Data in Brief 52 (2024) 109850

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Polarimetric instrument Global Navigation Satellite System - Reflectometry airborne data

Karin Dassas[a,*], Pascal Fanise[a], Michel Le Page[a], Emna Ayari[a,b], Philippe Baillion[a], Mateo Sige[a], Aaron Boone[c], Mehrez Zribi[a]

[a] CESBIO[1], University of Toulouse, CNES/CNRS/INRAE/IRD/UT3, 18 avenue Edouard Belin BPI 2801 31401 Toulouse Cedex 9, France
[b] National Agronomic Institute of Tunisia, Carthage University, 43 Av. Charles Nicolle, Tunis 1082, Tunisia
[c] CNRM University of Toulouse, Météo-France/CNRS, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France

ARTICLE INFO

Article history:
Received 25 October 2023
Revised 16 November 2023
Accepted 20 November 2023
Available online 26 November 2023

Dataset link: LIAISE GNSS-R GLORI CESBIO v1 2021 (Original data)
Dataset link: LIAISE Land Use CESBIO v1 2021 (Original data)
Dataset link: LIAISE In situ measurements CESBIO v1 2021 (Original data)

Keywords:
L1-band
Incoherent component
Soil moisture
Vegetation
Land use
LIAISE

ABSTRACT

In this paper, three datasets are described. The first dataset is a complete set of GNSS-R (GNSS-R: Global Navigation Satellite System – Reflectometry) airborne data. This dataset has been generated with the data acquired with the GLObal Navigation Satellite System Reflectometry Instrument (GLORI) developed at Centre d'Etudes Spatiales de la Biosphère (CESBIO), during the Land surface Interactions with the Atmosphere over the Iberian Semi-arid Environment (LIAISE) campaign in north-eastern Spain during the summer of 2021. It is the first time to our knowledge that a complete dataset of GNSS-R observables (reflectivity, incoherent component relative to the total scattering signal to noise ratio (SNR) for copolarized (right-right) and cross-polarized (right-left) measurements has been made available.
The two other datasets are ground truth sets of measurements which have been acquired simultaneously with the flights. The in-situ measurements dataset consists in soil measurements (surface soil moisture, surface roughness, Leaf Area Index (LAI)) over 24 reference fields). The land use dataset provides a land use map (along with 385 ground truth plots) over the studied site for GLORI data evaluation.

* Corresponding author.
E-mail address: karin.dassas@univ-tlse3.fr (K. Dassas).
[1] @CesbioLab.

https://doi.org/10.1016/j.dib.2023.109850

The combined datasets are particularly relevant for soil moisture and vegetation retrievals from GNSS-R observables, as well as studies for calibration and validation of bistatic empirical or physical models simulating coherent or incoherent components on agriculture sites, in the context of the preparation of future GNSS-R space missions, such as HydroGNSS, a European Space Agency mission, launch foreseen in 2024.
The entire database is archived in the AERIS LIAISE database. One DOI is available for each of the 3 datasets (airborne GLORI dataset, in situ measurements dataset and land use dataset).

Specifications Table

Subject	Earth and Planetary Science
Specific subject area	GNSS-R L1 polarimetric airborne data along with ground truth data (soil moisture, Land use)
Data format	**GLORI airborne dataset**: analysed and filtered data **In-situ measurements dataset**: analysed and filtered data **Land use dataset**: filtered raw and analysed data
Type of data	**GLORI airborne dataset**: one zip file containing three .csv data (one for each flight), one .csv with metadata and one .txt for flights information **In-situ measurements dataset**: one zip file with shapefiles (EPSG4326 - WGS84) for the plot locations and a csv file which gives in situ measurement data. plot_id is a common field which makes the link between the shapefiles and the csv fil **Land use dataset**: a zip file with shapefiles (EPSG:32631 - WGS 84 / UTM zone 31N) and a csv file corresponding to the identified Land use for the 385 fields, and a land use map GTiff format, EPSG:32631 - WGS 84 / UTM zone 31N
Data collection	**GLORI airborne dataset**: GNSS reflectometry is a bistatic radar remote sensing technology (transmitters and receivers are not in the same place) that uses microwave signals of opportunity from radio navigation constellations. This dataset has been generated with the data acquired with the GLORI reflectometer developed at CESBIO [1]. These data have been acquired during 3 flights carried out in July 2021 (flight 45/46/47 respectively on July 22nd, 27th and 28th) over the Urgell site. This dataset provides GNSS-R observables such as reflectivity, incoherent component relative to the total scattering, signal to noise ratio (snr) for copolarized (right-right) and cross-polarized (right-left) measurements. These products (Level 1-B) have been generated using a CESBIO software described with details in the data description section. Data provided in the delivered dataset are not normalized. They are then proposed for various incidence angles ranging between 0 and 60°. The noise floor has been removed. Outliers have not been removed. **In-situ measurements dataset**: the field datasets consist of punctual surveys for soil moisture acquired using a Delta-T ThetaProbe MI2x, surface roughness acquired using a pin profiler), and Leaf Area Index (LAI). The LAI was derived from hemispherical digital photography based on analysis of the canopy gap fraction with method proposed by Duchemin [2]. **Land use dataset:** the survey was carried (one driver and one surveyor) in two different cars, following the dirt roads around two GLORI transect lines. The central coordinates of each plot were taken from the road with the smartphone GPS and google maps application. Using QGIS, google maps and NDVI map, polygons have been drawn around the points of the original shape file (E. Ayari, K. Dassas, V. Dehaye). Polygons have also been drawn on water and urbain areas. Classes have been attributed to the different land uses (OCS). 1 = maize, 2 = grass and alfalfa, 3 = maize and apple trees / apple trees /pear trees, 4 = wheat cut / dry grass / bare soil, 5 = water, 6 = urbain. The .tif land use map available has been generated with the Orfeo Tool Box developed by CNES [3]. You'll find more details in the Land use map generation section

(*continued on next page*)

Data source location	The data were collected on the Urgell site in Spain (Leida) : EPSG:32631 - WGS 84 / UTM zone 31N. Please refer to Fig. 1 The data are stored in a French repository Data Terra Aeris. The AERIS atmosphere data and services hub federates national atmospheric data management and scientific expertise activities. Formed in 2014, it encompasses four integrated and increasingly connected data and services centres (DSCs). Data Terra is part of the Ministry for Higher Education, Research and Innovation's 2016, 2017 and 2021 RI/RI* roadmap.
Data accessibility	Direct URL to data: the 3 DOIs allow direct access to the data **LIAISE GNSS-R GLORI CESBIO v1 2021 dataset:** https://doi.org/10.25326/494 [4] Persistant identifier: 9329aad9-f539-fa26-72ac-db159c9deccb **LIAISE In situ measurements CESBIO v1 dataset:** https://doi.org/10.25326/493 [5] Persistant identifier: 8d0a51a0-0974-c787-7d09-2bffd92a9dfd **LIAISE Land use CESBIO v1 dataset**: https://doi.org/10.25326/495 [6] Persistant identifier: 6ac434da-ee64-268b-b71b-2ea3ae73029e There are no specific Instructions for accessing these data, apart agreeing with the Data Policy [7] The in situ and land use datasets corresponding to the airborne campaigns are also available through the EIS (environmental Information system at CESBIO) in the field surveys section, in the HILIAISE project. The purpose of the EIS [8] is to centralize, standardize and make available to internal and external users of the laboratory a (rich) set of environmental observations collected by the CESBIO and its partners, with fast access and quick look tools.
Related research article	M. Zribi, K. Dassas, V. Dehaye, P. Fanise, E. Ayari, M. Le Page, Analysis of Polarimetric GNSS-R Airborne Data as a Function of Land Use, IEEE Geosci. Remote Sensing Lett. 20 (2023) 1–5. https://doi.org/10.1109/LGRS.2023.3270730. 2 articles ([9] and [10]) have been published using these datasets, but we propose to choose the reference [10] as related research article. This article is related to the GLORI airborne dataset combined with the Land use dataset.

1. Value of the Data

- These data are valuable for the GNSS-R community in order to prepare the analysis of future GNSS-R Space mission such as ESA HydroGNSS that should propose polarimetric data. The GLORI dataset make polarimetric reflectivities and SNR LHCP (Left Hand Circular Polarization) and RHCP (Right Hand Circular Polarization) available, but also variables which give information on the incoherent part of the signals (incoherent ratio variables). This information is valuable to validate or invalidate common hypothesis on the predominance of the coherent component in the GNSS-R signal over continental surface.
- These data are valuable for electromagnetic modelling community to validate proposed scattering models for coherent and incoherent components and more generally simulators for GNSS-R measurements.
- The 3 flights data available in the GLORI airborne dataset correspond to 3 periods with different pluviometry (dry, after inundation event, and one during the drying process) as presented in Table 1. This is of great interest to use the data for SSM estimation. Other researchers can reuse this data to validate soil moisture model using the GNSS-R dataset along with the variable SSM in the situ measurement dataset. This data could also be proposed for inter-comparison of different SAR or passive microwave soil moisture products with GNSS-R estimations.

Table 1
Summary of the GLORI flights characteristics over the GLO zone.

Flight number	Date	Flight Duration	GLO Start (UTC)	GLO End (UTC)	Precipitation Over Urgell site
45	22/07/21	04:11:55	11:33:13	12:29:40	Dry
46	27/07/21	04:04:38	11:12:44	12:05:42	After inundation event
47	28/07/21	04:03:51	11:14:23	12:07:52	Drying process

- The Land use map can be used by any other researchers working on the Leida region (typical of the semi-arid region, of great interest for climate variables studies) but it can also be reused by researchers working on Land use models to compare their results using the available ground truth data.

2. Data Description

2.1. Context and study area

2.1.1. LIAISE campaign

All data needed to generate the L1b GLORI dataset, the Land Use dataset and the in-situ measurements dataset have been acquired during the Special Observation Period (SOP) of the last field campaign from the GEWEX (Global Energy and Water EXchanges)-supported LIAISE international project [11].

LIAISE project focuses on the Ebro basin in north-eastern Spain as its study domain. The basin is geographically delimited by the Pyrenees in the north and the Iberian System in the south. Over time, the presence of human society has led to increased surface heterogeneity in the region. This is primarily attributed to intensive agricultural practices, which have significantly impacted both the hydrological cycle and the landscape.

The SOP took place in July 2021, when contrasts between irrigated and rainfed agricultural fields are at their maximum.

2.1.2. Study area

The GLORI campaigns took place in Catalonia, Spain, specifically over the Urgell. This site is an agricultural area consisting of two distinct parts. The first part relies on intensive irrigation, with water coming from the Pyrenees via the Urgell canal. Within this region, three different irrigation methods are employed. The first method is flooding, which remains the dominant approach and is utilized for various crops such as fruit trees (apple and pear orchards) and cereals. The second method is sprinkler irrigation, including pivot, ramp, and integral systems, primarily employed for alfalfa and cereal crops (maize, barley, and wheat). The third method is local irrigation, in particular drip irrigation, which is commonly used for vegetable cultivation.

The second part of the area consists in rainfed agriculture and grassland. The climate in this region falls under the Koppen-Geiger classification, characterized as a cold, semiarid climate bordering the Mediterranean Sea. It experiences mild winters, very dry and warm summers, and two rainy seasons in autumn and spring.

The site exhibits a diverse land use pattern, encompassing both annual crops such as corn, cereals, and alfalfa, as well as orchards (apple and pear.). The specific locations of the land cover dataset and in situ measurements dataset can be observed in Fig. 1.

2.2. GLORI airborne dataset

2.2.1. Structure

In the downloaded LIAISE_GLORI_V1.zip file:

- one csv file with a time serie (L1b file) with a temporal resolution of 200ms (the incoherent integration time) for each flight on the GLO zone:

  LIAISE_GLORI-2021_SAFIRE-ATR42_Cesbio_05ms_200ms_as210045_v1.csv
  LIAISE_GLORI-2021_SAFIRE-ATR42_Cesbio_05ms_200ms_as210046_v1.csv
  LIAISE_GLORI-2021_SAFIRE-ATR42_Cesbio_05ms_200ms_as210047_v1.csv

- GLORI_transects_infos_flights45_46_47.txt file with all flights information, summarized in Table 1

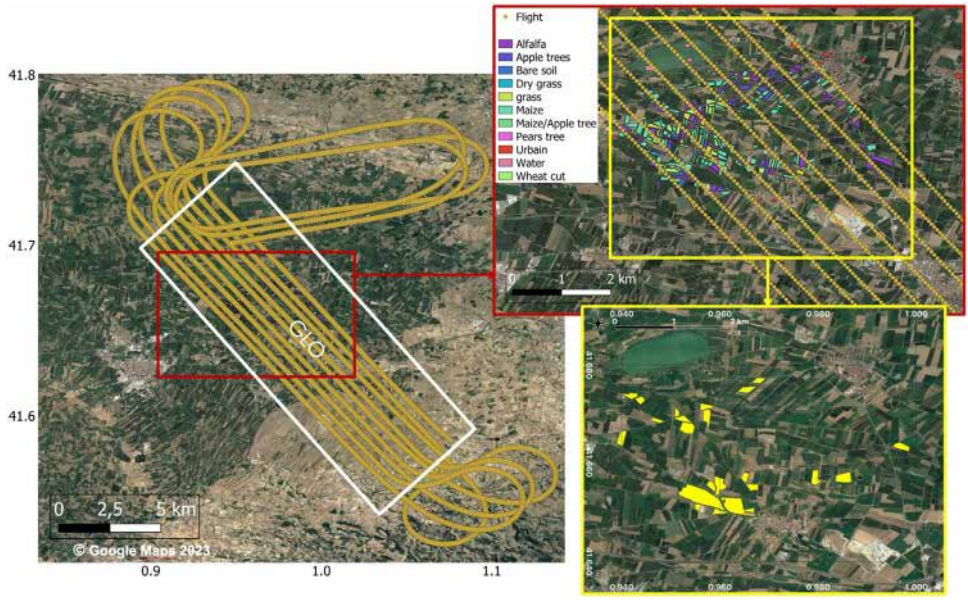

Fig. 1. Flight path with the GLO transects zone (white frame), (left panel), land use plots (upper right image) and plots with in situ measurements (yellow plots, bottom right image).

- LIAISE_GLORI-2021_SAFIRE-ATR42_Cesbio_05ms_200ms_metadata.csv gives information on the flights data variables, available in Table 2

2.2.2. GLORI flights

This airborne dataset has been generated with airborne raw data from 3 flights with French research ATR-42 aircraft carried out in July 2021 (flight 45/46/47, respectively, on July 22nd, 27th and 28th) over the Urgell site. Fig. 1 (left panel) shows the flight paths over the Urgell site, with the location of the GLO transects zone. These transects (9 by day) have been specially planned for the GLORI instrument study to fly both over an irrigated surface and over a dry surface. The altitude over ground of the flight over the GLO zone is approximately 1150m. The flights were carried out using 9 transects to ensure optimum coverage of the entire study site, taking into account the altitude of the aircraft and the incidence angles considered by the GNSS-R measurements. Choosing 200ms as incoherent integration time also respect the objective of having data at the scale of the agricultural plot with a good signal-to-noise ratio. This integration time takes into account the aircraft speed of 100m/s. This choice has made it possible to carry out analyses such as soil moisture mapping at a spatial resolution of 100m [9].

The airborne dataset only covers the flights over the GLO zone.

2.3. Variables description

2.3.1. GLORI L1b fields summary

Please refer to the GLORI processing chain section to have details on the products generation.

2.3.2. GLORI L1b reflectivities and snr statistics

The reflectivity histograms (Fig. 2) show reflectivity values comprised in the expected ranges with a mean of around -12 dB for Γ_l and -20 dB for Γ_r, with higher values on water specular surfaces as shown in Fig. 4 a) and b), with a zoom on a lake.

Table 2
GLORI airborne dataset variable.

Field	Decription	Format	Unit
dtime	Date and time	string	RFC3339
geometry	EPSG4326 - WGS84 polygons	float	
Platform			
h_msl	height over mean sea level (see step 3)	float	M
g_speed	velocity relative ground	float	m/s
theta_nad	incidence angle relative to reflected antenna boresight	float	degrees
theta_zen	incidence angle relative to direct antenna boresight	float	degrees
Satellites / Reflections			
prn	GPS satellite pseudo random noise identifier	int	n/a
azim	GPS satellite azimuth	float	degrees
elev	GPS satellite elevation	float	degrees
phi	GPS satellite azimuth relative to aircraft	float	degrees
s_lat	specular point latitude (see step 3 from pipeline)	float	degree north
s_lon	specular point longitude (see step 3 from pipeline)	float	degree east
s_dem	terrain elevation at specular point (see step 3 from pipeline)	float	M
GLORI Instrument			
gamma_l	LHCP reflectivity (Γ_l from step 2 from pipeline)	float	dB
gamma_r	RHCP reflectivity (Γ_r from step2 from pipeline)	float	dB
noise_fix	noise fix	float	dB
noise_nl	noise floor of reflected LHCP signal	float	dB
noise_nr	noise floor of reflected RHCP signal	float	dB
noise_zr	noise floor of direct signal	float	dB
incoherent_ratio_l_	incoherent ratio LHCP (α_l from step 2 from pipeline)	float	n/a
incoherent_ratio_r	incoherent ratio RHCP (α_r from step 2 from pipeline)	float	n/a
phase_l	LHCP phase difference relative to direct	float	Rad
phase_r	RHCP phase difference relative to direct	float	Rad
snr_nl	nadir LHCP signal to noise ratio	float	dB
snr_nr	nadir RHCP signal to noise ratio	float	dB
snr_zr	zenith RHCP signal to noise ratio	float	dB

Table 3
statistics L1b (-0.2% outliers on each side).

	Γ_ldB]		Γ_r[dB]		snr_nl [dB]		snr_nr [dB]		Nb samples
		σ	μ	σ	μ	σ	μ	σ	
Flight 45	-12.61	3.65	-20.45	5.30	14.01	4.15	8.195	2.78	47577
Flight 46	-11.18	3.48	-20.27	5.02	15.92	4.30	9.085	2.80	47415
Flight 47	-12.03	3.62	-20.77	5.25	15.18	3.96	8.70	2.82	46990

The Signal to Noise Ratio histograms (Fig. 3) show mean values around 15 dB for snr_nl and 9 dB for snr_nr. For each plot, outliers have been removed. We consider data in $\mu \pm 2\sigma$ *interval*, where 2σ represents the 95% quantile of the standard normal distribution.

Table 3 illustrates main statistics (mean and standard deviation values), considering a gaussian distribution of reflectivity and SNR distributions over GLO zone (with outliers removed using quantile 0.002 and 0.998).

2.3.3. GLORI L1b reflectivities and snr over the IVARS lake

Reflectivity and SNR maps (Fig. 4) presented below demonstrate the correct calibration of the delivered data, with the highest levels above the lake.

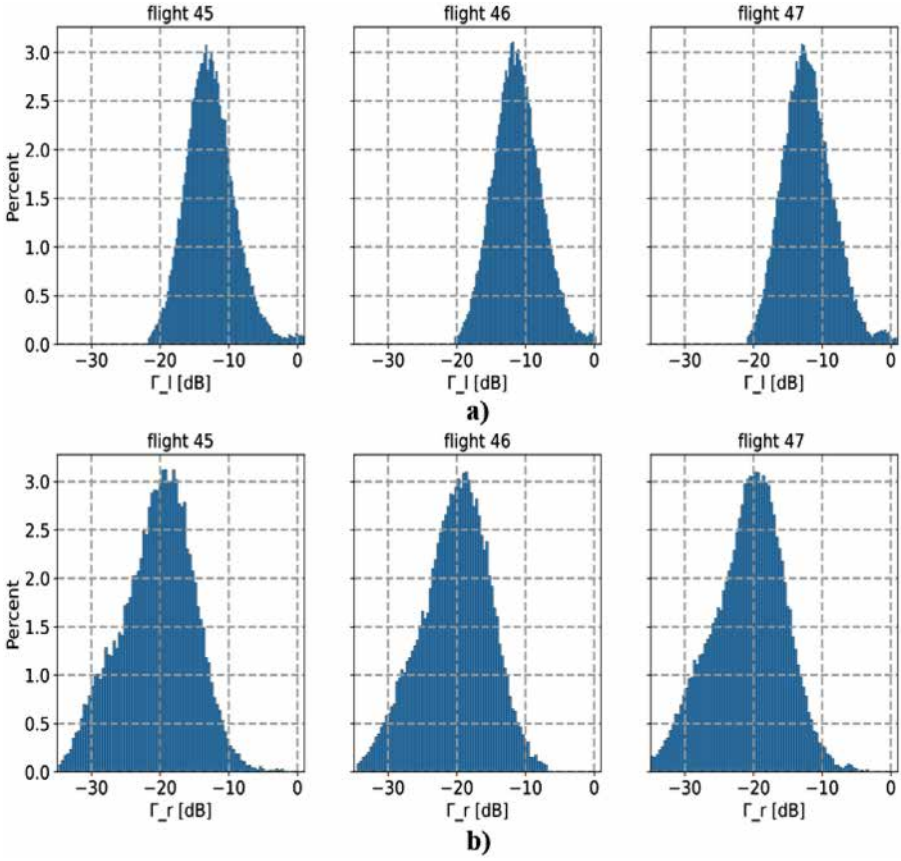

Fig. 2. (a) reflectivity Γ_l (b) Γ_r in the GLO zone.

3. In situ measurements dataset

3.1. Structure

In the downloaded LIAISE_SAFIRE_ATR42_INSITU_MEAS_L2_V1.zip file:

- HILIAISE_Leida_InSituMeasurements_July_2021.csv which gives in situ measurement data, whose characteristics are summarized in Table 4
- HILIAISE_Leida_InSituMeasurements_July_2021.cpg,.dbf, .prj, .qmd, .shp, .shx are shapefiles (EPSG4326 - WGS84)) which give the plot locations

3.2. Variables description

The in-situ measurements dataset consists of punctual surveys for soil moisture, surface roughness, and vegetation Leaf Area Index (LAI) over 24 reference fields (bare soil, alfalfa, maize, apple trees), cf Fig. 1.

The LAI is defined as the total one-sided area of leaf tissue per unit ground surface area, thus it is a dimensionless quantity characterizing the vegetation cover development.

The field plot_id is a common field which makes the link between the shapefiles and the csv file

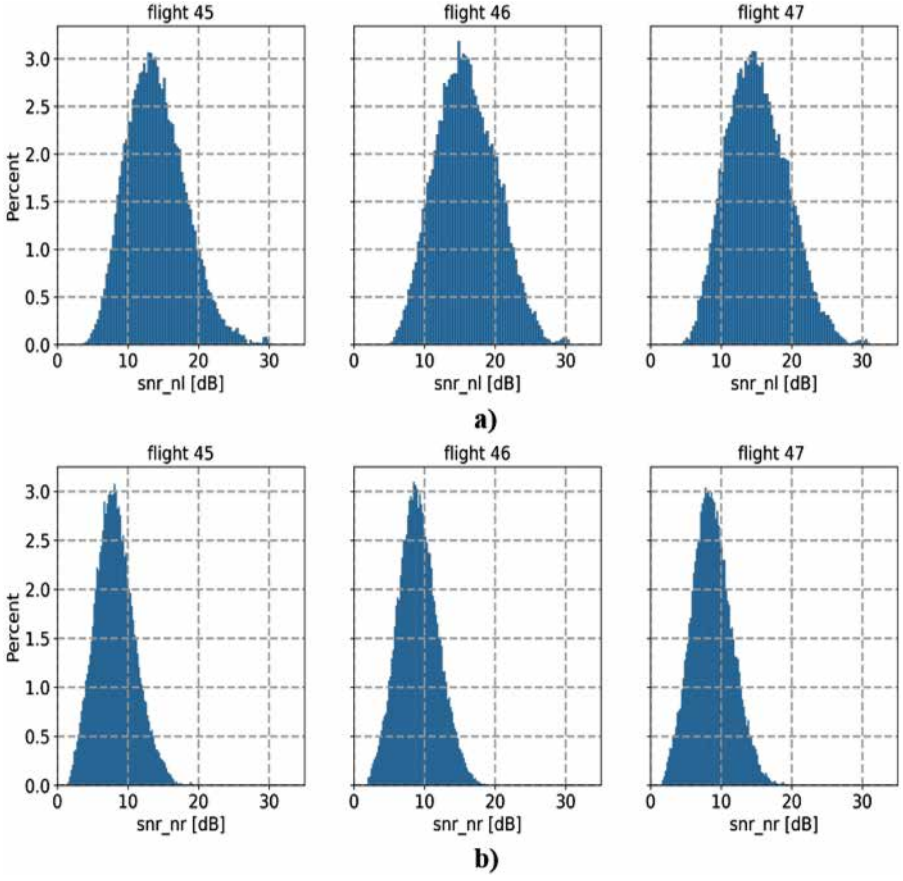

Fig. 3. SNR for left (a) and right (b) polarizations in the GLO zone.

3.3. Land use dataset

3.3.1. Structure

In the downloaded the LIAISE_SAFIRE_ATR42_LAND_USE_V1.zip file:

- HILIAISE_Leida_landcover_transect_25_07_2021_poly.zip is a zip file with shapefiles (EPSG:32631 - WGS 84 / UTM zone 31N) and a csv file corresponding to the identified Land use for the 385 fields.
- LeidaJuly2021LandUseMap_fromOTB.tif is a land use map (GTiff format, EPSG:32631 - WGS 84 / UTM zone 31N) generated with the Orfeo Tool Box (CNES, 2018) using the 385 ground truth samples previously cited and Level-2A (please refer to section Land use map generation to have more details on the method used).

3.3.2. Land use map visualization

Fig. 5 presents the delivered land use map with the six analysed classes. Irrigated land use classes are clearly identified in the western part of the study site. Note that the delivered map extent is bigger than the ground truth extent (Fig. 1)

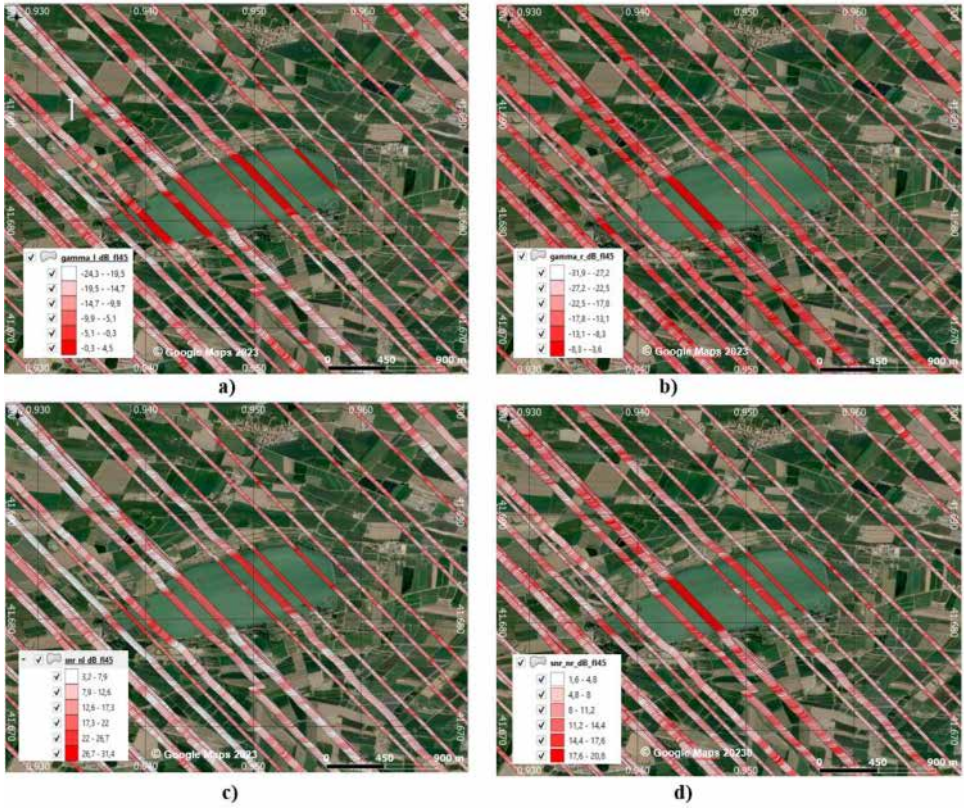

Fig. 4. (a) reflectivity left Γ_l (b) reflectivity right Γ_r and (c) snr_nl (d) snr_nr over the Ivars lake, flight 45.

Table 4
Summary of the in-situ measurements characteristics.

Date	*HRMS* (cm)	*Roughness_correlation_length* (cm)	SSM_{mean} (m^3/m^3)	*LAI_mean_57deg* (m^2/m^2)
15/07/2021			[0.1-0.41]	
16/07/2021			[0.09-0.39]	
17/07/2021			[0.07-0.46]	
19/07/2021				[1.07-2.54]
20/07/2021			[0.02-0.41]	
21/07/2021			[0.04-0.38]	
22/07/2021			[0.06-0.34]	
23/07/2021	[0.4-1.84]	[4.112.18]		[0-3.14]
27/07/2021	-	-	[0.12-0.41]	-
28/07/2021	-	-	[0.1-0.44]	
29/07/2021				[0.15-3.42]

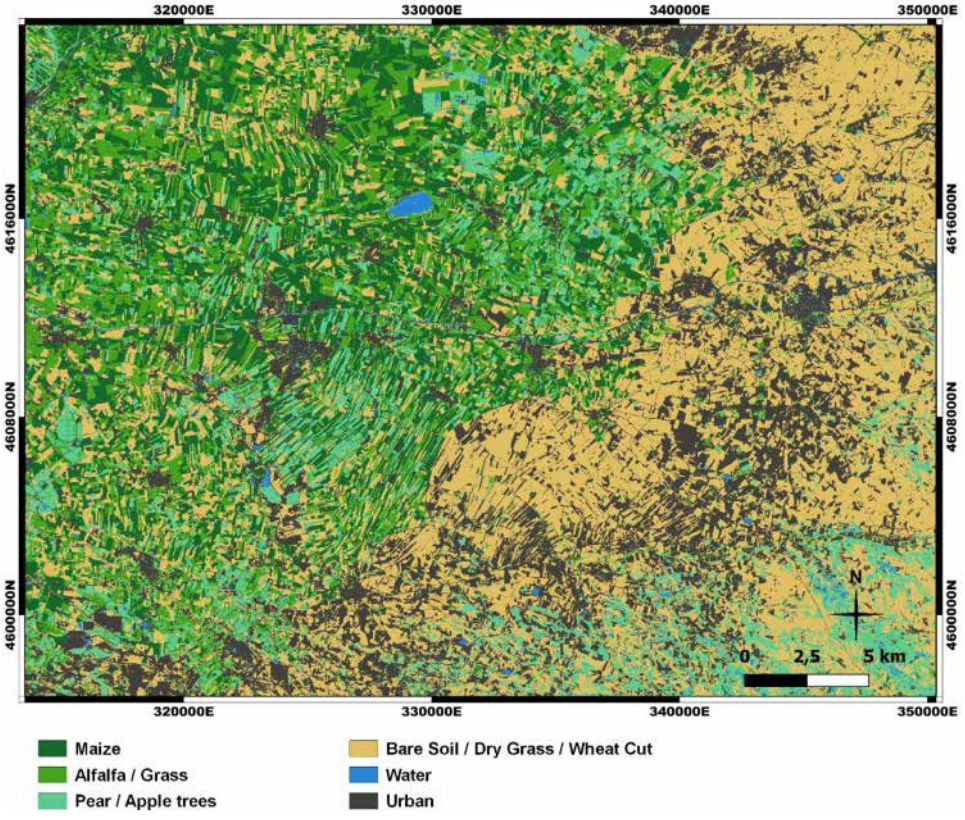

Fig. 5. Land use map (EPSG32631 UTM31N) using supervised classification with Sentinel2 data and Ground Truth data (July 2021).

Fig. 6. Location of the GLORI antennas on the ATR-42.

4. Experimental Design, Materials and Methods

4.1. GLORI airborne dataset

4.1.1. GLORI instrument description

The GLORI instrument is derived from the conventional GNSS-R (cGNSS-R) family. It is a 4-channel GNSS-R polarimetric receiver which allows simultaneous acquisition of reflected Left Hand Circular Polarization (LHCP) and Right Hand Circular Polarization (RHCP) signals (Fig. 6).

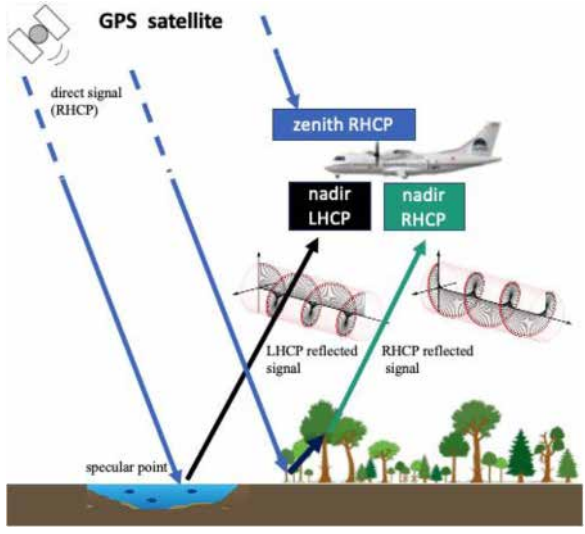

Fig. 7. GNSS-R principles.

For GLORI 2021, measurements of direct and reflected GNSS signals are realized by two passive hemispherical GPS L1 (L1 band = 1575.42 MHz) dual-polarization antennas. The zenith antenna (ZR) is looking up to the direct signal, and the nadir antenna is looking down to the nadir LHCP (NL) and RHCP (NR) as illustrated in Fig. 7.

The main and detailed technical specifications of the instrument are similar to the GLORI 2015 described by E. Motte [12], except for the zenith antenna which is a passive Cobham antenna (Cobham DS1563, the same model as the nadir antenna). The GLORI 2021 front end is presented in Fig. 8. The back end is exactly the same as the back end of GLORI 2015.

4.1.2. GLORI processing chain

The GLORI data have been processed after the flights using the GLORI pipeline developed at CESBIO. This pipeline includes 3 main steps described in Fig. 9, starting with the raw data acquired with the GLORI instrument described in the previous section.

Step 1 is the processing of the GNSS-R data from level RAW (RF signals down-converted to IF) to Level 0 (complex waveforms tagged in time and metadata). The direct and reflected GNSS signals, $u_{d,p}$ and $u_{r,p}$,respectively, were cross-correlated with pseudo-random noise (PRN) code replicas $a(t)$.

The cross-correlated waveform can be written in the following form:

$$Y(\tau, f) = \frac{1}{t_{t_{coh}}} \int_{t_{t_{coh}}} u_{d,p} a(t-\tau) e^{i(f_c+f)t} dt \tag{1}$$

where $t_{t_{coh}}$= 5 ms in the present study) is the coherent integration time. For the direct signal, τ corresponds to the signal propagation delay from transmitter to receiver, f_c is the GPS L1 frequency (1575,4 MHz) and f corresponds to the Doppler shift resulting from the velocity of the aircraft relative to that of the GNSS satellite. p stands for polarization and is equal to r for RHCP and l for LHCP. For the direct signal, p is always equal to r.

Step 2 is the processing step which yields to the following datasets variables: noise, gamma, incoherent_ratio, and snr.

When estimating the reflectivity (gamma), we utilize a calculation method that relies on the interferometric complex field (ICF). The ICF is defined as the ratio between the peak time series of the reflected waveform and the direct waveform. The tracking of the waveform peaks, which correspond to signal distributions generated by the area surrounding the specular point, and

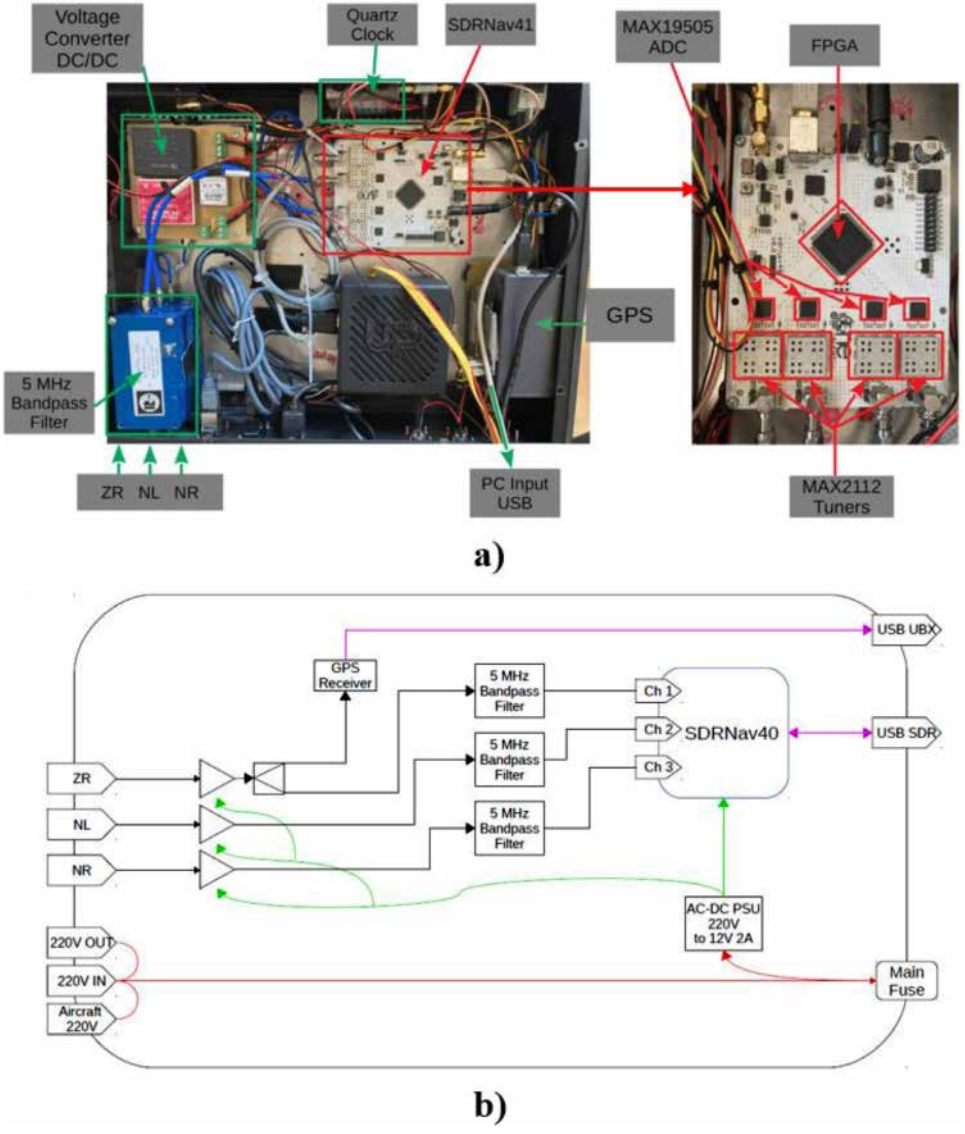

Fig. 8. GLORI 2021 Front End (a) picture and (b) synoptic view.

especially the direct signal mitigation algorithm employed in the GLORI pipeline, is described by E. Motte [12].

The general expression of ICF is as follows:

$$ICF_{corr} = \frac{|Y_{r,max}| - noise_{fix}}{|Y_d, max| - noise_{fix}} e^{j(\varphi_{r,max} - \varphi_{d,max})} \frac{G_d}{G_r} \tag{2}$$

where $noise_{fix}$ (3) is the instrumental noise as a function of the coherent integration time, G_d is the antenna gain for the direct channel, G_r is the antenna gain for the reflected channel, $\varphi_{r,max}$ and $\varphi_{d,max}$ are the phases, and $|Y_{r,max}|$ and $|Y_{d,max}|$ are the moduli of the reflected and direct waveform peaks, respectively.

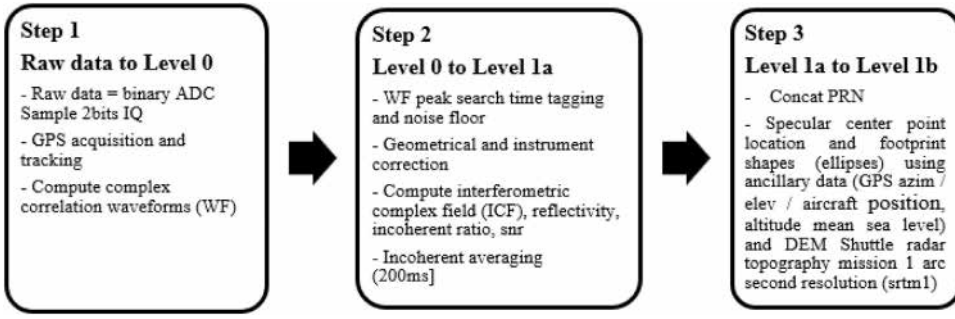

Fig. 9. GLORI pipelines steps.

The instrumental noise is defined as:

$$noise_{fix} = noise_{1ms}/t_{coh} \tag{3}$$

where noise1ms is the empirical noise value over 1 ms equal to 0.098 and t_{coh} is the coherent integration time in seconds.

B_{fix} is a good estimation of the direct and reflected waveform noise levels.

In the L1b dataset, the **noise_fix** variable is provided as well as the computed noise floor, **noise_nr, noise_nl**, and **noise_zr** (for comparison to noise_fix). The noise floor has been computed using the peak waveform mask.

Reflectivity: to estimate the reflectivity Γ'_p, we eliminate the incoherent contribution from the ICF, as proposed by Egido et al. [13]:

$$\Gamma_p = \left\langle |ICF_{p,corr}|^2 \right\rangle - \sigma^2_{ICF_{p,corr}} \tag{4}$$

where $\sigma^2_{ICF_{corr}}$ is the variance of the ICF over the incoherent averaging period (200ms), and p stands for polarization (right or left).

Incoherent component: in order to estimate the percentage of the incoherent component relative to the total scattering power, the incoherent ratio, α_p, is computed:

$$\alpha_p = \frac{\sigma^2_{ICF_{p,corr}}}{\left\langle |ICF_{p,corr}|^2 \right\rangle} \tag{5}$$

In the L1b dataset, Γ_l and Γ_r are provided as **gamma_l** and **gamma_r** variables, α_l and α_r are provided as the **incoherent ratio_l** and **incoherent_ratio_r** variables.

The total signal to noise-fix ratio is also provided for direct signal, for LHCP and RHCP reflected signals, respectively, as **snr_zr, snr_nr** and **snr_nr**.

Step 3: L1a files are then aggregated for each PRN.

Footprint shapes are computed the following way:

- compute the ellipse corresponding to the start of incoherent integration time
- compute the ellipse corresponding to the end of incoherent integration time.
- join the min and max footprints two by two
- compute convex hull as an approximation of the footprint shape.
- create a geopanda geoserie

compute the specular point ellipse centered variables: **s_lat, s_lon, s_dem, azim, elev, h_msl** provided in the L1b dataset are variables which corresponds to the center of the footprint shape, shown on Fig. 10. Each ellipse corresponds to the first Fresnel zone as described in [14] and [15].

At the end of the processing chain, the L1b files are time series with a temporal resolution of 200ms (the incoherent integration time), available in the GLORI airborne Dataset

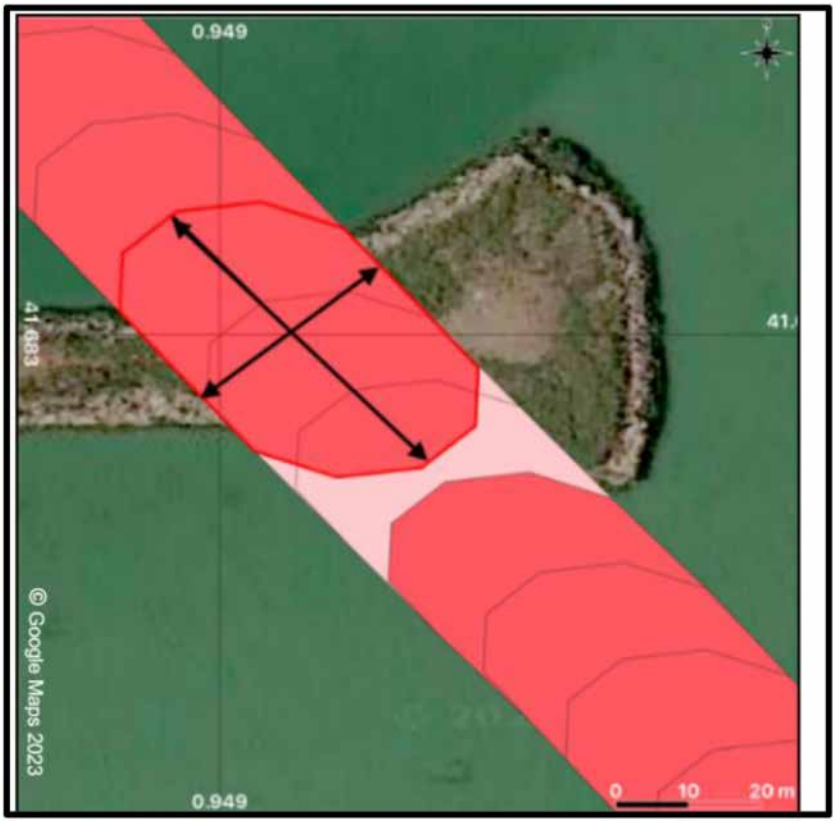

Fig. 10. footprint zoom of the L1b time serie flight 45.

Fig. 11. (a) soil moisture probe (b) a pin profiler picture, plot 25 (alfalfa), on July 23 (c) hemispherical photo maize.

4.2. In situ measurements dataset

4.2.1. Soil moisture

For each of the dates specified in Table 4, around 20 handheld measurements (Delta-T ThetaProbe Ml2x) were conducted in both the irrigated and non-irrigated reference fields. These measurements were taken at a depth of 5 cm and were collected from multiple locations within each reference field, within a four-hour time window coinciding with the airborne data acquisitions. To ensure accuracy, the thetaprobe (Fig. 11 a)) measurements were calibrated using gravi-

metric measurements obtained during previous campaigns ([16]). Given a homogeneous hypothesis for agricultural fields, these in situ measurements are expected to provide a 5 cm volumetric soil moisture content (SSM) mean accuracy of better than 0.02 m3/m3 (water fraction by volume)

4.2.2. Surface roughness

The soil roughness is characterized from measurements taken by a pin-profiler (Fig. 11b) with a 1m length and a spacing between needles of 2cm. We carried out between three and six measurements for each reference plot. We limited ourselves to three profiles for the smoothest plots. The profiles are digitized using a Matlab code which generates the height profile cantered around zero. From each profile, we calculate the height correlation function to deduce the two statistical parameters, the root mean square of the height (HRMS) and correlation length (Lc) [17]. These parameters are then averaged for all considered profiles. In the absence of any change in the surfaces, a single campaign of roughness measurements is carried out during the GLORI campaign, on July 23, 2023. HRMSs varied between 0.4 and 1.84 cm, and the Lc value varied between 4 and 12.18 cm (see Table 4)

4.2.3. Green Leaf Area index and cover fraction

During the GLORI campaigns, the LAI was derived from hemispherical digital photography based on analysis of the canopy gap fraction with method proposed by Duchemin [2]. Approximately 10 photos are taken for each campaign date and per reference field. They are taken using a Canon 6EOS 600D camera with SIGMA 4.5mm F2.8 EXDC circular fisheye HSM. It is essential to consider optimal lighting conditions to avoid shadow effects or overexposure phenomena in these photos. The classification algorithm used to extract the vegetation cover is simply based on the identification of two classes (soil, cover) in the hemispherical images with a thresholding of a green index. Image processing considers operator masking and high viewing angles ($>75°$). These measurements were applied to each reference field with vegetation cover, for three times. Table 4 illustrates range of retrieved levels.

4.3. Land use dataset

4.3.1. Identification of land use plots

Please refer to the Data Collection section of SPECIFICATIONS TABLE,

4.3.2. Land use map generation

We provide here more details on the method used to generate the Land use map with the Orfeo Tool Box with the 385 ground truth plots previously cited and Sentinel 2-data.

Sentinel-2 data are downloaded from the Theia site. Among the available products, only the products without clouds are used. To monitor the land use changes over time, the classification features include spectral indices and reflectance bands. Three indices are calculated, namely, the normalized difference vegetation index (NDVI), the normalized difference water index (NDWI) and the brightness index (BI). To better distinguish among the crop classes, the optical red edge bands are identified. The random forest classifier has been chosen to perform a supervised classification, with a training and a validation dataset, corresponding respectively to 70% and 30% of the reference data for the land covers (randomly divided). The overall accuracy of the land cover classification is approximately 0.93 with a kappa index value of 0.90.

Limitations

For the in-situ measurements dataset, there is a limitation to agricultural areas which does not allow all forms of analysis or validation of diffusion models. However, it is the most requested context for monitoring the water state of the soil or the dynamics of the cycle of the

plant cover. The duration is not very long for monitoring strong temporal variabilities, however with 24 reference plots, we considered different cover dynamics to allow the widest possible study.

Ethics Statement

The current work does not involve human subjects, animal experiments, or any data collected from social media platforms

Credit Author Statement

Pascal Fanise, Mehrez Zribi: conceptualization, methodology, investigation, validation for the GLORI airborne dataset, Michel Le Page was in charge of the in situ and land use ground truth data (methodology, resources, investigation, validation). **Emna Ayari:** has generated the land use map (using OTB software) Mateo Sige: has provided diagrams of the GLORI instrument (visualization). **Philippe Baillion:** is in charge of the EIS and has helped formatting the in-situ data (data curation). **Karin Dassas:** processed data for the airborne dataset (software), formal analysis, validation, visualisation, writing original draft. **Aaron Boon, Mehrez Zribi:** funding acquisition. All the co-authors have contributed in the review and editing of the manuscript.

Data Availability

LIAISE GNSS-R GLORI CESBIO v1 2021 (Original data) (DataTerra Aeris)
LIAISE Land Use CESBIO v1 2021 (Original data) (DataTerra Aeris)
LIAISE In situ measurements CESBIO v1 2021 (Original data) (DataTerra Aeris)

Acknowledgements

This work was supported by the French National Research Agency (HILIAISE ANR-19-CE01-0017 project) and the French Space Study Center (CNES/SCOMAG project). (Corresponding author: Mehrez Zribi.). We would like to thank all farmers of the Urgell site who have allowed the GLORI team to collect data on their fields. The authors also wish to thank the technical teams from the SAFIRE, as well as the SAFIRE pilots, for their support and contributions to the successful airborne campaigns. They also would like to thank the SEDOO team (Data Service at the Observatoire Midi-Pyrénées) in charge of SAFIRE+ and of the LIAISE database. Finally, Theia is acknowledged for providing free Sentinel-2 products (corrected from atmospheric effects).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. Motte, M. Zribi, P. Fanise, A. Egido, J. Darrozes, A. Al-Yaari, N. Baghdadi, F. Baup, S. Dayau, R. Fieuzal, P.-L. Frison, D. Guyon, J.-P. Wigneron, GLORI: a GNSS-R dual polarization airborne instrument for land surface monitoring, Sensors 16 (2016) 732, doi:10.3390/s16050732.

[2] B. Duchemin, P. Maisongrande, G. Boulet, I. Benhadj, A simple algorithm for yield estimates: evaluation for semi-arid irrigated winter wheat monitored with green leaf area index, Environ. Modell. Softw. 23 (2008) 876, doi:10.1016/j.envsoft.2007.10.003.
[3] CNES: the ORFEO tool box cook book. [online] Available from: https://www.orfeo-toolbox.org/CookBook/(last access November 2023).
[4] K. Dassas, M. Zribi, P. Fanise, LIAISE GNSS-R GLORI CESBIO v1 2021, (2023). https://doi.org/10.25326/494
[5] M. Le Page, M. Zribi, K. Dassas, LIAISE InSitu measurements CESBIO v1 2021, (2023). https://doi.org/10.25326/493
[6] K. Dassas, M. Le Page, M. Zribi, LIAISE land use CESBIO v1 2021, (2023). https://doi.org/10.25326/495
[7] Hymex Data policy, https://www.hymex.fr/liaise/LIAISE_docs/DataPolicy_HyMeX_2020-2025-1.pdf, 2020-2025 (last access November 2023).
[8] EIS, https://sie.cesbio.omp.eu/index.php (last access November 2023).
[9] M. Zribi, V. Dehaye, K. Dassas, P. Fanise, M.L. Page, P. Laluet, A. Boone, Airborne GNSS-R polarimetric multiincidence data analysis for surface soil moisture estimation over an agricultural site, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 15 (2022) 8432–8441, doi:10.1109/JSTARS.2022.3208838.
[10] M. Zribi, K. Dassas, V. Dehaye, P. Fanise, E. Ayari, M.Le Page, Analysis of polarimetric GNSS-R airborne data as a function of land use, IEEE Geosci. Remote Sens. Lett. 20 (2023) 1–5, doi:10.1109/LGRS.2023.3270730.
[11] A. Boone, J. Bellvert, M. Best, J. Brooke, G. Canut-Rocafort, J. Cuxart, O. Hartogensis, P. Le Moigne, J.R Miró, J. Polcher, J. J.Price, P. Quintana Seguí, M. Wooster, Updates on the international land surface interactions with the atmosphere over the Iberian Semi-Arid Environment (LIAISE) field campaign, GEWEX Q. 31 (4) (2021) 16–21.
[12] E. Motte, M. Zribi, Optimizing waveform maximum determination for specular point tracking in airborne GNSS-R, Sensors 17 (2017) 1880, doi:10.3390/s17081880.
[13] A. Egido, S. Paloscia, E. Motte, L. Guerriero, N. Pierdicca, M. Caparrini, E. Santi, G. Fontanelli, N. Floury, Airborne GNSS-R polarimetric measurements for soil moisture and above-ground biomass estimation, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7 (2014) 1522–1532, doi:10.1109/JSTARS.2014.2322854.
[14] K.M. Larson, F.G. Nievinski, GPS snow sensing: results from the EarthScope Plate Boundary Observatory, GPS Solut. 17 (2013) 41–52, doi:10.1007/s10291-012-0259-7.
[15] N. Roussel, Application de la réflectométrie GNSS à l'étude des redistributions des masses d'eau à la surface de la terre. Instrumentation et méthodes pour l'astrophysique [astro-ph.IM], Université Paul Sabatier - Toulouse III, Français. NNT, 2015 2015TOU30327.
[16] S. Bousbih, M. Zribi, C. Pelletier, A. Gorrab, Z. Lili-Chabaane, N. Baghdadi, N. Ben Aissa, B. Mougenot, Soil texture estimation using radar and optical data from sentinel-1 and sentinel-2, Remote Sens. 11 (2019) 1520, doi:10.3390/rs11131520.
[17] M. Zribi, N. Baghdadi, N. Holah, O. Fafin, C. Guérin, Evaluation of a rough soil surface description with ASAR-ENVISAT radar data, Remote Sens. Environ. 95 (2005) 67–76, doi:10.1016/j.rse.2004.11.014.