Work and data flows of the stellar L1/L2 pipeline

Prepared by	Date
Christian Renié, LESIA, Paris Observatory Kévin Belkacem, LESIA, Paris Observatory	15/12/2023
Checked by	**Date**
Rhita-Maria Ouazzani, LESIA, Paris Observatory Marie-Jo Goupil, LESIA, Paris Observatory Sylviane Chaintreuil, LESIA, Paris Observatory Serenelli, Institute of Space Science – Spain T. Morel, AGO, Liège University – Belgium A.F. Lanza, Observatorio Astrofisico di Catania, INAF – Italy M. Cunha, CAUP – Portugal Andrea Miglio (University of Bologna – Italy) J. Ballot, IRAP, Paul Sabatier University – France B. Mosser, LESIA, Paris Observatory – France W.J. Chaplin, University of Birmingham – UK	
Approved by	**Date**
Marie-Jo Goupil, LESIA, Paris Observatory	

Document Change Record

.

Issue	Date	N° change description
1.0	01/07/2021	Initial release
1.01	24/03/2021	Consistency check with PLATO-LESIA-PSPM-DRD-0007 MSAP5-2 and 3 updated
1.02	08/11/2021	MSAP1 major update, MSAP2,MSAP3, MSAP4, MSAP5 updated
1.03	15/11/2021	Consistency check with PLATO-LESIA-PSPM-DRD-0007
1.04	31/03/2022	Update of MSAP1 and MSAP3 workflows after first delivery
1.05	05/04/2022	Name of MSAP3_03, DeltaGAIAMagnitude Add PDP_B_121_MOD_STRUCTURE in MSAP5-part1 workflow
1.06	20/12/2022	MSAP1 and MSAP5 workflows
1.07	04/08/2023	Update of the work and data flows after first specification deliveries Replace WP number by 'SAS' Update MSAP5-2 and 3 leadership Add operating condition
1.08	15/12/2023	Add reference to Moving PDP to IDP_PFU

Contents

1. Introduction

1.1 Purpose of the document

This document describes the high-level overall architecture of the on-ground L1 to L2 processing pipeline related to the data products DP3, DP4, and DP5 (hereafter SAS pipeline) during operations. It thus defines and describes:

- the workflow. The present document establishes the common framework, and defines the links and the interfaces between the various components of the pipeline. This document thus defines the main modules and sub-modules. The baseline is that a sub-module corresponds to a functionality. Each individual component of the SAS pipeline will be described in details in dedicated specification documents.

- the data flow. This document describes the inputs and outputs data for each module and sub-module. The inputs and outputs data are defined in [RD1] and [RD2].

1.2 Scope of the document

This document shall be used as a reference document by WP12 and is the technical top-level document for the SAS data processing specifications.
This document will be updated regularly according to issues resolution and new specification items.

1.3 Reference document

1. [RD1] PLATO-LESIA-PSPM-DRD-0009, Data product description document of the stellar L1/L2 processing pipeline
2. [RD2] PLATO-LESIA-PSM-IRD-0004, Interface request document for the SAS pipeline
3. [RD4] PLATO-UWA-PSM-WPD-0001, PLATO Science Management: Work Package Descriptions - Development Phase
4. [RD5] PLATO-MPIA-PSPM-TN-0066
5. [AD1] ESA-PLATO-ESTEC-SCI-RS-001, PLATO Science Requirements Document (SCIRD)
6. [AD2] PTO-EST-SOC-RS-0247, PLATO Science Implementation Requirements Document (SIRD)
7. [AD3] PLATO-UWA-PSM-RS-0001, L2 Ground Data Processing User Requirements Document
8. [AD4] PLATO-MPSSR-PMC-SIP-0001, PMC Science Implementation Plan Document
9. [AD5] PLATO-ULG-PSM-DD-0035, Description of MSteSci1 and MSAP2 pipelines
10. [AD6] TBD, Description of MSteSci2 pipeline

1.4 Terminology

In this section, we define the terminology used throughout this document.

Term	Description
Asteroseismic parameters	Parameters derived from asteroseismic data that are related to pulsation properties (e.g. oscillation frequencies)

Classical stellar parameters	Stellar parameters that are *not* derived from asteroseismic data. Specifically, the stellar effective temperature, luminosity, surface chemical abundance, projected rotational velocity, and surface gravity derived from spectroscopy.
Global asteroseismic parameters	Parameters derived from asteroseismic data that are related to global pulsation properties. Specifically, the large separation, and the frequency of the maximum height in the power spectrum.
Data Products	The Data Products (hereafter DP) are the final outputs of the mission. They are submitted to requirements as stated in [AD1].
Preparatory Data Products	Preparatory Data Products (hereafter PDP) are any product used as input by the SAS stellar pipeline. These data are computed in advance, stored in the PDC-DB and not modified during the run of the SAS pipeline.
Intermediate Data Products	Intermediate Data Products (hereafter IDP) are mainly internal data products for the SAS pipeline, which are generated by one of the SAS module/sub-module and used as input by another SAS module/sub-module.
Additional Data Products	Additional Data Products (hereafter ADP) are additional output data products delivered by the SAS pipeline, which are not designated as DPs, nor are they used as inputs by another SAS module. They are to be stored in the PDC-DB because of their scientific interest.
Stellar properties	Global stellar intrinsic characteristics such as mass, radius, age, etc.
Level 0 module	high level processes named to as MSAP1 to MSAP5
Level 1 module	sub-module of a level 0 module
Quarters	Each observational run will be subdivided into quarters of 3 months. Quarter numbering is denoted Q=0,1,…

1.5 Abbreviations

DP Data Product

EAS Exoplanet Analysis System
ESA European Space Agency
FU follow-up
GO Guest Observer
GOP Ground-based Observation Program
LC light curve
PLATO PLAnetary Transits and Oscillations of Stars
PDC PLATO Data Centre
PDC-DB PDC-Database
PIC PLATO Input Catalogue
PMC PLATO Mission Consortium
ppm parts per million
PSM PLATO Science Management
SAS Stellar Analysis System
SNR Signal-to-Noise Ratio
TBC To Be Confirmed
TBD To Be Defined
TBS To Be Specified
TBW To Be Written
URD User Requirements Document
URJD User Requirements Justification Document

2. Overall description of the stellar pipelines

2.1 Main objectives of the SAS pipeline

The SAS will generate DP3, DP4, and DP5 from both the Level 1 light curves DP1, data products from the EAS pipeline, associated science preparatory data, and ground-based follow-up data [RD2]. Some of these science preparatory data will be generated by the PSM and PDC prior to the processing and stored in the PDC-DB (i.e., the classical stellar parameters and stellar properties) [RD2]. When available and where necessary, the SAS will generate DP3, DP4, and DP5 using Lg as inputs. All stars will be fully characterized, including stellar rotation, other activity properties, and global stellar properties (radius, mass, and age). For stars exhibiting solar-like oscillations, the pipeline will provide much more precise fundamental properties. The data products and associated data-levels are presented in the following table ([AD4]):

Product	Designation	Level
Validated imagettes, light curves, and centroid curves	DP0	L0
Calibrated light curves and centroid curves	DP1	L1
Planet transits and parameters	DP2	L2
Asteroseismic mode parameters	DP3	L2
Stellar rotation and activity	DP4	L2
Stellar radii, masses, and ages	DP5	L2

Product	Designation	Level
Living catalog of confirmed planetary systems and their characteristics using light curves and transit time variations	DP6	L2
Follow-up ground-based observations		Lg
Living catalog of confirmed planetary systems and their characteristics using new ground-based follow-up observations (Lg)	DP6+Lg	L3

2.2 Before operations

The initial stellar properties as well as the classical stellar parameters will be computed from preparatory observations before operations, in the following modules:

- Module MSteSci 1: Classical stellar parameters determination from preparatory data
- Module MSteSci 2: Stellar properties determination from preparatory data

The architecture of this pre-operation pipeline and the associated data processing are described in [AD5] and [AD6] and illustrated in Fig. 1.

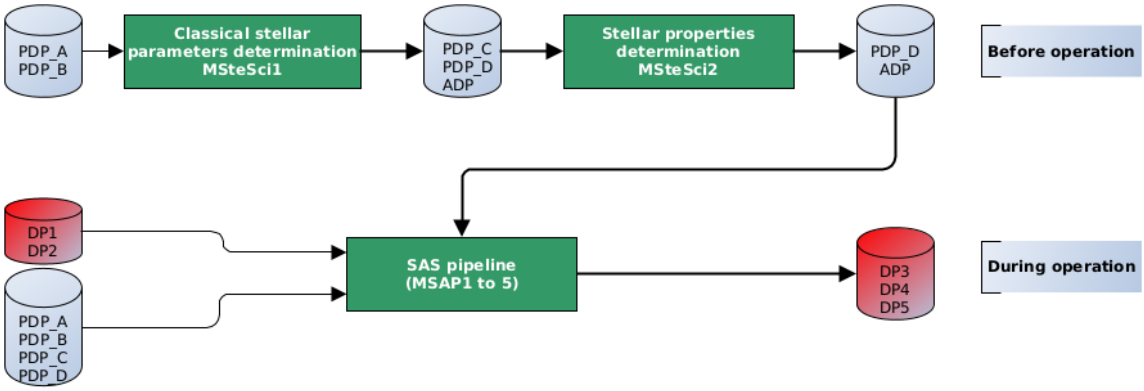

Fig 1: Illustration of the workflow for the overall data processing before and during operation

2.3 During operations: SAS pipeline

The overall SAS pipeline is designed as illustrated in Fig. 2. As a short summary, it is split into five main modules defined as follows:

- Module 1 of Stellar Analysis Pipeline (hereafter MSAP1): *Preparation of analysis-ready light-curves*
- Module 2 of Stellar Analysis Pipeline (hereafter MSAP2): *Classical stellar parameters determination*
- Module 3 of Stellar Analysis Pipeline (hereafter MSAP3): *Stellar oscillation modes detection and measurement* (this module provides DP3)

- Module 4 of Stellar Analysis Pipeline (hereafter MSAP4): *Stellar rotation and activity measurement* (this module provides DP4)
- Module 5 of Stellar Analysis Pipeline (hereafter MSAP5): *Stellar properties determination* (this module provides DP5)

Operating conditions:
The SAS pipeline can be triggered as soon as a new merged and stitched light curve is available from the Level 1 pipeline for a given star and the status of the transit flag has been set by the EAS pipeline. If this flag is true, waiting for the corresponding transit removal kit from EAS is recommended before launching the SAS pipeline. This is the baseline condition for triggering the SAS and it will be detailed in PDC requirements and design documents.

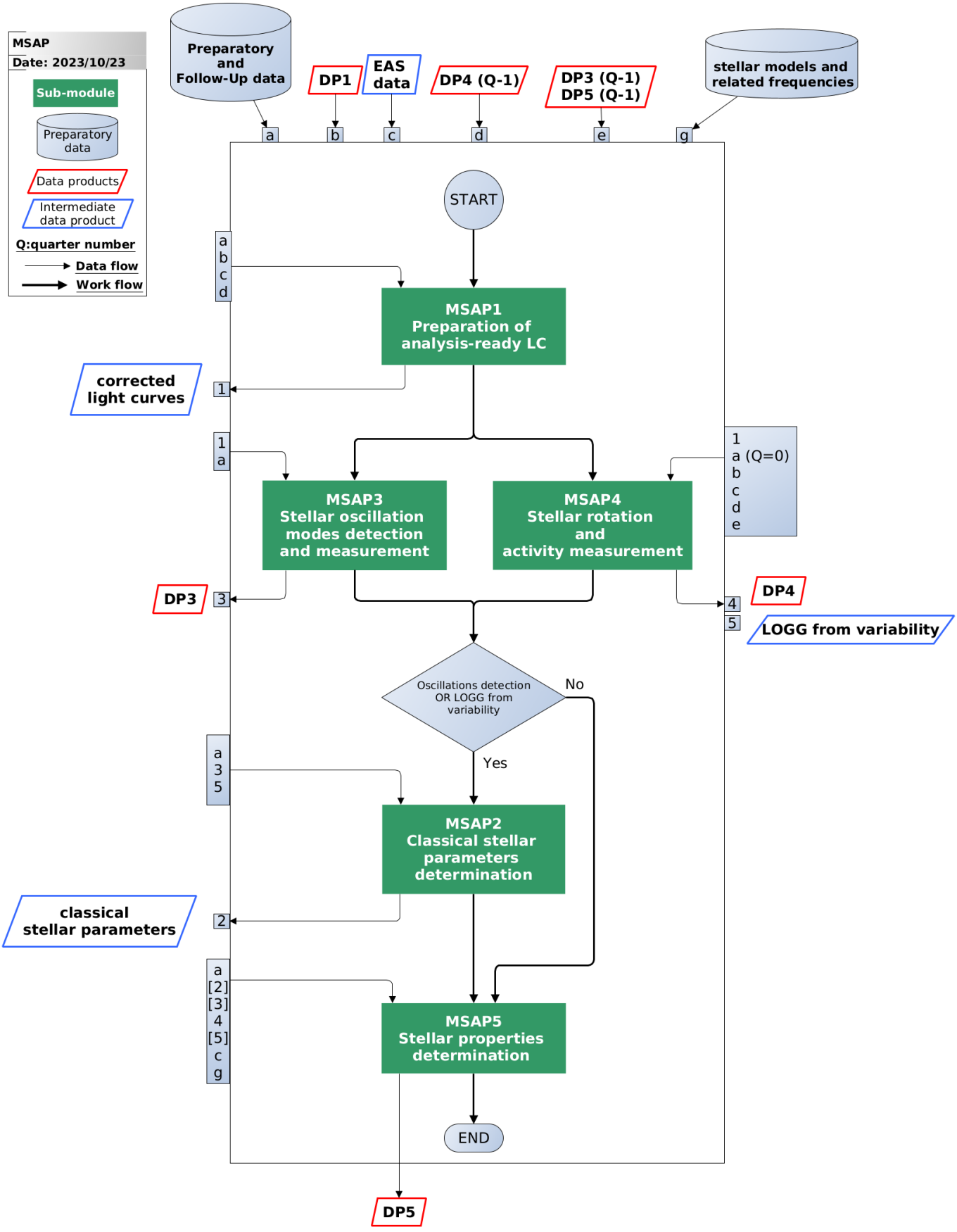

Fig 2: SAS overall work and data flow.

3. Top-level modules description

3.1 Diagrams legend

The following diagrams show at the same time the sequence of the different sub-modules and their inter-dependencies (workflow) and data exchange between sub-modules (dataflow). Data are connected to the different sub-modules through the little squares with letters (main inputs of the module) or numbers when they are the result of a sub-module. The data products, intermediate data products, preparatory data products, and additional data products nomenclature are defined and each data is described in [RD1].

3.2 MSAP1: Preparation of analysis-ready light-curves module

Responsibilities: A. Moya Bedon (University of Valencia, Spain) & A. F. Lanza (Observatorio Astrofisico di Catania, INAF, Italy) [RD4]

Synopsis:

This module encompasses all the necessary preliminary processes on the Level 1 merged and stitched light curves so as to produce analysis-ready light curves for deriving DP3, DP4, and DP5.

In a first step, a dedicated sub-module will detect and flag flares. Also, in the current mission design for PLATO, the detrending of systematics (on all timescales) is expected to be part of the Level 1 pipeline. However, it is anticipated that there will be residuals from this filtering process, and it will be necessary to filter these out to unambiguously detect astrophysical signals related to stellar oscillations and activity. In the case that those methods are not suitable for these purposes, specific methods will be applied for the detrending of the light curves.

In a second step, filtering and modelling of transits and/or eclipses will be present, if relevant. Transit models are expected to be produced by the EAS module for modelling of planetary candidates. This same sub-module must also include the capability of removing flares and then treat any gaps in the light curve, if necessary.

Finally, the power spectrum density is computed for each corrected light curve.

The MSAP1 overall workflow and data flow is illustrated in Fig. 3.

Current assumptions:

- Detrending is expected to be done by the Level 1 pipeline. An out of baseline detrending algorithm is planned as an option if necessary.
- EAS transit removal kit includes best fitting transit or eclipse model.
- Stitching is made by the Level 1 pipeline.

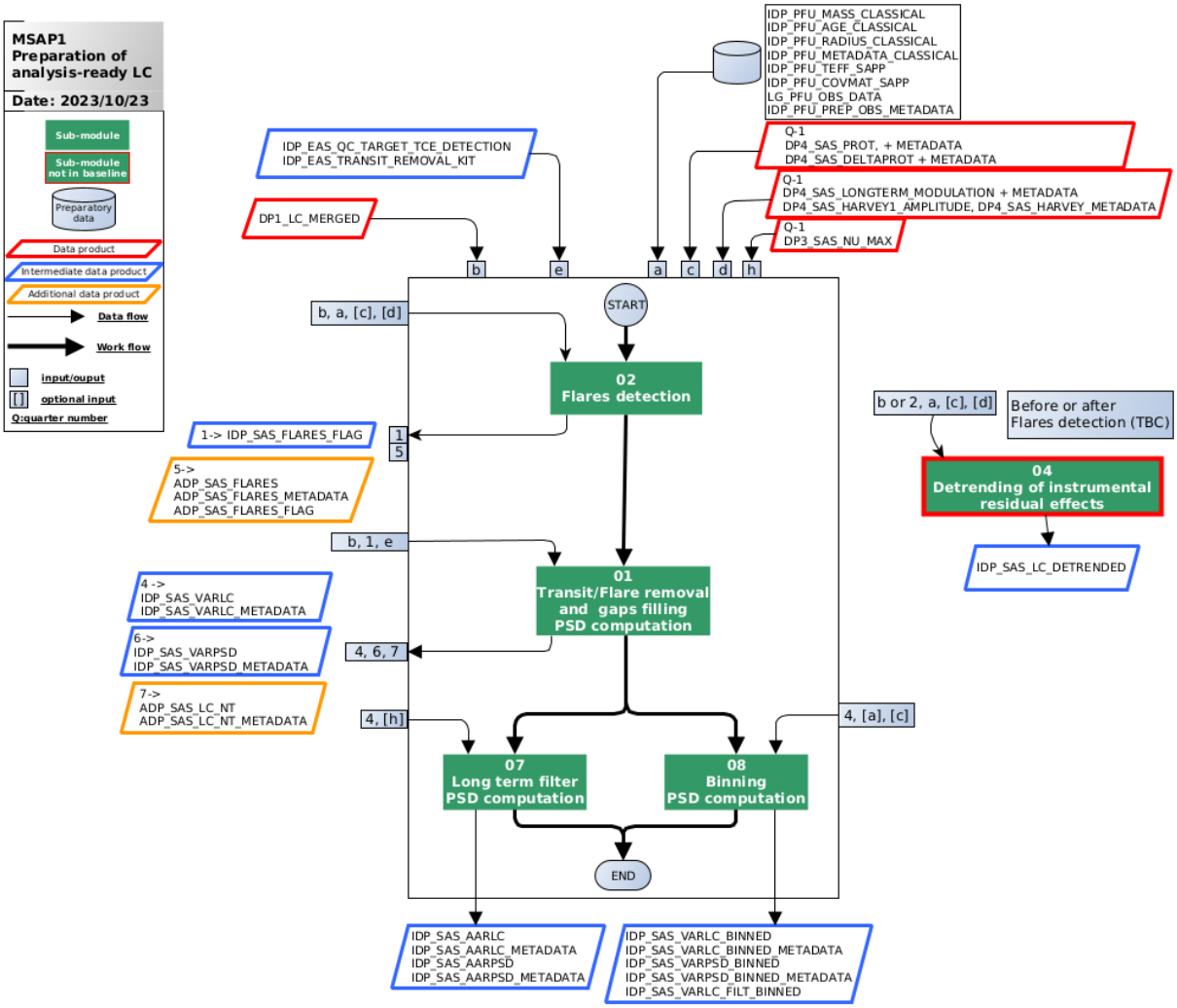

Fig 3: MSAP1 overall work and data flow.

3.3 MSAP2: Classical stellar parameters determination module

Responsibilities: T. Morel (University of Liège, Belgium)

Synopsis:

Although the determination of classical stellar parameters will primarily rely on science preparatory data, it will benefit during operations from the availability of quantities inferred from the PLATO light curve. Where available, precise surface gravities derived from seismic modelling and/or from convectively-driven brightness fluctuations will be used as input for the determination of the classical stellar parameters. Enforcing a prior on the surface gravity leads to improved classical parameters from spectroscopy and significantly increases the precision and accuracy of classical stellar parameters [RD5]. This motivates the inclusion of the module "classical stellar parameters" as part of the SAS pipeline. The determination must be sufficiently precise to permit seismic inferences satisfying the high-level science requirements that are defined in [AD2]. Some classical parameters (e.g. T_{eff}) will be determined from a variety of techniques before and during operations, but a single value based on Bayesian inference will be used for the determination of DP3 to DP5.

The MSAP2 overall workflow and data flow is illustrated in Fig. 4.

Current assumptions: none

Operating condition:

The SAS pipeline will run MSAP2 if surface gravity can be inferred by either MSAP3 (using stellar oscillations) or MSAP4 (using granulation). These two conditions are given by the flags IDP_SAS_DETECTION_FLAG / osc_detection_flag and IDP_SAS_LOGG_FLIPER_FLAG / logg_flag.

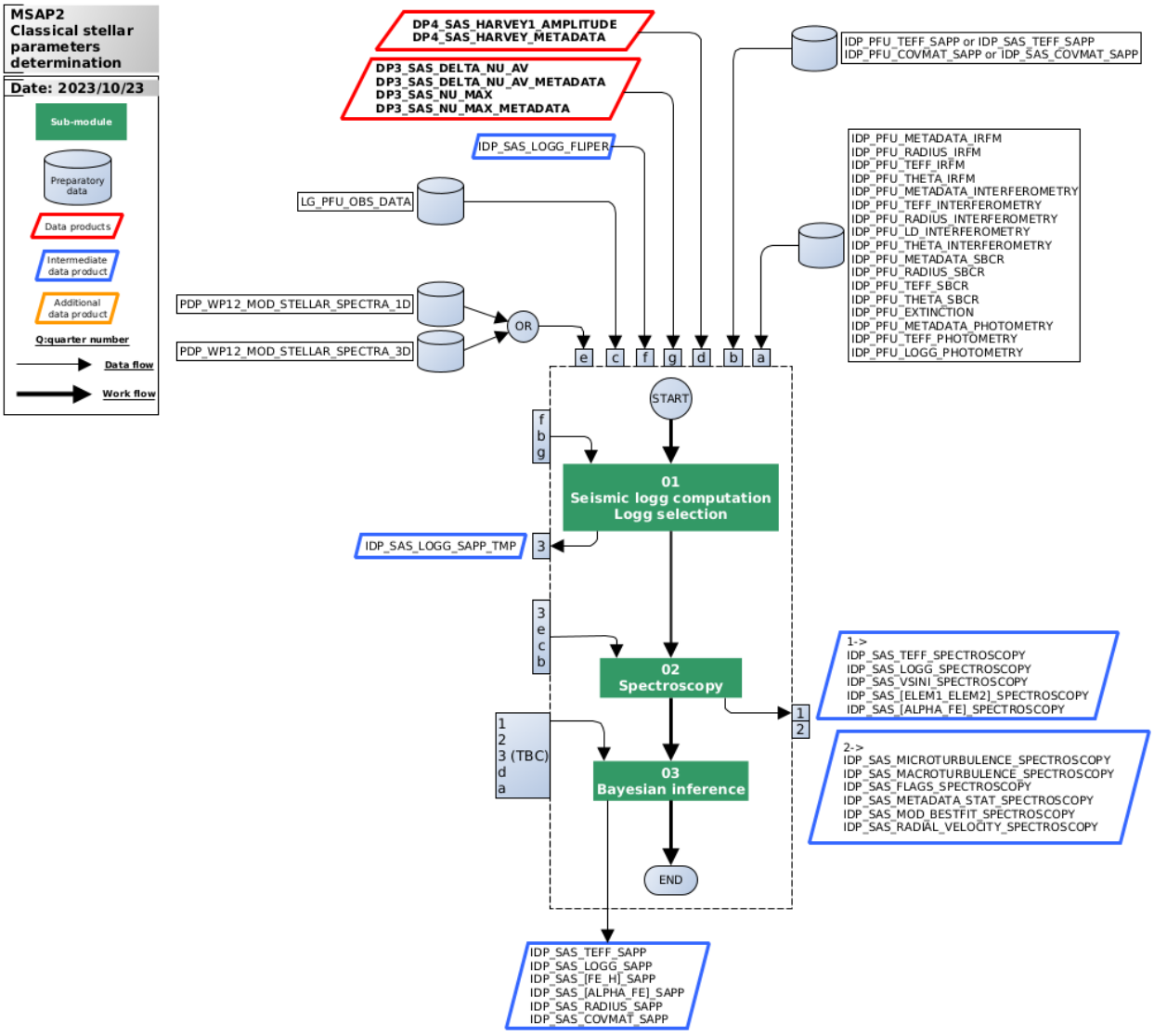

Fig 4: MSAP2 overall work and data flow.

3.4 MSAP3: Stellar oscillation modes detection and measurement module

Responsibilities: W.J. Chaplin (University of Birmingham, UK) [RD4]

Synopsis:

The objective of this module is to measure the seismic mode parameters with the precision required for satisfying PLATO specifications. This module is thus expected to provide DP3. The pipeline will include all of the standard procedures for detecting and measuring signatures of solar-like oscillations based on the legacy of CoRoT and Kepler. For cases where there is positive detection of oscillations, the module will provide average / global asteroseismic parameters and associated uncertainties. Whether or not the SAS can identify individual modes will depend on the quality of the power spectrum, and whether or not oscillations have been detected. In cases with both a detection and sufficient SNR, the module should provide individual frequencies and other individual mode parameters, covariance matrices, and associated uncertainties.

The MSAP3 overall workflow and data flow is illustrated in Fig. 5.

Current assumptions: none

Operation condition:
The MSAP3_04 sub-module will run only if IDP_SAS_DETECTION_FLAG/peakbagging_flag is true.

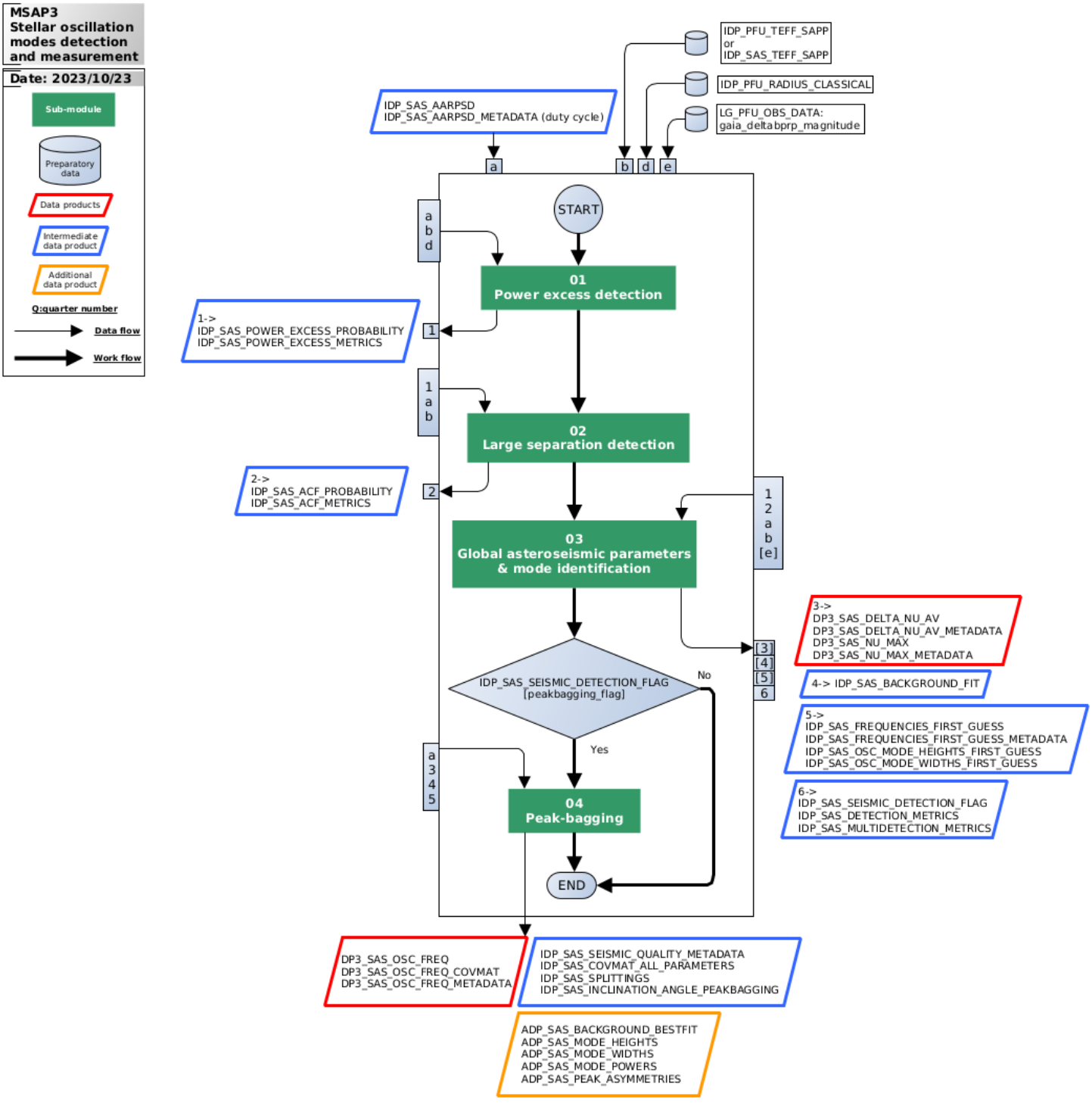

Fig 5: MSAP3 overall workflow and data flow

3.5 MSAP4: Stellar rotation and activity measurement module

Responsibilities: A. F. Lanza (Observatorio Astrofisico di Catania, INAF, Italy) [RD4]

Synopsis:

The objective is to provide measurements of surface rotation periods and, when possible, information on the latitudinal differential rotation and the activity level of the target stars. This module is thus expected to provide DP4.

This module will also provide the surface gravity computed from light curve fluctuations and power spectrum density amplitudes.

The pipeline will then include standard procedures for providing such information, adapted to accommodate the characteristics of PLATO Level 1 light curves. The MSAP4 overall workflow and data flow is illustrated in Fig. 6.

Current assumptions:

‘Spot modelling’ and ‘Fix rotation period’ sub-modules are out of the baseline version, they will be integrated at a later stage according to the maturity level of those algorithms.

Operating condition:
Decision to activate spot modelling or not : To be defined if spot modelling is included in the SAS.

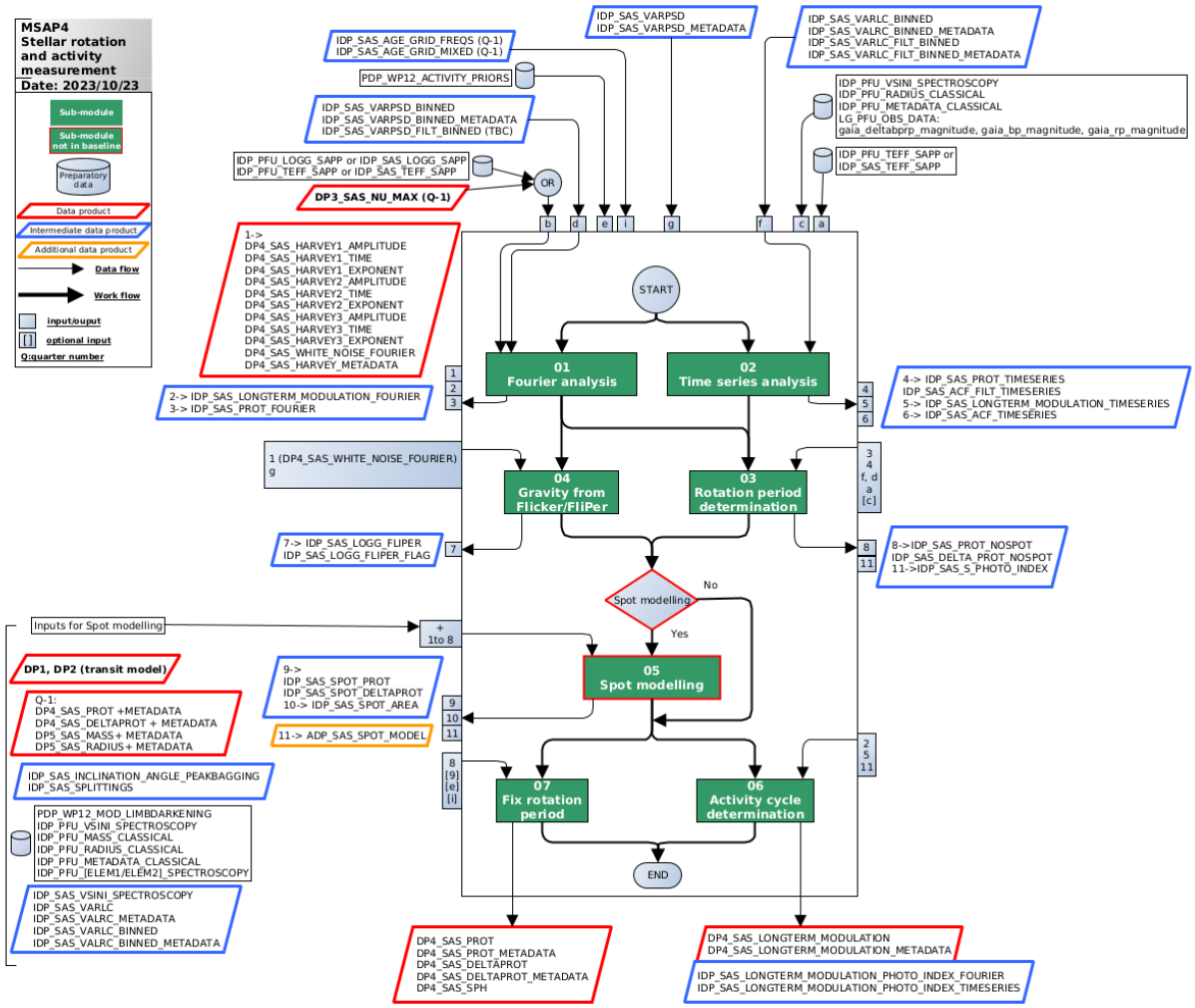

Fig 6: MSAP4 overall workflow and data flow

3.6 MSAP5: Stellar properties determination module

Responsibilities: M. Cunha (CAUP – Portugal) & Andrea Miglio (University of Bologna – Italy) [RD4]

Synopsis:

The objective of this module is to determine stellar properties, specifically radius, mass, and age, as precisely and as accurately as possible. This module is thus expected to provide DP5. Several optimization and grid searching processes will be used simultaneously in order to assess and remove biases and systematic uncertainties. Other information, if available, will be used as well; examples of such information include cluster membership, interferometric radius from MsteSci1, and binarity. Both seismic and non-seismic inferences will be done to eventually apply quality assessments and decision procedures that will permit to derive DP5, which comprises for a given target the mass, radius, and age (also called as MRA). For clarity, this module has been subdivided into three sub-modules, namely:

- MSAP5-1: MRA determination from seismic data
- MSAP5-2: MRA determination from non-seismic data
- MSAP5-3: Selection and validation

The MSAP5 overall workflow and data flow is illustrated in Fig. 7, while MSAP5-1, MSAP5-2, and MSAP5-3 are illustrated in Figs. 8, 9, and 10, respectively.

Current assumptions:

- Some sub-modules are out of the baseline version of the pipeline and they will be integrated according to the maturity level of the algorithms.
- Sub-modules 11, 13, 14, 15 can be run with or without interpolation depending on the quality of data. This decision will be implemented inside each sub-module according to tests done before operation.

Operating condition:

MSAP5-1:

- this part is activated if MSAP3 has detected oscillations in the light curve, so it depends on IDP_SAS_DETECTION_FLAG/ osc_detection_flag
- Individual frequencies: TBD, see issue [DD_0021_I08]
- Mixed modes: TBD, see issue [DD_0021_I09]
- Sophisticated method: at least one year of observation, not in the baseline
- Interpolation: not in the baseline
- Inversion: not in the baseline

MSAP5-2: decisions are made before running each independent sub-modules of this part and are based on proper value of DP4_SAS_PROT, DP4_SAS_SPH and the status of IDP_SAS_LOGG_FLIPER_FLAG.

MSAP5-3: TBD, see issue [DD_0021_I10]

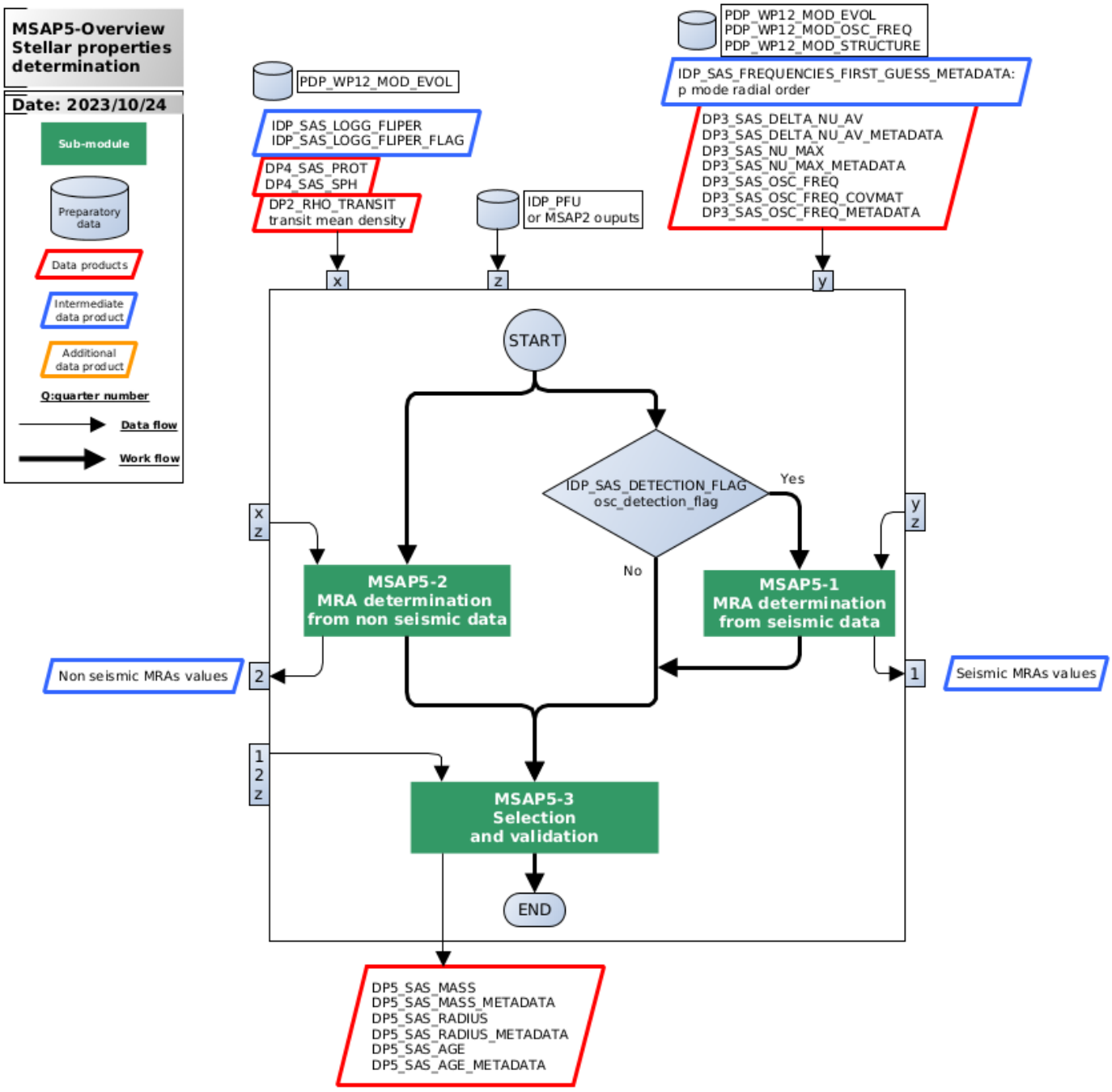

Fig 7: MSAP5 overall workflow and data flow

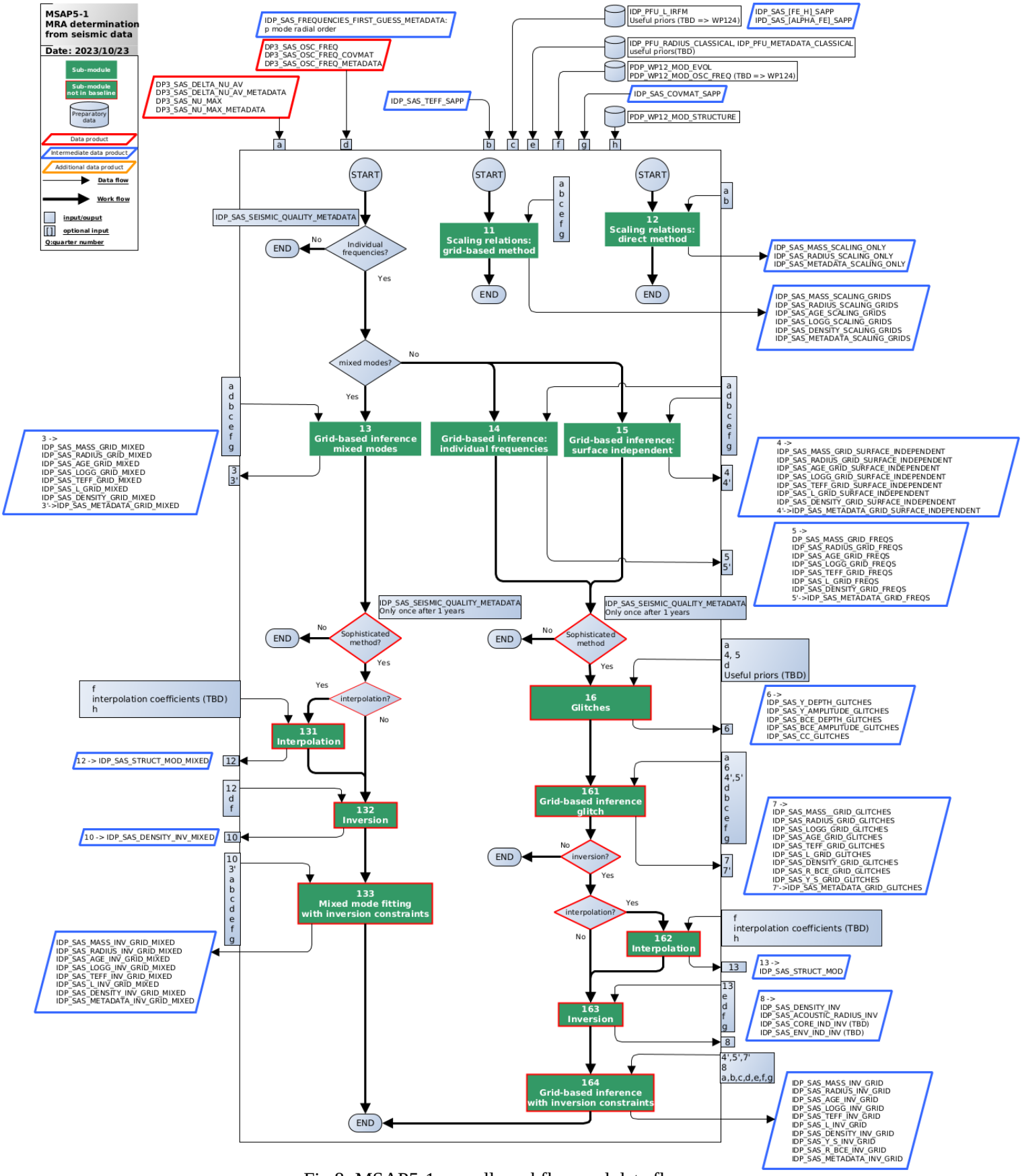

Fig 8: MSAP5-1 overall workflow and data flow

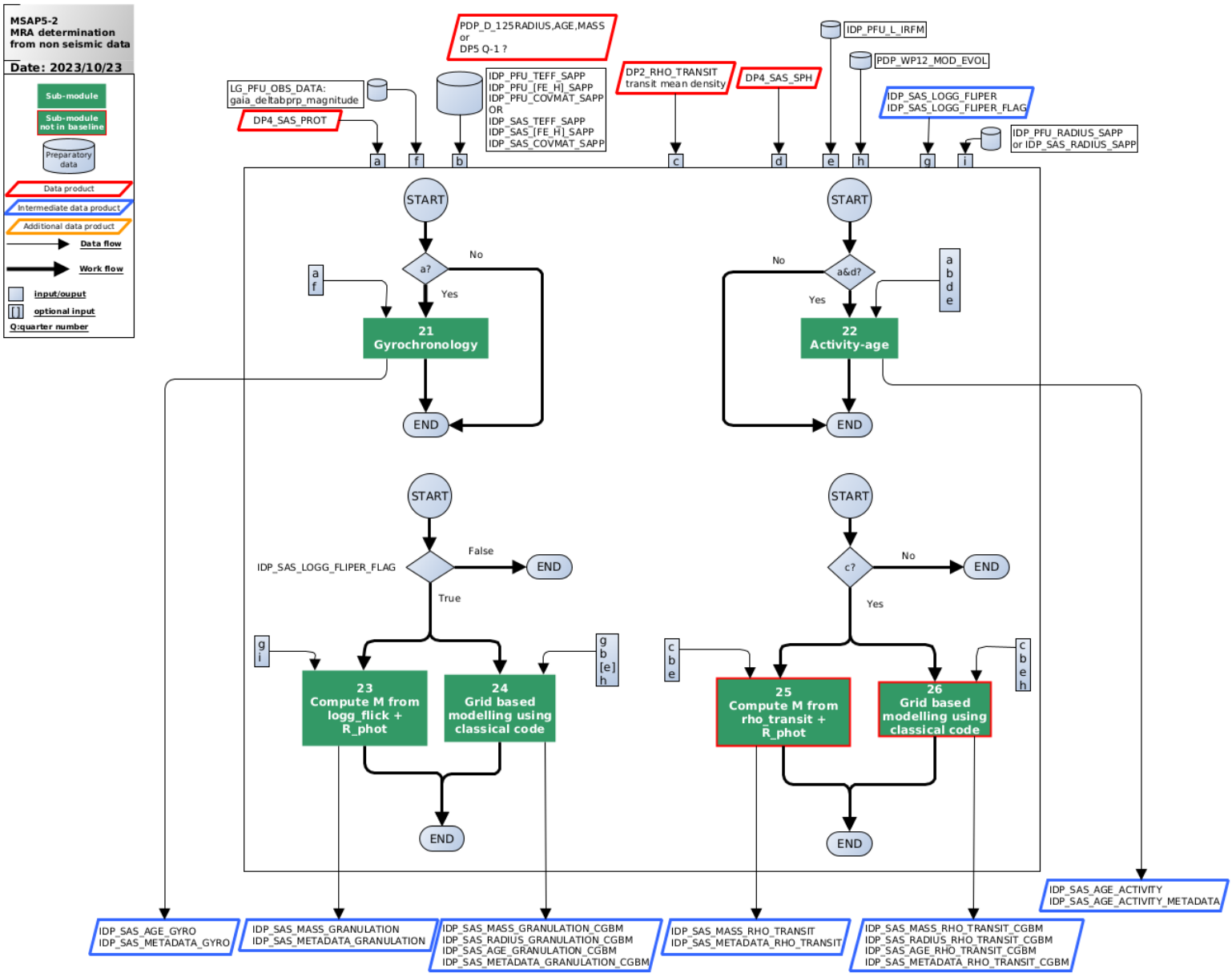

Fig 9: MSAP5-2 overall workflow and data flow

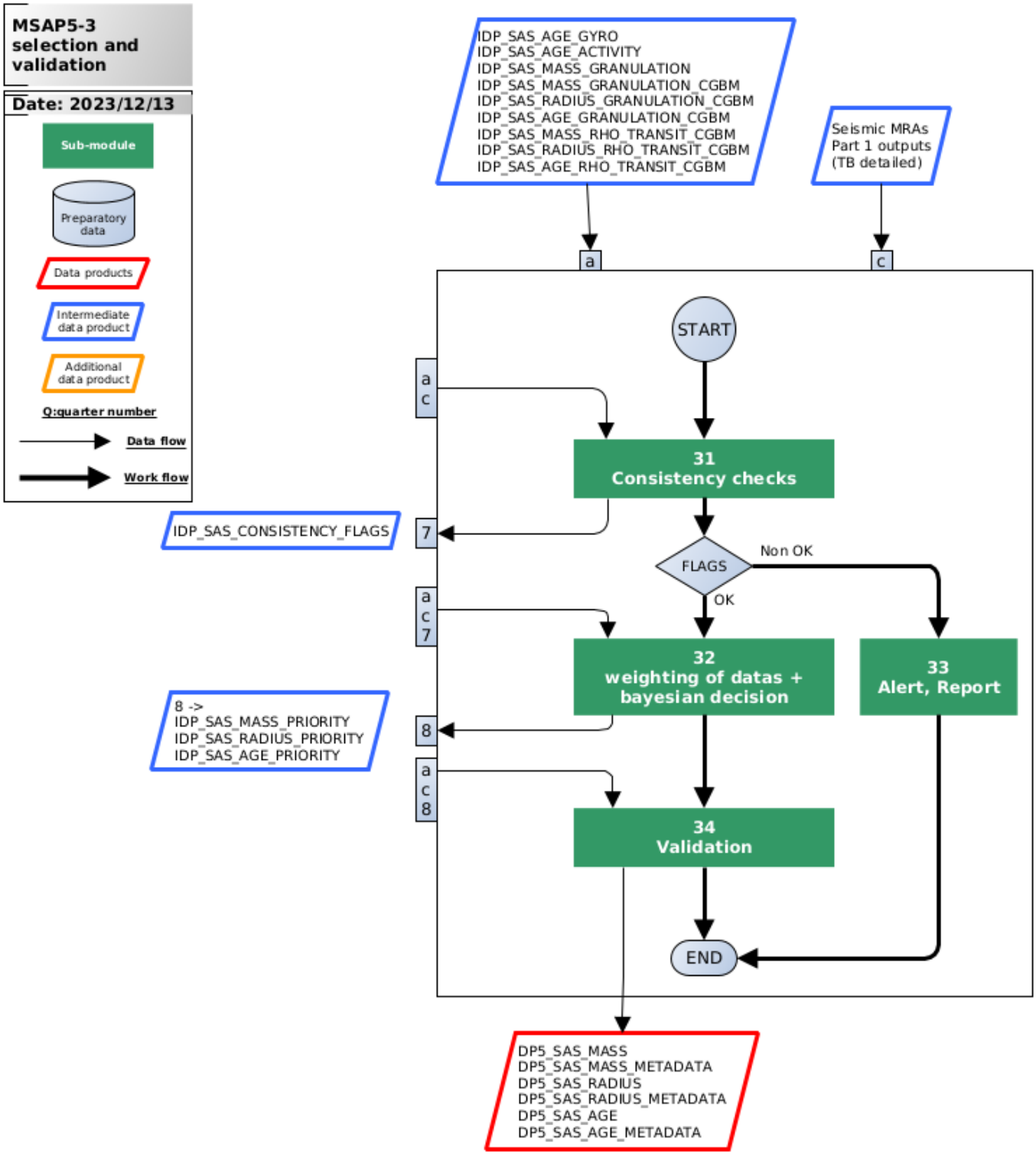

Fig 10: MSAP5-3 overall workflow and data flow

4. Pending issues

~~[DD_0021_I01] : MSAP1: Exact outputs of sub-modules 01 et 02, corrected light curves or flagged points~~. Flagged points for 02, full corrected LC for 01.

[DD_0021_I03] :MSAP2: Confirm that ~~PDP_C_122_F_IR_SED and~~ PDP_C_122_EXTINCTION are not inputs.

[DD_0021_I04] : ~~WP122 proposed to have metallicity proxy, [M/H], and mean abundance of α elements, [α/Fe], not delivered by MSteSci1 and MSAP2, but computed by MSteSci2 and MSAP5 instead.~~ [Fe/H] and [Alpha/Fe] are provided by MSteSci1 and MSAP2.

[DD_0021_I05] :~~MSAP5-2: consolidate sub-modules list and inputs/outputs~~

[DD_0021_I06] :MSAP5-3: consolidate sub-modules list and inputs/outputs

[DD_0021_I07] :Value of a data product when it is not generated.

[DD_0021_I08] :Decision boxes 'individual frequencies' in MSAP5-1: Can it be based on a lower value of IDP_SAS_QUALITY_METADATA/nmodes?

[DD_0021_I09] :Decision boxes 'mixed modes' in MSAP5-1: Can MSAP3 generate a flag for this test?

[DD_0021_I10] : MSAP5-33, sub-module 'Alert, report': do we need this process if IDP_SAS_CONSISTENCY_FLAGS is covering all possible inconsistency cases.

[DD_0021_I11] :

Behaviour depending on star samples:

It is expected that the SAS pipeline will not run fully for all the PLATO targets.

The table below gives, for each main modules of the SAS pipeline, an estimate of the percentage of stars expected to be fully processed by the module.

	Sample 1 >= 15000	Sample 2 <= 1000	Sample 4 >= 5000	Sample 5 >= 245000	Fast camera 300
MSAP1	100%	100%	100%	100%	100%
MSAP3	?% (TBD)	?% (TBD)	?% (TBD)	?% (TBD)	?% (TBD)
MSAP4	100%	100%	100%	100%	100%
MSAP2	?% (TBD)	?% (TBD)	?% (TBD)	?% (TBD)	?% (TBD)
MSAP5	?% (TBD)	?% (TBD)	?% (TBD)	?% (TBD)	?% (TBD)
MSAP5-2	?% (TBD)	?% (TBD)	?% (TBD)	?% (TBD)	?% (TBD)
MSAP5-3	100%	100%	100%	100%	100%