CCI Land Surface Temperature

Algorithm Theoretical Basis Document
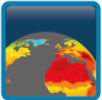
WP2.1 – DEL-LST-CCI-D2.2-ATBD

Ref.: LST-CCI-D2.2-ATBD
Date: 19-Sep-2019
Organisation: Consortium CCI LST

Signatures

	Name	Organisation	Signature
Written by	Emma Dodd	ULeic	
	Darren Ghent	ULeic	
	Carlos Jimenez	Estellus	
	Sofia Ermida	IPMA	
Reviewed by	Isabel Trigo	IPMA	
	Sofia Ermida	IPMA	
Approved by	Darren Ghent	ULeic	
Authorized by	Stephen Plummer	ESA	

Change log

Version	Date	Changes
1.0	20-Jun-2019	First version
1.1	19-Sep-2019	Updated version taking into account RIDs raised by ESA

Table of Content

List of Figures

List of Tables

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 will be 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 will be 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

| | Algorithm Theoretical Basis Document

WP2 – DEL-2.2 | Ref.: LST-CCI-D2.2-ATBD |
|---|---|---|
| land surface temperature cci | | Version: 1.1 |
| | | Date: 19-Sep-2019 |
| | | Page: 3 |

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 will aim 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, J. Geophys. Res., 89, 7231–7237.
[RD-2]	McMillin, L. M. and D. S. Crosby, 1984, Theory and validation of the multiple window sea surface temperature technique, J. Geophys. Res., 89, 3655–3661.
[RD-3]	Prata, A. J., 1993, Land surface temperatures derived from the AVHRR and the ATSR, 1, Theory, J. Geophys. Res., 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, J. Geophys. Res., 99(D6), 13,025–13,058.
[RD-5]	Prata, A. J., 1994b, Land surface temperature determination from satellites, *Advances in space research*, 14, 15-26.
[RD-6]	Sutherland, R. A., 1979, Broadband and spectral emissivities of natural sands and vegetation (2–20 µm), J. Atmos. Oceanic Technol., 3, 199–202.
[RD-7]	Salsbury, J. W., and D. M. D'Aria, 1992, Emissivity of terrestrial materials in the 8 – 14 µm atmospheric window, Rem. Sensing Environ., 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, Int. J. Remote Sensing, 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. IEEE Geoscience and Remote Sensing Letters, 11, 1840-1843
[RD-11]	Ghent, D., Corlett, G., Goettsche, F., & Remedios, J. (2017) Global land surface temperature from the Along-Track Scanning Radiometers. Journal of Geophysical Research – Atmospheres, 122, 12167-12193
[RD-12]	Wan, Z. and Dozier, J., (1996) A generalised spilt-window algorithm for retrieving Land-Surface Temperature from space, IEEE Trans. GeoSci. Remote Sens.
[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, J. Geophys. Res., 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, J. Geophys. Res. Atmos., 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), 5th August 2019. https://cds.climate.copernicus.eu/cdsapp#!/home
[RD-18]	Ramsay, B.H. (1998). The interactive multisensor snow and ice mapping system. *Hydrological Processes, 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. Journal of Geophysical Research: Atmospheres, 122
[RD-20]	Le Gleau, H. (2019) Algorithm Theoretical Basis Document for the Cloud Product Processors of the NWC/GEO. 5th August 2019. http://www.nwcsaf.org/Downloads/GEO/2018/Documents/Scientific_Docs/NWC-CDOP2-GEO-MFL-SCI-ATBD-Cloud_v2.1.pdf
[RD-21]	Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, 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. *Remote sensing of environment*, *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. *Remote Sensing of Environment*, *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. *Remote Sensing of Environment*, *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, Atmos. Chem. Phys., 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, Q. J. R. Meteorol. Soc., 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. *IEEE Transactions on Geoscience and Remote Sensing*, *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", *IEEE Trans. Geosci. Remote Sens.*, 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, 5th August 2019. https://www.ecmwf.int/en/elibrary/8174-era-interim-archive-version-20
[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. *Journal of Geophysical Research: Oceans, 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). *Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI).* 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, 5th August 2019. http://data.ceda.ac.uk/neodc/esacci/land_cover/docs/ESACCI-LC-PUG-v2.5.pdf
[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 Net- works*, 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, J. Geophys. Res., 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, J. Geophys. Res., 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, J. Geophys. Res.,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, Atmos. Meas. Tech., 12, 1531-1543, https://doi.org/10.5194/amt-12-1531-2019, 2019.
[RD-54]	Soria, G. and J.A. Sobrino, ENVISAT/AATSR derived land surface temperature over a heterogeneous region. Remote Sensing of Environment, 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). *Bulletin of the American Meteorological Society, 83*, 977-992
[RD-56]	Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkman, J., Rota, S., & Ratier, A. (2002b). Radiometric performance of SEVIRI. Bulletin of the American Meteorological Society, 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. International Journal of Remote Sensing, 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. *Journal of Applied Meteorology and Climatology, 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, Atmospheric Chemistry and Physics, 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, Atmospheric Chemistry and Physics, 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, Remote Sensing, 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. *IEEE Trans. Geosci. Remote Sens.* **46**, 307–315 (2008).
[RD-67]	Dee, Dick P., et al. "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system." *Quarterly Journal of the royal meteorological society* 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. Remote Sensing of Environment, 116, 62–78. https://doi. org/10.1016/j.rse.2011.02.028

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. **2005**, 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. https://doi.org/10.5194/amt-3-1005-2010
[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

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

Term	Definition
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
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 will be used to derive the following LST_cci products from:

- ❖ ATSR-2 and AATSR
- ❖ SLSTR
- ❖ MODIS
- ❖ SEVIRI
- ❖ 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 μm brightness temperatures (once a correction of order 0.2 K is applied to 12 μm 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 μm. Both ATSR-2 and AATSR have three additional visible channels at 0.55, 0.66 and 0.87 μm 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 ~55° 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 ion 16th February 2016, and Sentinel-3B was launched on 25th 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 $\lambda_{central}$ (m)	Dynamic Range	Radiometric Noise
VIS0.6	0.635	533 Wm-2sr-1λm-1	S/N 10 at 1% albedo
VIS0.8	0.81	357 Wm-2sr-1λm-1	S/N 7 at 1% albedo
NIR1.6	1.64	75 Wm-2sr-1λm-1	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-2sr-1λm-1	S/N 1.2 at 0.3% albedo

3.5. SSM/I and SSMIS

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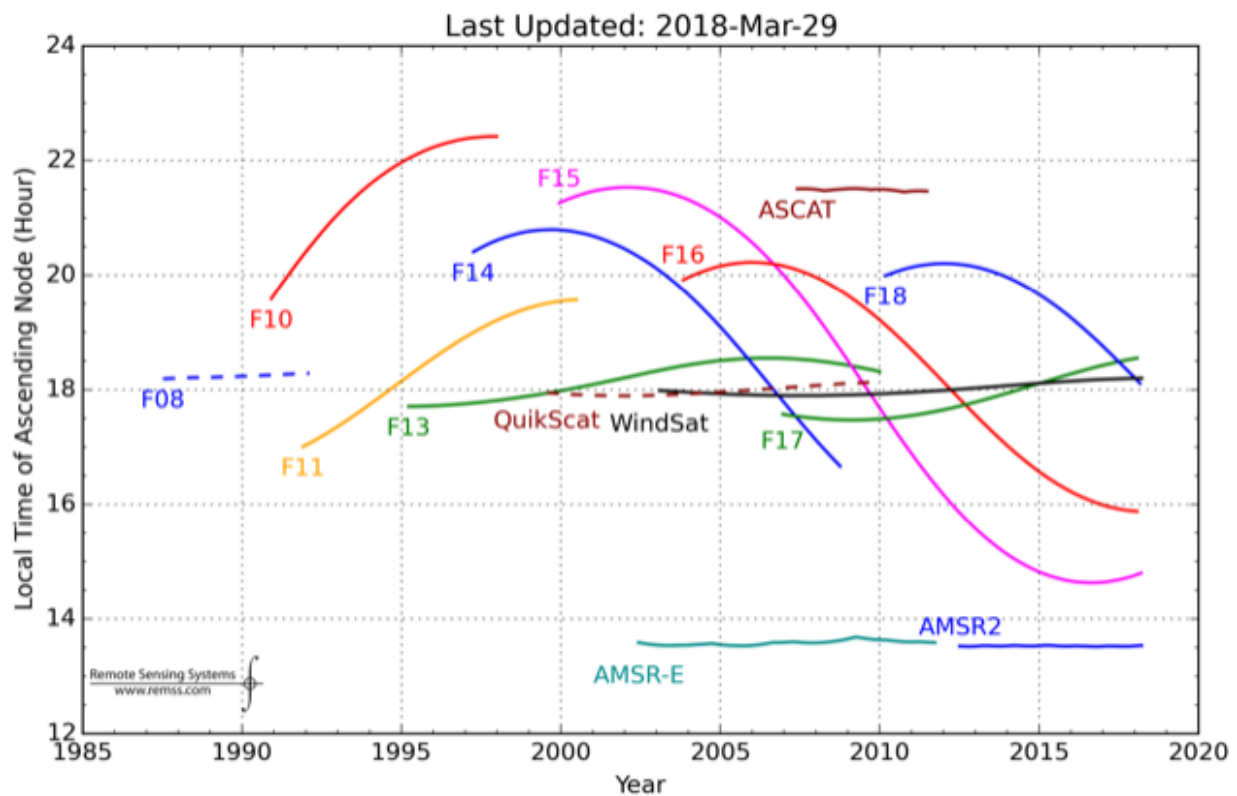
Figure 3: Local ascending equator crossing time for SSM/I and SSMIS (courtesy of Remote Sensing Systems, http://www.remss.com/support/crossing-times/). Only the FXX instruments are relevant here.

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.

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:
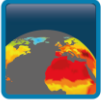
❖ the University of Leicester (UOL) Algorithm (TIR)
  ❍ for the Advanced Along-Track Scanning Radiometer (AATSR) LST ECV dataset
  ❍ for the AATSR / Sea and Land Surface Temperature Radiometer (SLSTR) / Moderate Resolution Imaging Spectroradiometer (MODIS) CDR

❖ 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 µm 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 µm 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

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 (≤ 1 km) and temporal (≤ 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 will provide pixel by pixel LSTs using only the nadir SW (11 and 12 μm) channels of the satellite instruments. The products will utilise the cloud-free top-of-the-atmosphere 11 and 12 μm 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,

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

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

$$I_v^{atmos} = \int_0^\infty B_v[T(z)] \frac{\partial \tau}{\partial z}(z, \infty) \mathrm{d}z$$

Where:

❖ I_v is the radiance at the radiometer,

❖ $I_v^{surface}$ is the surface leaving radiance,

❖ I_v^{atmos} is the radiance from the atmosphere,

❖ τ is the atmospheric transmittance,

❖ v 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,

❖ ϵ_v is the surface emissivity,

❖ B_v is the Planck function,

❖ ρ_v is the surface reflectance,

❖ I_v^{sky} is the downwelling sky radiance.

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

$$\int_{\Omega-} \mathbf{n} \cdot \mathbf{s} \epsilon_v(\mathbf{s}) \, \mathrm{d}\Omega = \int_{\Omega+} \mathbf{n} \cdot \mathbf{s} \left\{ 1 - \frac{1}{\pi} \int_{\Omega-} \mathbf{n} \cdot \mathbf{s}' \, \rho_v(\mathbf{s}, \mathbf{s}') \mathrm{d}\Omega' \right\} \mathrm{d}\Omega$$

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 ϵ_v and ρ_v are independent of direction,

$$\epsilon_v = 1 - \rho_v$$

The flux density of sky radiation is:
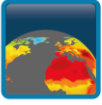
$$F_v^{sky} = \int_0^{2\pi} \int_0^{\pi/2} I_v^{sky} \cos\theta \sin\theta \, \mathrm{d}\theta \mathrm{d}\phi$$

Where θ is the satellite zenith view angle, and ϕ is the satellite azimuth view angle.

$$I_v^{surface} = \epsilon_v B_v[T_s] + \{1 - \epsilon_v\} L_v^{sky}$$

$$L_v^{sky} = \frac{F_v^{sky}}{\pi}$$

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

$$T_s = B_\nu^{-1}\left\{\frac{I_\nu^{surface} - (1 - \epsilon_\nu)L_\nu^{sky}}{\epsilon_\nu}\right\}$$

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 (≈ 1 μm 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_ν, ϵ_ν, L_ν^{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 μm channels. The relation takes account of atmospheric absorption (water vapour) and spectral emissivity effects.
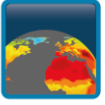
4.1.2. Emissivity

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 μm [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 ≈ ±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 ±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]:

$$\bar{I}_\nu^{atmos} = \frac{1}{1-\tau_\nu} \int_0^\infty B_\nu[T(z)] \frac{\partial \tau(z,z')}{\partial z} \mathrm{d}z$$

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,

$$\tau(z,\infty) = \exp\left\{-\int_z^\infty k_\nu(z')w(z')\sec\theta\,\mathrm{d}z'\right\}$$

Where:

❖ k_ν 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 μm channels and introduce appropriate subscripts):
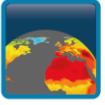
$$I_{11} = [\epsilon_{11}B_{11}(T_s) + (1-\epsilon_{11})L_{11}]\tau_{11} + (1-\tau_{11})\bar{I}_{11}^{atmos}$$

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

Linearise around ν_{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:

- ❖ $\gamma = \frac{1-\tau_{11}}{\tau_{11}-\tau_{12}}$,
- ❖ $\Delta\epsilon = \epsilon_{11} - \epsilon_{12}$,
- ❖ $\alpha = -\frac{(1-\tau_{11})\tau_{12}(1-\epsilon_{12})I_{12} - (1-\tau_{12})\tau_{11}(1-\epsilon_{11})I_{11}}{\epsilon_{12}\tau_{12}(1-\tau_{11}) - \epsilon_{11}\tau_{11}(1-\tau_{12})}$.
- ❖ I'_{11} is the radiance at $\nu = \nu_{11}$ that yields a temperature equal to T_{12}. Thus,
- ❖ $I'_{11} = B_{11}[T_{12}]$.

Below are some special cases which are worth considering:

- ❖ No spectral emissivity dependence:

$$\Delta\epsilon = 0$$

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

- ❖ Emissivity ≈ 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.

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

After some manipulation,

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

- ❖ $a_0 = \alpha \left(\frac{\partial B}{\partial T}\right)_{\overline{T}}^{-1}$,
- ❖ $b_0 = \frac{1+\gamma}{\epsilon_{11}}\left[\frac{1}{1+\gamma\tau_{12}\Delta\epsilon/\epsilon_{11}}\right]$,
- ❖ $c_0 = \frac{\gamma}{\epsilon_{12}}\left[\frac{1}{1+(1+\gamma)\tau_{11}\Delta\epsilon/\epsilon_{12}}\right]$.

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 µm channels linearly to the land surface temperature.

Although ϵ_{11} and ϵ_{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 will be used in LST_cci Cycle 1 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:

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

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].
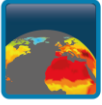
The retrieval parameters d and m are empirically determined from validation and control the behaviour of the algorithm for each zenith viewing angle (θ) across the nadir swath. The parameter d resolves increases in atmospheric attenuation as the water vapour increases, an effect that is accentuated with 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 μm than at 11 μm.

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 will be 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 (ΔLST).

4.2.3. Generalised Split Window (GSW) Algorithm

The GSW algorithm will be used in LST_cci Cycle 1 for the MODIS and SEVIRI LST ECV datasets as well as the Merged Dataset (AATSR / MODIS / SEVIRI). 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 μm 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 will be 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 λ_2 and λ_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 will be used in LST_cci. Here, ε_{mean} will be derived as:

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

Where ε_{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 μm 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 will be of width 15 kg·m^{-2} so that the first water vapour band is from [0,15) kg·m^{-2}. The bands for satellite zenith angle will be of width 5°. The retrieval coefficients are linearly interpolated between viewing angle and water vapour bands to minimise step changes.

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 will be 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. The profile data will be provided by ERA-Interim [RD-67] for Cycle 1, which provides a large number of input profiles which encompass the full range of atmospheres and surfaces observed by TIR instruments (Section 4.6). A land cover and atmospheric conditionally uniform random sampling strategy will be used to select a number of clear sky profiles for each land cover class. A large sample of locations are selected randomly across land and ice surface types over all latitude and longitude bands to represent the full range of surface types across land areas [RD-19]. A temporal sampling strategy ensures intra- and inter-annual coverage; for AATSR for example profiles can be selected from the an arbitrary day around the middle of each month (15th day of each month for instance) between 2002 and 2011 with identified profiles closest to the day and night overpass times of the satellite of interest. The temporal sampling strategy will be expanded to additional years in future Cycles to maximise representativeness inter-annually. Representative emissivity information is extracted for the locations of the profile data from auxiliary datasets (Section 4.6). These selected profiles are then inputs to the RTTOV forward model along with the various sensor spectral response functions. RTTOV then yields the brightness temperatures and LSTs for the given sensor and location and these are used to generate retrieval coefficients for all cases of land cover type, fractional vegetation and water vapour using linear regression. The Calibration Database comprises a global set of independent profiles and emissivity values covering all land cover types and distributed across all latitude and longitudes and capturing the seasonality of the land surface, as well as the coefficients generated from these profiles.

In LST_cci after Cycle 1 (in other words for Cycles 1.5 and 2) an extended version of the Benchmark database constructed for the Round Robin [RD-68] will be used to determine retrieval coefficients for TIR algorithms. This extended benchmark dataset will use ERA5 Atmospheric profile Data, CAMEL Emissivity Data and ESA CCI Land Cover data. ERA5 profiles and CAMEL emissivity data are used as an input to RTTOV. Representative profiles distributed across the globe are extracted, including simulated brightness temperatures, LST, elevation and other atmospheric information. Representative emissivity information is also extracted from CAMEL for the locations of the profile data. The rationale here is that the latest ECMWF data and emissivity data is exploited, which are anticipated to be more representative given the improved resolution. This also ensures consistency with other CCI. These parameters are inputs to the RTTOV forward model along with the various sensor spectral response functions. RTTOV then yields the

brightness temperatures and LSTs for the given sensor and location and the Calibration Database is generated as described above for Cycle 1.

The calibration of the GSW used within the LSA-SAF to retrieved LST for the GEOs (MSG, MTSAT and GOES series) relies on radiative transfer simulations of TOA brightness temperatures performed with the MODerate spectral resolution atmospheric TRANSsmittance algorithm (MODTRAN4) [RD-63]. The difference to using RTTOV is expected to be negligible, but any difference will be assessed and if required a change in model would be anticipated. The simulations are performed for the database of global profiles of temperature, moisture, and ozone compiled by SeeBor [RD-64] for clear sky conditions. This SeeBor database described above was split into two subsets – one used for the calibration of the LST GSW, and an independent one used for verification of the fitted algorithm. A full description of the methodology to select the calibration profiles may be found in [RD-65]. The parameters in the GSW algorithm are estimated for 8 different classes of total column water vapour (W), up to 6 cm, and for 16 classes of VZA, up to 75º, ensuring that all ranges of atmospheric attenuation within the thermal infrared are covered. In order to ensure that all W and VZA class have enough representative cases to provide robust parameter estimations, the radiative transfer simulations are performed over the all selected atmospheric profiles with the following settings: (i) surface temperature ranging between Tskin–15 and Tskin+15 K in steps of 5K; (ii) channel emissivities of both TIR channels covering the range $0.96<\varepsilon_{12}<0.995$ in steps of 0.0175 and ε_{12}-$0.030<\varepsilon_{11}<\varepsilon_{12}$+0.018 in steps of 0.006; and (iii) VZA ranging from nadir to 75º in steps of 5º. It is worth noting that the whole simulations cover a range of Tskin between 230 K and 341 K, and a range of [Tskin minus T2m] from -20 to +33 K. These are the optimum settings for representativeness when using the SeeBor database.

After Cycle 1, all GEO products will be reprocessed with a GSW calibrated with the benchmark dataset constructed for the Round Robin, in order to provided harmonized data for the Merged Product. A natural consequence of this is that all single-sensor GEO ECV Products will also be consistent with LEO ECV Products 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 being considered for use in LST_cci Cycle 1 for TIR LST Products are presented here:

- ❖ The UOL_3 algorithm (Section 4.5.1).
- ❖ Bayesian algorithm (Section 4.5.2)
- ❖ the NWCSAF 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. Either the Bayesian or the UOL_3 algorithm will be applied to the CDR depending on which is most appropriate. The NWC-SAF Cloud Mask Algorithm will be 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.

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] will be used for profiles in Cycle 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 5x5° grid cell (chosen to ensure sufficient representativeness) for each of the land covers and diurnal conditions (day/night) required by an offline enhanced LST retrieval algorithm [RD-58, RD-59]. In Cycle 1 this has been stratified by the 27 land covers of the ATSR Land Biome Classification (ALB2) [RD-19]. In future Cycles this will be replaced by LC_cci classes. 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 5x5° grid cell for the given month, land cover and diurnal state as shown in Figure 4 [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 μm brightness temperature or 11 - 12 μm 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 μm brightness temperature and the 11 - 3.7 μm 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.
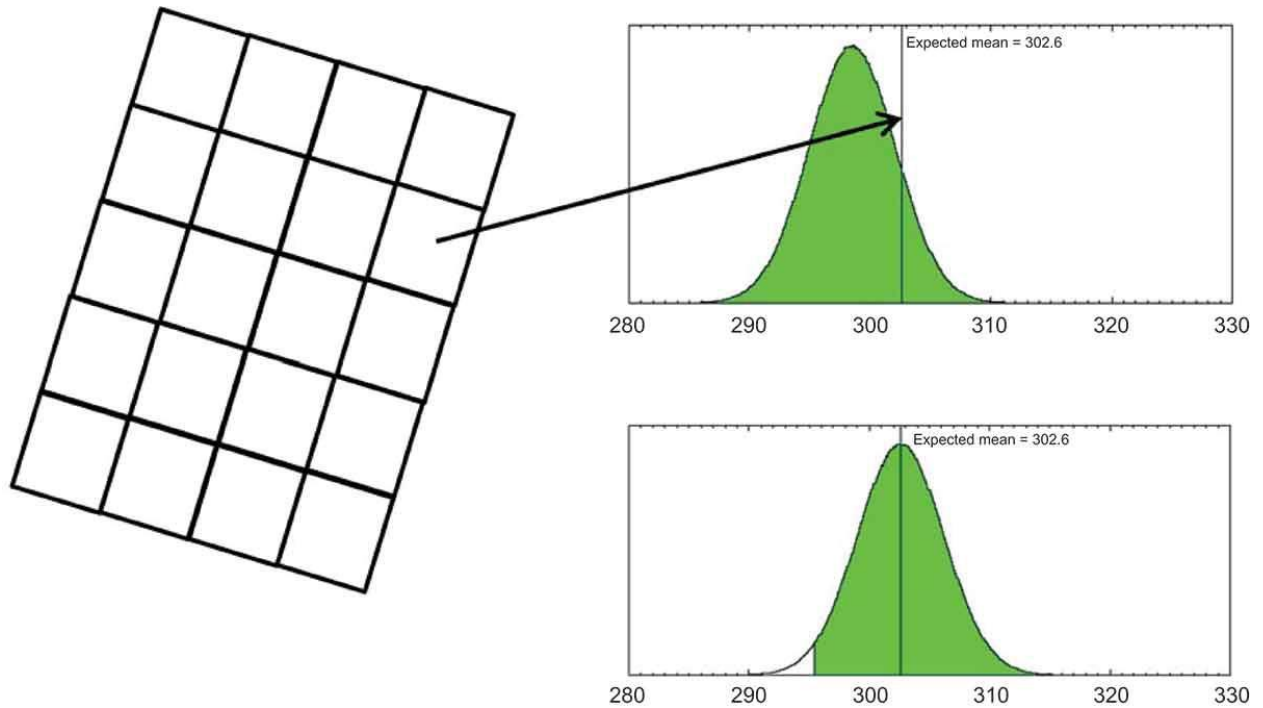
Figure 4: For each granule of an AATSR orbit (left), the expected 12 μm brightness temperature is simulated from coincident profiles. The PDF of observed 12 μm brightness temperatures for each land cover-diurnal condition, given the space and time position, is also determined (top-right in green). This 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].

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 ERA-Interim total cloud cover [RD-32] in Cycle 1. ERA5 will be used in Cycle 2.

$P(\mathbf{y}^o|\mathbf{x}_b)$ is the probability of the observations given the background state. For clear-sky observations this is calculated using the RTTOV version 11 fast forward model simulations in Cycle 1 (see Section 4.5.2; moving to RTTOV 12.3 in Cycle 2), 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. Full details will be presented in the next version

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

It should be noted that for initial LST products provided by LST_cci in Cycle 1, land cover information will be provided by the ATSR Land Biome Classification (ALB2) [RD-19]. Appendix A provides a table defining the ALB2 Land cover Classification.

Land cover information for LST_cci products will be provided by the ESA CCI Land Cover maps developed by the Land Cover CCI. These land maps 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 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]. CCI land Cover information will be used in the UOL retrieval algorithm, in combination with other variables, to determine the most appropriate coefficients to apply.

It will also be used in the UOL_3 and Bayesian cloud masking algorithms. Appendix A provides a table defining the 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 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 will be 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 will be 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 in Cycle 1. For future cycles we will investigate whether a consistent FCOVER dataset is needed to maximise consistency or if an appropriate external emissivity dataset is more suitable.

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 will be used after 2016 if regular updates of this dataset are not 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 µm 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 µm.

CAMEL emissivity data will be 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.

For initial LST_cci products in Cycle 1, the CIMMS Baseline Fit Emissivity Database [RD-60] will be used.

In the future it is expected that emissivity for the applications detailed in this section will be provided from the work being done as part of LST_cci Work Package 2.9: Temperature and Emissivity Separation from MODIS multispectral TIR data (CCN to Baseline Project).

CIMMS

For initial LST_cci products in Cycle 1, the CIMMS Baseline Fit Emissivity Database [RD-60] will be used. CIMMS is a monthly dataset 0.05° with emissivities available at ten wavelengths between 3.6μm and 14.3μm, - including emissivity at 10.8μm and 12.1μm. It has been derived using the MODIS operational land surface emissivity product and by applying a baseline fit method to fill in the spectral gaps between the six infrared emissivity wavelengths provided. The dataset is available as monthly filled files from 2003 to 2016 in netCDF format. A monthly climatology has been derived for use outside of the available data window. The data is spatially and temporally interpolated onto a 1 km grid for the given day of the satellite acquisition for use in LST_cci cycle 1.

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 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 ε_{veg} and ε_{bg} are the vegetation and bareground emissivities, respectively, per channel. Both ε_{veg} and ε_{bg} are estimated for land cover classes from spectral libraries.

VCM will be used in the single-sensor LST retrieval from SEVIRI 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.

It should be noted that ERA-Interim will be used instead of ERA-5 in Cycle 1 of LST_cci. ERA-Interim is the predecessor of ERA-5. It also provides hourly estimates of a significant number of land and atmospheric variables over the full globe from 1979 to present with a spatial resolution of 80 km with 60 atmospheric levels from the surface up to 0.1 hPa. ERA-Interim was based on a 2006 release of the IFS (Cy31r2) and is due to be replaced by ERA-5 in 2019. The use of ERA-Interim will be superseded by ERA5 for all subsequent cycles.

4.6.5. Snow masking

Snow masking information will initially be provided by the Interactive Multisensor Snow and Ice Mapping System (IMS) Daily Northern Hemisphere snow and ice analysis.

The IMS snow maps are daily maps of Northern Hemisphere land, sea, snow and ice on an equal area polar stereographic grid at 1 km, 4 km and 24 km resolution, depending on time period. The IMS product is manually created by an analyst using the previous day's IMS map, satellite imagery, automated snow mapping algorithms and other ancillary data [RD-18]. It is available from 1997 to present with higher resolution maps available for shorter time periods. For inclusion in LST_cci algorithms and products, 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 the Southern Hemisphere we use a channel ratio method based on [RD-74].
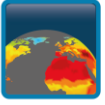
Snow masking is part of the land cover information used in the UOL retrieval algorithm to determine the most appropriate coefficients to apply. It is also utilised in the UOL_3 cloud masking algorithm. Further updates to LST_cci products will include a move to using ESA Snow Cover CCI products for snow 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 will be 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 $u_{ran}(y_c)$. 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 $u_{ran}(x)$ 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:
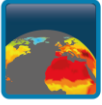
$$u_{ran,y}(x) = \sqrt{\sum_{c=1}^{n} \left(\frac{\partial R}{\partial y_c} u_{ran}(y_c) \right)^2}$$

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:

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

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.

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 LST_cci Cycle 1 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.
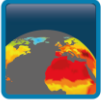
5.1. Physics of the Problem

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 v and polarization p assuming surface specular reflection and non-scattering atmosphere may be written:

$$Tb_{v,p} = T_v^\uparrow + \tau_v \left[\epsilon_{v,p} Te_v + (1 - \epsilon_{v,p}) T_v^\downarrow \right] =$$

$$\left(T_v^\uparrow + \tau_v T_v^\downarrow \right) + \epsilon_{v,p} (\tau_v Te_v - \tau_v T_v^\downarrow) =$$

$$A_v + \epsilon_{v,p} (B_v Te_v + C_v)$$

where τ_v 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, $\epsilon_{v,p}$ is the surface emissivity, Te_v is the effective emission temperature of the surface, and the A_v, B_v and C_v 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_v, B_v and C_v 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].

A MLP of one input layer of 14 nodes (the inputs of function F, i.e., the brightness temperatures and emissivities for the 7 MW channels), one hidden layer of 10 nodes, and one output node (the LST), will be used here. If the input vector of the MLP is called *i* and the output of the MLP *u*, the way the input signal propagates through the MLP is given by:

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

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 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:

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

where Z is the number of samples in the calibration database, $[\![\,]\!]$ 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.

When training the MLP, if the initial weights are slightly changed, or a new set of examples from the calibration database are selected, the minimization of the error function results in a new set of final weights and corresponding transfer function. In most cases the resulting transfer functions are very close, and for well-constrained inversion situations the variability in the output (here the retrieved LST) by applying the different transfer functions is small. Likewise, a large variability is an indication of inversion situations where the MLP has difficulties to solve the inverse problem. This is used in the retrieval algorithm as a form of quality control for the inversions. In practice, 100 neural networks with different initial conditions are trained, the estimated LST is the median of the 100 retrieved values, and the variability of the 100 estimates is monitored to capture cases where the inversion situation seems problematic.

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 100 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 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.

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

An initial assessment of the MW algorithm uncertainty is presented here. The calibration database for the MW algorithm is based on real SSM/I observed radiances and the corresponding LST estimates from an inversion of the previous observations, as described in Section 8.2, and it is also applicable to the SSMIS instrument. MW emissivities from climatology are also used as an input to the retrieval algorithm, together with the observed radiances. Given the peculiarities of this database, applying the uncertainty model described in Section 4.7 for the TIR algorithms to the MW LST estimates is not straightforward. An initial assessment of uncertainty is presented here for the MW product of the LST_cci Cycle 1, which will be further revised in the framework of the E2UB for the subsequent production cycles.

5.5.1. Theoretical Uncertainty

Theoretical uncertainty is estimated by looking at the retrieval errors of the validation database. Retrieval error is defined as the difference of the database LST and the algorithm retrieved LST from the corresponding brightness temperatures. For approximately 40(70)% of the database, the retrieval error is smaller than 1.0(2.0) K. To characterize this error and provide an estimate of uncertainty, the standard deviation of the retrieval error for different retrieval conditions is calculated. Over the whole validation database, the standard deviation of the retrieval error is 2.2 K.

In principle, given that the MW LST estimates come from a coefficient based retrieval algorithm, this uncertainty could be described as the retrieval ambiguity of the uncertainty model described in Section 4.7 for the TIR retrievals. However, the peculiarities of the MW calibration database imply that the described uncertainty is also likely to incorporate other components, such as the random components associated to L1 channel noise and emissivity variability. As the calibration database use real observations, the regression coefficients of the MW algorithm are calibrated with L1 radiances already containing instrumental noise. Therefore, the mapping to be approximated by the non-linear regression is already noisy regarding the radiances, and part of the retrieval ambiguity comes from this noise. A similar reasoning can be made concerning the emissivity. The inversion uses climatological emissivity, not an estimation of the true emissivity at the observation acquisition, which certainly affects how accurately the mapping between radiances and the LST can be approximated when calibrating the retrieval algorithm. In that sense, we can also justify that the derived theoretical error also includes an uncertainty component related to the emissivity variability. Also, it should be noticed that the target LST of the calibration database already comes from an inversion, i.e., it is not an error-free LST. In that sense, the derived theoretical uncertainty has to be considered as a lower limit. Moreover, this uncertainty analysis excludes any systematic components, which can typically only be inferred by comparisons with other sources of LST.
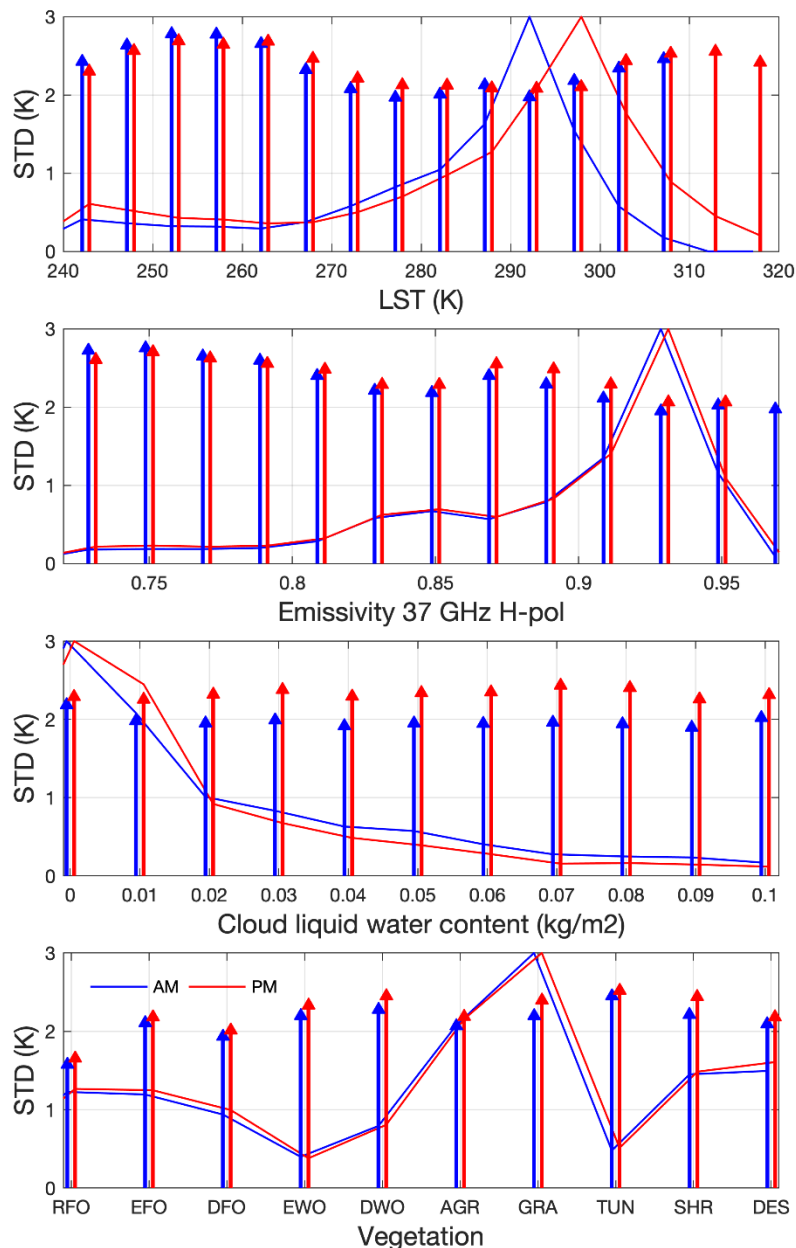
Figure 5: Retrieval uncertainty defined as the standard deviation of the difference of the database LST and the algorithm retrieved LST. The standard deviation is calculated over the LST estimates binned for different ranges of LST, 37 GHz vertically polarized emissivity, cloud liquid water, and vegetation class (rain forest (RFO), ever- green forest (EFO), deciduous forest (DFO), evergreen woodlands (EWO), deciduous woodlands (DWO), agriculture (AGR), grasslands (GRA), tundra (TUN), shrublands (SHR), and deserts (DES)). The standard deviation (y-axis) is plotted for the AM (PM) overpasses as blue (red) arrows. The normalized distribution of the variable stratifying the retrievals (x-axis) is also plotted as a solid line (numbered axis not displayed).

Figure 5 presents the retrieval uncertainty for different ranges of LST, emissivity, cloud liquid water content, and vegetation class. The retrieval uncertainty is larger for the warmer and colder LSTs. Larger values are also found for the lowest emissivity values associated to some of the challenging inversions,

with the difficulties typically caused by varying emissivity conditions not well captured by the climatological values (e.g., snow-covered areas or humid surfaces). Variations of the uncertainty with cloudiness are very small, showing the very small impact of cloud presence in the retrievals. Regarding vegetation conditions, the smallest averaged standard deviation happens for the rainforest. This is reflecting the impact of emissivity in the retrievals, with rainforest representing quite stable surface conditions and the emissivity climatology likely constraining the inversion problem well.

5.5.2. Retrieval variability

The retrieval variability is defined as the standard deviation of the LST estimates from the 100 MLP retrievals at each swath position (see Section 5.2.1). Figure 6 presents the average variability for the same ranges and variables as in Figure 5. When plotted as function of the LST and 37 GHz vertically polarized emissivity, a larger variability is observed for the lower LST and emissivity, coinciding with the largest retrieval uncertainty previously shown in Figure 5. Cold LSTs are in many cases associated with snow-covered areas, where the climatological emissivity applied in the retrieval can poorly represent the true conditions. Low emissivity values are associated with some of the places where difficult inversions are expected, such as snow-covered areas, humid surfaces, or coastal regions. But the larger retrieval variability for the low emissivity values can also be related to the smaller population representing those conditions in the database, as shown by the distribution of the emissivity values also displayed in Figure 6. As described in Section 5.2.1, the calibration of the retrieval algorithm minimizes a global cost function, and the minimization is likely to be more driven by the most represented conditions in the calibration database. Regarding the cloud liquid water content, the variability is close for clear and cloudy conditions, showing the negligible impact of cloud presence for most inversions. An exception is strong convection activity in the overlying atmosphere, where non-quantified scattering induced brightness temperature depressions can contaminate the retrievals as the inversion assumes that most of the microwave emission comes from the surface. Concerning the variability for different vegetation types, they are comparable apart from the tundra, possibly reflecting the inversion difficulties for large parts of the year related to the snow/inundation conditions.
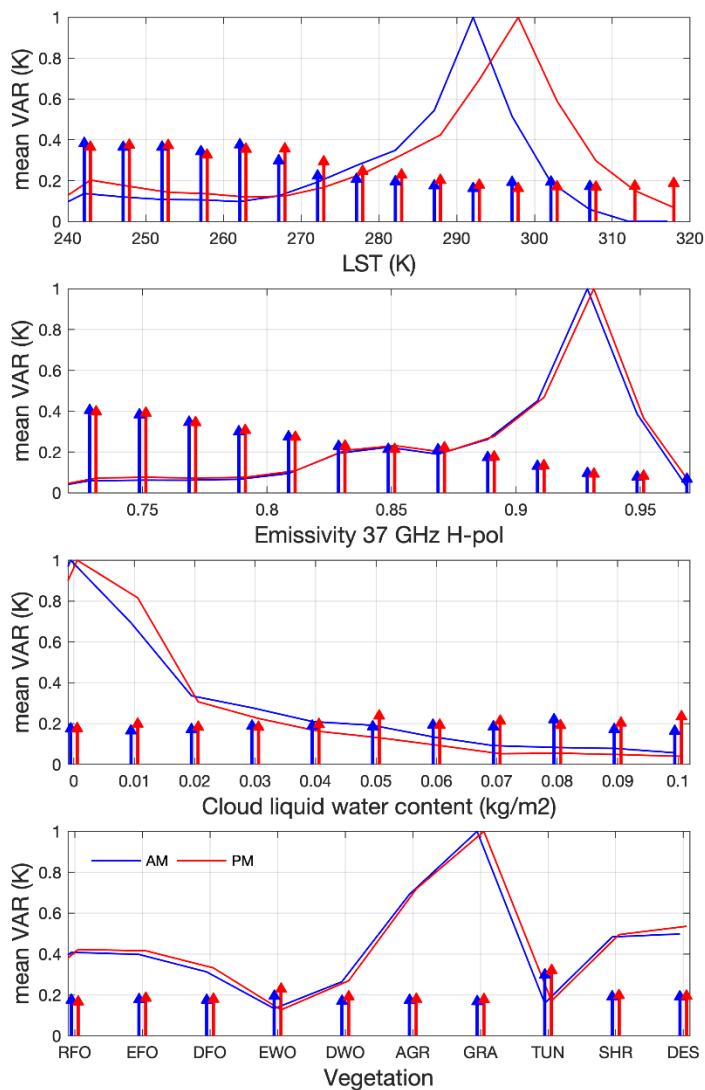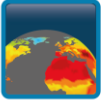
Figure 6: Retrieval variability defined as the standard deviation of the LST estimates from the 100 MLP inversions. The retrieval variability is calculated for LST estimates binned for the same conditions shown in Figure 5. See the text for more details.

5.5.3. Practical implementation

The previous uncertainty analysis is practically implemented by associating each retrieved LST estimate with:

❖ An uncertainty value derived from a lookup table built by parameterizing the estimated theoretical uncertainty presented in Section 5.5.1 as function of the corresponding emissivity, LST, and time of the overpass (~ 6 AM/PM).

❖ The retrieval variability value described in Section 5.5.2.

❖ A series of flags to describe the retrieval situation, signalling:

○ Snow-covered surfaces, by using a snow flag currently derived from the snow density and snow depth from the ERA-Interim reanalysis (expected to be replaced by the ESA CCI snow product when available).

○ Inundated surfaces, by using satellite estimates of surface water extent from the Global Inundation Extension from Multi-Satellites (GIEMS-2) [RD-51].

○ Coastal areas, using a database of distance to the coastline derived from [RD-73].

○ Surfaces with large microwave penetration depth where emission can be emanating from subsurface layers, by using a monthly climatology of radar backscattering from [RD-52].

○ Surfaces where there can be convection activity at the overlying atmosphere, by using the cloud flag described in [RD-53].

6. Merged Land Surface Temperature from Thermal Infrared and Microwave Sensors

To be described in next version.

Appendix A - Land cover Classification Definition

Table 2: Land Cover CCI land cover definition.

Land cover number	Definition
1	Cropland, rainfed
2	Herbaceous cover
3	Tree or shrub cover
4	Cropland irrigated or post-flooding
5	Mosaic cropland (>50%) / natural vegetation (Tree, shrub, herbaceous cover) (<50%)
6	Mosaic natural vegetation (Tree, shrub, herbaceous cover) (>50%) / cropland (<50%)
7	Tree cover, broadleaved, evergreen, closed to open (>15%)
8	Tree cover, broadleaved, deciduous, closed to open (>15%)
9	Tree cover, broadleaved, deciduous, closed (>40%)
10	Tree cover, broadleaved, deciduous, open (15-40%)
11	Tree cover, needleleaved, evergreen, closed to open (>15%)
12	Tree cover, needleleaved, evergreen, closed (>40%)
13	Tree cover, needleleaved, evergreen, open (15-40%)
14	Tree cover, needleleaved, deciduous, closed to open (>15%)
15	Tree cover, needleleaved, deciduous, closed (>40%)
16	Tree cover, needleleaved, deciduous, open (15-40%)
17	Tree cover, mixed leaf type (broadleaved and needleleaved)
18	Mosaic T and shrub (>50%) / herbaceous cover (<50%)
19	Mosaic herbaceous cover (>50%) / T and shrub (<50%)
20	Shrubland
21	Shrubland evergreen
22	Shrubland deciduous
23	Grassland
24	Lichens and mosses
25	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)
26	Sparse tree (<15%)
27	Sparse shrub (<15%)
28	Sparse herbaceous cover (<15%)
29	Tree cover, flooded, fresh or brakish water
30	Tree cover, flooded, saline water
31	Shrub or herbaceous cover, flooded, fresh/saline/brakish water
32	Urban areas
33	Bare areas
34	Consolidated bare areas
35	Unconsolidated bare areas
36	Water bodies
37	Permanent snow and ice
38	No data

Table 3: ATSR Land Surface Temperature Land cover classification V2 (ALB-2) land cover definition.

Land cover Number	Definition
1	Post-flooding OR irrigated croplands
2	Rainfed croplands
3	Mosaic Cropland (50-70percent) OR Vegetation (grassland shrubland forest) (20-50%)
4	Mosaic Vegetation (grassland shrubland forest) (50-70percent) OR Cropland (20-50percent)
5	Closed to open (more than 15percent) broadleaved evergreen AND OR semi-deciduous forest (more than 5m)
6	Closed (more than 40percent) broadleaved deciduous forest (more than 5m)
7	Open (15-40percent) broadleaved deciduous forest (more than 5m)
8	Closed (more than 40percent) needleleaved evergreen forest (more than 5m)
9	Open (15-40percent) needleleaved deciduous or evergreen forest (more than 5m)
10	Closed to open (more than 15percent) mixed broadleaved and needleleaved forest (more than 5m)
11	Mosaic Forest OR Shrubland (50-70percent) OR Grassland (20-50percent)
12	Mosaic Grassland (50-70percent) OR Forest OR Shrubland (20-50percent)
13	Closed to open (more than 15percent) shrubland (more than 5m)
14	Closed to open (more than 15percent) grassland
15	Sparse (more than 15percent) vegetation (woody vegetation shrubs grassland)
16	Closed (more than 40percent) broadleaved forest regularly flooded -Fresh water
17	Closed (more than 40percent) broadleaved semi-deciduous AND OR evergreen forest regularly flooded -Saline water
18	Closed to open (more than 15percent) vegetation (grassland shrubland woody vegetation) on regularly flooded or waterlogged soil -Fresh brackish or saline water
19	Artificial surfaces and associated areas (urban areas more than 50percent)
20	Bare areas of soil types not contained in land covers 21 to 25
21	Bare areas of soil type Entisols - Orthents
22	Bare areas of soil type Shifting sand
23	Bare areas of soil type Aridisols - Calcids
24	Bare areas of soil type Aridisols - Cambids
25	Bare areas of soil type Gelisols - Orthels
26	Water bodies (inland lakes rivers sea maximum 10km away from coast)
27	Permanent snow and ice
28	Sea ice

End of document