EUMETSAT Contract No. EUM/CO18/4600002180/JML

EVOLUTIONS STUDIES AUTOMATED OPTICAL SENSOR REGISTRATION MONITORING TOOL

Output Data Format, Reference Database and Diagnostic Datasets

Deliverable D9/D10/D11

Prepared by

Stefan SCHEIBLAUER and Thomas NAGLER

ENVEO IT, Innsbruck, AUSTRIA

Document Reference: ENVEO-OSMON_D9/10/11.ODF_RDB_DDS
Issue / Revision: 1/3
Date: 02.11.2020

Document controlled by Stefan Scheiblauer

EUMETSAT STUDY CONTRACT REPORT

CONTRACT NO: EUM/CO18/4600002180/JML	**SUBJECT:** Deliverable D9/D10/D11 - Output Data Format, Reference Database and Diagnostic Datasets		**CONTRACTOR:** ENVEO
	STAR CODE:	**NO OF VOLUMES**: 1 **THIS IS VOLUME NO**: 1	**CONTRACTOR'S REF:** Deliverable D9/D10/D11

ABSTRACT:

This document covers three deliverables of the OSMON project, D9 – Output data format, D10 – Reference data set and D11 Diagnostic Datasets. Output data of the geolocation accuracy of medium resolution optical imagers are stored in the OSMON database. The OSMON data base is implemented in a relational data base system consisting of several linked tables. The format and entries of each table are explained and the links between the individual tables is documented. The reference data base is generated from high resolution optical satellite images using image templates covering prominent features, which can be clearly identified in medium resolution satellite data. Each GCP includes spectral bands acquired by Landsat-8, including visible, near infrared and thermal infrared bands. The electronic reference database is available at ENVEO and at the EUMETSAT server.

Additionally, we describe the contents of the diagnostic data set, which is a subset of 17 GCPs of the reference data set and the corresponding MROI satellite images, which are needed to develop, implement and test the OSMON database.

The work described in this report was done under EUMETSAT Contract. Responsibility for the contents resides in the authors or organization that prepared it.

AUTHORS:
STEFAN SCHEIBLAUER AND THOMAS NAGLER

EUMETSAT STUDY MANAGER: JOHANNES MÜLLER	**EUMETSAT BUDGET HEADING**

This page is intentionally left blank.

DOCUMENT CHANGE LOG

Issue/ Revision	Date	Modification	Modified pages	Observations
1.0	30.10.2019	All new	All	NA
1.1	9.12.2019	Comments by EUMESAT		NA
1.2	30.04.2020	Update section 1, Output Data Format (D9)	All	NA
1.3	02.11.2020	Update section 1 and Definition of Terms	1, 6, 7, 13, 15, 17	Include EUMETSAT RIDs.1.E dated 12.08.2020

This page is intentionally left blank.

TABLE OF CONTENT

APPLICABLE AND REFERENCE DOCUMENTS

[RD-1]	GDAL/OGR contributors (2019) GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation. Available at: https://gdal.org.
[RD-2]	ENVEO IT, (2019), Evolution Studies - Automated Optical Sensor Registration Monitoring Tool Minutes of Technical Discussion Data needs for Development and Demonstration, Dated 14.03.2019
[RD-3]	Landsat-8 Data User Handbook (V4.0, April 2019). https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1574_L8_Data_Users_Handbook_v4.pdf
[RD-4]	ENVEO IT, (2019), Evolution Studies - Automated Optical Sensor Registration Monitoring Tool Requirements Baseline Document, Deliverable D05 ENVEO-OSMON_D05.RBD_1/3
[RD-5]	NASA, 2003. SRTM Water Body Data Product Specific Guidance. Version 2.0. 4pp
[RD-6]	ENVEO IT, (2019), Evolution Studies - Automated Optical Sensor Registration Monitoring Tool Algorithm Theoretical Basis Document, Deliverable D08 ENVEO-OSMON_D08.ATBD_1/3
[RD-7]	Kalnay, Eugenia, et al. "The NCEP/NCAR 40-year reanalysis project." Bulletin of the American meteorological Society 77.3 (1996): 437-472

Note: If not provided, the reference applies to the latest released Issue/Revision/Version

ACRONYMS

AVHRR	Advanced Very High Resolution Radiometer
DW	Data Window
GCP	Ground Control Point
MetOp	Meteorological Operational Satellite
MROI	Medium Resolution Optical Imager
NIR	Near Infrared
OLCI	Ocean and Land Colour Instrument
OLI	Operational Land Imager
SLSTR	Sea and Land Surface Temperature Radiometer
SWIR	Shortwave Infrared
TIR	Thermal Infrared
TIRS	Thermal Infrared Sensor
VIS	Visible

DEFINITION OF TERMS

GCP	In the OSMON project the term Ground Control Point (GCP) refers to an image template covering a prominent feature on the Earth Surface, which can be identified in medium resolution optical images. GCP image templates are extracted from VIS, NIR, SWIR, TIR spectral bands of Landsat-8 OLI and TIRS data. The geographic location of a GCP is exactly specified by the geographic map projection, ellipsoid and local datum, corner coordinates and pixel spacing in Easting and Northing direction.
MROI	Medium Resolution Optical Imagers is used as a general name of medium resolution optical satellites such as Sentinel-3 SLSTR, Sentinel-3 OLCI and MetOp/AVHRR.
MROI-DW	This term is used for MROI-satellite data covering the same area as the corresponding GCP.MROI-DW are raster files in the same map projection, ellipsoid and local datum, corner coordinates and pixel spacing in Easting and Northing direction. To achieve the same pixel spacing as GCP, oversampling of the native resolution of the MROI data is applied.
RasterBandSet	This term is used to refer to a subset of a product covering one GCP. It is a stack of bands cropped to the buffered envelope of the GCP and holds information applicable to all bands such as solar angles and viewing angles as well as percentage of snow/cloud coverage.

1. OUTPUT DATA FORMAT (DELIVERABLE D9)

In this section the output data format of the OSMON database is described. The OSMON database is implemented as a relational database supporting economical management, handling and selection of results and input reference data [RD-4]. The database is setup in PostgreSQL using the extension module PostGIS for efficient handling of geoinformation using the Django framework .

Figure 1 presents an overview of the relational OSMON database, showing the links between the different database tables. The attributes of each of the tables are described in detail in the following section, where keys in blue colour reference an other database table.

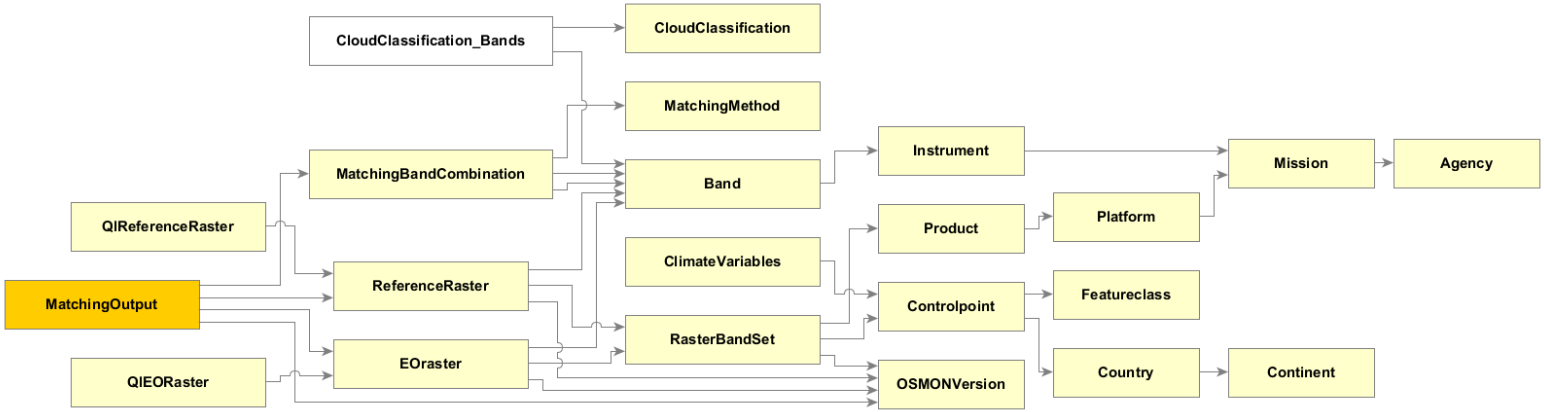


Figure 1: Data model of the OSMON database, showing the relation between database tables. *Note:* The table *CloudClassification_Bands* is created automatically by the Django Framework to handle the Many-to-many relations between the *Bands* table and the *CloudClassification* table.

1.1. OSMONVersion

Table holding the release notes of the OSMON software package. This table is created and filled during the OSMON setup process.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of software package	-	OSMON Configuration file	The software package name is defined in the OSMON configuration file.
major	integer	Software version major number	-	OSMON Configuration file	The software version major number is defined in the OSMON configuration file.
minor	integer	Software version minor number	-	OSMON Configuration file	The software version minor number is defined in the OSMON configuration file.
release_date	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of software release	-	OSMON Configuration file	The release date is defined in the OSMON configuration file.

Name	Data type	Description	Unit	Source	Comment:
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.2. Agency

Table holding the name of the agency providing the data.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of the agency providing the data	-	Derived	Derived from auxiliary data and metadata of the product
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.3. Mission

Table holding the mission name and agency operating the mission.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of the satellite mission.	-	Derived	Derived from auxiliary data and metadata of the product
agency	foreign key	Agency providing the data	-	Derived	References to: Agency
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.4. Platform

Table holding the platform and the mission it belongs to.

Name	Data type	Description	Unit	Source	Comment:
id	Integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of the platform.	-	Derived	Derived from auxiliary data and metadata of the product
mission	foreign key	Name of the satellite mission the platform belongs to.	-	Derived	References to Mission
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.5. Instrument

Table holding the description of the instrument acquiring the data as well as the mission employing the instrument. *Note: The abbreviated name 'acronym' must not contain the special character '/'.*

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Full name of the instrument collecting the data	-	Derived	Derived from auxiliary data and metadata of the product
acronym	character varying(32)	Abbreviated name of the instrument collecting the data.	-	Derived	Derived from auxiliary data and metadata of the product
mode	character varying(32)	Acquisition mode of instrument	-	Derived	Derived from auxiliary data and metadata of the product
mission	foreign key	Name of the satellite mission the platform belongs to.	-	Derived	References to Mission
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.6. Band

Table holding the dependence between band names and the instruments acquiring the data. *Note: Band name must not be a string of integers.*

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of the band	-	Derived	Derived from auxiliary data and metadata of the product
band_type	integer	Enumerated list of band types (visible, thermal, ...)	-	OSMON Configuration file	Definition of band type codes in 1.22.1
instrument	foreign key	Instrument collecting the data	-	Derived	References to Instrument
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.7. Product

Table representing the L1 product. Note: Product information not available in the metadata or auxiliary file is indicated by the string *NA = not available*.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Filename of the dataset	-	Derived	Derived from auxiliary data and metadata of the product
product_type	character varying(16)	Type of product	-	Derived	Derived from auxiliary data and metadata of the product
platform	foreign key	Platform carrying the product generating instrument	-	Derived	References to Instrument
acquisition_time	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of scene center	-	Computed	Mean acquisition timestamp computed form sensing start and stop time.
footprint	MultiPolygon	Footprint of satellite product	-	Derived	Derived from auxiliary data and metadata of the product
proc_time	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of product processing end time	-	Derived	Processing stop time of output level 1, derived from auxiliary data and metadata of the product
proc_facility	character varying(256)	Name of processing facility	-	Derived	Name of the processing facility, derived from auxiliary data

Name	Data type	Description	Unit	Source	Comment:
					and metadata of the product
proc_environment	character varying(256)	Name and version of the processing environment used to generate the L1 product.	-	Derived	Derived from auxiliary data and metadata of the product
proc_hardware	character varying(256)	Name and version of the processing hardware used to generate the L1 product.	-	Derived	Derived from auxiliary data and metadata of the product
proc_software	character varying(256)	Name and version of the processing software used to generate the L1 product.	-	Derived	Derived from auxiliary data and metadata of the product
osmon_version	foreign key	Version of the OSMON package	-	Derived	References to OSMONVersion
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.8. Featureclass

Table holding the type of imaged feature such ass coastline segment, lake segment, island, lake, reservoir or geological feature. The feature class is an attribute of the shapefile 'gcp.shp' and inserted to the table, created during the OSMON setup process.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Type of imaged feature	-	File: gcp.shp	Derived from attribute of shapefile containing GCPs

1.9. Continent

A table containing the names of the continents. This table is created and filled during the OSMON setup process.

Name	Data type	Description	Unit	Source	Comment:
id	Integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of continent	-	File: gcp.shp	Derived from attribute of shapefile containing GCPs

1.10. Country

A table containing the names of the countries. This table is created and filled during the OSMON setup process.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of country	-	File: gcp.shp	Derived from attribute of shapefile containing GCPs
continent	foreign key	Continent the country is located on	-	Derived	References to Continent

1.11. MatchingMethod

The table contains the implemented and available matching methods. This table is created and filled during the OSMON setup process.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of matching method	-	OSMON Configuration file	The name of the matching method is defined in the OSMON configuration file.
function	character varying(256)	List of image pre-processing functions.	-	OSMON Configuration file	Functions for the simple pre-processing steps are defined as Python expressions in the OSMON configuration file.
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.12. Controlpoint

A table representing the Ground Control Points (GCPs) and related geospatial information.

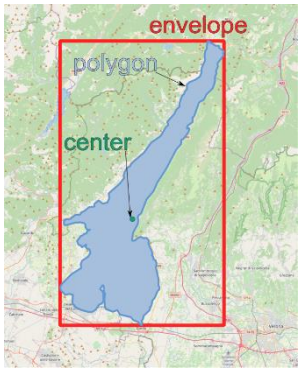


Figure 1.2: Definition of the envelope, polygon and center entries in the Controlpoint table, which are given in WGS-84 geographic (latitude, longitude) coordinates.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of GCP	-	File: gcp.shp	Derived from attribute of shapefile containing GCPs
cpid	character varying(32)	Unique GCP identifier derived from GCP location	-	File: gcp.shp	Derived from attribute of shapefile containing GCPs
featureclass	foreign key	Feature class of imaged GCP	-	Derived	References to Featureclass
country	foreign key	Country where the GCP is located	-	Derived	References to CountryFeatureclass
elevation	integer	Elevation of center point in respect to WSG84 ellipsoid [m]	-	Computed	Computed by intersecting GCP geometry with GETASSE30 Elevation Model
center	Point	Geographical location of GCP-feature center (WGS84)	-	Computed	Computed from GCP geometry

Name	Data type	Description	Unit	Source	Comment:
envelope	Polygon	Envelope of GCP-feature (WGS84)	-	Computed	Computed from GCP geometry
polygon	Polygon	Geometry of GCP-feature (WGS84)	-	File: gcp.shp	Geometry representing the GCP in the shapefile containing GCPs
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.13. ClimateVariables

A table holding climatological data linked to a Ground Control Point. The monthly minimum temperature at 2 m (K) is derived from NCEP reanalysis data [RD-7].

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
gcp	foreign key	GCP the climatological data relates to	-	Derived	References to Controlpoint
tmin_jan	real	Minimum 2m temperature in January	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_feb	real	Minimum 2m temperature in February	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_mar	real	Minimum 2m temperature in March	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_apr	real	Minimum 2m temperature in April	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_may	real	Minimum 2m temperature in May	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_jun	real	Minimum 2m temperature in June	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_jul	real	Minimum 2m temperature in July	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]

tmin_aug	real	Minimum 2m temperature in August	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_sep	real	Minimum 2m temperature in September	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_oct	real	Minimum 2m temperature in October	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_nov	real	Minimum 2m temperature in November	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
tmin_dec	real	Minimum 2m temperature in December	[K]	Computed	Computed by intersecting GCP geometry with NCEP reanalysis data [RD-7]
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.14. RasterBandSet

Table representing a stack of bands covering one GCP referred to as RasterBandSet which is a subset of a product. The table holds information applicable to all bands of the RasterBandSet such as solar angles and viewing angles as well as percentage of snow/cloud coverage.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
gcp	foreign key	GCP covered by the RasterBandSet	-	Derived	References to Controlpoint
product	foreign key	Product containing the RasterBandSet	-	Derived	References to Product
track_number	integer	Track number.	-	Derived	Derived from auxiliary data and metadata of the product
acquisition_time	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of GCP center	-	Computed	Acquisition timestamp computed for the GCP center
output_directory	character varying(1024)	Path to the storage location of the RasterBandSet	-	OSMON Configuration file	Path to the storage location of the RasterBandSet
sun_elevation_angle	real	Sun elevation angle at the GCP center	[°]	File: Geometry_t[no].nc	See ATBD, [RD-6]
sun_azimuth_angle	real	Sun azimuth angle at the GCP center	[°]	File: Geometry_t[no].nc	See ATBD, [RD-6]
view_elevation_angle	real	Viewing elevation angle at the GCP center	[°]	File: Geometry_t[no].nc	Automatically computed during import from satellite angles

view_azimuth_angle	real	Viewing azimuth angle at the GCP center	[°]	File: Geometry_t[no].nc	Automatically computed during import from satellite angles
heading_angle	real	Heading angle of groundtrack segment	[°]	Calculated	See ATBD, [RD-6]
cloud_coverage_pct	integer	Percentage of snow/cloud cover of RasterBandSet	[%]	Calculated	See ATBD, [RD-6]
cloud_class	foreign key	Applied cloud classification schema	-	Derived	See ATBD, [RD-6]
error_flag	integer	RasterbandSet error code	-	OSMON Configuration file	Definition of RasterbandSet /Error codes in 1.22.2
osmon_version	foreign key	Version of the OSMON package	-	Derived	References to OSMONVersion
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.15. ReferenceRaster

Table holding related information about a single band of a geocoded control point raster dataset (e.g. top of atmosphere reflectance of band 'B5') of a GCP.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
band	foreign key	Band stored in the ReferenceRaster.	-	Derived	References to Band
rasterband_set	foreign key	Parent RasterBandSet the ReferenceRaster belongs to	-	Derived	References to RasterBandSet
nodata_pct	integer	Percentage of nodata values within the ReferenceRaster	[%]	Calculated	See ATBD
raster_filename	character varying(256)	Basename of the reference raster	-	Calculated	Band name + '.tif'
osmon_version	foreign key	Version of the OSMON package	-	Derived	References to OSMONVersion
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.16. EORaster

Table holding related information about a single band of a geocoded EO raster dataset (eg. top of atmosphere reflectance of band 'CH2') of a GCP.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
band	foreign key	Name of the band stored in the EO raster.	-	Derived	References to Band
rasterband_set	foreign key	Parent RasterBandSet the EO raster belongs to	-	Derived	References to RasterBandSet
nodata_pct	integer	Percentage of nodata values within the EO raster	[%]	Calculated	See ATBD, [RD-6]
raster_filename	character varying(256)	Basename of the EO raster	-	Calculated	File name (and name + '.tif') of the band stored on the disk
osmon_version	foreign key	Version of the OSMON package	-	Derived	References to OSMONVersion
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.17. QlEORaster

Model holding the quicklook of the EO raster

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
src_raster	foreign key	Corresponding EO raster depicted in quicklook	-	Derived	References to the source dataset (EORaster)
quicklook	RasterField	Field containing the raster data of the quicklook	-	Computed	Generated during Processing
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.18. QlReferenceRaster

Model holding the quicklook of the reference raster

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
src_raster	foreign key	Corresponding reference raster depicted in quicklook	-	Derived	References to the source dataset (ReferenceRaster)
quicklook	RasterField	Field containing the raster data of the quicklook	-	Computed	Generated during Processing
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.19. MatchingBandCombination

Model holding the band combinations between reference and EO datasets. Stated bands are matched if parameters are within the limits of the sun elevation angle and within the range of allowed cloud/snow coverage. This table is created and filled during the OSMON setup process. The administrator can create new combinations or modify existing ones.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of the MatchingBandCombination	-	OSMON Configuration file	The initial band combinations are defined in the OSMON configuration file.
ref_band	foreign key	Reference raster band	-	Derived	References to Band
eo_band	foreign key	EO raster band	-	Derived	References to Band
matching_method	foreign key	Matching method to be applied	-	Derived	References to MatchingMethod
sun_elevation_angle_min	integer [-90, 90]	Minimum sun elevation	[°]	OSMON Configuration file	The initial value is defined in the OSMON configuration file.
sun_elevation_angle_max	integer [-90, 90]	Maximum sun elevation	[°]	OSMON Configuration file	The initial value is defined in the OSMON configuration file.

Name	Data type	Description	Unit	Source	Comment:
cloud_coverage_min	integer [0, 100]	Minimum snow and cloud coverage	[%]	OSMON Configuration file	The initial value is defined in the OSMON configuration file.
cloud_coverage_max	integer [0, 100]	Maximum snow and cloud coverage	[%]	OSMON Configuration file	The initial value is defined in the OSMON configuration file.
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.20. CloudClassification

Model holding the band relations for snow and cloud classification. Within the field 'equation' the mathematical expression to compute the snow/cloud cover is provided, where each band is addressed by its band name. It is foreseen to use accompanying flag bands to mask snow and cloud covered regions. This table is created and filled during the OSMON setup process. The administrator can create new combinations or modify existing ones.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
name	character varying(256)	Name of the Snow/Cloud screening algorithm	-	OSMON Configuration file	The initial band combinations are defined in the OSMON configuration file.
bands	many to many	Bands to be used in the snow and cloud screening algorithm. Multi selection allowed.	-	Derived	References to Band
equation	character varying(1024)	Equation to compute the snow and cloud cover. Equation must be stated in such a manner that it returns a Boolean array.	-	OSMON Configuration file	The initial equations are defined in the OSMON configuration file.
flag	integer	Flag defines if classification is applicable for day and/or night.	-	OSMON Configuration file	The flags are defined in the OSMON configuration file.
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.21. MatchingOutput

The MatchingOutput table contains the results of the matching procedure including calculated shifts and uncertainty estimates of Medium Resolution Optical Imager (MROI) data versus the reference data set at Ground Control Points.

Name	Data type	Description	Unit	Source	Comment:
id	integer	Primary key	-	Derived	Automatically generated during DB-insertion
eo_raster	foreign key	EO raster the matching was applied to	-	Derived	References to EORaster
ref_raster	foreign key	Shift is estimated wrt this reference raster	-	Derived	References to ReferenceRaster
matching_band_combination	foreign key	MatchingBandCombination used	-	Derived	References to MatchingBandCombination
shift_x	real	Shift estimate in x (easting) in meters	[m]	Calculated	See ATBD, [RD-6]
shift_y	real	Shift estimate in y (northing) in meters	[m]	Calculated	See ATBD, [RD-6]
shift_along_track	real	Shift estimate in along track direction in meters	[m]	Calculated	See ATBD, [RD-6]
shift_cross_track	real	Shift estimate in cross track direction in meters	[m]	Calculated	See ATBD, [RD-6]
snr	real	Signal to Noise Ratio	1	Calculated	See ATBD, [RD-6]
cor_coeff	real	Correlation coefficient	1	Calculated	See ATBD, [RD-6]
error_flag	integer	Matching error code	-	OSMON Configuration file	Definition of Matching/Error codes in 1.22.2
osmon_version	foreign key	Version of the OSMON package	-	Derived	References to OSMONVersion
last_insert	timestamp: (DD.MM.YYYY hh:mm:ss UTC)	Timestamp (UTC) of last database insertion	-	Derived	Automatically derived from computer system time during database ingestion

1.22. Numeric Codes

This section explains the numeric codes used in the database tables to indicate the spectral range covered by a band (BandType) and the exit status of the matching loop (MachingError).

1.22.1. BandTypes

The bands of satellite data sets are grouped into the basic categories:

Attribute	Numeric Code	Description
Unknown	0	Unspecified band
Visible	1	Bands covering the spectrum of visible light ranging from 0.380 to 0.740 µm
NIR (Near-infrared)	2	Bands covering the spectrum of near-infrared ranging from 0.75 to 1.4 µm
TIR (Thermal-Infrared, Long-wavelength infrared)	3	Bands covering the infrared spectrum ranging from 8 to 15 µm
Cloud	17	Bands holding cloud cover information.
Confidence	18	Bands holding confidence information.
Other	256	Any other bands which might be used within the processing.

1.22.2. RasterbandSetError

During the data import the reader performs a basic consistency check and logs possible errors. Error codes for the following situations are foreseen and relate to the entire RasterbandSet:

Attribute	Numeric Code	Description
OK	0	Rasterbandset has no data gaps
EmptySet	1	No valid bands are returned from the reader routines.
CloudSnowError	2	Marks an error during the cloud screening process
RegriddingError	4	An unspecified error occurred in the external SLSTR regridding routine (not maintained by OSMON developers)
Exception	256	Unspecific exception during data import

1.22.3. MatchingError

The matching algorithm performs a basic consistency and quality check at the beginning and at the end of the matching routine. Error codes for the following situations are foreseen:

Attribute	Numeric Code	Description
OK	0	Sufficient quality of matching result
CP_nodata	2	Reference raster contains pixels with nodata values
EO_nodata	4	EO raster contains pixels with nodata values

MATCH_outside	8	Shift estimate is larger than half the image width or half image height
MATCH_no_convergence	16	Maximal number of iterations reached
MATCH_unreliable	32	Current shift estimate is larger than shift estimate of previous iteration
MATCH_correlation_decreased	64	Correlation coefficient has decreased compared to the previous iteration
MATCHFUNC_Exception	256	Exceptions derived from Python's exception base class, raised when matching fails
MATCHFUNC_outofbounds	512	Exception is raised when the index values are outside the bounds of the array

2. REFERENCE DATASET (DELIVERABLE D10)

Reference data are the basis to generate a set of ground control points (GCPs) for determining the local geolocation accuracy of rectified satellite images. GCPs are extended targets such as lakes and islands which can be clearly identified in satellite images.

The reference database is extracted from Landsat-8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor). Sentinel-2 data are not used for this version of the reference database as Sentinel-2 does not provide TIR bands. Landsat-8 OLI and TIRS data in Level 1 Topographic Corrected (L1T) processing level are used.

The SRTM Water Body Data (SWBD) is a vector data base of open water bodies and is used as additional independent data source for visual inspection of the accuracy ground control points extracted from Landsat-8 L1T data and can be used for visualisation.

The generation of the reference data set is generated in three steps:

- STEP 1: selection of prominent features suitable for matching with MROI data.
- STEP 2: download of Landsat-8 image scenes from the archive and crop the image template covering the features (named ground control point).
- STEP 3: after processing a time series of MROI data, the successful matches of the GCPs are counted. Not suitable GCPs (no successful matches with MROI image templates) are replaced.

In the following sections in 2.1and section 2.2 the data sets used for generation of the reference data base are described. The steps for generation of the reference data is described 2.3 and section 2.4. The naming convention of the GCPs and structure of the reference data base are described in section 2.5 and section 2.6. An overview of the OSMON reference data base is shown in section 2.7.

2.1. Landsat L1T Products

The L1T correction process utilizes both ground control points (GCP) and digital elevation models (DEM) to attain absolute geodetic accuracy. The WGS84 ellipsoid is employed as the Earth model for the Universal Transverse Mercator (UTM) coordinate transformation. The L1T Product is geometrically rectified, free from distortions related to the sensor (e.g. jitter, view angle effects), satellite (e.g. attitude deviations from nominal) and Earth (e.g. rotation, curvature, relief). Geodetic accuracy of the product depends on the accuracy of the ground control points and the resolution of the DEM used. Ground control points used for Level 1T correction come from GeoCover which compiled a DEM from various sources including SRTM, GTOPO 30 [RD-3]. Characteristics of Landsat-8 L1T data products are listed in Table 1.

Table 1: Characteristics of Landsat-8 OLI and TIRS, L1T (Level 1 Topographic Corrected)
(from https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-8-oli-operational-land-imager-and?qt-science_center_objects=0#qt-science_center_objects visited 30/10/2019)

Processing:	• Level 1T - Terrain Corrected
Pixel Size:	• OLI Multispectral bands: 30 meters • OLI panchromatic band: 15 meters • TIRS Thermal bands: 100 meters (resampled to 30 meters to match multispectral bands)
Data Characteristics:	• GeoTIFF data format • Cubic Convolution resampling • North Up MAP orientation • Universal Transverse Mercator (UTM) map projection (Polar Stereographic for Antarctica) • World Geodetic System (WGS) 84 datum • **12 meter circular error, 90% confidence global accuracy for OLI** • **41 meter circular error, 90% confidence global accuracy for TIRS** • 16-bit pixel values

2.2. SRTM Water Body Data (SWBD)

The SRTM Water Body Data (SWBD) is a geographical dataset encoding high-resolution worldwide coastline outlines in a vector format, published by NASA. It was created by BAE Systems ADR for the US National Geospatial-Intelligence Agency as a complementary product during editing of the digital elevation model database of the Shuttle Radar Topography Mission (SRTM). SWBD data covers the Earth's surface between 56° southern latitude and 60° northern latitude. It is distributed in ESRI

shapefile format, divided into 12229 files, each covering one 1°-by-1° tile of the Earth's surface. SWBD data is in the public domain and is made available online for free download by NASA. [RD-8].

Specifications of SWBD:

- The horizontal datum is the World Geodetic System (WGS84 1984).
- The vertical datum is WGS84 Earth Gravitational Model 1996 (EGM 96) geoid.
- Absolute Horizontal Accuracy equivalent to SRTM DTED 2 is 20 m circular error, 90% confidence.
- Absolute Vertical Accuracy equivalent to SRTM DTED 2 is 16 m linear error, 90% confidence.

2.3. Selection of prominent features as reference image template (GCPs)

For the creation of the reference database, the SRTM water body data (SWBD), which have an resolution of 30 meters, has been used. From these water outlines, lakes and islands with a certain dimension have been pre-selected automatically. A minimum area of 5 km2 has been defined as threshold for removing too small water bodies.

In a next step, the remaining features have been manually checked for their shape, distribution and surroundings. GCPs are features which can be clearly identified in medium resolution optical mages, and which are stable in time. Typical features are

- lakes,
- islands
- coastlines (characteristic shape)
- peninsulas
- geological features

Thereby, some inappropriate features were rejected, whereas other, suitable features which had not been selected during the automatic procedure were manually added to improve the spatial coverage. It is intended to continuously adapt the set of GCPs in dependence on their performance during the matching procedure of a longer time series of satellite data from different sensors.

One important issue is to identify and exclude lakes or islands, which exhibit significant changes in their extent. This can be done by overlaying independent data sets of water bodies and manually reviewing the GCP candidates.

Figure 2.1 shows two examples of GCPs. We used Landsat-8 raster data and the SWBD vector data of open water. As the data sets were acquired at different dates, GCPs with temporarily variable extent (e.g. Figure 2.2) can be easily identified.

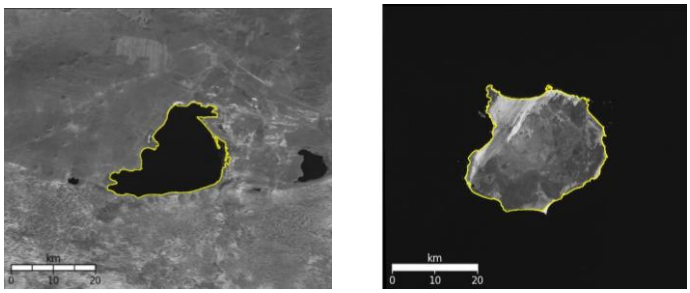


Figure 2.1: Examples of Ground Control Points, displayed by the SWBD water outlines laid over the Landsat band 4 (NIR) raster files: The lake Dali Nuoer in China (left), and the island Ilha da Boa Vista in Cape Verde, Africa (right).

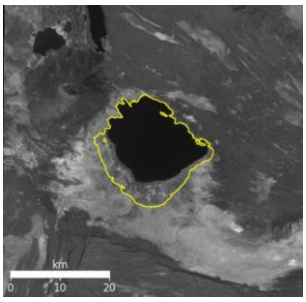


Figure 2.2: Example of lake with temporarily changing extent: Lake Abhe, Ethiopia. This type of GCP is not useful as GCP, as the lake extent changes significantly in time.

2.4. Automatic extraction of reference image templates (GCPs)

In the OSMON project the term Ground Control Point (GCP) refers to an image template covering a prominent feature on the Earth Surface, which can be identified in medium resolution optical images. GCP image templates are extracted from VIS, NIR, SWIR, TIR spectral bands of Landsat-8 OLI and TIRS data. The geographic location of a GCP is exactly specified by the geographic map projection, ellipsoid and local datum, corner coordinates and pixel spacing in Easting and Northing direction.

The automatic retrieval of the appropriate Landsat scene for each GCP candidate selected in step 1 (section 2.4), and the preparation of the Landsat image chips is outlined in Figure 2.3. The procedure utilizes Landsat-8 L1T OLI and TIRS data but can be also adapted for other orthorectified data sets.

For automatic pre-selection of GCPs the SWBD data set was screened for suitable targets according to their size and global distribution. The GCP selection was manually refined by taking the spatial density of points into account. For the selected shapefiles the center coordinates and the maximum extent in East and West direction are derived. The center points in latitude and longitude are used for creating the GCP id and filename, which makes sure that only unique GCP IDs and filenames are generated.

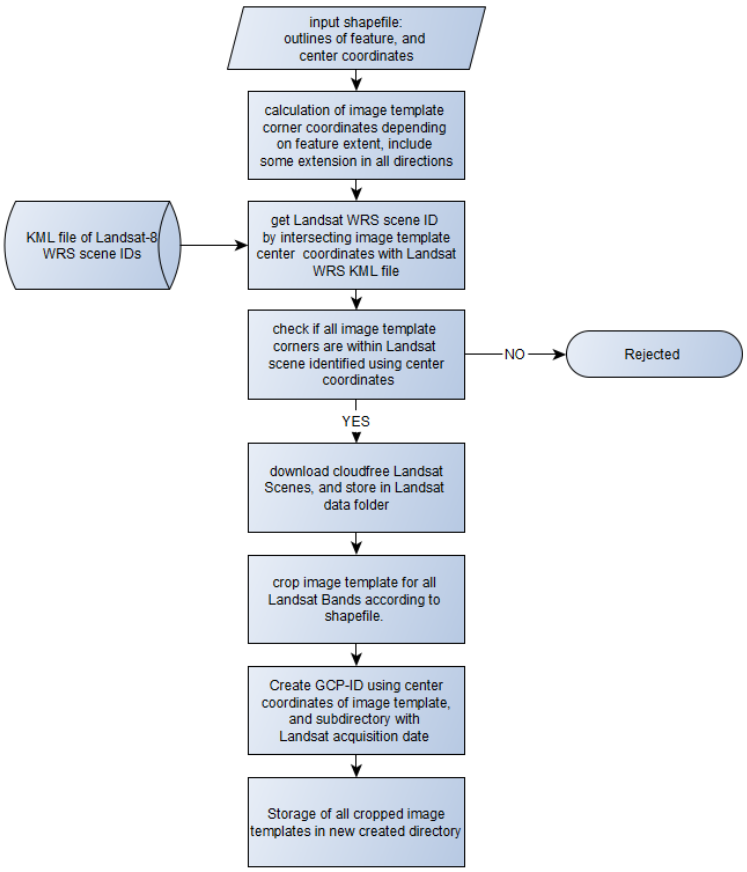


Figure 2.3: Procedure for automatic extraction of Landsat-8 image chips using centre coordinates of identified GCP, performed for all VIS, NIR, SWIR and TIR bands of Landsat-8.

Based on the center coordinates of the GCP the Landsat WRS Scene ID is calculated and the full coverage of the reference image chip by the LANDSAT scene is calculated.

If the image template of the GCP is only partly covered by the Landsat image (e.g. the GCP is located at the border of the Landsat Scene) the scene was not used for this GCP. . Otherwise it is checked if the scene has already been downloaded (e.g. by a previous GCP) from the data archive server.

Finally, the bounding box of the GCP shapefile is extended in all directions to include the surroundings of the feature. Using this bounding box image raster templates are cropped from the original Landsat-8 OLI product for all VIS, NIR, SWIR and TIR bands. Note that there is no resampling applied to crop the image templates from the full Landsat scenes. They are stored in the hierarchical reference data base using the GCP id as directory name (see section 2.5). The Creation of the GCP ID is described in section extracted GCPs raster files are overlaid with the outline of the feature of the SWBD in order to manually assess the accuracy and temporal stability of the GCP.

2.5. GCP Identification Name

The reference data set consist of Ground Control Points (GCP), which cover features which can be clearly identified in Medium Resolution Optical Imager (MROI) data.

Each GCP can be identified by a unique identification number. To guarantee a unique and automatic generation of IDs of GCPs we use the center coordinates of the feature. The ID is generated using the following the naming convention

HDDDdd**E**DDDdd

where

- H … describes the Northern (H=N) or Southern (H=S) hemisphere
- E … Easting; describes the longitude in degrees Easting
- DDD …. degrees (with proceeding zeros)
- dd …… decimal degrees

example: GCP center point is 47.43 North, and 52.34 East, the ID is given by N04743E05234.

2.6. Structure of the reference data set

The reference data set is organized in a hierarchical directory structure following the example given in Figure 4. For each GCP and acquisition date all spectral bands of Landsat-8 OLI are cropped and stored using a standard naming convention. The spectral bands and corresponding filenames are shown in Table 2. The reflective spectral bands are converted to top of atmosphere reflectance values (scaled between 0 and 1), the TIR bands are converted to brightness temperatures according to the Landsat-8 Data User Handbook [RD-3]

A GeoTIFF raster file contains also metadata on geoinformation according to [RD-1], which allows a full geolocation of each pixel of the raster file. The geolocation metadata information includes map projection, coordinate system, information on ellipsoid and local datum, corner coordinates and grid spacing of the raster file. Table 3:shows an example of the geographical information annotated in the GCP N045603E010671/20181214/ B5_toaref_utm.tif.

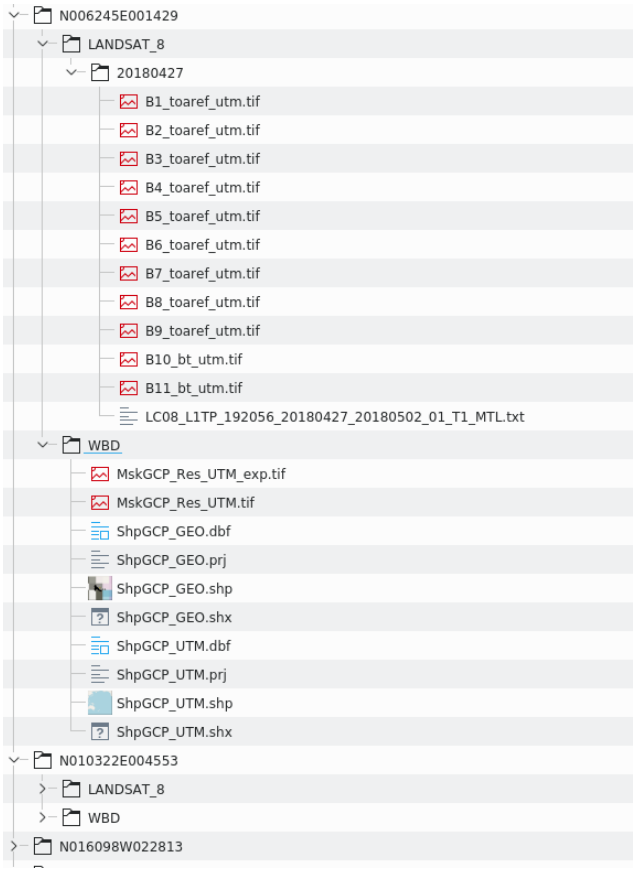


Figure 4: Reference data set directory structure and files.

Table 2: Full list of Landsat 8's bands: toaref- top of atmosphere reflectance; bt-brightness temperature; utm – Universal Transverse Mercator projection

Band Number	***Wavelength band (µm)***	***Resolution***	***Filename***	***Data type, Value Range; Unit***
1	0.433–0.453	30 m	B1_toaref_utm.tif	Float, 0-1 ; reflectance
2	0.450–0.515	30 m	B2_toaref_utm.tif	Float, 0-1 ; reflectance

Band Number	*Wavelength band (µm)*	*Resolution*	*Filename*	*Data type, Value Range; Unit*
3	0.525–0.600	30 m	B3_toaref_utm.tif	Float, 0-1 ; reflectance
4	0.630–0.680	30 m	B4_toaref_utm.tif	Float, 0-1 ; reflectance
5	0.845–0.885	30 m	B5_toaref_utm.tif	Float, 0-1 ; reflectance
6	1.560–1.660	30 m	B6_toaref_utm.tif	Float, 0-1 ; reflectance
7	2.100–2.300	30 m	B7_toaref_utm.tif	Float, 0-1 ; reflectance
8	0.500–0.680	15 m	B8_toaref_utm.tif	Float, 0-1 ; reflectance
9	1.360–1.390	30 m	B9_toaref_utm.tif	Float, 0-1 ; reflectance
10	10.6-11.2	100 m	B10_bt_utm.tif	Float, Kelvin
11	11.5-12.5	100 m	B11_bt_utm.tif	Float, Kelvin

Table 3: Example for Geographic information for N045603E010671/20181214/B5_toaref_utm.tif as extracted by gdalinfo [RD-1]

```
Driver: GTiff/GeoTIFF
Files: N045603E010671/LANDSAT/20181214/B5_toaref_utm.tif
Size is 1251, 1982
Coordinate System is:
PROJCS["WGS 84 / UTM zone 32N",
    GEOGCS["WGS 84",
        DATUM["WGS_1984",
            SPHEROID["WGS 84",6378137,298.257223563,
                AUTHORITY["EPSG","7030"]],
            AUTHORITY["EPSG","6326"]],
        PRIMEM["Greenwich",0,
            AUTHORITY["EPSG","8901"]],
        UNIT["degree",0.0174532925199433,
            AUTHORITY["EPSG","9122"]],
        AUTHORITY["EPSG","4326"]],
    PROJECTION["Transverse_Mercator"],
    PARAMETER["latitude_of_origin",0],
    PARAMETER["central_meridian",9],
    PARAMETER["scale_factor",0.9996],
    PARAMETER["false_easting",500000],
    PARAMETER["false_northing",0],
    UNIT["metre",1,
        AUTHORITY["EPSG","9001"]],
    AXIS["Easting",EAST],
    AXIS["Northing",NORTH],
    AUTHORITY["EPSG","32632"]]
Origin = (613111.165046000038274,5087748.132166000083089)
Pixel Size = (30.000000000000000,-30.000000000000000)
Metadata:
  AREA_OR_POINT=Area
Image Structure Metadata:
  INTERLEAVE=BAND
```

```
Corner Coordinates:
Upper Left  (  613111.165, 5087748.132) ( 10d27'32.56"E, 45d56' 2.36"N)
Lower Left  (  613111.165, 5028288.132) ( 10d26'42.75"E, 45d23'56.29"N)
Upper Right (  650641.165, 5087748.132) ( 10d56'34.46"E, 45d55'36.42"N)
Lower Right (  650641.165, 5028288.132) ( 10d55'28.16"E, 45d23'30.83"N)
Center      (  631876.165, 5058018.132) ( 10d41'34.36"E, 45d39'47.41"N)
Band 1 Block=1251x1 Type=Float32, ColorInterp=Gray
  NoData Value=3.4028234663852886e+38
```

2.7. Coverage of reference database

The reference dataset version 1.0 (Ref.: D10_ReferenceDataset_v1_0) includes globally distributed GCPs shown in Figure 5. These datasets contain the all bands of Landsat 8 converted to top of atmosphere reflectance and brightness temperature respectively and follows the naming convention and directory structure described in section 2.6.

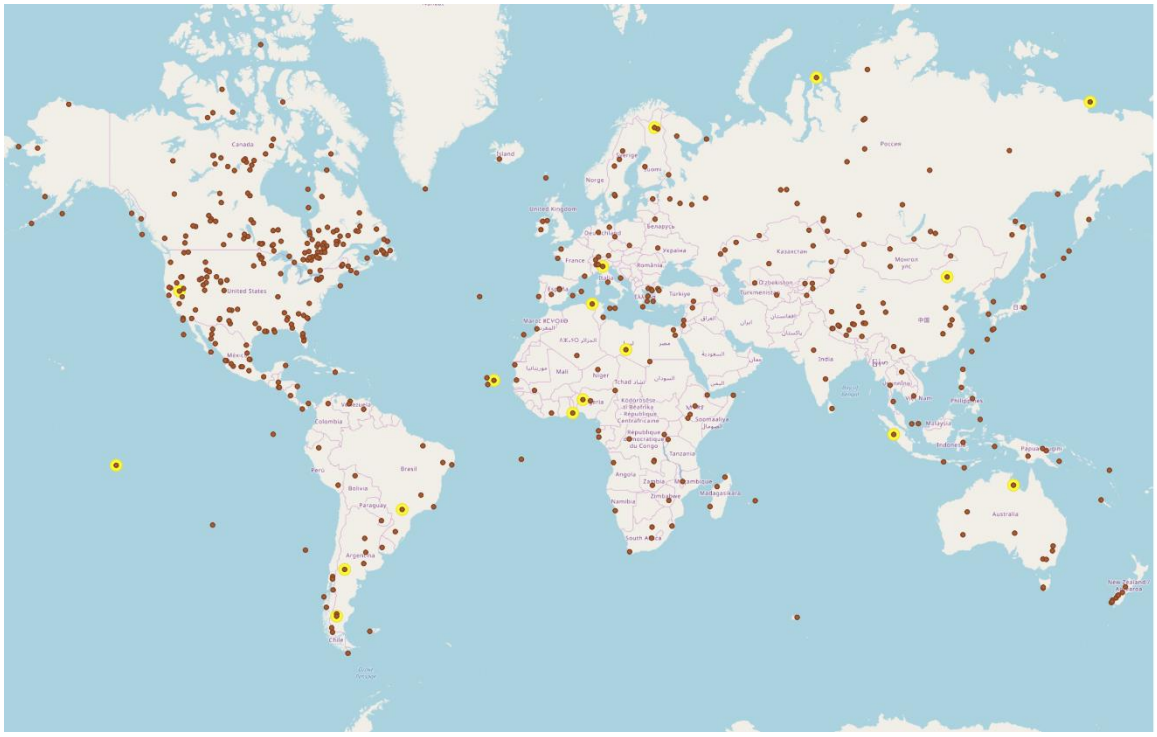


Figure 5: Overview of globally distributed GCPs. The yellow markers indicate the selected GCPs for the diagnostic dataset.

2.8. Quality of selected GCPs

To test the quality of the selected GCPs, a time series (at least several months) of MROI data is processed and the matching with GCPs is performed. The number of successful matches of GCPs with MTROI data is statistically evaluated. Unsuccessfully matches might have different reasons such as

permanent cloud cover, temporarily varying lake extent, too low contrast between feature and surroundings, etc. Unsuitable GCPs will be identified and replaced by other GCPs, which will be tested again.

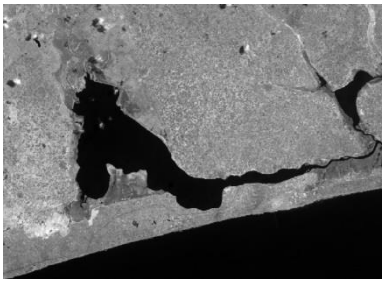
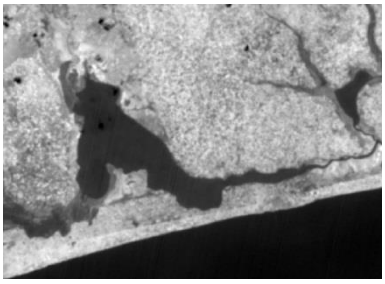
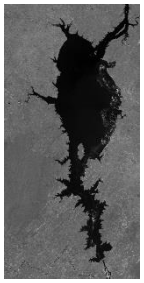
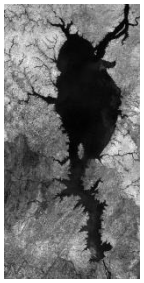
3. DIAGNOSTIC DATASET (DELIVERABLE D11)

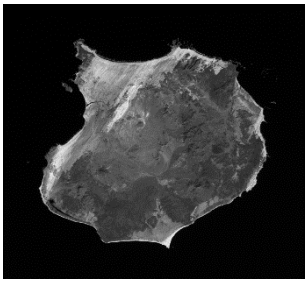
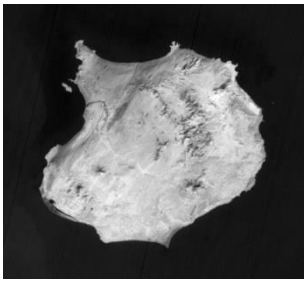
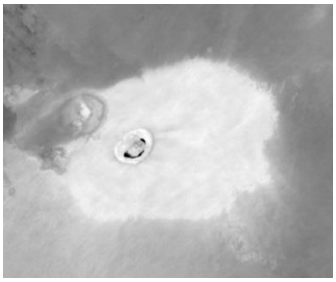
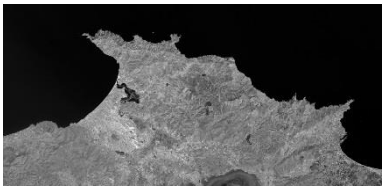
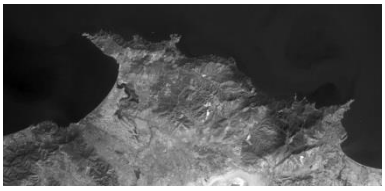
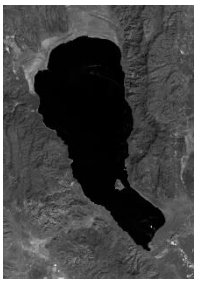
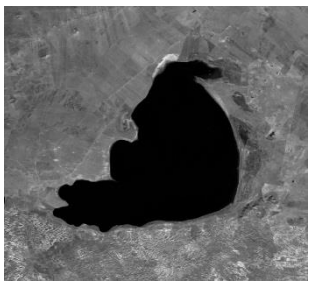
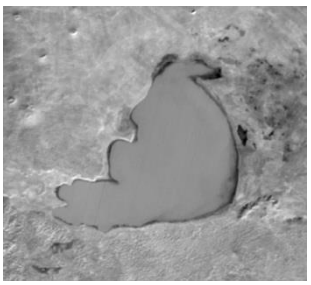
The diagnostic dataset version v1.0 [RD-2] includes all data required to develop, implement and test the OSMON tool. This includes a subset of ground control points of the reference data base, and a time series of Medium Resolution Optical Imager (MROI) data, and auxiliary data:

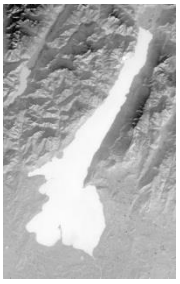
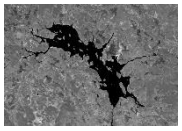
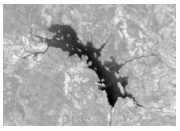
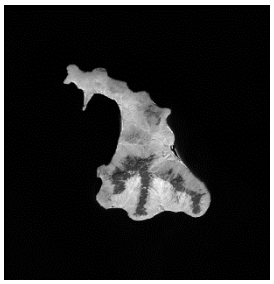
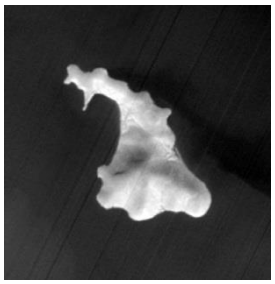
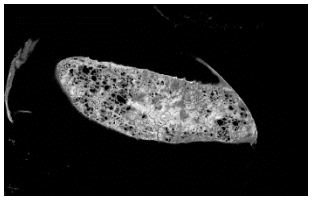
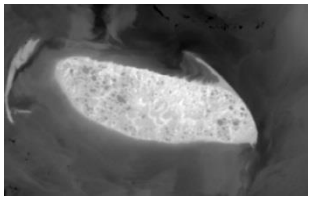
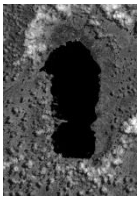
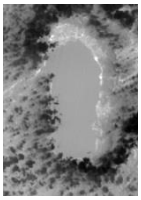
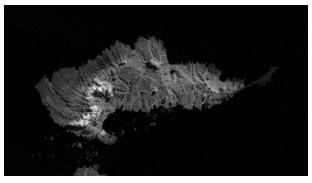
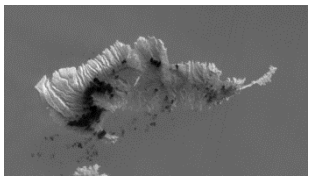
- a subset of 17 GCPs drawn from the reference dataset version 1.0 (Figure 5) which includes
  - Raster bands in GeoTiff format of Landsat-8 Spectral Bands (VIS; NIR, SWIR, TIR)
  - Water body vectors (SWB)
- 383069 SLSTR scenes acquired between 03/07/2017 and 09/07/2017 , Level-1B format
- 73727 OLCI scenes acquired between 03/07/2017 and 09/07/2017, Level-1B format
- 5208 AVHRR/3 Metop-B (M01) scenes acquired in 2017, Level-1B format
- 5203 AVHRR/3 Metop-A (M02) scenes acquired in 2017, Level-1B format

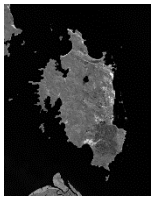
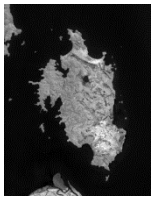
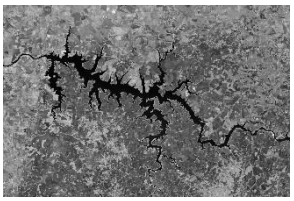
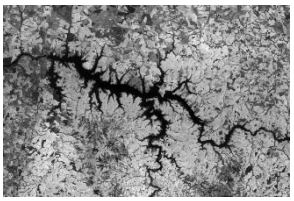
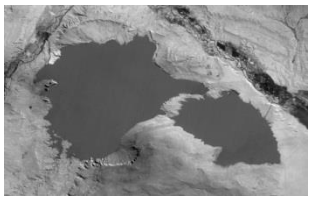
The electronic data set corresponding to Deliverable D11, includes the selected GCPs. In order to reduce the data volume of the diagnostic data set we did not upload the SLSTR, OLCI and AVHRR data to the EUMETSAT Server.

Table 4: The 17 GCPs of the diagnostic data set. The figure show band 5 and band 10 (TIR) of the GCPs, the electronic data set includes all bands organised in the hierarchical data structure as described in 2.6 [RD-1].

GCP ID	*B5*	*B10*
N006245E001429		
N010322E004553		

GCP ID	B5	B10
N016098W022813		
N024915E017797		
N036946E007424		
N040048W119565		
N043290E116634		

GCP ID	B5	B10
N045603E010671		
N068079E026594		
N070843E160606		
N073156E076398		
S000322E100193		
S009774W139013		

GCP ID	B5	B10
S015718E137025		
S022783W051052		
S038561W068741		
S048916W071215