EUROPEAN SPACE AGENCY

ESA Reference: ESA AO/1-11038/21/I-DT

ESA Contract Number: 4000139605/22/I-DT

Project Afri4Cast - Food Security and Safety In Africa

WP200

D7.1 Algorithm Theoretical Baseline Documents (ATBDs) and Product Specifications

Submitted by:

AgroApps

In collaboration with:

SAPIENZA
UNIVERSITÀ DI ROMA

23 September 2024

Document Information

ESA Contract Number	4000139605/22/I-DT		**Acronym**	AFRI4CAST
Full Title	EO AFRICA EXPLORERS – EXPRO+ Food Security and Safety In Africa			
Start Date	26th October 2022		**Duration**	24 months
Project URL	www.afri4cast.org			
Deliverable	D7 - Algorithm Theoretical Baseline Documents (ATBDs) and Product Specifications			
Work Package	WP200 – Prototype test and Outreach			
Date of Delivery	**Contractual**		**Actual**	Contractual
Nature	Report		**Dissemination Level**	Confidential
Lead Beneficiary	AgroApps			
Responsible Author	Agathoklis Dimitrakos			
Contributions from	SIA			

Document History

Version	**Issue Date**	**Pages**
V1	26/04/2024	118
V2	26/06/2024	123
V3	26/09/2024	185

Disclaimer

Contents

Table of Acronyms

Acronym	Full Term
AOI	Area of Interest
API	Application Programming Interface
AppEEARS	Application for Extracting and Exploring Analysis Ready Samples
AROSICS	Automated and Robust Open-Source Image Co-Registration Software
ASI	Agenzia Spaziale Italiana (Italian Space Agency)
BOA	Bottom Of Atmosphere
CDS	Climate Data Store
CNN	Convolutional Neural Network
CRS	Coordinate Reference System
DL	Deep Learning
DLS	Damped Least- Squares
DN	Digital Numbers
DOY	Day Of Year
DTC	Diurnal Temperature Cycle
ECOSTRESS	ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station
EnMAP	Environmental Mapping and Analysis Program
EO	Earth Observation
EODAG	Earth Observation Data Access Gateway

FIHS	Fast Intensity Hue Saturation
FRU	Feature Refinement Unit
FWHM	Full Width Half Maximums
GDD	Guided Deep Decoder
GLP	Gaussian Laplacian Pyramid
GPT	Graph Processing Tool
GRIB	GRIdded Binary or General Regularly-distributed Information in Binary form
GSA	Gram-Schmidt Adaptive
HR	High Resolution
HS	Hyperspectral
LAI	Leaf Area Index
LIDF	Leaf Inclination Distribution Function
LM	Levenberg-Marquardt
LR	Low Resolution
LST	Land Surface Temperature
LUT	Look Up Table
MAE	Mean Absolute Error
MODIS	Moderate Resolution Imaging Spectroradiometer
MS	Multispectral
MTF	Modulation Transfer Function
MSE	Mean Squared Error

MSI	Multispectral Instrument
MSSI	Mean Structural Similarity Index
NIR	Near Infrared
NLS	Nonlinear Least Squares
PRISMA	PRecursore IperSpettrale della Missione Applicativa
PSF	Point Spread Function
PSNR	Peak Signal to Noise Ratio
QA	Quality Assurance
QC	Quality Control
RAA	Relative Azimuth Angle
REA	Reanalysis
REP	Red Edge Position
RFR	Random Forest Regression
RMSE	Root Mean Squared Error
RTM	Radiative Transfer Model
SAM	Spectral Angle Mapper
SAR	Synthetic-Aperture Radar
SAIL	Scattering by Arbitrarily Inclined Leaf
SCL	Scene Classification Layer
SCM	Spectral Correlation Mapper
SLSTR	Sea and Land Surface Temperature Radiometer

SRF	Spectral Response Function
SSIM	Structural Similarity Metric
STTFN	Spatiotemporal Thermal Data Fusion Network
SWIR	Shortwave Infrared
SZA	Solar Zenith Angle
URU	Upsampling Refinement Unit
VNIR	Visible and Near-Infrared
VZA	Viewing Zenith Angle

1. Introduction

A general scope of AFRI4Cast project is to provide a proof of principle for the idea of using Thermal and Hyperspectral, enhanced and downscaled on spatial and temporal dimensions, as inputs to Crop Growth/Disease models, at crop parcel spatial scale. The feasibility of this concept is regarded specifically for the satellite missions of ECOSTRESS thermal and PRISMA hyperspectral data products as the input data.

Deliverable D7.1 represents a comprehensive compilation of all aspects detailed in the ATBD document, reflecting a holistic approach to presenting our work. This decision to expand the work on D7 into an extended document D7.1 from its original format, underscores our commitment to showcasing the entirety of our efforts in a cohesive manner. By consolidating all related tasks and findings into one document, the aim is to provide a thorough exploration of the ATBD's contents. This includes not only outlining the technical specifications and methodologies but also detailing the iterative processes and collaborative insights that have shaped the development of fusion algorithms.

The current ATBD document encapsulates a continuum of research and development efforts that converge towards the singular objective of enhancing fusion algorithms. Each section within D7.1 builds upon the preceding ones, fostering a comprehensive understanding of the complexities and nuances involved. By presenting the full spectrum of our work in this extended format, we ensure transparency and clarity, facilitating a deeper appreciation of the methodology evolution and the strategic decisions guiding its implementation. This approach not only enhances the document's utility as a reference but also reinforces our commitment to delivering robust and innovative solutions in the realm of algorithmic fusion.

For that cause several state-of-the-art algorithms that deal with the topics below have been adopted, modified and developed in order to be applied on the above-mentioned data products:

- EO data acquisition and pre-processing
- Curve Fitting
- Deep Learning Image Fusion
- RTM Inversion

The project's main proof of principle over the framework of algorithm development was to present:

- data pre-processing routines and solutions tailored to the input data products specificities
- evaluated performance on the data fusion tasks, comparable with that is achieved, by other operational thermal and hyperspectral products, over current technical and scientific literature

The demonstration of the project's proof of concept feasibility at that point, involved

- the determination input data product specifications
- the development of the input data acquisition tool components

- the experimental setting design
- the development of image processing algorithm tools
- the development of data fusion algorithm tools and models
- the development of RTM inversion models

The current ATBD aims to provide a detailed specification of the processing algorithms of the EO solution proposed by the project. Technical considerations justifying the selected methodologies are provided. It includes a functional description of each selected algorithm with a "high level" mathematical description of the problem formulation and a precise specification of all input data, processing steps, calibration parameters and output products. Finally, the ATBD provides the final validation results from the algorithm performance over the experimental setup obtained at the 2nd iteration of algorithm development, testing and analysis.

Algorithm Theoretical Baseline Document Structure

In the presented report a detailed description of the implemented algorithms for the processing and analysis of the EO data, which are related to the AFRI4Cast project, is attempted. In the respective chapters, the 3 main thematic areas are developed as separate subjects:

- Thermal Data Fusion
- Hyperspectral Data Fusion
- Inversion of RTMs for Biophysical Parameters Estimation

In each of the preceding chapters, the same sub-chapter organization is consistently followed throughout the text, describing:

1. the main task objectives
2. the features and specificities of the EO data products used
3. the overall data processing pipeline in a flowchart form
4. the secondary subtasks that involve image acquisition pipelines and the creation of a test dataset based on the experimental design of the project.
5. the subtasks dealing with satellite data pre-processing methods. Procedures such as quality masking in thermal images, evaluation of spectral channels of a Hyperspectral image, sharpening of channels of a Multispectral image to a common spatial resolution, or even registration of HS-MS image pairs, are described here.
6. the main subtasks of algorithm development for applications of Curve Fitting, Deep Learning Data Fusion, Machine Learning modeling for the inversion of RTMs, where the stronger emphasis is given to the description.
7. The updated results obtained at the 2nd iteration of algorithm development, testing and analysis

2. Thermal Data Fusion

2.1. Objectives

Land surface temperature (LST), or the skin temperature of the Earth's surface, is a major variable in climate systems and has been used as a key parameter to predict surface energy balance. Concerning the remote sensing-derived LST products, there are certain trade-offs between the spatial and temporal resolution of the available products and many studies have proposed various spatiotemporal downscaling approaches utilizing multi-sensor data to address these trade-offs.

In particular, thermal data fusion-based approaches determine the relationship between known image pairs of LST at fine and coarse resolution and downscale the coarse resolution image when a fine resolution LST image is not available. Within AFRI4Cast, ECOSTRESS LST data was intended to be sharpened using a deep learning-based Spatiotemporal Temperature Fusion Network (STFN) method aiming to generate fine resolution spatiotemporal LST products. In STFN, a multi-scale fusion convolutional neural network (CNN) was used to construct the complex nonlinear relationship between input and output LSTs.

ECOSTRESS and MODIS LST data were used as fine and coarse spatial resolution thermal data, respectively. Specifically, a pair of fine and coarse LST images before the target date, for which a coarse MODIS image was available, as well as another pair of fine and coarse images after the target date were collected. The objective of the Spatiotemporal Temperature Fusion Network was to predict a high resolution ECOSTRESS-like LST image using one MODIS LST image at the target date and two pairs of ECOSTRESS and MODIS images from nearby dates.

2.2. Data Products

The current chapter presents the designated thermal Earth Observation (EO) data products to be utilized within AFRI4Cast's framework. These products facilitate the data processing pipelines and some of them contribute to the validation of the results, as well (Table 1.). MODIS Terra/Aqua Day/Night LST, ECOSTRESS LST and Sentinel-3 SLSTR L2 LST observations are mentioned below along with the ERA5-Land REA (reanalysis) data product. For more detailed information, please refer to the Deliverable D3 “State of the art review report on analysis methods and algorithms”, Chapter “3. Input data products for AFRI4Cast EO Services”.

Table 1. Thermal data products and their use within AFRI4Cast.

Product	Purpose of Use
MODIS Terra/Aqua Day/Night LST	Calibration of DTC, STTFN reference layer

ECOSTRESS LST	STTFN target layer
Sentinel-3 SLSTR L2 LST	Validation of the DTC produced LST
ERA5-Land REA	DTC Initial values

MODIS Terra/ Aqua Day/Night LST

The Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) Land Surface Temperature/Emissivity (LST&E) Level 3 products, MOD11A1 and MYD11A1 respectively, Version 6.1 provide daily and global LST and emissivity observations in a spatial resolution of 1km. The layers of these data products which were used within AFRI4Cast, as well as their features are presented on Table 2.

Regarding MODIS Terra/ Aqua LST acquisition times, they are provided via the data product layers “Day_view_time” and “Night_View_Time” in hour units in local solar time. For more information regarding the time of acquisitions, please refer to the Chapter “3.4. Thermal Images Acquisition”.

Table 2. Layers/Variables of the MODIS LST Terra & Aqua data products and their features.

Product	Version	Layer	Layer Description	Units	Data Type	Scale Factor
MYD11A1	061	Day_view_time	Local solar time of day observation	Hours	8-bit unsigned integer	0.1
		LST_Day_1km	Day Land Surface Temperature	Kelvin	16-bit unsigned integer	0.02
		LST_Night_1km	Night Land Surface Temperature	Kelvin	16-bit unsigned integer	0.02
		Night_view_time	Local solar time of night observation	Hours	8-bit unsigned integer	0.1
		QC_Day	Daytime LST Quality Indicators	Bit Field	8-bit unsigned integer	N/A
		QC_Night	Nighttime LST Quality	Bit Field	8-bit unsigned	N/A

			Indicators		integer	
MOD11A1	061	Day_view_time	Local solar time of day observation	Hours	8-bit unsigned integer	0.1
		LST_Day_1km	Day Land Surface Temperature	Kelvin	16-bit unsigned integer	0.02
		LST_Night_1km	Night Land Surface Temperature	Kelvin	16-bit unsigned integer	0.02
		Night_view_time	Local solar time of night observation	Hours	8-bit unsigned integer	0.1
		QC_Day	Daytime LST Quality Indicators	Bit Field	8-bit unsigned integer	N/A
		QC_Night	Nighttime LST Quality Indicators	Bit Field	8-bit unsigned integer	N/A

ECOSTRESS LST

The ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) Land Surface Temperature and Emissivity Level 2 product, ECO2LSTE, Version 1, provides global LST and emissivity measurements in a spatial resolution of 70m. The layers of the ECO2LSTE data product to be used within AFRI4Cast as well as their features are presented on Table 3.

Regarding the ECOSTRESS LST acquisition time, a data product view time layer is not available. Thus, it was extracted from the product's filename assuming that for a small area the acquisition time is constant. ECOSTRESS LST acquisition time is given in hour units in UTC Greenwich time zone.

For example: "ECO2LSTE.001_SDS_LST_doy2023273195732_aid0001.tif" filename indicates that the image is acquired on the 273th day of 2023 at 19:57:32 UTC Greenwich time.

For more information regarding the frequency of ECOSTRESS LST acquisitions, please refer to the Chapter "3.4. Thermal Images Acquisition" of this document.

Table 3. Layers/Variables of the ECOSTRESS ECO2LSTE.001 data product and their features.

Product	Version	Layer	Layer Descripti	Units	Data Type	Scale Factor

			on			
ECO2LSTE	001	SDS_LST	Land Surface Temperature	Kelvin	16-bit unsigned integer	0.02
		SDS_QC	Quality Control for LST and emissivity	N/A	16-bit unsigned integer	N/A

Sentinel-3 SLSTR L2 LST

Sentinel-3 SLSTR Level-2 LST Non-Time Critical (NTC) product provides daily gridded land surface parameters at a spatial resolution of 1 km. Regarding Sentinel-3 SLSTR Level-2 LST acquisition time, a view time layer is not available. Therefore, the acquisition time is extracted from the product's filename making the assumption that the acquisition time is constant in a small area. Sentinel-3 SLSTR L2 LST acquisition time is provided in hour units in UTC Greenwich time zone.

For example the file name:
"S3B_SL_2_LST____20230125T072349_20230125T072649_20230126T001918_0179_075_220_2880_PS2_O_NT_004.SEN3"
indicates that the sensing start time is at 07:23:49 and the sensing end time is at 07:26:49 UTC Greenwich time.

For more information regarding the frequency of Sentinel-3 SLSTR Level-2 LST acquisitions, please refer to the Chapter "3.4. Thermal Images Acquisition" below.

ERA5-Land REA Skin Temperature

ERA5-Land provides information of land variables. It is composed of a detailed record from 1950, with a temporal resolution of 1 hour and a spatial resolution of 9km. ERA5-Land is not a remote sensing dataset, but a reanalysis (REA) dataset. Reanalysis combines model data with global observations to create a globally consistent dataset. Reanalysis produces data that cover several decades in the past, thus giving a precise description of the climate in the past.

Within AFRI4Cast, the Skin temperature variable was utilized, which depicts the temperature of the surface of the Earth in kelvin degrees. The skin temperature is the theoretical temperature needed to maintain the surface energy balance. It depicts the temperature of the uppermost surface layer, which has no heat capacity and thus, it responds instantly to variations in surface fluxes.

While working with the above-mentioned data products, some peculiarities were observed:

- Acquisition time of images: The time of images acquisition is given in different measurement time scales among the thermal data products. Specifically, MODIS Terra/ Aqua Day/Night LST acquisition time is given in local solar time, which is the time based on the sun's position in the sky in relation to a specific location on the ground. In contrast, the acquisition time of ECOSTRESS LST and Sentinel-3 SLSTR L2 LST products is provided in UTC Greenwich time zone. The models that are developed within AFRI4Cast require the acquisition time to be given in local solar time. Therefore, a conversion code between UTC Greenwich time zone and local solar time had to be created.

- Quality of images: The qualitative assessment of the thermal data products can't be conducted through photo interpretation since no specific spectral targets are visible, as it is in the case of multispectral and SAR images. In particular, the quality estimation of their geometric accuracy relies on the accuracy of the algorithms that carry out the reprojection, the co-registration and the resampling processes. Moreover, the qualitative assessment of their radiometry is based on the corresponding quality layers provided by each thermal data product. These quality layers encode in bits a set of quality criteria concerning the cloud coverage probability, the LST measurement uncertainty etc. Based on these layers and the defined methodology (Chapter 3.5. of this document), the quality mask of each thermal data product is derived, which is further utilized for the value masking of each product, i.e. the rejection of low-quality pixels.

- Data scaling: Thermal data products present a specific radiometric accuracy deriving from the corresponding sensor. In the case of the thermal data products that are used within AFRI4Cast, their radiometric accuracy is maintained through the use of the UInt16 as their data type. This means that their size is reduced but without information loss. During the data conversion from raw values to scaled integers (data scaling) there is a specific radiometric accuracy, which should be preserved during the multiple data conversions that take place in order to perform lossless conversions and support the functionality of the developed algorithms and models through the maintenance of data accuracy. Therefore, linear conversion equations are adopted in order to convert the LST data from UInt16 data type to Float32 data type based on the scaling factors that are provided for each thermal data product (Table 2. & Table 3.).

- ERA5-Land REA: Diurnal Temperature Cycle (DTC) equation modeling presented the peculiarity that the result of the curve fitting algorithm was prone to errors due to the special algebraic form of the equation. It was observed that the DTC equation modeling required good estimates of the parameters that are used as initial values. Thus, apart from the thermal radiance data, ERA5-Land REA Skin Temperature data were used as initial values for the parameters' optimization in iterative models since they present the advantage of covering the daily temperature cycle with a frequency of 1 hour, so a conclusion about the daily minimum and maximum temperature is easily derived despite their coarse spatial resolution. Moreover, it is worth mentioning that ERA5-Land REA Skin Temperature variable measurements are close enough to the LST measurements provided by the thermal images.

2.3. Data Processing Pipelines

The thermal data processing and fusion pipelines are depicted in Figure 1.

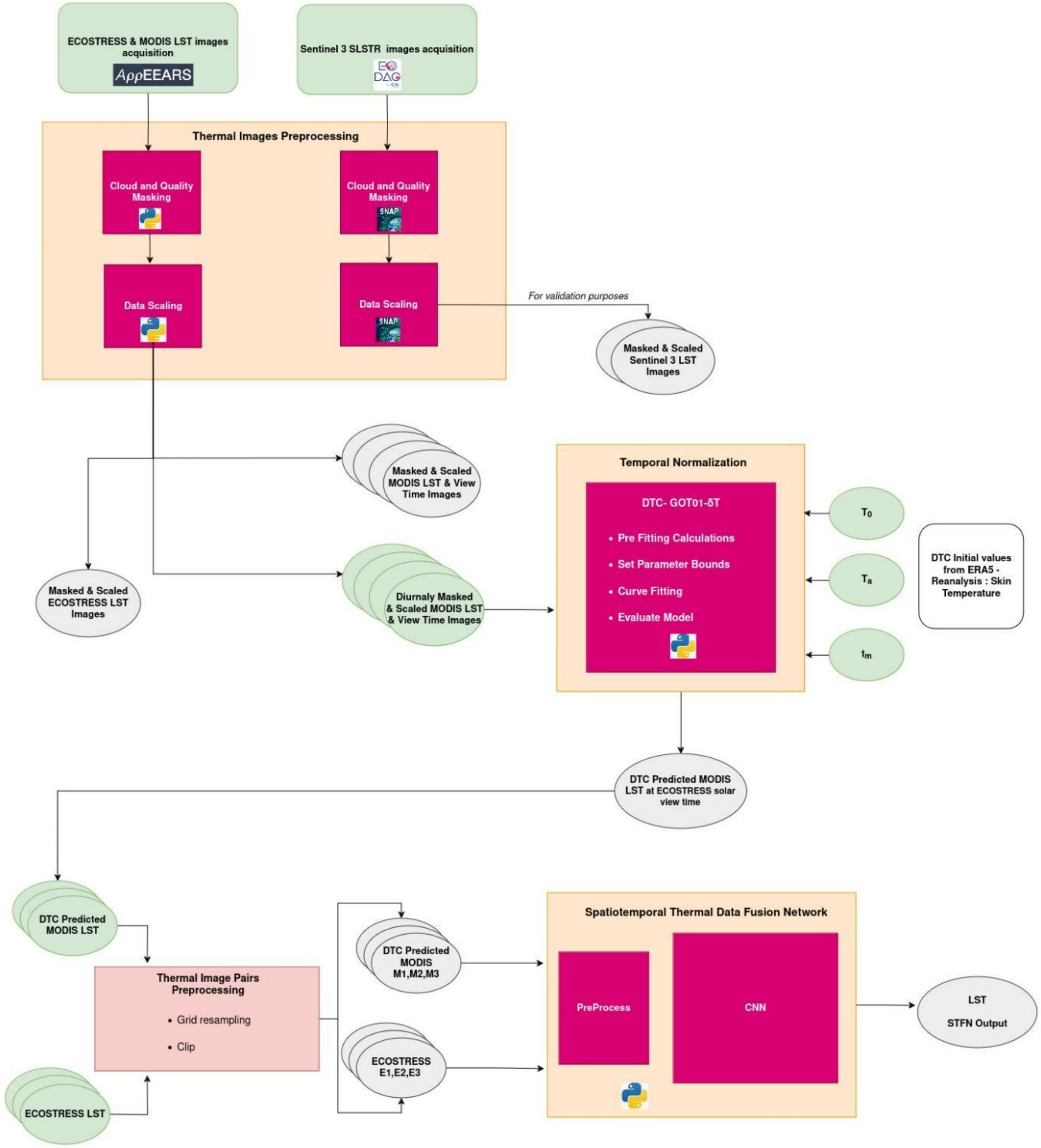


Figure 1. Thermal data processing and fusion pipelines.

2.4. Thermal Images Acquisition

Approach:

In this chapter, the tools adopted within AFRI4Cast for thermal images acquisition are presented. The tools refer to the distinct APIs that are exploited in order to download the thermal data products or their metadata supporting the realization of the data processing pipelines. Aspects such as image availability, acquisition tools capabilities and the finding of appropriate quality data for the creation of image pairs to serve as input in the Spatiotemporal Thermal Data Fusion network (STTFN) are displayed as well.

Acquisition Tools – AppEEARS API

Within AFRI4Cast, AppEEARS API, an Application Programming Interface allowing users to interact with AppEEARS data archives, was exploited for the acquisition of ECOSTRESS and MODIS Terra/ Aqua LST data products via custom Python scripting.

The Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) provides users with a convenient way to access and retrieve geospatial data from multiple Earth observation data sources. Through AppEEARS, users can subset geospatial satellite datasets with the provision of geographical, temporal and dataset layers' parameters. There are two different kinds of sample requests: point samples with the use of geographic coordinates and area samples for spatial areas via vector polygons, with the latter being used within AFRI4Cast.

The script parameters to be defined by the user include:

- The desired directory for the downloaded ECOSTRESS or MODIS Terra/ Aqua LST data in .tif format
- The product name & version in the format ProductName.Version
- The desired layers for the chosen product to be requested
- The directory in which the vector polygon file of the AOI is stored in a geojson format.
- The preferred task name
- The observation starts and end date in MM-DD-YYYY format

Regarding the requested product layers, they are presented in Table 2 and Table 3.

Auxiliary files are downloaded as well, such as:

- <ProductName>-granule-list.txt: The list of the requested product layers in .tif format that are downloaded
- <ProductName>-<ProductVersion>-QC-lookup.csv: Interpretation of the QC layer bit flag values that are found in the specific request

- <ProductName>-<ProductVersion>-QC-Statistics-QA.csv: Information about the pixels that belong in each bit flag value of the QC layer
- <ProductName>-<ProductVersion>-Statistics.csv: Statistics about the downloaded product layers -Min, Max, Mean, Standard Deviation, Variance, Quartiles-.

Acquisition Tools - EODAG

In the context of AFRI4Cast, EODAG, the Earth Observation Data and Access Gateway, served as the exploration and retrieval tool for Copernicus Sentinel-3 SLSTR Level-2 LST data products via custom Python scripting.

EODAG is both a command line tool and a Python package designed to ease the process of accessing and acquiring Earth observation data. EODAG facilitates the search, access and download of satellite imagery, while it offers a unified API for Earth observation data access independently of the data provider. Within AFRI4Cast, a custom developed Python script exploits the EODAG Python package and is configured to work with CreoDIAS, Copernicus Dataspace Ecosystem and ONDA DIAS API.

The script parameters that have to be determined by the user are:

- Workspace directory: The desired directory in which the data products will be downloaded
- Extent: The WKT string of the AOI polygon
- Start Date: Observation start Date in YYYY-MM-DD format
- End Date: Observation end Date in YYYY-MM-DD format

Acquisition Tools - Climate Data Store API

Copernicus Climate Data Store (CDS) API is a service providing programmatic access to CDS data archive. Part of this archive is the ERA5-Land REA data product, which was utilized within AFRI4Cast's pipelines. In particular, the product layer "Skin temperature", which depicts the temperature of the surface of the Earth as already mentioned in Chapter 3.2., is the one requested by defining:

- the AOI as a bounding polygon
- the product type
- the product variable
- the time period (year, month, day)
- the requested hours
- the format (GRIB)

Image Availability

Following the delineation of thermal and climate images acquisition tools, the image availability was examined. The image availability aspect involved both the frequency of acquisitions and the quality of images.

Frequency of acquisitions: MODIS LST data products have a daily and diurnal temporal coverage, with their acquisition times to be at around 10:30 (MOD Day View Time) and 22:30 (MOD Night View Time) for Terra and at 13:30 (MYD Day View Time) and 1:30 (MYD Night View Time) for Aqua (local solar time). Sentinel-3 SLSTR L2 LST have a daily and diurnal temporal coverage as well, with their acquisition times to be at around 7:00 – 8:00 and 20:00 – 21:00 (UTC Greenwich time). On the other hand, ECOSTRESS data products do not present a consistent pattern on temporal coverage over the study sites either regarding the frequency of acquisitions or the acquisition times.

Quality of images: ECOSTRESS and MODIS Terra/ Aqua LST image quality was examined with the most essential criterion to be the least possible cloud coverage in order to search and acquire images with the best possible quality for the creation of the test dataset (refer also to Chapter 1 & 3.6.). Image cloud coverage was examined in a manual way through the images metadata.

Issues :

As the zero or almost zero cloud coverage criterion was established (refer to Chapter 1.) in order to enhance the performance and the functionality of the SpatioTemporal Thermal Data Fusion Network, this had an effect on locating the available images that meet this criterion simultaneously. Consequently, the issue of finding available ECOSTRESS and MODIS Terra/ Aqua LST images that meet all the established quality criteria at the same time proved to be very important.

- AppEEARS and EODAG APIs don’t provide a cloud coverage criterion when it comes to requesting and acquiring ECOSTRESS and MODIS Terra/ Aqua LST images.
- The previous point led to examining the ECOSTRESS and MODIS Terra/ Aqua LST acquisitions’ metadata via manual inspection in order to locate the cloud-free images. However, the metadata of each acquisition concerns the whole image and not a subset of it, as it was our test area. Thus, the images had to be acquired, processed and cloud masked in order to draw a conclusion via visual inspection. This made the procedure time consuming and partially-automated.
- The swath of ECOSTRESS LST images was not consistent among the ECOSTRESS LST acquisitions over the defined AOI. This led to the need of manual intervention in order to find out the ECOSTRESS and MODIS Terra/ Aqua LST acquisitions that spatially coincide.
- Moreover, as the ECOSTRESS and MODIS Terra/ Aqua LST data products are provided in different time measurement scales, a time conversion had to take place in order to proceed.

2.5. Thermal Images Preprocessing

Challenge:

The challenge encountered during the thermal images preprocessing included the close examination of each data product, its specifications and further on, the delineation of the necessary preprocessing steps for each data product. Part of the challenge that was faced was the development of the algorithm codes for the realization of the preprocessing pipelines in order to make the thermal images compliant with the requirements of the models.

Approach:

The thermal images preprocessing steps included the quality-cloud masking based on the provided quality bit flags in order to eliminate pixels with illegitimate values and the data scaling based on the available scaling factors for each data product. Co-registration of the images is part of the preprocessing procedure as well.

The outline of the thermal images pre-processing approach is described through the below mentioned steps:

- Extraction of bit masks from the data products' quality layers (Bitwise raster algebra)
- Boolean union of the bit masks (Boolean raster algebra)
- Data masking implementation
- LST linear rescaling
- Grid resampling (Nearest neighbor method)

Methodology:

MODIS Terra/ Aqua LST and View Time

MODIS Terra/ Aqua LST and View Time layers acquired from AppEEARS API are already clipped to the extent of the defined AOI. Thus, solely a code for their cloud masking has to be developed.

A custom Python code for the cloud masking and the data scaling of MODIS Terra/ Aqua Day/ Night LST and View Time layers was developed. The cloud masking relied on the QA bit flags provided through the QC layers (Table 4.) that accompany each MODIS Terra/ Aqua Day/ Night LST data product and a bitwise extraction approach based on the decimal equal of the binary representation (key values) for each bit pair. The data scaling was based on the scaling factors that apply to each data product layer (Table 2.). Lastly, the spatial resampling to the ECOSTRESS LST grid system was conducted via the nearest neighbor method.

Table 4. The QA bit flags and their key values for MOD11A1.061 and MYD11A1.061 QC layers *(Source: https://lpdaac.usgs.gov/documents/715/MOD11_User_Guide_V61.pdf).*

Bit-No.	Parameter	Interpretation
1 & 0	Mandatory QA flags	00=LST produced, good quality, not necessary to examine more detailed QA
		01=LST produced, other quality, recommend examination of more detailed QA
		10=LST not produced due to cloud effects

		11=LST not produced primarily due to reasons other than cloud
3 & 2	Data Quality flag	00=good data quality
		01=other quality data
		10=TBD
		11=TBD
5 & 4	Emissivity Error flag	00=average emissivity error <= 0.01
		01=average emissivity error <= 0.02
		10=average emissivity error <= 0.04
		11=average emissivity error > 0.04
7 & 6	LST Error flag	00=average LST error <= 1K
		01=average LST error <= 2K
		10=average LST error <= 3K
		11=average LST error > 3K

The preprocessing steps for MODIS Terra/ Aqua Day/ Night LST and View Time layers' cloud masking, data scaling and spatial resampling are:

- Read the MODIS Terra/ Aqua Day/ Night QC and LST or View Time layers
- Extract specific bits from a value of the QC layer - Creation of bitwise extraction function
- Perform the bitwise extraction & create the bit masks based on the bit pairs and their key values:

  Bits 1 & 0: Mandatory QA mask
  Bits 3 & 2: Data Quality mask
  Bits 5 & 4: Emissivity Error mask
  Bits 7 & 6: LST Error mask

In order to conduct a proper cloud masking, all values bigger than the binary 10 in the bit pair 1 & 0 should be excluded:

- Combine the individual bit masks in a single mask
- Apply the mask to the LST or View Time layer

- Scaling: Multiply all the raster values by their scaling factor, i.e., 0.02 for LST and 0.1 for View Time layer
- Set the no data value of the masked layer
- Set the output file path
- Prepare the metadata for the masked raster file
- Write the masked and scaled array on a .tif raster file in Float32 data type
- Spatial resampling of the masked MODIS LST layers to the ECOSTRESS LST grid system following the nearest neighbor method as the reflectance values should be retained.

ECOSTRESS LST

Following the same approach that was adopted for MODIS LST and View Time layers' quality masking, a custom Python code for the cloud masking and the data scaling of ECOSTRESS LST layers was developed as well, as ECOSTRESS LST data products are downloaded already clipped from AppEEARS API.

The cloud masking was based on the QA bit flags provided through the QC layers (Table 5.) that accompany the ECO2LSTE.001 data product and a bitwise extraction approach that relied on the decimal equal of the binary representation (key values) for each bit pair. The data scaling was based on the scaling factor that applies to the LST product layer (Table 3.).

Table 5. The QA bit flags and their key values for ECO2LSTE.001 QC layers *(Source: https://lpdaac.usgs.gov/documents/423/ECO2_User_Guide_V1.pdf).*

Bit-No.	Parameter	Interpretation
1 & 0	Mandatory QA flags	00 = Pixel produced, best quality
		01 = Pixel produced, nominal quality. Either one or more of the following conditions are met:
		1. Emissivity in both bands 4 and 5 < 0.95, i.e. possible cloud contamination
		2. Low transmissivity due to high water vapor loading (<0.4), check PWV values and error estimates
		3. Pixel falls on missing scan line in bands 1&5, and filled using spatial neural net. Check error estimates. Recommend more detailed analysis of other QC information
		10 = Pixel produced, but cloud detected

		11 = Pixel not produced due to missing/bad data, or TES divergence, user should check data quality flag bits.
3 & 2	Data Quality flag	00 = Good quality L1B data
		01 = Missing stripe pixel in bands 1 and 5
		10 = not set
		11 = Missing/bad L1B data
5 & 4	Cloud/Ocean flag	Not set. Please check ECOSTRESS GEO and CLOUD products for this information.
7 & 6	Iterations	00 = Slow convergence
		01 = Nominal
		10 = Nominal
		11 = Fast
9 & 8	Atmospheric Opacity	00 = >=3 (Warm, humid air; or cold land)
		01 = 0.2 - 0.3 (Nominal value)
		10 = 0.1 - 0.2 (Nominal value)
		11 = <0.1 (Dry, or high altitude pixel)
11 & 10	MMD	00 = > 0.15 (Most silicate rocks)
		01 = 0.1 - 0.15 (Rocks, sand, some soils)
		10 = 0.03 - 0.1 (Mostly soils, mixed pixel)
		11 = <0.03 (Vegetation, snow, water, ice)
13 & 12	Emissivity accuracy (Average of all bands)	00 = >0.02 (Poor performance)
		01 = 0.015 - 0.02 (Marginal performance)
		10 = 0.01 - 0.015 (Good performance)
		11 = <0.01 (Excellent performance)

15 & 14	LST accuracy	00 = >2 K (Poor performance)
		01 = 1.5 - 2 K (Marginal performance)
		10 = 1 - 1.5 K (Good performance)
		11 = <1 K (Excellent performance)

The preprocessing steps for ECO2LSTE.001 cloud masking and scaling are:

- Read the ECOSTRESS QC and LST layers
- Extract specific bits from a value of the QC layer - Creation of bitwise extraction function
- Perform the bitwise extraction & create the bit masks based on the bit pairs and their key values (Table 5.)

Bits 1 & 0: Mandatory QA mask
Bits 3 & 2: Data quality mask
Bits 5 & 4: Cloud/Ocean mask
Bits 7 & 6: Iteration mask
Bits 9 & 8: Atmospheric opacity mask
Bits 11 & 10: MMD mask
Bits 13 & 12: Emissivity accuracy mask
Bits 15 & 14: LST accuracy mask

In order to conduct a proper cloud masking, all values different than 0 in the Bit pair 1 & 0 should be excluded.

- Combine the individual bit masks in a single mask
- Apply the mask to the LST layer
- Scaling: Multiply all the LST raster values by their scaling factor, i.e., 0.02
- Set the no data value of the masked layer
- Set the output file path
- Prepare the metadata for the masked raster file
- Generate the binary mask
- Set the output file path for the binary mask
- Write the masked and scaled LST array on a .tif raster file in Float32 data type

- Write the binary mask on a raster .tif file in Uint8 data type.

Despite the cloud masking, some low temperatures were noticed in the masked LST raster layer in some cases. These temperatures are probably associated with clouds and cloud shadows. Out of this it was inferred that using the Bit pair 1 & 0 may not be enough for a proper cloud masking. Therefore, the combined use of other key values in additional bit pairs should be considered as well.

Sentinel-3 SLSTR L2 LST

In opposition to the data products acquired from AppEEARS API which are provided already clipped to the extent of the defined AOI, Copernicus Sentinel-3 SLSTR Level-2 LST data products have to be processed from the beginning by firstly clipping the data and process them in an appropriate way afterwards. Therefore, an ESA SNAP Processing Graph was developed in order to implement the data clipping, the reprojection and the cloud masking of the data.

The cloud masking was performed by exploiting the bit 14 (summary_cloud) of the *confidence_in.nc* layer that accompanies the Sentinel-3 SLSTR L2 LST data product (Table 6.). It is worth mentioning that a data scaling wasn't applied in this case, in contrast to MODIS and ECOSTRESS LST data products, as the software that we worked on already applies the scaling factor to the LST values during their import. Moreover, depending on the extent of the study area, there is the chance that it intersects with two or more Sentinel-3 grid tiles. In this case, a mosaicking process of the distinct Sentinel-3 LST masked images took place. At last, as masked Sentinel-3 LST images were utilized for masked MODIS LST images validation purposes, they were resampled to the MODIS LST grid system. In this case, the Sentinel-3 LST image constituted the target raster layer and the MODIS LST layer played the role of the reference grid layer.

Table 6. confidence_in.nc layer word definitions and masks for each bit *flag (Source: Product Data Format Specification – SLSTR Level 2 Land Products).*

Bit	Text Code	Description
0	coastline	coastline in field of view
1	ocean	ocean in field of view
2	tidal	tidal zone in field of view
3	land	land in field of view
4	inland_water	inland water in field of view
5	unfilled	unfilled pixel (1 if this pixel is never tested or filled)

6	-	(spare)
7	-	(spare)
8	cosmetic	cosmetic fill pixel
9	duplicate	pixel has a duplicate
10	day	pxel in daylight
11	twillight	pixel in twilight
12	sun glint	sun glint in pixel
13	snow	snow
14	summary_cloud	summary cloud test
15	summary_pointing	summary pointing

The preprocessing steps for Sentinel-3 SLSTR L2 LST are:

- Subset: Spatial /Spectral subset of input image. Subset polygon boundaries given by user in WKT string
- Reproject
- Quality Masking based on the *confidence_in* metadata layer
- Write Output in .tif format in Float32 data type
- Mosaicking of the different Sentinel-3 LST masked images
- Spatial resampling of the masked and mosaicked Sentinel-3 LST image to the MODIS LST grid system with the nearest neighbor method, as the reflectance values should be maintained

ERA5-Land REA Skin Temperature

In order to prepare the ERA5-Land REA data for further use in the Diurnal Temperature Cycle modeling, some important preprocessing steps were followed. These steps included the AOI definition and the downloading of the data from the Copernicus Climate Data Store (CDS) as well as the retrieval of the Skin Temperature variable. An important step was the format conversion from GRIB to netCDF in order for the data to be easy for use in the next preprocessing steps. GRIB (GRIdded Binary, also known as General Regularly-distributed Information in Binary form) is usually used as a condensed data format in meteorology for storing historical and forecast weather data and GRIB files are collections of self-contained records of 2D data. After that, processes such as linear interpolation between the timesteps, calculation of key statistics and conversion between the different measurement time scales were performed.

The preprocessing steps of the ERA5 - Land REA data product, Skin temperature layer include:

- Download ERA5-Land Reanalysis data from CDS
- Define Area of Interest -> Get WKT -> <ymax>,<xmin>,<ymin>,<xmax>
- For the DTC study, download two consecutive days of data
- Retrieve Hourly data of Skin Temperature
- Inspect GRIB/GRIdded Binary file and translate format to netCDF. Create a relative time axis.
- Concatenate 2 days of data, to capture the full DTC
- Perform linear interpolation between timesteps
- Calculate Statistics, from which T Maximum and T Minimum layers of the DTC are computed
- Calculate Argmax of Temperature (tm), which is the actual time where T is maximum
- Time values in ERA5 are Greenwich UTC - tm converted to local solar time

Issues and Next Steps:

Issues:

- In some cases of ECOSTRESS LST images with cloud coverage presence, low LST values were observed in the masked LST raster layer despite the cloud masking process. Clouds and cloud shadows are probably connected to these values.
- Concerns regarding the effectiveness of the ECOSTRESS masking procedure were raised, as it seems that there are situations where the current masking technique, i.e., solely the use of Bit pair 1 & 0, fails to omit cloud pixels from the ECOSTRESS LST images.

Next Steps:

As the above issue is considered important with regard to the implementation of the rest procedures, the following next steps could take place:

- Thorough evaluation of the ECOSTRESS masking process in order to Identify flaws and weaknesses in the current masking algorithm.
- Implementation of modifications or fine-tuning of the masking parameters, i.e., the QC bit pairs and their key values, in order to improve the overall performance of the ECOSTRESS LST masking process. In particular, the use of different key values in additional bit pairs in combination with the key value of the Bit pair 1 & 0 will be considered.
- Validation of the updated ECOSTRESS LST masking procedure to ensure its efficacy in omitting cloud pixels accurately.

2.5.1. Thermal Images Acquisition/Preprocessing - Update iteration

Reviewing the results obtained during the development of tools and pipelines for the acquisition and processing of ECOSTRESS thermal imagery revealed issues related to the procedure of quality masking of these data which makes them an Analysis Ready Dataset. The approach taken in the first place was through the creation of a mask from the QA bit flags of the relevant layer. However, as is evident from the relevant table above (table 5), the use of the elements in the bit pair 1 & 0 is not sufficient to extract the cloudy pixels. There is a specialized bit pair in positions 4 & 5 related to the Cloud/Ocean flags but it is inactive in the Level 1 products used. This presumably happens to be the reason why no automatic cloud masking is performed by the Appears API. The following image (Figure XXX1) shows the acquisition of an ECOSTRESS L1 image dated on 26/07/2022 and at 10 : 48 UTC time. From the color scale of the map legend, as well as from the associated histogram of pixel values (Figure XXX2) it is obvious that , on a fairly warm day, some pixels with temperature values in the range of [0,10] degrees Celsius are present, which is unrealistic . Due to their characteristic shape and distribution it is obvious that these patterns represent clouds that were not masked during the application of QC extraction.

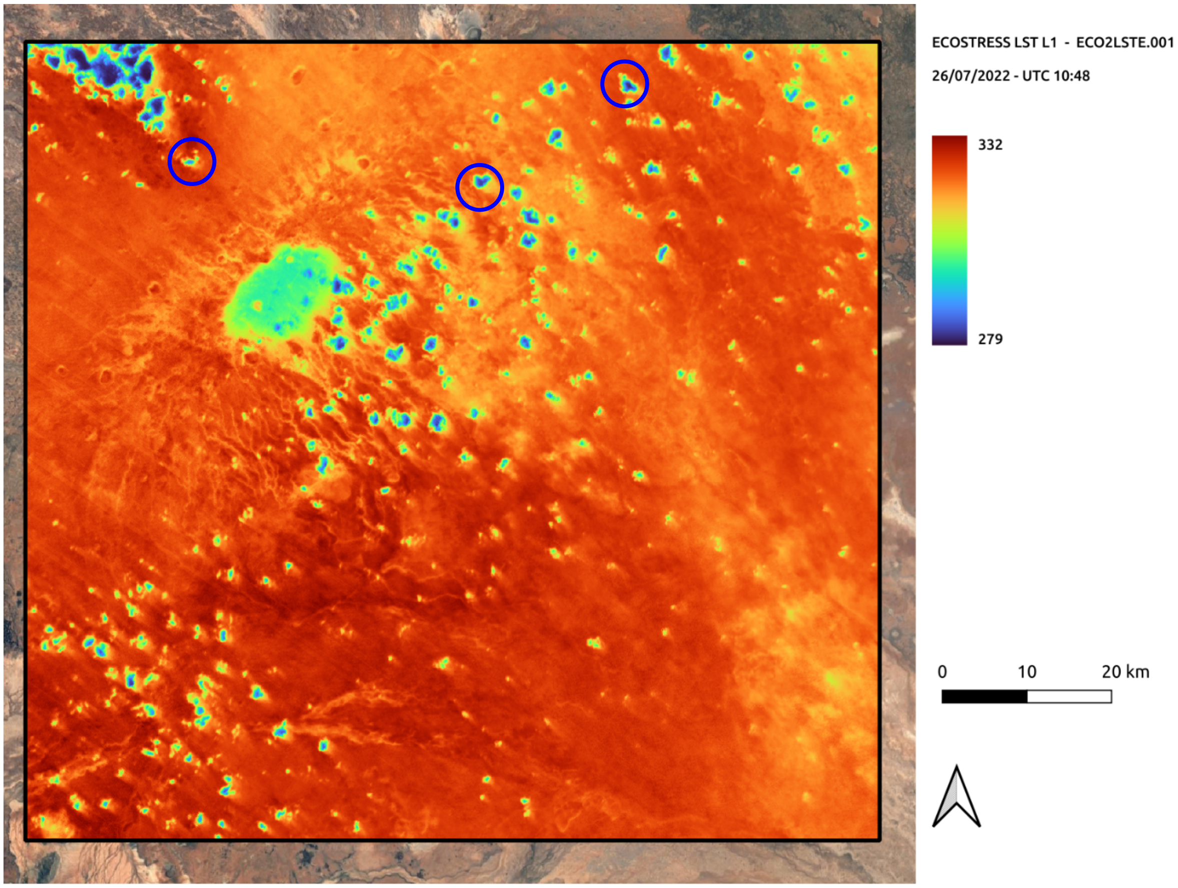


Figure XXX1. ECOSTRESS LST Level 1 image dated on 26/07/2022 and at 10 : 48 UTC time with cloudy pixels not properly masked.

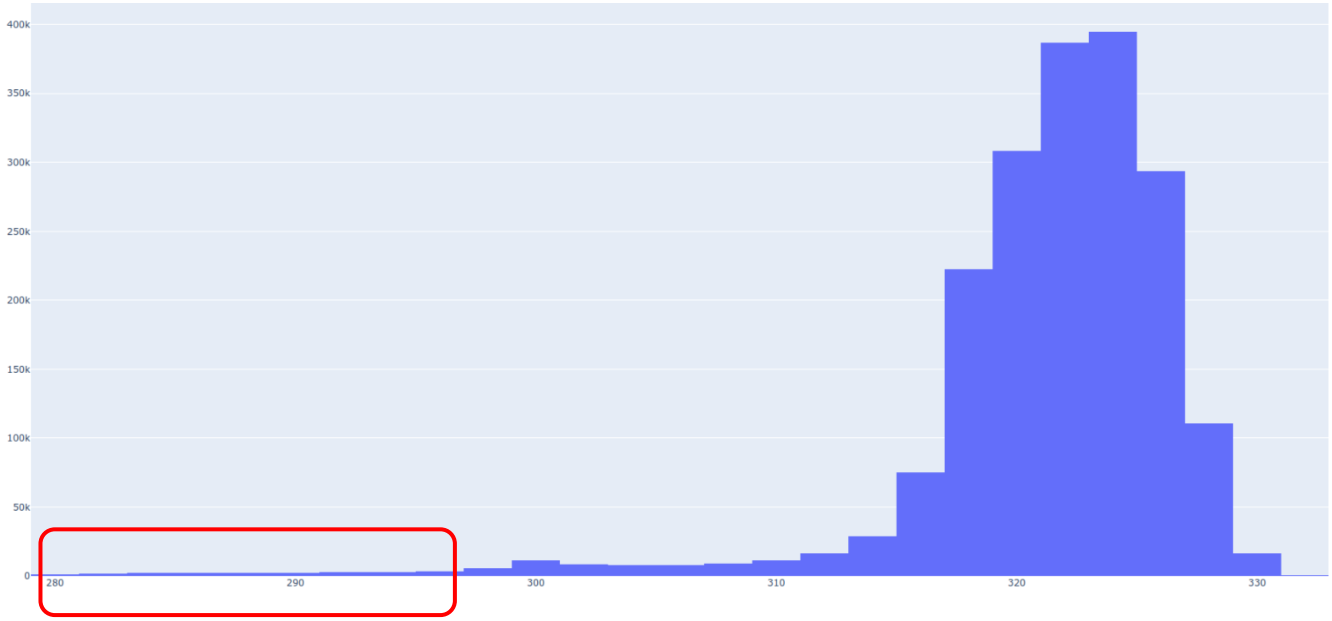


Figure XXX2. Pixel value Histogram of ECOSTRESS LST Level 1 image dated on 26/07/2022 and at 10 : 48 UTC time with cloudy pixels (noted in red frame) not properly masked.

Regardless of the effect cloudy pixels have on the DTC Fitting and Spatiotemporal Thermal Fusion procedures and processes, it was considered appropriate to address this issue by improving the data pipeline for the ECOSTRESS LST images for their generic use . The issue resolution turned out to be quite simple, as the product ECO_L2T_LSTE v002 was finally decided upon.

The ECOSTRESS Tiled Land Surface Temperature and Emissivity Instantaneous Level 2 Global 70 m (ECO_L2T_LSTE) Version 2 data product provides atmospherically corrected land surface temperature and emissivity (LST & E) values derived from five thermal infrared (TIR) bands. The ECO_L2T_LSTE data product was derived using a physics-based Temperature and Emissivity Separation (TES) algorithm. This tiled data product is subset from the ECO_L2G_LSTE data product using a modified version of the Military Grid Reference System (MGRS) which divides Universal Transverse Mercator (UTM) zones into square tiles that are 109.8 km by 109.8 km with a 70 meter (m) spatial resolution. The ECO_L2T_LSTE Version 2 data product is provided in Cloud Optimized GeoTIFF (COG) format, and each band is distributed as a separate COG. This product contains seven layers including LST, LST error, wideband emissivity, quality flags, height, and cloud and water masks.

Based on the Cloud Mask layer provided by the ECO_L2T_LST.002 product , masking on LST images is enabled, which is what the Appears API does by default. For that basic type of preprocessing performed there, the Cloud Mask layer is used which is given as an 8-bit unsigned integer with boolean values, denoting 0 as Clear and 1 as Cloudy. The following figure (Figure XXX3) illustrates the same acquisition, with properly masked cloud pixels. A "zoomed in" view of the image reveals pixels with lower LST values relative to their nearest neighbors, which are presumed to be cloudy but have not been labeled as such by the mask. The cloud mask is plotted in color black. The distribution of those pixels around the perimeter of the mask, and not in a particular direction with respect to orientation, suggests that they are not cloud shadows , but rather misclassified cloud pixels due to some kind of ambiguity that exists around the threshold value at which the cloud mask discretization is performed.

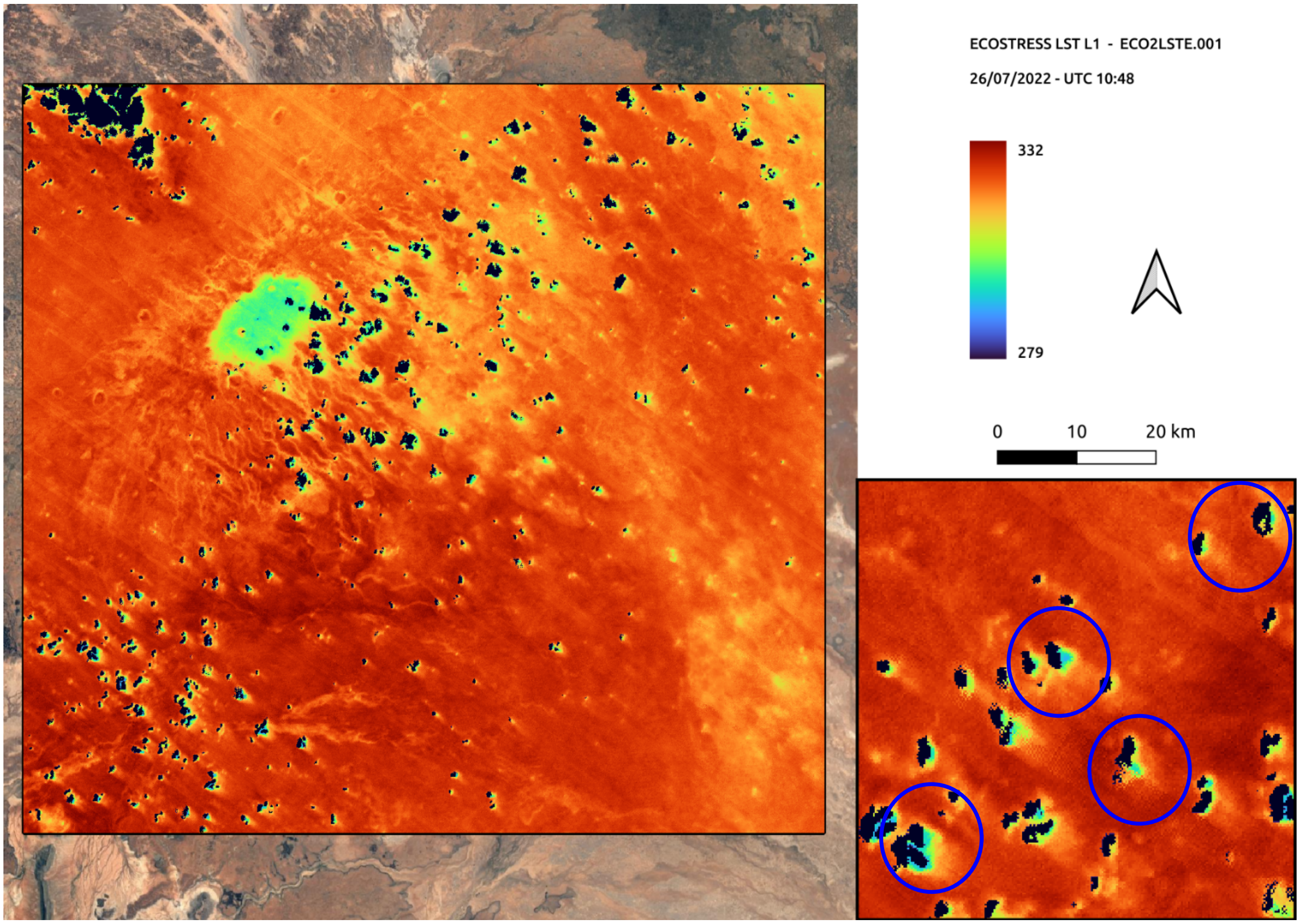


Figure XXX3. ECOSTRESS LST Level 2 image dated on 26/07/2022 and at 10 : 48 UTC properly masked from clouds. In detail : Cases of remnant cloudy pixels

It should be noted that all users of ECOSTRESS LST Level 2 products (ECO_L2T_LSTE, ECO_L2_LSTE, ECO_L2G_LSTE) should be aware that the cloud mask information previously available in the Quality Control (QC) layer in v001, is not available in the v002 QC layer. Instead, users should be using the 'cloud_mask' layer in the L2 LSTE product, or the cloud information in the standard cloud mask products (ECO_L2_CLOUD, ECO_L2T_CLOUD, ECO_L2G_CLOUD) to assess if a pixel is clear or cloudy.

To clarify the reasons which bring out this ambiguity in Cloud Masking, an analysis of the Quality Control for LST and emissivity layer values, which appear to occur in these marginal regions of the Cloud Mask, was performed. Cases of interest were thus identified in terms of Mandatory QA mask bits (0 & 1) and LST accuracy mask bits (14 & 15). Regarding the former, the general comment states that Pixel produced by the applied physics-based Temperature and Emissivity Separation algorithm. Certain conditions are met that may degrade the LST accuracy. Furthermore caution should be taken since either one of the following algorithmic conditions are met, and more detailed analysis of other QC bits is recommended:

- The retrieved emissivity in both longwave bands 4 (10.6 micron) and 5 (12 micron) is < 0.95 indicating possible cloud contamination.
- The pixel falls on a missing scan line in bands 1 and 5, in which the radiance was filled using a spatial neural net technique. The user should check error estimates for this pixel to see if they fall within tolerable bounds.
- The pixel had transmissivity less than 0.4 indicating either possible cloud contamination or high humidity, which would result in higher uncertainty in the LST retrieval. The user is encouraged to check error estimates before using this pixel for science analysis.

With respect to the LST accuracy mask, there has been in all pixels the alert of "marginal LST retrieval performance" with 1.5 - 2 Kelvin degrees error. In conclusion, for the cases of using ECOSTRESS LST data for agricultural applications, such as for example, Growing Degree Days or Crop Evapotranspiration computations, the application of Cloud Masking in the ECO_L2T_LSTE.002 (L2) product is recommended to avoid contamination with unacceptable pixels when calculating the average values per plot. To further discard the pixels at the mask margins, it is recommended to apply a Dilation Morphological filter to the Cloud Mask layer before the actual masking. For more specific cases such as the fitting of a Diurnal Temperature Cycle model to Thermal Image data or the Spatiotemporal Fusion of ECOSTRESS with MODIS LST images, the issues of Cloudy pixels are discussed in the relevant chapters that follow.

3.6. Creation of Test Dataset

The creation of a complete dataset for testing the Spatiotemporal Thermal Data Fusion Network (STTFN) involved the request and acquisition of three high/low spatial resolution image pairs (ECOSTRESS/MODIS) based on the established criteria (refer to Chapter 1. above) for example, the least possible cloud coverage.

Preliminary steps

Selection of a specific AOI in Kenya with 100 x 100 km size that belongs to a chosen ECOSTRESS swath. The data availability investigation was carried out for the whole time series of years 2022 and 2023.

Data Collection and Filtering

- *ECOSTRESS:* The investigation was firstly based on the availability of ECOSTRESS LST images for the defined AOI and years. By accessing the metadata of ECOSTRESS LST images, the initial stage involved filtering and selecting dates with cloud-free images in which the swath covers the defined AOI as well in order to ensure high-quality ECOSTRESS LST data (Table 7 & Table 8).

- *MODIS LST:* MODIS Terra and Aqua acquisitions were downloaded and examined closely for the candidate ECOSTRESS cloud-free dates in order for their quality to be assessed as well (Table 7 & Table 8).

- *Temporal normalization:* A crucial step was the consideration of the acquisition times of both ECOSTRESS and MODIS Terra/ Aqua LST images for the candidate dates. Selection criterion was established to identify images where the acquisition times of both ECOSTRESS and MODIS LST images coincided. ECOSTRESS time values, that are originally in Greenwich UTC time, were converted to local solar time.

Dataset Selection

- *Optimal Date Selection:* Based on the established criteria of cloud-free conditions in ECOSTRESS images and at least one corresponding cloud or semi-cloud -free MODIS Terra/ Aqua LST image at matching acquisition times on the same date, a subset of suitable dates was chosen.

For the year 2022, the dates in which ECOSTRESS LST acquisitions met the low cloud-coverage and swath coverage over the selected AOI criteria are presented in Table 7 on the column "Date". The acquisition time of each image in its original time measurement scale, i.e, UTC Greenwich time, and their acquisition time converted to local solar time is given as well. For each selected date, the quality of MODIS Terra/ Aqua Day/ Night LST acquisitions was also evaluated. It was noticed that the ECOSTRESS LST acquisition time coincided mostly with the one of the MODIS Aqua (MYD) Day acquisition. In the cases when time coincidence between the ECOSTRESS and the MODIS LST acquisitions was observed (marked with orange color) and as long as the corresponding MODIS acquisition was found to be cloud-free at the same time, that date was considered a candidate one for the creation of a high/ low resolution image pair.

Nevertheless, it was observed that cloud-free MODIS LST images were mostly provided by MODIS Terra (MOD) Night acquisitions for all the corresponding ECOSTRESS LST dates. However, as their acquisition times were significantly different than the ECOSTRESS ones, the need of temporal normalization of the low spatial resolution LST data to the time of the high spatial resolution data arised. In case it would have been decided to proceed with the year 2022 and based on the requirements placed for the temporal normalization methodology, the most suitable date would be the 27/01/2022 (marked with green color) as all of the four acquisitions for that date (ECOSTRESS and MODIS Terra/ Aqua Day/ Night LST) presented an optimal quality.

Table 7. The dates in which ECOSTRESS and MODIS LST acquisitions met the low cloud-coverage and swath coverage over the selected AOI criteria for the year 2022.

ECOSTRESS LST				MODIS TERRA LST				MODIS AQUA LST			
Date	DOY	UTC Greenwich Time	Local Solar Time	MOD Day Quality Check	Local Solar Time	MOD Night Quality Check	Local Solar Time	MYD Day Quality Check	Local Solar Time	MYD Night Quality Check	Local Solar Time
27/1/202	27	09:59	12.11	√	10.1	√	22.5	√	13.2	√	1.5

2		:57									
6/3/2022	65	18:50:43	20.96	X (no data)	X	√	21.8	√	14.2	X (no data)	X
11/2/2022	70	04:50:37	6.98	√	9.8	√	22.2	X (no data)	X	√	1.2
10/5/2022	130	17:16:21	19.65	√	10.1	√	22.5	√	13.2	√	1.5
26/7/2022	207	10:48:50	13.02	X (no data)	X	√	22	√	14.1	X (no data)	X
22/8/2022	234	11:49:34	14.08	√	10.9	√	21.6	√	13.6	X (no data)	X

Regarding the image availability of cloud or partially cloud-free images as well as the ones of full swath coverage over the defined AOI for the time series of 2023, the availability of ECOSTRESS images was significantly higher than the one of 2022 (Table 8.). The quality of MODIS Terra/ Aqua Day/ Night LST acquisitions for the dates of ECOSTRESS LST images availability was examined in order to create the three high/low spatial resolution image pairs. It was observed that the dates when the ECOSTRESS acquisition is cloud-free but also at least one MODIS Terra/ Aqua Day/ Night acquisition is also cloud-free and its acquisition time coincides with the one of ECOSTRESS are the 17/01/2023, 25/01/2023 and 5/12/2023. Consequently, the high/low spatial resolution image pairs for the creation of a complete dataset for testing the Spatiotemporal Thermal Data Fusion Network (STTFN) were formed for these dates:

- 17/01/2023: ECOSTRESS LST - MODIS Aqua (MYD) Day
- 25/01/2023: ECOSTRESS LST - MODIS Terra (MOD) Night
- 5/12/2023: ECOSTRESS LST - MODIS Terra (MOD) Night

Lastly, the date 25/01/2023 was chosen as the most suitable for the implementation of the temporal normalization methodology (refer to Chapter 3.7.) as all of the four acquisitions for that date (ECOSTRESS and MODIS Terra/ Aqua Day/ Night LST) presented the best quality in terms of cloud coverage.

Table 8. The dates in which ECOSTRESS and MODIS LST acquisitions met the low cloud-coverage and swath coverage over the selected AOI criteria for the year 2023.

ECOSTRESS LST	MODIS TERRA LST	MODIS AQUA LST

Date	DOY	UTC Greenwich Time	Local Solar Time	MOD Day Quality Check	Local Solar Time	MOD Night Quality Check	Local Solar Time	MYD Day Quality Check	Local Solar Time	MYD Night Quality Check	Local Solar Time
17/1/2023	17	13:29:59	15.56	X (no data)	X	√	21.7	√	14/14.1	√	2.3/2.4
25/1/2023	25	21:50:44	23.96	√	9.9	√	22.3	√	13.1/13.2	√	1.5
20/4/2023	110	00:53:59	3.23	√	10.7/10.8	X (no data)	X	√	14.1	X (no data)	X
21/5/2023	141	00:19:19	2.69	√	10.6	X (no data)	X	X (no data)	X	√	1.3
31/8/2023	243	7:53:57	10.19	√	9.7	√	22/22.1	√	13.9	√	2.2/2.3
4/9/2023	247	06:16:31	8.59	X (no data)	X	√	21.5	√	13.4	√	1.7/1.8
16/9/2023	259	01:30:34	3.9	X (no data)	X	√	21.5	√	13.6	√	1.9
1/12/2023	335	19:10:15	21.67	X (no data)	X	√	22.3	√	13.9/14	X (no data)	X
5/12/2023	339	17:32:59	20.02	√	9.4	√	21.8	√	13.4	√	1.7/1.8

Image Preparation

Both ECOSTRESS and MODIS Terra/ Aqua LST images were, lastly, preprocessed accordingly. Regarding the image preparation, please refer to Chapter 3.5.

3.7. Diurnal Temperature Cycle (DTC) Modeling

Challenge:

In the procedure of retrieving thermal image data for the application of Spatiotemporal Data Fusion algorithms, it was found that ECOSTRESS LST images do not, as a rule, exhibit a fixed acquisition time, but show a random repass pattern. There was some concern raised about their suitability as inputs to Deep Learning data fusion models. The main body of literature on which the STTFN component is based refers to operational LST data from Landsat 8 and MODIS satellites, which receive data almost simultaneously, whereas this is not the case for ECOSTRESS LST. It was considered as an important condition that the HR/LR image pairs were acquired at a sufficiently close time and therefore some computational method had to be applied for the temporal normalization of the Low Resolution LST data to the acquisition time of the High-Resolution data.

Approach:

As the temporal variation of the temperature during a diurnal cycle is not a linear function of time, all the more so because its parameters depend on particular geophysical features, it is not ideally accurate to derive LST values in a reliable way, by using simple data interpolation techniques such as General Linear Models or Splines. For this reason, it was decided to follow the approach of fitting the MODIS data to a Diurnal Temperature Cycle Model, extracting the parameters that optimize it, and then applying it to calculate LST at the time of ECOSTRESS data acquisition. As the need for the above approach arose during the data acquisition process, the methodology was not described in the related deliverable "D3 - State-of-the-art review report on the EO analysis methods and algorithms "and is therefore developed in more detail in this deliverable, in the upcoming chapter.

Methodology:

Diurnal Temperature Cycle Models have the potential to be used to obtain LST at an arbitrary time within the diurnal cycle, by interpolating LST derived from polar orbiting satellites to a diurnal cycle (Jin & Dickinson, 1999). By fitting a DTC model, a set of parameters which describe the thermal dynamics of the land surface are yielded. Once the parameters are determined, surface temperature at any time of a day is able to be predicted. Besides, DTC models are helpful for temporal normalization of remotely sensed LSTs (Jiang, 2007), which was the reason it was utilized in AFRI4Cast.

As a basis, the physical DTC model GOT01 suggested by Göttsche & Olesen (2001) was utilized. This model uses a cosine and an exponential term to describe the effect of the sun and the decrease of surface temperature at night. The exponential term is chosen because it is typical for natural decay-processes, as described by Newton's law of cooling. However, GOT01 describes the variation of LST with five parameters, which makes it less applicable to LST imagery at 2–4 times per day derived from polar orbiting satellite data, like MODIS. Four-parameter (FPD) modifications of DTC models can be applied to tandem polar orbiting satellite observations and have received special attention. Different approaches have been proposed to reduce the parameter number of DTC models to only four. Chang et al. (2019) evaluated 2 FPD modifications of the GOT01 model, which fixed a free parameter to a specified value, and more specifically the day to day change of Residual Temperature or the solar time, for which the free attenuation begins. In our current study the GOT01 - δT DTC modification was implemented, by fitting a Diurnal dataset of MODIS LST images over a defined area of interest.

GOT01 DTC Model Formulation - Goettsche and Olesen (2001)

$$T(t) = \left\{ T_0 + T_a \cdot cos\left(\frac{\pi}{\omega} \cdot (t - t_m)\right), t \right.$$
$$\left. < t_s \; (T_0 + \delta T) + \left[T_a \cdot cos\left(\frac{\pi}{\omega} \cdot (t_s - t_m)\right) - \delta T \right] \cdot exp\left(\frac{-(t - t_s)}{k}\right), t \geqslant t_s \right\}$$

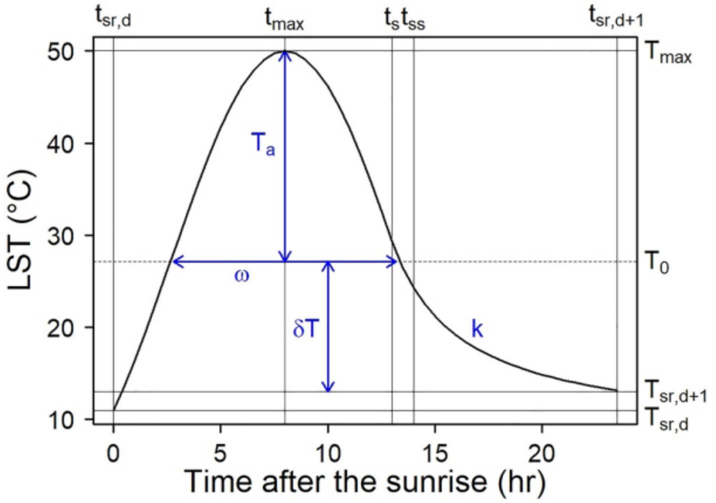


Figure 2. A schematic illustration of the diurnal temperature cycle (DTC) model (Wen, J. et al., 2022).

Free Parameters:

- *T_0: Residual Temperature, [Celsius, Kelvin]*
- *T_a: Temperature Amplitude, [Celsius, Kelvin]*
- *t_m: argmax T(t), [solar time]*
- *t_s: time when free attenuation begins, [solar time]*
- *δT: day to day change of Residual Temperature, [Celsius, Kelvin]*

Computed Parameters:

- *ω: half-period of oscillation of cosine*

$$\omega(\varphi, \delta) = \frac{2}{15} \cdot arccos\big(-tan(\varphi) \cdot tan(\delta)\big)$$

- *φ: geographic latitude*
- *δ: solar declination as a function of DOY*
- *DOY : Day of year*

$$\delta(DOY) = 23.45 \cdot sin\left(\frac{360}{365} \cdot (284 + DOY)\right)$$

- *k: attenuation constant*

$$k = \frac{\omega}{\pi} \cdot \left[arctan\left(\frac{\pi}{\omega} \cdot (t_s - t_m)\right) - \left(\frac{\delta T}{T_a} \cdot arcsin\left(\frac{\pi}{\omega} \cdot (t_s - t_m)\right)\right)\right]$$

GOT01 - δT DTC modification

In order to provide a well posed problem to the curve fitting algorithm, a restriction of at most 4 free parameters, equal to the number of available MODIS LST images in the Diurnal Cycle, and 0 degrees of freedom had to be implied. As a result, the δT parameter was fixed and the model was evaluated for a range of δT values.

Curve Fitting Algorithm

The procedure of fitting MODIS LST data to the GOT01 DTC is a nonlinear least squares (NLS) problem for which the application of Levenberg-Marquardt (LM) algorithm was decided. LM, also known as the damped least-squares (DLS) method, is a numerical optimization algorithm that balances the robustness of Gauss-Newton method with the softer constraints that the Gradient Decent alternative implies on the optimization cost function. If the Hessian of the modeled function is undefined in the function's root, then gradient descent could be applied on it, but not Newton's method. LM contribution is the adoption of a non-negative damping factor λ, which is adjusted at each iteration step. If reduction of the squared error is rapid, a smaller value can be used, bringing the algorithm closer to the Gauss–Newton algorithm, whereas if an iteration gives insufficient reduction in the residual, it can be increased, giving a step closer to the gradient-descent direction. In principle, LM is even more robust than the Gauss-Newton approximation, which means that in many cases it finds a solution even if it starts very far off the final minimum. It converges faster than first-order methods. However, like other iterative optimization algorithms, the LM finds only a local minimum, which is not necessarily the global minimum. This particular fact was something observed during DTC fitting with MODIS data. Even more, the DTC formula brought a range of ts, tm values for which the arcsin function, and subsequently the DTC could not be defined in. For these two particular reasons:

- Initial Values for DTC parameters were provided by ERA5 Reanalysis data, as the closest approximation of a proper initial guess for the algorithm, to avoid sticking in local minima of the error function.
- A modification of the Levenberg - Marquardt algorithm, with support for lower and upper parameter bounds was preferred.

Algorithm Evaluation

The standard procedure for evaluating the performance of a DTC fitting algorithm should entail the comparison of the result predicted by the model with the LST values obtained from an external dataset with higher accuracy. Usually, the data are derived from in situ ground truth measurements of upwelling and downwelling longwave radiation, measured by net radiometers mounted on Eddy Covariance Flux Tower stations. In the case of the Afri4Cast pilot studies no such data sources were found, especially in the areas where cloud free images were available. Therefore, alternatives had to be sought.

The case of evaluation with ERA5 Reanalysis data has the advantage of evaluation in many time intervals of the DTC, but beyond that it is limited by the very low spatial resolution of the data in comparison to MODIS imagery. In such a case, the accuracy of the DTC should have been tested over large areas of temperature homogeneity. In addition, the Skin Temperature variable is not identical to LST and it would not be appropriate to compare their values without applying some transformation.

Consequently, the evaluation with Sentinel 3 LST data from the Copernicus programme was chosen as the target alternative for the assessment. The DTC model prediction at the time of acquisition of the Sentinel 3 LST image was compared with the recorded value to derive the Validation RMS Error. Due to the fairly close acquisition time that the S3 and MODIS LST data had in the region, the Validation RMS Error was expected to be quite close to the model Calibration Residual for the acquisition times of the MOD LST Day and MYD LST Night images. Specifically, the Sentinel 3 LST data query yielded 2 available images, with the first one at 9.77 and the night one at 22.4, in the solar time scale. However, during image processing, all the pixels in the night image were masked out due to cloud cover conditions. For this reason, the RMS Error of Validation could certainly be compared with the Calibration Fitted Residual of the morning MOD Day image.

DTC Model Inputs

LST Calibration Data:

- MOD Day View Time
- MOD LST Day
- MYD Day View Time
- MYD LST Day
- MOD Night View Time
- MOD LST Night
- MYD Night View Time
- MYD LST Night

**It should be noted at this point, that the MODIS LST data and their corresponding View Time layers, have been value masked using the Diurnal Cloud Mask, which resulted as a union overlay of the individual binary masks. The DTC-δT model was tested for cloud free data at 100% of the AOI coverage.*

Ancillary Raster Data Layers:

- Land Cover
- Latitude
- Longitude
- DOY

Validation Raster Data Layers:

- Sentinel 3 LST DAY View Time
- Sentinel 3 LST DAY
- Sentinel 3 LST NIGHT View Time
- Sentinel 3 LST NIGHT

Initial Values Raster Data Layers:

- T_o: T minimum
- T_a: T maximum – T minimum
- δT: Set by user
- t_m: Computed as argmax of T in preprocessing
- t_s: Starting time of free attenuation computed as sunset time -1

DTC Pre-Fitting processing

- Computation of local Sunrise /Sunset time
- Computation of Daylight time (hs) - ω: DTC angular frequency
- Computation of Attenuation Time ts=sunset time-1

Setting of DTC Parameter Bounds

During the computation of the DTC error loss function Jacobian, value domain constraints are posed for the t_s, t_m parameters, in order to provide a real number range of values. The upper and lower bounds for the parameters are given by the system:

$$t_m(Upper) = \frac{\omega}{\pi} + t_s < t_m < t_s - \frac{\omega}{\pi} = t_m(Lower)$$

which has infinitely many solutions, thus tm bounds is a function of ts bounds. The ts bounds were decided relative to t_{ss} sunset time, by setting t_s in $[t_{ss}-2, t_{ss}]$ domain. Therefore:

t_s Lower Bound= $t_{ss}-2$

t_s Upper Bound=t_{ss}

while:

t_m Lower Bound =t_s Upper Bound – (ω/π)

t_m Upper Bound=t_s Lower Bound + (ω/π)

The acceptable value domain for the t_s, t_m parameters that produce non-infinity results is demonstrated on the graph that follows which represents the solution of the parameter inequality system (Figure 3.).

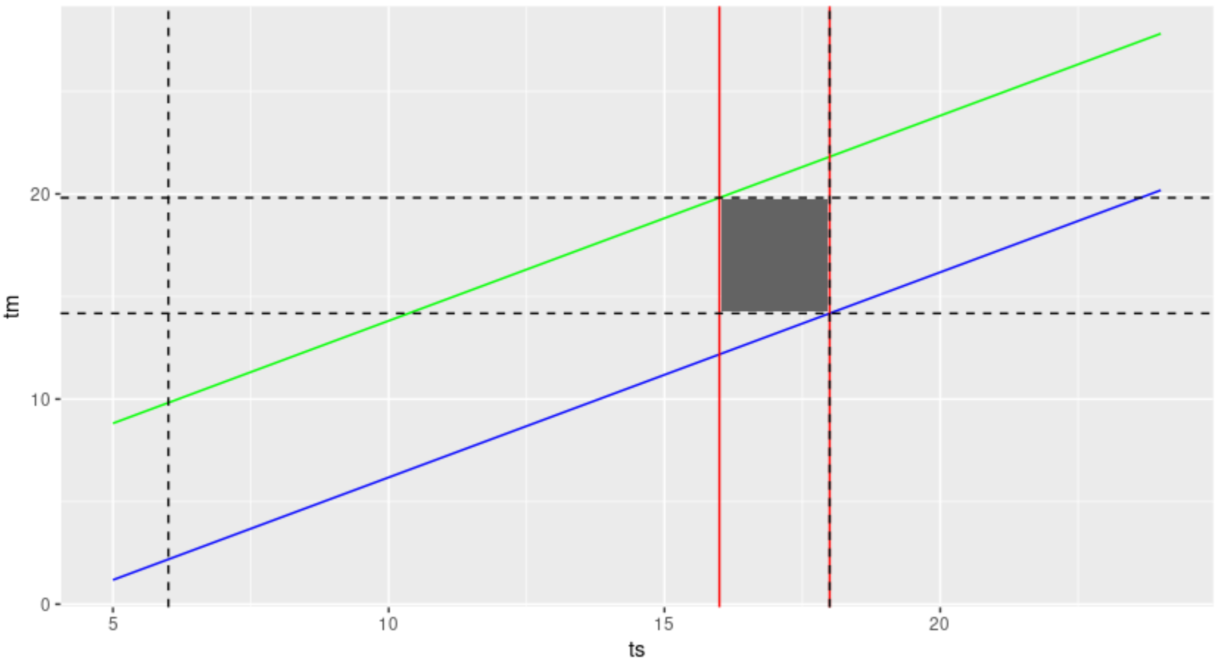


Figure 3. A graphing linear inequality system solution for parameter bound setting of the t_s and t_m.

DTC Model Outputs

- Estimated Model parameters: T_o, T_a, t_m, t_s, δT
- Residuals: Res MODISt1, Res MODISt2, Res MODISt3, Res MODISt4
- Residual Sum of Squares: RSS
- Root Mean Square Error of Calibration: Calibration RMSE
- Log Info Flags: termination of algorithm, convergence issues etc.
- Niter: Number of iterations

- Fitting Bias: Residuals normality - Shapiro-Wilk normality test ($p<0.05$)
- Root Mean Square Error of Validation with S3 LST: Validation RMSE

Experimental Results:

The experimental design for the development and evaluation of the DTC- δT model was determined by the availability of ECOSTRESS and MODIS LST data. The model was calibrated with 4 MODIS LST images of the DTC that occurred on 25 January of 2023 to predict a MODIS-DTC image at the ECOSTRESS acquisition time. As it is evident on the following table, the modeling of the DTC exponential decay was modeled with a past-midnight MYD LST image acquired on the 26th of January 2023.

Table 9. LST images and acquisition solar times of the experimental setting.

DOY		ECOSTRESS	MOD Day (TERRA)	MYD Day (AQUA)	MOD Night (TERRA)	MYD Night (AQUA)
25	***Local Solar Time***	23.96	9.9	13.1	22.3	25.5

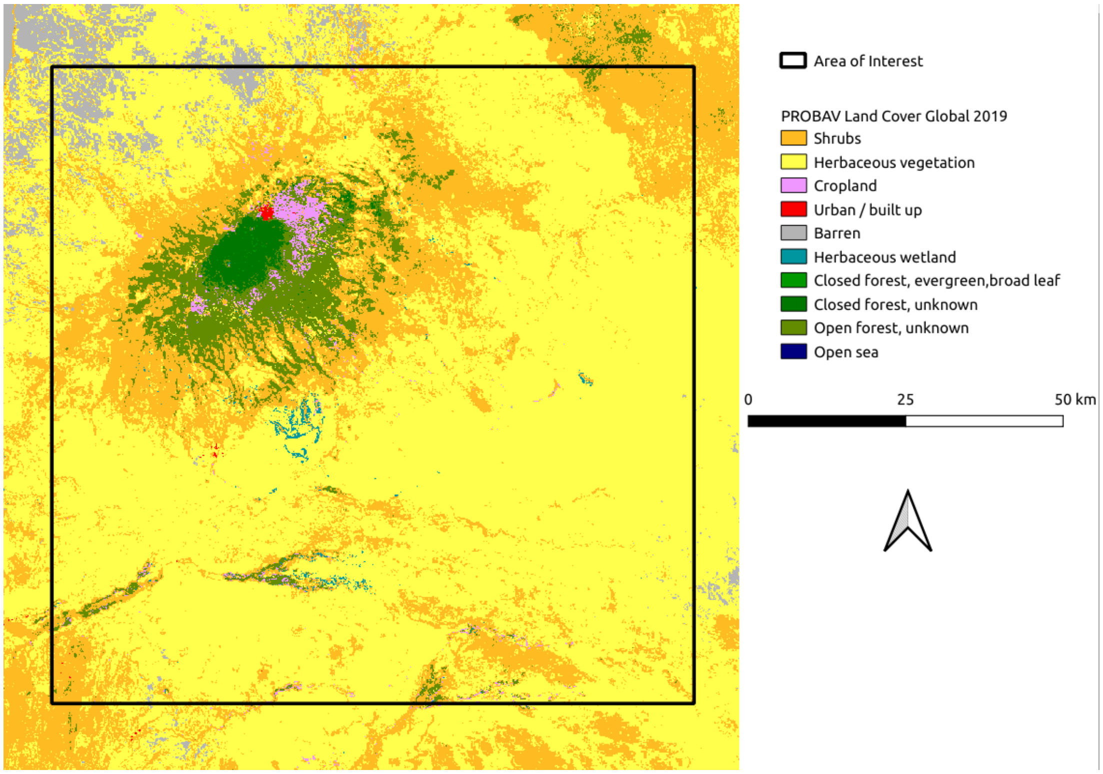


Figure 4. Land Cover map of the Area Of Interest.

The area coverage percentage of each land cover class in the area of interest is given on the following table, while the spatial distribution land cover map is presented on Figure 4.

Table 10. Area coverage percentage land cover classes in the area of interest.

Land Cover Class	Area Coverage Percentage
Barren	1.55
Closed Forest Evergreen Broad Leaf	0.01
Closed Forest Unknown	1.04
Cropland	0.85
Herbaceous Vegetation	64.09
Herbaceous Wetland	0.31
Open Forest Unknown	5.84

Shrubs	22.21
Urban	0.05

Regarding the fixed parameter δT, a grid search approach was applied throughout the model calibration. An assumption about a hidden relationship between land surface thermal inertia and the δT parameter, raised the need to investigate the response of DTC calibration errors with the variation of δT. For a given value, a different calibration error could be expected on the various land cover classes. Viewed from another perspective, a grid search on δT would tune the parameter to its optimal value for each land cover surface. As a result, DTC-δT was calibrated for a range of δT values [-16,8] with an increment of 0.5 Kelvin degrees.

The experimental case study results are presented as partial zonal statistics of the Calibration/Validation RMSE, and the Calibration Residual at each MODIS LST instance, for all the involved Land Cover classes. The zonal analysis involved the calculation of maximum, minimum, Q1/Q3 quantiles, average, and standard deviation metrics, over the whole image. Furthermore, in order to assess the model performance in a more general manner, a global average RMSE statistic is given for Calibration, as a goodness of fit metric, and Validation, as an external data evaluation. A DTC model performance summary is presented on the following Table 11.

Table 11. DTC performance metrics summary.

Land Cover	Herbaceous Vegetation	Herbaceous Wetland	Shrubs	Open Forest Unknown	Urban	Cropland	Closed Forest Unknown	Closed Forest Evergreen Broad Leaf	Barren
max Calibration RMSE	7.58	5.32	6.75	5.69	1.41	5.95	4.17	2.22	4.91
q75 Calibration RMSE	4.18	4.26	3.84	3.86	1.25	3.59	2.82	2.02	3.80
avg Calibration RMSE	3.53	3.85	3.12	3.11	1.16	2.96	2.17	1.82	3.25
std Calibration RMSE	0.92	0.77	0.98	1.02	0.21	1.02	1.02	0.57	0.70
q25 Calibration RMSE	2.86	3.50	2.36	2.48	1.13	2.22	1.56	1.62	3.00
min Calibration RMSE	0.73	1.46	0.42	0.00	0.85	0.64	0.00	1.42	1.22
avg Res MODISt1	6.55	7.38	5.78	5.83	2.25	5.50	2.55	2.45	4.48

avg Res MODISt2	-1.84	-1.71	-1.71	-1.51	-0.51	-1.40	-2.98	-2.67	-4.25
avg Res MODISt3	0.60	0.75	0.46	0.45	-0.13	0.14	-0.09	0.14	-0.04
avg Res MODISt4	0.44	0.52	0.18	0.37	0.19	0.38	-0.04	-0.05	0.89
max Validation RMSE	19.26	18.30	18.55	17.63	4.58	16.88	14.14	NA	13.86
q75 Validation RMSE	13.06	13.65	12.71	11.44	4.46	11.13	8.78	NA	10.39
avg Validation RMSE	10.79	12.38	10.24	9.43	3.97	9.54	6.51	NA	9.19
std Validation RMSE	2.77	2.57	2.99	2.92	0.58	2.77	3.72	NA	1.82
q25 Validation RMSE	8.57	10.48	7.85	7.82	3.38	7.92	3.69	NA	7.65
min Validation RMSE	2.61	7.93	2.20	0.25	3.36	3.21	0.27	NA	5.20
avg Res MODISt2	-1.84	-1.71	-1.71	-1.51	-0.51	-1.40	-2.98	-2.67	-4.25
avg Res MODISt3	0.60	0.75	0.46	0.45	-0.13	0.14	-0.09	0.14	-0.04

The argument of minima for the Calibration RMS Error metric, in δT values, provided insight about the grid search tuning procedure, while the tuning/sensitivity analysis results are presented on the following table, for the most dominant land cover class of Herbaceous Vegetation and the great importance category of Croplands (Table 12).

Table 12. Calibration RMSE argument of minima for the δT parameter.

Land Cover	**Herbaceous Vegetation**	**Herbaceous Wetland**	**Shrubs**	**Open Forest Unknown**	**Urban**	**Cropland**	**Closed Forest Unknown**	**Closed Forest Evergreen Broad Leaf**	**Barren**
δT	0.5	-1	-3	-8	-5	-6	-6.5	-4	-8.5

The distribution of the error trend for different δT values is represented on the screen plot figures that follow, where the land cover classes of major importance are selected. These representations are indicative of the model sensitivity for that parameter, underlying a trend towards negative values.

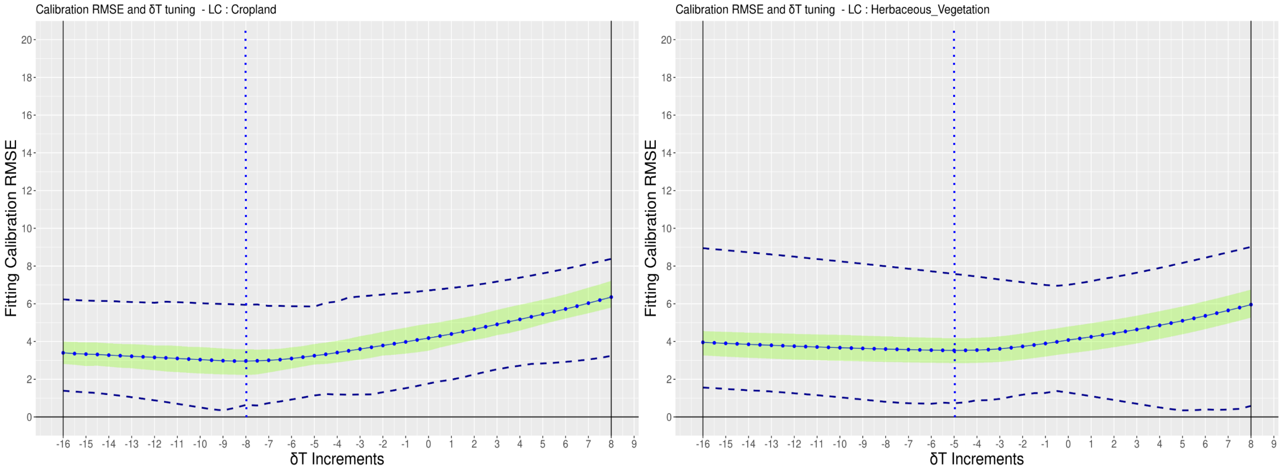


Figure 5. δT Scree Plot for Calibration RMSE. Cropland and Herbaceous Vegetation land cover.

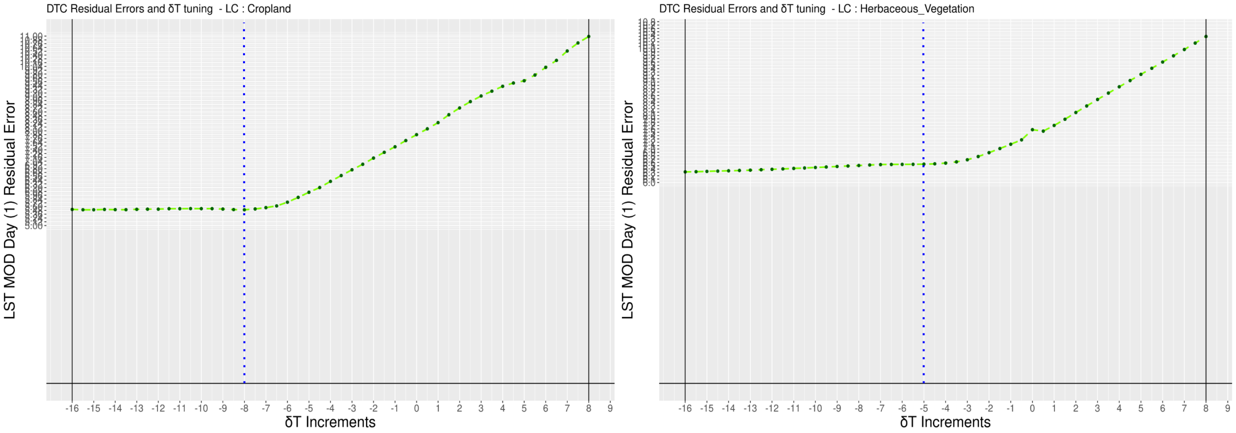


Figure 6. δT Scree Plot for MOD Day Calibration Residual. Cropland and Herbaceous Vegetation land cover.

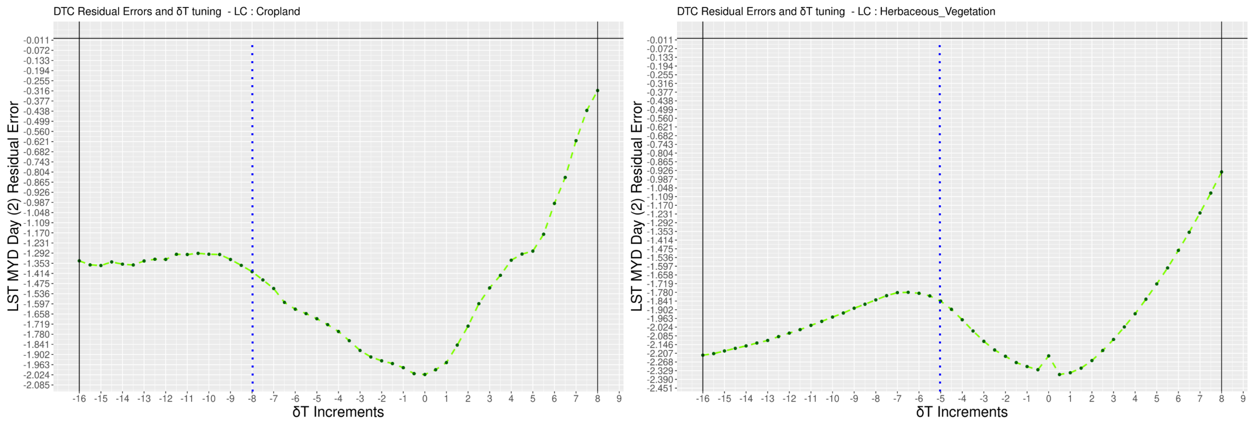


Figure 7. δT Scree Plot for MYD Day Calibration Residual. Cropland and Herbaceous Vegetation land cover.

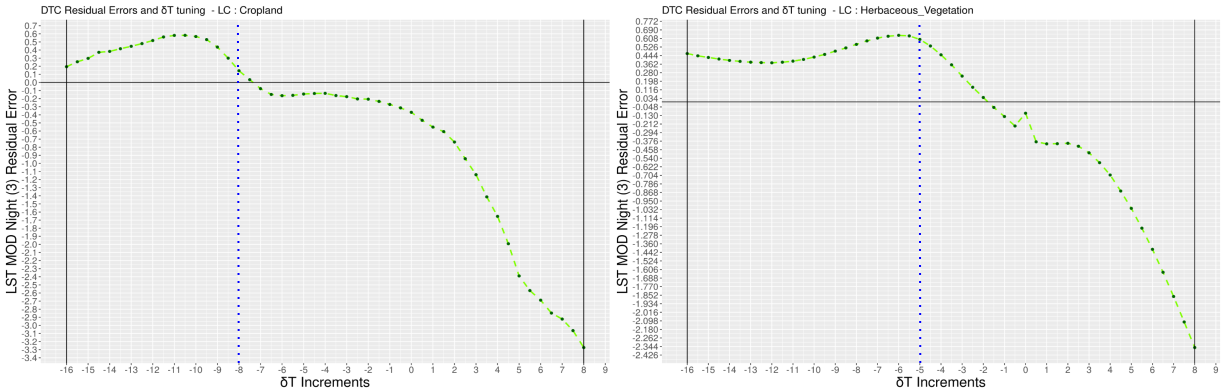


Figure 8. δT Scree Plot for MOD Night Calibration Residual. Cropland and Herbaceous Vegetation land cover.

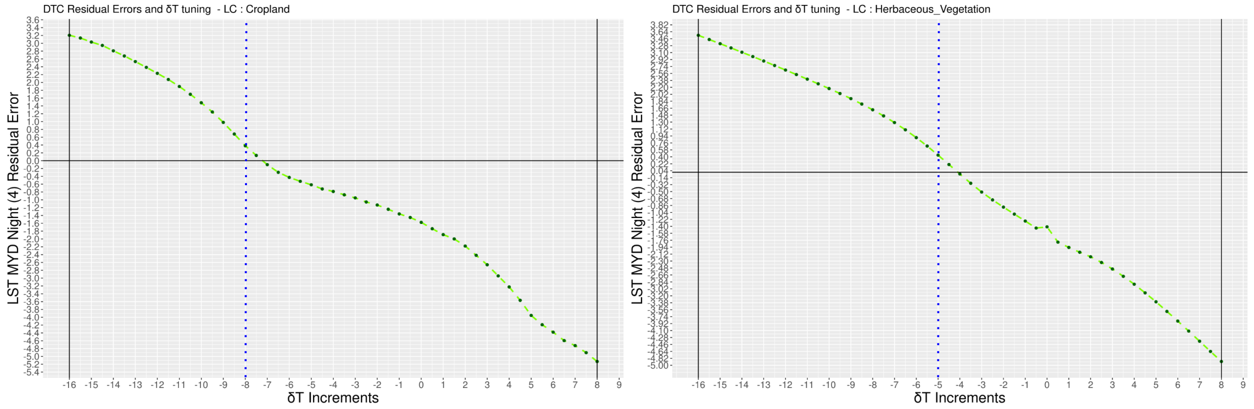


Figure 9. δT Scree Plot for MYD Night Calibration Residual. Cropland and Herbaceous Vegetation land cover.

What was evident, at this point, is that the selection of an optimal δT, as the argument of minima for Calibration RMSE, guaranteed an overall good model fit through all 4 MODIS acquisition times. When focusing, however, on the Night LST images, which also was the approximate time for the ECOSTRESS image, the selected δT provided the best fit to data, among all other values, with minimum residual error.

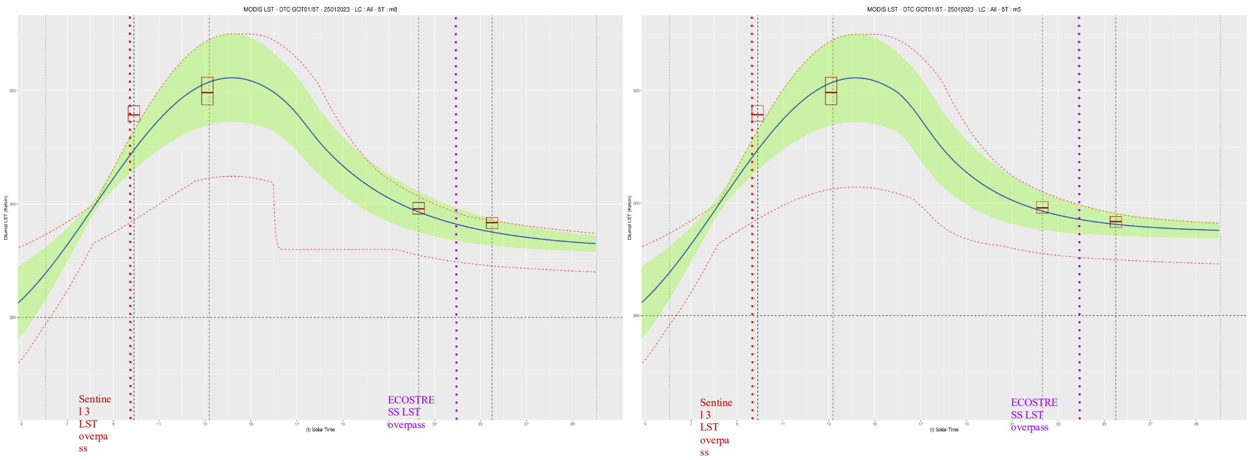


Figure 10. Average Fitted DTC models for δT equal to -8 and -5

By inspecting the average fitted DTC model plots (Figure 11), with the optimal δT values for Cropland and Herbaceous Vegetation respectively, inference can be drawn about under/overestimation qualities of the models. For both Land Cover classes, there was a clear underestimation on the MOD Day LST acquisition time, which is constant on all δT values. In a similar manner, on MYD Day time, the model somehow overestimated the LST values. Finally, on the MYD and MOD night times, the exponential decay part of the model fitted a lot better to the LST data.

Regarding the model Validation results, RMS errors followed a similar trend, but with the absolute values however being a lot higher.

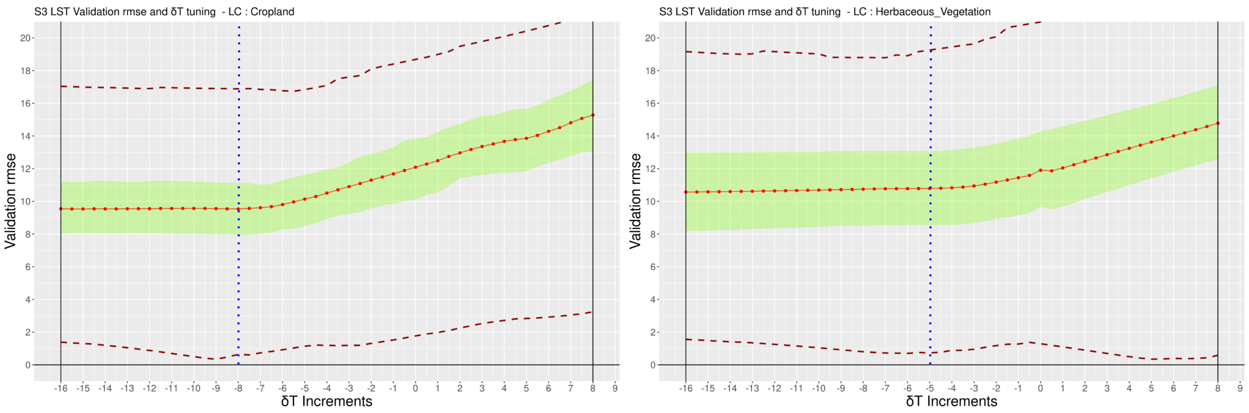


Figure 11. δT Scree Plot for Validation RMSE. Cropland and Herbaceous Vegetation land cover.

As mentioned previously the validation was eventually performed with data from the daytime Sentinel 3 LST acquisition. Given the close temporal coincidence of the MODIS and Sentinel 3 view times, and the underestimation observed in the model, a further emphasis was given in the comparison of the two LST datasets. It was found that the two images showed significant differences between them and that observation was not physically intuitive.

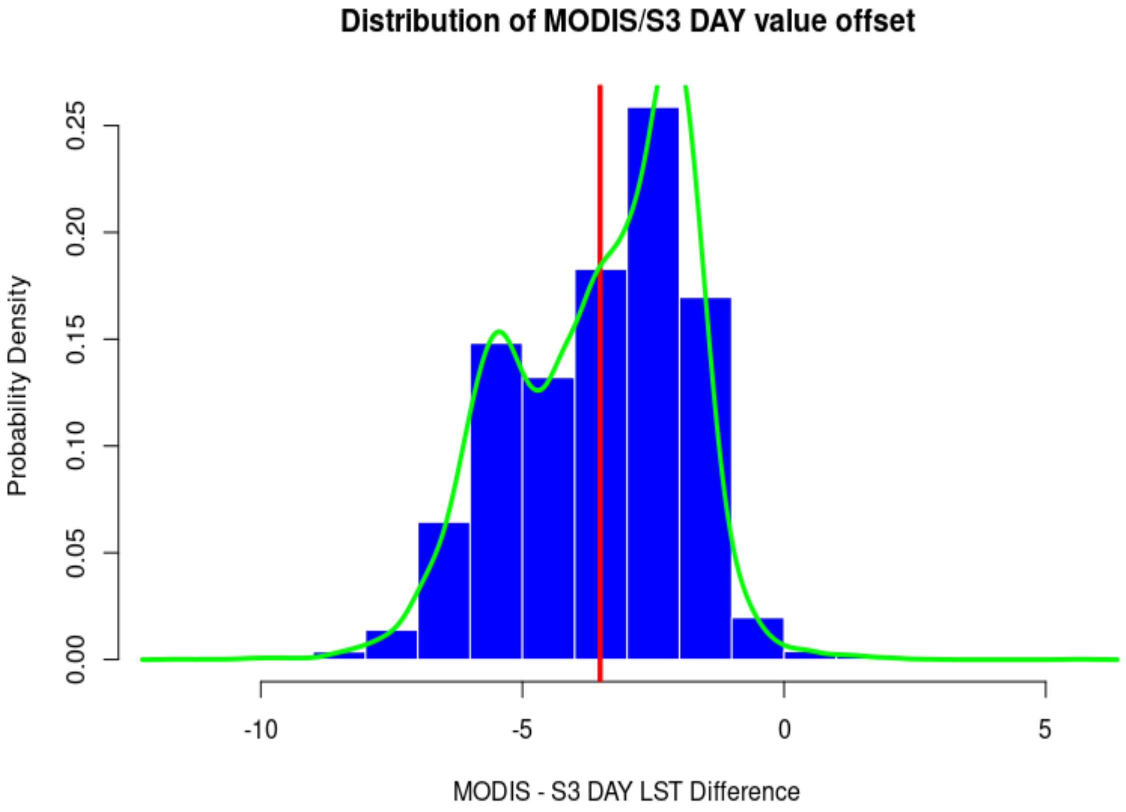


Figure 12. Distribution of the observed offset between MOD Day and Sentinel 3 LST images.

The MODIS image, although acquired a few minutes later than Sentinel 3, showed, according to the paired t test that was performed, values averaging -3.51 degrees Kelvin and the null hypothesis was rejected. In most of their distribution the differences in the values of the two images are negative, which is not expected, since one would expect, for a difference in acquisition time of only 8 minutes, the difference in LST values to be positive and much lower. Given the above remarks, and the inherent uncertainty of the LST values in the images (Shrestha, A. et al. 2018), usually quantified as measurement error, it finally appeared that the Validation method with an external dataset of a single Sentinel 3 LST image could not be reliable without taking into account the measurement errors of the images. It may even be assumed that the very high values in the Validation RMSE reflect the already known underestimation of the model during the daytime hours, for the reason that if the average difference between the two LST images, per land cover class, is added to the Residual Calibration Error of the MOD Day acquisition, then the values of the Fitting Calibration RMSE are closely approximated.

The average RMSE metric for Calibration and Validation respectively, considering all the land cover classes of the AOI, provides an overall performance indicator for the DTC model, was calculated as a weighted mean of the partial summary metrics of Table 12 and it is given by the following values (kelvin degrees):

- **Calibration RMSE: 3.387**
- **Validation RMSE: 10.5**

The mapped DTC prediction is a synthesis of the optimal DTC-δT predictions of each land cover class and it is presented on Figure 13. The provided map shows a visual evaluation of the DTC prediction, when compared between the ECOSTRESS LST image obtained on the target solar view time and the closest MODIS LST image obtained 1 hour 30' earlier. All images are mapped with a common value range color symbology to enable a reliable comparison.

A first conclusion that emerged is that the DTC Prediction image approximates well the values of the ECOSTRESS image, and in this respect, it is obviously significantly different from the values of the MODIS image. This means that at the pixel level, the simulation of the Diurnal Temperature Cycle achieves the temporal normalization, for which it was implemented, to a significant degree. However, image artifacts are still visible, which in the upper left part of the image are quite extensive. Because of their shape, it is suspected that in these blocks, the algorithm converged directly to the initial values given to it, and for this reason the shape of the artifacts resembles in block size the pixel size of the ERA 5 data. It was finally observed that all 3 images maintain the same spatial pattern.

Considering that in the case of the given experiment, the MODIS image is not temporally far from the acquisition of the ECOSTRESS image, different assumptions are made about the usefulness of the "DTC Fitting to MODIS LST data" process to the final task, which is SpatioTemporal Thermal Fusion, depending on the efficiency and the learning mode of the Deep Network. With the present data availability, however, it was not possible to examine this issue in depth for the reason that cases with target view time at different parts of the day need to be considered. The case studied was in any case favorable in the sense that the image acquisition times were close enough, and in a stage of the Diurnal Temperature Cycle during which temperature changes are rather gradual.

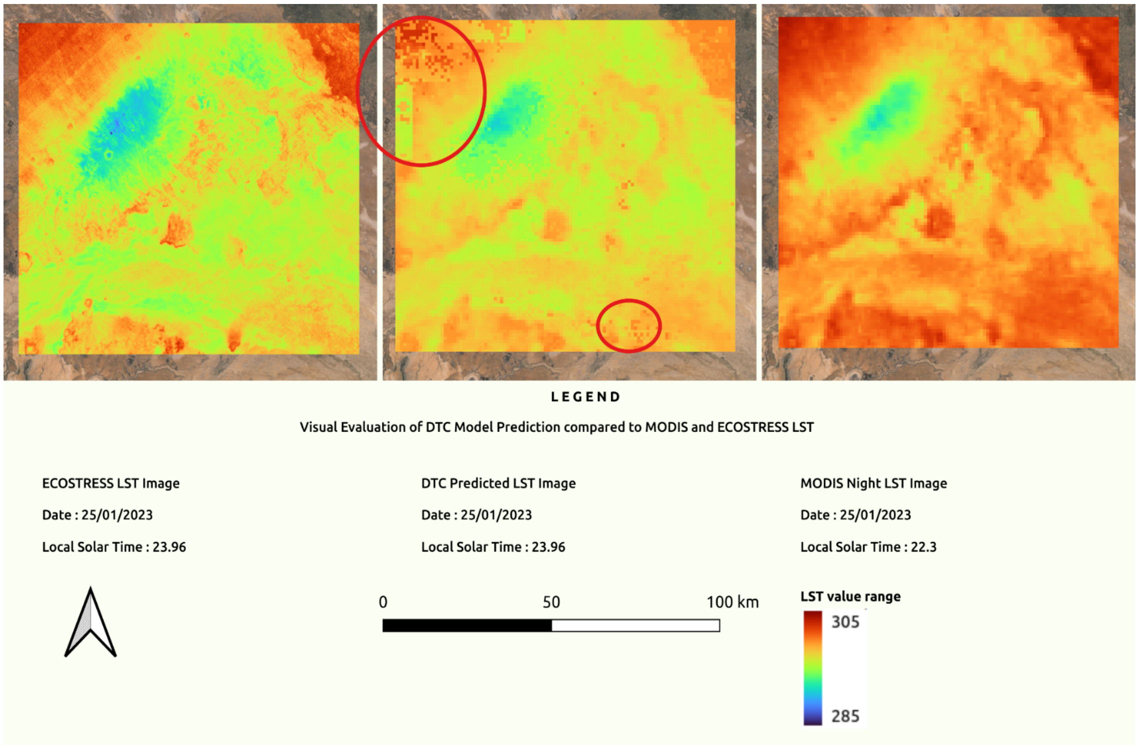


Figure 13. Evaluation of DTC prediction, as compared with MOD Night and ECOSTRESS LST images

Issues and Next Steps:

In conclusion, regarding the process of temporal normalization of MODIS LST thermal images, through their fitting to a DTC model, for their eventual use as training data in deep learning spatiotemporal fusion algorithms, the accuracy results were found to be within the acceptable framework defined by the relevant literature in terms of the parametrization and performance of curve fitting. However, an issue was found related to scene quality in the Sentinel 3 LST data used for independent validation and testing of the calibrated DTC model.

In next steps an extended study could be carried out that includes in its experimental design images at different seasons of the year to determine the stability of DTC accuracy across seasonal change. It will also be important during this 2nd iteration to investigate the existence of higher quality Sentinel 3 LST thermal image acquisitions to perform a relevant independent validation.

3.7.1. Diurnal Temperature Cycle (DTC) Modeling - Update iteration

The objective that was set in terms of the 2nd iteration for the current deliverable, was to perform a more extensive study and analysis on the results concerning the accuracy of the Diurnal Temperature Cycle (DTC) Modeling. In a new experimental design, a study area was selected at the Ahero site, with greater diversity in Land Use/Land Cover classes. Regarding the external validation with Sentinel 3 LST data, in order to avoid the errors that occurred in the 1st iteration, the selection of the AOI was determined by the criterion that it contained a single S3 LST image tile, while still the inherent uncertainty in the LST data used was determined by the Quality Metrics of the original S3 layers. The experimental design included the Calibration and Validation of DTC curve fitting models across different seasons of the year. Due to the geographical location, finding images with purely non-cloudy conditions throughout the day became impossible for all seasons of the year. Furthermore, since the analysis focused mainly on the quantitative evaluation of the error, zero cloud cover was not considered a critical issue, and the study focused on the non -cloudy pixels selected by sampling. In summary,

- 4 seasonal cases of variable Land Cover were selected.
- Minimal Cloud Coverage during day/night acquisitions was desired
- Creation of data frame via stratified random point sampling of 2500 instances among LC classes was implemented
- Dates :
  - Winter : 06 January 2023 ,
  - Spring : 28 May 2023,
  - Summer : 20 August 2023,
  - Autumn : 29 September 2023
- The 4-Parameter Diurnal Temperature Cycle Model GOT01 developed initially on the first iteration was used with a δT free parameter and initial values provided by ERA5 Land Temperature
- A Grid Search on the δT parameter was conducted for Sensitivity Analysis of the models.

Regarding the computational performance of the Levenberg-Marquardt Nonlinear Least-Squares algorithm, log info metadata were summarized providing insight about the reason for termination of algorithm, convergence issues etc. The accuracy of the Levenberg-Marquardt algorithm and its successful completion is controlled by the convergence parameters f_{tol}, p_{tol}, and g_{tol}. These parameters are used in tests which make three types of comparisons between the approximation par,a numeric vector of the initial conditions, and a solution. Algorithm terminates when any of the tests is satisfied. If any of the convergence parameters is less than the machine precision, then the algorithm only attempts to satisfy the test defined by the machine precision. Further progress is not usually possible. The tests assume that the fitting function (the DTC model) as well as its Jacobian matrix are reasonably well behaved. If this condition is not satisfied, then the algorithm may incorrectly indicate convergence.

Table TTT1.MODIS LST images and acquisition solar times of the Ahero experimental setting.

Date		MOD Day (TERRA)	MYD Day (AQUA)	MOD Night (TERRA)	MYD Night (AQUA)
06/01/2023	***Local Solar Time***	9.8	14.2	22.1	25.7
28/05/2023		10.3	14.4	22.7	25.8
20/08/2023		10.2	14.2	22.5	25.6
29/09/2023		9.6	14.2	21.9	25.6

The image Acquisition of MODIS images for the Ahero experimental study is presented on the above table (Table TTT1). The map that follows (Table XXX3) presents the geographic distribution of land cover classes in the Ahero study area for the year 2019.

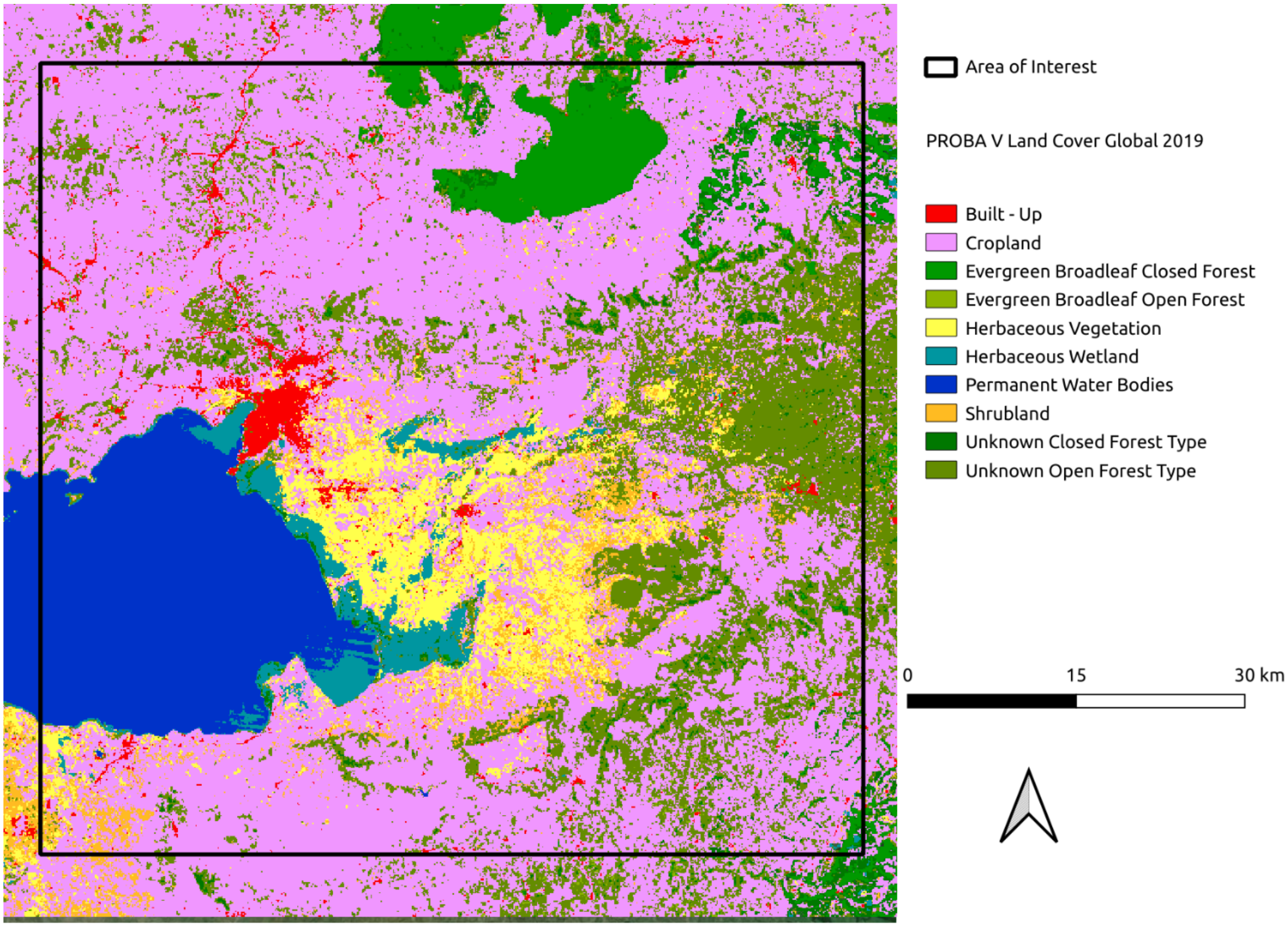


Figure XXX3. Land Cover map of the Ahero Area Of Interest.

Winter Case Study - 06 January 2023

The inherent uncertainties and errors of MODIS and Sentinel 3 LST data for the winter experimental study is summarized in Kelvin degree units,per Land Cover Class on the two following tables (TTT2, TTT3).

Table TTT2. Sentinel 3 image LST inherent errors per land cover class - Winter Case Study

	Sentinel 3 LST Uncertainty Error		
Land Cover	Minimum	Average	Maximum
Built-up	0.55	0.73	1.42
Cropland	0.53	0.66	0.99

Evergreen broadleaf closed forest	0.55	0.65	0.99
Evergreen broadleaf open forest	0.6	0.65	0.99
Herbaceous vegetation	0.6	0.71	2.15
Herbaceous wetland	0.59	0.86	1.76
Permanent water bodies	0.59	0.66	1.29
Shrubland	0.56	0.69	1.42
Unknown closed forest type	0.54	0.65	0.83
Unknown open forest type	0.53	0.66	1.01
Weighted by Area Average	**0.53**	**0.67**	**2.15**

Table TTT3. MODIS image LST inherent errors per land cover class - Winter Case Study

	MODIS LST Error		
Land Cover	**Minimum**	**Average**	**Maximum**
Built-up	1	1.27	2
Cropland	1	1.22	2
Evergreen broadleaf closed forest	1	1.39	1.75
Evergreen broadleaf open forest	1	1.25	1.75
Herbaceous vegetation	1	1.27	1.5
Herbaceous wetland	1	1.3	1.5
Permanent water bodies	1	1.12	1.5
Shrubland	1	1.28	1.75
Unknown closed forest type	1	1.22	1.5

Unknown open forest type	1	1.2	1.75
Weighted by Area Average	**1**	**1.22**	**2**

The experimental winter case study results are presented as partial zonal statistics of the Calibration/Validation RMSE, and the Calibration Residual at each MODIS LST instance, for all the involved Land Cover classes. The zonal analysis involved the calculation of maximum, minimum, Q1/Q3 quantiles, average, and standard deviation metrics, over the whole image. Furthermore, in order to assess the model performance in a more general manner, a global average RMSE statistic is given for Calibration, as a goodness of fit metric, and Validation, as an external data evaluation. A DTC model performance summary is presented on the following Table TTT4.

Table TTT4. Ahero - 06 January 2023 - DTC performance metrics summary.

Land Cover	**Built-up**	**Cropland**	**Evergreen Broadleaf Closed Forest**	**Evergreen Broadleaf Open Forest**	**Herbaceous Vegetation**	**Herbaceous Wetland**	**Permanent Water Bodies**	**Shrubland**	**Unknown Closed Forest Type**	**Unknown Open Forest Type**
max Calibration RMSE	6.26	6.04	5.74	5.52	5.49	4.58	4.82	4.53	6.83	5.41
q75 Calibration RMSE	2.17	2.67	4.61	3.02	1.87	2.56	3.99	1.63	3.2	2.24
avg Calibration RMSE	**1.97**	**2.16**	**3.41**	**2.87**	**1.71**	**2.09**	**3.61**	**1.55**	**2.49**	**1.88**
std Calibration RMSE	1.07	1.05	1.36	1.44	0.94	0.99	0.58	0.76	1.46	0.89
q25 Calibration RMSE	1.4	1.4	2.25	1.8	1.12	1.35	3.39	1.12	1.49	1.34
min Calibration RMSE	0.48	0	1.15	1.76	0.23	0.44	1.27	0.77	0.54	0.33
avg Res MODISt1	1.57	1.8	1.52	2.65	1.7	-0.11	-1.31	1.77	2.58	1.77
avg Res MODISt2	-2.16	-2.57	-5.52	-3.61	-0.74	-2.4	-4.98	-0.93	-3.14	-2.21
avg Res MODISt3	-0.72	-1.32	-1.41	-1.22	-1.43	-0.34	1.6	-0.92	-1.34	-1.32
avg Res MODISt4	1.38	0.63	-0.3	1.49	1.31	2.24	4.4	1.14	0.28	0.23

max Validation RMSE	8.7	10.99	7.42	8.51	9.45	10.79	6.14	7.71	12.5	7.53
q75 Validation RMSE	4.3	4.23	3.74	4.22	5.59	4.57	4.22	4.49	4.62	3.93
avg Validation RMSE	**3.45**	**3.47**	**3.02**	**4.05**	**4.44**	**3.35**	**3.22**	**3.84**	**3.99**	**3.35**
std Validation RMSE	1.52	1.24	1.37	1.98	1.73	1.99	1.32	1.05	1.92	1.04
q25 Validation RMSE	2.44	2.57	2.42	2.98	3.43	2.04	2.71	3.19	2.84	2.64
min Validation RMSE	0.69	0.62	0.08	1.87	0.03	0.1	0.11	1.43	1.62	0.25

The argument of minima for the Calibration RMS Error metric, in δT values, provided insight about the grid search tuning procedure, while the tuning/sensitivity analysis results for Calibration /Validation RMSE are presented on the following table (Table TTT5) and figures (XXX4,XXX5,XXX6), for land cover classes of Cropland, Shrubland and Herbaceous Vegetation.

Table TTT5. Ahero - 06 January 2023 - Calibration and Validation RMSE argument of minima for the δT parameter.

Land Cover	Built-up	Cropland	Evergreen Broadleaf Closed Forest	Evergreen Broadleaf Open Forest	Herbaceous Vegetation	Herbaceous Wetland	Permanent Water Bodies	Shrubland	Unknown Closed Forest Type	Unknown Open Forest Type
δT Calibration	-10.5	-11	-8	-5.5	-9	-3	4.5	-9	-10	-11
δT Validation	-14	-16	-16	-16	-14.5	.15	-6.5	-16	-16	-11.5
LC percentage %	**2.28**	**51.88**	**3.4**	**0.36**	**5.88**	**2.28**	**9.64**	**2.64**	**2.44**	**19.56**

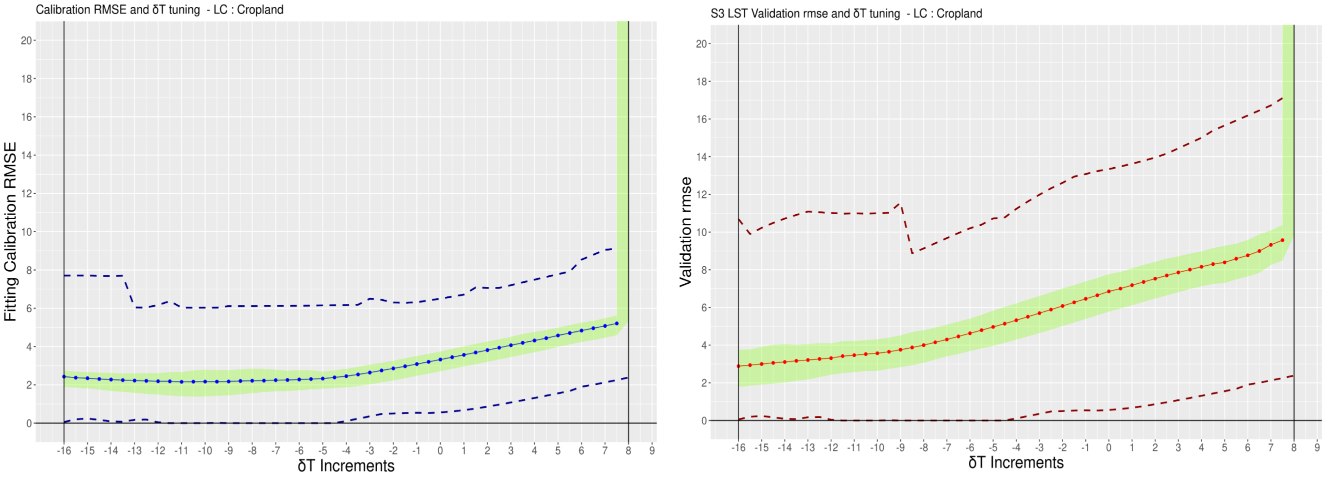


Figure XXX4. Ahero - 06 January 2023 - δT Scree Plot for Calibration and Validation RMSE. Cropland land cover.

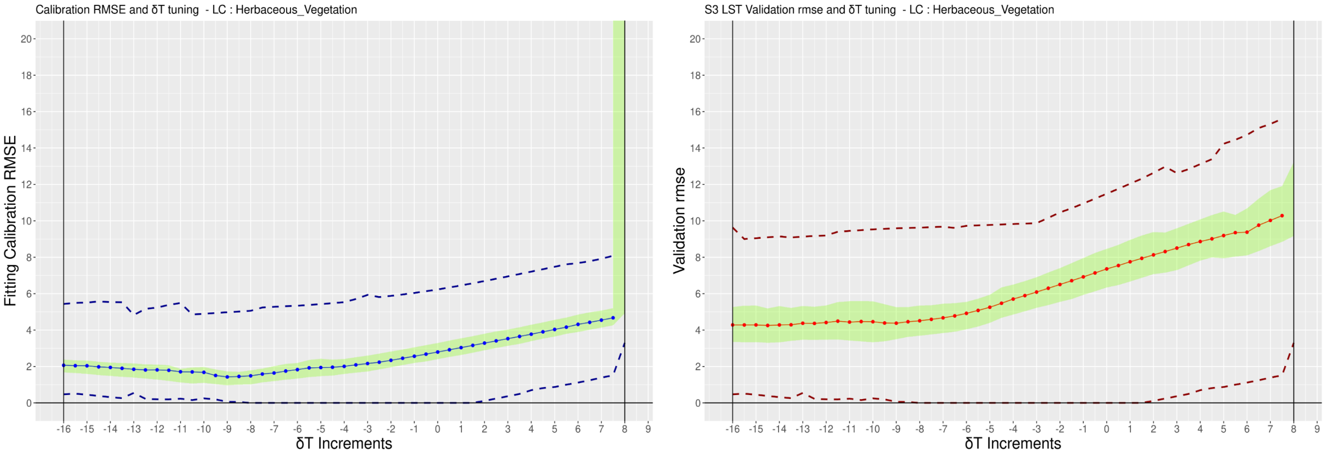


Figure XXX5. δT Ahero - 06 January 2023 - Scree Plot for Calibration and Validation RMSE. Herbaceous Vegetation land cover.

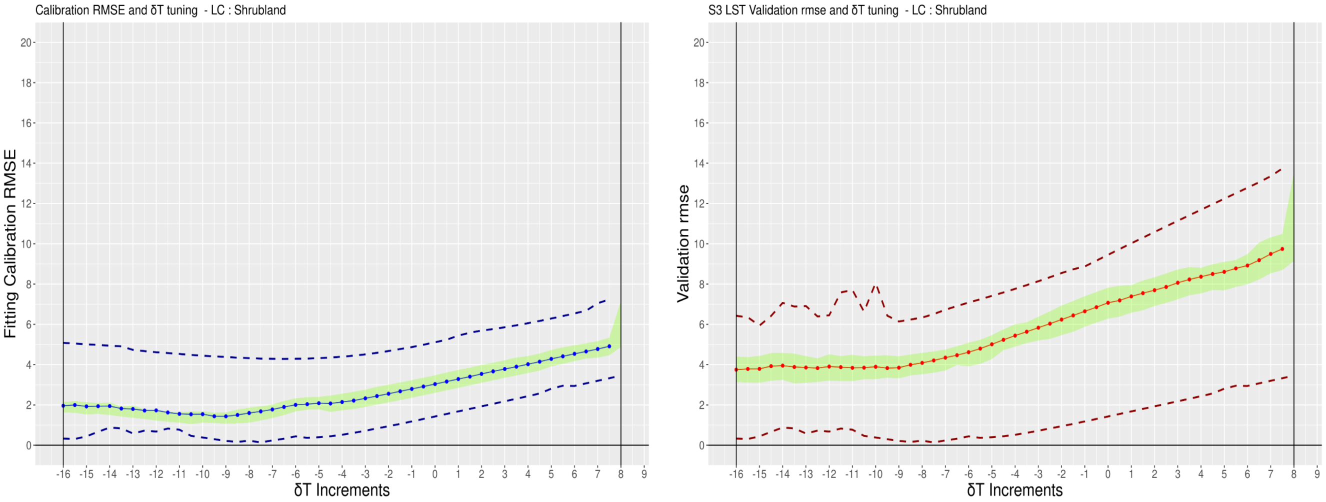


Figure XXX6. Ahero - 06 January 2023 - δT Scree Plot for Calibration and Validation RMSE. Shrubland land cover.

The selection of an optimal δT equal to -11, as the argument of minima for Calibration RMSE, guaranteed an overall good model fit through all 4 MODIS acquisition times. When focusing, however, on the Night LST images, the selected δT provided the best fit to data, among all other values, with minimum residual error.

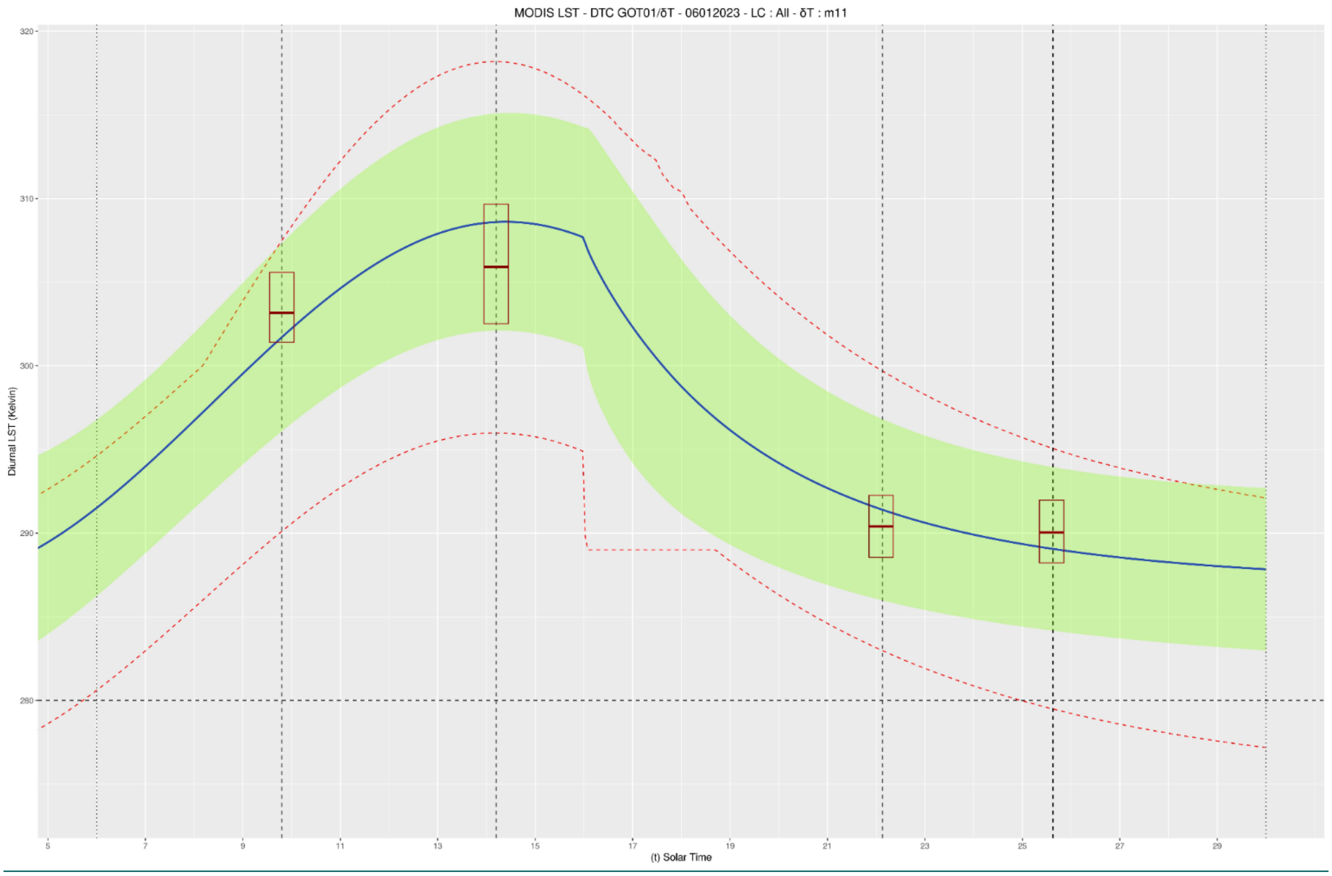


Figure XXX7. Ahero - 06 January 2023 - Average Fitted DTC models for δT equal to -11

The average RMSE metric for Calibration and Validation respectively, considering all the land cover classes of the AOI, provides an overall performance indicator for the DTC model, was calculated as a weighted mean of the partial summary metrics of Table TTT5 and it is given by the following values (kelvin degrees):

- **Calibration RMSE: 2.24**
- **Validation RMSE: 3.24**

In regards with the deviation of the fitted parameters from the ERA5 initial values the Mean Error was calculated for each of the 4 DTC parameters (Table TTT6).

Table TTT6. Ahero - 06 January 2023 - Fitted DTC parameters deviation from ERA5 Land Reanalysis Initial values

DTC Parameter Mean Error	Initial vs Fitted DTC Coefficient Deviation
T_0	8.85
T_a	0.2

t_m	0.98
t_s	-0.61

Regarding the algorithm's computational performance all 4 parameters converged for all 2500 data points.

Spring Case Study - 28 May 2023

The inherent uncertainties and errors of MODIS and Sentinel 3 LST data for the spring experimental study is summarized in Kelvin degree units,per Land Cover Class on the two following tables (TTT7, TTT8).

Table TTT7. Sentinel 3 image LST inherent errors per land cover class - Spring Case Study

	Sentinel 3 LST Uncertainty Error		
Land Cover	**Minimum**	**Average**	**Maximum**
Built-up	0.5	0.67	0.97
Cropland	0.51	0.66	1.22
Evergreen broadleaf closed forest	0.53	0.62	0.65
Evergreen broadleaf open forest	0.59	0.63	0.69
Herbaceous vegetation	0.59	0.7	0.98
Herbaceous wetland	0.59	0.71	1.1
Permanent water bodies	0.59	0.63	1.23
Shrubland	0.55	0.69	0.91
Unknown closed forest type	0.51	0.65	0.94
Unknown open forest type	0.57	0.62	0.75
Weighted by Area Average	**0.5**	**0.66**	**1.23**

Table TTT8. MODIS image LST inherent errors per land cover class - Spring Case Study

	MODIS LST Error		
Land Cover	**Minimum**	**Average**	**Maximum**
Built-up	1	1.31	1.75
Cropland	1	1.28	1.75
Evergreen broadleaf closed forest	1	1.56	1.75
Evergreen broadleaf open forest	1	1.31	1.5
Herbaceous vegetation	1	1.26	1.75
Herbaceous wetland	1	1.19	1.5
Permanent water bodies	1	1.06	1.25
Shrubland	1	1.25	1.5
Unknown closed forest type	1	1.33	1.75
Unknown open forest type	1	1.43	1.75
Weighted by Area Average	**1**	**1.27**	**1.75**

The experimental spring case study results are presented as partial zonal statistics of the Calibration/Validation RMSE, and the Calibration Residual at each MODIS LST instance, for all the involved Land Cover classes. The zonal analysis involved the calculation of maximum, minimum, Q1/Q3 quantiles, average, and standard deviation metrics, over the whole image. Furthermore, in order to assess the model performance in a more general manner, a global average RMSE statistic is given for Calibration, as a goodness of fit metric, and Validation, as an external data evaluation. A DTC model performance summary is presented on the following Table TTT9.

Table TTT9. Ahero - 28 May 2023 - DTC performance metrics summary.

Land Cover	Built-up	Cropland	Evergreen Broadleaf	Evergreen Broadleaf	Herbaceous Vegetation	Herbaceous Wetland	Permanent Water Bodies	Shrubland	Unknown Closed Forest Type	Unknown Open

			Closed Forest	Open Forest						Forest Type
max Calibration RMSE	4.99	7.79	3.38	3.03	8.36	5.71	4.7	8.35	5.94	6.48
q75 Calibration RMSE	1.99	1.91	2.14	2.67	3.31	3.77	4.14	2.82	2.38	2.36
avg Calibration RMSE	**1.5**	**1.59**	**1.91**	**1.89**	**2.42**	**2.8**	**3.57**	**2.21**	**2.05**	**1.8**
std Calibration RMSE	0.9	1.1	0.47	0.85	1.67	1.16	0.75	1.65	0.91	0.99
q25 Calibration RMSE	0.84	0.97	1.56	1.29	1.28	2.08	3.14	1.17	1.53	1.12
min Calibration RMSE	0	0	1.27	0.51	0.38	0.74	1.01	0.25	0.65	0.06
avg Res MODISt1	**1.88**	**1.62**	**2.39**	**2.42**	**2.31**	**0.92**	**-1.69**	**2.01**	**2.99**	**2.05**
avg Res MODISt2	-1.01	-1.12	-1.53	-1.24	-1.04	-2.02	-4.96	-0.86	-1.32	-1.02
avg Res MODISt3	**-1**	**-1.59**	**-1.59**	**-1.88**	**-3.3**	**-2.43**	**1.38**	**-3.2**	**-1.78**	**-2.06**
avg Res MODISt4	-0.06	0.43	0.65	0.38	0.45	2.59	3.8	-0.05	0.22	0.42
max Validation RMSE	8.54	11.27	9.77	7.89	13.18	7.1	5.66	9.3	9.89	12.13
q75 Validation RMSE	6.23	5.54	7.21	6.78	6.37	5.25	3.31	6.72	6.56	6.18
avg Validation RMSE	**4.41**	**3.83**	**5.35**	**4.76**	**5.16**	**3.67**	**2.68**	**5.29**	**5.53**	**4.81**
std Validation RMSE	2.35	2.3	2.42	2.64	1.98	1.85	1.27	2.15	1.79	2.14
q25 Validation RMSE	2.85	1.48	3.19	3.62	3.79	2.51	1.93	3.89	4.24	3.47
min Validation RMSE	0.13	0	0.19	0.34	0.86	0.02	0.13	0.75	2.53	0.04

The argument of minima for the Calibration RMS Error metric, in δT values, provided insight about the grid search tuning procedure, while the tuning/sensitivity analysis results for Calibration /Validation

RMSE are presented on the following table (Table TTT10) and figures XXX8, XXX9, XXX10, for land cover classes of Cropland, Shrubland and Herbaceous Vegetation

Table TTT10. Ahero - 28 May 2023 - Calibration and Validation RMSE argument of minima for the δT parameter.

Land Cover	Built-up	Cropland	Evergreen Broadleaf Closed Forest	Evergreen Broadleaf Open Forest	Herbaceous Vegetation	Herbaceous Wetland	Permanent Water Bodies	Shrubland	Unknown Closed Forest Type	Unknown Open Forest Type
δT Calibration	-4	-3.5	-5	-5.5	-4	-1	7	-3	-5.5	-4
δT Validation	-9.5	-10.5	-10.5	-11	-11	-16	7	-16	-12	-16
LC percentage %	**2.12**	**54.96**	**1.84**	**0.48**	**6.68**	**2.44**	**8.68**	**4.28**	**16.28**	**2.24**

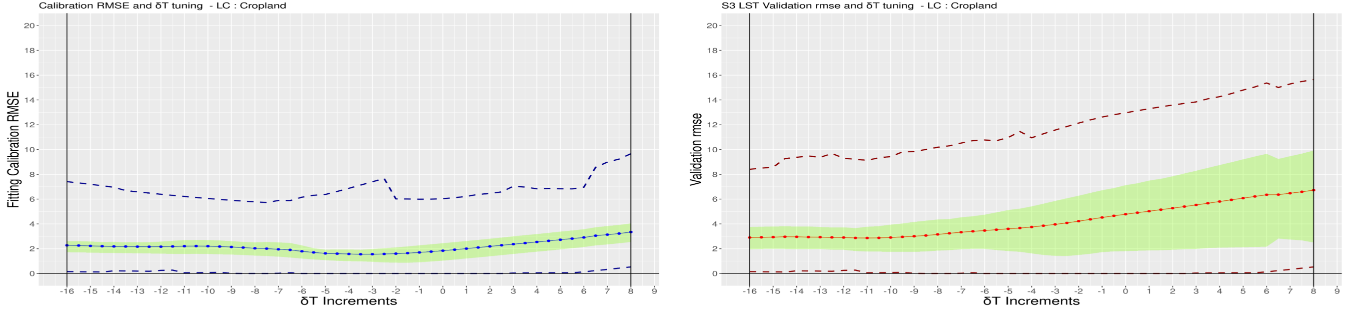


Figure XXX8. Ahero - 28 May 2023 - δT Scree Plot for Calibration and Validation RMSE. Cropland land cover.

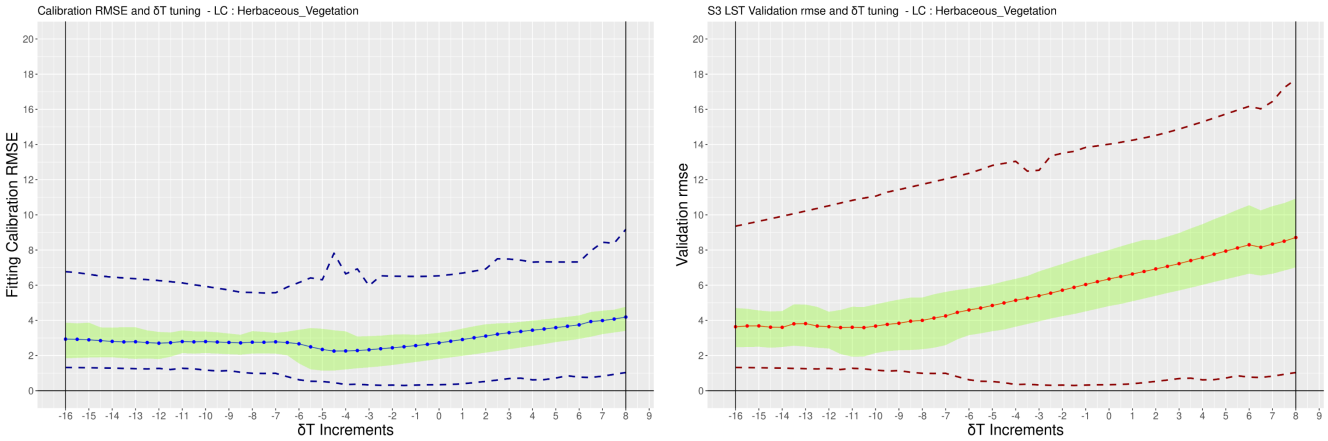


Figure XXX9. δT Ahero - 28 May 2023 - Scree Plot for Calibration and Validation RMSE. Herbaceous Vegetation land cover.

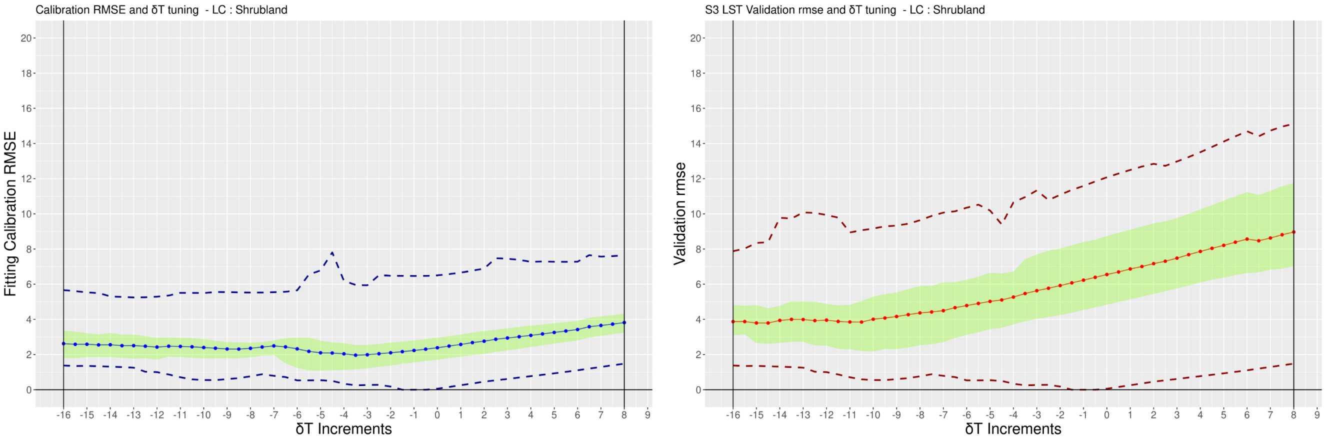


Figure XXX10. Ahero - 28 May 2023 - δT Scree Plot for Calibration and Validation RMSE. Shrubland land cover.

The selection of an optimal δT equal to -3.5, as the argument of minima for Calibration RMSE, guaranteed an overall good model fit through all 4 MODIS acquisition times. When focusing, however, on the Night LST images, the selected δT provided the best fit to data, among all other values, with minimum residual error.

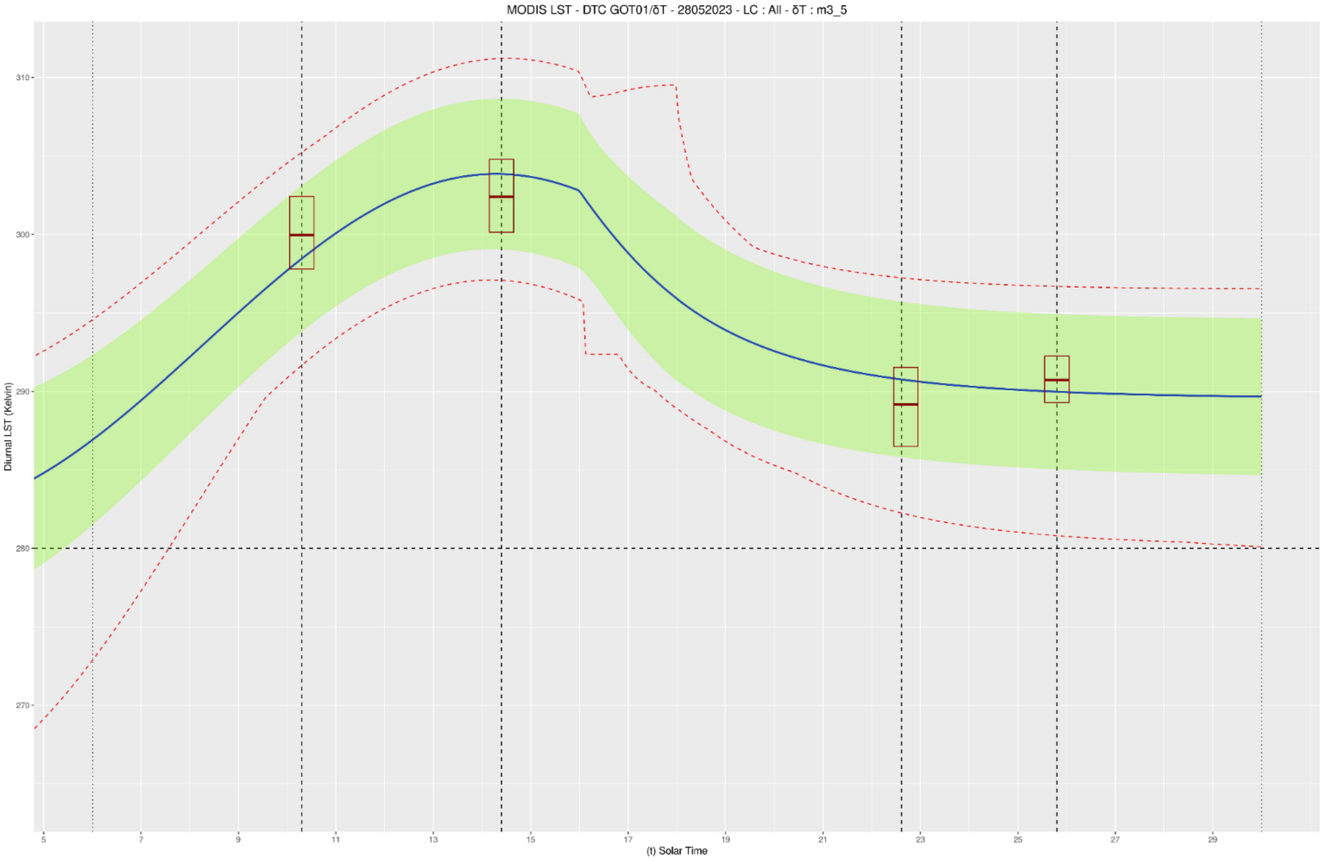


Figure XXX11. Ahero -28 May 2023 - Average Fitted DTC models for δT equal to -11

The average RMSE metric for Calibration and Validation respectively, considering all the land cover classes of the AOI, provides an overall performance indicator for the DTC model, was calculated as a weighted mean of the partial summary metrics of Table XXX and it is given by the following values (kelvin degrees):

- **Calibration RMSE: 1.96**
- **Validation RMSE: 4.22**

In regards with the deviation of the fitted parameters from the ERA5 initial values the Mean Error was calculated for each of the 4 DTC parameters (figure TTT11).

Table TTT11. Ahero - 28 May 2023 - Fitted DTC parameters deviation from ERA5 Land Reanalysis Initial values

DTC Parameter Mean Error	Initial vs Fitted DTC Coefficient Deviation
T_o	3.39
T_a	0.51
t_m	0.54
t_s	-0.6

Regarding the algorithm's computational performance all 4 parameters converged for all 2500 data points.

Summer Case Study - 20 August 2023

The inherent uncertainties and errors of MODIS and Sentinel 3 LST data for the summer experimental study are summarized in Kelvin degree units,per Land Cover Class on the two following tables (TTT11, TTT12).

Table TTT11. Sentinel 3 image LST inherent errors per land cover class - Summer Case Study

	Sentinel 3 LST Uncertainty Error		
Land Cover	**Minimum**	**Average**	**Maximum**
Built-up	0.38	0.66	1.35

Cropland	0.35	0.61	2.26
Evergreen broadleaf closed forest	0.6	0.6	0.71
Evergreen broadleaf open forest	0.5	0.57	0.6
Herbaceous vegetation	0.35	0.64	0.86
Herbaceous wetland	0.2	0.67	1.45
Permanent water bodies	0.2	0.32	0.99
Shrubland	0.37	0.63	0.81
Unknown closed forest type	0.36	0.61	0.84
Unknown open forest type	0.5	0.59	0.74
Weighted by Area Average	**0.2**	**0.61**	**2.26**

Table TTT12. MODIS image LST inherent errors per land cover class - Summer Case Study

	MODIS LST Error		
Land Cover	**Minimum**	**Average**	**Maximum**
Built-up	1	1.38	2
Cropland	1	1.32	1.75
Evergreen broadleaf closed forest	1	1.27	1.5
Evergreen broadleaf open forest	1.25	1.4	1.75

Herbaceous vegetation	1	1.16	1.75
Herbaceous wetland	1	1.11	1.5
Permanent water bodies	1	1.09	1.5
Shrubland	1	1.25	1.75
Unknown closed forest type	1	1.42	1.75
Unknown open forest type	1	1.38	1.75
Weighted by Area Average	**1**	**1.28**	**2**

The experimental summer case study results are presented as partial zonal statistics of the Calibration/Validation RMSE, and the Calibration Residual at each MODIS LST instance, for all the involved Land Cover classes. The zonal analysis involved the calculation of maximum, minimum, Q1/Q3 quantiles, average, and standard deviation metrics, over the whole image. Furthermore, in order to assess the model performance in a more general manner, a global average RMSE statistic is given for Calibration, as a goodness of fit metric, and Validation, as an external data evaluation. A DTC model performance summary is presented on the following Table TTT13.

Table TTT13. Ahero - 20 August 2023 - DTC performance metrics summary.

Land Cover	**Built-up**	**Cropland**	**Evergreen Broadleaf Closed Forest**	**Evergreen Broadleaf Open Forest**	**Herbaceous Vegetation**	**Herbaceous Wetland**	**Permanent Water Bodies**	**Shrubland**	**Unknown Closed Forest Type**	**Unknown Open Forest Type**
max Calibration RMSE	3.23	4.91	4.07	3.74	4.75	3.95	5.77	3.05	3.42	4.66
q75 Calibration RMSE	1.68	1.72	1.81	1.94	1.69	3	4.05	1.58	1.53	1.52
avg Calibration RMSE	**1.12**	**1.14**	**1.42**	**1.54**	**1.23**	**1.96**	**3.69**	**1.06**	**1.2**	**1.11**
std Calibration RMSE	0.8	0.9	0.77	1.13	0.81	1.18	0.69	0.68	0.87	0.86

q25 Calibration RMSE	0.47	0.39	1.06	1.07	0.6	1.11	3.53	0.44	0.59	0.49
min Calibration RMSE	0	0	0	0	0	0	0.02	0	0	0
avg Res MODISt1	1.01	0.82	0.19	0.56	1.79	-0.03	-2.29	1.54	0.78	0.75
avg Res MODISt2	-0.93	-1.23	-2.5	-2.29	-0.94	-2.36	-4.7	-0.73	-1.48	-1.17
avg Res MODISt3	-0.53	-0.56	0.28	-0.31	-0.09	0.1	1.71	-0.28	-0.33	-0.74
avg Res MODISt4	0.1	-0.23	0.7	0.79	-0.27	1.84	4.64	-0.24	0.19	-0.26
max Validation RMSE	4.95	5.9	2.63	2.07	6.4	6.2	8.92	5.31	3.5	3.85
q75 Validation RMSE	2.15	1.88	1.43	1.71	3.25	3.14	5.12	2.81	1.94	1.95
avg Validation RMSE	**1.6**	**1.38**	**1.14**	**1.14**	**2.49**	**2.49**	**4.29**	**2.09**	**1.49**	**1.35**
std Validation RMSE	0.98	0.86	0.53	0.56	1.11	1.48	1.62	1.02	0.75	0.83
q25 Validation RMSE	0.97	0.74	0.81	0.73	1.9	1.56	3.75	1.42	1.01	0.71
min Validation RMSE	0.1	0	0.07	0.46	0.12	0.1	0.14	0	0.1	0.01

The argument of minima for the Calibration RMS Error metric, in δT values, provided insight about the grid search tuning procedure, while the tuning/sensitivity analysis results for Calibration /Validation RMSE are presented on the following table (Table TTT14) and figures XXXX, for land cover classes of Cropland, Shrubland and Herbaceous Vegetation

Table TTT14. Ahero - 20 August 2023 - Calibration and Validation RMSE argument of minima for the δT parameter.

Land Cover	**Built-up**	**Cropland**	**Evergreen Broadleaf Closed Forest**	**Evergreen Broadleaf Open Forest**	**Herbaceous Vegetation**	**Herbaceous Wetland**	**Permanent Water Bodies**	**Shrubland**	**Unknown Closed Forest Type**	**Unknown Open Forest Type**

δT Calibration	-8.5	-7.5	-6	-7	-11.5	-1	4.5	-11.5	-9	-10
δT Validation	3.5	4.5	2	5	4	-16	6.5	4.5	5.5	5
LC percentage %	**2.32**	**49.1**	**2.16**	**0.4**	**8.04**	**3**	**14.81**	**4.08**	**14.53**	**1.56**

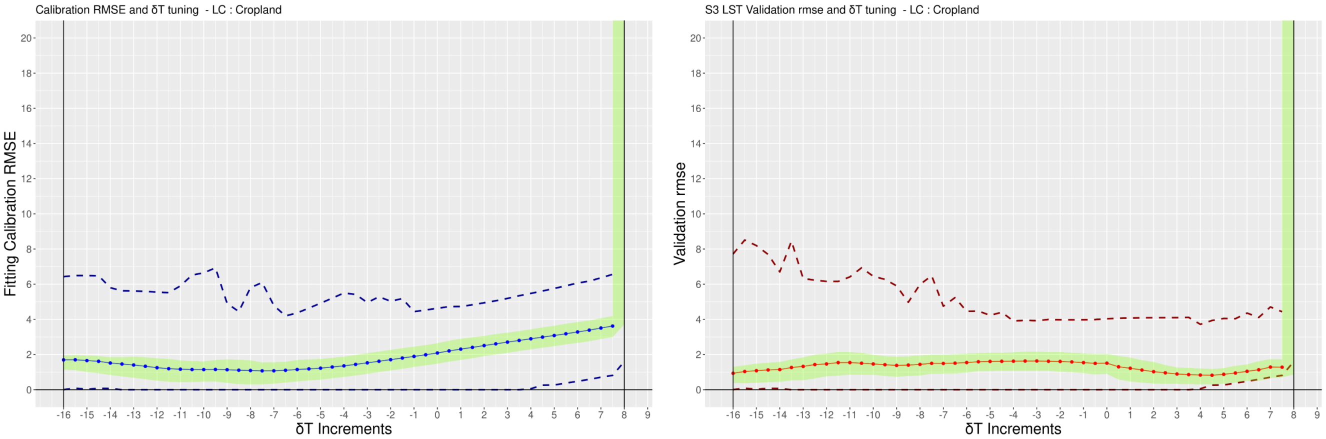


Figure XXX12. Ahero - 20 August 2023 - δT Scree Plot for Calibration and Validation RMSE. Cropland land cover.

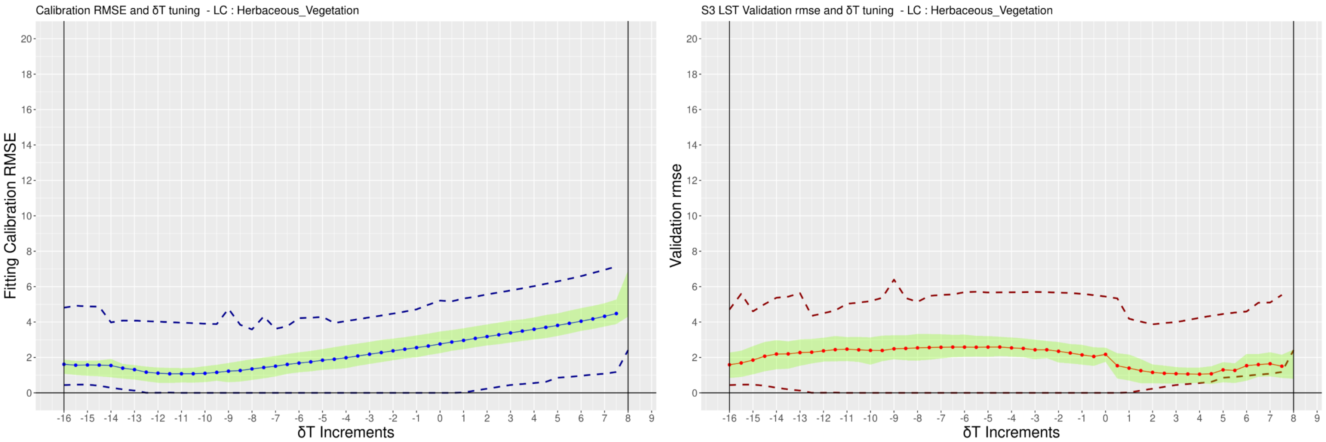


Figure XXX13. δT Ahero - 20 August 2023 - Scree Plot for Calibration and Validation RMSE. Herbaceous Vegetation land cover.

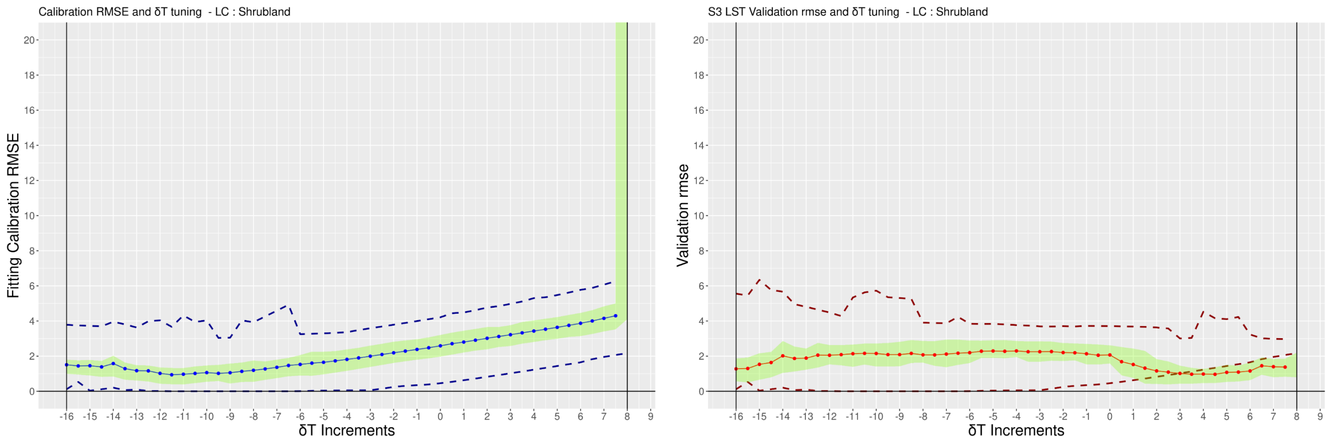


Figure XXX14. Ahero - 20 August 2023 - δT Scree Plot for Calibration and Validation RMSE. Shrubland land cover.

The selection of an optimal δT equal to -9, as the argument of minima for Calibration RMSE, guaranteed an overall good model fit through all 4 MODIS acquisition times. When focusing, however, on the Night LST images, the selected δT provided the best fit to data, among all other values, with minimum residual error.

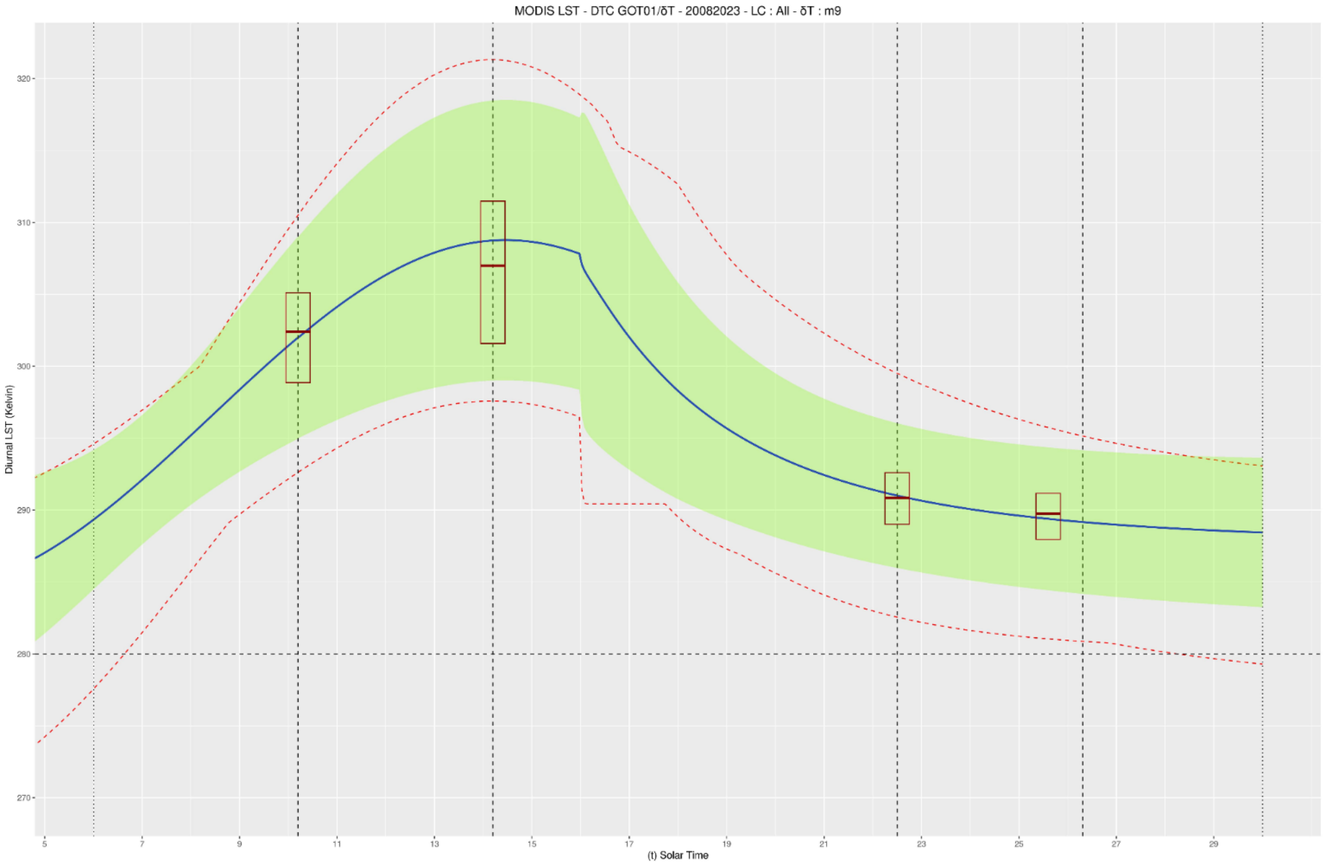


Figure XXX15. Ahero - 20 August 2023 - Average Fitted DTC models for δT equal to -9

The average RMSE metric for Calibration and Validation respectively, considering all the land cover classes of the AOI, provides an overall performance indicator for the DTC model, was calculated as a weighted mean of the partial summary metrics of Table TTT14 and it is given by the following values (kelvin degrees):

- **Calibration RMSE: 1.56**
- **Validation RMSE: 1.98**

In regards with the deviation of the fitted parameters from the ERA5 initial values the Mean Error was calculated for each of the 4 DTC parameters (TTT15).

Table TTT15. Ahero - 20 August 2023 - Fitted DTC parameters deviation from ERA5 Land Reanalysis Initial values

DTC Parameter Mean Error	Initial vs Fitted DTC Coefficient Deviation
T_o	6.95
T_a	1.92
t_m	0.99
t_s	-0.5

Regarding the algorithm’s computational performance nearly all 4 parameters converged for all 2498 data points.

Autumn Case Study - 29 September 2023

The inherent uncertainties and errors of MODIS and Sentinel 3 LST data for the autumn experimental study are summarized in Kelvin degree units,per Land Cover Class on the two following tables (TTT16, TTT17).

Table TTT16. Sentinel 3 image LST inherent errors per land cover class - Autumn Case Study

	Sentinel 3 LST Uncertainty Error		
Land Cover	**Minimum**	**Average**	**Maximum**
Built-up	0.55	0.63	0.7
Cropland	0.55	0.64	1.03

Evergreen broadleaf closed forest	0.64	0.64	0.64
Evergreen broadleaf open forest	0.64	0.67	0.77
Herbaceous vegetation	0.62	0.73	1.05
Herbaceous wetland	0.62	0.7	0.8
Permanent water bodies	0.67	0.68	0.71
Shrubland	-	-	-
Unknown closed forest type	0.55	0.63	0.7
Unknown open forest type	0.64	0.65	0.66
Weighted by Area Average	**0.55**	**0.66**	**1.05**

Table TTT17. MODIS image LST inherent errors per land cover class - Autumn Case Study

	MODIS LST Error		
Land Cover	**Minimum**	**Average**	**Maximum**
Built-up	1.5	1.63	1.75
Cropland	1.25	1.66	2
Evergreen broadleaf closed forest	1.75	1.75	1.75
Evergreen broadleaf open forest	1.5	1.75	2
Herbaceous vegetation	1.25	1.5	1.75
Herbaceous wetland	1.5	1.5	1.5

Permanent water bodies	1.25	1.37	1.5
Shrubland	-	-	-
Unknown closed forest type	1.5	1.71	2
Unknown open forest type	1.5	1.68	2
Weighted by Area Average	**1.25**	**1.62**	**2**

The experimental autumn case study results are presented as partial zonal statistics of the Calibration/Validation RMSE, and the Calibration Residual at each MODIS LST instance, for all the involved Land Cover classes. The zonal analysis involved the calculation of maximum, minimum, Q1/Q3 quantiles, average, and standard deviation metrics, over the whole image. Furthermore, in order to assess the model performance in a more general manner, a global average RMSE statistic is given for Calibration, as a goodness of fit metric, and Validation, as an external data evaluation. A DTC model performance summary is presented on the following Table TTT18.

Table TTT18. Ahero - 29 September 2023 - DTC performance metrics summary.

Land Cover	Built-up	Cropland	Evergreen Broadleaf Closed Forest	Evergreen Broadleaf Open Forest	Herbaceous Vegetation	Herbaceous Wetland	Permanent Water Bodies	Shrubland	Unknown Closed Forest Type	Unknown Open Forest Type
max Calibration RMSE	1.31	10.11	1.73	3.69	10.45	7.96	5.11	-	4.73	5.01
q75 Calibration RMSE	1.2	4.67	1.66	2.51	3.8	3.13	4.49	-	3.1	4.06
avg Calibration RMSE	**1.09**	**3.41**	**1.45**	**2.28**	**3.5**	**3.23**	**3.73**	**-**	**2.6**	**2.87**
std Calibration RMSE	0.31	1.81	0.37	0.98	2.49	2.17	0.94	-	1.26	1.48
q25 Calibration RMSE	0.98	1.8	1.31	1.72	1.7	2.01	2.92	-	1.87	1.38
min Calibration RMSE	0.87	0.66	1.03	1.49	1.31	1.57	2.26	-	1	0.6

avg Res MODISt1	0.91	4.68	2.64	1.9	4.51	2.71	-0.22	-	2.62	3.04
avg Res MODISt2	-1.11	-2.78	-1.11	-3.08	-1.3	-1.36	-5.35	-	-3.05	-3.18
avg Res MODISt3	-0.83	1.01	0.24	-0.37	-3.18	-3.64	-1.22	-	-0.57	0.25
avg Res MODISt4	0.72	-1.26	-0.07	-0.92	0.48	2.99	4.62	-	1.17	-0.82
max Validation RMSE	2.2	13.6	6.33	4.78	17.77	12.73	3.92	-	8.45	12.64
q75 Validation RMSE	2.11	7.03	4.91	4.54	12.25	7.93	1.1	-	6.45	7.17
avg Validation RMSE	**2.01**	**5.24**	**3.5**	**3.22**	**10.52**	**5.98**	**1.03**	-	**4.81**	**5.26**
std Validation RMSE	0.27	2.66	4.01	1.64	3.19	4.05	0.83	-	2.77	3.12
q25 Validation RMSE	1.92	2.87	2.08	1.97	8.3	3.15	0.62	-	3.78	2.27
min Validation RMSE	1.82	0.3	0.67	1.53	6.17	1.13	0.01	-	0.55	1.29

The argument of minima for the Calibration RMS Error metric, in δT values, provided insight about the grid search tuning procedure, while the tuning/sensitivity analysis results for Calibration /Validation RMSE are presented on the following table (Table TTT19) and figures XXX16, XXX17, for land cover classes of Cropland and Herbaceous Vegetation

Table TTT19. Ahero - 29 September 2023 - Calibration and Validation RMSE argument of minima for the δT parameter.

Land Cover	Built-up	Cropland	Evergreen Broadleaf Closed Forest	Evergreen Broadleaf Open Forest	Herbaceous Vegetation	Herbaceous Wetland	Permanent Water Bodies	Shrubland	Unknown Closed Forest Type	Unknown Open Forest Type
δT Calibration	-10.5	-13.5	-14	-14	-12	-9.5	1.5		-8.5	-14.5
δT Validation	-15	-16	-15.5	-12	-16	-13.5	-15		-13	-16

LC percentage %	1.15	57.47	1.72	2.3	8.05	4.02	10.92		10.34	4.02

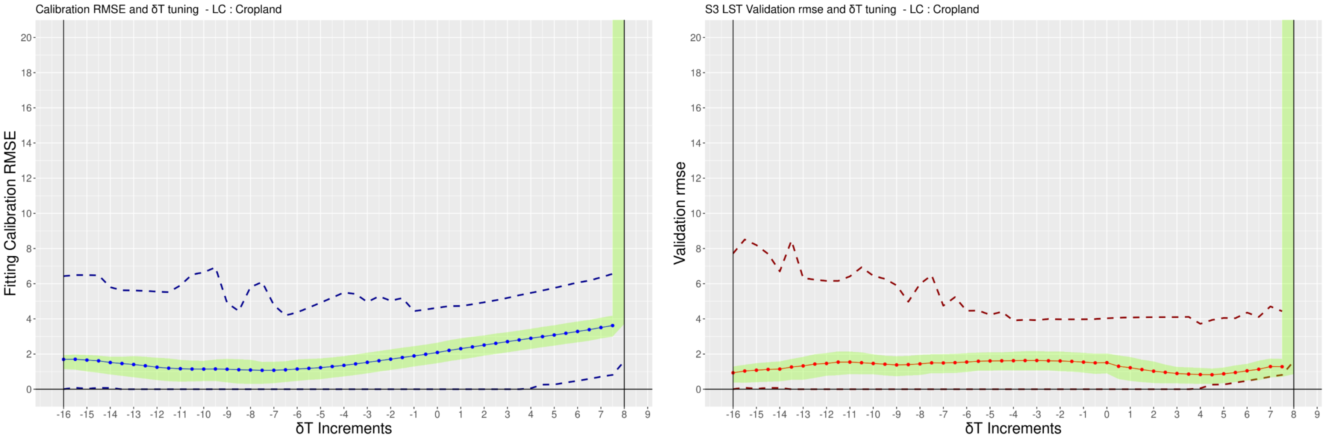


Figure XXX16. Ahero - 29 September 2023 - δT Scree Plot for Calibration and Validation RMSE. Cropland land cover.

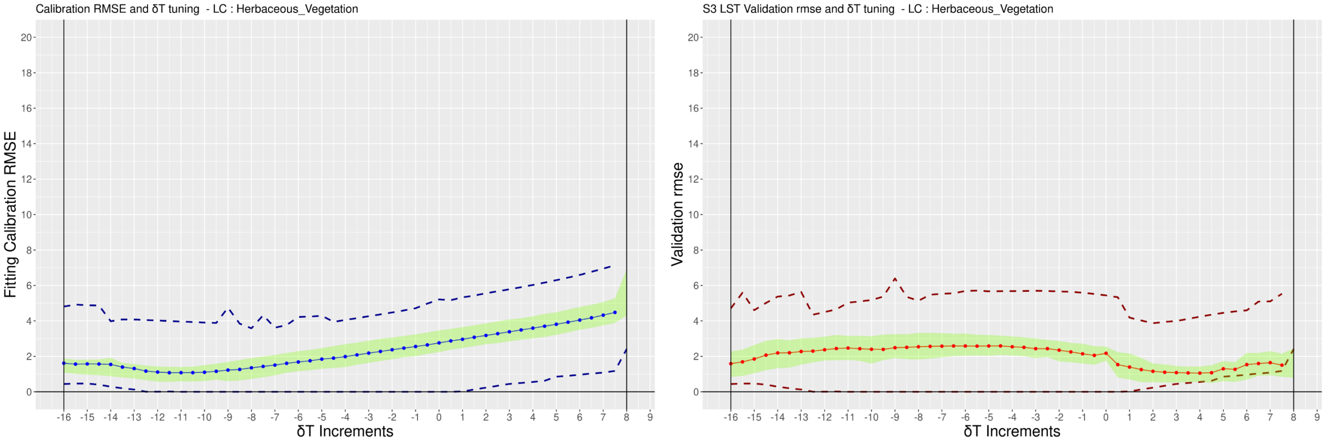


Figure XXX17. δT Ahero - 29 September 2023 - Scree Plot for Calibration and Validation RMSE. Herbaceous Vegetation land cover.

The selection of an optimal δT equal to -0.5, as the argument of minima for Calibration RMSE, guaranteed an overall good model fit through all 4 MODIS acquisition times. When focusing, however, on the Night LST images, the selected δT provided the best fit to data, among all other values, with minimum residual error.

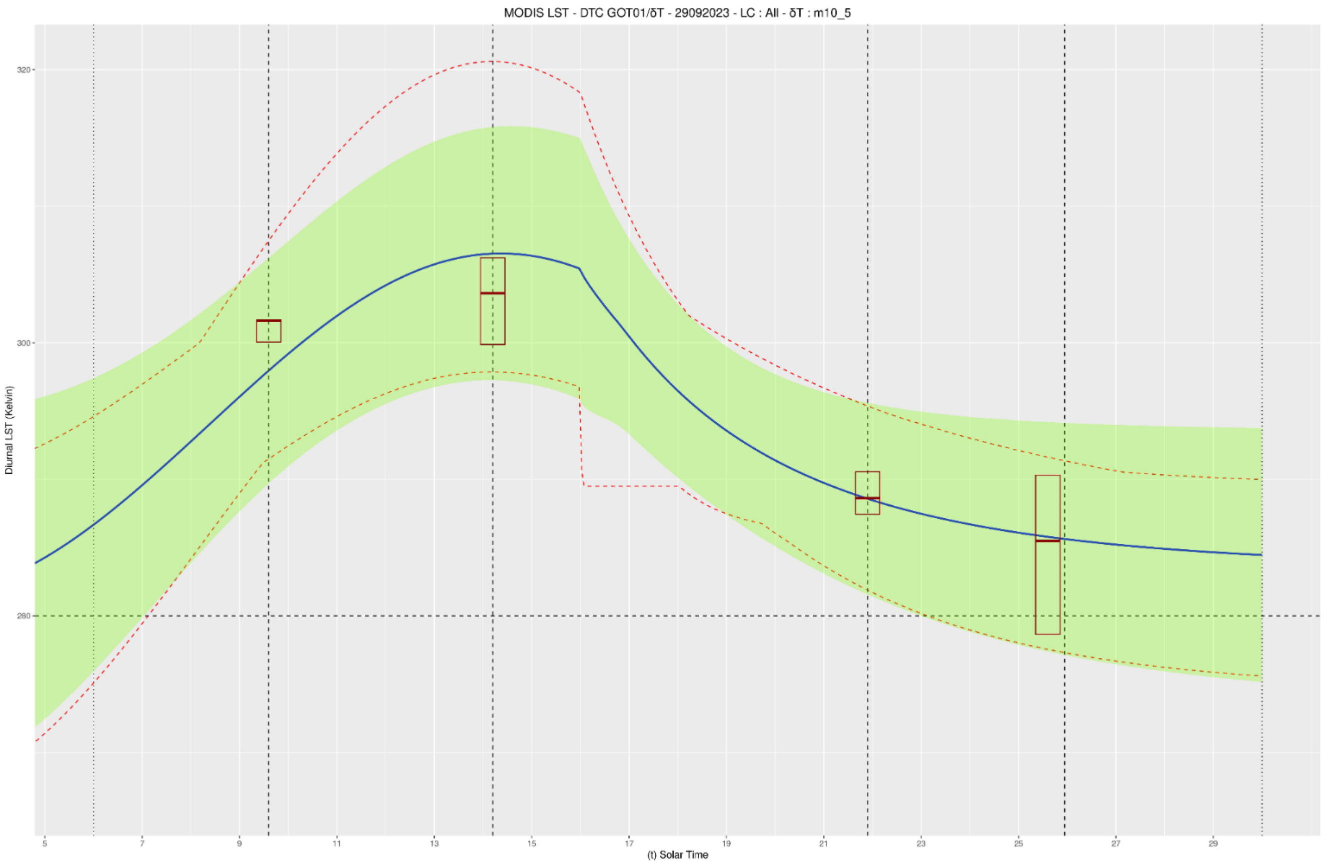


Figure XXX18. Ahero - 29 September 2023 - Average Fitted DTC models for δT equal to -0.5

The average RMSE metric for Calibration and Validation respectively, considering all the land cover classes of the AOI, provides an overall performance indicator for the DTC model, was calculated as a weighted mean of the partial summary metrics of Table XXX and it is given by the following values (kelvin degrees):

- **Calibration RMSE: 3.25**
- **Validation RMSE: 5.08**

In regards with the deviation of the fitted parameters from the ERA5 initial values the Mean Error was calculated for each of the 4 DTC parameters (Table TTT20).

Table TTT20. Ahero - 29 September 2023 - Fitted DTC parameters deviation from ERA5 Land Reanalysis Initial values

DTC Parameter Mean Error	Initial vs Fitted DTC Coefficient Deviation
To	6.56

T_a	1.7
t_m	0.61
t_s	-0.44

Regarding the algorithm's computational performance all 4 parameters converged for all 2500 data points.

The conclusions drawn from the data analysis performed in the update iteration can be summarized as the following:

- Regarding the influence of seasonality on the accuracy of DTC models, it was observed that it exists. The fitting models achieved the smallest errors in the spring and summer months. The worst performances were observed in the autumn months.

- As for the influence that the underlying Land Use, Land Cover (LU/LC) has on the DTC result, it has an impact on the accuracy of the models which is justified by the evident differences in RMSE between LU/LC classes. While the Cropland/ Unknown Open Forest classes are the most dominant ones in the area of the case study, their RMSE is close to the above average values. Throughout the year, the Permanent Water Bodies, and Wetland land cover classes exhibit the higher Calibration RMS errors, while the high validation RMS errors are more randomly distributed among classes and seasons. The above is also supported by the adjustable free parameter δT which is used to obtain the optimum results. Initially the tuned value varies to a small or larger extent between the LU/LV classes but where it has a very large difference is specifically in the Permanent Water Bodies class, where, in contrast to all other classes, it takes positive values. This fact could be related to the thermal lag that water bodies exhibit in their temperature variation due to their high thermal capacit

- With respect to the cloud cover regions it is obvious that they should not be included in DTC modeling because the errors in the input LSTs are large and contribute, for these pixels, to unacceptable final estimation errors. Also cloud masking procedures bring about nan values which results in insufficient data to converge the curve fitting algorithm at these locations. For the data pipeline involving DTC fitting, DTC prediction at ECOSTRESS time, and STTFN fusion of MODIS and ECOSTRESS, the rather tight constraint of having for the given Area of Interest 4 Cloud free images is evident.

- Finally, in the more general discussion concerning the fidelity of curve fitting to MODIS data during the Calibr;ation and the External Validation with Sentinel 3 LST data, it was observed from the average residuals of the DTC model to the actual 4 MODIS instances, across the different classes and seasons, that the model fits best in the afternoon and evening observations, while it gives the largest errors at noon. The above conclusion is consistent with the corresponding validation error since the Sentinel 3 LST image was acquired at 22.45 local time very close to MOD Terra LST view time.

3.8. Spatiotemporal Thermal Data Fusion with Deep Learning (STTFN)

Challenge:

The major challenges in the approach of thermal sharpening methods are discussed in detail in Deliverable D3 where the state-of-the-art review in this field is presented. In conclusion, these include:

1. The Uncertainty in LST retrievals
2. Low thermal contrast of downscaled LST
3. Non-Linearity of LST temporal change
4. Cloud contamination
5. Model generalization

In addition to these, there are the imagery specificities that exist and are characteristic of the thermal data under consideration for the current study.

- Satellite images from ECOSTRESS thermal sensors do not exhibit a stable spatio-temporal acquisition pattern. This fact combined with the condition for cloud-free imagery limits their availability for model training and development, with implications on the experimental design. The instability that characterizes their acquisition time can have a serious impact on the downscaling result, according to challenge 3, and for this reason the pre-processing step of DTC curve fitting, as mentioned in Chapter 3.7 was employed for the temporal normalization of thermal fusion image pairs. In addition, temporal normalization is only performed on the low spatial resolution data. In ECOSTRESS high spatial resolution data, the curve fitting with the use of DTC models is not possible for daily acquisitions, as these include only 1 to 2 images per day.
- The observed radiometric artifacts of oblique stripping that ECOSTRESS LST images exhibit. Despite the fact that for vertical/horizontal image descripting many algorithms have been developed and are available, with high potential for radiometric correction, there’s no significant variety of methods for oblique stripes.

Approach:

The decided approach within the course of AFRI4Cast is the application of a deep learning based SpatioTemporal Temperature Fusion Network (STTFN) method, as described by Yin et al. (2021), through adaptation of their proposed Deep Learning architecture for ECOSTRESS/MODIS imagery. The flexibility that this Deep Learning approach offers is a considerable advantage over other conventional fusion methods. The purpose is to provide a proof of concept for the possibility of spatiotemporal downscaling of ECOSTRESS LST data, given their operational specificities, as well as to give evaluation results for the downscaled product.

Methodology:

This section summarizes the methodological framework, which was described in detail in Deliverable D3 “State of the art review report on analysis methods and algorithms’” and finally was developed and applied to the experimental case study. Fusion-based methods such as the one applied, determine the relationship between known image pairs of LST, at fine and coarse resolution, and downscale coarse resolution (e.g., MODIS LST) when fine resolution (ECOSTRESS) LST is not available.

In STTFN, a multi-scale fusion Convolutional Neural Network is employed to perform the fusion task. It uses two fine-coarse spatial resolution LST image pairs observed before and after the target date to train two multi-scale fusion CNNs, and then combine their predictions using a Spatiotemporal-Consistency (STC)-Weighting function. It is composed of three parts:

1. a component for **super-resolution** of temporal change between the input target and neighboring coarse spatial resolution LST images
2. a component for high-level **feature abstraction** of neighboring fine spatial resolution data
3. a component of **integration** of the extracted multi-scale features

The aim is to obtain a pair of fine and coarse LST images that pre-date the target date, for which a coarse MODIS image is available, together with another pair of fine and coarse images that post-dating the target date. Then, from one MODIS LST image at the target date and two pairs of ECOSTRESS and MODIS images at the neighboring dates, a high resolution (ECOSTRESS -like) LST image is predicted.

For simplicity, the target date is denoted as t_2, and the dates pre- and post-dating t_2 are denoted as t_1 and t_3, respectively. Accordingly, the MODIS coarse resolution image (LST) at t_2 is denoted as M_2, and the ECOSTRESS-MODIS image pairs (LST) at dates t_1 and t_3 are denoted as E_1 and M_1 and E_3 and M_3, respectively. As depicted in Figure 14, the implemented STTFN generates a fine spatial resolution image via a three-stage process:

1. Forward and backward model training
2. Forward and backward prediction
3. Combination of forward and backward predicted high images to yield a final fine spatial resolution predicted image for t_2

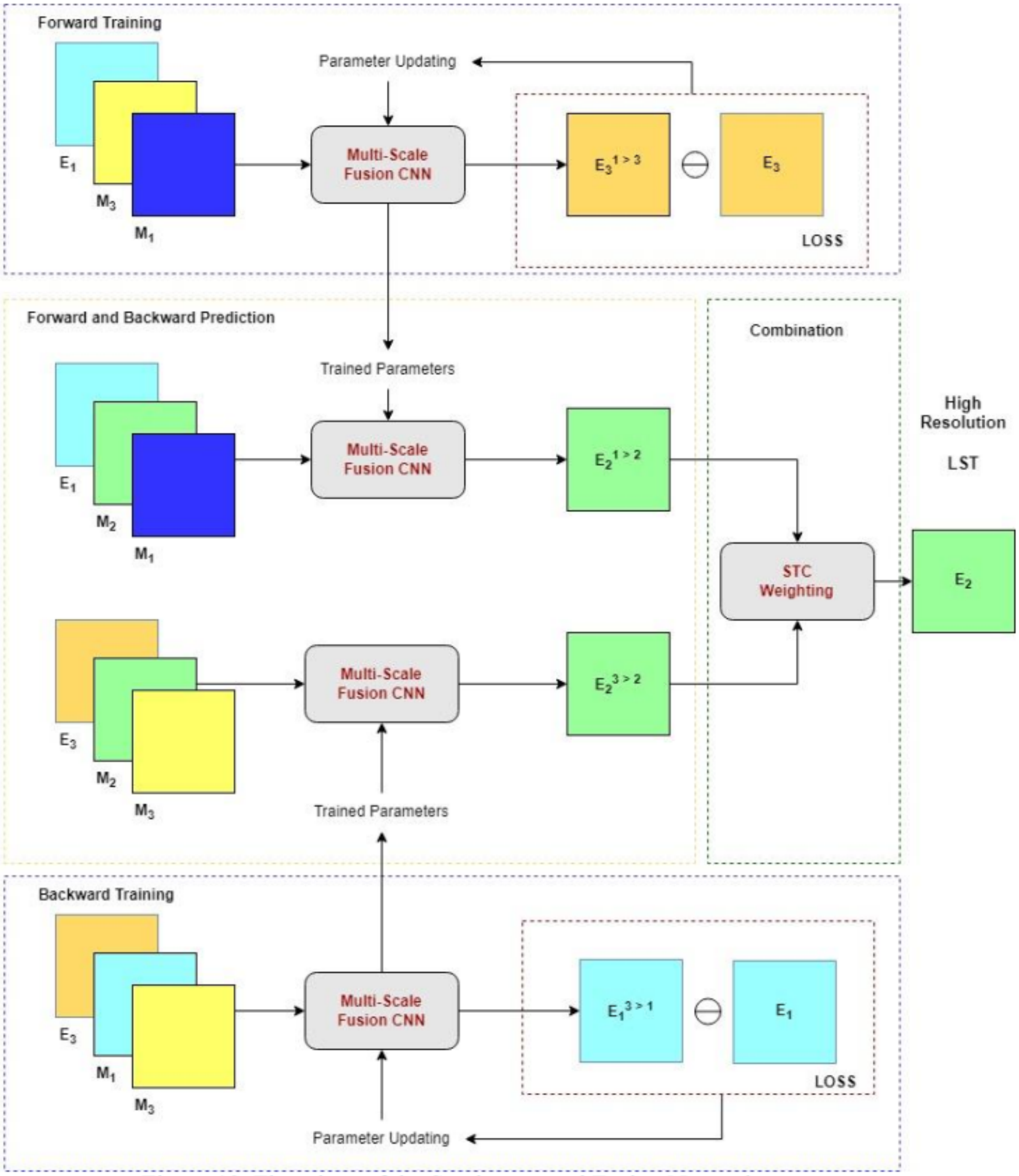


Figure 14. Flowchart of the Spatiotemporal Fusion Network.

The architecture of multi-scale fusion CNN is illustrated using the prediction of $E_2^{1>2}$ from E_1, M_1 and M_2 as example (Figure 15). The multi-scale fusion CNN is a fully convolution network, which enables an end-to-end mapping from three input images to an output high resolution image.

It first extracts high-level features of E1 and super resolves the change image between M_2 and M_1 at the same time, then fuses and retrieves the extracted feature maps to the fine spatial resolution result.

Therefore, the whole architecture of the multi-scale fusion CNN contains three major parts, termed here as:

- **Extraction-Net**, a two-layer convolution network employed to extract the high-level features of the ECOSTRESS image and provide fine abundant spatial pattern information for the prediction.
- **Super-Resolution-Net**. The focus in the fusion process is on locating temporal change between the target and neighboring dates. By using a super-resolution module, the coarse spatial resolution pixels of the temporal change (Subtraction) image are disaggregated into fine spatial resolution pixels, which can offer more spatial pattern information. To implement the super-resolution, a modified version of the super-resolution model “Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR)”, (Yu et al., 2018), is used, referred to here as slim-WDSR.
- **Integration-Net**, which is made up of three stacked convolutional layers. The generated feature maps are integrated and used to estimate the target image gradually through the Integration-Net.

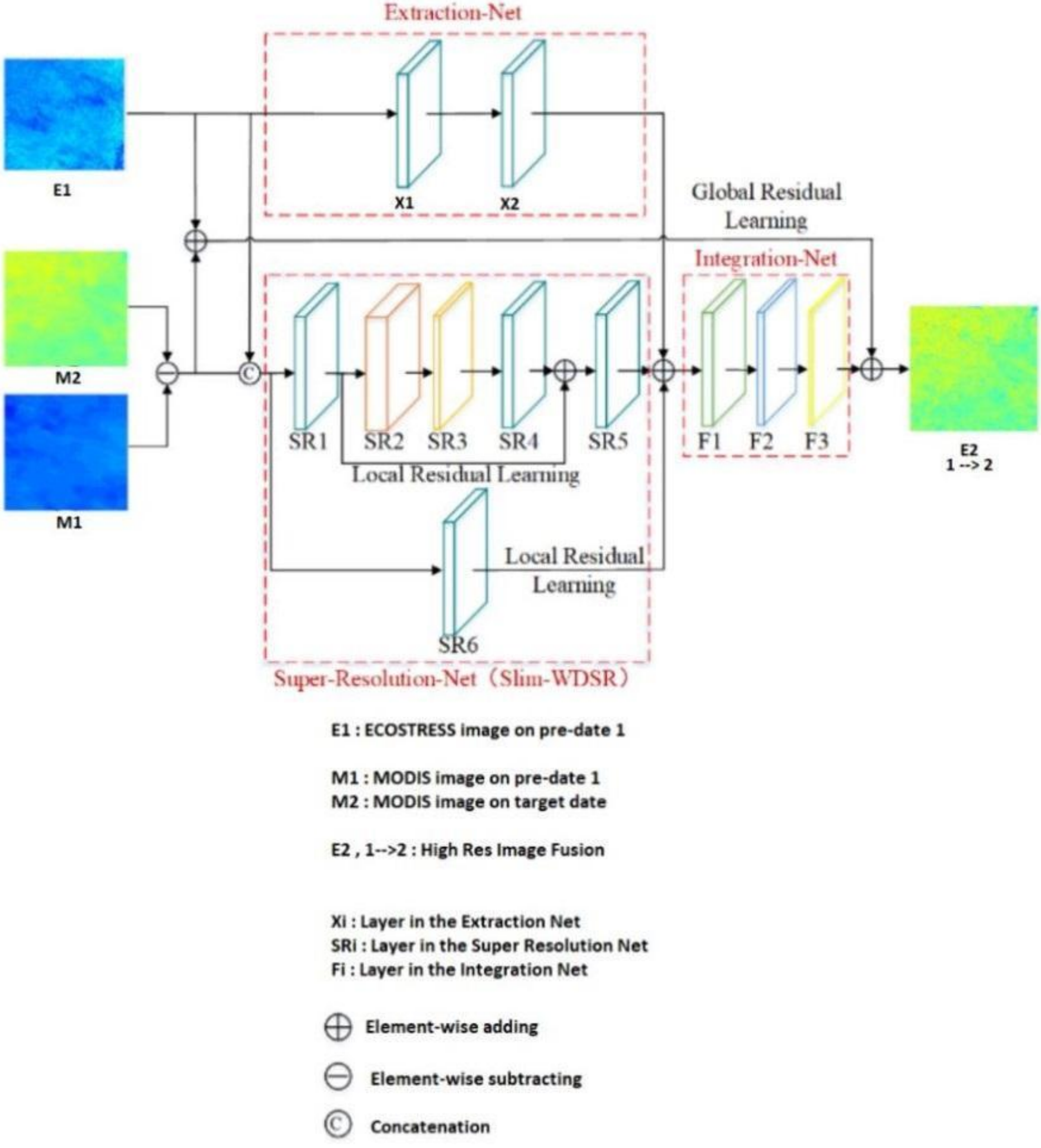


Figure 15. Architecture of Multiscale Fusion CNN.

The core components of the Multiscale Fusion CNN, such as the levels of each sub-network, as well as their respective information about the convolutional layers, kernel size, number of filters, activation functions, etc. are described on the following Table 13:

Table 13. Core components of the Multiscale Fusion CNN.

LAYERS OF THE MULTI-SCALE FUSION CNN					
Extraction-Net		**Super-Resolution-Net**		**Integration-Net**	
level	**layer: kernel size, filters**	**level**	**layer: kernel size, filters**	**level**	**layer: kernel size, filters**
X1	conv 3x3, 16,Batch Normalization,ReLU	**SR1**	conv 3×3, 32	**F1**	conv 3x3, 16,Batch Normalization,ReLU
X2	conv 3×3, 32	**SR2**	conv 1x1, 64, Batch Normalization,ReLU	**F2**	conv 3x3, 16,Batch Normalization,ReLU
		SR3	conv 1x1, 64, Batch Normalization,ReLU	**F3**	conv 3×3, 1
		SR4	conv 1x1, 32, Batch Normalization,ReLU		
		SR5	conv 3x3, 32, Batch Normalization,ReLU		
		SR6	conv 3x3, 32, Batch Normalization,ReLU		

Algorithm Training, optimization and prediction

Two multi-scale CNNs, a backward and a forward, were trained. A Huber Loss function was adopted instead of the mean squared error (MSE). Huber Loss or Smooth Mean Absolute Error is a loss function that takes the advantageous characteristics of the Mean Absolute Error and Mean Squared Error loss functions and combines them into a single loss function. The hybrid nature of Huber Loss makes it less sensitive to outliers, just like MAE, but also penalizes minor errors within the data sample, similar to MSE.

$$L_{\delta}(a) = \left\{ \frac{1}{2} \cdot a^2, for|a| \leq \delta \; \delta \cdot \left(|a| - \frac{1}{2} \cdot \delta \right), otherwise \right\}$$

This function is quadratic for small values of error a, the difference between the observed and predicted values, and linear for large values, with equal values and slopes of the different sections at the two points where absolute error equals δ. Huber Loss was selected because of its greater robustness to noise and outliers, possibly included to the network by unidentified cloudy pixels in input images. Loss function hyperparameter δ was set empirically to 1.

For the optimization, the weights of the network were initialized to small random values, which were drawn from a Gaussian distribution N(μ,σ)=(0,0.001).The Adam optimizer with standard back propagation was applied to minimize the loss and update the network weights until convergence, with a value of β_1=0.9 and β_2=0.999 for its settings. The learning rate α was initialized as 1×10-4 and was multiplied by a decaying factor 0.1 every 10 epochs to shrink the searching range of the parameters.

At prediction, two fine spatial resolution images $E_2^{1>2}$ and $E_2^{3>2}$ at the target date, t_2, were produced from the two trained networks. $E_2^{1>2}$ was produced from M_1, E_1 and M_2 images, from the forward trained CNN. $E_2^{3>2}$ was produced from M_3, E_3 and M_2 images, from the backward trained CNN. Finally, the two high resolution LSTs were combined in a weighted sum to obtain the final predicted LST.

$$\hat{E}_2 = p_1 \cdot E_2^{1\to 2} + p_3 \cdot E_2^{3\to 2}$$

For the determination of the weights p_1, p_3, the spatial consistency between the two predicted fine spatial resolution images and the corresponding coarse spatial resolution LST image was considered and a novel weighting strategy, STC-Weighting, was implemented. The weight magnitude in STC-Weighting depends on the difference between the predicted fine spatial resolution images and the coarse spatial resolution image at the target date, in such a way that a smaller difference results in a higher weight. The formula for the STC-Weighting is:

$$p_i = \frac{\frac{1}{\left|E_2^{i\to 2} - M_2\right|}}{\frac{1}{\left|E_2^{1\to 2} - M_2\right|} + \frac{1}{\left|E_2^{3\to 2} - M_2\right|}}, i = 1,3$$

Experimental Results:

The experimental design and selection of Area of Interest for the application of STTFN to MODIS & ECOSTRESS data came as a consequence of availability during image acquisition and as a follow-up to the preprocessing methods applied by fitting the MODIS data to a DTC model. Thus following the notation preceding in the methodology chapter the data of the experimental case study are listed in the following table (Table 14).

Table 14. Experimental case study image dataset for STTFN.

LST Image	Usage	Sensor	Date	Local Solar Time
E1	Training-Prediction	ECOSTRESS	17/01/2023	15.65

M1	Training-Prediction	MODIS	17/01/2023	14.1
E2	Ground Truth for Evaluation	ECOSTRESS	25/01/2023	23.96
M2	Training-Prediction	DTC Predicted MODIS	25/01/2023	23.96
E3	Training-Prediction	ECOSTRESS	05/12/2023	20.02
M3	Training-Prediction	MODIS	05/12/2023	21.8
<E2>	Predicted LST	STFN ECOSTRESS	25/01/2023	23.96

- Area Coverage: 100x100 km
- LST Acquisitions for the STTFN experiment only captured Winter phenological change on vegetation
- LST image pairs were resampled to common grid system with the Nearest Neighbor method
- Patching: Training data was fed into the STTFN as cropped patches of 32X32 pixels with an overlap stride equal to 16 (50 % overlap)
- Image Dimensions (pixels): 1416 X 1416
- Regarding quantitative evaluation this was performed by comparison with the available E2 image, which was kept aside for that particular reason. The mean absolute error (MAE), root mean square error (RMSE) and structural similarity (SSIM) were calculated using the ECOSTRESS high spatial resolution image at the target date as ground truth reference.

The output of the STTFN for the prediction image involved the reconstruction of the predicted patches into the initial image grid system. At the initial trials of the network, notable blocky artifacts (gridded pattern) were noticed on the reconstructed image, as shown clearly on Figure 16. This result posed a serious issue that had to be tackled. To mitigate the blocky artifacts, the followed approach involved:

- Training with a larger image patch size, keeping only a center region of the patch and removing the surrounding padding during reconstruction: This first modification, although it removed some lines, didn't have much impact on the reconstructed image. The gridded pattern was preserved.

- Apply smoothing and blending on the overlapping pixels between consecutive patches: With the application of filtering, most or all of the lines were removed but the image quality decreased a lot.

To gain a deeper understanding of the reason these artifacts occured, in the first place, the feature layers produced between convolutions had to be accessed and visualized, one by one. It was found that image blocks occurred even from the first convolution layer, and all the subsequent ones. An intuitive suggestion was that the reason for the generation of blocky artifacts could be the padding of the 2D convolutions. Padding is a technique used in convolutional neural networks (CNNs) to preserve the spatial dimensions of the input data and prevent the loss of information at the edges of the image. It involves adding additional rows and columns of pixels around the edges of the input data.

Initially padding was set to the “same” method, according to which additional rows and columns of pixels around the edges of the input data so that the size of the output feature map is the same as the size of the input data. This is achieved by adding rows and columns of pixels with a value of zero around the edges of the input data before the convolution operation. A first step was to try to remove these lines by cropping the surrounding border values and resizing the layer dimensions in order to feed the next layer. Doing so, indeed removed the blocky pattern, but as well degraded the reconstructed image quality severely. In a further modification, upsampling and/or transposing of convolution layers (and whenever it was needed, resizing and cropping layers to match the dimensions) in order to upscale the resolution, was performed, but this action didn’t enhance the prediction quality.

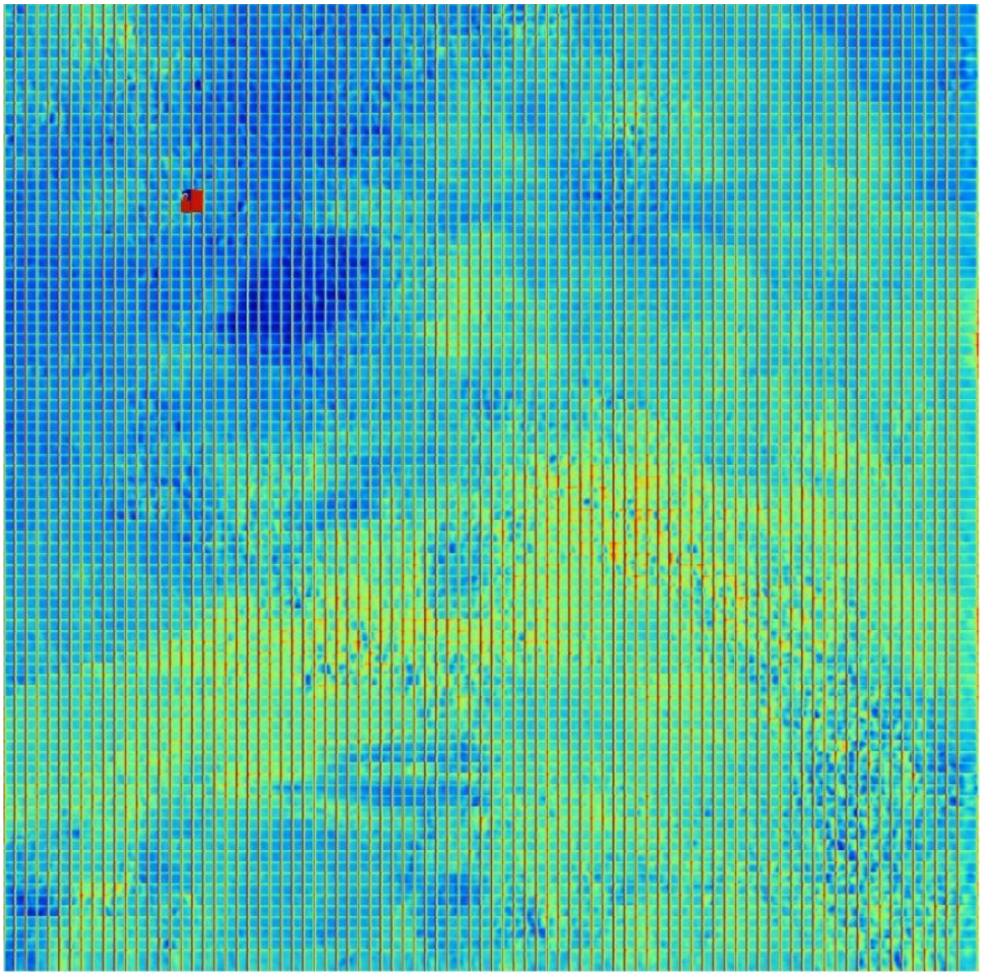


Figure 16. Blocky Artifacts created when reconstructing image from predicted patches. Notice the red spot indicating pixels with “no-data” values.

Finally, a “Valid” padding method was tested. This technique is used in CNNs to process the input data without adding any additional rows or columns of pixels around the edges of the data. This means that the size of the output feature map is smaller than the size of the input data. After changing to valid padding, the resulting image was free of lines and better quality than before but still not quite acceptable. After same further augmentations which comprised of:

- flip images, adjustment of brightness
- reducing the patch stride to 8 (25% overlap)

the predicted reconstructed result reached a good error result, acceptable in relation to the reference paper on which STTFN implementation was based on. Evaluation metrics of the study case for the predicted ECOSTRESS image was:

- **RMSE: 1.949**
- **MAE: 1.649**

- **SSIM: 0.974**

In summary, the modified setup of the model was:

- Patching: size 32 and stride 8 pixels
- Image dimensions: resized to 1408 X 1408
- Total number of training patches: 29926
- Data Split: 80 % for training, 20 % for validation
- Early stopping and Dropout Layers modes were used in order to prevent overfitting
- In every epoch, the best model weights were saved
- Usage of custom Keras generator to scale data to [0,1] range, perform the augmentation, and return a batch of data, with a size of 16
- The model was trained on an NVIDIA GeForce GTX1650 with 4GB memory and the training process takes around 50-60 minutes depending on early stopping.

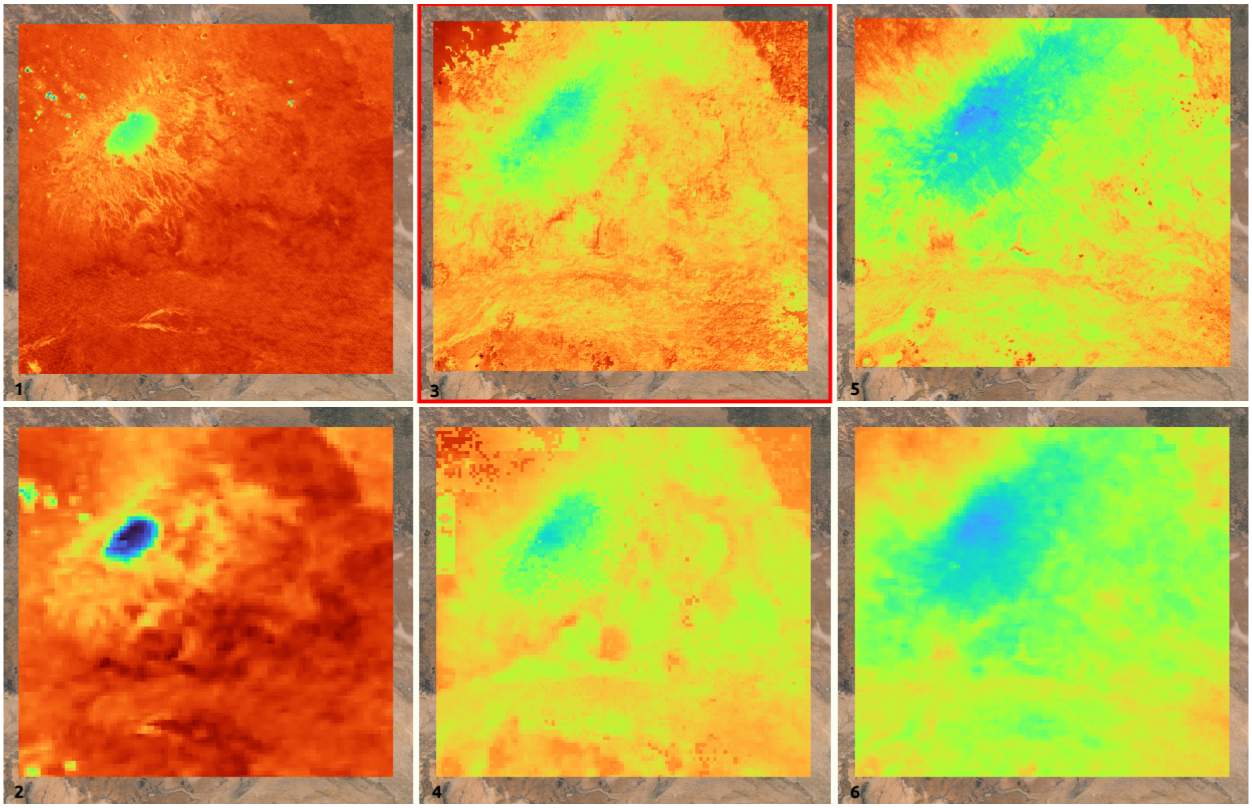


Figure 17. STTFN prediction results in regards with the input data: (1) E_1 ECOSTRESS, (2) M_1 MODIS, (3) Predicted STTFN E_2 ECOSTRESS, (4) Predicted DTC M_2 MODIS, (5) E_3 ECOSTRESS, (6) M_3 MODIS.

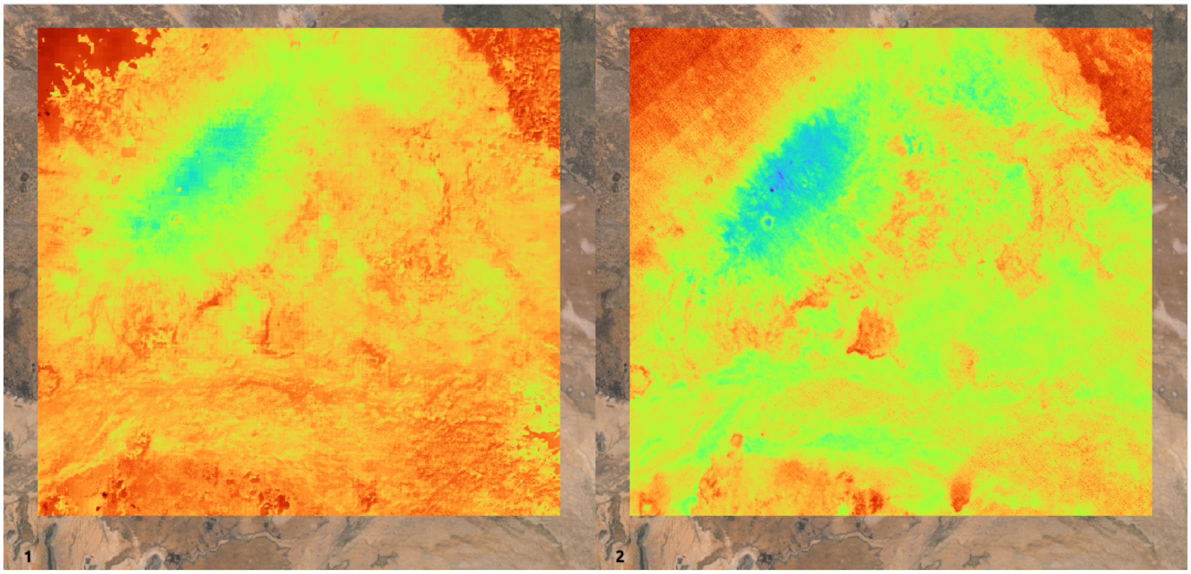


Figure 18. STTFN prediction results (1) in regards E_2 ECOSTRESS Ground Truth Data (2).

The mapped output prediction of the STTFN in the experimental study area is depicted along with the input data of the spatiotemporal model, E_1, M_1, M_2, E_3, M_3, in Figure 17. In Figure 18 the predicted output is compared with the ground truth image E_2.

Issues & Next Steps:

In regards with the process of Spatiotemporal Fusion of ECOSTRSS and MODIS LST thermal images, through a Deep Learning Framework application over the STTFN architecture, the evaluated accuracy results were found to be comparable and within the acceptable error range defined by the relevant thermal image fusion literature (ESTARFM, STTFN, StfNet methods), in terms of RMSE and SSIM metrics. However, the mapping and localization of error in the conducted study, indicated that some minor further improvement could be possible. Error observations are somehow related to the blocky artifacts issue of the Multiscale CNN, which was highly improved but not fully addressed. Besides the emergence of blocky artifacts in images produced by CNNs and patch training is a general issue by itself in Deep Learning research as noted by Xu et al. (2022) and Huang et al. (2019), where only general guidelines are provided and frequently the solutions are found in a heuristic manner.

- A further next step, in a STTFN algorithm development would be to test its performance in a more extended spatiotemporal context, involving a greater land cover variability, as well as a seasonal assessment study to verify the stability of the accuracy across the year timeframe.

Some more aspects that could be evaluated in a 2nd iteration might include:

- to test the sensitivity of the STTFN model to variable cloudy conditions. The requirement of cloudless conditions for the fused scene, has a huge impact on the available data that could be used and benefit from the fusion capabilities. It is important to assess to what extent cloud presence deteriorates the fusion result locally (pixelwise regarding the cloud mask) and focally (on cell locations within a neighborhood from the cloud mask).

- to test the sensitivity of the STTFN model to different local time of image acquisitions. While the DTC model is a highly useful tool in thermal data, by itself, it introduces to a greater or lesser extent some inherent error to the data used for training. Even more restrictions about cloud coverage are imposed for all the diurnal time-frames, which makes it even harder to find data with cloudless conditions on all 4 daily acquisitions. For that reason, it would be great knowledge to infer to what extent the temporal inconsistency of ECOSTRESS imagery affects the fusion result.

3.8.1. Spatiotemporal Thermal Data Fusion - Update iteration

Following the first version of the deliverable "D7 - Algorithm Theoretical Base Document and Product Specifications", further modifications and tests were carried out in an effort to upgrade the performance and accuracy of the Spatiotemporal Thermal Data Fusion results.

As demonstrated by the preceding approaches presented in the sections above, despite all the satisfactory performance and accuracy of the STTFN model, issues arose that were related to the feasibility of this methodology to be implemented as an operational service. The experimental design was carried out for zero cloud cover conditions, and by applying a temporal normalization of the participating MODIS LST images to the target time of the ECOSTRESS LST image acquisition, that was carried out by using a DTC model. The prior assumption is a condition not often encountered at the operational level, as cloud cover is an uncontrollable parameter that is a hurdle, and is often encountered in Thermal Remote Sensing. Moreover, the presence of cloud cover makes it impossible to calibrate the DTC model. If for a certain location, any one of the 4 MODIS LST daily exposures shows cloud cover, then for that pixel it is not possible to calculate the model coefficients, resulting in a no data value during prediction. For these reasons, it was decided to modify the experiment to test the STTFN under random cloud cover conditions, and at non-normalized time and synchronized hours of thermal data. The experimental design for the application of STTFN on this 2nd iteration remained on the same Area of Interest, but regarded thermal data pairs from different dates. The data of the experimental case study are listed in the following table (Table TTT21).

Table TTT21. Experimental case study image dataset for STTFN.

LST Image	Usage	Sensor	Date	Local Solar Time
E1	Training-Prediction	ECOSTRESS	27/01/2022	12.11
M1	Training-Prediction	MODIS	27/01/2022	13.2

E2	Ground Truth for Evaluation	ECOSTRESS	06/03/2022	20.96
M2	Training-Prediction	MODIS	06/03/2022	21.8
E3	Training-Prediction	ECOSTRESS	26/07/2022	13.12
M3	Training-Prediction	MODIS	26/07/2022	14.1
<E2>	Predicted LST	STFN ECOSTRESS	06/03/2022	21.8

- Area Coverage: 100x100 km
- LST image pairs were resampled to common grid system with the Nearest Neighbor method
- Patching: Training data was fed into the STTFN as cropped patches of 32X32 pixels
- Image Dimensions (pixels): 1416 X 1416
- Regarding quantitative evaluation this was performed by comparison with the available E2 image, which was kept aside for that particular reason. The mean absolute error (MAE), root mean square error (RMSE) and structural similarity (SSIM) were calculated using the ECOSTRESS high spatial resolution image at the target date as ground truth reference.

For the implementation the setup of the STTFN model was not modified from its initial specifications which included :

- Patching: size 32 and stride 8 pixels
- Image dimensions: resized to 1408 X 1408
- Total number of training patches: 29926
- Data Split: 80 % for training, 20 % for validation
- Early stopping and Dropout Layers modes were used in order to prevent overfitting
- In every epoch, the best model weights were saved
- Usage of custom Keras generator to scale data to [0,1] range, perform the augmentation, and return a batch of data, with a size of 16
- The model was trained on an NVIDIA GeForce GTX1650 with 4GB memory and the training process takes around 50-60 minutes depending on early stopping.

The thermal data inputs to the STTFN, ECOSTRESS and MODIS LST, as set for the 2nd experimental setup are presented on figure XXX19. Since MODIS LST data were provided preprocessed and masked from the API, spatial interpolation had to be performed in order to fill the no data cells at the cloudy locations. Cloudy coverage was interpolated, with a bilinear spline to the radiometric value of the most proximal pixels, at the boundaries of the cloud mask. LST values at those spots lie between the “real” LST and the much lower LST values mapped at the center of the cloud. This introduced some mismatch between the ECOSTRESS and MODIS cloudy pixel values.

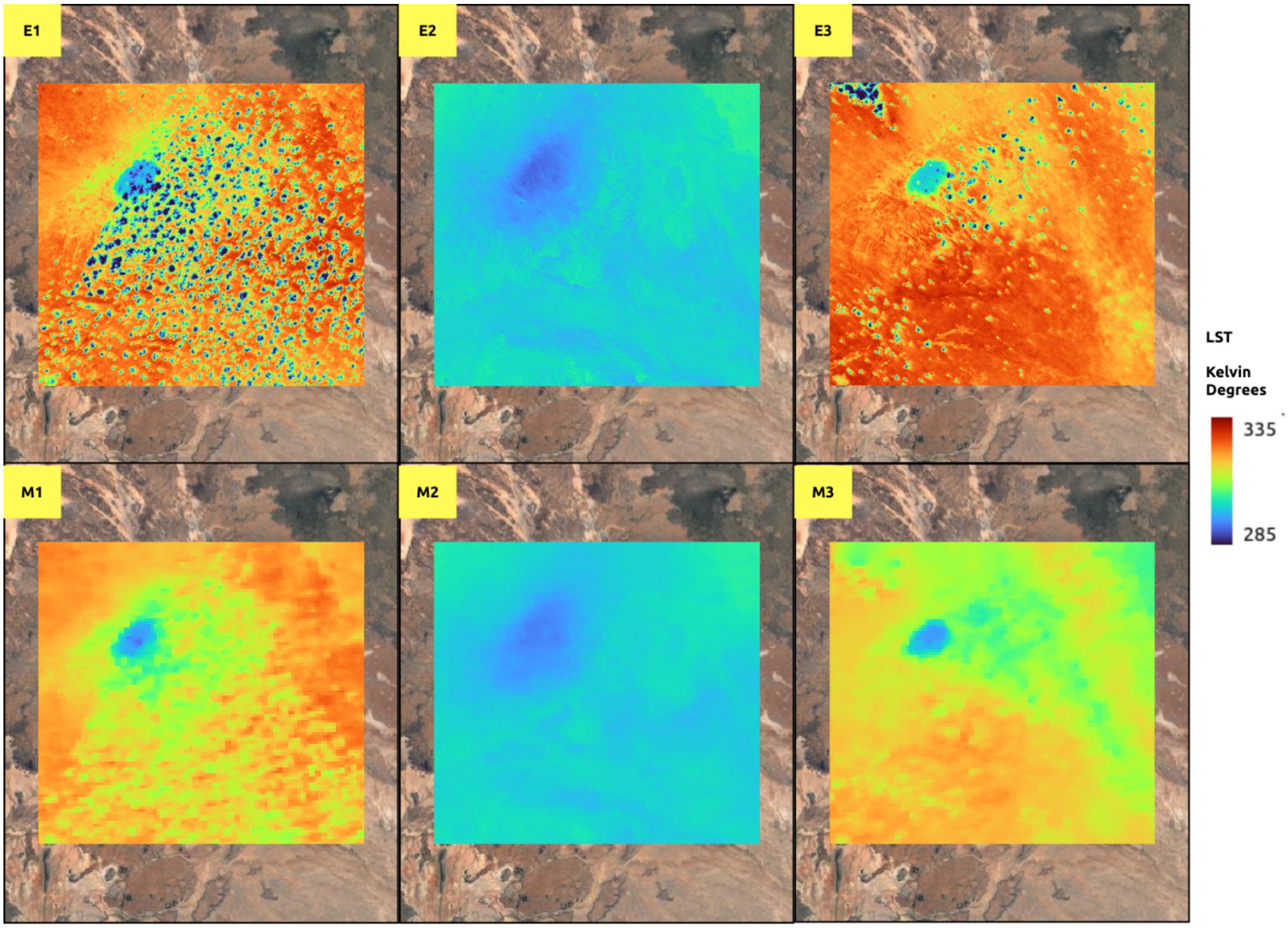


Figure XXX19. data inputs to the STTFN, ECOSTRESS and MODIS LST, as set for the 2nd experimental setup.

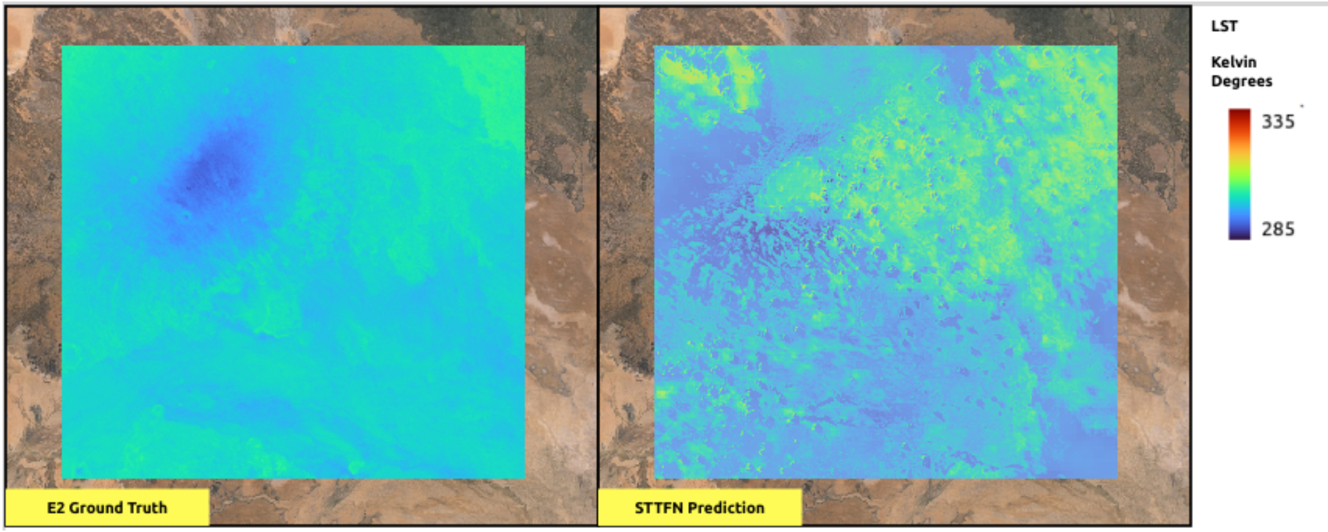


Figure XXX20. STTFN Validation, Ground Truth ECOSTRESS LST image on E2 time (left) and STTFN Prediction LST, as inferred from the 2nd experimental setup.

The prediction of the STTFN for the intermediate date (Figure XXX20) yielded unsatisfactory results, in terms of radiometric accuracy and spatial feature reconstruction. Evaluation metrics of this 2nd experimental study case for the predicted fused image were:

- **RMSE: 6.410**
- **MAE: 5.222**
- **SSIM: 0.898**

The absolute error between STTFN prediction and the E2 Ground truth was further analyzed over its value distribution and its spatial distribution. The absolute error histogram is given in figure XXX21.

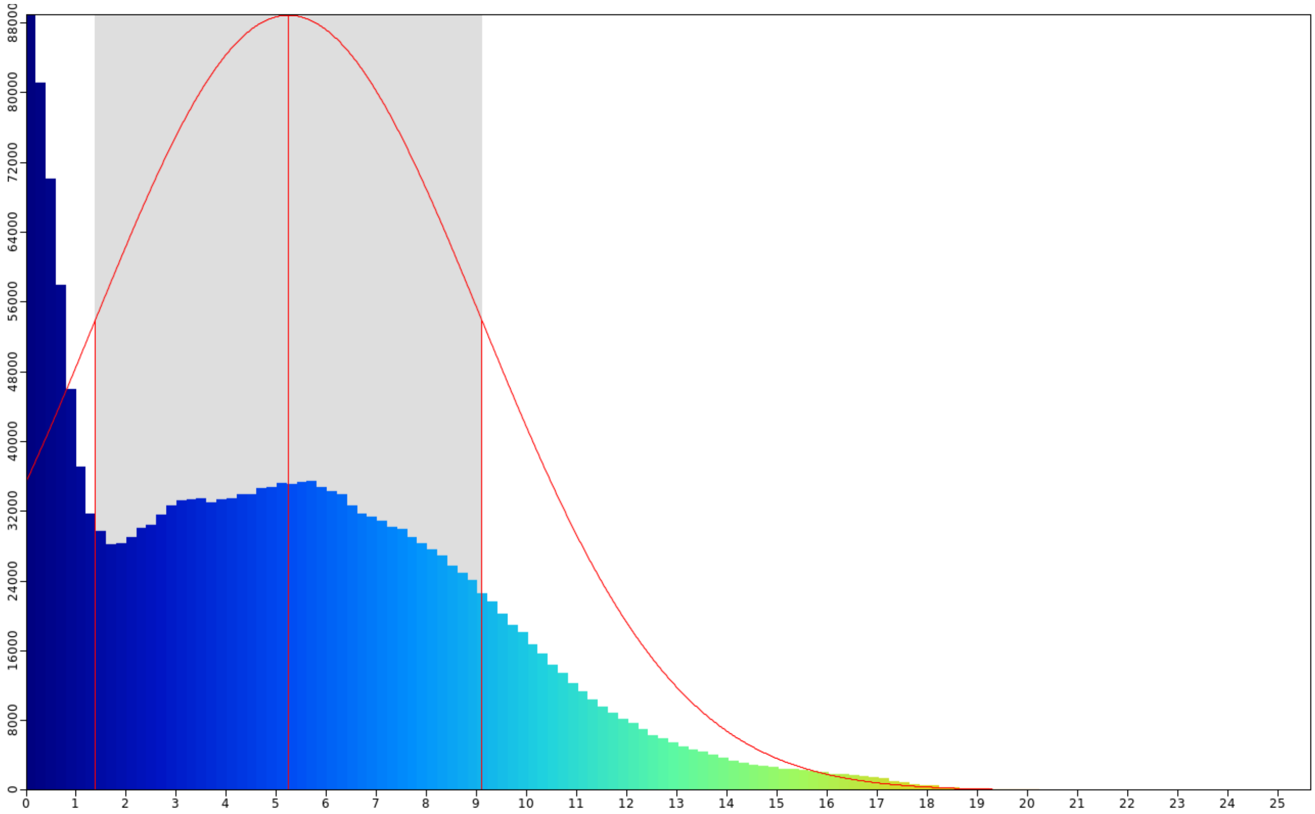


Figure XXX21. STTFN absolute error histogram. 2nd experimental setup.

For the spatial analysis, a union (Boolean OR) of the ECOSTRESS and MODIS cloud masks was created to be used as a categorical feature for the zonal analysis of error. Zonal Statistics of Absolute Error was calculated for cloudy and non cloudy coverage and are presented on the following table (TTT22). It is evident that the average absolute error in clouded pixels is much higher with a larger variance.

Table TTT22. Zonal Statistics of STTFN Absolute Error in cloudy and non cloudy coverage.

Coverage	Count	Minimum	Maximum	Average	St. Deviation
Cloudy	347263	0	19.63	8.47	6.12
Non Cloudy	1590882	0	9.87	3.56	2.12

Even more than the low global metrics, the confirmation about the reduced capability of STTFN for thermal data fusion on cloudy conditions also becomes evident from the qualitative evaluation of the result. First of all, it is obvious that the fused result exhibits artifacts related to the cloudy coverage on the dates before and after the prediction. The spatial pattern of the artifacts introduced in the prediction is clearly similar to the cloudy coverage present in the E1,M1, and E3, M3, images. The temporal mismatch between MODIS and ECOSTRESS images in the training pairs is very likely to contribute to the radiometric error, but this hypothesis is not quantifiable since in all three pairs the time difference is relatively small, and especially since in the prediction date we are in the evening hours, during which there is slow thermal attenuation on the DTC curve, and thus the temperature deviations for an hour time difference are expected to be even smaller.

The conclusions derived from the experimental design carried out confirm the working hypothesis that the Spatiotemporal Thermal Fusion tools developed can only produce reliable results under conditions of non cloudy coverage. In cases where the temporal difference between MODIS and ECOSTRESS images is large, it is necessary to preprocess the data via DTC Curve Fitting for the temporal normalization of the inputs. In this case the non-cloudy conditions must hold for all 4 MODIS images. From the experience gained with data acquisition and preprocessing of thermal data, it became clear that the requirement for zero cloud cover conditions, dramatically limits the available images that could be used. The application training and prediction of thermal data with STTFN, could be operationally done by user selection of those Areas of Interest that have non cloudy conditions. In this case the size of this area would be provided by the spatial intersection of non cloudy AOIs, through all available days and will be relatively small. This would have an effect on the available patches for training, and the overall accuracy of the algorithm.

4. Hyperspectral Data Fusion

4.1. Objectives

Hyperspectral and multispectral images data fusion is a powerful technique in the field of remote sensing providing a comprehensive view of the Earth’s surface through the combination of the advantages of different spectral and spatial resolutions. Hyperspectral sensors excel in capturing images with detailed spectral information across multiple narrow bands, which allows for exact characterization of materials and processes. However, the spatial resolution is sacrificed, which poses an implication in their applicability in tasks that require finer spatial details. Multispectral sensors, on the other hand, provide imagery with a higher spatial resolution but with fewer spectral bands, which limits their ability to distinguish between materials with small spectral differences.

The fusion of hyperspectral and multispectral data targets on addressing these limitations by integrating the detailed spectral information of hyperspectral imagery with the high spatial resolution of multispectral images. Through this approach, the combined spectral information of hyperspectral data with the spatial information of multispectral data results in a fused dataset that provides both high spectral and spatial resolution. Therefore, more precise and detailed insights regarding land cover, vegetation health, environmental conditions and a variety of applications are provided.

Within AFRI4Cast, the primary objective of hyperspectral-multispectral data fusion was the enhancement of the PRISMA hyperspectral images spatial resolution by using higher resolution multispectral imagery provided by Sentinel-2 MSI. By fusing these datasets following different methods, such as hyper pansharpening and deep learning fusion, AFRI4Cast targets on creating enhanced hyperspectral images that are able to provide accurate vegetation indices and canopy biophysical parameters at a finer spatial scale as well as contribute to crop disease models.

4.2. Data Products

The current chapter presents the Earth Observation (EO) hyperspectral and multispectral data products that are used within AFRI4Cast as inputs to the data processing pipelines that regard the hyperspectral/ multispectral data fusion. PRISMA Level 2D (L2D) as well as Sentinel-2 MSI Level 2A data products are mentioned in this chapter. For more detailed information, please refer to the Deliverable D3 “State of the art review report on analysis methods and algorithms,” Chapter 3. “Input data products for AFRI4Cast EO services”.

• PRISMA Level 2D

PRISMA Images Features

In the terms of AFRI4Cast, PRISMA Level 2D hyperspectral imagery at a spatial resolution of 30m is used in order to perform the hyperspectral/ multispectral data fusion pipeline. Each PRISMA image consists of 239 spectral bands ranging from 400 to 2500 nm.

PRISMA images are comprised of three data cubes:

- The VNIR Cube containing 66 spectral bands in the spectral range from 400 to 1010 nm,
- the SWIR Cube containing 173 spectral bands in the spectral range from 920 to 2500 nm and
- the PAN Cube containing 1 panchromatic band at 5m spatial resolution

PRISMA images are distributed via ASI’s PRISMA platform and they are delivered in HDF-EOS5 format (.he5). Each data product is divided in three data cubes which can be accessed via the following paths in the HDF5 file:

- VNIR Cube: ‘/HDFEOS/SWATHS/PRS_L2D_HCO/Data Fields/VNIR_Cube'
- SWIR Cube: ‘/HDFEOS/SWATHS/PRS_L2D_HCO/Data Fields/SWIR_Cube'
- PAN Cube: ‘/HDFEOS/SWATHS/PRS_L2D_PCO/Data Fields/PAN_Cube'

Based on the spectral range of bands that are contained in the VNIR and SWIR cubes, an overlap between these two cubes is observed. VNIR and SWIR cubes overlap is noticed at around 943.35 to 979.22 nm (Table 11.).

Table 15. VNIR and SWIR PRISMA cubes overlapping spectral areas – bands.

VNIR Cube bands (nm)	SWIR Cube bands (nm)
923.95	**943.35**
934.11	**951.40**

944.62	**959.97**
956.27	**969.84**
967.02	**979.22**
977.36	988.91
	998.90
	1008.64

Data Scaling

Concerning the PRISMA data scaling aspect, all data cubes are stored in the L2D products as Digital Numbers (DN) with a uint16 pixel depth. In order for a data scaling to be performed and a conversion from uint16 to Float32 data type to take place, the following conversion formula has to be applied *(Source: ASI, "PRISMA ATBD", 14/12/21)*. Through this formula, the cubes' data are converted from Uint16 to Float32 reflectance units.

$$XXX = L2ScaleXXXMin + \frac{XXX_{DN} * L2ScaleXXXMax - L2ScaleXXXMin}{65535}$$

where <XXX> refers to "PAN", "VNIR" or "SWIR" dataset.

The conversion/ scaling factors are apparent in the metadata in the tags:

L2ScalePanMax

L2ScalePanMin

L2ScaleSwirMax

L2ScaleSwirMin

L2ScaleVnirMax

L2ScaleVnirMin

A significant finding that was observed while working with PRISMA data products was that conversion/ scaling factors were found to be inconsistent throughout all PRISMA acquisitions. The inconsistency of PRISMA scaling factors posed an implication on the fully automated methodology approach, because each separate image acquisition should have been scaled using its own individual scaling factors, which had to be retrieved beforehand. Moreover, it was noticed that the cloud coverage information provided by PRISMA images metadata is not indicative of the images' actual cloud coverage. Thus, the evaluation of PRISMA acquisitions based on their cloud coverage was carried out through visual inspection, since a quality control layer is not available.

- **Sentinel-2 MSI Level 2A**

Sentinel-2 MSI Level 2A data products are part of the AFRI4Cast's data fusion component, concerning the aspect of PRISMA spatial enhancement through the data fusion approach of hyperspectral and multispectral imagery. For more detailed information regarding this data product, please refer to the Deliverable D3, Chapter "3. Input data products for AFRI4Cast EO services".

4.3. Data Processing Pipelines

The hyperspectral/ multispectral data processing and fusion pipelines are depicted in Figure 19.

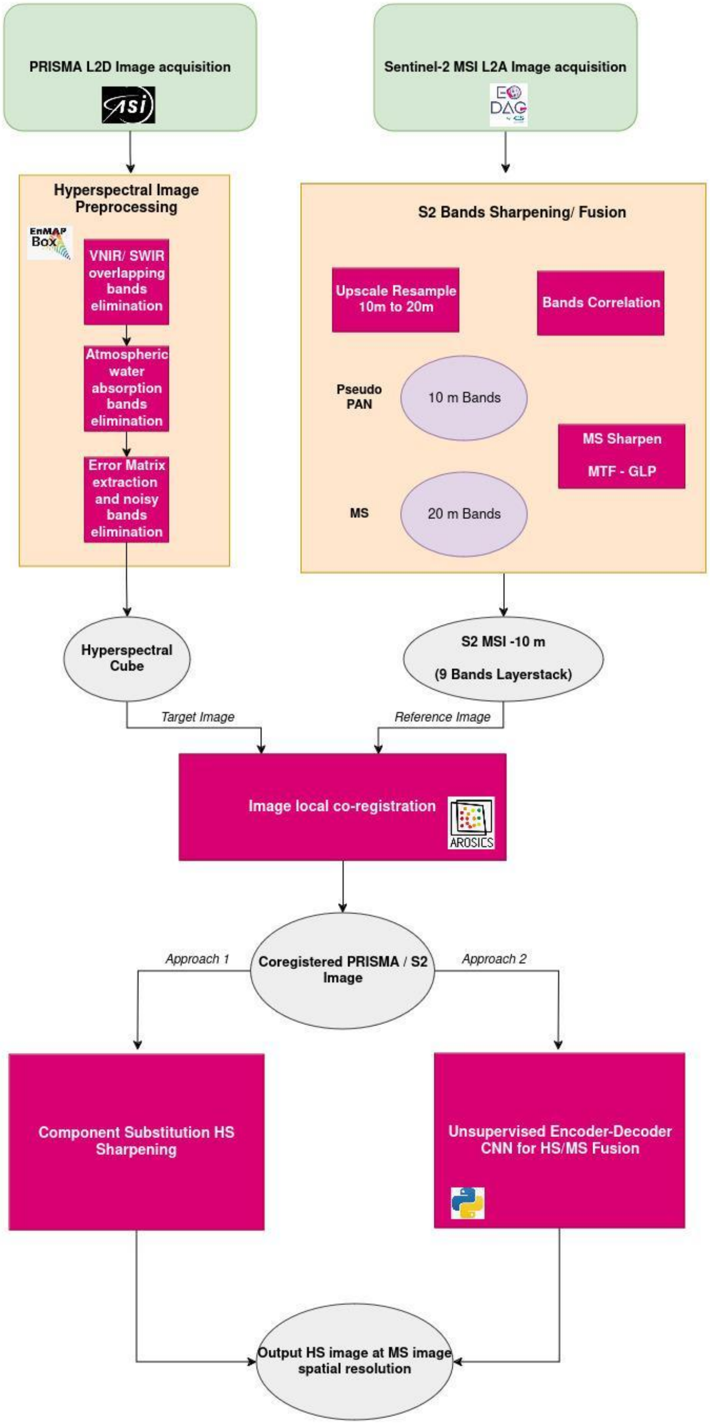


Figure 19. Hyperspectral/ multispectral data processing and fusion pipelines.

4.4. Hyperspectral – Multispectral Image Acquisition

Approach:

In this chapter, the ways of acquiring the images to be used for the spatial enhancement pipeline of hyperspectral (HS) imagery through the data fusion of hyperspectral and multispectral (MS) images within AFRI4Cast are presented. Image availability of hyperspectral and multispectral imagery as well as the examination of data quality in the sense of cloud coverage in order to form HS/ MS image pairs to serve as input in the data fusion approach are discussed.

Prisma L2D

In the terms of AFRI4Cast, PRISMA Level 2D imagery is acquired via ASI's PRISMA Platform (https://prisma.asi.it) through the tool "Mission Planning Operation". That approach was adopted due to the lack of PRISMA observations in the platform's archive over the study sites, thus acquisition requests had to be submitted.

The bare minimum settings for the submission of a new PRISMA acquisition through the "Mission Planning Operation" tool are specified by:

1. the latitude and longitude of the image center (decimal degrees),
2. the maximum permitted percentage of cloud coverage (%)
3. the minimum and maximum Look Angle
4. the minimum and maximum Sun Zenith Angle
5. the type of product which shall be generated at image acquisition completion.

The acquired images for the placed requests over the study sites during 2023 – 2024 present a sparse but rather adequate availability.

Further on, their quality in terms of cloud coverage was examined. It was observed that the images' metadata don't provide reliable information regarding the cloud coverage and a quality control layer is not provided. Consequently, the quality was evaluated through visual inspection.

Finally, four PRISMA images during 2023 were selected as optimal for our studies. The list of acquired images along with the selected ones as optimal acquisitions (marked with tick symbol) are presented in Table 16.

Sentinel-2 MSI L2A

Sentinel-2 MSI Level 2A imagery was retrieved via EODAG for the same or close date to each optimal PRISMA image (based on their cloud coverage evaluation). A custom developed Python script, that uses the EODAG Python package capabilities and is pre-configured to work with CreoDIAS, Copernicus Dataspace Ecosystem and ONDA DIAS API, is employed in order to explore the availability and retrieve this data product.

Sentinel-2 MSI L2A imagery quality, regarding the cloud coverage categories, was examined through the SCL layer provided along with the specific data product.

Table 16. PRISMA L2D available acquisitions over the study sites and corresponding Sentinel-2 MSI L2A image availability on the same or close dates to good quality PRISMA images.

Order No	Date of Acquisition	Center Y	Center X	Study Site Coverage	Product Name	Quality Check	S2 Availability and Date	S2 Quality Check
54897	**26/05/2020**	**0,05**	**35,97**	**Perkerra**	**PRS_L2D_STD_20200526080833_20200526080837_0001**	✓	**✓ (26/05)**	**Good Quality**
54902	02/06/23	-0,23	35,95	Nakuru	PRS_L2D_STD_20230602080857_20230602080902_0001\\	X		
54903	**22/08/2023**	**-0,23**	**35,95**	**Nakuru**	**PRS_L2D_STD_20230822080544_20230822080548_0001**	✓	**✓ (24/08)**	**Good Quality**
54708	**28/08/2023**	**0,47**	**36,04**	**Perkerra**	**PRS_L2D_STD_20230828080855_20230828080859_0001**	✓	**✓ (29/08)**	**Good Quality**
20386	07/10/23	-0,70	37,29	Mwea	PRS_L2D_STD_20231007075908_20231007075912_0001\\	X		
20382	08/10/23	-0,13	34,90	Ahero	PRS_L2D_STD_20231008081539_20231008081544_0001\\	X		
19869	13/10/2023	0,47	36,03	Perkerra	PRS_L2D_STD_20231013080218_20231013080223_0001\\	X		
54899	14/10/2023	0,03	34,84	Ahero	PRS_L2D_STD_20231014081900_20231014081905_0001\\	X		
54901	**14/10/2023**	**-0,22**	**34,78**	**Ahero**	**PRS_L2D_STD_20231014081905_20231014081909_0001**	✓	**✓ (13/10)**	**Good Quality**
20389	25/10/2023	-0,21	35,95	Nakuru	PRS_L2D_STD_20231025080857_20231025080901_0001\\	X		
20387	06/11/23	-0,13	34,90	Ahero	PRS_L2D_STD_20231106081553_20231106081557_0001\\	X		
21572	08/01/24	-0,11	34,90	Ahero	PRS_L2D_STD_20240108080224_20240108080228_0001\\	X		
21576	19/01/2024	-0,72	37,31	Mwea	PRS_L2D_STD_20240119075232_20240119075237_0001\\	X	✓ (23/01)	Medium Quality

21573	20/01/2024	0,46	36,04	Perkerra	PRS_L2D_STD_20240120080848_20240120080852_0001\	X		

PRISMA / Sentinel-2 Image Pairs

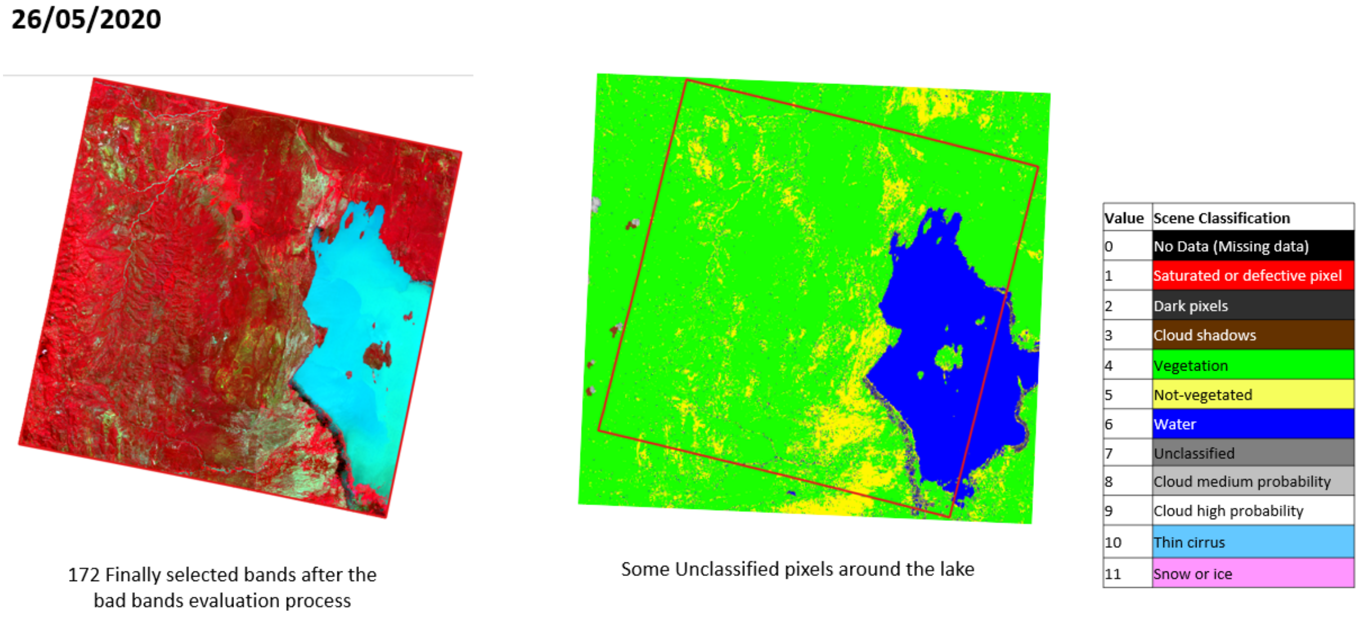
In order to fulfill the hyperspectral/multispectral data fusion pipeline, the possible PRISMA/ Sentinel-2 image pairs with the minimum flaws and cloud coverage were examined. In order to realize that, Sentinel-2 images were acquired in the same or near dates to those that PRISMA images evaluated as optimal exist (Table 16.).

The HS/ MS image pairs were further examined in both ways:

a. PRISMA L2D images underwent the bad bands evaluation process, as it is described in detail in Chapter 4.6.2.

b. Sentinel-2 MSI L2A images were evaluated via the SCL layer in the area that is within the swath of the corresponding PRISMA image on a specific date. The presence of clouds, dark pixels, shadows and unclassified pixels was determined.

These possible PRISMA/ Sentinel-2 image pairs that the optimal quality coincided on both images were observed on 26/05/2020, 22/08/2023, 28/08/2023 and 14/10/2023. These image pairs are also presented on Figures 20 - 23 and Table 16. (marked with bold font and orange selection lines).


26/05/2020

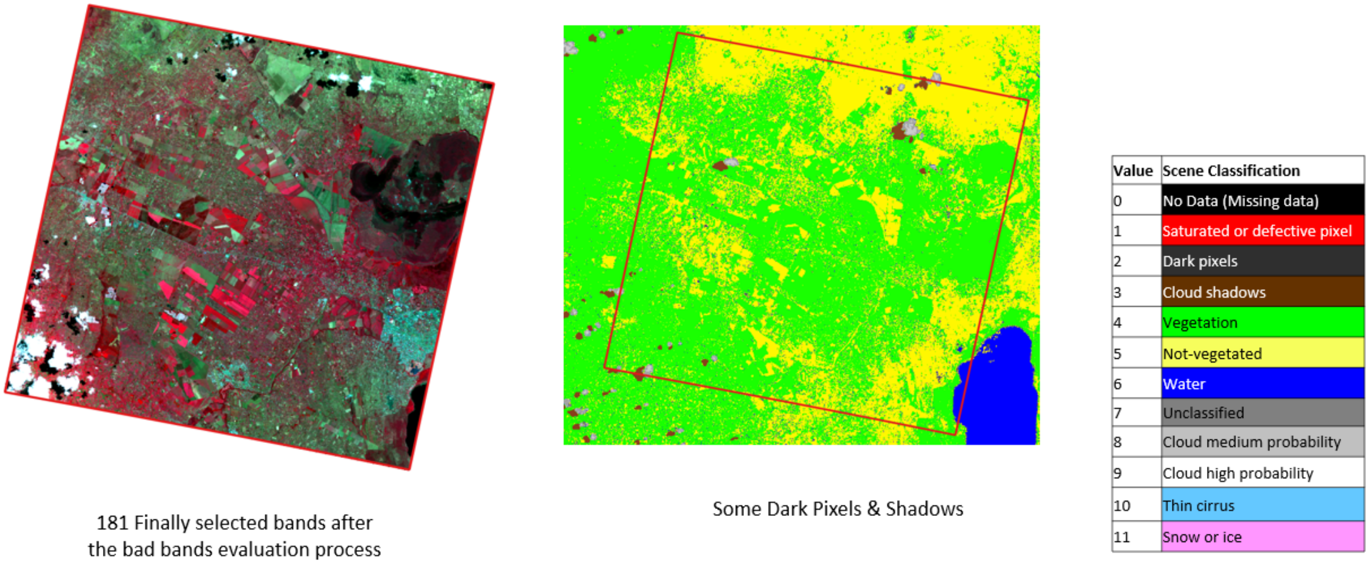
172 Finally selected bands after the bad bands evaluation process

Some Unclassified pixels around the lake

Value	Scene Classification
0	No Data (Missing data)
1	Saturated or defective pixel
2	Dark pixels
3	Cloud shadows
4	Vegetation
5	Not-vegetated
6	Water
7	Unclassified
8	Cloud medium probability
9	Cloud high probability
10	Thin cirrus
11	Snow or ice

Figure 20. PRISMA and Sentinel-2 SCL images on 26/05/2020.

22/08/2023



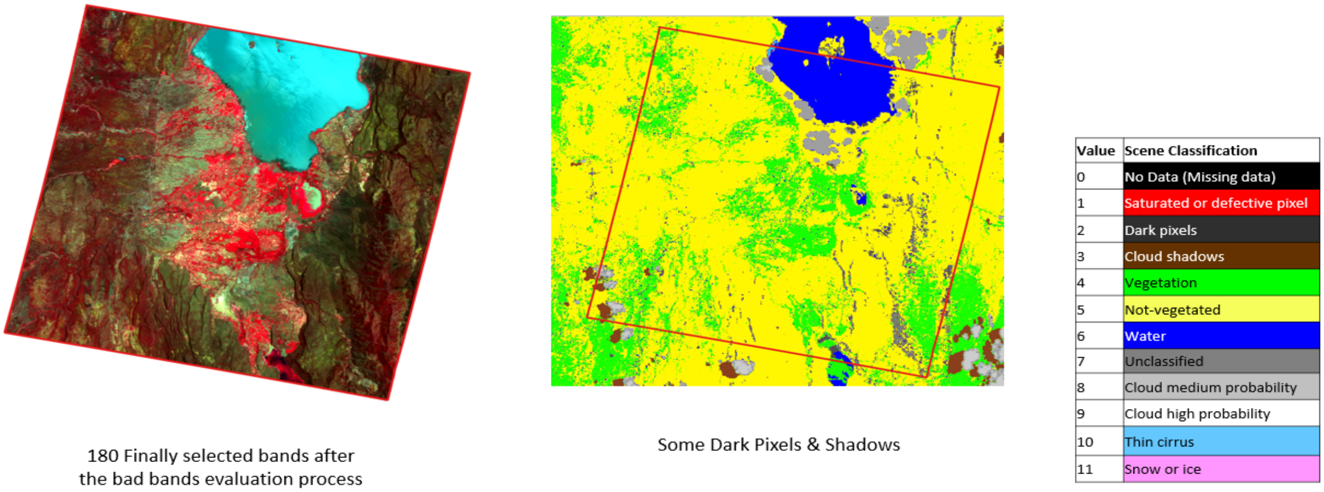
181 Finally selected bands after
the bad bands evaluation process

Some Dark Pixels & Shadows

Value	Scene Classification
0	No Data (Missing data)
1	Saturated or defective pixel
2	Dark pixels
3	Cloud shadows
4	Vegetation
5	Not-vegetated
6	Water
7	Unclassified
8	Cloud medium probability
9	Cloud high probability
10	Thin cirrus
11	Snow or ice

Figure 21. PRISMA and Sentinel-2 SCL images on 22/08/2020.

28/08/2023

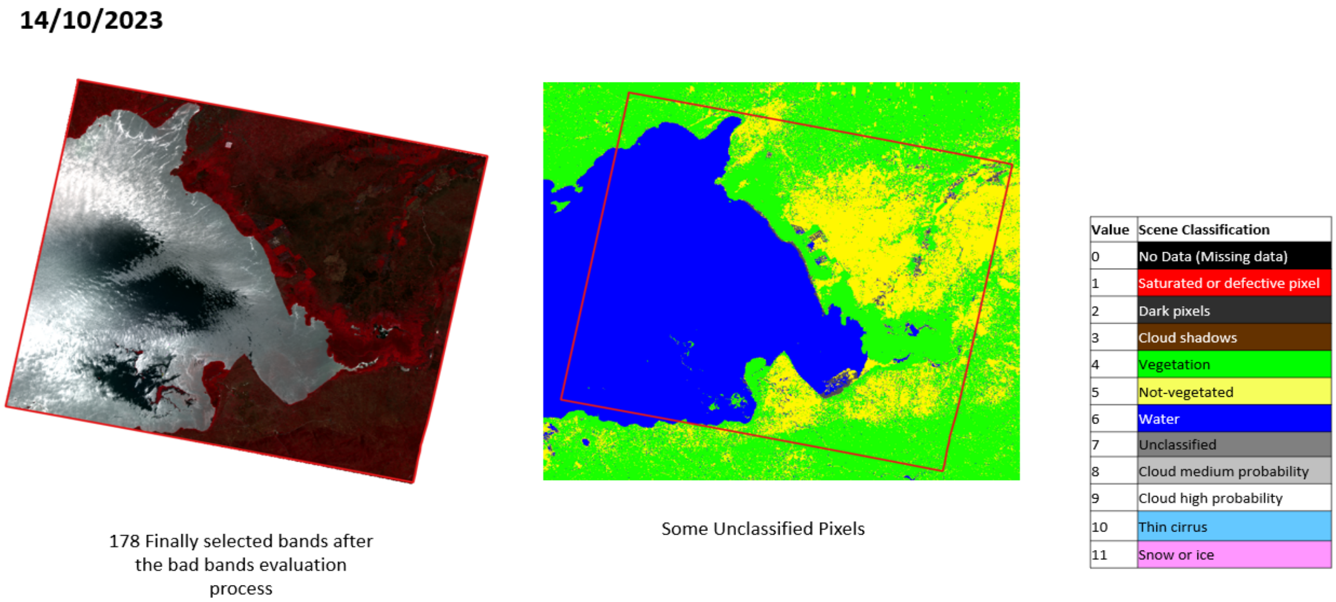


180 Finally selected bands after
the bad bands evaluation process

Some Dark Pixels & Shadows

Value	Scene Classification
0	No Data (Missing data)
1	Saturated or defective pixel
2	Dark pixels
3	Cloud shadows
4	Vegetation
5	Not-vegetated
6	Water
7	Unclassified
8	Cloud medium probability
9	Cloud high probability
10	Thin cirrus
11	Snow or ice

Figure 22. PRISMA image on 22/08/2023 and Sentinel-2 SCL image on 24/08/2023.



14/10/2023

178 Finally selected bands after the bad bands evaluation process

Some Unclassified Pixels

Value	Scene Classification
0	No Data (Missing data)
1	Saturated or defective pixel
2	Dark pixels
3	Cloud shadows
4	Vegetation
5	Not-vegetated
6	Water
7	Unclassified
8	Cloud medium probability
9	Cloud high probability
10	Thin cirrus
11	Snow or ice

Figure 23. PRISMA image on 14/10/2023 and Sentinel-2 SCL image on 13/10/2023.

The image pair on the 26/05/2020 was chosen to proceed with. That was due to the coincidence of both PRISMA and Sentinel-2 good quality images on that specific date.

Issues and Next Steps:

Upon assessing our activities, some issues have emerged:

- The lack of an API for automatically obtaining PRISMA images poses a significant challenge, especially within the context of developing automated services. The EO processing pipelines concerning the hyperspectral/ multispectral data fusion could become interrupted or inefficient as a result of this absence.
- During the PRISMA images evaluation in terms of cloud coverage, a discrepancy between the information provided by PRISMA metadata and the visual inspection results was observed. It was concluded that the PRISMA metadata are not a reliable source of information for the evaluation of cloud coverage. Therefore, as a quality control layer is not provided, the PRISMA acquisitions evaluation was performed via visual inspection.

4.5. Image Preprocessing

4.5.1. Multispectral Data

Challenge:

For the realization of a HS-MS Fusion scheme, the use of a Sentinel 2 MS image as the high spatial resolution component was considered. As reported in deliverable D3 "State-of-the-art review report on the EO analysis methods and algorithms", these images cover the Visible, Red Edge, NIR and SWIR regions of the spectrum, with 11 channels, out of which, those related to the Blue, Green, Red and Near-Infrared regions are mapped at a spatial resolution of 10m, while the Red Edge and Short Wave Infrared regions with 3 and 2 channels respectively have a spatial resolution of 20m. A reasonable prerequisite for the fusion of S2 and PRISMA images was the specification of the spatial resolution of the final HS-MS Fusion products which meant that all S2 MSI layers should be processed to a 10m resolution.

Approach:

The full exploitation of the potential provided by multispectral data implied the downscaling to a spatial resolution of 10m. The existence of 10 m bands covering the same scene offers excellent opportunities for downscaling the 20 m bands to 10 m spatial resolution to provide more detailed spatial information. As a result, the need for a S2 bands fusion, by spatial downscaling of the Red Edge and SWIR channels to the specified pixel size of VIS-NIR, was therefore raised. At this point, the application of an image fusion algorithm to the 10 m and 20 m S2 channels, in order to obtain a 11 multiband MS image with the highest spatial resolution which would be used to bring the spatial resolution of the PRISMA image to 10 m, was the followed approach. The approach adopted was based on the operational feasibility of the fusion scheme, and the computational efficiency of the algorithm, since the design assumes its application to entire S2 MSI tiles, which imposes a high computational burden on a task that is merely a sub-process MS Fusion.

Methodology:

As to the design of the S2 data fusion problem, it involved four fine spatial resolution bands at distinct, narrow and localized parts of the spectrum, which is different to, and more complex than, the common pan-sharpening fusion problem which involves only one fine band that however covers a much wider spectral range. Following the work described by Wang et al. (2016) and Selva et al. (2015), two fusion schemes were proposed as alternatives for the panchromatic type of fusion. The synthesized and the selected band scheme. The synthesized band variation aims to compute a single fine band as a weighted average of the initial high resolution S2 bands, where the weight coefficients are inferred by a multilinear regression model between the considered 20m band and the, spatially degraded versions of the four 10m bands. On the other hand, the selected band scheme aims to use the mostly correlated 10m band as the panchromatic image alternative. Out of these two schemes, the latter one, i.e. the selected band fusion scheme was finally decided, as the least complex method out of the two. The rationale of this approach was the fact that the Red Edge and SWIR band stacks are located on either side of the NIR image spectral range, and thus they are expected to have a very high Pearson correlation.

For the main S2 image fusion part, a hyper-sharpening (HP) adaptation of the Modulation Transfer Function (MTF) – Gaussian Laplacian Pyramid (GLP), belonging to the MRA class of methods was chosen because of its computational efficiency and of the good fusion performance experienced on real

S2 data. In the MRA approach, spatial detail is injected by multiresolution decomposition of the fine band by using high pass filtering as proposed by Chavez et al. (1991).

The S2 algorithm evaluation was performed in a quantitative manner, following the principles of Wald's protocol, according to which, in image fusion, it is important that any sharpened image, once degraded to its original spatial resolution, should be as close as possible to the original image. The same principles were also considered and applied on the initialization of the selected band scheme, for computing the correlation matrix between 10m and 10m bands. The quantitative metrics, also computed on a degraded reference image were, the Peak Signal to Noise Ratio (PSNR), the Structural Similarity Metric (SSIM), and the Spectral Angle Mapper (SAM). PSNR is given by the ratio of gray level numbers in the image to the corresponding pixels in the source input image and combined output image. A higher value of PSNR denotes better fusion quality. SSIM is a spatial performance fusion metric to compare the pixel intensity local patterns in the source image and fused output image. If the SSIM value is close to 1, that means higher spatial similarity between the source image and fused image. Spectral Angle Mapper (SAM) is a cosine distance metric in the image feature space of the MS and Pan channels, that denotes the spectral radiometric distortion of the fused result. If the SAM value is zero, that means "no spectral distortion".

Experimental Results:

The data processing pipeline for the S2 fusion involved the following steps:

- Extraction of Sentinel 2 MSI bands and conversion to BOA reflectance values
- Degradation of VIS-NIR 10m bands to a spatial resolution of 20m. This step was performed in order to compute the Pearson correlation matrix between the MSI channels, in a pixel wise manner. 2 degradation methods were tested and were evaluated in a qualitative way. A weighted average resampling, as well as a variation of it, in which the 10m bands are initially smoothed by a Gaussian Low Pass filter, and then resampled to their average. The degradation process is controlled by the sensor's Point Spread Function (PSF) which in the latter variation, is simulated by a Gaussian kernel filter, of 12 pixels radius. The filter's Standard Deviation parameter was calculated for each HR band to be degraded, according to the following formula

  $$SD = R \cdot \sqrt{\frac{-2 \cdot ln(MTF)}{\pi^2}}$$

  where R the degradation scale factor (equal to 2) and MTF the Modulation Transfer Function of Sentinel 2 MSI, provided for each band in the table that follows (TTT23).

Table TTT23. Modulation Transfer Function of S2 MSI for the simulation of PSF through Gaussian filtering.

	B2	B3	B4	B8
	VIS			***NIR***
MTF	0.26	0.28	0.24	0.26

- Layer Stacking of the 20m bands
- Computation of the Bands Correlation. The produced Pearson Correlation Matrix for the 26/05/2020 S2 MSI image is given below (Table 17).

Table 17. Pearson correlation matrix of S2 MSI layer stack. VIS-Red Edge-NIR-SWIR bands.

r	B2	B3	B4	B5	B6	B7	B8	B8A	B11	B12
	VIS			*Red Edge*			*NIR*		*SWIR*	
B2	1.000	0.996	0.982	0.984	0.907	0.890	0.889	0.881	0.920	0.923
B3	0.996	1.000	0.985	0.991	0.918	0.903	0.901	0.893	0.925	0.924
B4	0.982	0.985	1.000	0.984	0.866	0.846	0.845	0.836	0.915	0.938
B5	*0.984*	*0.991*	*0.984*	1.000	0.938	0.923	*0.921*	*0.915*	0.957	0.956
B6	*0.907*	*0.918*	*0.866*	0.938	1.000	0.998	*0.998*	*0.997*	0.947	0.900
B7	*0.890*	*0.903*	*0.846*	0.923	0.998	1.000	*0.999*	*0.999*	0.937	0.885
B8	0.889	0.901	0.845	0.921	0.998	0.999	1.000	0.999	0.940	0.887
B8A	0.881	0.893	0.836	0.915	0.997	0.999	0.999	1.000	0.939	0.884
B11	*0.920*	*0.925*	*0.915*	0.957	0.947	0.937	*0.940*	0.939	1.000	0.986
B12	*0.923*	*0.924*	*0.938*	0.956	0.900	0.885	*0.887*	0.884	0.986	1.000

The correlation results indicated the following LR-HR pairs

- B5 (Red Edge 1): B4 (Red)
- B6 (Red Edge 2): B8 (NIR broadband)
- B7 (Red Edge 3): B8 (NIR broadband)
- B11 (SWIR1): B8 (NIR broadband)
- B12 (SWIR2): B4 (Red)

- Data Normalization. Each image in the LR-HR pairs was normalized with the use of the Linear Transformation method of data scaling from integer (uint16) to float (float32). From image metadata, for all S2 bands:

```
<BOA_QUANTIFICATION_VALUE unit="none">10000</BOA_QUANTIFICATION_VALUE>
<BOA_ADD_OFFSET band_id="1:12">-1000</BOA_ADD_OFFSET>
```

thus, going from DN (uint16) to BOA (float32) the following transformation was applied:

$$BOA = \frac{DN + OFFSET}{QUANTIFICATIONVALUE}$$

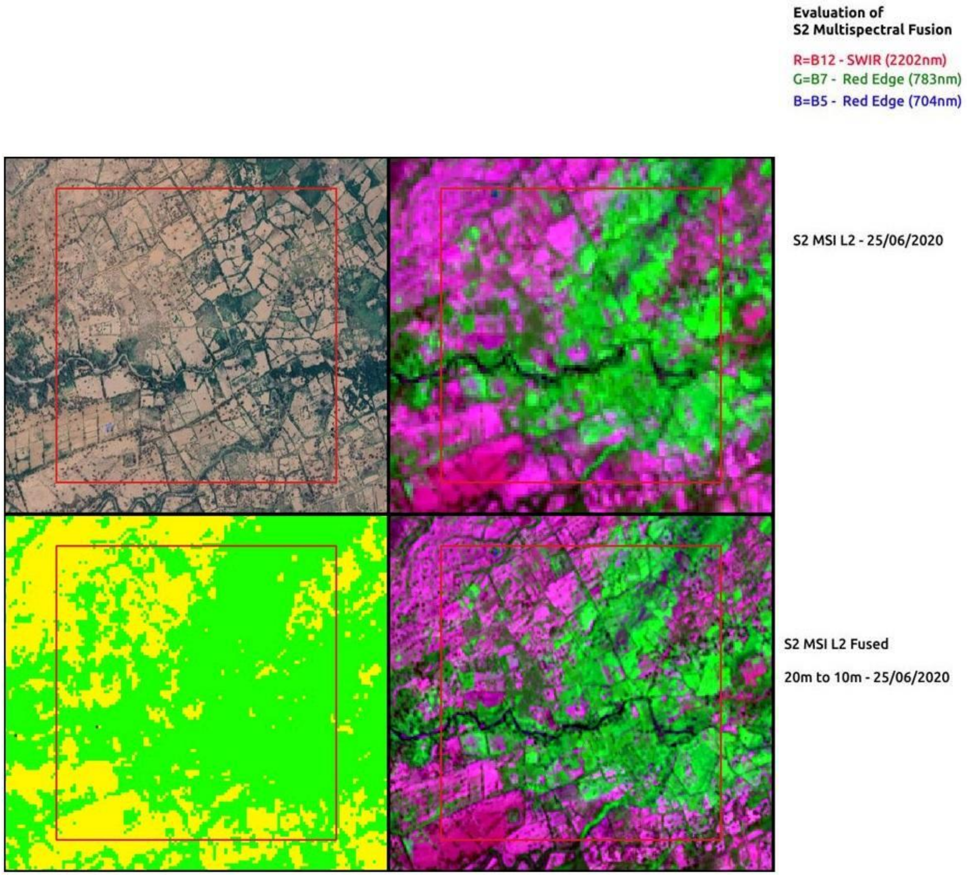
- Reference image creation for the application the S2 Fusion scheme and the subsequent quantitative evaluation. HR-LR image pairs were initially smoothed with Gaussian Kernel convolution filters and then degraded by a factor of 2.
- Set the parameters for High Pass 2D Convolution filtering. For the HP adaptation of the MTF-GLP algorithm, a kaiser kernel window was used, with default values (beta=0.5).
- Application of the HP – MTF – GLP fusion algorithm to the degraded image pairs for the calculation of the quantitative metrics
- Application of the HP – MTF – GLP fusion algorithm to the original HR-LR image pairs for the downscaling of LR to 10m spatial resolution.

The summary of the quantitative evaluation of the S2 fusion result is presented on the following table (Table 18):

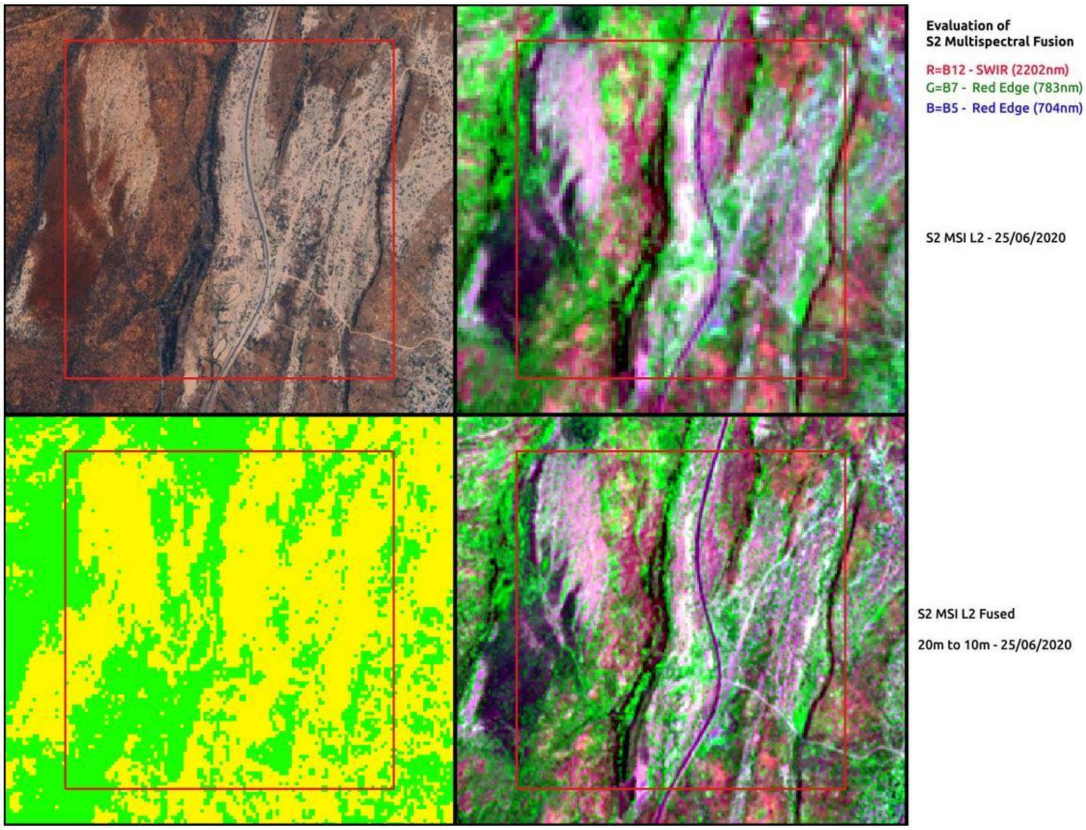
Table 18. HP- MTF-GLP algorithm quantitative evaluation metrics.

HP-MTF-GLP - Evaluation Metrics			
	PSNR	***SSIM***	***SAM***
B5	47.48	0.987	0.37
B6	43.82	0.975	0.39
B7	42.15	0.966	0.39
B11	38.25	0.922	0.39
B12	45.54	0.982	0.37

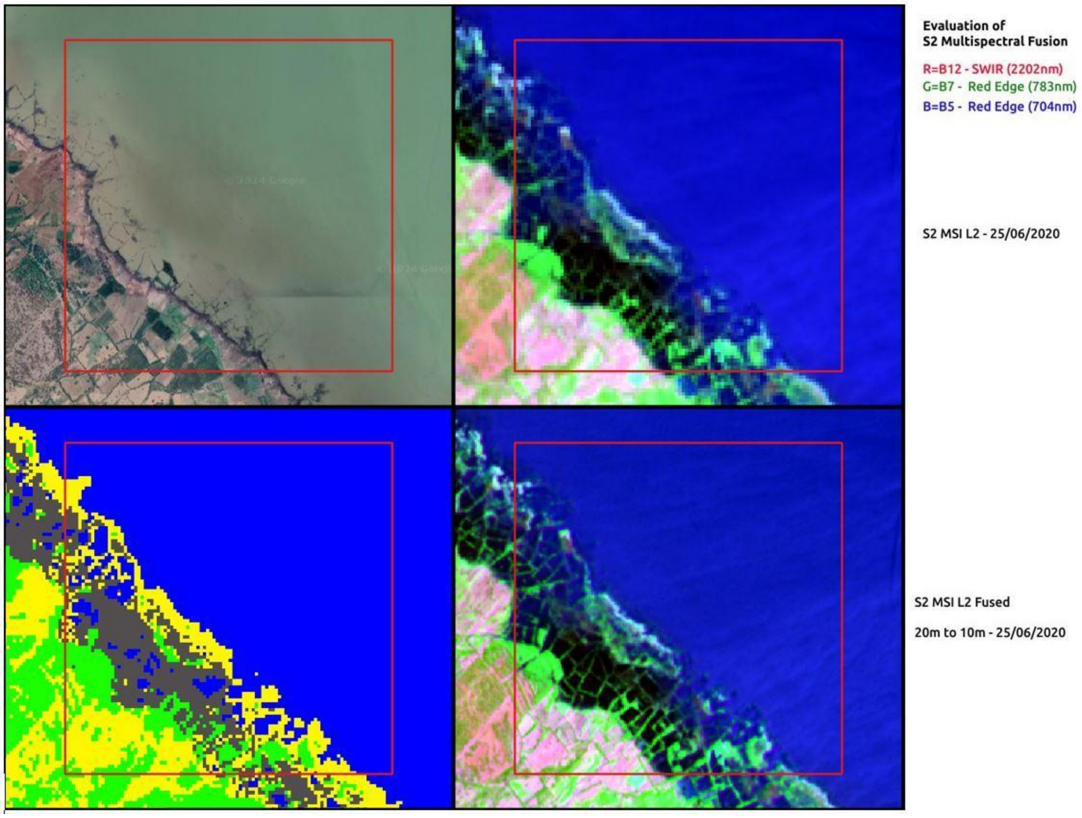
A complementary qualitative evaluation could be assessed by the observation and analysis of Reference and Fused images visually. Some illustrative examples of this assessment of the pilot experiment sub regions can be viewed on Figure 24 (a,b,c,d,e,f).


Evaluation of
S2 Multispectral Fusion
R=B12 - SWIR (2202nm)
G=B7 - Red Edge (783nm)
B=B5 - Red Edge (704nm)
S2 MSI L2 - 25/06/2020
S2 MSI L2 Fused
20m to 10m - 25/06/2020

(a)


Evaluation of
S2 Multispectral Fusion
R=B12 - SWIR (2202nm)
G=B7 - Red Edge (783nm)
B=B5 - Red Edge (704nm)
S2 MSI L2 - 25/06/2020
S2 MSI L2 Fused
20m to 10m - 25/06/2020

(b)


Evaluation of
S2 Multispectral Fusion
R=B12 - SWIR (2202nm)
G=B7 - Red Edge (783nm)
B=B5 - Red Edge (704nm)
S2 MSI L2 - 25/06/2020
S2 MSI L2 Fused
20m to 10m - 25/06/2020

(c)

(d)

Figure 24. Qualitative evaluation of the S2 fused result. Comparison of RGB=(B12,B7,B5) pseudocolor composite of SWIR- Red Edge bands, locally histogram stretched. Selected sites (a,b,c,d) presented, along with SCL mask and basemap.

A spectral signature sampling has been implemented to the Reference and Fused images, in a stratified manned dictated by the Sentinel 2 MSI Quality Scene Classification Layer SCL, for the qualitative assessment of spectral distortion, over the major land cover classes of Barren Land, Vegetation, and Water. The radiometric deviation between the two images (Figures 25-27), over the Red Edge and SWIR bands is a subjective but nevertheless clear indicator of the spectral alteration (over/under estimation) introduced by the S2 fusion.

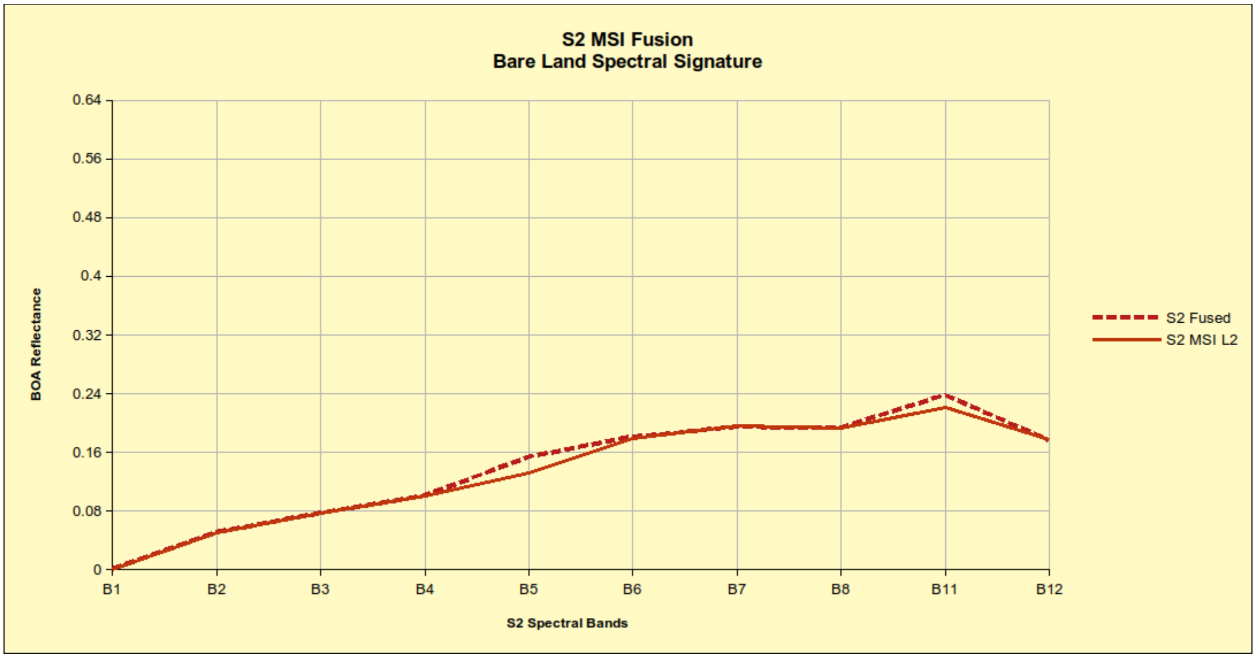


Figure 25. Averaged Spectral Signature of Barren Land for Fused and Reference S2 MSI image.

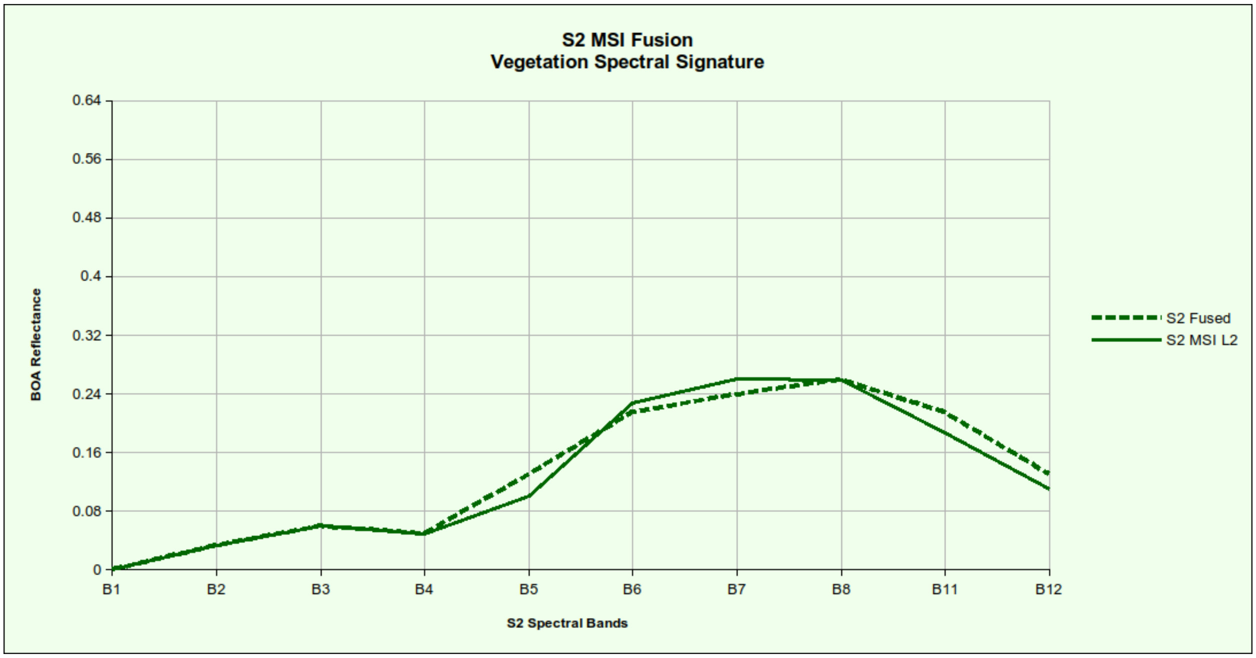


Figure 26. Averaged Spectral Signature of Vegetation for Fused and Reference S2 MSI image.

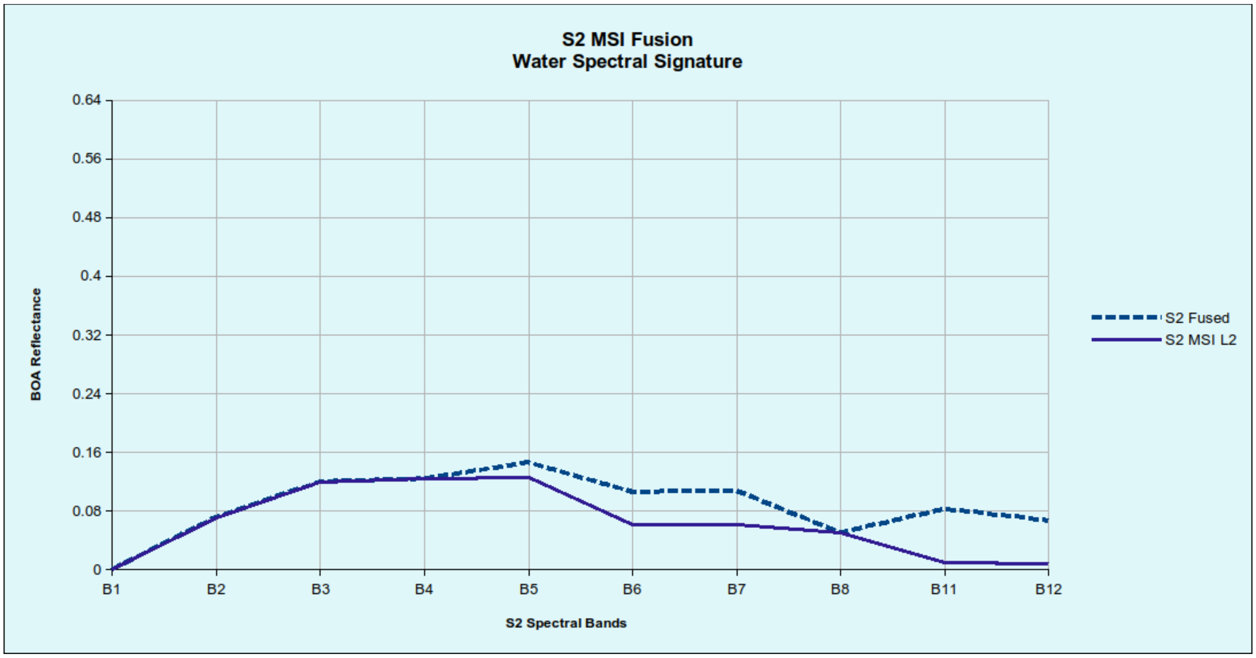


Figure 27. Averaged Spectral Signature of Water for Fused and Reference S2 MSI image.

The observed spectral disambiguation is expressed as an average relative error of radiometric BOA reflectance, on the following summary table. Relative errors expressed as % percentage are summarized for each Red Edge and SWIR band, over the selected SCL classes.

Table 19. Average Relative Error between Reference and Fused S2 Image spectral signatures, for the SCL classes of Barren Land, Vegetation, and Water.

	Average Relative Error of SCL spectral signatures (%)				
	B5 – Red Edge 1	**B6 – Red Edge 2**	**B7 – Red Edge 3**	**B11 – SWIR 1**	**B12 – SWIR 2**
Barren	16.56	0.79	-1.19	7.01	-0.82
Vegetation	28.35	-5.34	-8.01	15.26	18.21
Water	16.54	73.13	73.27	716.39	769.25

The overall evaluation of the procedure, a combination of qualitative and quantitative assessments, provided the following conclusions about the S2 Fusion preprocessing:

- The performance of the algorithm on the spatial aspects of the fused image was quite satisfying, something that was supported both by visual observation and the SSIM quantitative assessment.

The lowest value of SSIM was observed on the SWIR 1 band, which could partially be related to the relatively lower correlation value between the SWIR1 and NIR bands.

- On the spectral aspects of S2 fusion, the performance was suboptimal. Spectral distortion is an inevitable fact of image fusion, that is pronounced on the bands, whose spectral region is not covered by the HR image or it is least correlated with. That was also a fact during the evaluation of S2 Fusion with a selected band scheme, where high deviation was observed, especially for B5 (Red Edge) and the SWIR bands. The drawn conclusions are supported by a high SAM value, as well as by the evaluation of the relative spectral error, especially on both SWIR bands.

Issues and Next Steps:

During the S2 band fusion procedure, no significant issues were observed at the functional level as both fusion schemes were easily recreated on algorithm code. However, there were performance issues which need further improvement since the spectral distortions observed are expected to degrade the overall performance of HS-MS Fusion to some degree. For this reason, the following next steps are proposed in the S2 band fusion part:

- Application of the synthesized pan image fusion scheme. This approach, while involving more analysis and evaluation steps, like the assessment of a regression model for the production of a synthetic “pan” image that spans the whole spectral range of interest, could provide more valid spectral information to the fusion, and thus enhance the relevant metrics

4.5.1.1. S2 Data Fusion - Update iteration

In the period following the first version of the deliverable "D7 - Algorithm Theoretical Base Document and Product Specifications", further modifications and tests were carried out in an attempt to upgrade the performance and fidelity of the Sentinel 2 Band Fusion to 10 m spatial resolution results. As was evident from the preceding approaches presented so far, there was the issue of spectral distortions observed in the Red Edge and SWIR channels, with different strengths per land cover type. The reasons why it was considered important to try to improve the characteristics of the Fused image are the potential propagation of spectral errors to the overall HS-MS Fusion accuracy, but not only. The technique itself for Sentinel 2 Band Fusion with pansharpening methods is useful for extracting vegetation indices and Biophysical parameters, sensitive to the Red Edge or SWIR spectrum, in 10 m spatial resolution, and therefore an effort to improve it is of standalone value.

In terms of the methodological framework, there was no significant difference in terms of the algorithmic part. The approach of the HP-MTF-GLP algorithm for pansharpening of Multispectral images has comparative advantages over Deep Learning Fusion, because it is simpler to implement, does not require data for supervised learning, is faster and scalable. However, a different method was decided to create the "pseudo panchromatic" image band required to implement the algorithm. In order to sharpen a Sentinel-2A image, a specific multispectral band should be defined as an optimal panchromatic band with a 10 m spatial resolution.

During the preceding analysis a selected band scheme had been applied which selects a fine band from the VIS - NIR for each coarse band, which is determined as the one with the largest correlation with the visited coarse band. As a next step towards the enhancement of the Sentinel 2 Fusion, the application of the synthesized pan image fusion scheme was considered. The synthesized band scheme synthesizes a "pseudo panchromatic" image (i.e., fine multispectral image), as a weighted linear combination of all high resolution multispectral bands.This approach, while involving more analysis and evaluation steps, like the assessment of a multi - linear regression model for the production of a synthetic "pan" image that spans the whole spectral range of interest, could provide more valid spectral information to the fusion, and thus enhance the relevant metrics.

For the implementation of the synthesized band scheme, a stratified random sampling approach was applied on the Sentinel 2 image, fetching 10000 samples, or 1/1000 of the image pixels. The 10 m bands of VIS-NIR channels were degraded to a spatial resolution of 20m. The bands were initially smoothed by a Gaussian Low Pass filter, and then resampled 2 by 2 pixel average. A linear model for each band from the Red Edge and SWIR spectral regions was trained and cross validated with a 2 fold data partitions. The evaluation of the synthetic pan image models is presented for each 20m band , on the tables that follow (TTT24 to TTT33).

Table TTT24. Multilinear Regression Model for Sentinel 2 Red Edge 1 channel (B5)- Model Coefficients.

	Red Edge 1 Panchromatic regression model - Model Coefficients					
	Coefficient	**r**	**R^2**	**Adjusted R^2**	**Std Error**	**t -test**
B2 Blue	-0.678	-0.646	0.417	0.417	0.002	-262.648
B3 Green	1.019	0.617	0.381	0.381	0.004	243.356
B4 Red	0.663	0.830	0.689	0.689	0.0014	461.891
B8 NIR	0.094	0.638	0.407	0.407	0.0003	257.209

Table TTT25. Multilinear Regression Model for Sentinel 2 Red Edge 1 channel (B5)- Model Evaluation.

Red Edge 1 Panchromatic regression model - Model Evaluation									
	R^2	**Adjusted R^2**	**Std Error**	**SSR**	**SSE**	**SST**	**MSR**	**MSE**	**F - test**
ANOVA	0.994	0.994	0.007	869.161	5.007	874.168	217.290	0.00005	4170999.19
	CV MSE		**CV RMSE**		**CV NRMSE**		**CV R^2**		
Cross Validation	0.000052		0.00722		0.00429		0.994268		

Table TTT26. Multilinear Regression Model for Sentinel 2 Red Edge 2 channel (B6)- Model Coefficients.

	Red Edge 2 Panchromatic regression model - Model Coefficients					
	Coefficient	**r**	**R^2**	**Adjusted R^2**	**Std Error**	**t -test**
B2 Blue	-0.250	-0.240	0.057	0.057	0.003	-76.765
B3 Green	0.647	0.367	0.135	0.134	0.005	122.48
B4 Red	-0.126	-0.219	0.048	0.048	0.001	-69.7664
B8 NIR	0.751	0.982	0.964	0.964	0.0004	1621.687

Table TTT27. Multilinear Regression Model for Sentinel 2 Red Edge 2 channel (B6)- Model Evaluation.

Red Edge 2 Panchromatic regression model - Model Evaluation									
	R^2	**Adjusted R^2**	**Std Error**	**SSR**	**SSE**	**SST**	**MSR**	**MSE**	**F - test**
ANOVA	0.994	0.994	0.009	1489.62	7.96	1497.58	372.40	0.00003	4493966.77
	CV MSE		**CV RMSE**		**CV NRMSE**		**CV R^2**		
Cross Validation	0.00008		0.0091		0.005		0.994		

Table TTT28. Multilinear Regression Model for Sentinel 2 Red Edge 3 channel (B7)- Model Coefficients.

	Red Edge 3 Panchromatic regression model - Model Coefficients					
	Coefficient	**r**	**R^2**	**Adjusted R^2**	**Std Error**	**t -test**
B2 Blue	-0.086	-0.088	0.007	0.007	0.003	-27.54
B3 Green	0.247	0.155	0.024	0.024	0.005	48.68
B4 Red	-0.141	-0.254	0.064	0.064	0.001	-81.44
B8 NIR	0.965	0.989	0.979	0.979	0.0004	2170.13

Table TTT29. Multilinear Regression Model for Sentinel 2 Red Edge 3 channel (B7)- Model Evaluation.

Red Edge 3 Panchromatic regression model - Model Evaluation									
	R²	**Adjusted R²**	**Std Error**	**SSR**	**SSE**	**SST**	**MSR**	**MSE**	**F - test**
ANOVA	0.996	0.996	0.008	1948.28	7.342	1955.63	487.07	0.00007	6375629.29
	CV MSE		**CV RMSE**		**CV NRMSE**		**CV R²**		
Cross Validation	0.00007		0.0087		0.0053		0.9962		

Table TTT30. Multilinear Regression Model for Sentinel 2 SWIR 1 channel (B11)- Model Coefficients.

	SWIR 1 Panchromatic regression model - Model Coefficients					
	Coefficient	**r**	**R²**	**Adjusted R²**	**Std Error**	**t -test**
B2 Blue	-1.555	-0.429	0.184	0.184	0.010	-147.35
B3 Green	-0.011	-0.002	0.000	0	0.017	-0.69
B4 Red	2.150	0.763	0.582	0.582	0.005869	366.35
B8 NIR	0.332	0.581	0.337	0.337	0.0015	221.44

Table TTT31. Multilinear Regression Model for Sentinel 2 SWIR 1 channel (B11)- Model Evaluation.

SWIR 1 Panchromatic regression model - Model Evaluation									
	R²	**Adjusted R²**	**Std Error**	**SSR**	**SSE**	**SST**	**MSR**	**MSE**	**F - test**
ANOVA	0.954	0.954	0.029	1751.33	83.570	1834.902	437.832	0.0008	503556.6
	CV MSE		**CV RMSE**		**CV NRMSE**		**CV R²**		
Cross Validation	0.00087		0.029488		0.020499		0.954451		

Table TTT32. Multilinear Regression Model for Sentinel 2 SWIR 2 channel (B12)- Model Coefficients.

	SWIR 2 Panchromatic regression model - Model Coefficients					
	Coefficient	**r**	**R^2**	**Adjusted R^2**	**Std Error**	**t -test**
B2 Blue	-1.088	-0.350	0.123	0.123	0.009	-116.12
B3 Green	0.181	0.03	0.001	0.001	0.01	11.93
B4 Red	1.606	0.705	0.497	0.497	0.005	308.48
B8 NIR	0.081	0.192	0.037	0.037	0.001	60.91

Table TTT33. Multilinear Regression Model for Sentinel 2 SWIR 2 channel (B12)- Model Evaluation.

SWIR 2 Panchromatic regression model - Model Evaluation									
	R^2	**Adjusted R^2**	**Std Error**	**SSR**	**SSE**	**SST**	**MSR**	**MSE**	**F - test**
ANOVA	0.931	0.931	0.026	901.58	65.811	967.393	225.395	0.0006	329183.2
	CV MSE		**CV RMSE**		**CV NRMSE**		**CV R^2**		
Cross Validation	0.0006		0.026		0.021		0.931		

The model evaluation in the individual channels shows an increase in the value of the coefficient of determination compared to the simpler selected band scheme approach, which is even significant in the SWIR channels. HP – MTF – GLP fusion algorithm was applied with the synthetic pan bands created by the described models. The summary of the quantitative evaluation of the S2 fusion result is presented on the following table (Table TTT34):

Table TTT34. HP- MTF-GLP algorithm quantitative evaluation metrics for synthetic pan band scheme.

HP-MTF-GLP - Evaluation Metrics			
	PSNR	***SSIM***	***SAM***
B5	47.9548	0.99687	0.363
B6	43.82	0.975	0.388
B7	42.15	0.966	0.381
B11	42.075	0.9681	0.38

B12	49.1832	0.98691	0.364

A slight improvement has been observed to the SWIR bands fusion results, which were also the most problematic ones.

A spectral signature sampling has been implemented to the Reference and Fused images, in a stratified manned dictated by the Sentinel 2 MSI Quality Scene Classification Layer SCL, for the qualitative assessment of spectral distortion, over the major land cover classes of Barren Land, Vegetation, and Water. The radiometric deviation between the two images (Figures XXX22 - XXX24), over the Red Edge and SWIR bands is a subjective but nevertheless clear indicator of the spectral alteration (over/under estimation) introduced by the S2 fusion.

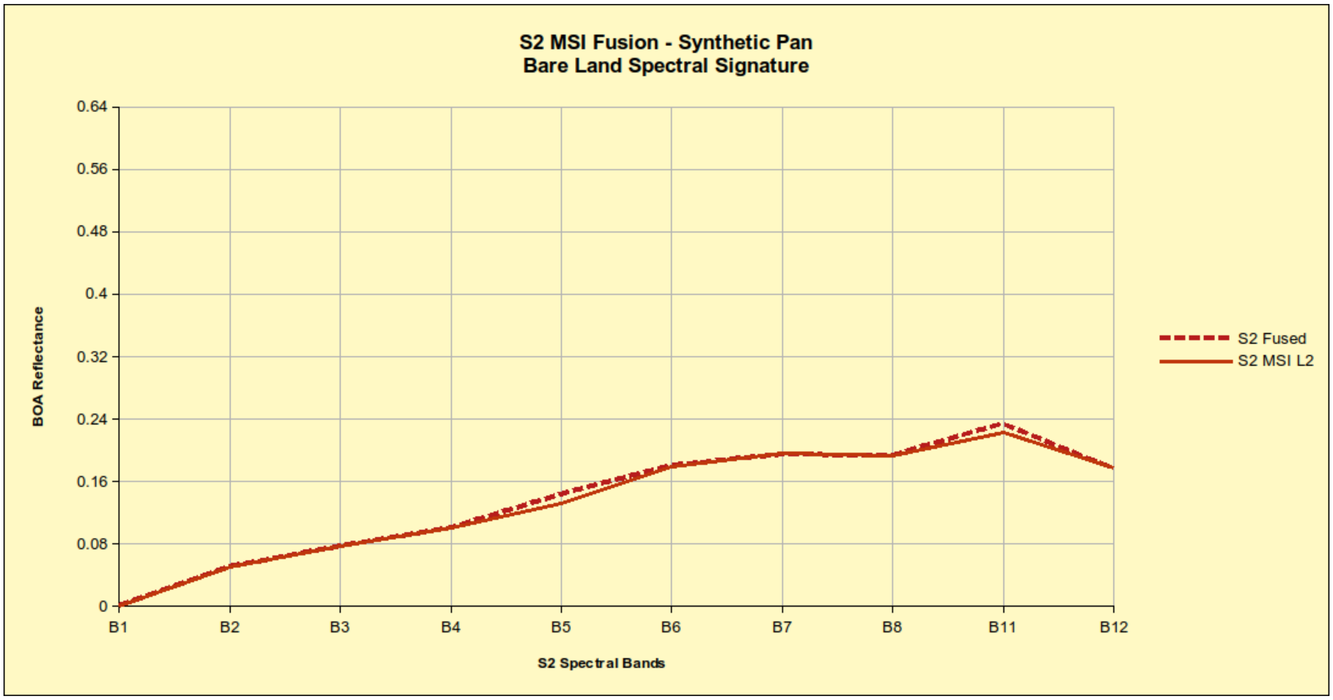


Figure XXX22. Averaged Spectral Signature of Barren Land for Fused and Reference S2 MSI image.

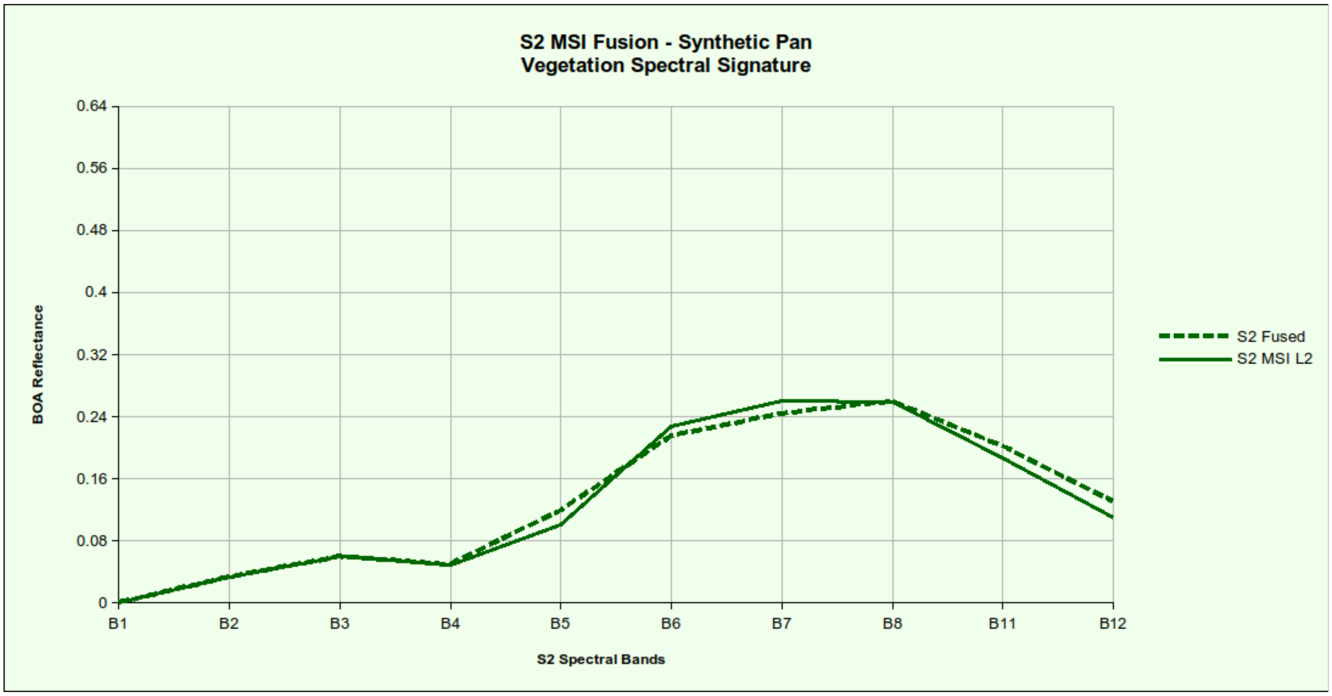


Figure XXX23. Averaged Spectral Signature of Vegetation for Fused and Reference S2 MSI image.

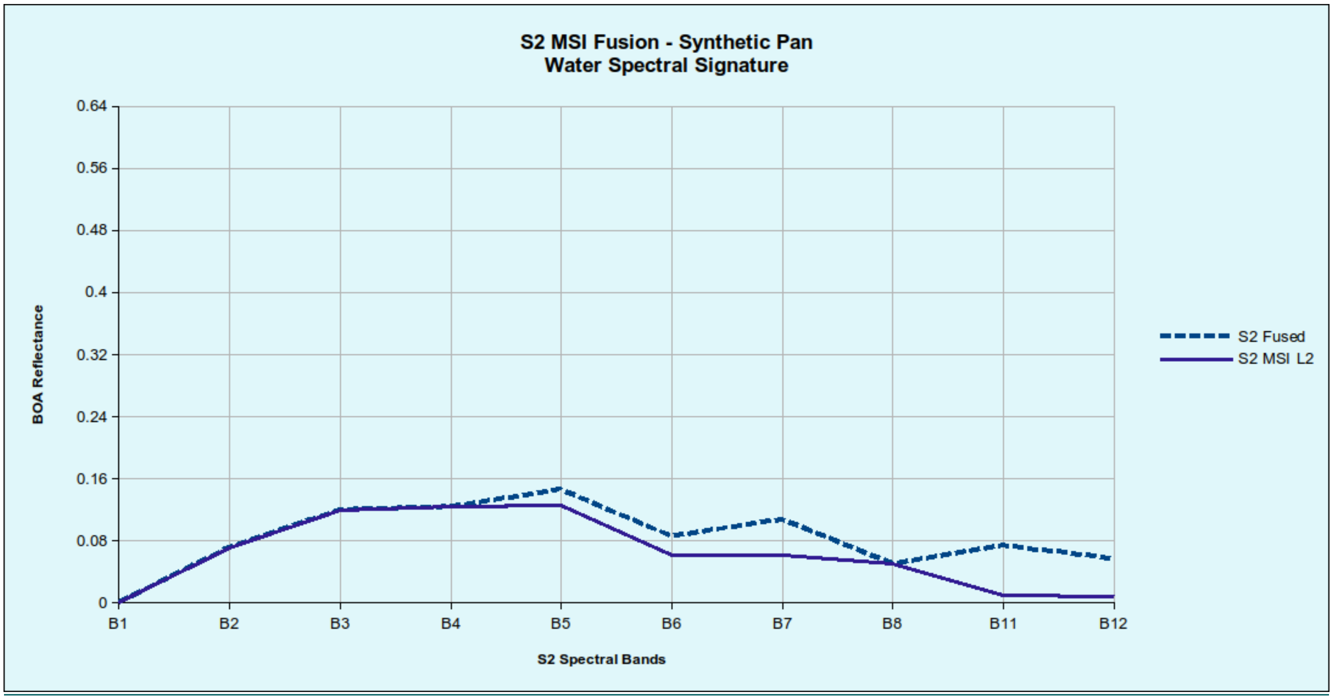


Figure XXX24. Averaged Spectral Signature of Water for Fused and Reference S2 MSI image.

The observed spectral disambiguation, is expressed as an average relative error of radiometric BOA reflectance, on the following summary table. Relative errors expressed as % percentage are summarized for each Red Edge and SWIR band, over the selected SCL classes.

Table TTT35. Average Relative Error between Reference and Fused S2 Image spectral signatures (synthetic pan band scheme), for the SCL classes of Barren Land, Vegetation, and Water.

	Average Relative Error of SCL spectral signatures (%)				
	B5 – Red Edge 1	**B6 – Red Edge 2**	**B7 – Red Edge 3**	**B11 – SWIR 1**	**B12 – SWIR 2**
Barren	16.40	0.79	-1.19	5.96	-0.81
Vegetation	28.06	-5.34	-8.01	12.97	16.39
Water	16.38	73.13	73.27	644.75	692.32

From the summary evaluation of the S2 image fusion results using the HP - MTF - GLP fusion algorithm with a synthetic pan bands scheme it was concluded that the quality metrics were improved, with emphasis on the problematic bands SWIR 1 , SWIR 2, which was also evident when comparing the spectral signatures in the 3 main land cover categories. To use the fine spatial resolution from the four 10 m bands, two schemes, the synthesized band scheme and the selected band scheme, were considered to obtain a single fine resolution band for matching with each 20 m multispectral image or 20 m band. In the current iteration, analysis of the results demonstrated that the synthetic panchromatic band scheme provided enhanced results compared to the more simple approach of the selected band scheme.

4.5.2. Hyperspectral Data

Challenge:

One of the most significant challenges in the PRISMA data preprocessing process and the preparation of the hyperspectral cube for fusion with the Sentinel-2 data is locating, identifying, evaluating and handling the bad bands within the PRISMA hyperspectral cube. Bad bands can reduce the quality of hyperspectral data and they are characterized by spectral inconsistencies such as sensor noise, atmospheric interference or an insufficient signal-to-noise ratio.

Approach:

The evaluation of bad bands in PRISMA imagery encompasses a process that aims to improve the hyperspectral data quality for further analysis and use in the data fusion pipeline. This bad bands evaluation process involves several steps, one of which is the elimination of overlapping bands between the visible near-infrared (VNIR) and shortwave infrared (SWIR) data cubes (Table 15.) Targeted elimination of water absorption bands is also necessary in order to minimize the atmospheric interference distortions and maximize data clarity. Moreover, image error matrix and error matrix statistics are used, which allows for the accurate identification, evaluation and removal of the noisy bands from the hyperspectral cube.

Methodology:

As the HS/MS images pair on the 26/05/2020 is chosen to proceed with (Chapter 4.4.), the PRISMA image on 26/05/2020 (PRS_L2D_STD_20200526080833_20200526080837_0001) underwent the preprocessing and bad bands evaluation process.

Tools

PRISMA data were imported, explored and processed manually in the EnMAP-Box plugin of QGIS. EnMap-Box is able to import the VNIR and SWIR cubes as well as the PAN cube as raster layers in Float32 data type. VNIR and SWIR error matrices can also be imported in 8-bit data type.

VNIR/SWIR overlapping bands elimination

As already mentioned, the VNIR and SWIR data cubes composing the PRISMA image present a band overlap at around 943.35 to 979.22 nm (Table 15.). It is observed that SWIR cube bands cover a wider spectral area as the bands present a more even distribution concerning their wavelengths. Thus, at this step, the 4 VNIR cube bands in the overlapping area (i.e, 944.62 nm, 956.27 nm, 967.02 nm and 977.36 nm) were eliminated from the hyperspectral cube.

Atmospheric water absorption bands elimination

The atmospheric water absorption bands are located in the wavelength ranges of 1340 - 1440 nm, 1780 - 1970 nm and 2300 - 2500 nm (Figure 28.) and they are apparent at an image's spectral profile. Other peaks can be found at different rates in the wavelengths around 920 and 1120 nm (Minor absorption zones). These spectrum regions constitute a part of the bands characterized as bad bands, which were removed from the PRISMA hyperspectral cube as well.

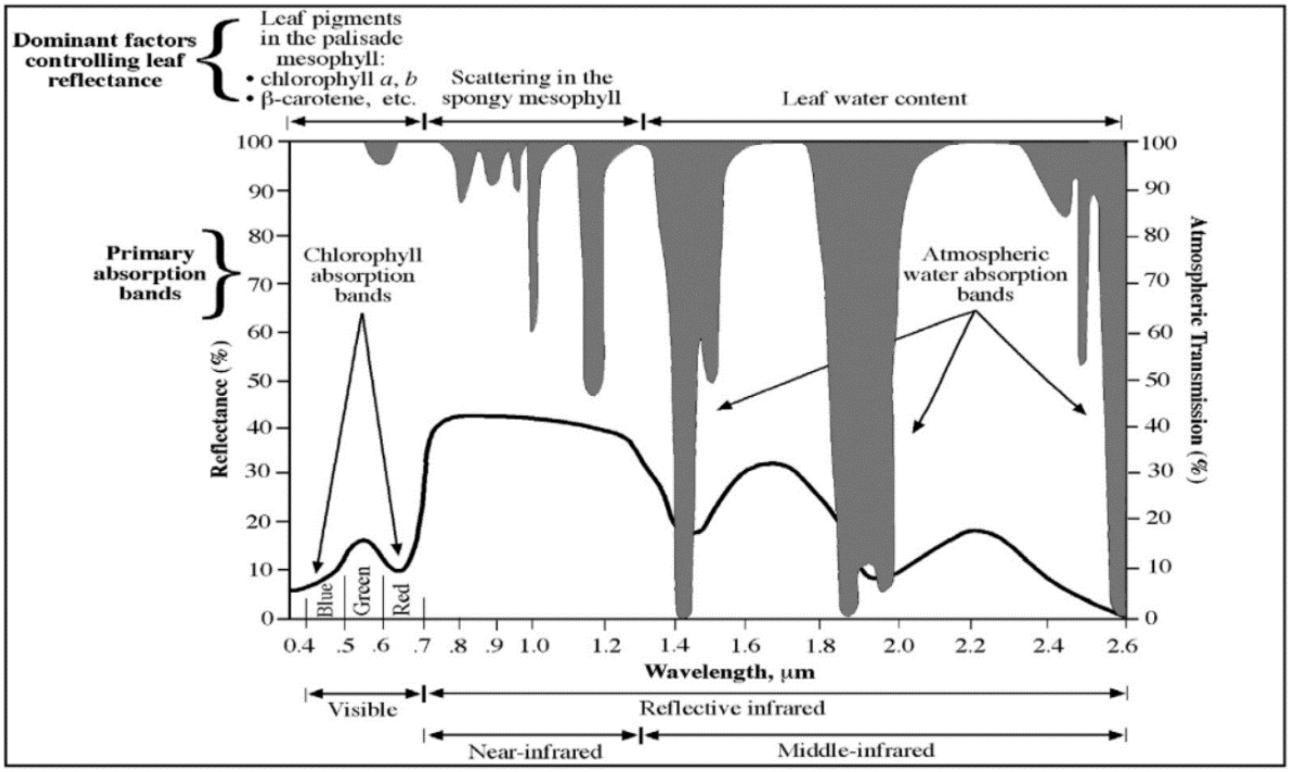


Figure 28. Atmospheric water absorption bands (Mohamed et al., 2018).

Bad (noisy) bands evaluation and elimination

The VNIR/ SWIR overlapping bands and the atmospheric water absorption bands elimination procedures were followed by the image noise removal, which constituted the last step in the PRISMA data preprocessing and a necessary procedure as well. Within this context, the VNIR and SWIR data cubes were processed for bad/noisy bands identification and extraction, where a bad/noisy band was a band that presented reflectance values close to zero or anomalies along the distribution and therefore it was not desirable for subsequent processing of the hyperspectral image in the hyperspectral/ multispectral fusion. The identification and evaluation of noisy bands was carried out through the error matrix extraction and its statistics, specifically the error matrix maximum values, as well as through the simultaneous examination of the spectral plots in order to gain information about the image noise through visual inspection.

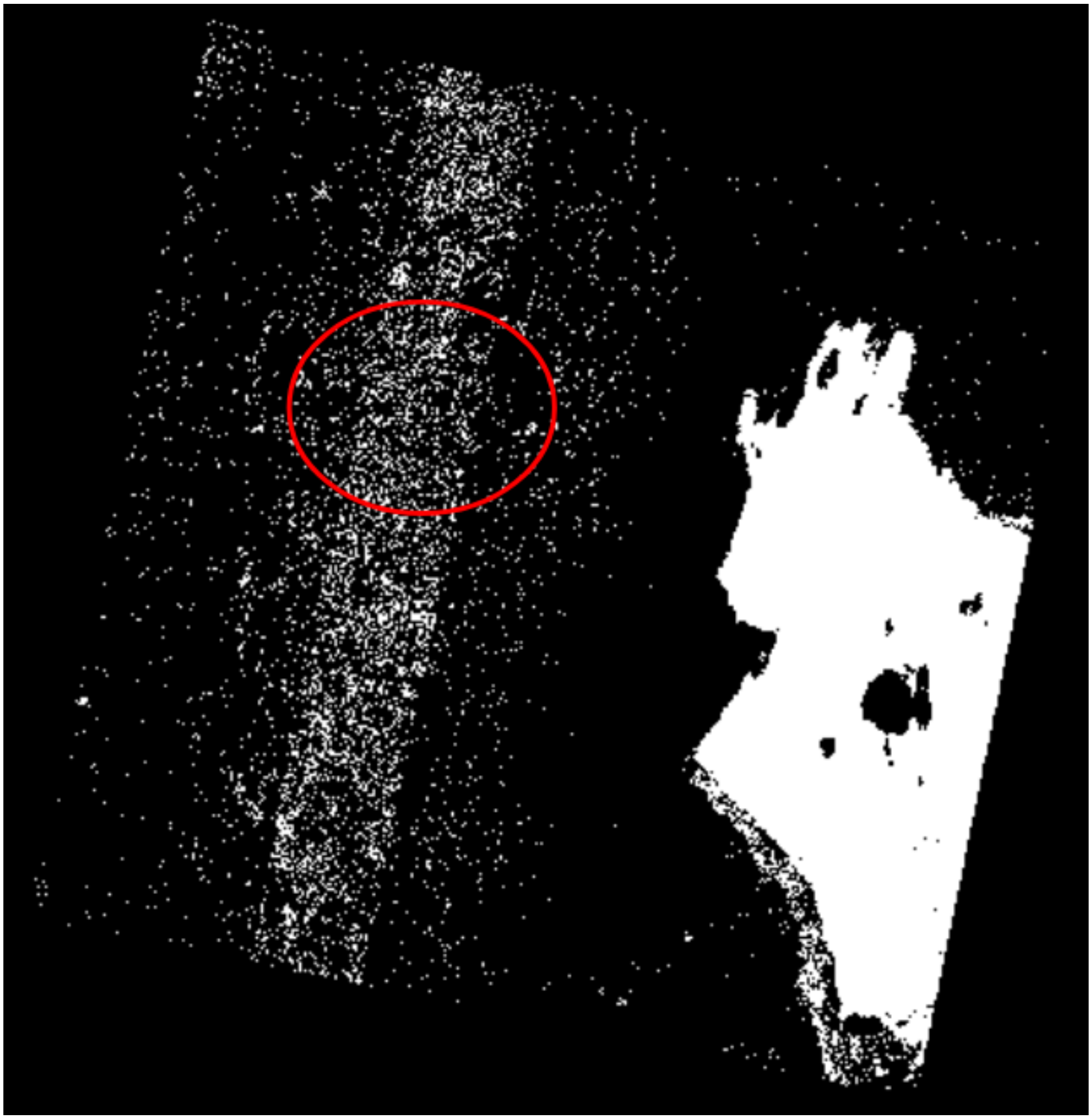


Figure 29. Error matrix maximum values, in which a typical "salt and pepper" effect indicating the noisy bands is present.

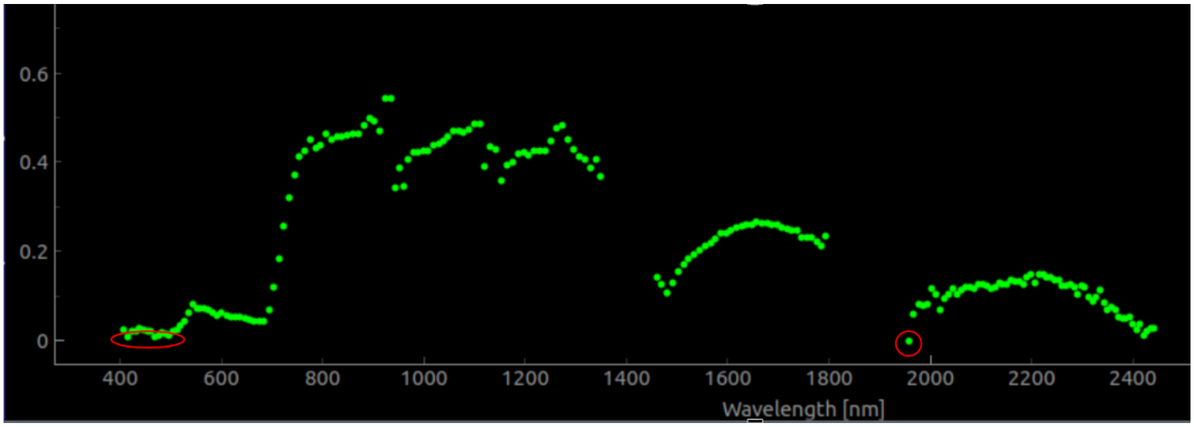


Figure 30. Spectral plot of a pixel in the hyperspectral cube after the overlapping and the water absorption bands removal, but before the noisy bands removal. It is observed that several bands have reflectance values at zero or close to zero (bad/noisy bands).

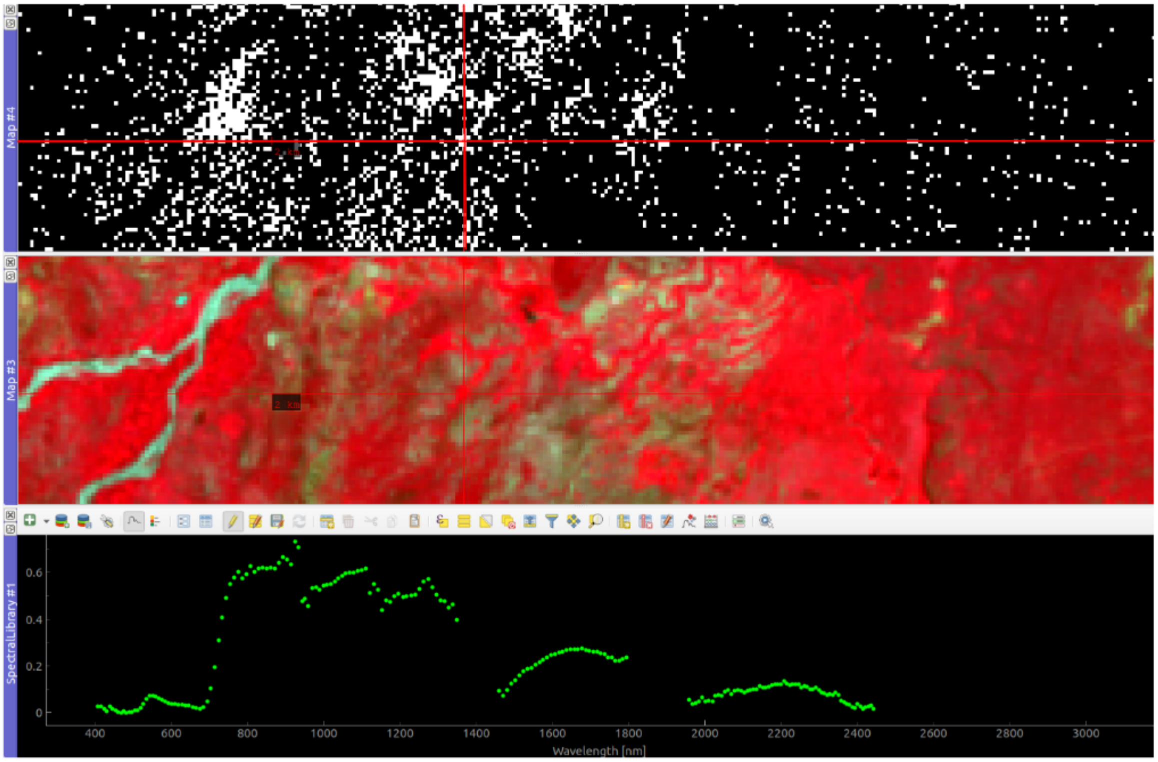


Figure 31. From top to bottom: Error matrix maximum values, hyperspectral cube after the removal of overlapping and water absorption bands, spectral plot of a pixel with noisy bands in the area of the typical “salt and pepper” effect.

Experimental Results:

The selected PRISMA Image on 26/05/2020 (PRS_L2D_STD_20200526080833_20200526080837_0001) underwent the preprocessing and bad band evaluation process, after which 172 spectral bands were finally retained and saved as a new hyperspectral cube in a GeoTIFF format image.

The metadata of the new hyperspectral cube are:

Data Type = float32
CRS = WGS 84 / UTM zone 36N
Units: BOA Reflectance units, 0-1
No-data value: -99999
File type = TIFF
Samples = 1191
Lines = 1184
Bands = 172 (Remained after the bad band evaluation process - data cleaning)
Wavelength units = Nanometer
Wavelength = {453.38947, 460.73175, 468.09842, 475.31885, 482.54816, 489.79486, 497.05865, 504.51172, 512.04639, 519.54376, 527.3053, 535.05255, 542.88513, 550.91461, 559.02026, 567.20612, 575.48682, 583.84412, 592.33899, 601.0144, 609.95819, 618.71997, 627.77844, 636.67627, 645.96381, 655.41876, 664.8941, 674.46436, 684.13727, 694.12836, 703.737, 713.72687, 723.87994, 733.9552, 744.14954, 754.4696, 764.85645, 775.2735, 785.65955, 796.12701, 806.71106, 817.31104, 827.91949, 838.52722, 849.20996, 859.97314, 870.74255, 881.45605, 892.08093, 902.80164, 913.44507, 923.9502, 934.11206, 943.35791, 951.40143, 959.974, 969.84491, 979.224, 988.91791, 998.9082, 1008.6443, 1018.5357, 1029.344, 1037.9878, 1047.675, 1057.5737, 1067.7948, 1078.2161, 1088.761, 1099.2776, 1109.8894, 1120.6759, 1131.3048, 1142.0703, 1152.6501, 1163.676, 1174.7142, 1185.5884, 1196.3394, 1207.2737, 1217.8635, 1229.1852, 1240.2145, 1250.9799, 1262.5322, 1273.4963, 1284.4878, 1295.4218, 1306.218, 1317.2566, 1328.2993, 1339.1294, 1459.3157, 1469.9308, 1480.8422, 1491.4292, 1502.0236, 1512.6333, 1523.2222, 1533.7764, 1544.2262, 1554.8168, 1565.3688, 1575.6274, 1585.8597, 1596.2454, 1606.4913, 1616.8336, 1627.021, 1637.0919, 1647.2316, 1656.933, 1667.1851, 1677.3193, 1687.4269, 1697.2943, 1707.0945, 1716.8589, 1726.6516, 1736.4883, 1746.2192, 1755.833, 1765.5127, 1775.1178, 2027.7267, 2036.2607, 2044.6809, 2053.0078, 2061.3787, 2069.7957, 2077.9915, 2086.3823, 2094.6252, 2102.8213, 2111.0391, 2119.2314, 2127.3372, 2135.5103, 2143.4656, 2151.3862, 2159.564, 2167.4849, 2175.3442, 2183.4202, 2191.1003, 2199.1353, 2206.843, 2214.625, 2222.4263, 2230.0076, 2237.9041, 2245.4485, 2253.1104, 2260.8665, 2268.2883, 2276.0537, 2283.4934, 2290.8267, 2298.6094, 2305.7227, 2313.2007, 2320.8955, 2327.8242, 2335.5264, 2342.8228, 2349.7915, 2357.2937, 2364.5945, 2371.5522, 2378.771, 2386.0618, 2393.0388}

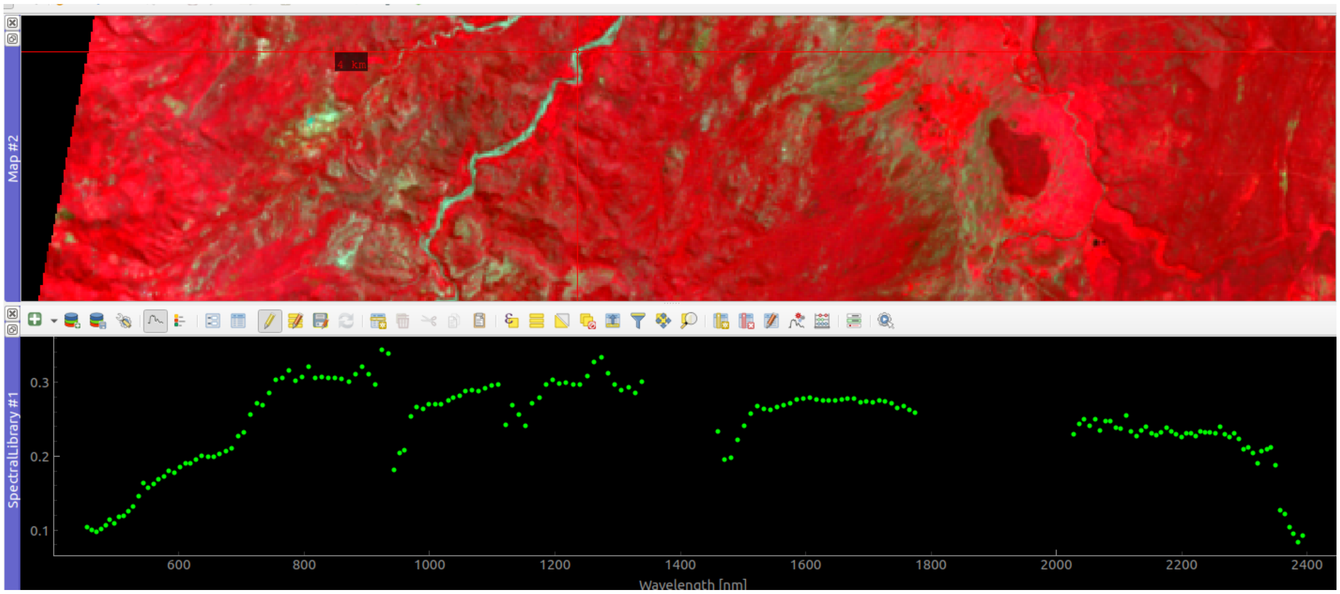


Figure 32. Spectral plot of a pixel in the hyperspectral cube after the removal of overlapping, water absorption and noisy bands. It is noticed that the bands at zero or close to zero have been eliminated.

Issues and Next Steps:

Issues

During the preprocessing and bad bands evaluation and elimination process, the following issues were observed:

- The conversion/ scaling factors are not consistent among the PRISMA acquisitions. This fact raised a concern regarding the implementation of a fully automated methodology approach, as each PRISMA acquisition should be scaled by utilizing its own unique scaling factors, which have to be retrieved in advance.

- In some cases, the conversion/ scaling factors are different even within the same image, as each one is the union of the VNIR and SWIR data cubes, which may have different scaling factors. This was observed in the case of the acquisition on 26/05/2020, in which the scaling factors between the VNIR and the SWIR data cubes were different. In particular, they were:

  L2ScaleSwirMax=0.686
  L2ScaleSwirMin=0

  L2ScaleVnirMax=0.606
  L2ScaleVnirMin=0

- Thus, in order to perform the data scaling on such a hyperspectral cube, the PRISMA data scaling equation has to be applied separately in the VNIR and the SWIR data cubes.

- A different number of spectral bands remains on each different hyperspectral cube after the bad bands evaluation and elimination process. This fact has an implication on the development of a fully automated approach as each time a distinct number of bands composing the processed hyperspectral image serves as input in the hyperspectral/ multispectral data fusion algorithm.

- Regarding the cloud coverage evaluation of each PRISMA acquisition, it was found that the cloud coverage information provided by PRISMA images metadata is not indicative of the images' actual cloud coverage. Since a quality control layer is not available, the evaluation of acquisitions had to be performed via visual inspection, a fact that poses an implication on a fully automated approach.

4.5.3. Hyperspectral – Multispectral Image Co-registration

Challenge:

PRISMA hyperspectral (HS) and Sentinel-2 MSI multispectral (MS) imagery compose the input data of AFRI4Cast's hyperspectral/ multispectral data fusion component targeting on the spatial enhancement of PRISMA hyperspectral bands. However, these datasets present inherent differences in spatial and spectral resolutions. The accurate detection and correction of geospatial misalignments between these datasets presents a challenge and it was also the case in the current HS and MS selected images. This image misalignment constitutes an obstacle and significantly complicates the image fusion process.

Approach:

The geospatial misalignment of PRISMA/ Sentinel-2 images was solved through an image co-registration process. Image co-registration is a technique for minimizing the shifts between two different images at a sub-pixel level by fixing the displacement of an image with respect to a reference image. This technique was the chosen approach as the HS/MS images had to be aligned in order to serve as input in the image fusion process and avoid critical errors.

Methodology:

Within AFRI4Cast, in order to proceed with the HS/ MS image co-registration, AROSICS was utilized. AROSICS is an automated and robust open-source image co-registration software for multi-sensor satellite data, offering two interfaces, i.e., a Python API and a command line interface. It detects and corrects local and global misalignments between two input images at a sub-pixel scale.

Generally, there are two groups of image registration methods: intensity based and feature based. In this context, the primary technology utilized by AROSICS to identify sub-pixel displacements is phase correlation, an intensity-based registration method. In this manner, the independence from discrete image features that may be unevenly distributed or absent altogether, especially regarding the dynamic surface coverages that are present in multi-temporal satellite data, is maintained (Scheffler et. al, 2017).

Phase correlation allows for the fast computation of relative translational offsets between two images in the same dimension by relying on the derivation of image displacements in the frequency (Fourier) domain. Phase correlation offers far higher accuracy than other intensity-based methods, such classical cross-correlation, due to a distinct peak in the cross-power spectrum that indicates the registration point. It produces excellent co-registration results even when there are significant variations in ground cover across the different images and low signal-to-noise ratios (Scheffler et. al, 2017). To sum up, AROSICS is based on a phase correlation approach and makes use of Fourier shift theorem, enabling the determination of precise X/Y offsets at a given geographical position.

According to Scheffler et. al (2017), the AROSICS workflow consists of three main processing steps:

- Input data preparation, that includes: (1) automatic selection of the spectral bands from target and reference image to be used for co-registration; (2) no-data mask generation and combination with user-provided masks (e.g., containing clouds and cloud shadows) to separate bad-data masks for reference and target image; (3) calculation of the respective footprint polygons and corresponding overlap area; (4) pixel coordinate grid equalization; and (5) adjustment of matching window positions and sizes.
- Detection of geometric shifts, either in a moving-window manner for each point of a dense grid (local co-registration approach) or for only a single coordinate position (global co-registration approach), including validation.
- Correction of displacements, by warping the target image under the use of a list of tie points (local co-registration approach) or a single X/Y shift vector (global co-registration approach).

The presented workflows for correcting local or global displacements are different algorithms/modes and both are components of AROSICS. The **local co-registration method** estimates X/Y translations for each point inside the overlap area of the input images by applying phase correlation in a moving-window manner to a regular grid of coordinate points. Afterwards, a set of quality metrics are used in order to validate the resulting shift vector grid. After passing all validation stages, the remaining points are utilized as a list of tie points to fit an affine transformation model that warps the target picture and corrects any local geometric misalignments that were detected (Scheffler et. al, 2017).

On the other hand, the **global co-registration method** calculates the displacements within a small image subset (from here: matching window) at only one coordinate position. Then, a single displacement vector is used to shift the target image. Compared to the local co-registration approach, the global method is computationally inexpensive and, therefore, very fast. Additionally, by adjusting the geocoding metadata, the deactivation of image warping to address sub-pixel X/Y shifts without resampling is allowed. This does not match the coordinate grid of the reference image, but it does keep the initial pixel values (Scheffler et. al, 2017). However, the global co-registration method assumes that a constant (global) translation in the X and Y dimensions may characterize the entire displacement between the reference and target images.

In conclusion, the local co-registration approach identifies the spatial shifts over the whole overlap area of the input images. The spatial shifts are computed for every point in the grid of which the parameters can be changed by using keyword arguments. Then, shift correction uses the estimated shifts of each grid point as ground control points to execute a polynomial transformation. As a result, locally varying geometric distortions of the target image can be corrected utilizing this approach. In the case of the global co-registration approach, the global changes in X and Y between a reference image and the target image are recognized and fixed by the algorithm performing a global shifting in X or Y direction. Within AFRI4Cast, both local and global co-registration approaches were tested and their performance in the selected hyperspectral and multispectral images was evaluated.

For the accomplishment of the co-registration approach, the reference and the target images (inputs) were defined.

- **Reference Image:** The multispectral layerstack (Sentinel-2) in BOA reflectance values and Float32 data type
- **Target Image:** The hyperspectral cube (PRISMA) in BOA reflectance values and Float32 data type

AROSICS framework also allows the fine-tuning of the results through the provision of user-defined data. In the terms of HS/MS image co-registration within AFRI4Cast, the grid resolution (in case of local co-registration) and the matching window size were used as tuning arguments. The grid resolution defines the distance between the tie points in pixels of the target image grid. At this point, it is worth mentioning that small point distances enable the complex distortions correction, although they increase the computational load. As for the matching window size, it has a default value of 256 x 256 pixels or a user-provided setting and it remains constant within the overlap area used for co-registration. Since the accuracy of the estimated displacement varies with the size of the matching window, it is necessary that the results of neighboring displacement vectors remain comparable. Several tests were performed, in which the tuning parameters received the following values:

- **Grid resolution (grid_res)** (float): {5, 15, 25}
- **Window size (window_size)** (Tuple[int, int]): {64, 128, 256, 512}

Apart from the arguments that were used for the algorithm tuning, there were also the fixed arguments which maintained a constant value throughout the various tests that were performed. These arguments were the following:

- **path_out** (str) – target path of the coregistered image
- **fmt_out** (str) – raster file format for output file. Can be any GDAL compatible raster file format (e.g. ‘ENVI’, ‘GTIFF’; default: ENVI)
- **r_b4match** (int) – band of reference image to be used for matching
- **s_b4match** (int) – band of shift image to be used for matching
- **max_shift** (int) – maximum shift distance in reference image pixel units (default: 5 pixel)
- **align_grids** (bool) – True: align the input coordinate grid to the reference (does not affect the output pixel size as long as input and output pixel sizes are compatible
- **out_gsd** (float) – output pixel size in units of the reference coordinate system
- **v** (bool) – verbose mode

Experimental Results:

Co-registration method

Various tests were performed in the selected hyperspectral and multispectral images both with the global and the local co-registration methods for the delineation of the appropriate approach for our test case.

It was observed that locally varying spatial displacements existed in our test case. Therefore, a co-registration approach that is able to identify the spatial shifts by utilizing every point in the grid (tie points) in order to carry out the shift correction resulting in the correction of the locally varying displacements (local co-registration method) was preferred compared to a method that performs global shifting in X or Y direction so as to fix the spatial misalignments. Moreover, the performance of the several tests presented better results in the local co-registration mode trials than the global ones. Hence, the local co-registration method was chosen to proceed with as it was proven to be the most appropriate in our case.

Tuning arguments evaluation

For the determination of the best values for the tuning parameters (grid resolution and matching window size), several tests including all the possible combinations between the grid resolution and the matching window size values were performed. As mentioned above, within these tests, the tuning parameters' values that were tested were:

- **Grid resolution (grid_res):** {5, 15, 25}
- **Window size (window_size):** {64, 128, 256, 512}

These parameters were evaluated based on the co-registration overall statistics extracted after the performance of local co-registration for each parameter combination that was carried out. The selected evaluation approach compares the similarity of the subset image content within the matching window before and after shift correction. It is based on the assumption that a successful shift correction results in an increased similarity between the reference and the target image (Scheffler et. al, 2017). Thus, the Mean Structural Similarity Index (MSSIM) was chosen as an image similarity metric since it is known to be highly sensitive to even minor image displacements. MSSIM can have a value between 0 and 1, where 0 indicates no match and 1 indicates a perfect match.

In this context, "MSSIM before" and "MSSIM after" constitute the metrics that were used for the tuning arguments evaluation. "MSSIM before" expresses the mean structural similarity index within each matching window before co-registration, while the "MSSIM after" expresses the mean structural similarity index within each matching window after co-registration.

Relying on these metrics, the best results were indicated when the grid resolution was equal to 25 pixels (grid_res = 25) and the matching window size equal to 64 pixels (window_size = 64). That was the case in which the biggest difference between the MSSIM index before and after the co-registration as well as the highest result of the MSSIM index after the co-registration were observed, compared to the same metrics of all the other tests that were performed.

Tuning and fixed arguments configuration

Considering the tuning and fixed arguments configuration, the values of these parameters, that were used during the local co-registration mode, were formed as follows:

- grid_res : 25,
- window_size: (64,64),
- path_out : '/path/to/store/output/coregistered/image/cor.tif',
- fmt_out : 'GTIFF',
- r_b4match: 7,
- s_b4match: 45,
- max_shift: 15,
- align_grids': True,
- out_gsd : [30,30],
- v: True

Regarding the band of the reference image (r_b4match) and the band of the target image (s_b4match) that were used for matching, it was decided to proceed with the bands that belong to the NIR spectral range. These bands corresponded to Band 7 in the multispectral composite and Band 45 in the hyperspectral cube.

In the images below, at first the relative geometric shifts between the Sentinel-2 and the PRISMA images are presented (Figure 33). Figures 34 and 35 exhibit the absolute geometric shift vectors calculated on the tie point grid, on which the shift correction was based. It was concluded that on average the geometric shift was 167.5 m. The absolute geometric shift vectors are presented in Figure 36, as well. The calculated shift angle which is equal to 173.6 degrees is also displayed in Figure 36. Lastly, Figure 37. exhibits the HS/ MS image co-registration result.

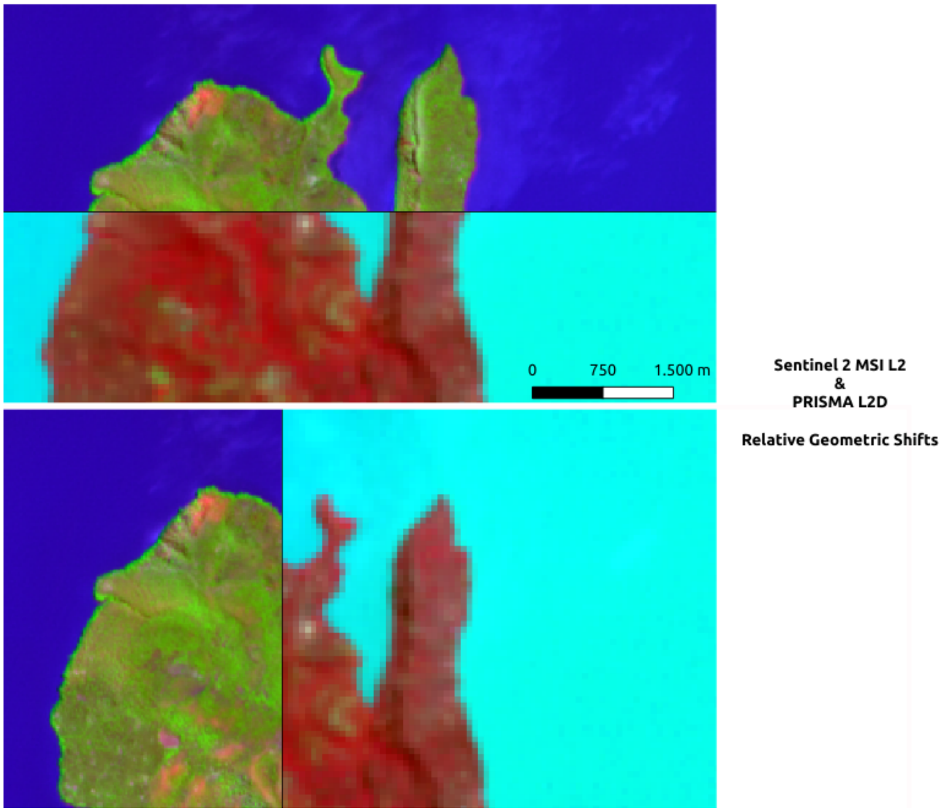


Figure 33. Relative geometric shifts between the Sentinel-2 and the PRISMA images. Both images are presented in RGB pseudocolor visualization.

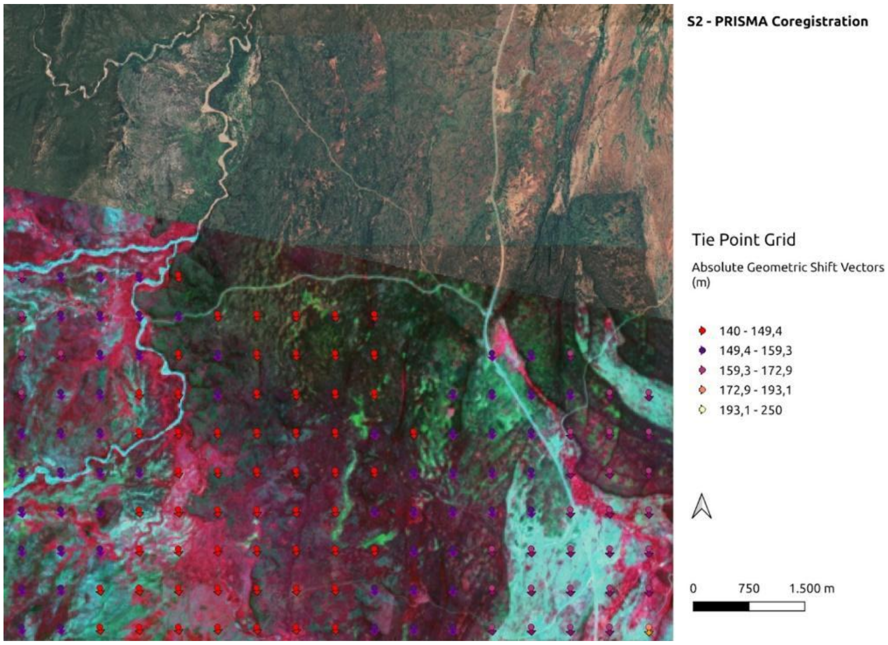


Figure 34. Absolute geometric shift vectors calculated on the tie point grid, on which the shift correction was based. The co-registered PRISMA image in RGB (850,650,550) visualization serves as a base image.

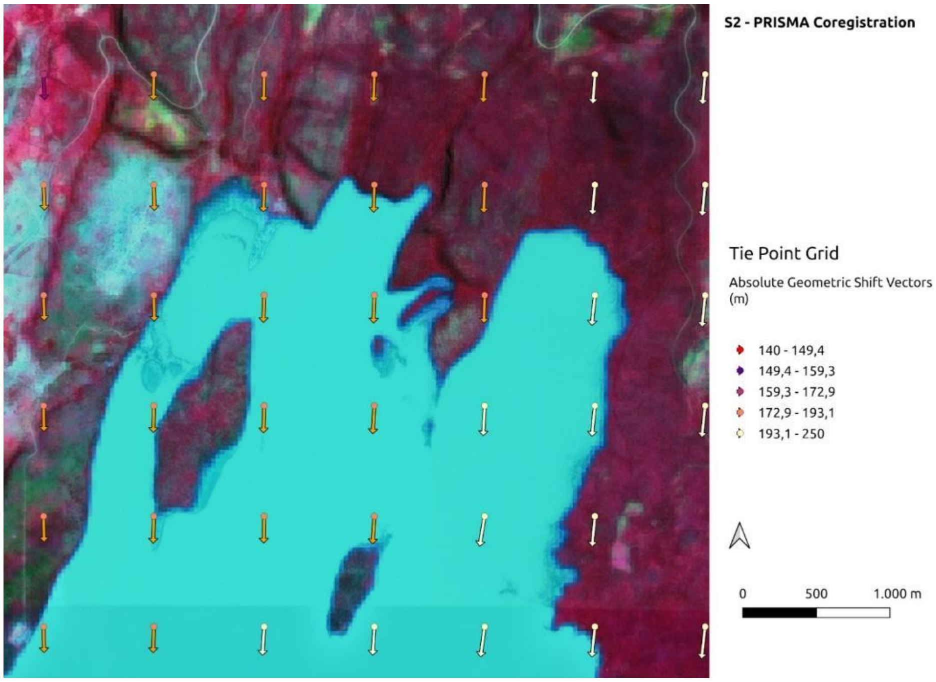


Figure 35. Geometric shifts calculated on the tie point grid, on which the shift correction was based. The co-registered PRISMA image in RGB (850,650,550) visualization serves as a base image.

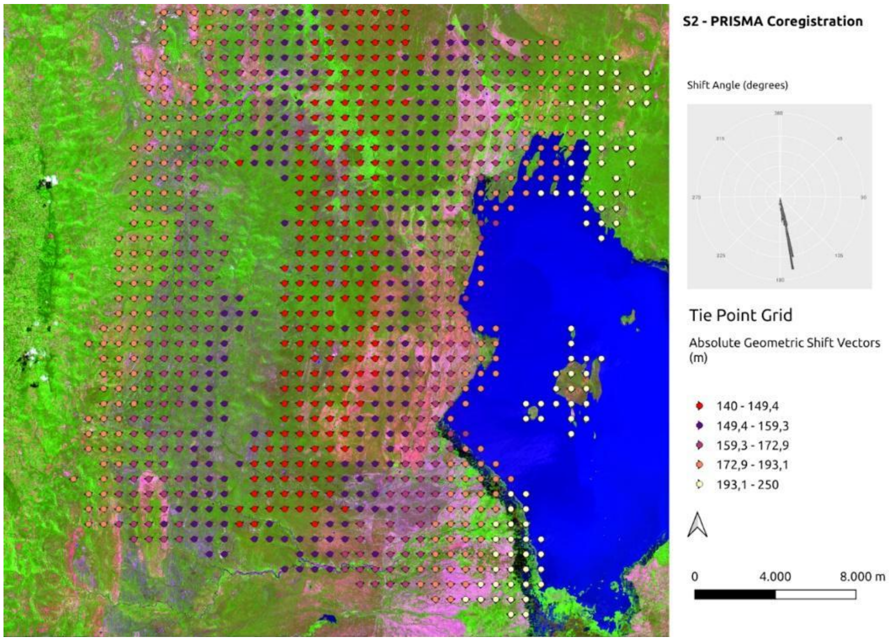


Figure 36. Absolute geometric shift vectors calculated on the tie point grid, on which the shift correction was based. The Sentinel-2 image in RGB pseudocolor (SWIR-RedEdge3-RedEdge1) visualization serves as a base image.

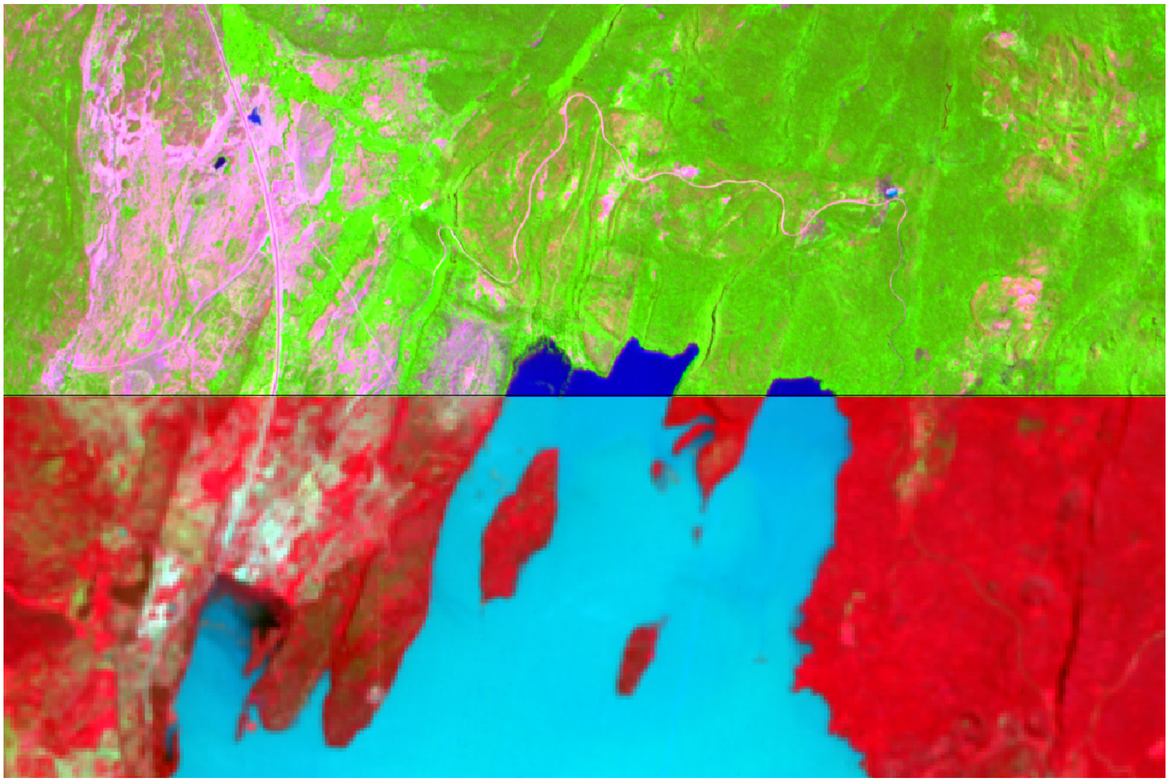


Figure 37. HS/ MS image co-registration result. Above: Sentinel-2 image, Below: Co-registered PRISMA image.

Issues and Next Steps:

The co-registration of PRISMA and Sentinel-2 images using the AROSICS local co-registration method produced satisfactory results that met our expectations, as determined by the evaluation of metrics outlined in the previous chapter. Thus, given the successful alignment accomplished between the images, no immediate necessity for additional actions is considered.

4.6. Hyperspectral - Multispectral Data Fusion

Challenge:

According to the findings documented in the relevant deliverable D3, the task of HS - MS Fusion, considers a pair of images that cover the same region of interest but also pose a tradeoff between their dimension resolutions. This tradeoff regards the well-known and inevitable inverse-reciprocal relationship between spatial and spectral resolution, in remote sensing systems. HS Imagery with a large number of bands usually has a low spatial resolution to ensure high SNR. On the contrary, MS sensors can obtain an image with a higher spatial resolution but with a small number of spectral bands. An image fusion algorithm is designed in order to recreate a synthetic image that overcomes that specific trade-off to be generated, by extracting image priors, i.e. features, that outline the relative advantages of each of the HS-MS representations (Nanni et al, 2017). For example, the low-rankness of HS image, or the spatial smoothness of MS, have been considered as useful priors, for that cause. At this point it is worth mentioning that an important factor in the flexibility of an algorithm is the way in which the image priors it uses are calculated and extracted.

Traditional image fusion frameworks emerging from classical computer vision methods, design explicitly defined handcrafted image priors. The parametrization of latent variable selection in component substitution algorithms is the simplest example of it, while the more advanced Multiresolution Analysis methods involve a thorough tuning of their filtering parameters in order to approximate intrinsic image priors of the HS sensor, such as the MTF. In that sense, these features are considered as handcrafted because they presume expert knowledge in their design, in order to overcome specific issues like occlusions and variations in image scale and illumination, while balancing between accuracy and computational efficiency. The frameworks that involve the use of handcrafted image features can reconstruct the enhanced image without the need of a guidance image as a training dataset, by explicitly assuming prior knowledge. However, the priors designed for one task may not be as effective when applied to other tasks. This has an effect on the flexibility of the framework, to handle, for example, data from different sensors (Uezato et al, 2020).

Recently, Deep Learning Image Fusion frameworks have received more and more attention in many processing applications due to its high efficiency and promising performance. In the HS and MS image fusion, DL methods use LR-HS and HR-MS images as an input and an HR - HS image as an output, by learning the mapping function between the inputs and the output. Deep learning (DL) approaches avoid the assumption of explicit priors for each specific fusion task. Although network architectures themselves need to be handcrafted, properly designed network architectures have shown to solve various problems. However, they have the disadvantage that they need training data, which limits their use in practical remote sensing applications. It is usually difficult, or even impossible to collect a large size of training data including reference, i.e., HR-HS or HR-MS (Uezato et al, 2020). Most studies, like the ones presented on deliverable D3, used synthetic data emerging from HR-HS imagery, from missions such as the AVIRIS, that exist as experimental datasets. For that reason, they have limited generalization performance on “real data” and have to be modified in order to be applicable.

In consideration of the particularities of the HS-MS Fusion task in general, outlined above, combined with the constraints that justify it in the case of Afri4Cast, such as the intrinsic specifications of Sentinel 2 MSI and PRISMA imagery, or even the generalization capability of the algorithm on an operational level, the following challenges are highlighted:

- The need of handcrafted image feature extraction that requires expert knowledge for parameterization, in the case of traditional computer vision approaches
- The requirement of High Resolution Hyperspectral data for training a Deep Learning Fusion Network.

Approach:

The decided approach that was adopted in Afri4Cast for the HS-MS image fusion task was, initially, to develop and evaluate models derived from traditional fusion methodological frameworks, prevalent in the remote sensing field, such as Component Substitution algorithms, and use them as benchmarks. Subsequently the implementation of a Deep Learning Fusion Framework, for which various qualitative and quantitative properties such as the accuracy, generalization, or its computational efficiency would be evaluated in comparison to traditional methods. The experimental design involved implementation and evaluation on a common dataset, from the pilot regions of interest of Afri4Cast, which had undergone the same preprocessing steps.

Hyper Pansharpening – Components Substitution Methods

The starting point for the pan-sharpening process was the acquisition of the basic data set for the analysis. In particular, a hyperspectral (HS) image (e.g. PRISMA with 30 m resolution) and one or more multispectral images (e.g. Sentinel-2 with 10 m resolution) that completely covered the same area as the HS, in the same time period, was used. A number of methods will be described below, which have been used and compared to identify the one with the best performance. The methods used are:

- Brovey Fusion Methodology;
- FHIS (Fast Intensity Hue Saturation);
- Gram-Schmidt;
- Gram-Schmidt Adaptive.

Deep Learning Fusion

In order to bridge the gap between the traditional and supervised DL approaches, an unsupervised DL approach has been adopted, as a way to bypass the necessity of HR-HS training data. But could it be possible to use DL approaches without training data? The possible way to solve this obstacle was by adopting a zero-shot learning scheme, which trains the deep network from the data to be fused, by employing Wald's Protocol principle. According to that, the performance of data fusion models is independent of the scale, provided that certain conditions hold. In their seminal work Wald et al. (1997) suggested first degrading the input image according to a factor k, thus creating LR-HR image pairs, and proceed to design a model tasked to downscale the degraded HS image to the original resolution. Then the developed method can be transferred to downscale the original image into one of much higher resolution according to the same downscaling factor k. The suggested approach was the development of an unsupervised DL method. This network architecture was proposed as a regularizer for unsupervised image fusion tasks that can be applied in a zero shot learning scheme, with a degrading factor equal to 3, by using the multiscale semantic features from a guidance image, which for the specified task was the PRISMA image at its native spatial resolution.

Methodology:

4.6.1. Hyper Pansharpening – Components Substitution Methods

Brovey Fusion

The Brovey method, a widely recognized pan-sharpening technique in remote sensing, uses the following approach: the pan-sharpened image is acquired by changing the intensity values of each multispectral band according to the proportion of the corresponding pixel values in the panchromatic band. This method is typically applied individually to each band and the resulting pan-sharpened bands are subsequently merged to generate the final high-resolution image.
The Brovey pan-sharpening equation for a specific band can be mathematically represented as:

$$F_i = \frac{B_i}{\left(\sum_{\square}^{\square} \square\, of all bands\right)} \cdot Sen2$$

Where F_i is the ith fused band, B_i is the single band of the hyperspectral image (PRISMA), Sen2 is the image of the multispectral band (Sentinel-2) while the sum is taken over all the bands considered for the region. Through this normalization technique, the spectral attributes of multispectral data are retained, while the spatial resolution is enhanced using detailed information provided by the panchromatic band. The panchromatic band in this case is represented by the S2 wider band channel. Well-known for its simplicity and effectiveness, the Brovey method efficiently preserves spectral accuracy throughout the pan-sharpening process.

FIHS (Fast Intensity Hue Saturation)

The Fast Intensity Hue Saturation (FIHS) pan-sharpening technique is a cutting-edge method within the realm of remote sensing data fusion. Through FIHS, a pan-sharpened image is skillfully created by carefully adjusting both the intensity and hue aspects of multispectral bands, taking cues from the pixel values present in the panchromatic band. This goes beyond traditional approaches that mainly concentrate on intensity adjustments, as FIHS thoughtfully incorporates hue, providing a comprehensive enhancement to the combined imagery. During the FIHS procedure, each band undergoes a thorough transformation, with intensity changes made in relation to pixel values in the panchromatic band, like the Brovey technique. Concurrently, the hue element is delicately adjusted to ensure accurate representation of spectral attributes. This simultaneous alteration guarantees that not just spatial details, but spectral subtleties are meticulously preserved within the resulting pan-sharpened image.

In situations where the harmonious fusion of spatial and spectral data is vital, the FIHS method proves to be highly effective. This makes it a significant tool for applications such as land cover classification and environmental monitoring. As FIHS expands upon traditional pan-sharpening practices by incorporating hue adjustments, it distinguishes itself as an advanced approach capable of producing high-resolution images with enriched spectral accuracy.

For each band B_i, the fused band F_i is calculated using the formula:

$$F_i = B_i + (Sen2 \cdot L)$$

Where B_i is the single band of the hyperspectral image (PRISMA), Sen2 is the image of the multispectral band (Sentinel-2) and L is the sum of all the bands divided for the number of bands. The sum is taken over all the bands considered for the region.

Gram-Schmidt

It is a method for orthogonalizing a set of vectors. In the specific context of data fusion, particularly pan-sharpening in remote sensing, it helps enhance spatial resolution by combining high-resolution panchromatic imagery (high-resolution multispectral data) with lower-resolution multispectral data (lower-resolution hyperspectral data). Following a simplified explanation of how the Gram-Schmidt algorithm works for data fusion.

In the data initialization, we have two sets of data one is multispectral (MS) bands and a panchromatic (PAN) band. The Gram-Schmidt process begins by considering the MS bands one at a time. The first MS band is chosen as the initial vector for the orthogonal basis. The chosen MS band is projected onto the subspace defined by the PAN band. This projection yields the component of the MS band that aligns with the PAN band. Simultaneously, the orthogonal component, which is perpendicular to the PAN band, is identified. The sum of the orthogonal and projection components is equal to the zero-mean version of the original MS band. This ensures that the MS bands are adjusted in a way that preserves their spectral information. The algorithm proceeds iteratively, considering each MS band in turn. For each band, it finds the projection on the subspace defined by the PAN band and its orthogonal component with respect to previously processed bands. Once the orthogonalization is complete, the pan-sharpening is achieved by combining the orthogonalized MS bands with the original high-resolution PAN band.

Gram-Schmidt Adaptive

The Gram-Schmidt Adaptive (GSA) algorithm builds upon the Gram-Schmidt process by introducing adaptability to better handle variations in spectral characteristics. Here's a breakdown of how the Gram-Schmidt Adaptive algorithm works for data fusion. Similar to Gram-Schmidt, we have two sets of data: multispectral (MS) bands and a panchromatic (PAN) band. MS bands are ordered and interpolated to the PAN scale, and the mean of each band vector is subtracted.

- Synthetic Low-Resolution Approximation (IL): GSA starts with a synthetic low-resolution approximation of the PAN image (IL) as the initial vector in the orthogonal basis.
- Orthogonalization Procedure: Like Gram-Schmidt, GSA proceeds by orthogonalizing one MS vector at a time. The chosen MS band is projected onto the subspace defined by the PAN band, and its orthogonal component is identified.
- Adaptive Scaling: The key innovation in GSA is the adaptive scaling factor for pan-sharpening. Instead of a fixed scaling factor, GSA adjusts this factor based on the maximum intensity values

within the hyperspectral bands. This adaptability ensures better preservation of spectral information.

- Pan-sharpening and Inverse Transformation: Pan-sharpening is accomplished by replacing the synthetic low-resolution approximation (IL) with a histogram-matched version of the PAN image (P). The inverse transformation completes the pan-sharpening process.

Methods discussion

The Gram-Schmidt Adaptive (GSA) algorithm builds upon the Gram-Schmidt process by introducing adaptability to better handle variations in spectral characteristics. These techniques were originally designed for enhancing the spatial resolution of multispectral imagery. However, this approach involves testing these methods beyond their conventional application. Rather than pan-sharpening high-resolution multispectral data, their efficiency was explored in the fusion of high-resolution (10 meters) multispectral data from Sentinel-2 with lower-resolution (30 meters) hyperspectral data from PRISMA. This unconventional application assesses the adaptability and performance of these methods in the context of combining diverse spatial and spectral characteristics.

The selected fusion methods were evaluated according to their ability to preserve the spectral information of the original PRISMA hyperspectral data, while improving their spatial resolution. Specific spectral bands corresponding to the blue, green, red and near-infrared (NIR) regions were considered in the analysis. The choice of these specific regions was based on the availability of a resolution of 10 meters in these bands within the Sentinel-2 dataset. For the fusion process, the Sentinel-2 dataset provided the blue band (B2) ranging from 460.2 nanometres to 525.2 nanometres. This band was merged with the PRISMA dataset, specifically bands 9 to 17, which cover the same spectral range. The same selection process was followed for all four regions (blue, green, red and NIR) (Table 20). By combining these selected bands, the fusion process integrated 29 bands of the PRISMA hyperspectral data with the 4 bands of Sentinel-2. Using these fusion techniques, the study aims to demonstrate the effectiveness of the selected methods in enhancing the spatial resolution of PRISMA hyperspectral data for improved analysis and classification.

Table 20. Sentinel-2 and PRISMA bands selection for the pan-sharpening algorithm comparison.

Region	Sentinel 2	PRISMA
Blue Region	Band_2 (460.2-525.2) nm	B9-B17 (468-527.3) nm
Green Region	Band_3 (542.3-577.3) nm	B20-B24 (550.9-583.8) nm
Red Region	Band_4 (649.6-679.6) nm	B32-B35 (655.4-684.1) nm
NIR	Band_8 (780.3-885.3) nm	B45-B55 (785.6-892) nm

Total No. Bands	4	29
Resolution	10 meters	30 meters

4.6.2. Deep Learning for HS-MS Image Fusion

The followed approach for solving the HS-MS data fusion problem for PRISMA and Sentinel 2 images with an unsupervised Deep Learning framework, regarded the architecture proposed by Uezato et al. (2020) of a Guided Deep Decoder Network for Hyperspectral Super Resolution (GDD), applied on a Zero Shot Learning scheme.

Problem Formulation

The task denotes:

- a low-resolution input image (PRISMA)**:** $Y \in R^{C \times w \times h}$
- a guidance image (Sentinel 2 MSI) : $G \in R^{c \times W \times H}$

where **C, W**, and **H** represent the number of channels, the image width, and the image height, respectively. For HS Super Resolution **w<<W, h<<H, c<<C**

In an unsupervised image fusion approach, the corresponding output can be estimated by solving the following optimization problem:

$$\theta min \; L(f_{\theta}(Z), Y, G)$$

where **X** is estimated by a convolutional neural network (CNN)-based mapping function

$$X = f_{\theta}(X)$$

where $f_{\theta}(Z)$ represents the mapping function with the network parameters θ, and Z is the input representing the random code tensor. In this formulation, only one input image Y and a guidance image G are used for the optimization problem, thus, training data are not required. X is regularized by the implicit prior of the network architecture.

Deep Network Architecture – The Guided Deep Decoder (GDD)

GDD is composed of an encoder-decoder network with skip connections and a deep decoder network, as shown in Figure 38. The encoder-decoder network is similar to the architecture of U-net and produces the features of a guidance image at multiple scales. The multi-scale features represent hierarchical semantic features of the guidance image from low to high levels. The semantic features are used to guide the parameter estimation in the deep decoder.

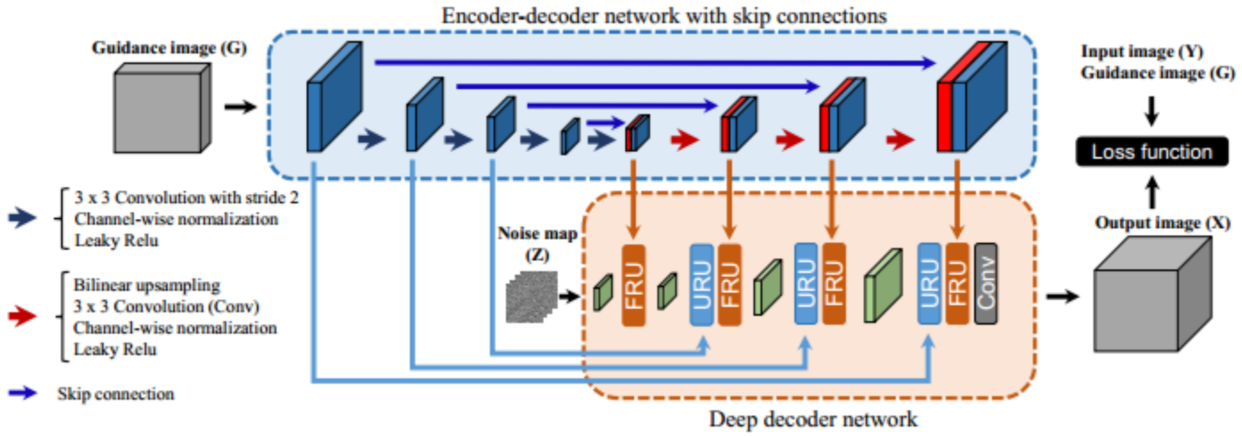


Figure 38. GDD network architecture. The blue layers represent the features of the encoder. The red layers represent the features of the decoder. The green layers represent the features of the deep decoder network. The semantic features of G are used to guide the features of the deep decoder in the upsampling and feature refinement units, URU and FRU (from Uezato et al., 2020).

Upsampling refinement unit (URU): Upsampling is a vital part of GDD. URU incorporates an attention gate for weighting the features derived after upsampling and channel-wise normalization (CN) in the deep decoder. The bilinear upsampling used in GDD causes a strong bias to promote piecewise smoothness and tends to wash away the small objects or boundaries. GDD uses an attention gate to weight the features derived by the upsampling. The features from the guidance image are gated by a 1 × 1 convolution (Conv), a leaky rectified linear unit (LeakyRelu), and a sigmoid activation layer (Sigmoid) to preserve the spatial locality of the features and generate the conditional weights (Uezato et al. 2020).

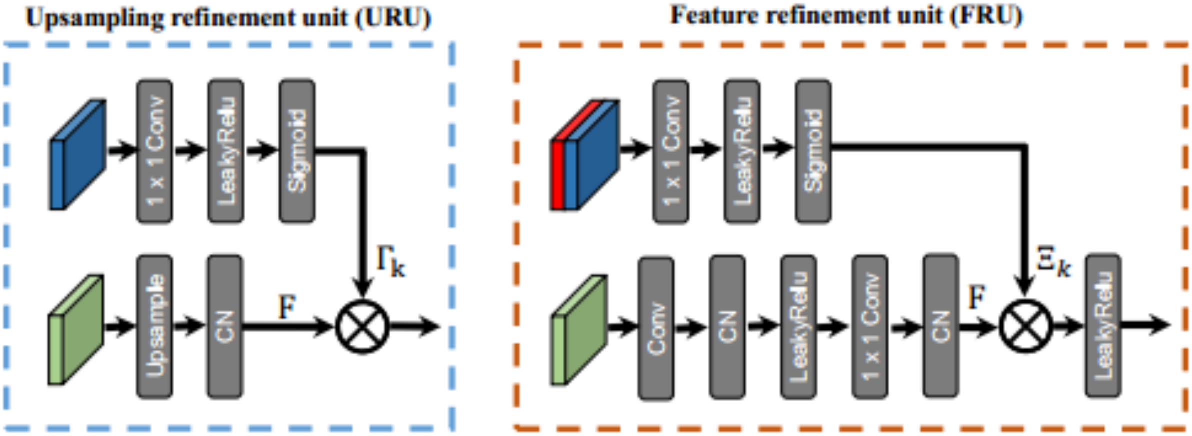


Figure 39. The internal structure of upsampling and feature refinement units (from Uezato et al., 2020).

Feature refinement unit (FRU): FRU is different from URU in that the features of the deep decoder are weighted by the high-level semantic features of the guidance image. FRU promotes semantic alignment with the features of the guidance image, while URU promotes similar spatial locality. Using an attention gate, the high-level features are gated by a 1 × 1 convolution, a leaky rectified linear unit, and a sigmoid activation layer to generate the conditional weights. GDD enables the features of each layer in the decoder part to be conditioned on the semantic features from the guidance image at multiple scales. The attention gates at multiple scales emphasize salient features within each layer, leading to the semantic alignment between the output image and the guidance image (Uezato et al. 2020).

HS Super Resolution Loss Function: The loss function of the task was designed to preserve the spectral information from the HS image while keeping the spatial information from the MS image. For simplicity, the matrix forms of **X, Y, G** image arrays are denoted as $\hat{X}$, $\hat{Y}$, $\hat{G}$, respectively. Given the estimated HR - HS $\hat{X}$ image, the loss function can be defined as:

$$L(X, Y, G) = \mu \cdot \|X^{\wedge} \cdot S - Y^{\wedge}\|_F^2 + \|R \cdot X^{\wedge} - G^{\wedge}\|_F^2$$

Where $\|.\|_F$ is the Frobenius norm, S is the spatial downsampling with blurring and R is the spectral response function that integrates the spectra into the MS channels. The first term encourages the spectral similarity between the spatially downsampled X and Y. The second term encourages the spatial similarity between the spectrally downsampled X and G. μ is a scalar controlling the balance between the two terms (Uezato et al. 2020).

GDD Network Preprocessing

Zero Shot Learning Scheme: The application of a Zero Shot Learning scheme in an Unsupervised Deep Learning HS-MS fusion framework involves the spatial downgrading of the HS-MS image pair by a factor equal to the quotient of their respective spatial resolutions. For the PRISMA HSI - S2 MSI fusion this factor f equals to 3, and the degradation model is the 3X3 block average. Thus, the derived datasets are:

$HSI \in R^{C\times\left(\frac{w}{f}\right)\times\left(\frac{h}{f}\right)}$: The low-resolution HS image, spatially degraded PRISMA.

$MSI \in R^{c\times\left(\frac{W}{f}\right)\times\left(\frac{H}{f}\right)}$: The Guidance-high resolution MS image, spatially degraded Sentinel 2

which feed the GDD network, along with the Ground Truth image:

$GT \in R^{C\times w\times h}$: The initial PRISMA HS image, considered as the HR-HS image over which GDD is evaluated

SRF resampling from Sentinel 2 to PRISMA: Spectral Response Functions - often abbreviated to 'SRF' - describe how sensitive an instrument's photosensor is to radiation of a particular wavelength. Different photosensors have varying sensitivities to different wavelengths that can vary across the spectrum. This is the case in remote sensing instruments. SRF is a sparse matrix, in that for each row representing a sensor's channel, it has non-negative values only on the columns that regard the wavelengths that correspond to this channel. In the formulation of the Loss function, R denotes the Spectral Response Function (SRF) which degrades output of the GDD Net, spectrally, in order to compare it with the Guidance image, the HR MS. It can be seen as a mapping from MS to HS channels space, i.e. $C \ \leftrightarrow c$

$$R \in R^{c\times C}$$

In HS-MS fusion problems, the SRF is simulated by the resampling of the SRF of the MS sensor, to the spectral resolution of the HS sensor. Regarding the applied fusion task, c equals the number of PRISMA HSI spectral channels, while C equals the number of Sentinel 2 MSI spectral channels. The spectral resampling function transforms the MSI instrument response, usually given in 1nm resolution, to an aggregated interval, the HSI spectral resolution, given by the full width half maximums (fwhm) of the new band positions. The function uses gaussian models defined by fwhm values to resample the high-resolution data to new band positions and resolution. It assumes that band spacing and fwhm of the input data is constant over the spectral range. The interpolated values are set to 0 if input data fall outside by 3 standard deviations of the gaussian densities defined by fwhm. The Sentinel 2 SRF, as resampled to simulate PRISMA spectral resolution, is depicted on the following Figure 40.

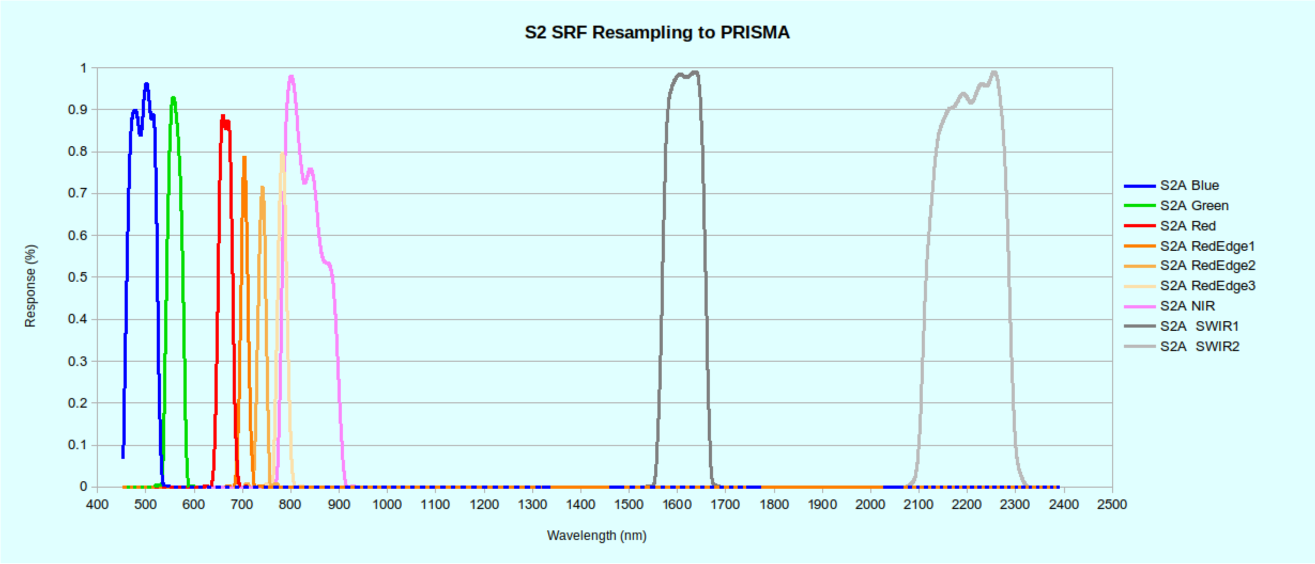


Figure 40. Spectral Resampling of the S2A SRF to the PRISMA HSI spectral resolution.

Algorithm Outline

1. **Provide Inputs**
   - MSI
   - HSI
   - GT
   - R

2. **Set Parameters:** the most important of them being,
   - factor : the spatial degradation factor
   - param_balance: the μ parameter for the spatial-spectral balance in the loss function
   - method = '2D'
   - pad = 'reflection'
   - OPT_OVER = 'net'
   - num_c : the number of channels
   - LR: the learning rate
   - OPTIMIZER = 'adam': Adam optimization is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments.
   - num_iter = 30000
   - reg_noise_std = 0

3. **Set net:** define GDD number of channels, network structure, squared L2 norm. downsampler etc

4. **Define closure:** Define Loss Function, and Evaluation Metrics

Experimental Results:

Hyper Pansharpening

Various types of methodologies and approaches for increasing the spatial resolution of hyperspectral images while maintaining spectral information and resolution have previously been described. By analyzing and applying these methods of fusing PRISMA images with Sentinel-2 images, the results shown below were obtained.

To determine the efficiency of various methods for improving spatial resolution and maintaining spectral details, different land covers, such as forests, crops, urban areas, and barren land, have been examined. The primary focus was to study the performance of data fusion methods in these distinct environments under a range of circumstances. To initiate the analysis, homogeneous regions within each land cover type have been identified and a 90x90 meter square grid considered with 9x9 pixels representing each landscape section's overall character. Within this confined area, the average values of these pixels were calculated in order to extract the core attributes of each land cover.

The detailed investigation delved into the complexity of the spectral signatures, using the graphs below to visualize the reflectance patterns in the various bands (Figure 41, Figure 42). Each spectral signature on these graphs represented spatial enhancement and spectral richness, giving an overview. The research extended beyond visual inspection, presenting charts with measurements such as Spectral Correlation Mapper (SCM), Spectral Angle Mapper (SAM) and the useful space between two spectral signatures. These numerical analyses provided a quantitative perspective on the performance of each method.

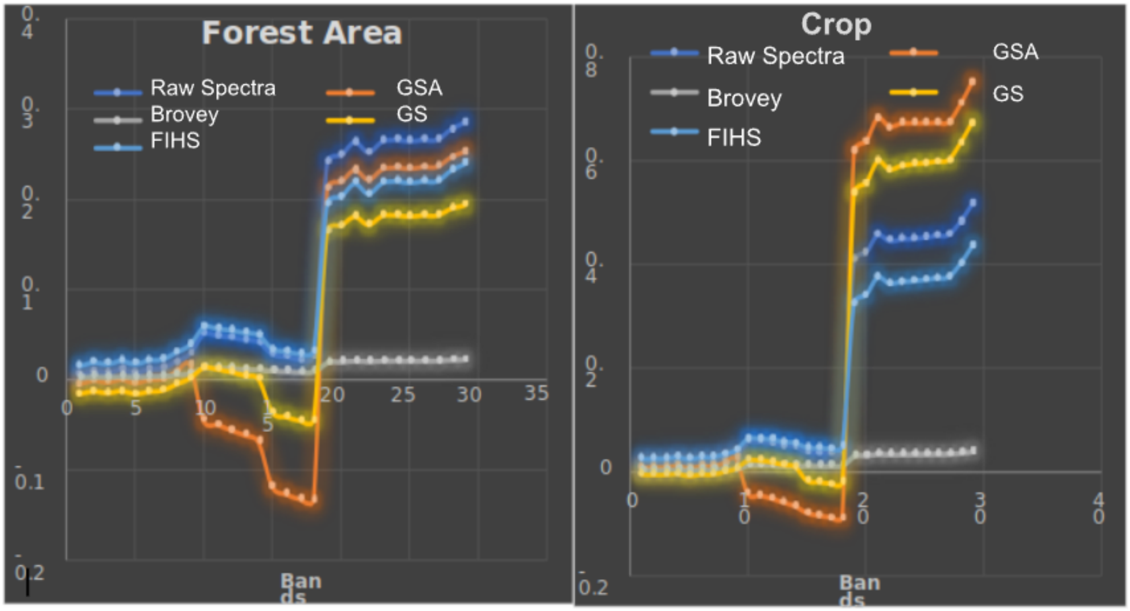


Figure 41. Left: Before and After Fusion for Forest area. Right: Before and After Fusion for Crop area.

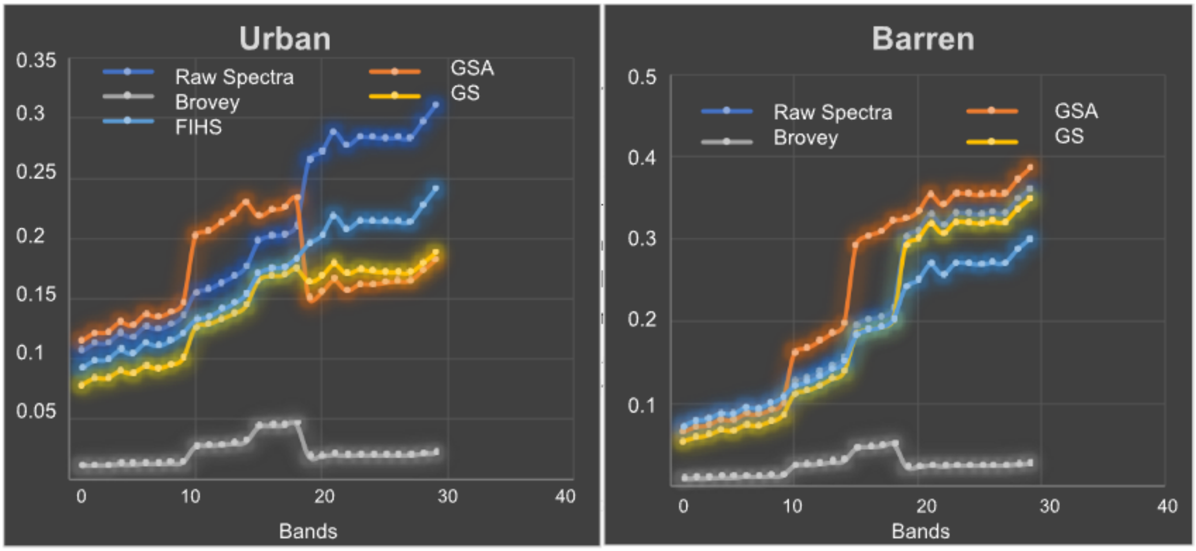


Figure 42. Left: Before and After Fusion from Urban area. Right: Before and After Fusion for Barren Land area.

All implemented algorithms take the two datasets as input and generate fused data with enhanced spatial resolution. Preliminary results demonstrate that the Gram-Schmidt Adaptive fusion algorithm exhibited promising results, providing more accurate and detailed information while preserving spectral information compared to other fusion methods. Finally, it can be stated that the performance of the fusion techniques was assessed using quantitative and qualitative evaluation metrics. Spectral Angle Mapper (SAM), Spectral Correlation Mapper (SCM), and other spectral similarity measures were employed to compare the spectral signatures of the original and fused datasets. Spatial analysis techniques, including land cover classification, were also utilized to evaluate the enhanced mapping accuracy achieved by the fused dataset.

The introduction of suitable metrics allowed to quantitatively estimate the performances of the different methods tested. Ultimately, the FIHS method emerged as the most impressive technique in our findings. It delivered a balanced performance across bands while creating a harmonious representation of reflectance that surpassed other methods. In terms of spectral correlation and angle mapping, FIHS achieved better results. FIHS method not only proved efficient but also demonstrated an enriched understanding of these complex relationships.

An in-depth investigation of spatial improvement and spectral conservation, research thoroughly assessed the performance of data fusion techniques—Brovey, FIHS, Gram-Schmidt, and Gram- Schmidt Adaptive (GSA). The primary goal was to enhance spatial resolution while valuing the spectral subtleties found in PRISMA's hyperspectral data by incorporating Sentinel-2's multispectral information.

Visual representations effectively displayed the subtle interplay of reflectance across bands, presenting a clear picture of spatial and spectral dynamics. At the same time, the quantitative metrics Spectral Correlation Mapper (SCM), Spectral Angle Mapper (SAM) and the area between two spectral signatures provided valuable numerical indications, observable in Tables 21-24. As a result of this comprehensive evaluation, the FIHS method surfaced as the most proficient approach in striking a balance between enhanced spatial resolution and spectral accuracy. In measurable terms, FIHS has formed with an average value of SCM: 0.9969, SAM: 0.0738, Area: 0.8719. These numerical validations emphasized FIHS's method has dominated other data fusion methods. In conclusion, this study not only deepened the understanding of these methods but also reasserted the crucial role of data fusion in harnessing the full power of remote sensing applications.

Table 21. Different methods performance for a Forest Area.

Forest Area			
Methods	**SCM**	**SAM**	**Area**
GSA	0.8858	0.4779	1.54
Brovey	0.9569	0.2936	2.88
GS	0.9689	0.2468	1.54
FIHS	0.9971	0.0753	0.63

Table 22. Different methods performance for a Crop Area.

Crop Area			
Methods	**SCM**	**SAM**	**Area**
GSA	0.9823	0.1858	3.39
Brovey	0.9728	0.2262	4.82
GS	0.9948	0.1004	2.09
FIHS	0.998	0.0564	1.02

Table 23. Different methods performance for a Urban Area.

Urban Area

Methods	SCM	SAM	Area
GSA	0.9108	0.3951	1.69
Brovey	0.8817	0.4899	5.01
GS	0.9541	0.213	1.71
FIHS	0.9959	0.0817	1.07

Table 24. Different methods performance for a Barren Land.

Barren Area			
Methods	**SCM**	**SAM**	**Area**
GSA	0.9893	0.1402	0.85
Brovey	0.8771	0.5008	5.00
GS	0.9988	0.0448	0.36
FIHS	0.9966	0.0818	0.77

Deep Learning for HS-MS Fusion

The experimental design was carried out in order to obtain a proof of concept for the feasibility of using unsupervised Deep Learning Image Fusion on "real" Earth Observation data, with an evaluation carried out in a zero-shot learning approach. In order to study and evaluate the above concept in many different configurations, given the increased computational time required for training in these cases, the experimental implementation on patches of located and restricted areas of interest was chosen.

Thus, for the HS-MS, PRISMA and Sentinel 2 image pair with an acquisition date of 26/05/2020, the preprocessing steps outlined in the previous Chapter 4.5. were implemented. For the dimensions of the image patches, further consideration had to be given to the condition imposed by the Deep Learning architecture to be implemented. The spatial dimensions of the patches should be integer multiples of 32. In the case of our data the scale factor equal to 3 could not meet the above condition and therefore the Sentinel 2 data were resampled to a spatial resolution of 7.5 m in order to apply a scale factor equal to 4.

The synthetic data for zero shot learning were created by spatial degradation of the original HS and MS images. For the understanding of image degradation at a lower resolution, significant is the contribution of the imager Point Spread Function (PSF). The blurring that occurs through the degradation process is formulated as an image convolution with the PSF kernel. According to Wang et.al 2020, experimental results showed that the PSF effect influences downscaling greatly, and that downscaling can be enhanced obviously by considering the PSF effect. However, the actual PSF of the satellite image sensor is an unknown matrix that needs to be simulated in the degrading process, by the means of a low pass convolution. The idea of Gaussian smoothing is to use this 2D distribution as a simulated PSF. As a result, the image degradation scheme included:

- low pass filtering through an application of a Gaussian convolution, with a sigma value equal to 2. Gaussian smoothing simulates the Point Spread Function (PSF) effect in image downscaling.
- resampling to a lower spatial resolution for HSI and MSI, with a degradation factor equal to 4.

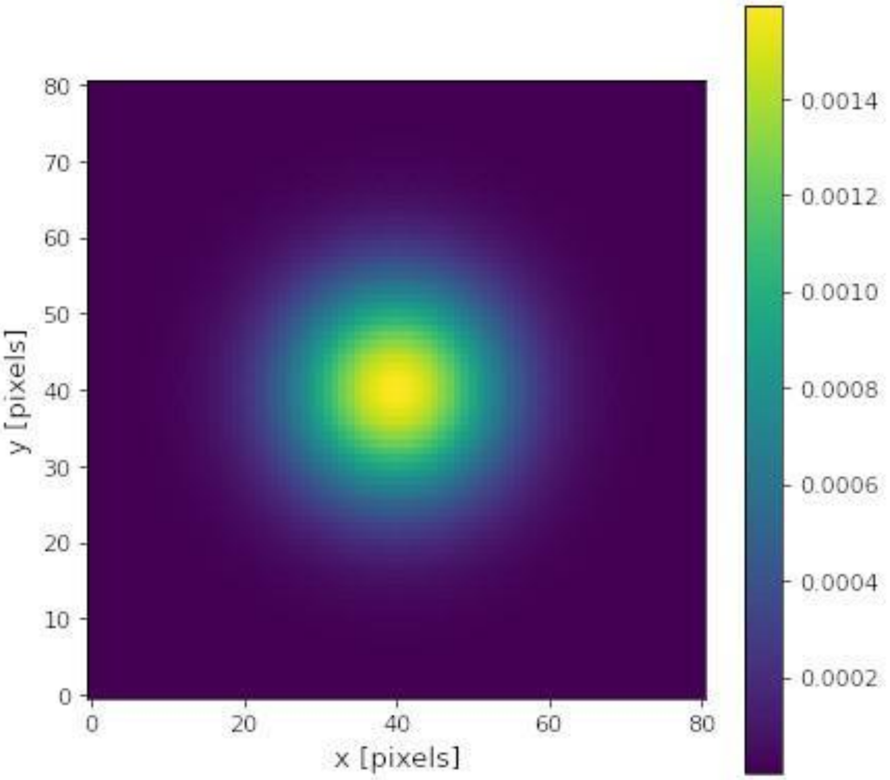


Figure 43. Gaussian 2D convolution kernel. A PSF simulation for zero shot degradation scheme.

As a result, the following image patch representations were created (Figure 44):

1. $MSI \in R^{9\times512\times512}$: at 7.5 m spatial resolution, not used in GDD
2. $Degraded\ MSI \in R^{9\times128\times128}$: at 30 m spatial resolution, the synthetic Guidance image(G), input to GDD network
3. $HSI \in R^{172\times128\times128}$: at 30 m spatial resolution, the ground truth image for evaluation of GDD model

4. $Degraded\ HSI \in R^{172\times32\times32}$: at 120 m spatial resolution, the synthetic HSI (Y) used in GDD model loss function optimization during training

Furthermore, a matrix representation of the Spectral Response Function: $R \in R^{9\times172}$,was used for the spectral downsampling of the GDD output, at the loss function calculation.

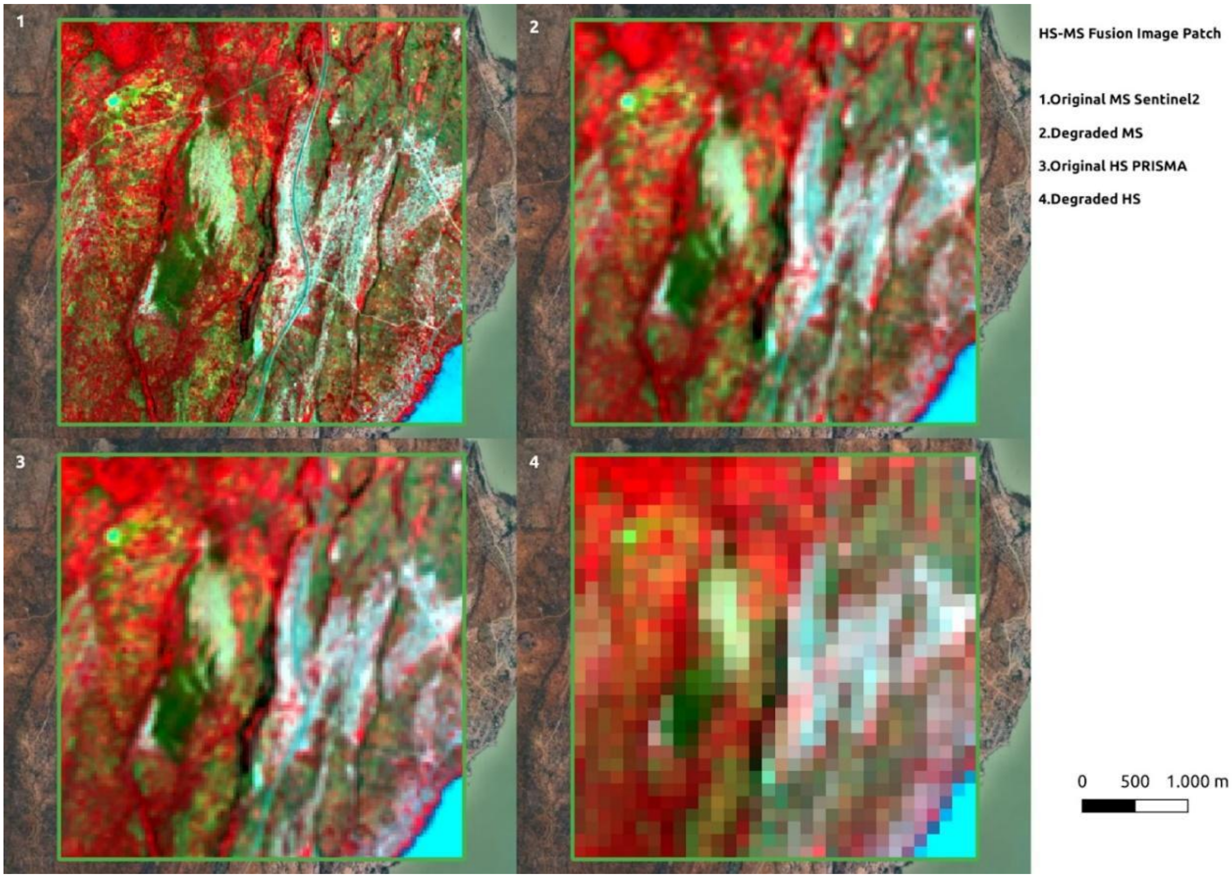


Figure 44. Image patch data for GDD training - evaluation.

The GDD model experiments for HS-MS image fusion were conducted on an NVIDIA GeGorce GTX1650 with 4GB memory. Depending on the parameterisation of each take, the training task duration ranged from 1,5 to 2 hours of computational time.

Initial trials of the GDD code were tested with the following parameters:

- factor = 4 (spatial degradation factor)
- param_balance = 0.4 (spatial vs spectral balance in loss optimization)
- number of network channels = 80
- Learning Rate = 0.01
- Data was fed into the model in batches (BATCH_SIZE=4) by using a data generator.
- number of iterations = 2000

- reg_noise_std = 0.03
- thresh_v = 0.00001 (loss function threshold)
- n_layer = 5 ($\log_2 32$)

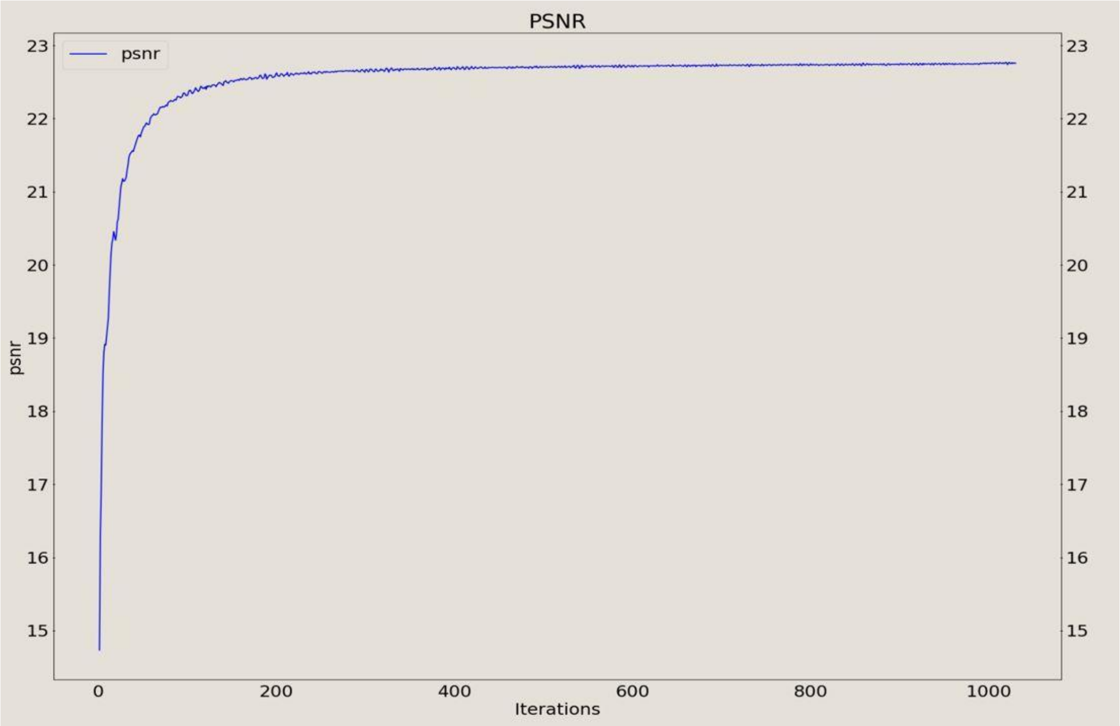
The evaluation of the fusion process regarded the following metrics:

- the Peak Signal to Noise Ratio (PSNR), an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation. The higher the PSNR, the better a degraded image has been reconstructed to match the original image and the better the reconstructive algorithm performed. This would occur in the process of minimizing the MSE between images with respect to the maximum signal value of the image.
- the Mean Squared Error (MSE) or else the optimisation error, as a loss function measure during the training process, comparing the output of the GDD with the Degraded HS image input
- the Root Mean Square Error (RMSE), or else an independent data validation error, which is obtained by the comparison of the fused image of zero shot learning, with the original HSI at 30 m spatial resolution (HR-HS image). RMSE can be evaluated against the ground truth image bands statistics, mean and standard deviation, of reflectance values, to gain insight about the relative error of the fusion reconstruction in regards with the HR-HS image dynamic range.
- the Structural Similarity Index (SSIM), another validation error measure for the evaluation of the similarity of the fused image of zero shot learning, compared with the original HSI at 30 m spatial resolution. In this aspect, image degradation is considered as the change of perception in structural information, with a given emphasis on the spatially close pixels. It also collaborates with some other important perception-based facts such as luminance masking and contrast masking.

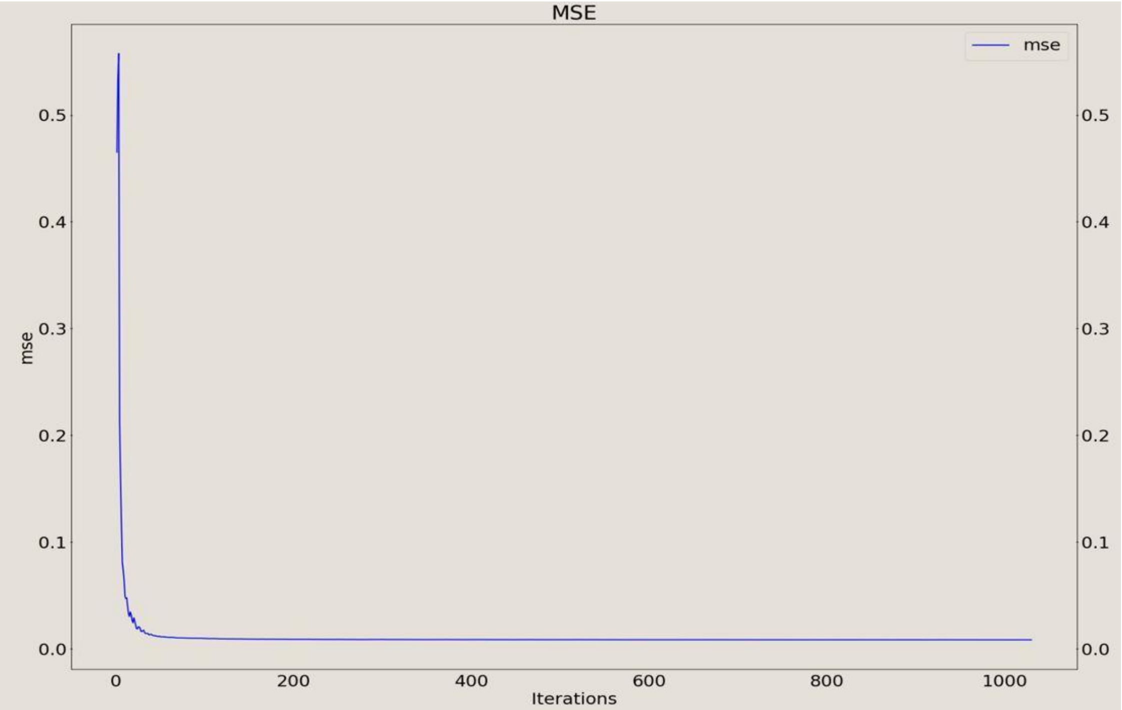
The experimental results from this initial testing of the GDD algorithm for HS-MS image fusion with a zero-shot learning scheme, provided evaluation for reconstructing a degraded HS image to its original spatial resolution, with the aid of a degraded MS guidance image. The performance of the task, still after a lot of algorithm parameter tweaking, was disappointing. The evaluation metrics history along training iterations is shown on Figure 45 while their final values were:

PSNR=22.755
MSE=0.0084
RMSE = 0.1454
SSIM = 0.5823

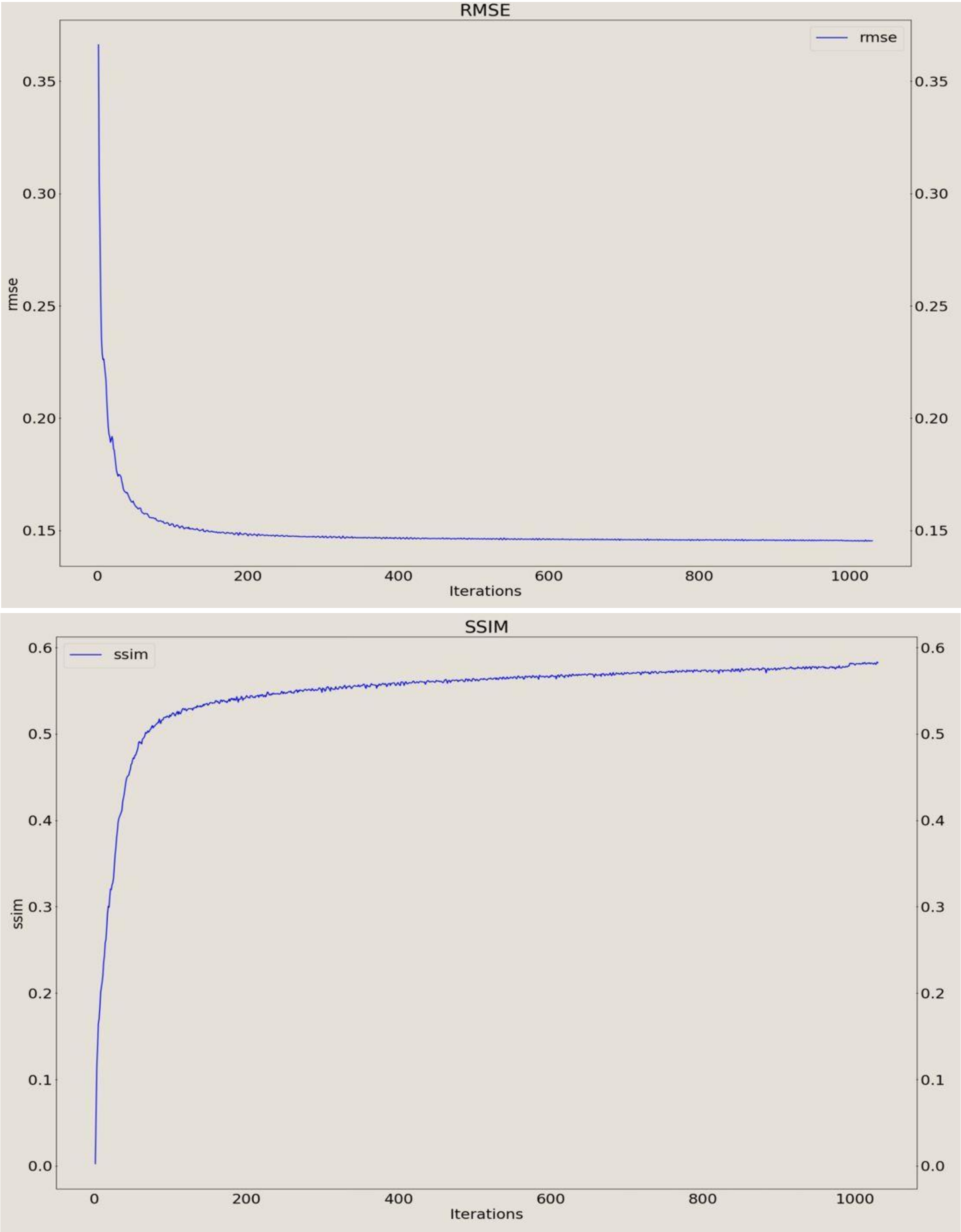
Loss optimization converged quickly to a minimum, at mere 1000 iterations, but still the HS image didn’t achieve a faithful reconstruction while it was mainly dominated by noise, something evident by the low PSNR values. Structural similarity between result and ground truth was equally low, while the perceived error was high relatively to the HS reflectance statistical distribution.


PSNR
psnr
psnr
23
22
21
20
19
18
17
16
15
0
200
400
600
800
1000
Iterations

(a)


MSE
mse
mse
0.5
0.4
0.3
0.2
0.1
0.0
0
200
400
600
800
1000
Iterations

(b)



(c)

(d)

Figure 45. Evaluation metrics of GDD image fusion, at the first take of experiment parameterisation and training. PSNR (a), MSE (b), RSME (c), SSIM (d)

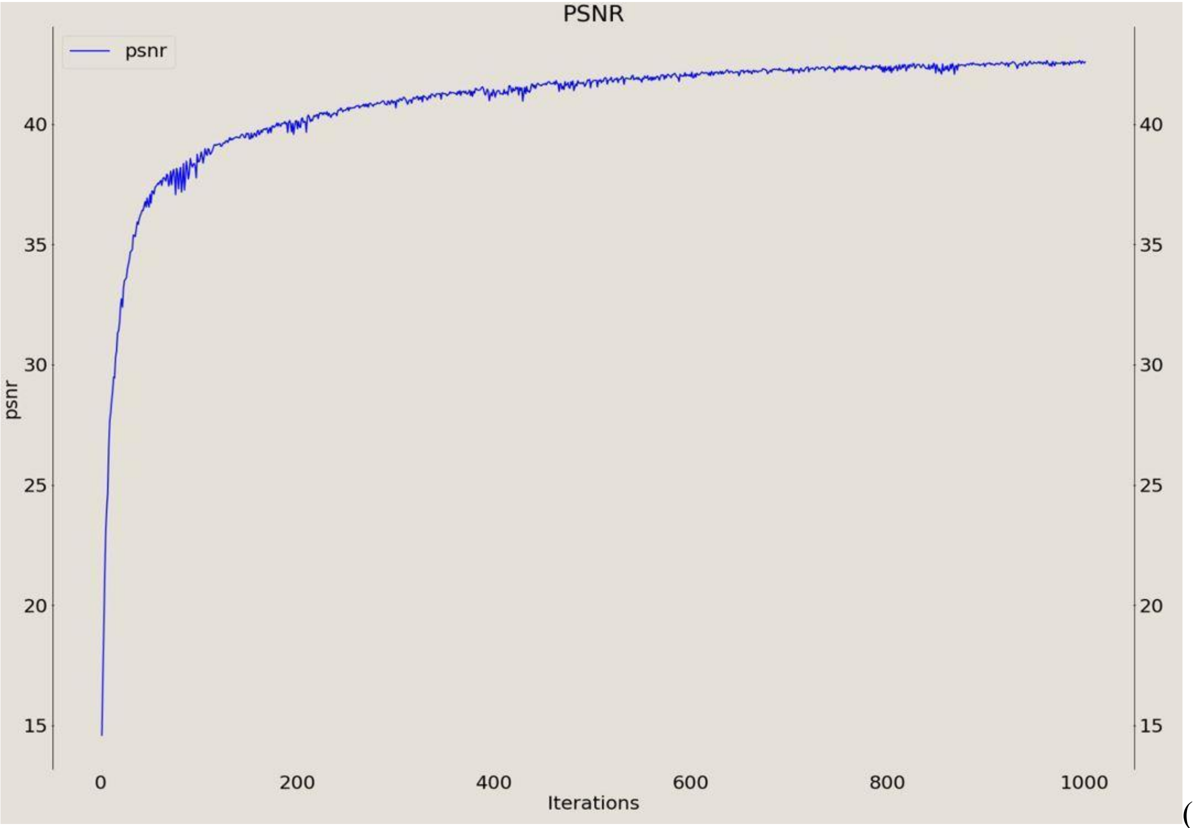
On a second iteration of the fusion experiment, no modifications were made to the model parameters, but evidence from the former evaluation suggested a slight tweaking on the calculation of the Loss function. Providing a strong emphasis on the first term the Loss L. Immediately, the first attempts on this 2nd take of GDD training showed a great improvement on the HSI reconstruction. Evaluation results of this second take of training were:

PSNR=42.591
MSE=1.34 * 10-6
RMSE = 0.0147
SSIM = 0.9663

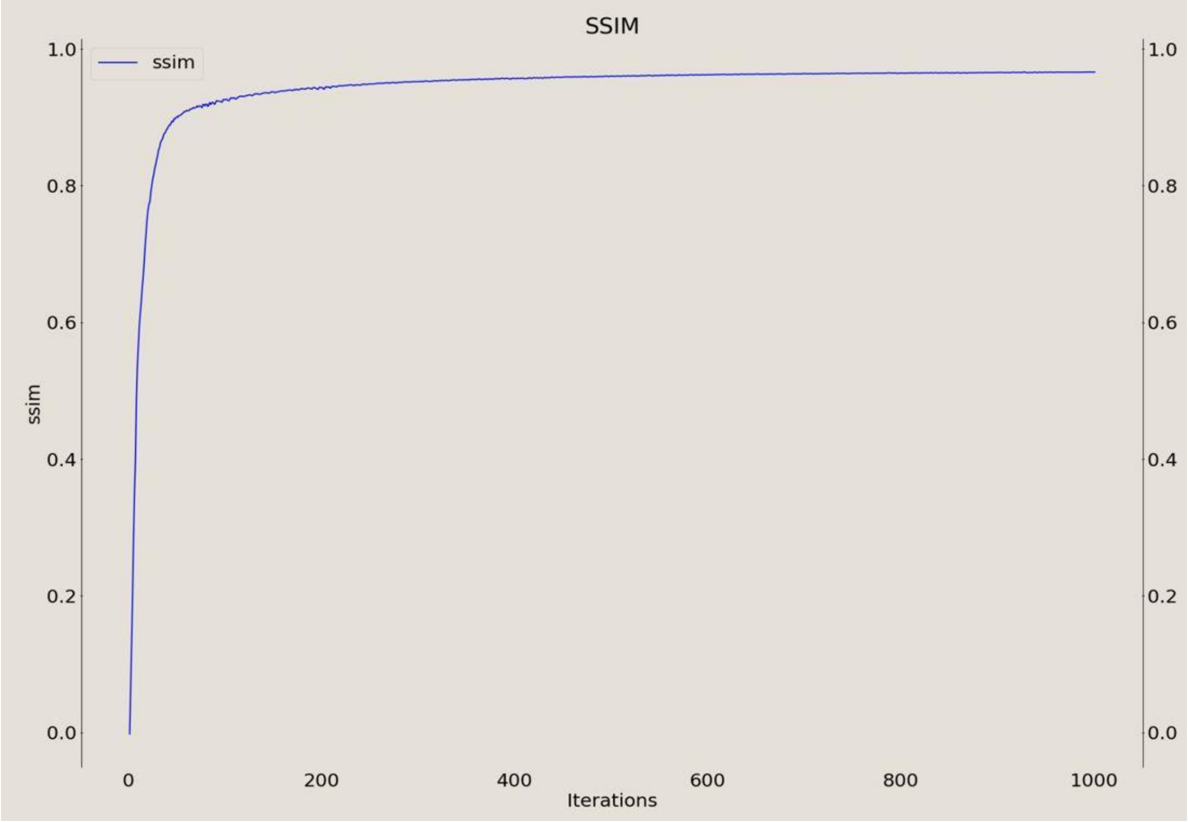
while the metrics history along training iterations is shown on Figure 46. The performance of the deep network was still considered suboptimal, especially in regards with the validation RMSE, which had a relatively high value compared with the standard deviation of the reflectance values, along ground truth HS image bands. On the following table, the Fused HS image bands statistics are presented, aggregated on contiguous spectral regions of visible (RGB), Near InfraRed (NIR and NIR-SWIR), and Shortwave InfraRed (SWIR1, SWIR2), in comparison with the band statistics of the input HS PRISMA image. As concluded from the HS band statistics summary, the evaluation RMSE was considered high, especially in regards with the dynamic range of reflectance values at the Visible HS bands.

Table 25. HS image aggregated bands statistics.

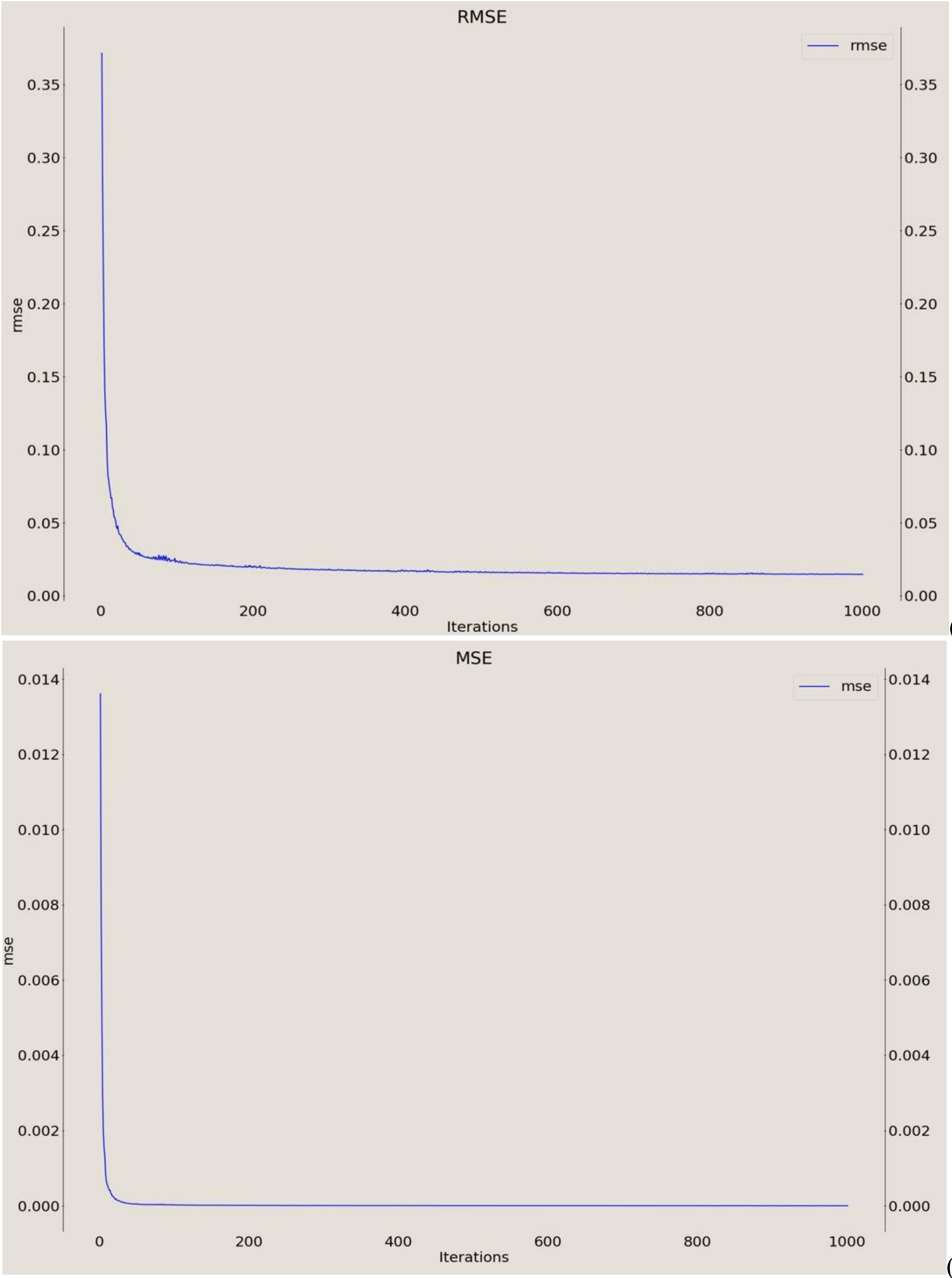
	FUSED HSI-MSI		**PRISMA HSI**	
HS Reflectance Bands Aggregation (nm)	Mean	Stdv	Mean	Stdv
BLUE (400-500)	0.026	0.011	0.045	0.019
GREEN (500-600)	0.068	0.019	0.084	0.027
RED (600-700)	0.118	0.031	0.112	0.037
RED EDGE (700-800)	0.385	0.054	0.298	0.041
NIR (800-1200)	0.506	0.069	0.368	0.054
NIR-SWIR (1200-1350)	0.565	0.080	0.389	0.060
SWIR1 (1470-1780)	0.454	0.071	0.294	0.055
SWIR2 (2000-2380)	0.332	0.055	0.209	0.051


PSNR
psnr
psnr
15
20
25
30
35
40
0
200
400
600
800
1000
Iterations

(a)


SSIM
ssim
ssim
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
Iterations

(b)



(c)

(d)

Figure 46. Evaluation metrics of GDD image fusion, at the second take of experiment parameterisation and training. PSNR (a), SSIM (b), RSME (c),MSE (d)

At the prediction stage of the fusion process, HS-MS images at their original spatial resolutions were used as inputs for the inference of the HR-HS image at 7.5 m ground sampling distance. The prediction results of the HS-MS Fusion, for an image patch from the experimental case study, is depicted in Figure 47(4).

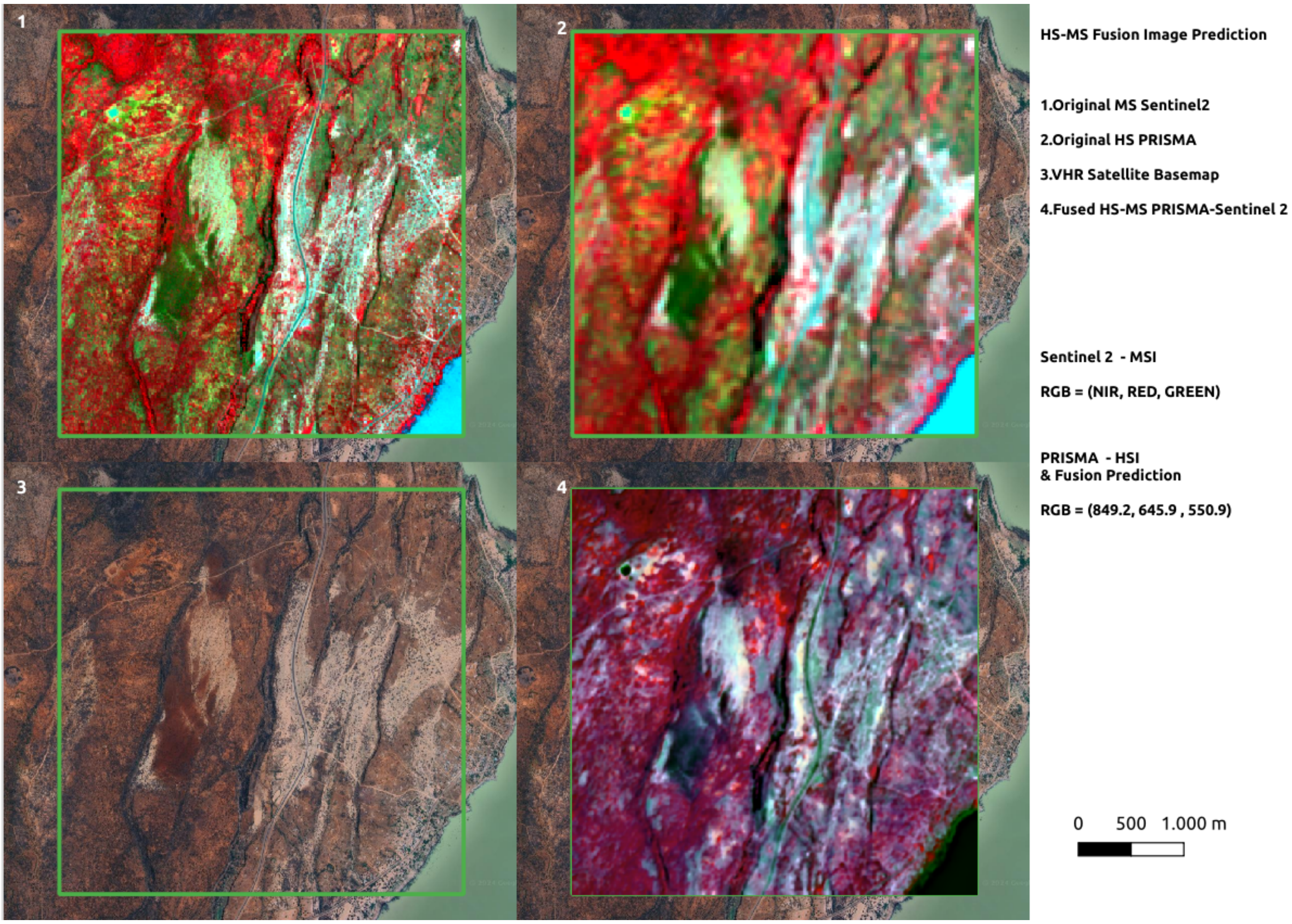
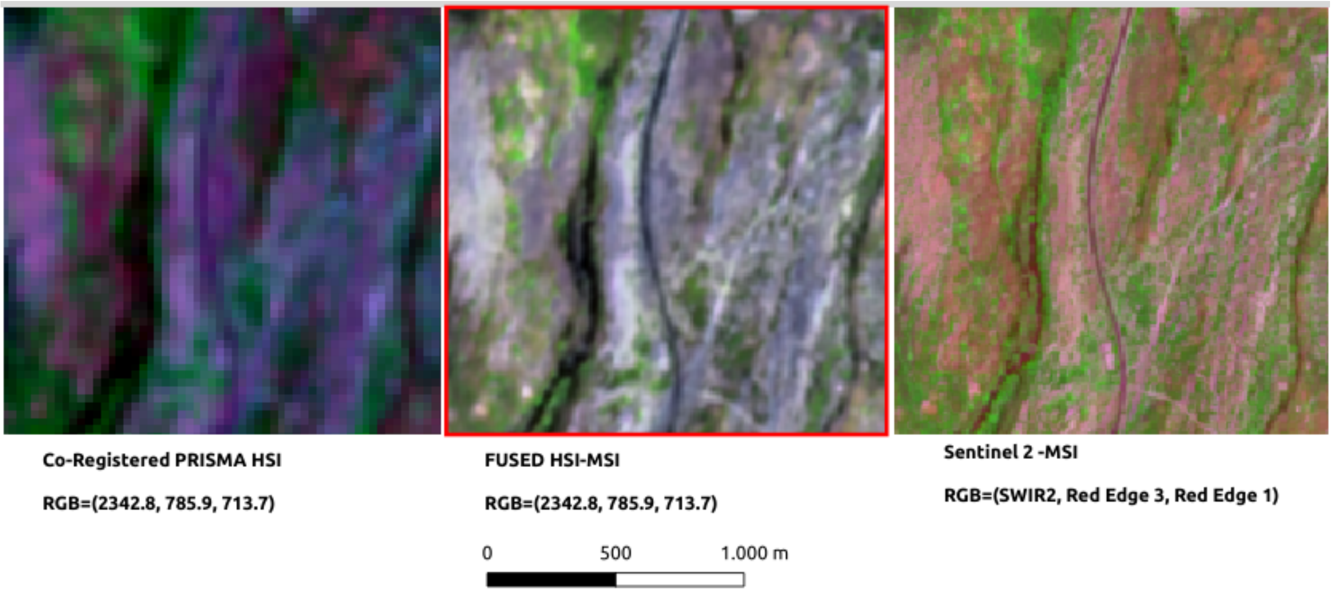


Figure 47. Image patch data for GDD prediction.(1) Multispectral Sentinel 2, (2) CoRegistered Hyperspectral PRISMA, (3) VHR Satellite Basemap, (4) HS-MS image fusion output.

An observable interpretation which is supported as well by the reflectance statistics of Table 25 can be derived by Figure 48, which shows RGB composites at the SWIR and Red Edge part of the spectrum, from the input HS and MS imagery in a visual comparison with the Fused HS-MS image reconstruction. Visible bands show a spectral distortion with lower reflectance values, while the Red Edge, NIR and SWIR regions are overestimated, with higher reflectance values in general. In regards with the spatial detail reconstruction, the fused image has embodied a great deal from the spatial features of the high resolution MS image, showcasing however a blurring effect, present probably because of the bilinear interpolation applied on the upsampling 3x3 convolutions of the GDD encoder -decoder network.


Co-Registered PRISMA HSI

RGB=(2342.8, 785.9, 713.7)

FUSED HSI-MSI

RGB=(2342.8, 785.9, 713.7)

Sentinel 2 -MSI

RGB=(SWIR2, Red Edge 3, Red Edge 1)

Figure 48. RGB composite representations, from the input HS and MS imagery in a visual comparison with the Fused HS-MS image reconstruction, at the SWIR and Red Edge part of the spectrum.

Issues and Next Steps:

In conclusion, the evaluation from the HS-MS image fusion experiments performed on PRISMA-Sentinel 2 data that were obtained from the case study area of interest, indicated that some further modification of the Deep Network GDD architecture originally developed, is required for the enhancement of its performance. The aspects of both spectral and spatial similarity between reference ground truth and the HS-MS fused images, is obviously improvable, by a reduction of the RMSE to acceptable levels, and an increase of the fidelity of the SSIM.

- A first step is the continuation of testing and tuning on the existing GDD architecture, with targeted modifications to the calculation of the Loss function, as well as to the values of associated hyper parameters, in order to improve the values of the evaluation metrics.
- In a next step, an assessment of the spectral content of the channels of the resulting HS-MS Fused image is legitimate, by obtaining spectral signatures from defined land cover targets, and a comparison between primary data HS-MS data.
- A final important step that follows the previous ones, once the results of the process are improved and become significant, is the configuration of the code for training and prediction of the GDD model with parallel processing for the purpose of using it in HS-MS Fusion applications on whole image tiles.

4.6.2.1. Deep Learning for HS-MS Image Fusion - Update iteration

During the evaluation of the accuracy achieved by the HS-MS Fusion method using neural networks, the need for further modifications was identified in order to obtain improved results, mainly in the part of the spectral fidelity with which the radiometric reflection is reproduced.It was decided that the modification should not only concern the tuning of the hyperparameters of the model, since this was already tested beforehand with very small improvements in the final result, but moreover to include changes in the Loss Function of the Guided Deep Decoder. Thus the original relationship was modified to

$$L(X,Y,G) = \mu \cdot \|\hat{X} \cdot S - \hat{Y}\|_F^2 + \sigma \cdot \|R \cdot \hat{X} - \hat{G}\|_F^2$$

allowing parameterization of the relative weighting between the two terms related to spectral and spatial similarity. A range of values m and σ were then tested to evaluate the optimal combination that minimizes the rmse error in test validation.

During the hyperparameter tuning of the GDD the following parameters, provided optimal error:

- factor = 4 (spatial degradation factor)
- param_balance μ = 0.62
- param_balance σ = 0.34
- number of network channels = 80
- Learning Rate = 0.01
- Data was fed into the model in batches (BATCH_SIZE=4) by using a data generator.
- number of iterations = 10000
- reg_noise_std = 0.08
- thresh_v = 0.00001 (loss function threshold)
- n_layer = 5 ($\log_2 32$)

There was a slight improvement in the global evaluation metrics, which were as follows :

PSNR=44.83
MSE=1.02 * 10-6
RMSE = 0.0108
SSIM = 0.9841

At the prediction stage of the fusion process, HS-MS images at their original spatial resolutions were used as inputs for the inference of the HR-HS image at 7.5 m ground sampling distance. The prediction results of the HS-MS Fusion, for an image patch from the experimental case study, is depicted in Figures XXX25 and XXX26 using a pseudocolor composite representation at different parts of the VIS - NIR - SWIR spectrum, in comparison with the initial fusion results that emerged before the iteration updates to the GDD.

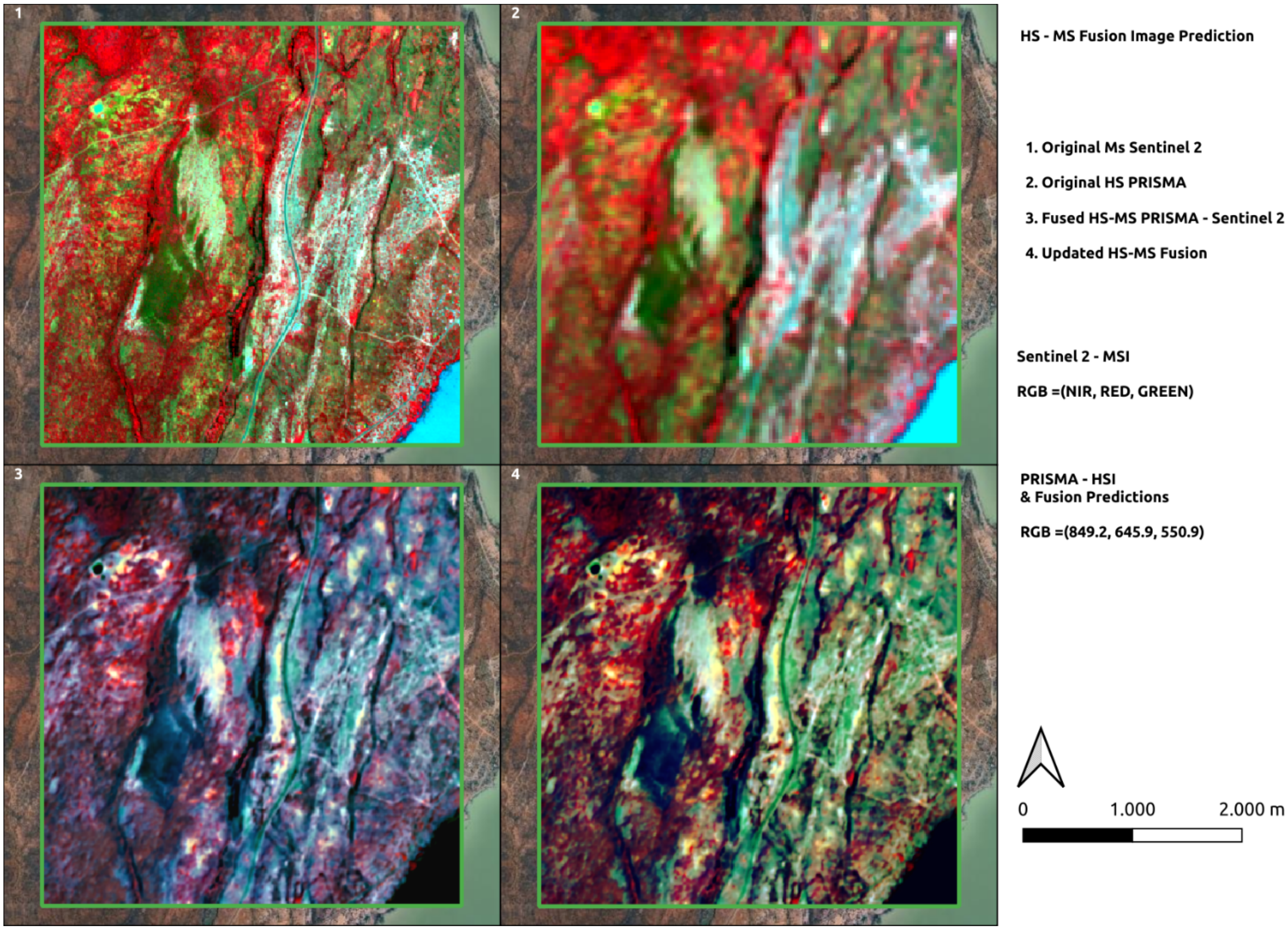


Figure XXX25. Image patch data for GDD prediction of 2nd iteration. Pseudocolor composite rendering using NIR, RED, Green channels for RGB.(1) Multispectral Sentinel 2, (2) CoRegistered Hyperspectral PRISMA, (3) HS-MS image fusion output 1st iteration, (4) HS-MS image fusion output 2nd iteration.

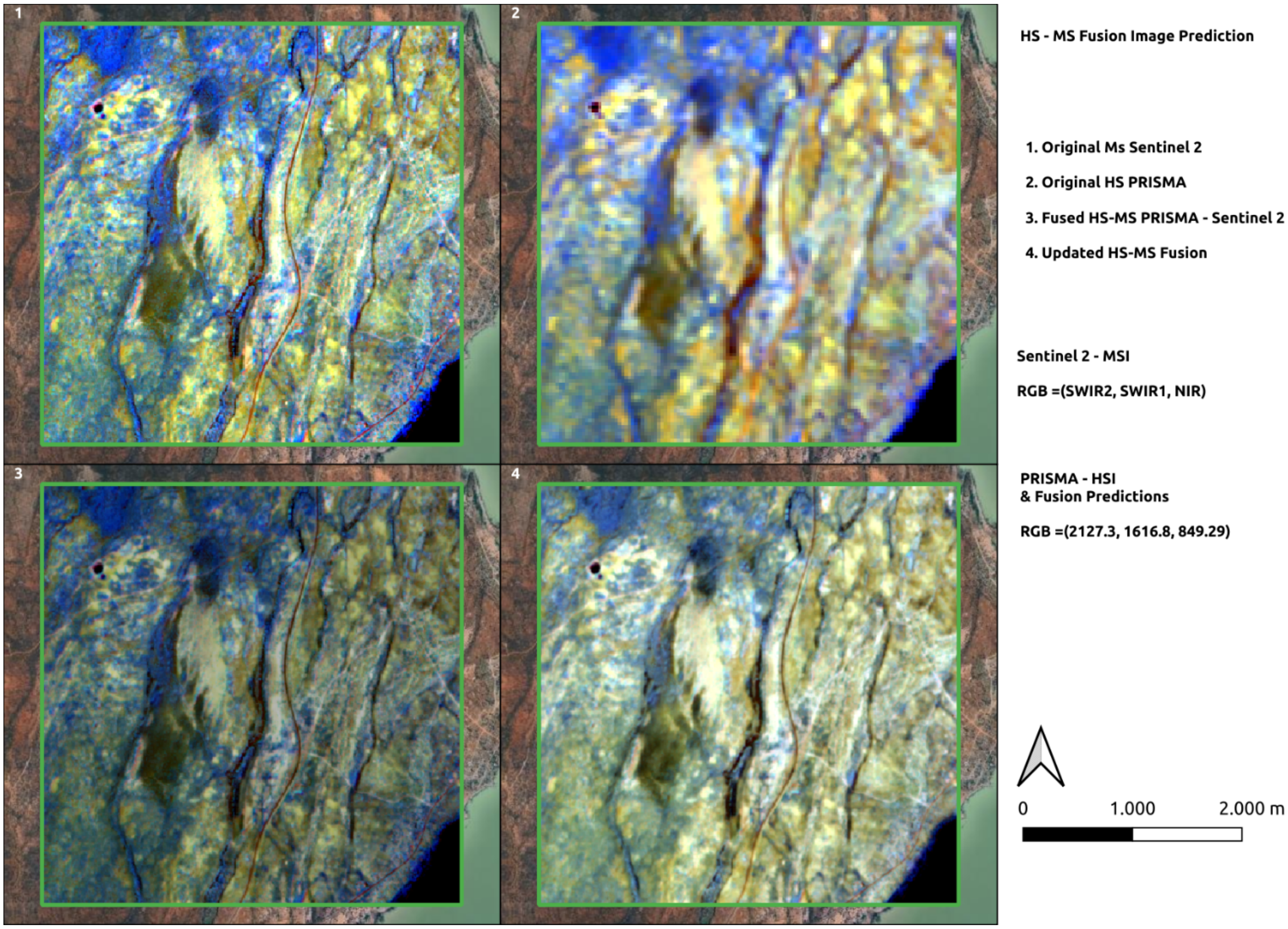


Figure XXX26. Image patch data for GDD prediction of 2nd iteration. Pseudocolor composite rendering using SWIR2,SWIR1,NIR channels for RGB.(1) Multispectral Sentinel 2, (2) CoRegistered Hyperspectral PRISMA, (3) HS-MS image fusion output 1st iteration, (4) HS-MS image fusion output 2nd iteration.

Spectral Signatures

In relation to the above evaluation of the performance of the HSI - MSI fusion process on the data, the global metrics obtained as well as the qualitative inspection of the result on RGB pseudocolor composites provide a limited overview of the reconstruction fidelity. It was considered important at this point to further investigate the performance of the data fusion on the individual spectral channels in order to draw some conclusions about the usefulness of the resulting product in remote sensing mapping applications. It is a fact that the fusion process introduces spectral distortions and it is important to see in which parts of the spectrum these occur, to what extent they are significant, and how they may be varied in the spectral targets of the earth's surface.

Initially, the RMSE error obtained in the zero shot validation was calculated per spectral band. To allow comparison between channels, the normalized mode was calculated in terms of the standard deviation of each channel. Thus, for example, normalized RMSE values in the range [0.25,0.5] indicate that the active range of reflectance values in the channel in question is, double to quadruple the range of the error, and this means that the spectral reconstruction is quite satisfactory. Normalized RMSE values in the range (0.5,0.75) suggest that the active range of the reflectance values in this channel is 1.3 to 2 times larger relative to the range of the standard deviation of the error, which characterizes the performance in this channel as good to satisfactory. In the lower limit, the performance is therefore satisfactory, while in the upper limit it is considered moderate. Finally for Normalized RMSE values greater than 0.75 the range of error tends to the same range of image reflectance values, which indicates a small signal to noise ratio and unsatisfactory image fusion performance for this channel.

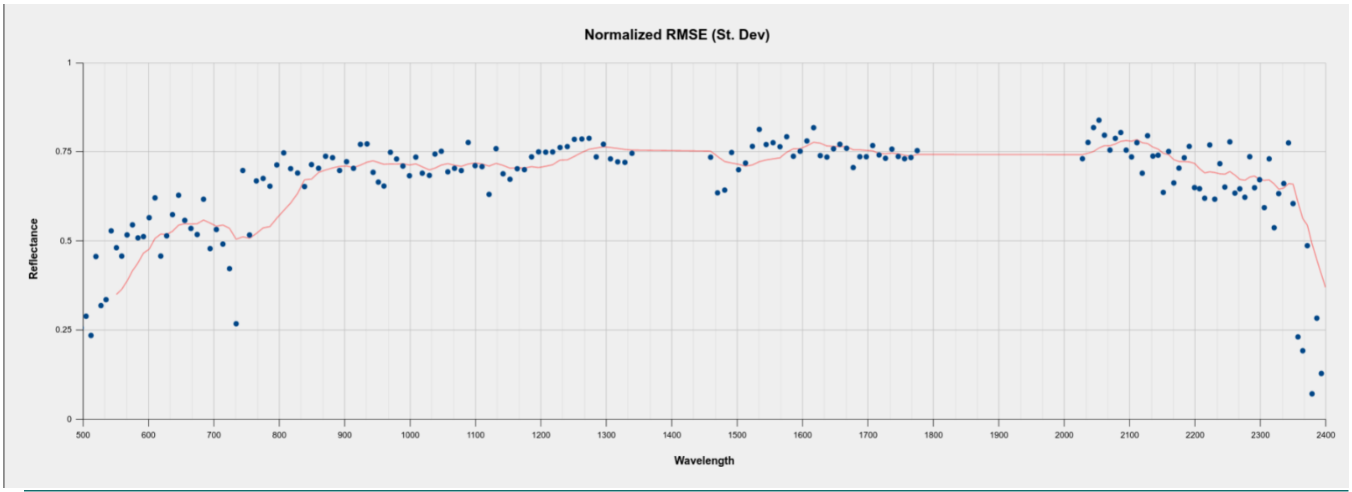


Figure XXX27. Normalized RMSE of HS-MS Fusion across spectral channels in BOA Reflectance units

The outcome of this analysis is given in figure (XXX27), which shows the distribution of Normalized RMSE in the hyperspectral channels. It is obvious that the performance is not homogeneous, but shows different qualitative characteristics in the different regions of the spectrum. For example, the HSI- MSI fusion seems to perform quite well in the green, red, red edge regions, to be judged satisfactory in the NIR and outermost SWIR2 regions, while in the intermediate region (1000 - 2100 nm) the performance is judged moderate. A possible explanation for the reasons for this difference in Normalized RMSE could be the spectral coverage provided by the multispectral data of Sentinel 2 MSI, as the 900 - 1600 nm region is missing, as well as the relatively larger error introduced by the Sentinel 2 band fusion preprocessing at 10m spatial resolution, in the SWIR1 channel (1610 nm) compared to SWIR 2 (2190 nm).

On the level of hyperspectral remote sensing applications, the result of the HSI-MSI fusion could be used with relative safety in the calculation of vegetation radiometric indices in regions where the Normalized RMSE is small. Applications involving Carotenoid and Anthocyanin Content mapping focus on the green and red edge regions of the spectrum, and to this extent, the hyperspectral data resulting from the fusion can be useful, with reasonable accuracy. This is particularly true for indices such as CRI (Carotenoid Reflectance Index), which is calculated exclusively in narrow bands within the green region, and where the use of multispectral data is not sufficient in terms of spectral resolution. The same conclusion can be drawn for leaf/canopy chlorophyll content estimation, and Dry Matter Pigment mapping, where the radiometric indices focus on the green, red and red edge regions. The applications related to the properties of the leaf/canopy structure further involve the NIR region, which has a higher error. Thus the calculation of radiometric indices , with the typical example of the well-known NDVI varieties would be better done directly from the multispectral Sentinel 2 image. Finally applications involving Leaf Water Content mapping use radiometric indices in the NIR and SWIR regions and due to the unsatisfactory performance of fusion in these channels, it would be advisable to avoid the use of fused hyperspectral products.

In a more in-depth quality control, it was considered appropriate to study the spectral signatures presented by targets of interest from the Earth's surface, to determine the divergence they present between the primary hyperspectral image and the final product of the HSI - MSI fusion, for all their channels. In the absence of ground truth field data on the type of land cover in the image under study an unsupervised classification approach was followed to identify classes of interest after using photo interpretation. Principal Component Analysis (PCA) was applied to the PRISMA hyperspectral image for a linear dimensionality reduction using Singular Value Decomposition of the data to project it to a lower dimensional space. The input data is centered and scaled to its standard deviation, for each feature before applying the SVD.

The selection of Principal Components for unsupervised analysis was based on the explained variance criterion that they cumulatively contribute. According to figure XXX28 that follows below, it was found that the use of the first three PCs, in the case of the studied image, explains 98.6 of the observed dispersion of values, condensing the set of useful spectral information, while as found all the rest PCs depict large noise contributions to the data.

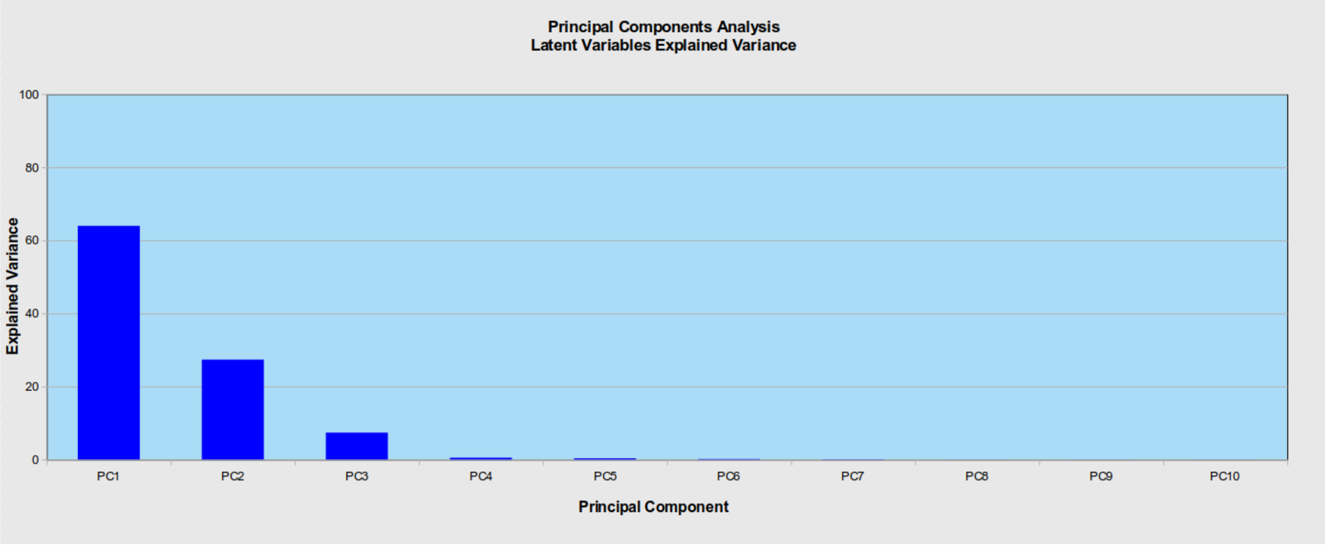


Figure XXX28. Principal Components Analysis of the PRISMA HS image. Explained Variance of the first 10 latent variables (PCs)

The classification of the PRISMA hyperspectral image into land cover classes, in an unsupervised manner, was performed using the k-means clustering method. The adjustable hyperparameter was the number k of optimally separable clusters, which was tuned through the use of silhouette analysis. Silhouette analysis can be used to study the separation distance between the resulting clusters. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters and thus provides a way to assess parameters like number of clusters visually. This measure has a range of [-1, 1]. Silhouette coefficients (as these values are referred to as) near +1 indicate that the sample is far away from the neighboring clusters. A value of 0 indicates that the sample is on or very close to the decision boundary between two neighboring clusters and negative values indicate that those samples might have been assigned to the wrong cluster.

As a result , the classification into 8 unsupervised classes was implemented, for which a label was then given through photo interpretation, using the very high resolution google basemap. Among others, the following classes of interest were recognized: :

- Bright Soil
- Mixed Bright Soil and Artificial Surfaces
- Dark Brown Soil
- Dense Vegetation
- Mixed Soil and Vegetation
- Mixed Barren and Broadleaf vegetation

Accordingly, sampling locations were selected to obtain the spectral profiles for the above classes from the 2 images under study, the PRISMA and the fused hyperspectral image. In the figures that follow (XXX28 to XXX31) the spectral profiles obtained for the most important classes of interest are presented in comparison. The spectral curve of the raw PRISMA image is shown in red and that of the HSI-MSI fused image in green. The y-axis shows the reflectance values, and the x-axis the spectral channels, in ascending number, Visible (bands 1-31), Red Edge (bands 32 - 40), Near InfraRed (bands 41 - 70), Short Wave Infrared (bands 107 - 173). According to the observed pattern, the same conclusions as in the Normalized RMSE per band analysis hold true. In addition, it is clear that there is a difference in the observed deviation between the 2 HS images, that depends on the type of spectral target under consideration. Spectral targets with mixed land cover appear to show smaller discrepancies between the PRISMA raw acquisition and the HSI-MSI Fused image,which can be considered expected given the smoothing effect observed during the fusion process and the larger pixel size of the PRISMA image.

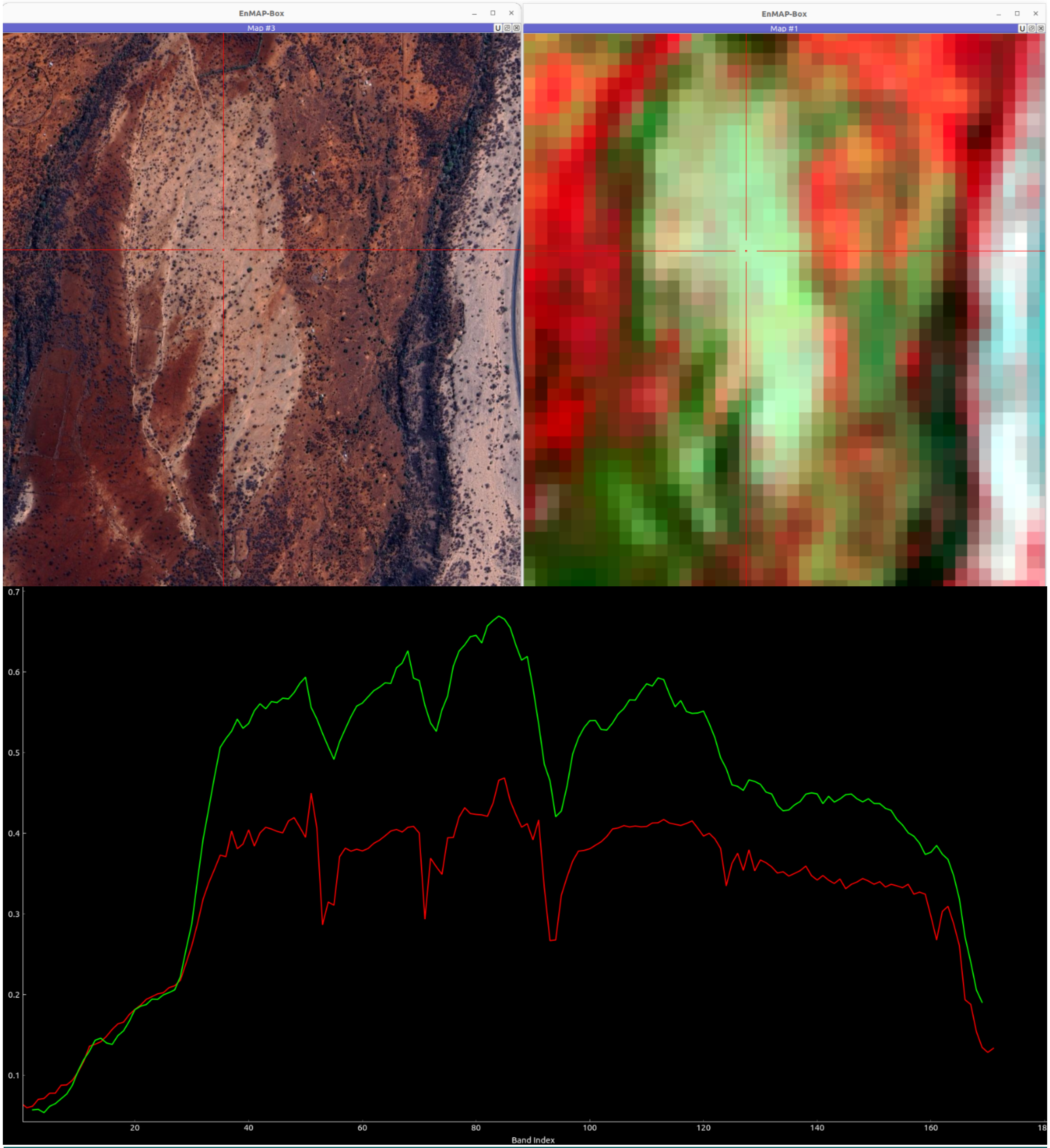


Figure XXX28. Spectral Profile curves comparison between PRISMA (red) and the Fused HS-MS image (green), on High Albedo Soil ground target. The spectral sampling location on Google Basemap and the PRISMA imagery.

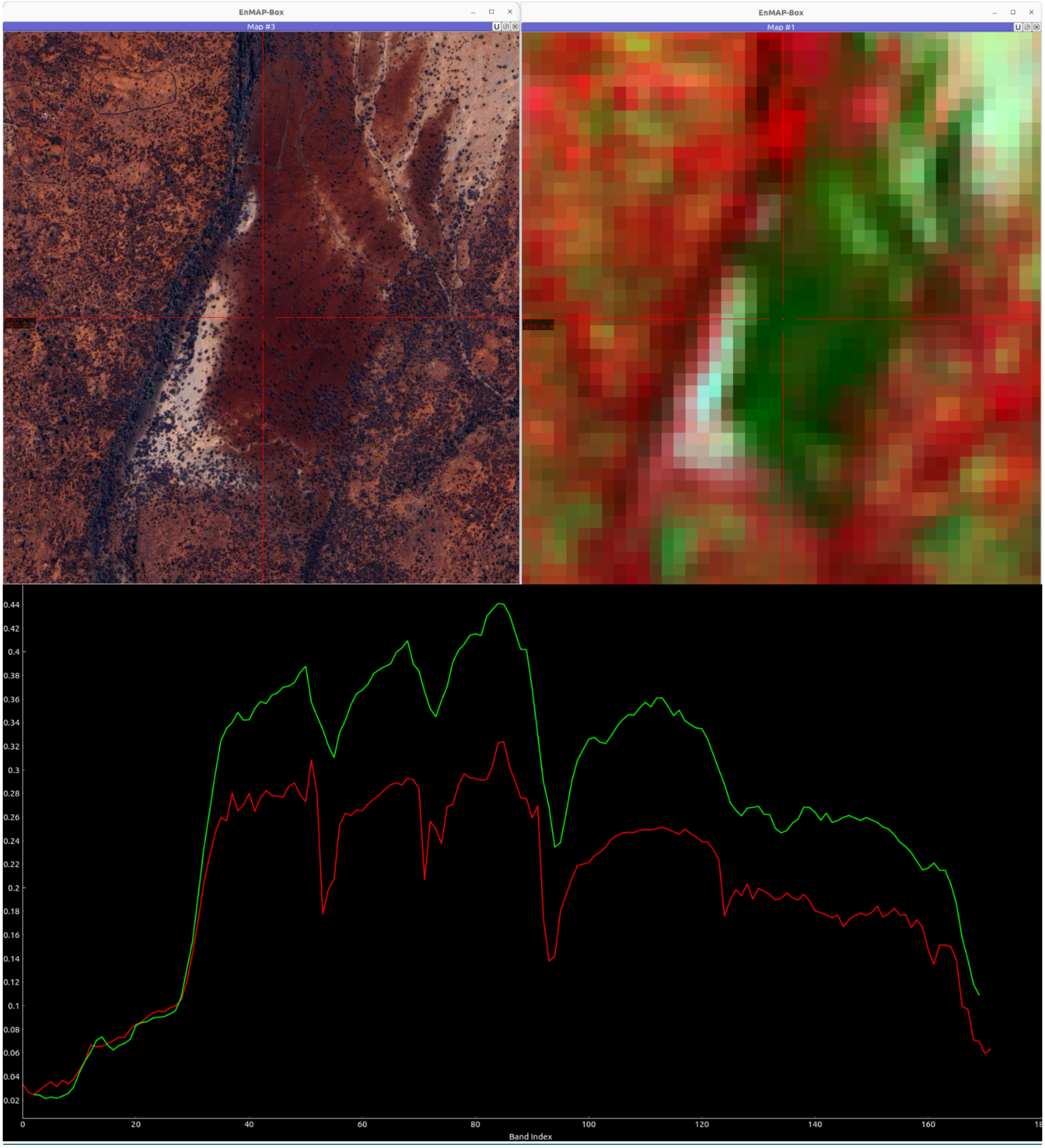


Figure XXX29. Spectral Profile curves comparison between PRISMA (red) and the Fused HS-MS image (green), on Low Albedo Soil ground target. The spectral sampling location on Google Basemap and the PRISMA imagery.

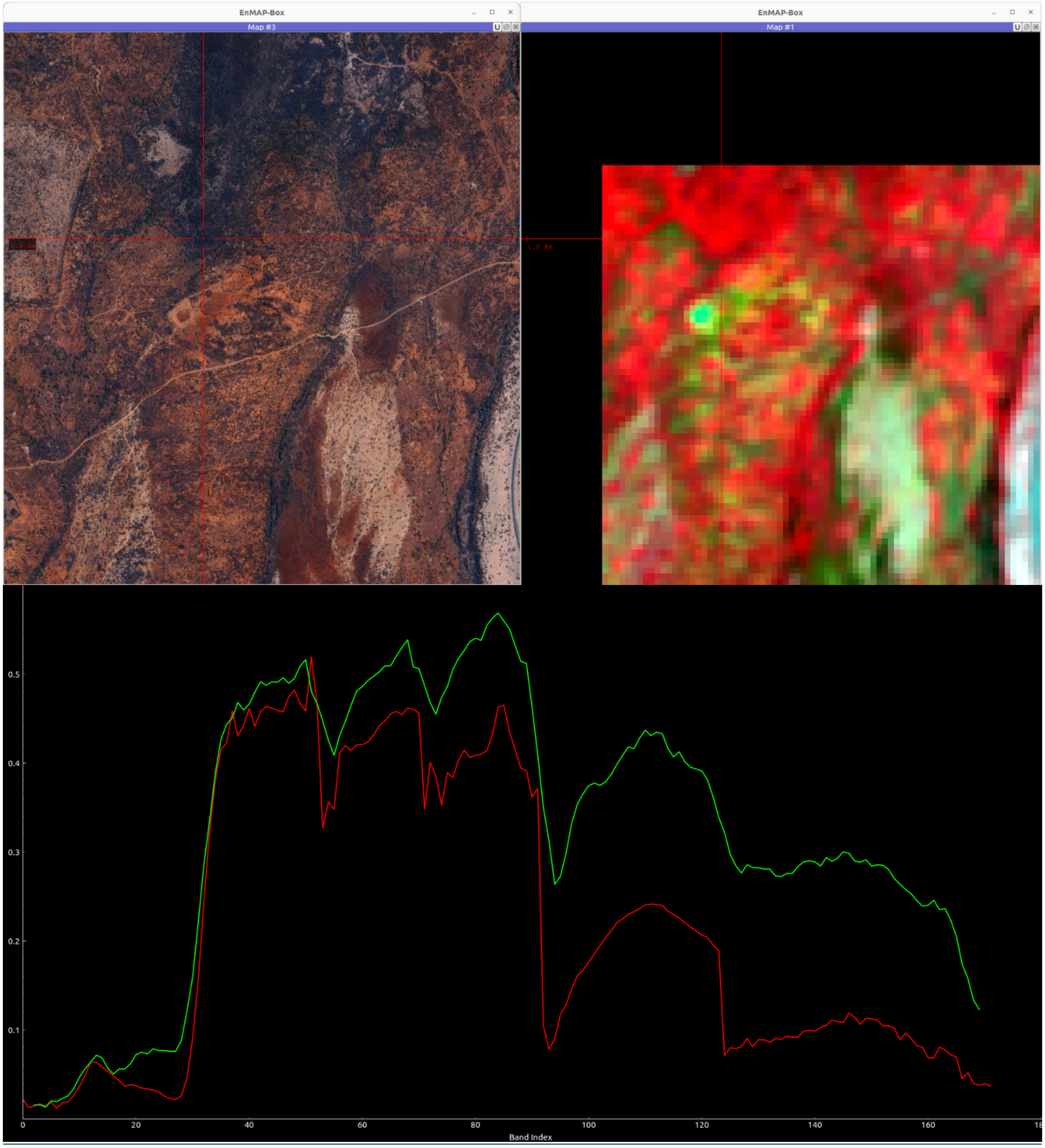


Figure XXX30. Spectral Profile curves comparison between PRISMA (red) and the Fused HS-MS image (green), on Dense Vegetation ground target. The spectral sampling location on Google Basemap and the PRISMA imagery.

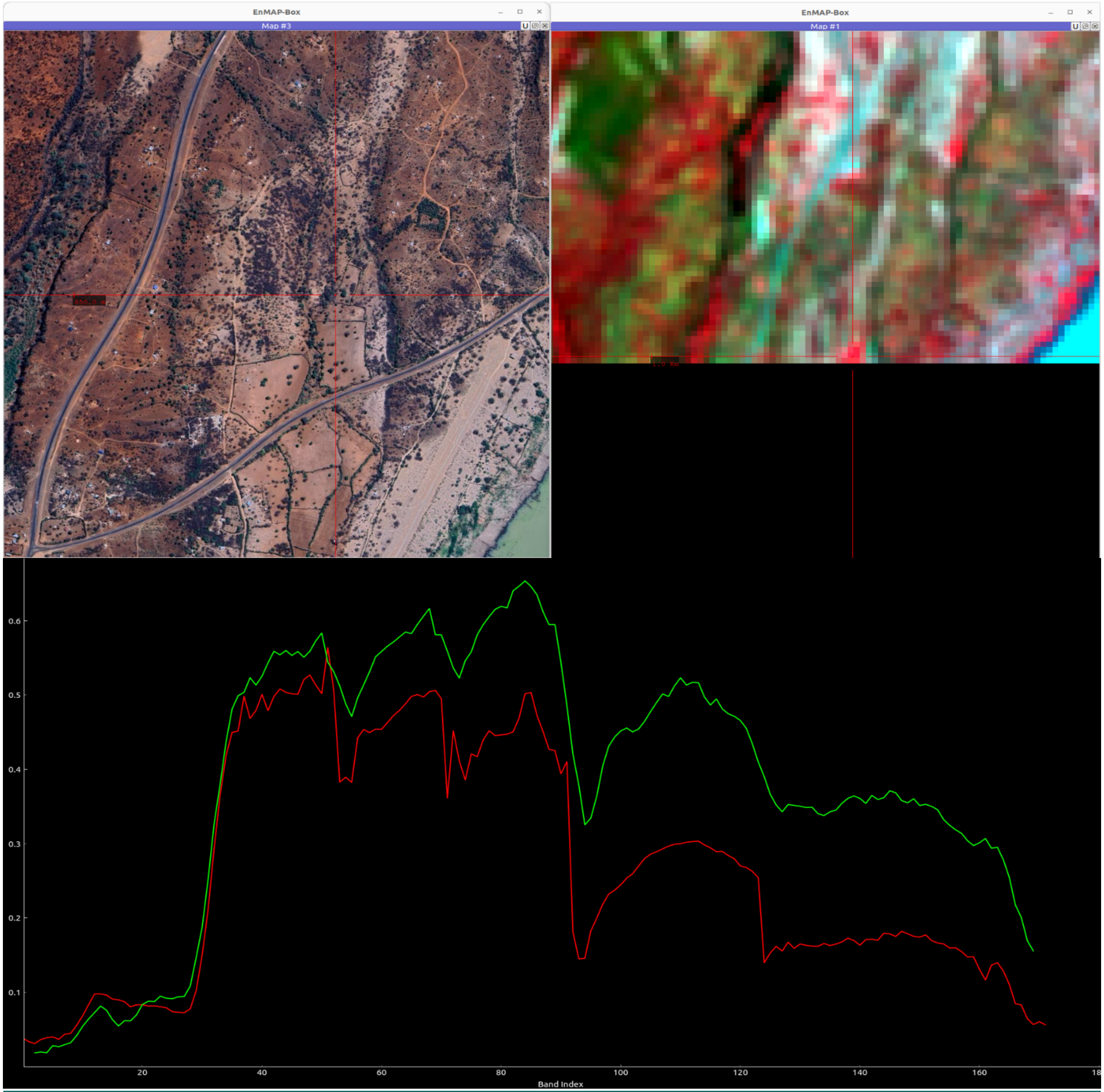


Figure XXX31. Spectral Profile curves comparison between PRISMA (red) and the Fused HS-MS image (green), on Mixed Soil and Broad leave Vegetation ground target. The spectral sampling location on Google Basemap and the PRISMA imagery.

From the evaluation of the spectral profiles, the following conclusions can be drawn regarding the usefulness of the fused product in agricultural applications :

- The potential that the HSI-MSI Fused image can offer for quantitative mapping of biophysical parameters of cultivation in high resolution, through the use of PROSAIL inversion models, does not seem to be particularly realistic. This is concluded from the relatively large deviation that the curves show in their overall spectral dimension. So in the field of crop traits, the most appropriate solution is the use of radiometric vegetation indices described above and their reclassification into quality classes (high, mid, low) based on empirical value limits that characterize them.

- From the point of view of thematic mapping applications, such as image segmentation or even supervised classification of land cover and crops at object or pixel level, the prospects of HSI-MSI Fused data could be more promising since the greater spectral resolution and higher spatial resolution are still advantages, which make their use more efficient than multispectral in terms of separability of spectral signatures. The above hypothesis, in order to be validated, needs to be established through classification benchmarking and the use of reliable ground truth data.

5. Inversion of RTMs for Biophysical Parameters

Objectives:

Radiative Transfer Models (RTMs) have gained considerable attention in several fields, including remote sensing (Kganyago et al., 2022), atmospheric and environmental science. RTMs simulate the interaction of electromagnetic radiation with different mediums, providing valuable information on the physical and optical properties of these mediums. Inversion techniques based on RTMs have a critical role in retrieving these properties from measured data.

Biophysical variables refer to measurable physical properties of living organisms or ecosystems that are used to quantify and understand their biological processes and interactions with the environment. These variables often encompass characteristics such as leaf area index (LAI), vegetation cover, biomass, and canopy structure. Biophysical variables play a crucial role in remote sensing and ecological studies as they provide valuable insights into the health, productivity, and dynamics of vegetation and ecosystems (Elwadie et al., 2005). They are widely used in applications such as vegetation monitoring, land cover classification, carbon sequestration estimation, and climate change studies. RTMs allow to estimate the relationship between spectral reflectance and some vegetation characteristics as:

- Leaf Area Index (LAI) is a parameter used to describe the intrinsic properties of the canopy. It is strictly related in a non-linear way to the reflectance and is, however, unrelated to the observation conditions. (Duan et al., 2014).

5.1. In Situ Data of Biophysical Parameters

Experimental Setting:

In order to be able to test and validate the algorithms and data produced by the project, a case study was defined that included a collection and analysis of ground data. The basic requirements of this data collection were:

- Site location
- Time period
- Data type

Given previous collaborations with some local partners (University of Nairobi and ESIPPS), Kenya and Uganda were selected as the pilot countries for the measurement campaigns, identifying Narok and Ahero as the main areas of interest for Kenyan country (Figure 49), and the eastern part of Uganda. This choice was mainly related to the availability of observation of the crops identified by the project (i.e. wheat, maize and rice) and that these were logistically reachable and not too widely dispersed across the territory.

The time period for measurements is closely linked to the phenological cycle of the crops mentioned, identified between the months of May and August (between sowing and harvesting). Several measurement campaigns were carried out, at different times within the identified time period, in order to be able to acquire data at different phenological phases of the crops.

Different types of data were needed in order to validate and compare satellite data and data products. Mainly we can divide them from a spatial and a physical point of view: the former necessary in order to identify the pixels corresponding to the ground data on the image, the latter in order to compare the data produced by the algorithms developed in Afri4Cast with the physical data acquired in the field.
position, perimeter and classification of the fields;

- LAI measurements;
- chlorophyll content measurements;
- collection of vegetation samples for visual and laboratory analysis.

Three measurement campaigns were carried out in Kenya, in mid-June, mid-July and mid-August of 2023 respectively, in which 151 maize fields, 128 wheat fields and 252 rice fields were mapped and considered for the Crop Growth Model. When available, information about the crop variety, sowing and harvest dates, crop density, presence of irrigation, fertilizer use and final field yield was acquired for each field.

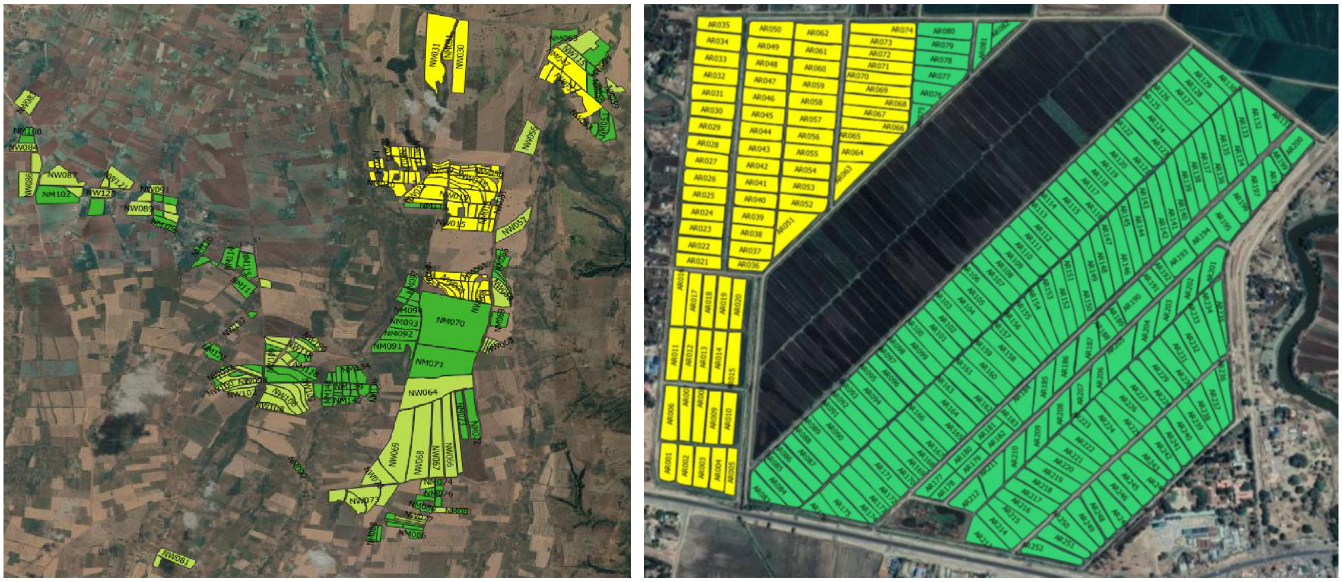


Figure 49. Field campaign sites respectively in Narok (left) and Ahero (right) areas.

As for the measurement campaigns in Uganda (Figure 50), three missions were carried out: between June and September 2023, in July 2023 and at the end of January 2024.

Figure 50. Field campaign sites in the East part of Uganda.

The field measurements performed by the different teams used specific instrumentation, namely: the LI-COR LAI 2200c for LAI measurements and the Dualex for chlorophyll content (Figure 51).
While for the Dualex the measurement procedure is rather simple, it is sufficient to insert a leaf inside the instrument clip to obtain a measurement, for the LAI it will be necessary to follow a more accurate procedure: for each measurement the instrument has to be calibrated against the current lighting conditions.

Figure 51. Instruments used for field campaigns. Respectively: on right, the LI-COR LAI2200c for LAI measurements and on left, the Dualex for chlorophyll content measurements.

In order to be used in the validation of the satellite data, all measurements performed, both instrumental and sample collection, had to meet certain requirements. Given the characteristics of satellite images in terms of pixel spatial resolution, all fields that were considered in the campaign had to have a minimum size of 40x40m, so as to be sure of obtaining pixels completely within the field under examination. Considering also a potential misregistration of the Sentinel-2 image of 1 pixel. This requirement also serves to avoid contamination from objects outside the field, such as roads, buildings, adjacent fields with different crops, etc.

This premise translates into the need to carry out the three different measurements/samples in three spots sufficiently inside the field and not close to the perimeter. Furthermore, these should be acquired in areas as homogeneous and dense as possible. Each of the instruments used, in addition to measuring the physical data, also records the geographical position of the measurement thanks to the built-in GPS.

All the data acquired in the various field campaigns were subsequently reprocessed in order to extract and put together only the information strictly necessary for the project, so that it could be easily used during the analysis and validation phases.

Issues and Next Steps:

Following the measurement campaign, the necessary satellite images were obtained to process and validate the algorithms developed by the project teams. During this process, a major (well known in remote sensing) problem arose: all the images acquired for the two sensors selected for this study (PRISMA and Sentinel-2) turned out to be unusable due to cloud cover. This problem mainly afflicted the Narok area where the data on maize and wheat had been collected (Figure 52).

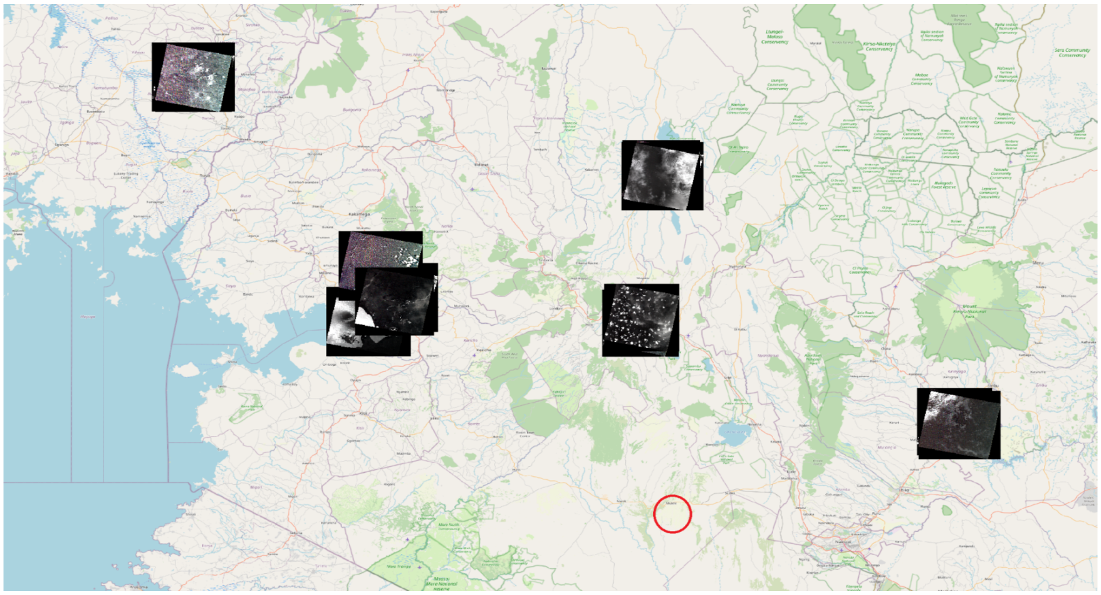


Figure 52. PRISMA acquired images on the AOI. In the red circle the Narok area with no acquisitions.

This led to other solutions being analyzed in order to continue the analysis. Other sensors with similar characteristics to PRISMA and Sentinel-2 were considered that could be used for the extraction of the various indices. Specifically, EnMAP imagery, to compensate for the lack of PRISMA images, and Planet, to compensate for the lack of Sentinel-2 images.

Unfortunately, even for EnMAP no cloud-free images were found in the areas of interest during the period of the ground campaigns, but it was still taken into account for future ones and as support for PRISMA. For this purpose, the project was submitted on the EnMAP platform in order to be able to request specific acquisitions in the areas of interest, and the first images were already obtained.

5.2. PROSAIL to LAI Inversion

PROSAIL, as already mentioned above, is a fusion of two widely used models: PROSPECT and SAIL. PROSPECT provides a biochemical representation of an individual leaf in the spectral range between 400 and 2500 nanometers. This model, which originally included a handful of parameters, has been extended over the years to include more leaf pigments such as chlorophyll a and b (Cab), carotenoids (Ccx), anthocyanins (Canth), brown pigments (Cb), in addition to the leaf water content (Cw) and dry matter content (Cdm) and to the leaf structural parameter (N).

Scattering by Arbitrarily Inclined Leaf (SAIL) is a turbid medium model which allows us to represent the interaction of electro-magnetic radiation with the leaf canopy and the underlying soil represented as infinitely extending 1-D layers. In addition to the parameters which describe the canopy, such as the leaf area index (LAI), and leaf inclination distribution function (LIDF), and to the parameters that describe the sun-viewer geometry (Sun Zenith Angle, Viewer Zenith Angle, Relative Azimuth Angle), SAIL requires the leaf reflectance and transmittance in order to run. These two are provided by the PROSPECT model as outputs, which is what allows the two models to run as a combination. In addition, SAIL also allows to freely model the underlying soil layer. This can be done either by using a custom soil model (i.e., by providing directly the soil reflectance spectrum ρsoil in the range 400-2500 nm), or by mixing pre-loaded dry and wet soil spectra using a soil brightness (αsoil) and a soil dryness parameter (psoil).

With up to 16 input parameters, PROSAIL calculates the bidirectional reflectance of the canopy in the previously described range, in 1 nm increments, considering biophysical and biochemical parameters of the plant and geometric factors related to the observation mode. It has been widely used in forward mode, where canopy reflectance is calculated with various input parameter combinations for different research objectives. Studies employing PROSAIL in forward mode have explored sensitivity to factors like red edge position (REP) of vegetation, leaf optical properties, observation geometry, canopy architecture, soil background reflectance, and atmospheric conditions. The model has also been leveraged for designing new vegetation indices. Its usage in inverse mode, i.e., to infer the value of biophysical parameters from observed reflectances, is rarer but of great interest. Nevertheless, the inversion problem is heavily ill-posed and implies significant challenges both in terms of performance and in terms of model parametrization.

Challenge:

The core idea is based on a coupling scheme of PROSAIL as a well-known RTM with fast and an efficient machine learning regression algorithm. Using ML methods directly to extract biophysical parameters is rarely viable, due to the scarcity of in situ observations which can be used to appropriately train the models.

Vegetation modeling strategy was based on Danner et al., (2021) and Pampanoni et al., (2022), which can be considered the state of the art regarding LUT-based approach and ML inversion for LAI from satellite images, based on the PROSAIL-adjacent RTMs. Coupling a classical LUT-based approach with ML inversion allows us to keep the underlying physical foundation provided by the RTM, and to greatly simplify the inversion procedure by making use of the ML inversion. This has also the added benefit of greatly shortening the inversion times, since most of the heavy calculations are made during the training of the model. The combination of PROSPECT-D and 4SAIL is used to model croplands, which are assumed to be homogeneous vegetated pixels. To model the cropland understory, the pyprosail package was tested with dry and wet soil spectra in different proportions using the psoil parameter.

The RTM inversion strategy applied to these two model combinations is based on the use of LUTs, again similarly as Danner et al., (2021). Before describing the methodology adopted to create and invert the LUTs, a sensitivity analysis of the two model combinations was performed in order to verify that the model is sensitive to the variables that allow us to calculate the LAI, and in particular to identify the ideal Sentinel-2 and PRISMA channels to use to successfully perform the inversion.

Approach:

Before attempting to retrieve the biophysical parameters from satellite observations using a model inversion technique, it was mandatory to verify the feasibility of the retrieval on a variable-by-variable basis. If a variable does not have any significant effect on the variation of the model output, it would be very difficult to invert it, because the model will barely react even to large variations of the variable. To this end, Global Sensitivity Analysis (GSA) was applied (Sobol, 2001, 1993), often used to quantify the effect of the uncertainty of a variable on the uncertainty of the model output (Saltelli, 2002). In particular, the Sobol method allowed an overall evaluation with first-order sensitivity indices, which quantify the contribution of each variable per se, i.e. while all the other variables are kept constant, and higher-order sensitivity indices that account for the interactions between variables.

The possibility to take into account the interactions between variables is fundamental, because some variables that may appear to be not significant at the first-order, may become significant through interactions. The total-order sensitivity index of the i-th variable is the sum of the first-order sensitivity index of that variable and of all the higher-order sensitivity indices associated with its interactions with the other variables.

These calculations have been carried out using the SALib Python library (Herman and Usher, 2017), which makes it straightforward to construct the necessary sequences for each model parameter through the Saltelli module, requiring only the definition of the variation interval of each variable as an input. The module returns quasi-random, low-discrepancy sequences calculated in order to sample the parameter space as uniformly as possible by exploiting Saltelli's sampling scheme, which is an improvement of Sobol's (Herman and Usher, 2017). The sequences are then used as inputs for the pyprosail Python library, and the resulting outputs are run through the sobol.analyze function which performs the Sobol Global Sensitivity Analysis returning both first-order indices and total-order indices.

Methodology:

Generation of the LUTs

To generate the LUT, the PROSAIL model was run in forward mode to simulate canopy reflectance for an appropriate number of parameter combinations. The LUT calculation process took place as follows:

- Definition of the LUT target size: LUT sizes were tested, ranging from 1.000 to 200.000 samples. A LUT size of 10.000 parameter combinations were found to achieve a good compromise between the computer resource requirement and the accuracy of canopy variable estimation (Weiss et al., 2004). The 10,000 parameter combinations were randomly generated with uniform or normal distributions and specific ranges for the variables.
- Definition of the LAI minimum and maximum values.
- Creation of first-guess distributions for each variable: after setting a seed for the random number generator, independent distributions of values were created for each model parameter.
- Elimination of unlikely combinations based on ecological rules (Quan, Yebra 2021).
- Sample selection: with an adequate number of samples guaranteed, in order to match the LUT size to the target sample size, a selection among the available profiles of those that generate an equal number of LAI samples in a number of 10%-wide LAI bins between the minimum and maximum target LAI values. This ensured that an equal number of profiles in each of these bins was held, and therefore profiles that generate different ranges of LAI were sampled equally.

- LUT calculation: the reflectance spectra outputted by the model were convolved using the Spectral Response Function (SRF) of the target sensor.

ML Inversion

Machine learning methods saw a surge of interest in the past decade for their ease of implementation and their computational speed after the training phase. Random forest regression (RFR) is an ensemble algorithm based on multiple decision trees with frequent use for functional traits retrieval in remote sensing (Izquierdo-Verdiguier and Zurita-Milla, 2018). The mixture of bootstrapping and random feature selection makes it less sensitive to outliers and noise (Waske et al., 2009). RFRs are out-of-bag predictors, meaning that they can only predict values within the range they were trained (Breiman, 2001). Initially conjoint pre-processing, fitting, prediction and evaluation of target variables with RFR. Following (Danner et al., 2021), for RFR, neither the variation of max. features, nor the number of estimators, nor the minimum number of samples per leaf had a notable impact on the model's performance. Best results for estimation of LAI were obtained when 75% of the features were used for training (max. feature = 0.75). A higher number of estimators in the random forest turned out to be slightly beneficial for the training success. The minimum number of samples per leaf for each decision tree turned out to be quite insensitive. The best results with a 2000 combinations LUT (rRMSETest = 0.25, R_2Test = 0.82) were obtained (Danner et al., 2021) with the following parameterization:

- Max. features: 0.75
- Min. samples per leaf: 1
- NEstimators: 100+

Max. features did not have any impact on model size, but the minimum number of samples per leaf scaled with inverse potency and could therefore be increased to reduce model size.

Experimental Results:

Before comparing the inversion results with field data, the performance of the RF models was evaluated by comparing the results of the RF LAI estimated through the inversion of the LUT in the training/testing phase with the original LUT LAI values. Specifically, the simulation was performed starting from the Sentinel-2 image over the Narok area of 12/6/2023 (the only partially cloud-free image). In order to make a comparison, the same observation conditions as in the S2 image were included as input, i.e. Solar Zenith Angle (SZA)=34, Viewing Zenith Angle (VZA)=8 and Relative Azimuth Angle (RAA)=60.

The LUT inversion parameters defined for the cropland class were as follows (Table 26):

Table 26. PROSAIL ranges parameters for LUT inversion.

Parameters	Units	Used ranges for Cropland class

Leaf structure	-	1 – 2.5
Equivalent water thickness	cm	0.0001 – 0.07
Dry matter content	µg/m2	0.001 – 0.01
Chlorophyll a+ b content	µg/m2	1 – 80
Total carotenoid content	µg/m2	1 – 15
Total anthocyanin content	µg/m2	0 – 5
Brown pigments	-	0
Leaf area index	m2/m2	0.5 – 8
Hot spot parameter	-	0.01 – 0.5
Average leaf inclination angle	degrees	30 – 70
Soil reflectance	-	0 – 1
Soil brightness factor	-	0 – 1

The Random Forest LAI inversion performance in the training/test phase is shown in (Figure 53), which shows a very high coefficient of correlation (R2=0.91).

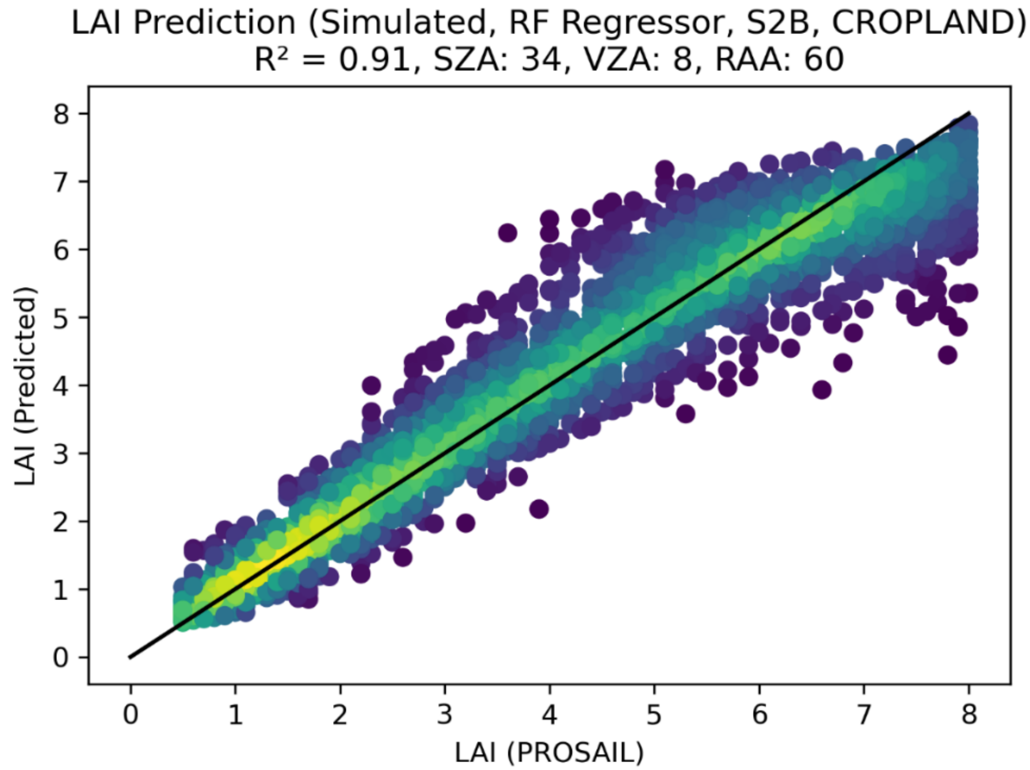


Figure 53. Comparison between Random Forest and PROSAIL simulations for LAI estimation.

Subsequently, on the same Sentinel-2 image, comparisons were made between the LAI measured instrumentally in the field campaigns with the LAI obtained from the SNAP Biophysical Processor tool and with that obtained as output from PROSAIL. Specifically, the differences between the LAI values measured in the field (for both maize and wheat) and the LAI values, at the respective GPS coordinates, of the images obtained through the different software mentioned were calculated directly. The averages and standard deviations of these differences are shown, respectively. The analysis is still in progress, but from an initial assessment the results provided by PROSAIL seem to be more in line with ground truth data from the field, than with those obtained by SNAP (Table 11).

Table 27. Comparison between LAI measured values and LAI computed by SNAP (up) and PROSAIL.

Maize: In Situ Vs. SNAP					
15/06/2023		**16/06/2023**		**17/06/2023**	
Mean	0.4064	**Mean**	1.1372	**Mean**	0.4115
STD	0.4991	**STD**	1.166	**STD**	0.867
Wheat: In Situ Vs. SNAP					

Mean	0.1439	**Mean**	-2.0411	**Mean**	-0.2578
STD	0.5007	**STD**	0.9895	**STD**	0.8168
Maize: In Situ Vs. PROSAIL					
Mean	0.0971	**Mean**	0.0226	**Mean**	-0.4174
STD	0.6774	**STD**	1.1353	**STD**	1.2567
Wheat: In Situ Vs. PROSAIL					
Mean	-0.618	**Mean**	1.2426	**Mean**	-0.8644
STD	0.9651	**STD**	0.9907	**STD**	0.9871

Issues & Next Steps:

As already mentioned before, the main difficulty arises from the reduced number of cloud free images available in the area of interest. The results of Table 27 show that the estimate of the LAI done by using the tool available in SNAP is not particularly accurate. To improve the LAI prediction model, we can use the approach described above, that is computing a LUT based on the spectral channels of Sentinel-2 and/or a selected number of PRISMA spectral channels (the selection could be based on a sensitivity analysis or a Principal Component Analysis approach). Then, using the ground measurements of LAI, a more or less complicated model, starting from simple regression to Neural Network approach, could be developed. For instance, one of the advanced methods used to retrieve quantitative terrestrial bio-geophysical variables is the Hybrid approach (Figure 54) (Verrelst et al., 2015). This method involves using an inverse mapping technique with a non-parametric model (e.g. Machine learning Regression Algorithms) that is trained using simulated data generated by Radiative Transfer Models (RTMs).

If a large number of images, both PRISMA and Sentinel-2, is collected in the coming season, the above-mentioned approach could be attempted.

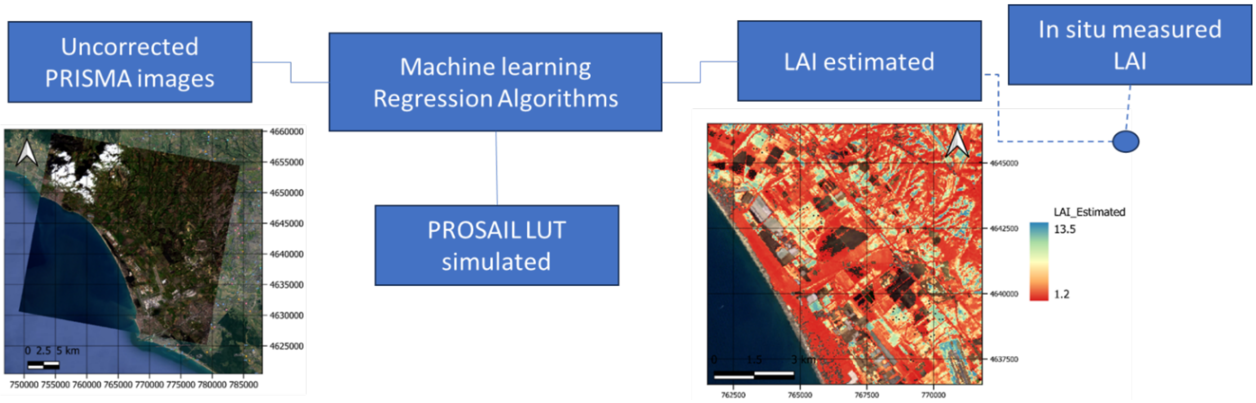


Figure 54. Example of the hybrid approach, applied at an agricultural area in Italy

6. References

Breiman, L., 2001. Random forests. Machine learning, 45, pp.5-32.

Chang, Y. et al, 2020. *A comprehensive evaluation of 4-parameter diurnal temperature cycle models with in situ and MODIS LST over alpine meadows in the Tibetan Plateau*. Remote Sensing, 12. doi: 10.3390/RS12010103.

Chavez Jr., P.S., Sides, S.C., Anderson, J.A., 1991. *Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic.*Photogramm. Eng. Remote. Sens. 57, 295–303.

Chen, W., Sakai, T., Moriya, K., Koyama, L. and Cao, C., 2013. *Estimation of vegetation coverage in semi-arid sandy land based on multivariate statistical modeling using remote sensing data*. Environmental Modelling & Assessment, 18, pp.547-558.

Danner, M., Berger, K., Wocher, M., Mauser, W. and Hank, T., 2021. *Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops*. ISPRS Journal of Photogrammetry and Remote Sensing, 173, pp.278-296.

Darvishzadeh, R., 2008. *Hyperspectral remote sensing of vegetation parameters using statistical and physical models*.

Darvishzadeh, R., Skidmore, A., Schlerf, M. and Atzberger, C., 2008. *Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland*. Remote sensing of environment, 112(5), pp.2592-2604.

Djamai, N. and Fernandes, R., 2018. *Comparison of SNAP-derived Sentinel-2A L2A product to ESA product over Europe*. Remote Sensing, 10(6), p.926.

Duan, S.B., Li, Z.L., Wu, H., Tang, B.H., Ma, L., Zhao, E. and Li, C., 2014. *Inversion of the PROSAIL model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data*. International Journal of Applied Earth Observation and Geoinformation, 26, pp.12-20.

Elwadie, M.E., Pierce, F.J. and Qi, J., 2005. *Remote sensing of canopy dynamics and biophysical variables estimation of corn in Michigan*. Agronomy Journal, 97(1), pp.99-105.

Göttsche, F.-M., & Olesen, F. S., 2001. *Modelling of diurnal cycles of brightness temperature extracted from METEOSAT data*. Remote Sensing of Environment, 76, 337–348. https://doi.org/10.1016/S0034-4257(00)00214-5

Guerif, M., Houlès, V., & Baret, F., 2007, October. *Remote sensing and detection of nitrogen status in crops. Application to precise nitrogen fertilization*. In the 4th International Symposium on Intelligent Information Technology in Agriculture (pp. 26-29).

Herman, J. and Usher, W., 2017. *SALib: An open-source Python library for sensitivity analysis*. Journal of Open Source Software, 2(9), p.97.

Huang, H. et al, 2019. *Image restoration from patch-based compressed sensing measurement*. Neurocomputing, 340. doi: 10.1016/j.neucom.2019.02.036.

Izquierdo-Verdiguier, E. and Zurita-Milla, R., 2018, July. *Use of guided regularized random forest for biophysical parameter retrieval*. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 5776-5779). IEEE.

Jacquemoud, S. and Baret, F., 1990. *PROSPECT: A model of leaf optical properties spectra*. Remote sensing of environment, 34(2), pp.75-91.

Jiang, G.M., 2007. *Retrievals of land surface emissivity and land surface temperature from MSG1-SEVIRI data*. (Doctoral dissertation, Université Louis Pasteur (Strasbourg)(1971-2008)).

Jin, M., and R. E. Dickinson, 1999. *Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle: 1. Without clouds*, J. Geophys. Res., 104(D2), 2105–2116, doi:10.1029/1998JD200005.

Kganyago, M., Adjorlolo, C. and Mhangara, P., 2022. *Exploring Transferable Techniques to Retrieve Crop Biophysical and Biochemical Variables Using Sentinel-2 Data*. Remote Sensing, 14(16), p.3968.

Kimes, D.S., Knyazikhin, Y.P.J.A.A.G.F., Privette, J.L., Abuelgasim, A.A. and Gao, F., 2000. *Inversion methods for physically-based models*. Remote Sensing Reviews, 18(2-4), pp.381-439.

Lu, L. and Zhou, X. M, 2021. *A Four-Parameter Model for Estimating Diurnal Temperature Cycle From MODIS Land Surface Temperature Product*. Journal of Geophysical Research: Atmospheres, 126. doi: 10.1029/2020JD033855.

Miraglio, T., Adeline, K., Huesca, M., Ustin, S. and Briottet, X., 2020. *Joint use of PROSAIL and DART for fast LUT building: Application to gap fraction and leaf biochemistry estimations over sparse oak stands*. Remote sensing, 12(18), p.2925.

Nanni, Loris & Ghidoni, Stefano & Brahnam, Sheryl., 2017. *Handcrafted vs Non-Handcrafted Features for computer vision classification*. Pattern Recognition. 71. DOI:10.1016/j.patcog.2017.05.025

Pampanoni, V., Laneve, G. and Santilli, G., 2022, July. *Evaluating Sentinel-3 Viability for Vegetation Canopy Monitoring and Fuel Moisture Content Estimation*. In IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium (pp. 5634-5637). IEEE.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É., 2011. *Scikit-learn: Machine learning in Python*. The Journal of machine Learning research, 12, 2825-2830.

Quan, X., Yebra, M., Riaño, D., He, B., Lai, G., & Liu, X., 2021. *Global fuel moisture content mapping from MODIS*. International Journal of Applied Earth Observation and Geoinformation, 101, 102354.

Saltelli, A., 2002. *Making best use of model evaluations to compute sensitivity indices*. Computer physics communications, 145(2), pp.280-297.

Scheffler, D., Hollstein, A., Diedrich, H., Segl, K., Hostert, P. , 2017: *AROSICS: An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data*. - Remote Sensing, 9, 7, 676. https://doi.org/10.3390/rs9070676

Shrestha, A., Angal, A. and Xiong, X, 2018. *Evaluation of MODIS and Sentinel-3 SLSTR thermal emissive bands calibration consistency using Dome C*. in. doi: 10.1117/12.2303987.

Soboĺ, I.M., 1993. *Sensitivity estimates for nonlinear mathematical models*. *Math*. Model. Comput. Exp., 1, p.407.

Sobol, I. M., 2001. *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*. Mathematics and computers in simulation, 55(1-3), 271-280.

Thorp, K. R., Wang, G., West, A. L., Moran, M. S., Bronson, K. F., White, J. W., & Mon, J. A. R. A. I., 2012. *Estimating crop biophysical properties from remote sensing data by inverting linked radiative transfer and ecophysiological models*. Remote Sensing of Environment, 124, 224-233.

Uezato, T. et al, 2020. *Guided Deep Decoder: Unsupervised Image Pair Fusion*. in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics. Springer Science and Business Media Deutschland GmbH, pp. 87–102. doi:10.1007/978-3-030-58539-6_6

Uezato, T. et al, 2020. *Guided Deep Decoder: Unsupervised Image Pair Fusion Supplementary Material*. Eccv2020. Verhoef, W., 1984. *Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model*. Remote sensing of environment, 16(2), pp.125-141.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G., & Moreno, J., 2015. *Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–A review*. ISPRS Journal of Photogrammetry and Remote Sensing, 108, 273-290.

Wald, L., Ranchin, T. and Mangolini, M., 1997. *Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images*. Photogrammetric engineering and remote sensing, 63(6), pp.691-699.

Wang, Q., Tang,Y., Atkinson,P. M., 2020 : *The effect of the point spread function on downscaling continua*, ISPRS Journal of Photogrammetry and Remote Sensing,V. 168, pp. 251-267

Wang, Q. et al, 2016. *Fusion of Sentinel-2 images*. Remote Sensing of Environment. 187. doi: 10.1016/j.rse.2016.10.030.

Waske, B., Fauvel, M., Benediktsson, J.A. and Chanussot, J., 2009. *Machine learning techniques in remote sensing data analysis*. Kernel methods for remote sensing data analysis, pp.3-24.

Xu, Y., Hu, S. and Du, Y, 2022. *Research on Optimization Scheme for Blocking Artifacts after Patch-Based Medical Image Reconstruction*. Computational and Mathematical Methods in Medicine, 2022. doi: 10.1155/2022/2177159.

Yin, Z et al, 2021. *Spatiotemporal Fusion of Land Surface Temperature Based on a Convolutional Neural Network*. IEEE Transactions on Geoscience and Remote Sensing 59, 1808–1822. doi:10.1109/TGRS.2020.2999943

Yu, J et al, 2018. *Wide Activation for Efficient and Accurate Image Super-Resolution*. Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 2621-2624

END OF THE DOCUMENT