CCI Land Surface Temperature

LST CCI System Requirements Document

WP3.1 – DEL-D3.1

Ref.: LST-CCI-D3.1
Date: 13-Mar-2019
Organisation: Consortium CCI LST

Signatures

	Name	Organisation	Signature
Written by	Jérôme Bruniquel	ACRI-ST	
	Olivier Sardou	ACRI-ST	
	Hille Nabédrik	ACRI-ST	
Reviewed by			
Approved by	Darren Ghent	ULeic	
Authorized by	Stephen Plummer	ESA	

Change log

Version	Date	Changes
1.0	17-Dec-2018	First version
1.1	13-Mar-2019	Updated version to include ESA RIDs

List of Changes

Version	Section	Changes
1.1		

Table of Content

List of Figures

List of Tables

1. Executive Summary

This document summarises the processing system requirements for the Land Surface Temperature (LST) Climate Change Initiative (CCI) project aiming at producing climate data records.

LST, which is one of the Essential Climate Variables (ECV), will be used by climate user communities for long-term analysis, and the LST production system shall allow a generation of consistent datasets over a long time period (more than 20 years) and from different data sources. It shall also allow the reprocessing of these datasets when necessary and deliver LST climate data records (CDR).

The set of LST CCI requirements has been derived from the following documents :

- The LST CCI ITT, more specifically the statement of work and its annex G (specific to LST);
- The LST CCI URD document (ref. LST-CCI-D1.1), prepared in the frame of the LST CCI project by UKMO;
- The Data Standards Requirements for CCI Data Producers document.

These 3 documents define various requirements, but not all are linked to the processing system or only in an indirect way. The analysis of these 3 documents has allowed to identify all requirements which are relevant to the processing system.

A naming convention has been decided (applicable also in other LST CCI documents) to identify all requirements, on the model **LST-SYS-REQ-NN-DOC** where:

- NN is a double digit number
- DOC is the original document
    - SOW
    - URD
    - CCI

The next table lists all system requirements which will be the basis for deriving the system specifications of the LST CCI processing system.

Table 1: List of requirements

Id	Title	Description
LST-SYS-REQ-01-SOW	Production of time-series	The LST CCI processing system shall generate all LST ECV products as requested in the SOW
LST-SYS-REQ-02-SOW	Building of processing system	The LST CCI processing system shall be built and documented in order to fulfil the needs for processing and reprocessing
LST-SYS-REQ-03-SOW	Legacy of CCI projects	The LST CCI processing system shall make use, as much as possible, of the best practises and existing developments of CCI projects
LST-SYS-REQ-04-SOW	System adaptation to the production of LST CCI products	The LST CCI processing system shall be scalable in order to include new datasets and new processing chains

Id	Title	Description
LST-SYS-REQ-05-SOW	Integration of Copernicus and key satellite missions	The LST CCI processing system shall be able to include Copernicus Sentinel missions and other key satellite missions as input data
LST-SYS-REQ-06-SOW	System sizing to account for large volume of data	The LST CCI processing system shall be able to process or reprocess large amount of data, with expecting growing volumes.
LST-SYS-REQ-07-SOW	Specifications of products to process	The LST CCI processing system shall process all data listed in the SOW
LST-SYS-REQ-08-SOW	Specifications of IR based LST products	The LST CCI processing system shall process all IR data listed in the SOW
LST-SYS-REQ-09-SOW	Specifications on NOAA AVHRR products	The LST CCI processing system shall process AVHRR products listed in the SOW
LST-SYS-REQ-10-SOW	Specifications on global merged products	The LST CCI processing system shall process global merged products as specified in the SOW, in particular over the period 1998-2020
LST-SYS-REQ-11-SOW	Specifications on passive MW based LST products	The LST CCI processing system shall process all MW data listed in the SOW
LST-SYS-REQ-12-SOW	Temporal Coverage of MW based products	The LST CCI processing system shall generate MW based products over the period 1998-2020
LST-SYS-REQ-13-SOW	DOI assigned to LST products	The LST CCI processing system shall assign DOI to LST CCI products
LST-SYS-REQ-14-SOW	Heritage from past and current projects	The LST CCI processing system shall build on past and current projects, in particular on GlobTemperature project heritage
LST-SYS-REQ-15-SOW	Sizing of the processing systems to account for large amount of data	The LST CCI processing system shall be able to process large amount of data
LST-SYS-REQ-16-SOW	Availability of LST CCI products	The LST CCI processing system shall make available the LST CCI products, in particular on the CCI Data Portal
LST-SYS-REQ-17-SOW	Acquisition of all EO and in-situ data	The LST CCI processing system shall ensure that all inputs data which are need for the processing are effectively available and accessible
LST-SYS-REQ-18-URD	CCI standard format	The LST CCI processing system shall comply with CCI standard format described in the CCI product guidelines
LST-SYS-REQ-19-URD	Access to data	The LST CCI processing system shall ensure an easy access to the LST CCI products
LST-SYS-REQ-20-CCI	CCI data standards requirements	The LST CCI processing system shall comply with CCI standard format described in the CCI product guidelines

Id	Title	Description
LST-SYS-REQ-21-CCI	INSPIRE compliant metadata records	The LST CCI processing system shall comply with INSPIRE directive for the metadata associated to each product
LST-SYS-REQ-22-CCI	Terms from CCI vocabulary tables used in netCDF	The LST CCI processing system shall use the same terms as those included in the CCI vocabulary tables in netCDF products
LST-SYS-REQ-23-CCI	Identification of key variables	The LST CCI processing system shall ensure that the key primary variables in the file and their related ancillary variables are identified, and the range of their expected values is indicated.
LST-SYS-REQ-24-CCI	Variables shall include at least time, longitude and latitude	The LST CCI processing system shall include, at least, these 3 variables in product: time, longitude, latitude
LST-SYS-REQ-25-CCI	Use of common directory structure	The LST CCI processing system shall comply with the structure proposed in the CCI data portal for the data
LST-SYS-REQ-26-CCI	Application of CCI filenaming convention	The LST CCI processing system shall comply with the filename conventions defined by CCI

2. Purpose and scope

This Land Surface Temperature (LST) System Requirements Document defines a set of requirements for the processing system of LST data which will be developed and deployed in the frame of the Climate Change Initiative (CCI) from the European Space Agency.

LST, which is now one of the Essential Climate Variables (ECV), will be used by climate user communities for long-term analysis, and the LST production system shall allow a generation of consistent datasets over a long time period (more than 20 years) and from different data sources. It shall also allow the reprocessing of these datasets when necessary and deliver LST climate data records (CDR).

Initial requirements were expressed in the statement of work of the ITT [AD-1] and some of these were clearly linked to the processing system. The prepared proposal, in answer to this ITT, took these requirements into account and a preliminary technical solution has been drafted.

Concerning the system, it is worth mentioning that the ITT requested (see R-17) "*each CCI project shall ensure that the system is adequately dimensioned to accommodate the growing volumes of input and output data, and the increasing computational loads needed to process, re-process, quality control, validate, and disseminate multi-decadal, global, ECV data products, of the required climate quality, in a timely manner.*" while "*CCI does not build operational processing systems. CCI develops software systems for pre-operational ECV production in a research context. Within CCI the system development work comprises products specification, algorithm development and improvement, uncertainty characterisation, prototype product generation, system definition, sizing and demonstration. The aim is that the resulting specifications will subsequently be handed over to be further developed into operational systems in a non-ESA context.*"

The proposed LST processing system will thus not be fully operational, but all pre-operational requirements will be fulfilled in order to meet the cardinal requirement [CR-4] *("Generate and fully document a production system capable of processing and reprocessing the data in CR-2[1], with the aim of supporting transfer to operational activities outside CCI (such as C3S)*".

The specificities of the LST CCI project which have an impact on system design are multiple and can be summarized as follow:

- The LST CCI project relies on the strong heritage built in the frame of the GlobTemperature project during which a large number of satellite data were collected and stored in different locations: mainly on the JASMIN workspaces managed by University of Leicester and at IPMA for infrared (IR) sensors, and at ESTELLUS for microwave (MW) sensors. All LST products generated during this project were also stored at ACRI-ST who hosted the data portal;
- Due to these multiple sources of data storage, and to avoid useless time to transfer all data sources in a unique location, the technical solution proposed is built around distributed processing centres, each in charge of running dedicated processing chain(s); this allows the consortium to fully benefit from the work done during GlobTemperature and to avoid restarting from scratch;
- There is a great deal of independence between IR-derived products and MW-derived products and, at this stage, it was too premature to have the processing of these two types of datasets completed in the same location. It is nonetheless the purpose of this document to propose how

[1] [CR-2] Produce, validate and deliver consistent time series of multi-sensor global satellite ECV data products for climate science.

we go about fool-proofing our system to minimise the risks associated with a distributed system – this is particularly important for preparing for Phase 2.

This document thus defines all requirements which are related to the processing system that we shall set up in order to appropriately generate the LST ECVs for the climate community. The Statement of Work (SOW) defines the expected content of the SRD document:

Create a System Requirements Document (SRD) specifying the requirements of a Processing System capable of generating ECV data products as specified in the applicable technical annex (A-I). The SRD shall include verifiable requirements on the following: Data processing function of each step of its processing chain, including data volumes; Platform specification; Compliance to all processing needs defined by the Task 3 inputs.

Fully capitalise on existing European assets through their reuse, particularly Open Source scientific tools and prototype ECV processing systems from prior projects. All investment in tools and technology shall be optimal and justified.

The document is structured as follow:

- Section 3 provides the concept and a few assumptions
- Section 4 defines all the requirements to which the processing system shall adhere
- Section 5 details the operations concept, the production workflow and the processing chains which will be developed and installed in the processing system(s)
- Section 6 gives all system requirements and defines the verification means.

Though this initial version of the SRD is considered to be stable, it is foreseen to update this document as necessary to fully capture all needs and requirements expressed by the users.

2.1. References

The references used in this document are listed in document “References and Acronyms” (ref. LST-CCI-D6.5-REFACR).

In addition, we have

- LST CCI User Requirement Document, ref. LST-CCI-D1.1-URD (still in review)

2.2. Acronyms

The list of references and acronyms is given in document "References and Acronyms" whose reference is LST-CCI-D6.5-REFACR.

Acronym	Full name
AATSR	Advanced Along-Track Scanning Radiometer
ADF	Auxiliary DataFile
AMSR	Advanced Microwave Scanning Radiometer
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
AMSU	Advanced Microwave Sounding Unit
ARC	ATSR Reprocessing for Climate
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
ATSR	Along-Track Scanning Radiometer
ATSR-2	Along-Track Scanning Radiometer - 2
ATSR-SLSTR CDR	ATSR-SLSTR CDR Product
AUX	Auxiliary datafile
AVHRR	Advanced Very High Resolution Radiometer
BT	Brightness Temperatures
C3S	Copernicus Climate Change Service
CCI	Climate Change Initiative
CCI+	Climate Change Initiative +
CDR	Climate Data Record
CEDA	Centre for Environmental Data Analysis
CF	Climate and Forecast metadata
CPU	Central Processing Unit
CR	Cardinal Requirements
CRDP	Climate Research Data Package
CRG	Climate Research Group
DMSP	Defense Meteorological Satellite Program
DOI	Digital Object Identifier
DUE	Data User Element
E3UB	End-to-End ECV Uncertainty Budget
ECV	Essential Climate Variable
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA-Interim	ECMWF Reanalysis - Interim

Acronym	Full name
FOV	Field Of View
FRAC	Full Resolution Area Coverage
FRM	Fiducial Reference Measurements
GAC	Global Area Coverage
GCOS	Global Climate Observing System
GEO	Geostationary Earth Orbit
GIEMS	Global Inundation Extension from Multi-Satellites
GlobTemperature	Global Temperature
GOES	Geostationary Operational Environmental Satellite
GSICS	Global Space-based Inter-Calibration System
IASI	Infrared Atmospheric Sounding Interferometer
IPMA	Instituto Português do Mar e da Amosfera
IR	InfraRed
L1	Level-1 data
L2	Level-2 data
L3	Level-3 data
L3C	Level-3 collated data
L3U	Level-3 uncollated data
LECT	Local Equator Crossing Times
LEO	Low Earth Orbit
LSA SAF	EUMETSAT Land Surface Analysis Satellite Applications Facility
LST	Land Surface Temperature
LST CCI	Land Surface Temperature Climate Change Initiative
MMDB	Multi-sensor Matchup Database
MODIS	MODerate-resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
MTSAT	Multi-Functional Transport Satellite
MVIRI	Meteosat Visible and InfraRed Imager
MW	Microwave
NetCDF	Network Common Data Form
NN	Neural Network
NOAA	National Oceanic and Atmospheric Administration
NRT	Neal Real Time
Obs4MIPs	Observations for Model Intercomparisons Project
RFS	Research File Store

Acronym	Full name
RTTOV	Radiative Transfer for TOVS
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SEN4LST	Sentinel-3 for synergistic use with Sentinel-2
SLSTR	Sea and Land Surface Temperature Radiometer
SNO	Simultaneous Nadir Overpasses
SOW	Statement Of Work
SRD	System Requirements Document
SSD	System Specification Document
SSM/I	Special Sensor Microwave Imager
SSMIS	Special Sensor Microwave Imager/Sounder
SST	Sea Surface Temperature
SST CCI	Sea Surface Temperature Climate Change Initiative
UKMO	UK Met Office
ULeic	University of Leicester
UL_MDB	University of Leicester Matchup Database
UOL_V3	University of Leicester Version 3 cloud mask
URD	User Requirements Document
UTC	Coordinated Universal Time
VNIR	Visible and Near-InfraRed
WCRP	World Climate Research Programme
WP	Work Package
WPD	Work Package Description

3. LST concepts and assumptions

3.1. LST system context

The system requirements are defined with regard to several elements and constraints:

- Production workflow, which takes into account the skills of each partners and the existing bricks, in particular those developed in the frame of GlobTemperature;
- The capacity to process and to reprocess very large amounts of data;
- The volume of data which could become bottlenecks in terms of transfer time and data storage;
- The verification steps to ensure that the production is done as expected;
- The validation with the development of a multi-sensor matchup system.

Considering the heritage of past projects on which LST CCI builds, the approach proposed to meet the requirements for LST CCI is:

- Analysis of all requirements which shall be used to specify the processing system(s) – this is the purpose of this SRD document
- Guidance will be then given to data provider teams by the system engineering team on how to adapt their processing chains to conform to all system specifications which will be described in the System Specification Document (SSD).

Activities to be carried out include: ensuring processing software acquires necessary source data, and can process and re-process LST ECV Products in a seamless End-to End system; implementation of development risk mitigation practices; assessment of the licencing restrictions to tools and technologies; implementation of version control for all software; ensuring bugs and issues are tracked and prioritised providing full traceability of the coding development activities; ensuring the data providers are responsible for regular backups of all software code and key resources.

Reciprocally, all modifications to existing systems by the EO Science Team at each processing cycle will be planned, communicated and iterated with the System Engineering Team, so as to have a comprehensive picture of the various systems used to perform the processing, and this will be traced in the Algorithm Development Plan.

Two main processing cycles for public availability have been proposed. In addition, there will be an intermediate offline processing for internal product development, testing, validation and assessment. The foreseen planning is:

- Cycle 1: this will run from M09 to M12 following development of the prototype ECV Production System, and selection and enhancement of the retrieval algorithms.
- *Cycle 1.5: this will run from M13 to M24 improving the algorithms and processing environment further.*
- Cycle 2: this will run from M13 to M36 following improvements to the algorithms taking into account lessons learnt from the validation, intercomparison and climate assessment of the products from Cycle 1.

Note: it shall be noted that Cycle 1.5 is intended to be internal to the project.

Verification of the output products will be carried out during the final processing stages of each processing cycle with results being communicated directly to the ECV production teams. Lessons learnt from Cycle 1 production, validation, verification and climate assessment will be documented as updates to the System Requirements Document (SRD) and System Specification Document (SSD) respectively in preparation for Cycle 2 processing. The results of the system verification will be documented in the System Verification Report (SVR).

The System Engineering Team will support the implementation of the processing system, ensuring it is designed and implemented with clear specification of its components, interfaces and dependencies. Software will be implemented to monitor data throughput and CPU load so as to provide guidance to the data providers on maximising efficiency. This is essential for the design of an operational-ready system. Thus, we aim to demonstrate a full end-to-end system.

The ECV Production System will generate both Level-2 swath data in harmonised format, and Level-3 collated (L3C) data at the spatial and temporal resolutions defined from the User Requirements document (ref. LST-CCI-D1.1-URD) and the traceability of product specification back to the URD is detailed in the Product Specification Document. All L3C data will be generated directly from intermediate, internal only (unless otherwise required as an output product from the User Requirements), Level-3 uncollated (L3U) data. The L3U data allows for customisable selection and processing of the required L3C resolutions. At this preliminary stage, it is expected that L3C data will be produced on daily, monthly, seasonal and annual bases as a minimum. DOIs shall be assigned to all ECV data sets made publicly available.

3.2. LST data products

LST is integral to a radiative energy budget of the surface as it provides the best approximation to the thermodynamic temperature that drives the outgoing longwave flux from surface to atmosphere and space. The theoretical basis for remote sensing of LST is that the total radiative energy emitted by the surface increases with an increase in temperature according to Planck's radiation law, the surface being whatever is viewed by the sensor at a nominal depth (referred to as the skin).

LST can be determined from thermal emission at wavelengths in the IR and MW atmospheric windows. Retrievals in the IR are generally more accurate than MW retrievals due to smaller variation of surface emissivities, better instrument calibration, independence of measurements from other temperature datasets, and stronger dependence of the radiance on temperature. Nevertheless, MW measurements have been shown to complement those in the IR due to their lower sensitivity to non-precipitating clouds, offering increased coverage in cloudy conditions albeit with a usually larger spatial resolution per pixel. Geostationary Earth Orbit (GEO) platforms (providing IR measurements), which allow regional coverage at high temporal resolution, provide frequent observations which can resolve the diurnal cycle since surface temperature changes vary significantly over periods ranging from hours to years and beyond. Low Earth Orbit (LEO) satellites provide global observations up to twice-daily with increased sampling at high latitudes.

The LST CCI project aims to deliver a significant improvement on the capability of current satellite LST data records to meet the GCOS requirements for climate applications of LST data. As reminded in the introduction, work from previous studies and projects (in particular ESA DUE GlobTemperature) provides a strong heritage which will allow the consortium to meet the objectives of the project in the defined timeframe (3 years of Phase 1, plus potentially another 3 years for Phase 2).

The system we propose for the generation of consistent time series of multi-sensor global satellite LST ECV data products for climate science is in line with [CR-2] and [CR-4] cardinal requirements.

The LST ECV Products which will be generated in each processing cycle are given thus:

- Cycle 1 products to be generated are as follows:
  - ATSR-2 on ERS-2 (1995-2003)
  - AATSR on Envisat (2002-2012)
  - MODIS on Terra (2000-2017)
  - MODIS on Aqua (2002-2017)
  - SLSTR on Sentinel-3A (2016-2017)
  - SEVIRI on MSG-1 to 4 (2004-2017)
  - SSM/I on DMSP F-13, 17 (1998-2016)
  - ATSR2-AATSR CDR (1995-2012)
  - Prototype Merged IR CDR for 1-year (2010)
- Cycle 1.5 products to be internally generated are as follows:
  - AVHRR/3 on NOAA-15 to 19 (1998-2018) [from GAC data]
  - AVHRR/3 on Metop-A to C (2007-2018) [from FRAC data]
  - ATSR-2 on ERS-2 (1995-2003)
  - AATSR on Envisat (2002-2012)
  - MODIS on Terra (2000-2018)
  - MODIS on Aqua (2002-2018)
  - SLSTR on Sentinel-3A (2016-2018)
  - Imager on GOES-12 to 16 (2004-2018)
  - JAMI on MTSAT-2 to 3 (2009-2015)
  - SEVIRI on MSG-1 to 4 (2004-2018)
  - SSM/I on DMSP F-13, 17 (1998-2018)
  - Time Series of SSM/I on DMSP F-13, 17 (1998-2020)
  - Prototype Merged IR CDR for 1-year (2010)
- Cycle 2 products to be generated are as follows:
  - AVHRR/3 on NOAA-15 to 19 (1998-2020) [from GAC data]
  - AVHRR/3 on Metop-A to C (2007-2020) [from FRAC data]
  - ATSR-2 (ERS-2: 1995-2003) and AATSR (Envisat: 2002-2012)
  - SLSTR on Sentinel-3A and 3B (2016-2020)
  - MODIS on Terra (2000-2020)
  - MODIS on Aqua (2002-2020)
  - SEVIRI on MSG-1 to 4 (2004-2020)
  - Imager on GOES-12 to 16 (2004-2020)
  - JAMI on MTSAT-2 to 3 (2009-2015)
  - Time Series of SSM/I on DMSP F-13, 17 (1998-2020)
  - Merged IR CDR for (2009-2020)
  - ATSR-SLSTR CDR (1995-2020)
  - Prototype all-sky Merged product (2010)

Note: High Resolution LST estimates, such as from Landsat or ASTER are out of scope for the current framework of the LST CCI

4. LST CCI Requirements

The set of LST CCI requirements has been derived from the following documents :

- The LST CCI ITT (ref. ESA-CCI-PRGM-EOPS-SW-17-0032, v1.3, 22/08/2017), more specifically the statement of work and its annex G (specific to LST);
- The LST CCI URD document (ref. LST-CCI-D1.1-UR, v1.1, 22/02/2019), prepared in the frame of the LST CCI project by UKMO;
- The Data Standards Requirements for CCI Data Producers document (ref. CCI-PRGM-EOPS-TN-13-0009, v2.0, 17/09/2018).

These 3 documents define various requirements, but not all are linked to the processing system or only in an indirect way.

All requirements defined in these 3 documents were reviewed and those which are relevant to the processing system are listed below.

In Table 2, Table 3 and Table 4 which summarise these requirements, we have displayed 2 Ids for the requirements: the first one is its Id in the document where the requirement was defined; the second, for the purpose of our project, is a new Id in which we include the origin of the requirement, on the model **LST-SYS-REQ-NN-DOC** where:

- NN is a double digit number
- DOC is the original document
  - SOW
  - URD
  - CCI

The analysis of the requirements from the 3 documents listed above have allowed the defininition of high-level categories which are displayed in Figure 1:

- The **Processing and reprocessing requirements (P)**
  - All requirements which are directly linked to the processing system and its capacity to reprocess large amount of data
- The **Design, evolution and scalability of the processing system (D)**
  - All requirements linked to the system design, its evolution (e.g., to include new processors) and its scalability (e.g., to handle additional dataset)
- The **Verification and Validation (V)**
  - All requirements linked to the verification and validation of the processing system (i.e., is it fit for purpose? Does it generate expected outputs?); the validation of the products is external to the processing system.
- The **Access to data (A)**
  - All requirements expressing the need to make the LST CCI products available to users communities.

The capital letter is used in the following tables to map each requirement with this high-level category.

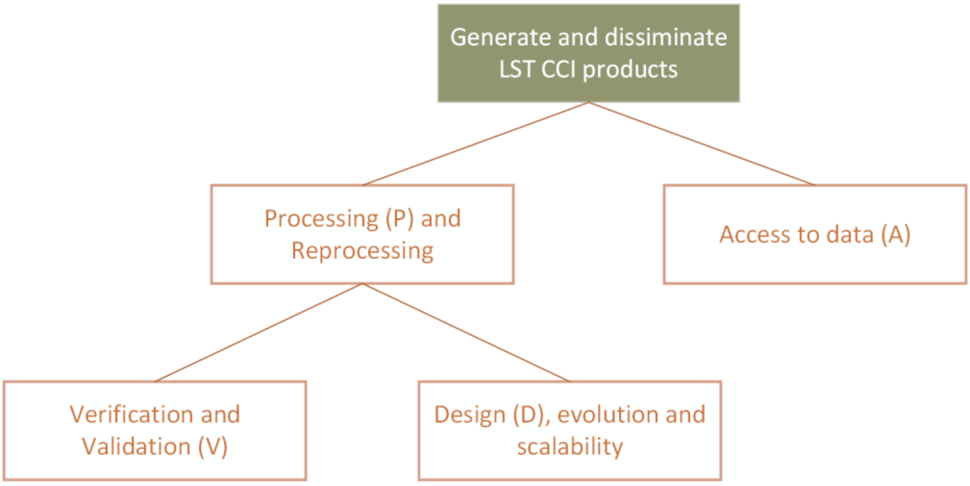

Figure 1: High-level LST CCI requirements

4.1. Requirements from the SOW

The requirements from [AD-1] which are applicable to the processing system are listed in Table 2.

Table 2: Requirements from the SOW applicable to the processing system

Original Id	LST CCI Requirement Id	SOW requirements	High-level category
CR-2	LST-SYS-REQ-01-SOW	Produce, validate and deliver consistent time series of multi-sensor global satellite ECV data products for climate science.	P
CR-4	LST-SYS-REQ-02-SOW	Generate and fully document a production system capable of processing and reprocessing the data in CR-2, with the aim of supporting transfer to operational activities outside CCI (such as C3S).	P
R-7	LST-SYS-REQ-03-SOW	Each CCI project (the contractor) shall take into account the legacy of the CCI in their projects. This involves: • adopting the community consensus concept • learning from, understanding and building upon the success achieved in CCI • espousing the implicit request to contribute actively to the core elements of the CCI Programme, in particular Colocation and Integration meetings, working groups, cross-programme activities (Open Data Portal, Toolbox, Knowledge Exchange) and cross-project initiatives.	D

Original Id	LST CCI Requirement Id	SOW requirements	High-level category
R-8	LST-SYS-REQ-04-SOW	Each CCI+ project (the contractor) shall assess the relevant documents, systems and data generated in CCI as the basis for their CCI+ project with the emphasis on the value and applicability of the algorithms, quality of outputs, specification of uncertainties and capability for reprocessing.	V
R-16	LST-SYS-REQ-05-SOW	Each CCI project team (the contractor) shall integrate data from the Copernicus Sentinels and other key satellite missions within the relevant CCI processing systems and ECV data products.	P
R-17	LST-SYS-REQ-06-SOW	Each CCI project team (the contractor) shall ensure that the system is adequately dimensioned to accommodate the growing volumes of input and output data, and the increasing computational loads needed to process, re-process, quality control, validate, and disseminate multi-decadal, global, ECV data products, of the required climate quality, in a timely manner.	D
TR-5	LST-SYS-REQ-07-SOW	The LST ECV Products delivered by LST_cci shall include: • LST derived from polar, geostationary, infrared and passive microwave satellite instruments. • All products shall: o cover the full globe (land and land-based ice surfaces) o cover the full mission lifetimes of the satellite missions selected in [TR-6] and [TR-9]. o include daily, monthly, seasonally and annually aggregated versions o be provided to users as single-mission Level-2 and Level-3 product versions, multi-mission Level-3 products, and potentially as higher-level derived products if required by the users.	P

Original Id	LST CCI Requirement Id	SOW requirements	High-level category
TR-6	LST-SYS-REQ-08-SOW	Infrared split-window based LST ECV products shall be derived from at least the following missions (and fully covering at least the time periods specified in parenthesis): • AVHRR/3 on board NOAA-15, 16, 17, 18, 19 (1998-2016) • AVHRR/3 on board MetOp-A, B, C (2006-2020) • ATSR-2 on board ERS-2 (1995-2003) • AATSR on board Envisat (2002-2012) • SLSTR on board Sentinels-3A and 3B (2016-2020) • MODIS on board TERRA and AQUA (1999-2020) • SEVIRI on board MSG-1, 2, 3 and 4 (2003–2020) • Imager on GOES 12–16 (2003–2020) • JAMI on MTSAT-2, 3 (2006–2018)	P
TR-7	LST-SYS-REQ-09-SOW	NOAA AVHRR products shall be derived from the reduced resolution but globally available GAC data type (4km). MetOp AVHRR products shall be derived from the full resolution (1km) data.	P
TR-8	LST-SYS-REQ-10-SOW	A global merged product covering the 1998–2020 period shall be delivered from the combination of all infrared polar and geostationary instrument data listed in [TR-6]. The merged infrared LST ECV shall provide global information on the diurnal cycle of LST derived from the combination of available polar orbiter overpasses and geostationary observation time slots in each day (see also [TR-23]).	P
TR-9	LST-SYS-REQ-11-SOW	Passive microwave based LST ECV products shall be derived from at least the following missions (and fully covering at least the time periods specified in parenthesis): • SSM/I on board DMSP F-8, 11, 13, 17 (1987–2016)	P
TR-10	LST-SYS-REQ-12-SOW	These products shall be merged together to provide a gap-free global passive microwave LST ECV covering the full period from 1987–2020. The merged LST ECV shall provide information on the diurnal cycle of LST derived from the combination of available orbital overpasses in each day (see also [TR-23]).	P
TR-13	LST-SYS-REQ-13-SOW	DOI's shall be assigned to all ECV data sets made publicly available.	P

Original Id	LST CCI Requirement Id	SOW requirements	High-level category
TR-19	LST-SYS-REQ-14-SOW	The Contractor shall build on the existing best practice, including that already developed in projects such as DUE GlobTemperature, LSA-SAF, CM-SAF, and the CCI programme to date, and made available in reports of the projects, colocation meetings, CCI working groups, and the published scientific literature.	D
TR-36	LST-SYS-REQ-15-SOW	Given the large amounts of data to be processed, the Contractor shall develop an automated high performance processing chain. This processor shall be implemented on a sufficiently powerful (possibly distributed) computing infrastructure that is capable of processing all the required products within the required project schedule.	D
TR-39	LST-SYS-REQ-16-SOW	The suite of LST ECV products produced annually shall be made publicly available before the end of the corresponding project Year, together with the associated validation results.	A
TR-45	LST-SYS-REQ-17-SOW	It shall be the Contractor's responsibility to acquire all EO and in-situ data required to carry out the project tasks.	P

4.2. Requirements for the URD

The vast majority of requirements expressed in the URD are not directly linked to the processing system, which is expected: the users have provided requirements that focus on the products they want to have in terms of contents, accuracy, temporal period, etc rather than how they want them to be processed. These requirements will be taken into account for the definition of the products which will be generated, but the link with the system definition is rather indirect at this stage.

Nevertheless a few requirements addressing the data format and delivery shall be taken into account in this document and are listed in Table 3.

Table 3: Requirements from the URD applicable to the processing system

Original Id	LST CCI Requirement Id	URD requirements	High-level category
A-01-O	LST-SYS-REQ-18-URD	CCI standard format is recommended for LST CCI products (req. LST-URD-ADV-01-O)	P
A-04-LI	LST-SYS-REQ-19-URD	Long term, easy access to data (req. LST-URD-ADV-04-LI)	A

Regarding the timeliness, it is noted from URD (in section 8.1.1) that "*For reasons of continuity and consistency, LST CCI data will be provided in either the GlobTemperature harmonised format, or the CCI standard format used elsewhere within the CCI project. Relevant requirements from the GlobTemperature RBD relating to data format, metadata specification and data access are summarised in Table 31. GlobTemperature requirements relating to timely delivery of data (e.g. near-real-time or 'NRT') are excluded here, as it is not a current objective of LST CCI to provide Interim Climate Data Records (ICDR) or NRT data. The user needs gathered in LST CCI aim to build on this existing information and focus on establishing the impact for LST users of choosing either the GlobTemperature harmonised format, or the CCI standard format, for the LST CCI products.*"

4.3. Requirements from CCI project guidelines

The Data Standards Requirements document provides the minimum requirements for CCI data producers to ensure consistency between output products from CCI projects. It was updated to version 2.0 in September 2018.

Table 4: Requirements from the CCI guidelines applicable to the processing system

Original Id	LST CCI Requirement Id	CCI requirement	High-level category
R-1 and R-2	LST-SYS-REQ-20-CCI	Projects who have commitments to produce data in other formats, shall do this in addition to the standardized products, and shall ensure these products comply as much as possible to the CCI Data Standards (e.g. filenames, metadata) : • a. Produced in netCDF-4 (classic) format 3 • b. Conform with the CF (Climate and Forecasting) convention 4 (currently version 1.7) • c. CF standard names used for the main variables • d. Include the global attributes listed in section 2.5.1 below	D
R-4	LST-SYS-REQ-22-CCI	The CCI projects shall ensure that INSPIRE compliant metadata records are created for each dataset. For datasets held in the CCI Open Data Portal, this requires the CCI projects to provide information to the Data Portal Team.	D
R-5	LST-SYS-REQ-23-CCI	CCI Data Producers shall use terms from the CCI vocabulary tables in the netCDF global attributes, or if terms are missing they shall request that they are added to the tables.	D

Original Id	LST CCI Requirement Id	CCI requirement	High-level category
R-6	LST-SYS-REQ-24-CCI	The key primary variables in the file and their related ancillary variables (e.g. uncertainty) shall be identified, and the range of their expected values shall be indicated.	D
R-7	LST-SYS-REQ-25-CCI	For consistency across CCI gridded products, variables shall have, as a minimum, the following dimensions: time, latitude, longitude (or alternative horizontal grid)	D
R-8	LST-SYS-REQ-26-CCI	CCI Data Producers shall use the common directory structure for all output data made available to users	D
R-9	LST-SYS-REQ-27CCI	CCI Data Producers shall use the CCI filenaming convention for all output data made available to users	D

5. Operations concept

The operations concept has been developed in order to fulfil specific requirements which are detailed in sections 4 and 6.

The main drivers are:

- The capability to process long-time series of LST ECV data
- The processing time, especially for reprocessing
- The volume of data to circulate
- The configuration control of the processing baselines which will be used to process the datasets
- The capacity to include updated or new processing chains and to process new datasets
- Monitoring of production

It is acknowledged that regular iterations will take place between the System Engineering Team and the EO Science Team.

Figure 2 describes how the operations will be carried out and what the workflow is:

- The LST processing chains (wherever they are located) use as inputs satellite L1 data and auxiliary data
- A verification process is included in production to check that generated datasets are fulfilling the requirements
- The LST CDRs are validated by the Validation team, which is external to the EO science team and the System engineering team – and thus formally speaking <u>outside</u> the processing system. Validation consists of comparing LST CDRs with other datasets (product intercomparison) and with in-situ measurements (product validation); comparison and consistency with other ECVs will also be included, firstly performed by the CRG and then by the CMUG;
- Once verified and validated, the LST CDRs are uploaded to the CCI Data Portal. They are then easily accessible to Climate User communities and to any other users who have an interest to such datasets.

Though not displayed in Figure 2, the data portal set up for the GlobTemperature project, which is still maintained, may be used for internal data transfer purposes.

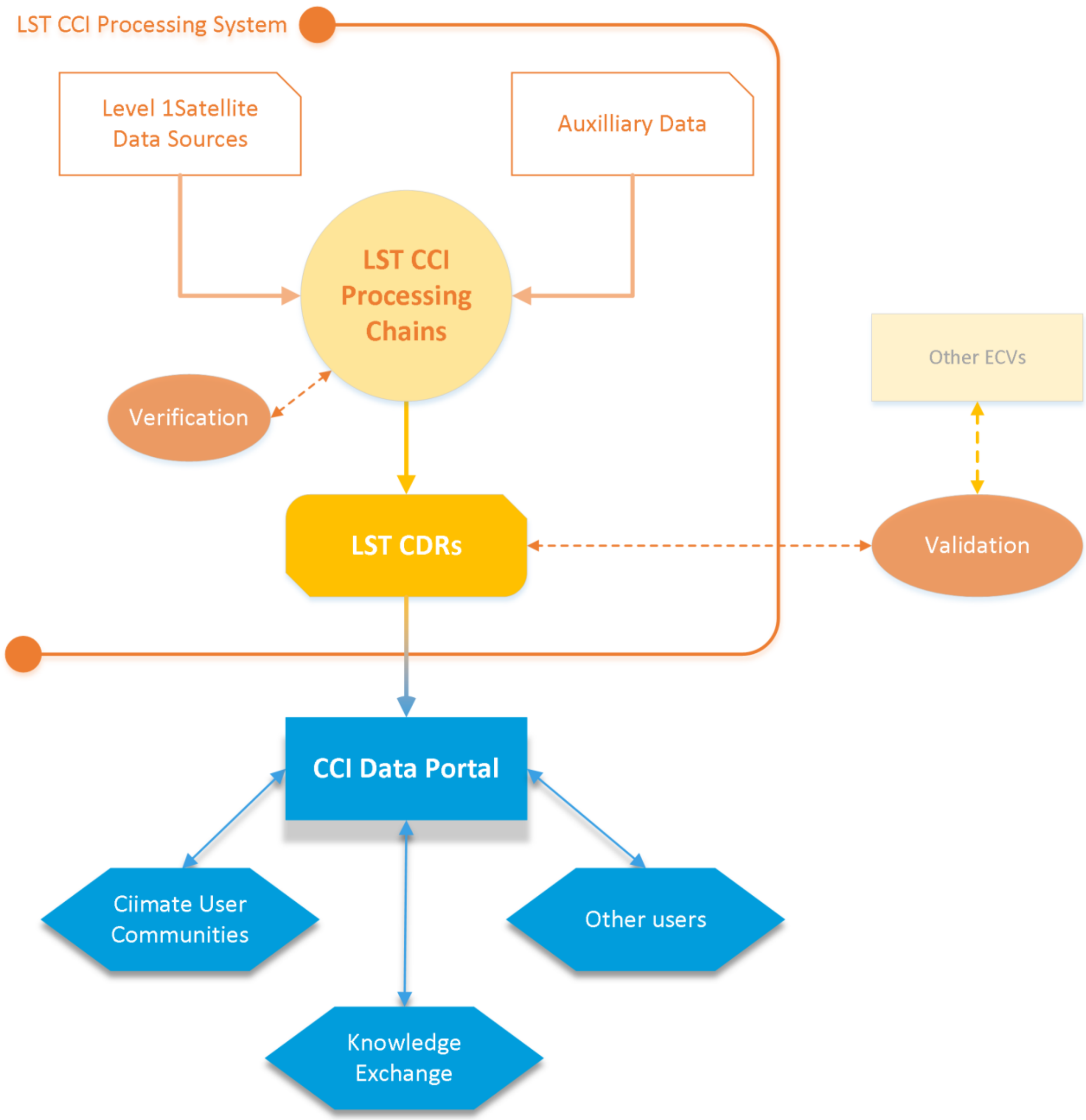

Figure 2: Operations concept of the LST CCI Processing System. The other ECVs referred to here include as a first estimate: i) snow cover; ii) land cover; iii) permafrost; iv) lakes; v) soil moisture; vi) clouds.

5.1. Prototype LST production workflow

5.1.1. Motivation for approach

The System for LST ECV Production will be distributed across three different centres:

- UK JASMIN will host and process all the LEO IR Products, the ATSR-SLSTR CDR, and the Merged GEO+LEO IR CDR. In addition it will also host the MMDB and processing associated with it.
- IPMA will host and process all the GEO IR Products, and the Merged GEO IR data prior to merging with the LEO data.
- Estellus will host and process all SSM/I and SSMIS MW data.

A fourth centre at ULeic can be exploited for additional offline development, testing and archival of intermediate internal data. This has a fast access to the UK JASMIN facility for transfer of data; and has a large storage capacity for internal datasets during the processing cycles. Finally, the GlobTemperature Data Portal, which is hosted by ACRI-ST, is planned to be used for storage and dissemination of some LST ECV Products produced during the project.

The distributed system also takes advantage of the processing nodes being next to both the output workspaces and the input Level-1 data. On UK JASMIN the central LOTUS processing system is able to directly access Level-1 data for ATSR-2, AATSR, SLSTR, MODIS and AVHRR, and output the Level-2 data onto project workspaces. The CEDA Archive is host to the ATSR-2, AATSR, SLSTR and Metop-AVHRR Level-1b data. MODIS Level-1b data (both Terra and Aqua) is directly accessible from Fast Tape, and NOAA AVHRR Level-1c is directly accessible from the SST CCI workspace, which resides as a workspace on JASMIN for which access can be requested.

While we choose this distributed approach as the most cost-effective in terms of both infrastructure costs and manpower effort, we will also review other possibilities and present in the system documentation the advantages of our chosen approach. In order to demonstrate the end-to-end processing of large quantities of data in an pre-operational environment we will consider not only ease of implementation, algorithm traceability, and end-user documentation, but also platform dependency, use of software libraries, licensing constraints etc. As such, this will provide a solid baseline for the Phase-II activities.

5.1.2. UK JASMIN

We aim at utilising for Phase II JASMIN as the prime centre for all processing of output products. However, during this Phase I, it was optimal, considering the time and budget constraints, in terms of both system design and cost-effectiveness, to rely on the existing distributed system since it builds directly off existing infrastructure, data availability and resource allocation. In Phase I, we will explore ways of migrating aspects of the distributed system onto the shared JASMIN resource.

The UK JASMIN/CEMS facility is a “super-data-cluster” scientific data analysis environment administered by CEDA. The default development environment is Linux Red-Hat Enterprise operating system. It provides access to interactive computing, a batch processing cluster (LOTUS), short-term project storage (Group Workspaces), a cloud computing service (JASMIN Cloud), and virtual machines. It is also linked to long-term storage archives; CEDA, BADC, NEODC and Met Office MASS.

Interactive computing on JASMIN/CEMS comprises login servers, scientific analysis servers (for interactive and low resource data analysis), data transfer servers (for larger data transfer tasks) and project-specific servers (virtual machines). Batch processing on JASMIN/CEMS is provided by LOTUS. LOTUS is a cluster of 218 hosts with over 4000 cores. It provides five public queues for batch and parallel processing with different run length and resource limits. The LOTUS cluster is used for longer running or resource intensive analysis jobs which are not suitable for running on the interactive scientific analysis servers. Short-term project data storage on the JASMIN/CEMS system is provided by Group Workspaces.

Group Workspaces can be backed up to the Elastic Tape system on request but are not backed up automatically. More than 12 petabytes of short term storage is currently allocated for Group Workspaces. Further Group Workspaces can be requested for new projects. For small short-term data storage requirements there are generic workspaces for members of NCAS, NCEO and CEH. There is also a limited amount of space in user’s home directories (10GB), which is fully backed up, and a “scratch” area for temporary small data volume storage (4TB shared across all users).

The shared project workspace on UK JASMIN will be that used for GlobTemperature. This workspace has now been extended to 50 Tb with NCEO resources to support the storage of additional output data for the Merged IR CDR, ATSR-SLSTR CDR, and the MMDB. The name of the workspace has now changed to "esacci_lst". All project partners have access to this workspace and are already using this as their main facility for data and software sharing.

5.1.3. IPMA

The IPMA operational processing chains deliver operational data for both the LSA SAF and Copernicus Land Services and thus have direct access to the required Level-1 GEO data for SEVIRI, GOES and MTSAT.

The existing processing chains of IPMA's Remote Sensing group for product development, testing and validation, are composed of 2 processing servers with: 20 cores; 96 GB Memory; 2.5 GHz velocity. Although the archive of LST CCI will be ensured outside IPMA, we will exploit a local storage with an average of 3 Tb/year/satellite to maintain all input and output data involved in the processing.

5.1.4. Estellus

The microwave processor will be constructed around a fast inversion algorithm based on multi-layer perceptron neural network (NN) architecture, and neither complex-powerful computing system architectures, nor parallel processing is required.

A Quad-Core Intel Xeon 5 3.7 GHz with 12 Gb of RAM running a macOS desktop operating system, associated to a RAID-enable system of 24 Tb, will be dedicated to the processing of microwave Brightness Temperatures (BTs) into LST. If the initial building of the NN architecture requires very intense calculations, e.g., for the initial weight calibrations, a CentOS based computing server with 12 Quad-Core AMD 2.1 GHz processors and 132 GB of RAM can also be used for the initial computations.

5.1.5. ULeic

The University of Leicester facility (ALICE) is a High Performance Computing (HPC) cluster providing several login nodes and several types of compute nodes. The default development environment is Linux Red-Hat Enterprise operating system. There are 170 standard compute nodes for job execution, 2 large memory nodes and 4 GPU nodes. ALICE provides 1.5 petabytes of shared storage for home directories, data and scratch. Of this, around 900TB is provided as storage for user data. This storage is accessible to all compute nodes on the cluster. There an additional Research File Store (RFS), which is not accessible from the compute nodes. The University of Leicester Surface Temperature group has access to around 80TB of available storage space for data processing, much of this has been utilised for GlobTemperature and will be available for additional LST CCI processing.

5.2. Preliminary Output Design

As described above, the processing system will build on existing systems but with the objective to centralize as much as possible the infrastructure. In this regards, each project team member (including non-UK partners as long as they are part of a project led a UK institute), once registered, will have access to the facilities offered by JASMIN: products developers, the validation team, and the Climate Research

Group will benefit from a direct access to data and to processing resources. The concept of the distributed system for processing and storage is illustrated in Figure 3.

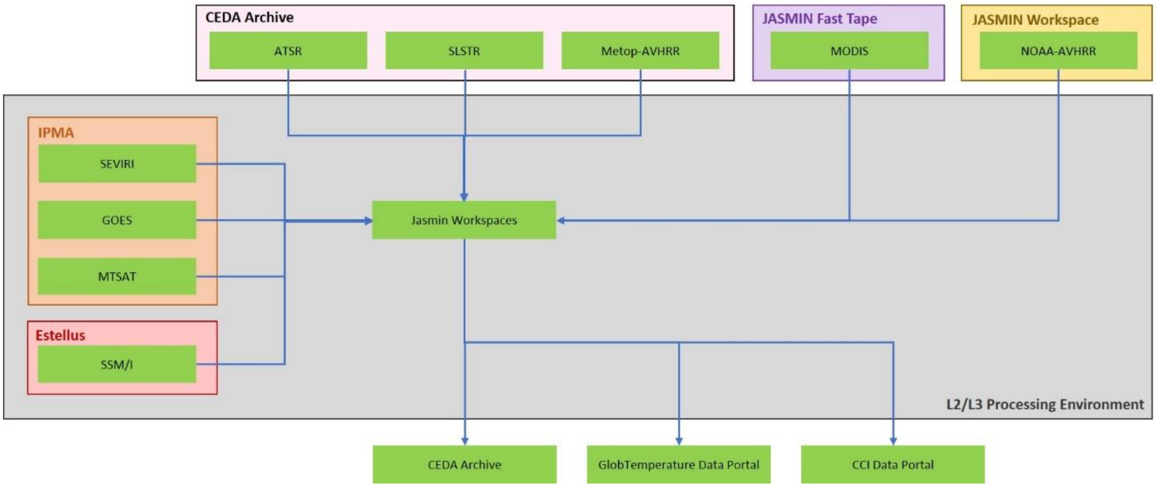

Figure 3: Distributed processing and storage solution for LST ECV Production.

Existing processing chains on the proposed distributed system, which will be advanced to support LST ECV production, are not pre-operational in terms of seamless delivery of climate quality LST ECV Products. In the first processing cycle we propose to specify, prototype, and demonstrate the additional components, built on these systems, needed to create multi-sensor LST ECV Products in a pre-operational capacity.

We propose to enhance, where necessary, these systems to deliver climate quality LST ECV Products. This includes systems designed to deliver the requirements of the climate user community such as cycles of validation, intercomparison and verification of algorithm improvements, and dissemination of final products in Obs4MIPs format.

For the final output data format of the LST ECV Products to be delivered through the Climate Research Data Package (CRDP) we will adhere to the CCI Data Standard. The LST harmonised format [RD-66] will need only minor modifications to be compatible with the CCI Data Standard. The fact that the LST harmonized format is close to the CCI is an important element for the LST community who has widely accepted it, this format becoming the standard both for climate users of gridded global LST data and users of high resolution LST Level-2 data in local applications.

All datasets provided by GlobTemperature are in netCDF-4 format [RD-133] using the CF-1.6 metadata convention [RD-39]. This is a self-describing, portable, scalable, appendable, sharable, archivable, and machine-independent data format. It is supported by all major data analysis and visualisation packages, and programming interfaces exist for a wide variety of other programming languages. The data are provided as individual variables in the netCDF files. For CCI datasets we propose moving to the latest version of CF-compliance.

The GlobTemperature format only includes essential variables as mandatory. These are held in the primary (LST) datafiles which ensure datafile size is at a minimum, and thus is not restrictive to wide user uptake. Optional fields are included in accompanying auxiliary (AUX) datafiles. Where information is available for a given sensor then these are to be provided as mandatory in these auxiliary datafiles.

For datasets where the uncertainty budget is explicitly broken down into the different components these are available in the AUX datafiles. The detailed contents of the GlobTemperature datafiles are presented in full in [RD-66].

The main differences between the GlobTemperature harmonised format and the CCI Data Format can be summarised as follows:

- CCI Data Format requires a few additional mandatory metadata
- CCI Data Format dimensions are identified as “time”, “latitude” and “longitude” with associated 1d coordinate variables. The harmonised LST format Level-2 format dimensions are “time”, “ni” and “nj” with associated 2d coordinate variables for the latter two. The harmonised LST format Level-3 format dimensions are “time”, “lat” and “lon” with associated 1d coordinate variables.
- File naming conventions are similar in content but in a different order

These minor changes to the existing harmonised format to meet the CCI Data Standard will minimise any impact on the large community of LST users of the harmonised format. Finally, the necessary formatting changes will be implemented for conversion of the CDR Products into Obs4MIPs format with final ECV products submitted to the WCRP Data Advisory Council for inclusion.

Furthermore, the output data format applied to the datasets generated in production will be compliant with the obs4MIPs initiative, as requested in [AD-1]. This involves products in NetCDF format including metadata compliant with CF (Climate and Forecast) standards. This point will be part of the verification process explained hereafter.

Addresses Requirements: LST-SYS-REQ-01-SOW, LST-SYS-REQ-03-SOW, LST-SYS-REQ-04-SOW, LST-SYS-REQ-18-URD, LST-SYS-REQ-20-CCI

5.2.1. LEO IR Products

The LEO IR ECV Products will be processed on UK JASMIN utilising as much as possible the existing infrastructure from pre-cursor projects. The products generated include those for ATSR-2, AATSR, SLSTR, MODIS, Metop-AVHRR and NOAA-AVHRR. The processing chains for ATSR-2, AATSR, SLSTR and MODIS will be prototyped in the first processing cycle (Cycle 1) and lessons learnt will be implemented for re-processing of these and processing of the AVHRR products in Cycle 1.5 and Cycle-2.

All the LEO IR processing is performed on Linux-based systems. All processing carried out will be parallelised. Experience from pre-cursor projects suggests multi-mission re-processing is most time and cost-effective by parallelising on daily sub-divisions. For these processing systems estimates of processing CPU time for 1-month of data of the order of 600 hours. Parallelisation means that in real-time full mission data can be re-processed in a matter of days.

The processing software will be built on the existing processing chains and practices developed in the precursor projects, in this case ESA DUE GlobTemperature, which addresses LST-SYS-REQ-14-SOW. All existing software code for the existing processing chains is version controlled using Github. This will be maintained for LST CCI starting with a new Github repository linked to JASMIN. All LEO IR processing code will be Open Source. For deployment of the output products to the CCI Data Portal the corresponding version of the processing software will be added to the common CCI github code repository.

The System Engineering team will develop and deploy verification tool(s) for the processing chains to automatically check that the production is done accordingly to the specifications. This tool will be further detailed in the SSD.

Addresses Requirements: LST-SYS-REQ-07-SOW, LST-SYS-REQ-08-SOW, LST-SYS-REQ-09-SOW, LST-SYS-REQ-14-SOW

5.2.1.1. ATSR-2 and AATSR

The ATSR-2 and AATSR ECV Products will be created using the existing GlobTemperature processing chains, with any necessary modifications, on UK JASMIN for these data. The processor has direct access to the Envisat-format Level-1b data from the CEDA archive. The processor takes as other input various auxiliary datasets, the algorithm formulation and associated coefficient ADFs, and the cloud mask. The processor itself is modular based in which pre-processing stages can be run independently or integrated with the Level-2 and Level-3 processing modules.

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. For example, in the GlobTemperature CDR Trade-off-Analysis [RD-65] three different algorithms were assessed: i) the ATSR-SLSTR algorithm; ii) the Generalised Split Window algorithm; and iii) the SEN4LST algorithm.

Another pre-processing stage is the derivation of the cloud mask. For probabilistic-based approaches this includes the generation of probabilities from simulated data. In the case of the UOL_V3 cloud mask this involves the generation of expected BTs from coincident profile data using a radiative transfer model. The profile data is acquired from ECMWF and interpolated onto the ATSR tie-point grid. The proposed model proposed here is RTTOV-11 which offers fast simulation in line with the expectations of an operational system. Indeed, this approach with the same model is being implemented in the Sentinel-3 IPF for per-pixel probabilistic cloud masking. Consistency with other ECVs e.g. land cover which does cloud clearing using MERIS will be checked. However, the LST CCI cloud masks use additional channel data so some differences will exist.

The output datasets will 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 is shown in Figure 4.

The processing software will be built on the existing processing chains and practices developed in the precursor projects, in this case ESA DUE GlobTemperature, which addresses LST-SYS-REQ-14-SOW. A key feature of all the existing processing chains is semi-automated generation of matchups for both validation and intercomparison. Here we propose fully automating this functionality. Finally, the existing processing chains include generation of per-pixel uncertainty budgets using a 3-component model similar in principle to the expected approach in the ECV Production here. Nevertheless, the software for uncertainty characterisation will be modified to incorporate any new components as defined for the LST ECV Products.

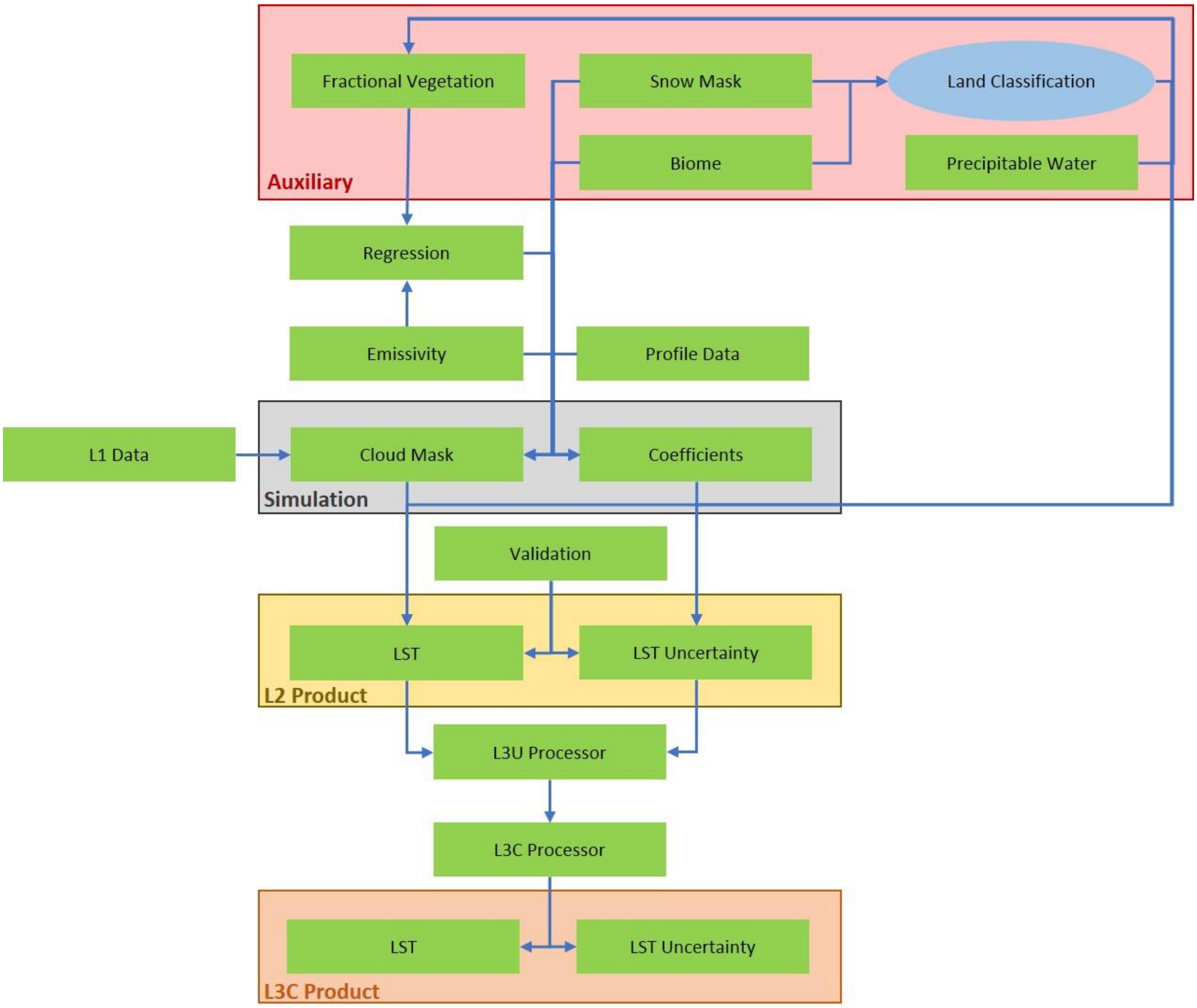

Figure 4: Data flows for the AATSR LST ECV prototype production system.

Addresses Requirements: LST-SYS-REQ-08-SOW, LST-SYS-REQ-14-SOW,LST-SYS-REQ-17-SOW

5.2.1.2. SLSTR

The SLSTR ECV Products for Sentinel-3A and 3B will be created using the existing GlobTemperature processing chain for SLSTR. Although the GlobTemperature system is implemented on an ACRI-ST virtual machine with direct access to the S3MPC data, the same system has been prototyped on UK JASMIN following the recent availability of Level-1b SLSTR on the CEDA Archive. The implementation on UK JASMIN has direct access to the NetCDF-format Level-1b data in addition to the other various input auxiliary datasets as per the ATSR processing chains.

For the cloud mask pre-processing stage the generation of probabilities here will utilise the meteorological data available with the SLSTR level-1b data on the tie-point grid. In the first processing cycle (Cycle 1) SLSTR data from Sentinel-3A will be processed. It is then expected that Sentinel-3B post-commissioning Level-1b data will be available for Cycle 2.

As for ATSR, the output datasets are to be processed onto the permanent Leicester workspace on JASMIN which has all the necessary auxiliary data for fast and cost-effective processing. The rationale for processing SLSTR data on UK JASMIN in LST CCI is colocation of all the input data streams for the ATSR-SLSTR CDR and the merging of the GEO and LEO data thus minimising the transfer of large data volumes.

Addresses Requirements: LST-SYS-REQ-05-SOW, LST-SYS-REQ-08-SOW

5.2.1.3. MODIS

The Terra and Aqua MODIS LST ECV Products will be created using existing processing chains created for pre-cursor projects (GlobTemperature and H2020 EUSTACE). These are implemented on UK JASMIN with direct access to Collection-6 Level-1b HDF-EOS MODIS data on Fast Tape. This includes the necessary radiances, geolocation data, satellite angles, and quality control information such as low confidence, aerosol affected pixels and operational cloud mask.

As for ATSR the derivation of the optimised probabilistic-based cloud mask is a pre-processing stage in the chain. This requires generation of probabilities from simulated expected BTs from coincident ECMWF profile data using RTTOV-11 at the MODIS tie-points (every 5 pixels).

MODIS is currently processed onto the permanent NCEO workspace on JASMIN and will continue to for LST CCI for cost-effectiveness. AVHRR is on SST CCI and MODIS L1c is being ingested onto the CEDA Archive All workspaces are directly accessible from the LOTUS processing infrastructure, and as for ASTSR and SLTR these will be directly accessible for the merging of the GEO and LEO data.

Addresses Requirements: LST-SYS-REQ-08-SOW

5.2.1.4. Metop-AVHRR

The AVHRR/3 LST ECV Products for Metop-A, B and C will be created using an existing prototype processing chain created during GlobTemperature. This is implemented on UK JASMIN with direct access to EPS native format AVHRR FRAC data on the CEDA Archive. This includes the necessary radiances at full 1km resolution, geolocation data, satellite angles, and quality control information such as low confidence (L1 information on poor calibration, saturation, pointing, etc.).

As for ATSR the derivation of the optimised probabilistic-based cloud mask is a pre-processing stage in the chain. This requires generation of probabilities from simulated expected BTs from coincident ECMWF profile data using RTTOV-11 at the AVHRR tie-points (every 16 pixels).

Metop-AVHRR is currently processed onto the Leicester workspace on JASMIN and will continue to for LST CCI for cost-effectiveness.

Addresses Requirements: LST-SYS-REQ-08-SOW

5.2.1.5. NOAA-AVHRR

The AVHRR/3 LST ECV Products from the NOAA-15 to NOAA-19 will be created by adapting the existing prototype processing chain for Metop-AVHRR created during GlobTemperature. This will be implemented on UK JASMIN with direct access to value-added Level-1c AVHRR GAC data on the SST CCI Workspace. This includes the necessary BTs, geolocation data, satellite angles, and quality control information at 4km.

As for ATSR the derivation of the optimised probabilistic-based cloud mask is a pre-processing stage in the chain. This requires generation of probabilities from simulated expected BTs from coincident ECMWF profile data using RTTOV-11 per each AVHRR pixel. The output products will be processed onto the Leicester workspace on JASMIN.

Addresses Requirements: LST-SYS-REQ-09-SOW

5.2.1.6. VIIRS

Though there is no requirement related to generate LST ECV products from VIIRS, we see this as an important instrument into the future being the successor to MODIS. Therefore, we propose to build a prototype processor for VIIRS, and to run the selected algorithm from the Round Robin Algorithm Intercomparison on extractions for the MMDB. This will allow us to assess the potential of a future LST ECV Product from VIIRS. This is of benefit going forward into Phase II, where we would then be better positioned to roll out such a development in replacing the MODIS Products beyond 2020.

5.2.2. GEO IR Products

The generation of LST ECV Products derived from geostationary satellites will be performed at IPMA through modification of one of the existing processing chains. 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 5).

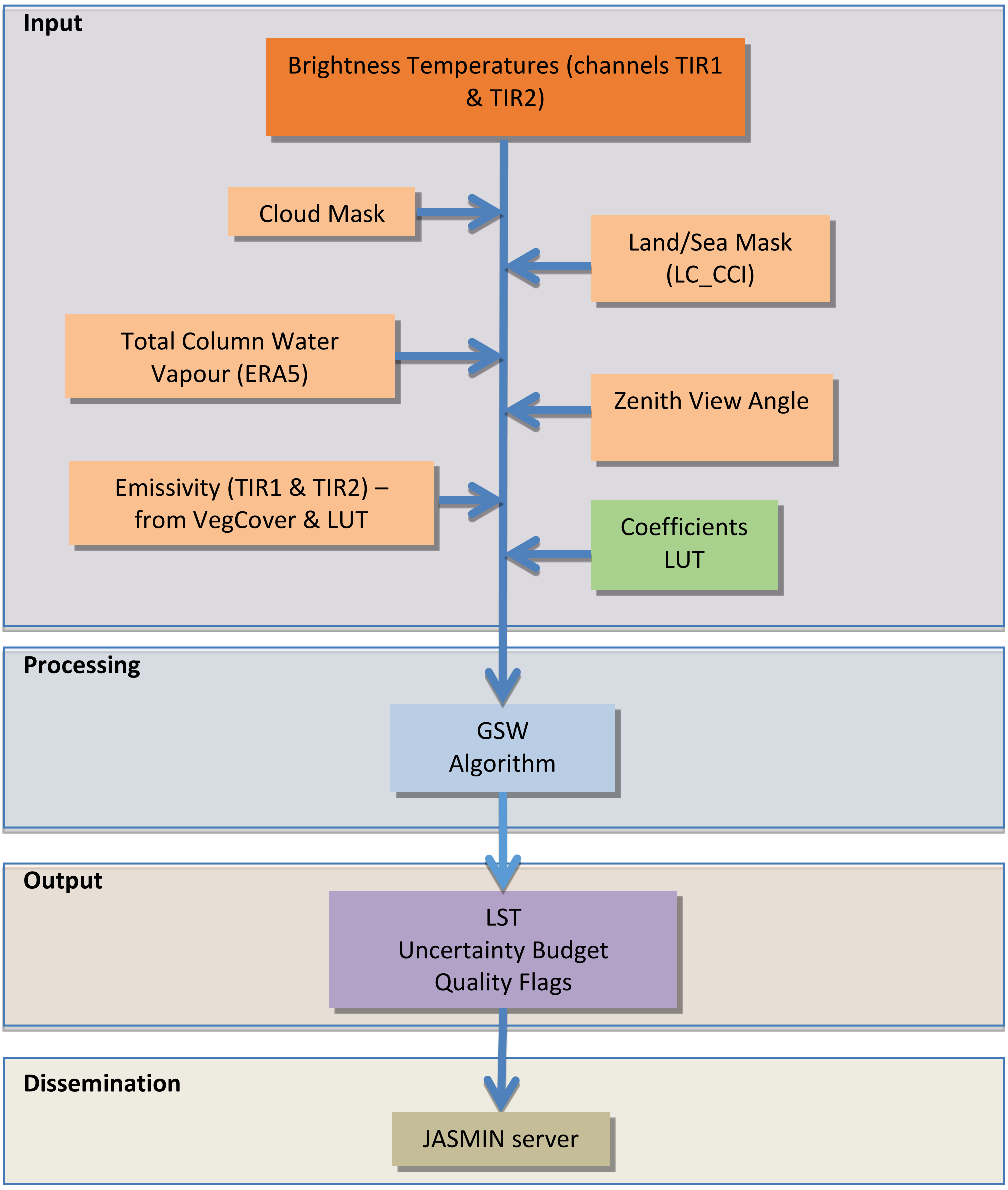

Figure 5: Flowchart for LST production from sensors with split-window such as SEVIRI/Meteosat.

For sensors/data with a single channel in the thermal infrared window and one middle IR channel (e.g., GOES-E, MTSAT data archived at IPMA up to 2012) a "Dual Algorithm" variant on the Generalised Split-Window approach is applied (Figure 6). For the single-senor GEO products, we intend to utilise the best cloud clearing algorithm for that sensor. Currently, the operational cloud clearing algorithm for SEVIRI for example is optimal. If this proves still to be the case for this single-sensor product, compared to the probabilistic cloud test (UOL_V3) proposed for the Merged IR CDR, then we will retain this approach rather than risk any degradation. The processing chains include generation of per-pixel uncertainty budgets, and software for uncertainty characterisation will be modified to incorporate any new components as defined for the LST ECV Products.

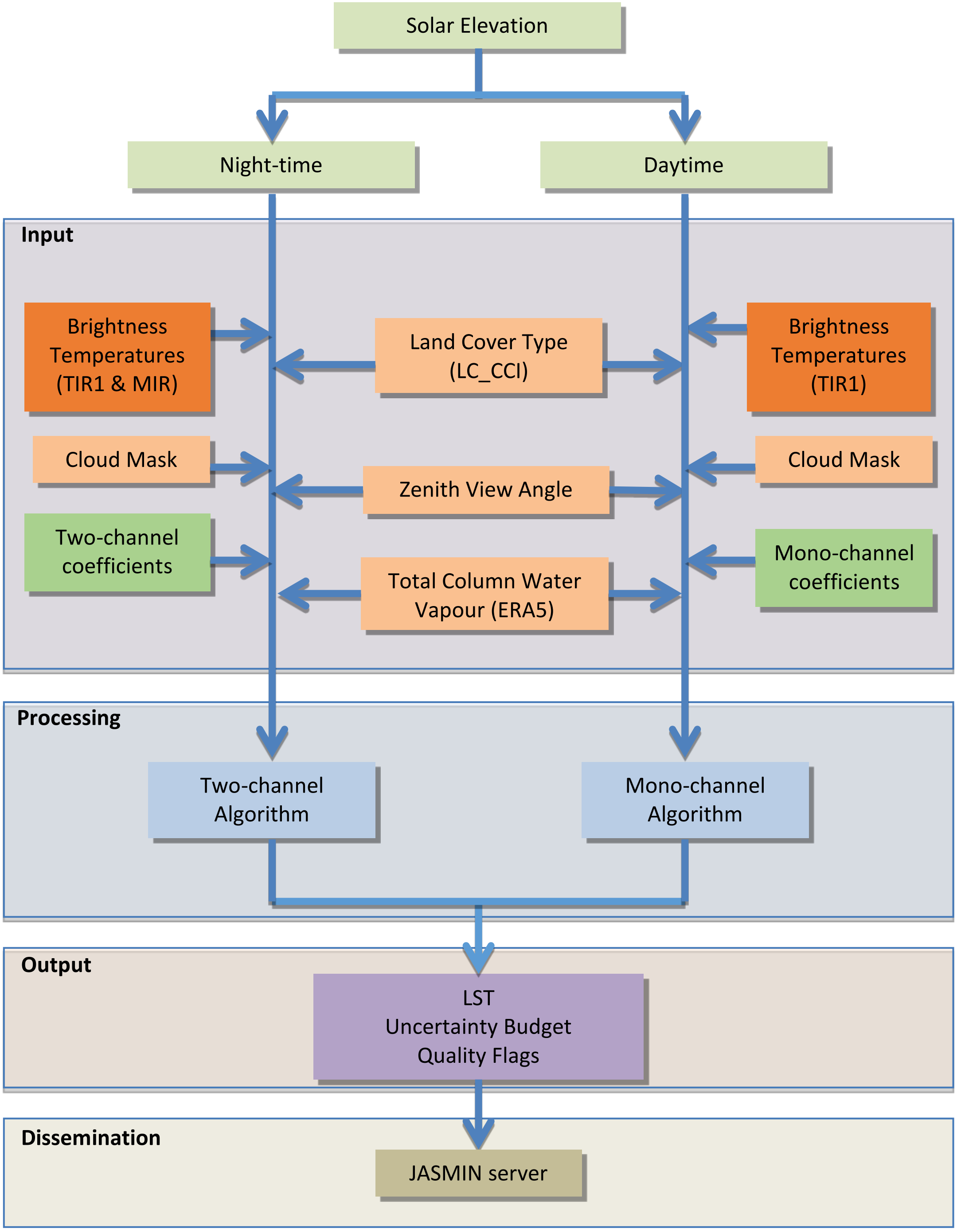

Figure 6: Flowchart for LST production from sensors with a single-channel such as GOES-E/MTSAT up to 2012.

Addresses Requirements: LST-SYS-REQ-07-SOW, LST-SYS-REQ-10-SOW

5.2.3. MW Products

The Passive Microwave ECV Products will be processed on Estellus infrastructure utilising as much as possible the existing resources and capabilities from pre-cursor projects. The products generated include those for SSM/I and SSMIS. The processing chains will be prototyped in the first processing cycle (Cycle 1) and lessons learnt will be implemented for re-processing in Cycle-2.

The processing of the SSM/I and SSMIS data is illustrated in Figure 7. The processing chain consists of three main processors: the L1 processor, the Uncertainty processor, and the L3 processor.

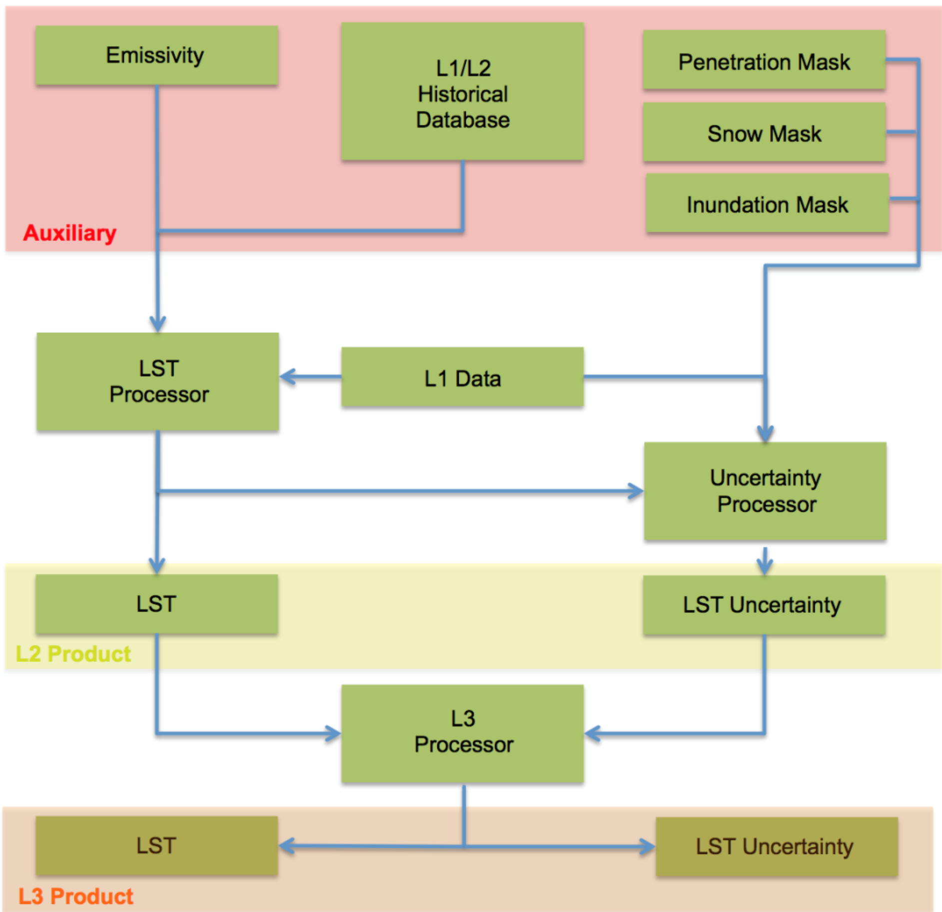

Figure 7: Data flows for the SSM/I and SSMIS LST ECV prototype production system.

The L1 processor converts L1 BTs to LST, and operates by means of a multi-layer perceptron NN calibrated to regress LST on BTs. Apart from the BTs, the NN also takes climatological monthly surface emissivities at the BT channels as an input to further constrain the inversion problem. A calibration of the coefficients of the NN is first required, and it is achieved by training the NN with a historical database of observed BTs and retrieved LSTs.

The Uncertainty Processor characterizes the L1 inversion uncertainty. The L1 processor is applied to a new subset of the training database not seen by the NN during the calibration phase, and the differences between the target LST in the database and the newly retrieved LST is used to infer an uncertainty estimate. To speed the Uncertainty processor, this calculation is done off-line for a range of emissivity and LSTs, and the results stored in a lookup table. For a given L1 processing of BTs, the lookup table is searched and the found value used as an estimation of the inversion uncertainty for the corresponding LST retrieval.

The Uncertainty processor also derives a series of flags to signal potentially difficult inversions. For the SSM/I and SSMIS frequencies, BTs inversions can be difficult for snow-covered, water-covered, or very arid surfaces. The L1 processor also operates for these conditions, but the inversion error is likely to be larger. Identifying these situations is based on masks derived from external products available in the auxiliary data. The masks identify: (a) snow conditions, using a daily snow flag derived from ERA-5 atmospheric reanalysis; (b) inundated regions using a using a monthly climatology from the Global Inundation Extension from Multi-Satellites (GIEMS), and; (c) areas with large microwave penetration depth, using a monthly climatology of radar backscattering. A flag to identify atmospheric convection activity is also produced by monitoring BT depressions in the high frequency channels.

The L3 processor aggregates in space and time the LSTs retrieved at the sensor swath grid by the L1 processor, and propagates inversion uncertain and flags generated by the Uncertainty processor to the new spatial and time resolutions.

The output datasets are to be transferred onto the shared project workspace on UK JASMIN for final product validation and climate assessment by the CRG prior to delivery to the CCI Data Portal.

Addresses Requirements: LST-SYS-REQ-07-SOW, LST-SYS-REQ-11-SOW, LST-SYS-REQ-12-SOW

5.2.4. Long-term IR Climate Data Records

This section describes the CDRs which will be produced as part of the LST CCI project. The algorithms have been developed in previous projects and consolidated during GlobTemperature.

Addresses Requirements: LST-SYS-REQ-01-SOW, LST-SYS-REQ-10-SOW, LST-SYS-REQ-19-URD

5.2.4.1. Merged IR CDR

The Merged IR CDR will be processed across the UK JASMIN – IPMA distributed systems. The processing chain will be prototyped in the first processing cycle (Cycle 1) for select instruments over 1-year (nominally chosen as 2010); these being AATSR, Terra-MODIS, SEVIRI, GOES, and MTSAT. This period is sufficient to demonstrate the principle of the system and create sufficient data for assessment of the quality of output over a complete annual cycle. Lessons learnt will be implemented for re-processing, and the addition of new sensors, for the CDR in Cycle1.5 and Cycle-2.

For the Merged IR products this 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 8.

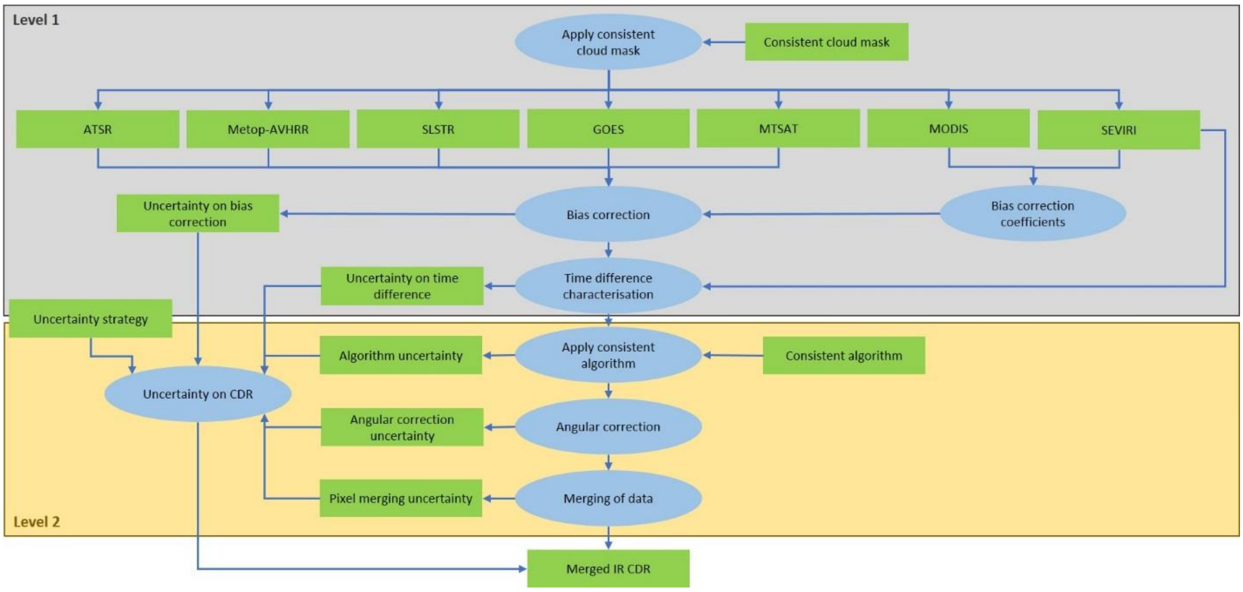

Figure 8: Schematic representation of the development and implementation of the Merged IR CDR.

The work builds on the advances made for LST product merging in GlobTemperature [RD-67; RD-41]. 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 consistent uncertainty budget according to the model detailed in E3UB. The Merged IR CDR will be produced on a global scale eight times daily (00h, 03h, 06h, 09h, 12h, 15h, 18h, and 21h UTC times). The nominal spatial resolution will a 0.05° equal angle latitude-longitude grid. The individual stages are described in detail below.

Stage 1: bias correction

First, we will develop bias corrections between the different input Level-1 data. Following the approach developed in GlobTemperature, bias corrections to a single reference sensor will be determined. This is, in essence, the removal of systematic differences between instruments. The bias correction is carried out by computing linear regressions between BTs from a reference sensor and those of the sensor to be bias-corrected. The computation is done for satellite viewing angles within a limit range (± 5º for zenith and ± 5º for azimuth angle) and, in a first approach, for night-time only. The matching between LEO and GEO data will be performed for a select number of days for each month in a given year.

The reference sensor chosen is MSG-SEVIRI, which is the only sensor in the ensemble which has an operational maturity level in terms of Global Space-based Inter-Calibration System (GSICS) calibration to the standard reference source (IASI-A) [RD-83; RD-25]. The LEO sensors will thus be corrected to SEVIRI. For the remaining GEOs we will use MODIS as the transfer sensor. The choice of MODIS is justified given the wider swath for this sensor compared with AATSR for example. The possible impacts of MODIS calibration issues will be assessed as part of the development. Given the viewing angle limitation we must impose on the LEO-GEO matching, MODIS allows for a higher number of superposition matchups with the GEO data. The associated uncertainty from these bias correction adjustments will be determined in accordance with the strategy developed in the E3UB.

Stage 2: time difference characterisation

Each GEO has a different scanning time of its full disk, and each has a different delay relative to the nominal time given in the respective file (schematically represented in Figure 9).

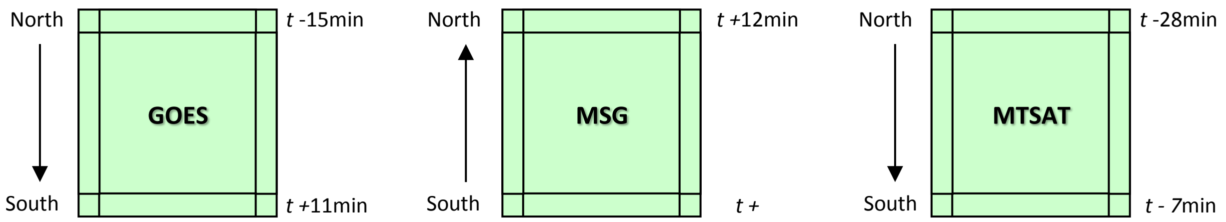

Figure 9: Scanning delays for each GEO relative to their respective time stamp, and scan direction.

In order to better relate the nominal time stamp of the merged IR CDR product and the individual sensing times, and taking into account the temporal sampling of the different GEO products (15-minute in the case of SEVIRI; hourly in the case of GOES and MTSAT LST), we will apply the following criteria: for any given hour (e.g., 03 UTC), we select SEVIRI/MSG data with nominal time 15 minutes before (i.e., 02:45), to be merged with GOES and MTSAT corresponding to the same nominal time-slot (i.e., 03:00). In this manner, we aim to minimize the time differences between the GEO's actual observation times (as illustrated in Figure 10).

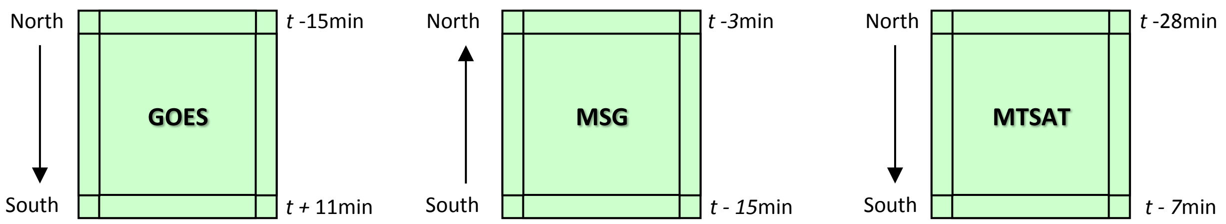

Figure 10: Same as Figure 9, except for MSG (middle), with retrieval time of 15 minutes prior to MTSAT and GOES.

The same assessment of the time differences will be made for LEOs. On a global scale, each UTC time step is only sparsely filled with LEO data. Nevertheless, the LEO data in many cases fills in areas of the globe unobserved by the GEO satellites. As more satellites are added (in processing Cycle 2 for instance) then coverage will increase further. On a per-pixel scale the input Level-1 LEO data is included at a given time step if their corresponding observation time is within a minimal threshold of the available GEO data in line with the minimisation of the temporal difference with the output UTC time as described above.

Time corrections to a common UTC are not straight forward. This is because the diurnal variation in LST may vary significantly from place to place and from day to day. We will therefore consider building models of the diurnal cycle with geostationary data (following the strategy already used within the LSA SAF for the derived LST product; [RD-75]) to estimate the expected rate of change of LST for any time of the day. These models will be applied to 10-day synthesis of geostationary LST (median LST per observation time-slot). The estimated rate will be compared with that obtained from a linear fitting between two consecutive geostationary observations to assess if 10-day adjustment deviates significantly from the specific day. We expect that the time adjustments obtained with hourly geostationary data will hold for the polar-orbiter LST as well.

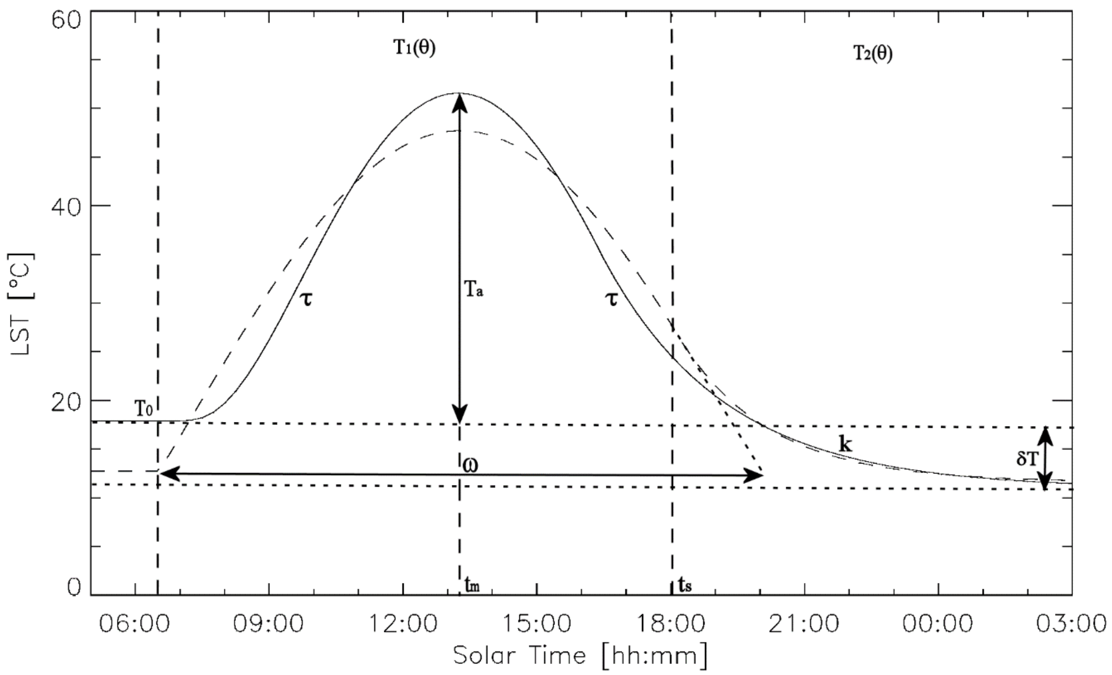

Figure 11: Idealized model of the LST diurnal cycle for a clear sky day, following the parameterizations described in [RD-75].

The merged product shall be made available together with a dataset describing the expected LST deviation from that expected from a reference (nadir) view. This will be based on the parametric modelling of LST directional effects described above.

The understanding of the impact of temporal differences between sensors is key information for use elsewhere in the project. Based on requirements from climate users documented in the URD and experience from pre-cursor projects a strategy for including diurnal information in the ECV products can be formulated. The associated uncertainty corresponding to this time difference to the nominal time will be characterised.

Stage 3: apply consistent Level-2 algorithms

The best performing algorithm from the Round Robin exercise will be applied to the bias corrected Level-1 data to ensure consistency from an algorithm perspective across the Merged IR CDR.The processing cycles for these are as follows:

- Cycle 1 (Prototype Merged IR Product for 1-year [2010]) from the following:
  - MODIS on Terra
  - AATSR on Envisat
  - SEVIRI on MSG-2
  - Imager on GOES-13
  - JAMI on MTSAT-2
- Cycle 1.5 (Prototype Merged IR Product for 1-year [2010]) from the following:
  - AVHRR/3 on Metop-A
  - AATSR
  - MODIS on Terra
  - MODIS on Aqua
  - SEVIRI on MSG-2
  - Imager on GOES-13
  - JAMI on MTSAT-2
- Cycle 2 (2009-2020) from the following:
  - AVHRR/3 on Metop-A to C (2009-2020)
  - AATSR (Envisat: 2009-2012)
  - SLSTR on Sentinel-3A and 3B (2016-2020)
  - MODIS on Terra (2009-2020)
  - MODIS on Aqua (2009-2020)
  - SEVIRI on MSG-1 to 4 (2009-2020)
  - Imager on GOES-12 to 16 (2009-2020)
  - JAMI on MTSAT-2 (2009-2015)

The developments will be transfered to the Merged IR CDR ensuring a consistent approach to cloud clearing across all input sensors. The UOL_V3 cloud mask, which has been shown to perform very well in the GlobTemperature Cloud Clearing Round Robin, will be adapted and implemented for the GEO inputs to the Merged IR CDR. For Cycle 2 this cloud mask will be applied to the Level-1 data for the individual sensors.

Stage 4: angular correction

Satellite LST retrievals correspond to an integrated radiometric surface temperature within the sensor footprint and, therefore, remotely sensed LST is in general highly anisotropic. As such, a scene viewed by the same sensor from different viewing angles would lead to different LST retrievals. This effect contributes to increase the LST variability among datasets. Several methodologies have been developed to simulate LST directionality, either through physically-based formulations such as those relying on radiative transfer modelling, geometric and 3-D models, or by means of parametric models [RD-153]. Since physical models require detailed knowledge of surface characteristics, which is not readily available at the continental or global scale, we aim the use parametric models to describe the variability of LST estimates with viewing and illumination geometries. The latter are also computationally more efficient and require few input data, which makes them particularly appropriate for operational use in satellite LST production.

We propose to consider two methodologies to simulate LST anisotropy based on the parametric approach [RD-154] and on a modified version of the model developed by [RD-104]. In the former LST directionality is modelled by two kernels composed of trigonometric functions of the viewing and sun geometries, and its potential to be applied to existing LST products has been tested in several studies [RD-42; RD-139; and RD-154]. The other approach based on the hot spot formulation proposed by [RD-137] for the optical domain where reflectance is formally replaced by surface temperature has proven to be more successful in representing in situ observations [RD-36; RD-104; and RD-167], although its calibration may be more sensitive to uncertainties in LST estimations.

➜ Vinnikov et al. (2012)

$$\frac{T(\theta_v,\theta_i,\Delta\phi)}{T_0} = 1 + A\Phi(\theta_v) + D\Psi(\theta_v,\theta_i,\Delta\phi)$$

Emissivity kernel:

$$\Phi(\theta_v) = 1 - \cos(\theta_v)$$

Solar kernel:

$$\Psi(\theta_v,\theta_i,\Delta\phi) = \sin(\theta_v)\cos(\theta_i)\sin(\theta_i)\cos(\Delta\phi)\cos(\theta_i - \theta_v)$$

➜ Lagouarde and Irvine (2008) – Modified version

$$T(\theta_v,\theta_i,\Delta\phi) - T_0 = (\boldsymbol{T_H} - \boldsymbol{T_0})\frac{e^{-kf} - e^{-kf_0}}{1 - e^{-kf_0}}$$

$$\boldsymbol{\Delta T_{HS}} = \mathrm{B} \cdot \boldsymbol{Rad^{*}_{TOA}} \cdot \mathbf{sin(2\theta_i)}$$

Distance between sun and view vectors:

$$f = \sqrt{\tan^2\theta_v + \tan^2\theta_i - 2\tan\theta_v\tan\theta_i\cos\Delta\phi}$$

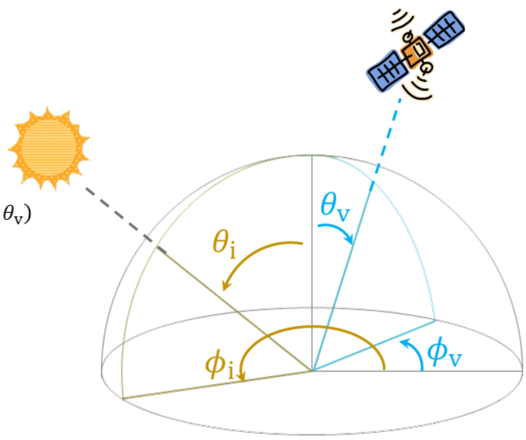

Figure 12: Parametric Models to be considered in the modelling of LST directional effects.

In both methodologies, the parametric models are calibrated using collocated LST products in space time, namely LST estimates from geostationary platforms (e.g., MSG, GOES) collocated with LST values retrieved from polar-orbiters (e.g. MODIS, SLSTR). The retrieval algorithms and auxiliary input data used to derive those datasets must be as close as possible, in order to ensure that the differences being modelled are indeed due to the different viewing geometries. The parametric models are then calibrated by pixel categories, classified via a cluster analysis of land-cover type (based on LC_cci), vegetation seasonality and pixel height. The validation of the methodology is based on in situ observations (limited to sites where ground measurements allow inferring LST from different view angles, [RD-167]) and by inter-comparison of LST products (as shown in [RD-42]).

Stage 5: GEO + LEO Merging

There are different approaches to dealing with superposition of pixels from the individual data streams. One could simply average these, or take the "best" pixel only. The "best" pixel could be determined either from the uncertainty estimates or more simply from the lowest satellite zenith angle.

For the GEOs the principle area of overlap is over South America. In this case all data will be taken from only one instrument to minimise possible discontinuities. For LEO vs. GEO superposition it is proposed that the selection criterion will be founded on the respective uncertainties of the different pixels. Since algorithm, cloud masking and anisotropic differences have been accounted for any residual discontinuities are expected to be minimised.

Experimental 1-year merged IR-MW LST demonstration

We will utilise the Level-2 data from the Merged IR CDR for 1-year (nominally chosen as 2010 - this being the prototype year in the Merged IR CDR in Processing Cycle 1) and Level-2 SSM/I data to combine into an experimental all-sky product. This will use information on the differences between the products from the validation and intercomparison activities to correct in LST space, to first order, the least accurate product. We will also utilise knowledge acquired from previous and ongoing projects on the relationship between LST and LSAT to improve our understanding of the clear-sky bias for an experimental product from IR and MW data.

5.2.4.2. ATSR-SLSTR CDR

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-61].

To fill the gap between the end of the Envisat mission and the end of the Phase E1 commissioning of the Sentinel-3A mission requires an instrument of equivalent spatial resolution with a LECT (Local Equatorial Crossing Time) 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 described in detail below, and schematically represented Figure 13. 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. Note, we also propose to deliver a first CDR from the ATSR-2 and AATSR instruments in Cycle 1 to allow sufficient time for evaluation by the climate User Case Studies.

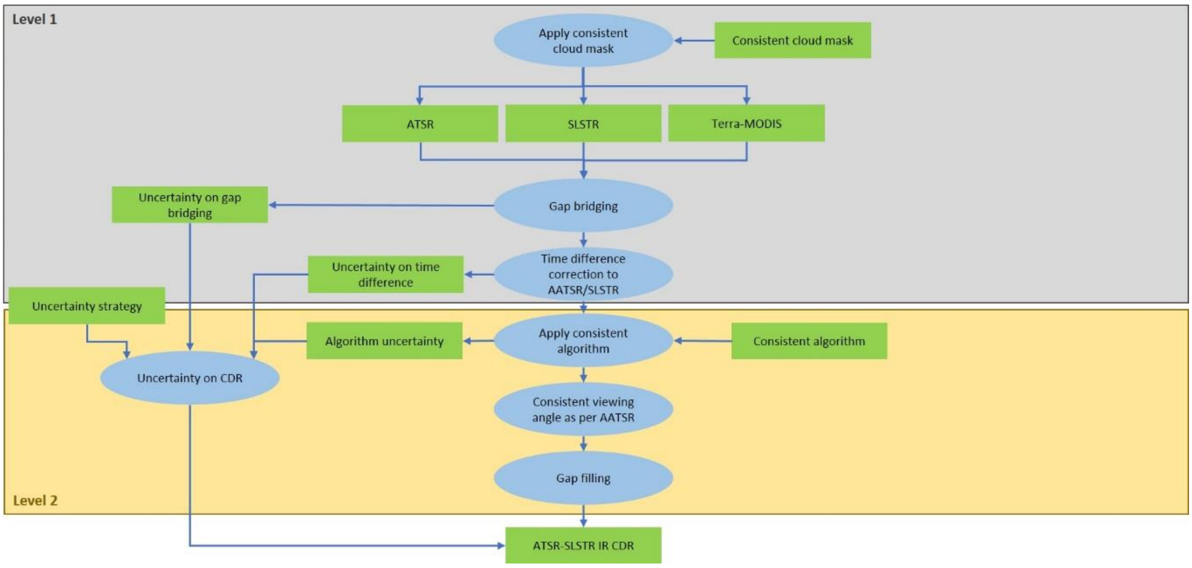

Figure 13: Schematic representation of the development and implementation of the ATSR-SLSTR CDR.

<u>Stage 1: gap bridging</u>

A first step for bridging the gap between the end of Envisat and the start of routine data availability from Sentinel-3 is to intercalibrate the radiances between AATSR and SLSTR. This necessarily also includes intercalibration of the other sensors in the CDR: ATSR-2 and Terra-MODIS.

Two approaches are to be used for intercalibrating the radiances between AATSR, Terra-MODIS and SLSTR:

- Utilising matched observations from in situ Fiducial Reference Measurements (FRMs):
  - These will bridge the data gap between Envisat and Sentinel-3A and provide a key reference of traceable observations
  - The site at Gobabeb, Namibia will be a key source of FRM data
  - The radiometry here has been traced to SI standards as part of the FRM4STS project [RD-54]:
    - Laboratory-based comparison of calibration processes for the participating radiometers
    - Laboratory-based comparison to verify the blackbody sources
    - Following best practice protocols
    - Full data characterisation of uncertainties
  - The application of these FRMs in the field are demonstrated on the LST Field Inter-comparison Experiments (FICE) [RD-121]
- Generating simultaneous nadir overpass (SNO) matchups between each of AATSR, Terra-MODIS and SLSTR, and a reference sensor:
  - IASI is the chosen reference sensor which is a GSICS reference sensor due to the stability and characterisation [RD-70]
  - The matchup approach can be summarised as follows:
    - Generate AATSR vs. IASI, SLSTR vs. IASI, and Terra-MODIS vs. IASI SNOs at high latitudes (Greenland and Antarctica) and where possible elsewhere
    - Minimise for spatial, temporal and geometry differences

- Convolve IASI radiance spectra with AATSR, SLSTR and Terra-MODIS spectral response functions
- Aggregate AATSR, SLSTR and Terra-MODIS pixels onto IASI FOV

- The calibration coefficient derivation [RD-49] is illustrated in Figure 14, which shows how the difference between the operation target radiance and corrected radiance (L_{bias}) is related to the newly calculated calibration coefficients (a_g, b_g) that are applied to the target instrument digital counts (C_{target}).

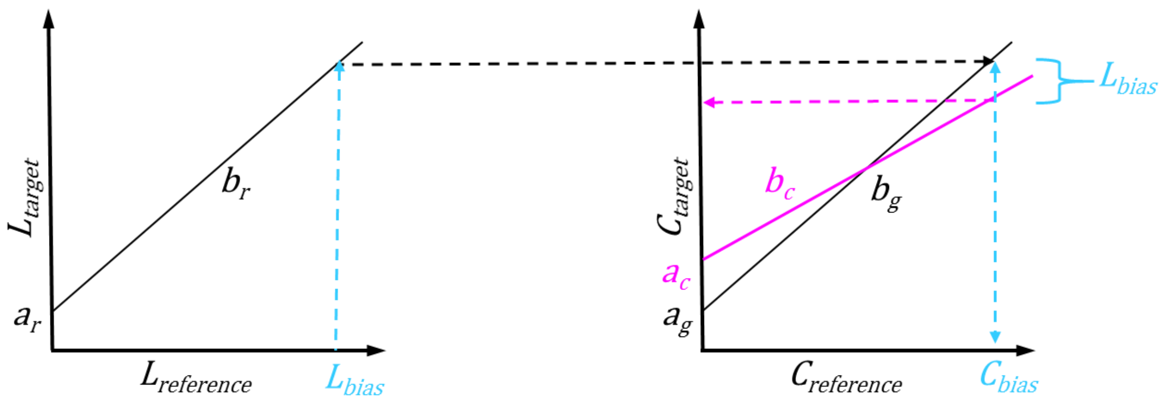

Figure 14: Adaptation of figure from [RD-49] showing the relationship between radiances observed by the target instrument (Ltarget) and those observed by reference instrument (Lreference).

A motivation for utilising both approaches is firstly that it permits more than one independent assessment of the radiances, and secondly that a wider scope of the dynamic range of the instruments can be assessed. Due to the strict 5-minute temporal threshold for SNO matchups as recommended by GSICS [RD-70], matchups between IASI and AATSR / Terra-MODIS / SLSTR will be principally in the mid-to-high latitudes and therefore confront the intercalibrations at the low to mid temperatures of the dynamic range. The use of FRMs from Gobabeb permit intercalibrations at the mid to high temperatures.

For harmonisation of ATSR-2 we will utilise existing harmonisation information developed in the ATSR Reprocessing for Climate (ARC) project for SST [RD-113].

Stage 2: time difference correction

A key consideration in the derivation of a continuous CDR from ATSR-2 through to SLSTR is to harmonise the temporal differences between the instruments. This difference is more complex than for SST with the impact on surface temperature due to the temporal offset in observation times an order of magnitude greater than radiometric calibration and spectral filter response variations. The 30-minute difference in LECT between Envisat and Sentinel-3, and ERS-2 and Terra has significant implications for stability of the long-term LST CDR at the levels required by GCOS.

To correct ATSR-2 and Terra-MODIS to the common reference LECT of AATSR and SLSTR it is crucial to characterise the diurnal information for the different land covers encountered for a global mission. In extreme cases, such as in arid environments, a 30-minute difference between observations during local mid-morning can result in several kelvin in surface temperature Figure 15 [RD-74].

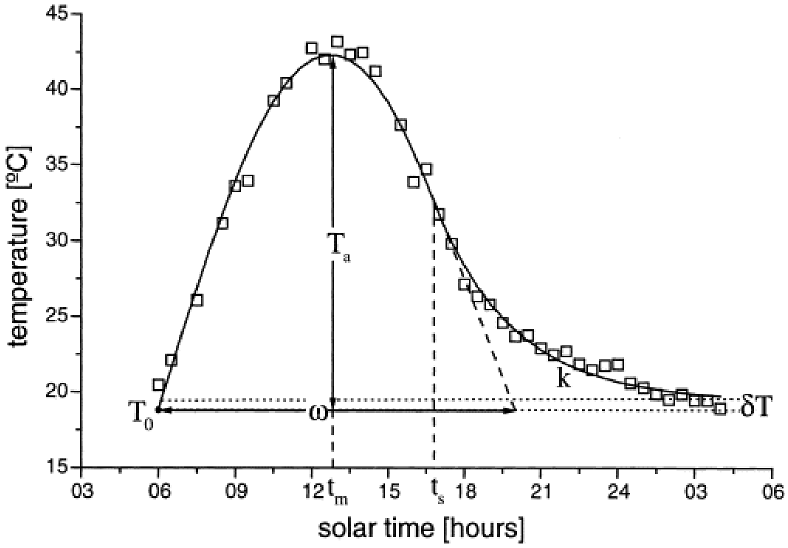

Figure 15: Plot from [RD-74].

Examination of AATSR-ATSR2 differences in surface temperature as a result of the 30-minute difference in LECT, performed for the GlobTemperature Prototype ATSR CDR [RD-73], shows that these differences are primarily correlated with the land surface class. The largest differences are found for arid and semi-arid regions. For daytime (Figure 16) these differences are predominantly negative since the 30 minute temporal difference between the earlier AATSR and later ATSR-2 coincides with the local morning warming; the opposite is true for night-time.

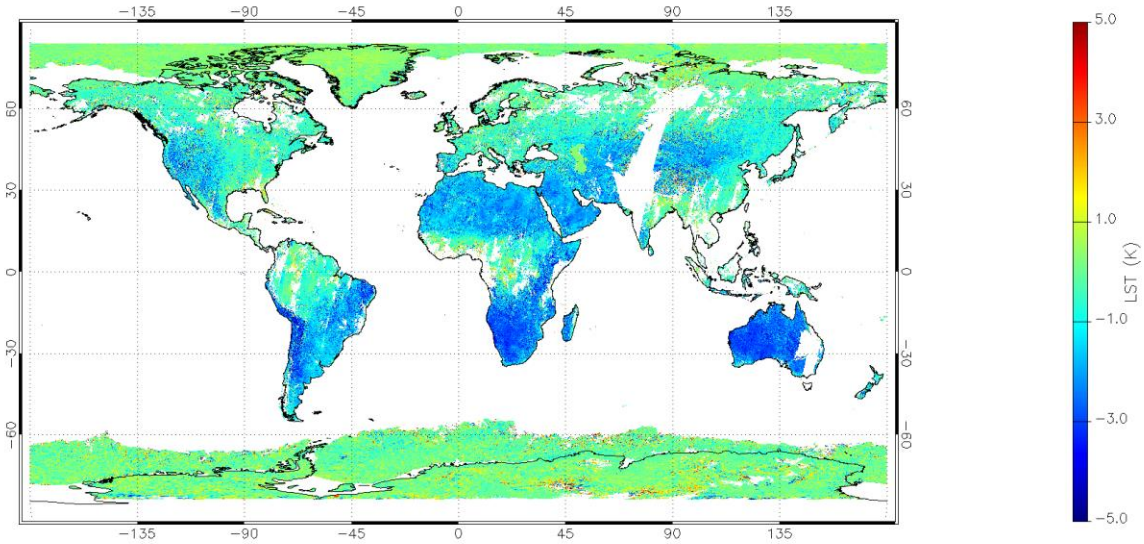

Figure 16: AATSR – ATSR2 differences from matchups during September 2002 used in correcting CDR LST data from ATSR-2. This does not show differences for "lcc" classes per block or globally for missing data.

To derive a dataset of corrections to apply to one or more of the input datasets to alleviate the problem of combining time series that have a 30 minute difference in LECT we will utilise two sources of diurnal information:

- An LST overlap analysis between ATSR-2, AATSR, SLSTR and a reference sensor during the overlap period of the respective missions (in this case Terra-MODIS).
- Diurnal information from the high temporal resolution geostationary satellites for each land cover class

The overlap analysis will be performed separately for day and night retrievals based on the solar zenith angle of the input Level-2 data. Only months where full records of both sensors in the analysis are to be processed. Only pixels over land and where BTs exists with high confidence, and are cloud-free according to the consistent algorithm from WP2.4, are to be included in the overlap analysis.

Stage 3: apply consistent Level-2 algorithms

The best performing algorithm from the Round Robin exercise (WP2.1) will be applied to the intercalibrated Level-1 data to ensure consistency from an algorithm perspective across the ATSR-SLSTR CDR. The best performing cloud mask from WP2.4 developments and assessment will be applied to the Level-1 data for the individual sensors to ensure consistency.

Stage 4: restrict satellite viewing angle

To maintain consistency of the CDR from ATSR-2 through to SLSTR in terms of data coverage and to minimise anisotropic differences in the data record as we switch from one instrument to another the satellite zenith angle (SZA) will be restricted to the lowest common denominator. In the case, both ATSR-2 and AATSR have a maximum SZA of approximately 22°. Thus, only MODIS and SLSTR pixels with SZAs less or equal to this threshold will be processed through the CDR chain.

Stage 5: gap filling

The gap filling process consists of the production of the CDR using the appropriate data stream for a given day. The modularisation of the LEO processing chains, for which we will enhance the current systems developed for the Prototype ATSR CDR in GlobTemperature [RD-73], allows for the ingestion of multiple satellite Lcvel-1 data through module plug-ins.

Since a key temporal interval is expected to be monthly data, the switch between sensors will occur at the monthly interchange. So the following sensors are to be the input over the full CDR time window:

- ATSR-2: 08/1995 to 07/2002
- AATSR: 08/2003 to 03/2012AATSR
- Terra-MODIS: 04/2012 to 07/2016
- SLSTR: 08/2016 to 12/2020

Note, we propose the full period from ATSR-2 even though this means incompletely observed months. This includes a missing period of 6 months from January 1996 to June 1996 inclusive due to a scan mirror failure, and a further period from January 2001 to June 2001 inclusive due to a gyro failure. Further periods lasting up to a few days occur throughout the data record due to instrument decontaminations or anomalies. Months where these occur are not excluded from the data record.

A final point to note here is that ATSR-2 suffers from a systematic data gap over central Asia. The data for these missing parts of the orbit is not recoverable since it was never downlinked from the satellite. This is in comparison with the full global coverage for AATSR (Figure 17).

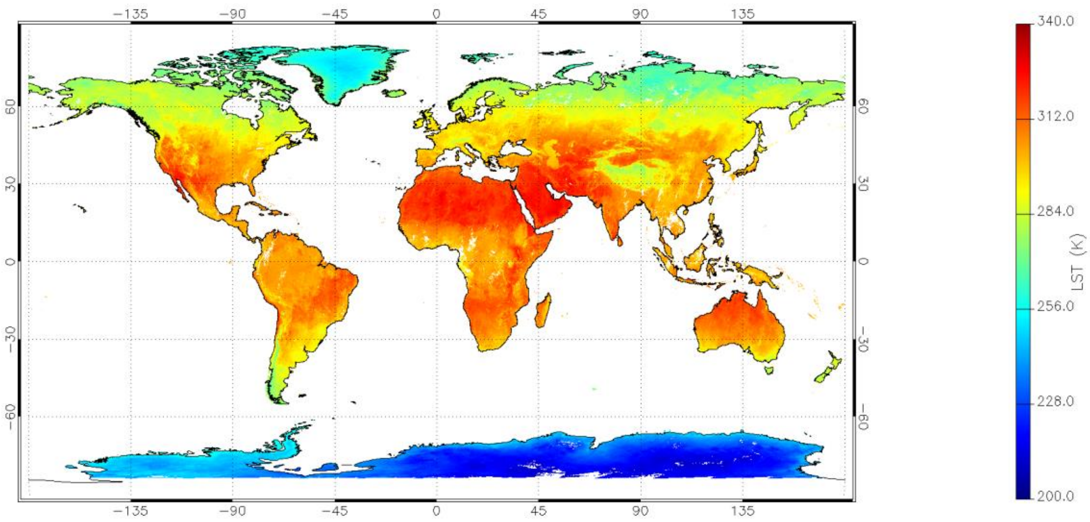

Figure 17: Example images from the GlobTemperature Prototype LST CDR for September 2002 from AATSR.

5.2.4.3. CDR Processing System

For the Merged IR CDR we propose a distributed system between UK JASMIN and IPMA. The merging of the GEO data will be carried out on the IPMA infrastructure taking full advantage of the ease of access to the Level-1 GEO data and an optimised processing chain built on the foundations of the operational GEO infrastructure for the LSA SAF and the Copernicus Global Land Services LST Products. The outputs of these will be transferred to the shared project workspace on UK JASMIN/CEMS for which all project partners will have full access to.

This fundamental of this proposed system has been demonstrated in GlobTemperature as part of the Merged NRT Product. In this demonstration, automated transfer of the Merged GEO Product is supported from a dedicated FTP server at IPMA onto the ULeic ALICE system. Here, we propose to follow this baseline concept and demonstrate a system where the Merged GEO Products are transferred to UK JASMIN for the final merging of the LEO and GEO data taking advantage of existing processing infrastructure from pre-cursor projects and access the LEO Level-1 data stream on the CEDA Archive.

As for the above IR processing chains, the LEO and GEO merging is performed on Linux-based systems, with all processing parallelised on these systems. All software codes for the existing processing chains are version controlled, which are to be maintained for LST CCI. All GEO and LEO processing code for the Merged IR CDR will be Open Source with the final version to be added to the common CCI GitHub code repository. Full access to the relevant workspaces for the merging processes will be provided to the System Engineering Team who will be deploy and monitor the processing with verification tools.

For the period 2009 – 2020, we propose a Merged IR CDR product of high scientific value, using the identified input data stream. The processing of the implemented algorithms will be distributed between UK JASMIN and IPMA (Figure 18). This Figure also shows the processing chain for the ASTR-SLSTR CDR. This latter concept is designed to meet the GCOS climate requirement [RD-61] for >20 years of consistent satellite-based LST estimates

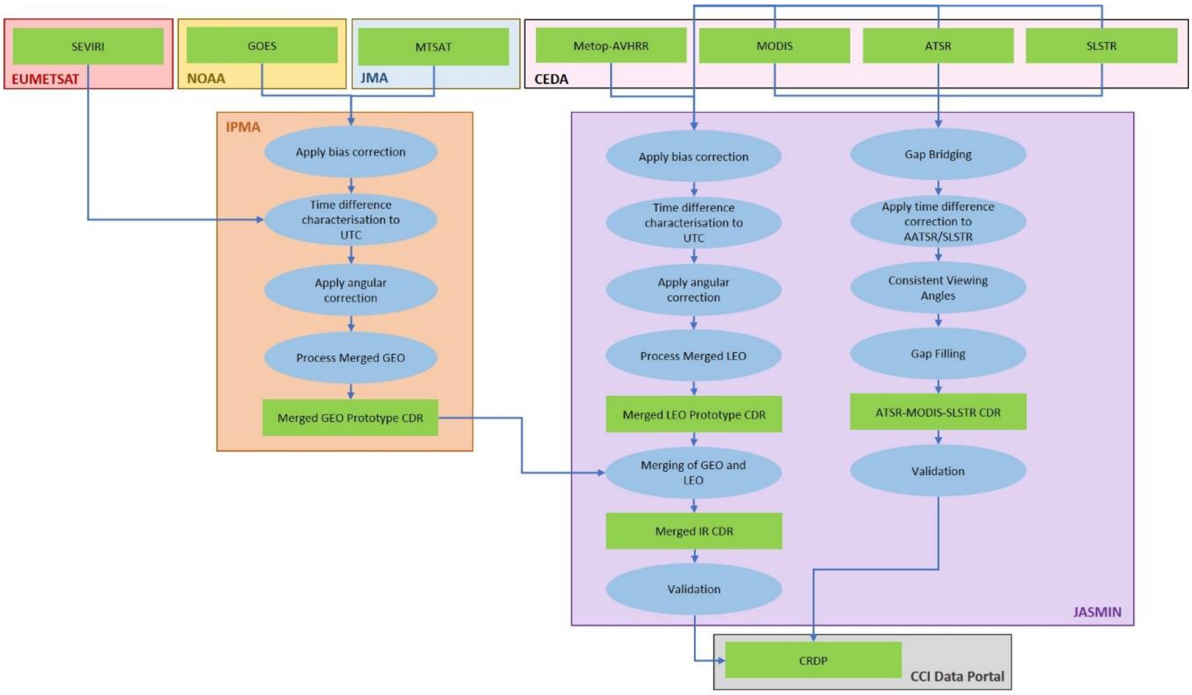

Figure 18: Data flows for the Merged IR CDR system and ATSTR-SLSTR CDR.

The ULeic processing toolbox is modular in design with functionality for data merging, extraction and aggregation. It provides the capacity to create L3C products of any user required spatial and temporal resolution from the base L2 products via the intermediate L3U datafiles.

These prototype systems will be composed of both pre-existing and new components building effectively on existing software tools and libraries which have proven maturity in pec-cursor and complementary projects [RD-47; RD-106; RD-29]. This approach maximises cost-effectiveness and is the basis for operational readiness targets in Phase-II.

Addresses Requirements: LST-SYS-REQ-01-SOW, LST-SYS-REQ-05-SOW, LST-SYS-REQ-06-SOW, LST-SYS-REQ-08-SOW, LST-SYS-REQ-15-SOW, LST-SYS-REQ-19-URD

5.2.5. Verification activities in production

The verification step in the processing chain ensures that all data products are CF compliant and that the NetCDF files contain the required variables and metadata (variable and global attributes). It also ensures that the variable values lie in the valid range. For cost-effectiveness, we will use components inherited from previous projects. Such software was developed for the verification of the GlobTemperature formatted products and can easily be adapted for verification of products in CCI.

The verification process involves a number of steps. The first step takes place during development of the processing chain. Product files are checked for CF compliance using an online checker such as "CF-Convention Compliance Checker for NetCDF Format" developed by the Hadley Centre for Climate Prediction and Research, UK Met Office and available at http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl. At this stage, any non-compliance can be corrected in the processor. When CF compliance is met a script is run to produce a template that can be used for the testing of all the files of the output dataset.

Then, in the post-production phase, the main verification process is performed by a Python program which checks each file against the template and produces a report for each file detailing any errors. Any corrupted files are reported as are any LST files that are missing a matching AUX file, or vice versa. Checks are made for the presence of each variable listed in the template. Both the absence of required variables and the presence of non-required variables are reported. The variable attributes, both the tag names and the attribute values are checked, again any discrepancies are reported. For each variable, the variable values are tested to see if they lie within the valid range given by the valid minimum and valid maximum attributes. Finally, the global attributes are checked, again any errors in tag names or attribute values, any missing or superfluous attributes are reported.

The final step in the verification is the analysis of the individual file reports. This step produces a summary report giving the number of files processed through the verification, the type and number of errors found, and lists of those files containing each type of error. Thus, any errors can easily be investigated and the products reprocessed and the verification process repeated until a complete and clean dataset is ready for dissemination to users. In addition, regular visual checks will be done by the EO team to verify that output products have no major issues.

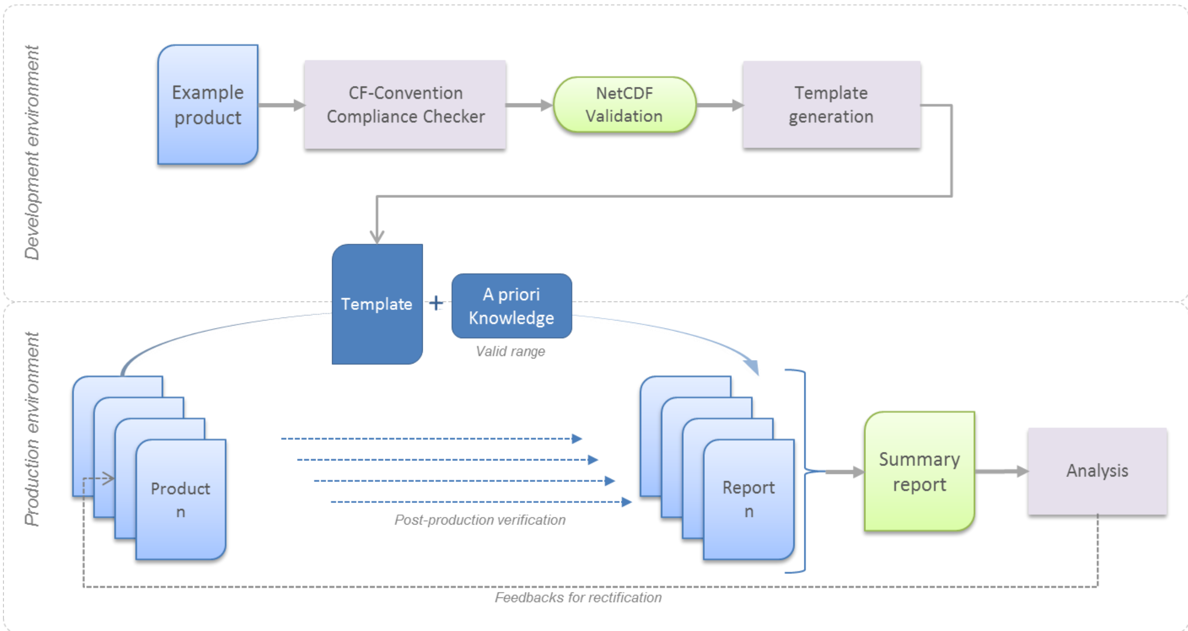

Figure 19: Flow chart of metadata verification.

Addresses Requirements: **LST-SYS-REQ-04-SOW**

6. Detailed System Requirements

The requirements applicable to the processing system which have been listed in Table 2, Table 3 and Table 4 are further described in this section.

Figure 20 displays all requirements classified per category defined in section 4. This figure shall be read as follow:

- The origin of the requirement is indicated by a colour code
  - Requirements from the SOW are in 'brown'
  - Requirements from the URD are in 'red'
  - Requirements from the CCI data requirements are in 'blue-gray'
- The dashed gray arrows indicate when 2 requirements relates which each other
  - It is noted that the overlap between 2 linked requirements may be very important
  - We could have then decided to merge them to reduce the overall number of requirements
  - However, in to keep the traceability with the origin of the requirement, we have decided not to merge these "similar" requirements

All these requirements are summarised in Table 5 in which we have indicated:

- The ID
- The title
- The description
- The link with other requirement(s) (if any)
- The verification mean

Concerning the verification, three different aspects are foreseen to assess that the requirement has been correctly implemented:

- **Review:** consists of reviewing the documentation associated with the system or the sub-system in charge of fulfilling the function(s) described in the requirement
- **Test:** consists of performing tests at different levels (unitary tests or end-to-end tests) done before the processing system is operationally deployed – an important step is the rehearsal during which the finalised version of the system is tested in operational conditions.
- **Monitoring:** consists of regularly verifying, when the system is operating, that the requirement is still being met – this will be particularly important for the evolution of algorithms or ingestion of new datasets

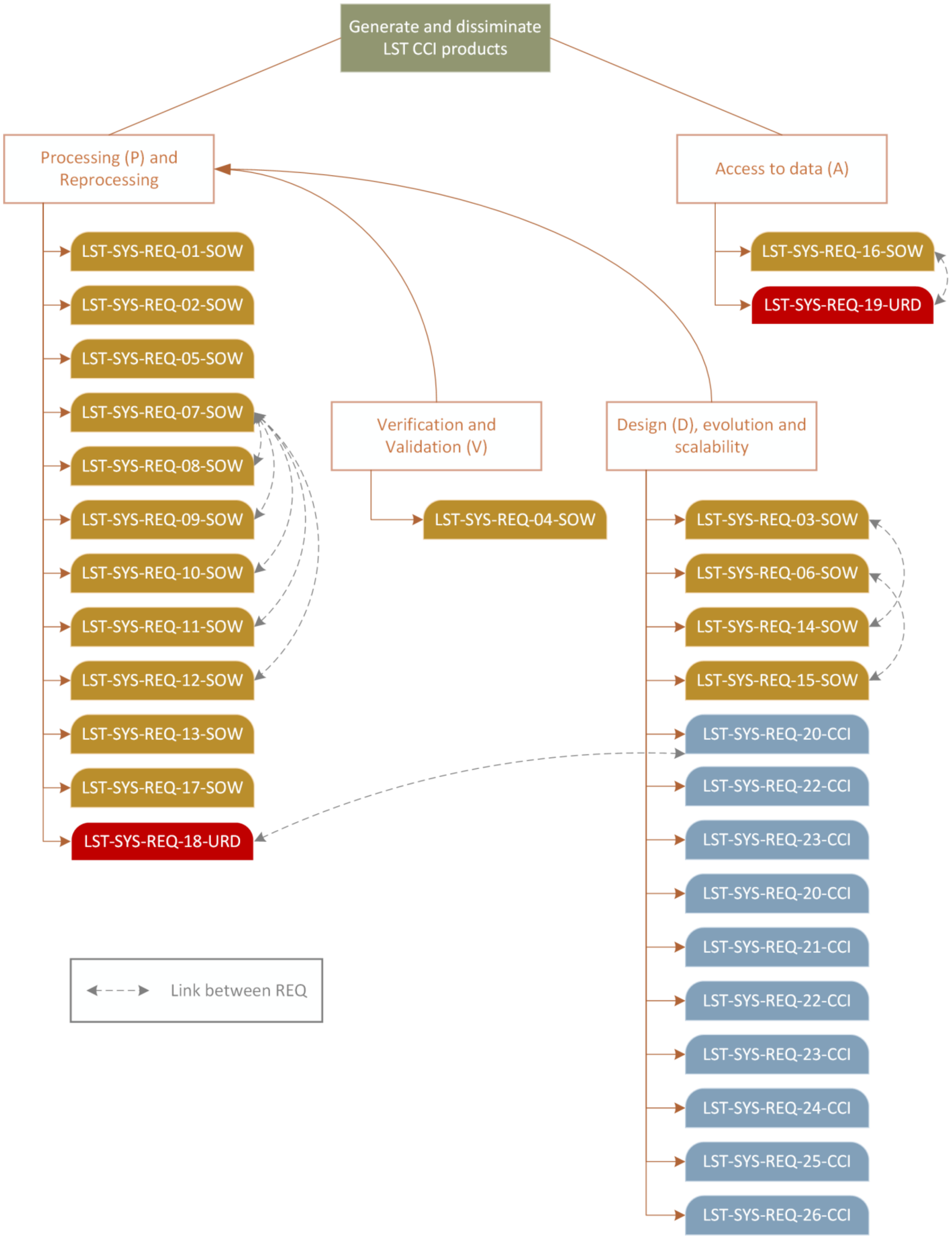

Figure 20: LST CCI processing system requirements

Table 5: List of all requirements

Id	Title	Description	Relates to REQ(s) (if any)	Verification
LST-SYS-REQ-01-SOW	Production of time-series	The LST CCI processing system shall generate all LST ECV products as requested in the SOW		Review Test
LST-SYS-REQ-02-SOW	Building of processing system	The LST CCI processing system shall be built and documented in order to fulfil the needs for processing and reprocessing		Review
LST-SYS-REQ-03-SOW	Legacy of CCI projects	The LST CCI processing system shall make use, as much as possible, of the best practises and existing developments of CCI projects	LST-SYS-REQ-14-SOW	Review
LST-SYS-REQ-04-SOW	System adaptation to the production of LST CCI products	The LST CCI processing system shall be scalable in order to include new datasets and new processing chains		Review Test
LST-SYS-REQ-05-SOW	Integration of Copernicus and key satellite missions	The LST CCI processing system shall be able to include Copernicus Sentinel missions and other key satellite missions as inputs data		Review
LST-SYS-REQ-06-SOW	System sizing to account for large volume of data	The LST CCI processing system shall be able to process or reprocess large amount of data, with expecting growing volumes.	LST-SYS-REQ-15-SOW	Review Test
LST-SYS-REQ-07-SOW	Specifications of products to process	The LST CCI processing system shall process all data listed in the SOW	LST-SYS-REQ-08-SOW LST-SYS-REQ-09-SOW LST-SYS-REQ-10-SOW LST-SYS-REQ-11-SOW LST-SYS-REQ-12-SOW	Review
LST-SYS-REQ-08-SOW	Specifications of IR based LST products	The LST CCI processing system shall process all IR data listed in the SOW	LST-SYS-REQ-07-SOW	Review
LST-SYS-REQ-09-SOW	Specifications on NOAA AVHRR products	The LST CCI processing system shall process AVHRR products listed in the SOW	LST-SYS-REQ-07-SOW	Review

Id	Title	Description	Relates to REQ(s) (if any)	Verification
LST-SYS-REQ-10-SOW	Specifications on global merged products	The LST CCI processing system shall process global merged products as specified in the SOW, in particular over the period 1998-2020	LST-SYS-REQ-07-SOW	Review
LST-SYS-REQ-11-SOW	Specifications on passive MW based LST products	The LST CCI processing system shall process all MW data listed in the SOW	LST-SYS-REQ-07-SOW	Review
LST-SYS-REQ-12-SOW	Temporal Coverage of MW based products	The LST CCI processing system shall generate MW based products over the period 1998-2020	LST-SYS-REQ-07-SOW	Review
LST-SYS-REQ-13-SOW	DOI assigned to LST products	The LST CCI processing system shall assigned DOI to LST CCI products		Test
LST-SYS-REQ-14-SOW	Heritage from past and current projects	The LST CCI processing system shall build on past and current projects, in particular on GlobTemperature project heritage	LST-SYS-REQ-03-SOW	Review
LST-SYS-REQ-15-SOW	Sizing of the processing systems to account for large amount of data	The LST CCI processing system shall be able to process large amount of data	LST-SYS-REQ-06-SOW	Monitoring
LST-SYS-REQ-16-SOW	Availability of LST CCI products	The LST CCI processing system shall make available the LST CCI products, in particular on the CCI Data Portal	LST-SYS-REQ-19-URD	Test
LST-SYS-REQ-17-SOW	Acquisition of all EO and in-situ data	The LST CCI processing system shall ensure that all inputs data which are need for the processing are effectively available and accessible		Test
LST-SYS-REQ-18-URD	CCI standard format	The LST CCI processing system shall comply with CCI standard format described in the CCI product guidelines	LST-SYS-REQ-20-CCI LST-SYS-REQ-21-CCI	Test
LST-SYS-REQ-19-URD	Access to data	The LST CCI processing system shall ensure an easy access to the LST CCI products	LST-SYS-REQ-16-SOW	Test
LST-SYS-REQ-20-CCI	CCI data standards requirements	The LST CCI processing system shall comply with CCI standard format described in the CCI product guidelines	LST-SYS-REQ-18-URD	Review

Id	Title	Description	Relates to REQ(s) (if any)	Verification
LST-SYS-REQ-21-CCI	INSPIRE compliant metadata records	The LST CCI processing system shall comply with INSPIRE directive for the metadata associated to each product		Review Test
LST-SYS-REQ-22-CCI	Terms from CCI vocabulary tables used in netCDF	The LST CCI processing system shall use the same terms than those included in the CCI vocabulary tables in netCDF products		Review Test
LST-SYS-REQ-23-CCI	Identification of key variables	The LST CCI processing system shall ensure that the key primary variables in the file and their related ancillary variables are identified, and the range of their expected values is indicated.		Review Test
LST-SYS-REQ-24-CCI	Variables shall include at least time, longitude and latitude	The LST CCI processing system shall include, at least, these 3 variables in product: time, longitude, latitude		Test
LST-SYS-REQ-25-CCI	Use of common directory structure	The LST CCI processing system shall comply with structure proposed in the CCI data portal for the data		Test
LST-SYS-REQ-26-CCI	Application of CCI filenaming convention	The LST CCI processing system shall comply with the filename conventions defined by CCI		Review Test

End of document