Scout-2 HydroGNSS

Algorithm Theoretical Baseline Document (ATBD) – L2FT, L3FT and L3FTR processors

ESA Contract No:	4000135686/21/NL/AD
Subcontract No:	MS000433
Issue / Revision:	3/1
Date:	8 November 2024
FMI reference:	FMI-ATBD-012

Prepared by

Kimmo Rautiainen, Juval Cohen, Manu Holmberg

Finnish Meteorological Institute (FMI)

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Export Control Document Status

UK Export Control	This document and any hardware referred to within has been UK Export Control Rated as:		
Hardware UK Export Rating Number:	N/A		Date of rating:
Information UK Export Rating Number:	N/A		08/11/2024
I confirm I am an approved SSTL Export Rater and this rating is correct as of the date above:			Martin Unwin
US Export Control	No US Content		
ITAR EAR	This document contains information controlled by the U.S. government and authorized for export only to the country of ultimate destination for use by the ultimate consignee or end-user(s) herein identified. It may not be resold, transferred, or otherwise disposed of, to any other country or to any person other than the authorized ultimate consignee or end-user(s), either in its original form or after being incorporated into other items, without first obtaining approval from the U.S. government or as otherwise authorized by U.S. law and regulations.		
ITAR	The information contained herein is controlled under US ITAR and is authorised for export under the provision of TAA/DSP:		
EAR	The information contained herein is controlled under US EAR and has been classified as ECCN:		

TABLE OF CONTENTS

LIST OF FIGURES

DOCUMENT REVISION STATUS

Last Edited Date	Revision / Release Number	Status	Edited By	Pages / Paragraphs Affected	Change Ref
02/12/2021	0.1		Kimmo Rautiainen	First draft version	
24/03/2022	0.2		Kimmo Rautiainen, Juval Cohen	All, based on H_ATBD_RIDS v0_3	
14/06/2022	0.3		Kimmo Rautiainen, Juval Cohen	Sections 3.2.1, 3.2.6, 4	
26/10/2022	1.0		Kimmo Rautiainen	All	
16/12/2022	1.1		Kimmo Rautiainen, Juval Cohen	All, based on ESA and SSTL comments	
09/03/2023	1.2		Kimmo Rautiainen	All, based on ESA comments	
01/11/2023	2.0		Kimmo Rautiainen, Juval Cohen	All, major update	
05/01/2024	2.1		Kimmo Rautiainen, Juval Cohen	Updated based on SSTL comments	
07/03/2024	2.2		Kimmo Rautiainen	Updated based on ESA comments	
29/05/2024	2.3		Kimmo Rautiainen	Added section 2.1.1.1	
01/08/2024	3.0		Kimmo Rautiainen	All sections revised and updated	
08/11/2024	3.1		Kimmo Rautiainen	Sections 2.2.2.3., 2.2.3.4, 3	

Applicable Documents

AD#	Title	Doc #	Revision	Date
AD-1	Scout-2 HydroGNSS PDGS Algorithm ICD	00637207	14.0	21/5/2024
AD-2	Scout-2 HydroGNSS – System Requirement Document Annexe	00635466	3	1/10/2021
AD-3	Scout-2 HydroGNSS Science Plan	00637068	5	22/4/2022
AD-4	User Guide for L2 and L3 FT beta version	FMI-UGu-017	6.0	29/8/2024
AD-5	Scout 2 HydroGNSS Calibration Validation Plan	00644845	6.0	26/9/2023

Reference Documents

RD#	Title	Journal	Rev.	Date
RD-1	Campbell et al., Intercomparison of Electromagnetic Scattering Models for Delay-Doppler Maps Along a CYGNSS Land Track With Topography	IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2022.3210160		Vol. 60, 2022
RD-2	Melebari et al. Improved Geometric Optics with Topography (IGOT) Model for GNSS-R Delay-Doppler Maps Using Three-Scale Surface Roughness	Remote Sensing, https://doi.org/10.3390/rs15071880		2023
RD-3	Dente et al., Space-Borne GNSS-R Signal Over a Complex Topography: Modeling and Validation	IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, doi: 10.1109/JSTARS.2020.2975187		Vol. 13, 2020
RD-4	Ren et al., Analytical Kirchhoff Solutions (AKS) and Numerical Kirchhoff Approach (NKA) for First-Principle Calculations of Coherent Waves and Incoherent Waves at P Band and L Band in Signals of Opportunity (SoOp)	Progress In Electromagnetics Research, doi:10.2528/PIER21050607		Vol. 171, 2021
RD-5	Comite et al. Monitoring freeze-thaw state by means of GNSS reflectometry: An analysis of TechDemoSat-1 data	IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens, doi: 10.1109/JSTARS.2020.2986859		Vol. 13, 2020
RD-6	Rautiainen et al. Freeze-Thaw Detection over High-Latitude Regions by Means of GNSS-R Data	IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2021.3125315		Vol. 60, 2022
RD-7	ICDR land cover, Product user guide and specification	https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=doc	v1.1	21/06/2021
RD-8	USGS EROS Archive - Digital Elevation - Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010)	https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-multi-resolution-terrain-elevation		
RD-9	H. Carreno-Luengo and C. S. Ruf Retrieving Freeze/Thaw Surface State From CYGNSS Measurements	IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2021.3120932		Vol. 60, 2022
RD-10	H. Carreno-Luengo and C. S. Ruf (2022) Mapping Freezing and Thawing Surface State Periods With the CYGNSS Based F/T Seasonal Threshold Algorithm	IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, doi:0.1109/JSTARS.2022.3216463		Vol 15, 2022
RD-11	Wu et al. First Measurement of Soil Freeze/Thaw Cycles in the Tibetan Plateau Using CYGNSS GNSS-R Data	Remote Sensensing, https://doi.org/10.3390/rs12152361		Vol 12, 2020

Acronyms and Abbreviations

The following abbreviations are used within this document:

CCI	Climate Change Initiative
DDM	Delay doppler map
DEM	Digital elevation model
E2ES	End to end simulator
EASE	Equal Area Scalable Earth
ECMWF	The European Centre for Medium-Range Weather Forecasts
EPSG	European Petroleum Survey Group
ESA	European Space Agency
FMI	Finnish Meteorological Institute
FT	Freeze and thaw
GMTED	Global Multi-resolution Terrain Elevation Data
GNSS	Global Navigation Satellite System
HRES	High resolution
IEEC	Institut d'Estudis Espacials de Catalunya
IMS	Interactive Multisensor Snow and Ice Mapping System
ISMN	International Soil Moisture Network
L2FT	Level 2 Freeze and Thaw Processor
L3FT	Level 3 Freeze and Thaw Processor
L3FTR	Level 3 Freeze and Thaw Reference Processor
LC	Land cover
LHCP	Left hand circular polarised
NOC	National Oceanography Centre
NRT	Near real time
NSDIC	National Snow and Ice Data Center
PDGS	Payload Data Ground Segment
SP	Specular point
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Ocean Salinity
SSTL	Surrey Satellite Technology Limited
TBD	To be determined
TDS	TechDemoSat
USGS EROS	United States Geological Survey, Earth Resources Observation and Science Center
UTC	Universal Time Coordinated

1 BACKGROUND AND SCIENTIFIC JUSTIFICATION

The physical basis for detecting freezing and thawing (FT) of soils using GNSS-R observations is based on the high permittivity contrast between unfrozen and frozen soils. The presence of free liquid water in unfrozen soils increases the effective dielectric constant compared to the frozen situation. The reflectivity value (reflection coefficient at the specular point) observed with bistatic radar (transmitter and receiver at different locations) is lower in frozen soils than in thawed soils. However, there are also other factors that contribute to the observed reflectivity value, such as surface roughness, topography, and vegetation layer (RD-1: Campbell et al., 2022; RD-2: Melebari et al., 2023). In general, reflectivity is stronger over flat terrain compared to terrain with higher surface roughness or more variable slopes. The vegetation layer attenuates the microwave signal and causes volume scattering, which reduces the overall measured reflectivity. The L-band signal can still partially penetrate the canopy layer even in dense forest. A coherent signal is received from perfectly flat surfaces, but surface roughness and vegetation increase the incoherent component of the received signal (RD-3: Dente et al., 2020; RD-4: Ren et al., 2021).

The freeze and thaw detection algorithm is based on previous studies conducted with TechDemoSat-1 observations at high latitudes (RD-5: Comite et al. 2020 and RD-6: Rautiainen et al. 2021), including the work performed during the six-month HydroGNSS system consolidation study.

2 DESCRIPTION OF THE ALGORITHM

The Freeze and Thaw (FT) module consists of two operational processors: Level 2 FT (L2FT) and Level 3 FT (L3FT) processors. The Scout-2 HydroGNSS PDGS Algorithm ICD (AD-1) refers to these two processors as L2OP-FT and L3OP-FT respectively. The L2 and L3 processors are compatible with each other. The L2 processor reads the L1b data and produces the track-based initial estimates of the ground state. L2FT also stores all relevant L1b variables and attributes in an intermediate data folder in Matlab format. The L3FT processor reads the stored data from the intermediate data folder to avoid duplication of effort. Level 3 FT estimates are provided at the polar EASE-2 grid for the Northern Hemisphere, in 25km x 25km pixel size. The FT module also contains a non-operational processor, the FT Reference Processor. The FT Reference Processor provides the frozen and thawed ground references used in the L2FT and L3FT processors. The high-level block diagram of the FT module is shown in Figure 2-1. The frozen and thawed ground references used in the released versions of the L2FT and L3FT processors are determined based on available TechDemoSat-1 (TDS-1) observations from 2018.

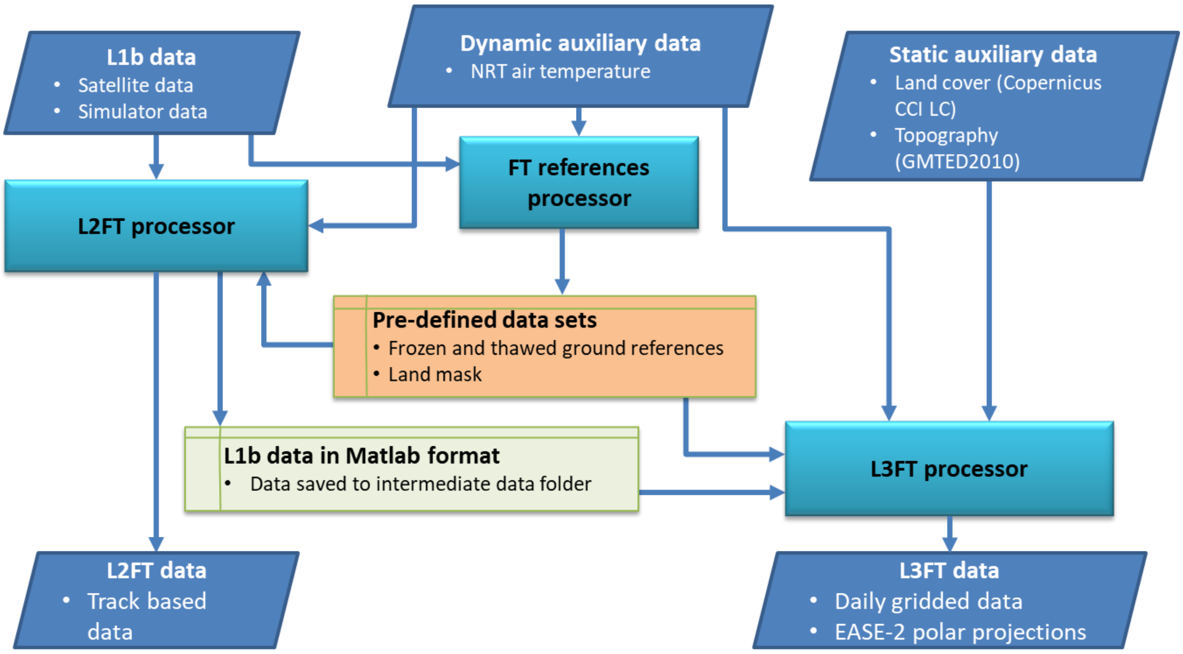

Figure 2-1. High level block diagram of the FT module. L2FT processor reads the L1b data, produces the track-based FT estimates, and stores all the relevant input variables to intermediate data folder in Matlab format. L3FT processor uses these intermediate data files as input data and provide daily level 3 FT estimates at the polar EASE-2 grid, in 25km x 25km pixel size. The FT Reference Processor is not operational and provides pixel-wise references for the frozen and thawed ground conditions.

2.1 INPUT DATA

2.1.1 L1b data

The FT processors use L1b data from the merged L1 data file (named as metadata_L1_merged.nc). More specifically, the FT processors use the incoherent LHCP channel, both E1/L1 and E5/L5 frequencies. The input variable used to estimate the soil FT state is the reflection coefficient at the specular point: <*ReflectionCoefficientAtSP*>. In addition, several other variables are required from the merged L1 data file. These are all listed in Table 1.

Table 1. L1b variables used in FT processors.

Variable name	Short description
<ReflectionCoefficientAtSP>	The calibrated reflection coefficient at the specular point
<SpecularPointLat>	The location of the specular point at the integration mid-point time (latitude)
<SpecularPointLon>	The location of the specular point at the integration mid-point time (longitude)
<SPIncidenceAngle>	The incidence angle of the reflection
<IntegrationMidPointTime>	The mid-point time between the start and the end of the DDM incoherent measurement's integration (UTC time stamp)
<LandType>	Land type at the specular point location retrieved from ESA CCI Land Cover
<SnowIceCover>	Snow and Ice type at the specular point location retrieved from NSIDC
<LowOverallQuality>	Low quality summary flag
<LowSNR>	Quality flag based on Signal to Noise Ratio
<LowAGSP>	Quality flag based on Antenna Gain towards Specular Point
<RFIFlag>	Radio Frequency Interference flag provided in NOC L1b_ext
<CFFlag>	Correlator anomaly flag provided in NOC L1b_ext
<Coherency>	Coherence flag provided in IEEC L1b_ext

2.1.1.1 L1b Quality flags

The quality flags of the L1b merged data are read and propagated through to the L2 output data files. They are not used to remove any data points from the L2 data.

2.1.2 Static auxiliary data

The FT processor requires two static auxiliary data sets: land cover and topography data. Additionally, two pre-defined data sets are required: land mask and references for the frozen and thawed ground. These two pre-defined data sets are provided by the processor developer. The data sets are listed in Table 2.

Table 2. Static auxiliary data for FT Processors.

Dataset	Processor	Description
Land cover data	L3	Copernicus CCI Land Cover product (RD-7)
Topography	L3	the USGS EROS Archive - Digital Elevation - Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) at ~1 km spatial resolution (30-arc-second spatial resolution), (RD-8),
Land mask	L2 & L3	Derived from the Copernicus CCI Land Cover product
Frozen and thawed ground references	L2 & L3	Determined from the TDS-1 observations over the Northern Hemisphere (RD-6).

The L3FT processor uses the CCI Land Cover product at a spatial resolution of 300m. The land cover data are reprojected to the EASE-2 projection of the Northern Hemisphere. To save computing time, the original land cover classes are aggregated into six classes and resampled to 900m pixel size using the majority value. The original and aggregated classes are listed in Table 3.

Table 3. Original CCI Land Cover product classes and the aggregated ones used in L3FT processor.

Aggregated land class	Original class id	Original classes
Croplands	10 - 40	• Cropland, rainfed • Cropland, irrigated or post-flooding • Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) • Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland
Low vegetation	100 - 150	• Mosaic tree and shrub (>50%) / herbaceous cover (<50%) • Mosaic herbaceous cover (>50%) / tree and shrub • Shrubland • Grassland
Forests	50 - 90	• Tree cover, broadleaved, evergreen, closed to open (>15%) • Tree cover, broadleaved, deciduous, closed to open (>15%) • Tree cover, needleleaved, evergreen, closed to open (>15%) • Tree cover, needleleaved, deciduous, closed to open (>15%) • Tree cover, mixed leaf type (broadleaved and needleleaved)
Wetlands	160 - 180	• Tree cover, flooded, fresh or brackish water • Tree cover, flooded, saline water • Shrub or herbaceous cover, flooded, fresh/saline/brackish water
Open water	210	• Water bodies
Others (urban area, bare, permanent snow and ice)	190, 200, 220	• Urban areas • Bare areas • Permanent snow and ice

The FT algorithm uses the GMTED2010 product with a spatial resolution of 30 arc seconds (~1 km). This included layers of mean and standard deviation of ground elevation in the lat-lon geographic grid. These data layers have been reprojected to the EASE2 projection for the Northern Hemisphere using the mean interpolation method, i.e. the weighted average of input pixels within a 1 km output pixel area. The output is restricted to latitudes between 26 and 80 degrees.

The land mask was derived from the land cover data. To produce the land mask, the CCI Land Cover product was first converted to the polar EASE-2 grid with a pixel size of 25 km x 25 km. Then all classes except open water were marked as land.The land mask is shown in Figure 2-2.

The frozen and thawed ground references are described in section 2.2.1

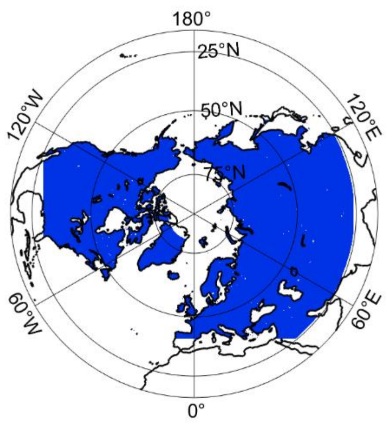

Figure 2-2. Land mask derived from the used land cover data and converted to the polar EASE-2 grid (25km x 25km pixel size).

2.1.3 Dynamic auxiliary data

The current processors use two metre air temperature data from the ECMWF Atmospheric Model high-resolution 10-day forecast dataset (Set I - HRES). In addition, the processors use snow extent data from the IMS Daily Northern Hemisphere Snow and Ice Analysis. The snow extent data is added as a snow/ice flag to the merged L1b data (AD-1). Dynamic auxiliary data are listed in Table 4.

Table 4. Dynamic auxiliary data for FT Processors.

Dataset	Processor	Description
Air temperature	L2 & L3	Air temperature (2-meter) data: ECMWF Atmospheric Model high resolution 10-day forecast (Set I - HRES) https://www.ecmwf.int/en/forecasts/datasets/set-i
Snow extent	L2 & L3	IMS Daily Northern Hemisphere Snow and Ice Analysis data provided by NSDIC, ftp://sidads.colorado.edu/DATASETS/NOAA/G02156/GIS/4km

The ECMWF Atmospheric Mode data provides new 10-day forecast data every day. FT processor uses only the first day forecast runs at 6 hour intervals (00/06/12/18 hh).

2.2 IMPLEMENTATION

The implementation of the FT algorithm is described in this section.

The FT algorithm is based on threshold detection; a set of new observations is compared to the pre-defined frozen and thawed ground references from the same area. The target ground is estimated to be in frozen or thawed condition if the observed variable is closer to the frozen or thawed ground reference, respectively. The functionality of the L2 and L3 processors is described in detail in this section.

2.2.1 Frozen and thawed ground references

The well-established frozen and thawed ground references are essential for the threshold-based FT algorithm. As discussed in Section 1, the reflection coefficient varies depending on the target surface. Conventionally, the frozen and thawed references are determined for grid-based data and for each grid cell independently (e.g. SMAP and SMOS FT algorithms). With GNSS-R, the changing geometry of the transmitter-receiver system adds a further challenge to the gridding of the data - no two observations are from the same target area or geometric configuration. To mitigate these challenges, spatial and temporal averaging is required when determining reflection coefficient maps on frozen and thawed ground conditions. The initial HydroGNSS references were derived from the TDS-1 data observed in 2018, with measurements taken in February and August for the frozen ground and thawed ground references, respectively. The reference maps are shown in Figure 2-3. The L2FT and L3FT processors are currently using these references derived from the TDS-1 observations. The final references need to be derived from the HydroGNSS observations. The reflection coefficient values vary spatially, with the highest values recorded over Scandinavia, eastern Canada and central Russia for both the thawed and frozen ground references.

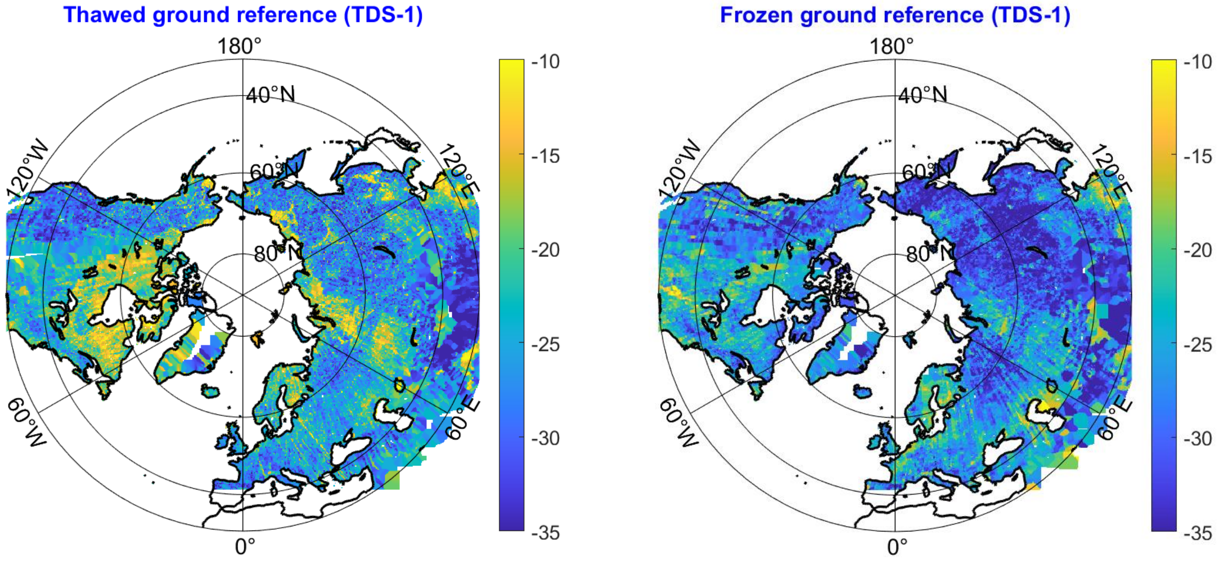

Figure 2-3. Thawed (left panel) and frozen (right panel) ground reference reflectivity maps in EASE2 projection (25 km x 25 km resolution) derived from TDS-1 observations.

2.2.2 L2FT processor

The L2FT processor is presented in this section. The functional diagram of the L2FT processor is shown in Figure 2-4.

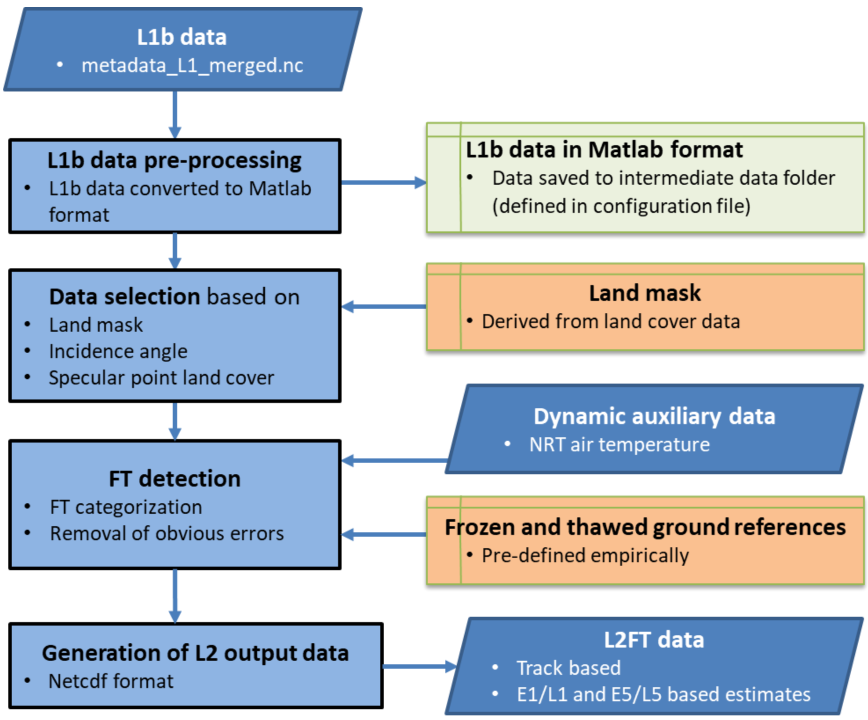

Figure 2-4. Functional diagram of the L2FT processor.

2.2.2.1 Data pre-processing

The L2FT processor reads the L1b data, converts all relevant variables to Matlab format and saves them for further use with the L3FT processor.

2.2.2.2 Data selection

Observations are selected based on the angle of incidence <*SPIncidenceAngle*>, the land type <*LandType*> and the location of the specular point. The location is compared with the land mask derived from the land cover data (Section 2.1.2). If the specular point is not over land, the observation is not included. A summary of the criteria used to filter the data is given in Table 5.

Table 5. L2FT data selection criteria

Description	Criteria	Notes
Specular point incidence angle	Observations filtered out if <*SPIncidenceAngle*> is over 40°	This criterion can be changed in the configuration file
Specular point land type	All data having <*LandType*> information over open water and permanent ice are filtered out.	
Specular point location	If the given SP location (<*SpecularPointLat*>, <*SpecularPointLon*>) is outside the land mask, observation is filtered out.	

2.2.2.3 FT detection and removal of obvious errors

By default, observations are categorized into two levels – frozen or thawed soil. The selection is based on comparing the observed reflection coefficient at the specular point (<*ReflectionCoefficientAtSP*>) with the frozen and thawed ground references and applying a predefined FT threshold value. The default method consists of the following steps:

1. Identify the EASE-2 grid cell from which the <ReflectionCoefficientAtSP> observation originated.
2. Collect the frozen and thawed soil reference values from the predefined reference data for the identified EASE-2 grid cell.
3. Scale the observed <ReflectionCoefficientAtSP> to the reference values, with a value of 1 corresponding to the thawed soil reference and a value of 0 corresponding to the frozen soil reference.
4. Compare the scaled value with the threshold value.
5. Determine the soil state status.

The default FT threshold is set to 0.5 so that the ground is considered frozen or thawed if the observed <ReflectionCoefficientAtSP> is closer to the frozen or thawed ground reference respectively. The threshold value is provided in the configuration file and can be changed if another value proves more appropriate. There is also an option to use two separate FT thresholds to provide three levels of ground conditions: thawed, frozen and an intermediate condition that can be interpreted as partially frozen ground over the pixel. The final threshold(s) will be determined when the product is validated.

The removal of obvious errors is performed on the categorised FT data in order to avoid obvious false detections during the summer period. The error removal procedure is as follows:

1. Identification of potential summer period using snow extent data: (a) If the SnowIceCover flag is NOT present in the L1b data, only default summer months (defined in the configuration file) are considered as potential summer period. By default, these are June, July and August. (b) If the SnowIceCover flag is available, obvious error removal is NOT performed if snow on ground is true. If the observation is from a land area without snow, it is considered to have been taken during the potential summer period.
2. If the potential summer period is not considered, error removal is NOT applied

3. For the potential summer period observations, dynamic air temperature data are checked. If all three of the following criteria are met, the observations are assumed to be from summer soil conditions and the frozen soil category is not allowed. The criteria are
    a. The most recent air temperature observation from the area is above Th_1
    b. The average of the X most recent temperature observations is above Th_2
    c. The ratio of the number of observations above zero to the number of total observations is greater thanTh_3.

    The three thresholds Th_1, Th_2 and Th_3 are defined in the configuration file. The default values are 3°C, 3°C and 0.95 respectively. The default number of days of ECMWF air temperature observations to be averaged - symbol X in criterion (b) - is 10. This parameter is also defined in the configuration file.

2.2.2.4 Generation of L2 FT output data

The output data are produced in NetCDF format with the same structure as in the L1b data file. Separate FT estimates are provided for the E1/L1 and the E5/L5 channel data if both are available. The NetCDF file also contains the time and location information for the specular points and the data quality flags derived from the L1b data. All L2FT output variables are listed in Table 6. More detailed information on the file structure of the output data can be found in the User Guide (AD-4).

Table 6. The output data of the L2FT product

Variable	Description	Output values
<FreezeThawState>	The estimated soil freeze and thaw state using E1/L1 channel data	1 – thaw 3 – frozen 255 – no data 2 – currently not used, reserved for partially frozen/unknown state
<FreezeThawState ChannelL5E5>	The estimated soil freeze and thaw state using E5/L5 channel data	Same as above
<IntegrationMidPointTime>	The UTC timestamp of the Mid-point location between the start and the end of the DDM incoherent measurement's integration (AD-1)	The number of days from January 0, 0000
<SpecularPointLon>	The longitude of the specular point at *<IntegrationMidPointTime>*	Degrees
<SpecularPointLat>	The latitude of the specular point at *<IntegrationMidPointTime>*	Degrees
<LowOverallQuality>	Low quality summary flag, provided in L1b data for the E1/L1 channel.	1 – poor quality 0 – good quality
<LowOverallQualityL5E5>	Low quality summary flag, provided in L1b data for the E5/L5 channel.	1 – poor quality 0 – good quality
<LowSNR>	Quality flag for the low SNR value, threshold provided in L1b data for the E1/L1 channel.	1 – poor quality 0 – good quality
<LowSNRL5E5>	Quality flag for the low SNR value, threshold provided in L1b data for the E5/L5 channel.	1 – poor quality 0 – good quality
<VeryLowSNR>	Quality flag for the very low SNR value, threshold provided in L1b data for the E1/L1 channel.	1 – poor quality 0 – good quality
<VeryLowSNRL5E5>	Quality flag for the very low SNR value, threshold provided in L1b data for the E5/L5 channel.	1 – poor quality 0 – good quality
<LowAGSP>	Quality flag for the low AGSP value, threshold provided in L1b data for the E1/L1 channel.	1 – poor quality 0 – good quality

<LowAGSPL5E5>	Quality flag for the low AGSP value, threshold provided in L1b data for the E5/L5 channel.	1 – poor quality 0 – good quality

2.2.3 L3FT processor

The L3FT processor is introduced in this section. The functional diagram of the L3FT processor is shown in Figure 2-5.

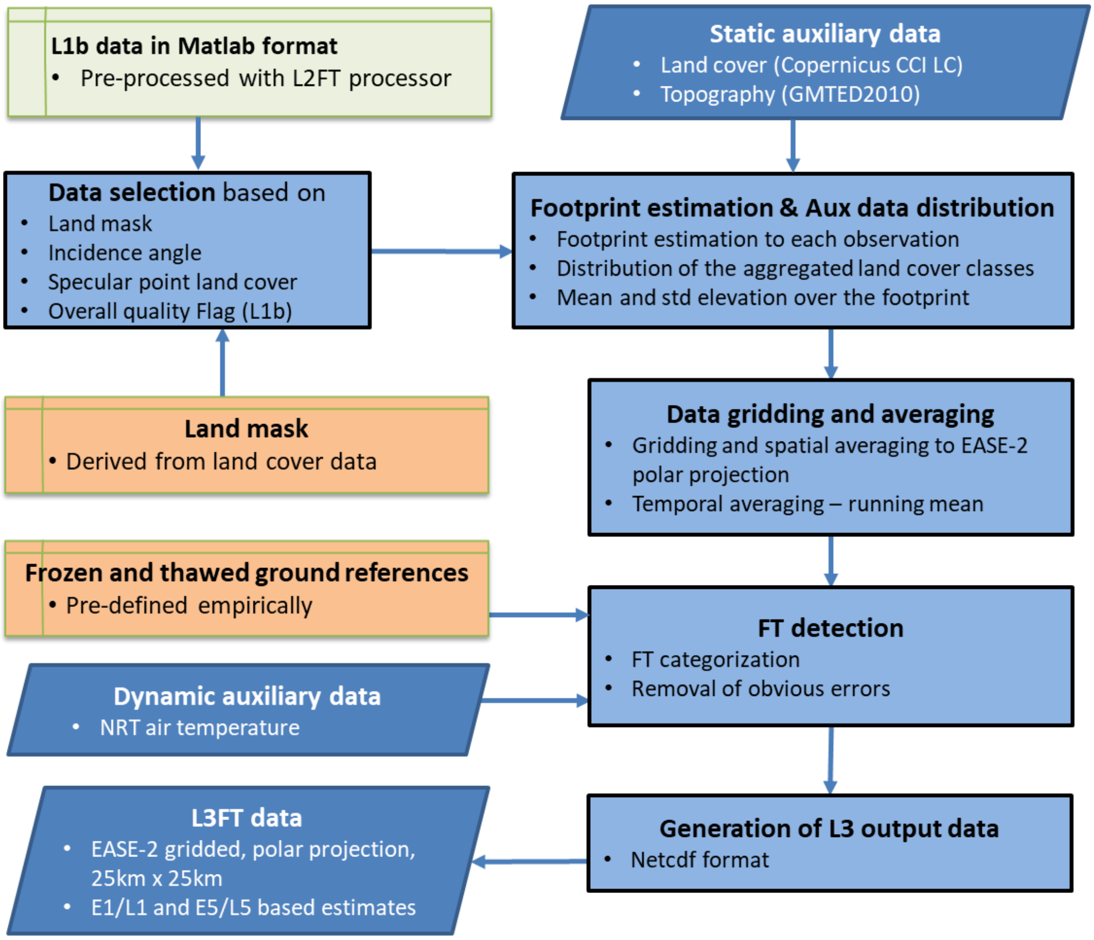

Figure 2-5. Functional diagram of the L3FT processor.

2.2.3.1 Data selection

Observations are selected using the same criteria as in the L2FT processor with additional criteria based on the L1b quality flags. Four quality flag inputs can be used: *<LowOverallQuality>, <VeryLowSNR>, <LowSNR>, and <LowAGSP>*. The use of these flags is defined in the configuration file. By default, only *<LowOverallQuality>* flag information is used. All data that are flagged as low quality and the flag defined to be used in the configuration file will be discarded. A summary of the criteria used to filter the data is given in Table 7.

Table 7. L3FT data selection criteria

Description	Criteria	Notes
Specular point incidence angle	Observations filtered out if <*SPIncidenceAngle*> is over 40°	This criterion can be changed in the configuration file
Specular point land type	All data having <*LandType*> information over open water and permanent ice are filtered out.	
Specular point location	If the given SP location (<*SpecularPointLat*>, <*SpecularPointLon*>) is outside the land mask, observation is filtered out.	
Low overall quality	If the <*LowOverallQuality*> flag in the L1b data indicates low quality (value 1), the observation is filtered out	The use of this flag can be changed in the configuration file, **by default it is used**
Very low SNR	If the <*VeryLowSNR*> flag in the L1b data indicates low quality (value 1), the observation is filtered out	The use of this flag can be changed in the configuration file, **by default it is NOT used**
Low SNR	If the <*LowSNR*> flag in the L1b data indicates low quality (value 1), the observation is filtered out	The use of this flag can be changed in the configuration file, **by default it is NOT used**
Low AGSP	If the <*LowAGSP*> flag in the L1b data indicates low quality (value 1), the observation is filtered out	The use of this flag can be changed in the configuration file, **by default it is NOT used**

2.2.3.2 Footprint estimation and the distribution of the auxiliary data

The effective footprint area of any given <*ReflectionCoefficientAtSP*> is not clearly known. It depends on the properties of the target area, mainly surface roughness, but also terrain topography and vegetation. For coherent reflections (smooth surfaces) the signal originates mainly from the first Fresnel zone (FFZ). For incoherent reflections (rough surfaces), the affected area is much larger. Proper estimation would require accurate information on the coherence/incoherence of the scattered field. For data gridding purposes, the footprint area of each <*ReflectionCoefficientAtSP*> is estimated (by default) to be a circular area with a radius of 5 km around the given specular point location (defined for the <*IntegrationMidPointTime*>). The satellite moves during the integration of the observations, which extends the footprint area along the direction of motion (note that FT processors use the incoherent channel). This is illustrated in Figure 2-6.

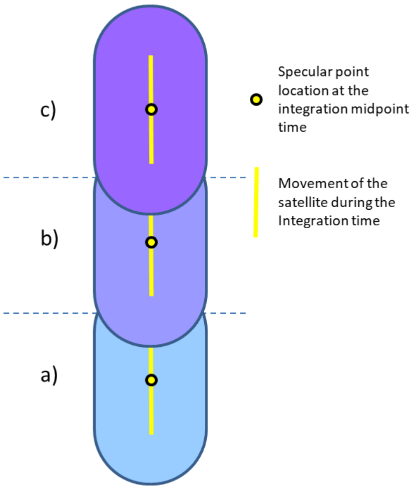

Figure 2-6. Simplified illustration on the estimated footprint area considering the movement of the satellites during the integration time.

The distribution of land cover classes is determined for each estimated footprint. Land cover data at 300m spatial resolution are first aggregated into 6 classes and resampled to 900m pixel size using the majority value. The aggregated land classes are cropland, low vegetation, forest, wetland, open water and other. The land cover fraction of each aggregated land class is defined for each footprint using SAGA software. Similarly, the mean elevation and its standard deviation are calculated for each Footprint area using data from the GMTED2010 digital elevation model.

The land cover distribution of each footprint is used to determine how well each footprint represents the EASE-2 grid cell in which it is located. The land cover distribution of each EASE-2 grid cell was similarly determined using aggregates of 6 land classes. Representativeness is defined using Jensen-Shannon divergence, a method that measures the similarity between two probability distributions.

2.2.3.3 Data gridding and averaging

Data are spatially averaged to the EASE-2 polar grid with a pixel size of 25 km x 25 km. Each observation is localised to the correct EASE-2 grid cell based on the *<SpecularPointLat>*, *<SpecularPointLon>* information at *<IntegrationMidPointTime>*.

In addition, the data is averaged over a 10 day window (default, can be changed in the configuration file). Rolling averaging of the last 10 days is applied to produce daily maps of soil state conditions with minimal latency (typically one day latency is expected).

The data gridding and averaging process consists of the following steps:

1. Data localisation check: All data within the temporal averaging window are localised to EASE-2 grid cells.
2. Data representativeness check: The similarity between the footprint and the EASE-2 grid cell land cover distributions is determined using the Jensen-Shannon (JS) divergence method. All data that do not pass the similarity test are discarded. In addition, all data from footprints with too large a proportion of open water are discarded. The default JS similarity threshold is 0.1 and the default maximum value for the open water fraction is 5%. These can be changed in the configuration file.
3. Data are averaged over the EASE-2 polar grid with a pixel size of 25 km x 25 km. All observations are equally weighted. Reflection coefficient, mean and standard deviation are calculated. The number of observations and the time of the last observation are stored.

2.2.3.4 FT detection and removal of obvious errors

By default, the soil state is categorized into two levels – either the soil was frozen or thawed. The selection is based on comparing the reflection coefficients of the gridded data with the predefined frozen and thawed ground references. The gridded reflection coefficient values are scaled to the reference values, with a value of 1 corresponding to the thawed soil reference and a value of 0 corresponding to the frozen soil reference. The scaled value is compared to a predefined FT threshold. By default, the FT threshold is set to 0.5 so that the soil is estimated to be frozen or thawed if the reflection coefficient is closer to the frozen or thawed soil reference, respectively. The threshold value is specified in the configuration file and can be changed if another value is more appropriate. There is also an option to use two separate FT thresholds to provide three levels of soil states: thawed, frozen, and an intermediate state that can be interpreted as partially frozen soil over the pixel. The final threshold value(s) will be determined when the product is validated.

The procedure for removing obvious errors from L3FT data is the same as for L2FT data. Error removal is performed on the categorized L3 FT data. The error removal procedure is as follows:

1. Identification of potential summer period using snow extent data: (a) If the SnowIceCover flag is NOT present in the L1b data, only default summer months (defined in the configuration file) are considered as potential summer period. By default, these are June, July and August. (b) If the SnowIceCover flag is available, obvious error removal is NOT performed if snow on ground is true. If the observation is from a land area without snow, it is considered to have been taken during the potential summer period.
2. If the potential summer period is not considered, error removal is NOT applied
3. For the potential summer period observations, dynamic air temperature data are checked. If all three of the following criteria are met, the observations are assumed to be from summer soil conditions and the frozen soil category is not allowed. The criteria are

a. The most recent air temperature observation from the area is above Th_1

b. The average of the X most recent temperature observations is above Th_2

c. The ratio of the number of observations above zero to the number of total observations is greater thanTh_3.

The three thresholds Th_1, Th_2 and Th_3 are defined in the configuration file. The default values are 3°C, 3°C and 0.95 respectively. The default number of days of ECMWF air temperature observations to be averaged – symbol X in criterion (b) – is 10. This parameter is also defined in the configuration file.

2.2.3.5 Generation of L3 FT output data

The output data are produced in NetCDF format, including all parameters defined in AD-1. Data will be provided in the EASE-2 polar grid at 25 x 25 km pixel size (equivalent to 720 x 720 cells) for the Northern Hemisphere for all areas subject to seasonal soil freezing. The data will consist of:

- FT data in discrete levels, at least two levels: grid cell is estimated to be in either thawed or frozen condition. Data matrix size: 720 x 720, values: 1 – thawed, 3 – frozen, 255 – no data. Value 2 is reserved for partially frozen ground state.
- Date stamp for the gridded product.
- The number of days since the last observation was collected from each EASE-2 grid cell. The size of the data matrix is 720 x 720. No data value is 255.
- A binary coded quality flag (only dummy data provided with the current processor). Data matrix size is 720 x 720. The quality information is coded in an 8-bit unsigned integer and contains the information described in Table 8.

The number of days since the last observation is a parameter important since the HydroGNSS satellite does not provide daily coverage. The L3 FT data are averaged using a 10-day moving window, meaning that each daily FT map includes observations from the past 10 days. Without knowing the age of the most recent observation for a grid cell, users might mistakenly assume the FT status is current, when in fact, it could be several days old. This information is especially critical during transitional periods, when soil is freezing or thawing. For example, if there is a gap of 7 days between two observations in a grid cell, the exact timing of a freeze-thaw transition could be uncertain by up to 7 days.

Table 8. The information contents and the format of the L3FT quality flag

The quality flag is an 8-bit unsigned integer: **eddccbba**, where a is the least significant bit and e the most significant bit		
Bit	Name	Description
a	Data available	1 – Data available 0 – No data available
bb	Days since last observation	11 – Maximum of one day since last observations was acquired 10 – Two to four days since last observation was acquired 01 – Five to eight days since last observation was acquired 00 – More than eight days since last observation was acquired
cc	Total number of observations averaged	11 – At least 25 observations 10 – Between 15 and 24 observations 01 – Between 5 and 14 observations 00 – Less than 5 observations
dd	Std of the reflection coefficient with respect to the difference between the thawed and frozen reference values.	11 – Std is 0.1 or less than the frozen – thawed ground difference 10 – Std is between 0.33 and 0.1 01 – Std is between 0.67 and 0.33 00 – Std is larger than 0.67
e	Reflection coefficient value in respect to the frozen and thawed reference levels	1 – Reflection coefficient value is closer to the reference value than to the threshold value. 0 – Reflection coefficient value is closer to the threshold value than to the reference value.

A quality flag value of 237 coverts to bit string '11101101'. The information it contains is as follows:

- a = '1', there are data available.
- bb = '10', the latest observation on the pixel is acquired at maximum four days ago but at least two days ago (user can get the exact information when the data was last time received from the *<dnum>* variable).
- cc = '01', total number of observations is between 5 and 14.
- dd = '11', the standard deviation of the reflection coefficient is less than 0.1 with respect to the difference between the thawed and frozen reference values. If the difference between the reference values is 8 dB, the std has been 0.8 dB or better. This is given in respect to the reference values, as for areas where the thawed and frozen ground conditions provide larger dynamics, larger variation in observations are acceptable.
- e = '1', The acquired reflection coefficient value is closer to the reference value than to the threshold value.

The L3 data will be updated daily. Input data are provided every six hours. When all data from one day are available, the data are processed. All output variables from the L3FT processor are listed in Table 9.

Table 9. The output data of the L3FT product

Variable	Description	Output values
<FreezeThawState>	Estimated soil freeze and thaw state using E1/L1 channel data provided in the polar EASE-2 grid at 25 x 25 km pixel size (720 x 720 pixels)	1 – thaw 3 – frozen 255 – no data value 2 – currently not used, reserved for partially frozen/unknown state
<FreezeThawStateChannelL5E5>	Estimated soil freeze and thaw state using E5/L5 channel data provided in the polar EASE-2 grid at 25 x 25 km pixel size (720 x 720 pixels)	Same as above
<dnum>	Days since last E1/L1 channel observation provided in the polar EASE-2 grid at 25 x 25 km pixel size (720 x 720 pixels)	Full number of days since last observation: 0-254 255 – no data value
<dnumChannelL5E5>	Days since last E5/L5 channel observation provided in the polar EASE-2 grid at 25 x 25 km pixel size (720 x 720 pixels)	Full number of days since last observation: 0-254 255 – no data value
<QualityFlag>	A binary coded quality flag for L3FT product using E1/L1 channel	See Table 8
<QualityFlagChannelL5E5>	A binary coded quality flag for L3FT product using E5/L5 channel	See Table 8
<date>	Date stamp of the daily gridded product	Date format 'yyyy-mm-dd'
<crs>	The coordinate reference system for the EASE-2 projection	
<x>	x coordinate of the EASE-2 projection	Units is meters
<y>	y coordinate of the EASE-2 projection	Units is meters

2.2.4 L3FT reference processor

The L3FT reference processor (L3FTR) is introduced in this section. The functional diagram of the L3FTR processor is shown in Figure 2-7. The purpose of the L3FTR processor is to provide the frozen and thawed soil references for the operational L2 and L3 processors. The L3FTR processor is not operational, but is operated manually when sufficient data is available from summer and/or winter conditions. Ideally, at least one full year of HydroGNSS data should be available to obtain reliable references. The initial references for

HydroGNSS were derived from GNNS-r data from the TDS-1 mission. The L3FTR processor consists essentially of the same functional blocks as the L2FT and L3FT processors.

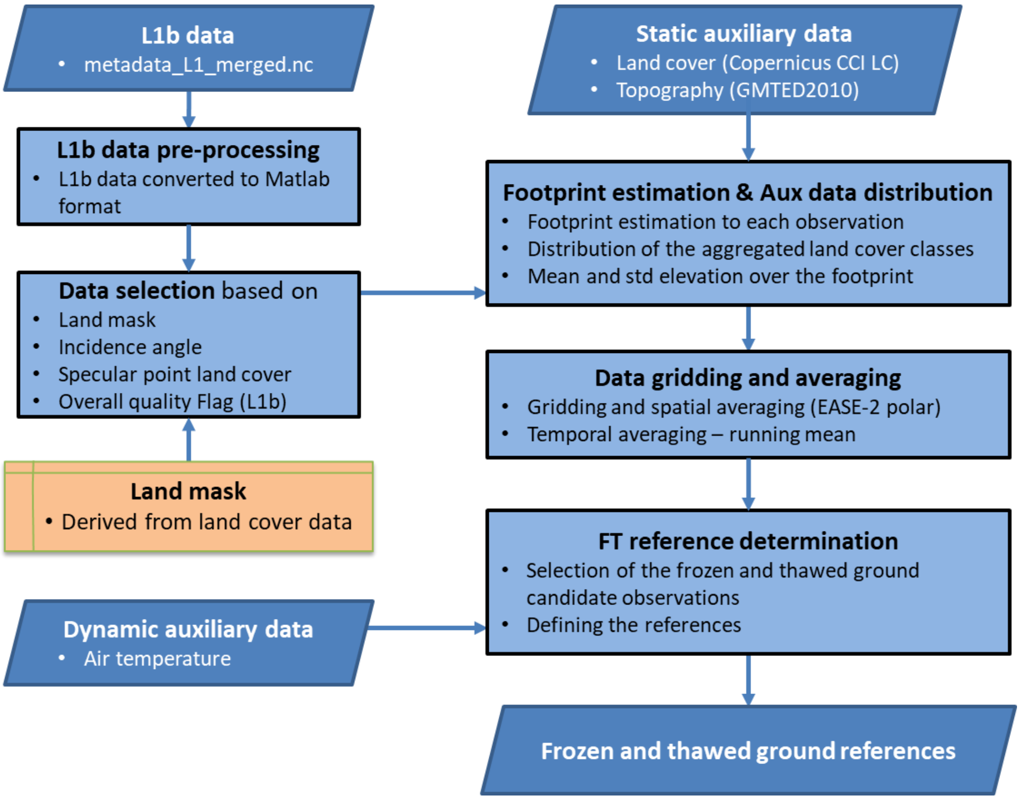

Figure 2-7. Functional diagram of the Reference Processor.

2.2.4.1 Data pre-processing

L1b data pre-processing is identical with the one in the L2FT processor. L1b data is read and converted to Matlab format and saved to intermediate data folder.

2.2.4.2 Data selection

Data selection is identical with the L3FT processor.

2.2.4.3 Footprint estimation and the distribution of the auxiliary data

The footprint estimation and the distribution of the auxiliary data is identical with the L3FT processor.

2.2.4.4 Data gridding and averaging

Data gridding and averaging is identical to the L3FT processor. Note that similarly to the L3FT processor, a data representativeness check is performed during this step; the similarity between the footprint and the EASE-2 grid cell land cover distributions is determined using the Jensen-Shannon (JS) divergence method. A stricter JS threshold for the references can be set in the configuration file.

2.2.4.5 FT reference determination

The frozen and thawed ground references are empirically determined unique values for each EASE-2 pixel. The references are determined from the gridded and averaged FT data. Before launch, they are defined from the GNSS-R measurements of the TDS-1 mission. After launch, they are redefined using HydroGNSS data.

The FT algorithm uses dynamic auxiliary data to select all data that potentially originate from the frozen and thaw ground conditions. These auxiliary data are two metre air temperature and snow cover extent (provided in the L1b data). The logic for selecting the candidate data to define the references is illustrated in Figure 2-8 and described below:

a) New data available – Only pixels that have been measured are included. Daily coverage is much less than 100%. However, as the algorithm uses a rolling average, the freeze and thaw map covers most of the northern hemisphere. If the latest measurement of a pixel is not from the current day, it has already been checked and evaluated as a candidate observation.

b) Correct period – The first estimate of whether the observation is potentially from thawed or frozen ground is based on the season. By default, the observation is potentially from thawed ground conditions in July, August and September. Similarly, frozen ground conditions are measured in January, February and March.

c) Snow condition – The snow information provided in the L1b data is used to further determine the ground conditions: snow on the ground indicates a potentially frozen ground condition and snow off the ground indicates a potentially thawed ground condition.

d) Air temperature threshold – The latest air temperature (daily average) and the average of the last 10 days air temperature within an EASE-2 pixel must be below -5°C to define the measurement as potentially frozen ground. Similarly, the latest air temperature (daily average) and the average of the last 10 days air temperature within an EASE-2 pixel must be above +5°C to define the measurement as potentially thawed ground.

The L3FTR processor collects all data defined as potentially originating from thawed or frozen ground conditions. Theoretically, the measured reflection coefficient will be at its lowest value in solidly frozen ground conditions as the permittivity of the ground is very low. Similarly, the reflection coefficient will be highest in unfrozen and moist soil conditions when the permittivity of the soil is high due to a high concentration of free liquid water in the soil. The most extreme values for thawed and frozen soil conditions are selected and the 50 most extreme values are used as a base. From these selected values, the median is chosen to represent the frozen and thawed soil references. The first good quality references are expected after the HydroGNSS satellites have been operational for a full summer and winter period. It is unlikely that enough data will be collected for a full set of 50 extreme values for one year. The references will need to be checked and updated after each full summer and winter period during the mission

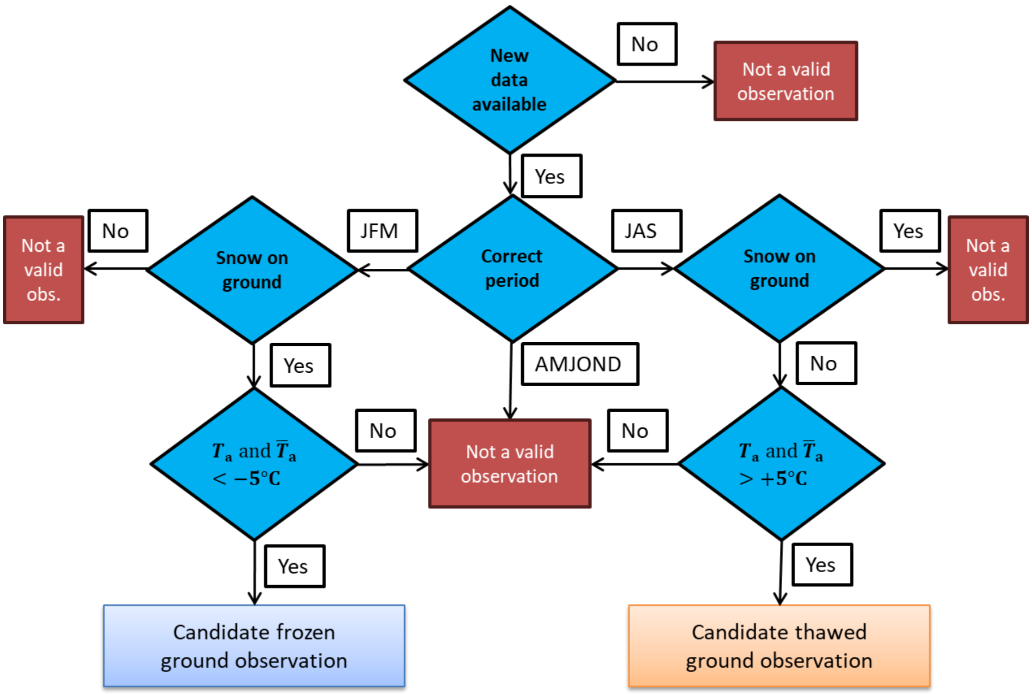

Figure 2-8. Logic for selecting data to define pixel-wise ground references. The threshold temperatures (+5°C and -5°C) are defined in the configuration file and can be adjusted. The initial decision for the correct period is based purely on the season, the letters on the boxes indicate the months as described in the text above.

3 VALIDATION

3.1 PREVIOUS WORK SUPPORTING GNSS-R FOR FREEZE/THAW DETECTION

The use of GNSS reflectometry (GNSS-R) for detecting soil freeze/thaw (F/T) states is based on well-established principles of dielectric permittivity change between frozen and thawed soils. This change is significant because ice and liquid water have distinct dielectric properties: frozen soils, which contain ice, exhibit much lower permittivity compared to thawed soils, which contain liquid water. This permittivity contrast translates into a detectable shift in GNSS-R reflectivity, enabling soil freeze/thaw transitions to be identified.

Although GNSS-R data over freeze/thaw-affected land areas is limited, studies using data from missions like TechDemoSat-1 (TDS-1) and CYGNSS have demonstrated the feasibility of this approach. Notably, CYGNSS observations over the Tibetan Plateau and the Andes (RD-9, RD-10, RD-11) and TDS-1 data from high-latitude regions (RD-5, RD-6) confirmed GNSS-R's sensitivity to soil freeze/thaw transitions. Each study successfully identifies surface soil state changes and provides evidence on using GNSS-R measurements for this purpose.

Publications (RD-6, RD-9, RD-10) demonstrate the feasibility of GNSS-R-based F/T detection algorithms relying on threshold-based approaches that use predefined reference values for frozen and thawed soil conditions. Publications (RD-6, RD-9) have compared the GNSS-R based FT results with SMOS and SMAP FT products showing good correlation results, particularly on homogeneous landscapes. The results and findings of the above-mentioned studies (1) suggest that GNSS-reflectometry is in general a very promising method for monitoring surface soil freeze and thaw cycles and (2) justify the chosen approach presented in this document.

3.2 PRE-LAUNCH VALIDATION

The FT pre-launch calibration activities are based on the use of simulated data from the end-to-end simulator (E2ES) over the Sodankylä core site and the use of TDS-1 observations from February 2018 and August 2018. Considering the nature of the threshold-based detection algorithm, the pre-launch validation focuses on two aspects:

1. **Dynamic Range in Reflectivity Due to Soil Freezing and Thawing**: The key to successful F/T detection lies in determining the dynamic range in reflectivity caused by the freeze/thaw transitions. It is essential to quantify how well the reflectivity values differentiate between frozen and thawed soil conditions.
2. **Variability in Measurements Due to Changing Geometry**: Another important factor is the variation in measurements originating from the same grid cell but different locations within the grid cell due to the changing geometry of the GNSS-R transmitter-receiver system. When gridding the data, it is important to understand how these variations affect the final product and ensure consistency across the grid.

The E2E simulator is used to generate frozen and thawed ground references over the core site under different input parameter conditions. However, the E2E simulator is limited in its ability to account for the variability of all the influencing input parameters and thus provide good initial estimates of the expected dynamic range of reflectivity. With the E2E simulator, the main input parameters (soil temperature, vegetation and roughness) are varied using the Monte Carlo method to obtain the simulated frozen and thawed ground references.

In RD-6, the variability of measurements in general was investigated for forested and lightly vegetated areas over large regions. The variability due to changing geometry within a grid cell can be further investigated using the TDS-1 data, although the amount of data is limited. The February 2018 and August 2018 TDS-1 data have been converted to HydroGNSS format. Using this data, grid cells with the most data and from different transmitting satellites and orbits will be analysed to determine variability. The focus is on the permafrost tundra and boreal forest zone.

3.3 POST-LAUNCH VALIDATION

Post-launch validation will emphasize verifying the algorithm's performance using real satellite data, complemented by comparisons with in-situ observations and satellite-derived reference products.

3.3.1 Comparison with In-situ Data

Core Validation Sites: Validation will employ in-situ measurements from core sites such as Sodankylä. These sites provide comprehensive data on soil moisture, temperature, and freeze/thaw states, which will be compared against satellite-derived outputs.

Seasonal Transitions: Emphasis will be placed on capturing the critical seasonal transitions in soil states (autumn and spring). This will involve gathering comprehensive soil moisture and temperature data from sources such as the International Soil Moisture Network (ISMN).

3.3.2 Cross-validation with Other Satellite Products

SMAP and SMOS Data: Use data from other satellite missions, such as SMAP and SMOS, which provide freeze/thaw and soil moisture products. Cross-validation will help verify the performance of the HydroGNSS F/T algorithm across various geographic regions and soil conditions.

3.3.3 Assessment of Measurement Geometry Variations

Gridded Data Consistency: Post-launch validation will include further assessment of how varying measurement geometries within the same grid cell impact the final freeze/thaw detection product. The consistency of gridded data from different satellite passes will be analysed.

4 ROADMAP FOR FUTURE IMPLEMENTATIONS

This section presents future implementations that will need to be further investigated as HydroGNSS data become available. Both are identified as future evolutionary steps.

4.1.1 Use of footprint DEM distribution

The L3 FT processor estimates the footprint size for each incoherent reflection coefficient value. It then determines the distribution of land cover classes and the mean and standard deviation of elevation within each footprint (Section 2.2.3.2). The land cover information is used to determine how well the footprint represents the EASE-2 grid cell with which it overlaps (Section 2.2.3.3). However, the elevation information is not used. Further studies are needed to define if the elevation information adds value to the soil condition detection and how best to use it.

4.1.2 Use of Coherency data from IEEC

The <*Coherency*> provided in IEEC L1b data can provide information on the footprint area; the footprint area is expected to be the largest for completely incoherent observations and the smallest for coherent observations. The footprint radius can be adjusted based on the <*Coherency*> information. However, it is unclear how the <*Coherency*> information relates to the footprint size and this needs to be further investigated when HydroGNSS data are available.

4.1.3 Use of NOC L1b_ext processor output products

The NOC L1b processor output products <*CFFlag*> and <*RFIFlag*> could provide additional information on the RFI contaminated or bad quality observations. The use of these data needs to be investigated when the satellite is providing data.

Appendix A – List of the FT processor versions.

Table 10 lists the FT processor versions and their differences regarding the main algorithm functions. All processor versions are compatible with the E2ES and PDGS data. The source of the data, whether it is satellite or simulator, needs to be indicated in the configuration files. Some options and functions may not be available, if E2ES data are used, details are provided in the User Guide documentation.

Table 10. The placeholder, alpha, and beta processor version difference regarding the main algorithm functions. The placeholder and alpha processors are the last release versions and the beta processor the latest release.

Main algorithm functions	Placeholder processor (PP)	Alpha processor (AP)	Beta Processor (BP)	Operational version	Future evolutionary steps
Output product	Only L3 gridded product	Only L3 gridded product	Two processors, two output products: **L2FT** processor: L2 track product, **L3FT** processor: L3 gridded product	Three processors, three output products: **L2FT** processor: L2 track product, **L3FT** processor: L3 gridded product, **L3FTR** processor: Gridded frozen and thaw ground references	No changes
Input data selection	Input data selected only based on incidence angle and specular point location	Same as PP	Land mask added to select observations. Specular point land cover used to remove obs. over open water. Bad quality data removed based on L1b data overall quality flag.		Use of NOC L1b_ext processor output products: <*CFFlag*> and <*RFIFlag*> (Section 4.1.3).[(2)]
Footprint area estimation	Circular area of 5 km radius around each specular point	Consider also the satellite along track movement and the integration time	Same as AP		Footprint size estimated based on the <*Coherency*> information (Section 4.1.2). [(3)]
Distribution of auxiliary data within the footprint area	Land cover fractions calculated from CCI LC data over each footprint area	Same as PP	Topography information (elevation mean and std) added to the footprint area.	No changes	No changes
Representa-tiveness check	Observations localised to the EASE-2 grid cells based on specular point location. Observations with open water fraction higher than 5 % are discarded	Same as PP	No changes.	Land class data used for the similarity check between the estimated footprint area and the EASE.2 grid cell area.	Use of topography data (Section 4.1.2). [(1)]

Spatial averaging (gridding)	All qualified observations within each EASE-2 grid cell averaged	Same as PP	Same as PP	No changes	No changes
Temporal averaging	Temporal averaging using running mean. The output are daily gridded files	Same as PP	Same as PP	No changes	No changes
Frozen and thaw ground references	Defined separately for Sodankylä test site using simulated data with E2ES	Same as PP	Defined empirically from the TDS-1 observations to land areas above 55°N latitude.	Defined using L3FTR processor. Possible to use also E2ES data, if following instructions given in the User Guide.	
Removal of obvious errors	Not implemented, not relevant with simulated E2ES data	Same as PP	No changes.	Dynamic auxiliary data are used to remove the obvious error	
FT detection and error estimation	FT state defined in two levels: either frozen or thaw, threshold in the middle between frozen and thaw reference reflectivity. No FT data quality flag information implemented	Same as PP	FT estimates are produced also using the E5/L5 channel input data Number of days since last observation added to the L3FT product.	L3FT quality flag (section 2.2.3.5, Table 8).	

Appendix B Timeline for the future implementations

Below are given the timeline of the future evolutionary steps indicated in Section 4 and Table 10.

The similarity check between the estimated footprint and the EASE-2 grid cell **(1)**. The similarity check using topography data will be investigated with HydroGNSS data and implemented in the future evolutionary processor versions.

NOC L1b output products: <*CFFlag*> and <*RFIFlag*> **(2)**. The NOC L1b outputs may provide additional information on the input data quality. The possibility to use them will be included in the future processor versions, when this input data is available.

Use of IEEC output product: <*Coherency*> **(3)**. The <*Coherency*> information may be used to estimate the size of the footprint on ground. This will be tested when there are live HydroGNSS data available after the launch.