



land surface  
temperature  
cci



CCI Land Surface Temperature

## Algorithm Theoretical Basis Document

WP2.1 – DEL-LST-CCI-D2.2-ATBD

Ref.: LST-CCI-D2.2-ATBD

Date: 25-Nov-2021

Organisation: Consortium CCI LST





## Signatures

	Name	Organisation	Signature
Written by	Darren Ghent	ULeic	
	Emma Dodd	ULeic	
	Karen Veal	ULeic	
	Mike Perry	ULeic	
	Carlos Jimenez	Estellus	
	Sofia Ermida	IPMA	
Reviewed by	Carlos Jimenez	IPMA	
	Sofia Ermida	IPMA	
Approved by	Darren Ghent	ULeic	
Authorized by	Simon Pinnock	ESA	

## Change log

Version	Date	Changes
1.0	28-Feb-2109	First version
1.1	7-May-2019	RIDs addressed
2.0	15-May-2020	Added details of improvements to surface classification (biome and snow masking) and temporal correction methodology
2.1	27-Jul-2020	RIDs addressed
3.0	25-Nov-2021	Additions made for new products

## List of Changes

Version	Section	Changes
2.0	4.4	Update to the description of the Calibration Database to account for new developments
2.0	4.6.1	Description of new implementation of biome dataset from LC_cci
2.0	4.6.5	Description of improvements to snow masking approach
2.0	4.8.1	New section describing the intercalibration of L1 data between different sensors in a Climate Data Record
2.0	4.8.2	New section describing the temporal correction of LST between different sensors in a Climate Data Record
2.1	All	RIDs addressed
3.0	3.5	New section for description of GOES
3.0	3.6	New section for description of MTSAT
3.0	4.2.4	New section on single-channel (SMW) algorithm
3.0	4.7.2	Updates for TCWV uncertainty
3.0	4.9.1	New section describing the ATSR-SLSTR IR CDR
3.0	4.9.2	New section describing the Merged IR CDR
3.0	5.5	Rewritten section detailing the new uncertainty model for microwave algorithms

## Table of Content

<b>1. EXECUTIVE SUMMARY</b>	<b>1</b>
<b>2. INTRODUCTION</b>	<b>2</b>
2.1. Purpose and Scope	3
2.2. Reference Documents	3
2.3. Glossary	10
<b>3. SENSOR DESCRIPTIONS</b>	<b>12</b>
3.1. ATSR-2 and AATSR	12
3.2. SLSTR	13
3.3. MODIS	13
3.4. SEVIRI	13
3.5. GOES	14
3.6. MTSAT	14
3.7. SSM/I and SSMIS	15
<b>4. RETRIEVAL OF LAND SURFACE TEMPERATURE FROM THERMAL INFRARED SENSORS</b>	<b>17</b>
4.1. Physics of the Problem	17
4.1.1. Mathematical description	18
4.1.2. Emissivity	20
4.2. Algorithm Descriptions	21
4.2.1. Split-Window (SW) Approximation	21
4.2.2. UOL Algorithm	23
4.2.3. Generalised Split Window (GSW) Algorithm	23
4.2.4. Single-channel (SMW) Algorithm	24
4.3. Radiative Transfer Modelling	25
4.4. Calibration Database for Determining Retrieval Coefficients for the TIR Algorithms	25
4.5. Identification of Observations Valid for Land Surface Temperature Estimation from Thermal Infrared Sensors	27
4.5.1. UOL_3 Algorithm	28
4.5.2. Bayesian Algorithm	29
4.5.3. NWCSAF Cloud Mask Algorithm	30
4.6. Auxiliary Datasets for Thermal Infrared Retrievals	30
4.6.1. Land cover	30
4.6.2. Fractional Vegetation	31
4.6.3. Emissivity	32
4.6.4. Atmospheric Variables	33
4.6.5. Snow masking	33
4.7. Uncertainty Model for Thermal Infrared Algorithms	35
4.7.1. Random	35
4.7.2. Locally systematic	36
4.7.3. Systematic	37
4.8. Harmonisation for Climate Data Records	37
4.8.1. Cross-calibration of BTs using IASI spectra	37
4.8.2. Adjustment to common reference time	41
4.9. Multi-sensor Climate Data Records	43
4.9.1. ATSR-SLSTR CDR	43
4.9.2. Merged IR CDR	44
<b>5. RETRIEVAL OF LAND SURFACE TEMPERATURE FROM MICROWAVE SENSORS</b>	<b>46</b>
5.1. Physics of the Problem	46
5.1.1. Mathematical description	46

	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: v
---	---	---

5.1.2. Emissivity-----47

**5.2. Algorithm Description----- 48**

5.2.1. NNEA Algorithm-----48

**5.3. Calibration Database for Determining Retrieval Coefficients for the MW Algorithm ----- 49**

**5.4. Auxiliary Datasets for Microwave Retrievals ----- 50**

5.4.1. Emissivity-----50

**5.5. Uncertainty Model for Microwave Algorithms----- 51**

### List of Figures

Figure 1: Data flows for LST\_cci ECV single-sensor product prototype production system for thermal infrared sensors. For merged products the algorithms are applied to harmonised L1 data processed through to L3U and then merged to form L3S Products ----- 2

Figure 2: Data flows for the SSM/I and SSMIS LST ECV prototype production system. ----- 3

Figure 3: Local ascending equator crossing time for SSM/I and SSMIS (courtesy of Remote Sensing Systems, <http://www.remss.com/support/crossing-times/>). Only the FXX instruments are relevant here. ----- 16

Figure 4: Flow chart for the generation of the simulated true Top-Of-Atmosphere Brightness Temperature values for the dataset.----- 26

Figure 5: For each granule of an AATSR orbit (left), the expected 12 μm brightness temperature is simulated from coincident profiles. The PDF of observed 12 μm brightness temperatures for each land cover-diurnal condition, given the space and time position, is also determined (top-right in green). This PDF is moved so that the mean equals the expected mean for the granule and the new PDF represents the expected clear-sky conditions (bottom-right in green). Figure 1 from [RD-58]. ----- 29

Figure 6: A schematic showing how IMS NH data is incorporated into biome and cloud information for the LST products using biome information. ----- 34

Figure 7: 5x5 box of SEVIRI pixels that align with IASI pixel (taken from “ATBD for Prototype GSICS SEVIRI- IASI Inter-Calibration”) ----- 38

Figure 8: Effect of collocation by histogram binning: if IASI pixel (blue) is centred in corner of lat/lon grid (green) then matching SEVIRI pixels in adjacent grid boxes are ignored and only those pixels (orange) in the same gridbox are used in the analysis. ----- 40

Figure 9: Change of shape of IASI pixel across track (taken from IASI Level 1: Product Guide) – note axes have different scales – pixels are circular at nadir.----- 40

Figure 10: Envisat AATSR minus Terra MODIS LST difference v. time difference: average for January over all years, for latitude band 40-50 °N, and landcover class 6 (Closed\_(>40percent)\_broadleaved\_deciduous\_forest\_(>5m)), for daytime (top) and night-time (bottom). Top panel each plot shows mean (circle) and mean +/- 1 sd (bars), bottom panel show number of samples. ----- 42

Figure 11. Illustration of a single-channel regression model and calibration dataset. A neural network is trained to retrieve LST from the SSM/I 37 GHz vertically polarized brightness temperatures (TB). Red dots are the samples in the calibration dataset, and the neural network output is represented in blue. The horizontal axis has been divided into 5 K TB bins, and for each bin the mean (black circles) and +/- one standard deviation (black line centred around the mean) are given. See the text for more details. ----- 52

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: vi
--	---	--

Figure 12. Illustration of (A) a traditional LST estimation with a Neural Network (NN), and (B) the proposed scheme where both the LST and an estimation of the LST error are provided by the NN. See the text for more details. Diagram adapted from Figure 2 in [RD-9]. ----- 53

**List of Tables**

Table 1: Characteristics of the SEVIRI instruments onboard Meteosat Second Generation. ----- 14

Table 2: Mapping logic for replacement of LC\_cci indexes: 200, 201 and 202 with ALB-2 bare soil information.----- 31

Table 3: ULeic / LandCover\_cci (LCCS) hybrid biome definition.----- 54

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 1
--	---	---

## 1. Executive Summary

The European Space Agency Climate Change Initiative on Land Surface Temperature (hereafter LST\_cci) aims to provide Land Surface Temperature (LST) LST Essential Climate Variable (ECV) products and validate these data to provide an accurate view of temperatures across land surfaces globally over the past 20 to 25 years.

This Algorithm Theoretical Basis Document (ATBD) provides a detailed definition of the Land Surface Temperature (LST) valid observation identification and retrieval methodologies to be used for LST data products provided by LST\_cci. The algorithms described in this document have been identified as the best algorithms for a future climate quality operational system during an open algorithm intercomparison round-robin. This document describes retrieval algorithms selected for use in deriving LST from Thermal Infrared and Microwave sensors. These are the University of Leicester (UOL) algorithm and Generalised Split Window (GSW) algorithm for thermal infrared data, and the Neural-Network-Emissivity-All-channels (NNEA) algorithm for microwave data.

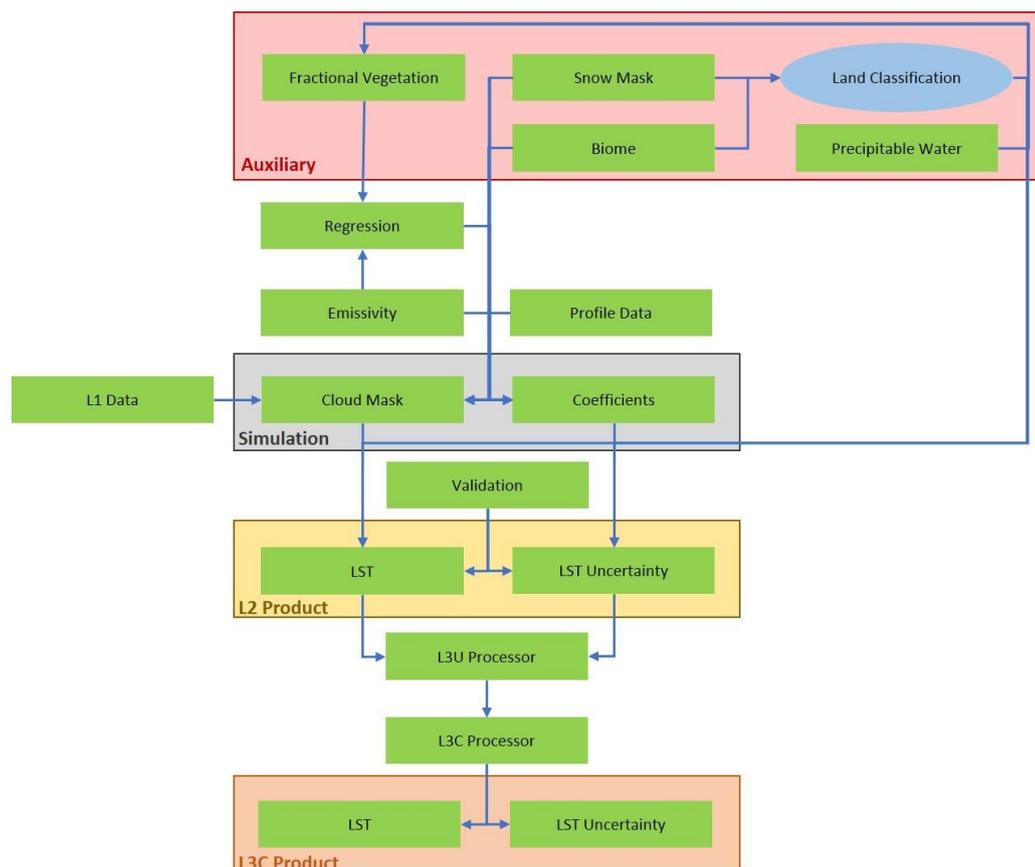
Information is also provided in this document for any cloud clearing methods used; auxiliary datasets; uncertainty models and propagation of uncertainties; and calibration datasets. The methods outlined in this document are implemented in an end-to-end system to generate the first LST\_cci climate data records.

## 2. Introduction

The European Space Agency Climate Change Initiative on Land Surface Temperature (hereafter LST\_cci) aims to provide Land Surface Temperature (LST) LST Essential Climate Variable (ECV) products and validate these data to provide an accurate view of temperatures across land surfaces globally over the past 20 to 25 years.

This Algorithm Theoretical Basis Document (ATBD) provides a detailed definition of the Land Surface Temperature (LST) clear sky detection and retrieval methodologies to be used for LST data products provided by LST\_cci. The algorithms described in this document have been identified as the best algorithms for a future climate quality operational system. The retrieval algorithms were selected during an open algorithm intercomparison round-robin which assessed the performance of a number of different LST retrieval algorithms for a set of specific thermal infrared and microwave satellite sensors [RD-68].

The methods outlined in this document are implemented in an end-to-end system to generate the first LST\_cci climate data records. A flow chart summarising the algorithm processing is provided in Figure 1 for thermal infrared sensors and in Figure 2 for microwave sensors.



**Figure 1: Data flows for LST\_cci ECV single-sensor product prototype production system for thermal infrared sensors. For merged products the algorithms are applied to harmonised L1 data processed through to L3U and then merged to form L3S Products**

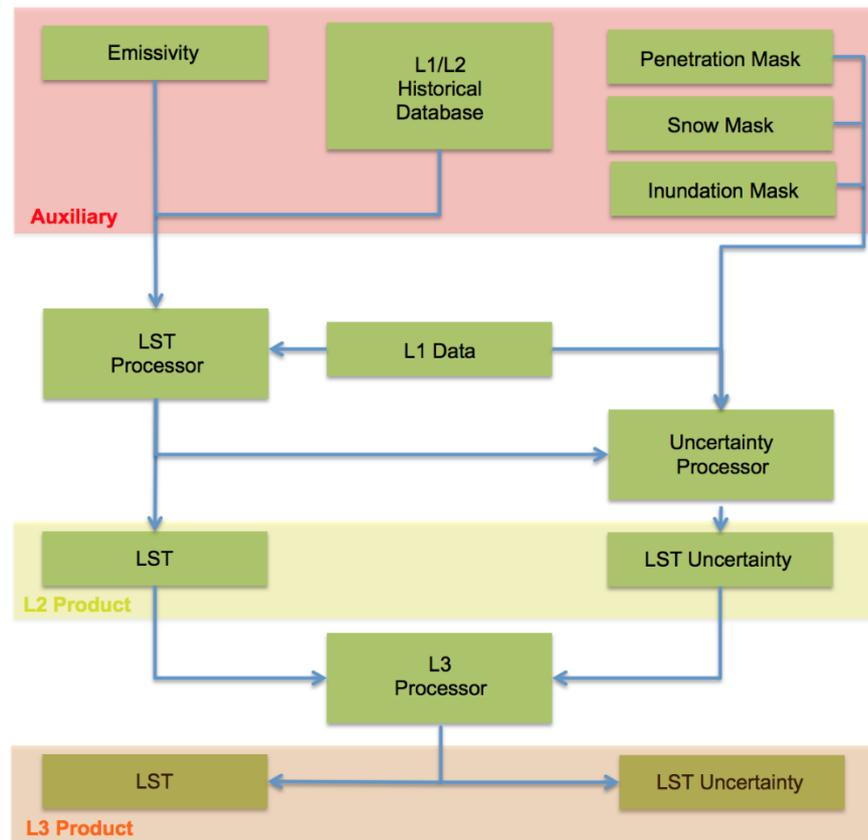


Figure 2: Data flows for the SSM/I and SSMIS LST ECV prototype production system.

It is expected that ongoing algorithm assessment will be carried out for each subsequent reprocessing to ensure the best performing algorithm is always implemented. This aims to produce the most accurate LST retrieval for each LST\_cci product. Therefore, this document will be updated as necessary.

Note, for merging TIR and MW an experimental approach will be taken not necessarily utilising the existing algorithms.

## 2.1. Purpose and Scope

This document presents the algorithm theoretical basis of retrieval methodologies to be used for LST data products provided by LST\_cci.

## 2.2. Reference Documents

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

Id	Reference
[RD-1]	Price, J. C., 1984, Land surface temperature measurements from the split-window channels of the NOAA- 7/AVHRR, <i>J. Geophys. Res.</i> , 89, 7231–7237.
[RD-2]	McMillin, L. M. and D. S. Crosby, 1984, Theory and validation of the multiple window sea surface temperature technique, <i>J. Geophys. Res.</i> , 89, 3655–3661.
[RD-3]	Prata, A. J., 1993, Land surface temperatures derived from the AVHRR and the ATSR, 1, Theory, <i>J. Geophys. Res.</i> , 98(D9), 16,689–16,702.
[RD-4]	Prata, A. J., 1994a, Land surface temperatures derived from the AVHRR and the ATSR, 2, Experimental results and validation of AVHRR algorithms, <i>J. Geophys. Res.</i> , 99(D6), 13,025–13,058.
[RD-5]	Prata, A. J., 1994b, Land surface temperature determination from satellites, <i>Advances in space research</i> , 14, 15-26.
[RD-6]	Sutherland, R. A., 1979, Broadband and spectral emissivities of natural sands and vegetation (2–20 $\mu\text{m}$ ), <i>J. Atmos. Oceanic Technol.</i> , 3, 199–202.
[RD-7]	Salsbury, J. W., and D. M. D’Aria, 1992, Emissivity of terrestrial materials in the 8 – 14 $\mu\text{m}$ atmospheric window, <i>Rem. Sensing Environ.</i> , 42, 83–106.
[RD-8]	Snyder, W. C., Wan, Z., Zhang, Y., and Y.-Z. Feng, 1998, Classification-based emissivity for land surface temperature measurement from space, <i>Int. J. Remote Sensing</i> , 19(4), 2753–2774.
[RD-9]	Prata, F., Land Surface Temperature Measurement from Space: AATSR Algorithm Theoretical Basis Document. 2002.
[RD-10]	Jiménez-Muñoz, J.C., Sobrino, J.A., Skoković, D., Mattar, C., & Cristóbal, J. (2014). Land Surface Temperature Retrieval Methods From Landsat-8 Thermal Infrared Sensor Data. <i>IEEE Geoscience and Remote Sensing Letters</i> , 11, 1840-1843
[RD-11]	Ghent, D., Corlett, G., Goettsche, F., & Remedios, J. (2017) Global land surface temperature from the Along-Track Scanning Radiometers. <i>Journal of Geophysical Research – Atmospheres</i> , 122, 12167-12193
[RD-12]	Wan, Z. and Dozier, J., (1996) A generalised split-window algorithm for retrieving Land-Surface Temperature from space, <i>IEEE Trans. GeoSci. Remote Sens.</i>
[RD-13]	Holmes, T. R. H., R. A. M. De Jeu, M. Owe, and A. J. Dolman (2009), Land surface temperature from Ka band (37 GHz) passive microwave observations, <i>J. Geophys. Res.</i> , 114, D04113, doi:10.1029/2008JD010257.
[RD-14]	Prigent, C., C. Jimenez, and F. Aires (2016), Toward “all weather,” long record, and real-time land surface temperature retrievals from microwave satellite observations, <i>J. Geophys. Res. Atmos.</i> , 121, 5699–5717, doi:10.1002/2015JD024402.
[RD-15]	Ghent D, Trigo I, Pires A, Sardou O, Bruniquel J, Gottsche F, Martin M, Prigent C, Jimenez, C, and Remedios. J., 2016. ESA DUE GlobTemperature Product User Guide V2
[RD-16]	Merchant, C., Ghent. D., Kennedy, J., Good., E., and Hoyer, J. 2016. Common approach to providing uncertainty estimates across all surfaces, H2020 EUSTACE Report
[RD-17]	Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), 5 <sup>th</sup> August 2019. <a href="https://cds.climate.copernicus.eu/cdsapp#!/home">https://cds.climate.copernicus.eu/cdsapp#!/home</a>
[RD-18]	Ramsay, B.H. (1998). The interactive multisensor snow and ice mapping system. <i>Hydrological Processes</i> , 12, 1537-1546

Id	Reference
[RD-19]	Ghent, D., Corlett, G., Gottsche, F., & Remedios, J. (2017). Global land surface temperatures from the Along-Track Scanning Radiometers. <i>Journal of Geophysical Research: Atmospheres</i> , 122
[RD-20]	Le Gleau, H. (2019) Algorithm Theoretical Basis Document for the Cloud Product Processors of the NWC/GEO. 5 <sup>th</sup> August 2019. <a href="http://www.nwcsaf.org/Downloads/GEO/2018/Documents/Scientific_Docs/NWC-CDOP2-GEO-MFL-SCI-ATBD-Cloud_v2.1.pdf">http://www.nwcsaf.org/Downloads/GEO/2018/Documents/Scientific_Docs/NWC-CDOP2-GEO-MFL-SCI-ATBD-Cloud_v2.1.pdf</a>
[RD-21]	Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholczyk, P. and Smets, B., 2013. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production. <i>Remote sensing of environment</i> , 137, pp.299-309.
[RD-22]	Verger, A., Baret, F. and Camacho, F., 2011. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. <i>Remote Sensing of Environment</i> , 115(2), pp.415-426.
[RD-23]	Camacho, F., Cernicharo, J., Lacaze, R., Baret, F. and Weiss, M., 2013. GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products. <i>Remote Sensing of Environment</i> , 137, pp.310-329.
[RD-24]	Combined ASTER and MODIS Emissivity database over Land (CAMEL) Emissivity Monthly Global 0.05Deg V001. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MEaSURES/LSTE/CAM5K30EM.001
[RD-25]	Saunders, R., RTTOV-7 - SCIENCE AND VALIDATION REPORT, 2001, EUMETSAT.
[RD-26]	Matricardi, M.: Technical Note: An assessment of the accuracy of the RTTOV fast radiative transfer model using IASI data, <i>Atmos. Chem. Phys.</i> , 9, 6899-6913, doi:10.5194/acp-9-6899-2009, 2009.
[RD-27]	Aires, F., C. Prigent, F. Bernardo, C. Jimenez, R. Saunders, and P. Brunel (2011), A Tool to Estimate Land-Surface Emissivities at Microwave frequencies (TELSEM) for use in numerical weather prediction, <i>Q. J. R. Meteorol. Soc.</i> , 137, 690–699.
[RD-28]	Hocking, J., Rayer, P., Rundle, D., Saunders, R., Matricardi, M., Geer, A., Brunel, P., & Vidot, J. (2015). RTTOV v11 Users Guide. EUMETSAT Satellite Application Facility on Numerical Weather Prediction. NWPSAF-MO-UD-028. Version 1.4
[RD-29]	Trigo, I.F., Peres, L.F., DaCamara, C.C. and Freitas, S.C., 2008. Thermal land surface emissivity retrieved from SEVIRI/Meteosat. <i>IEEE Transactions on Geoscience and Remote Sensing</i> , 46(2), pp.307-315.
[RD-30]	The IGBP-DIS Global 1 km Land Cover Data Set (DISCover) Proposal and Implementation Plans, 1996.
[RD-31]	L. F. Peres, C. C. DaCamara, "Emissivity maps to retrieve land-surface temperature from MSG/SEVIRI", <i>IEEE Trans. Geosci. Remote Sens.</i> , vol. 43, no. 8, pp. 1834-1844, Aug. 2005.
[RD-32]	Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kaliberg, P., Kobayashi, S., Uppala, S., Simmons, S. (2011). The ERA-Interim archive Version 2.0. ERA Report Series, 5 <sup>th</sup> August 2019. <a href="https://www.ecmwf.int/en/elibrary/8174-era-interim-archive-version-20">https://www.ecmwf.int/en/elibrary/8174-era-interim-archive-version-20</a>
[RD-33]	Merchant, C.J., Embury, O., Rayner, N.A., Berry, D.I., Corlett, G.K., Lean, K., Veal, K.L., Kent, E.C., Llewellyn-Jones, D.T., Remedios, J.J., & Saunders, R. (2012). A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers. <i>Journal of Geophysical Research: Oceans</i> , 117

Id	Reference
[RD-34]	Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S., and Donlon, C. (2014). <i>Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI)</i> . Geoscience Data Journal, 1 (2), pp. 179-191. ISSN 2049-6060 doi: 10.1002/gdj3.20.
[RD-35]	Maturi, E., Harris, A., Merchant, C. J., Mittaz, J., Potash, B., Meng, W., Sapper, J. (2008). NOAA's sea surface temperature products from operational geostationary satellites. Bulletin of the American Meteorological Society, 89, 12, 1877-1888.
[RD-36]	Merchant, C. J., Harris, A. R., Maturi, E., Embury, O., MacCallum, S. N., Mittaz, J., 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, 570-581.
[RD-37]	Jiménez, C., C. Prigent, S. L. Ermida, and J.-L. Moncet (2017), Inversion of AMSR-E observations for land surface temperature estimation: 1. Methodology and evaluation with station temperature, J. Geophys. Res. Atmos., 122, 3330–3347, doi:10.1002/2016JD026144.
[RD-38]	Aires, F., C. Prigent, F. Bernardo, C. Jimenez, R. Saunders, and P. Brunel (2011), A Tool to Estimate Land-Surface Emissivities at Microwave frequencies (TELSEM) for use in numerical weather prediction, Q. J. R. Meteorol. Soc., 137, 690–699.
[RD-39]	Nguyen, D., and B. Widrow (1990), Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptative weights. Paper presented at 1990 International Joint Conference on Neural Networks (IJCNN), pp. 21–26, IEEE, New York, 17–21 Jun.
[RD-40]	Hagan, M. T., and M. Menhaj (1994), Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Networks, 5, 989–993.
[RD-41]	Kirches, G., Brockman, C., Boettcher, M., Peters, M., Bontemps, S., Lamarche, C., Schlerf, M., and Santoro, M. (2016), Land Cover CCI PRODUCT USER GUIDE VERSION 2, 5 <sup>th</sup> August 2019. <a href="http://data.ceda.ac.uk/neodc/esacci/land_cover/docs/ESACCI-LC-PUG-v2.5.pdf">http://data.ceda.ac.uk/neodc/esacci/land_cover/docs/ESACCI-LC-PUG-v2.5.pdf</a>
[RD-42]	Prigent C., W. B. Rossow, E. Matthews, Microwave land surface emissivities estimated from SSM/I observations, Journal of Geophysical Research, 102, 21867-21890, 1997.
[RD-43]	Van der Baan, M. and Jutten, C. (2000) Neural networks in geophysical applications. Geophysics, 65 (4). pp. 1032-1047, doi.org/10.1190/1.1444797
[RD-44]	Hagan, M., and M. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Networks, 5, 989–993, 1994.
[RD-45]	Prigent, C., E. Jaumouille, F. Chevallier, and F. Aires, A parameterization of the microwave land surface emissivity between 19 and 100 GHz, anchored to satellite-derived estimates, IEEE TGRS, 46, 344-352, 2008
[RD-46]	Rossow, W. B., and R. A. Schiffer, Advances in understanding clouds from ISCCP, Bull. Am. Meteorol. Soc., 80(11), 2261–2287, 1999.
[RD-47]	Prigent, C., F. Aires, and W. B. Rossow, Land surface skin temperatures from a combined analysis of microwave and infrared satellite observations for an all-weather evaluation of the differences between air and skin temperatures, Journal of Geophysical Research, 108, 4310-4321, 2003
[RD-48]	Prigent, C., F. Aires, and W. B. Rossow (2003b), Retrieval of surface and atmospheric geophysical variables over snow from microwave satellite observations, J. Appl. Meteorol., 42, 368–380, doi:10.1175/1520-0450.

Id	Reference
[RD-49]	Catherinot, J., C. Prigent, R. Maurer, F. Papa, C. Jimenez, F. Aires, and W. B. Rossow (2011), Evaluation of “all-weather” microwave-derived land surface temperatures with in situ CEOP measurements, <i>J. Geophys. Res.</i> , 116, D23105, doi:10.1029/2011JD016439.
[RD-50]	Fennig, K., Schroder, M., and Hollmann, R.: Fundamental Climate Data Record of Microwave Imager Radiances, Edition 3, 10.5676/EUM SAF CM/ FCDR MWI/V003, 2017.
[RD-51]	Prigent, C., Jimenez, C, and P. Bousquet, Satellite-derived global surface water extent and dynamics over the last 25 years, <i>J. Geophys. Res.</i> , in review.
[RD-52]	Prigent, C., I. Tegen, F. Aires, B. Marticorena, and M. Zribi (2005), Estimation of the aerodynamic roughness length in arid and semi-arid regions over the globe with the ERS scatterometer, <i>J. Geophys. Res.</i> ,110, D09205, doi:10.1029/2004JD005370.
[RD-53]	Favrichon, S., Prigent, C., Jimenez, C., and Aires, F.: Detecting cloud contamination in passive microwave satellite measurements over land, <i>Atmos. Meas. Tech.</i> , 12, 1531-1543, <a href="https://doi.org/10.5194/amt-12-1531-2019">https://doi.org/10.5194/amt-12-1531-2019</a> , 2019.
[RD-54]	Soria, G. and J.A. Sobrino, ENVISAT/AATSR derived land surface temperature over a heterogeneous region. <i>Remote Sensing of Environment</i> , 2007. 111(4): p. 409-422.
[RD-55]	Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkman, J., Rota, S., & Ratier, A. (2002a). An introduction to Meteosat Second generation (MSG). <i>Bulletin of the American Meteorological Society</i> , 83, 977-992
[RD-56]	Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkman, J., Rota, S., & Ratier, A. (2002b). Radiometric performance of SEVIRI. <i>Bulletin of the American Meteorological Society</i> , 83, ES50-ES51
[RD-57]	Sembhi, H., and Ghent, D. (2017). University of Leicester Thermal Infrared Probabilistic Cloud Detetion for Land Algorithm Theoretical Basis Document for Sentinel-3, ESA document, S3-ATBD_UOL_CLOUD_V3
[RD-58]	Bulgin, C. E., Sembhi, H., Ghent, D., Remedios, J. J. and Merchant, C. (2014) Cloud clearing techniques over land for land surface temperature retrieval from the Advanced Along Track Scanning Radiometer. <i>International Journal of Remote Sensing</i> , 35 (10). pp. 3594-3615. ISSN 0143-1161
[RD-59]	Ghent, D. 2012. “Land Surface Temperature Validation and Algorithm Verification.” ESA Report, pp.1-17.
[RD-60]	Seemann, S.W., Borbas, E.E., Knuteson, R.O., Stephenson, G.R., & Huang, H.L. (2008). Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multispectral satellite radiance

Id	Reference
	measurements. <i>Journal of Applied Meteorology and Climatology</i> , 47, 108-123
[RD-61]	Steinke, S., S. Eikenberg, U. Löhnert, G. Dick, D. Klocke, P. Di Girolamo, and S. Crewell (2015), Assessment of small-scale integrated water vapour variability during HOPE, <i>Atmospheric Chemistry and Physics</i> , 15(5), 2675-2692.
[RD-62]	Vogelmann, H., R. Sussmann, T. Trickl, and A. Reichert (2015), Spatiotemporal variability of water vapor investigated using lidar and FTIR vertical soundings above the Zugspitze, <i>Atmospheric Chemistry and Physics</i> , 15, 3135–3148, doi:10.5194/acp-15-3135-2015.
[RD-63]	Berk, A., G.P. Anderson, P.K. Acharya, J.H. Chetwynd, L.S. Bernstein, E.P. Shettle, M.W. Matthew, and S.M. Alder-Golden, 2000: MODTRAN4 Version 2 User's Manual Air Force Res. Lab., Space Vehicles Directorate, Air Force Material Command, Hanscom AFB, MA.
[RD-64]	Borbas, E., S. W. Seemann, H.-L. Huang, J. Li, and W. P. Menzel, 2005: Global profile training database for satellite regression retrievals with estimates of skin temperature and emissivity. Proc. of the Int. ATOVS Study Conference-XIV, Beijing, China, 25-31 May 2005, pp763-770, 2005.
[RD-65]	Martins, J. P. A., I. F. Trigo, V. A. Bento, and C. da Camara, 2016: A Physically Constrained Calibration Database for Land Surface Temperature Using Infrared Retrieval Algorithms, <i>Remote Sensing</i> , 8, 808; doi:10.3390/rs8100808.
[RD-66]	Trigo, I. F., Peres, L. F., Dacamara, C. C. & Freitas, S. C. Thermal land surface emissivity retrieved from SEVIRI/Meteosat. <i>IEEE Trans. Geosci. Remote Sens.</i> <b>46</b> , 307–315 (2008).
[RD-67]	Dee, Dick P., et al. "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system." <i>Quarterly Journal of the royal meteorological society</i> 137.656 (2011): 553-597.
[RD-68]	LST CCI (2019) Product Validation and Algorithm Selection Report, Reference LST-CCI-D2.1-PVASR
[RD-69]	LST CCI (2019) End-to-End ECV Uncertainty Budget Report, Reference LST-CCI-D2.3-E3UB
[RD-70]	Wan, Z. MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD), (1999), Reference: NAS5-31370
[RD-71]	Embury, O., Merchant, C. J., & Corlett, G. K. (2012). A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability effects. <i>Remote Sensing of Environment</i> , 116, 62–78. <a href="https://doi.org/10.1016/j.rse.2011.02.028">https://doi.org/10.1016/j.rse.2011.02.028</a>

Id	Reference
[RD-72]	Merchant, C.J.; Harris, A.R.; Maturi, E.; MacCallum, S. Probabilistic physically-based cloud screening of satellite infra-red imagery for operational sea surface temperature retrieval. Q. J. R. Meteorol. Soc. <b>2005</b> , 131, 2735–2755
[RD-73]	T. R. Loveland, B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang & J. W. Merchant (2000) Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, International Journal of Remote Sensing, 21:6-7, 1303-1330, DOI: 10.1080/014311600210191
[RD-74]	Istomina, L. G., von Hoyningen-Huene, W., Kokhanovsky, A. A., & Burrows, J. P. (2010). The detection of cloud-free snow-covered areas using AATSR measurements. Atmospheric Measurement Techniques, 3, 1005–1017. <a href="https://doi.org/10.5194/amt-3-1005-2010">https://doi.org/10.5194/amt-3-1005-2010</a>
[RD-75]	W. C. Snyder and Z.Wan, (2005). BRDF models to predict spectral reflectance and emissivity in the thermal infrared. IEEE Transactions Geosci. Remote Sens., 36(1), 214–225
[RD-76]	Duguay-Tetzlaff, A., Bento, V. A., Göttsche, F. M., Stöckli, R., Martins, J. P. A., Trigo, I., Olesen, F., Bojanowski, J. S., da Camara C., and Kunz, H. (2015) Meteosat land surface temperature climate data record: Achievable accuracy and potential uncertainties, Remote Sensing
[RD-77]	Freitas, S., Trigo, I., Macedo, J., Barroso, C., Silva, R., and Perdigao, R. (2013) Land surface temperature from multiple geostationary satellites, International Journal of Remote Sensing, 34
[RD-78]	F. Aires and V. Pellet, "Estimating Retrieval Errors From Neural Network Inversion Schemes--Application to the Retrieval of Temperature Profiles From IASI," in IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2020.3026944.
[RD-79]	G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," Psychometrika, vol. 50, no. 2, pp. 159–179, Jun. 1985.
[RD-80]	WMO (2012). Guidelines on Ensemble Prediction Systems and Forecasting. WMO-No.1091
[RD-81]	Perry, M., Ghent, D., Jimenez, C., Dodd, E., Ermida, S., Trigo, I. F., and Veal, K. (2020). Multi-Sensor thermal infrared and microwave land surface temperature algorithm intercomparison. Remote Sensing, 12(24), 4164

## 2.3. Glossary

The following terms have been used in this report with the meanings shown.

Term	Definition
ATSR	Along-Track Scanning Radiometer
ATSR-2	Along-Track Scanning Radiometer-2
AATSR	Advanced Along-Track Scanning Radiometer
ALB2	ATSR Land Biome Classification
ATBD	Algorithm Theoretical Basis Document
BT	Brightness Temperature
C3S	Copernicus Climate Change Service
CAMEL	Combined ASTER and MODIS Emissivity for Land
CCI	Climate Change Initiative
CDR	Climate Data Record
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
Envisat	Environmental Satellite
ERA5	ECMWF Re-analysis 5
ERS	European Remote-Sensing Satellite
ESA	European Space Agency
GEO	Geostationary Orbit
GSW	Generalised Split Window
IGBP	International Geosphere–Biosphere
ISRF	Instrument Spectral Response Function
LEO	Low Earth Orbit
LSE	Land Surface Emissivity
LST	Land Surface Temperature
LST_cci	ESA CCI on LST
MODIS	Moderate Resolution Imaging Spectroradiometer
MW	Microwave
NN	Neural-Network
NWC SAF	Satellite Application Facility on Support to Nowcasting & Very Short Range Forecasting
RTM	Radiative Transfer Model
RTTOV	Radiative Transfer for TOVS
SEVIRI	Spinning Enhanced Visible and InfraRed Imager

Term	Definition
SLSTR	Sea and Land Surface Temperature Radiometer
SSM/I	Special Sensor Microwave/Imager
SSMIS	Special Sensor Microwave Imager Sounder
SST	Sea Surface Temperature
SW	Split Window
TCWV	Total Column Water Vapour
TIR	Thermal Infrared
UOL	University of Leicester
VCM	Vegetation Cover Method

## 3. Sensor Descriptions

The algorithms described in this ATBD is used to derive LST\_cci products from the following low earth orbit (LEO) and geostationary orbit (GEO) sensors:

- ❖ ATSR-2 and AATSR
- ❖ SLSTR
- ❖ MODIS
- ❖ SEVIRI
- ❖ GOES
- ❖ MTSAT
- ❖ SSM/I and SSMIS

Descriptions and summary information for the satellite sensors noted above are provided in the following sections.

### 3.1. ATSR-2 and AATSR

---

The Along Track Scanning Radiometer (ATSR) series of instruments include ATSR-2 and AATSR (Advanced Along-Track Scanning Radiometer). These were launched on board European Space Agency (ESA) sun synchronous, polar orbiting satellites ERS-2 in April 1995, and Envisat (Environmental Satellite) in March 2002, respectively. The last of these instruments – AATSR – provided its final data on 8th April 2012. These ATSRs therefore provide approximately 17 years of data. Continuation of this sensor series occurred, albeit with a data gap, with the launch of the Sea and Land Surface Temperature Radiometer (SLSTR) sensors on board Sentinel-3 satellites (see Section 3.2).

All ATSR instruments used similar orbits and equator crossing times ensuring a high level of consistency. With a swath width of 512km, AATSR is able to provide approximately 3-day global LST coverage with a repeat cycle of 35 days. The overpass of AATSR is 10:00 (local solar time) in its descending node and 22:00 (local solar time) in its ascending node. For ATSR-2 the overpass times are 10:30 and 22:30 in the descending and ascending nodes respectively. The orbit of the ATSRs was very stable in local crossing times and no notable orbital drifts occurred.

AATSR has good radiometric accuracy of less than 0.1 K in the mid-range of surface temperatures for both 11 and 12  $\mu\text{m}$  brightness temperatures (once a correction of order 0.2 K is applied to 12  $\mu\text{m}$  brightness temperatures [RD-19]), based on two blackbodies scanned on each scan cycle for calibration and using Stirling Cycle coolers to maintain the infrared detectors at low noise. All three ATSRs have similar specifications with near-infrared (NIR) / infrared (IR) channels at 1.6, 3.7, 11 and 12  $\mu\text{m}$ . Both ATSR-2 and AATSR have three additional visible channels at 0.55, 0.66 and 0.87  $\mu\text{m}$  for extending the application of ATSR data into the land domain. A distinguishing feature of the ATSRs was the dual-angle (DA) capability (nadir and forward at an angle of  $\sim 55^\circ$  to nadir). However, only the nadir view is generally utilised in LST retrievals, LST\_cci included. The rationale on the use of the nadir view only is provided in [RD-54] which assessed both SW and DA over topographically flat and homogeneous rice fields and found DA algorithms to be less accurate. Further information on the algorithm to be used for ATSR-2 and AATSR LST retrievals in LST\_cci is given in Section 4.

### 3.2. SLSTR

---

The Sea and Land Surface Temperature Radiometer (SLSTR) –which is based on the principles of AATSR – on board the Sentinel satellites 3-A and 3-B comprises a space element of Copernicus programme. This responds to the requirements for an operational and near-real-time monitoring of the Earth surface over a period of 15 to 20 years. Sentinel-3A was launched on 16<sup>th</sup> February 2016, and Sentinel-3B was launched on 25<sup>th</sup> April 2017.

SLSTR is designed to retrieve global sea-surface temperatures to an accuracy of better than 0.3 K and global land surface temperature to an accuracy of less than 1 K. Like AATSR a dual view capability is maintained with SLSTR - the nadir swath being 1420 km, and the backward view having a swath width of 750 km. This supports a maximum revisit time of 4 days in dual view and 1 day in single view. There are nine spectral channels including two additional bands optimised for fire monitoring and improved cloud detection. The spatial resolution of SLSTR is 500 m in the visible and shortwave infrared channels and 1 km in the thermal infrared channels. The baseline retrieval for the operational ESA SLSTR LST product consists of a nadir-only split-window algorithm with classes of coefficients for each land cover-diurnal (day/night) combination.

### 3.3. MODIS

---

MODIS (Moderate Resolution Imaging Spectroradiometer) instruments were launched on board two sun-synchronous, near-polar orbiting satellites Terra (EOS AM-1) launched on 18 December 1999 and Aqua (EOS PM-1) launched on 4 May 2002, respectively. Each instrument provides a pair of observations each day acquiring data in 36 spectral bands. Terra-MODIS acquires data at approximately 10:30am (local solar time) in its descending node and at approximately 10:30pm (local solar time) in its ascending node; while Aqua-MODIS observes the Earth at approximately 1:30pm (local solar time) in its ascending node; and at approximately 1:30am (local solar time) in its descending node. The swath width of these instruments, 2330km, enables these satellites to view almost the entire surface of the Earth every day. The spatial resolution of the thermal bands is 1 km; with both land surface temperature and land surface emissivity being core products from these instruments.

### 3.4. SEVIRI

---

The Spinning Enhanced Visible and Infrared Imager (SEVIRI) is the main sensor on board Meteosat Second Generation (MSG), a series of 4 geostationary satellites operated by EUMETSAT. SEVIRI was designed to observe an earth disk over Africa, most of Europe and part of South America with a temporal sampling of 15 minutes. Satellite view angles for SEVIRI range from 0° to 80°. The first MSG satellite was launched in August 2002, and operational observations are available since January 2004.

SEVIRI spectral characteristics and accuracy, with 12 channels covering the visible to the infrared [RD-55, RD-56] were unique among sensors on board geostationary platforms, for several years since the launch of MSG-1. The High Resolution Visible (Table 5) channel provides measurements with a 1 km sampling distance at the sub-satellite point (SSP); for the remaining channels the spatial resolution is reduced to 3 km at SSP. Level 1.5 data are disseminated to users after being rectified to 0° longitude, which means the satellite viewing geometry varies slightly with the acquisition time (satellite zenith angles typically differ by less than 0.25° between consecutive observations).

**Table 1: Characteristics of the SEVIRI instruments onboard Meteosat Second Generation.**

Channel	Central wavelength $\lambda_{\text{central}}$ (m)	Dynamic Range	Radiometric Noise
VIS0.6	0.635	533 Wm <sup>-2</sup> sr <sup>-1</sup> μm <sup>-1</sup>	S/N 10 at 1% albedo
VIS0.8	0.81	357 Wm <sup>-2</sup> sr <sup>-1</sup> μm <sup>-1</sup>	S/N 7 at 1% albedo
NIR1.6	1.64	75 Wm <sup>-2</sup> sr <sup>-1</sup> μm <sup>-1</sup>	S/N 3 at 1% albedo
IR3.9	3.92	335 K	0.35K at 300K
WV6.2	6.25	300 K	0.75K at 250K
WV7.3	7.35	300 K	0.75K at 250K
IR8.7	8.70	300 K	0.28K at 300K
IR9.7	9.66	310 K	1.50K at 255K
IR10.8	10.80	335 K	0.25K at 300K
IR12.0	12.00	335 K	0.37K at 300K
IR13.4	13.40	300 K	1.80K at 270K
HRV	Broadband (about 0.4 – 1.1)	460 Wm <sup>-2</sup> sr <sup>-1</sup> μm <sup>-1</sup>	S/N 1.2 at 0.3% albedo

### 3.5. GOES

The GOES series are part of the National Ocean and Atmosphere Administration (NOAA) earth observation mission and provide coverage over the American continent. GOES-12 was launched in July 2001 and GOES-13 in May 2006. GOES-13 replaced GOES-12 as the operational satellite in April 2010. The operational location of the satellites renamed GOES-East (GOES-E) is 75°W. In contrast to SEVIRI, the Imager onboard the GOES satellites (from GOES-12 onwards, <http://www.oso.noaa.gov/goes/goes-calibration/change-channels.htm>) does not include two channels in the thermal atmospheric window of the spectrum, only the one centred at 10.8 μm. This channel has a spatial resolution of 4 km at SSP and a 3-hourly observation frequency for the full disk (hourly frequency being available only for North America).

NOAA's latest generation of GOES satellites, known as the GOES-R Series, were recently launched. GOES-16 of that series replaced its predecessor as GOES-E in December 2017. The GOES-R payload includes the Advanced Baseline Imager (ABI), which incorporates a wider range of channels in the thermal infrared split-window region of the spectrum, namely centred at 10.3, 11.2 and 12.3 μm. The spatial resolution of these channels is 2 km at SSP and the temporal resolution is 5-15 minutes.

The raw data from these satellites was obtained through the Copernicus Global Land Service.

### 3.6. MTSAT

The Multifunction Transport SATellite (MTSAT) series was operated by the Japanese Meteorological Association (JMA), providing coverage over the eastern part of Asia and over Australia. MTSAT-1R was launched in February 2005 and it was located at 140°E. MTSAT-2 was launched in February 2006 and it

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 15
--	---	--

was located at 145°E. MTSAT-2 replaced MTSAT-1R as the operational satellite in July 2010 and was in operations until December 2015.

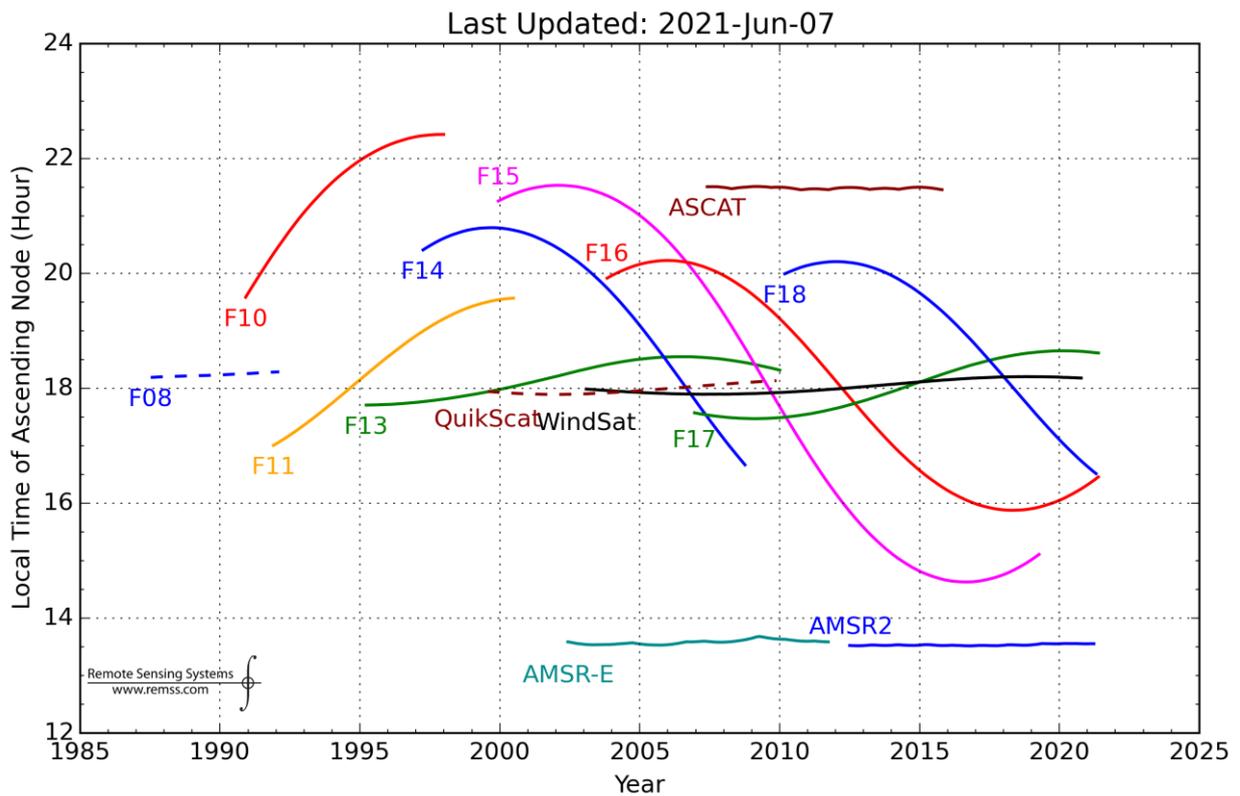
The Japanese Advanced Meteorological Imager (JAMI) onboard MTSAT-2 has one visible near-infrared (IR) channel (0.55-0.90  $\mu\text{m}$ ), two medium IR channels (3.5-4.0  $\mu\text{m}$  and 6.5- 7.0  $\mu\text{m}$ ), and two thermal IR (10.3-11.3  $\mu\text{m}$  and 11.5-12.5  $\mu\text{m}$ ). Unfortunately, the last channel was not provided to the Copernicus Global Land Service, and thus LST can only be retrieved based only on one TIR channel (centred around 10.8  $\mu\text{m}$ ). The spatial resolution of this TIR channel is 4 km at SSP, being available with a 3-hourly frequency of observation for MTSAT-1R and hourly for MTSAT-2.

The raw data from these satellites was obtained through the Copernicus Global Land Service.

### 3.7. SSM/I and SSMIS

---

The MW LST product is built using radiances observed by the Special Sensor Microwave/Imagers (SSM/I since 1987) and its more recent version the Special Sensor Microwave Imagers Sounder (SSMIS since 2003). This family of instruments flies on board Defense Meteorological Satellite Program (DMSP) near-polar orbiting satellites, and provides passive microwave observations twice a day at 19.35, 22.235, 37.0, and 85.5(SSM/I)/91.665(SSMIS) GHz with an incident angle of 53 degrees resulting in ground resolutions for SSM/I (SSMIS) of 69x43 (73x41), 50x40 (73x41), 37x28 (41x31), and 15x13 (14x13) km, respectively. Vertically and horizontally polarized BTs are available at all frequencies, apart from the 22.235 GHz channel, which is only vertically polarized. Instrument swath widths are close to 1400 (SSM/I) and 1700 (SSMIS) km, providing a 1-2 days revisiting time depending on acquisition latitude. The source of brightness temperatures are the Fundamental Climate Data Record of Microwave Imager Radiances [RD-50], where the brightness temperatures from the different SSM/I and SSMIS instruments have been inter-calibrated to reduce changes related to inter-sensor differences.



**Figure 3: Local ascending equator crossing time for SSM/I and SSMIS (courtesy of Remote Sensing Systems, <http://www.remss.com/support/crossing-times/>). Only the FXX instruments are relevant here.**

At a given time, a number of these instruments can be flying with slightly different local overpassing times, with a local time of the ascending node around 18 hours. Figure 3 (courtesy of Remote Sensing Systems) shows the local ascending equator crossing time. The variations in overpassing time are related to the original orbit injection of the DMSP satellites, and the subsequent orbital drift, which is not corrected during the lifetime of the missions. This means that when deriving climate data records (CDRs) from these observations diurnal cycle aliasing will be present, and need to be taken into account when using the CDRs. As the F13 SSM/I and F17 SSMIS are the instruments having the smaller drift, those are the instruments selected to produce the MW LST data record.

## 4. Retrieval of Land Surface Temperature from Thermal Infrared Sensors

The retrieval algorithms used in LST\_cci are either TIR or Microwave (MW) algorithms, which exploit different parts of the electromagnetic spectrum to derive LST.

In the LST\_cci open algorithm intercomparison round-robin, the performance of different LST retrieval algorithms for a set of specific thermal infrared and microwave satellite sensors was assessed to identify the best algorithms for a future climate quality operational system. The algorithms chosen for TIR were:

- ❖ the University of Leicester (UOL) Algorithm (TIR)
  - for the Advanced Along-Track Scanning Radiometer (AATSR) LST ECV dataset
  - for the AATSR / Sea and Land Surface Temperature Radiometer (SLSTR) / Moderate Resolution Imaging Spectroradiometer (MODIS) CDR
- ❖ the Generalised Split Window (GSW) Algorithm (TIR)
  - for the MODIS and Spinning Enhanced Visible and InfraRed Imager (SEVIRI) LST ECV datasets
  - for the Merged Dataset (AATSR / MODIS / SEVIRI)

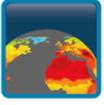
In the following section a description of TIR retrievals is presented, along with a description of the retrieval algorithms chosen for use in LST\_cci. The MW retrievals are described in Section 5.

### 4.1. Physics of the Problem

---

In the TIR region of the electromagnetic spectrum, absorption and emission effects, mainly due to the presence of water vapour, are responsible for attenuation of the surface signal observed by a satellite radiometer. As such, instruments on satellites designed for retrieval of the earth's surface temperature use spectral windows where absorption and emission effects are minimised and the surface emission signal is higher. Generally the 10.5 to 12.5  $\mu\text{m}$  window (commonly referred to in the literature as the split-window region) is used for LST retrieval purposes, rather than the 3.5 to 4.2  $\mu\text{m}$  window, as it is not as subject to solar signal effects. Even in this region of high transmission, correcting for atmospheric attenuation is still a necessity for accurate LST retrievals. Accurate LST retrievals also require algorithms which correct for emissivity effects.

Both of the TIR retrieval algorithms chosen for use in LST\_cci are so-called split-window (SW) algorithms, which utilise the radiances reaching the sensor in two channels whose band centres are close in wavelength. This SW method provides an estimate of the surface temperature from two brightness temperature measurements and assumes that the linearity of the relationship results from linearisation of the Planck function (which is generally a good assumption), and linearity of the variation of atmospheric transmittance with column water vapour amount as the most important trace gas (sometimes a poor approximation). For retrieval of LST, where the emissivity over land can be low and where emissivity varies significantly with surface cover and type, compared to Sea Surface Temperature retrievals over open ocean, the surface and atmosphere must be treated as a coupled system. There are two approaches to solving the problem of LST determination using the SW channels. The first assumes that the effects due to the land and atmosphere can be decoupled and the method is then to separate out the surface effects (emissivity) from the atmospheric effects (water vapour). The second approach is to accept that the surface and atmosphere are coupled, solve the problem without taking explicit account of either emissivity or water vapour, but to allow for their effects simultaneously. The difficulty of the first approach is that an estimate of the emissivity must be provided or retrieved and validated. Robust methods to

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 18
--	---	--

retrieve both LST and emissivity simultaneously require multiple thermal channels so are not feasible for instruments with only two channels in the 8-14 micron range. There is also a lack of global high spatial ( $\leq 1$  km) and temporal ( $\leq$  monthly) resolution emissivity datasets covering the entire range required for multi-decadal data records. These restrictions make the first approach very challenging and the recommendation is to follow the second approach.

The approach used in LST\_cci is the second approach which is outlined mathematically in the following section. Having established that there is a linear relation between the surface leaving radiance and the two SW radiances for the land, the problem is reduced to one of multiple, linear regression. The retrieval coefficients, derived by regression, have physical meaning and physical constraints can be utilised to ensure their validity. The temperature that is retrieved using the algorithm is a radiative surface temperature; it is appropriate for use as the temperature corresponding to the radiative flux density from the surface (i.e. Stefan-Boltzmann law). When used in modelling studies care must be taken to ensure that the model output temperature corresponds to the LST product definition below.

The definition of LST from TIR is the effective radiometric temperature of the Earth’s surface “skin” in the instrument field of view. “Skin” temperature here refers to the temperature of the top surface in bare soil conditions and to the effective emitting temperature of vegetation “canopies” as determined from a view of the top of a canopy. For mixed scenes skin temperature is the aggregated radiometric surface temperature of the ensemble of components within the sensor field of view.

#### 4.1.1. Mathematical description

The mathematical development of the problem of determining LST from a satellite radiometer with SW channels follows closely that of [RD-1], [RD-2] and [RD-3, RD-4, RD-5]. These papers show that under certain assumptions (these are introduced at each step below), it is possible to formulate the surface leaving radiance in terms of a linear combination of radiances reaching the satellite sensor in two channels close in their respective central wavebands.

The proposed LST products provide pixel by pixel LSTs using only the nadir SW (11 and 12  $\mu$ m) channels of the satellite instruments. The products utilise the cloud-free top-of-the-atmosphere 11 and 12  $\mu$ m brightness temperatures and ancillary information to correct for water vapour absorption and spectral emissivity effects. The product is generated using a regression relation and look-up tables that accommodate global and seasonal variations in the main perturbing influences. The mathematical basis for the formulation is provided here.

The starting point for any LST algorithm is a consideration of the thermal radiative transfer equation for monochromatic radiation emitted and reflected from a surface that is assumed homogenous, and received by a spaceborne radiometer. The homogeneous area is defined by the angular field-of-view of the radiometer. The radiance received at the satellite-borne radiometer may be written,

$$I_{\nu}(s) = \int_{\nu} \mathbf{F}_{\nu} \{ \tau_{\nu}(s) I_{\nu}^{\text{surface}}(s) + I_{\nu}^{\text{atmos}}(s) \} d\nu$$

$$I_{\nu}^{\text{surface}} = \epsilon_{\nu} B_{\nu}[T_s] + \frac{1}{\pi} \int_{\Omega_{-}}^{\text{sky}} \mathbf{n} \cdot \mathbf{s} \rho_{\nu}(s, s') I_{\nu} d\Omega$$

$$I_{\nu}^{atmos} = \int_0^{\infty} B_{\nu}[T(z)] \frac{\partial \tau}{\partial z}(z, \infty) dz$$

Where:

- ❖  $I_{\nu}$  is the radiance at the radiometer,
- ❖  $I_{\nu}^{surface}$  is the surface leaving radiance,
- ❖  $I_{\nu}^{atmos}$  is the radiance from the atmosphere,
- ❖  $\tau$  is the atmospheric transmittance,
- ❖  $\nu$  is wavenumber,
- ❖  $z$  is height
- ❖  $F$  is the filter response function of the radiometer,
- ❖  $s$  is a unit vector defining the view direction,
- ❖  $s'$  is a unit vector defining the sun's direction,
- ❖  $T_s$  is the surface temperature,
- ❖  $T$  is the atmospheric temperature,
- ❖  $\epsilon_{\nu}$  is the surface emissivity,
- ❖  $B_{\nu}$  is the Planck function,
- ❖  $\rho_{\nu}$  is the surface reflectance,
- ❖  $I_{\nu}^{sky}$  is the downwelling sky radiance.

If the surface is in thermodynamic equilibrium with the atmosphere, then according to Kirchhoff's law:

$$\int_{\Omega^-} n \cdot s \epsilon_{\nu}(s) d\Omega = \int_{\Omega^+} n \cdot s \left\{ 1 - \frac{1}{\pi} \int_{\Omega^-} n \cdot s' \rho_{\nu}(s, s') d\Omega' \right\} d\Omega$$

We assume that the surface is Lambertian. This assumption is valid since the Lambertian approximation of the surface reflection does not introduce a significant error in thermal infrared region [RD-3, RD-70]. Then  $\epsilon_{\nu}$  and  $\rho_{\nu}$  are independent of direction,

$$\epsilon_{\nu} = 1 - \rho_{\nu}$$

The flux density of sky radiation is:

$$F_{\nu}^{sky} = \int_0^{2\pi} \int_0^{\pi/2} I_{\nu}^{sky} \cos \theta \sin \theta d\theta d\phi$$

Where  $\theta$  is the satellite zenith view angle, and  $\phi$  is the satellite azimuth view angle.

$$I_{\nu}^{surface} = \epsilon_{\nu} B_{\nu}[T_s] + \{1 - \epsilon_{\nu}\} L_{\nu}^{sky}$$

$$L_{\nu}^{sky} = \frac{F_{\nu}^{sky}}{\pi}$$

This leads to the definition of surface temperature as sensed by a space-borne infrared radiometer:

$$T_s = B_v^{-1} \left\{ \frac{I_v^{surface} - (1 - \epsilon_v)L_v^{sky}}{\epsilon_v} \right\}$$

This definition has the attribute that  $T_s$  is directly measurable from space (e.g. from the AATSR), is valid at any scale, and for a homogeneous surface it is equivalent to the thermodynamic temperature.

The definition is only strictly true for monochromatic radiation. For sufficiently narrow channels ( $\approx 1 \mu\text{m}$  width) with relatively smooth filter response functions, the variation of the Planck function with wavenumber is small. Thus an integration of the various quantities ( $I_v$ ,  $\epsilon_v$ ,  $L_v^{sky}$ , etc.) over the filter function is appropriate.

The definition is only strictly valid under the assumptions outlined above over typical surface temperature ranges. Under most circumstances we expect the assumptions to remain valid and violation are weak so that the definition (and hence derivation of the surface temperature) is approximately correct.

Determination of the quantities in (9) can be done by various means. The approach we have taken follows [RD-4] and [RD-5], and shows that the surface temperature may be written as a regression relation involving the brightness temperatures in the 11 and 12  $\mu\text{m}$  channels. The relation takes account of atmospheric absorption (water vapour) and spectral emissivity effects.

#### 4.1.2. Emissivity

It is well-known that variations in surface properties cause variations in the emission of radiation from natural surfaces and this complicates LST retrieval. One major source of variation is due to the structural properties of the surface and this affects the efficiency of emission and reflection of thermal radiation from the surface.

There are substantial variations in surface emissivity, which is unitless, over the globe. The lowest values occur in sandy regions where the emissivity may be as low as 0.92 at 11  $\mu\text{m}$  [RD-6]. Over highly vegetated surfaces (e.g. closed-canopy trees) the emissivity is known to be spectrally uniform and high ( $\epsilon_{11} > 0.98$ , e.g. [RD-7]). Within a particular surface type the variation of emissivity is not well known, but measurements suggest it is small  $\approx \pm 0.01$ , except when structural changes occur as in senescent vegetation. Thus the greatest concern for deriving LSTs is the variation between surface types rather than the variation within surface types.

The scheme for accounting for emissivity variations between surface types relies on a surrogate measure of the surface structure; in this case we have used fractional vegetation cover and vegetation type. [RD-8] suggests using a classification based emissivity system for MODIS LST products. Their system uses 17 IGBP 'static' land cover classes. This is applicable to Land Cover CCI (LC\_cci) by way of transfer functions from one classification system to the other. Also of concern is the directional variation of emissivity. Generally, the variation is strongest with view angles greater than 50° or so. Little is known of the variation with azimuth angle, although over real surfaces emissivity angular effects are likely to be associated with induced changes in the actually observed scene.

While it is important to note the role that emissivity plays in determining the emission and reflection of thermal radiation from the land surface, it must be stressed that few field measurements of emissivity at scales appropriate to (for example) the AATSR pixel size have been made, although increasingly becoming available. Thus while it is possible to retrieve an emissivity from thermal satellite measurements, its

validation is problematic. Moreover, none of the emissivity schemes proposed can claim accuracies better than  $\pm 0.02$ . It is likely that the retrieval errors and biases are re-mapped from atmospheric transmittance errors, since the radiative transfer problem shows that the surface emissivity and atmospheric transmittance always appear as a product. Separating their effects accurately suggests that the atmospheric transmittance must be known at least to the same accuracy.

These factors should be borne in mind in considering the derivation of the LST algorithm itself in the following sections. Details of the auxiliary emissivity products used in LST\_cci are given in Section 5.4.1.

## 4.2. Algorithm Descriptions

### 4.2.1. Split-Window (SW) Approximation

By utilising the mean value theorem, it can be shown that as in [RD-2]:

$$\bar{I}_v^{atmos} = \frac{1}{1 - \tau_v} \int_0^\infty B_v[T(z)] \frac{\partial \tau(z, z')}{\partial z} dz$$

Where:

- ❖  $T(z)$  is the atmospheric temperature profile,
- ❖  $z$  is height,
- ❖  $\tau(z, z')$  is the transmittance profile between two heights.

The transmittance may be written,

$$\tau(z, \infty) = \exp \left\{ - \int_z^\infty k_v(z') w(z') \sec \theta dz' \right\}$$

Where:

- ❖  $k_v$  is the absorption coefficient,
- ❖  $w(z)$  is the vertical profile of the absorber amount.

This leads directly to the SW formulation. Consider two wavelengths (e.g. SLSTR 11 and 12  $\mu\text{m}$  channels and introduce appropriate subscripts):

$$I_{11} = [\epsilon_{11} B_{11}(T_s) + (1 - \epsilon_{11}) L_{11}] \tau_{11} + (1 - \tau_{11}) \bar{I}_{11}^{atmos}$$

$$I_{12} = [\epsilon_{12} B_{12}(T_s) + (1 - \epsilon_{12}) L_{12}] \tau_{12} + (1 - \tau_{12}) \bar{I}_{12}^{atmos}$$

Linearise around  $\nu_{11}$  and then manipulate:

$$B(\nu, T) = B(\nu_{11}, T) + \left( \frac{\partial B}{\partial \nu} \right)_{\nu_{11}} (\nu - \nu_{11})$$

$$B(T_s) = \left[ \frac{1 + \gamma}{\epsilon_{11} + \gamma \tau_{12} \Delta \epsilon} \right] I_{11} - \left[ \frac{\gamma}{\epsilon_{12} + (1 + \gamma) \tau_{11} \Delta \epsilon} \right] I'_{11} + \alpha$$

Where:

$$\diamond \gamma = \frac{1-\tau_{11}}{\tau_{11}-\tau_{12}},$$

$$\diamond \Delta\epsilon = \epsilon_{11} - \epsilon_{12},$$

$$\diamond \alpha = -\frac{(1-\tau_{11})\tau_{12}(1-\epsilon_{12})I_{12}-(1-\tau_{12})\tau_{11}(1-\epsilon_{11})I_{11}}{\epsilon_{12}\tau_{12}(1-\tau_{11})-\epsilon_{11}\tau_{11}(1-\tau_{12})}.$$

$\diamond I'_{11}$  is the radiance at  $\nu = \nu_{11}$  that yields a temperature equal to  $T_{12}$ . Thus,

$$\diamond I'_{11} = B_{11}[T_{12}].$$

Below are some special cases which are worth considering:

$\diamond$  No spectral emissivity dependence:

$$\Delta\epsilon = 0$$

$$B_s = \frac{1+\gamma}{\epsilon} I_{11} - \frac{\gamma}{\epsilon} I'_{11} + \alpha$$

$\diamond$  Emissivity  $\approx 1$  (e.g. sea surface):

$$B_s = (1+\gamma)I_{11} - \gamma I'_{11}$$

By linearising the Planck function about a mean atmospheric temperature, the algorithm can be formulated in terms of brightness temperatures.

$$B(\nu, T) = B(\nu, \bar{T}) + \left(\frac{\partial B}{\partial T}\right)_{\bar{T}} (T - \bar{T})$$

After some manipulation,

$$LST = a_0 + b_0 T_{11} + c_0 T_{12}$$

$$\diamond a_0 = \alpha \left(\frac{\partial B}{\partial T}\right)_{\bar{T}}^{-1},$$

$$\diamond b_0 = \frac{1+\gamma}{\epsilon_{11}} \left[ \frac{1}{1+\gamma\tau_{12}\Delta\epsilon/\epsilon_{11}} \right],$$

$$\diamond c_0 = \frac{\gamma}{\epsilon_{12}} \left[ \frac{1}{1+(1+\gamma)\tau_{11}\Delta\epsilon/\epsilon_{12}} \right].$$

This mathematical development shows that under the assumptions highlighted at each step in the process it is possible to relate the brightness temperatures in the 11 and 12  $\mu\text{m}$  channels linearly to the land surface temperature.

Although  $\epsilon_{11}$  and  $\epsilon_{12}$  are non-unity for land surface emissivities, from a radiative transfer point-of-view, they are sufficiently close to one for the approach to be appropriate.

#### 4.2.2. UOL Algorithm

The UOL algorithm is used in LST\_cci for the ATSR and SLSTR series LST ECV datasets and the AATSR / SLSTR / MODIS CDR. Below is the description of this algorithm as provided in the PVASR [RD-68].

The standard algorithm ([RD-9], for (A)ATSR and SLSTR) uses a nadir-only (SW) algorithm with classes of coefficients for each combination of land cover-diurnal (day/night) condition. The physics in principle are the same as for other SW algorithms, such as [RD-10] which also applies coefficients to a combination of emissivity, water vapour and BT differences. For [RD-10] non-linearity is accounted for in the quadratic term, where here it is parametrised across the swath. The full form of the algorithm is presented as follows:

$$LST = d(\sec(\theta) - 1)pw + (fa_{v,i} + (1 - f)a_{s,i}) + (fb_{v,i} + (1 - f)b_{s,i})(T_{11} - T_{12})^{1 / (\cos(\theta / m))} + ((fb_{v,i} + (1 - f)b_{s,i}) + (fc_{v,i} + (1 - f)c_{s,i}))T_{12}$$

where the six retrieval coefficients  $a_{s,i}$ ,  $a_{v,i}$ ,  $b_{s,i}$ ,  $b_{v,i}$ ,  $c_{s,i}$  and  $c_{v,i}$  are dependent on the land cover (i), fractional vegetation cover (f) - the retrieval coefficients  $a_{s,i}$ ,  $b_{s,i}$  and  $c_{s,i}$  relate to bare soil (f = 0) conditions, and  $a_{v,i}$ ,  $b_{v,i}$  and  $c_{v,i}$  relate to fully vegetated (f = 1) conditions. The fractional vegetation cover (f) and precipitable water (pw) are seasonally dependent whereas the land cover (i) is invariant [RD-11].

The retrieval parameters d and m are empirically determined from validation and control the behaviour of the algorithm for each zenith viewing angle ( $\theta$ ) across the nadir swath. The parameter d resolves increases in atmospheric attenuation as the water vapour increases, an effect that is accentuated with increasing zenith viewing angle. The parameter m is supported by previous studies [RD-11], which suggest a non-linear dependence term on the BT difference T11 - T12 would elicit improvement in the accuracy of the LST retrievals. The rationale here is that the BT difference increases with increasing atmospheric water vapour, since attenuation due to water vapour is greater at 12  $\mu$ m than at 11  $\mu$ m.

The nature of the algorithm means that land surface emissivity is implicitly dealt with through the regression of retrieval coefficients to land cover and bare soil / fully vegetated states. In other words, while LSE is not an estimated output the algorithm still uses LSE knowledge, any uncertainty of which is propagated in the LST derivation. This knowledge is passed to the algorithm through the land cover and fractional vegetation states, which themselves are regressed to emissivity states in the coefficient generation. Dynamic Fractional Vegetation Cover (FVC) ancillary data are retrieved from auxiliary data.

For the generation of the retrieval coefficients for each land cover–diurnal (day/night) combination vertical atmospheric profiles of temperature, ozone, and water vapour, surface and near-surface conditions and the surface emissivities are required. These are input, in addition to specifying the spectral response functions of the instrument, into a radiative transfer model in order to simulate TOA BTs. Retrieval coefficients are determined by minimizing the l2-norm (Euclidean norm) - which is calculated as the Euclidean distance from the origin - of the model fitting error ( $\Delta$ LST).

#### 4.2.3. Generalised Split Window (GSW) Algorithm

The GSW algorithm is used in LST\_cci for the MODIS and SEVIRI LST ECV datasets as well as the Merged Dataset (AATSR / MODIS / SEVIRI). It is also used for the GOES-16 LST ECV dataset. The generalised split window algorithm is a view-angle dependent split-window algorithm proposed for LST retrieval by [RD-12]. It is based around channels in the 11 and 12  $\mu$ m regions.

The success of the generalized split-window LST algorithm depends on knowledge of the band emissivities for real land surfaces. In the LST\_cci GSW method, emissivity information is used explicitly rather than incorporating this information implicitly through land cover coefficients. For example, in the operational MODIS implementation band averaged emissivities for each of the two channels are used:

$$\varepsilon_i = \frac{\int_{\lambda_1}^{\lambda_2} \psi(\lambda) \varepsilon(\lambda) B(\lambda, T_s) d\lambda}{\int_{\lambda_1}^{\lambda_2} \psi(\lambda) B(\lambda, T_s) d\lambda}$$

Where  $\lambda_2$  and  $\lambda_1$  are the upper and lower bounds of the channel, and  $T_s$  is the surface temperature. This parameter is assigned on a pixel basis according to land cover class. In cases of mixed pixels this term is recalculated based upon the proportion of the pixel assigned to each classification. A similar method is used in LST\_cci. Here,  $\varepsilon_{mean}$  is derived as:

$$\varepsilon_{mean} = 0.5 (\varepsilon_{11} + \varepsilon_{12})$$

Where  $\varepsilon_{mean}$  is the mean emissivity of the two thermal channels used in the GSW algorithm.  $\Delta\varepsilon$  is the difference between the two thermal channels, calculated as:

$$\Delta\varepsilon = \varepsilon_{11} - \varepsilon_{12}$$

Having determined the emissivity of the pixel coefficients these can be applied to derive an LST estimate similar to that given below:

$$T_s = C + \left( A_1 + A_2 \frac{1 - \varepsilon_{mean}}{\varepsilon_{mean}} + A_3 \frac{\Delta\varepsilon}{\varepsilon_{mean}^2} \right) \frac{T_1 + T_2}{2} + \left( B_1 + B_2 \frac{1 - \varepsilon_{mean}}{\varepsilon_{mean}} + B_3 \frac{\Delta\varepsilon}{\varepsilon_{mean}^2} \right) \frac{T_1 - T_2}{2}$$

Where C, A and B are coefficients derived from linear regression using simulated data as done for the UOL algorithm (Section 4.2.2) but adapted for the GSW. T1 and T2 are the 11 and 12  $\mu\text{m}$  brightness temperatures. The coefficients for GSW are dependent on satellite viewing angle and water vapour. Error analysis [RD-70] shows that viewing angle and atmospheric column water vapour must be considered in the retrieval to achieve highest accuracy over the wide atmospheric and surface conditions. The bands for water vapour are of width 15  $\text{kg}\cdot\text{m}^{-2}$  so that the first water vapour band is from [0,15)  $\text{kg}\cdot\text{m}^{-2}$ . The bands for satellite zenith angle are of width 5°. The retrieval coefficients are linearly interpolated between viewing angle and water vapour bands to minimise step changes.

#### 4.2.4. Single-channel (SMW) Algorithm

For some single-sensor products (early GOES and MTSAT) only one thermal infrared channel can be used.

The specific algorithm chosen is that used by the CM SAF for MVIRI retrievals and has been successfully applied to SEVIRI with good results [RD-76] and is close to the formulation for GOES [RD-77]:

$$LST = A + \frac{T_x}{\varepsilon_x} + B \frac{1}{\varepsilon_x} + C$$

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 25
--	---	--

Where A, B and C are coefficients derived from linear regression using simulated data as done for the split window algorithms but adapted for the SMW.  $T_x$  is the single channel brightness temperature and  $\epsilon_x$  is the spectral emissivity for this channel.

The coefficients for SMW are dependent on satellite viewing angle and water vapour. The bands for water vapour are of width  $15 \text{ kg}\cdot\text{m}^{-2}$  so that the first water vapour band is from  $[0,15) \text{ kg}\cdot\text{m}^{-2}$ . The bands for satellite zenith angle are of width  $5^\circ$ .

### 4.3. Radiative Transfer Modelling

Radiative Transfer for TOVS (RTTOV) is a fast Radiative Transfer Model (RTM) from the NWP-SAF [RD-25]. It is an efficient radiative transfer forward model for the visible, infra-red and microwave wavelengths. In contrast to models using a line-by-line methodology, RTTOV conceptualizes the simulation in terms of channel radiances. It therefore requires both an Instrument Spectral Response Function (ISRF) and a pre-calculated set of coefficients relating the channel to sensitivities to various atmospheric parameters. These coefficients parameterize the gas contributions to transmittances associated with the profile. These requirements allow significantly increased computational speed in RTTOV compared to the line-by-line methodology [RD-26]. Yet this increase in computational speed leads to a reduction in the accuracy [RD-27], although negligible [RD-71], and spectral resolution [RD-28] of the simulated radiances. The choice of RTTOV facilitates fast processing of sufficient numbers of profiles to adequately characterize the entire range of potential atmospheric states representative of each land cover class for highest accuracy [RD-11]. RTTOV Version 12.3 is used in LST\_cci.

RTTOV is used in the UOL\_3 (Section 4.5.1) and Bayesian (Section 4.5.2) cloud masking algorithms to calculate the probability of cloud cover in the observations given the background state. Retrieval coefficients are derived using forward modelling. Specifically, regressions between the skin temperature and the TOA radiances are used to populate a Calibration Database for determining retrieval coefficients. RTTOV is also used in the threshold tests employed in the NWC-SAF cloud masking algorithm (Section 4.5.3).

### 4.4. Calibration Database for Determining Retrieval Coefficients for the TIR Algorithms

Globally robust, traceable retrieval coefficients for both the GSW and UOL approaches are generated using RTTOV, which allows fast processing of sufficient numbers of profiles to adequately characterise a wide range of potential atmospheric states representative of each land cover. Simulated brightness temperatures and LSTs are derived from RTTOV given inputs of vertical atmospheric profiles, surface and near-surface conditions, surface emissivities, and the spectral response function of the sensor of interest.

A calibration database consisting of a range of LST, Top-Of-Atmosphere Radiance, Atmospheric profiles and varied emissivity is required to both train and assess the different algorithms used with in LST\_cci. This database must be representative and reusable to provide consistency between both algorithms and consortium partners. Building on the Phase-1 Round Robin work the LST\_cci Calibration Database (CCICDB) is a much expanded database of surfaces and atmospheres designed to be the principle environment in which all thermal IR algorithms are trained and tested.

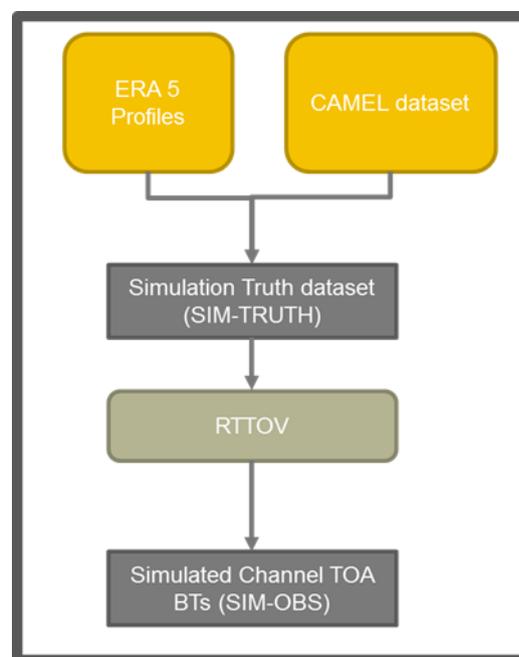
The use of simulated data necessitates the use of modelled atmospheric data and ancillary sources for emissivity, fractional vegetation and biome information. All of these inputs must be used with a common

forward model to result in useful and robust data at both the surface LST level and for the observed top of atmosphere brightness temperatures.

The calibration dataset is constructed from the following data sources:

- ❖ ECMWF Re-analysis 5 (ERA5) atmospheric profile data [RD-17]
- ❖ Combined ASTER and MODIS Emissivity for Land (CAMEL) emissivity data (v2) [RD-24]
- ❖ RTTOV Forward Model (v12.3) [RD-25]
- ❖ ULeic / LandCover\_cci (LCCS) hybrid biome map (Appendix A)

In order to construct a robust and representative dataset a rigorous procedure is developed to ensure the quality. The flow of input data and the significant processing steps are outlined in Figure 4.



**Figure 4: Flow chart for the generation of the simulated true Top-Of-Atmosphere Brightness Temperature values for the dataset.**

The process starts with the ERA5 profile data. The initial parameter considered is the atmosphere, primarily focussing on the water vapour as the most significant influence on the retrieval of LST. All selected profiles are drawn from cloud free data. The ERA5 data is gridded at 0.25° and contains both full water vapour profiles as well as the Total Column Water Vapour (TCWV) values. Averaging the full profile can result in an overly smoothed product, which might not be representative of real atmospheres. To avoid this the ERA5 data is gridded to cells of 2.5x2.5° and the mean TCWV values calculated. The mean total column water vapour is then compared to all the TCWV values in the 2.5x2.5° grid cell. This grid cell size is significantly finer than the resolution used in the Round Robin assessment [RD-81], and as such result in a higher number of points spatially. A further selection criterion involves only 10% of the points above 60°N and below 60°S are selected to avoid polar samples having an unrepresentative influence on the dataset.

The profile that corresponds to the closet matching TCWV value is selected as the representative profile for the grid cell. This methodology ensures that not only is the profile physically realistic but that it represents the most likely conditions for the majority of the grid cell, this is important when comparing

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 27
--	---	--

to other parameters such as biome and emissivity which are sourced from different datasets. Any variation in the profiles within the grid cell is explored in the sensitivity on water vapour. A climatology for each grid cell is derived and the standard deviations of the TCWV compared; the range of variation is in line with the variation seen in the selected dataset. All other information such as LST, elevation, and other gas profiles are also selected using the index of this profile.

The second stage of processing involves emissivity from either CAMEL or a Tool to Estimate Land Surface Emissivity in the Microwave (TELSEM) [RD-45, RD-38]. The latitude and longitude information at the locations selected from the ERA5 data is used to find the corresponding emissivity in both cases. In the case of CAMEL, the emissivity data is a higher resolution than the 0.25° ERA5 data. To address this the all emissivity data in a 0.25° box is averaged to get an emissivity value that is representative at the spatial resolution of the simulation. This does reduce the impact of the most extreme emissivity values, but results in values more representative of those expected at the 1 km resolution of the LEO sensors. All of the data is collated and the biome and fractional vegetation information added. All parameters are input to the RTTOV v12 forward model along with the sensor ISRF. RTTOV then simulates the BTs for each sensor for the given simulated truth.

The Calibration Database requires two datasets (“Training” and “Testing”) each with the capacity for perturbations of parameters to be representative. The multi-year data is sub-divided so that odd years are assigned to the “Training” partition of the Calibration Database and even years to the “Testing” partition. The “Training” set are the profiles and surface conditions used to develop the algorithm coefficients and train the retrievals. The “Testing” set is used to assess the performance and to derive robust uncertainty estimates for the retrievals. For both subsets an intra-annual temporal sampling method is based on three daily samples per month on the 5th, 15th and 25th day.

All thermal IR LST ECV Products (LEO, GEO and their resultant multi-sensor CDRs) are consistent with respect to coefficient generation.

#### 4.5. Identification of Observations Valid for Land Surface Temperature Estimation from Thermal Infrared Sensors

---

Cloud screening is a fundamental step for Thermal Infrared (TIR) LST retrieval. For LST\_cci products the cloud mask is given, or applied to, Level 2 and Level 3 LST products.

Traditionally, threshold based techniques have been used to detect cloud but these often fail under difficult circumstances -- for example, in the detection of thin cirrus or low-level fog. Three cloud detection algorithms which are used in LST\_cci are presented here:

- ❖ The UOL\_3 algorithm (Section 4.5.1).
- ❖ Bayesian algorithm (Section 4.5.2)
- ❖ the NWC SAF Cloud Mask Algorithm (Section 4.5.3)

The UOL\_3 Algorithm is expected to be applied to single sensor products produced from sensors on Low Earth Orbit (Low Earth Orbit) platforms. In LST\_cci Phase-1 the UOL\_3 algorithm is applied to the CDR. The NWC-SAF Cloud Mask Algorithm is applied to single sensor products produced from sensors on Geostationary (GEO) platforms. Updates will be made to this document as the algorithms are developed and the best algorithm for each LST\_cci product is identified.

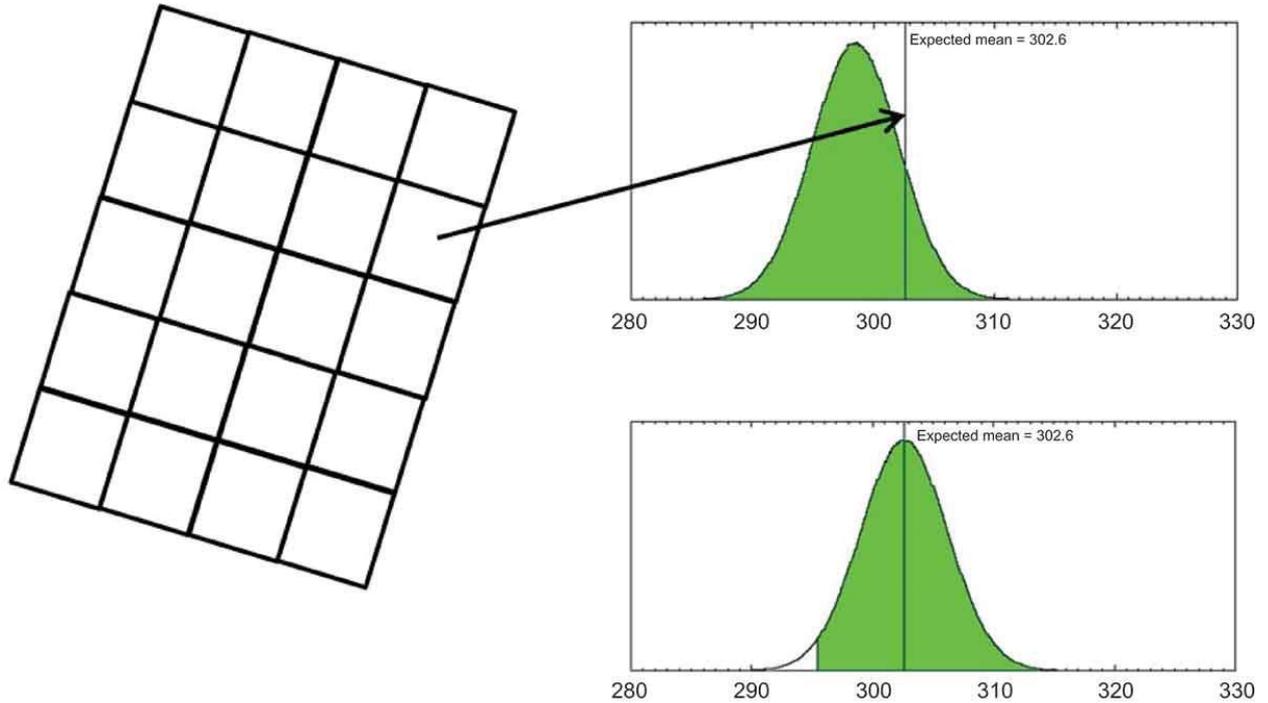
 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 28
--	---	--

#### 4.5.1. UOL\_3 Algorithm

The UOL\_3 algorithm is a semi-Bayesian cloud masking approach using the probability of clear-sky conditions which has been developed at University of Leicester [RD-19]. A pixel-level cloud mask is derived using a combination of simulated brightness temperatures and observational climatology. The approach is equally valid for both day and night-time retrievals as this method is independent of visible wavelength information. It has been implemented in the ESA DUE GlobTemperature project previously for ATSR data records [RD-19] and is being implemented operationally for SLSTR [RD-57].

This cloud masking algorithm uses atmospheric profile data to predict clear-sky conditions for the coincident space and time of a given satellite sensor observation. Coincident clear-sky brightness temperatures are derived by bilinear interpolation between surrounding ECMWF profile locations and a temporal interpolation between the 6-hourly analysis fields. ERA-Interim data [RD-67] is used for profiles in Phase-1 of LST\_cci, moving to ERA-5 data subsequently. The coincidence is modelled through bilinear interpolation of surrounding profiles and temporal interpolation between 6-hourly analysis fields. On a spatial plane these modelled profile data correspond to the tie-point grid of the respective instrument and orbit granules. For example, when applied to AATSR orbit granules, which are orbit subsets of pixels every 25 km across track and 32 km along track, are used [RD-19].

An observational climatology is acquired for each  $5 \times 5^\circ$  grid cell (chosen to ensure sufficient representiveness) for each of the land covers and diurnal conditions (day/night) required by an offline enhanced LST retrieval algorithm [RD-58, RD-59]. This is stratified by the 43 land covers of the ULeic / LandCover\_cci (LCCS) hybrid biome map (Appendix A). The mean and standard deviations for clear-sky conditions are stored in a LUT. Using RTTOV (see Section 4.5.1), expected clear-sky brightness temperatures / brightness temperature differences are simulated for these profile data. To calculate the clear-sky probability at each pixel location a probability density function (PDF) assuming a normal distribution is constructed from the simulated mean brightness temperatures for the corresponding granule and the standard deviation of the brightness temperature from the observational climatology from the corresponding  $5 \times 5^\circ$  grid cell for the given month, land cover and diurnal state as shown in Figure 5 [RD-58]. A per-pixel cloud mask is generated from comparing the pixel brightness temperatures/brightness temperature differences with the pixel probability density functions. Pixels are identified as cloudy if the combined probabilities are less than a set of confidence thresholds. For daytime observations, the cloud flag is set if either the observed  $12 \mu\text{m}$  brightness temperature or  $11 - 12 \mu\text{m}$  brightness temperature difference fall outside of the 95% confidence levels of the corresponding simulated PDFs. The thresholds themselves are simply for converting the probabilities into a binary mask. Users can choose to rather use the probabilities. For night-time observations, the  $12 \mu\text{m}$  brightness temperature and the  $11 - 3.7 \mu\text{m}$  differences are used. The actual tests relate to the optimum criteria in which clouds can be distinguished. For granules where insufficient profile data are available to simulate the expected brightness temperatures, or where incompatibilities between the atmospheric and surface states result in an RTTOV error (which is a rare occurrence) then the individual pixel cloud flags are instead derived from other cloud masking routines, specifically operational flags.



**Figure 5:** For each granule of an AATSR orbit (left), the expected  $12\ \mu\text{m}$  brightness temperature is simulated from coincident profiles. The PDF of observed  $12\ \mu\text{m}$  brightness temperatures for each land cover-diurnal condition, given the space and time position, is also determined (top-right in green). This PDF is moved so that the mean equals the expected mean for the granule and the new PDF represents the expected clear-sky conditions (bottom-right in green). Figure 1 from [RD-58].

#### 4.5.2. Bayesian Algorithm

The Bayesian cloud mask, which was developed at the University of Reading [RD-72], calculates the probability of clear-sky  $P(c|\mathbf{y}^o, \mathbf{x}_b)$  given the observation vector ( $\mathbf{y}^o$ ) and prior knowledge of the background state ( $\mathbf{x}_b$ ):

$$P(c|\mathbf{y}^o, \mathbf{x}_b) = \left[ 1 + \frac{P(\bar{c})P(\mathbf{y}^o|\mathbf{x}_b, \bar{c})}{P(c)P(\mathbf{y}^o|\mathbf{x}_b, c)} \right]^{-1}$$

Where  $\bar{c}$  and  $c$  denote cloud and clear conditions respectively. The prior probabilities of clear and cloudy conditions ( $P(c)$  and  $P(\bar{c})$ ) are defined using ECMWF ERA5.

$P(\mathbf{y}^o|\mathbf{x}_b)$  is the probability of the observations given the background state. For clear-sky observations this is calculated using the RTTOV version 12.3 fast forward model simulations, with cloud properties specified using an empirical PDF as these are computationally expensive to calculate. This cloud detection algorithm has been successfully applied to the Sea Surface Temperature (SST) CDR for ATSR instruments [RD-33; and RD-34] and for the AVHRR data record in Phase 2 of the SST CCI project. Previous work has applied these techniques to GOES instruments [RD-35; and RD-36] demonstrating its applicability to geostationary sensors. The mask is also used operationally for SST products from SLSTR SST and C3S. In LST\_cci this cloud mask is being investigated for adaptation to terrestrial surfaces. Such investigation includes optimised use of VNIR/SWIR information and appropriateness of *a priori* data.

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 30
--	---	--

### 4.5.3. NWCSAF Cloud Mask Algorithm

The cloud mask to be applied to all GEO single sensor products is the Satellite Application Facility on Support to Nowcasting & Very Short Range Forecasting (NWC SAF) cloud mask algorithm (also known as CMA). This cloud mask has been designed to be applicable to imagers on board meteorological geostationary satellites [RD-20]. It aims to support nowcasting applications as well as remote sensing of continental and oceanic surfaces, including identification of cloud free areas for LST products. This cloud mask algorithm also provides information on the presence of snow/sea ice, dust clouds and volcanic plumes [RD-20].

This algorithm is based on a series of satellite dependent threshold tests [RD-20]. The first step in the process aims to identify most pixels containing cloud and snow using a series of multispectral threshold tests based on factors such as viewing geometry, surface temperature and atmospheric water content (from Numerical Weather prediction fields), elevation, and climatological data. A second, optional step uses a smaller series of multispectral tests on thresholds computed from RTTOV applied on-line to NWP vertical profiles. This allows a more accurate threshold computation a detection of low or thin high clouds that remained undetected in the first set of tests. Then an analysis of the temporal variation in a short time period (around 15 minutes) of a combination of channels allows the detection of rapidly moving clouds. Fourthly, a specific treatment combining temporal coherency analysis and region growing technique allows the improvement of low clouds detection in twilight conditions. There is then another optional step which involves an analysis of solar channels at high spatial resolution to detect sub-pixel clouds inside pixels at default horizontal resolution. Finally, a spatial filtering is applied to cold areas, cloud edges (over ocean), isolated cloud pixel (land) and snow-area edge. For the additional information on dust clouds and ash clouds there are further processes to identify these features, which are applied to all pixels and stored in separate flags.

## 4.6. Auxiliary Datasets for Thermal Infrared Retrievals

---

The following section gives a description of the auxiliary datasets used for cloud detection and thermal infrared retrieval algorithms utilised in LST\_cci. This section also described how these auxiliary datasets are applied in each algorithm.

### 4.6.1. Land cover

Land cover information for LST\_cci products is provided by the ULeic / LandCover\_cci (LCCS) hybrid biome map. The baseline for this from Land Cover CCI (LC\_cci) are produced mainly from the MERIS FR time series, but also the MERIS RR dataset and SPOT Vegetation (SPOT-VGT) [RD-41]. Land cover maps in Land Cover CCI are derived using a classification model based on the GlobCover unsupervised classification chain. The processing chain was developed with the aim of being globally consistent, but also regionally tuned. In order to do this, the GlobCover processing chain was improved by including machine learning classification steps and developing a multi-year strategy [RD-41].

Previously the UOL retrieval scheme used the ATSR Land Surface Temperature Biome classification V2 (ALB-2) biome map. This takes account of the main bare soil classes where emissivity values are highly variable [RD-19], and is important for algorithms that have no explicit emissivity terms. The use of these additional biomes (20 to 25 in the ALB-2 scheme) ensures an algorithm is exposed to the full range of emissivity variation in the same manner as algorithms that directly assimilate emissivity.

For LST\_cci our aim is consistency across the CCI projects, while also retaining the most appropriate system for thermal infrared data. In that respect we modified the LC\_cci system to sub-divide the bare soil classes into distinct sub-classes based on soil taxonomy. As per the ALB-2 scheme, the rationale here is that emissivity variability is highest for different types of bare soil and for biome-based algorithms distinguishing between different bare soil types is crucial for improving robustness of the retrieval algorithms. The LCCS hybrid product merges the LC\_cci classification with the bare soil differentiation of the ALB-2. This maintains consistency across the CCI programme in the use of a single land cover dataset, while optimising its implementation in LST\_cci.

The LC\_cci dataset is gridded on a global 1/360°. Since this is higher spatial resolution than the majority of the LST\_cci sensors, the data was re-binned to 1/120° to match the ALB-2 dataset. The re-gridding was performed using a 2-dimensional histogram approach, first sub dividing the LC\_cci data into the 1/120° grid cells, and then finding the median land classification index for the LC\_cci pixels in each 1/120° cell. Bare soil values in the LC\_cci dataset were identified and the corresponding pixel in the ALB-2 data found. The replacement was done using the logic shown in Table 2.

**Table 2: Mapping logic for replacement of LC\_cci indexes: 200, 201 and 202 with ALB-2 bare soil information.**

		LC_cci pixel values		
		200	201	202
ALB-2 pixel values	20	200	201	202
	21	203	203	203
	22	204	204	204
	23	205	205	205
	24	206	206	206
	25	207	207	207

These new values 200 – 207 have been incorporated into the LCCS Hybrid scheme. The resulting map has been further re-gridded using the same 2-dimensional histogram technique as described previously to resolutions of 0.01° and 0.05° for ease of implementation in Level-3 LST\_cci products.

Overall the LCCS Hybrid map is consistent with the CCI programme in general while allowing for an extra dimension to support the specific needs of LST\_cci. Biome information is used in the UOL retrieval algorithm, in combination with other variables, to determine the most appropriate coefficients to apply. It is also used in the UOL\_3 cloud masking algorithm. Appendix A provides a table defining the LST\_cci Land Covers.

#### 4.6.2. Fractional Vegetation

Fractional vegetation cover information for LST\_cci is provided by the Copernicus Global Land Cover Services FCOVER dataset V2.0 (<https://land.copernicus.eu/global/products/fcover>). This global dataset is available at 1/112° resolution every 10 days from 1999 onwards. It is acquired using a moving temporal window of around 30 days [RD-21, RD-19]. FCOVER is generated from normalized nadir reflectances in the red, near-infrared, and shortwave infrared wavebands of SPOT-4 and SPOT-5 vegetation sensors using a neural network trained with the 1-D radiative transfer models SAIL and PROSPECT [RD-22]. Data values range from 0.0 (no vegetation or snow/water surface types) to 1.0 (full vegetation). Validation of this

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 32
--	---	--

product shows that it is good quality with a spatially consistent global distribution of retrievals [RD-23]. For use in LST\_cci processing, an FCOVER value is assigned to each sensor pixel via a nearest neighbour approach. For any pixel where no FCOVER values exist in a given 10-day window (either through missing or poor quality data) the pixel is filled from a climatology [RD-19]. The climatology is constructed from a complete temporal window of the same 10-day period across all years where the FCOVER dataset is available.

Fractional vegetation is used in the UOL retrieval algorithm, in combination with other variables, to weight the appropriate retrieval coefficients applied in the algorithm. For the SEVIRI product, fractional vegetation is used as follows.

Within the LSA-SAF, the fractional vegetation cover is used to derive TIR emissivity [RD-66] (see Section 4.6.3). FCOVER is important to weight the emissivities between bare soil and fully vegetated states. The same fundamental approach is employed for both the SEVIRI single-sensor product and the UOL algorithm.

### 4.6.3. Emissivity

#### CAMEL

The Combined ASTER and MODIS Emissivity for Land (CAMEL) database is a global monthly mean emissivity dataset spanning the years 2000 – 2016. A climatology of CAMEL data is used after 2016 until regular updates of this dataset become available. It assimilates both ASTER Global Emissivity Database retrieved values and University of Wisconsin-Madison MODIS Infra-red Emissivity dataset values. The CAMEL dataset contains 12 emissivity values at different wavelengths from 3.6 to 14.3  $\mu\text{m}$  at a resolution of 0.5° [RD-24]. Due to the dataset originating from satellite observations, it is highly relevant to realistic materials observed from space and should remove materials in spectral libraries, which are too fine a scale to be useful. The benchmark dataset and the retrievals testing in this study use wavelengths: 10.8, 11.3 and 12.1  $\mu\text{m}$ .

CAMEL emissivity data is used in the SW approximation algorithms (Section 4.2.1), explicitly in the case of the GSW algorithm (Section 4.2.3) to calculate LST. Emissivity from CAMEL is employed in a Calibration Database for determining retrieval coefficients for all SW approximation algorithms.

#### VCM

The Vegetation Cover method (VCM) uses pixel fraction of vegetation cover to derive Land Surface Emissivity (LSE) [RD-29]. In the VCM, LST is considered to be a combination of the emissivity from vegetation and bare ground across the land surface. The method for VCM is summarised below and can be found in detail in [RD-29] and [RD-31].

The vegetation and bare ground emissivities per channel are estimated for land classes within the International Geosphere–Biosphere Program (IGBP) database [RD-30]. For each IGBP class, the typical vegetation and bare ground components of that class are identified. Then laboratory spectral reflectances are used for the difference surface types within that class (taking into account sensor channel response functions). Then appropriate bidirectional reflectance distribution function models [RD-75] are applied to the channel emissivities to generate the LSE for the structured land surfaces. VCM assumes that the surface is Lambertian and ignores the influence of shadow and double-scattering processes [RD-29]. It is possible to transform this to other land classifications and this will be looked at in future cycles to ensure

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 33
--	---	--

consistency. For vegetated land covers, emissivity is considered to be the result of the contribution from vegetation and bareground proportions, following the VCM ([RD-29] and [RD-31]):

$$\varepsilon_{eff} = \varepsilon_{veg}FVC + \varepsilon_{bg}(1 - FVC)$$

here FVC is the pixel fractional vegetation cover, and  $\varepsilon_{veg}$  and  $\varepsilon_{bg}$  are the vegetation and bareground emissivities, respectively, per channel. Both  $\varepsilon_{veg}$  and  $\varepsilon_{bg}$  are estimated for land cover classes from spectral libraries.

VCM is used in the single-sensor LST retrieval from GEO sensors as this is expected to be optimum for this product.

#### 4.6.4. Atmospheric Variables

Atmospheric variables (for example Total Column Water Vapour (TCWV), precipitable water and atmospheric temperature) which is used as an input to TIR retrieval algorithms, is provided in LST\_cci by the ECMWF Re-analysis 5 (ERA5) [RD-17]. ERA5 is a re-analysis dataset which provides hourly estimates of a significant number of land and atmospheric variables over the full globe at a spatial resolution of 30km grid. It is the successor to the widely used ERA-Interim Re-analysis dataset [RD-67]. ERA5 currently has a temporal coverage similar to other reanalyses (from 1979 to present), but more years are due to be added to extend this dataset back to 1950.

Precipitable water is used in the UOL retrieval algorithm, along with coefficients selected using land cover and fractional vegetation information, to derive LST. Water vapour information and atmospheric temperature are inputs required to determine retrieval coefficients for the GSW algorithm. Atmospheric profile data from ERA5, which is resolved with 137 atmospheric levels from the surface up to a height of 80km., is used in the UOL\_3 cloud masking algorithm to derive clear sky probability information. Furthermore, ERA5 is employed to create a Calibration Database for determining retrieval coefficients.

#### 4.6.5. Snow masking

Snow masking is part of the biome information used in the UOL retrieval algorithm to determine the most appropriate coefficients to apply.

The NOAA IMS snow maps are daily maps of land, sea, snow and ice for the Northern Hemisphere (NH), derived from both satellite and in situ data and produced by the National Snow and Ice Data Center (NSIDC). IMS maps are provided on an equal area polar stereographic grid at three nominal resolutions: 24 km (available from 1997 to present), 4 km (from 2004 to present) and 1 km (from 2014 to present). IMS is considered an operational product (<https://nsidc.org/data/G02156>) and both ASCII text and GeoTIFF file formats are available.

The IMS product is manually created using several different sources of data [RD-18]. The map generated from the previous day is used as a first approximation. Then data from a combination of satellite imagery, automated snow mapping algorithms and other ancillary data is overlaid from which an analyst creates the new day's map. The satellite data primarily comprises visible imagery from the Polar Operational Environmental Satellites (POES) as well as the Geostationary Operational Environmental Satellite (GOES), Geostationary Meteorological Satellites (GMS, also known as Himawari), Meteosat and MODIS instruments. Microwave products from SSM/I are also used to provide data in cloudy or night-time conditions, but at a lower spatial resolution. The ancillary data is mostly from ground weather observations.

For inclusion in most LST\_cci products using the UOL retrieval algorithm (ATSRs and ATSR-MODIS-SLSTR CDR) daily IMS maps of snow and ice presence in the northern hemisphere at a resolution of 0.01° are produced by nearest neighbour interpolation of 4km IMS data [RD-19]. Prior to 2004, when 4km IMS data became available, a climatology is used. For SLSTR single sensor products 1km IMS is used to benefit from the improved information at higher spatial resolution. The snow data is mapped onto sensor granules or orbits from the IMS data. The conversion is supported by flipping the IMS latitude and longitude data from NSIDC to produce a look up table (LUT). Indices from the LUT provide the information required to interpolate IMS data to the 0.01° equal angle grid.

IMS data, which has been mapped to a granule / orbit is used to update static auxiliary biome data with dynamic snow cover. Figure 6 provides an illustration of the incorporation of IMS NH into the processing chain. For the Southern Hemisphere (SH), an approach combining the methods of [RD-71] and [RD-72] which exploits visible and thermal channels is used to produce a snow mask. The approach attempts to account for factors affecting the spectral signature of snow such as aerosols, ground contamination, and snow grain size [RD-19]. The criteria for the visible and near-infrared channels screen optically thick warm clouds but have difficulty with cirrus and optically thin clouds as they do not significantly disturb the spectral signature of snow in those spectral regions, therefore the TIR channels provide the necessary additional screening in these cases.

Information from these snow masking datasets (in NH) and techniques (in SH) are used for snow cover quality flags. Specifically, where a surface is identified as transient snow/ice retrieval coefficients for permanent snow/ice biome are applied instead of the static biome for the given pixel when implementing the UOL retrieval algorithm.

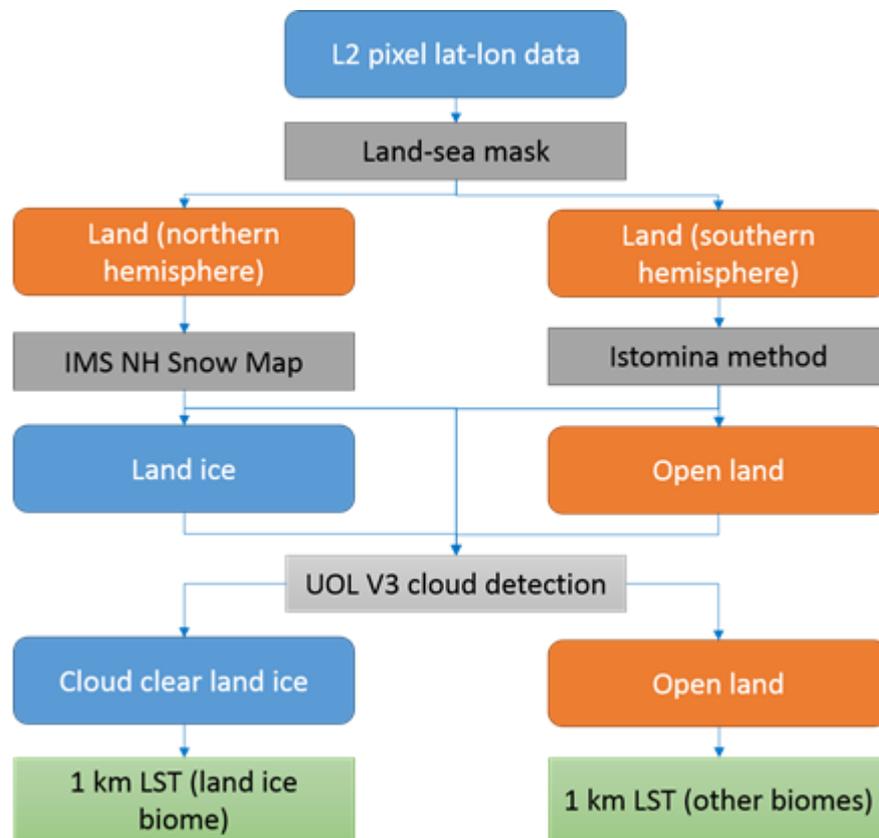


Figure 6: A schematic showing how IMS NH data is incorporated into biome and cloud information for the LST products using biome information.

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 35
--	---	--

Snow masking is part of the land cover information used in the LST retrieval algorithm to determine the most appropriate coefficients to apply. As described above the static biome is replaced by the land class for permanent snow/ice. In addition to the impact on the LST retrieval, it is also utilised in the probabilistic cloud masking algorithm (UOL\_3). Where the static biome is replaced by the land class for permanent snow/ice the corresponding cloud coefficients are also changed, so the permanent snow/ice coefficient from the cloud coefficients ADF are instead used to determine the probabilities for the probabilistic cloud mask. This is expected to improve the accuracy of the masking over these regions, in addition to the expected improvement in the LST retrieval over these regions experiencing transient snow and ice.

Further updates to LST\_cci products will include a move to using ESA Snow Cover CCI products for snow masking once they become available.

## 4.7. Uncertainty Model for Thermal Infrared Algorithms

Following the agreed approach being undertaken in other projects such as ESA DUE GlobTemperature [RD-15] and H2020 EUSTACE [RD-16], whereby SST, LST and IST all conform to a standardised uncertainty model. For LST this has been implemented this for AATSR, MODIS and SEVIRI data, which are the sensors of interest here.

Generally, for each pixel, three components of uncertainty are provided, representing the uncertainty from effects whose errors have distinct correlation properties:

- ❖ random (no correlation of error component between cells);
- ❖ locally systematic (correlation of error component between “nearby” pixels);
- ❖ [large-scale] systematic (correlation of error component between “distant” pixels).

Locally correlated errors are modelled via spatio-temporal correlation length scales that determine how an observation influences the analysis in the vicinity of its time-space location. Systematic errors are accounted for by allowing a bias to be determined within the analysis procedure between different sources of data, whose magnitude is conditioned by the uncertainty attributed to systematic effects.

This approach is both a necessary minimum, since locally systematic effects are significant, and preclude use of a simple random/systematic model and an approximation, in that there are several effects that have a systematic aspect, and all of these are required to be partitioned into either the locally systematic or systematic component. This is though, a significant advance on what has generally been done for LST datasets to date. Moreover, this three-component model applies to all satellite processing levels (L1, L2, L3, and L4). Full details are presented in the End-to-End ECV Uncertainty Budget [RD-69]. Here we only present what is specifically included in the output products.

### 4.7.1. Random

The random component of L1 channel uncertainty can be denoted as  $u_{\text{ran}}(y_c)$ . The effect of this combined across all channels needs to be propagated through the retrieval to give a contribution to the estimate of uncertainty from random effects  $u_{\text{ran}}(x)$  in the retrieved surface temperature. The assumption is that the radiance noise is sufficiently Gaussian and small that the law of propagation of uncertainty is adequate for this propagation, which means:

$$u_{ran,y}(x) = \sqrt{\sum_{c=1}^n \left( \frac{\partial R}{\partial y_c} u_{ran}(y_c) \right)^2}$$

Emissivity is an auxiliary input to all estimates of thermodynamic temperature from BTs, whether explicit or implicit. For LST, there is a potentially significant random error component caused by the pixel-to-pixel variations in emissivity not captured in emissivity auxiliary information because it is related to variability on the ground that is not captured in emissivity atlases/models. The associated uncertainty can be estimated as:

$$u_{ran,\varepsilon}(x) = \sqrt{\sum_{c=1}^n \left( \frac{\partial R}{\partial \varepsilon_c} u_{ran}(\varepsilon_c) \right)^2}$$

Where, some estimate of the uncertainty in emissivity per channel is required. In practice, this is estimated as the magnitude of pixel-to-pixel scale emissivity variability within areas that, based on same land cover classes being treated as having a common emissivity. Emissivity errors are estimated per land class based on both existing literature and validation studies. Full details are included in [RD-69].

The total random component is the acquired by adding the individual components in quadrature.

#### 4.7.2. Locally systematic

Atmospheric fields are correlated on timescales >1 day and length scales >100 km, and it is assumed that errors in estimates of these fields from NWP are correlated on the same scales. For coefficient based retrieval methods the retrieval ambiguity is a contributor of residuals in the fit. For radiative-transfer based retrieval coefficients, simulated-retrieved and simulation-input surface temperatures can be compared. The standard deviation of this input and output difference is an estimate of the magnitude of this locally correlated form of uncertainty. The calculation of the uncertainty can be done on stratified data to parameterise the variations in magnitude of this form of uncertainty. For each range of satellite viewing angle and water vapour (being the primary sources of variability), the uncertainty is estimated as:

$$u_{loc,fit}(x) = \sqrt{Var(\hat{x} - x_{in})}$$

In the GSW algorithm, the model coefficients are calibrated for bands of satellite zenith angle and bands of total column water vapour. As such, an estimate of the state TCWV is required for the LST retrieval, which is used to select which coefficients are applied in the model. To characterize this effect, it's necessary to propagate the uncertainty in TCWV through the LST model. A commonly used measure of uncertainty in Numerical Weather Prediction Models is the ensemble spread that is generally available together with the variable best estimate. The ensemble systems consist in a set of model runs with perturbed initial conditions; some systems also include perturbations to the model physics, more than one model within the ensemble or different physical parametrization schemes [RD-80]. Processing of ensemble data can be quite demanding and, therefore, currently we approximate the instantaneous TCWV spread to a climatology of this spread that depends on actual TCWV, latitude and month.

LST retrieval assumes an emissivity which may be driven by auxiliary land classification information and/or and observed vegetation indices. Across a particular land class area, there may be a mean difference between the assumed and true mean emissivity. This is thus a locally correlated effect on the scales of emissivity variability. The form of the propagation to L2 uncertainty is estimated as:

$$u_{loc,\varepsilon}(x) = \sqrt{\sum_{c=1}^n \left( \frac{\partial R}{\partial \varepsilon_c} u_{loc}(\varepsilon_c) \right)^2}$$

This locally correlated component is based on pixels for the same land cover having the same error characteristics. This does not capture sub-pixel variability for any given pixel within a land cover, which is captured above in the random component, and for which high resolution emissivity data are used to quantify the error properties. The correlation length scale is dependent on the source of the uncertainty. As a result, atmospheric and surface related uncertainties are considered and provided separately and propagated as either correlated or uncorrelated uncertainties as appropriate for a given product. The total locally correlated component is the acquired by adding the individual components in quadrature.

### 4.7.3. Systematic

This includes components such as the uncertainty in the radiative transfer model. It is assumed here that known corrections have been applied by data producers, either at L1 or in the retrieval process to L2, and that what remains is describable as an uncertainty in the bias of the satellite surface temperatures (i.e. the skin temperature of the surface the satellite sees) relative to other data sources of temperature. Knowledge of the satellite engineering specifications and/or validation performance may allow a reasoned estimate of the likely magnitude of residual biases.

Since the different components are independent of each other they are combined in quadrature for a total uncertainty per pixel in the product.

## 4.8. Harmonisation for Climate Data Records

---

### 4.8.1. Cross-calibration of BTs using IASI spectra

Intercalibration differences between sensors mean that different BTs would be derived even in the hypothetical situation where they were observing the same field of view, within a similar spectral band, at the same time and from the same view angle.

Harmonisation of L1 data across all sensors for the LST\_cci Merged Product CDR and ATSR-MODIS-SLSTR CDR entails adjustment of the BTs to a reference sensor. The Global Space-based Inter-Calibration System (GSICS) have used the Infrared Atmospheric Sounding Interferometer (IASI) as the reference sensor in a calibration of SEVIRI radiances. IASI is a Fourier transform spectrometer and provides infrared spectra with high resolution ( $0.5 \text{ cm}^{-1}$  after apodisation, L1C spectra) between  $645 \text{ cm}^{-1}$  and  $2760 \text{ cm}^{-1}$  ( $3.6 \mu\text{m}$  to  $15.5 \mu\text{m}$ ). The IASI spectra are multiplied by the SEVIRI spectral response function. These IASI radiances are then compared to matched radiances from SEVIRI, the resulting differences analysed and a bias and uncertainty in the bias calculated. The SEVIRI radiances are then aligned with IASI [RD-69].

In LST\_cci access to radiances for all instruments is not available so the inter-calibration is performed with BTs. The steps required are:

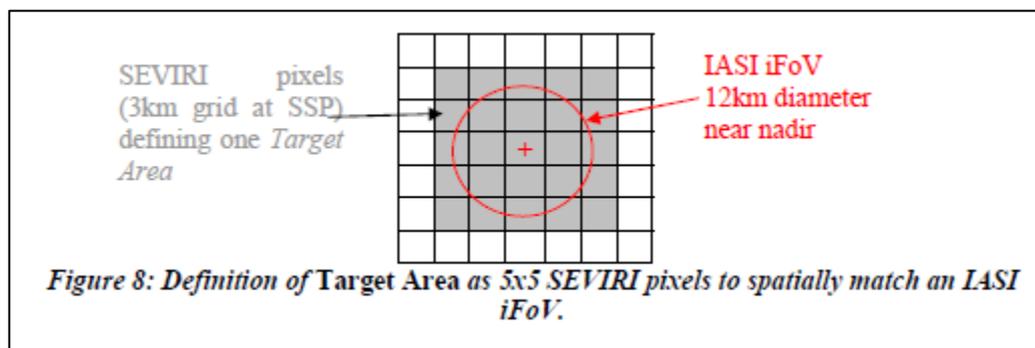
1. Collocate IASI pixels with instrument pixels
  - a. Use L1c IASI swath data
  - b. Use instrument data on Level-3 Uncollated (L3U) grid
2. Multiply IASI spectra with instrument spectral response function to produce instrument equivalent IASI BTs
3. Analyse differences – use results of analysis to calibrate instrument BTs and propagate uncertainty

IASI is an across-track scanning system with scan range of  $\pm 48^\circ 20'$ , symmetrically with respect to the nadir direction. A nominal scan line covers 30 scan positions towards the Earth and two calibration views. One calibration view is into deep space, the other is observing the internal black body. The scan starts on the left side with respect to the flight direction of the spacecraft.

The elementary (or effective) field of view (EFOV) is the useful field of view at each scan position. Each EFOV consists of a  $2 \times 2$  matrix of so-called instantaneous fields of view (IFOV). Each IFOV at nadir is circular with a diameter of 12 km at nadir. At the edge of swath the IFOV is elliptical with size 39 km across track and 20 km along track [RD-68].

#### GSICS method of collocation used for collocating IASI and SEVIRI

- ❖ Collocation in space
  - IASI FoV for nadir pixels defined as circle with diameter 12 km
  - SEVIRI FoV defined nominally as square pixels 3x3 km if close to nadir
  - Array of 5x5 SEVIRI centred on pixel with centre nearest to IASI pixel centre represents IASI FoV.
  - SEVIRI and IASI pixels are selected that fall in same bin of histogram of  $0.125^\circ$  lat/lon grid between  $\pm 35^\circ$  lat/lon. The SEVIRI pixels in each collocation are then averaged.

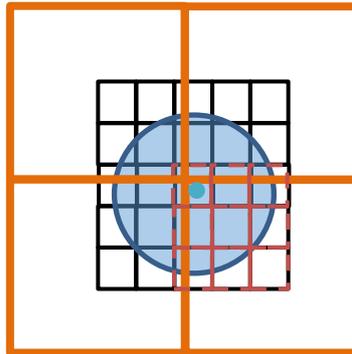


**Figure 7: 5x5 box of SEVIRI pixels that align with IASI pixel (taken from “ATBD for Prototype GSICS SEVIRI-IASI Inter-Calibration”)**

- ❖ Collocation in time
  - Time difference  $\leq 900$ s
  - GEO image where observation time at the equator is closest to LEO equator crossing time is selected and time differences neglected, observations assumed simultaneous and time difference does not contribute to uncertainty, so in practice this means  $< 7.5$ min
- ❖ Viewing geometry
  - $\left| \frac{\cos zen_{GEO}}{\cos zen_{LEO}} - 1 \right| < 0.01$
  - Although acknowledgement that 0.05 for window channels could be acceptable
  - Recommend over land limit intercalibration to night-time only as less effect of differences in azimuth angle.
- ❖ Scene uniformity
  - Option to filter where scene radiance is  $>5\%$  of the standard radiance (see below)
- ❖ Outliers
  - Include outliers since they have equal chance to appear in either sensor's FoV – shouldn't affect bias but will increase noise.
- ❖ Standard Radiances
  - For each channel, bin BT for collocated pixels into 5K wide bins from 200 to 300 K – use mode (or mean of modes, if bimodal)
- ❖ Uncertainty methods associated with the intercalibration are detailed in [RD-70].

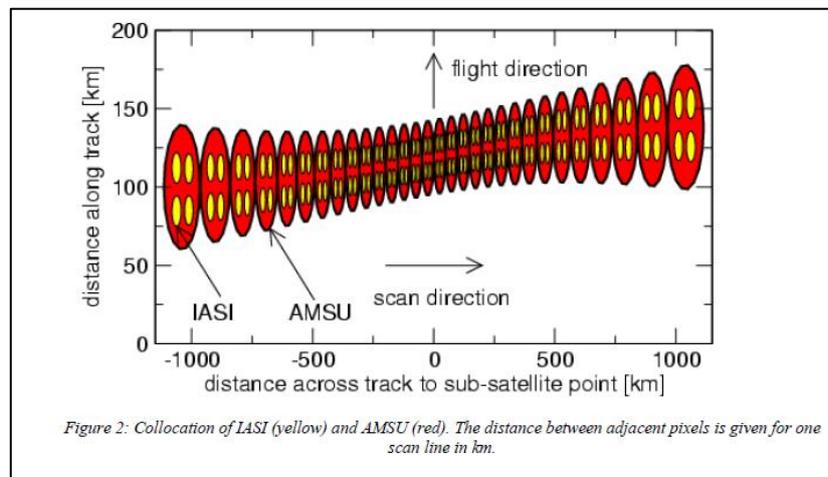
#### Modifications to GSICS for LST\_cci

- ❖ By utilising the input L3U for the LST\_cci dataset we can more readily stay in line with the matchups depicted in Figure 7 above. In other words for a IASI pixel at nadir then a 3x3 array of L3U grid cells wholly covering the IASI pixel would be used in an idealised case.
- ❖ At the edge of the IASI swath then the array would be approximately 4x8 grid cells in an idealised case.
- ❖ In other words L3U pixels covering the min/max lon/lat of the IASI pixel.
- ❖ Applying the standard time difference between sensor matchups of 300s as implemented in the study by EUMETSAT with S3MPC between SLSTR-A and IASI.
- ❖ These modifications are to address certain challenges with applying the standard GSISC methodology:
  - Binning of lat/lons means that if IASI pixel is centred towards corner of lat/lon box then collocated pixels in other grid boxes are ignored (Figure 8).



**Figure 8: Effect of collocation by histogram binning: if IASI pixel (blue) is centred in corner of lat/lon grid (green) then matching SEVIRI pixels in adjacent grid boxes are ignored and only those pixels (orange) in the same gridbox are used in the analysis.**

- IASI pixels change shape across swath: at nadir they are circular with diameter 12 km, at the edge of the swath they are elliptical with an along-track size of 20 km and an across track size of 39 km (Figure 9).



**Figure 9: Change of shape of IASI pixel across track (taken from IASI Level 1: Product Guide) – note axes have different scales – pixels are circular at nadir.**

- Analysis limited between +/- 35° lat/lon and only at night over land, which reduces the dynamic range of the matchups

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 41
--	---	--

#### 4.8.2. Adjustment to common reference time

Differences in overpass time between sensors of a series results in step changes in LST. This occurs for example between ATSR-2, overpass time 10:30 and 22:30, and AATSR with overpass time of 10:00 and 22:00. Since there is a data gap between the end of AATSR mission and start of SLSTR mission, LST from Terra-MODIS can be used to fill the gap but with the caveat that there are differences in overpass time also.

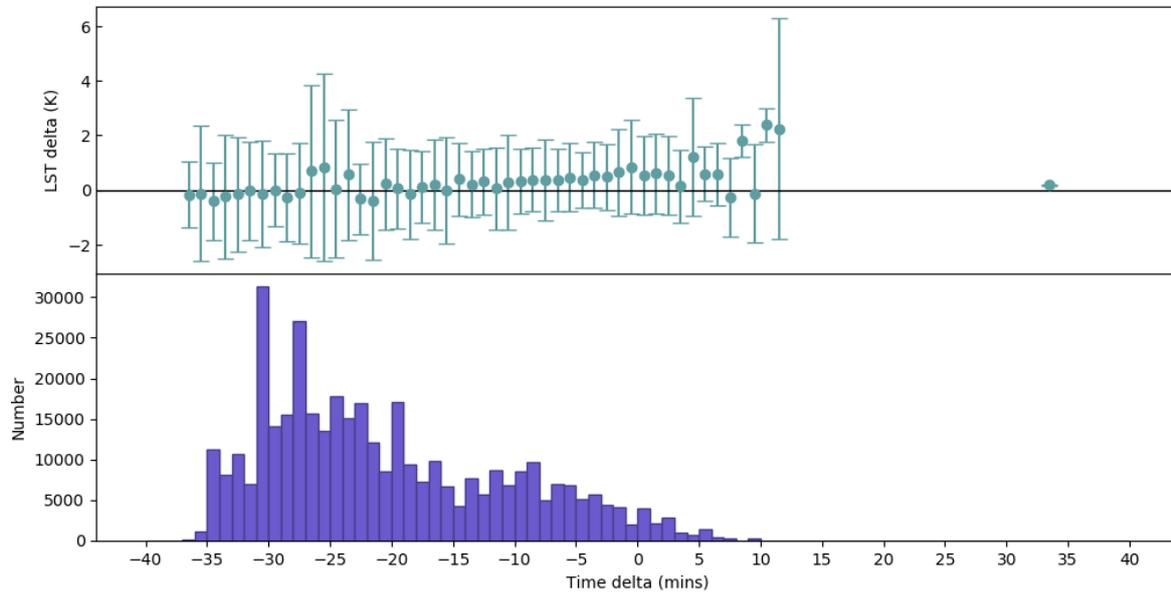
We can estimate a correction to the LST due to the time difference using Terra-MODIS as the reference sensor and another sensor as the sensor to be corrected.

It should be noted that a correction to the LST for Climate Data Records should only be carried out after the procedures in Section 4.8.1 have been implemented to harmonise across the L1 brightness temperatures.

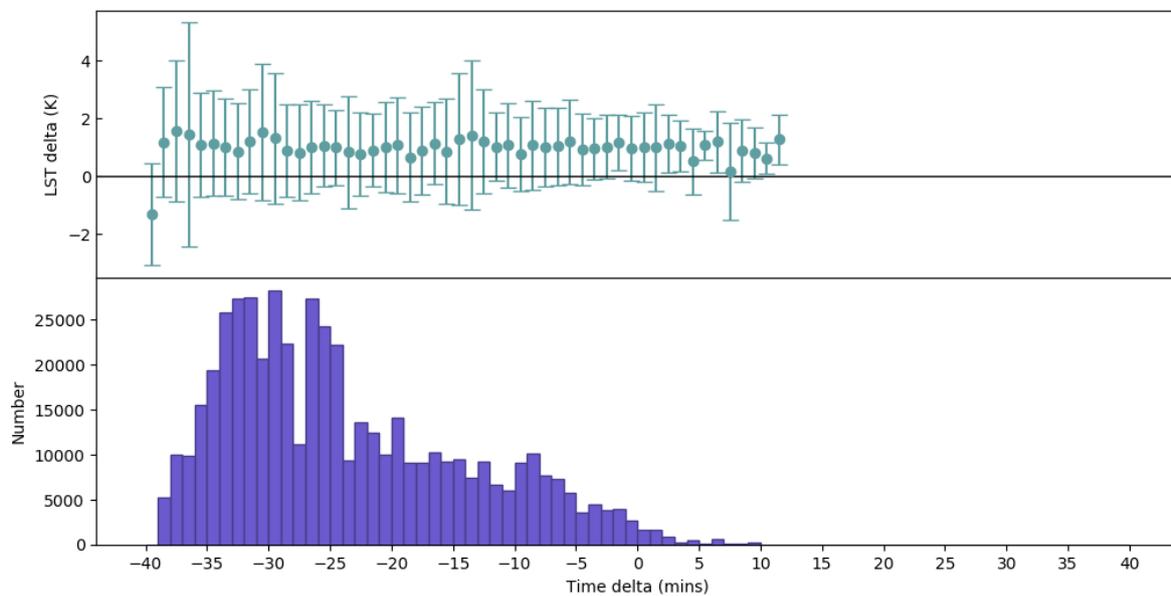
The procedure for implementing the time correction can be summarised thus:

- ❖ Match ups are produced from gridded data (L3U) using a time window of 40 minutes. Matchups are restricted to data where the difference in satellite zenith angle was less than 10°.
- ❖ The matchup LST differences are binned to multi-dimensional histograms: day or night (on reference solar angle), observation time difference (1 minute bins), latitude (10° bins), and landcover class (biome). Multi-year histograms for each calendar month are calculated.
- ❖ Statistics are determined for each bin (mean, standard deviation, number of samples) and look-up tables (LUTs) produced. Statistics are plotted for example histograms in Figure 10.
- ❖ The LST correction is read from the LUT (given the time of day, location, landcover class, and observation time difference between the acquisition time and the nominal time for the CDR for the given pixel) then applied at the L3U level on the 0.05° grid to the sensor to be corrected.
- ❖ It is expected that some instability will occur for this correction, but is expected to be minimal and cosmetic filling is avoided.
- ❖ Cloud contamination is also likely to cause some artefacts, although a common approach is taken in the cloud masking between the sensors to minimise this.
- ❖ Note, currently there is no uncertainty propagated through the correction, but this will be documented in due course in version 2.0 of the E3UB [AD-2].

ATSR\_3 - MODIST: day: Jan: 40.0 to 50.0 Latitude (degrees N): Landcover class 6.0



ATSR\_3 - MODIST: night: Jan: 40.0 to 50.0 Latitude (degrees N): Landcover class 6.0



**Figure 10: Envisat AATSR minus Terra MODIS LST difference v. time difference: average for January over all years, for latitude band 40-50 °N, and landcover class 6 (Closed\_(>40percent)\_broadleaved\_deciduous\_forest\_(>5m)), for daytime (top) and night-time (bottom). Top panel each plot shows mean (circle) and mean +/- 1 sd (bars), bottom panel show number of samples.**

## 4.9. Multi-sensor Climate Data Records

### 4.9.1. ATSR-SLSTR CDR

To fill the gap between the end of the Envisat mission and the end of the Phase E1 commissioning of the Sentinel-3A mission requires an instrument of equivalent spatial resolution with a LECT close in time to AATSR and SLSTR and of sufficient high quality. The choice made in Phase-1 was Terra-MODIS. Not only does it meet these minimum requirements, but it also spans all three of ATSR-2, AATSR and SLSTR, and moreover has a common LECT (10:30 and 22:30) with ATSR-2 so knowledge of the temporal correction between ATSR-2 and AATSR is also applicable for Terra-MODIS and AATSR. Note, AATSR and SLSTR have the same LECTs (10:00 and 22:00).

Five steps have been taken in the development of the ATSR-SLSTR CDR:

- ❖ Stage 1: gap bridging
- ❖ Stage 2: time difference correction
- ❖ Stage 3: apply consistent Level-2 algorithms
- ❖ Stage 4: restrict satellite viewing angle
- ❖ Stage 5: gap filling

The approaches for each of these steps are summarised here:

- ❖ Stage 1:
  - Intercalibration work for each input sensor vs Metop-IASI, through Quasi Simultaneous Nadir Overpasses (QSNOs) at high latitudes (Greenland, Antarctica, and over sea-ice) and SNOs where possible elsewhere.
- ❖ Stage 2:
  - We characterise the effect of the temporal mismatch between instruments on the LST time series and evaluate and tune the resulting correction.
  - The reference sensor here is Terra-MODIS since this spans all of ATSR-2, AATSR and SLSTR-A and -B and has a common LECT with ATSR-2. As Terra-MODIS is the reference sensor to determine the corrections, we correct AATSR and SLSTR to the overpass time of ATSR-2 and Terra-MODIS, so that the much of the time series is uncorrected (i.e. at its native overpass time).
  - The look-up table (LUT) of corrections is applied to the respective input datasets to address the problem of combining time series that have a 30 minute difference in local equator crossing time (LECT).
- ❖ Stage 3:
  - A common algorithm is applied across all sensors, in this case the UOL algorithm as described in Section 4.2.2.
  - The uncertainty components described in Section 0 are implemented across all sensors consistently.
  - Cloud masking is consistent, with the probabilistic cloud mask applied for each of the sensors.
- ❖ Stage 4:
  - We restrict the satellite viewing angles of the sensors to that of the lowest common denominator – the ATSRs to avoid discontinuities in the time-series.
- ❖ Stage 5:

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 44
--	---	--

- We will apply the consistent approach in calibration database, retrieval algorithm, uncertainty characterisation, and cloud detection to the gap-filling instrument Terra-MODIS between the end of Envisat and the in orbit commissioning review (IOCR) of SLSTR.

#### 4.9.2. Merged IR CDR

The aim of the Merged IR CDR development is to integrate polar and geostationary infrared-based LST to deliver information on the diurnal variation of LST at a global scale. The Merged IR CDR includes all available IR geostationary sensor data and IR polar orbiting sensor data over the period 2009 – 2020 with the output LST ECV product being delivered with a 3-hourly temporal resolution. The processing chain of the final product is based on the foundation of the Merged GEO product. The LEO data are then added in one at a time (AATSR, MODIS (Terra + Aqua), SLSTR (A + B)).

We take the approach of the best algorithm per number of available channels. So all instruments with two channels used the GSW algorithm (Section 4.2.3), whereas the instruments with only one available channel, such as on MTSAT and early GOES used the SMW algorithm (Section 4.2.4).

Consistency between input data includes all retrieval and uncertainty coefficients being generated using the LST\_cci Calibration Database (Section 4.4). The following steps are taken in the development:

- ❖ Step 1 (merging strategy):
  - Temporal mismatches between the individual sensors with respect to the nominal UTC time are characterised.
  - The data is merged at specified 3-hourly UTC time slots throughout the day consistent with the observation times of the geostationary satellites, with the available data from the polar-orbiters being integrated by time matching.
- ❖ Step 2 (algorithm consistency):
  - Application of intercalibrations between sensors with respect to Metop-IASI.
  - Generation of all IR algorithm coefficients and accompanying uncertainty coefficients using the LST\_cci Calibration Database (Section 4.4) to ensure consistency across all input sensors.
  - In each case a consistent algorithm is applied across all sensors where no degradation in performance is encountered:
    - ◆ For single channel sensors the single channel SMW algorithm is consistently applied.
    - ◆ For split-window sensors the GSW algorithm is applied that maximises the information in the two channels rather than loss of information through use of a single channel algorithm.
- ❖ Step 3 (merging logic):
  - The LEO IR and GEO IR data are combined using all the building blocks in place for consistency across products.
  - The basis of the Merged IR CDR is the Merged GEO with the LEOs infilling the gaps at higher latitudes and over Central Asia:
    - ◆ The viewing geometry of the GEOs is restricted to satellite zenith angles  $\leq 60^\circ$ .
    - ◆ In regions where data from multiple GEOs overlap only data from one GEO is used.
    - ◆ LEO data is not ingested inside the coverage of the GEO discs.
    - ◆ For the GEOs, the time window depends on the respective GEO scans (starting at -28 minutes for MTSAT in relation to the nominal UTC time, through to +12 minutes for MSG-SEVIRI).
    - ◆ To maintain consistency with the GEO time windows, LEO data is added to the 3-hourly UTC product where the LEO data falls within  $\pm 30$  minutes of the nominal UTC time.



- ◆ If data from more than LEO sensor is available for a grid-cell then the data nearest to the nominal UTC time is used.

## 5. Retrieval of Land Surface Temperature from Microwave Sensors

In the LST\_cci open algorithm intercomparison round-robin, the performance of different LST retrieval algorithms for a set of specific thermal infrared and microwave satellite sensors was assessed to identify the best algorithms for a future climate quality operational system. The algorithms chosen for the LST\_cci MW products is:

- ❖ the NNEA Algorithm (MW)
  - for the Special Sensor Microwave/Imager (SSM/I) LST

In the following section a description of MW retrievals is presented, along with a description of the retrieval algorithm chosen for use in LST\_cci.

### 5.1. Physics of the Problem

---

As in the TIR region of the electromagnetic spectrum, MW instruments on satellites designed for retrieval of the earth's surface parameters also use spectral windows with large atmospheric transmission. Typical measuring channels are placed around 6, 10, 18, 37, and 89 GHz. For the low frequencies the atmospheric transmission is very high and the impact of the atmospheric attenuation in the accuracy of the LST retrieval is small. This is also valid for atmospheres with clouds, as the emission from liquid water acts as a weak source when compared with the large emission from the land surface. This is a major difference with respect to the TIR, where the presence of clouds prevents the instrument from seeing the radiation from the surface. At the higher frequencies, especially at 89 GHz, cloud contribution can be significant. The scattering depressions in the radiation emitted by the surface can be important for clouds containing large ice particles, affecting the accuracy of the LST retrieval. Emissivity varies more in the MW region than in the TIR, showing spatial and temporal variability that needs to be accounted for in the retrieval.

Given the limited impact of clouds, the MW LST can complement the TIR estimations. However, retrieving LST from microwaves is more challenging than in the TIR. First, at the large wavelengths of the MW region the Rayleigh-Jeans law is valid and the emitted radiation from the surface is the direct product of the emissivity and the surface temperature. Compared with the TIR, this results in a stronger dependence on the emissivity of the MW signal. Furthermore, the MW emissivity varies more with surface properties such as soil moisture, vegetation cover, or the presence of snow, compared with the TIR emissivity. In addition, as the antenna aperture is function of the wavelength, the spatial resolution of the MW observations degrades with decreasing frequencies, whereas the atmospheric contribution to the signal tends to increase with frequencies. The spatial resolution of the current MW imagers is typically 10 to 20 km at the frequencies of interest, a factor 10 to 100 coarser than the current TIR observations. Moreover, MW observations are only available from polar satellites, contrarily to the TIR also observed from geostationary orbits, thus limiting the time sampling of the MW observations. Lastly, the microwave radiation can emanate from the subsurface, not only from the surface skin: the lower the frequency, the larger the wavelengths, and the larger the penetration into the subsurface.

#### 5.1.1. Mathematical description

In the MW region, the intensity of the radiation is commonly expressed in terms of brightness temperature, i.e., the temperature a black body would have to match the intensity emitted by a given source. In this region, the Rayleigh-Jeans approximation of the Planck's law is valid, and with the emissivity defined as the ratio of the radiance of a grey body with respect to the radiance of a black body

at the same physical temperature, the brightness temperature becomes the direct product of the emissivity and the physical temperature. Using that relationship, the brightness temperature observed by the MW space-borne radiometer at frequency  $\nu$  and polarization  $p$  assuming surface specular reflection and non-scattering atmosphere may be written:

$$\begin{aligned}
 Tb_{\nu,p} &= T_v^\uparrow + \tau_\nu [\epsilon_{\nu,p} T e_\nu + (1 - \epsilon_{\nu,p}) T_v^\downarrow] = \\
 &= (T_v^\uparrow + \tau_\nu T_v^\downarrow) + \epsilon_{\nu,p} (\tau_\nu T e_\nu - \tau_\nu T_v^\downarrow) = \\
 &= A_\nu + \epsilon_{\nu,p} (B_\nu T e_\nu + C_\nu)
 \end{aligned}$$

where  $\tau_\nu$  is the total atmospheric transmittance along the sensor line of sight,  $T_v^\uparrow$  and  $T_v^\downarrow$  represent the upwelling and downwelling atmospheric emission, respectively,  $\epsilon_{\nu,p}$  is the surface emissivity,  $T e_\nu$  is the effective emission temperature of the surface, and the  $A_\nu$ ,  $B_\nu$  and  $C_\nu$  terms are abbreviations of the corresponding terms. Dependence of these variables on sensor viewing angle is omitted since we observe at a constant zenith angle close to 53 deg.

The effective emission temperature of the surface will be the parameter to be estimated as the MW LST. If the terms  $A_\nu$ ,  $B_\nu$  and  $C_\nu$  and the emissivity can be properly characterized, the LST can then be retrieved. In some specific locations, especially in very dry and sandy areas, the effective emission temperature can be significantly different from the LST, because of the penetration of the microwave radiation into the sub-surface. These pixels will be duly associated with larger error estimates.

### 5.1.2. Emissivity

As in the TIR, variations in surface properties change the surface emissivity and the emitted MW radiation. The variations are larger than in the TIR. For instance, at 37 GHz and horizontal polarization, emissivity values can get below 0.8 for some outcrops in arid regions, and even lower in regions where standing water is in the footprint of the observation. Without significant changes in surface conditions, the emissivity is quite stable for a given location and month, with a reported variability  $< 0.02$  for day-to-day variations from satellite emissivity retrievals [RD-42]. There are noticeable emissivity variations for different surface types, with higher (lower) emissivity for vegetated (arid) regions. Different to the TIR, there can be significant variability also within a given surface type due to changes in surface conditions, mainly related to variation in soil moisture and vegetation cover. Especially challenging are the snow-covered surfaces, where melting processes and snow metamorphisms can result in large emissivity variations, and regions where seasonal flooding can occur, due to the large changes in the electromagnetic permittivity associated to the presence of water.

Details of the auxiliary emissivity products used in LST\_cci are given in Section 5.4.1.

## 5.2. Algorithm Description

### 5.2.1. NNEA Algorithm

From the formulation given in Section 5.1.1, we can express the LST as:

$$Te_v = \frac{Tb_{v,p} - A_v - \epsilon_{v,p}C_v}{\epsilon_{v,p}B_v}$$

In principle, one single channel could be used to derive the LST assuming that all the terms in the equation can be properly characterized. But given the uncertainties associated to those calculations, retrieving LST using simultaneously a larger number of frequency channels further constrains the inversion problem, resulting in more accurate LST retrievals, as shown in the LST\_cci open algorithm inter-comparison round-robin. This is similar to the TIR, where the double-channel algorithms are more suitable to take into account the atmospheric absorption and emissivity effects.

Comparable to the TIR, the MW retrieval algorithm needs to deal with emissivity and atmospheric variations. To deal with emissivity changes, pre-calculated microwave monthly mean emissivity estimates available from the Tool to Estimate Land Surface Emissivity in the Microwave (TELSEM, see Section 5.4.1) are used as inputs to the retrieval algorithm, together with the brightness temperatures. Concerning the atmosphere, no temperature or water vapour information is used as input, but the information is introduced into the retrieval by also including the 22 GHz channel, which is close to a water vapour line and therefore sensitive to changes in atmospheric conditions.

The function given by the previous equation is approximated by a non-linear regression between the LST and the combination of the brightness temperatures and emissivity values, with the coefficients of the regression determined with a calibration database (see Section 6.8). The non-linear regression is built by a standard multi-layer perceptron (MLP) as in [RD-14]. MLPs are a type of neural network commonly used to reproduce transfer functions between observations and related geophysical parameters given their proven capability to approximate any continuous function with an arbitrary precision [RD-43].

MLPs contain one input layer with as many nodes as required inputs (in this case the 14 inputs of function F, i.e., the brightness temperatures and emissivities for the 7 MW channels), one hidden layer with a number of nodes, and as many output nodes as the number of variables to be predicted. In principle, the MLP should be set to retrieve only LST. However, the implementation of an uncertainty model for the MW LST will also require to set a second MLP outputting not only the retrieved LST, but also the LST retrieval uncertainty. This is described in the coming Section 5.5 when the uncertainty model for the MW algorithm is described.

If the input vector of the MLP is called  $i$  and one of the outputs of the MLP is called  $u$ , the way the input signal propagates through the MLP is given by:

$$u = f_o(W^o i^o + b^o) = f_o(W^o f_h(W^h i + b^h) + b^o)$$

where  $f_i$  is the activation function,  $W^j$  the weighting matrix,  $b^j$  the bias, and  $i^j$  the input at layer  $j$ , in this case  $o$  is for the output layer and  $h$  for the hidden layer. Hyperbolic tangent and linear activation

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 49
--	---	--

functions are used for the hidden and output neurons, respectively.

The weight and biases can be considered as the regression coefficients of the non-linear model provided by the MLP. These are determined during a learning phase, called training, where the weights and biases that minimize a cost function, determined by a set of input-output examples, are found. Here the examples are provided by the calibration database described in Section 6.8, while the cost function can be expressed as:

$$C = \sum_{l=1}^Z \|t^l - u(y^l)\|$$

where Z is the number of samples in the calibration database,  $\| \cdot \|$  is the standard 2-norm, and  $u(y^l)$  is the output vector of the MLP for the corresponding input vector. In other words, we minimize the mean sum of squares of the difference between targets (the training LSTs of the calibration database) and current outputs of the MLP to the corresponding input vectors (the training brightness temperatures and emissivities). The initial weights of the neural network are randomly initialized by the Nguyen-Widrow algorithm [RD-39], and the final weights are assigned by a Marquardt-Levenberg back-propagation algorithm [RD-44]. To prevent over-fitting to the training data set, a cross-validation technique is used to monitor the evolution of the training error function.

Permanently ice-covered surfaces have a distinct range of LST values and surface emissivities. Tests have shown that the LST accuracy improves if one regression is dedicated to invert observations over Antarctica and Greenland, while a second one is devoted to the remaining continental land [RD-14]. Therefore, we adopt the same strategy here, and two sets of MLPs are trained separately.

The MW retrieval algorithm is applied to the brightness temperatures at the original locations of the sensor swath acquisitions. Given the different channel footprints, the retrieval combines information at different spatial resolutions. As the 19.35 GHz channels have a resolution of ~60 km, information from up to ~ 60 km affects the LST retrievals. However, retrieval tests show that the 37.0 GHz channels are the ones having more weight in the retrieval and as such the effective spatial resolution may be considered to be of the order of ~30 km, corresponding to the resolution of those channels.

### 5.3. Calibration Database for Determining Retrieval Coefficients for the MW Algorithm

A clear-sky calibration database was generated for the LST\_cci open algorithm inter-comparison round-robin, in a similar way to the database described in Section 6.8 for the TIR but simulating SSM/I observations instead. This achieved the objective of evaluating algorithm performance both in the TIR and MW regions within a common inversion setup. However, for the final MW retrieval algorithm, a larger database already exists, based on real SSM/I observations for both cloudy and clear atmospheres, and it is preferred as it describes in a more comprehensive way the relationship between the LST and the SSM/I and SSMIS observations.

The database used for the final retrieval algorithm is based on the inversions of SSM/I observations described in [RD-38]. Atmospheric and surface parameters were retrieved for clear and cloudy conditions with a relatively complex inversion setup. To constrain the inversions, a large range of ancillary

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 50
--	---	--

observations were used, including cloud and surface parameters from the International Satellite Cloud and Climatology Project (ISCCP) [RD-46], and atmospheric information from the National Centre for Environmental Prediction (NCEP) meteorological analysis [RC]. This ancillary information was required because the original inversions of [RD-38] were first developed to estimate the atmospheric parameters over land, with caution being necessary given that the atmospheric signal is rather small as compared to the surface contribution.

The retrievals also provided LST estimates under clear and cloudy conditions. They have been thoroughly evaluated [RD-47, RD-48, RD-49], and are the reference for the calibration of the simplified retrieval algorithm of the LST\_cci MW product. As described in Section 5.2.1, the new retrieval algorithm only targets LST and removes the need of ancillary information, making it more robust than the original inversion algorithm regarding the objective of providing a seamless climate data record.

The final database includes the global LST estimates, together with the corresponding SSM/I observations at the different frequencies, and monthly climatological emissivity sourced from TELSEM (see Section 5.4.10). Four years (2000, 2003, 2005, and 2007) are included to provide sufficient land and atmospheric variability, with the pairs of LST and brightness temperatures quality-controlled to assure that only pairs where the difference between the observations and the simulated brightness temperatures for the given atmospheric and land state is within an acceptable noise, typically of the order of half of the instrumental noise. The final database consists of ~4.5 million cases, which are divided into a validation database of 1 million cases, while the remaining 3.5 million are sampled to provide a training database of 1 million cases equalized in LST and clear/cloudy occurrence. Both training and validation databases are further divided into sub-databases covering Antarctica and Greenland, and the remaining continental land, for the separate retrievals described in Section 5.2.1.

## 5.4. Auxiliary Datasets for Microwave Retrievals

---

The following section gives a description of the auxiliary datasets used for microwave retrieval algorithms utilised in LST\_cci. This section also described how these auxiliary datasets are applied in each algorithm.

### 5.4.1. Emissivity

#### TELSEM

TELSEM (a Tool to Estimate Land Surface Emissivities at Microwave frequencies) [RD-45, RD-38] provides a global monthly mean emissivity dataset. It is currently integrated to the Radiative Transfer for TOVS (RTTOV) forward model, but a general version that can be interfaced with other radiative transfer codes is also available. The dataset is based on SSM/I observations at 0.25° resolution from the period 1993–2004, but observations from the Tropical Rainfall Measuring Mission (TRMM) and Advanced Microwave Sounder Unit-A (AMSU-A) for selected months in the same period are also used to provide interpolation routines capable to generate the emissivity at different frequencies and angles. The emissivity data are produced here at frequencies of 19, 22, 37 and 85 Hz in both vertical and horizontal polarisations. These channels are used in the radiative transfer calculations of the MW calibration database and in the microwave LST retrievals. In the retrievals, they provide a reasonable guess of the surface emissivity, which helps to improve the accuracy of the LST estimates as demonstrated in the LST\_cci open algorithm inter-comparison round-robin.

 <b>land surface temperature</b> cci	<b>Algorithm Theoretical Basis Document</b>  <i>WP2 – DEL-2.2</i>	Ref.: LST-CCI-D2.2-ATBD Version: 3.0 Date: 25-Nov-2021 Page: 51
--	---	--

The TELSEM dataset will be used in the microwave LST retrieval algorithm NNEA.

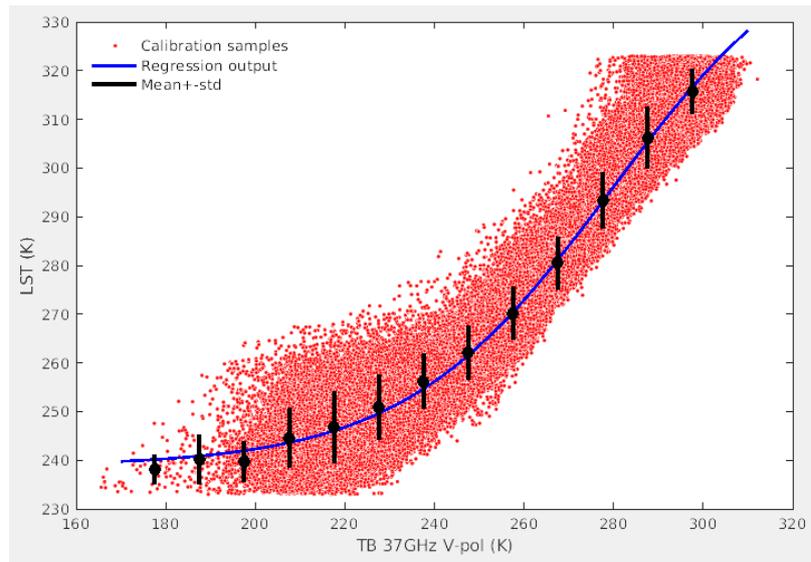
For more information on this dataset see: <http://www.estellus.fr/index.php?static12/microwave-emissivity>.

## 5.5. Uncertainty Model for Microwave Algorithms

The microwave calibration dataset is built with real observed brightness temperatures, retrieved LST estimates from a detailed inversion of these observations. Given the nature of this calibration dataset, it can be argued that most of the sources of uncertainty are included in this dataset. Therefore, this information can be used not only to calibrate the MLP regression model, but also to estimate an LST uncertainty based on inspecting the dispersion of values in the input-output space of the calibration dataset [RD-78].

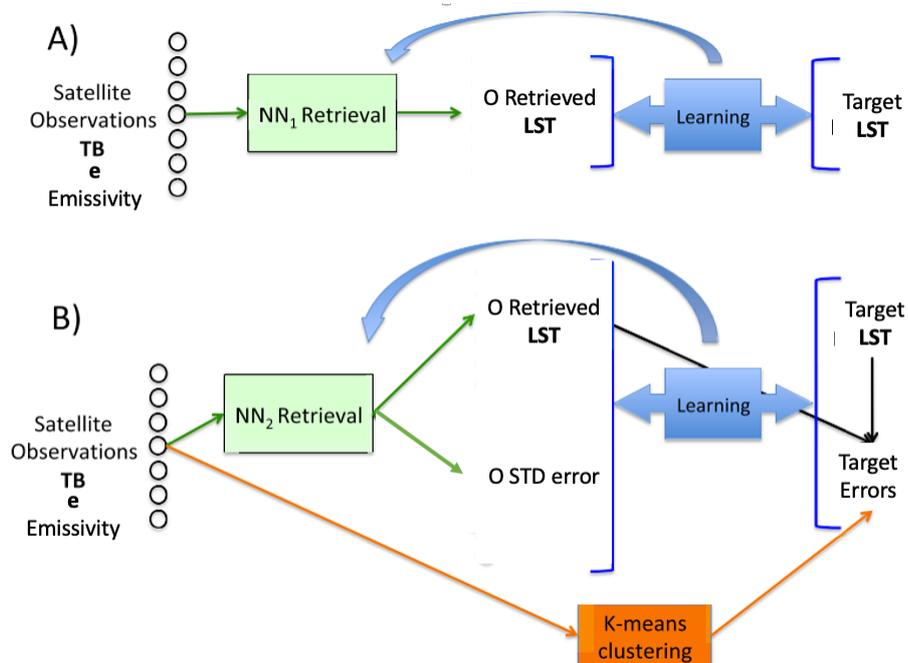
For each combination of brightness temperatures and emissivities, the deterministic function of the MLP regression model will output a single LST value. In order to associate an uncertainty value to that LST estimation, a small subset of the input space around that combination can be selected, and the corresponding dispersion in LST values in the calibration output space used as a measure of uncertainty. In practical terms, this can be done by dividing the multi-variable input space in a number of narrow bins, followed by the calculation, for each selected bin in the input space, of the standard deviation of the LST error (i.e., the difference between the LST values associated to that input bin and the corresponding LST estimates from the regression model). A lookup table storing bins and uncertainties can be created, and searched for each LST retrieval to associate an uncertainty to each LST estimation.

This approach is illustrated in Figure 11. For illustration purposes the inversion problem is simplified by regressing only the brightness temperatures of the 37 GHz vertically polarized channel on the corresponding LST of the calibration dataset. The blue lines show the deterministic output of the regression model, the black circles the mean value of the LST for a number of bins of brightness temperatures, and the black horizontal lines the  $\pm$  one standard deviation of the LST for that bin, centred around the mean value. As expected, most of the LST mean values coincide with the regression model output, i.e., the regression model performs well and is able to find the expected LST value for each given brightness temperature. The standard deviations characterize the vertical dispersions representing the uncertainty in the brightness temperature – LST relationship. This dispersion is a characteristic of the calibration dataset, and it is the result of (1) expected variability, i.e., the observed brightness temperatures are affected by variations in other atmospheric and surface parameters, not just the LST, and (2) uncertainty in the calibration brightness temperatures (e.g., the instrumental noise) and associated LST values (e.g., the retrieval error of these LST estimates). This dispersion is independent of the regression model, and cannot be overcome unless additional information is added to the inversion scheme.



**Figure 11. Illustration of a single-channel regression model and calibration dataset. A neural network is trained to retrieve LST from the SSM/I 37 GHz vertically polarized brightness temperatures (TB). Red dots are the samples in the calibration dataset, and the neural network output is represented in blue. The horizontal axis has been divided into 5 K TB bins, and for each bin the mean (black circles) and +/- one standard deviation (black line centred around the mean) are given. See the text for more details.**

In practice, the multi-dimensionality of the input space complicates the bin selection. Also, storing the dispersion values in a lookup table results in a discrete uncertainty estimation. A more convenient implementation is to adopt the technique of [RD-78], where a clustering technique is used to facilitate the bin selection, and the uncertainty estimation is incorporated to the MLP regression model itself, to provide a continuous mapping also for the uncertainty. This approach is illustrated in Figure 12. First, a MLP is calibrated to provide a first regression model and allow a first estimation of the retrieval error, calculated as the difference of the retrieved LST and the corresponding target LST. Then, a k-mean clustering technique [RD-79] is used to classify the samples of the calibration dataset into a number of coherent clusters, and the retrieval error is estimated for each cluster. This is followed by calibration of a new MLP regression model having as output the LST and the retrieval error, using the same calibration dataset but with the associated errors estimated with the clustering technique added. This procedure can be iterated until there is a convergence on the estimated error, i.e., the error estimation from the output of the neural network agrees with the error derived from the difference between retrieved and target LST.



**Figure 12. Illustration of (A) a traditional LST estimation with a Neural Network (NN), and (B) the proposed scheme where both the LST and an estimation of the LST error are provided by the NN. See the text for more details. Diagram adapted from Figure 2 in [RD-9].**

A limitation of this approach is that it provides a total uncertainty value, without the possibility of propagating uncertainties individually to produce an uncertainty breakdown as in the IR. However, this total uncertainty derivation produces a more robust and useful estimation than a partial breakdown of uncertainties, where only a small number of error effects could have been taken into account with the present MW algorithm.

## Appendix A - Land cover Classification Definition

*Table 3: ULeic / LandCover\_cci (LCCS) hybrid biome definition.*

Biome number	Definition
0	No data
10	cropland rainfed
11	cropland rainfed herbaceous cover
12	cropland rainfed tree or shrub cover
20	cropland irrigated
30	mosaic cropland
40	mosaic natural vegetation
50	tree broadleaved evergreen closed to open
60	tree broadleaved deciduous closed to open
61	tree broadleaved deciduous closed
62	tree broadleaved deciduous open
70	tree needleleaved evergreen closed to open
71	tree needleleaved evergreen closed
72	tree needleleaved evergreen open
80	tree needleleaved deciduous closed to open
81	tree needleleaved deciduous closed
82	tree needleleaved deciduous open
90	tree mixed
100	mosaic tree and shrub
110	mosaic herbaceous
120	shrubland
121	shrubland evergreen
122	shrubland deciduous
130	grassland
140	lichens and mosses
150	sparse vegetation
151	sparse tree
152	sparse shrub
153	sparse herbaceous
160	tree cover flooded fresh or brakish water
170	tree cover flooded saline water
180	shrub or herbaceous cover flooded
190	Urban
200	bare areas of soil types not contained in biomes 21 to 25
201	unconsolidated bare areas of soil types not contained in biomes 21 to 25
202	consolidated bare areas of soil types not contained in biomes 21 to 25
203	bare areas of soil type Entisols Orthents
204	bare areas of soil type Shifting sand
205	bare areas of soil type Aridisols Calcids
206	bare areas of soil type Aridisols Cambids
207	bare areas of soil type Gelisols Orthels
210	Water
220	Snow and ice



***End of document***