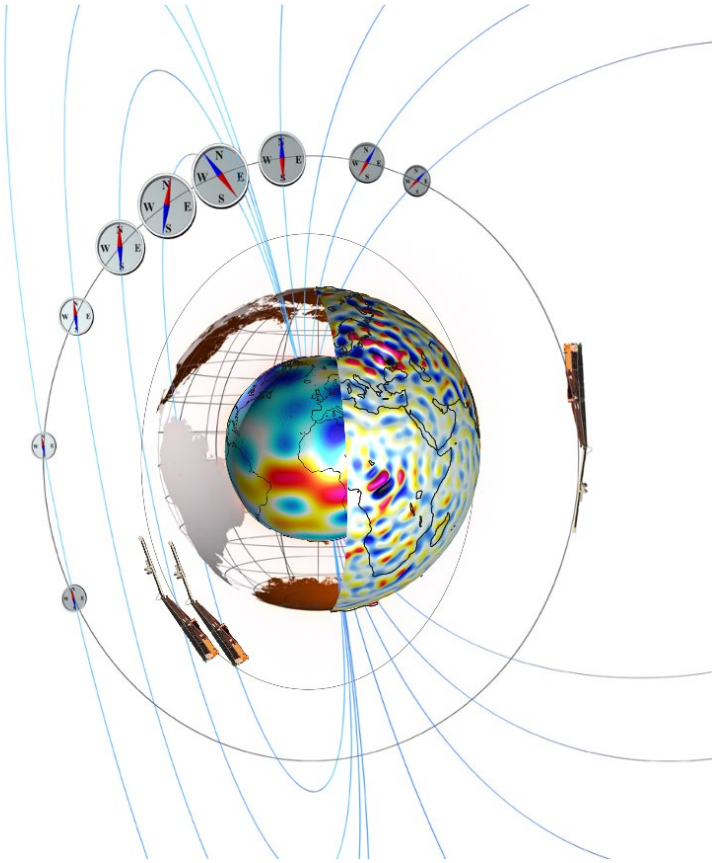
Statement of Work

Swarm DISC ITT 4.2

"Toolbox for Correlation and Maximum Variance Analysis of Swarm Time-series Data Products"



Doc. no: SW-SW-DTU-GS-128, Rev: 1

Record of Changes

Reason	Description	Rev	Date
Initial version	Approved by the TEB	1 dA	2 June 2021
Final version	Approved for publication by ESA	1	8 June 2021

Table of Contents

List of Figures

1 Introduction

This Invitation to Tender is issued by the Swarm DISC consortium on behalf of ESA within the reference frame of ESA contract 4000109587/13/I-NB, under the Swarm DISC Procurement Procedure described [RD-1].

1.1 Scope and Applicability

This document describes the activity to be executed and the deliverables required under the Swarm DISC ITT 4.2 – "Toolbox for Correlation and Maximum Variance Analysis of Swarm Time-series Data Products".

It will become part of the contract and shall serve as an applicable document throughout the execution of the work (with possible amendments recorded during the negotiation meeting).

The document is structured as follows:

- Chapter 2 quotes reference documents (including links and applicable standards).
- Chapter 3 introduces the background and main objectives of the work, and presents the constraints on the system to be produced.
- Chapter 4 defines the work to be performed in the contract to produce the required output.
- Chapter 5 contains the requirements on deliverables and on general project management aspects.
- Chapter 6 contains schedule and milestones.

2 Applicable and Reference Documentation

2.1 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-RS-DTU-GS-003 rev. 2, Swarm DISC Procurement Procedure

[RD-2] Bunescu, C., Vogt, J., Marghitu, O., and Blagau, A. (2019). Multiscale estimation of the field-aligned current density, *Annales Geophysicae,* 37(3), 347-373, https://doi.org/10.5194/angeo-37-347-2019

[RD-3] Laundal, K. M., Hatch, S. M., and Moretto, T. (2019). Magnetic effects of plasma pressure gradients in the upper F region. Geophysical Research Letters, 46, 2355– 2363. https://doi.org/10.1029/2019GL081980

[RD-4] Sonnerup, B.U.O. and Scheible, M. (1998). Minimum and Maximum Variance Analysis, Analysis Methods for Multi-Spacecraft Data, Götz Paschmann and Patrick W. Daly (Eds.), ISSI Scientific Report SR-001 (Electronic edition 1.1), http://ankaa.unibe.ch/forads/sr-001-08.pdf

[RD-5] Yang, J. -Y., Dunlop, M. W., Lühr, H., Xiong, C., Yang, Y.-Y., Cao, J. -B., et al. (2018). Statistical correlation analysis of field-aligned currents measured by Swarm. Journal of Geophysical Research: Space Physics, 123, 8170– 8184. https://doi.org/10.1029/2018JA025205

2.2 Links

- Swarm Mission page
  https://earth.esa.int/eogateway/missions/swarm
- VirES
  https://earth.esa.int/eogateway/tools/vires-for-swarm

- VRE
  https://earth.esa.int/eogateway/tools/swarm-vre
- viresclient
  https://viresclient.readthedocs.io
- Swarm data products
  https://earth.esa.int/eogateway/missions/swarm/data
- Swarm Data Handbook
  https://earth.esa.int/eogateway/missions/swarm/product-data-handbook
- Swarm DISC GitHub Organisation
  https://github.com/Swarm-DISC/

2.3 Terminology

In this document the term '*shall*' indicates requirements which the proposed effort must meet, while '*should*' indicates a desirable feature.

2.4 Abbreviations

A frequently updated Acronyms and Abbreviations list for Swarm and related projects can be found here in the Swarm Data Handbook.

3 Background and Objectives

3.1 Background

With an increasing portfolio of products available in the Swarm L1b and L2 catalogues, the focus is changing from delivering fixed file types in a continuously updated manner to tools that support easy exploitation of the scientific data. Such tools (typically as software) will allow scientists to interact more readily with datasets from Swarm and other missions and provide the capability to recreate and reproduce published scientific results as well as help facilitate new discoveries by extension of existing results or exploration of new ones.

For this ITT, we wish to develop a software toolbox that allows cross-correlations and maximum (or minimum) variation analyses with single or multiple satellite datasets to be computed. Ideally, the software will allow users to select L1b datasets, different physical measurements and correlation parameters in order to extract geophysical features. As an example, the toolbox will be able to compute the cross-correlation of the magnetic field between Swarm A and C allowing the orientation of a field aligned current sheet (with respect to the plane of the orbit) to be determined [RD-5].

As a general software toolbox, it shall be able to accept both scalar (e.g. ion temperature, field-aligned currents) and vector (magnetic field) parameters, and allow the user to select the time window, correlation length, and data cadence and perform the analysis for rolling time windows. The interface shall be orientated towards scientific data exploration. It should be written in Python and be aligned with existing tools like viresclient (https://viresclient.readthedocs.io) in order to improve ease of use. There shall be ample documentation and use-case examples.

3.1.1 Use-case example

The aim of the toolbox is to be flexible and help new science discoveries. As an example of how the tool may be used, figure 1 presents results of a correlation analysis performed on data from Swarm A and Swarm C [RD-5]. The red lines represent the alignments of field-aligned currents and show that, contrary to common assumptions, the currents are not always aligned in the magnetic east-west direction. While this analysis was performed on field-aligned current estimates using a specific sliding window, the toolbox shall allow different windows and parameters, including both scalar and vector measurements. This would allow, for example, analyses of differences in the alignment of current structures and structures in plasma density [RD-3]. Other relevant techniques, like minimum / maximum variance analysis based on single-satellite measurements [RD-4], or autocovariance calculations, shall also be included in the toolbox.

In addition to the correlation analysis example above, the toolbox shall also extend to other parameters. The most obvious additions would be electron density and temperature, ion temperatures and cross-track ion velocities. Furthermore, in the example of the case of FAC orientation analysis, the toolbox should also support maximum variance analysis. The toolbox shall also be capable of generating products based on autocorrelation from each satellite separately for different time windows. Such analysis from Swarm A, C and B would, for example, give two-point information about altitude variations in ionospheric structuring.

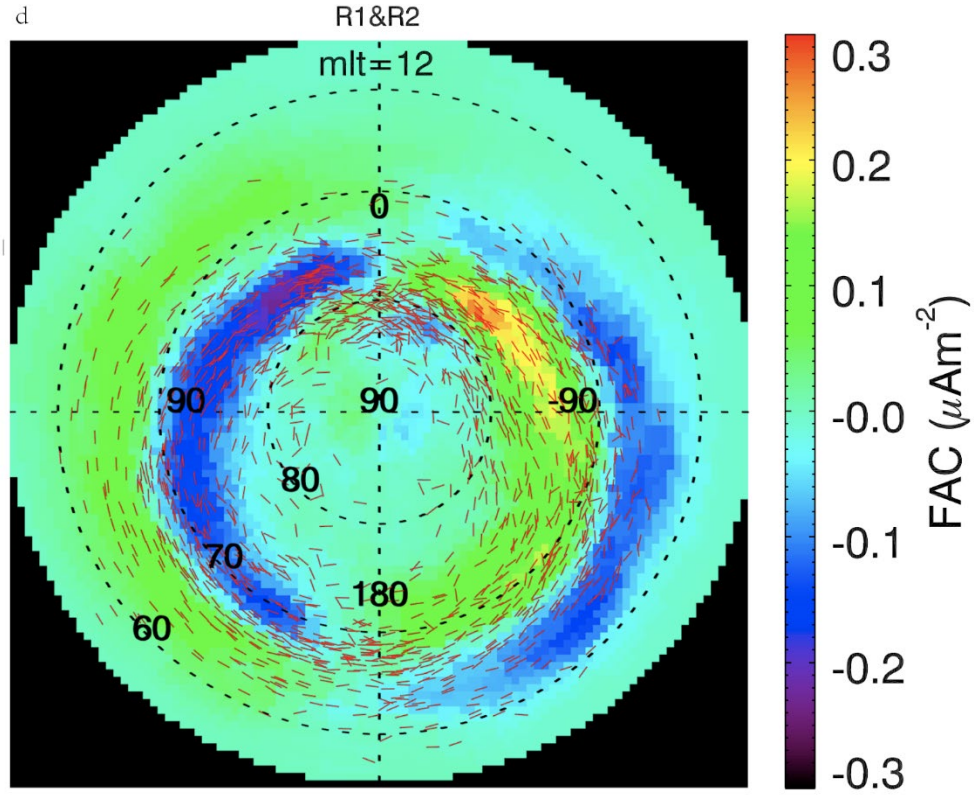


Figure 1 - Average field-aligned current strength (colours) and field-aligned current sheet orientations (red lines), derived from a correlation analysis of Swarm data [RD-5].

3.1.2 Swarm data and software environment

The Swarm mission, as well as existing data products, services, and projects are described at https://earth.esa.int/eogateway/missions/swarm and included links.

The Swarm Data Handbook describes the contents and scientific background of the data products. Many of these products can be accessed and visualised through the interactive VirES for Swarm web tool (https://vires.services), as well as through the Virtual Research Environment (https://earth.esa.int/eogateway/tools/swarm-vre) and associated Python package, viresclient (https://viresclient.readthedocs.io). The toolbox that this ITT is seeking should take the existing data infrastructure into account. In particular, tools that can integrate well with the Virtual Research Environment will add value to the service and increase accessibility of the toolbox itself to end users.

The software toolbox will aim to empower scientists to more easily answer research questions. This can be done by providing a consistent set of processing tools that can be applied to Swarm data in a flexible and configurable manner, i.e. tools that can use different Swarm data as inputs, and whose operating parameters are free to be controlled by the user. The tools will also, where reasonable, reduce the burden of data engineering on the user. The software should enable the reproducible creation of scientifically meaningful outputs (figures, statistics, etc.) with significantly fewer lines of code than are currently required. The toolbox shall be designed to accept generic satellite time-series data as input but can rely on viresclient to provide automated access to a chosen set of Swarm products (one-way integration). In this way, the software allows users to create portable analysis code using the toolbox, while data requirements are resolved by VirES.

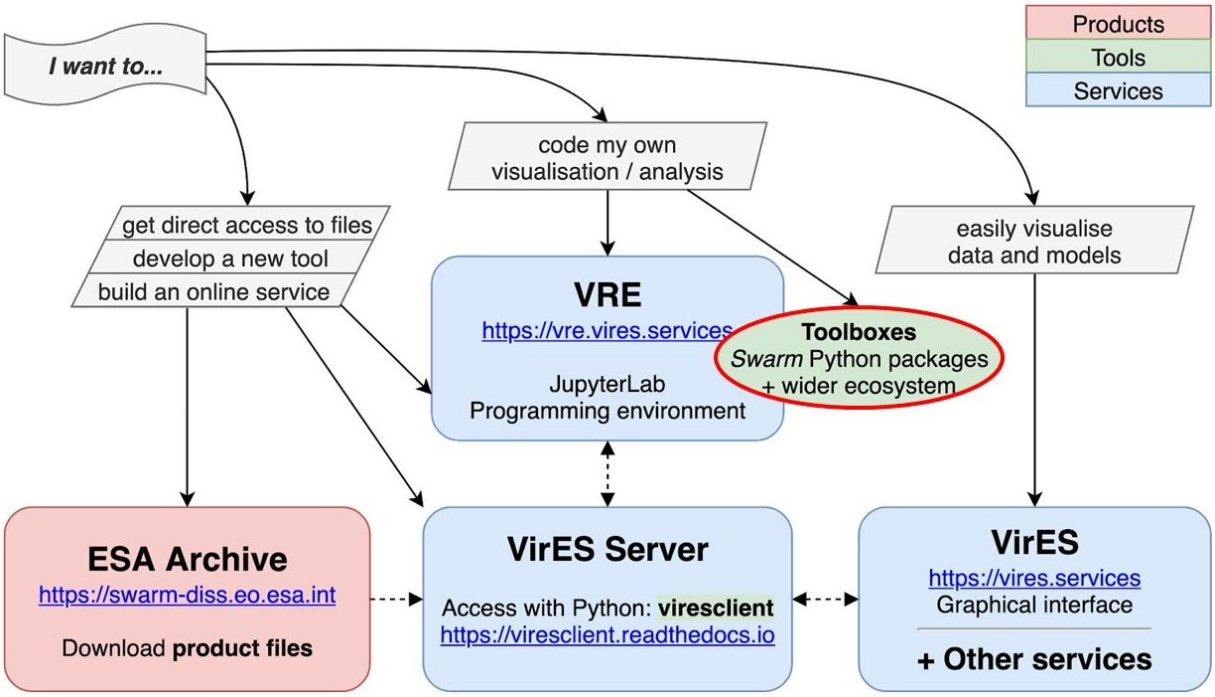


Figure 2 - Flowchart depicting how a software tool fits into the end-user view of the Swarm data and software land-scape. The position of "toolboxes" indicates that they are portable so that they are installable on the VRE but can also be used on other systems.

3.2 Objectives of the Activity

The main objective of this activity is to produce a software toolbox that can produce correlation and maximum/minimum variance analyses between satellite datasets of the user's choosing.

This activity *shall*:

- allow users to input their choice of satellite or satellites for cross-correlation and minimum/maximum variance analysis, the physical parameter they wish to use, the time and date range of the correlation period, the window length of the correlation between datasets and the data cadence on which to correlate
- output the necessary values to allow the user to create plots similar to figure 1 above (see [RD-5]) using correlation and/or MVA techniques
- include quantification of the uncertainty of the output
- generate a small validation dataset for other users to test against
- give examples and use-cases in the documentation
- provide a maintainable, open-source software package (or be part of an existing package), with a plan for how the software can be maintained in the future
- give a talk at a Swarm Data Quality workshop or similar event.

This activity *should*:

- be written in Python (to enable easier interfacing with the existing software ecosystem for Swarm and future maintenance)

- allow command line interface and scripted calls to extract data
- include visualisation tools that allow reproduction of plots such as those given in the examples
- also allow application of the algorithms to datasets beyond Swarm (though testing and validation of this is not expected)
- in addition to the uncertainty quantification above, provide validation checks of the input and output data to warn users about potential problems (e.g. missing data, data quality flags, ill-advised combinations of source data cadence and user-chosen analysis sampling rate).

3.3 Assumptions and Constraints

Approval of deliverables will normally require 14 days for review by Swarm DISC Project Office. Approval of payment milestones is subject to approval of the related deliverables.

The tenderer shall show that they have access to any input data and that the project outputs preferably can be distributed in accordance with the ESA data policies.

Software developed within DISC+ projects should integrate with a wider ecosystem preferably in the form of a Python package in order to ensure sustainability, portability, and independence from paid licenses. If the tenderer does not feel qualified to produce a Python package, DISC can offer either to perform the necessary translation or to help guide the proposer towards a suitable external collaborator. Depending on the scope of the translation task the tenderer shall be prepared to allocate a percentage of the ITT budget to this task to be determined during the contract negotiation meeting.

4 Work to be Performed

All deliverables described here will require an informal review – reviewer to be appointed by the Swarm DISC technical representative – and subsequent written approval.

The following sections describe the tasks anticipated to complete this project. Required output deliverables are listed in chapter 5.

4.1 Work Logic

The work to be performed shall, as a guideline, cover the following tasks: development and validation of the underlying algorithms, followed by implementation of the algorithms as a toolbox with documentation and use-case examples, dissemination of the tools, preparation of plan for future maintenance, and the presentation of the results at an international conference. The compilation into a citable scientific publication is encouraged.

The Contractor is expected to provide a brief summary of the project (about 200 words) to be published on the Swarm mission website a week after Kick Off at the latest.

4.2 Implementation

4.2.1 Task 1: Develop the underlying algorithms

4.2.1.1 *Input*

- Statement of Work (this document)
- Proposal

4.2.1.2 *Task description*

The contractor shall develop an algorithm that can take Swarm (and possibly other) satellite data from any choice of time and date range and carry out a correlation analysis and minimum/maximum variance analysis on a number of variables in the dataset. Parameters specific to the analysis such as window length and data cadence shall be able to be freely chosen within realistic limits.

The algorithm shall be able to produce values that allow plotted outputs similar to the results from the original studies noted in the reference.

4.2.1.3 *Deliverables*

- Description of the processing algorithm (DPA) (Project Report, section 2)

4.2.2 Task 2: Validation of algorithms

4.2.2.1 *Input*

- Statement of Work (this document)
- Output of Task 1

4.2.2.2 *Task description*

In order to make the validation easily reproducible, the contractor shall define an example dataset (e.g. from a real or preferably a synthetic satellite dataset) for all validation purposes. This test data set shall be citable and usable by any others wanting to apply the toolbox. This will allow different methodologies to be compared and contrasted and an uncertainty analysis of the outputs to be made.

Additional validation strategies may be suggested and implemented by the contractor.

4.2.2.3 *Deliverables*

- Validation Report (Project Report, section 3)
- Citable test data set used for validation (DL-01)

4.2.3 Task 3: Development of the software toolbox

4.2.3.1 *Input*

- Outputs of Tasks 1 and 2

4.2.3.2 *Task description*

We would encourage use of the existing ecosystem of Python tools within VRE. The proposer is free to decide the logical arrangement of the software, but it may be separated into elements (modules) which can be used either independently or in composition so as to allow more ways to interact with the toolbox. For example, it may comprise modules providing tools to:

- flexibly run the algorithms given arbitrary inputs [*algorithms*]
- systematically apply the algorithms to given Swarm products [*data pipeline*]
- generate catalogues of results from application of algorithms over long time periods [*data pipeline*]
- visualisation routines to show the outputs of the algorithms and of statistical outputs synthesised from such catalogues [*visualisation*]

The software shall be developed openly online under the Swarm-DISC GitHub Organisation (https://github.com/Swarm-DISC) so that progress may be monitored and influenced during development, both by DISC and any interested external parties.

The code shall include unit tests, where appropriate, and reproduce the expected outputs from the validation test set.

4.2.3.3 *Deliverables*

- Code repository on GitHub or similar with installation instructions (DL-02)
- Brief overview of GitHub repository (Project Report, section 4).

4.2.4 Task 4: Documentation and use-cases

4.2.4.1 *Input*

- Developed code

4.2.4.2 *Task description*

The software shall be accompanied by machine-readable documentation logically linked to the software itself and published to the web to provide users a smooth on-boarding experience to the toolbox and to act as a reliable reference. It shall contain:

- installation and setup/configuration instructions
- tutorial for new users
- cookbook of example use-cases (to include the generation and analysis of the validation dataset)
- systematic description of each component (API reference)

4.2.4.3 *Deliverables*

- Version-controlled documentation published on the web (e.g. using Sphinx and readthedocs.io) with use-case examples (DL-03)
- Brief overview of online documentation for use-cases and test data. (Project Report, section 5)
- Plan for code maintenance and support (Project Report, section 6).

4.2.5 Task 5: Dissemination and Final Presentation

4.2.5.1 *Input*

- All project outputs

4.2.5.2 *Task description*

- Presentation of project achievements at a Swarm Data Quality Workshop or similar event to be agreed with the Swarm DISC Project Office towards the end of the project.
- Delivery of all documentation to Swarm DISC

4.2.5.3 *Deliverables*

- Presentation of project achievements made during a Swarm DQW or similar event (DL-04)
- Final project documentation delivered electronically to the Swarm DISC Project Office in searchable PDF format (DL-05)

5 Requirements for Management, Reporting, Meetings, and Deliverables

The following are the requirements for management, reporting, meetings and deliverables applicable to the present activity.

5.1 Management

MG-01	The Contractor shall assign a responsible project manager as point of contact with the DISC Project Office / ESA.
MG-02	A point of contact shall be assigned for each subcontractor, if any, but generally any correspondence with the project will be via the project manager assigned in MG-01
MG-03	All correspondence between the project and ESA must be via – or if agreed by DTU in copy to – the Swarm DISC Project Office, with attention to the DISC project manager by email or letter post: Klaus Nielsen klausn@space.dtu.dk: Swarm DISC Project Office DTU Space Centrifugevej, Building 356 DK-2800 Kgs. Lyngby Denmark Fax: +45 4525 9701

5.2 Reporting

GR-01	The Contractor shall submit all documents to the DISC Project Office in a searchable, non-protected PDF format, as well as their native format.
GR-02	The Contractor shall ensure that electronic documents do not contain any harmful code (e.g. virus)
GR-03	The Contractor shall produce a short quarterly progress report (or at other interval as agreed), communicated to the Swarm DISC Project Office via SVN and email. This report shall contain highlights of recent achievements, status on work progress, references to publications or presentations, new challenges, etc. Swarm DISC will provide a reporting template.

5.3 Meetings

ME-01	The Contractor shall organize a Kick Off meeting via telecon where key persons are introduced and the project schedule is presented.
ME-02	The Contractor shall at the Mid Term Review present highlights of recent achievements, status on work progress, and plan for the remaining part of the project to the Swarm DISC Project Office via telecon. The presentation should preferably be comprised of a limited number of slides provided to DTU one week before the telecon. ESA reserves the right to participate.
ME-03	The Contractor shall prepare a presentation of the final result (DL-05) and present it to the Swarm DISC community at a suitable event (Data Quality Workshop or conference) in Europe to be agreed with the Swarm DISC Project Office.

ME-04	The Swarm DISC Project Office and ESA reserves the right to call up ad hoc meetings at any time for justified reasons.

5.4 Technical Documentation

The individual deliverables referred to in the task descriptions above and listed below shall be submitted as individual sections in a combined Project Report. This will take the form of a living document to be submitted in revisions according to the schedule outlined in section 6.2 below. A template will be provided by the DISC Project Office. Please note that the following sections will be made publicly available in the Swarm Data Handbook: DPA, Validation Report.

Section	Content
1	Work plan
2	Description of the Processing Algorithms (DPA)
3	Validation report
4	Brief overview of GitHub repository
5	Brief overview of online documentation for use-cases and test data.
6	Plan for code maintenance and support

5.5 Other Deliverables

DL-01	Test data set for validation (preferably including a minimum viable working code to do a set of test correlations)
DL-02	Code reposit
DL-03	Version-controlled documentation published on the web (e.g. using Sphinx and readthedocs.io) with use-case examples
DL-04	Presentation of project achievements to Swarm DQW or similar event.
DL-05	Final Project Report, presentations, publications, and other relevant project documentation delivered electronically to the Swarm DISC Project Office in searchable PDF format.

6 Schedule and Milestones

6.1 Schedule

SC-01	The Contractor shall establish a schedule that is consistent with the planned start of work and the milestones in section 6.2. Any deviation shall be identified and duly justified.
SC-02	The Contractor shall during execution monitor the major milestone schedule. Deviations shall be reported with justification to the DISC Project Office as soon as identified.
SC-03	In the event that delays to milestone deliveries are anticipated, this shall be reported to the Swarm DISC Project Office as soon as possible.

6.2 Milestones

Milestone	Description	Suggested timeline (months)
MIL-01	Project Kick Off • Updated work plan (Project Report, section 1)	KO
MIL-02	Delivery 1 • Description of the Processing Algorithm (DPA) draft • Validation Report draft or plan for validation (Project Report, section 2-3)	KO+4
MIL-03	Delivery 2 – Mid Term Review • DPA • Validation Report • Test data set for validation (DL-01)	KO+6
MIL-04	Delivery 3 – Final delivery • Code reposit (DL-02) • Version-controlled documentation (DL-03) • Plan for maintenance (Project Report, section 4-6)	KO+11
MIL-05	Final Delivery and Presentation • DL-04 and DL-05	KO+12