Plan

Survey Management Plan: Back-end Operations

Document No.: VIS-PLA-4MOST-47110-9220-0003

Issue No.: 3.00

DocuShare URL: https://ds-web.aip.de/docushare/dsweb/Services/Document-26665

Prepared by	Jakob Walcher Digital unterschrieben von Jakob Walcher Datum: 2024.10.16 17:50:06 +02'0
	Jakob Walcher
Approved by	
	Jon Loveday
Released by	Digitally signed by Roelof Sybe de Jong Date: 2024.10.17 16:23:43 +02'00'
	Roelof de Jong

Change Record

Issue	Date	Page, Section, Paragraph Affected	Reason, Remarks
0.01	2023-10-05	All	New document from earlier single comprehensive SMP and Google Doc input
1.00	2023-10-16	Complete lists of contributors, acronyms	Release version prepared by RdJ
1.01	2024-06-11		Implementation of RIXes from ESO review
2.00	2024-07-18		Release version
3.00	2024-10-15	Sec 10.3	Updates based on answers to ESO RIXes, update for IWG4, release

List of Contributors

Name	First Name	Institute	Comments
de Jong	Roelof	AIP	PI
Walcher	Jakob	AIP	Operations Manager
Worley	Clare	UC SPCS	Operations System Engineer/Data Release Manager
Couto	Guilherme	AIP	Operations System Engineer
Norberg	Peder	Durham	Operations Scientist
Murphy	David	IoA	DMS developer
Irwin	Jonathan	IoA	DMS developer
Irwin	Mike	IoA	DMS developer
Molaeinezhad	Alireza	IoA	DMS developer
Herpich	Fabio	IoA	DMS developer
Walton	Nicholas	IoA	DMS WP lead

Galkin	Anastasia	AIP	4PA lead
Makan	Kirill	AIP	4PA developer
Knoche	Jörg	UHH	FROG & 4PA developer
Tampieri	Simone	UHH	4PA developer
Alexander	David	Durham	IWG1 lead
Kacharov	Nikolay	AIP	IWG1 lead
Norberg	Peder	Durham	IWG2 lead
Storm	Jesper	AIP	IWG2 lead
Valentini	Marica	AIP	IWG3 lead
Gadotti	Dimitri	Durham	IWG3 lead
Martell	Sarah	UNSW – SoP	IWG4 lead
Sharma	Sanjib	STScI	IWG4 lead
Tempel	Elmo	UT	IWG4 lead
Alam	Shadab	Tat Institute	IWG4 lead
Pelisoli	Ingrid	Warwick	IWG6 lead
Wiseman	Philip	Southampton	IWG6 lead
Church	Ross	Lund	IWG7 lead
Kordopatis	Georges	OCA	IWG7 lead
Croom	Scott	SIFA	IWG8 lead
Krogager	Jens-Kristian	CRAL	IWG8 lead
Davies	Luke	UWA	IWG8 lead
Scaringi	Simone	Durham	IWG9 lead
Heneka	Caroline	UHH	IWG9 lead
Hayden	Michael	Sydney Uni	Deputy Project Scientist

Table of contents

List of Figures

List of Tables

1 Scope

This document describes the Back-end Operations part of the overall Survey Management Plan for the Surveys conducted during the first five years of operations of 4MOST. It first provides an overall overview of the data-flow and the hardware involved in the processing of the data produced by the 4MOST instrument. It then continues with a detailed description of the various pipelines, first the L1 pipeline that removes the instrumental profiles and extracts the 1D spectra from the 2D images coming from the spectrograph, followed by the different L2 pipelines to determine physical parameters from Galactic and extragalactic target spectra, to determine the selection function, an automatic classification pipeline, and a quality control pipeline. The document concludes with a description of the overall data quality control, data curation, data release process, and that data archiving and publication. This document is intended to be read in conjunction with the accompanying Survey Management Plan documents on the Organisation, the Front-end Operations and the Individual Surveys.

2 Applicable Documents (AD)

The following applicable documents (AD) of the exact issue shown form a part of this document to the extent described herein. In the event of conflict between the documents referenced herein and the contents of this document, the contents of this document are the superseding requirement.

AD ID	Document Title	Document Number	Issue	Date
[AD1]	Data Release Plan	VIS-PLA-4MOST-47110-9210-0003	1.00	2023-10-06

3 Reference Documents (RD)

The following reference documents (RD) contain useful information relevant to the subject of the present document.

RD ID	Document Title	Document Number	Issue	Date
[RD1]	4MOST Survey Management Plan – Organisation	VIS-PLA-4MOST-47110-9220-0001	3.00	2024-10-17
[RD2]	4MOST Survey Management Plan – Front-end Operations	VIS-PLA-4MOST-47110-9220-0002	3.00	2024-10-08
[RD3]	4MOST Survey Management Plan - Individual Surveys	VIS-PLA-4MOST-47110-9220-0004	3.00	2024-10-17
[RD4]	Science Team Policies	VIS-POL-4MOST-47110-9213-0001	6.00	2024-10-13
[RD5]	ESO Science Data Product Standard	ESO-044286	8	2022-03-15
[RD6]	Operations Plan	VIS-PLA-4MOST-47110-9720-0001	5.00	2019-09-24

RD ID	Document Title	Document Number	Issue	Date
[RD7]	Data Management System Requirements Specification	VIS-SPE-4MOST-47110-1400-0001	3.06	2023-09-20
[RD8]	Design and Analysis Report QC and Level-1 Science Data Reduction Pipeline Description	VIS-DER-4MOST-47110-1410-0002	2.82	2023-09-15
[RD9]	Template Reference Manual	VIS-MAN-4MOST-47110-9850-0001	2.00	2023-06-26
[RD10]	Data Reduction Library Validation and Test Plan	VIS-PLA-4MOST-47110-1410-0001	1.42	2020-05-26
[RD11]	IWG7 - Galactic Pipeline (4GP) Requirements Specification	VIS-SPE-4MOST-47110-9237-0001	2.00	2023-11-20
[RD12]	IWG7 Management Plan Galactic Analysis Pipelines	MST-PLA-PSC-20307-9237-0001	1.01	2023-04-26
[RD13]	IWG8 - Extragalactic Pipeline (4XP) Requirements Specification	VIS-SPE-4MOST-47110-9238-0001	3.00	2024-07-11
[RD14]	IWG8 Management Plan Extra-Galactic Analysis Pipelines	VIS-PLA-4MOST-47110-9238-0001	1.01	2023-04-27
[RD15]	IWG9 - Object Classification Subsystem Requirement Specification	VIS-SPE-4MOST-47110-9239-0001	2.00	2023-11-20
[RD16]	Definition and provenance of target and object identifiers in the 4MOST data flow	VIS-SPE-4MOST-47110-9700-0001	1.00	2024-10-14
[RD17]	4MOST DMS dataflow design	VIS-DER-4MOST-47110-1400-0002	0.6	2022-10-12
[RD18]	4MOST Acronym List	VIS-LIS-4MOST-47110-9350-0001	6.00	2020-07-08
[RD19]	Definition of the shared-target algorithm	VIS-SPE-4MOST-47110-9700-0002	1.00	2024-10-13
[RD20]	Operational Rehearsal 3 Phase3 plan	VIS-PLA-4MOST-47110-9710-0007	1.00	2024-10-17

4 Definitions

- **Data levels**:
    - Level 0: Raw data with associated meta-data

- Level 1: Calibrated 1D spectra, catalogue of targeted objects
- Level 2: Derived scientific parameters, e.g., elemental abundances, redshifts, selection functions, etc.

- **L2 Data Release Products**
  - DL2 or DL2-IWG: ESO Deliverable-L2 products generated by the common 4MOST pipelines of the IWGs on the agreed schedule in the Data Release Plan .
  - DL2-SURV: ESO Deliverable-L2 products generated by a dedicated pipeline developed by a Survey to be delivered to ESO Science Archive Facility (SAF) on a previously agreed schedule.
  - AL2-IWG: Additional Data Product that will be generated by an IWG on a best effort basis and will only be delivered to the 4PA on a flexible schedule.
  - AL2-SURV: Additional Data Product that will be generated by a Survey on a best effort basis and will only be delivered to the 4PA on a flexible schedule.

5 Pre-amble

An efficient use of 4MOST is maximal if the Consortium's surveys and all other approved Participating Community Surveys are executed in parallel observing mode (ESO-287481). Parallel observing mode means that targets from several surveys (Consortium and Participating Community Surveys) with clearly different target classes and science aims are targeted simultaneously by different fibres of the 4MOST instrument.

Consortium and Participating Community Surveys will join a common Science Team and this document is part of a series of documents that together provide a coherent and unique Survey Management Plan (SMP). The other documents are:

- *4MOST SMP: Organisation* [RD1] containing the organisational structure of the project including conflict resolution pathways, the work-flow and data-flow concepts, the work breakdown structure, and the schedule including data releases.
- *4MOST SMP: Front-end Operations* [RD2] describing the survey strategy concept, the front-end operations plan including the feedback loop, special survey strategy considerations like deep fields and poor observing conditions program, and the calibration plan.
- *4MOST SMP: Individual Surveys* [RD3] describing in more detail the management aspects of the individual Surveys in the 4MOST Project, highlighting in particular those aspects where a Survey deviates from the general plan in terms of target scheduling, data analysis and data products, and/or data publication schedule.
- *4MOST SMP: Survey Simulation Prediction* a restricted access web site showing the input catalogues and simulation predictions for both the overall survey as well as for the individual (sub-)Surveys.

In agreement with the Science Team Policies ([RD4], ESO-286592), the 4MOST PI will represent the full Science Team towards ESO and is responsible for the delivery of the 4MOST survey programme. The SMP should capture the full details of the delivery of one dimensional, calibrated, science-ready spectra extracted from the raw data, which are in common to all surveys (Level-1 data), as well the delivery of the additional science products that may differ from survey to survey, including further processed Level-1, aka stacked 1D spectra to different depth/sensitivities, and the Level-2 data, aka catalogues of physical measurements for the targeted objects/science aims.

Both Level-1 and Level-2 data will be ingested into the ESO Archive and shall adhere to the ESO Science Data Product Standard (ESO-044286, [RD5]).

The 4MOST Survey Management Plan will be made public via the ESO Public surveys web pages.

6 Operations process overview

Figure 1 provides a top-level overview of the work and data flow within the 4MOST Project. All Participating Surveys in the joint Science Team submit their target catalogues with (sub)Survey success criteria to the Front-end Operations System (OpSys) to create a joint database for the survey. The 4MOST Facility Simulator (4FS) is used to optimise the survey strategy parameters before the start of the operations phase. Once operations start, the OpSys selects during the night the next field to visit, assigns science and calibration targets to fibres according to an optimised weighting probability scheme, and then creates an Observing Block (OB) with telescope and instrument instructions to be executed for the next observation(s). An OB is accompanied with an FIBINFO file giving unique IDs to each target/spectrum to be obtained and their sky coordinates. Using the designated Visitor Mode (dVM) OpSys sends the OB to the ESO OB database, which is then used by the Telescope Operator (TO) to start the next observation as soon as the previous one is finished. Sufficient and regularly updated backup OBs are stored at the observatory to carry on observations for two weeks in case of network failure. After the exposure is finished and checked against technical failures by the TO using the QC0 process, the obtained data are sent by the end of the night to the ESO Science Archive Facility (SAF).

On the back-end, the Data Management System (DMS) fetches the raw data (L0), performs the data reduction and creates calibrated 1D spectra (L1) which are stored in the 4MOST Operational Repository (4OR). Spectral Success Criteria are evaluated by DMS and the results retrieved by OpSys to inform on the necessity of further observations of individual targets and to report overall survey status through the Survey Progress Monitor system. The various science pipelines fetch the L1 spectra, determine physical parameters, and feed the resulting data products (L2) back to the 4OR.The data are now transferred to the 4MOST Pre-Release Access Point (4PRAP), where a subset are visually inspected for quality control purposes. The 4PRAP also serves to prepare the different data release candidates (internal, public, ESO). Data Release packages are then handed over to the Public Archive (4PA) for project internal and to 4PA and ESO SAF for public data releases at pre-agreed dates. At very limited intervals during the five years duration of the survey, feedback from the science analysis of the Surveys may be used to adjust the survey strategy parameters to improve the science goal. As this complicates the calculation of the selection function, such modification shall be minimised as much as possible.

Derived from this work-flow, Figure 2 shows an overview of the data-flow structure in the Project. It highlights the central role of data repositories and processing nodes for the front-end at MPE, Garching, and for the back-end at CASU, IoA, Cambridge. The node interfaces with ESO are managed through separate front-end and back-end Interface Control Documents

(ICDs). The Project internal interfaces at the back-end are managed through Data eXchange Unit (DXU) documents.

Further details of the operations concepts of 4MOST can be found in the 4MOST Operations Plan document [RD6].

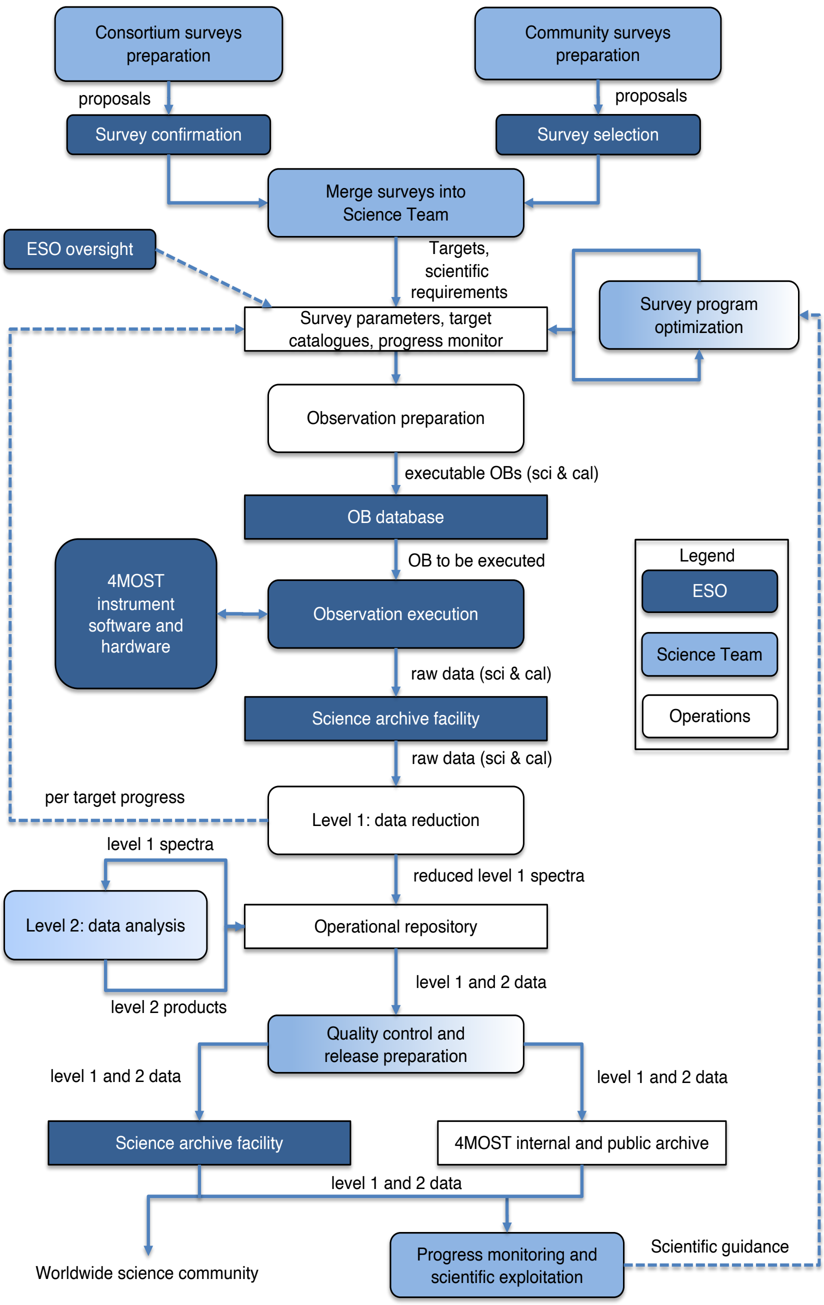

Figure 1: Highly simplified, linear overview of 4MOST operations, with some critical feedback loops indicated with dotted lines.

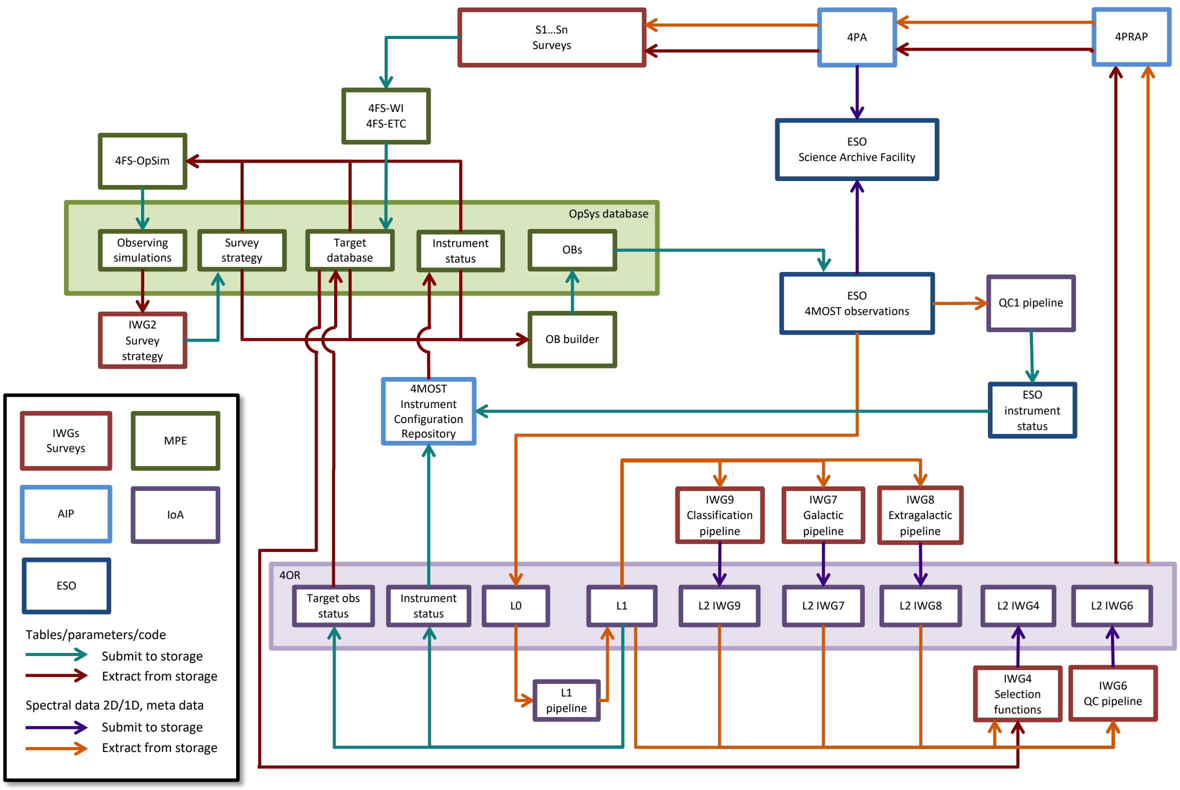

Figure 2: Data flow between Surveys, IWGs and operations compute and storage nodes developed by the different WPs.

7 Hardware

The processing, analysis and storage of 4MOST data products is centralised within a dedicated 4MOST Data Processing Centre (4DPC) located in Cambridge. All nodes have access to high-performance computing facilities with which to run software on data held in the centralised data store. Data availability, status and cross-node signalling is performed via access to 4OR centralised database, permitting inter-node communication via a common interface (4DAP) as described in the 4MOST DMS dataflow design document . The 4DPC provides a dedicated platform for "backend modules" (the IWG9 classifier, TiDES transient analysis, 4XP redshift analysis) with the same (via 4DAP) levels of access to the centralised database.

As a broad overview, the 4DPC comprises:

- 4 compute servers with a total of:
  - 320 cores / 640 threads
  - 3TB RAM
  - 96TB local RAID6 local cache
- A database server:
  - Offering local and external 4DAP access

    - Serving the 4OR DB
    - Hosting Virtual Machines providing the 4OR web frontend and other services such as the transient classifier web frontend
- 1.7PB RAID6 shared centralised disk store
- A dedicated server for backend modules, providing:
    - 20 cores / 40 threads
    - 182GB RAM
    - 96TB RAID6 local cache (shared with 4OR machine)
- Infiniband 100Gbps switch
- Access to the Cambridge Service for Data-Driven Discovery (CSD3) resources
    - 2PB off-site archive at a secured facility
    - Cluster-based usage of 768 CPU cores with 10GB RAM per core
    - 8 A100 GPU cards

Figure 3 shows an overview of the 4DPC, including machine allocations that we will detail in the following subsections.

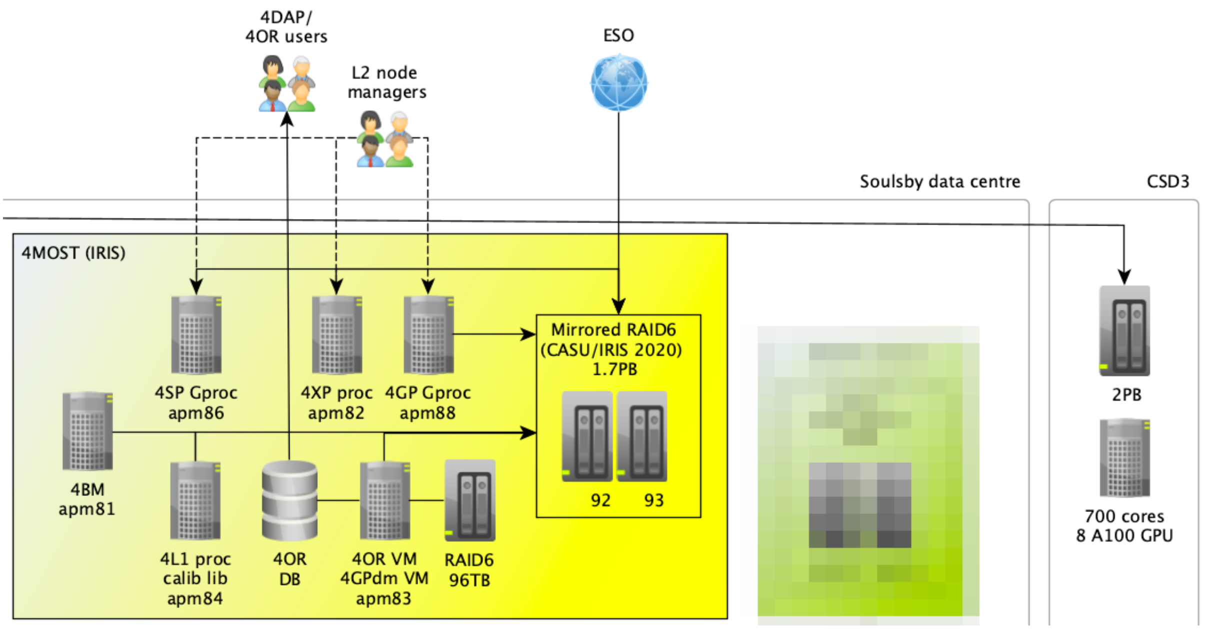

Figure 3: Overview of computational resources available to the 4MOST pipelines under the 4MOST Data Processing Centre (4DPC).

Whilst specific resources have been ringfenced to each pipeline, the 4DPC operates a system of interoperability - namely that any 4MOST pipeline supported under 4DPC should run on any machine (subject to certain exclusions for operational reasons).

7.1 Raw data storage and ingestion into the 4OR

Raw data are retrieved by the 4OR database machine via authenticated TAP query followed by parallelised download of results from the ESO RAW SAF. The FITS files are stored on the 4DISK - a fast NFS shared storage space accessible to all 4DPC stakeholders. The 4DISK is managed by two dedicated disk servers, apm92 and apm93, each with the same specification (jointly serving 4DISK):

The apm92 and apm93 4DISK disk servers		
Component	Specification	Notes
CPU	2x Intel Xeon 6226 2.7G	Gold-class CPU 2.7GHz. 24(48) cores(threads)
RAM	64GB	DDR4 ECC running at 3.2GHz
Disk	1TB local storage	RAID6
NAS	1.7PB exported storage	LVM RAID6 & NFSv4, ACL implemented

4DPC end-users do not have access to the 4DISK servers, only the disk itself.

The 4DPC machine tasked with downloading and ingesting raw data from the SAF is “apm83”. This is the 4MOST Operational Repository machine, and has the following specifications:

apm83 Operational Repository machine		
Component	Specification	Notes
CPU	2x Xeon 6230 2.1G	Gold-class CPU 2.1GHz. 20(40) cores(threads)
RAM	192 GB	DDR4 ECC running at 3.2GHz
Disk	96TB local storage	RAID6
VM	Virtualbox 6.1	Provisioning OR front-end & transient classification
DB	PostgreSQL	v12, content deployed onto local RAID

This machine further acts as the coordinating node for ingest of 4MOST pipeline data products and signalling of data downstream to the archive for wider dissemination. As with the 4DISK servers, the sensitive operational nature of this machine means it is not accessible to end-users.

7.2 L1 processing hardware

The 4L1 pipeline will run at the 4DPC using the "apm84" machine. As with all machines under the 4DPC, apm84 has authenticated TCP/IP access to the 4OR database for the purposes of coordinating dataflow via the 4DAP software package. The specification of this machine is the following:

apm84 L1 pipeline machine		
Component	Specification	Notes
CPU	4x Xeon 6230 2.1G	Gold-class CPU 2.1GHz. 80(160) cores(threads)
RAM	768 GB	DDR4 ECC running at 3.2GHz
Disk	12TB local storage	RAID6, SAS disks

The hardware is leveraged to support the efficient multi-core processing techniques adopted by the pipeline, ensuring sufficient memory footprint per process to handle the data volumes 4MOST generates at each exposure.

Access to this machine is granted to all 4DPC users.

7.3 L2 processing hardware

7.3.1 Galactic, extragalactic, and classification pipelines

The IWG7 (4GP; apm88), IWG8 (4XP; apm82), and IWG9 (4CP; apm81) pipelines will run at the 4MOST Data Processing Centre (4DPC) at the CASU, IoA, Cambridge. The 4DPC hardware is designed to cater for the needs of all L2 pipelines. Specifications of these machines are the following:

The 4XP machine provides a large number of cores in a similar manner to the L1 pipeline.

apm82 4XP pipeline machine		
Component	Specification	Notes
CPU	4x Xeon 6230 2.1G	Gold-class CPU 2.1GHz. 80(160) cores(threads)
RAM	768 GB	DDR4 ECC running at 3.2GHz
Disk	12TB local storage	RAID6, SAS disks

The 4GP analysis pathway potentially includes use of routines optimised for GPU processing. As such, it has been provisioned with CUDA-compatible GPUs:

apm88 4GP pipeline machine		
Component	Specification	Notes
CPU	4x Xeon 6230 2.1G	Gold-class CPU 2.1GHz. 80(160) cores(threads)
RAM	768 GB	DDR4 ECC running at 3.2GHz
Disk	12TB local storage	RAID6, SAS disks
GPU	2x Tesla V100S 32GB	CUDA-compatible NVIDIA Tensor Cores

The 4MOST Backend Module machine is designed to provide computational support to a range of software packages designed to run within the 4L1 pipeline environment (though are not specifically 4L1 modules themselves). These include coarse-grain classification, estimation of target redshifts for use by OpSys, and triaging of transient sources for rapid delivery to the Lasair transient broker. The following hardware supports these modules:

apm81 Backend module machine		
Component	Specification	Notes
CPU	2x Xeon 6230 2.1G	Gold-class CPU 2.1GHz. 20(40) cores(threads)
RAM	192 GB	DDR4 ECC running at 3.2GHz
Disk	96TB local storage	RAID6
Cache	1.6 TB local storage	Fast SAS SSD disk for performance-critical data

7.3.2 Selection function pipeline

The computationally-intensive segment of the Selection function pipeline, namely the generation of ~10^4 simulations over a period of a few weeks, results in a bursty use of resources. A 10GB/core allowance for each simulation provides a throughput of approximately 5,000 simulations over a 2-week period using the “apm86” 4DPC machine assigned to the 4SP:

apm86 4SP pipeline machine		
Component	Specification	Notes
CPU	4x Xeon 6230 2.1G	Gold-class CPU 2.1GHz. 80(160) cores(threads)
RAM	768 GB	DDR4 ECC running at 3.2GHz
Disk	12TB local storage	RAID6, SAS disks
GPU	2x Tesla V100S 32GB	CUDA-compatible NVIDIA Tensor Cores

Due to the bursty usage mode, this machine is likely to be used in support of the 4GP where GPU use is required, as well as for parallel reprocessing work for any pipelines requiring re-analysis of the data.

4SP is provided access to the Cambridge Service for Data-Driven Discovery (CSD3) resources under IRIS, with a throughput of 10,000 simulations in approximately 70 hours.

7.4 Quality Control Pipeline

The 4MOST Quality Control Pipeline (4QP) is developed by IWG6. It runs within the 4DPC environment as an additional step after the L2 pipelines. In that respect its development requirements and in particular hardware requirements are very similar to those of the other L2 pipelines. The 4DPC hardware is designed to cater for the needs of all L2 pipelines including 4QP.

7.5 4MOST Pre-Release Access Point (4PRAP)

The 4MOST Pre-Release Access Point (4PRAP) will be running in AIP. The 4PRAP needs to support about 100 users during survey operations. User management will be provided through the project-wide 4MOST User Management System. User support can be handled by the 4MOST Helpdesk. Over the project lifetime 4PRAP will require at least 500Tb of disk space, however with very little backup facilities, as backups of all basic data are stored elsewhere. The computing power will be provided by a modern compute node with 48 CPUs with 800 Gb RAM. 4PRAP will run Linux for compatibility with other 4MOST systems.

The 4PRAP will run a collaborative environment to allow for online discussions of all user groups (e.g. IWG6, release management, etc.). Common software tools will be developed and run on the system, with software stack archiving on the 4MOST gitlab instance. Common tools will likely include a spectral viewer for QC purposes. Another specific tool to be run on 4PRAP is the interface for the Data Release Manager to apply release flags to targets.

7.6 4MOST Public Archive (4PA) archiving and publication

The 4MOST Public Archive (4PA) will be running in AIP. The compute facilities will consist of a DB server, maintenance node, file system raids and a web application server.

Over the lifetime of 4MOST 4PA will require up to 4PB storage space, including backup facilities. The computing power for the DB server will be provided by a modern dedicated DB server with 32 CPU and 800GB RAM. The web application servers will be run on the virtualized infrastructure of AIP, presumed allocated resources consist of 4 CPU and 16GB RAM. The maintenance node is used to ingest the data into the DB, run 4PA internal QC and maintenance scripts. The maintenance node will be included into the 4PRAP environment.

The design lifetime for 4PA is 15 years.

8 Level-0 (raw data) products delivery time-line to the ESO archive

The raw data (Level 0 or L0 data) for the 4MOST Public Surveys can be accessed via the ESO User Portal as soon as they become available in the ESO Science Archive Facility (SAF). The raw data will become immediately public after observation via the ESO SAF web pages.

9 Level-1 Operations

9.1 Level-1 data reduction process

All the requirements for the L1 pipeline and associated activities to be executed by CASU at IoA are described in the Data Management System Requirements Specification document [RD7].

The L1 and QC1 pipelines are described in full detail in the QC and Level-1 Science Data Reduction Pipeline Description document [RD8]. The main principles are described here, but [RD8] will always contain the most up-to-date status of the L1 pipeline.

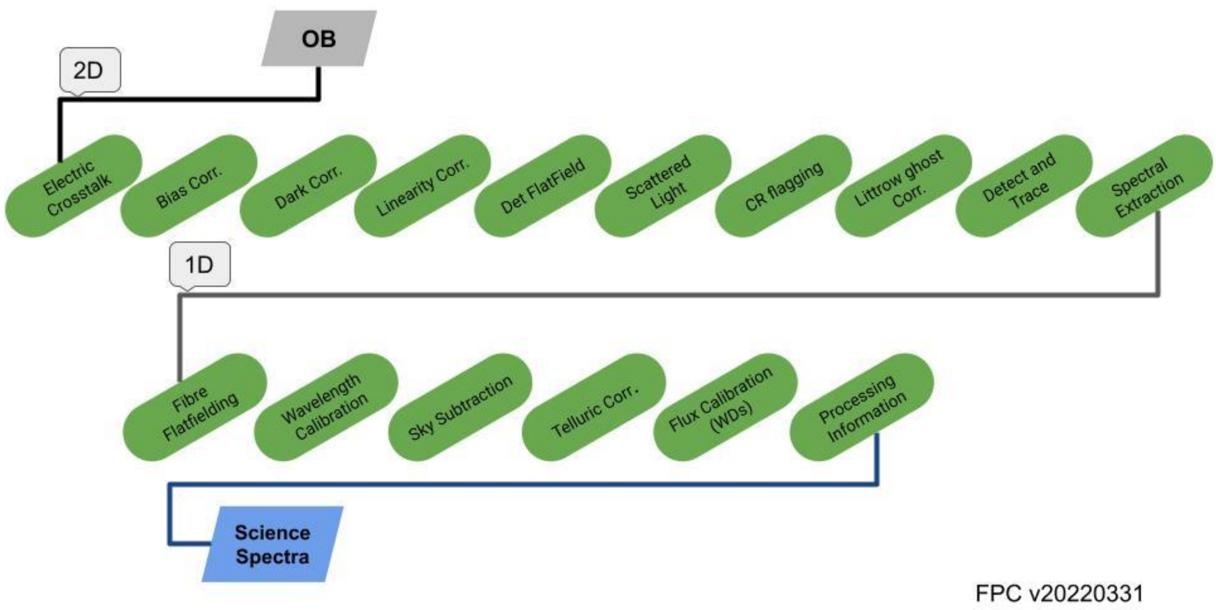

Figure 4: Flow diagram showing the calibration steps required to produce science spectra

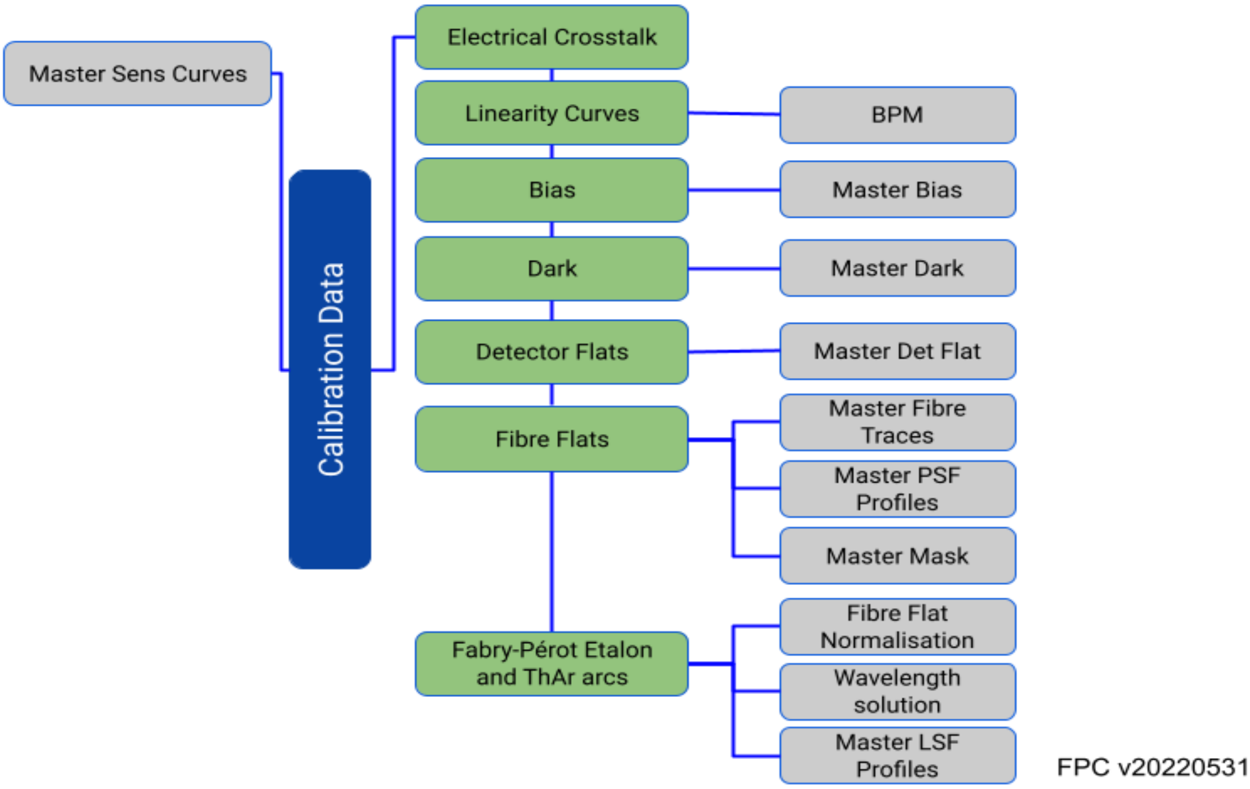

Figure 5: Flow diagram showing how the calibration data are processed to create the Master files used in the data reduction process.

Figure 4 shows the basic steps that will need to be employed to produce the processed spectra. Figure 5 is a similar diagram showing how the calibration data are processed to create the Master files used in the data reduction process. Each one of the different calibration configurations has their reference template in the Template Reference Manual [RD9]. In the following we give a short description of each of the steps in these diagrams.

9.1.1 Electrical Crosstalk Correction

Electrical crosstalk is a common artefact of multi-readout electronics in detectors. Electronic ghosts from one detector may appear on other detectors in the focal plane or there may be ghosting between channels within a single detector that are being read out by separate amplifiers. Typically, what one gets is a scaled image of an originating detector onto a second 'victim' detector. The scale factors are typically quite small in magnitude and can be either positive or negative, hence only the brightest features of the originating image are noticeably replicated. The scale factors are also in general very stable over time. Laboratory spot tests can help determine whether the electronics are susceptible to crosstalk and if so what sort of scaling we can expect, thus providing us with the opportunity to characterise it and investigate ways to minimise its effect. But it is far more important to work out the scaling factors once the whole camera is assembled and on the telescope in a production environment.

Bright arc lines appear as circular spots on a two-dimensional spectral image and hence can be useful for working out the crosstalk between two detector elements. For each combination of detector elements, we can work out a scale factor and hence form a crosstalk matrix C_{jk} that can be applied at the start of any reduction procedure. Provided the terms in the crosstalk matrix are small (i.e. < 1%) then the following single-pass additive correction scheme can be used to correct for this problem.

$$I_j' = I_j - \sum_{k \neq j} I_j C_{jk}$$

The typical error in making a single pass correction is approximately$\langle C_{jk} \rangle^2_{j \neq k}$, which governs the requirement on the magnitude of the crosstalk terms. Note also that in general the C matrix will not be symmetric. It is important to monitor this, especially after there has been some sort of intervention with the instrument.

9.1.2 Bias Correction

To mitigate against the effect of random noise from the readout amplifier a low level DC bias is added to the analogue signal prior to the (unsigned 16-bit) ADC digitisation. The level of this DC bias may vary and hence is monitored by over clocking the CCD readout producing strips of overscan/underscan pixels. In principle a bias frame of the active pixels should register the same average value as recorded in the overscan or underscan sections. In principle an estimate of the DC bias for each readout amplifier is all that is required to bias-correct a detector image. However, the zero-level charge in the active pixels may also have some low level repeatable spatial structure. A master bias frame generated from a stack of individual bias exposures is used to do a full two-dimensional bias level correction to allow for this possibility. The overscan regions in any CCD output can be used to monitor the DC offset together with the readout noise. The master bias image formed by the recipe **qmost_bias_combine** will simply be a two-dimensional image of the light sensitive part of the detector (i.e. the overscan regions will be removed). It will be in the form of a multi-extension FITS file where each image extension will represent the master bias for a given detector. The primary header unit will contain general information about the bias exposures. The statistical information regarding the overscan regions that have been trimmed away will be stored in the image extension header. Each image extension will have a value of EXTNAME that will indicate the detector/arm to which it refers.

The Master Bias images are expected to be created weekly and will be subtracted from science exposures using the overscan areas to determine an appropriate offset value.

9.1.3 Dark Correction

Thermally generated electrons in the detector system are indistinguishable from those generated by detected photons. Cooling the detector to cryogenic temperatures helps to ensure that thermal current (dark noise) is minimal. It is important to monitor the dark current in a detector, as it can be an indicator of problems in the cooling system. Although the amount of dark current is generally small, it is usually worth correcting for it, especially when working at faint magnitudes. Dark current tests usually consist of a series of long (~ an hour) exposures with the shutter closed so there are no external sources of light. These are combined with rejection to remove the cosmic rays and then divided by the exposure time to give an image in units of ADU/s. These can be compared to previous master dark images to look for changes in the DC level or to look for the emergence of spatial structure in the dark current. The master dark image will be formed by the recipe **qmost_dark_combine** and will be a simple two-dimensional image of the relevant detector. These will be stored in a multi-extension FITS file where each master dark image will be an image extension and the statistical information about each will be stored in the relevant image extension header. The primary will contain general information about the observations. Each image extension will have a value of EXTNAME that will indicate the detector/arm to which it refers.

The Master Dark images are expected to be created monthly and will be subtracted from science exposures after being scaled appropriately by the exposure time.

9.1.4 Linearity Correction

The quantum efficiency (QE) of a CCD detector will vary from pixel to pixel and will also depend upon wavelength. The Science CCDs may also have a slight non-linearity in their response function (possibly negligible unless very close to full well capacity). To be able to model the inter-pixel QE variation and non-linearity, a number of LED lamps have been installed near the detector vessel entrance window, which can illuminate the detectors with nearly monochromatic light. The detector non-linearity can then be calibrated and by taking a series of LED detector flats with increasing illumination length and subsequentially fitting a low-order polynomial to the flux scaled by the illumination time. It is most likely sufficient to do this per full detector, but could be done, if necessary, per quadrant or even per pixel. Because 4MOST has no detector shutters, this will be done by switching the detector flatfield LEDs on and off, if necessary, by an increasing (or decreasing) set of short pulses, such that the LEDs do not change in illumination strength due to the increasing (or decreasing) exposure time. Furthermore, the sequence could be interspersed with a set of exposures with constant illumination time to monitor any drift in the illumination strength.

Linearity corrections are expected to be determined on a monthly time scale and corrections will be applied to all science frames.

9.1.5 Bad pixel masking

The exposure sequence obtained for non-linearity corrections can also be used to identify any pixels that strongly and systematically deviate from the expected nearly linear behaviour as function of illumination length. This is done by dividing all (bias and dark corrected and exposure time normalised) frames in the linearity sequence by the Master Flatfield (see also next section). These ratio images should exhibit nearly uniform values. A bad pixel mask is then created by identifying all pixels that consistently in the sequence fall a threshold amount above or below that level. Those that only occasionally fail may have been hit by cosmic rays.

Bad pixel masks are expected to be determined on a monthly time scale and the identified pixels will in all frames be flagged and ignored in subsequent analysis.

9.1.6 Detector Flatfielding

The pixel-to-pixel quantum efficiency (QE) variations of the science detector will also be created from the linearity sequence described in Section 6.7.4. Master detector flats will be formed by the recipe **qmost_detector_flat_analyse** from multiple exposures of LEDs and combined with rejection. The normalized master flat for each detector and its associated variance array will be stored in image extensions of a single multi-extension FITS file. All of the detector level header and QC information will be stored in the extension headers, whereas the primary header will store information about the flat field observations.

Detector Master Flatfield images are expected to be determined on a monthly time scale and divided into all (bias and dark corrected) data frames.

9.1.7 Cosmic Ray Flagging

Here we restrict ourselves to non-invariant random events that cause excess localised counts on the detector. Such affected pixels are located by making use of something akin to a Bayesian Classifier, *i.e.* are the counts in this pixel more probably a cosmic ray event than an expected pixel count level. Generically these methods make use of a prior, *e.g.* the distribution of cosmic ray event properties derived from, say a series of long exposure dark frames, combined with the likelihood of seeing this level of excursion due purely to noise (Poisson and readout). The information on thus identified pixels are then during spectral extraction combined with several other sources of information to attempt to flag or recognise pixels as most probably being affected by cosmic rays and then bypass their use in spectral. During extraction, the full 2D information is directly combined with a more precise 1D subset derived from the model fitting inherent in the spectral extraction process. In other words, the prior probabilities of which pixels are likely to be affected by cosmic rays is updated with the model fitting information to yield a better a-posteriori probability of the presence of cosmic ray-affected pixels. These pixels are then subsequently ignored by the fitting process.

Initial CR flagging is performed on each science image.

9.1.8 Scattered Light Correction

All spectrographs suffer from scattered light to some degree. While this may not be a problem when observing reasonably bright objects, it can be catastrophic for observations of faint objects. Adding to the problem is that the amount and distribution of the scattered light will depend on what is being observed and hence will be different for every exposure. Most of the focal plane in 4MOST will be covered by the overlapping fibre profiles and hence there will only be small parts of the detector, which will be free from fibre coverage and where an estimate of the scattered light distribution can be obtained.

After mapping out the location of all the fibres during the fibre tracing and PSF fitting, a mask of the detector can be made showing locations where there are regions that can be used to measure the scattered light. A coarsely sampled background grid is then formed (with the grid specified by the recipe parameter **scattered.nbsize)**. Within each grid pixel, an iterative k-sigma clipped median value is computed based on the histogram of the flux within the grid pixel using only those image pixels flagged by the mask as suitable for scattered light estimation. The value of "sigma" used for clipping is calculated using the median absolute deviation (MAD).

The resulting grid is then smoothed using median and boxcar filters and interpolated back onto the full input pixel grid of the image to form a scattered light map predicting the scattered light at each pixel. The resulting scattered light map is then subtracted from each 2D science frame.

9.1.9 Littrow Ghost Removal

For the low-resolution observations the Littrow ghosts will probably appear as an extra component to the background scattered light, which can be removed along with the rest of the scattered light as described in section 4.5. The situation is different in the high-resolution modes where the ghost takes on a more concentrated and well-defined structure. If the ghost does not fall off the detector it may affect certain spectral features that are important for some of the survey science. As this is an additive issue and because the shape is well defined, it might be possible to model the shapes of the features using calibration images. However, the strength of the feature will depend on the integrated flux incident at the fibres on the slit *i.e.* it will change with each observation. In principle the ghosts can be modelled on a per exposure basis to remove the features, or if that is not possible by flagging suspect spectral regions. Whether or not this is necessary will be examined during commissioning and LAR using calibration source exposures.

9.1.10 Fibre Tracing

Because of the instrumental optics, the spectra that are imaged on the detectors will not in general co-align with the rows or columns of the CCD. In order to facilitate extracting science spectra, it is necessary to have a measurement of the location of the peak of each fibre's profile as a function of wavelength. This needs to be done using reasonably high signal:to:noise observations of a fibre flat field. The variable signal:to:noise of the science observations as well as the fact that fibres will be allocated to a range of source magnitudes during a science exposure means that this operation cannot be reliably attempted using the science data.

To trace the fibres, we take blocks of rows of data across the dispersion axis in turn and locate all of the local maxima. As the position of each fibre will change very slowly in the spatial domain, even at the edges of the field, it is a simple task to associate the local maxima detected in each row block to a given fibre. Once this is done, a low order polynomial or spline fit can be used to model the curvature of the fibre position on the spatial axis of the image as a function of its position on the wavelength axis.

The instrumental stability should ensure that any spatial shifts of the fibres should be very small in the time between the science and flat field observations. Initially, this will mean that each science exposure will have an attached flat-field exposure. If after commissioning and start of observations it turns out that the instrument is stable enough such that, using the location of the simulcal spectra as reference, it is sufficient to take only flats each visit / a few times at night / only during daytime, we will reduce the number of night time flat field exposures.

9.1.11 Spectral Extraction

Doing a spectral extraction in a scenario where the fibre spatial profiles overlap (as they will with the fibre packing of 4MOST) depends critically on knowledge of the shape and extent of the spatial profile for each individual fibre as a function of its location along the spatial axis and as function of wavelength. The method employed by 4MOST to obtain this so-called Line Spread Function (LSF) from FPE and ThAr arc observations is described in Section 6.5.1, which will be combined with lab information to describe outer the outer wings of the LSF.

Once we have an estimate of the normalised profile as a function of the spatial coordinate *x* and wavelength coordinate λ we can model a row of data across the slit as a combination of the PSFs for all of the fibres together, where each has been scaled by some constant value. That is,

$$d(x,y) = \sum_{n=1}^{Nfib} S_n(y)\psi_n(x,y) + \epsilon(x,y)$$

where *d* is a row of data along the spatial axis (*x*) at a given location along the spectral axis (*y*), ψ_n is the spatial profile of the *n*th fibre, S_n is a scale factor for each profile to be determined and is also the value for the flux of the *n*th spectrum at that particular wavelength and $\epsilon(x,y)$ is the error or noise with expected variance $\sigma^2(x,y)$ from both Poisson and readout noise. This yields a standard system of equations for solution of the form:

$$[C_{11}\ C_{12}\ C_{21}\ C_{22}\ \ C_{13}\ \cdots\ C_{1m}\ C_{23}\ \cdots\ C_{2m}\ \ C_{31}\ C_{32}\ \cdots\ C_{m1}\ \ \cdots\ C_{m2}\ \ C_{33}\ \cdots\ C_{3m}\ \cdots\ C_{m3}\ \ \cdots\ \cdots\ \ \cdots\ C_{mm}\ ]\ [S_1\ S_2\ S_3\ \cdots\ S_m\ ] = [D_1\ D_2\ D_3\ \cdots\ D_m\ ],$$

where

$$C_{ij}(y) = \langle \psi_j(x,y) \rangle_x$$

is the weighted overlap integral of the PSFs for fibres *i* and *j*, and

$$D_i(y) = \langle \psi_i(x,y) \rangle_x$$

is the weighted overlap integral of the *i*th PSF with the data in row *y*.

At first glance this looks rather daunting to deal with, but in practice the matrix is very sparsely populated with most of the non-zero elements in a limited diagonal band representing the PSF overlap from adjacent fibres on the detector array. Also, we would expect the overlap matrix elements C_{ij} to be only slowly varying as a function of wavelength. In practice this means that the solution for the cross-talk corrected spectra $S_k(y)$ can be generated using a simple recursive procedure based, for example, on Gauss-Seidel iteration. Providing the off-diagonal elements C_{ij} in general satisfy $C_{ij}^2 \ll C_{ii}C_{jj}$, which they should, this then provides a fast stable inversion. In addition, the goodness of fit can be used iteratively in conjunction with a cosmic ray mask defined in Section 9.1.7 to remove cosmic ray artefacts from the extracted spectra.

Using this method of extraction reduces the amount of cross-talk with neighbouring fibres by a factor of 10–100 depending on circumstances, while preserving nearly all information available in the 2D image and not adding aliasing or unnecessary noise effects.

9.1.12 Wavelength Calibration

The wavelength solution for the science exposures will be tied to exposures of both arc lamps and Fabry-Perot (FP) sources. FP exposures provide a large number of high S/N emission features, spaced at regular wavelength intervals that are known to a high degree of accuracy. Such exposures can be used to map out the spectral distortion of each fibre across the whole detector. The ThAr arc lamps will only be able to illuminate the simucal fibres. Hence, Fabry-Perot and arc lamp exposures using the simucal fibres will be used to put the relative wavelength calibration done by the FP exposure onto an absolute wavelength scale. If the instrument proves stable enough, then good high signal/noise FP exposures can be done during the daylight hours when the daytime flat exposures are done. It will be a simple task to ensure that at least one good FP is done for each observing mode used during the previous night. If instrumental drift is an issue, then it will be necessary to update the science fibre wavelength solution with FP exposures taken during the night, possibly within each OB. Alternatively, the FP illuminated simucal fibres within each science observation can be used to monitor wavelength distortions across the detector. It is likely that a high-quality solution based on a daytime arc exposure can be adjusted by a low order polynomial wavelength offset correction obtained by either analysis of short FP exposure taken before or after a set of science exposures or ideally using the simucal fibres within each science observation.

The emission spectra will be extracted and the positions of the features in each spectrum on the detector will be mapped to their wavelengths using a separate polynomial fit for each fibre. We will rebin the final spectra onto a heliocentric linear wavelength scale. All spectra will have the same bin size, which will be chosen to be the natural bin size close to the centre of the spectra for the given instrumental setting. This will help to minimise the covariance in the final rebinned spectral variance array. All spectra for a given wavelength setting will be given the same spectral endpoints and bin size which will facilitate stacking and joining spectra from different arms at a later stage.

After commissioning it will be determined whether any further wavelength optimisations based on sky emission lines are possible or necessary.

9.1.13 Fibre Flat Calibration

The shape of spectra that are extracted in any spectrograph is a convolution of the true spectral energy distribution of the object with atmospheric effects, the telescope, corrector, and spectrograph optics, the focal plane vignetting profile, the fibre throughput, and the detector QE, all of which are wavelength dependent. There is a further effect that results from the Aesop fibre positioner and which causes the fibre ends to be slightly off set from the focal plane, resulting in a variation in the throughput of the fibre depending on the angle of the spine. During this step all extracted, 1D spectra of one exposure will be brought to the same relative count level at a given wavelength, for instance to ensure that accurate sky subtraction is possible.

This is done using daytime master facility fibre flats to define the variation of the throughput as a function of wavelength, normalised to unity per fibre. The correct relative normalisation of the fibres is then determined using the measured counts in twilight sky fibre flats, and further adjusted for the throughput changes caused by any differences in spine tilt between the twilight flat and the science fibre configuration using an attached OB-level facility fibre flat taken with the same fibre configuration as the science frames.

9.1.14 Background Correction

All science spectra need to have a detailed background (sky) correction applied. The observations for 4MOST will be designed such that there will be a significant number of fibres, usually 5 - 10%, allocated to monitoring the sky at locations spread spatially around the field of interest. As these are observed at the same time as the science spectra, they should provide an indication of the spatial variability of both the sky background continuum and the sky emission lines. The number of sky fibres should not go below 5% per spectrograph, and will need to be evenly distributed both spatially over the field-of-view and along the slit. The former is to monitor possible sky background gradients and the latter to account for varying fibre LSFs as a function of location on the detector.

However, optimising the sky background correction in multi-object fibre spectroscopy is one of the hardest problems to solve. The idea, in a nutshell, is to form an estimate of the background emission from both astronomical and atmospheric sources that is affecting the spectrum of a science object. This estimated background spectrum is then subtracted from the input science spectrum to give an output spectrum that consists only of emission from the science source. Automatically estimating the background spectrum and accounting for its variation over a large field-of-view is extremely challenging. For example, it may be possible to form a good average master sky spectrum, but if the sky continuum and the sky emission lines vary over the field, knowing exactly how to scale the master sky to account for these variations is difficult. If the atmospheric emission lines vary coherently with the sky continuum, and if we have emission lines in the sky spectrum, we can compare the scale of the emission line peaks between the master sky and the target spectrum to automatically work out a scaling factor. At bluer wavelengths there are insufficient sky emission lines and we have no other choice but to assume that the internal scaling obtained from the analysis of fibre flats has determined the relative throughput of the fibres to sufficient accuracy. In reality sky emission lines do not necessarily vary coherently with the sky continuum so even at redder wavelengths scaling from fibre flats is likely to be the only reliable solution. The data from the three arms of each spectrograph (blue, green and red) is obtained simultaneously and is included in the same FITS file. Hence an analysis of the red arm sky emission lines can in principle provide a scaling factor prior for the green or blue arm data where there will be a paucity of sky emission lines. The optimum solution is most likely to involve using the internal scaling of the fibre throughput to determine the correct amount of sky continuum to subtract, and independently to use sky line scaling, or fibre throughput scaling, to bulk correct for atmospheric emission lines, followed by PCA analysis to deal with the emission line residuals from this process.

The details of the sky subtraction process employed by 4MOST is beyond the scope of this document and the interested reader is encouraged to see the details in [RD01].

9.1.15 Telluric Correction

Correcting for telluric absorption can be done in a number of ways using variants of the methods in the ESO tool ‘molecfit’ (e.g., Smette *et al.* 2015 A&A, 576, 77). Three options have been implemented in the pipeline:

- Use *molecfit* 'as is'.
- Compare the spectra of known white dwarfs, A stars or B stars in each observation with a grid of model atmospheres. An approximate of 10 SED-known White Dwarf will be needed per spectrograph, for each exposure. Each model will be shifted in wavelength space to match the recessional velocity of the observed star under consideration and binned to a resolution appropriate for the observations. The models will then be divided into the observed spectrum. The ratio of the two will be flattened, removing the continuum. The RMS of the region affected by the telluric lines will be determined for each spectrum/model combination. The model that yields the best result will be chosen and used to do the telluric correction for all of the stars in the exposure.
- The *molecfit* spectral library can be used to construct a set of PCA eigenvectors to model the telluric absorption features. A reconstruction can then be done and divided into each spectrum.

The final decision on which method or methods to use will depend on the performance tests using on-sky measurements with 4MOST and for instance the amount and quality of suitable standard stars.

9.1.16 Flux calibration

The processing places all the fibres as close as possible to a common internal flux system in ADUs. This initial calibration pass is based on the analysis of twilight/calibration flats and effectively on the assumption that a known relative amount of light goes down each fibre. The next stage is to make use of any flux calibrators in the field (e.g. WDs and Gaia Bp/Rp targets with known flux distributions and broadband magnitudes) to compute an "average" conversion from ADUs to f_λ as a function of wavelength. Any sources with reliable broadband magnitudes will be used to update the overall flux conversion and hence place those sources on an absolute flux scale. This method also ensures that spectra of targets with no available photometry or extended sources with an uncertain surface brightness distribution on the fibre will still have an approximate calibrated flux available.

The method will follow the below steps:

1. The first pass average absolute flux calibration will be calculated from selected white dwarfs (WDs). This will be the main correction, with an expected high number of points to create the model.
2. Assuming at least Gaia DR3 spectrophotometry is available, a second order correction for atmospheric extinction and wavelength-dependent seeing effects will be performed. This will allow more reliable overall relative and absolute flux calibration. It must be kept in mind that Gaia Bp/Rp data will have very sparse sampling compared to 4MOST.
3. Sources without reliable Gaia spectrophotometry will use an average conversion to absolute flux, derived from step 1.

We will be basing the zero-points on the AB-mag system as defined in e.g. Fukugita et al. 1995. Particularly useful are Tables 5 and 9 from that paper giving the flux of Vega at the effective wavelength(s) of various commonly used passbands and the conversion from Johnson-Cousins to AB mags.

The flux calibration information will be written into a separate extension in the L1 pipeline product – one extension per arm. A static version of the flux calibration sensitivity curve will be used for the QC pipeline.

A long-term goal (years-scale after start of operations) is the characterisation of fibre throughputs for a large number of configurations involving spine tilt-angle orientation, position in the slitlets, position in the fields, atmospheric conditions, airmass, among others variables. Such an ambitious plan could only be achieved after observing stars with known SED in a considerable large number of configurations. Having this information, a Look Up Tensor will need to be constructed, as a source for the conversions. Note that a Lookup Table would not contain all the necessary information, hence the need for a Tensor.

9.1.17 Joined spectra

Given the limited spectral coverage of the High-Resolution Spectrograph arms, there is no clear requirement to stitch the blue, green, and red arms together. The spectral coverage gaps could in theory be filled using say black body modelling or simply flagged as NULL values, but this offers no advantage over using the individual arm spectra.

The Low Resolution Spectrograph contains enough overlap between blue, green, and red arms to run routines to combine the spectra into one spectrum. However, this can only be done after the flux calibration has been applied. All the individual arm spectra will share a common wavelength system avoiding the need for any further resampling of the data. Therefore, non-overlapping regions can simply be slotted into the proto-joined spectrum. In regions where the spectra overlap pixels will be optimally combined using a weight proportional to the relative signal in each pixel and inversely proportional to the noise variance of each pixel. The spectral variance or error array will be populated in a similar way.

9.2 Level-1 data quality assessment

A series of quality assessment tests are performed on each data product generated by the L1 pipeline, irrespective of stack level (single exposure, OB-level to deep stacks). These tests generate both diagnostic plots and data tables that are used in an automated way (with the option of additional manual intervention) to flag any issues with the processed data.

Procedurally, once the L1 pipeline has completed processing a night, a signal is transmitted to the Operational Repository. This signal initiates the OR's multistage modular validation process, culminating in the ingestion of metadata into the database. QA efforts consist of 2 distinct modules within the OR validation environment.

9.2.1 *qa* - generation of QA plots and tables

This module handles the creation of plots and data tables based on a series of tests developed by CASU to gauge (among other things) the performance of the L1 pipeline when applied to data taken under a variety of conditions. The QA tests are designed to complement the diagnostic QC parameters produced during the L1 processing. For example: using skylines, when present, as an independent check on the wavelength calibration; and analysing the sky-subtracted sky spectra to check both the continuum subtraction accuracy and the sky emission line suppression with respect to predicted photon noise residuals

Figure 6 shows a montage of examples of these diagnostic plots for the 812 spectra from the L1 output of a simulated science observation exposure for the LRS-B green arm. The plots are described in more detail in Table 1.

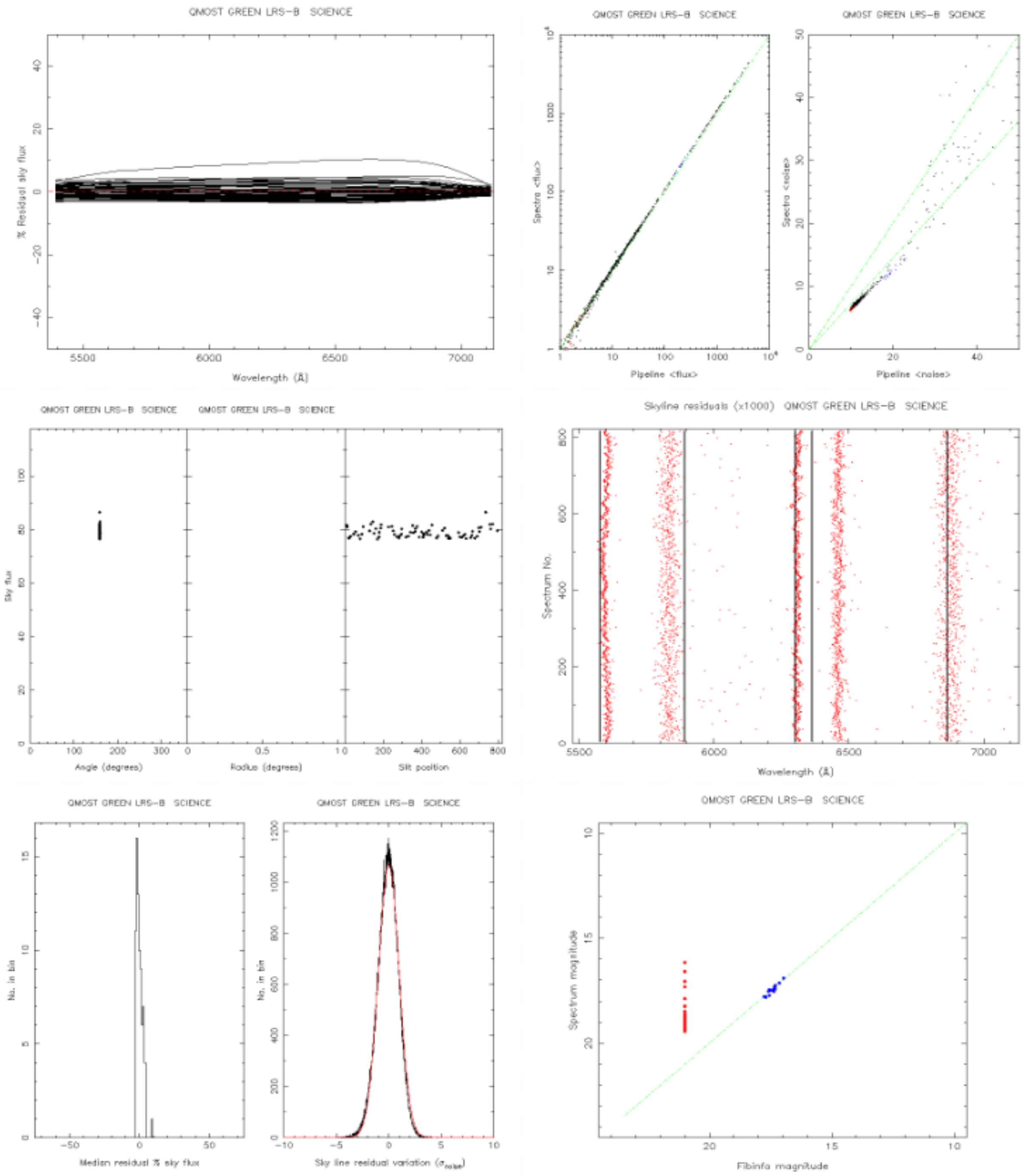

Figure 6: Example diagnostic plots. Refer to the table in this section for the description of each plot.

These plots are available to users via the 4MOST Operational Repository when either searching for specific observations or presented as an ensemble when investigating data taken over the course of a night.

The main features of these diagnostics are listed in Table 1 below.

Table 1: Description of diagnostic plots in Figure 6.

Diagnostic	Figure 10 location	Description
delskys.gif	Top left	Heavily smoothed and resampled plots of the run of sky-subtracted sky fibre flux as a function of wavelength. These are expressed as a % of the average (median) sky level. This works well as a diagnostic of issues with the master sky subtraction. HR data taken in dark sky would look worse as a %, but the sky-level variation there would be typically +/-1ADU.
flxnois.gif	Top right	Black points are targets and, red are sky-subtracted sky spectra, blue are WD calibrators. The pipeline <flux> is estimated in the nearest passband, whilst the y-axis average is a median of the 10%-90% range of the spectra. The pipeline pixel-to-pixel <noise> estimate is based on the inverse variance array and is reported as sqrt(median variance). The spectrum <noise> is estimated directly from the pixel-to-pixel noise in the spectrum. Note that noise estimates direct from spectra are generally lower due to noise covariance caused by resampling but can be corrupted (higher) by "forests" of absorption lines, particularly in blue LR. Dashed lines are 1:1 and predicted slope from noise covariance analysis.
skyflux.gif	Middle left	Distribution of average (median) sky flux in sky fibres as a function of radial distance from the centre and azimuthal angle. These can be used to assess issues with sky continuum variation or badly contaminated sky spectra. Note that the Ra Dec coordinates in this example were incorrect affecting the radial and azimuthal plots.
skyline.gif	Middle right	Measured wavelengths of selected isolated skylines (red) compared to fiducial values (black) for all allocated fibres. The differences have been scaled by x1000 for better visualisation. For example, an apparent offset of 100A corresponds to a 100mA shift. The typical rms error in measuring a skyline position is around 15mA (see QA table below for a summary).

skystat.gif	Bottom left	Left panel shows the median sky-subtracted sky level as a % of the average sky level. With a robust master sky and good subtraction, the distribution should be centred on zero. Outliers usually correspond to contaminated sky spectra but may also be caused by a poor master sky. The right panel shows the sigma (noise) normalised distribution of sky residuals leftover at locations of sky emission lines. Black is the actual distribution, and red is theoretical if Poisson noise limits are attained.
magplot.gif	Bottom right	Y-axis: magnitudes computed directly from spectra after multiplying by the sensitivity function (extension#10 for green arm) in g, r or i SDSS passband as appropriate, and based on the SDSS sensitivity curves. X-axis: closest magnitude available in fibinfo table (extension#16) - in this case SDSS r. Red points are estimates of sky brightness in band from all-sky fibres compared to reference sky brightness value in PHDU blue - WD calibrators black - science targets

In addition to these diagnostic plots a data table is generated, providing numerical metrics for determining the performance of the pipeline as applied to that particular data product. An example of this can be seen below in Figure 11, where we have clipped out most of the data entries for reasons of brevity.

```
QMOST GREEN LRS-B  SCIENCE

No.   Mag    <flux>s  <flux>p  <flux>t  <sigma> sqrt<var> s:n fibuse nused     RA       Dec

Sky covariance: -0.019 -0.012  0.010  0.443  1.000  0.443  0.010 -0.015 -0.015
Noise factor  :  1.380

log10(KS) stats sky -v- targets =   -0.00    0.00   -0.21
Average sky flux and no. of sky =   78.53      77
Residual median sky sigma and % =    2.56    3.26
No. of fibres with extremely low flux  =   0
Skyline residual rms & kurtosis =    0.95    1.06

   Skyline   <offset>  rms   <FWHM>   rms
[OI] 5577.338  0.022  0.010  1.012  0.020
[Na] 5889.950 -0.055  0.023  0.999  0.044
[OI] 6300.304  0.007  0.008  0.988  0.030
[OI] 6363.780  0.097  0.012  0.973  0.035
72Q1 6863.955  0.010  0.026  0.973  0.058

No. of pixels with -ve IVAR and IVAR >1  =         0         0
No. of pixels with -ve NOSS out of total =      1003   5820582
```

Figure 7: Example QA table output. Information from this file is used to record requirement compliance for L1 data products. This file is a sample and not representative of the full data content of a QA output table.

Data stored in this table are used as an input into the qa_analyse module as a means to test against a series of requirements on the L1 pipeline.

9.2.2 *qa_analyse* - analysis and assessment of QA data

Following successful generation of QA outputs from the *qa* module, control is passed to the *qa_analyse* module designed to ingest data in the 4OR QA tables and (where appropriate) flag data measured as outside of the defined bounds.

The qa_analyse module creates entries into the 4OR QA table describing the source of data being tested, the thresholds needed to meet the requirements and goal, the outcome of the test and an assessment of whether this data source meets this requirement. The schema for this database table is:

Table 2: 4OR QA table schema

The 4OR qa table		
Column	Type	Description
id	integer	Primary key (used only for index)
nightobs	integer	Night of the data (YYYYMMDD; indexed)
ob_id	integer	ID for the Observing Block
specid	bigint	CASU SPECUID for the file analysed
testid	bigint	Identifier for the test performed
req	float	The REQUIRED value this test must produce
goal	float	The IDEAL value this test should produce
val	float	The result of the test
state	integer	Assessment of the QC test (0:failed, 1:passed requirement, 2:passed goal)
info	char(255)	Text field for comments on test (255char max)
timestamp	timestamp	Timestamp of row creation

We note that many of the thresholds for testing are TBD during commissioning, and once we understand the overall stability of the instrument. Where possible, we derive the test requirements directly from requirements placed on DMS from [RD7].

In cases where an entry is flagged with state=0 (failed) this status propagates to the overall validation state of the night. Email alerts are issued to notify DMS where requirements have not been satisfied, for escalation and potential re-processing of the data. In cases where the requirement has passed but the goal has not been met (state=1), a warning is added to the validation log for that night. Formally, this blocks downstream flow of the L1 data (to L2 pipelines) until the full validation process is re-run and either the output products improved or (where the quality of the data precludes this) exemption flags are added to permit completion of the validation process subject to appropriate flagging.

9.2.3 Pipeline QA measures

The full details of the L1 analysis code quality assurance procedures are described in the Data Reduction Library Validation and Test Plan document [RD10]]. The quality control parameters are given in the DRPD [RD8]. All quality control processes will be run on a night basis. QC metrics will be stored in the QC database and made available via the 4OR web frontend for inspection and trend analysis.

9.3 Staff effort devoted to Level-1 data reduction and quality assessment

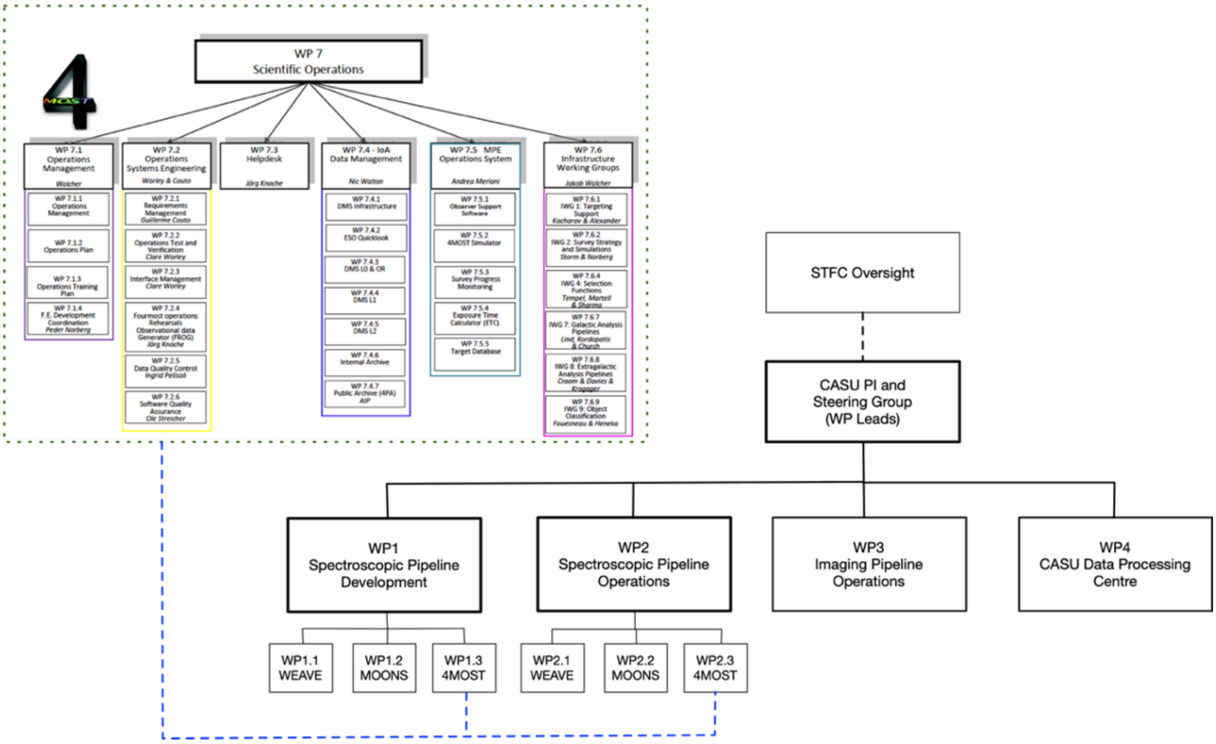

Figure 8: Organisational structure of the DMS within the context of the overall 4MOST project.

Figure 8 shows the WP7 Scientific Operations WBS, with WP7.4 indicated. The 4PA activity is based at the AIP, the rest of WP7.4 is located at CASU, Cambridge. The diagram to the right shows the mapping to the activities for WEAVE, MOONS and 4MOST carried out support through the UK STFC grant.

The staff effort breakdown is provided in Table 3.

Table 3: Staff effort breakdown for DMS at CASU, University of Cambridge.

Name	%FTE	WP	Role
Nicholas Walton	0.1-0.2	7.4	Coordination and reporting. DMS LPM
David Murphy	0.5-0.75	7.4.1, 7.4.3, 7.4.6	DMS infrastructure, DMS L0, L1, OR
Alireza Molaeinezhad	0.25-0.5	7.4.3, 7.4.5	OR front end feedback, L2 integration
Giuseppe D'Ago	0.25-0.75	7.4.1, 7.4.5	Back end modules, L2 integration
Fabio Herpich	0.25-0.75	7.4.1, 7.4.3, 7.2.1	DMS Requirements management, OR, DB model
Mike Irwin	0.25-0.5	7.4.4	L1 pipeline
Jonathan Irwin	0.25-0.75	7.4.2, 7.4.4	QC and L1 pipelines

Staff effort % 's are given as ranges reflecting the evolving situation during the operations phase (e.g. commissioning peaks).

10 Level-2 Operations

This section only describes the L2 data products produced by the common pipelines of the Project maintained and operated by the Infrastructure Working Groups (IWGs). The Deliverable L2 of individual Surveys (DL2-SURV) are described in individual Survey sections in the Survey Management Plan – Individual Surveys document [RD3], as are any optional Additional L2 products (AL2).

The 4MOST Project has four common pipelines maintained by the IWGs: the Galactic Pipeline (4GP), the Extragalactic Pipeline (4XP), the Selection Function Pipeline (4SP), and the Classification Pipeline (4CP). Each of them contains several modules, with the basic components summarised in the following subsections. The format of each data product created is described in the Data eXchange Units (DXUs) document available for each pipeline and checked at each data ingest in the 4MOST Operation Repository (4OR) at CASU. Each pipeline also implements its own set of QC measures. On top of these pipelines IWG6 runs its own Quality Control pipelines for overall QC checks.

The most up-to-date memberships of IWGs are always available on the 4MOST web pages (https://www.4most.eu/qmostauth/team/).

10.1 Galactic Pipeline (4GP)

10.1.1 The Galactic Pipeline (4GP) data production

The requirements for the 4GP are specified in the Galactic Pipeline (4GP) Requirements Specification document [RD11]. L2 data-products from 4GP will include radial velocities and stellar parameters (effective temperature, log *g*, etc.) as well as abundances of selected elements. The requirements represent the minimal acceptable capability needed to achieve the scientific objectives of 4MOST. In a number of cases goals are specified as well, that indicate the level up to which the scientific applications identified for 4MOST can benefit from a better capability. It is, therefore, desirable that such goals are met but the corresponding efforts should not significantly impact the cost or schedule.

The requirements (and goals) are split into precision and accuracy. Precision is method dependent and as 4GP will use more than one method to derive the data products the precision way precision is assessed will be documented for each data release (be it internal or external).

The accuracy will be assessed against extant data for well understood stars. Such comparisons provide relative accuracy as for some L2 data products derived by 4GP no "ground truth" exists.

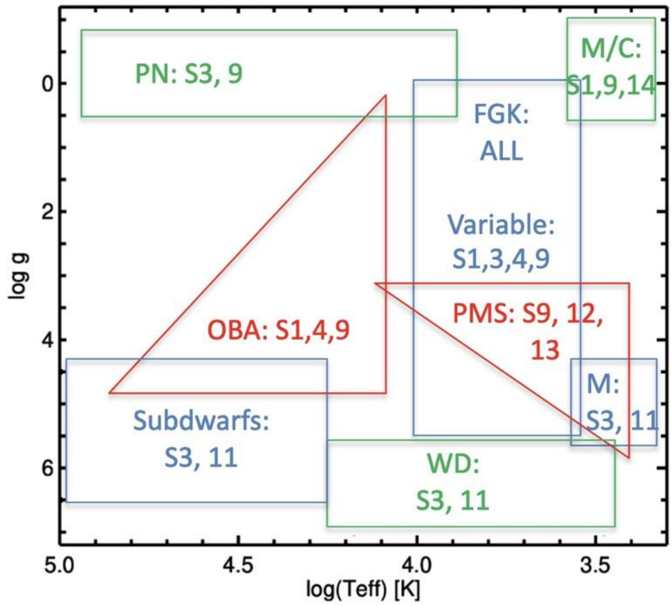

Figure 9: Stellar types observed by the various 4MOST Surveys.

The 4GP is required to deal with a large number of very different types of stars. A list of the minimal set of types can be found in Table 4 with a graphical representation in Figure 9. Such a variety of stars cannot be analysed by a single algorithm nor can the same set of parameters be obtained for all stellar types and with the same accuracy and precision. The 4GP requirements document has therefore a significant number of Appendices tabulating the accuracy and precision (and goals) to be obtained for the different parameters as function of stellar type and for LRS and HRS separately. A summary is provided in Table 5.

Table 4: Evolutionary and stellar types to be handled by the 4GP.

Type	T_{eff} [K]	Log g	[M/H]	Surveys
MS	3000 – 7000	4.0 – 5.5	> -2.0	1,2,3,4,11,12,13
TO	4500 – 7000	3.5 – 4.2	> -2.0	1,2,3,4,13
RGB	3500 – 5500	0.0 – 3.8	> -2.0	1,2,3,4,9,13,14
AGB	2500 – 3500	-1.0 – 0.5	> -2.0	1,2,3,4,9,13,14
HB/RC	4000 – 6500	2.0 – 2.8	> -2.0	1,2,3,4,9,13,14
OBA	10000 – 55000	log g >= -15.027 + 4*log T_{eff}	--	1,2,3,4
WR	30000 – 220000	--	--	1,2,3,4
WD	3000 – 60000	5.5 – 9.5	--	1,2,3,4,11
Subdwarfs	15000 - 100000	4.5 – 6.5	--	1,3
Cepheids	5500 – 6500	1.0 – 3.0	--	
C stars				9,13
Metal-poor			< -2.0	1,2,13

Table 5: Summary of precision requirements on IWG7 (note that precision on elemental abundances depends on nucleosynthetic channel, and the metallicity regime, and not all of the surveys require individual abundances. More details in the Galactic Pipeline (4GP) Requirements Specification [RD11].

Type of target	RV [km/s]	Teff (%)	Logg (dex)	[Fe/H] dex	[X/Fe] (N)	[X/Fe] (dex)
FGK HR	1	1	0.06	0.03-0.05	15-20	0.05-0.15
FGK LR	1-2	2	0.1	0.05-0.07	10	0.1-0.2
WD	5	1	0.04			
OBA	5	5-10	0.1-0.3			
Cepheids & RRLyr		1	0.2	0.05-0.1	α, Iron-peak	0.1-0.3
Hot Subdwarfs	5	2	0.1			

Figure 10 gives an overall visualisation of 4GP work and data flow. 4GP contains many modules performing the analysis and the infrastructure to manage the data-flow. The main characteristics and the used tools are described in the next subsections.

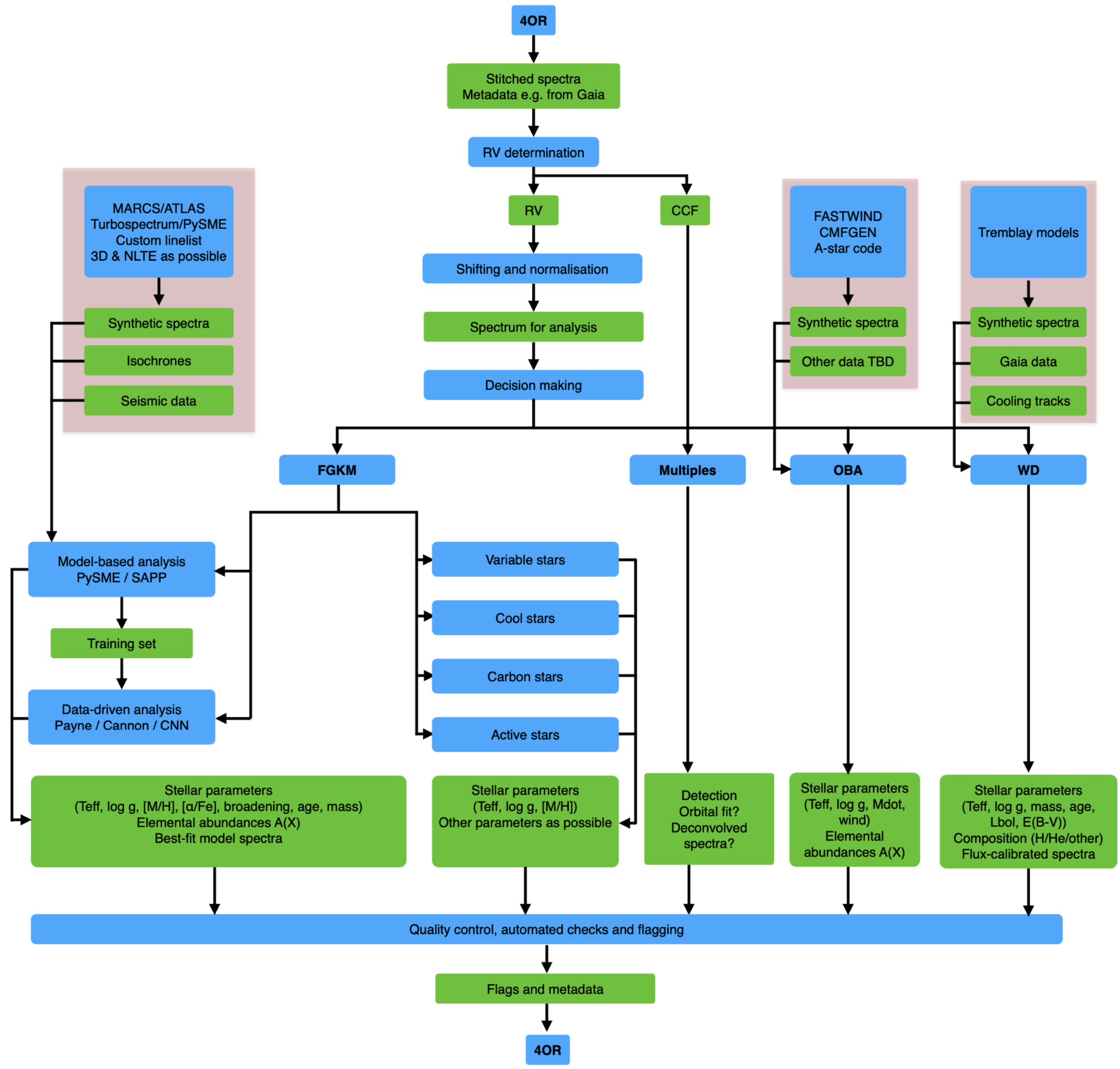

Figure 10: Chart summarising the structure of 4GP and the different codes involved. The blue boxes highlight parts of 4GP involving code, whereas the green boxes contain input or output data. Red shaded sections take place outside operations (i.e. as preparatory work).

10.1.1.1 4GP Skeleton

This is the part of 4GP that provides the input/output functionality and distribution of work via tasks per spectrum. The CPU time that the skeleton code consumes is negligible compared to the analysis modules. The skeleton is divided into two types of logical units:

- one MASTER node which contains: ingestion of L1 files, output of L2 results, internal database storage for intermediate results and bookkeeping, analysis module selection, communication with worker nodes.
- multiple worker nodes, each of which contains all of the analysis modules.

The master and worker nodes run inside Docker containers on a single machine in the 4MOST Data Processing Center (4DPC) at CASU, IoA.

Current Status: Mostly complete. Will need minor updates for new DXUs etc. Code to extract auxiliary information from input catalogues / flags and determine pipeline tasks needs to be written.

10.1.1.2 Radial velocity and V_{rot} module

The radial velocity (RV) submodule is designed to determine the radial velocity in a heliocentric reference frame with a precision of 2 km/s for low-resolution spectra and 1 km/s for high-resolution spectra.

This module consists of various template spectra, including FGKM and OBA type stars, in low and high resolutions. The input spectra are corrected for the blaze function, and both the input (observed) spectra and the templates are scaled to have a mean of 0 and standard deviation of 1. Then cross-correlation functions (CCFs) between the input (observed) spectra and each template are calculated. The peak position of each CCF is extracted by fitting a few pixels around the peak, and then translated to the RV scale. The final radial velocity is the weighted mean of all the RVs, with weights determined by the maximum value of each CCF. The CCF of the best-fitting template is retained for use in the multiplicity module.

The templates and code have recently been updated to cover a wider range of stellar parameters, as well as improving the computational time. The templates now cover the ranges of: T_{eff}=[3000, 7500] K, log *g*=[0, 5] and [Fe/H]=[-3, 0]. The whole process of RV determination takes ~1.2s with 663 templates, similar to the old version which only contained 75 templates. The performance of the module is derived by applying the RV code to a set of random-sampled 4mostified (through 4FS) synthetic spectra: Though the standard deviation is a bit larger than previously reported, it is smaller than the required RV precision for 4MOST Galactic surveys. Further, the RV code can also provide constraints on stellar parameters (i.e., Teff , logg and [Fe/H]), with a precision of ~270K in T_{eff} , 0.75 in logg and 0.53 in [Fe/H]. We have not found a significant RV performance degradation in current stellar parameter ranges.

A few improvements of the module are still planned, including 1) check and decrease the number of spectra giving exceptionally large RV, 2) test the performance for spectra with large Vsin*i*, 3) provide a more precise estimate on stellar parameters if necessary and 4) test the performance of high T_{eff} (OBA) templates.

Current Status: The part of the code that determines the RV has been ingested in 4GP. Nothing is implemented for V_{rot} (though a code is being developed). RV is determined within spec for the majority of targets in both LRS and HRS. Failures are detected while analysing OpR2.5 spectra that require further investigation (10% for HRS; more for LRS). RV determination for WDs is carried out with a separate code that works to requirements. For OBA stars, updated templates and testing are in progress. For cool (M), active and variable stars no specific templates have yet been implemented but in principle the current routine is expected to work.

10.1.1.3 Normalisation module

To create normalised spectra, having corrected for the blaze function we fit low-order polynomials, or splines, to the upper envelope of the stellar spectra. It is unclear whether a more sophisticated approach is required, and this will highly depend on the quality of the L1 data.

Current status: A code that can continuum-normalise the spectra exists, which can either work blindly (i.e. without having any idea of the parameters of the stars) or with a template (assuming a rough estimation of the parameters, coming for example from the RV code). The development of the module has been paused due to issues with OPR 2 data, which showed flux wiggles due to the extraction routine. Further development will take place using OpR 3.5 spectra, or during commissioning if OpR 3.5 does not take place.

10.1.1.4 Multiplicity module

This module is designed to detect and characterise spectroscopic binaries (SB) with two or more radial velocity (RV) components (SBn, n≥2).

The first submodule is the implementation of the DOE code (Detection Of Extrema, Merle et al. 2017) which takes a cross-correlation function (CCF) of one spectrum with a template as input, and provides the number of detected RV components, individual RVs and formal errors and broadening parameters, including $v_{\rm rot}$. To automatically assess the multiple or single nature of the CCF, the three successive derivatives of the CCF are computed. The ascending zeros of the third derivative determine the number of components, provided that the local minima of the second derivatives are above a given threshold. This implementation allows detecting SB with small RV differences, and specifically for CCFs that do not show local minima. Since the CCF is provided from the RV code it is anticipated that the code will be fast.

The module for the characterisation of the individual atmospheric parameters of SBs will be based on multi-component fits using template spectra generated by The Payne but will not be available for DR1.

For single-lined binaries (SB1s) with multiple observations where the RV is large enough that the lines move appreciably between exposures, we will provide custom-stacked spectra. The details of how these will be computed and ingested into the IWG data flow are still being developed but this will only become relevant later in the project.

Current Status: DOE is being developed and ingestion into 4GP and testing is underway. DOE requires an input cross correlation function (CCF) and outputs the number of stellar components identified, RVs and errors from (multi-) Gaussian fits, and broadening estimations. Flux ratios could also be provided. DOE was tested against a non-4MOST CCF, hence unclear yet if requirements are met.

10.1.1.5 FGK modules: Payne, Cannon, CNN, SAPP, PySME

Five codes are still being considered for FGK stars: The Cannon, The Payne, CNN, SAPP, and PySME. The first three can be trained either on real-data or on pre-computed templates. SAPP and PySME determine parameters from first principles (physics driven codes). In particular, we expect PySME to be used, amongst others, to determine the parameters of the training sets of data-driven codes. Comparisons between the codes (performance as a function of S/N and stellar type) are underway. The Cannon and the Payne are already ingested in 4GP and can handle OpR data. SAPP has been implemented but requires testing. CNN has been run on mock 4MOST spectra but not yet integrated into 4GP. All of these codes are fast and can analyse a full night's worth of spectra in a couple of hours (requirement < 8 hours for one night worth of data). Not all elements have been implemented yet, and specific focus needs to be made for metal-poor stars. Performance testing with synthetic spectra degraded using 4FS suggests that performance is approaching requirements for some but not all elemental abundances. Optimisation of the training sets is ongoing. Tests on OpR 2.5 data are less good but this is thought to be largely due to spectral artefacts which have been eliminated in pipeline development. There may be some extra work to implement additional abundances. Implementation of PySME has been delayed due to challenges porting code from IDL, and decisions need to be taken to focus its scope, as it is relatively slow.

Current Status: Even though a lot of testing and code implementation has taken place, currently it is not clear which code/method (or codes/methods) will be used for DR0. We foresee that multiple analysis codes will be run and a decision on which parameters to release will be made by the Release Manager.

10.1.1.6 White Dwarfs (WD) module

This submodule will handle the spectra of all objects which have been targeted as WD (part of the S3 survey). WDs will only be targeted as low-resolution targets so this pipeline only deals with LR spectra. The required inputs for the pipeline (aside from the spectrum) are:

- A flag indicating that the object was targeted as a WD
- The target ID from the input catalogue.
- The Gaia G-band magnitude of the target.
- The Gaia parallax of the target.

The pipeline produces 4 outputs:

1. A WD spectral sub-classification into classes: DA, DAB, DAZ, DAB, DZA, DB, DBA, DBAZ, DABZ, DZB, DC, DO, DZ, DQ - For all spectra.
2. Effective temperature (Teff) - For spectra classified as DA, DAB, DB, DBA, DAZ.
3. Surface gravity (log (g)) - For spectra classified as DA, DAB, DB, DBA, DAZ.
4. Hydrogen abundance (log(H/He)) - For spectra classified as DAB, DB, DBA.

The considered parameter ranges for DA, DAZ are: Teff range: 5.000 K- 140.000 K; log (g) range 6.0 - 9.5.

The considered parameter ranges for DB, DBA, DAB are: T_{eff} range: 9.000 - 40.000 K; log (*g*) range: 6.0 - 9.5; H/He range: 0 - 10^{-1}.

The WD pipeline is divided in 2 main steps:

1. A scikit-learn based, machine learning algorithm that classifies the WD spectra in the different spectral subclasses. This step uses a decision tree classifier trained on ~30.000 SDSS spectra (previously classified by human inspection) to distinguish between WD spectral subclasses. The supervised algorithm looks for presence and prevalence of specific spectral lines. Only spectra classified as DA, DAB, DAZ, DB, DBA (expected to be 80%-90% of all

WD spectra) are further analysed in the second step. The only output for all other spectral types is this classification.

A version of this algorithm is already implemented in 4GP and it was tested to be ~85% reliable in classifying all WD sub-classes. Specifically for the 5 sub-classes that are further analysed the algorithm is up to 95% reliable. The remaining misclassified spectra consists mostly of instances where weak and/or noisy features are not recognised or, in rarer cases, where artefacts are erroneously recognised as spectral features. Spectra which are, erroneously, not recognized as DA, DAB, DAZ, DB or DBA cannot be further analysed in the second step .The modelling of such spectra with noisy and/or weak features would, anyway, result in unreliable stellar parameters.

An updated version of the algorithm is currently being developed and will be able to distinguish WD binaries (necessary for the inclusion of the community survey S11) and will improve the overall reliability of the classifier.

2. In the second step T_{eff}, and log(*g*) is estimated for spectra classified as DA and DAZ; and Teff, log(g) and log(H/He) is estimated for spectra classified as DB, DAB, and DBA.

The spectra to be analysed are firstly continuum-normalised and flattened. This is done via a spline fit to the continuum, using predetermined anchor points placed in spectral regions known to be featureless. Two different sets of anchor points are used for white dwarfs classified as DA and DAZ, and for those classified as DB, DAB, and DBA.

The H and/or He absorption lines are then cropped and compared to appropriate models. The models are stored as pre-computed grids where each grid point corresponds to a specific Teff-log(g)-H/He combination. The models in the grid are interpolated to produce new intermediate models corresponding to any parameter combination between grid points. Chi squared minimization routines within the scipy modules are then used to find the best-fitting model to the observed spectral lines and so estimate the best-fit parameters and corresponding statistical uncertainties. Due to degeneracies in the line profiles this fitting procedure can produce two solutions with approximately equally small chi2. Therefore, the fitting needs to be repeated twice for each object creating 2 best fitting models and so 2 sets of solutions.

The true “best” solution is then identified by calculating a synthetic Gaia G-band magnitude from both models (scaled using the Gaia parallax of the source) and comparing it with the observed Gaia magnitude. The synthetic magnitude closer to the observed one is taken to correspond to the true “best” solution.

The version of the pipeline currently implemented in 4GP can only determine T_{eff} and log(*g*) for spectra classified as DA and DAZ.

A new version is currently being developed which will include the appropriate models for spectra classified as DB, DAB and DBA and a dedicated fitting routine. This new version will also not use interpolation to generate intermediate models from the grid, but it will instead rely on a principle component analysis approach which will be faster and more reliable. Finally, the new version will use pre-recorded Teff and log(g) values based on Gaia photometry as priors in the fitting procedure. This will break the degeneracy in solutions described above and will remove the need to produce two solutions for each target and to calculate synthetic magnitudes.

Performance:

In OpR2 the version of the pipeline currently implemented in 4GP performed at an average of 30s per star.

The planned upgrades to the code are not expected to negatively impact the performance and should, instead, speed up the pipeline by a small margin.

Current Status: Ingested in 4GP. Performance has been adequate in tests so far although development continues.

10.1.1.7 OBA module

This module employs a least squares method and calculates the maximum likelihood function for a grid of stellar atmospheres for all input spectra on the basis of the entire spectral range provided by HRS/LRS including the observational error (error spectrum from L1). In addition, stellar models are de-idealised to compensate for assumptions/missing physics, which is also included in the derived parameter uncertainties. The model error is calculated using the information provided by the observational data. Therefore, all sources must be analysed in parallel rather than star by star. This is an iterative approach until the error matrix is sufficiently converged.

Pre-processing: As all observational data are analysed at once the pipeline requires that all spectra are read at the beginning to determine iteratively the stellar parameters and model uncertainties. For a representative and meaningful model error similar objects should be grouped together (still needs to be implemented). After a spectrum is loaded it is corrected for the potential radial velocity shift, transformed to the wavelength grid of the synthetic spectra grid and decomposed into principal components using the decomposition matrix based on the stellar atmosphere grid. Grid of stellar atmosphere: The grid of synthetic model spectra has been computed with the non-LTE stellar atmosphere and radiative transfer code FASTWIND v10.6 (e.g. Puls et al. 2005). The spectra are pre-convolved with combinations of varying broadening parameters of projected rotational velocity and macro-turbulent velocity and then decomposed to reduce the size.

Deliverables: Bolometric luminosities using photometric data and distances and calculating reddening and extinction parameters, effective temperatures, surface gravities, wind strengths and approximated chemical abundances of He and CNO abundances as well as estimates of line broadening parameters like projected rotational and micro/macro-turbulent velocities.

Current Status: Ingested in 4GP. The pipeline is able to analyse a few thousand massive stars in less than half a day, which should be sufficient to analyse all massive stars observed in a single night. Effective temperatures are derived to specification, but uncertainties of logg, mass-loss rates and Helium abundances are too large due to degeneracy between those parameters; it is unlikely that this problem will be solved before operations start. Determination of the broadening parameters (v sin i, v mac & vmic) are still rough. Additional support is required to test and optimise the pipeline, determination of broadening parameters and computation of the stellar atmosphere grid for Milky Way metallicity.

10.1.1.8 RR Lyrae module

This module processes data from previously identified RR Lyrae stars observed by the S9 Magellanic Clouds (1001MC) Survey. It is designed to estimate the metallicity from low-resolution (R~2000) spectra, i.e. lower than the 4MOST spectral resolution. It utilizes the ΔS method originally presented by Preston (1959) and later developed by, for instance Layden (1994) and Crestani et al. (2021). The ΔS method uses broad spectra features (Balmer lines Hβ, Hγ, and Hδ) and Ca II K line to estimate the metallicity of a given RR Lyrae star. The normalization and measurements of the pseudo-equivalent widths (pW) of the lines (mentioned above) are done over a predefined wavelength range (set by Crestani et al., 2021). The resulting metallicity is estimated through empirical relations calibrated by Crestani et al. (2021) using high-resolution data.

As an input, the module takes 4MOST low-resolution spectra (R ≈ 4000 – 7500) at the rest-frame. The module then normalizes the spectrum and estimates the pWs of Ca II K, Hβ, Hγ, and Hδ. To estimate uncertainties in individual pWs, we implemented the Monte Carlo error simulations where we vary the flux within its uncertainties. The resulting metallicity is then estimated from the distributions of values for pWs using empirical relations from Crestani et al. (2021, see their Equation 1 and Table 8 for reference).

The code, as of now, can process 10 000 stars in under 80 minutes (~0.5 seconds ~ 1 star) on an eight-core machine. The method's accuracy varies between 0.1-0.3(0.4) dex depending on the quality of the spectra. The precision fluctuates around 0.1 dex, when compared with a different implementation of the same method.

The pipeline outputs: metallicity, pWs of Ca II K, Hβ, Hγ, and Hδ, and all associated uncertainties.

Current Status: in the process of ingesting the module in the 4GP pipeline.

10.1.1.9 Cool stars module

This module determines fundamental parameters of cool stars (emphasis on M-type stars, but not limited to that spectral type). It was originally conceived for stars in clusters, that is, chemically homogeneous sets of spectra. As such, no individual abundance determinations are done, only a general 'metallicity' [M/H]. Under these assumptions, the module aims at high-precision determinations of Teff and log *g*, over the ranges: 2500 K < T_{eff} < 8000 K, 0 < log *g* < 5, -3 < [Fe/H] < +0.5. With young star clusters in mind, the module has provision for dealing with rotating stars, with *v* sin*i* up to 100 km/s. The *v* sin*i* of each target star is an input parameter (it might be eventually determined internally by the module, if no better option is available). RV is an additional input parameter if spectra are not in their rest-frame. Formal statistical errors will be provided for each output parameter (T_{eff}, log *g*, [Fe/H]), while "final" errors including systematic effects will only be available after analysis of performance verification data (instrumental signatures etc.).

The module is based on an extended set of spectral indices developed after extensive experience with Gaia-ESO data (on a smaller wavelength range than 4MOST LRS), see Damiani+ (2014). The new procedure employed here will be published in a forthcoming paper. The 4MOST M-star module was originally written in R language, and developed/calibrated using a set of spectral templates obtained using Kurucz/Castelli model atmospheres and the SPECTRUM v2.76 program by R.O Gray (with linelists also available at https://www.appstate.edu/~grayro/spectrum/spectrum.html). The module was recently translated from R to a stand-alone Python code (at the moment, without the ability to deal with rotating stars), and is now being integrated in an in-development M-star branch of the 4GP pipeline. Tests of the Python pipeline made using the same set of simulated spectra as the original R code, in the 3500-8000K range, gave almost identical results to this latter. For S/N=100/pixel we recovered the input parameters within:

- std.dev.: 40 K in T_{eff}, 0.09 dex in log *g*, and 0.1 dex in [Fe/H] (non-rotating stars and input Vsin *i* = 0).
- mean(input-output): -4 K in Teff , -0.009 dex in log g, and -0.004 dex in [Fe/H].

Current Status: When applying the same pipeline to current IWG7 templates based on MARCS and Turbospectrum with up-to-date line lists, results are not good. Therefore, the M-star module is currently being re-calibrated using new templates (estimated timescales of several weeks). The current M-star module implementation is applicable to 4MOST LRS spectra only, but we plan to extend to HRS after the new calibration template grid will be in place. The Python version of the code processes one LRS spectrum in about 0.5 sec, but it can probably do better after optimization, since the original R version analysed one spectrum in about 0.05 sec. The output precision may also be improved (according to tests already made) if sub-pipelines are used (e.g. separately for solar-like and sub-solar metallicities, or for high-gravity and low-gravity stars), provided that a pre-classification of the input spectra is available (e.g. from the RV code).

10.1.1.10 Active stars

The activity submodule will provide calculations of 7 activity related indices, with all 7 returned for LR spectra, and 4 of the 7 returned for the HR spectra.

The indices provided will be as follows:

- Hα excess equivalent width (HR and LR), measured from spectral subtraction of inactive library templates. The residual is then integrated using scipy.integrate.trapezoid. Currently this utilises a fixed bandwidth of 6556.8 Å 6568.8 Å, but will be updated to take account of vsini broadening.
- Hα core flux ratio (HR and LR), measured using the mean on and off-band flux ratios in the core of the H line. The core region is defined as 6560.8 Å – 6864.8 Å, and continuum bands are defined as 10 Å wide windows centred around 6537.8 Å and 6587.8 Å. The index is defined as follows:
  $$\text{alpha\_c} = \bar{f}_{core}/((\bar{f}_{(cont,1)} + \bar{f}_{(cont,2)})/2) ,$$
  where cont,1 and cont,2 refer to the first and second continuum bands respectively.
- Hα wings flux ratio (HR and LR), measured using the mean on and off-band flux ratios in the wings of the H line. The wings are defined as 4 Å windows centred at 6558.8 Å and 6566.8 Å, the same continuum bands as in core are used. The index is defined similarly to core as follows:
  $$\text{alpha\_w} = (\bar{f}_{(wing,1)} + \bar{f}_{(wing,2)})/(\bar{f}_{(cont,1)} + \bar{f}_{(cont,2)}) ,$$

where wing,1 and wing,2 refer to the first and second wing bands respectively.

- A pseudo-S index based on Mt Wilson CaII H&K indices (HR and LR), measured using a triangular weighting function around the H & K lines and taking the ratio to the mean continuum flux either side of the lines. In the high-res spectrum, only the longer-wavelength continuum band is within the spectrum band. The weighted average H & K fluxes are taken by applying a triangular weight function 1.09 Å either side of the cores of each line. The continuum bands are defined as 20 Å wide windows centred on 3801 Å and 4101 Å respectively. The index is defined as follows:

  $$pseudo\_s = (\tilde{f}_h + \tilde{f}_k)/(\bar{f}_{cont,1} + \bar{f}_{cont,2}),$$

  where $\tilde{f}$ refers to the triangular weighted mean flux.
- A similar flux ratio index in the Ca infrared triplet lines at 8498 Å, 8542 Å, and 8662 Å (LR only). This is measured similarly to the pseudo-S index, however, it currently takes only the mean flux values 1.09Å either side of each line, with no weight function applied. The continuum bands are defined as 20 Å windows centered on 8450 Å and 8700 Å. The index is defined as follows:

  $$r\_IRT = (\bar{f}_{8498} + \bar{f}_{8542} + \bar{f}_{8662})/(1.5(\bar{f}_{cont,1} + \bar{f}_{cont,2}))$$
- An excess equivalent width in the IRT lines is given (LR only), measured similarly to the H excess equivalent width, currently using the RV submodule's templates. The bands used are 1.5 Å either side of the central line, as in the Gaia index (Lanzafame, A. C., Brugaletta, E., et al. 2022, arXiv:2206.05766).
- Finally, a pseudo Gaia- index is defined as follows using the same IRT bands as the excess equivalent width above (LR only):

  $$alpha\_IRT = \int_{-\lambda}^{+\lambda} \left(\frac{r_\lambda^{obs}}{r_\lambda^{temp}} - 1\right) d\lambda,$$

  where r_λ^{obs} and r_λ^{temp} refer to the continuum-normalised fluxes in the observed and template spectra respectively.

Current Status: Currently all indices, apart from the IRT, are producing results from test spectra. The IRT code is being developed, and initial development is underway to implement uncertainty calculations for each of the indices. Further work is needed to confirm the best selection of template spectra, and the addition of *v* sin*i* broadening in the equivalent width indices to ensure a reasonable wavelength band is chosen for the indices.

10.1.1.11 Ages/Masses module

No ages module has been implemented in 4GP yet. However, such a code exists, based on Kordopatis et al. (2023) or Sahlhodt & Lindegren (2021). This module is expected to take into account the spectroscopic T_{eff}, log *g*, [Fe/H], and as many other magnitudes desired, and project them on isochrones. The K23 code is fast, and can achieve precision of the order of 20 per cent or better for turn-off stars, provided small uncertainties in the input parameters.

The SAPP also provides ages and masses, and its estimations may be added or be complementary to the ones derived from isochrone fitting.

Current status: Code exists, but has not yet been integrated into 4GP.

10.1.2 Input physics and data

4.1.1.1 Atomic physics and non-LTE departure coefficients

The principal input physics for the pipeline are atomic physics of transitions: term identifications, wavelengths, oscillator strengths and information about hyperfine splitting and isotopic effects. The pipeline maintains a single central linelist which is based on an extraction from the Vienna Atomic Line Database (VALD). The linelist is being manually curated element-by-element to identify the most reliable and up-to-date source of atomic data for each transition, following a similar procedure to that adopted for the Gaia-ESO survey but extending to the larger wavelength range of 4MOST and including literature published since the Gaia-ESO linelist was developed. The resulting linelist is being used for all FGKM template spectra generation and other analyses.

4.1.1.2 Model template production

We will rely on pre-computed grids of model atmospheres, i.e. new models will not be generated at run-time. For 1D model atmospheres of late-type stars we use MARCS, which is developed and maintained in Uppsala and has the advantage of using spherical symmetry instead of the plane-parallel assumption for giants. The 3D STAGGER model grid covers a significant portion of the FGK star parameter space, but needs to be extended at low-surface gravities to be well suitable for 4MOST analysis; we intend to use STAGGER for later data releases. All available model grids for FGK type stars assume LTE. As the default for most OB stars we will use TLUSTY, an extensive grid of non-LTE models without winds. For stars with strong winds e.g. Wolf-Rayet stars, we will use FASTWIND and CMFGEN. For white dwarfs we will use the 1D grids of Koester (2009) and 3D COBOLD models as available.

The analysis will be carried out using spectra calculated without using the assumption of local thermodynamic equilibrium (LTE) to the maximum extent possible, by using grids of departure coefficients (NLTE/LTE population ratios), when synthesising synthetic spectra. Two groups within 4GP that develop departure coefficients, to be used with two different spectral synthesis codes (Turbospec & PySME). For Turbospectrum, the elements that are considered are H, O, Na, Mg, Si, Ca, Ti, Mn, Fe, Co, Ni, Sr, Ba; Y and Eu are foreseen to be ready by start of operations. For PySME, the elements are H, Li, C, N, O, Na, Mg, Al, Si, K, Ca, Mn, Fe, Sr, Ba (with Ti and Cu potentially available by start of operations). A taskforce including people from the two groups is currently combining and homogenising the available departure coefficients and computing grids of non-LTE synthetic spectra for test and verification, and ultimate inclusion in the analysis pipeline.

10.1.3 Galactic Pipeline (4GP) Quality Control

Many modules within 4GP contain internal quality checks; e.g. whether the quality of a spectral fit is as good as expected. These quality flags will be returned by the modules as part of the output data, as specified in the DXUs. Initially all flagged outputs will be subject to a manual review until the pipeline is sufficiently mature that it is clear that the flagging is reliable.

Each module task in 4GP will be followed by an associated per-task QC check, which will perform completeness and bounds checking on the output data. These checks will output flags, which will be attached to the output data as it is passed through 4GP's internal data flow process. Ultimately these flag values will be added to the output FITS tables returned to 4OR. Output data that fails bounds checks will be returned with flags as described rather than not returned.

Finally, per-night and per-OB QC tasks will run before data is returned to 4OR. These will produce a set of diagnostic plots that are expected to have a well-known shape, such as T_{eff}-log*g* diagrams, [X/Fe] vs [Fe/H] diagrams and histograms. These plots will be manually inspected on a per-night cadence, after data has been returned to 4OR, to monitor pipeline stability and performance. If data are found to be clearly problematic it will be possible to manually retract them using standard 4OR procedures. These plots are expected to be extremely valuable for the first weeks of operations, and the necessity of producing them after that will be evaluated accordingly.

Current status: Internal quality checks already exist for all modules that will be run up to and including DR1. Development of the bounds check and per-night QC routines will take place during Q4 2023.

10.1.4 Galactic Pipeline (4GP) Organisational Structure and Staff Effort

The details of the organisational structure and the available staff effort for the 4GP pipeline is described in the IWG7 management plan [RD12].

The organisational structure within IWG7 is shown in Figure 11. The main effort within the working group in the run up to commissioning has been in developing the input physics and pipeline modules and so these areas are well developed. Atomic and molecular data associated wit

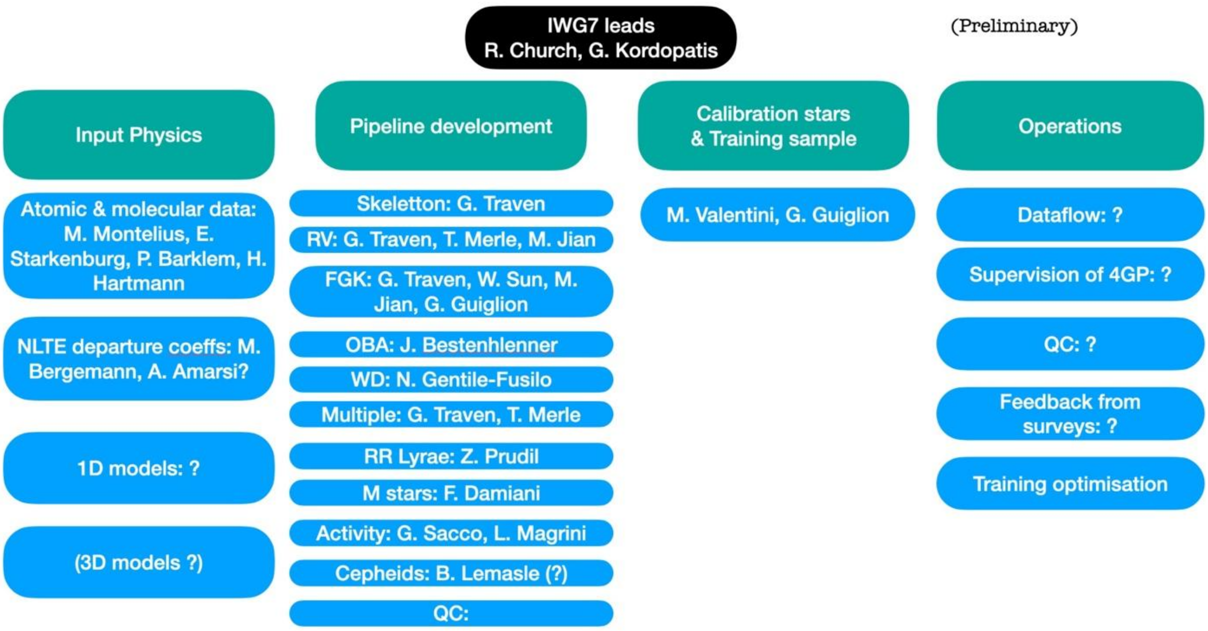

Figure 11: Organisational structure and WP distribution among members in IWG7.

The available IWG7 staff effort is listed in Table 6. We are regularly reviewing the adequacy of the FTEs in this table in conjunction with the surveys.

Table 6: Staff effort available in IWG7.

Name	Institute	Main Contribution to IWG7	FTE to IWG7 until start operations	Survey membership
D. Aguado	IAC	Templates, linelists.	0.1	S1, S13
A. Amarsi	Uppsala University	Non-LTE models	<0.1	S2, S14
S. Antoniucci	INAF-Roma	Accretion from emission lines	<0.1	S12
P. Barklem	UU	Line list, non-LTE models	0.1	S2, S4
J. Bestenlehner	U. Sheffield	OBA module	0.3-0.4	S3, S9
K. Biazzo	INAF - Rome	Accretion from emission lines	<0.1	S12
S. Buder	ANU	Cannon, training set	<0.1	S2, S11
E. Caffau	Observatoire de Paris	Representative for the S2	<0.1	S2
A. Casey	Monash		<0.1	S3, S4
R. Church	Lund University	Co-lead	0.2-0.3	S3, S4
F. Damiani	INAF - Palermo	Cold-star pipeline for fundamental parameters	0.4-0.5	S12, S13
D. Feuillet	Lund University	OpR2, Documentation, ages	0.1	S1, S4, S14
M. Gent	MPIA	SAPP module	0.2-0.3	S4
N. Gentile Fusillo	ESA	WD pipeline	0.2-0.3	S3, S11
G. Guiglion	MPIA	CNN module	0.2-0.3	S1, S2, S3, S4
C. Hansen	Institute for Applied Physics (IAP)	Wavelength coverage definition and early tests.	0.1	S1, S2, S14

Name	Institute	Main Contribution to IWG7	FTE to IWG7 until start operations	Survey membership
M. Jian	Stockholm University	Development of 4GP	>0.6	S13
G. Kordopatis	OCA	IWG7 co-lead	0.2	S1, S2, S3, S4, S9
K. Lind	SU	Previous co-lead, advising and supervising	<0.1	S2, S3, S4, S14
L. Magrini	INAF- Arcetri	analysis of young stars	0.1	S12, S13
J. Maldonado	INAF - Palermo	M dwarf's pipeline.	0.1	S11
S. Monty	Cambridge	Training and testing the 4GP Cannon version	<0.1	S2, S13, S14
B. Nisini	INAF - Roma	Accretion from emission lines	<0.1	S12
T. Nordlander	ANU	Synthetic spectrum grids	0.1	S2, S14
Z. Prudil	ESO Garching	RR Lyrae pipeline	0.1	S9
G. Sacco	INAF - Arcetri	Coordination of the pipeline for activity	0.1	S12
E. Starkenburg	University of Groningen	Linelist	<0.1	S1, S2, S3
W. Sun	AIP	Payne module for FGK	0.2-0.3	S3
S. Taibi	AIP	RV testing	<0.1	S9
T. Merle	ULB/ROB	multiplicity pipeline, templates for composite spectra	0.1	S4
G. Traven	U. Ljubljana	4GP development	0.1	S3, S4
M. Valentini	AIP	Discussion, IWG3 (validation/training) , seismic log(g)	0.1	S1, S2, S3, S4

Name	Institute	Main Contribution to IWG7	FTE to IWG7 until start operations	Survey membership
M. Van der Swaelmen	INAF - Arcetri	Templates	0.1	S9, S12
C. Worley	UC SPCS	Representative for S1	0.1	S1, S4
K. Youakim	SU	Representative for S14	<0.1	S4, S14

10.2 Extragalactic Pipeline (4XP)

10.2.1 Extragalactic Pipeline (4XP) Data Production

The requirements for the 4XP are specified in the Extragalactic Pipeline (4XP) Requirements Specification document [RD13]. L2 data-products from 4GP will include redshifts and galaxy and AGN parameters, including uncertainties and reliability parameters.

10.2.1.1 4XP Data flow and Interfaces Overview

The 4XP internal dataflow is shown in Figure 12. L1 data for one night is queried from the 4OR via the 4DAP interface. The target metadata from the input catalogue is tagged along with the L1 data as it is passed through the pipeline modules. Each block of modules is executed in series one after another, whereas all modules in the same block are executed in parallel. The output of each module is saved to temporary files (marked as L2*) and is available to subsequent blocks 'down-stream', i.e., the output of block 0 is available to block 1 and higher. All temporary L2* products are collected and validated before being ingested back into the 4OR via the 4DAP interface signalling the end of 4XP for that given night.

The content of the different modules is described in the next subsections.

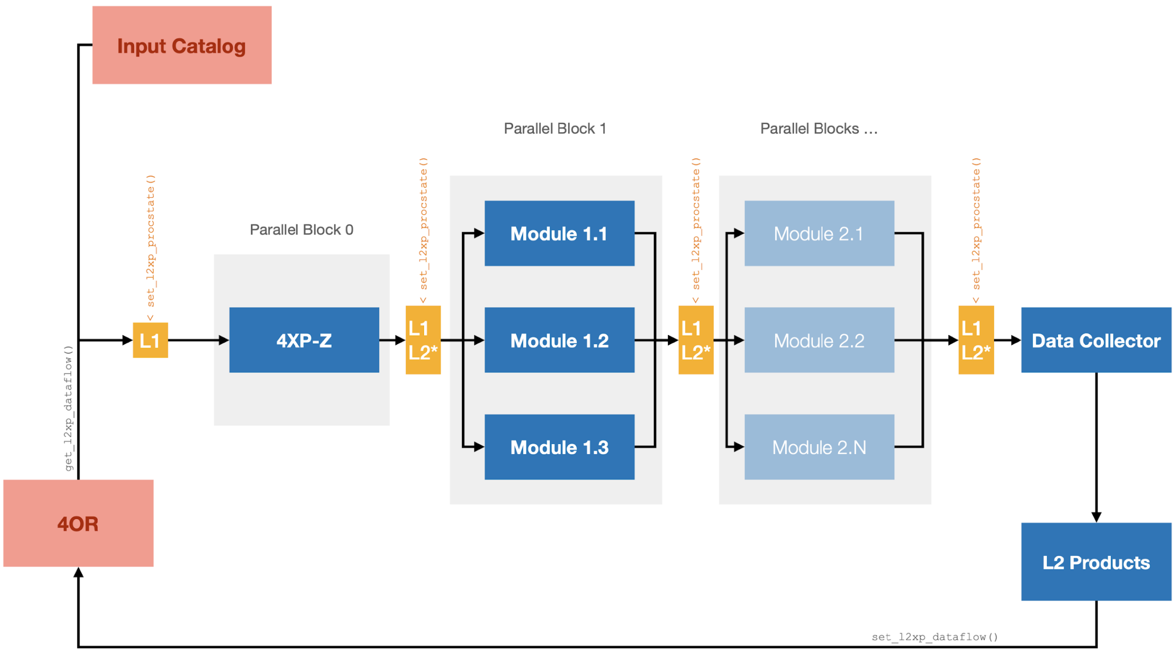

Figure 12: Current top-level outline of 4XP data flow.

10.2.1.2 4XP-Z (Redshifting Algorithm)

This module is the primary component of the IWG8 developed codebase. It will essentially take in individual spectra and super-stacks of the L1 data products and a series of source templates to generate:

Redshifts with $\sigma_z \leq 0.0001\ (1 + z)$ precision for galaxies and $\sigma_z \leq 0.003\ (1 + z)$ precision for AGN/QSOs

- Numeric confidences on these redshifts based on how secure the redshift is deeded to be and descriptions of these confidences such that:
    - Survey teams can easily assess the validity of the redshift
    - Any redshift feedback loop can automatically assess the success/failure of the redshift assignment produced.
- Source type of the best-fit template used in the redshift assignment
- The four next best redshift solutions and confidences for alternate fits.

In addition, the module will also be able to include redshift priors (i.e. from photometric redshifts) in its analysis and weight confidences accordingly. This module will also include wrapper code to read the L1 spectra in the correct format, unit tests (see below) and code to generate output data product in the format defined by the IWG8 DXU.

Previous spectroscopic surveys have utilized a range of approaches to determine spectroscopic redshifts in an automated fashion. These are generally split into two distinct categories: 1) Fourier template fitting using cross-correlation or 2) direct χ^2 template fitting. Both approaches have strengths and weaknesses as detailed below. Since 4MOST will observe a very wide range of objects, we need an algorithm that performs well in all cases. For this purpose, we have designed an algorithm that combines both these approaches, the so-called xPCA algorithm. This is the core algorithm of the 4XP-Z module.

In parallel with this more traditional approach, we are also investigating a machine learning algorithm which would provide a significant improvement in terms of computational cost.

Below we will give a brief presentation of strengths and weaknesses of previous approaches before presenting the new xPCA algorithm. Lastly, we will summarise the machine learning approach.

10.2.1.2.1 Fourier Cross-correlation with Spectral Templates

Examples of codes using the cross-correlation approach are:

- **Auto-z (Baldry et al 2014):** used in the Galaxy and Mass Assembly Survey (GAMA, e.g. Driver et al 2022) and modified version Deep Extragalactic Visible Legacy Survey (DEVILS, Davies et al 2018, 2021).
- **MARZ (Hinton et al 2016):** used in OzDES (Lidman et al 2020)

This methodology essentially applies a pre-processing step to both template spectra and target spectra. The benefit here is that the template pre-processing can be done a-priori, increasing the speed of the fitting. Pre-processing typically takes the form of:

- Scaling to a common flux scale
- Flagging bad pixels of high noise regions
- Continuum subtraction (typically using smooth spline fits)
- Resampling onto an (over-sampled) logarithmically-binned wavelength scale

The Fourier transform on all templates and target spectra are then taken and cross-correlated. The cross-correlation is inverse transformed back into real space and peaks identified in the cross-correlation. These peaks correspond to the shift (in log-wavelength) between the template and target spectrum, where there is a good match in phase (e.g. the redshift). These codes will identify the strongest unique peaks over all templates and find the best-matching template/redshift.

The confidence of any given redshift can then be calculated using the strength of the cross-correlation peak in comparison to a weighted sum of the 2nd, 3rd, 4th… peaks (essentially how unique the solution is).

This method works very well for spectra with narrow emission/absorption features such as galaxies and stars, however, for quasars and AGN the method does not recover the redshifts very accurately due to the very broad features that may easily be filtered out during the pre-processing, or due to the very strong quasar emission lines that introduce strong spurious peaks in the cross-correlation functions leading to incorrect redshift solutions and poorly estimated redshift confidence.

10.2.1.2.2 Direct Fitting with PCA Templates

Examples of codes using approach:

- **Redmonster (Hutchinson et al 2016):** used in SDSS DR14 eBOSS LRG samples (e.g. Bautista et al 2018).
- **Redrock (Bailey & Schlegel 2017):** used for DESI in combination with post-processing codes (so-called afterburners).

These codes all essentially take the similar approach where:

- Target spectra have bad pixels flagged.
- A redshift grid is defined for each template type with logarithmically spaced sampling.
- At each bin in the redshift grid, principal component analysis (PCA) templates are redshifted and interpolated onto the observed wavelength scale of the target spectra.
- The χ^2 is calculated for each bin in the redshift grid for all PCA templates (stars, galaxies, quasars).
- Local optimization is performed on a more finely sampled grid near the minimum χ^2.
- Best-fit redshifts are calculated using the minimum χ^2 over the full parameter space of all templates and redshifts. Further refinement can be obtained by including low-order polynomial functions in the template fitting (e.g., Redmonster).

Redshift confidences are calculated by comparing the χ^2 values across the full parameter space. These codes take much longer to run than the Fourier cross-correlation codes due to the high number of interpolations required together with a large number of redshift samples in order not to skip over narrow minima in χ^2 space (typically of the order 10^4 grid samples per template type for extragalactic templates). The direct fitting methods perform better than the cross-correlation methods for spectra dominated by continuum features or broad emission lines. For spectra with only narrow emission features, this technique requires very finely sampled redshift grids which increase computation time (linearly with number of samples).

In the case of 4MOST data, the high sampling rate of 0.25Å per spectral bin in the calibrated LR spectra (around 3 times higher than DESI) means that the direct fitting on a brute-force redshift grid becomes computationally very expensive.

10.2.1.2.3 Fourier Cross-Correlation combined with PCA Template Fitting

Given the challenging setup of 4MOST with a large range of different object types observed simultaneously and a large number of pixels per spectrum, we have designed a new approach which combines the two methods of cross-correlation and direct PCA fitting.

The algorithm performs a fast cross-correlation with a set of PCA templates (using only the first component) to identify peaks in the correlation function that correspond to valleys in χ^2 space. We then identify the 5 strongest peaks and only run the PCA fitting on a fine redshift grid around those predetermined tentative solutions. For a given target, the following steps are taken:

- All PCA templates are interpolated onto an oversampled logarithmically spaced wavelength grid.
- The target spectrum is interpolated onto the same log-wavelength grid.
- The cross-correlation function (CCF) between the target and the first template spectrum is calculated using Fast Fourier Transforms.
- The 5 most significant peaks for each template type are identified.
- A logarithmically refined redshift grid is constructed around each CCF peak.
- The PCA fitting, including a low-order Chebyshev polynomial, is performed on the original target spectrum on the linear wavelength grid. This way we do not introduce further correlations in the variance spectrum.
- The minimum χ^2 and the depth of the χ^2 ($\Delta\chi^2$) with respect to the surrounding χ^2 space are calculated around each peak for all template types.
- All redshift solutions are ranked using a combination of the following metrics: min. χ^2, $\Delta\chi^2$, and the significance of the CCF peak.
- The 5 best solutions and their corresponding spectral class are kept, and uncertainties are calculated using the curvature of χ^2 in redshift space.

We calculate the redshift probability using the three metrics, min. χ^2, $\Delta\chi^2$, and the significance of the CCF peak. However, this can only be done after comparing our solutions to a sample of already known redshifts. This is true for all the redshift methods described here. This last step is being done using training data (simulated and empirical) and will be fine-tuned during the Pipeline Validation phase.

10.2.1.2.4 Machine Learning

Examples of machine learning codes being investigated:

- **GaSNet:** Current development code for 4MOST (Fucheng, Nicola)

Very briefly this approach works in the same way to any other machine learning problem:

- Splitting the sample into training, validation and testing sets
- Training the algorithm on the training set with known redshifts to identify key properties of samples at a given redshift.
- Producing an algorithm which can identify common properties in the test sample.
- Using this algorithm to assign redshifts and uncertainties to the test sample.

This approach has currently only been studied using simulated 4MOST data and SDSS spectra. One of the main drawbacks of the machine learning approach is the need for large training data sets. It is therefore not possible to have a fully trained algorithm on real 4MOST data before survey operations but we are planning to test and develop this option in parallel once real data arrive. The main advantage of this algorithm is a significant improvement in computation time.

10.2.1.3 4XP-Fit (Spectral Fitting Algorithm)

This module will take the L1 data products and the redshifts obtained by QXP-Z to derive galaxy properties. The core deliverables of this module are:

- Stellar continuum fits providing best fit properties of the stellar population synthesis models and stellar velocity dispersion.
- Emission and absorption line measurements based on Gaussian line fits to the data. These include line fluxes, EWs and line-widths for a fixed list of lines.

4XP-Fit will be based on the code PyPARADISE (Husemann & Walcher). This code broadly performs the following steps:

- Target spectra have bad pixels and high noise regions flagged.
- Spectral emission line features are masked
- An MCMC fitting approach is applied to fit stellar population synthesis models to the remaining stellar continuum (using the X-shooter Spectral Library, Verro et al. 2022). The fitting includes a flexible low-order polynomial to account for variable spectral slopes.
- The best-fit continuum is then subtracted and the emission lines are fitted using Gaussian profiles.

In order to cross-validate the results from PyPARADISE, we are also performing a comparison to the code pPXF (Cappellari et al. 2022).

10.2.2 Extragalactic Pipeline (4XP) Quality Control

Regular quality control will be handled at two stages within the extragalactic pipeline. The first step happens on a nightly basis as part of the individual pipeline modules. The second step is a long-term monitoring where performance metrics will be verified against known targets (the definition of these fields of known cross-calibration targets are part of IWG3 and will cover targets also observed by SDSS, DESI, and WEAVE).

Nightly QC criteria

Each pipeline module produces a 16-bit integer flag per spectrum, where 0 means everything passed nominally. Each of the 16 bits corresponds to a unique error message if set to 1.

As part of the spectral preprocessing in 4XP-Z, the number of 'bad pixels' is calculated based on the noise properties, i.e., statistical outliers most likely caused by cosmic rays or errors in the L1 pipeline. If the number of bad pixels in a spectrum exceeds 230 (corresponding to 1% of all pixels), a warning flag is raised. Similarly, if the median signal-to-noise per pixel is less than 0.5 a warning flag is raised. (The actual value of this S/N criterion will be determined during the pipeline validation phase.)

During the actual processing of the spectra, the following warnings are raised: If the cross-correlation function does not yield enough peaks; If the PCA fitting does not converge; If the local Chi^2 optimization around cross-correlation peaks does not yield a well-determined minimum; If the redshift uncertainties could not be determined; If more than one redshift solution has similar confidence; If the best solution has high reduced Chi^2; Or if the best solution has less than 3-sigma significance from the cross-correlation function. The last bits (9-15) are currently not used but are kept for future implementations.

The 4XP-Fit module will provide the following QC metrics: a global chi^2 for the continuum fit, and individual chi^2 for the emission lines that are fitted. These individual values will allow us to flag emission lines that show anomalous features, such as asymmetric or broad non-Gaussian wings.

Long-term QC criteria

The warning flags detailed above will be used to monitor the pipeline performance leading on longer time-scales (every 3 months). On top of these diagnostics from the pipeline itself, we will cross check solutions for repeat observations of the same objects within a given 3-month period. Specifically, we will verify that individual observations (if sufficient S/N is reached) and stacked spectra yield consistent results within the uncertainties.

Furthermore, known targets will be used to cross-check that the redshifts are consistently recovered to the required level of precision as a function of signal-to-noise ratio. A subset of these targets will also be part of the pipeline validation phase before survey operations begin.

10.2.3 Extragalactic Pipeline (4XP) Organisational Structure and Staff Effort

The details of the organisational structure and the available staff effort for the 4XP pipeline is described in the IWG8 Management Plan [RD14].

The WBS organisational structure within IWG8 is shown in Table 7 and the staff effort contributions of the IWG8 members is tabulated in Table 8.

Table 7: Overview of the Work Packages and their leads in IWG8.

WP	Name	Lead	FTE (Total person yrs)	Time	Resources & Coordination
0	Management	JKK, LD, SC	0.7	2022-2029	ODG
1	Redshifting Algorithm	LD & JKK	1.0	2022-2029	-
2	Galaxy Spectral Fitting Algorithm	GC	1.5	2022-2029	-
3	Definition of Atomic Line Lists	MS	0.1	2016	-
4	L1 Simulation Tool	LD, IC	0.5	2022-2023	IWG7, IWG9, QMOST-ETC
5	Pipeline Validation	SMC, JKK	1.5	2022-2024	IWG3
6	Quality Control / Visual Inspection	IC	1.5	2024-2029	IWG6, 4PA
7	Documentation / DXU	JKK	1.0	2022-2029	-
8	Redshift Feedback Loop	LD	1.0	2022-2029	OpSys, DMS
9	Operations / Maintenance	JKK	1.0	2024-2029	OpSys, DMS, Helpdesk

Table 8: IWG8 staff commitments for the preparation and operations phases. The WPs are 1: Redshifting Algorithm; 2: Galaxy Spectral Fitting Algorithm; 3: Definition of Atomic Line Lists; 4: L1 Simulation Tool; 5: Pipeline Validation; 6: Quality Control and Visual Inspection; 7: Documentation / DXU; 8: Redshift Feedback Loop; 9: Operations / Maintenance.

Member	Work Package	2023	2024	2025	2026	2027	2028	2029
JK Krogager	0, 1, 7, 9	0.3	0.5	0.5	0.3	0.3	0.3	0.3
Luke Davies	0, 1, 4, 8	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Scott Croom	0, 1, 7	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Guilherme Couto	2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Bianca Poggianti	2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Giuseppe D'Ago	1, 2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Young-Lo Kim	1, 2	0.2	0.2	0	0	0	0	0
Matthias Kluge	2	0.1	0.1	0.1	0	0	0	0
Stijn Wuyts	1, 2	0.1	0.1	0.1	0.1	0	0	0
Don Terndrup	2	0.1	0.2	0.2	0	0	0	0
Marcella Longhetti	2, 6, 7	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Alireza Molaeinezhad	1, 2, 8	0.1	0.1	0.1	0	0	0	0
Carolin Villforth	1, 2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Igor Chilingarian	4	0.15	0.1	0	0	0	0	0
Kirill Grishin	4	0.15	0.1	0	0	0	0	0
Karen Leighly	1, 5	0	0.1	0.1	0.1	0	0	0
TOTAL	1, 2, 4, 6, 7, 8, 9	1.95	2.2	1.9	1.3	1.2	1.1	1.1

10.3 Selection Function Pipeline (4SP)

10.3.1 Selection Function Pipeline (4SP) Data Production

All large scale surveys suffer from selection effects that raise potential issues in the analysis and interpretation of results. Have we observed all there is to observe? Have we omitted observations that are crucial for the scientific questions posed? Many science cases require correction for the survey selection function and the answers to these questions are very important for the interpretation of the observed data. The biases from the selection function come in two main forms: the object selection bias (WP1) and the geometric & observational bias (WP2). The 4SP (4MOST Selection functions Pipeline) will calculate the posterior selection functions or the probability that a target and its physical properties was (or was not) observed/determined given the input catalogues, the target priorities, instrument and observing conditions, extraction software, etc. The selection effects due to the generation of the individual target catalogs are the responsibilities of each individual survey. An overview of the dataflow for 4SP is shown in Fig. 13, with inputs of target catalogs, simulation suites, and observed data sets, while the 4SP module(s) determine the outputs which are the probabilistic and counting-based selection functions.

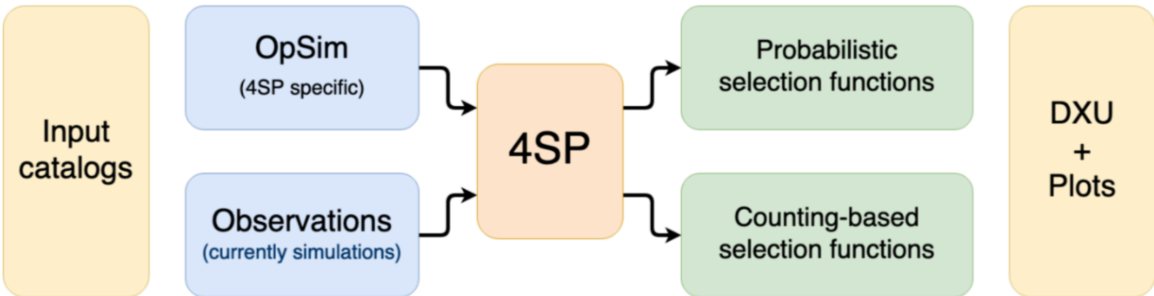

Figure 13 Basic dataflow for 4SP

WP1 – Object Selection Function (OSF)- It is not possible to derive attributes for all observed objects. This can be due to reasons ranging from data quality to instrumental setup and intrinsic object properties and we discuss this in detail later. The inability to derive object attributes for certain objects introduces a selection function which we call the object selection function (OSF). The main function of WP1 is to calculate the probability to obtain physical parameters of an object.

There are various reasons for the existence of the object selection function in 4MOST. Typical sources of incompleteness are the observed spectrum having large noise, the gradient of the spectral flux (with respect to attributes y such as Teff or logg) being weak, the use of data quality flags used to cull unreliable measurements or when the measurement uncertainty is too high, or because most theoretical models cannot make accurate predictions over the entire range of the attribute space.

Characterizing the object selection function is critical. In the initial planning stages of a survey, knowing the object selection function allows us to better optimize the survey strategy to satisfy our main science goals. For example, if certain attributes cannot be measured for a certain type of objects, then we can decide to alter our input catalog selection criteria to either minimize observing such objects (if they are not required for our science) or increase the exposure time so that we can measure the required attributes (if they are required for our science). Once the survey is done, the object selection function is needed in almost any science that one wishes to do using the survey.

The OSF is defined by two main properties. The first is the set of attributes of interest y from the spectrum f and whose completeness we are interested in. The attributes for stars could be surface gravity g, effective temperature Teff, iron abundance [Fe/H] and other elemental abundances like [Ba/Fe], [Mg/Fe] and attributes for galaxies could be redshift z and strength of certain emission lines. The second is the set of variables x that are used to characterize or model the spectrum, for stars these could be g, Teff, [Fe/H] and SNR and for galaxies z and SNR. The OSF is then the probability of measuring y over the space explored by the set of variables x.

The OSF is concerned with attribute measurements which are performed by L2 pipelines. Each measurement made by an L2 pipeline has an OSF associated with it. We have three L2 pipelines: 4GP measuring stellar attributes, 4XP measuring redshift of extragalactic objects, and 4CP that classifies objects into different classes. For each given object type and attribute one different OSF depending upon the measurement method and spectral resolution (high-res vs low-res). In order to determine the OSF:

- each survey will provide a list of attributes y they are interested in exploring the OSF for and the space x over which they will explore it.
- Other working groups (iWG7, iWG8) will provide synthetic spectra for each grid point of space x.

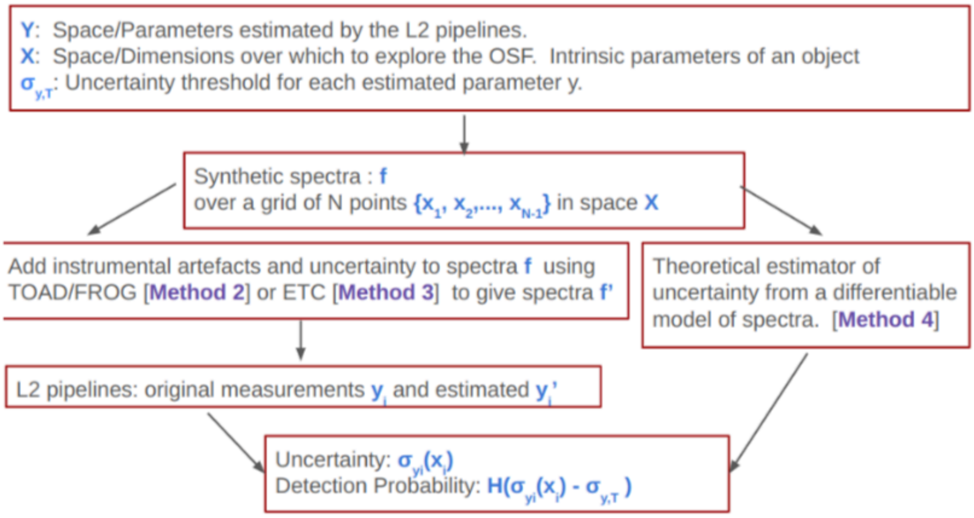

Figure 14 An outline of the approaches for the OSF

Several approaches to estimate the OSF will be used (Fig. 14)- **1)** based on a large number of observed spectra with a range of S/N ratios uncertainties in the derived parameters will be determined using pipeline values, and **2)** by generating a large number of artificial spectra with known properties using the TOAD/FROG instrument simulators and by running the L1 and L2 pipelines a large known parameter space can be covered, **3)** is the same as 2) but using 1D ETC generated spectra which is computationally much faster/cheaper but includes more approximations. **4)** Compute theoretical confidence intervals y (upper and lower) based on the simulated spectra by using the Fisher information. Approaches 2) and 3) are already being used by IWG7 and IWG8 for testing and validation of their L2 pipelines. A straightforward extension of that can be used by IWG4 to estimate the OSF.

The OSF for each target observed with 4MOST will be computed and updated for each data release.

WP2 – Geometric Selection Functions (GSF)- The principal task of WP2 is to track survey completeness as a function of position on the sky, accounting for observational effects such as fibre placement constraints and observing conditions. Note that survey completeness as a function of redshift (radial selection function) is a task left to the individual surveys, as it depends on the luminosity function, visibility limits (due to flux, surface brightness, emission line strength, etc.) of each class of target. Surveys are also responsible for the generation of target catalogues which is the first part of selection functions. To derive the angular selection function, IWG4 will compute the fraction of targets that were successfully observed in each pixel or polygon on the sky.

The effective geometric selection function for 4MOST will be a significantly more complex task than selection functions for previous surveys. Each fiber has a fixed "base" position within the focal plane and a circular patrol area that is limited by some maximum tilt angle. The locus of potential positions where a fiber can be placed is further constrained by the need to avoid collisions with neighboring fibers, which have overlapping patrol circles. The situation is further complicated by several factors:

- **i.** Approximately 1/3 of the fibers feed the high-resolution spectrograph and the remainder feed the low-resolution spectrograph, and the focal plane may contain immobile fiducial fibers, broken fibers etc.
- **ii.** The throughput of individual fibers will vary from fiber to fiber
- **iii.** The throughput of fibers depends on their instantaneous tilt angles
- **iv.** Fibers are assigned to targets in finely graduated steps in object probability, typically with objects from several different surveys targeted within each 4MOST tile
- **v.** Any field on the sky is likely to be visited several times with different focal plane configurations for each tile (and different observing conditions for each visit)
- **vi.** The plate scale varies across the FoV (and possibly as a function of airmass) and hence the effective density of fiber positioners also varies.
- **vii.** Duplicated targets between (sub)surveys.

The net result of all these considerations is that the probability for some target to have been observed is highly dependent on its position on the sky, its location within the focal plane, its requested spectrograph resolution and its place within the object probability hierarchy. The scale on which the selection function varies is of order the separation of fibers (or potentially even smaller when multiple visits to the same field are considered). This work package will account for this myriad of factors and produce the products that describe the completeness of 4MOST surveys as a function of position on the sky, for different object classes. For each data release, there will be an update on these files. There are several sub-tasks for the geometric selection function:

1. Input catalogue definition. ID, RA, DEC and observational priority; survey boundaries; mask (excluded regions).

2. Update survey geometry after each set of observations. Each exposure will generate the following records: a. Observational parameters formatted into a fits file (airmass, cloud coverage, moon phase, distance to the moon, seeing, background brightness). If the parameter is not constant across the field of view, these should be given as a function of a location in the field of view or for each fiber separately. b. Geometry of the observation in a polygon file, holes in the field of view (e.g. broken fibers).

3. Input from databases and pipelines. a. Input A: tiling of the targets - what targets obtained a fiber and what fibers did not. Priorities of all targets in a given field of view. Additional input: fiber position on the optics, fiber tilt angles, spectrograph number. b. Input B: pipeline reduces the data at the pixel level and assigns a likelihood quantifying if the observation be successful in the sense each survey has defined. c. Input C: output from IWG7 and IWG8 pipelines.

4. Calculate completeness as a function of position on the sky. a. Convolution inputs (3.a,b,c) with the 2.a and 2.b of the observation to assign weights to the observations. b. Output: For each survey, a completeness map as a function of RA, DEC, and the observational parameters.

Summary of steps for completeness function calculation/modelling

The completeness in a single field can be estimated when running target allocation algorithms for a given field multiple times. If the targeting algorithm is applied multiple times for real targets, it will give us the mean fraction of observed objects in a field and the variation of this in a field of view. The steps described below are applied to each individual target catalogues (subsurveys). In each field: estimate the fraction of observed/targeted objects. Fraction of objects that get fibers. This can be done using 4FS when the targeting algorithm is run repeatedly for the same field. 2. In each field: estimate the field coverage, i.e. how uniform is the target allocation in a given field. This should take into account the fixed fiber pattern, fibers that are allocated for higher priority targets, and plate scale variations. 3. In each field: the fraction of targets that are successfully observed. 4. Combining the estimated completeness in all fields to generate full-sky completeness map.

10.3.2 Selection Function Pipeline (4SP) Quality Control

QC measures for 4SP are still being developed. The OSF can be QCed on a nightly basis, as it produces information per target. Expected correlations can then be explored in a similar way to the 4GP. For example the OSF probabilities should be correlated with spectral signal-to-noise. These plots and correlations will be manually inspected on a per-night cadence, after data has been returned to 4OR, to monitor pipeline stability and performance. If data are found to be clearly problematic it will be possible to manually retract them using standard 4OR procedures. These plots are expected to be extremely valuable for the first weeks of operations, and the necessity of producing them after that will be evaluated accordingly.

The GSF will be produced less often, typically for a data release. The main tool to estimate that the GSF is scientifically sensible, should again be to recover expected correlations. The probability of observing a target should be anti-correlated with the local density of available targets, but be positively correlated with the number of visits, for example. GSF calculations will be repeated until the sky map of

Current status: QC measures are still at the design stage at this point.

10.3.3 Selection Function Pipeline (4SP) Organisational Structure and Staff Effort

We are currently working on improving the staff commitments for IWG4. However, with the first delivery of the selection functions scheduled for a later DR, this is not the first priority for the moment.

Table 9 IWG4 committed staff efforts for development and survey operations phases

Name	Institute	FTE	Main Contribution to iWG4
Shadab Alam	Tata Institute of Fundamental Research	0.25	IWG4 Extragalactic Lead
Elmo Tempel	Tartu	0.25	WP2 lead
Sanjib Sharma	STScI	0.15	WP1 lead
Sarah Martell	UNSW	0.05	
TBD	N/A	0.25	IWG4 Galactic Lead

10.4 Classification Pipeline (4CP)

10.4.1 Classification Pipeline (4CP) Data Production

The requirements for the 4CP data production are specified in the Classification Pipeline (4CP) Requirements Specification document [RD15]. IWG9 will produce one FITS file per OB, to be ingested by the 4OR database, implying a daily cadence for both retrieval of input data products from the 4OR and delivery of output data products to the 4OR. It will contain the probabilistic classifier tags determined by the 4CP and will be delivered as a 2D Binary Table. Each class tag can be in the range [0,1]. Input data are (stacked) L1 spectra retrieved from the 4OR.

The probabilistic multi-classifiers developed by IWG9 are based on data-driven machine learning methods. These multi-classifiers will be a) a coarse classifier (star, galaxy, AGN, quasar) and b) fine classifier that match sub-module requirements by 4XP and 4GP; these are concerning 4GP modules for FGKM stars, OBA stars, and white dwarfs (WD), as well as probabilities for binary and variable (Cepheids, RR Lyrae) galactic objects. Currently, classifiers with FGKM, OBA and WD classes have been tested. In addition to galactic and extragalactic classes, the sub-class 'unknown' is defined to provide an anomaly score for each spectrum ingested. Sub-classes to match 4XP requirements are star forming galaxies as well as AGN sub-classes. AGN sub-class requirements are still in development in line with module developments under consideration for 4XP. The structure of 4CP is summarised in Figure 15.

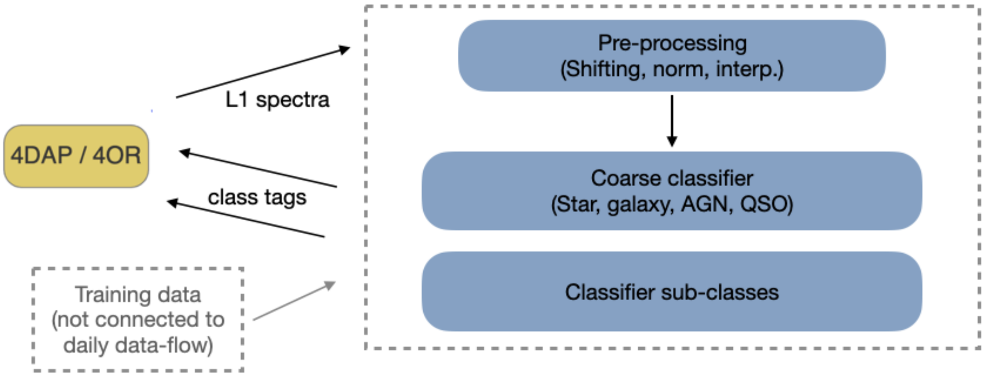

Figure 15: Outline of 4CP classification modules.

The classifier models themselves are currently based on neural network models; also random forests and similar trained classifiers are possible to be added based on their performance. Different variants of convolutional neural networks (CNNs) are benchmarked for the coarse classifier; this includes the GasNet model also tested in 4XP for the redshift estimator. Comparisons between the network models (performance as a function of class) are underway. Fine classifier models as well are tested based on CNN variants. For each class and sub-class probability an uncertainty is to be provided. To this end, Bayesian networks have been tested and will be further developed.

Training is performed offline from the daily dataflow and with lower cadence, while evaluation is performed daily. Currently network models are trained on SDSS spectra, SDSS superset spectra, and have also been tested for OpR2.5 data. Data augmentation for training with mock 4MOST spectra based on spectral templates is explored in close exchange with extragalactic and galactic pipelines. As soon as 4MOST commissioning data is available, extensive re-training is foreseen. To improve classifier performance, active learning that involves human intervention will ideally be included for re-training.

Concerning the daily evaluation schedule, both the coarse and fine classifier models are, as trained and supervised machine learning models, fast codes. They can classify one night of spectra at the order of seconds to minutes, with the main bottleneck being data ingestion and writing. Basic steps for the evaluation of coarse and fine classifier models include after data loading the normalisation (different normalisation choices depending on the classifier model, as defined by the model-loader pipeline sub-module) of spectra, their interpolation to the desired 4MOST wavelength grid if needed and ingestion to the classifier model itself, which returns the normalised probability class tags.

10.4.2 Classification Pipeline (4CP) Quality Control

In training, accuracy and confusion matrix between classes serve to judge the classifier performance and further augmentation of training data needed. Besides high accuracy for single classes a balance with low confusion between different classes is a goal when verifying performance. During training therefore always at least ten percent of the labelled data available is retained for testing. Re-training and validation with test data is performed offline in irregular intervals (weeks to months), especially when novel training data becomes available.

For unlabelled spectra unseen by the models, quality control can entail finding spectra with undecided probabilistic class tags across classes (for example 50:50 classification), as well as class tags that contradict metadata expectation. In addition, high class uncertainties will point to difficult to classify spectra. Finally, a high anomaly score will point towards spectra to be flagged for potential human inspection. Class uncertainties are currently under development and will be provided alongside class tags on a nightly basis, as is the anomaly score. Optional class tag behaviour across classes and comparison to metadata expectation is fast and can either be evaluated on a nightly basis or in regular intervals every few days.

10.4.3 Classification Pipeline (4CP) Organisational Structure and Staff Effort

The details of the organisational structure and the available staff effort for the 4CP pipeline is described in Table 10. List of work packages:

WP.1: Spectral classification. Provide probabilistic class membership, class uncertainties, test supervised learning methods. [‘ML classifier’]

WP.2: External data and training data. Explore the possibility to use all-sky surveys and catalogues, for information on templates and classes, active learning. [‘training data’]

WP.3: Unsupervised feedback, anomaly scoring. [‘Classification inspection’]

Table 10: Overview planned staff effort in IWG9.

Name	Institute	Survey	Contribution, WP	FTE
Simon Barton	Uppsala U.	S8	BNN classifier, WP1, WP2	0.4
Maciej Bilicki	Center for Theoretical Physics Warsaw	S7, S18	classification, photo-zs	0.05
Hyunseop (Joseph) Choi	Montreal U.	S6	BAL quasars, WP1, WP2	0.1
Andreas Korn	Uppsala U.	S2	BNN classifier	0.1
Karen Leighly	University of Oklahoma	S6	BAL quasars, WP1	0.1

Name	**Institute**	**Survey**	**Contribution, WP**	**FTE**
Alfred Castro Ginard	Leiden U.	S13	open stellar clusters, WP2	0.1
Caroline Heneka	Heidelberg U.	S7	IWG co-lead, classifier models, all WP1-3	0.15-0.2
Domitilla de Martino	INAF	S11	classification compact binaries, WP2	0.01
Nicola R. Napolitano	Sun Yat-Se U.	S7, S10	classification extragalactic, WP1-3	0.2
Anna Francesca Pala	ESA/ESAC	S11	classification of white dwarf binaries, WP1-2	0.01
Arthur Alencastro Puls	Frankfurt U.	S14	late-type (FGK) stars, WP2	0.05
Martin Sahlén	Uppsala U.	S8	BNN classifier	0.1
Simone Scaringi	Durham U.	S11	compact binaries, classification, WP1, WP2	0.15
Paula Szkody	University of Washington	S11	classification variables, WDs, WP1, WP2	0.05
Fucheng Zhong	Sun Yat-Se U.	S7	classification, anomalies, WP1-3	0.2

11 Project Level Quality Control and Data Delivery

11.1 Archiving, Publication and ESO Delivery overview

4MOST raw data will be archived in the ESO SAF. While the project will keep copies of the raw data, fault protection will be minimal under the assumption that raw data can always be re-downloaded from SAF.

4MOST L1 and L2 data that are part of a public data release will be archived in the ESO SAF and the 4MOST Public Archive (4PA). Both of these systems will provide the worldwide user community with access to the data, as well as search capabilities.

4MOST L1 and L2 data will be internally archived in the 4MOST Operational Repository (4OR). This serves as the central hub for distribution to and collection of data from the various pipelines. Processing will be based on chunks that correspond to one night.

Data will be trickled from 4OR to 4PRAP on a daily basis. Data in 4PRAP are subject to retraction at any time, which will be made clear to all stakeholders. As described below, 4PRAP will be used for quality control purposes and data release preparation. Data release preparation consists of the creation of an exclusion list. This list excludes either certain targets, or certain parameters from a release, based on QC parameters or considerations that follow from the data release agreements within the project.

Once an internal data release is defined it can be ingested from the 4PRAP into the 4PA. The internal data release candidate will be checked by the DRM, OM, PI, and others, before finally being made internally available. All papers published by 4MOST scientists shall be based on identifiable internal data releases.

Public data releases will be based on an internal data release, with additional exclusions to be added as defined jointly by the project and ESO. Public data releases will be packaged in 4PRAP and delivered to the SAF.

Further details on how the project aims to collaborate with ESO towards data releases can be found in the OpR3 Ph3 plan [RD 20]. We emphasize that this is a development plan that will change over time, as we adapt to what we learn and the resources that are available.

11.2 TARGET/OBJECT names and unique identifiers

The 4MOST TARGET/OBJECT names are defined in the document Definition and provenance of target and object identifiers in the 4MOST data flow [RD16].

Overall, the following target identifiers are used within 4MOST:
TARG_ID: unique target identifier.
U_OBJ_ID: unique object identifier, discriminated based on spectroscopic resolution mode (LR or HR). Note it is referred to as OBJ_UID in the FIBINFO table.
CNAME: internal object identifier based on celestial coordinates. Note it is referred to as OBJ_NME in the FIBINFO table.
SPECUID: unique object spectrum identifier.
SPECUID_P: SPECUID but with additional provenance-related identification information.
IAU_NAME: IAU-approved object name used in published catalogues.

Target catalogues are created by the survey users. They are uploaded to the 4FS_WI on a 6-monthly cadence for non-transient targets and to 4FS_API on a daily cadence for transient targets. All target catalogues are then concatenated. To ensure that each target is uniquely identifiable within the concatenated catalogue, an identifier (TARG_ID) is calculated for each target. The shared-target procedure is then used to identify all unique objects (U_OBJ_ID) from the concatenated target catalogue, which are stored along with the associated targets in the object catalogue. The object catalogue grows each time the shared-target procedure is run following a target-catalogue ingest. Consequently, all targets uploaded through 4FS, from the first ingest of target catalogues through to the latest target-catalogue ingest, are linked to objects, providing not only a catalogue of unique objects but also the provenance of each target. Each object also has an associated name, which encodes the object position using sexagesimal coordinates (CNAME).

Targets are selected for nightly observations with the key object properties stored within the FIBINFO table. All target spectra are given a unique identifier (SPECUID) which is linked to each object through DMS@CASU.

The 4MOST project has been registered with the IAU who have approved the following naming format for 4MOST targets:

4MOST JHHMMSSss+DDMMSSs (example "4MOST J12290568+0512035"),

where coordinates shall be created based on the reference EPOCH of Gaia DR3 (2016). Any proper motion will be corrected such that the IAU_NAME references the position on the sky in the year 2016. Targets are identified in publishable 4MOST catalogues using the IAU-approved object name calculated from the astrometric properties of the observed object. OpSys is responsible for the creation of the IAU-approved object name.

11.3 Project Level Quality Control process

Quality control on the aggregate L1 and L2 data comes in two flavours in the 4MOST project. IWG6 Quality Control is in charge of this process.

First, the 4MOST Quality Control Pipeline (4QP) will produce quality flags for each dataset. This includes for example flagging missing calibration data, unexpectedly low signal-to-noise, high background, etc. Most of these flags will describe properties of the data that result from the observations process, i.e. are not related to a bug or possible improvement in the pipeline. The 4QP will also provide some basic checks on data integrity, such as parameter ranges, but this is not the main task.

Second, IWG6 will carry out visual inspection of the data using the 4PRAP environment. Visual inspection may consist of looking at spectra to find unexpected features. It may also consist in plotting trends using derived parameters to check for expected or unexpected correlations. This activity cannot be planned very well, as it will depend on the relative science goals of involved scientists, as well as their ability to come up with useful tests. Visual quality checking may lead to improvements in the 4QP pipeline in case it is found that a useful quality parameter could be attached to the data. Issue reports will be streamlined to the relevant parties (e.g. DMS, L2 pipelines) via a ticketing service, overseen by the IWG6 co-leads.

The 4QP pipeline will run on the data nightly, i.e. on the same frequency as any other L2 pipeline. The output of the 4QP is another DXU that can be read by the 4MOST archives and can be used by any science user downstream of 4QP. The visual inspection will be carried out on a somewhat more random schedule during working hours of IWG6 personnel. It is envisaged that all nights get some level of visual checking, even if this checking happens a few days later than the data were taken.

The Data Release Manager (DRM) oversees the selection of those data that will be contained in a data release from the data available for the period associated with that data release. The filtering of data into the data release will be carried out using 'rules of exclusion'. Examples of exclusions are: a specific target and all its associated spectra and measurements or; a specific measurement (parameter) across a selection of targets. Reasons for exclusion can be either bad quality (catastrophic failures in data quality, based on the quality flags as set by 4QP), or "political", i.e. a target does not belong in a release for reasons foreseen in the Survey Management Plans release schedule. The 4PRAP will support the DRM such that the filter can be applied automatically for a large number of targets to create the data release.

11.4 4MOST Public Archive (4PA)

In preparation of a data release, 4PA will collect the Data Release Candidate (DRC) data (L1 and relevant L2, Catalogues, ancillary data) from 4PRAP. On a more technical level, the DRC consists of files and tabular data. The tabular data is ingested into the database, while the files are unpacked and copied to the 4PA storage area.

The DRC tabular data will be curated with necessary information for easier access and integration with other datasets. The Data Release Managers (DRM) have then the opportunity for final checks on the DRC. When the DRM approves the DRC, it becomes a Data Release and is made available through the 4PA interfaces.

The 4PA will provide access to the data products of the 4MOST consortium, serving them to the worldwide community of professional and amateur astronomers. The 4PA will provide a set of ways to perform queries against the database. The 4PA web and the API interfaces will allow logged in users (science team members) access to the internal data releases and anonymous users (without account) access to the public releases.

The tabular data will be hosted in the open-source database management system PostgreSQL.

The 4PA web interface and associated services will be implemented using the existing Daiquiri framework. This framework, based on the popular Django framework written in Python, was developed at AIP to create highly customizable web applications for astronomical data publication. The code is Open Source, easily maintainable and well documented.

The 4PA will provide a set of ways to perform queries against the database that follows the standards of the Virtual Observatory as well as SQL. The DB access will provide a keyword search against the Name and the FITS headers of the files in the archive, but also a full ADQL interface to perform queries against the main catalogue of sources as well as the different L2 products. There will be a set of form-based queries for common queries, e.g. cone search for objects in a given right-ascension-declination range, redshift selections, or selections based on different physical parameters." The API access will include a TAP API to perform ADQL queries. This includes the possibility to perform synchronous and asynchronous queries. TAP

is implemented in several popular client-side programs and libraries (e.g. TOPCAT, astropy). Additionally, 4PA will include a spectral viewer for the spectra.
Furthermore, the 4PA will include features that enable the use of the 4MOST data with other datasets. Users will be able to upload tabular data of limited size and use them within a query, e.g. for cross-matching. The 4PA will also contain common external catalogues (e.g. Gaia) that can be used to formulate queries and for cross-matching purposes.

11.5 Level-1 Data delivery and phase 3 compliance

Concerning spectra, four different data products will be delivered:

- Single spec: The 1D single spectra for all observations
- OB-level stack: 1D spec resulting from the combined repeated spectra in a single OB.
- Night-level stack: 1D spec resulting from the combined spectra over a full night.
- Deep-level stack: 1D spec resulting from the combined spectra over all observations for a single target.

All low resolution stacked products can be either joined or not. However, all combinations will be joined at the moment, meaning the spectra are stacked, and the blue, green and red portions are joined in a single spec.

All single spec and stacked product deliveries will be compressed in MEC files and inflated at ESO. After inflation, the spectra must pass the tests during ESO ingestion, when all header keys are checked against their document, and the values attributed to them are validated. Any inconsistency makes the spectra to be failed and not ingested in the database.

Additonally, the catalogue of objects included in a public data release is part of the L1 dataset.

The project is aware that Phase 3 compliance will be a major task. Therefore, Operational Rehearsals have been planned and are underway to ensure early communication with ESO, in particular the SAF.

11.6 Level-2 Data delivery and Phase 3 compliance

For all common L2 pipelines, adherence to Phase 3 standards is checked by DMS before ingestion in 4MOST Operational Repository (4OR) at CASU. This is enforced by employing strict interface management in the form of Data eXchange Units (DXUs) definitions against which the data is checked at each ingest. Any change of DXUs is overseen by the Data Model Curator and has to be agreed by the archives (4OR, 4PA) to ensure continued compliance to standards.

The L2 pipelines will check on a daily basis whether new L1 data is available for processing. This ensures that, every time previously generated L1 data is updated due to a change in the L1 pipeline, the L2 pipelines will automatically reprocess these updated data and also update their outputs.

All newly generated L2 data are ingested into the 4OR. From there they are trickled on a daily basis to the staging area in the 4PRAP for Quality Control and release preparation as described in Section 11.3.

11.6.1 4GP delivery and compliance

The full details of the 4GP data format are described in the IWG7 Level 2 DXU (MST-TNO-4MOST-47110-9237-0001). Compliance to the DXU will be verified at ingestion into the 4OR and the 4PA. Additionally, all DXUs are under common control by the Data Model Curator, who will provide high level guidance to develop SAF-compliant DXUs and data model versioning.

11.6.2 4XP delivery and compliance

The full details of the 4XP data format are described in the IWG8 Level 2 DXU (MST-TNO-4MOST-47110-9238-0001). Compliance to the DXU will be verified at ingestion into the 4OR and the 4PA. Additionally, all DXUs are under common control by the Data Model Curator, who will provide high level guidance to develop SAF-compliant DXUs and data model versioning.

11.6.3 4SP delivery and compliance

The full details of the 4SP data format are described in the IWG4 Level 2 DXU (MST-TNO-4MOST-47110-9234-0001). Compliance to the DXU will be verified at ingestion into the 4OR and the 4PA. Additionally, all DXUs are under common control by the Data Model Curator, who will provide high level guidance to develop SAF-compliant DXUs and data model versioning.

11.6.4 4CP delivery and compliance

The full details of the 4CP data format are described in the IWG9 Level 2 DXU (MST-TNO-4MOST-47110-9239-0001). Compliance to the DXU will be verified at ingestion into the 4OR and the 4PA. Additionally, all DXUs are under common control by the Data Model Curator, who will provide high level guidance to develop SAF-compliant DXUs and data model versioning.

11.7 Time-line delivery of Level-1 and Level-2 data products to the ESO archive

The Data Release Plan [AD1] describes in detail the timeline of the public Data Releases (DRs) of the 4MOST Project. A summary can be found in the companion 4MOST Survey Management Plan: Organisation [RD1] document. Further details on how the project aims to collaborate with ESO towards preparation of data releases can be found in the OpR3 Ph3 plan [RD 20]. We emphasize that the latter is a development plan that will change over time, as we adapt to what we learn and the resources that are available.

Reduced data volume: 248 kb per singlespec. While spectra are transferred to ESO in MEC files, the MEC is unpacked in a staging area. Therefore, what matters for SAF ingestion is the singlespec size.

Raw data volume 227 MB per ccd image for each spectrograph.

4GP will provide normalized spectrum and best fit spectrum for the vast majority of targets in 4MOST. Whether these are part of the ESO delivery is subject to further discussion with ESO. These will be in the same FITS file as the table of derived properties.

11.8 Staff effort devoted to Project level quality assessment and data delivery

Project level QC is carried out in IWG6 and in the Data Release Management work package. IWG6 has two co-leads (Phil Wiseman and Ricardo Carrera), as well as one representative per survey in the 4MOST project. The IWG6 WP is under the direct authority of the Operations Manager.

The DRM work package is staffed by two experts, namely Clare Worley (20% of time) for the Data Release Manager and Ole Streicher (50% of time) for the Data Model Curator. The DRM work package is part of the 4PA work package, but DRM contributors regularly attend ODG/JOG meetings, meaning communication is very direct.

Appendix A List of Acronyms

List of Acronyms	
4MOST	4-metre Multi-Object Spectroscopic Telescope
4DAP	4MOST Data Access Point
4DPC	4MOST Data Processing Centre
4GP	4MOST Galactic Pipeline
4OR	4MOST Operational Repository
4PA	4MOST Public Archive
4PRAP	4MOST Pre-Release Access Point
4SP	4MOST Selection Function Pipeline
4XP	4MOST eXtragalactic Pipeline
AD	Applicable Document
AGB	Asymptotic Giant Branch
AL2	Additional Level 2
CCD	Charge Coupled Device
CCF	Cross-correlation Function
CNN	Convolutional Neural Network
DL2	Deliverable Level 2
DL2-SURV	Deliverable Level 2 - Survey (data product not delivered by an IWG, but by a 4MOST Survey)
DMS	Data Management System
DOE	Detection Of Extrema
DR	Data Release

List of Acronyms	
DRC	Data Release Candidate
DRM	Data Release Manager
DRPD	Design Report Pipeline Description
dVM	designated Visitor Mode
DXU	Data eXchange Unit
ETC	Exposure Time Calculator
FITS	Flexible Image Transport System
FP	Fabry-Pérot
FTE	Full Time Equivalent
HB/RC	Helium Burning/Red Clump
HRS	High Resolution Spectrograph
ICD	Interface Control Document
IDR	Internal Data Release
IWG	Infrastructure Working Group
JOG	Joint Operations Group
L1	Level 1
L2	Level 2
LRS	Low Resolution Spectrograph
MS	Main Sequence
OB	Observation Block
ODG	Operations Development Group
OpSys	Operations System

List of Acronyms	
OSP	Operations Start-up Phase
PCA	Principal Component Analysis
PDR	Public Data Release
PN	Planetary Nebulae
QA	Quality Assurance
QC	Quality Control
QE	Quantum Efficiency
RD	Reference Document
RGB	Red Giant Branch
RV	Radial Velocities
S/N	Sign-to-Noise ratio
SAF	Science Archive Facility
SB1	Single Lined Binary stars
SCB	Science Coordination Board
SED	Spectral Energy Distribution
simulcal	Simultaneous Calibration
SMP	Survey Management Plan
SPV	Survey Programme Validation
ST	Science Team
TBC	To Be Confirmed

List of Acronyms	
TBD	To Be Determined
ThAr	Thorium Argon
TO	Turn-Off
VALD	Vienna Atomic Line Database
WD	White Dwarf
WP	Work Package
WR	Wolf Rayet