enhancement (for what these terms mean see Table 1 and Table 2). Each of these parameters have a slightly different impact on the DW time series (see Figure 16 to Figure 18).

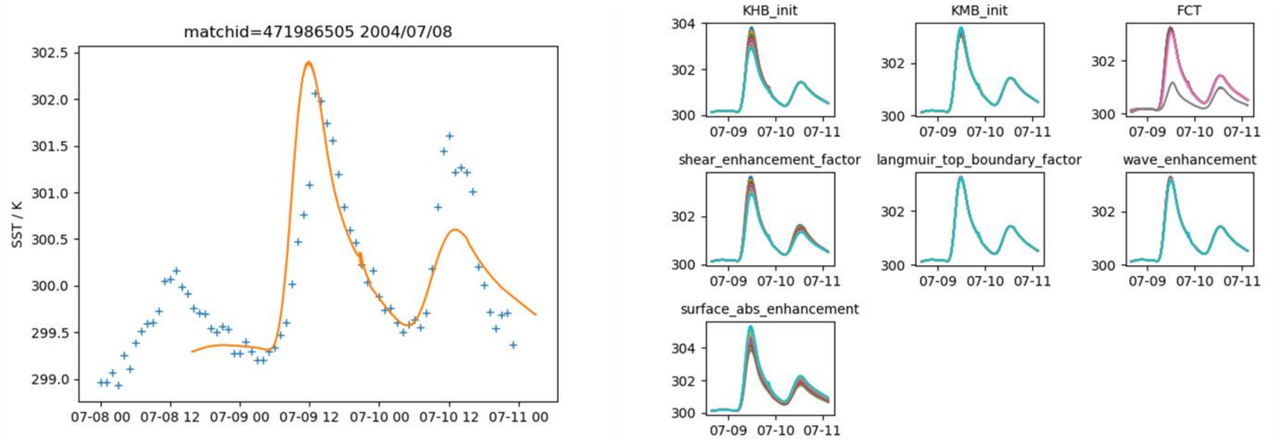

Figure 14 LH plot shows the SEVIRI SST with the default DW model plotted in orange for matchup id=471986505. The RH plot shows a variation in the model output for changes in the parameter type from Table 2. The largest DW signal seen in all the locations is quite well represented by the DW model (second peak).

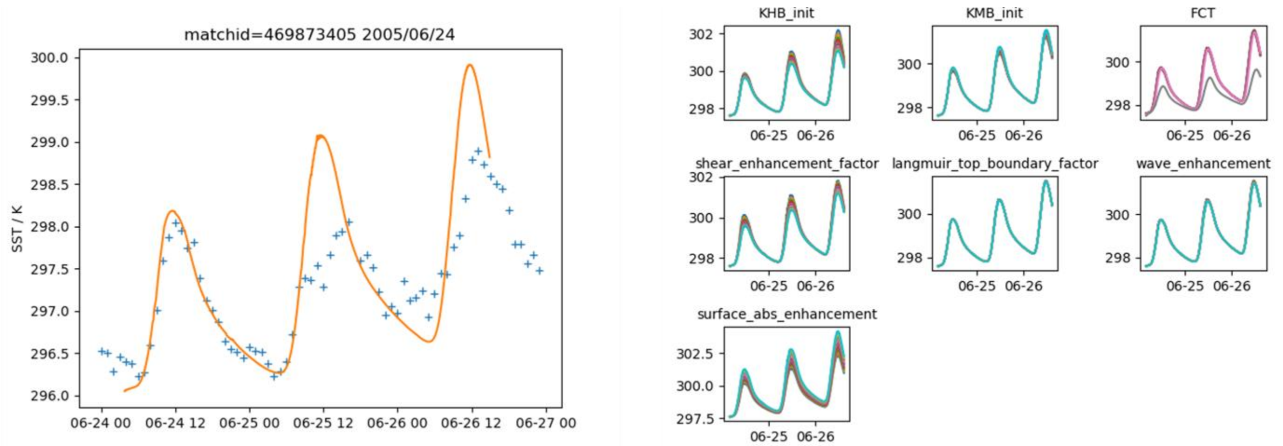

Figure 15 LH plot shows the SEVIRI SST with the default DW model plotted in orange for matchup id=469873405. The RH plot shows a variation in the model output for changes in the parameter type from Table 2. The first peak lines up very well with the observed DW signal.

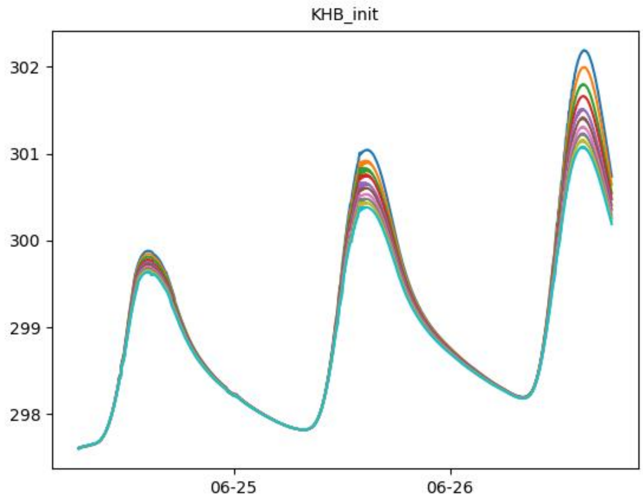

Figure 16 Diurnal model variations as the KHB_init parameter is varied.

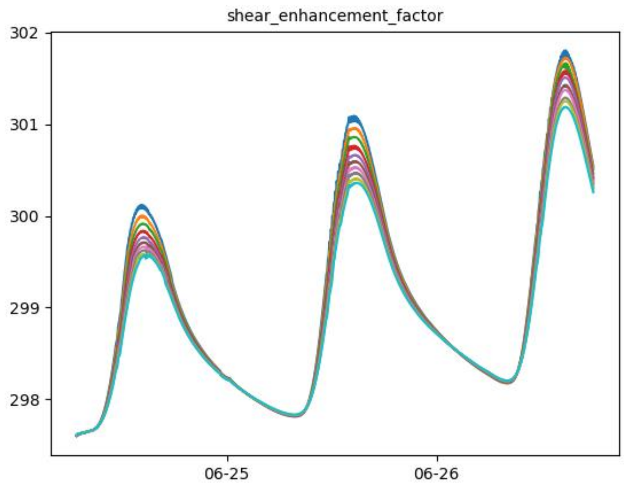

Figure 17 Diurnal model variations as the shear_enhancement_factor parameter is varied.

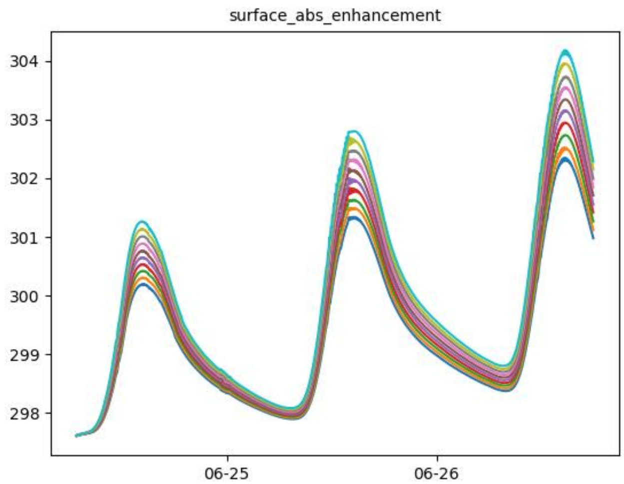

Figure 18 Diurnal model variations as the surface_abs_enhancement parameter is varied.

In the case of those parameters which do lead to model variations it should be possible to discriminate between the different cases because there are differences in the asymmetry of the observed changes. For example, the KHB_init and surface_abs_enhancement cases are both more asymmetric than the shear_enhancement_factor case with the surface_abs_enhancement case looking like it is the most asymmetric as well as impacting the next day dawn signal which does not happen with the other parameters. There are also differences in how the variation evolves across the days although it would only be the second day that would be used for subsequent analysis, leaving the first day to let the model spin up. We also note that there are differences between the how the parameters operate in the sense that the suface_abs_enhancement variable is essentially a correction to the modelled net solar flux (defined by the NWP) whereas the other corrections are more related to the physical constants used in the model itself. One other NWP related variable that can change the size of a diurnal warming signal is, of course, the wind speed so looking at observed diurnal signals compared to the model will help us understand the pattern of error that exists in the NWP data itself.

A7.7. Proposed Tasks

Given the results shown above there is definitely scope for tuning the NOAA diurnal model to better match the model to observed diurnal signals.

For future work we would recommend improving upon the AATSR matchup dataset described above and would take as a starting point the TRUSTED buoy network; this provides both highly accurate SSTs and some buoys can also provide a high time resolution time series.

Our initial survey of the TRUSTED data considering approximately two years of data shows ~7% of the days have an SST variation of greater than 1 K over a 24 hour period (2215 from 30979 observed days). The overall distribution of probable warming events (> 1 K over a day) from the TRUSTED buoy data is shown in the left-hand panel of Figure 19. Also shown is an example of a diurnal warming events from a single buoy in the right-hand plot of Figure 20.

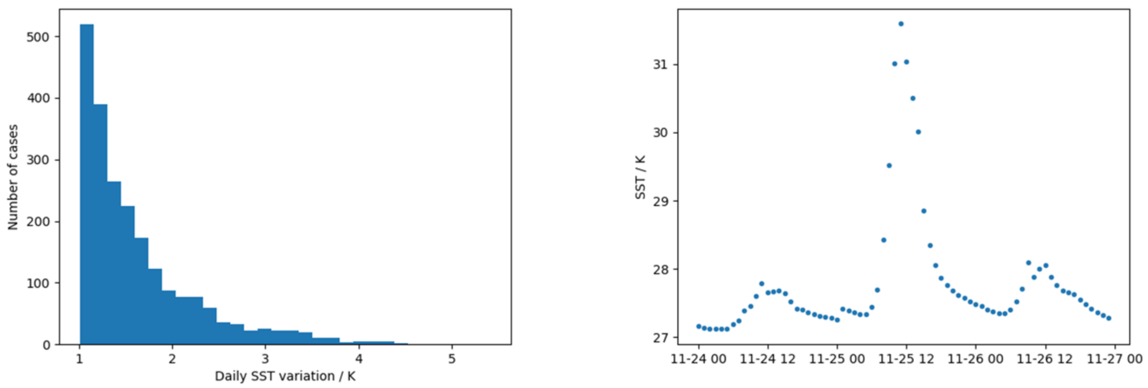

Figure 19 Left hand plot shows the distribution of the 24 hour variation of SST (filtered to events where the variation is > 1K) from the trusted dataset and the right hand plot shows an example of the time variation in SST over a three-day period from a trusted buoy from 2019/11/24.

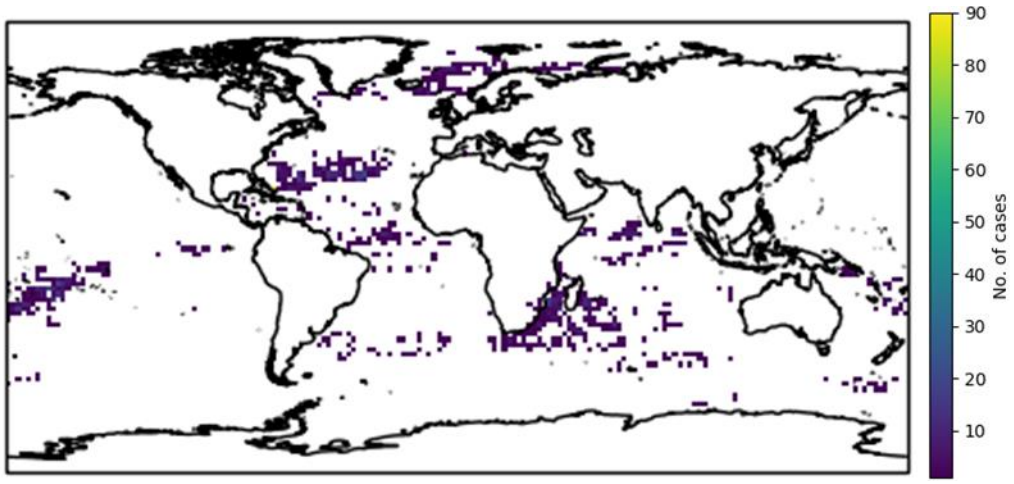

Figure 20 Spatial density of TRUSTED buoys with a daily variance of 1 K binned into 2 degree boxes.

Therefore, the TRUSTED dataset seems to have an adequate distribution of diurnal events including some large warming events that can be used to constrain the NOAA DW model. It should be pointed out, however, that while the TRUSTED data samples a quite a large range of latitudes it does not have global spatial coverage (as shown in Figure 20).

Therefore we recommend in addition to look at some other data sources including data from the GEO Ring to ensure the tuned model can be used globally. Geostationary SST data will also be used to understand the NWP error characteristics (both spatial and temporal) by varying NWP fields (such as the wind speed etc.) to match the observed diurnal variation and studying the spatial structure of NWP field error to better understand the uncertainties in the NWP data. The combination of analysing the TRUSTED data together with studies using geostationary SST data will then provide both DW model tuning together with a better understanding of NWP related model uncertainties. As a final step we will compare the tuned model with the SLSTR matchup data and do a final scaling of the model if appropriate.

The recommended work plan is then as follows:

1. Create a version of the TRUSTED data which includes the relevant NWP fields (based on the operational SLSTR met fields) and run the diurnal warming model using default parameters. Assess the default behaviour of the DW model.
2. Study the variations in the differences between the observed and modelled signal to look for cases which can be used to identify possible NWP field errors (e.g. wind speed). Obtain associated geostationary and NWP data for the identified events.
3. Study the spatial variations of the diurnal warming events using the geostationary SST data compared to the modelled diurnal signals and modify NWP data (such as wind speed) to improve the match between model and signal to understand the location and size of possible NWP errors. Also obtain geostationary data where the NWP wind speed is low and relatively cloud clear in regions missed by the TRUSTED data to further understand possible NWP errors.
4. Using the improved understanding of NWP error characteristics, select TRUSTED locations where the final NWP error should be small (including possible adjustments) to tune model parameters to the TRUSTED time series.
5. Use an SLSTR matchup dataset to check the observed diurnal signals and look for any overall scaling that may be needed compared to the TRUSTED fits. Note that as the TRUSTED data is at depth, some extra adjustments may be needed to scale to the observed SLSTR skin SST.
6. Rewrite the DW model code to allow for different input parameter values and set up the required parameter files.

Development Task	Tuning/improvements of the NOAA Diurnal model to SLSTR
Task Number	DV
Objective	To tune the NOAA Diurnal Model to better match DW events seen from TRUSTED data and normalise to the SLSTR in-situ MMD signals.
Required Research	Obtain TRUSTED data, add associated NWP fields and run the NOAA diurnal model. Study the observation to model differences. Get geostationary data for TRUSTED warming events as well as for regions with no TRUSTED data. Understand the range of diurnal variability and how well errors in selected NWP fields map spatially to understand NWP error terms. Select TRUSTED data with improved NWP data based on previous study. Fit appropriate DW model parameters to the selected TRUSTED data and compare with geostationary SST signals. Compare tuned model to the SLSTR matchup data and scale if appropriate for final model.
Required Algorithm Development	Modify the NOAA DW code to allow for different input parameters that may be a function of time/location.

Comments on implementation	The DW model will be a standalone executable with associated files
Benefits beyond state of the art	Improvements to any diurnal correction/estimation for SLSTR.

A8. Tools helpful for algorithm development

At the second project meeting, it was agreed that it was also relevant to document here any concepts for tools that could in general support the improvement and development of SLSTR SST products.

A8.1. L1 and L2 image exploration tool

During the BCDS a simple HTML tool has been developed and employed to help compare instances of cloud detection (Figure 21).

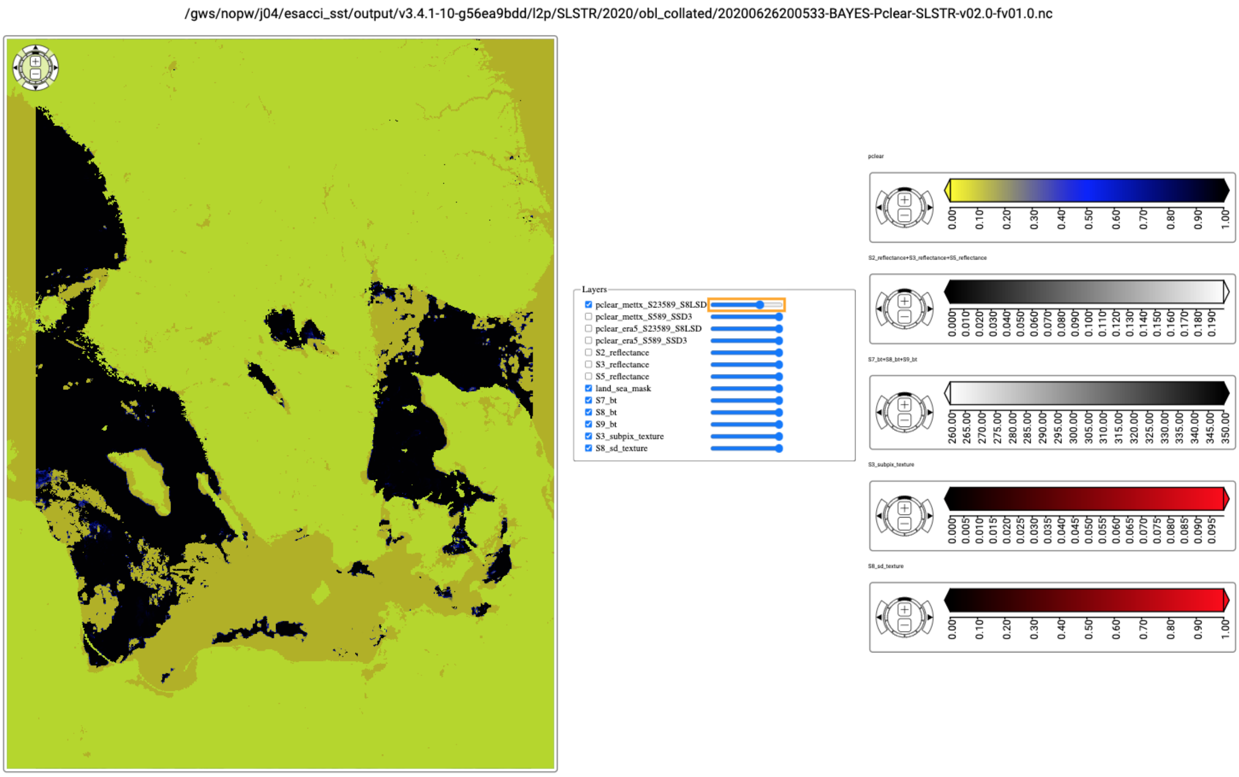

Figure 21 Screenshot of BCDS “slider” tool.

In summary the functionality is as follows:

- Select layer(s) to display and its(their) transparency. As in the figure, this enables an image to be seen in relation to the SLSTR cloud mask (for example), or Bayesian probability values to be overlaid on a reflectance channel.
- Drag and zoom the image within its window.
- Drag and zoom the colour bars applicable for a given layer.

The SLSTR data (input L1 and layers output from the PP) are preprocessed to select and create the layers available within the tool, and to create an html page. The page is opened in a browser (tested only for Chrome) and is used in this way.

Developments that would enhance the tool (unprioritised):

- extend domain of the “tool” to include calling a configured run of the PP (modified to appropriately interface its output layers with the selections from tool call), and generate the html (or other) interface
    - extended set of selectable output layers from PP:

  - as well as clear-sky probability, intermediate Bayesian quantities (ADF values, prior fields, likelihoods, etc.), any channel and (in case of reflectance) the sub-pixel standard deviation of the channel from the preprocessor,
  - control to run the PP more than once to integrate alternative configurations of clear-sky probability,
  - SST fields (all channel/view combinations),
- in the context of the above, create automatic metadata for the PP run and configuration associated with the HTML page, for traceability,
- ability to switch land-mask colours, including ability to choose "transparent" for either land or ocean areas,
- ability to hover cursor over image to read x, y and all layer data in a pop window, and/or store/download the set of values for offline detailed investigation,
- ability to save the image window as an image with metadata for its creation,
- greater flexibility in applying colour bars and scale limits,
- ability to add other user-furnished layers.

The approach of the current tool raises performance issues for large numbers of layers which will also need to be addressed as part of the development.

Of course, all this capability could be done in other contexts (perhaps SNAP, for example) but (1) workflow is streamlined considerably when a more limited but more tailored tool is available, and (2) the traceability options and PP integration would also be significant advantages.

A further extension of the principle would be to enable 2 SLSTR images to be available, projected on a common image grid. This could be of value in developing night-time cloud detection over sea ice, where inter-orbital coherence could be a clue to clear/cloud determination in the absence of reflectance imagery.

A8.2. Integration of DV model

Tools that help integrate the DV model into validation strategies using satellite-in situ matchup datasets. The first step should be a python wrapper for the model that users can integrate easily.