Doc. no: SW-ASMV-DD-IPGP-0012
Rev: 1E
Date: 2023-03-24

Swarm Experimental ASMV Project
ASM Whistler Product Definition for Operational Release of Data

Institut de Physique du Globe de Paris (IPGP)
Commissariat à l'Energie Atomique (CEA)
Centre National des Etudes Spatiales (CNES)

Prepared:

Pierdavide Coïsson Date 2023-03-24
Associate Physicist

Checked and authorized:

Gauthier Hulot Date 2023-03-24
Team Leader

Prepared:

Louis Chauvet Date 2023-03-24
Research engineer

Record of Changes

Reason	Description	Rev	Date
Initial vers.	First version	1	2021-03-19
Mails from Poul Erik 2021-03-30, Lars Toffner-Clausen and Enkelejda Qamili 2021-04-12	Finalisation of the file formats, minor corrections and editorial changes in sections 1.1, 2.2, 4.2, 4.3, 5.4,5.5, 6, 7 and Table 4.1.	1A	2021-04-29
Mail from Lars Toffner-Clausen 2021-05-05 and Enkeleida Qamili 2021-05-06	Update of Table 4.1 and inclusion of a footnote.	1B	2021-05-10
Mail from Lars Toffner-Clausen 2021-05-19	Addition of an overview table in Section 3.	1C	2021-06-08
Typo corrections and additions to known limitations	Additional information provided in section 1.1 specifying that the current version of the document applies to WHIxEVT_2_ version 0102 data product. Correction of units for PSD in Table 4.1, for whistler intensity in Section 5.7 and Figures 3-2, 5-2, 5-4 and 6-1 (now 6-3). Additional explanations on limitations of whistler characterisation in Section 6, Figures 6-1 and 6-2 added. Updated RD list with final report of ILGEW project.	1D	2022-01-31
Updated version corresponding to generic delivery of 0103 and 0201 data files	Section 1.1 updated Section 4.1 updated Note on high precision time added in section 4.3 Justification for update to version 0103 and 0201 provided in section 6. New documents for reference [AD-8], [AD-9], [AD-10], [RD-8] and [RD-9]	1E	2023-03-24

Table of Contents

1 Introduction 5
1.1 Purpose 5
1.2 Scope 5
2 Applicable and Reference Documentation 5
2.1 Applicable Documents 5
2.2 Reference Documents 6
2.3 Abbreviations 7
3 General description 8
3.1 Product overview 8
3.2 Scientific description 9
4 Output product 12
4.1 Product file types 12
4.2 Filenames 12
4.3 Product file content 12
5 Algorithms 16
5.1 Large-scale trends removal 17
5.2 PSD 17
5.3 Whistler dispersion and time characterisation 18
5.4 Geographic localisation of the data 18
5.5 Whistler dispersion reference uncertainty 19
5.6 Whistler t_0 and t_0 uncertainty 20
5.7 Whistler Intensity 20
6 Known limitations 21
7 Examples 25

Table of Figures

Figure 3-1 Schematic representation of the physical processes involved in the generation of a whistler signal detected by Swarm. 9
Figure 3-2 Example of strong whistler event recorded in burst data. 10
Figure 3-3 – Input/output of the whistler identification and characterization processing. 11
Figure 5-1 Example of large-scale trend removal for a whistler event. 17
Figure 5-2 Example of whistler with the theoretical dispersion curve superposed in white. 18
Figure 5-3 Comparison between dispersion obtained from a whistler wave fit on the ASM burst time-series D_{TS} and the dispersion D_{PSD} obtained from the PSD 19
Figure 5-4 Example of whistler characterization for time and intensity parameters. 20
Figure 6-1 Example of ambiguous whistler activity 22
Figure 6-2 Example of whistlers occurring in rapid succession. 23
Figure 6-3 Example of outlier in ASM burst data occurring during a whistler event. 24
Figure 7-1 Whistler events detected on 22/09/2020: geographical location and whistler dispersion. 26

List of Tables

Table 3-1 – Overview of WHIxEVT_2_ product 8
Table 4-1 – ASM whistler product file description. 14

List of TBDs

None.

1 Introduction

1.1 Purpose

This document describes the procedures and algorithms used to produce L2 Whistler Detection in ASM Burst Mode (BM), 250 Hz scalar data. This document also provides a description of the L2 product files WHIxEVT_2_ and presents an example of detected whistler events. This document is to be updated to a new version only when changes are introduced in the procedures and algorithms to produce updated data files (in which case the corresponding updated or new data files would bear a new version number). This version is in line with the production of WHIxEVT_2_ data product version 0201 (with a note on previous versions provided in section 6). Note that in contrast, [AD-9] is to be updated every time a new set of version 0201 data is released. Furthermore, all available WHIxEVT_2_ products in all released versions are listed in [AD-10]. It is recommended to use the latest available version for any WHIxEVT_2_ product. Please make sure to always look for the latest version of [AD-9] and [AD-10].

1.2 Scope

This document is produced as part of the Work Package IE30 of the Swarm DISC-3 proposal (see [AD-2] section 2.1.10.2.4). It is a support document to the release of the corresponding WHIxEVT_2_ data product, which was referred as ASMxWHI_2_L2 in [AD-2].

2 Applicable and Reference Documentation

2.1 Applicable Documents

The following documents are applicable to the definitions within this document.

[AD-1] SW-OF-DTU-GS-006 SE, Swarm xDISC Proposal SE, rev. 1dD, 2018-04-21

[AD-2] SW-OF-DTU-GS-009, Swarm DISC-3 Proposal, rev 1, 2020-07-20

[AD-3] SW-DS-DTU-GS-0001, Swarm Level 2 Processing System, Product specification for L2 Products and Auxiliary Products, rev 2Y, 2019-09-05

[AD-4] SW-IF-EAD-GS-00017, Level 0 Data Products, rev. 14.0, 07/06/2018

[AD-5] SW-IF-EAD-SY-0004, SWARM Packet Utilization Standard (S-PUS), rev. 6.0, 20/10/2008

[AD-6] PE-TN-ESA-GS-0001, Earth Explorer Ground Segment, File Format Standard, rev. 1.4, 2003-06-13

[AD-7] SW-TN-ESA-GS-0074, Tailoring of the Earth Explorer File Format Standard for the Swarm Ground Segment, rev. 1.5, 2009-08

[AD-8] SW-ASMV-DD-IPGP-0018, List of available 0103 ASM Whistlers Product files, rev. 1, 2023-03-24

[AD-9] SW-ASMV-DD-IPGP-0019, List of available 0201 ASM Whistlers Product files, rev. 1, 2023-03-24

[AD-10] SW-ASMV-DD-IPGP-0020, List of all available ASM Whistlers Product files, rev. 1, 2023-03-24

2.2 Reference Documents

The following documents contain supporting and background information to be taken into account during the activities specified within this document.

[RD-1] SW-CEA-MAG-PRO-509, ASM User Manual, rev. 1.7, 26/11/2013

[RD-2] SW-ID-GMV-GS-0006, CDF Data Format Swarm L1B, rev. 3.5, 2012-02-17

[RD-3] In-flight performance of the Absolute Scalar Magnetometer vector mode on board the Swarm satellites, Léger et al. (2015), Earth Planets and Space, 67:57

[RD-4] Swarm Absolute Scalar Magnetometers first in-orbit results, Fratter et al. (2016), Acta Astronautica 121:76-87

[RD-5] SW-ASMV-DD-IPGP-0008, ASM Burst Product Definition for Operational Release of Data, rev. 2D, 2023-02-28

[RD-6] SW-ASMV-DD-IPGP-0009, List of available 0104 ASM Burst Product files, rev. 1A, 2020-01-27

[RD-7] SW-ASMV-DD-IPGP-0010, List of available 0105 ASM Burst Product files, rev. 1, 2020-03-26

[RD-8] SW-ASMV-DD-IPGP-0013, List of available 0201 ASM Burst Product files, rev. 1E, 2022-09-23

[RD-9] SW-ASMV-DD-IPGP-0016, List of available 0302 ASM Burst Product files, rev. 1, 2023-02-28

[RD-10] SW-ASMV-DD-IPGP-0017, List of all available ASM Burst Product files, rev. 1, 2023-02-28

[RD-11] SW-ILGEW-IPGP-D15 Investigating Lightning-Generated Whistlers to improve ionospheric models, Final Report, 2021-11-16

[RD-12] Eckersley, T., Musical atmospherics, *Nature*, *135* (3403), 104–105, doi: 10.1038/135104a0, 1935.

[RD-13] SW-CEA-AUT-ALGO1B-001, ASM Level 1B algorithms, rev. 6.0, 2013-10-18

[RD-14] SW-TN-DSC-SY-0005, Level 1b Processor Characterization and Calibration Data Base, rev. 4.11, 2012-03-01

[RD-15] SWARM-GSEG-EOPG-05-001, Swarm L0 Product Format, rev. 1.7, 13/01/2010

[RD-16] SW-RS-DSC-SY-0002, Level 1b Processor Algorithms, rev. 6.10b, 2015-12-07

[RD-17] SW-RS-DSC-SY-0007, Level 1b Product Definition, rev. 5.18, 2016-04-20

[RD-18] Space Physics Guidelines for CDF, ISTP/IACG Variable Attributes, https://spdf.gsfc.nasa.gov/istp_guide/vattributes.html (accessed on 2023-03-17)

2.3 Abbreviations

Acronym	*Description*
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ASM	Absolute Scalar Magnetometer
BM	Burst Mode
CDF	Common Data Format
ELF	Extremely Low Frequency
LEO	Low Earth Orbit
PPS	Pulse Per Second
UTC	Universal Time Coordinated
VFM	Vector Field Magnetometer

3 General description

3.1 Product overview

The WHIxEVT_2_ characteristics are summarized in the table below.

Product Identifier	WHIxEVT_2_
Definition	The WHIxEVT_2_ product contains information about the Extremely Low Frequency (ELF) portion of whistlers detected during ASM Burst Mode sessions
Input Data	ASMxBUR_1B
Input Time Span	One day (daily files)
Spatial representation	N/A
Time representation	Time (start of whistler event in UTC)
Units	• Whistler dispersion: $\sqrt{s}$ • Estimated initial time of whistler propagation inside the ionosphere: s • Whistler intensity: pT^2/Hz
Resolution	• Whistler dispersion: 0.1 $\sqrt{s}$ • Estimated initial time of whistler propagation inside the ionosphere: ms • Whistler intensity: 0.1 pT^2/Hz
Uncertainty	• Whistler dispersion: provided with each daily file: Whistler_Dispersion_reference_ uncertainty variable. • Estimated initial time of whistler propagation inside the ionosphere: provided with each event: Whistler_t0_uncertainty variable; • Whistler intensity: not provided
Quality Indicator	N/A
Data Volume	5 MB per daily file
Data Format	Both CDF and Matlab
Output Data	See Table 4-1 of present document
Output Time Span	One day
Update Rate	Bi-monthly
Latency	Two months
Notes	Uncertainty for intensity not yet assessed

Table 3-1 – Overview of WHIxEVT_2_ product

3.2 Scientific description

The WHIxEVT_2_ product contains information about the ELF portion of whistlers detected during ASM BM sessions. BM sessions have been acquired in early 2014, a few days in 2018 and early 2019 and every month since summer 2019 [RD-6] [RD-7] [RD-8] [RD-9] [RD-10]. These whistlers are detected randomly when the satellites approach regions of active thunderstorms. Therefore, this product contains a variable number of events for each daily data file.
The duration of a whistler event is variable and depends on its dispersion. The dispersion is a parameter measuring the relation between the instantaneous frequency of the whistler signal and its time of arrival as will be detailed in section 5.3.

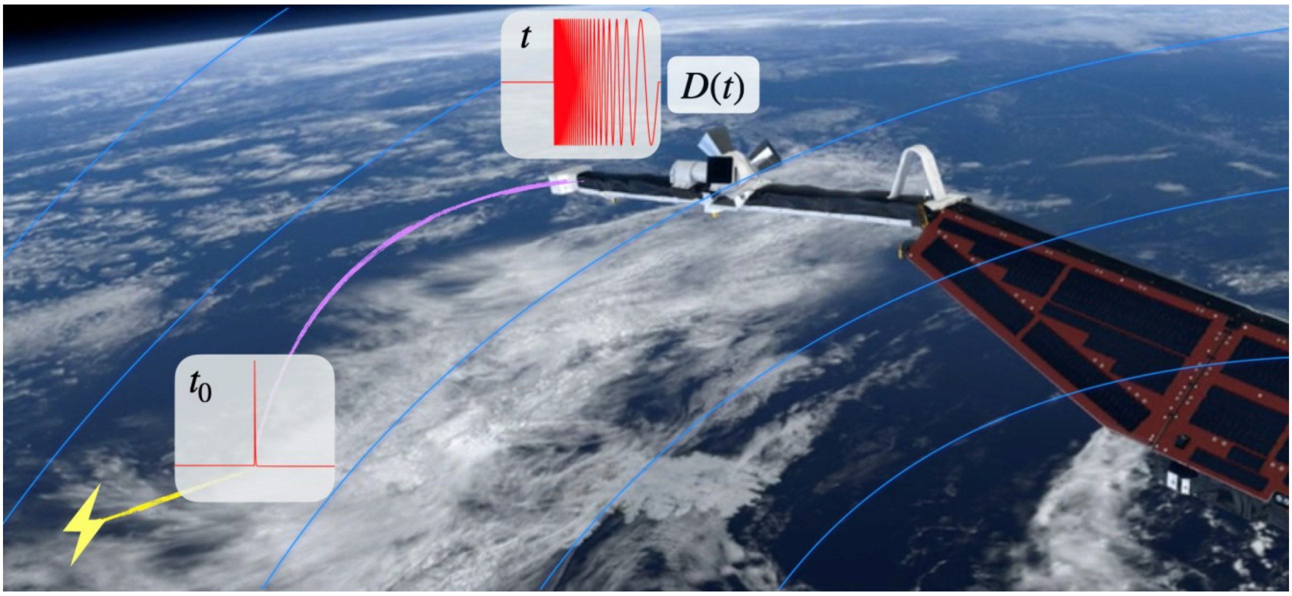

Figure 3-1 Schematic representation of the physical processes involved in the generation of a whistler signal detected by Swarm. t_0 corresponds to the time when the lightning electromagnetic signal enters the ionosphere and t the time when it reaches Swarm and its dispersion is $D(t)$. The blue lines represent the Earth's magnetic field lines.

A lightning strike in the troposphere generates a short impulsive signal that propagate at far distance in the ground-ionosphere wave-guide. At the interface between the neutral atmosphere and the ionosphere, part of this signal can penetrate into the ionosphere, where the refractive index depends on the charged particles distribution and on the direction of the magnetic field. Progressively the frequency content of the signal gets dispersed, before reaching Swarm satellites, see Figure 3-1. We define t_0 the time when the signals enter the ionosphere, as derived from the observed dispersion $D(t)$. Since the propagation in the neutral atmosphere is close to the speed of light in vacuum, t_0 can be used to identify the lightning strike that originated the observed whistler.
Note that the frequency band sampled by Swarm is the ELF part of the electromagnetic spectrum, below the cutoff frequency of the ground-ionosphere waveguide. This implies that most of the lightning signal in this frequency band is trapped inside the waveguide and only the strongest lightning strikes leak into the ionosphere and produce a signal that can be detected in the ELF at Swarm altitude.

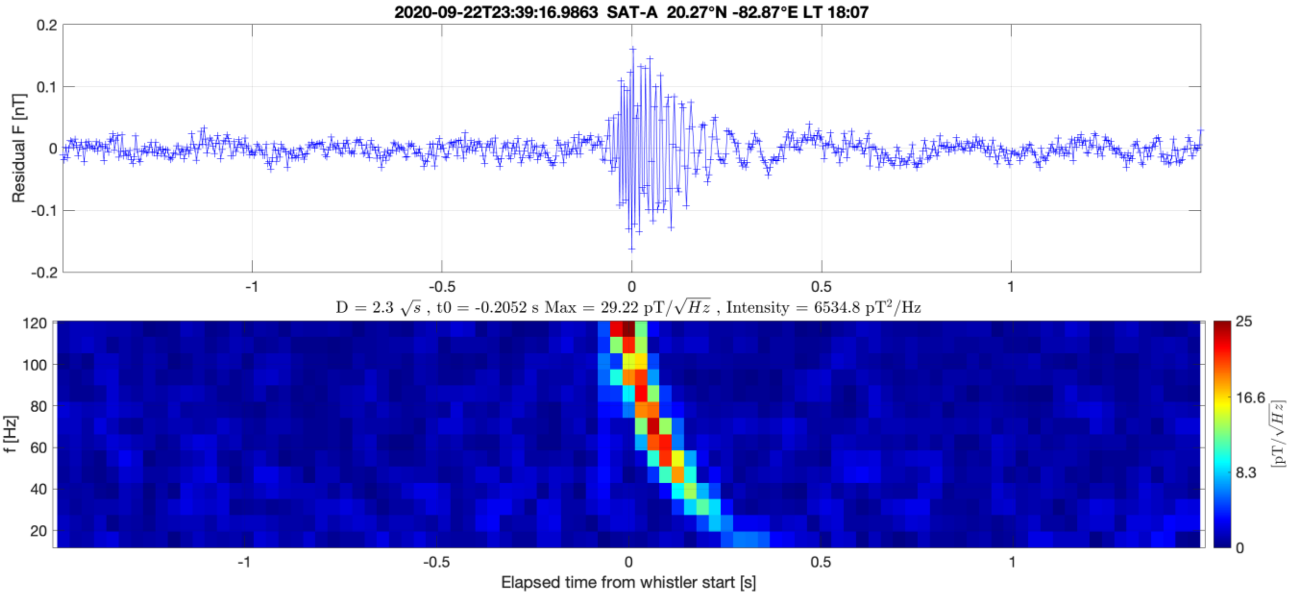

Figure 3-2 Example of strong whistler event recorded in burst data. Top panel: time series of magnetic field intensity, after removal of large-scale trends.
Bottom panel: Power spectral density plot showing the frequency evolution of the whistler signals and the parameters computed for its characterization.

The amplitude of the magnetic component of a whistler in the ELF at LEO altitude is on the order of a fraction of nT and its duration is generally shorter than 1 s, depending on whistler dispersion, see Figure 3-2.

Specific filtering algorithms have been designed to identify them in the ASM BM data, which contain the total field recorded at the position of the satellite. The measurement of the ASM intrinsically is the sum of all magnetic field effects at the measurement location. The dominant part is the Earth main field, of the order of several tens of thousands of nT to which all other effects add up.

All whistlers can be characterized by their dispersion curve, but they do not always present an identical spectral characteristic at ELF and their amplitude is sometimes comparable to the background or instrumental noise, making it difficult to establish an automatic procedure to identify and fully characterize all events. The current algorithm involves automatic and manual processing steps. Artificial intelligence algorithms are applied to speed up the process of finding the parts of the orbits where whistler activity is detected. Following this detection, manual screening is needed to validate or reject the candidate signals. Whistlers are characterized in terms of: dispersion, intensity and origin time, which corresponds to the time when the lightning signal entered the ionosphere. Estimated uncertainties on the dispersion and origin times are also provided.

The processing is done in various steps:

1. An initial time-window over 1 minute is selected from ASM L1b burst data in order to apply a polynomial high-pass filter and compute the power spectral density using FFT.
2. Some statistical parameters over a shorter time window of 3 s are computed from both the filtered time series and the corresponding PSD.

3. The statistical parameters and the flags contained in ASM burst data are given as input to an ANFIS network, to decide if this time-window contains a candidate whistler or not.
4. The time-window is shifted by 2.5 s and steps 2 and 3 are performed on the new time-window in order to obtain a list of times occurrences of possible whistler events. This is subsequently applied to the complete L1b burst data file.
5. The minutes of ASM burst data containing whistler events are reanalysed manually to characterize the detected whistlers and reject the false positive events:
    a. A dispersion curve is cross correlated with the PSD values to obtain the dispersion D and the origin time t_0.
    b. The intensity of the whistler event is obtained by summing up the squares of the PSD intensities at the crossing with the dispersion curve.
6. The whistler parameters are saved both in matlab and CDF formats and the corresponding header file is generated.

The algorithms for whistler's identification use as input only the Level 1b ASM BM data files ASMxBUR_1B [RD-5], which already contain all the information needed for characterization and georeferencing of the whistler data. The corresponding data flow to generate WHIxEVT_2_ data files is summarized in Figure 3-3

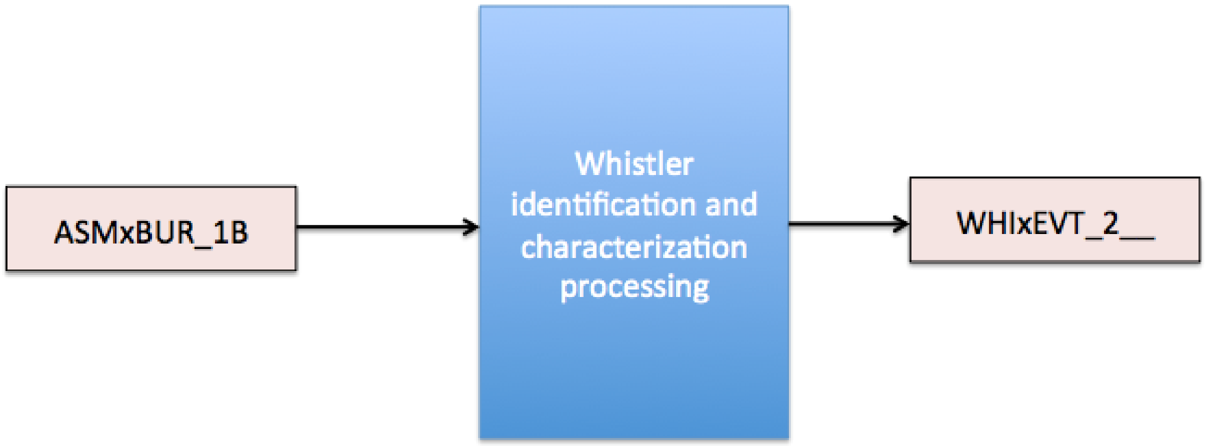

Figure 3-3 – Input/output of the whistler identification and characterization processing.

4 Output product

4.1 Product file types

The output product files of whistler's parameters consist of three daily files per satellite:

- A CDF product file
- A matlab product file containing the same data as the CDF file
- A xml header file

The header file contains some information about the data generation, such as the input files and the software used to generate the data. Please make sure to the get to correct ASMxBUR_1B input file from the <List_of_Input_File_Names> field when using the WHIxEVT_2_ product along with its input as multiple versions of ASMxBUR_1B are available, each with its specific processing ultimately impacting the WHIxEVT_2_ product content.

4.2 Filenames

Since the output product is not part of the original Swarm L2 products, its filename is not defined in [AD-3]. Thus, we define here our own format compatible with the official one (see section 4.4. of [AD-3]):

`MM_CCCC_TTTTTTTTTT_yyyymmddThhmmss_YYYYMMDDTHHMMSS_vvvv.EXT`

Where:

- `MM` (Mission Id) = "SW"
- `CCCC` (File Class) = same as the L1b input file: 'OPER'
- `TTTTTTTTTT` (Product name) = `FFFFDDDDDD` with `FFFF` = 'WHIx' and `DDDDDD` = 'EVT_2_'
- `yyyymmddThhmmss` = validity start time
- `YYYYMMDDTHHMMSS` = validity end time
- `vvvv` = version number

We introduce new File Category and Semantic Descriptor for the Product name (see section 5.1 of [AD-3]).
The File Category 'WHIx' stands for ASM BURST MODE WHISTLER data for satellite "x" (x = A, B or C).
The Semantic Descriptor 'EVT_2_' indicates that the file contains whistler Events and this is a final Level 2 product.
The validity start and end times are those of the corresponding L1b ASM burst file used to generate this product. This time validity is also an indication that no events have been detected between the time of the beginning of the L1b file and the time of the first event in the L2 whistler file and between the time of the last event and the end of the L1b file.
The version number indicates the revision number of the Whistler processing chain consisting of a two-digit baseline followed by a two-digit file-counter.

4.3 Product file content

The output data are stored in both CDF and MATLAB formats. The table below describes the different fields of data contained in these files.

For each variable listed below, the scope is given in the Scope column and is equal to G for Global and V for Variable. The data files contain also some additional Global parameters, which are also included in the header file. They contain metadata about data production and temporal coverage of the data file. They are not detailed in Table 4-1, they follow the definitions provided in [AD-3].

Field Name	Type	Dim	Units	Scope	Description
TITLE	CDF_CHAR	-	-	G	Filename of the product (without extension)
CREATOR	CDF_CHAR	-	-	G	Creator of the product (set to «WHI »)
Outlier_threshold	CDF_UINT1	-	nT/s	G	Threshold for outliers detection (currently set to 120 nT/s)
Outlier_threshold_description	CDF_CHAR	-	-	G	Description of the Outlier_threshold field
Whistler_Dispersion_reference_uncertainty	CDF_DOUBLE	-	$\sqrt{s}$	G	Reference uncertainty on the estimated whistler's dispersions
Whistler_Dispersion_reference_uncertainty_description	CDF_CHAR	-	-	G	Description of the Whistler_Dispersion_reference_uncertainty field
Timestamp_Whistler	CDF_EPOCH	N	Epoch	V	Time of start of whistler signal (UTC)
Latitude	CDF_DOUBLE	N	Degree	V	Position in ITRF - Geocentric latitude at Timestamp_Whistler
Longitude	CDF_DOUBLE	N	Degree	V	Position in ITRF - Geocentric longitude at Timestamp_Whistler
Radius	CDF_DOUBLE	N	m	V	Position in ITRF - Geocentric Radius at Timestamp_Whistler
LT	CDF_DOUBLE	N	h	V	Local Time at satellite position
Whistler_Dispersion	CDF_DOUBLE	N	$\sqrt{s}$	V	Dispersion of whistler signal
Whistler_t0	CDF_EPOCH	N	Epoch	V	Estimated initial time of whistler propagation inside the ionosphere
Whistler_t0_uncertainty	CDF_DOUBLE	N	s	V	Estimated uncertainty on Whistler_t0
Intensity	CDF_DOUBLE	N	pT^2/Hz	V	Intensity of whistler signal in the frequency range between 16 and 125 Hz

Field Name	Type	Dim	Units	Scope	Description
Timestamp	CDF_EPOCH	N×m	Epoch	V	Time of the time series in the whistler time-window
TimeFrac	CDF_UINT4	N×m	ns	V	Fractional time of observation: ns within second in the whistler time-window[1]
F_analysed	CDF_DOUBLE	N×m	nT	V	High-pass filtered magnetic field intensity in the whistler time-window
Flags	CDF_UINT1	N×m	-	V	Binary combination of the values taken by Flag_outlier, Flag_magnetic_condition, Flag_heater and Flag_step (See explanation below)
Timestamp_PSD	CDF_EPOCH	N×n	Epoch	V	Time of the PSD over the whistler time-window
Frequencies_PSD	CDF_DOUBLE	p	Hz	V	Frequencies of PSD[2]
PSD	CDF_DOUBLE	N×n×p	$nT/\sqrt{Hz}$	V	Power Spectral Density values at Timestamp_PSD over Frequencies_PSD

Table 4-1 – ASM whistler product file description. In the Dim column N indicates the number of whistler events contained in the data file, m the number of points of the corresponding time-series, n the number of points of the corresponding PSD, p the number of frequencies of the PSD.

The quantities provided include parameters for whistler event characterization and the corresponding time series and PSD from which these parameters have been obtained. Each parameter is stored in an array, which dimensions depend on the kind of parameter: single values for whistlers' events are stored in unidimensional arrays of size N. All the time series associated with these events are contained in arrays of dimensions N×m or N×n. Each row i of these arrays corresponds to the time series associated to the whistler event i. The corresponding magnetic field values F_analysed are stored in a 2D array, and each row i is associated to the whistler event i. Similarly, a three-dimensional array PSD of size N×n×p contains all the 2D PSDs corresponding to each event i. To allow identification of the proper support data for the various variables, specific attributes are included: DEPEND_0 for unidimensional variables, DEPEND_0 and DEPEND_1 for 2D arrays, DEPEND_0, DEPEND_1 and DEPEND_2 for 3D arrays, following the guidelines provided in [RD-18]. Support variables are those that define the reference axis for other variables.

[1] The length of the selected whistler time-window is of 3 s, centered around the beginning of the whistler signal.

[2] Note that Frequencies_PSD is the support frequency axis for the variable PSD, it should not be interpreted as a time series, even for the case N=p.

Note that these support variables do not include a DEPEND_i attribute (with i=0,1 or 2) for the dimension they support to other variables.

- **Timestamp_Whistler**, **Latitude**, **Longitude**, **Radius** provide the localization in time and space of the beginning of the detection of the whistler: the position of the Swarm satellite when the whistler signal appeared. These events occur randomly, therefore there is no regular spacing for these values.

LT indicates the local time at Swarm position at the beginning of the detection of the whistler. It is calculated from:

$$\mathrm{LT} = mod\left(\mathrm{UT} + \frac{\mathrm{Longitude}}{15}, 24\right) \quad (4\text{-}1)$$

UT is the Universal Time (hours) obtained from Timestamp_Whistler and the function mod(x,n) in the modulus of x/n.

- **Whistler_Dispersion** is the dispersion of the whistler, measuring the relation between time and frequency. It is measured in $\sqrt{\mathrm{s}}$. It is obtained following the algorithm described in section 5.3.
- **Whistler_Dispersion_reference_uncertainty** is the statistical uncertainty of the Whistler_Dispersion obtained as described in section 5.5. A unique value is provided, valid for all whistler events.
- **Whistler_t0** is the estimated time when the whistler signal entered the ionosphere. It is computed from the dispersion curve as described in section 5.6.
- **Whistler_t0_uncertainty** is the temporal uncertainty associated with the Whistler_t0 value, as described in section 5.6.
- **Intensity** is a measure of the intensity of the whistler in the ELF between 20 and 125 Hz obtained from the PSD as described in section 5.7.
- **Timestamp** is the time over 3 s of ASM burst data centred at Timestamp_Whistler, reproduced from the corresponding L1b data.
- **TimeFrac** is the additional time fraction needed to obtain a time representation for ASM data with a precision higher than the ms.
- **F_analysed** is the value of the magnetic field intensity over the whistler time window, after large-scale trends have been removed, as described in section 5.1.
- **Flags** are reproduced from the corresponding ASMxBUR_1B file. As defined in [RD-5] the flags are compactly stored as an 8 bits unsigned integer and can nominally take values from 0 to 15. This number can be written in a binary form such as Flags = ‘0000abcd’ where:
  - ‘a’ stands for Flag_outlier value,
  - ‘b’ stands for Flag_magnetic_condition value,
  - ‘c’ stands for Flag_heater value,
  - ‘d’ stands for Flag_step value.

In the same way, the Timestamp field for the time series is reproduced from corresponding ASMxBUR_1B file. This field is in the CDF epochs format which is internally represented as milliseconds since 1 January 0000 at 00:00. Since the date of the time series is available at a higher precision, using the TimeFrac field, it is possible to recover a high precision time expressed in elapsed UTC second from January 0000 at 00:00 using the same formula indicated in [RD-5]:

$$\mathrm{t} = floor(\mathrm{Timestamp} * 10^{-3}) + \mathrm{TimeFrac} * 10^{-9} \quad (4\text{-}2)$$

where Timestamp is expressed in elapsed UTC ms from January 0000 at 00:00 and $floor(x)$ returns the greatest integer less than (or equal to) x.

IMPORTANT: please note that the high precision time, described as is, requires more than 16 digits of precision which is the default precision of most numeric computing environment.
A secure way for recovering high precision time from Timestamp and TimeFrac without precision loss may be, using MATLAB and (4-2):

```
>> t = datetime(floor(Timestamp*1e-3),'ConvertFrom','epochtime','Epoch','0000-01-01') + seconds(double(TimeFrac)*1e-9);
```

- **Timestamp_PSD** is the time stamp associated to the PSD computed over the whistler time-window, as described in section 5.2.
- **Frequencies_PSD** is the frequency scale associated to the PSD computed over the whistler time-window, as described in section 5.2.
- **PSD** is the PSD computed over the whistler time-window Timestamp_PSD and the frequencies Frequencies_PSD as described in section 5.2.

5 Algorithms

The processing chain to detect whistler in the ASM BM data has been developed in the framework of the ILGEW project [RD-11]. The entire daily time series of ASM scalar magnetic field values is screened for detecting candidate events using an Artificial Neural Network (ANN) algorithm of type ANFIS. The ANN selects a list of time-windows where the whistler signals are found. The details of this processing of selection are described in [RD-11]. This document presents the algorithms to obtain the whistlers' parameters contained in the WHIxEVT_2_ data files. The complete processing chain involves both automatic calculations and manual operations, particularly to validate or rejects events detected automatically and to allow additional inclusion of events missed by the automatic algorithm. Over time the performances of the neural network have improved, reducing the screening time. However, the characterization of whistler still requires manual operations.

5.1 Large-scale trends removal

In order to remove slow trends on the scalar magnetic field time series, a polynomial fit is realized using one minute of ASM L1b burst mode data containing a detected event. The current processing chain uses a time window of 70 seconds of data, including 5 seconds of data before and after the UT minute, when available. The overlap between successive time-windows is required in order to guarantee proper characterization of whistler events occurring near their limits. The manual inspection of these 1-minutes time windows is necessary to appreciate the noise level in the region crossed by Swarm satellites and support the validation or rejection of the ANN detected events.
A polynomial of degree 4 is fitted to the selected burst data, the residuals are successively used for the frequency analysis. For each whistler event, 3 seconds of residual magnetic field values, centred at the Timestamp_Whistler are stored in the F_analysed array of the WHIxEVT_2_ files in units of nT.

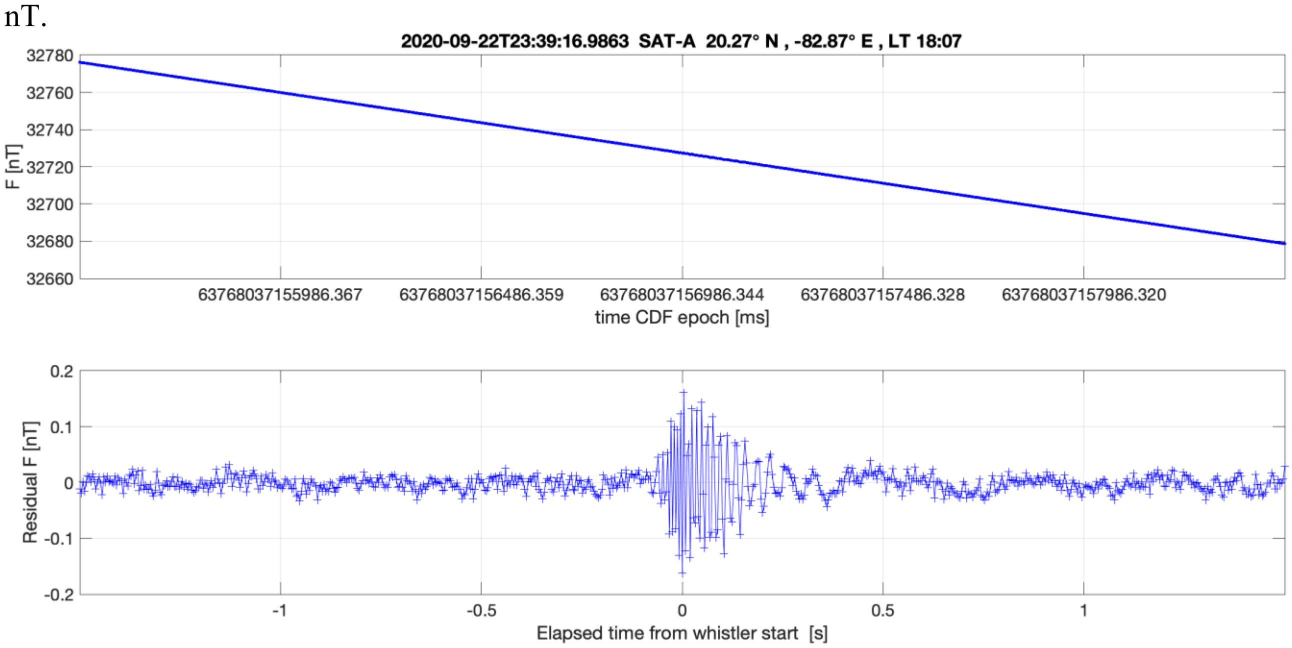

Figure 5-1 Example of large-scale trend removal for a whistler event. Top panel: ASM burst total intensity over a time window of 3 s. Note that the time scale following the CDF epoch scale is obtained combining Timestamp and TimeFrac variables. Bottom panel: the corresponding F_analysed; in this panel the time is shown relative to Timestamp_Whistler, indicated in the figure title in the form of a calendar date and UTC time.

5.2 PSD

The Power Spectral Density in the ELF band of sensitivity of the ASM scalar measurement during burst session is computed in order to identify the typical spectral signature of a whistler: a short transient signal presenting a decrease in frequency which can be characterized by the dispersion parameter.
The time series F_analysed is used for computing the evolution of the Power Spectral density with the following parameters:

- FFT order = 32
- Time window length: 0.13 s
- 75% overlap between two consecutive time windows

It is necessary to define a time stamp for each computed FFT: the time at the centre of the time window is selected and stored in the corresponding position of the variable Timestamp_PSD**.** A total of 14 frequencies between 16 and 125 Hz as defined in the variable Frequencies_PSD are stored in the array PSD. The two lowest frequencies computed by the FFT algorithm have been discarded since they are affected by instrumental noise. See [RD-3] for details.

5.3 Whistler dispersion and time characterisation

Under the hypothesis of propagation through a dispersive medium, corresponding to the Earth ionosphere under the influence of the Earth magnetic field, it is possible to obtain a simple relation between the instantaneous frequency $f(t)$ of the whistler and its time of arrival t [Eckersley, 1935], which can be expressed as:

$$D = \sqrt{f(t)}(t - t_0) \tag{5-1}$$

with t_0 the reference time of the beginning of the propagation of the whistler signal inside the ionosphere. This equation can be rewritten to obtain the whistler frequency as function of time:

$$f(t) = \frac{D^2}{(t - t_0)^2} \tag{5-2}$$

The dispersion of the whistlers contained in the ASM burst data has been obtained by correlating this theoretical curve with the values of the PSD as computed following the algorithm of 5.2. It is first estimated by comparing the curve obtained for a value of D with the PSD values. Several values of dispersion D are tested, varying in step of 0.1 $\sqrt{\text{s}}$. The resulting curve is also shifted in time, in order to obtain the parameters providing the highest correlation between the ASM signal and the theoretical dispersion curve, see Figure 5-2. In the current processing chain, a manual intervention is required. The values for D is obtained by this maximisation and the Timestamp_Whistler is taken as the time of the first tile of the PSD plot crossed by the selected dispersion curve. It corresponds to the arrival time of the 117 Hz component of the whistler at the satellite position, in the upper frequency band of the PSD calculation.

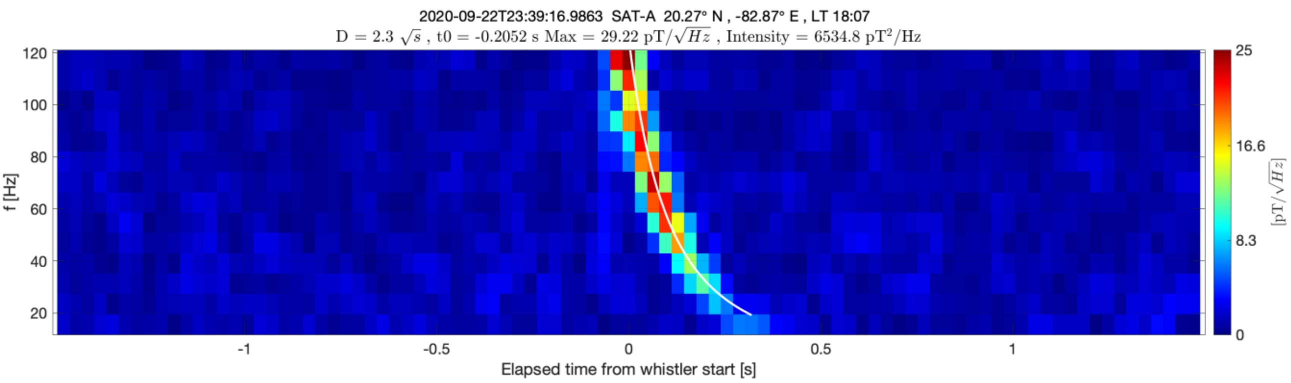

Figure 5-2 Example of whistler with the theoretical dispersion curve superposed in white. This curve is used to obtain the dispersion and timestamp values for the event.

5.4 Geographic localisation of the data

Since the whistler signal has a certain duration and a larger spectral band than the one to which the ASM instruments are sensitive, it has been decided to attribute the detection to the location where Swarm satellite was at the time Timestamp_Whistler, thus corresponding to the location at the arrival of the 117 Hz component of the whistler signal. Since the duration of a whistler is variable and of the order of a fraction of a second, the satellites move some km during the reception of a whistler, based on a typical satellite velocity of about 8 km/s. Note that no correction is implemented to account for it.

5.5 Whistler dispersion reference uncertainty

The dispersion has been obtained from the PSD plot, by comparison with a theoretical dispersion curve. There is an intrinsic uncertainty related to the size of the time-windows used for the FFT calculation and to the overlap between successive time-windows. In order to understand the bounding limits of the dispersion value obtained, a dedicated study was performed taking advantage of the availability of the whistler signals time series to obtain an independent estimation of the whistler dispersion. A synthetic waveform was computed and its parameters fitted by least square approach with the measured signals. This method can be applied to most of the whistlers, but there are about 6% of cases for which it is not converging, therefore it is not robust enough for routine whistler dispersion estimation. A quality flag was defined for the estimation of the dispersion from the time series to evaluate the goodness of the fit with values ranging from 0 to 2, from the most to the least reliable results. For the statistical comparison between these two methods, only cases for which this flag was equal 0 or 1 were retained. 1207 whistlers presenting nominal spectral coverage (covering all frequencies between 20 and 125 Hz) detected during burst mode campaigns between 2014 and the fall of 2020 were selected for this purpose.

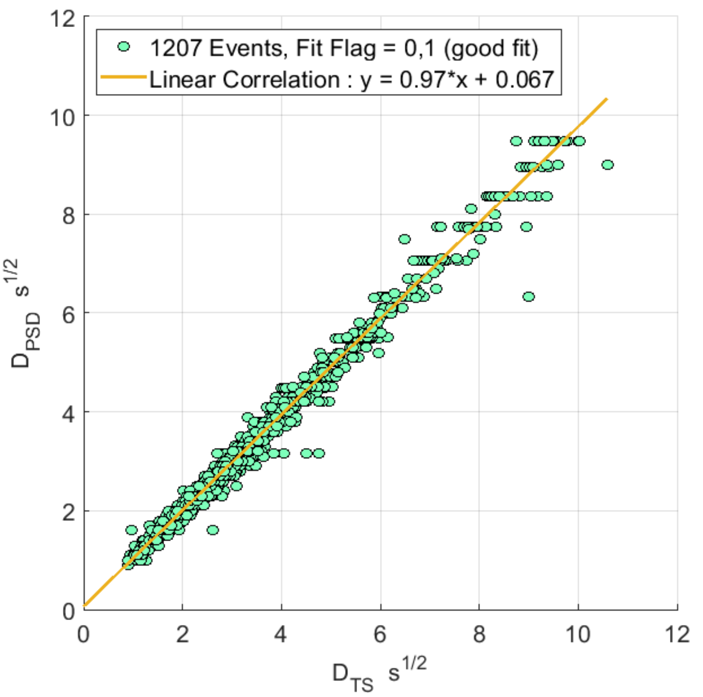

Figure 5-3 Comparison between dispersion obtained from a whistler wave fit on the ASM burst time-series D_{TS} and the dispersion D_{PSD} obtained from the PSD for 1207 selected events, for which the quality flag defined for the D_{TS} indicated a good fit.

By comparing the two independent estimates of whistler dispersion it was possible to obtain a very good linear relation between their values, see Figure 5-3. The differences between the two values of dispersion are generally below 0.4 $\sqrt{s}$ and this value has been chosen as the bounding limit for whistler dispersion, defined as the reference uncertainty. Note, however, that for the largest dispersions, the uncertainty can also be larger.

5.6 Whistler t_0 and t_0 uncertainty

A measure of the whistler's origin time t_0 and its uncertainty is obtained from the analysis of the points at the intersection between the theoretical dispersion curve (5-1) and the tiles composing the PSD plot. For the purpose of keeping the time error small, it was chosen to have a low resolution in frequency and a fine resolution in time, therefore for a specific frequency range, several successive tiles can be crossed by this curve. Their number increases with increasing dispersion, particularly for frequencies around 20 Hz, see Figure 5-4. Using the central time and central frequency of each tile, various dispersion curves compatible with the observed PSD can be obtained, along with their corresponding origin times, which are all different. The whistler's origin time t_0 has been defined as the average of these times.
Similarly, t0_uncertainty has been defined as the time range between the first and the last tile crossed by the dispersion curve.

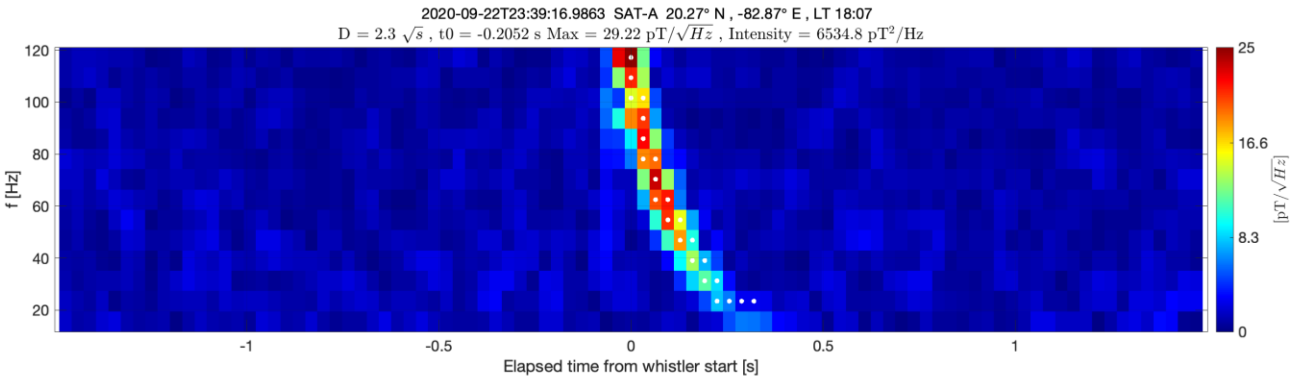

Figure 5-4 Example of whistler characterization for time and intensity parameters. The tiles on the PSD plot that are crossed by the dispersion curve are indicated by white dots. They are used to compute t0, its uncertainty and the intensity of the whistler in this frequency range.

5.7 Whistler Intensity

The measure of whistler intensity depends on the portion of the electromagnetic spectrum that is sensed. A whistler can cover a very large part of the electromagnetic spectrum, while the various instruments generally used for their measure cover a limited portion of it. Nevertheless, a value allowing to compare events recorded by the same instrument can be defined. It was chosen to use for this purpose the same theoretical curve obtained during the characterisation of whistler dispersion, following equation (5-1). This curve crosses a certain number of the tiles appearing on the PSD plot, see the white dots on Figure 5-4. The values in the frequency range between 20 and 125 Hz are selected and the corresponding whistler intensity is obtained from:

$$I = \sum_{j=1}^{N} p_j^2 \tag{5-3}$$

Where N is the total number of selected tiles, p_j is the PSD values for a tile at the time and frequency crossed by the dispersion curve. This quantity is expressed in pT^2/Hz. Note that if an outlier event occurs simultaneously with a whistler event, an unrealistic large value of intensity is recorded as shown later in Figure 6-3.

6 Known limitations

For WHIxEVT_2_ data in version 0103 (and all later versions), TimeFrac field data is directly copied from the ASMxBUR_1B file's data. Timestamp and TimeFrac field data in WHIxEVT_2_ version 0102 were previously computed in a way (now deprecated) that required a higher level of precision than MATLAB's provided default precision which led to precision loss and discrepancy between ASMxBUR_1B and WHIxEVT_2_ Timestamp and TimeFrac data where they should be equal.

For WHIxEVT_2_ data in version 0201, computations are based on ASMxBUR_1B version 0302 file's data which include new temporal calculations (see section 6 of [RD-5]) and take into account the new correction of the so-called dBSun perturbation slightly affecting the ASM scalar measurements (see section 5.4 of [RD-5]). Note that not all WHIxEVT_2_ files in version 0103 have been reprocessed in version 0201 as it has been investigated that the dBSun perturbation had little to no effect on the magnetic field intensity F_analysed data as slow trends are filtered out. Thus, to reduce the manual processing time, the production of WHIxEVT_2_ data in version 0201 data started when production of WHIxEVT_2_ data in version 0103 data stopped.

Whistler events identification

The algorithm to obtain whistler dispersion involves manual operations, primarily to validate the selections made by the ANN algorithm and to adjust the dispersion to obtain the whistler parameters. The ambient noise in the ELF frequency band is variable and there can be regions where its amplitude is similar to whistler amplitudes, making it difficult both for the ANN and for the operator to decide if a weak signal should be interpreted as a whistler. In many instances, whistler spectral signature is not complete increasing the uncertainty in deciding if the specific spectral signature is part of a whistler signal. There is also an intrinsic human bias in the appreciation if a fragmented signal with amplitudes similar to the background noise might be a whistler. The ANFIS algorithm is also affected by this limit: in many instances noisy time-windows are selected and provide false positive events. Other artefacts contained in BM data, like the heater perturbations, see [RD-3], [RD-4] or [RD-5], are flagged in L1b data and they don't produce false detections. The amplitude limit for whistler detection is depending on the background noise and on the ASM instrument sensitivity and it is about 4 $pT/\sqrt{Hz}$ on the PSD. It should also be noted that ASM burst data cover only a portion of the whistler spectrum and in this ELF portion the whistler amplitude is generally much smaller than at VLF, where a much larger number of events is expected. In some cases it is not possible to confidently characterise the dispersion of fragmented events. These events have not been included in the data files. Therefore, whistler contained in the WHI product cannot be considered as comprehensive of the total number of whistlers occurring at the satellite location.

Whistler characterisation ambiguities

Sometimes the task of characterising whistlers is challenging for various reasons. When several whistlers occur during a short interval of time, there is an intrinsic ambiguity in distinguishing each of them. For cases when it is not possible to obtain a reliable dispersion value, no whistler events have been included in the data files, see an example in Figure 6-1.

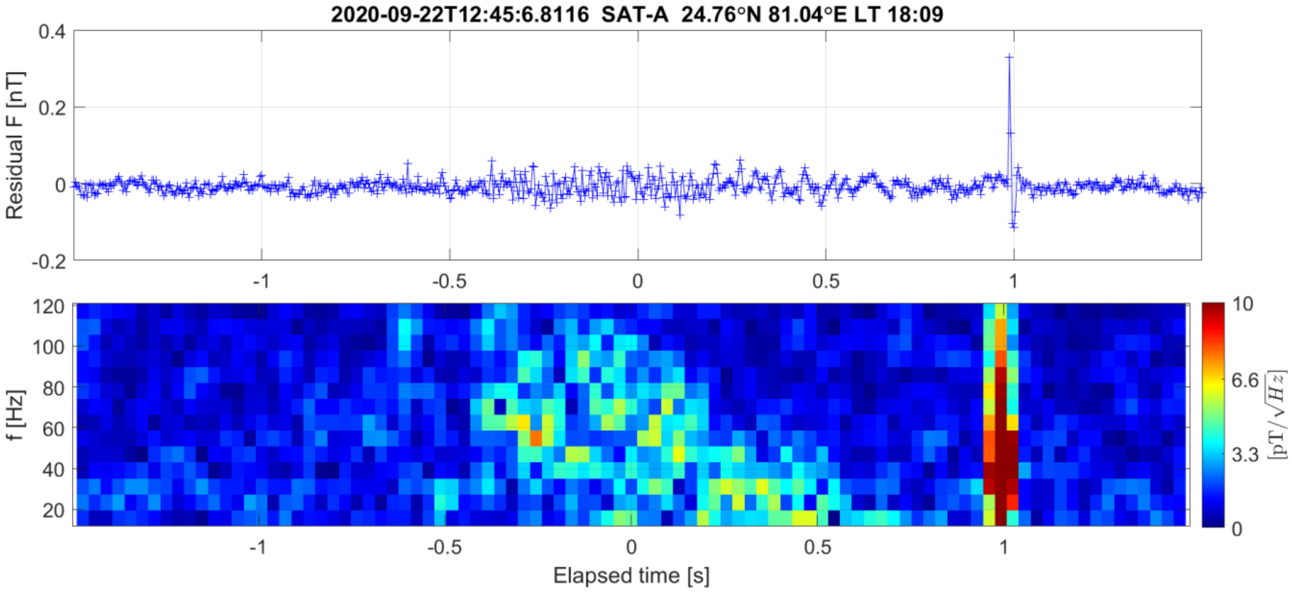

Figure 6-1 Example of ambiguous whistler activity for which is not possible to obtain a reliable value of dispersion.

Another challenging situation is the case of events with limited clear spectral coverage of the ELF band accessible during ASM burst sessions. This is particularly difficult for uncomplete whistlers that present large dispersion, exceeding 10 $\sqrt{s}$, see an example in Figure 6-2.

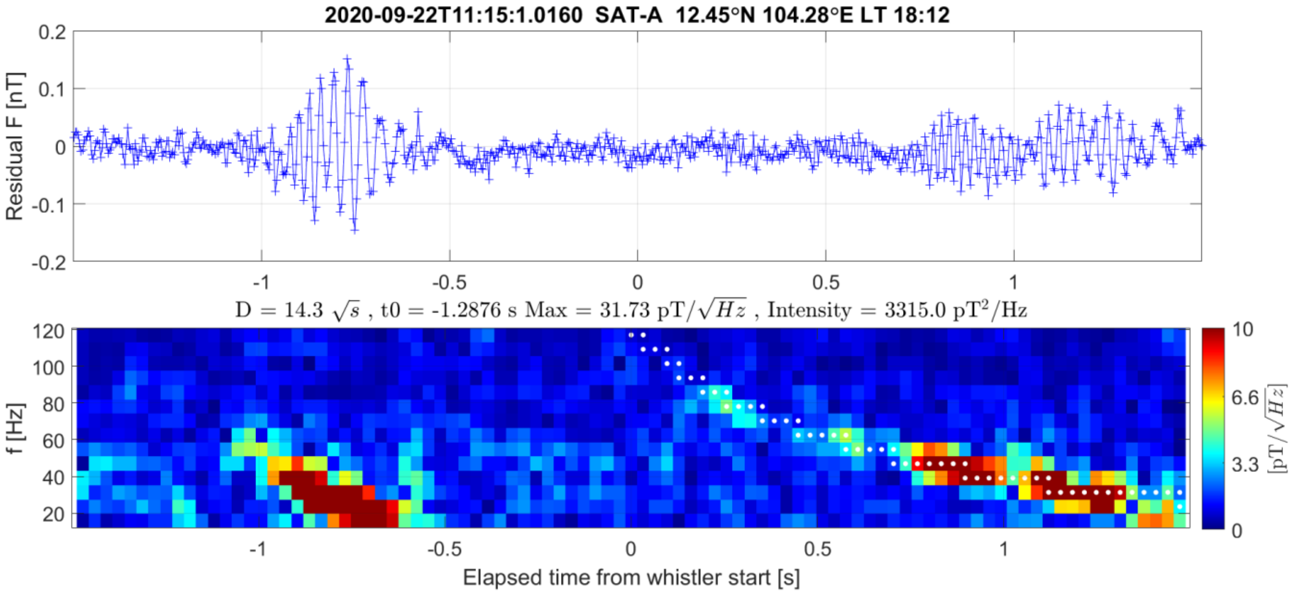

Figure 6-2 Example of whistlers occurring in rapid succession. The event identified with the white dots has a dispersion of 14.3 $\sqrt{s}$ and does not present a signature above 90 Hz. Its characterization is affected by ambiguity.

To mitigate this issue, most of the events presenting a large dispersion have been manually screened by multiple team members to agree on the characterisation. It is indeed expected that sometimes whistlers with large dispersion occur near the magnetic equator. Ambiguous events might be eliminated or separated into several whistlers during future data reprocessing activities, a new version of WHIxEVT_2_ data files will be produced accordingly.

Whistler intensity contamination

The measure of whistler intensity provided does not take into account incomplete whistler signal on the frequency band analysed. Adding intensities with low values does not affect much the total value.

However, there are cases when an outlier in the ASM burst data occur during a whistler event. In this case the resulting PSD value are extremely large and the intensity value is unrealistic. These cases can be identified by checking the presence of outliers simultaneous to whistler occurrence, as shown in Figure 6-3. Outliers are flagged in Flag_outlier, see [RD-5].

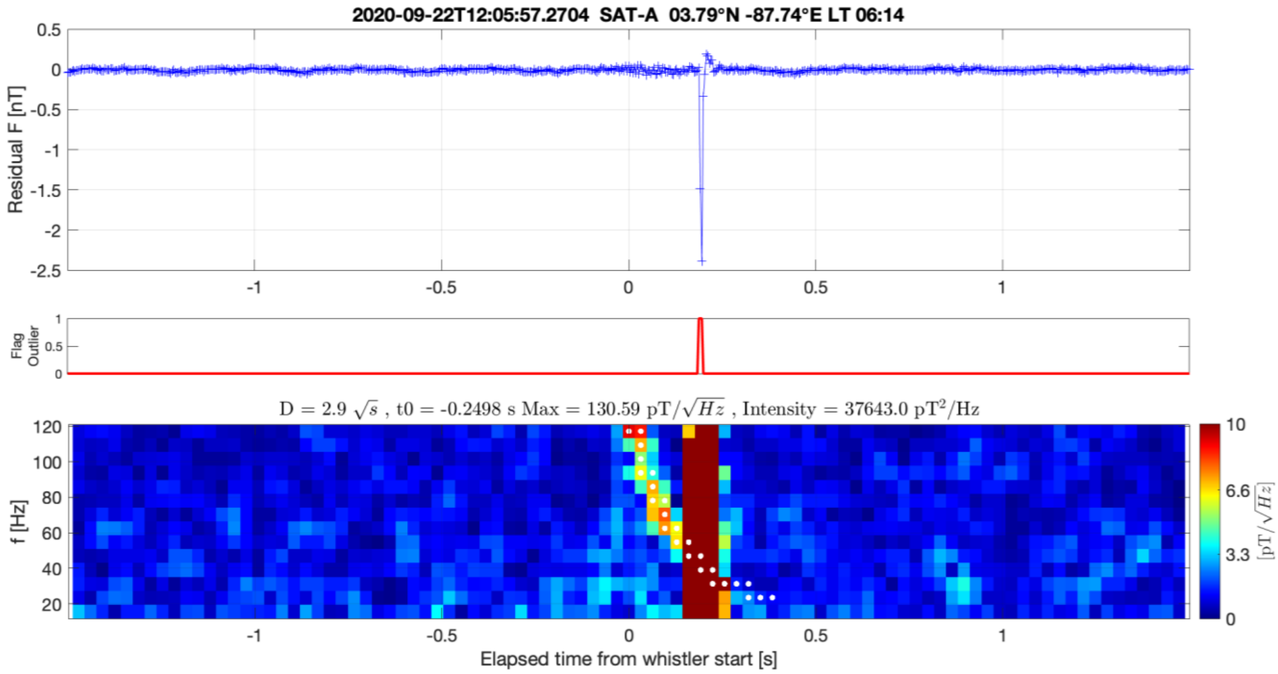

Figure 6-3 Example of outlier in ASM burst data occurring during a whistler event. The resulting Intensity value is unrealistically large. Top panel: F_analysed value. Middle panel: flag outlier. Bottom panel: PSD of the whistler event. The white dots indicate the PSD values that are used to compute the intensity.

7 Examples

The whistler events in the ELF occur randomly when Swarm satellites cross regions within few thousands of km from strong active storms. Since lightning activity depends strongly on the local time, an asymmetry in whistler occurrences is expected during the ascending and descending parts of the satellite orbits, see Figure 7-1 for an example. The maps show the locations of whistler detections and the colour scale indicates their dispersion. These events occur predominantly in regions of the world where active storms produce powerful lightning. Swarm can detect whistler in a radius of few thousands of km from the lightning strike location.

The day-to-day variability in whistler events is also large, therefore it was decided to conduct burst sessions lasting one week to obtain a good statistical significance of the detected whistler activity.

Usually, it is observed that the dispersion is higher at lower geomagnetic latitude, following the increase of ionisation in the ionosphere equatorial anomaly region. In the same way, dispersion is higher during the day-time than during night-time.

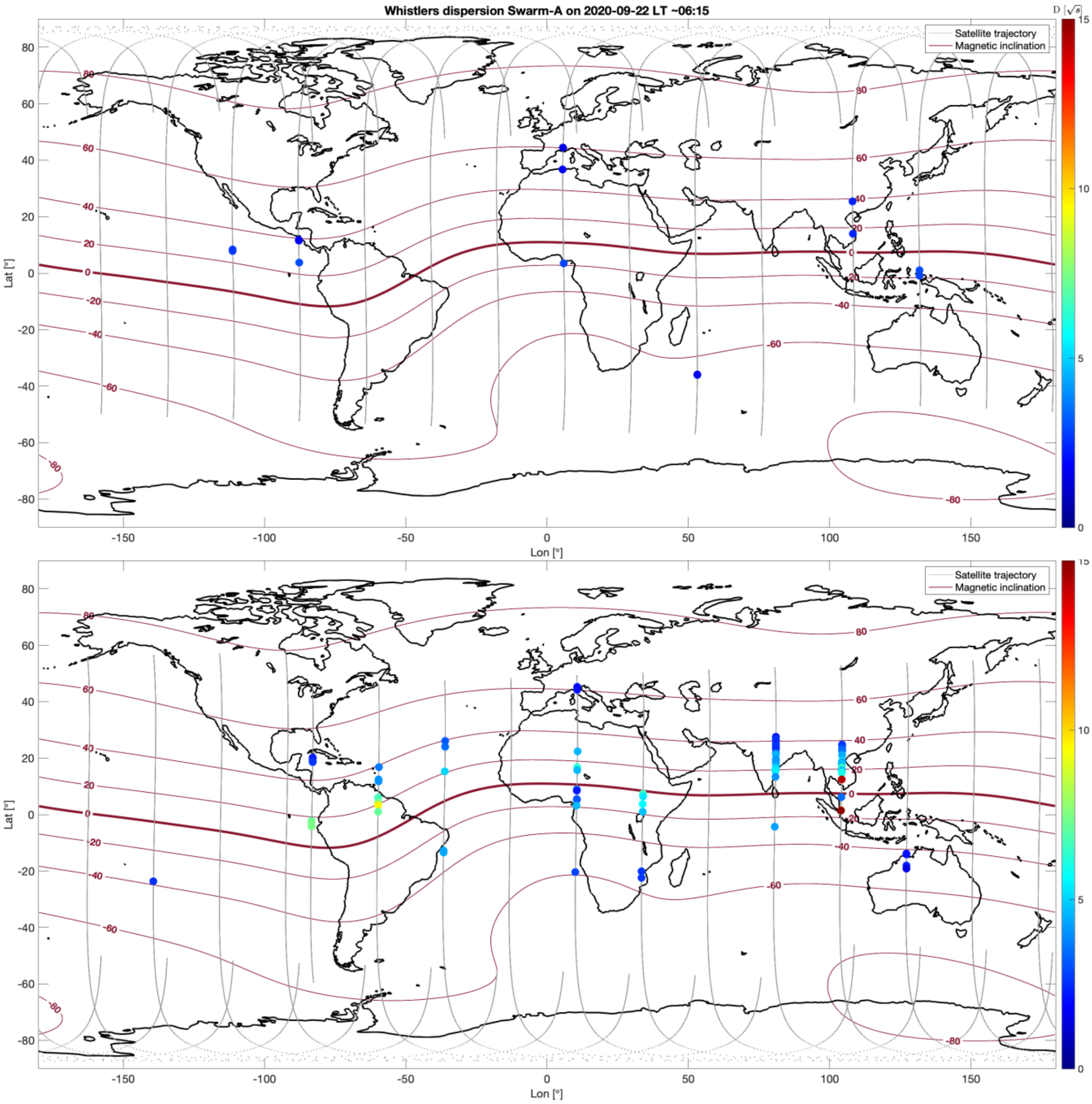

Figure 7-1 Whistler events detected on 22/09/2020: geographical location and whistler dispersion. Top panel: whistlers detected during local times between 6 and 18; bottom panel whistlers detected during local times between 18 and 6. The dark red isolines indicate magnetic inclination. The gray line indicates the satellite suborbital position in the corresponding local time interval. Note that on this example the satellite orbit is near the solar terminator.