CCI Land Surface Temperature

Product Validation Plan (PVP)

WP4 – DEL-2.5

Ref.: LST-CCI-D2.5-PVP

Date: 10-Nov-2022

Organisation: LST_cci consortium

Signatures

	Name	Organisation	Signature
Written by	Lluis Perez-Planells	KIT	
	Maria Martin	KIT	
Reviewed by	J. Bruniquel	ACRI-ST	
Approved by	Darren Ghent	ULeic	
Authorized by	Simon Pinnock	ESA	

Change log

Version	Date	Changes
1.0	17-Jun-2019	First version
1.1	10-Nov-2022	Updated version taking into account RIDs raised by ESA
1.2	10-Sep-2020	Updated version, including changes to the data file structure following changes in the Product Specification Document and addition of an in situ station that is used for validation in the first validation cycle
2.0	10-Nov-2022	Updated version, including changes for the validation of CCI_lst Phase 2 products.

Table of Content

List of Figures

List of Tables

0. Executive Summary

This document details the validation work that will be carried out as part of the land surface temperature (LST) Climate Change Initiative project (LST_cci) within the European Space Agency (ESA) Climate Change Initiative. Validation is an important part of the development of new LST data sets, as it can give insights into the quality of the derived products.

Validation work within LST_cci will follow protocols from the Committee on Earth Observation Satellites (CEOS), within its Working Group on Calibration and Validation (WGCV), as well as guidelines for calibration and validation provided by CEOS-WGCV under the Quality Assurance Framework for Earth Observation (QA4EO), which are endorsed by CEOS, and the LST Validation Protocols published in [RD-1] and [RD-2]. Validation in LST_cci will be carried out by the LST_cci validation team, working separately from the LST_cci EO Science Team who produce the LST data sets. Level-2 satellite data products are validated against independent in situ data from several stations, while Level-3 satellite data sets will be intercompared with external operational data sets and models.

The first validation approach is straight forward, where satellite LST data are validated against LST from in situ station data. The second involves intercomparisons of LST_cci data sets with external LST products. All data sets and software used for the validation are stored in the multi-sensor Matchup Database (MMDB). This will have a common, harmonized file format confirming to the CCI data format [RD-14]. The in situ data are split into two sections; one will be accessible to the data set producers for algorithm testing, and the second, larger part, will be accessible only to the validation team to ensure independence from the data set producers.

The in situ data sets cover stations from several different networks which are distributed globally. They include sites from the Karlsruhe Institute of Technology (KIT), Atmospheric Radiation Measurement (ARM), Surface Radiation Budget Network (SURFRAD), Baseline Surface Radiation Network (BSRN), Heihe River Basin network, Copernicus LAW network, as well as a station over Lake Constance. The possibility of adding further stations throughout the project will be explored.

The Low Earth Orbit (LEO) satellite data sets developed in LST_cci include data derived from the Along Track Scanning Radiometer - 2 (ATSR-2) and Advanced Along Track Scanning Radiometer (AATSR), Advanced Very High Resolution Radiometer (AVHRR), Moderate Resolution Imaging Spectrometer (MODIS), Sea and Land Surface Temperature Radiometer (SLSTR), visible Infrared Imaging Radiometer Suite (VIIRS), Special Sensor Microwave / Imager (SSM/I) and Advanced Microwave Scanning Radiometer (AMSR) microwave satellite sensors. The Geostationary Orbit (GEO) Satellite Data Sets include those derived from Geostationary Operational Environmental Satellite (GOES) data, Multi-functional Transport Satellite (MTSAT) and Himawari (HMWR) data, and Spinning Enhanced Visible and Infrared Imager (SEVIRI) data. Two Climate Data Records will also be developed and validated in LST_cci. For satellite – satellite intercomparisons, the following external satellite data sets will be used: operational MODIS, ATSR-2, AATSR, SLSTR, and SEVIRI data, the CM SAF SEVIRI Thematic Climate Data Record (TCDR), International Satellite Cloud Climatology Project (ISCCP) LST and ERA5 surface skin temperature data.

The results of all validations will be fed back to the data set producers for discussions on the quality of the data set algorithms and on possibilities for improvements. It will also be documented in the LST_cci Product Validation and Intercomparison Report (PVIR) for data users to explore the differences between the developed LST data sets.

1. Introduction

1.1. Purpose and Scope

Within the land surface temperature (LST) Climate Change Initiative project (LST_cci), which is part of the European Space Agency (ESA) Climate Change Initiative (CCI), different satellite LST products are developed. These data can be used for various applications within climate research.

Researchers who want to use the data products need reliable information about the product quality. This information is gained by validation of the data sets, which evaluates the quality of the products and is a critical step towards high credibility of the LST products. It is furthermore a crucial requirement for users to better understand the data they are exploiting and its limitations.

Information about data quality can also help to promote the uptake of LST data by the climate user community. Therefore, prior to delivery of the final products to the climate community the LST data sets are validated and assessed in terms of climate requirements within LST_cci. The validation will be carried out independently from the data producers of the EO Science Team by the Validation Team to gain impartial validation results and interpretations. This document describes details of the methodology used for the validation, their compliance to LST validation protocols, and the key resources used for validation in order to define a clear and transparent protocol for assessing the various LST datasets.

There are different ways to validate LST data set. [RD-1] defines four validation categories, where in situ validation is the most traditional and conclusive approach. This involves a direct comparison of satellite-derived LST with collocated and simultaneously acquired LST from ground-based radiometers that are independent of the algorithm selection and product generation. This should be done in a representative and consistent way over several stations covering major land cover types for each data set to inform users of the absolute performance of the LST products.

However, as sufficient high-quality in situ measurement stations covering different climate zones, land covers and time periods are globally not available, in situ validations can be supplemented by satellite – satellite intercomparisons. These intercomparisons of LST_cci data products to a defined set of external operational and reference LST products can give insights in the relative algorithm performance of the validated data products over different land covers. It is conducted in addition to in situ validation within LST_cci. Furthermore, the stability of climate data records (CDR) developed in LST_cci needs to be assessed, which will be done by validating the CDRs against data from a long-time running in situ measurement station at Lake Constance.

The results of all of these validation exercises will be documented and discussed in the LST_cci Product Validation and Intercomparison Report (PVIR).

1.2. Structure of the Document

The document is structured as follows: First, the validation protocols that will be followed in LST_cci are introduced. Next, the process of the in situ validation is explained, followed by the explanation of the process of the satellite - satellite intercomparisons. After that, the Multi-sensor Matchup Database (MMDB) is introduced, where all data and software needed for the product validation is stored, followed by a listing of the in situ data used, the CCI Satellite Data Sets and Climate Data records validated, and the external data sets used for satellite - satellite intercomparisons. Finally, the content of the document is summarized.

1.3. Applicable Documents

Ref Id	Applicable document
AD-1	LST CCI (2021) User Requirements Document, Reference LST-CCI-D1.1-URD - i2r0
AD-2	CCI LST Product Specification Document (LST-CCI-D1.2-PSD)

1.4. Reference Documents

The following is a list of reference 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.

Ref Id	Reference document
RD-1	Schneider, P.; Ghent, D.; Corlett, G.; Prata, F.; Remedios, J. AATSR Validation: LST Validation Protocol. ESA Report, Contract No.: 9054/05/NL/FF, European Space Agency (ESA), 2012. UL-NILU-ESA-LST-LVP Issue 1 Revision 0. http://lst.nilu.no/Portals/73/Docs/Reports/UL-NILU-ESA-LST-LVP-Issue1-Rev0-1604212.pdf
RD-2	Guillevic, P.; Göttsche, F.; Nickeson, J.; Hulley, G.; Ghent, D.; Yu, Y.; Trigo, I.; Hook, S.; Sobrino, J.; Remedios, J.; Román, M.; Camacho, F. Land Surface Temperature Product Validation Best Practice Protocol, Version 1.1 2018. https://lpvs.gsfc.nasa.gov/PDF/CEOS_LST_PROTOCOL_Feb2018_v1.1.0_light.pdf
RD-3	GCOS. The Global Observing System for Climate: Implementation Needs, GCOS-200/GOOS-214. Implementation plan, 2016. https://library.wmo.int/opac/doc_num.php?explnum_id=3417
RD-4	GCOS. The 2022 Global Observing System for Climate Implementation Plan, GCOS-244, 2022. https://library.wmo.int/doc_num.php?explnum_id=11317
RD-5	GCOS. The 2022 Global Observing System for Climate ECVs Requirements, GCOS-245, 2022. https://library.wmo.int/doc_num.php?explnum_id=11318
RD-6	JCGM. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. Technical Report JCGM 100, IEC BIPM, ILAC IFCC, and IUPAC ISO, Joint Committee for Guides in Metrology, 2008. GUM 1995 with minor corrections. https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

Ref Id	Reference document
RD-7	Göttsche, F.; Olesen, F.S.; Høyer, J.L.; Wimmer, W.; Nightingale, T. Fiducial Reference Measurements for Validation of Surface Temperature from Satellites (FRM4STS), Technical Report 3 - A Framework to Verify the Field Performance of TIR FRM. Reference: OFE- D120-V1-Iss-3-Ver-1-ISSUED, ESA Contract No. 4000113848 15I-LG. Techreport, ESA, 2017. http://www.frm4sts.org/wp-content/uploads/sites/3/2018/04/FRM4STS_D120-TR3_24Apr17_v3-signed.pdf
RD-8	LPV homepage, https://lpvs.gsfc.nasa.gov/. Assessed on 16/05/2019.
RD-9	Hulley, G.; Malakar, N.; Freepartner, R. Moderate Resolution Imaging Spectroradiometer (MODIS) Land Surface Temperature and Emissivity Product (MXD21) Algorithm Theoretical Basis Document Collection-6. Agorithm Theoretical Basis Document, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Californa, 2016. https://lpdaac.usgs.gov/documents/107/MOD21_ATBD.pdf
RD-10	Martin, M.; Göttsche, F.M.; Ghent, D.; Trent, T.; Dodd, E.; Pires, A.; Trigo, I.; Prigent, C.; Jimenez, C. ESA DUE GlobTemperature Project: Satellite LST Intercomparison Report. Technical report, ESA, 2016. http://www.globtemperature.info/index.php/public-documentation/deliverables-1
RD-11	Pearson, R.K. Outliers in Process Modeling and Identification. IEEE Transactions On Control Systems Technology 2002, 10, 55–63.
RD-12	Martin, M.; Göttsche, F.M. ESA DUE GlobTemperature Project: Satellite LST Validation report. Technical report, ESA, 2016. http://www.globtemperature.info/index.php/public-documentation/deliverables-1
RD-13	Martin, M.A.; Ghent, D.; Pires, A.C.; Göttsche, F.M.; Cermak, J.; Remedios, J.J. Comprehensive In Situ Validation of Five Satellite Land Surface Temperature Data Sets over Multiple Stations and Years. Remote Sensing 2019, 11.
RD-14	ESA Climate Office (2018) CCI data standards 2.0 (CCI-PRGM-EOPS-TN-13-0009), available from http://cci.esa.int/sites/default/files/filedepot/CCIDataStandards_v2-0_CCI-PRGM-EOPS-TN-13-0009.pdf
RD-15	Caselles, E.; Valora, E.; Abadb, F.; Caselles, V. Automatic classification-based generation of thermal infrared land surface emissivity maps using AATSR data over Europe. Remote Sensing of Environment 2012, 124, 321–333.
RD-16	Li, Z.L.; Tang, B.H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. Satellite-derived land surface temperature: Current status and perspectives. Remote Sensing of Environment 2013, 131, 14–37.
RD-17	Goldberg, M.; Ohring, G.; Butler, J.; Cao, C.; Datla, R.; Doelling, D.; Gärtner, V.; Hewison, T.; Iacovazzi, B.; Kim, D.; Kurino, T.; Lafeuille, J.; Minnis, P.; Renaut, D.; Schmetz, J.; Tobin, D.; Wang, L.; Weng, F.; Wu, X.; Yu, F.; Zhang, P.; Zhu, T. The Global Space-Based Inter-Calibration System. Bull. Amer. Meteor. Soc. 2011, 92, 467 – 475.
RD-18	Jallego, F.J. Stratified sampling of satellite images with a systematic grid of points. ISPRS Journal of Photogrammetry and Remote Sensing 2006, 59, 369–376.
RD-19	Rebhan, H.; Goryl, P. Sentinel-3 Calibration and Validation Plan, ESA Report S3-PL-ESA-SY-0265. Technical report, ESA, 2013.
RD-20	Merchant, C. J., G.D.K.J.G.E.H.J. Common approach to providing uncertainty estimates across all surfaces. EUSTACE (640171). Deliverable 1.1. Technical report, 2015.
RD-21	G. S. Team. The Recommended GHRSST Data Specification (GDS) 2.0, document revision 5. Technical report, GHRSST, 2012. https://www.ghrsst.org/governance-documents/ghrsst-data-processing-specification-2-0-revision-5/
RD-22	Bork-Unkelbach, A. Extrapolation von in-situ Landoberflächentemperaturen auf Satellitenpixel. Phd thesis, Karlsruher Institut für Technologie, 2012. DOI: 10.5445/IR/1000032489

Ref Id	Reference document
RD-23	Morris, V.R. Infrared Thermometer (IRT) Handbook (ARM TR-015). Technical report, Atmospheric Radiation Measurement, Climate Research Facility, U. S. Department of Energy, 2006.
RD-24	CIMMS Baseline Fit Emissivity Database, http://cimss.ssec.wisc.edu/iremis/
RD-25	Seemann, S.W.; Borbas, E.E.; Knuteson, R.O.; Stephenson, G.R.; Huang, H.L. Development of a Global Infrared Land Surface Emissivity Database for Application to Clear Sky Sounding Retrievals from Multispectral Satellite Radiance Measurements. Journal of Applied Meteorology and Climatology 2008, 47, 108–123.
RD-26	Stoffel, T. Ground Radiation (GNDRAD) Handbook (ARM TR-027) (Draft). Technical report, Atmospheric Radiation Measurement, Climate Research Facility, U. S. Department of Energy, 2006.
RD-27	König-Langlo, G.; Loose, B. The Meteorological Observatory at Neumayer Stations (GvN and NM-II) Antarctica. Polarforschung 2007, 76, 25–38.
RD-28	König-Langlo, G.; Augstein, E. Parameterization of the downward long-wave radiation at the Earth's surface in polar regions. Meteorologische Zeitschrift 1994, p. 343 – 347.
RD-29	Li, X.; Cheng, G.; Liu, S.; Xiao, Q.; Ma, M.; Jin, R.; Che, T.; Liu, Q.; Wang, W.; Qi, Y.; Wen, J.; Li, H.; Zhu, G.; Guo, J.; Ran, Y.; Wang, S.; Zhu, Z.; Zhou, J.; Hu, X.; Xu, Z. Heihe Watershed Allied Telemetry Experimental Research (HiWATER): Scientific Objectives and Experimental Design. Bulletin of the American Meteorological Society 2013, 94, 1145–1160.
RD-30	Liu, S., X. Li, Z. Xu, T. Che, Q. Xiao, M. Ma, Q. Liu, R. Jin, J. Guo, L. Wang, W. Wang, Y. Qi, H. Li, T. Xu, Y. Ran, X. Hu, S. Shi, Z. Zhu, J. Tan, Y. Zhang, and Z. Ren. The Heihe Integrated Observatory Network: A Basin-Scale Land Surface Processes Observatory in China. Vadose Zone J. 2018, 17:180072.
RD-31	Göttsche, F.M.; Olesen, F.S.; Trigo, I.; Bork-Unkelbach, A.; Martin, M. Long Term Validation of Land Surface Temperature Retrieved from MSG/SEVIRI with Continuous in-Situ Measurements in Africa. Remote Sensing 2016, 8, 410.
RD-32	Göttsche, F., F.O.; Bork-Unkelbach, A. Validation of Operational Land Surface Temperature Products with Three Years of Continuous In-Situ Measurements. EUMETSAT Meteorological Satellite Conference, 2011.
RD-33	Göttsche, F., Olesen, F., Poutier, L., Langlois, S., Wimmer, W., Garcia Santos, V., Coll, C., Niclos, R., Arbelo, M., and Monchau, J-P., Report from the Field Inter-Comparison Experiment (FICE) for Land Surface Temperature. http://www.frm4sts.org/wp-content/uploads/sites/3/2018/10/FRM4STS_LST-FICE_report_v2017-11-20_signed.pdf
RD-34	Duan, S.-B.; Li, Z.-L.; Li, H.; Göttsche, F.-M.; Wu, H.; Zhao, W.; Leng, P.; Zhang, X.; Coll, C. Validation of Collection 6 MODIS land surface temperature product using in situ measurements. Remote. Sens. Environ. 2019, 225, 16–29, doi:10.1016/j.rse.2019.02.020.
RD-35	Li, H.; Li, R.; Yang, Y.; Cao, B.; Bian, Z.; Hu, T.; Du, Y.; Sun, L.; Liu, Q. Temperature-Based and Radiance-Based Validation of the Collection 6 MYD11 and MYD21 Land Surface Temperature Products Over Barren Surfaces in Northwestern China. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59 (2), 1794-1807, doi: 10.1109/TGRS.2020.2998945.
RD-36	Yang, J.; Zhou, J.; Göttsche, F.-M.; Long, Z.; Ma, J.; Luo, R. Investigation and validation of algorithms for estimating land surface temperature from Sentinel-3 SLSTR data. Int. J. Appl. Earth Obs. Geoinf. 2020, 91, 102136, doi:10.1016/j.jag.2020.102136.
RD-37	Göttsche, F.M.; Hulley, G.C. Validation of six satellite-retrieved land surface emissivity products over two land cover types in a hyper-arid region. Remote Sensing of Environment 2012, 124, 149–158.

Ref Id	Reference document
RD-38	Peres, L.F.; Libonati, R.; DaCamara, C.C. Land-Surface Emissivity Retrieval in MSG-SEVIRI TIR Channels Using MODIS Data. IEEE Transactions on Geoscience and Remote Sensing 2014, 52, 5587–5600.
RD-39	Trigo, I.F.; Peres, L.F.; DaCamara, C.C.; Freitas, S.C. Thermal Land Surface Emissivity Retrieved From SEVIRI/Meteosat. IEEE Transactions on Geoscience and Remote Sensing 2008, 46, 307–315.
RD-40	Peres, L.; DaCamara, C. Emissivity maps to retrieve land-surface temperature from MSG/SEVIRI. IEEE Transactions on Geoscience and Remote Sensing 2005, 43, 1834–1844.
RD-41	Guillevic, P.C.; Bork-Unkelbach, A.; Gottsche, F.M.; Hulley, G.; Gastellu-Etchegorry, J.P.; Olesen, F.S.; Privette, J.L. Directional Viewing Effects on Satellite Land Surface Temperature Products Over Sparse Vegetation Canopies – A Multisensor Analysis. IEEE Geoscience and Remote Sensing Letters 2013, 10, 1464–1468.
RD-42	Ermida, S.L.; Trigo, I.F.; DaCamara, C.C.; Göttsche, F.M.; Olesen, F.S.; Hulley, G. Validation of remotely sensed surface temperature over an oak woodland landscape – The problem of viewing and illumination geometries. Remote Sensing of Environment 2014, 148, 16–27.
RD-43	Xu, H.; Yu, Y.; Tarpley, D.; Gottsche, F.; Olesen, F.S. Evaluation of GOES-R land surface temperature algorithm using SEVIRI satellite retrievals with in situ measurements. IEEE Transactions on Geoscience and Remote Sensing 2014, 52, 3812–3822.
RD-44	Freitas, S.; Trigo, I.; Bioucas-Dias, J.; Gottsche, F.M. Quantifying the Uncertainty of Land Surface Temperature Retrievals From SEVIRI/Meteosat. IEEE Transactions on Geoscience and Remote Sensing 2010, 48, 523–534.
RD-45	Feltz, M.; Borbas, E.; Knuteson, R.; Hulley, G.; Hook, S. The Combined ASTER MODIS Emissivity over Land (CAMEL) Part 2: Uncertainty and Validation. Remote Sensing, 2018, 10 (5), 664, doi: 10.3390/rs10050664.
RD-46	Borbas, E.; Hulley, G.; Feltz, M.; Knuteson, R.; Hook, S. The Combined ASTER MODIS Emissivity over Land (CAMEL) Part 1: Methodology and High Spectral Resolution Application. Remote Sensing, 2018, 10 (4), 643, doi: 10.3390/rs10040643.
RD-47	Riffler, M.; Lieberherr, G.; Wunderle, S. Lake surface water temperatures of European Alpine lakes (1989–2013) based on the Advanced Very High Resolution Radiometer (AVHRR) 1 km data set. Earth System Science Data 2015, 7, 1–17.
RD-48	Hulley, G.C.; Hook, S.J. Generating Consistent Land Surface Temperature and Emissivity Products Between ASTER and MODIS Data for Earth Science Research. IEEE Transactions on Geoscience and Remote Sensing 2011, 49, 1304–1315.
RD-49	Lieberherr, G.; Wunderle, S. Lake Surface Water Temperature Derived from 35 Years of AVHRR Sensor Data for European Lakes. Remote Sensing 2018, 10.
RD-50	Göttsche, F.M.; Olesen, F.S.; Trigo, I.F. Radiometric in-situ measurements over Lake Constance for validating land surface temperature products. Proceedings for the 2016 EUMETSAT Meteorological Satellite Conference, 26-30 September 2016, Darmstadt, Germany, 2016.
RD-51	Hulley, G.; Göttsche, F.; Rivera, G.; Hook, S. J.; Freepartner, R. J.; Martin, M. A.; Cawse-Nicholson, K.; Johnson, W. R. Validation and Quality Assessment of the ECOSTRESS Level-2 Land Surface Temperature and Emissivity Product. IEEE Transactions on Geoscience and Remote Sensing, 2021, 1–23, doi: 10.1109/TGRS.2021.3079879.
RD-52	Göttsche, F-M., Wimmer, W., Martin, M., and Pérez-Planells, L., Report on the Thermal Infra-red Product Inter-comparison and Validation Study on Lake Constance. https://www.eumetsat.int/media/48683

Ref Id	Reference document
RD-53	Yu, Y.; Tarpley, D.; Privette, J.L.; Flynn, L.E.; Xu, H.; Chen, M.; Vinnikov, K.Y.; Sun, D.; Tian, Y. Validation of GOES-R Satellite Land Surface Temperature Algorithm Using SURFRAD Ground Measurements and Statistical Estimates of Error Properties. IEEE Transactions on Geoscience and Remote Sensing 2012, 50, 704–713.
RD-54	Augustine, J.A.; DeLuisi, J.J.; Long, C.N. SURFRAD—A National Surface Radiation Budget Network for Atmospheric Research. Bulletin of the American Meteorological Society 2000, 81, 2341–2357.
RD-55	Wang, K.; Liang, S. Evaluation of ASTER and MODIS land surface temperature and emissivity products using long-term surface longwave radiation observations at SURFRAD sites. Remote Sensing of Environment 2009, 113, 1556–1565.
RD-56	Li, S.; Yu, Y.; Sun, D.; Tarpley, D.; Zhan, X.; Chiu, L. Evaluation of 10 year AQUA/MODIS land surface temperature with SURFRAD observations. International Journal of Remote Sensing 2014, 35, 830–856.
RD-57	Guillevic, P.C.; Biard, J.C.; Hulley, G.C.; Privette, J.L.; Hook, S.J.; Olioso, A.; Göttsche, F.M.; Radocinski, R.; Román, M.O.; Yu, Y.; Csiszar, I. Validation of Land Surface Temperature products derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) using ground-based and heritage satellite measurements. Remote Sensing of Environment 2014, 154, 19–37.
RD-58	Hulley,G. C.; Hook,S. J.; Abbott, E.; Malakar, N.; Islam, T.; Abrams, M. The ASTER Global Emissivity Dataset (ASTER GED): Mapping Earth's emissivity at 100 meter spatial scale. Geophysical Research Letters, 2015, 42 (19), 7966–7976, doi: 10.1002/2015GL065564.
RD-59	Augustine, J.A.; Dutton, E.G. Variability of the surface radiation budget over the United States from 1996 through 2011 from high-quality measurements. Journal of Geophysical Research: Atmospheres 2013, 118, 43–53.
RD-60	Wan, Z.; Dozier, J. A Generalized Split-Window Algorithm for Retrieving Land-Surface Temperature from Space. IEEE Transactions on Geoscience and Remote Sensing 1996, 34, 892–905.
RD-61	Aires, F.; Prigent, C.; Rossow, W.; Rothstein, M. A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations. Journal of Geophysical Research 2001, 106, 14887–14907.
RD-62	Yu, Y.; Tarpley, D.; Privette, J.; Goldberg, M.; Rama Varma Raja, M.; Vinnikov, K.; Xu, H. Developing Algorithm for Operational GOES-R Land Surface Temperature Product. IEEE Transactions on Geoscience and Remote Sensing 2009, 47, 936–951.
RD-63	Trigo, I.F.; Monteiro, I.T.; Olesen, F.; Kabsch, E. An assessment of remotely sensed land surface temperature. Journal of Geophysical Research 2008, 113, 1–12.
RD-64	Ermida, S.L.; DaCamara, C.C.; Trigo, I.F.; Pires, A.C.; Ghent, D.; Remedios, J. Modelling directional effects on remotely sensed land surface temperature. Remote Sensing of Environment 2017, 190, 56–69.
RD-65	Prata, F. Land surface temperature measurement from space: AATSR algorithm theoretical basis document. Technical report, CSIRO Atmospheric Research, Aspendale, Australia, 2002.
RD-66	Ghent, D. GlobTemperature Project Technical Specification Document. Technical report, ESA, 2016.
RD-67	Garcia-Haro, F.; Sommer, T. Simultaneous land surface temperature-emissivity retrieval in the infrared split window. International Journal of Remote Sensing 2012, 26, 2135 – 2162.
RD-68	CM SAF homepage, http://www.cmsaf.eu/EN/Home/home_node.html

Ref Id	Reference document
RD-69	Duguay-Tetzlaff, A.; Bento, V.A.; Göttsche, F.M.; Stöckli, R.; Martins, J.P.; Trigo, I.; Olesen, F.; Bojanowski, J.S.; da Camara, C.; Kunz, H. Meteosat Land Surface Temperature Climate Data Record: Achievable Accuracy and Potential Uncertainties. Remote Sensing 2015, 7, 13139–13156.
RD-70	Jiménez, C.; Prigent, C.; Catherinot, J.; Rossow, W.; Liang, P.; Moncet, J.L. A comparison of ISCCP land surface temperature with other satellite and in situ observations. Journal of Geophysical Research: Atmospheres 2012, 117.
RD-71	Rossow, W.B.; Schiffer, R.A. Advances in Understanding Clouds from ISCCP. Bulletin of the American Meteorological Society 1999, 80, 2261–2288
RD-72	European Centre for Medium-Range Weather Forecasts. IFS DOCUMENTATION – Cy43r1 Operational implementation 22 Nov 2016 PART IV: PHYSICAL PROCESSES. Technical report, European Centre for Medium-Range Weather Forecasts, 2016.
RD-73	ILSTE homepage, https://lpvs.gsfc.nasa.gov/LSTE/LSTE_home.html

1.5. Glossary

Term	Definition
AATSR	**Advanced ATSR**
ARM	**Atmospheric Radiation Measurement**
ATSR	**Along Track Scanning Radiometer; the ATSR series of instruments comprises ATSR-1, ATSR-2 and AATSR**
AVHRR	**Advanced Very High Resolution Radiometer**
BSRN	**Baseline Surface Radiation Network**
CAR	**Climate Assessment Report**
CCI	**Climate Change Initiative**
CDR	**Climate Data Record**
CEOS	**Committee on Earth Observation Satellites**
CEOS – WGCV	**Committee on Earth Observation Satellites - Working Group on Calibration and Validation**
CGLOPS	**Copernicus Global Land Operational Service**
CIMSS	**Cooperative Institute for Meteorological Satellite Studies database of monthly land surface emissivity**
ECMWF	**The European Centre for Medium-Range Weather Forecasts**
ECV	**Essential Climate Variable**
EO	**Earth Observation**
ERA-5	**ECMWF Reanalysis 5**
ERA-Interim	**ECMWF Reanalysis – Interim**

Term	Definition
ESA	European Space Agency
GCOS	Global Climate Observing System
GEO	Geostationary orbit
GOES	Geostationary Operational Environmental Satellite
HMWR	Himawari
ILSTE	International Land Surface Temperature and Emissivity
IS	In Situ Files
IST	Ice Surface Temperature
IR	Infrared
JAMI	Japanese Advanced Meteorological Imager
KIT	Karlsruhe Institute of Technology
L2	Level 2 data
L3	Level 3 data
L3U	Level 3 uncollated data; gridded version of L2P data.
L3C	Level 3 collated data; L2P data from a single instrument that have been combined and mapped onto a space-time grid.
LCC	Land Cover Class
LEO	Low Earth Orbit
LSA SAF	Land Surface Analysis Satellite Application Facility
LST	Land Surface Temperature
LST_cci	Land Surface Temperature Climate Change Initiative
LWST	Lake Water Surface Temperature
MMDB	Multi-sensor Matchup Database
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
MTSAT	Multifunctional Transport Satellites
MW	Microwave
NetCDF	Network Common Data Format
NDVI	Normalized Difference Vegetation index
PSD	Product Specification Document
PVIR	Product Validation and Intercomparison Report
PVP	Product Validation Plan
QA4EO	Quality Assurance Framework for Earth Observation

Term	Definition
SE	**Satellite Extraction Files**
SEVIRI	**Spinning Enhanced Visible Infra-Red Imager**
SI	**Satellite – In situ match ups**
SLSTR	**Sea and Land Surface Temperature Radiometer**
SNO	**Simultaneous Nadir Overpass**
SS	**Satellite – Satellite match ups**
SSM/I	**Special Sensor Microwave – Imager**
STD	**Standard Deviation**
SURFRAD	**Surface Radiation Budget Network**
TOA	**top-of-atmosphere**
ULeic	**University of Leicester**

2. Requirements from Validation Protocols and User Needs

There are several protocols about LST validation that formulate guidelines for a consistent validation process, which are introduced in the following [RD-1, RD-2].

The Committee on Earth Observation Satellites (CEOS), within its Working Group on Calibration and Validation (WGCV), has done much work in this field with the goal to "ensure long-term confidence in the accuracy and quality of Earth Observation data and products, and to provide a forum for the exchange of information about calibration and validation activities." [RD-2]. LST_cci will conform with the guidelines for calibration and validation provided by CEOS-WGCV under the Quality Assurance Framework for Earth Observation (QA4EO) and endorsed by CEOS. Another important LST Validation Protocol was published by [RD-1], which specifies the different validation categories. This is also considered in the LST_cci validation work. Furthermore, the approaches endorsed by the LST community through the International Land Surface Temperature and Emissivity (ILSTE) Working Group and the CEOS-WGCV subgroup on Land Product Validation (LPV) are combined in the "Land Surface Temperature Product Validation Best Practice Protocol" [RD-2]. They mention that validation should "be capable of testing products for compliance with Global Climate Observing System (GCOS) requirements". A consistent approach to LST validation taking [RD-2] as a base line is promoted by [RD-3].

It will be investigated within LST_cci how far the CCI satellite LST data products (Section 7.1 and Section 7.2), as well as the merged CCI satellite products (Section 7.3), comply with the requirements specified by GCOS for accuracy, precision, measurement uncertainty and stability of the record. These requirements are listed in Table 1 [RD-3] and [RD-5], as well as the requirements that were defined within LST_cci by a user assessment [AD-1]. The requirements specified by GCOS in [RD-3] were updated in [RD-5], however, in LST_cci the accuracy and precision requirements are kept as they provide useful information for users and is consistent within the project. In addition, the measurement uncertainty requirement specified in [RD-5] is also investigated. The terms used are also consistent with the LST_cci End-To-End Uncertainty Budget.

Table 1: Requirements for LST data sets

Requirement for	GCOS Threshold [RD-3], [RD-5]	GCOS Target [RD-3], [RD-5]	LST_cci User Threshold [AD-1]	LST_cci User Target (Objective) [AD-1]
Accuracy	< 1 K	< 1 K	1 K	0.3 K
Precision	< 1 K	< 1 K	1 K	0.3 K
Measurement Uncertainty	< 1 K	< 1 K	-	-
Stability	0.3 K / decade	< 0.1K / decade	0.3 K / decade	0.1 K / decade

The terms accuracy and stability are defined in [RD-3] as follows, they are based on [RD-6]:

- Accuracy: '"closeness of the agreement between a measured quantity value and a true quantity value of the measurand". The concept "measurement accuracy" is not a quantity and is not given a numerical quantity value."' [RD-3]

[AD-1] add to that definition, that it "is theoretical, as the true value cannot be known due to measurement error." [RD-3]

- Stability: "Assessment of whether a monotonic trend exists with respect to ground- based Fiducial Reference Measurements or related ECV datasets (such as near- surface air temperature)" [RD-5]

  [RD-3] defines it as "the extent to which the uncertainty of measurement remains constant with time" and, in addition, refers to it as "the maximum acceptable change in systematic error, usually per decade."

  [AD-2] defines it as the "consistency of LST measurements from a given satellite product over time (including a product comprised of multiple satellite instruments)"

- Precision is not defined by [RD-3], but [AD-2] introduce it as "The closeness of agreement between independent measurements of a quantity under the same conditions."
- For measurement uncertainty, [RD-2] and [RD-4] give the definition from [RD-6], which is a "parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand". Furthermore, [RD-5] refers to it as the "required total uncertainty per pixel combining the four groups of uncertainty components: random, locally correlated atmospheric, locally correlated surface, and large scale systematic".

The different ways to validate LST data are subdivided by [RD-1] into four categories, which are introduced here:

- Category A: Comparison of satellite LST with in situ measurements
  - This is the traditional and most straightforward validation type, where a satellite LST product is directly compared with in situ LST data from ground-based radiometers. The in situ data needs to be collocated and simultaneously acquired. Following [RD-7], this is the most accurate validation category, if the in situ data is taken with infrared (IR) radiometers. The measurements should be "calibrated traceably to SI units, generally through a reference radiance blackbody" [RD-7]. A fully traceable link to SI units also makes sure that the data "are robust and can claim the status of a "climate data record" " [RD-7]. The CEOS-WGCV subgroup LPV mentions that this is the reference validation method, "if the measurements are performed with well maintained, high quality instrumentation and are representative for the satellite sensor footprint" [RD-8].
  - [RD-2] adds that also the in situ LST uncertainty needs to be assessed, and that the spatial representativeness of the station needs to be considered. [RD-2] further recommends validating each season separately to resolve seasonal differences in e.g. leaf-on/leaf-off periods of deciduous forests. Results from different stations should be investigated by classes of ancillary data, e.g. by using different land cover types or total atmospheric water vapour content.
  - As this approach is often more labour intensive than other approaches as a result of finding suitable sites, operating validation stations and obtaining the required ancillary information from additional measurements, there are not many well-suited LST validation sites worldwide. [RD-1] lists over land surface only the validation sites from the Karlsruhe Institute of Technology (KIT) in the best in situ validation class.
  - In general, the protocol for the in situ measurements utilised in the LST_cci validation will conform as closely as possible to the following criteria defined in [RD-1]:

- The in situ measurements should be continuous and frequent to allow a close temporal matching between satellite and in situ LST data.
- At least minimum ancillary data relating to the near surface conditions should be included.
- The validation site should be uniform at several kilometres at best.
- The in situ instruments should be calibrated to as near an accuracy of ±0.1 K as possible considering the restrictions of the respective instrument and be traceable to a PTB (Physikalische Technische Bundesanstalt, Germany) blackbody.
- At sites where directional radiances are measured, measuring down-welling sky radiances as well allows for corrections of reflected radiance.
- At sites where the measurements are made with hemispherical broadband radiance sensors, measuring up-welling and down-welling radiances simultaneously also allows for correction of reflected down-welling radiance.
- Sun photometer measurements of aerosol optical depth are in general desirable since these provide information on the expected LST retrieval quality.
- LST retrieval from satellite measurements and the derivation of in situ LST from directional radiance measurements require knowledge of directional spectral (8-14 µm) land surface emissivity (LSE) for the respective target areas.
- If these LSE data are not available, the use of in situ emissivities derived from MOD21 LSE [RD-9] is recommended, since it is the only physically based satellite emissivity product and reflects the actual situation during satellite acquisitions.
- The fractional cover of the end-members (main land cover types, ideally observed by separate radiometers) at the validation sites should be estimated; the emissivities of the relevant end-members should be acquired from literature.
- Obtaining in situ LST from broadband radiance measurements requires knowledge of the associated broadband emissivities (BBE). The standard approach to obtain BBE is via regression relationships with multi-spectral emissivities [RD-24, RD-25].
- All-sky cameras are desirable to provide an objective measure of cloudiness during the daytime.

- Category B: Radiance-based validation
  - For this method, no ground LST measurements are required. It thus offers an alternative to in situ validation, in regions where measurements are scarce or for long-term LST evaluation at a global scale. Radiance-based validation means that top-of-atmosphere brightness temperatures are simulated with a radiative transfer model, which uses surface emissivities and atmospheric profiles of air temperature and water vapour content as input [RD-1]. This method is not applied in LST_cci, since it requires in-situ emissivity data for static sites and radiative transfer calculations, which demand considerable additional capacities. However, it is recommended to include it in Phase 2 of LST_cci.
- Category C: Inter-comparison with similar LST products
  - Operational LST products from thermal infrared data are provided by many different airborne and spaceborne instruments [RD-1]. These data can be used for satellite – satellite intercomparisons of LST products, when both data sets are temporarily and spatially matched. Although, according to [RD-1], these intercomparisons cannot provide an absolute validation as the Category A validation with in situ stations, they are a good supplement to it. The CEOS-WGCV subgroup LPV adds, that this “method is particularly valuable for

identifying disagreements between LST products over large areas and different land cover types." [RD-8]

- For satellite – satellite intercomparisons, the methodology in LST_cci follows that applied in GlobTemperature [RD-10], which is consistent with [RD-1]. It can be summarised as follows:
  - The original Level-2 data are converted to harmonised Level 3 uncollated (L3U) data.
  - Simultaneous nadir overpasses (SNO) of the Low Earth Orbit (LEO) instruments within the reference Geostationary Orbit (GEO) disk or with respect to the reference LEO orbit are determined.
  - These SNO matches are composited by a monthly aggregator creating monthly median fields in same L3U format from which monthly statistic are derived.
  - The final stage aggregates the new monthly L3U SNO data to seasonal statistics for December, January, February (DJF); March, April, May (MAM); June, July, August (JJA); and September, October, November (SON).
  - The analysis processor is used to generate results and visualisations from the monthly and seasonal data.

- The intercomparison processing chain is shown in Figure 1.

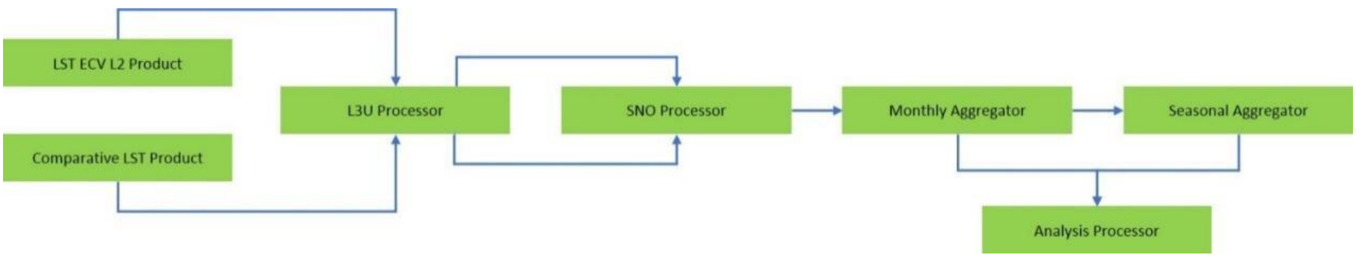

Figure 1: Intercomparison matchup and analysis processing chain

- Category D: Time series analysis
  - When satellite data are analysed over a longer time period over a temporally stable target site, potential calibration drift or other issues of the instrument can be identified, if they manifest themselves over time [RD-1]. These time series analysis might also help to detect problems with cloud contamination from artefacts. However, attention must be paid to geophysical changes of the target site or the atmosphere. The CEOS-WGCV subgroup LPV mentions that this validation approach needs "relatively long time series of observations over temporally stable targets, e.g. inland water bodies or deserts." [RD-8]. This approach can also be used to assess the long term stability of satellite LST CDRs, which will be done in LST_cci. As the target sites for Category D validation need to be on a very homogeneous land cover, usually water bodies or well-characterized barren surfaces are used [RD-2]. The lake site used in LST_cci will be on Lake Constance, where a long time series of in situ data is available (see Section 6.6).

The CEOS-WGCV subgroup LPV defined four validation stages for their focus areas, which correspond "to increasing spatial and temporal representativeness of samples used to perform direct validation" [RD-2]. The stages range from no validation at all (stage 0) to the best validation category (stage 4), which includes systematically updated results, spatial and temporal consistency, characterized uncertainties, and publication of results. The CEOS-WGCV LPV Focus Area on Land Surface Temperature & Emissivity Product Validation [RD-73] mentions that the highest validation stage currently reached for LST validation is only stage 1. For higher validation stages, more global sites over different land cover types and an extended temporal coverage would be needed, as well as further satellite - satellite intercomparisons. They expect that satellite and in situ LST data products agree within 1 – 2 K with each other over homogeneous sites. The number of globally available LST in-situ validation stations remains limited. With them, Validation Stage 2 is achieved. Radiance-based validation (Cat B) can extend the number of sites to reach Stage 3. However, due to limited resources, it is not performed within LST_cci and the objective is to reach Validation Stage 2.

Concerning the analysis of LST validation, [RD-1] advises not to use classic statistical methods such as the mean for statistical presentation of the validation results as LST data can contain outliers. This can lead to misinterpretation, if outliers are actually present and influence the results. [RD-1] recommends the use of robust statistical methods instead which are less influenced by outliers. In the validation for LST_cci, robust statistics such as the median and robust standard deviation (see e.g. [RD-11]) will be used. These methods were already applied during the LST validation in the GlobTemperature project [RD-12, RD-13]. The measurement uncertainty will be assessed by analysing the total uncertainty in the satellite-in situ matchup, which considers the uncertainty contribution of four different components: the satellite product uncertainty, the in situ LST uncertainty, the spatial and temporal mismatching (more details are given in Section 3.2).

Once the validation results are analysed, they need to be reported in a well-structured and transparent manner. It is also important to feedback identified issues in the satellite data sets to the data set producers so that they can refine their algorithms [RD-2]. This will be done within LST_cci in the Product Validation and Intercomparison Report (PVIR), and the validation team will also be in constant discussion with the data set developers to help them improve their data.

Within the LST_cci project, requirements of users have been assessed during the Joint Land Workshop in Lisbon during 2018 and the online LST_cci User Workshop in 2020 and 2022and furthermore in an online questionnaire. The outcomes of these surveys are documented in [AD-1], where users were also asked about their expectations from LST validation. Most users state there that they do consider validation results in their work and require a range of validation and intercomparison data. The most useful validation and intercomparison information for users is the comparison of the satellite LST product with in situ measurements, followed by the satellite – satellite intercomparisons, the summary of accuracy and precision per product, and an overview of the best performing products. Slightly less important is the time series analysis. However, interviews conducted with the climate research group [AD-1] depict a clear requirement for the provision of homogenised LST time series, which are free from non-climatic effects. As this implies a time series analysis is of benefit, it will be addressed in LST_cci. Reported barriers in using validation results were due to “a lack of understanding of how to incorporate the data, and a lack of available information” [AD-1]. Within LST_cci validation we aim to avoid these barriers by providing thorough documentation of the validation process in this document as well as a description and discussion of the results and its implications in the PVIR.

The Level-2 satellite data products produced by the EO Science Team in LST_cci are validated against in situ data from several, globally distributed stations. This validation work is done independently from the EO Science Team by the validation team led by KIT. This ensures that the validation is not influenced by the developers of the satellite data sets. This is a key requirement in [AD-1] for the validation and user assessment activities. However, there will be regular interactions with the data producers from the EO Science Team to discuss the results of the validations, which will help to improve the satellite data sets.

The results of the validation work will be documented in the PVIR, which will be growing throughout the project. It will include the results of each validation cycle and will also be fed into the corresponding version of the Climate Assessment Report (CAR). In the PVIR also recommendations for the data producers stemming from the validation results will be included.

All in situ data sets that are needed for the validation are produced by KIT. The stations used are set in different regions, cover different land covers and time periods. For all produced satellite data sets at least one station is available, against which it will be validated. Where possible, given temporal and spatial availability of suitable stations, several stations with different land covers will be used for in situ validation. The stations utilised are described in detail in Section 6.

3. Product In Situ Validation Methodology

3.1. Data Preparation and Matching

Existing ground-based validation data has been reviewed and a list of suitable sources of this data has been created. For selection of the in situ stations, the temporal and spatial coverage of the in situ data sets has been taken into account, as well as their suitability for validation. The latter is mainly dependent on the particular measurement devices used, the temporal resolution of the measurements and their availability, as well as on additional geographical information such as surface homogeneity or orography around the station. Based on these characteristics, several in situ LST data sets have been identified as well-suited for validation within LST_cci. These include the stations run by KIT in Africa and Europe, the SURFRAD network and ARM stations in the USA, the Neumayer station in Antarctica and the Copernicus LAW stations. For time series analysis of the climate data records, it is planned to use lake water surface temperature (LWST) data from Lake Constance as it has a long temporal availability over a homogeneous surface. An identified data gap that needs to be filled in LST_cci is a suitable in situ site on the MTSAT disk. A possibility to fill this gap is by using in situ data from the Heihe River Basin, China, as well as the Copernicus LAW station in Robson Creek Forest, Australia. These data will be tested in LST_cci. Several of the data sets mentioned above have already been used successfully in the GlobTemperature project [RD-12]. It is envisaged, wherever possible, to validate each satellite data set over several in situ stations with different land covers, as the quality of satellite LST data set can depend on land cover.

Three different data files are needed for the in situ validation, which are:

- In situ files (IS)
- Satellite extraction files (SE)
- Matched satellite – in situ files (SI)

All these data files are going to be produced in a common harmonized data format that confirms to the CCI data format [RD-14]. These data formats are described in Section 5.2.

The in situ data is stored in IS files while satellite data for an area centred on each station is extracted and stored in SE files. The temporarily and spatially matched satellite extractions and in situ data is stored in SI files, which are the files utilised in the validation. In order to make the results of different data sets comparable to each other, it is generally attempted to validate IR satellite data on a 0.05° x 0.05° grid against the in situ data. However, during the GlobTemperature project it was found that that was not always possible due to surface heterogeneity or difficult orography at certain stations [RD-12]. If this is the case for a certain station, it will be mentioned in Section 6 for the respective station, and also in the results described in the PVIR.

The matching itself is done in the same way as in the GlobTemperature project [RD-12, RD-13], and is described briefly here. For the GEO and microwave (MW) data sets, the satellite pixel directly over the in situ station (station pixel) is used for the spatial matching as the GEO data sets are already on a 0.05° x 0.05° grid. For the LEO data sets, which have a resolution of an equal angle grid of 0.01° x 0.01° around the globe, 25 pixels in a square (5 x 5) around the pixel above the station are chosen to derive the desired 0.05° x 0.05° grid. This grid was chosen as it is the same grid that the GEO data sets are on, which makes in situ validation of LEO and GEO data sets comparable to each other over stations set in a large homogeneous area. Furthermore, from these 25 pixels only those that have the same combined land cover class (LCC) as the station pixel are averaged to avoid influences from different land covers. Combined land cover class is defined as in [RD-15], where the GLOBCOVER classes were reduced to ten classes with similar properties. LST_cci land cover classes will be used. The median of the LST of these pixels is taken as merged satellite LST for the validation process. For two stations (GBB_W and TBL), another satellite pixel than the station pixel is chosen as centre of the validation to avoid spatial mismatching (see Section 6 for details).

Using the 5 x 5 pixels area around the station was not possible at all sites due to surface heterogeneity around the station or changing orography. At some sites, either 3 x 3 pixels are merged or only a single pixel is taken for the validation (see Section 6). Only grid points where at least 80 percent of the pixels are not flagged as cloudy are considered for the validation, because in areas where a larger fraction is flagged the possibility of not masked cloud edges was thought to be too high. The uncertainty of the averaged LEO LST is calculated according to the following formula:

$$(AveragedUncertainty)^{2} = \frac{\sum(ClearPixelUncertainty)^2}{NumberClearPixels} + \frac{NumberCloudyPixels \, * \, VarianceClearPixels}{NumberClearPixels \, + \, NumberCloudyPixels}$$

where ClearPixelUncertainty is the uncertainty from the LEO LST values for the single pixels that are not flagged, NumberClearPixels and NumberCloudyPixels are the number of the clear and flagged pixels, respectively, and VarianceClearPixels is the variance of the clear pixels.

Temporal matching between satellite LST and in situ LST is achieved by linear interpolation between the two station measurements that are closest to the satellite overpass time. The temporal resolution varies between the different stations, but the maximum temporal interval is 3 minutes. If there are gaps in the in situ data and the closest in situ data was acquired more than 3 minutes away from the acquisition time of the satellite data, the data point is disregarded.

3.2. Validation

The validation itself is done in the same way for all data sets. This is possible as all data sets will be in the same harmonized data format. Differences between satellite and in situ LST are analysed both statistically and visually. Statistical results are presented in terms of bias, here the median average of the difference of satellite LST minus in situ LST (following RD-2), and by the robust standard deviation (RSTD) (see e.g. [RD-11]). RSTD is connected to the precision (median absolute deviation, MAD) introduced in Section 2 by a constant factor of 1.48, i.e. RSTD = 1.48 * MAD.

Daytime and night-time differences, differences between the single satellite data sets, seasonal variations and differences at several stations for one data set are investigated. Possible influences from satellite observation geometry, cloud contamination, surface elevation and land cover at the station, as well as possible seasonal cycles will be examined.

In addition to validation of the satellite LST data, the satellite LST uncertainty will also be validated. The EO Science Team developing the uncertainty budget will work closely with the Validation Team to ensure that the appropriate knowledge of the uncertainty budget construction is available. There are four main factors that can cause differences between satellite and in situ LST data. These are:

- Satellite LST uncertainty, which includes measurement uncertainty, the retrieval algorithm, uncertainty in the atmospheric correction of the data, and inaccurate land cover classification or emissivities [RD-16].

- In situ LST uncertainty, which includes uncertainty from the measurement device, and an uncertainty due to the land surface emissivity, which was used to calculate in situ LST.
- Spatial mismatching due to upscaling problems. This is an important factor. If the land surface around the station is heterogeneous, as for all in situ validation a point measurement is compared to a satellite measurement covering a larger area.
- Temporal mismatching, if the temporal difference of the satellite LST data and in situ LST data is large. However, this point can be ruled out in the validation within LST_cci, as all in situ data sets have a small sampling interval of 3 minutes or less.

The first three factors will be combined to give a total uncertainty of the matched in situ LST and satellite LST. This can then be compared against the standard deviation of the differences between satellite and in situ LST. The resulting differences between both values (total uncertainty, STD of differences between satellite and in situ LST) will then indicate if the estimated uncertainties are realistic (similar to STD), if uncertainty components are missing (smaller than STD) or if they are overestimated (larger than STD).

3.3. Interpretation of Results, Feedback to Producers

Significant statistical differences between satellite and in situ LST data found during the validation process will be investigated in detail and possible reasons for larger deviation from the GCOS requirements for accuracy, precision and total uncertainty will be investigated with consideration of the main reasons for uncertainties introduced above. During this process, the data set producers will be consulted, as they have the most knowledge about their data sets. This consultation will include presentations of the validation results to the EO Science Team and discussions of the possible implications.

A similar approach was already used successfully within the GlobTemperature project and resulted in the following feedback for in situ validations [RD-10, RD-12]:

- Validation over different times and land covers is important.
- Good information on how the LST data set was produced, e.g. which algorithms were used or how the uncertainty was derived is needed, for example by using the NetCDF data format to store the metadata with the data is required.
- Additional analyses of LST datasets under conditions of sub-optimal algorithm performance give useful information.
- Comparisons of the results obtained over one station for different satellites are instructive.
- Comparison of the results obtained for one sensor, but with different algorithms, are informative.

Detailed statistics and figures from all the analyses will be included in the PVIR. It is furthermore intended to publish the key findings in journal articles.

4. Product Intercomparison Methodology

Level-3 satellite data sets from LST_cci will be intercompared with external operational data sets and models (introduced in Section 8). Satellite - satellite intercomparisons are defined in LST_cci as the investigation of the difference and co-variability between two or more surface temperature datasets at the scales of the lower spatial resolution data set. This means primarily satellite - satellite intercomparisons, but also includes satellite vs. model intercomparisons.

Satellite – satellite intercomparisons will be carried out, similar to the in situ validations, by the validation team independently of the EO Science Team using the Multi-sensor Matchup Database (MMDB, http://gws-access.ceda.ac.uk/public/esacci_lst/) as base for data sets and code. The results will also be reported in the PVIR and will be discussed with the EO Science Team to improve the satellite data sets. It is the intention of the consortium to target a journal article on the results.

The Validation Team will utilise the functionality developed in pre-cursor and complementary projects for multi-sensor long-term global matching and analysis [RD-10].

Satellite - satellite intercomparisons are a good addition to in situ validation, since there are globally only few in situ validation sites dedicated to long-term LST validation, with a limited number of additional sites that can be utilised but have known challenges.

Intercomparisons of LST data sets over different land covers are particularly relevant for algorithms which have different sets of coefficients per land cover class, which are usually not totally covered by the in situ stations available.

4.1. Data Preparation and Matching

The intercomparison of the products generated here with external products will cover the respective time windows of the LST ECV products. This allows assessing the stability of the data records.

Since the fields-of-view of any two satellite instruments, even when nominally coincident, will not be the same, it is useful to perform the intercomparison at lower, consistent spatial resolutions (ideally on the same grid). In LST_cci, this is done by using L3U data, which are orbit / granule level data gridded in space but not in time. They therefore preserve the acquisition times. All data fields (including uncertainties, quality flags, cloud/aerosol information, and auxiliary data) are propagated within the L3U product.

These L3U data files are internal to the processing team but represent the baseline for all matchups and for higher-level collated (L3C) products. All matchups within a 5 minutes temporal threshold, consistent with the GSICS criteria [RD-17], are generated at the L3U product level, and then temporally collated into daytime and night-time composites. Observations are categorised as “day” or “night” based on their respective solar zenith angles. Once converted to the L3U format the new files are read in by the SNO processor.

For a useful intercomparison of datasets from different satellites, the spatial variability within the field of view of each satellite needs to be accounted for. This can be achieved by re-gridding the data onto a common spatial grid by averaging all geo-referenced, cloud free pixels weighted by their respective fractional area overlap with the corresponding common grid cell. For matchups between datasets with different spatial resolutions or at different orbital tracks, the spatial resolution of the coarsest instrument is used. For IR LEO vs. LEO and IR LEO vs. GEO the spatial resolution for re-gridding is defined as 0.05° x 0.05°; for MW vs. IR it is defined as 0.25°.

There are several methods for the spatial matching of two satellite datasets. The so-called nearest neighbour approach is an effective and relatively straightforward method. The images of the two considered satellites are overlaid with each other by shifting one set of pixels so that they match the other set. The advantage of this method is that data is not averaged or weighted, thus the original data remains unchanged and the two original datasets are compared. However, the larger the shift gets, the further apart the compared LST values are spatially. A second approach is the averaging of the data by polygon weighting. A polygon tessellation is formed. To account for the fact that the pixel area and the area of interest are often not exactly the same, the data in each polygon is weighted according to the proportion of the area of interest in the polygon to its total area (see [RD-18]). A third possibility is a simpler version of the second approach. In this case, each point in space is assigned to the pixel to which midpoint it is closest, and these are then averaged without weighting. Although this method is easy to implement, it has disadvantages: i) the original pixels may not be representative of the final sampling grid; and ii) for coarser resolution pixel gaps occur where the assignment of original pixels means some grid cells remain unfilled even for a continuous field of original pixels.

The optimal approach taken here is to apply the polygon weighting. The rationale being that a high spatial resolution matchup grid would be highly sensitive to LEO orbit tracks and their pixel nearest neighbour binning. This is particularly the case at the edge-of-swath of wide-swath instruments such as MODIS where pixel sizes are similar in size to the common matchup grid.

The high temporal variability of LST means that intercomparison of different LST products is a challenging prospect. In order to minimise the impact on the intercomparison results, LST differences due to different observation times have to be minimised. This can be achieved by limiting the data to close temporal matchups. To maintain consistency no interpolation between adjacent GEO LSTs that temporally bracket a specific LEO overpass time, as has been applied in some studies, is carried out here. Moreover, interpolation between less frequent GEO observations increases the risk that any assumption of a linear relationship between bracketing LST observations becomes invalid.

The defined temporal matchup threshold for the Quasi SNOs is set to 5 minutes to facilitate increased matchups between LEOs and between a LEO and the nearest GEO acquisition while limiting the potential for significant temperature change. Larger thresholds increase the risk of LST differences representing actual ground temperature changes rather than that they are attributable to the products themselves.

Outputs from the SNO processor are first aggregated to monthly median statistics on the common matchup grid. Initially all valid individual SNO matches are collected and stored as ΔLST (differences between the satellite LST values) values along with their associated uncertainty, satellite viewing angles, cloud, time, and geolocation information. The collated variables are filtered according to day or night statistics. The outputs are passed to the seasonal aggregator and analysis processor.

The seasonal aggregating processor is identical to the monthly aggregator with the exception that it takes in all the data from the whole period where the two sensor time series overlap. Once the seasonal Level-3 matchup composite files have been produced they are also passed on to the analysis processor.

The analysis processor is split into three components. It produces a range of analytical plots and result tables. By design, the complexity of the analysis performed at each stage increases. This allows for the control of the volume of output content. Visualisation of the spatial distribution and time series of variables for each test will also be produced. Excluded from the analysis will be data, where the uncertainties in the input data are very high due to very low clear to cloud ratios. An appropriate threshold for this ratio will be set as the project evolves. The matched satellite – satellite data will be stored in the MMDB in a harmonized data format, called the Satellite – Satellite match up files (SS), as introduced in Section 5.

4.2. Intercomparisons

The median and RSTD of the differences of the non-reference LST product to the reference LST product per pixel will be analysed. This means that the LST products are going to be analysed statistically in terms of their differences, i.e. the median of satellite 1 LST minus satellite 2 LST and the corresponding RSTD (see e.g. [RD-11]). Data are composited over each month and over the four seasons. The influence of satellite viewing geometry, sun angle, orography, elevation classes, and land cover classes on the resulting differences will be investigated. In the assessment by satellite viewing geometry, differences are binned and analysed against the product of the satellite zenith angle (satze) times the sign of the satellite azimuth angle (sataz) ($satze * \frac{|sataz|}{sataz}$). Also the number of averaged pixels will be considered, as this can influence the statistical significance of the results. A similar approach has already been used successfully in the GlobTemperature project [RD-10].While no quantitative assessment can be made based on differences in cloud contamination since “true” manual masks have only been produced for a few orbit scenes, interpretation of the results will include discussion on the likely impact of cloud contamination.

4.3. Interpretation of Results, Feedback to Producers

Similar to the in situ validation, when differences considerably larger than 2 K between the intercompared satellite data sets are found, they will be investigated further. Feedback on the results of the intercomparisons will be given to the data set producers. All results will also be reported in the PVIR.

During the GlobTemperature project, the satellite-satellite intercomparison exercise [RD-10] resulted in the following feedback to the producers:

- Information on how the LST data set was produced, e.g. which algorithms were used or how the uncertainty was derived, needs to be provided. The netCDF format is a good tool for making this kind of information available together with the data.
- The importance of satellite view angles, emissivity, topography, and land cover class of the investigated area for intercomparisons should be integral to the analysis when interpreting the results
- The influence of the homogeneity of the surface, satellite viewing geometry, of clouds and cloud masking should be explored further. Many of the factors are regionally dependent. Thus, to obtain an optimised evaluation of the quality of a data set, it should be intercompared with different instruments over different regions and at best also at different times. To help facilitate this, the provision of an intercomparison tool is recommended.
- Concerning the intercomparison of LST data sets from TIR sensors with LST derived from MW sensors, the initial intercomparison has shown some significant differences at higher and at

lower temperatures. While cloud contamination, MW penetration of the surface, and temporal matching methodology may be responsible for some of these differences, further investigation needs to be carried out for better understanding of them, which includes totally clear-sky and partially cloudy conditions.

These points will be considered during the satellite-satellite validations in LST_cci: the same netCDF data format is used in LST_cci, with an improved data file structure based on the GlobTemperature file structure. The difference factors influencing the validation results that were found to be relevant in GlobTemperature will be further investigated in LST_cci validations.

5. Multi-sensor Matchup Database (MMDB)

5.1. MMDB Structure

The MMDB is the tool for carrying out all validation exercises in LST_cci, as it consists of both the data and the core software for performing the extractions, matchups and analysis of all validation and intercomparison exercises. The MMDB will be physically located on the UK JASMIN server.

Four different types of data in a harmonized format each are included in the MMDB:

- Satellite Extraction files (SE)
- Satellite – Satellite match ups (SS)
- In situ files (IS)
- Satellite – In situ match up files (SI)

All data files are in netCDF format and confirm to the CCI data format [RD-14]. They are described in detail in Section 5.2. The SE and SS files are produced by the EO Science Team, and the IS and SI files by the Validation Team. All software to validate and intercompare the data is written by KIT in Python.

All IS data will be partitioned into two sections, the “development” and “validation” section. The “development” data will cover at least one full seasonal cycle of in situ data at all but two stations (GBB_W, Section 6.4, and the Lake Constance site, Section 6.6) and, wherever possible, the same data will be provided to all producers of satellite LST. The “development” data allows the producers to test their algorithms over different land surfaces and for different seasons. However, this cannot be a single “development” dataset (i.e. one specified year for all stations), since several of the relevant satellite sensors do not temporally overlap. Thus, the year of matchup data will be tailored for the individual products. The larger part of the IS data (all remaining years) - the “validation” dataset, will be in a restricted access section exclusively available to the Validation Team within LST_cci. These datasets will be used to validate the final satellite LST datasets provided by the producers. The complete time series of matched-up data allows a validation of the various LST products over several years, thereby covering repeatedly all seasons, providing a more complete assessment and validation over a broad range of atmospheric conditions. The two stations held back for the “validation” dataset only, are to ensure complete independence of the validation data from the product development over an unbroken time period for stability assessment. The division of the data is schematically shown in Figure 2, the years open for the developers are introduced in Section 6, and also in Figure 4 and Figure 5.

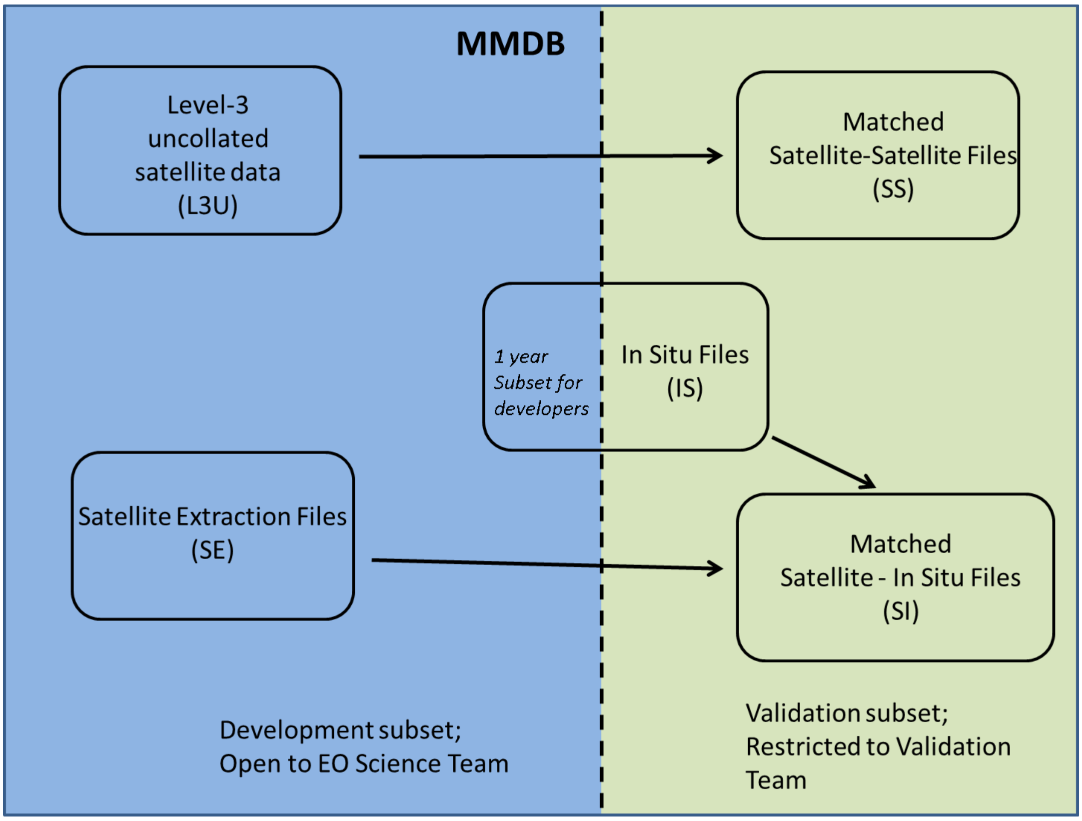

Figure 2: Schematic of the MMDB. SE and L3U files are produced by the EO Science Teams in the end-to-end processing chains.

Extensive experience and infrastructure from pre-cursor and complementary projects as well as existing MDB software and infrastructure from ULeic (UL_MDB) and from KIT (GT_MDB) will be integrated to build the MMDB. Thus, GlobTemperature [RD-10, RD-12], Sentinel-3 MPC [RD-19] and EUSTACE [RD-20] will be utilised.

The overall responsibility for the MMDB will by the KIT validation team, but both KIT and ULeic will work on the integration of the software code and population with data. For instance, the LEO IR and CDR ECV processing chains developed by ULeic on UK JASMIN will automatically populate the MMDB with satellite extractions for the specified in situ locations, and satellite vs. satellite extractions. All GEO and MW data transferred to the shared project workspace on JASMIN will be processed for these satellite extractions.

The design concept for the integration of the MMDB with the data production is presented in Figure 3. Where cross-project activities occur we will share data and resources to avoid duplication of effort.

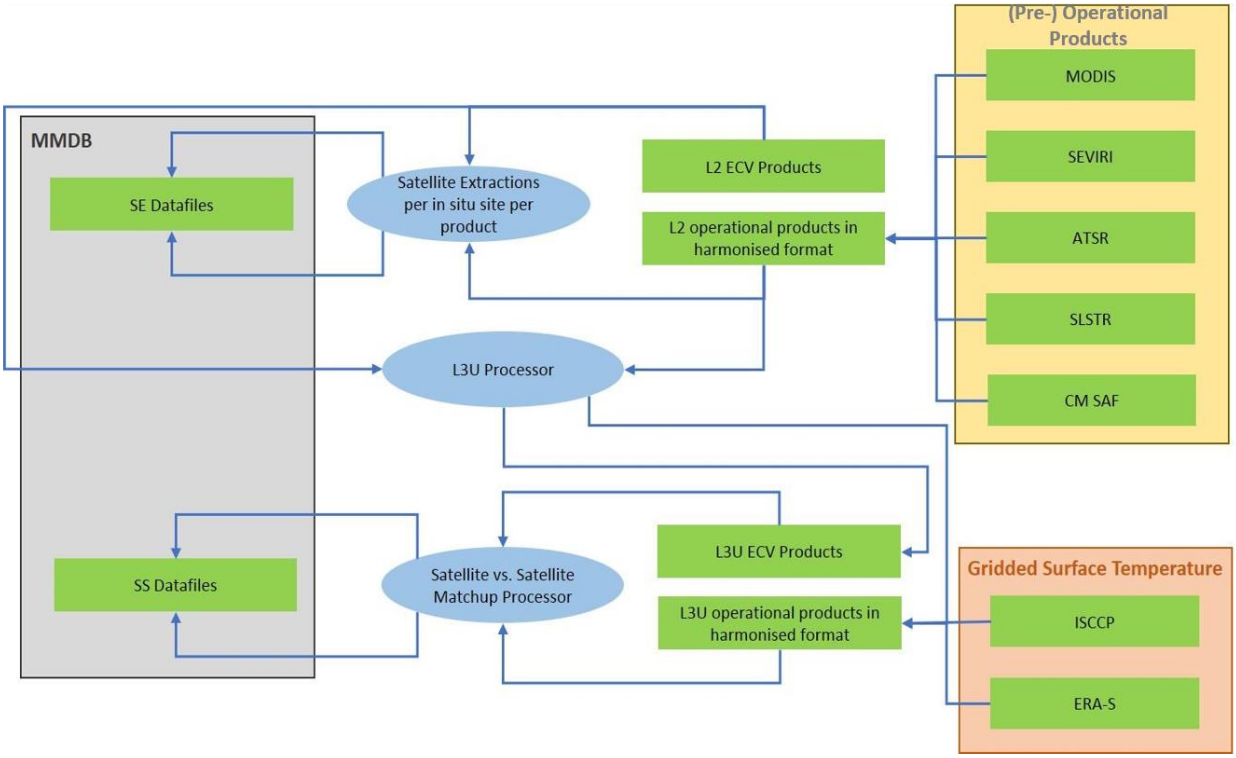

Figure 3: Schematic illustrating where the MMDB is integrated with the ECV processing chains. SLSTR data is available from Copernicus, MODIS from NASA DAAC, and SEVIRI data from Land Surface Analysis Satellite Application Facility (LSA SAF). These are operational data products, whereas the CM SAF and ATSR products are pre-operational. The CM SAF TCDR LST product is derived from a combination of MVIRI and SEVIRI data.

5.2. Data File Structure

An important point for consistent validation in the LST_cci project are the harmonized NetCDF data formats that are used throughout the validation work for the data files. Using the same format for all data sets makes the single validation exercises directly comparable to each other. Furthermore, any further validations can be produced fast, as the same code can be re-used. This approach has already been used successfully in the GlobTemperature project [RD-12].

Common to all LST_cci data files are the global attributes, which are metadata stored in the NetCDF files. They contain information that applies to the whole contents of the NetCDF file. The CCI Data Standards document defines a set of global attributes that is included in LST_cci files as well as recommend attributes for data discovery [RD-21]. A variable in a NetCDF file refers to an array of data stored within the file, with associated attributes. The structure of the LST_cci files is fully specified in [AD-2]. Also, the file naming follows a harmonized concept [AD-2], its structure is explained for the four data files used in the validation below, the components varying for the MMDB data files are introduced in Table 2.

Table 2: Components of the MMDB file names (from [AD-2])

Name	Definition	Description	LST_cci definition
<Product String>	A character string identifying the LST product set.	Each ECV team defines the Product String they will use for their data and make this information available in their documentation. The product string must not include hyphens, but can include underscores.	*See Table 10 of AD-2*
<Indicative Date>	YYYYMMDD	The identifying date for this data set. Format is YYYYMMDD, where YYYY is the four digit year, MM is the two digit month (either 00 for annual files, or from 01 to 12) and DD is the two digit day of the month (either 00 for annual and monthly files, or from 01 to 31). The date used should best represent the observation date for the data set.	As stated in the description column.
<Indicative Time>	HHMMSS	The identifying time for this data set in UTC. Format is HHMMSS where HH is the two digit hour from 00 to 23, MM is the two digit minute from 00 to 59 and SS is the two digit second from 00 to 59.	As stated in the description column.
<FileVersion>	x.xx	File version number in the form n{1,}[.n{1,}] (That is 1 or more digits followed by optional . and another 1 or more digits.) Each external cycle will increment main digit by 1. Internal cycle will increment first digit after decimal point to 5. Each minor release will increment by second digit after decimal point. For example: 1.53 would be the Year 2 internal cycle 3rd minor release (such as due to a bug fix).	As stated in the description column.

The global attributes that are common to all data files are introduced in Table 3, the different storage types in Table 4.

Table 3: Global attributes common to all data files [from AD-2]. The orange coloured entries are only mandatory for some data files.

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	Beginning of global attributes					
Title	Insert satellite sensor name and data level into the string to create a descriptive name for the data.	ESA LST CCI <Product String> <Data Level> product		st		
Institution	Name of institution at which file was created			st		
Source	Comma separated list of all source data present in this file. List LST sources first, followed by auxiliary sources.			st		
History	To contain a history of applications that have been used to process the data. For example this can be used to record the version of the retrieval algorithm applied to the data, the identity of intermediate files used in the processing etc.			st		
References	Published or web based references.					
tracking_id	Universally Unique Identifier. For example see http://www.ossp.org/pkg/lib/uuid/			st		
Conventions	The version of the NetCDF conventions followed.	"CF-1.8"		st		
product_version	Product version			st		
Summary	A paragraph describing the dataset.			st		
keywords	A comma-separated list of key words and phrases.	Typical keywords include: Earth Science, Land		st		

Element name	Description	Range/value	Unit	T	D	Bytes
		Surface, Land Temperature and Land Surface Temperature				
Id	The filename of the file.			st		
naming_authority	The naming authority.	"ESA"		st		
keywords_vocabulary	The guideline being followed for the words/phrases in the "keywords" attribute.	"NASA Global change Master Directory (GCMD) Science Keywords"		st		
cdm_data_type	The THREDDS data type appropriate for this dataset.	"swath" for L2P and "grid" for L3 files.		st		
Comment	Miscellaneous information about the data or methods used to produce it.	"These data were produced as part of the ESA LST CCI+ project."		st		
date_created	File creation	Format: yyyymmddThhmmssZ		st		
creator_name	Provide a name and email address for the most relevant point of contact, as well as a URL relevant to this data set.			st		
creator_url				st		
creator_email				st		
Project	The scientific project that produced the data.	"Climate Change Initiative - European Space Agency"		st		
geospatial_lat_min	Southernmost latitude in decimal degrees north, range -90 to +90.			fl		
geospatial_lat_max	Northernmost latitude in decimal degrees north, range -180 to +180.			fl		
geospatial_lon_min	Westernmost longitude in decimal degrees north, range -180 to +180.			fl		

Element name	Description	Range/value	Unit	T	D	Bytes
geospatial_lon_max	Easternmost longitude in decimal degrees north, range -180 to +180.			fl		
geospatial_vertical_min	Assumed to be in metres above ground unless geospatial_vertical_units attribute defined otherwise.			fl		
geospatial_vertical_max	Assumed to be in metres above ground unless geospatial_vertical_units attribute defined otherwise.			fl		
time_coverage_start	Date and time of the first measurement in the data file.	Format: yyyymmddThhmmssZ		st		
time_coverage_end	Date and time of the last measurement in the data file.	Format: yyyymmddThhmmssZ		st		
time_coverage_duration	An ISO8601 string of the difference between time_coverage_start and time_coverage_end.	Format: PdDThHmMsS where d is the number of days, h is the number of hours, m is the number of minutes, s is the number of seconds, omitting dD etc. if the number is zero.		st		
time_coverage_resolution	An ISO8601 string of the time coverage resolution for the data in the file.	'satellite_orbit_frequency' for L2P data. ISO8601 strings for L3 data, format: PdDThHmMsS where d is the number of days, h is the number of hours, m is the		st		

Element name	Description	Range/value	Unit	T	D	Bytes
		number of minutes, s is the number of seconds, omitting dD etc. if the number is zero.				
standard_name_vocabulary	The name of the controlled vocabulary from which variable standard names are taken.	Set to "NetCDF Climate and Forecast (CF) Metadata Convention version 1.8".		st		
License	Describes the data license.	"ESA CCI Data Policy: free and open access"		st		
Platform	Satellite names from the CCI common vocabulary list. Comma-separated if more than one and angled brackets for a platform series.	See AD-2 for the platforms relevant to the LST_cci products.		st		
Sensor	Sensor names from the CCI common vocabulary list. Comma-separated if more than one.	See AD-2 for the sensors relevant to the LST_cci products.		st		
spatial_resolution	String describing the approximate resolution of the product For example, “1.1km at nadir”.	Value depends on the product.		st		
key_variables	A comma-separated list of the key primary variables in the file i.e. those that have been scientifically validated and are appropriated for display in the CCI Open Data Portal and CCI Toolbox.	"land_surface_temperature"		st		
sense	sensor zenith angle; mandatory for IS data files			fl		

Element name	Description	Range/value	Unit	T	D	Bytes
Senaz	sensor azimuth angle; mandatory for IS data files			fl		
insitu_lat	in situ latitude in decimal degrees north; mandatory for SI data files			fl		
insitu_lon	in situ longitude in decimal degrees east; mandatory for SI data files			fl		
Sensor	sensor name; mandatory for SS data files					
Attributes	End of global attributes					

Table 4: Abbreviations used for storage types [AD-2].

Type	Description	Common name	Comment
sc	8-bit signed integer	Byte	The NetCDF data type names are the same as the common names mentioned in “Common name” for these types.
ss	16-bit signed integer	Short	
sl	32-bit signed integer	int (or long)	
fl	32-bit floating point	Float	
db	64-bit floating point	double	
st	Character array	String	To be stored in the NetCDF file as an array of characters (8-bit unsigned integers)

5.2.1. Satellite Extraction Files (SE)

Satellite extraction files are satellite data sets centred over the area of the respective ground-based validation station for matching with in situ LST data. The structure of the LST_cci SE data file name is:

ESACCI-LST-SE_-LST-SE_-LST-<Product String>-<Indicative Date>[<Indicative Time>]-fv<FileVersion>.nc

Thus, the file name for an exemplary MODIS Terra SE file would be: ESACCI-LST-SE_-LST-MODIST-20180401011500-fv1.00.nc

The dimensions and variables used in the CCI LST SE data files are described below in Table 5, presented in the same way as in [AD-2].

Table 5: Specification of the LST_cci SE files. T contains the storage type (see Table 4 for abbreviations), D the dimensionality, Bytes the storage requirement.

Element name	Description	Range/value	Unit	T	D	Bytes
Dataset	Beginning of dataset					
Dimensions	Beginning of dimensions					
time	Time coordinate. Reference time of file					
lat	Latitude coordinate					
lon	Longitude coordinate					
Dimensions	End of dimensions					
Variables	Beginning of variables					
Variable	Beginning of variable					
time	Time coordinate. Reference time of file start as seconds since 1981-01-01 00:00:00		s	fl	1	4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	reference time of file		st		
standard_name	Unique descriptive name for data.	time		st		
units	Text description of the units.	seconds since 1981-01-01 00:00:00		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
calendar	Defines the calendar used to define the times.	gregorian		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lat	Centre latitude in decimal degrees north	[-90, 90]	degrees_north	fl	1	lat x 4
attributes	Beginning of attributes					

Element name	Description	Range/value	Unit	T	D	Bytes
long_name	A free-text descriptive variable name.	latitude_coordinates		st		
standard_name	Unique descriptive name for data.	latitude		st		
units	Text description of the units.	degrees_north		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-90		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	90		fl		4
reference_datum	Information about the coordinates.	geographical coordinates, WGS84 projection		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lon	Centre longitude in decimal degrees east	[-180, 180]	degrees_east	fl	1	Lon x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	longitude_coordinates		st		
standard_name	Unique descriptive name for data.	longitude		st		
units	Text description of the units.	degrees_east		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-180		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	180		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
reference_datum	Information about the coordinates.	geographical coordinates, WGS84 projection		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
dtime	Time difference from reference time.		s	fl	1	time x lat x lon x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	time difference from reference time		st		
units	Text description of the units.	seconds		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	86400		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
qual_flag	Quality flags.	[0, 1023]		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	Quality flags		st		
flag_meanings	Meaning attached to each flag value.	day_or_night-1_is_night summary_cloud-1_is_cloudy summary_confidence-		st		

Element name	Description	Range/value	Unit	T	D	Bytes
		1_is_low_confidence aerosol_mask-1_is_aerosol_detected ocean_flag land_flag lake_flag coast_flag tidal_flag seaice_flag				
flag_masks	Bit masks corresponding to the flags described in flag_meanings.	1, 2, 4, 8, 16, 32, 64, 128, 256, 512		ss		2
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	1023		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lst	Land Surface Temperature.	[-8315, 7685] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land surface temperature		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	273.15		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-8315		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	7685		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
variable	End of variable					
variable	Beginning of variable					
lst_uncertainty	Land Surface Temperature Total Uncertainty.	[0, 10000] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land surface temperature total uncertainty		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
satze	Satellite zenith angle.	[0, 18000] (after scaling)		ss	1	time x lat x lon x 2
attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	satellite zenith angle		st		
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
sataz	Satellite azimuth angle.	[-18000, 18000] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	satellite azimuth angle		st		
units	Text description of the units.	degrees		st		

Element name	Description	Range/value	Unit	T	D	Bytes
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-18000		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variables	End of variables					
Attributes	Beginning of global attributes					
Global attributes that have values that are consistent between data levels are defined in ***Table 3***						
Attributes	End of global attributes					
Dataset	End of Dataset					

Table 6: Detailed specifications of optional variables for SE data files

Element name	Description	Range/value	Unit	T	D	Bytes
Variables	Beginning of variables					
Variable	Beginning of variable					
channel	Channel coordinates.					
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	channel		st		
Units	Text description of the units.	**Channel wavelength in microns**		st		

Element name	Description	Range/value	Unit	T	D	Bytes
_FillValue	A value used to indicate array elements containing no valid data.	-32678		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	15000		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
emis	Land Surface Emissivity	[0, 10000] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	surface emissivity		st		
units	Text description of the units.	unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.0001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
variable	Beginning of variable					
bt	channel brightness temperature		K			
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	brightness temperature by sensor channel		st		
standard_name,	Unique descriptive name for data.	brightness temperature		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	273.15		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-8315		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	7685		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lcc	Land cover class.	[1,27]		ss	1	time x lat x x 2

Element name	Description	Range/value	Unit	T	D	Bytes
attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land cover class		st		
units	Text description of the units.	1		st		
flag_meanings	Meaning attached to each flag value.	Post-flooding_/_irrigated_croplands Rainfed_croplands Mosaic_Cropland_(50-70%)_/_Vegetation_(grassland,_shrubland,_forest)_(20-50%) Mosaic_Vegetation_(grassland,_shrubland,_forest)_(50-70%)_/_Cropland_(20-50%) Closed_to_open_(>15%)_broadleaved_evergreen_and/or_semi-deciduous_forest_(>5m) Closed_(>40%)_broadleaved_deciduous_forest_(>5m) Open_(15-40%)_broadleaved_deciduous_forest_(>5m) Closed_(>40%)_needleleaved_evergreen_forest_(>5m) Open_(15-40%)_needleleaved_deciduous_or_evergreen_forest_(>5m) Closed_to_open_(>15%)_mixed_broadleaved_and_needleleaved_forest_(>5m) Mosaic_Forest/Shrubland_(50-70%)_/_Grassland_(20-50%) Mosaic_Grassland_(50-70%)_/_Forest/Shrubland_(20-50%) Closed_to_open_(>15%)_shrubland_(<5m) Closed_to_open_(>15%)_grassland Sparse_(>15%)_vegetat		st		

Element name	Description	Range/value	Unit	T	D	Bytes
		ion_(woody_vegetation,_shrubs,_grassland) Closed_(>40%)_broadleaved_forest_regularly_flooded_-Fresh_water Closed_(>40%)_broadleaved_semi-deciduous_and/or_evergreen_forest_regularly_flooded_-Saline_water Closed_to_open_(>15%)_vegetation_(grassland,_shrubland,_woody_vegetation)_on_regularly_flooded_or_waterlogged_soil_-Fresh,_brackish_or_saline_water Artificial_surfaces_and_associated_areas_(urban_areas_>50%) Bare_areas_of_soil_types_not_contained_in_biomes_21_to_25 Bare_areas_of_soil_type_Entisols_–_Orthents Bare_areas_of_soil_type_Shifting_sand Bare_areas_of_soil_type_Aridisols_-_Calcids Bare_areas_of_soil_type_Aridisols_-_Cambids Bare_areas_of_soil_type_Gelisols_-_Orthels Water_bodies_(inland_lakes,_rivers,_sea:_max_10km_away_from_coast) Permanent_snow_and_ice_(<5m), Closed_to_open_(>15%)_grassland, Sparse_(>15%)_vegetation_(woody_vegetation,_shrubs,_grassland), Closed_(>40%)_broadleaved_forest_regularly_flooded_-Fresh_water, Closed_(>40%)_broadleaved_semi-				

Element name	Description	Range/value	Unit	T	D	Bytes
		deciduous_and/or_evergreen_forest_regularly_flooded_-Saline_water, Closed_to_open_(>15%)_vegetation_(grassland,_shrubland,_woody_vegetation)_on_regularly_flooded_or_waterlogged_soil_-Fresh,_brackish_or_saline_water, Artificial_surfaces_and_associated_areas_(urban_areas_>50%), Bare_areas_of_soil_types_not_contained_in_biomes_21_to_25, Bare_areas_of_soil_type_Entisols_–_Orthents, Bare_areas_of_soil_type_Shifting_sand, Bare_areas_of_soil_type_Aridisols_-_Calcids, Bare_areas_of_soil_type_Aridisols_-_Cambids, Bare_areas_of_soil_type_Gelisols_-_Orthels, Water_bodies_(inland_lakes,_rivers,_sea:_max_10km_away_from_coast), Permanent_snow_and_ice				
flag_values	Values corresponding to the flags described in flag_meanings.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27		ss		2
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	1		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	27		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	End of attributes					
variable	End of variable					
Variable	Beginning of variable					
fv	Fractional vegetation cover.	[0, 10000] (after scaling)				
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	Fractional vegetation cover.		st		
units	Text description of the units.	Unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.0001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	20000		fl		4
source	Auxiliary data source.	CGLPS FCOVER dataset:		st		
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
tcwv	total column water vapour	[0, 20000] (after scaling)		ss	1	time x lon x lat x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	total column water vapour		st		
units	Text description of the units.	kg m-2		st		

Element name	Description	Range/value	Unit	T	D	Bytes
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.004		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	20000		fl		4
source	Auxiliary data source.	ECMWF ERA-Interim dataset: http://www.ecmwf.int/ (Cycle 1), ECMWF ERA-5 dataset: (Cycle 2 onwards)				
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
solze	Solar zenith angle.	[0, 18000] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	solar zenith angle		st		
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
solaz	solar azimuth angle.	[-18000, 18000] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	solar azimuth angle		st		
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-18000		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					

Element name	Description	Range/value	Unit	T	D	Bytes
Variable	End of variable					
Variables	End of variables					

5.2.2. In Situ Data Files (IS)

All in situ data files are produced by KIT validation team, and consist each of yearly data of in situ LST measurements over one station. The file name convention for the IS files is:

ESACCI-LST-IS_-<Product String>-<Indicative Date>[<Indicative Time>]-fv<FileVersion>.nc For the components of the file name see Table 2.

For an exemplary IS data file over GBB_W station the file name would be: ESACCI-LST-IS_-LST-GBB_W_-20180401011500-fv1.00.nc

The dimensions and variables of the IS data files are specified in Table 7.

Table 7: Specification of the LST_cci IS files. T contains the storage type (see Table 4 for abbreviations), D the dimensionality, Bytes the storage requirement.

Element name	**Description**	**Range/value**	**Unit**	**T**	**D**	**Bytes**
Dataset	Beginning of dataset					
Dimensions	Beginning of dimensions					
sensor	Sensor name					
time	Time coordinate. Reference time of file					
lat	Latitude coordinate					
lon	Longitude coordinate					
sr_length	Length of sensor name					
Dimensions	End of dimensions					
Variables	Beginning of variables					
Variable	Beginning of variable					
time	Time coordinate. Reference time of file start as seconds since 1981-01-01 00:00:00		s	fl	1	4
Attributes	Beginning of attributes					

Element name	Description	Range/value	Unit	T	D	Bytes
long_name	A free-text descriptive variable name.	Reference time of file		st		
standard_name	Unique descriptive name for data.	time		st		
units	Text description of the units.	seconds since 1981-01-01 00:00:00		st		
calendar	Defines the calendar used to define the times.	gregorian		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lat	Centre latitude in decimal degrees north	[-90, 90]	degrees_north	fl	1	lat x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	latitude_coordinates		st		
standard_name	Unique descriptive name for data.	latitude		st		
units	Text description of the units.	degrees_north		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-90		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	90		fl		4
reference_datum	Information about the coordinates.	geographical coordinates, WGS84 projection		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lon	Centre longitude in decimal degrees east	[-180, 180]	degrees_east	fl	1	lon x 4

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	longitude_coordinates		st		
standard_name	Unique descriptive name for data.	longitude		st		
units	Text description of the units.	degrees_east		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-180		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	180		fl		4
reference_datum	Information about the coordinates.	geographical coordinates, WGS84 projection		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
dtime	Time difference from reference time.		s	fl	1	time x lat x lon x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	time difference from reference time		st		
units	Text description of the units.	seconds		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	31536000		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
coordinates	Identifies coordinate variables.	lon lat		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
elevation	Elevation of land surface					time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	surface elevation		st		
Units	Text description of the units.	m		st		
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-700		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lst	Land Surface Temperature.	[-8315, 7685] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land surface temperature		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying	273.15		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
	by the scale factor to recover the original value.					
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-8315		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	7685		fl		4
coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lst_uncertainty	Land Surface Temperature Total Uncertainty.	[0, 10000] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land surface temperature total uncertainty		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variables	End of variables					
Attributes	Beginning of global attributes					
Global attributes that have values that are consistent between data levels are defined in ***Table 3***.						
Attributes	End of global attributes					
Dataset	End of Dataset					

Table 8: Detailed specifications of optional variables for IS data files

Element name	Description	Range/value	Unit	T	D	Bytes
Variables	Beginning of variables					
Variable	Beginning of variable					
endmember	name of end-member observed by sensor				1	sensor
Attributes	Beginning of attributes					
units	Text description of the units.	1		st		
long_name	A free-text descriptive variable name.	name of end-member observed by sensor				
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
coordinates	Identifies coordinate variables.	sr_length sensor		st		
Variable	End of variable					
Variable	Beginning of variable					
bt	brightness temperature of end-member observed by sensor.					time x sensor x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	brightness temperature by sensor channel.		st		

Element name	Description	Range/value	Unit	T	D	Bytes
standard_name	Unique descriptive name for data.	Brightness_temperature		st		
Units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	273.15		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-8315		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	7685		fl		4
coordinates	Identifies coordinate variables.	sensor time		st		
source	Data source			st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
weight	relative weight of measured BT; sum over sensors has to equal 1.0					time x sensor x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	relative weight of measured BT		st		
units	Text description of the units.	unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	100		fl		4
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
emis	emissivity of end-member observed by sensor					time x sensor x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	emissivity of end-member observed by sensor		st		
standard_name	Unique descriptive name for data.	Surface_longwave_emissivity		st		
units	Text description of the units.	unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.0001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
coordinates	Identifies coordinate variables.	time sensor lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
bt_sky	sky brightness temperature					time x lat x lon x 2

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	sky brightness temperature		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	273.15		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	- 8315		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	7685		fl		4
coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
bt_unc	brightness temperature uncertainty					time x sensor x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	brightness temperature uncertainty		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
coordinates	Identifies coordinate variables.	time sensor lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
bt_sky_unc	Sky brightness temperature uncertainty					time x sensor x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	sky brightness temperature uncertainty		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	1000		fl		4
coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					

Element name	Description	Range/value	Unit	T	D	Bytes
solze	Solar zenith angle.	[0, 18000] (after scaling)				time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	solar zenith angle		st		
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
solaz	Solar azimuth angle.	[-18000, 18000] (after scaling)				time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	solar azimuth angle		st		
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying	0		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
	by the scale factor to recover the original value.					
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-18000		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
t2m	2m Surface Air Temperature					time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	2m air temperature		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	273.15		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-8315		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	7685		fl		4
coordinates	Identifies coordinate variables.	time lat lon		st		

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
sh2m	2 m specific humidity					time x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	2m specific humidity		st		
units	Text description of the units.	unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.0001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
Source	Auxiliary data source	ECMWF ERA-5 dataset: https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era5				
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
ws2m	2m wind speed					time x lat x lon x 2
Attributes	Beginning of attributes					

Element name	Description	Range/value	Unit	T	D	Bytes
long_name	A free-text descriptive variable name.	2m wind speed		st		
Units	Text description of the units.	m s-1		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.004		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	20000		fl		4
coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
tcwv	total column water vapor					time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	total column water vapour		st		
units	Text description of the units.	kg m-2		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
scale_factor	To be multiplied by the variable to recover the original value.	0.004		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	20000		fl		4
Coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lcc	Land cover class.	[1,27]		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land cover class		st		
units	Text description of the units.	unitless		st		
flag_meanings	Meaning attached to each flag value.	Post-flooding_/_irrigated_croplands Rainfed_croplands Mosaic_Cropland_(50-70%)_/_Vegetation_(grassland,_shrubland,_forest)_(20-50%) Mosaic_Vegetation_(grassland,_shrubland,_forest)_(50-70%)_/_Cropland_(20-50%) Closed_to_open_(>15%)_broadleaved_evergreen_and/or_semi-deciduous_forest_(>5m) Closed_(>40%)_broadleaved_deciduous_forest_(>5m) Open_(15-40%)_broadleaved_deciduous_forest_(>5m) Closed_(>40%)_needlel		st		

Element name	Description	Range/value	Unit	T	D	Bytes
		eaved_evergreen_forest_(>5m) Open_(15-40%)_needleleaved_deciduous_or_evergreen_forest_(>5m) Closed_to_open_(>15%)_mixed_broadleaved_and_needleleaved_forest_(>5m) Mosaic_Forest/Shrubland_(50-70%)_/_Grassland_(20-50%) Mosaic_Grassland_(50-70%)_/_Forest/Shrubland_(20-50%) Closed_to_open_(>15%)_shrubland_(<5m) Closed_to_open_(>15%)_grassland Sparse_(>15%)_vegetation_(woody_vegetation,_shrubs,_grassland) Closed_(>40%)_broadleaved_forest_regularly_flooded_-Fresh_water Closed_(>40%)_broadleaved_semi-deciduous_and/or_evergreen_forest_regularly_flooded_-Saline_water Closed_to_open_(>15%)_vegetation_(grassland,_shrubland,_woody_vegetation)_on_regularly_flooded_or_waterlogged_soil_-Fresh,_brackish_or_saline_water Artificial_surfaces_and_associated_areas_(urban_areas_>50%) Bare_areas_of_soil_types_not_contained_in_biomes_21_to_25 Bare_areas_of_soil_type_Entisols_–_Orthents Bare_areas_of_soil_type_Shifting_sand Bare_areas_of_soil_type_Aridisols_-_Calcids				

Element name	Description	Range/value	Unit	T	D	Bytes
		Bare_areas_of_soil_type_Aridisols_-_Cambids Bare_areas_of_soil_type_Gelisols_-_Orthels Water_bodies_(inland_lakes,_rivers,_sea:_max_10km_away_from_coast) Permanent_snow_and_ice_(<5m), Closed_to_open_(>15%)_grassland, Sparse_(>15%)_vegetation_(woody_vegetation,_shrubs,_grassland), Closed_(>40%)_broadleaved_forest_regularly_flooded_-Fresh_water, Closed_(>40%)_broadleaved_semi-deciduous_and/or_evergreen_forest_regularly_flooded_-Saline_water, Closed_to_open_(>15%)_vegetation_(grassland,_shrubland,_woody_vegetation)_on_regularly_flooded_or_waterlogged_soil_-Fresh,_brackish_or_saline_water, Artificial_surfaces_and_associated_areas_(urban_areas_>50%), Bare_areas_of_soil_types_not_contained_in_biomes_21_to_25, Bare_areas_of_soil_type_Entisols_–_Orthents, Bare_areas_of_soil_type_Shifting_sand, Bare_areas_of_soil_type_Aridisols_-_Calcids, Bare_areas_of_soil_type_Aridisols_-_Cambids, Bare_areas_of_soil_type_Gelisols_-_Orthels, Water_bodies_(inland_lakes,_rivers,_sea:_max_10km_away_from_c				

Element name	Description	Range/value	Unit	T	D	Bytes
		oast), Permanent_snow_and_ice				
flag_masks	Values corresponding to the flags described in flag_meanings.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27		ss		2
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	1		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	27		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
fv	fractional vegetation cover					time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	fractional vegetation cover		st		
units	Text description of the units.	unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.0001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_max	Maximum valid value for this variable once they are packed (in storage type).	20000		fl		4
source	Auxiliary data source.	CGLPS FCOVER 1 km dataset v2.0, which has been brokered to C3S: https://land.copernicus.eu/global/products/fcover		st		
coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variables	End of variables					

5.2.3. Matched Satellite – In Situ Data Files (SI)

The SE data files are matched with the corresponding IS files and the resulting matches are stored in a matched satellite – in situ data file (SI). These data files are yearly data files over one station and needed for the actual validation.

The file naming convention for SI files is:

ESACCI-<CCI Project>-<Processing Level>-<Data Type>-<Product String>[-<Additional Segregator>]-<Indicative Date>[<Indicative Time>]-fv<FileVersion>.nc.

An exemplarily data file name for a MODIS Terra satellite data set matched with in situ data from GBB_W station would be: ESACCI-LST-SI_-LST-MODIST-GBB_W_-20180401011500-fv1.00.nc.

The dimensions and variables of the SI files are listed below in Table 9.

Table 9: Specification of the LST_cci SI files. T contains the storage type (see Table 4 for abbreviations), D the dimensionality, Bytes the storage requirement.

Element name	Description	Range/value	Unit	T	D	Bytes
Dataset	Beginning of dataset					
Dimensions	Beginning of dimensions					
sensor	Sensor name, size 2: 1 = satellite, 2 = in situ					
time	Time coordinate. Reference time of file					

Element name	Description	Range/value	Unit	T	D	Bytes
lat	Latitude coordinate					
lon	Longitude coordinate					
channel	Channel description					
Dimensions	End of dimensions					
Variables	Beginning of variables					
Variable	Beginning of variable					
sensor	sensor name					sensor x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	sensor name		st		
units	Text description of the units.	Unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
time	Time coordinate. Reference time of file starts as seconds since 1981-01-01 00:00:00		s	fl	1	Time x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	Reference time of file start as seconds since 1981-01-01 00:00:00		st		
standard_name	Unique descriptive name for data.	time		st		
units	Text description of the units.	seconds since 1981-01-01 00:00:00		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
calendar	Defines the calendar used to define the times.	gregorian		st		

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lat	Centre latitude in decimal degrees north	[-90, 90]	degrees_north	fl	1	1
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	latitude_coordinates		st		
standard_name	Unique descriptive name for data.	latitude		st		
units	Text description of the units.	degrees_north		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-90		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	90		fl		4
reference_datum	Information about the coordinates.	geographical coordinates, WGS84 projection		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lon	Centre longitude in decimal degrees east	[-180, 180]	degrees_east	fl	1	1
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	longitude_coordinates		st		
standard_name	Unique descriptive name for data.	longitude		st		
units	Text description of the units.	degrees_east		st		

Element name	Description	Range/value	Unit	T	D	Bytes
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-180		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	180		fl		4
reference_datum	Information about the coordinates.	geographical coordinates, WGS84 projection		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
dtime	Time difference from reference time.		s	fl	1	time x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	time difference from reference time, Seconds since start of file (reference time)		st		
units	Text description of the units.	seconds		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	31622400		fl		4
coordinates	Identifies coordinate variables.	time		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					

Element name	Description	Range/value	Unit	T	D	Bytes
lst	satellite / in situ Land Surface Temperature.	[-8315, 7685] (after scaling)		ss	1	time x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	satellite / in situ land surface temperature		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	273.15		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-8315		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	7685		fl		4
coordinates	Identifies coordinate variables.	time sensor		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lst_uncertainty	Satellite / In Situ Land Surface Temperature Total Uncertainty.	[0, 10000] (after scaling)		ss	1	time x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	satellite / in situ land surface temperature total uncertainty		st		
Units	Text description of the units.	kelvin		st		

Element name	Description	Range/value	Unit	T	D	Bytes
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
coordinates	Identifies coordinate variables.	time sensor		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
satze	Satellite zenith angle.	[0, 18000] (after scaling)		ss	1	time x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	satellite zenith angle		st		
Units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	time		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
sataz	Satellite azimuth angle.	[-18000, 18000] (after scaling)		ss	1	time x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	satellite azimuth angle		st		
Units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-18000		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	time		st		
Attributes	End of attributes					
Variable	End of variable					
Variables	End of variables					
Attributes	Beginning of global attributes					
Global attributes that have values that are consistent between data levels are defined in Table 3						
Attributes	End of global attributes					

Element name	Description	Range/value	Unit	T	D	Bytes
Dataset	End of Dataset					

Table 10: Detailed specifications of optional variables for SI data files

Element name	Description	Range/value	Unit	T	D	Bytes
Variable	Beginning of variable					
n	Number of pixels flagged as clear sky					time x 2
Attributes	Beginning of attributes					
units	Text description of the units.	Unitless		st		
long_name	A free-text descriptive variable name.	Number of pixels averaged		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		ss		2
valid_max	Maximum valid value for this variable once they are packed (in storage type).	75000		ss		2
Coordinates	Identifies coordinate variables.	time		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
ncld	Number of pixels flagged as cloudy					time x 2
Attributes	Beginning of attributes					
Units	Text description of the units.	unitless		st		
long_name	A free-text descriptive variable name.	Number of pixels excluded		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		ss		2
valid_max	Maximum valid value for this variable once they are packed (in storage type).	75000		ss		2
Coordinates	Identifies coordinate variables.	time		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
daynight	Day / Night index					time x 2
Attributes	Beginning of attributes					
units	Text description of the units.	unitless		st		
long_name	A free-text descriptive variable name.	day/night indicator		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
coordinates	Identifies coordinate variables.	time		st		
flag_meanings	Meaning attached to each flag value.	day = 0, night = 1		st		
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	1		fl		4
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
solze	Solar zenith angle.	[0, 18000] (after scaling)				time x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	solar zenith angle		st		

Element name	Description	Range/value	Unit	T	D	Bytes
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	time		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
solaz	Solar azimuth angle.	[-18000, 18000] (after scaling)				time x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	solar azimuth angle		st		
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	Time		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lcc	Land cover class.	[1,27]		ss	1	time x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land cover class		st		
units	Text description of the units.	unitless		st		
flag_meanings	Meaning attached to each flag value.	Post-flooding_/_irrigated_croplands Rainfed_croplands Mosaic_Cropland_(50-70%)_/_Vegetation_(grassland,_shrubland,_forest)_(20-50%) Mosaic_Vegetation_(grassland,_shrubland,_forest)_(50-70%)_/_Cropland_(20-50%) Closed_to_open_(>15%)_broadleaved_evergreen_and/or_semi-deciduous_forest_(>5m) Closed_(>40%)_broadleaved_deciduous_forest_(>5m) Open_(15-40%)_broadleaved_deciduous_forest_(>5m) Closed_(>40%)_needleleaved_evergreen_forest_(>5m) Open_(15-40%)_needleleaved_deciduous_or_evergreen_forest_(>5m)		st		

Element name	Description	Range/value	Unit	T	D	Bytes
		Closed_to_open_(>15%)_mixed_broadleaved_and_needleleaved_forest_(>5m) Mosaic_Forest/Shrubland_(50-70%)_/_Grassland_(20-50%) Mosaic_Grassland_(50-70%)_/_Forest/Shrubland_(20-50%) Closed_to_open_(>15%)_shrubland_(<5m) Closed_to_open_(>15%)_grassland Sparse_(>15%)_vegetation_(woody_vegetation,_shrubs,_grassland) Closed_(>40%)_broadleaved_forest_regularly_flooded_-Fresh_water Closed_(>40%)_broadleaved_semi-deciduous_and/or_evergreen_forest_regularly_flooded_-Saline_water Closed_to_open_(>15%)_vegetation_(grassland,_shrubland,_woody_vegetation)_on_regularly_flooded_or_waterlogged_soil_-Fresh,_brackish_or_saline_water Artificial_surfaces_and_associated_areas_(urban_areas_>50%) Bare_areas_of_soil_types_not_contained_in_biomes_21_to_25 Bare_areas_of_soil_type_Entisols_–_Orthents Bare_areas_of_soil_type_Shifting_sand Bare_areas_of_soil_type_Aridisols_-_Calcids Bare_areas_of_soil_type_Aridisols_-_Cambids Bare_areas_of_soil_type_Gelisols_-_Orthels Water_bodies_(inland_lakes,_rivers,_sea:_max_1				

Element name	Description	Range/value	Unit	T	D	Bytes
		0km_away_from_coast) Permanent_snow_and_ice_(<5m), Closed_to_open_(>15%)_grassland, Sparse_(>15%)_vegetation_(woody_vegetation,_shrubs,_grassland), Closed_(>40%)_broadleaved_forest_regularly_flooded_-Fresh_water, Closed_(>40%)_broadleaved_semi-deciduous_and/or_evergreen_forest_regularly_flooded_-Saline_water, Closed_to_open_(>15%)_vegetation_(grassland,_shrubland,_woody_vegetation)_on_regularly_flooded_or_waterlogged_soil_-Fresh,_brackish_or_saline_water, Artificial_surfaces_and_associated_areas_(urban_areas_>50%), Bare_areas_of_soil_types_not_contained_in_biomes_21_to_25, Bare_areas_of_soil_type_Entisols_–_Orthents, Bare_areas_of_soil_type_Shifting_sand, Bare_areas_of_soil_type_Aridisols_-_Calcids, Bare_areas_of_soil_type_Aridisols_-_Cambids, Bare_areas_of_soil_type_Gelisols_-_Orthels, Water_bodies_(inland_lakes,_rivers,_sea:_max_10km_away_from_coast), Permanent_snow_and_ice				
flag_masks	Values corresponding to the flags described in flag_meanings.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27		ss		2

Element name	Description	Range/value	Unit	T	D	Bytes
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	1		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	27		fl		4
source	Auxiliary data source.	**ESA CCI Land Cover Data** **https://www.esa-landcover-cci.org/**		st		
coordinates	Identifies coordinate variables.	Time		st		
Attributes	End of attributes					
Variable	End of variable					
Variables	End of variables					

5.2.4. Matched Satellite – Satellite Data Files (SS)

For the satellite – satellite intercomparisons, two satellite data sets are temporarily and spatially matched. The matches are saved in matched satellite – satellite data files (SS), where each file contains one match-up.

The file naming convention for SS files is:
ESACCI-<CCI Project>-<Processing Level>-<Data Type>-<Product String>[-<Additional Segregator>]-<Indicative Date>[<Indicative Time>]-fv<FileVersion>.nc.
An exemplary file name for a SS data file, where MODIS Terra and SEVIRI satellite LST data are matched, would be: ESACCI-LST-SS_-LST-MODIST-SEVIRI-20180401011500-fv1.00.nc.

The dimensions and variables from the SS data files are summarized in Table 11.

Table 11: Specification of the LST_cci SS files. T contains the storage type (see Table 4 for abbreviations), D the dimensionality, Bytes the storage requirement.

Element name	Description	Range/value	Unit	T	D	Bytes
Dataset	Beginning of dataset					
Dimensions	Beginning of dimensions					
sensor	Sensor name					
time	Time coordinate. Reference time of file					
Lat	Latitude coordinate					
Lon	Longitude coordinate					
Channel	Channel description					
ch_length	Length of channel name					
Dimensions	End of dimensions					
Variables	Beginning of variables					
Variable	Beginning of variable					
time	Time coordinate. Reference time of file start as seconds since 1981-01-01 00:00:00		S	fl	1	time x sensor x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	reference time of file		st		
standard_name	Unique descriptive name for data.	time		st		
units	Text description of the units.	seconds since 1981-01-01 00:00:00		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
calendar	Defines the calendar used to define the times.	gregorian		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lat	Centre latitude in decimal degrees north	[-90, 90]	degrees_north	fl	1	lat x 4

Element name	Description	Range/value	Unit	T	D	Bytes
attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	latitude_coordinates		st		
standard_name	Unique descriptive name for data.	latitude		st		
units	Text description of the units.	degrees_north		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-90		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	90		fl		4
reference_datum	Information about the coordinates.	geographical coordinates, WGS84 projection		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lon	Centre longitude in decimal degrees east	[-180, 180]	degrees_east	fl	1	lon x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	longitude_coordinates		st		
standard_name	Unique descriptive name for data.	longitude		st		
units	Text description of the units.	degrees_east		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-180		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	180		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
reference_datum	Information about the coordinates.	geographical coordinates, WGS84 projection		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
dtime	Time difference from reference time.		s	fl	1	time x sensor x lat x lon x 4
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	time difference from reference time		st		
units	Text description of the units.	seconds		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	86400		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
n	Number of pixels flagged as clear sky		fl			time x sensor x lat x lon x 2
Attributes	Beginning of attributes					
units	Text description of the units.	unitless		st		
long_name	A free-text descriptive variable name.	Number of clear-sky pixels		st		

Element name	Description	Range/value	Unit	T	D	Bytes
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		ss		2
valid_max	Maximum valid value for this variable once they are packed (in storage type).	75000		ss		2
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
ncld	Number of pixels flagged as cloudy		fl			time x sensor x lat x lon x 2
Attributes	Beginning of attributes					
units	Text description of the units.	unitless		st		
long_name	A free-text descriptive variable name.	Number of cloudy pixels		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		ss		2
valid_max	Maximum valid value for this variable once they are packed (in storage type).	75000		ss		2
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lst	Land Surface Temperature.	[-8315, 7685] (after scaling)		ss	1	time x sensor x lat x lon x 2

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land surface temperature		st		
units	Text description of the units.	kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	273.15		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-8315		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	7685		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
variable	End of variable					
variable	Beginning of variable					
lst_uncertainty	Land Surface Temperature Total Uncertainty.	[0, 10000] (after scaling)		ss	1	time x sensor x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land surface temperature total uncertainty		st		
units	Text description of the units.	Kelvin		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying	0		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
	by the scale factor to recover the original value.					
scale_factor	To be multiplied by the variable to recover the original value.	0.001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
satze	Satellite zenith angle.	[0, 18000] (after scaling)		ss	1	time x lat x lon x 2
attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	satellite zenith angle		st		
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		

Element name	Description	Range/value	Unit	T	D	Bytes
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
sataz	Satellite azimuth angle.	[-18000, 18000] (after scaling)		ss	1	time x sensor x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	satellite azimuth angle		st		
units	Text description of the units.	Degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-18000		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variables	End of variables					
Attributes	Beginning of global attributes					
Global attributes that have values that are consistent between data levels are defined in Table 3.						
Attributes	End of global attributes					
Dataset	End of Dataset					

Table 12: Detailed specifications of optional variables for SS data files

Element name	Description	Range/value	Unit	T	D	Bytes
Variables	Beginning of variables					
Variable	Beginning of variable					
channel	Channel coordinates.					
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	channel		st		
units	Text description of the units.	Channel wavelength in microns		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
emis	Land Surface Emissivity	[0, 10000] (after scaling)		ss	1	time x sensor x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	surface emissivity		st		
units	Text description of the units.	unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		Fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.0001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					

Element name	Description	Range/value	Unit	T	D	Bytes
Variable	End of variable					
Variable	Beginning of variable					
solze	Solar zenith angle.	[0, 18000] (after scaling)				time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	solar zenith angle		st		
units	Text description of the units.	degrees		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
solaz	Solar azimuth angle.	[-18000, 18000] (after scaling)				time x lat x lonx 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	solar azimuth angle		st		
units	Text description of the units.	degrees		st		

Element name	Description	Range/value	Unit	T	D	Bytes
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.01		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	18000		fl		4
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
Fv	Fractional vegetation cover.	[0, 10000] (after scaling)		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	fractional vegetation cover		st		
units	Text description of the units.	unitless		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.0001		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_max	Maximum valid value for this variable once they are packed (in storage type).	20000		fl		4
source	Auxiliary data source.	CGLPS FCOVER 1 km dataset v2.0, which has been brokered to C3S: https://land.copernicus.eu/global/products/fcover		st		
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
tcwv	Total Column Water Vapour	[0, 20000] (after scaling)		ss	1	time x sensor x lon x lat x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	total column water vapour		st		
units	Text description of the units.	kg m-2		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0.004		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	0		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	20000		fl		4
Source	Auxiliary data source.	**ECMWF ERA-Interim dataset:** http://www.ecmwf.int/ (Cycle 1), ECMWF ERA-5				

Element name	Description	Range/value	Unit	T	D	Bytes
		dataset: (Cycle 2 onwards)				
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
lcc	Land cover class.	[1,27]		ss	1	time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	land cover class		st		
units	Text description of the units.	unitless		St		
flag_meanings	Meaning attached to each flag value.	Post-flooding_/_irrigated_croplands Rainfed_croplands Mosaic_Cropland_(50-70%)_/_Vegetation_(grassland,_shrubland,_forest)_(20-50%) Mosaic_Vegetation_(grassland,_shrubland,_forest)_(50-70%)_/_Cropland_(20-50%) Closed_to_open_(>15%)_broadleaved_evergreen_and/or_semi-deciduous_forest_(>5m) Closed_(>40%)_broadleaved_deciduous_forest_(>5m) Open_(15-40%)_broadleaved_deciduous_forest_(>5m) Closed_(>40%)_needleleaved_evergreen_forest_(>5m) Open_(15-40%)_needleleaved_deciduous_or_evergreen_forest_(>5m) Closed_to_open_(>15%)_mixed_broadleaved_and_needleleaved_forest_(>5m)		st		

Element name	Description	Range/value	Unit	T	D	Bytes
		Mosaic_Forest/Shrubland_(50-70%)_/_Grassland_(20-50%) Mosaic_Grassland_(50-70%)_/_Forest/Shrubland_(20-50%) Closed_to_open_(>15%)_shrubland_(<5m) Closed_to_open_(>15%)_grassland Sparse_(>15%)_vegetation_(woody_vegetation,_shrubs,_grassland) Closed_(>40%)_broadleaved_forest_regularly_flooded_-Fresh_water Closed_(>40%)_broadleaved_semi-deciduous_and/or_evergreen_forest_regularly_flooded_-Saline_water Closed_to_open_(>15%)_vegetation_(grassland,_shrubland,_woody_vegetation)_on_regularly_flooded_or_waterlogged_soil_-Fresh,_brackish_or_saline_water Artificial_surfaces_and_associated_areas_(urban_areas_>50%) Bare_areas_of_soil_types_not_contained_in_biomes_21_to_25 Bare_areas_of_soil_type_Entisols_–_Orthents Bare_areas_of_soil_type_Shifting_sand Bare_areas_of_soil_type_Aridisols_-_Calcids Bare_areas_of_soil_type_Aridisols_-_Cambids Bare_areas_of_soil_type_Gelisols_-_Orthels Water_bodies_(inland_lakes,_rivers,_sea:_max_10km_away_from_coast) Permanent_snow_and_ice_(<5m), Closed_to_open_(>15%)				

Element name	Description	Range/value	Unit	T	D	Bytes
		grassland, Sparse(>15%)_vegetation_(woody_vegetation,_shrubs,_grassland), Closed_(>40%)_broadleaved_forest_regularly_flooded_-Fresh_water, Closed_(>40%)_broadleaved_semi-deciduous_and/or_evergreen_forest_regularly_flooded_-Saline_water, Closed_to_open_(>15%)_vegetation_(grassland,_shrubland,_woody_vegetation)_on_regularly_flooded_or_waterlogged_soil_-Fresh,_brackish_or_saline_water, Artificial_surfaces_and_associated_areas_(urban_areas_>50%), Bare_areas_of_soil_types_not_contained_in_biomes_21_to_25, Bare_areas_of_soil_type_Entisols_–_Orthents, Bare_areas_of_soil_type_Shifting_sand, Bare_areas_of_soil_type_Aridisols_-_Calcids, Bare_areas_of_soil_type_Aridisols_-_Cambids, Bare_areas_of_soil_type_Gelisols_-_Orthels, Water_bodies_(inland_lakes,_rivers,_sea:_max_10km_away_from_coast), Permanent_snow_and_ice				
flag_masks	Values corresponding to the flags described in flag_meanings.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27		ss		2
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4

Element name	Description	Range/value	Unit	T	D	Bytes
valid_min	Minimum valid value for this variable once they are packed (in storage type).	1		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	27		fl		4
source	Auxiliary data source.	ESA CCI Land Cover Data https://www.esa-landcover-cci.org/		st		
coordinates	Identifies coordinate variables.	lat lon		st		
Attributes	End of attributes					
Variable	End of variable					
Variable	Beginning of variable					
elevation	Elevation of land surface					time x lat x lon x 2
Attributes	Beginning of attributes					
long_name	A free-text descriptive variable name.	surface elevation		st		
units	Text description of the units.	m		st		
_FillValue	A value used to indicate array elements containing no valid data.	-32768		fl		4
add_offset	To be added to the variable after multiplying by the scale factor to recover the original value.	0		fl		4
scale_factor	To be multiplied by the variable to recover the original value.	0		fl		4
valid_min	Minimum valid value for this variable once they are packed (in storage type).	-700		fl		4
valid_max	Maximum valid value for this variable once they are packed (in storage type).	10000		fl		4
coordinates	Identifies coordinate variables.	time lat lon		st		
Attributes	End of attributes					

Element name	Description	Range/value	Unit	T	D	Bytes
Variable	End of variable					
Variables	End of variables					

6. In Situ Data Sets

The basis of the in situ validation that will be performed within LST_cci is the set of globally distributed measurement stations described here. A key source of data are the four stations set up and operated by KIT, which are dedicated to LST validation and were defined by [RD-1] as currently the only class A2 sites for LST validation over solid land surfaces worldwide. A narrow band, narrow FOV radiometer with similar spectral characteristics and view direction as the overpassing satellite sensor (as deployed at KIT's sites) is the ideal instrument for validating satellite-derived LST. Following KIT's station set up, the Copernicus LAW project recently deployed five new LST validation stations in five previously unrepresented forest biomes. However, such instruments are not everywhere available. Therefore, in order to have a more global set of validation matchups, some stations measuring broadband radiances are also included in the validation. These sites are those from the SURFRAD network (Surface Radiation Budget Network, https://www.esrl.noaa.gov/gmd/grad/surfrad/index.html), and an ARM (Atmospheric Radiation Measurement, https://www.arm.gov/) site in Alaska. Another ARM site measuring narrowband radiances in Oklahoma is also included.

As described in Section 5, one year of in situ data from all stations except GBB_W and Copernicus LAW stations is made available to the data producers to test their algorithms. Which year is chosen at each station depends on the temporal availability of the in situ and satellite data. The main goal was to have some in situ data sets for all satellite data, which cover different time periods. The time periods of the LEO and GEO satellite data as well as the in situ station data are shown in Figure 4 for the LEO satellite data sets and in Figure 5 for the GEO satellite data sets. According to these graphics, three years are needed to get in situ data for testing of all satellite data sets, which are going to be 1999, 2011 and 2017.

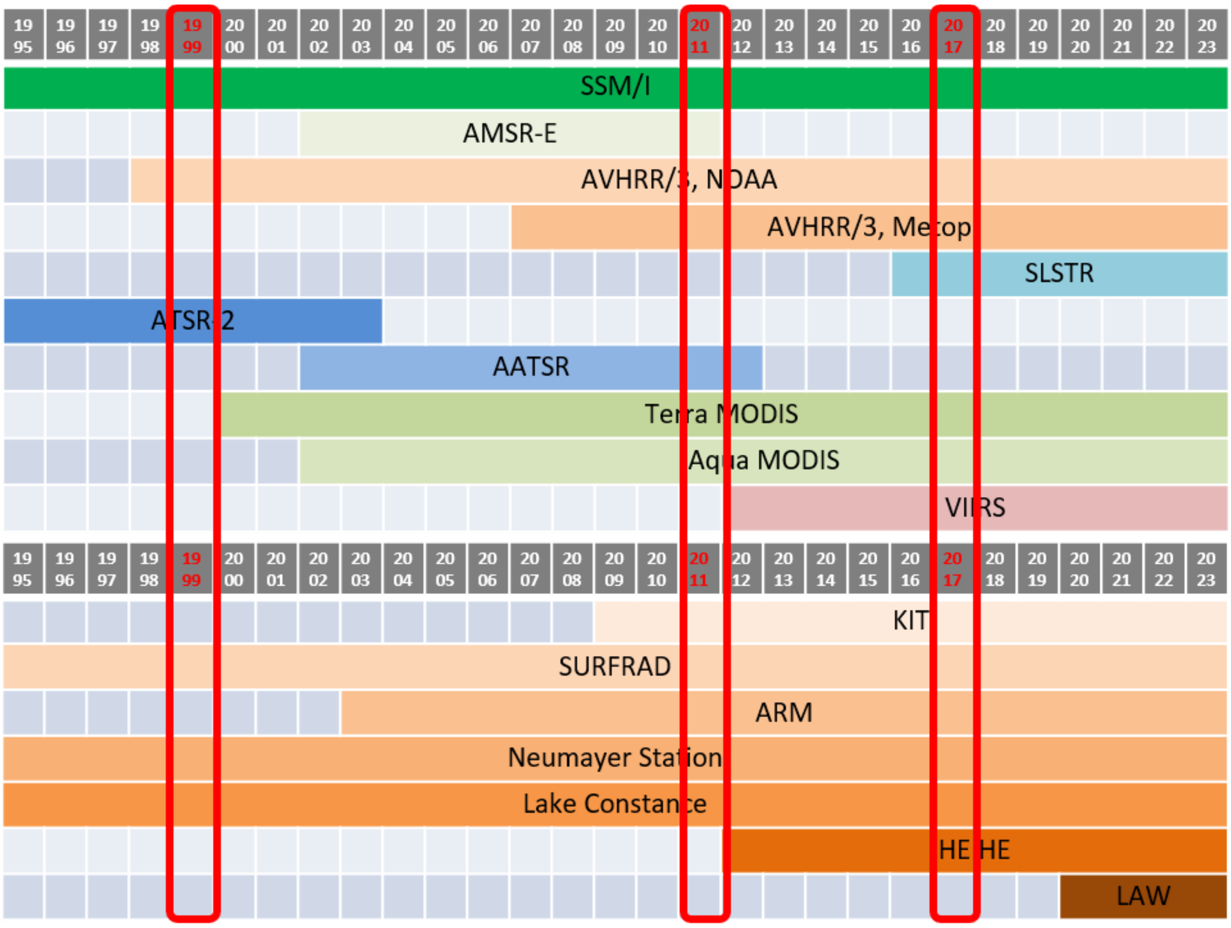

Figure 4: Temporal windows for the IR and MW LEO satellite LST data to be produced in LST_cci, as well as for the in situ station LST data. The red bars indicate the years, where the in situ data will be available for the data set producers.

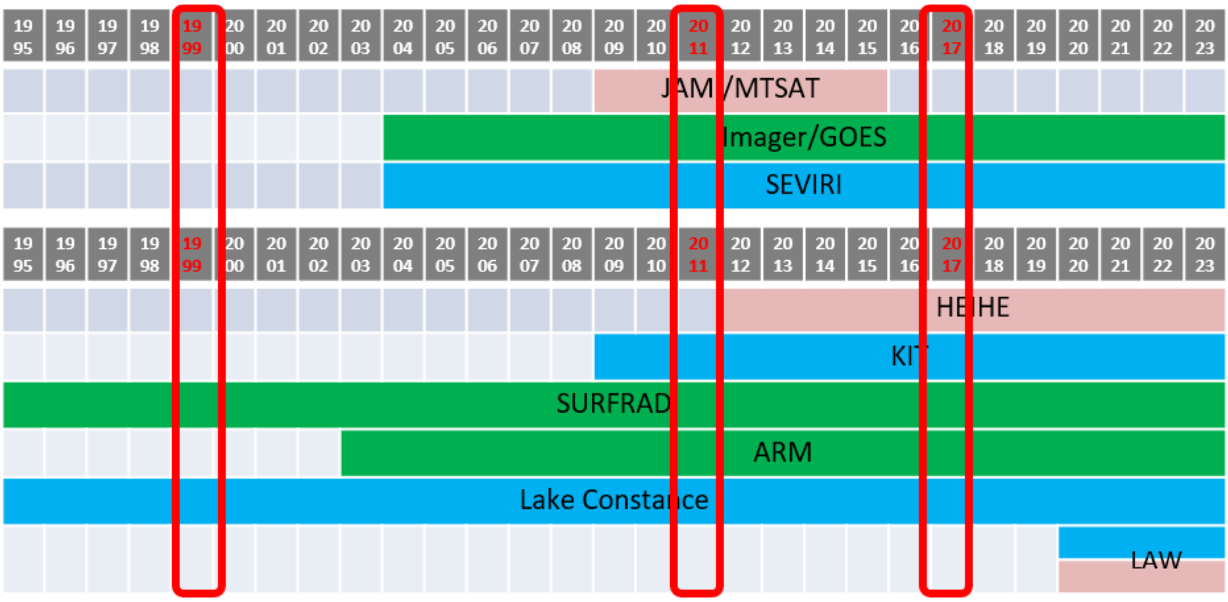

Figure 5: Temporal windows for the GEO satellite LST data to be produced in LST_cci, as well as for the in situ station LST data. The red bars indicate the years, where the in situ data will be available for the data set producers. Same colour indicates same spatial coverage of satellite and in situ data.

An overview of all used in situ stations is given in Table 13, also indicating the year that will be available for data set developers.

Table 13: Overview of the used in situ stations.

Code	Network	Name	Latitude	Longitude	Elevation	Surface type	Temporal availability
SGP___	ARM	Southern Great Plains Facility, Oklahoma	36.605° N	97.485° W	318 m	rural (mixture of grassland pasture / wheat fields / bare soil [RD-22])	2003 – now
NSA___	ARM	North Slope of Alaska, Barrow	71.320° N	156.600° W	8 m	mainly snow	2003 - now
GVN___	AWI	Georg-von-Neumayer station, Antarctica	70.67° S	8.270° W	40 m	Ice / Snow	1995 – now
EVO___	KIT	Evora, Portugal	38.540° N	8.003° W	230 m	Savannas, woody savanna; 32% tree, 68% grass	2010 – now

DAH_T_	KIT	Dahra tree mast, Senegal	15.402° N	15.433° W	90 m	Grassland; 96% grass, 4% tree	2010 – 2017
GBB_W_	KIT	Gobabeb wind tower, Namibia	-23.551° N	15.051° E	406 m	Bare ground; 75% gravel, 25% dry grass	2010 – now
KAL_R_	KIT	Rust mijn Ziel (RMZ) Farm, Kalahari, Namibia	23.011° S	18.353° E	1450 m	Shrub land; 85% grass / soil, 15% tree	2010 - 2011
KAL_H_	KIT	Farm Heimat, Kalahari, Namibia	22.933° S	17.992° E	1380 m	Shrub land; 37% tree / bush, 63% grass	2011 – 2018
CNS___	LUBW/KIT	Lake Constance	47.605° N	9.444° E	396 m	Water	1995 - now
BND___	SURFRAD	Bondville, Illinois	40.052° N	88.373° W	230 m	Grassland	1995 - now
TBL___	SURFRAD	Table Mountain, Boulder, Colorado	40.126° N	105.238° W	1689 m	Sparse grassland	1995 - now
DRA___	SURFRAD	Desert Rock, Nevada	36.623° N	116.020° W	1007 m	Arid shrub land	1998 - now
FPK___	SURFRAD	Fort Peck, Montana	48.308° N	105.102° W	634 m	Grassland	1995 - now
GCM___	SURFRAD	Goodwin Creek, Mississippi	34.255° N	89.873° W	98 m	Grassland	1995 - now
PSU___	SURFRAD	Penn. State Univ., Pennsylvania	40.720° N	77.931° W	376 m	Cropland	1998 - now
SXF___	SURFRAD	Sioux Falls, South Dakota	43.734° N	96.623° W	473 m	Grassland	2003 – now
BGB___	Heihe Integrated Observatory Network	Heihe River Basin, China	38.899° N	100.282° E	1562 m	Desert	2013 - 2015
DMN___	Heihe Integrated Observatory Network	Heihe River Basin, China	38.860° N	100.370° E	1556 m	Maize fields	2013 - now
HZZ___	Heihe Integrated Observatory Network	Heihe River Basin, China	38.746 ° N	100.290° E	1731 m	Desert	2013 - now
KIT_F_	Copernicus LAW	KIT Forest, Germany	49.091° N	8.425° E	104 m	Forest	2020 - now
HYY_F_	Copernicus LAW	Hyytiälä Forest, Finland	61.846° N	24.296° E	180 m	Forest	2021 - now

PUE_F_	Copernicus LAW	Puéchabon Forest, France	43.741° N	3.596° E	275 m	Forest	2021 - now
ROB_F_	Copernicus LAW	Robson Creek Forest, Australia	17.118° S	145.63° E	710 m	Forest	2021 - now
SVA_F_	Copernicus LAW	Svartberget Forest, Sweden	64.171° N	19.747° E	160 m	Forest	2021 - now

6.1. ARM

The U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) site Southern Great Plains (SGP___) Facility in Oklahoma (http://www.arm.gov/sites/sgp/C) will be used for LST validation. The station is equipped with Infrared Thermometers (IRT) Wintronics (Heitronics KT15) and it is located in a large area with cattle pasture and wheat fields. The upwelling IR radiance measurements are undertaken at SGP___'s central facility C1, where the instrument is located at a height of 10 m. The downwelling IR radiance measurements needed for correcting reflected sky radiance as well as 2m air temperature were taken at SGP___'s extended facility E13, which is located close to C1.

Another ARM site that can be used for validation purposes is the North Slope of Alaska (NSA___) site, which is covered by snow in winter. Thus, the NSA___ site can be used to validate Ice Surface Temperature (IST). At this station, upwelling and downwelling radiances are measured using Eppley Precision Infrared Radiometers. The location of the two ARM sites is displayed in Figure 6.

Figure 6: Locations of the two ARM sites (Southern Great Plains and North Slope of Alaska) climate research facilities, which will be used for LST validation in the project. Source: Adapted from www.flickr.com/photos/armgov/12483406053/in/album-72157625275280553/

All ARM data are freely available, providing long-term continuous atmospheric measurements. For more details see https://www.arm.gov/. The spectral characteristics of the radiometer at the ARM SGP___ site are identical to the KT15.85 instruments used at the KIT sites. Thus, the calculation of the LST and its uncertainty was performed in analogy to the KIT sites (see Section 6.4). However, one difference is that the uncertainty of the measured brightness temperatures was set to ±0.5 K following [RD-23]. Furthermore, the monthly Combined Aster and Modis Emissivity for Land (CAMEL) emissivity at the hinge point of 10.6 µm was taken [RD-46] as channel effective emissivity, as this hinge point falls into the center wavelength of the KT19.85 radiometer. The uncertainty is composed of the uncertainty in the CAMEL emissivities and the maximum difference between the emissivity at the hinge point at 10.6 µm and those at the two neighbouring hinge points. The ARM NSA___ site uses the same pyrgeometers as the SURFRAD sites; thus the same method for calculating LST and its uncertainty was used (see Section 6.5). The only difference is that the uncertainty of the IR radiance data was set to ± 4 Wm^{-2} following [RD-26].

6.2. Georg-von-Neumayer Station

A long time validation station is the Georg-von-Neumayer–Station (GVN___) on Antarctica, which is operated by the Alfred-Wegener-Institute (Bremerhaven, Germany). IST is one of the most important components of the surface-atmosphere energy balance of ice/snow covered surfaces. IST data over sea-ice will be processed within the work of LST_cci and thus this station can be used to validate the satellite IST data. Measurements of radiation are carried out on a large scale as part of a global observation network at the Neumayer station to detect long-term changes in the Earth's radiation budget and their impacts on climate. Since 1998 derived surface temperatures at 1 min interval are available. These are obtained from broadband hemispherical longwave fluxes (PIR and CGR4 pyrgeometers from Eppley and Kipp & Zonen), which are part of Neumayer's Baseline Surface Radiation Network (BSRN) station. For calculating LST at the Neumayer station, emissivity was set to 1.0, which is a good approximation for the snow surface at the station [RD-27]. Thus, IST can be obtained from measured upward broadband radiances alone using the Stefan-Boltzmann law. The uncertainty of the measured radiances was set to ± 5 Wm^{-2} [RD-28]. Additionally, a value of 0.01 was assumed as the uncertainty in emissivity. However, this uncertainty only has a minor effect on total LST uncertainty. The location of the station in Antarctica is displayed in Figure 7.

Figure 7: Antarctica with the location of the Neumayer Station III. Source: www.awi.de.

6.3. Stations at Heihe River Site

The Heihe Watershed Allied Telemetry Experimental Research (HiWATER) program set up several stations to monitor hydrological and ecological processes in the region of the Heihe river basin (China) in a variety of landscapes, including deserts, forest, and cropland [RD-29, RD-30]. The Heihe stations are equipped with broadband sensors [RD-34, RD-35, RD-36] and the most suitable ones for LST validation will be selected based on the published literature. These sites not only increase the coverage over additional land cover types but also provide a basis for validating GEO LST ECV Products derived from MTSAT and Himawari.

Three stations were chosen for validation within LST_cci, covering data since 2012 until now:

- BGB___: Bajitan Gobi, CNR1, Kipp&Zonen, 10 min (until April 2015), Reaumuria desert
- DMN___: Daman, PIR Eppley, unc 5 W/m^2, 10 min, maize
- HZZ___: Huazhaizi Desert Steppe, CNR 1, Kipp&Zonen, 10 min (after June 2015)/ 30 min (before June 2015), Kalidium foliatum desert

6.4. KIT stations

KIT station measurements are combined with detailed knowledge of the heterogeneity of the validation sites. The stations are located in different climate zones and represent stable land covers that experience seasonal variation only.

The KIT stations are located in in Africa and Europe (Figure 8) within the following climatic regions:

- Temperate Mediterranean climate, cork-oak trees and grass: Evora (EVO___), Portugal, since 2005:

At Evora, radiometers observe grass and the crown of an oak tree. A static tree crown cover of 32% (determined from high-resolution satellite imagery [RD-22]) is used for the current validation scheme of satellite LST with in situ data from Evora.

- Semi-arid climate, subtropical bush: Dahra (DAH_T_), Senegal, 2008 - 2017:

DAH_T station is located in the subtropics with a strong seasonal vegetation cycle. The rainy season is between June and November, where the atmospheric water vapour content is very high. This makes it difficult to evaluate the data then and can lead to high differences between satellite and in situ LST [RD-31].

- Arid Desert hot climate, gravel desert: Gobabeb (GBB_W_), Namibia, since 2007, since 2012 part of BSRN:

GBB_W station is located on the large and homogeneous gravel-plains of the Namib Desert at 405 m asl. in Namibia [RD-31]. Performing measurements along a 40 km track, [RD-32, RD-33] showed that the station LST is representative for an area of several 100 km^2.

- Arid desert hot climate at high elevation, Kalahari bush: Kalahari Farms (Rust Mijn Ziel; KAL_R_ and Heimat; KAL_H_, ~1450m asl), 2008 – 2019:

  The two farms are located in the Kalahari semi-desert in Namibia. The climate there is hot and arid, with two rainy seasons. The small rainy season is from September to November with little rain and the big one from January to March. Except for the time of the big rainy season, the grass in the Kalahari desiccates quickly and the region is dry [RD-31]. The station LST is measured at Farm Rust Mijn Ziel (KAL_R_) until February 2011, and from then on at Farm Heimat (KAL_H_).

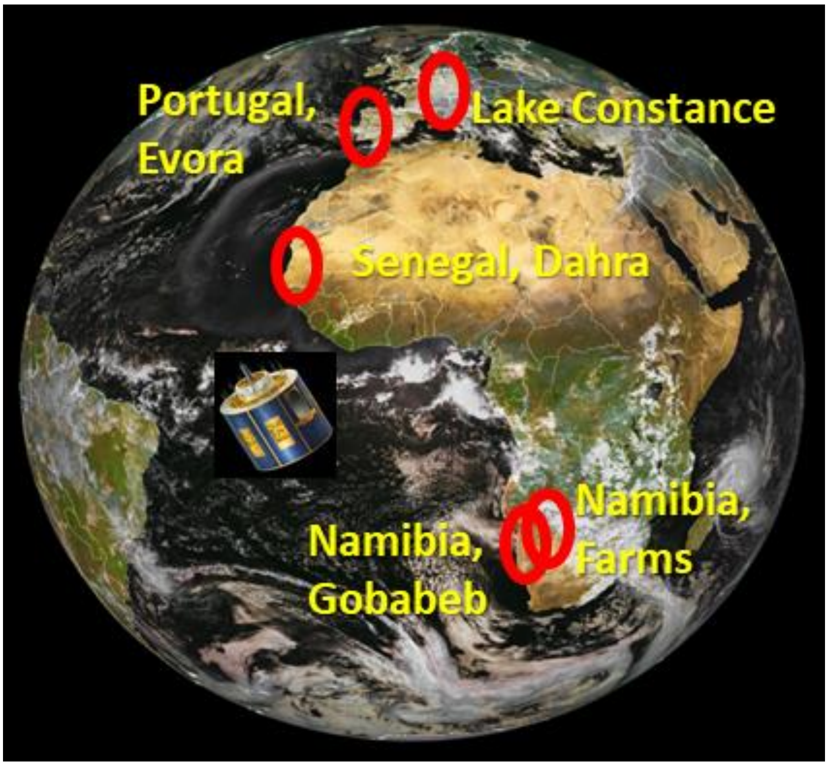

Figure 8: Locations of KIT`s validation stations.

The validation stations have a sampling rate of 1 minute, and measure directional TIR radiance from the ground and the sky. They provide standard meteorology, and two stations additionally measure upwelling and downwelling hemispherical shortwave and longwave broadband radiances.

The main instrument for the in situ determination of land surface temperature at KIT's validation stations is the precision radiometer 'KT15.85 IIP' produced by Heitronics GmbH, Wiesbaden, Germany. These instruments measure thermal infrared radiances between 9.6 µm and 11.5 µm with a temperature resolution of 0.03 K and an accuracy of ± 0.3 K over the relevant temperature range. The KT15.85 IIP has a drift of less than 0.01 % per month: the high stability is achieved by linking the radiance measurements via beam-chopping (a differential method) to internal reference temperature measurements.

The KT15.85 IIP radiometers are typically mounted at 25 m altitude and have a FOV of about 10 m². Due to the radiometer's narrow spectral response function in the atmospheric window and the small distance between the radiometer and the surface, the atmospheric attenuation of the surface-leaving TIR radiation is negligible. However, the measurements contain radiance emitted by the surface, i.e. the target signal, as well as reflected downwelling TIR radiance from the atmosphere. Therefore, an additional KT-15.85 IIP measures downwelling long wave IR radiance from the atmosphere at 53° zenith angle pointing towards an azimuth that is never reached by the sun. Measurements under that specific zenith angle are directly related to downwelling hemispherical radiance so that no ancillary data for deriving in situ LST are needed. The corrections for the reflected component in the measurements are performed at KT-15's centre wavelength of 10.55 µm.

For the hyper-arid and quasi-static validation site Gobabeb (GBB_W_) a constant emissivity of 0.940 ± 0.015 is used, which was determined from a combination of in situ measurements and laboratory emissivity spectra of soil samples [RD-37]. In contrast, the other KIT sites exhibit considerable emissivity variation due to changing soil moisture content and vegetation cover. Therefore, for these sites the operational LSE product provided for SEVIRI ch10.8 by LSA SAF is used to derive in situ LST, since ch10.8 is spectrally similar to the KT15.85 IIP radiometer [RD-37]. The overall product accuracy of LSA SAF ch10.8 LSE is estimated as 0.025 when compared to the current MODIS MOD11C3 LSE [RD-38].

The uncertainty of the in situ LST obtained from the KT15 radiometer measurements is estimated via uncertainty propagation, considering random uncertainties of ±0.3 K of the measured land surface and sky brightness temperatures (BT_L and BT_S). For the Gobabeb site the uncertainty in emissivity was estimated as ± 0.015 [RD-37]. For the other stations, emissivity uncertainty is calculated with equation 3 given in [RD-39] and using the constants given in [RD-40]. Furthermore, a systematic uncertainty is introduced by the protective window in front of the upward looking KT15 radiometer, which was found to degrade over time by up to 5% from its initial value.

Considering the above uncertainties for the year 2010 at GBB_W_ station and combining them via uncertainty propagation leads to a median random uncertainty of $0.80 \pm 0.12\ K$ and a median systematic uncertainty of $-0.08 \pm 0.01\ K$ [RD-31]. The main contribution to total uncertainty stems from uncertainty in emissivity, followed by the uncertainty in BT_L. In contrast, the uncertainties in BT_S and the protective window have smaller contributions, which was expected since for the relatively high surface emissivity and the corresponding small surface reflectivity BT_S only makes a minor contribution to surface leaving radiance. The median of ΔLST_{total} for 2010 is estimated as 0.80 ± 0.12 K, which is identical to random uncertainty and shows that the contribution due to systematic uncertainty is negligible [RD-31].

The IR radiance measurements from KIT stations have been successfully used to validate several satellite LST products within the GlobTemperature Project [RD-12] and also in several other studies [RD-31, RD-37, RD-41, RD-42, RD-43, RD-44].

6.5. SURFRAD stations

The SURFRAD network was established in 1993 through the support of NOAA's Office of Global Programs and has been operational since 1995 [RD-53, RD-54]. Its primary objective is to support climate research with accurate, continuous, long-term measurements pertaining to the surface radiation budget over the United States [RD-55]. The main advantage of the SURFRAD sites is the long-term availability of data and its high quality. SURFRAD stations have already been successfully utilised to validate satellite LST [RD-53, RD-55, RD-56, RD-57]. For the location of the SURFRAD stations, see Figure 9.

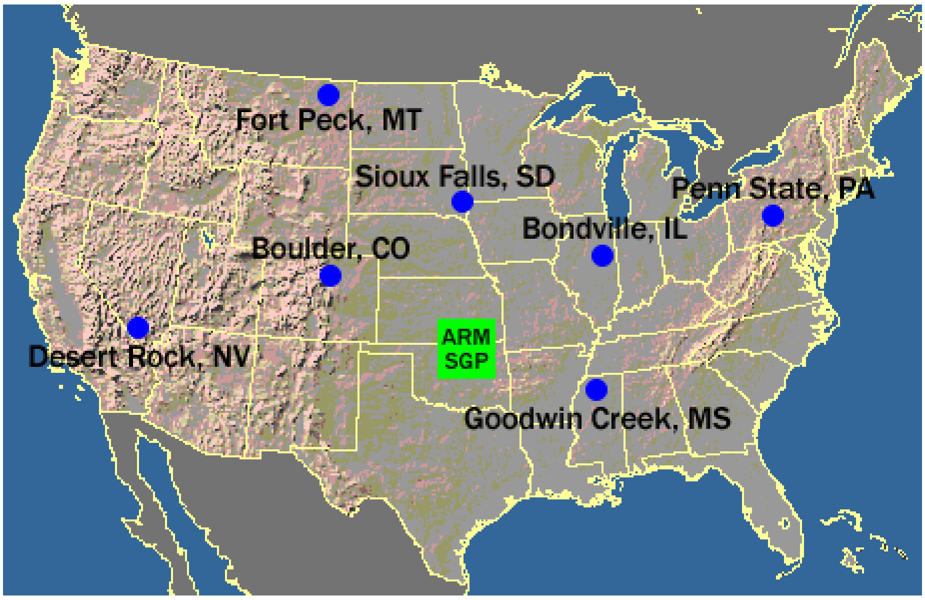

Figure 9: Locations of SURFRAD stations (Source: www.esrl.noaa.gov/gmd/grad/surfrad/sitepage.html).

All SURFRAD stations measure downwelling and upwelling broadband solar and thermal infrared irradiances by facing upward and downward Eppley Precision Infrared Radiometers. The instruments measure hemispherical IR radiances in a wavelength range from 4 - 50 µm and have a spatial representativeness of around 70 m × 70 m [RD-57]. The pyrgeometers on the SURFRAD stations are exchanged annually and are run in parallel to a reference instrument some days prior to that for further quality control [RD-54]. They are calibrated using three standards maintained at NOAA's Field Test and Calibration Facility at Table Mountain near Boulder, CO, which are traceable to world standards.

These sites have been extensively used within the GlobTemperature Project [RD-12] to validate LST data sets. The results indicate that care needs to be taken at some of these stations due to issues with the seasonal heterogeneity of the land surface or elevation around the station. Therefore, matchups from these will be carefully analysed to ensure only matchups where representativeness can be assured are used. The considered SURFRAD stations are the following:

- BND___ (Bondville) station: The station is located in an agricultural area in Illinois, where mainly corn is grown. The LCC of the station pixel is "Mosaic Cropland / Vegetation". Publicly available Google Earth data (earth.google.com) show that the station is located in a ~200 x 200 m^2 area of grass surrounded by agricultural fields. Analysis of monthly AATSR LST accuracies over BND___ station (Figure 10) [RD-13] shows strong positive values for the daytime data with peaks in May and October. These maxima could stem from harvesting: in May, the fields get prepared, before the crop starts to grow, AATSR observes mainly bare soil and its LST is considerably higher than the LST of the green grass observed by the station. The same applies in October, when the crop has been harvested and the fields are brown. From this, it was concluded that the area observed by the station radiometer is unrepresentative of the area covered by the corresponding AATSR satellite pixel: therefore, BND___ daytime data was disregarded completely for LST validation within GlobTemperature. Night-time data were only considered between November and May, when the influence of the seasonal cycle is minimal. An average of 5 x 5 pixels around the station pixel was taken.

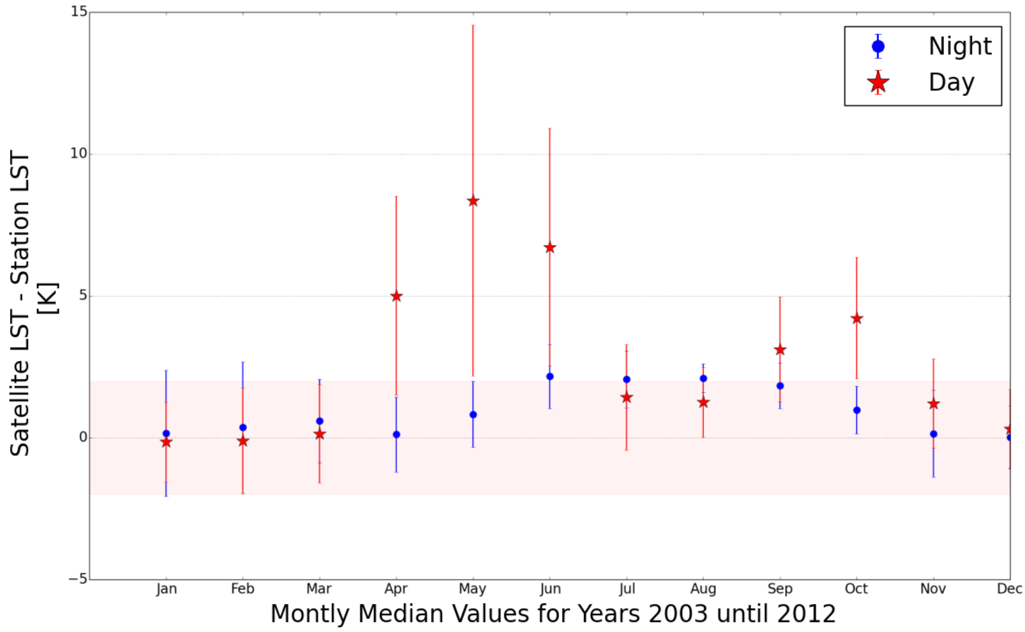

Figure 10: Monthly median for AATSR LST - in situ LST at BND___ station [RD-13]

- DRA___ (Desert Rock) station: Desert Rock station is located in Nevada. The LCC of the station pixel is "Closed to open shrubland", and 20 x 20 pixels around the station fall into this LCC class, i.e. the area is very homogenous in terms of land cover. GoogleEarth imagery shows that the station is located in a valley (approx. altitude 1000 m asl.) and the elevation around it is about 300 m higher. Analysis of AATSR accuracies analysis was performed by [RD-13] for different numbers of satellite pixels around the station: all results showed a large STD during daytime. To further investigate this, the monthly mean accuracies for the pixels in the 5 x 5 area around the station were obtained, which gave an obvious gradient with higher accuracies in the NE than in the SW (Figure 11). These differences were observed throughout the year: it was negative in summer and positive (but not as pronounced) in winter. This is probably caused by the local orography (i.e. depending on sun angles and shadows). The daytime AATSR overflight is in the morning, when the sun illuminates the scene from the East. To avoid this influence, daytime validation is performed using only an average of 3 x 3 pixels around the station pixel, as the results were in close agreement with the 1 x 1 pixel analysis.

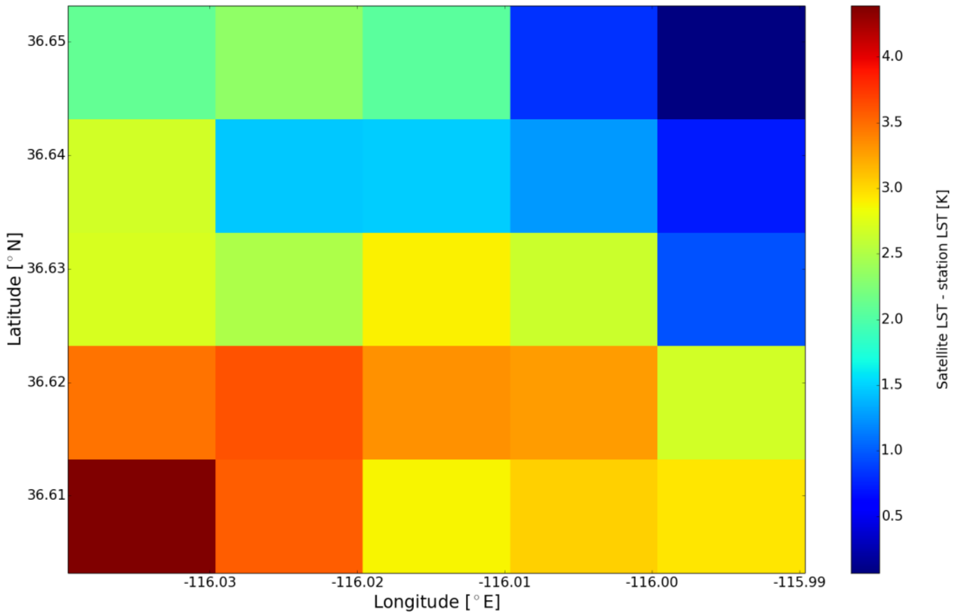

Figure 11: Median daytime accuracies of AATSR LST–in situ LST for June 2010 for the individual pixels in a 5 × 5 pixel area around DRA___ station, covering an area of 0.01° x 0.01° each. The station is located on the pixel in the centre [RD-13]

- FPK___ (Fort Peck) station: the station is located in the Fort Peck Tribes Reservation and the LCC of the station pixel is “Mosaic Cropland / Vegetation”. The area around the station it is a mixture of forest, shrublands and grassland. A pronounced seasonal cycle of monthly mean daytime AATSR LST accuracy was found for FPK___ station by [RD-13]. It is suspected that this cycle is caused by the small fence (approx. 20 m x 30 m) protecting the station: the station is surrounded by grassland where bison herds graze, which could lead to a different colour/length of the grass inside and outside the fence. Therefore, daytime data are not considered in the analysis. During night-time, the 5 x 5 pixels around the station pixel are averaged, which is justified by the fact that at night-time AATSR LST are usually close to isothermal; this is also supported by the small seasonal cycle found during night.
- GCM___ (Goodwin Creek) station: the station is located in rural pasture land in the state of Mississippi. While the station radiometer observes an area covered by grass, the LCC of the satellite pixel centred on the station is “Closed broadleaved deciduous forest”. The area around the station consists of a heterogeneous mixture of grassland and woods. Averaging the satellite AATSR LST over different numbers of pixels around the station by [RD-13] did not significantly alter the accuracy and STD; thus, the average of 5 x 5 pixels around the centre pixel is taken for validation.
- PSU___ (Pennsylvania State University) station: it is located in a broad Appalachian valley and the LCC of the station pixel is “Mosaic Cropland / Vegetation”. The downward looking sensor of PSU___ station observes a mixture of (short) grass and crops. The GoogleEarth image of the area (Figure 12) shows that the surface around the station is very heterogeneous and includes

fields, forests and small villages. It was not possible to find a larger area representing the same LCC as the station pixel [RD-13]. Thus, only the centre pixel directly above the station is taken for validation.

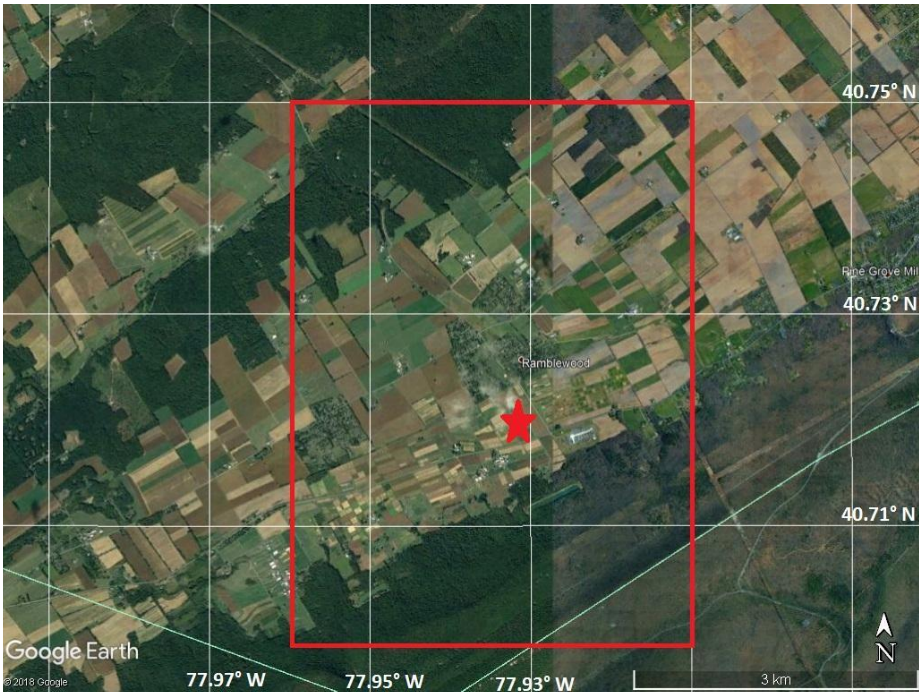

Figure 12: Google Earth true-colour satellite image of the area around PSU___ station (indicated by the red star). The red rectangle indicates the approx. position of the pixel used for validation [RD-13].

- TBL___ (Table Mountain) station: the station is located on the plateau of Table Mountain, north of Boulder. The LCC of the station pixel is “Mosaic Vegetation / Cropland”. However, the station pixel is near the edge of the mountain (Figure 13) and all pixels on the plateau have the combined LCC class “broadleaved/needleleaved deciduous forest”. Since the station is located on the mountain plateau and, thus, measures an area covered by the “broadleaved/needleleaved deciduous forest” class, a pixel with the same class directly north-west of the station is used for validation [RD-13]. Only this pixel is chosen for the validation in order to exclude mixed pixels near the “edges” of the mountain.

Figure 13: Google Earth true-colour satellite image of TBL___ station (indicated by the red star). The red rectangle indicates the nominal position of the pixel used for validation [RD-13].

For validation of LST at the SURFRAD stations, broadband emissivities were obtained from the CAMEL global broadband infrared product [RD-45]. The CAMEL product [RD-46] provides emissivity values with a spatial resolution of 0.05° at 13 spectral 'hingepoints' between 3.6 and 14.3 µm by combining data from the MODIS baseline fit emissivity database (i.e., CIMMS database) [RD-25] with the ASTER Global Emissivity database version 4 (ASTER GEDv4) [RD-58]. Using a high spectral resolution (HSR) emissivity algorithm, the 13 hingepoints form the basis for reproducing emissivity in 417 spectral channels covering the same spectral range. Finally, BBE values are estimated by numerically integrating the HSR emissivities between 8 µm and 13.5 µm. As monthly CAMEL BBEs are only available between 2000 and 2016, in situ LSTs before 2000 and after 2016 were obtained using monthly mean CAMEL emissivities. The mean emissivities were obtained for each station separately by averaging the BBEs over the years 2001 – 2010 for the data acquired before 2000 and over 2007 – 2016 for data acquired after 2016. The determined BBE together with the measured upwelling and downwelling broadband IR radiances can then used to obtain LST via Stefan-Boltzmann's law.

Uncertainties of the LST values were established from the statistically independent BBE uncertainty and the measurement uncertainties of the radiances using error propagation. [RD-45] determined the BBE uncertainty between 0.004 and 0.01 during the BBE retrieval at five specific sites. Thus, a BBE uncertainty of ±0.01 is considered for SURFRAD stations. However, the BBE uncertainty has only a minor influence on the total uncertainty. For the field pyrgeometers, both for upwelling and downwelling measurements, the uncertainty is estimated to be ±5 Wm^{-2} [RD-59]. The resulting overall LST uncertainty lies in a range between 0.6 – 2 K for all stations, which translates to a relative LST uncertainty of less than 1%.

6.6. Lake validation

The Validation Team will explore the use of LWST data to validate long-time series of satellite LST data. Various researchers successfully used in situ LWST to validate LST, e.g. [RD-47, RD-48, RD-49]. The Validation Team has acquired long-term water temperature data from Lake Constance, where since 1962 a German Regional Office continuously measures water temperatures near the middle of the lake surface. These water temperature measurements can be intercompared with the radiometric in situ LWST acquired by KIT and then be used to evaluate historic satellite LST products. Since 2014, KIT obtains LWST from two Heitronic KT15 IIP radiometers set up on the car ferry Friedrichshafen operated by the Bodensee-Schiffsbetriebe (BSB) GmbH, Germany, which crosses Lake Constance regularly between Friedrichshafen (Germany) and Romanshorn (Switzerland) up to 16 times per day [RD-36, RD-50, RD-51] (Figure 14). These radiometers take continuous measurements of LWST and sky measurements at the complementary angle during each ferry crossing. A recent field inter-comparison campaign [RD-52] between the KT15 IIP radiometer and the Infrared SST Autonomous Radiometer (ISAR), considered as the 'gold standard FRM' for water bodies, showed an agreement between both instruments better than ±0.2 K.

Figure 14: Measurements obtained from radiometry installed on the Lake Constance car ferry.

6.7. Copernicus LAW stations

Five new LST stations were deployed in the "Copernicus Space Component Validation for Land Surface Temperature, Aerosol Optical Depth and Water Vapor Sentinel-3 Products (LAW)" project to validate the SLSTR LST product at five different under-represented forest biomes (Figure 15). The LAW stations are located in different forest biomes representing the following LST_cci land cover classes:

- KIT forest (Germany). Instruments are set up on a 200 m high meteorological tower on the premises of KIT Campus Nord. The representative pixel is classified as "closed broadleaved decidious forest". This station was deployed in August, 2020.

- Hyytiälä (Finland). The instruments are set up at the CEOS Land Validation supersite SMEAR II. The representative pixel is classified as "closed to open mixed broadleaved and needle leaved forest". Data from Hyytiälä station are available since October, 2021.
- Svartberget (Sweden). The radiometers were deployed in October 2021 in the CEOS Land Validation Supersite at Svartberget Experimental Forest. The site is classified as "open needleleaved decidious or evergreen forest".
- Robson Creek (Australia). The instruments were deployed in October 2021 on a tower at Robson Creek in a rainforest known as the Wet Tropics Bioregion of Australia. This site is classified as "closed to open broadleaved evergreen" and/or "semi-deciduous forest".
- Puéchabon (France). The instruments were set up on a tower at the CEOS LPV validation site in Puéchabon in October 2021. Although the station is located in a dense holme oak forest (https://puechabon.cefe.cnrs.fr), according to the ALB2 land cover classification the 3x3 km^2 around the site is known to be covered by a mix of surface types (forest, shrub and grassland), while the 1 km pixel that includes the station is classified as "sparse vegetation" land cover.

The five new LST stations set up for the LAW project are equipped with two thermal infrared (TIR) Heitronics KT15.85 IIP radiometers each following the same configuration as the KIT stations. In addition, Hyytiälä and Svartberget sites are usually covered by snow during winter season. Thus, these sites can also be considered to validate non-permanent snow/ice surface temperature.

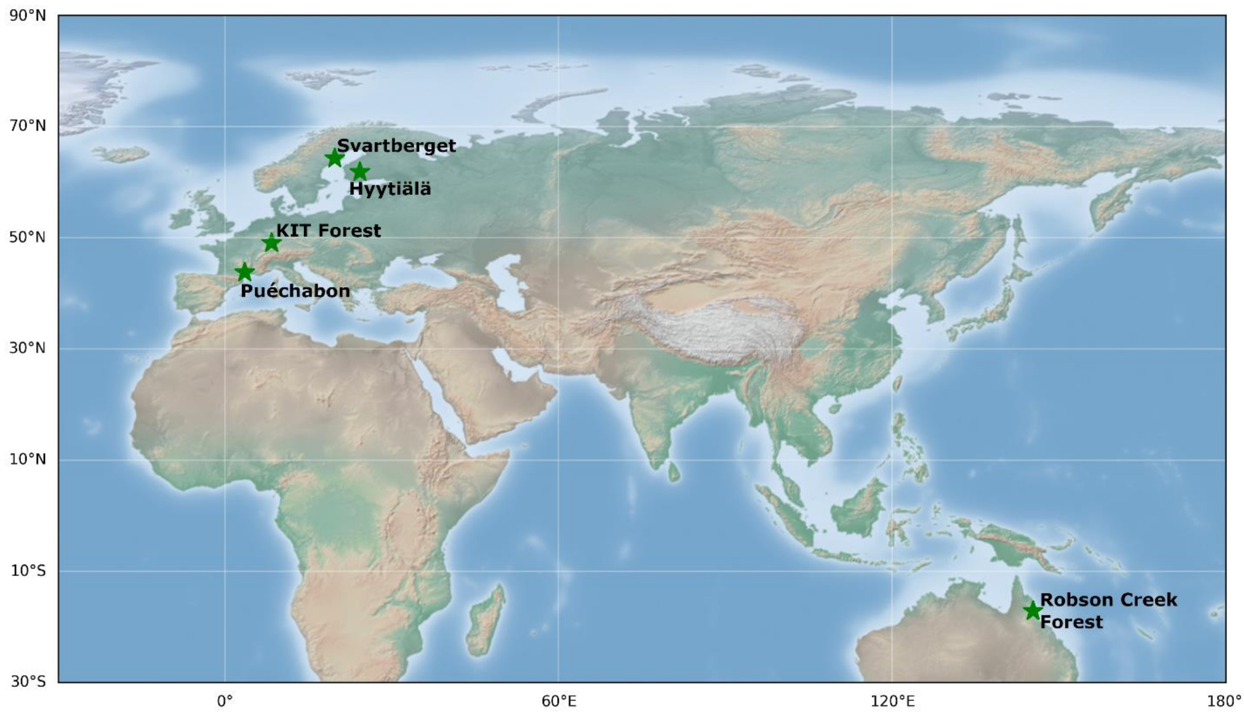

Figure 15: New LST validation stations set up within the ESA Copernicus LAW Project.

7. CCI Satellite Data Sets

All introduced satellite data sets will be produced within LST_cci.

7.1. LEO data sets

The LEO ECV Products will be processed on UK JASMIN utilising the existing infrastructure from pre-cursor projects as much as possible. LEO data sets produced in LST_cci are introduced in the following sections.

7.1.1. Along Track Scanning Radiometer -2 (ATSR-2) and Advanced Along Track Scanning Radiometer (AATSR)

ATSR-2 is the second of a series of instruments (ATSR-1, ATSR-2 and AATSR). It was on board the European Space Agency's (ESA) sun-synchronous, polar orbiting satellite ERS-2 which was launched in April 1995. The AATSR is the third in the series. It was on board ESA's sun-synchronous, polar orbiting satellite Envisat which was launched in March 2002 and stopped operating in April 2012. Both had similar orbit and equator crossing times, which ensures a high level of consistency. Together all three instruments cover approximately 20 years of data. As only satellite data from ATSR-2 and AATSR will be used in the project, the considered data length is 17 years.

The ATSR-2 and AATSR ECV Products will be created using the existing GlobTemperature processing chains, with any necessary modifications for these data on UK JASMIN.

The choice of input auxiliary datasets is dependent on the selection of the final algorithm for implementation in the ECV Production System. These may include land cover data, fractional vegetation cover, water vapour, emissivity, and snow cover masks. Since the LEO IR processing systems created in pre-cursor projects are modular in design different algorithms can be switched in and out.

The output datasets are currently, and are proposed to be, processed onto the permanent Leicester workspace on JASMIN which has all the necessary auxiliary data for fast and cost-effective processing. As an example, the data flows for the LST ECV prototype system for AATSR are shown in Figure 16.

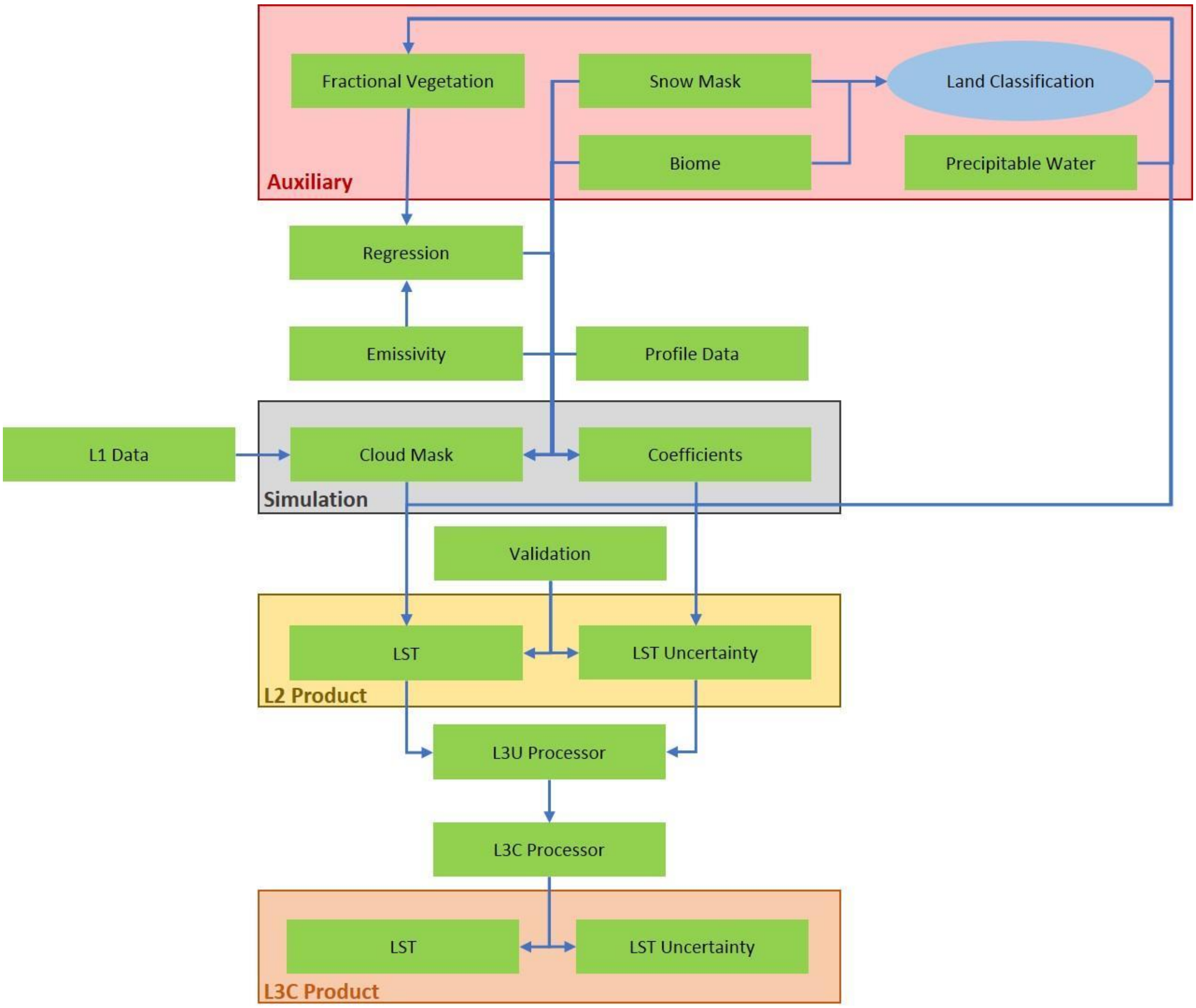

Figure 16: Data flows for the AATSR LST ECV prototype production system. (Figure from AD-2).

7.1.2. Advanced Very High Resolution Radiometer (AVHRR)

AVHRR/3 is an across track scanner that senses the Earth's outgoing radiation from horizon to horizon: with a scan range of ±55.37° and a swath width of about ±1447 km. It provides global observations in the visible and infrared bands. These measurements are available twice a day (higher frequency may be achieved over the Polar Regions), at 1.1 km resolution at the sub-satellite point. Although AVHRR/3 is a six-channel radiometer, only five channels are transmitted to the ground at any given time, since channels 3a and 3b do not operate simultaneously. For Metop satellites, channel 3a is operated during the daytime portion of the orbit and channel 3b during the night-time portion. The LSA SAF adopted a procedure for LST retrieval from AVHRR split-window channels similar to that for SEVIRI data, but taking into account AVHRR response functions. Emissivity estimates are based on the so-called Vegetation Cover Method and atmospheric correction takes into account hourly forecasts obtained from the operational global ECMWF model. The AVHRR/Metop LST is first generated on a pixel-by-pixel basis. A daily composite is then generated from the former, where LST is re-projected onto a regular (0.01°) grid to form two regular fields (65°S –65°N) with local daytime and night-time passages, respectively.

7.1.3. Moderate Resolution Imaging Spectrometer (MODIS)

The LST_cci MODIS products employ a generalized split-window (GSW) approach [RD-60], similar to the method used for AVHRR data, to estimate LST as a linear function of clear-sky top-of-atmosphere (TOA) brightness temperatures from bands 31 and 32 centred on 11 µm and 12 µm, respectively. It provides data on LST and its associated full pixel-level uncertainty breakdown. It further provides auxiliary information that has been used for the LST retrieval, such as emissivity and quality control flags. Retrieval coefficients are categorised into classes of satellite viewing angle and water vapour.

Land surface emissivity (LSE) for MODIS LST is estimated from the CAMEL database of monthly land surface emissivity [RD-46]. This is available at ten wavelengths between 3.6 μm and 14.3 μm, - including emissivities at 10.8 µm and 12.1 µm – at a spatial resolution of 0.05°. The data itself are spatially and temporally interpolated onto the ~1 km grid for the given day of the satellite acquisition.

7.1.4. Sea and Land Surface Temperature Radiometer (SLSTR)

The ATSR series continues with 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, which responds to the requirements for an operational and near-real-time monitoring of the Earth surface over a period of 15 to 20 years.

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 being 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. 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 split-window algorithm with classes of coefficients for each biome-diurnal (day/night) combination.

The SLSTR ECV Products for Sentinel-3A and 3B will be created using the existing GlobTemperature processing chain for SLSTR.

7.1.5. Visible Infrared Imaging Radiometer Suite (VIIRS)

The visible Infrared Imaging Radiometer Suite (VIIRS) is on board Suomi National Polar-orbiting Partnership (S-NPP) and on first Joint Polar Satellite System (JPSS-1) platforms, which were launched in October 2011 and November 2017, respectively. Both are sun-synchronous polar orbit satellites. VIIRS has 22 channels between 0.4 and 12.5 µm, with three thermal infrared channels centred at 8.55, 10.8 and 12.0 µm. The spatial resolution of VIIRS is of 750 m for the thermal bands, and a swath of 3000 km, providing a full coverage of Earth observation twice per day.

The LST is estimated in LST_cci using a GSW algorithm [RD-52], following the same procedure than for MODIS data. The LST is estimated as a linear function of TOA brightness temperatures from bands centred at 11 and 12 µm, which coefficients are dependent on the satellite viewing angle and water vapour. The LSE for VIIRS LST is obtained from CAMEL database.

7.1.6. Special Sensor Microwave / Imager (SSM/I)

SSM/I sensors have been on board the Defense Meteorological Satellite Program (DMSP) polar satellites since 1987. They observe the Earth twice daily at 19.35, 22.235, 37.0, and 85.5 GHz with both vertical and horizontal polarizations, with the exception of 22 GHz, which is for vertical polarization only. The observing incident angle is close to 53°, and the elliptical fields of view decrease in size proportionally with frequency, from 43 x 69 km^2 to 13 x 15 km^2. The local times of their descending and ascending modes are early morning and late afternoon, respectively. There are up to 4 SSM/I instruments in space at the same time with similar overpassing times.

A methodology has been developed to estimate the LST, along with atmospheric water vapour, cloud liquid water, and surface emissivities over land, from passive microwave imagers [RD-61]. It is based on a neural network inversion, trained on a large data set of simulated radiances, using real atmospheric and surface information over the globe. The method has been applied to the Special Sensor Microwave/Imager (SSM/I) measurements, and LST have been estimated with a spatial resolution of 0.25° x 0.25°, at least twice daily, depending on the number of SSM/I instruments in space.

Theoretical uncertainties have been estimated in [RD-61]. The uncertainty varies mostly with cloud liquid water content and with microwave surface emissivity.

7.1.7. Advanced Microwave Scanning Radiometer for EOS (AMSR-E)

The Advanced Microwave Scanning Radiometer for EOS (AMSR-E) was launched in May 2002 on board EOS-Aqua satellite and operated until December 2011. The EOS-Aqua satellite has a sun-synchronous polar orbit and crosses the equator at 1:30 pm/am. The sensor measured the Earth-emitted radiation in twelve microwave channels: 6.9, 10.6, 18.7, 23.8, 36.5 and 89.0 GHz with both vertical and horizontal polarizations. The observing incident angle was 55° and the swath 1445 km, while spatial resolution increased with frequency from 75 x 43 km^2 (6.925 GHz) to 6 x 4 km^2 (89.0 GHz).

The LST is estimated with the neural network inversion methodology proposed for passive microwave imagers in [RD-61] with a spatial resolution of 0.25° x 0.25°. The theoretical uncertainties are estimated according to [RD-61]. The cloud liquid water content and the surface emissivity are the main contributors to the LST uncertainty.

7.2. GEO Data Sets

The generation of LST ECV Products derived from geostationary satellites is performed at IPMA through modification of one of the existing processing chains. The GEO data sets produced in LST_cci are GOES, MTSAT, Himawari and SEVIRI data sets. They are introduced below.

There are two separate processing chains dependent on whether the satellite sensors support single-channel or split-window retrievals. In the case of SEVIRI/Meteosat the Generalised Split-Window approach is used (Figure 17).

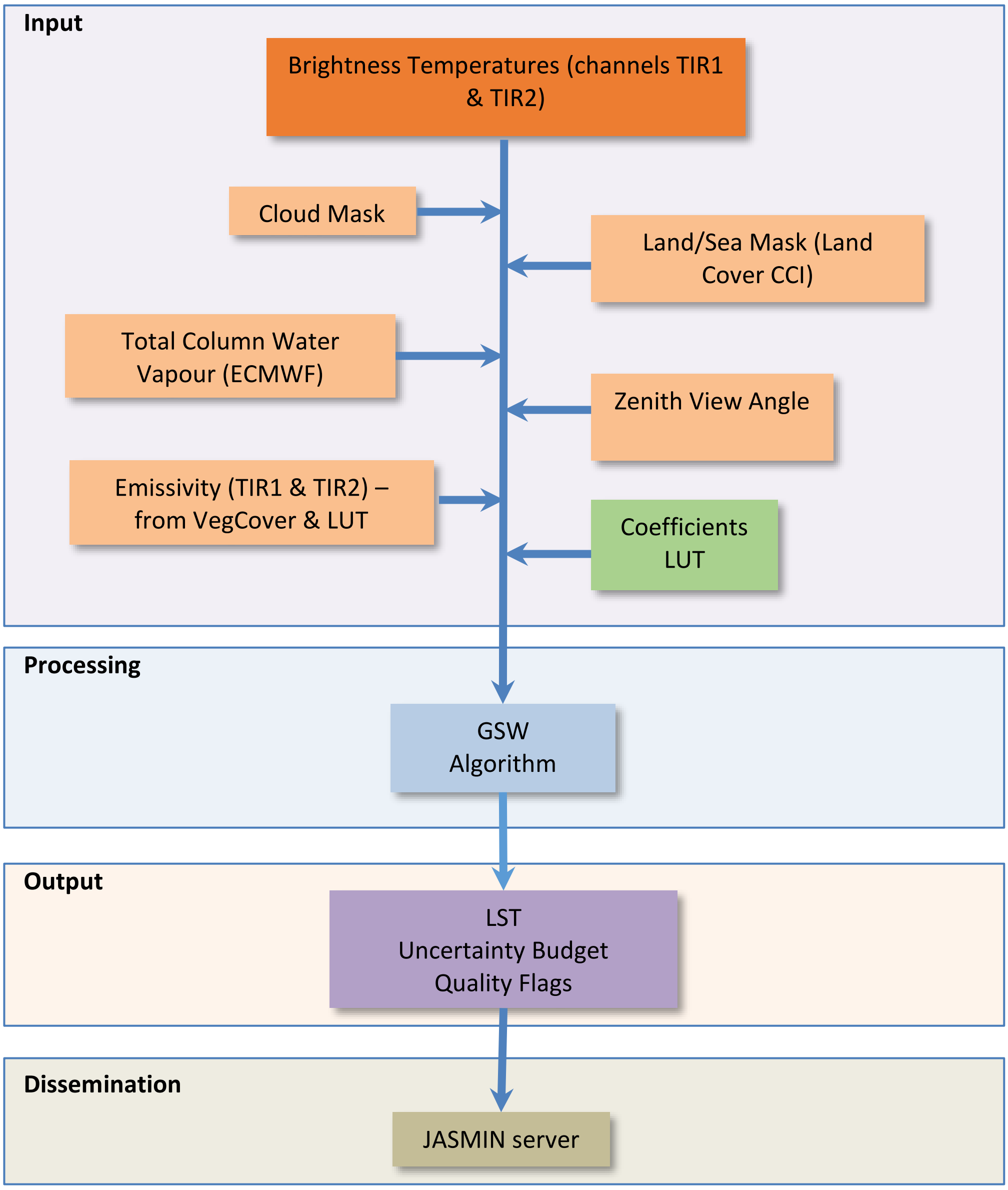

Figure 17: Flowchart for LST production from sensors with split-window such as SEVIRI/Meteosat.

For sensors/data with a single channel in the thermal infrared window and one middle IR channel (e.g., GOES 12-16, MTSAT-2 data archived at IPMA up to 2012) a “Dual Algorithm” variant on the Generalised Split-Window approach is applied (Figure 18).

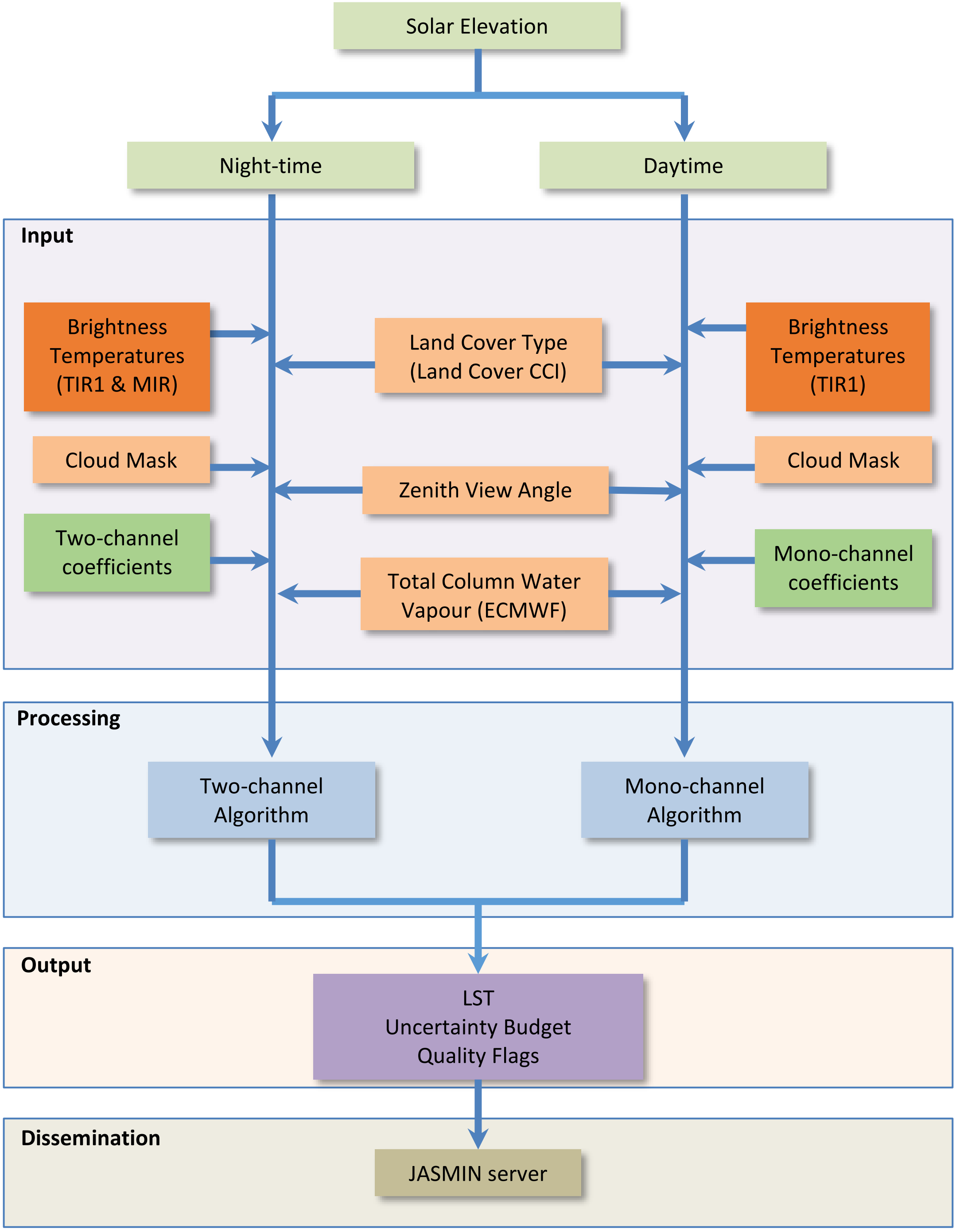

Figure 18: Flowchart for LST production from sensors with a single-channel such as GOES 12-16/MTSAT-2 up to 2012.

7.2.1. Geostationary Operational Environmental Satellite (GOES)

The Copernicus Global Land Service generates GOES hourly LST data. These data are available in near real time and off-line, covering the GOES disk centred at 75°W and with a spatial resolution of about 4 km at the sub-satellite point. The LST algorithm used for GOES accounts for the fact that the most recent imager on this platform does not have the two split-window channels.

The applied methodology, named Dual-Algorithm (DA) [RD-62], implies two LST algorithms, which are used for daytime and night-time, respectively: a two-channel algorithm is applied to night-time observations, making use of one thermal infrared – around 11 µm – and one middle infrared – around 3.9 µm; and a mono-channel algorithm is applied for daytime cases, using the available thermal infrared channel for atmospheric attenuation and surface emissivity. The middle-infrared is discarded for daytime cases to avoid the correction of solar radiation reflected by the surface. The methodologies mentioned above are all based on semi-empirical formulations, where LST is expressed as a regression function of TOA brightness temperatures.

7.2.2. Multi-functional Transport Satellite (MTSAT) and Himawari (HMWR)

The Copernicus Global Land Operational Service (CGLOPS) generates Multi-functional Transport Satellite (MTSAT) –replaced by Himawari-8 (HMWR) in 2015– hourly LST data. These data are available in near real time and off-line, covering the MTSAT disk centred at 140°E (MTSAT-1) and 145°E (MTSAT-2/HMWR), with a spatial resolution of about 4 km at the sub-satellite point for the MTSAT series and 3 km for HMWR.

The LST algorithm used for MTSAT is not the split-windows approach as in the case of HMWR, because although having both, one of the split-window channels (12 µm) was not available for Near Real Time processing (EUMETCast relay) until the beginning of 2013, when CGLOPS started receiving both channels. Therefore, until the end of 2012 for MTSAT and HMWR the LST retrieval is carried out by the Dual-Algorithm (DA) methodology [RD-62] as described above in Section 7.2.1.

7.2.3. Spinning Enhanced Visible and Infrared Imager (SEVIRI)

SEVIRI is the main sensor onboard Meteosat Second Generation (MSG), a series of 4 geostationary satellites to be operated by EUMETSAT. SEVIRI was designed to observe the Earth disk with view zenith angles (SZA) ranging from 0º to 80º at a temporal sampling rate of 15 minutes. SEVIRI's spectral characteristics and accuracy, with 12 channels covering the visible to the infrared, are unique among sensors onboard geostationary platforms. The first MSG satellite was launched in August 2002, and operational observations are available since January 2004. The High Resolution Visible (HRV) channel provides measurements with a 1 km sampling distance at sub-satellite point (SSP); for the remaining channels the spatial resolution is 3 km at SSP. The nominal SSP is located at 0° longitude and therefore the MSG disk covers Africa, most of Europe and part of South America.

Land Surface Analysis Satellite Application Facility (LSA SAF) produces LST based on SEVIRI. It is obtained by correcting TOA radiances for surface emissivity, atmospheric attenuation along the path and reflection of downward radiation. The LST algorithm follows closely the generalized split window proposed by [RD-60] for AVHRR and MODIS, but adapted to SEVIRI response functions [RD-44, RD-63].

7.3. Climate Data Records

Two climate data records (CDR) will be developed within LST_cci. The first is based on data from ATSR, MODIS, and SLSTR, the second will be a merged IR product (Section 7.3.2), including also AVHRR, GOES, MTSAT and SEVIRI data in addition to the data sets from the first CDR. For climate applications, the availability of LST CDRs covering a long period of time will be very valuable. The length of the first CDR will be 25 years, the length of the second will be 12 years.

7.3.1. ATSR – MODIS – SLSTR

The finest temporal resolution of the final product will be daily (day / night). The diurnal dimension determined from the solar zenith angle. The spatial resolution is to be 0.05°. This spatial-temporal combination is intended to meet the corresponding GCOS requirements [RD-3].

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

There are several steps to generating a CDR from ATSR-2 through to SLSTR. These are schematically represented in Figure 19. At each step the associated uncertainty for the step is to be characterised and propagated through to the next step according to the model detailed in E3UB.

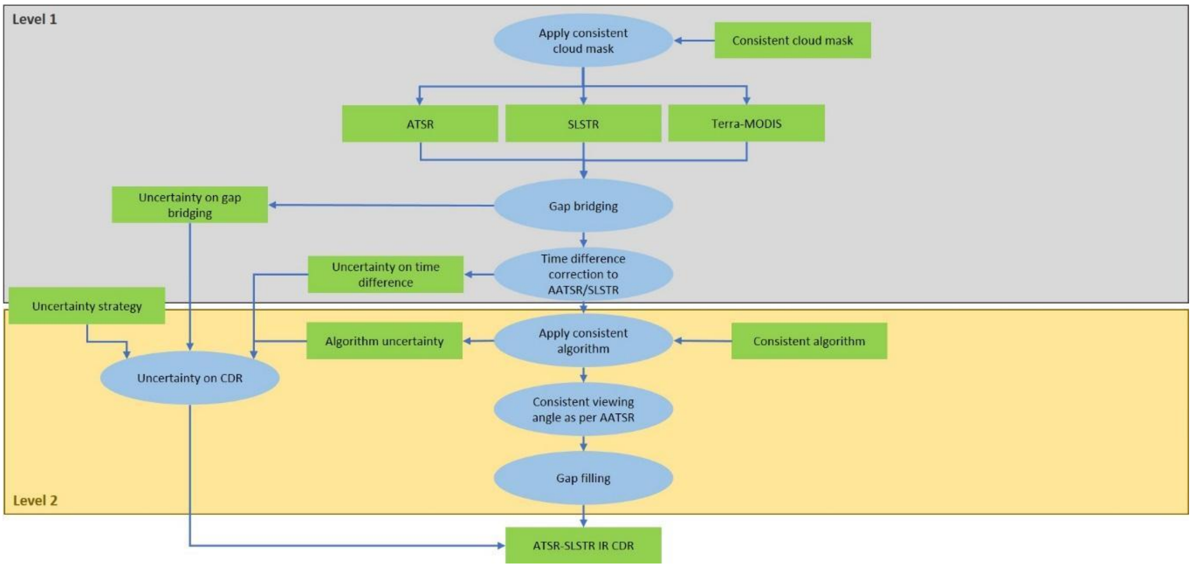

Figure 19: Schematic representation of the development and implementation of the ATSR-SLSTR CDR.

7.3.2. Merged IR

The Merged IR CDR will be processed across the UK JASMIN – IPMA distributed systems. The benefit here is to utilise as much as possible the existing infrastructure from pre-cursor projects and that being utilised in the single-sensor ECV production. The processing chain utilises the following LEO and GEO data sets produced by LST_cci: ATSR, Metop-AVHRR, SLSTR, MODIS, SEVIRI, GOES and MTSAT.

For the Merged IR products, the algorithm implementation will take the form of LST ECV data at specified UTC time slots throughout the day consistent with the nominal observation times of the geostationary satellites and the availability of the polar-orbiters. A schematic representation of the development and implementation of the Merged IR CDR is given in Figure 20.

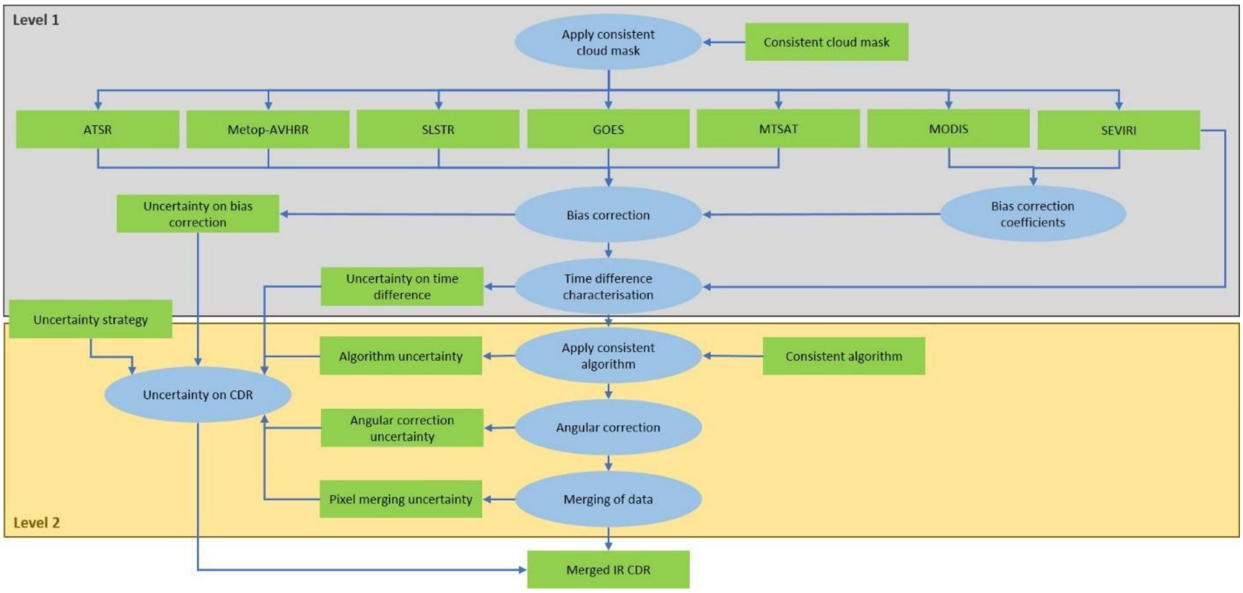

Figure 20: Schematic representation of the development and implementation of the Merged IR CDR.

The work builds on the advances made for LST product merging in GlobTemperature [RD-64, RD-66]. The significant step forward proposed here includes bias correction of the Level-1 data, consistency in algorithm and cloud clearing, advancements in correction for anisotropy, and full characterisation of the uncertainty budget. The Merged IR CDR will be produced on a global scale eight times daily (00h, 03h, 06h, 09h, 12h, 15h, 18h, 21h, UTC times). The nominal spatial resolution will be a 0.05° equal angle latitude-longitude grid.

8. External Satellite Data Sets

For the satellite – satellite intercomparisons within LST_cci, several external satellite data sets will be used. They are introduced in the following.

8.1. Operational MODIS (MOD11/MYD11)

The standard operational MODIS LST products MOD11_L2 (Terra) and MYD11_L2 (Aqua) uses the generalized split-window algorithm to retrieve LST of clear-sky pixels from BTs in bands 31 and 32. This implementation of the algorithm operates with classification-based emissivities. The retrieval coefficients are determined by interpolation on a set of multi-dimensional look-up tables; these are obtained by linear regression of the simulation data generated by radiative transfer over a broad wide range of surface and atmospheric conditions. The look-up tables incorporate several improvements for the algorithm such as view-angle dependence, water vapour dependence, and atmospheric lower boundary temperature dependence.

8.2. Operational ATSR-2, AATSR, and SLSTR

All ATSRs had a dual-angle capability (nadir and forward at an angle of ~55° to nadir), but only the nadir view is used for the standard ESA LST retrieval. This is a split-window algorithm [RD-65], where LST is retrieved with 1 km resolution based on the infrared channels at 11 and 12 µm [RD-66]. Continuation of the ATSR data record is sustained with the launch of SLSTR on board Sentinel-3, which also has a 1 km measurement grid using two infrared channels for LST retrieval and a split-window approach.

8.3. Operational SEVIRI

The LSA SAF operational LST algorithm applied to SEVIRI data takes the GSW formulation. The coefficients are calibrated for classes of viewing angle and total column water vapour (available from ECMWF as operational hourly forecasts), and calculated using a regressive analysis technique, whereby simulated T_{11} ($T_{10.8}$ to be precise in the case of SEVIRI) and $T_{12.0}$ are obtained from radiative transfer simulations performed over a wide variety of atmospheric profiles. When compared with an independent dataset of radiative transfer simulations the algorithm and its coefficients were found to be bias free, although random errors tended to increase with increasing water vapour content and at high view-angles [RD-39].

The channel surface emissivity is calculated from an average of bare-ground and fully vegetated emissivities weighted by the fractional vegetation cover (FVC) and by the fraction of water within the pixel. The emissivity extremes are selected from a look-up-table in accordance to the land cover classification. The FVC data used is another product produced by the LSA SAF from SEVIRI/Meteosat and corresponds to a 5-day composite updated daily [RD-67].

The uncertainty of LST values is also estimated and distributed. This depends on a number of factors [RD-40]: (i) the retrieval conditions in terms of atmospheric water content and viewing geometry, since the total optical path will strongly influence the performance of the generalized split-window algorithm; (ii) input uncertainties, including sensor noise, errors in the estimation of surface emissivity and in atmospheric total water vapour obtained from ECMWF forecasts. The final error estimation is given by the square root of the sum of the squares of each individual error, including the GSW model error. For most regions within the MSG disk these values are below 2 K and pixels with error bars higher than 4 K, generally near the edge of the MSG disk, are masked out.

Despite the strong impact of non-identified clouds on LST retrievals, this source of error is not currently being accounted for in the uncertainty estimation of SEVIRI LST. The error introduced by misclassification of cloudy pixels as clear-sky may be dealt with by flagging the pixels adjacent to clouds as contaminated.

8.4. CM SAF SEVIRI TCDR

The CM SAF have produced a first version of a Thematic Climate Data Record (TCDR) for LST derived from the Meteosat Visible and InfraRed Imager (MVIRI) on board the Meteosat First Generation (MFG) and SEVIRI [RD-68]. The time period of the TCDR ranges from January 1991 to December 2015, and therefore includes MFG 4 to 7 and MSG 1 to 3. Original thermal radiances have been intercalibrated by EUMETSAT using the High Resolution Infrared Radiation Sounder (HIRS) as a reference. The LST is derived from Meteosat by use of single-channel LST retrieval algorithms suitable to generate LST data across the Meteosat satellite generations. This TCDR includes two different LST products: the Statistical Land Surface Temperature product generated with a statistical LST retrieval approach and a Physical Land Surface Temperature product based on radiative transfer calculations. The LST is presented as hourly data and as monthly averaged diurnal cycle composites on a 0.05° x 0.05° grid covering the entire Meteosat disk [RD-69].

8.5. ISCCP LST

Within the International Satellite Cloud Climatology Project (ISCCP), IR measurements from several polar and geostationary satellites are combined to derive a global LST product from 1983 onwards. A single-channel algorithm is used to produce this multisensor data set [RD-70]. All surfaces are treated by ISCCP as blackbodies with unit emissivity [RD-70]. ISCCP data have a spatial sampling of 30 km, and a temporal resolution of 3 h [RD-71].

8.6. ERA5 surface skin temperature

ERA5 will be the fifth generation of ECMWF atmospheric reanalysis of the global climate expected to replace ERA-Interim early in the project lifetime of LST_cci.

Each independent surface functional unit, called tile, has an own surface energy balance and skin temperature. The skin temperature is calculated by linearizing the surface energy balance equation for each tile. The equation is solved implicitly together with the vertical turbulent transport in the boundary layer. In the approach used, the skin temperatures of the different tiles are dependent on each other, and therefore the equations cannot be solved independently [RD-72].

9. Summary

This document details the plan for the validation work within LST_cci. Validation is an essential part of developing LST data sets that are employed by users, as it is the means of giving information on the quality of the data sets. It thus increases the trust of data set users in the data sets provided.

All validation carried out within LST_cci is based on several, well-acknowledged validation protocols in the LST community, which give guidelines for a consistent, meaningful validation. Following these protocols will ensure that all validations carried out within LST_cci are trustworthy and comparable with other validation work. It is of particular interest that the requirements formulated by the Global Climate Observing System (GCOS) regarding accuracy, precision and stability for LST data sets are met by the validated LST data sets. Furthermore, requirements from the LST user community that were assessed in a questionnaire will be considered [AD-1].

The LST_cci validation work will be mainly divided into two activities: a validation against in situ stations, and intercomparisons against operational satellite LST data sets or against models. In situ validation is the most straightforward type of validation [RD-1], where satellite data sets are directly compared to in situ LST measurements from ground-based instrumentation. For this, it is used a set of globally distributed stations which covers different time periods, climate zones and land covers. These stations are introduced in the document. Ideally, each satellite data set is validated over several stations with different land covers as this can influence the performance of the data set. It is also best to cover several yearly cycles to assess the quality of the data set for different seasons.

Satellite – satellite intercomparisons are a good supplement to in situ validation, as the in situ station network does not globally cover all regions and land covers. These intercomparisons can give valuable insights in the relative algorithm performance of LST data sets over regions where in situ stations are scarce.

All data needed for the validation within LST_cci are stored in the Multi-sensor Matchup Database (MMDB). The MMDB also encompasses all software for validation. All data files are in a common harmonized data format, which allows consistent validation for different data sets. These data formats are described in this document. The in situ data in the MMDB are divided in two parts, the smaller part covering one year per station is accessible to the EO Science Team to test their algorithms. The larger part of in situ data files, however, is only accessibly by the Validation Team to ensure an independent validation. Results of all validation work will be detailed and discussed in the Product Validation and Intercomparison Report and will also be fed back to the producers.

The in situ station network used for the validation work in LST_cci is introduced in this document, as well as the satellite data sets that will be developed and the satellite data sets from external data providers used for satellite – satellite intercomparisons. These data sets form a good database for carrying out meaningful validations, which will help to improve the satellite data sets produced within LST_cci.

End of document