ESO Phase 3 Data Release Description

Data Collection	GIRAFFE
Release Number	2
Data Provider	ESO, Quality Control Group
Document Date	2023-01-09
Document version	2.2
Document Author	Isabelle Percheron
Changes with respect to the previous version: - some faulty stacks have been removed and replaced by single exposure products (see p.29) *You may want to update your already downloaded version 1 spectra if cosmics have been an issue in the past, or if you want to benefit from the new stacked products.* -Acknowledgment section	

Abstract

This is the release of reduced 1D spectra from the FLAMES/GIRAFFE[1] spectrograph, taken in the MEDUSA mode. This is the most frequently used multi-object spectroscopy mode of GIRAFFE, as opposed to its two IFU modes. FLAMES is the multi-object, intermediate and high resolution spectrograph of the VLT, having two components: GIRAFFE and UVES. The GIRAFFE spectrograph allows the observation of up to 130 targets in one shot with intermediate resolution.

This release is an open stream release, it includes the so far observed GIRAFFE MEDUSA data and will be continued into the future operational lifetime. The data content will grow with time as new data are being acquired and processed (approximately with monthly cadence and with a short delay of 1 or 2 months). The processing scheme is as homogeneous as possible.

The release covers all GIRAFFE MEDUSA data starting from the begin of operations, 2003-04-13, until now. For the early data until 2005-05, the master calibrations have been reprocessed using the same pipeline version as used for the science products.

The data have been reduced with the GIRAFFE pipeline, version `giraf-2.16.3` and higher. The data have their instrument signature removed: they have been cleaned for cosmics, de-biased, flat-fielded, extracted, and wavelength-calibrated. Their wavelength scale has been corrected to the heliocentric reference system. The sky has not been subtracted.

Recently we have added stacked 1D spectra to the release. For those OBs[2] where more than one exposure exists, the individual 1D spectra are stacked in a second step, this data product being the intended output of the observation. The spectra from single exposures have recently been reprocessed with a preprocessing step for cosmic ray cleaning. These data are now available as version 2 of the release, and users who have downloaded the uncleaned data (version 1) at earlier times may want to consider downloading this new version.

The processing is performed by the Quality Control Group in an automated process. The pipeline processing uses the archived, closest-in-time, quality-controlled, and certified master calibrations. It is important to note that the reduction process itself is automatic, while the quality assessment and certification of the master calibrations is human-supervised.

The pipeline products come in the ESO 1D standard binary table[3], along with a set of ancillary files. The data products cover many different science cases (defined by their parent programmes, in the case of GIRAFFE several hundreds).

---

[1] http://www.eso.org/sci/facilities/paranal/instruments/flames.html

[2] OB = Observing block, a single pointing on the sky and the fundamental unit of the VLT observations.

The data format follows the ESO 1D spectroscopic standard[3] for science data products. Each spectrum is a multi-column binary table. There are N product files for each input science raw file, where N corresponds to the number of science fibres. N can be as high as 130 (the number of science or sky fibres in each Medusa fibre system)[4]. If an OB contains M exposures identical in instrument mode and setup, each of the M products from a given fibre is also combined into a stacked data product.

This data release offers science-grade data products, with the instrumental signature removed to a large extent. The spectra come with error estimates and a signal-to-noise column (SNR), and with a list of known shortcomings. They are considered to be ready for scientific analysis. They are expected to be useful for any kind of medium resolution spectroscopic research, including abundance and line profile studies, and radial velocity studies. There has been no attempt to correct for sky background. We provide the signal from all SKY fibres but leave it to the user to select the appropriate ones for a given target. See more under 'Data reduction and calibration'.

Disclaimer. Data have been pipeline-processed with the best available calibration data. However, please note that the adopted reduction strategy may not be optimal for the original scientific purpose of the observations, nor for the scientific goal of the archive user.

Release Content

The GIRAFFE release is a stream release. The overall data content is not fixed but grows with time as new data are being acquired and processed. The data are tagged "GIRAFFE" in the ESO archive user interface[5].

The first data have been published under GIRAFFE in April 2015. New data taken since then are being added at roughly monthly intervals and with a delay of 1 or 2 months after acquisition, when all master calibrations for the corresponding time interval are certified and available in the archive. With the recent reprocessing, all GIRAFFE Medusa data are either processed into stacked spectra (if the OB contains more than one SCIENCE exposure), or into single-exposure spectra (if not). No stacking beyond the single OB execution is done. The OB stacks are new data products added in 2019. The single exposures have been reprocessed until 2019-01-14 until which date they come as version 2 (with version 1 being the old data products uncleaned for cosmics). After that date, there are no old data products, and the cleaned products come as version 1.

The single-exposure data products are described in Part I, the OB-stacked spectra in Part II.

---

[3] http://www.eso.org/sci/observing/phase3/p3sdpstd.pdf
[4] Five more fibres can be used for the simultaneous calibration, so that the total number of fibres is 135.
[5] http://archive.eso.org/wdb/wdb/adp/phase3_spectral/form

Part I: Single-exposure data products

Data Selection

Data selection is rule-based. The following rules apply:

- instrument=GIRAFFE;
- observing technique (FITS key DPR.TECH) = MOS, INS.MODE = MED, INS.SLIT.NAME = Medusa1 or Medusa2;
- category (DPR.CATG) = SCIENCE;
- type (DPR.TYPE) = OBJECT,OzPoz or OBJECT,SimCal;
- 'normal', 'slow' and 'fast' read modes supported; however, as non-standard modes, both the 'slow' and the 'fast' read modes frequently have incomplete master calibrations and then could not be processed into data products.

No selection is made on the basis of the observing mode (visitor or service), nor on settings.

Settings. GIRAFFE MOS settings are defined by the combination of INS.SLIT.NAME (*Medusa1* or *Medusa2*) which is actually the fibre system; central wavelength (in nm); and grating name (HR or LR). For some central wavelengths, two different order separation filters are offered for the observations. In those cases the setting is made unique by appending a letter A or B (e.g., H525.8B). In all other cases the setting is marked as L543.1 or H665.0. The setting is stored in the FITS key INS.EXPMODE.

Input files. Each spectral data product is based on a single input raw file. Stacks from setup-identical raw files from the same OB execution are also available, see the Part II of this release description. We did not combine multiple exposures with different wavelength setups within the same OB. We did not combine multiple OBs.

SimCal vs. OzPoz. For the DPR.TYPE, two different values exist:
- OBJECT,OzPoz: this is the classical GIRAFFE Medusa observation, using the OzPoz fibre positioner on plate 1 (with the fibre system Medusa1) or plate 2 (Medusa2). Both are logically equivalent (one plate is used to configure the next OB while the other one is still observing) but need their own fibre and calibration system and are therefore always distinguished. An OzPoz observation has *always* been processed into a set of products, unless, very rarely, the processing failed or an OzPoz observation has no science fibre configured.
- OBJECT,SimCal: this is also a GIRAFFE Medusa observation taken with OzPoz, with additional simultaneous arclamp calibrations ("SimCal") illuminating the 5 SimCal fibres. The purpose is to provide a reference for incremental shifts against the daytime arc-lamp calibrations. This observing mode is chosen for high-precision radial-velocity studies.

Beyond about 700 nm there are some very strong arc lines which easily saturate and contaminate the signal of neighbour fibres. Therefore many observers have designed their OBs such that they contain multiple exposures, with one or more short SIMCAL observations and one or more longer OZPOZ observations, taken with the same fibre configuration and setting. The short SIMCAL observations are designed to deliver the signal of the SIMCAL lamp (5 fibres) without saturation, but the signal of the science fibres is just noise. These SIMCAL data can be considered as *calibrations*. Unfortunately they are not formally marked as such but still carry the DPR.CATG value SCIENCE.

Below about 700 nm, there exist many SIMCAL observations with long exposure times, in settings where the SIMCAL emission lines do not saturate. These data are clearly science-type observations, with the SIMCAL fibres adding RV precision.

We have made an attempt to recognize this complex pattern empirically. For those OBs with an OZPOZ observation and (one or more) SIMCAL observations, we have always treated the SIMCAL observations as *calibration-type* and attached them to the corresponding OZPOZ products, with their names starting with GI_SSIM. With this delivery pattern, we intend to avoid offering pseudo-science products from the short SIMCAL exposures which would actually contain only noise in the

science fibres.

A more complex handling is required for stand-alone SIMCAL observations. These come typically as single observations with longer exposure times, but some observers (typical for Visitor Mode) have designed special OBs with complex SIMCAL observation patterns (several settings combining short and long exposure times). With the goal in mind to distinguish calibration-type and science-type SIMCAL data, we have normally rejected stand-alone SIMCAL observations, unless the following criteria apply:

- Exposure time is longer than 120 sec.
- Wavelength setting is not contained in a list of configured rejection values that mostly contain the red settings beyond 700 nm (but 881.7 is accepted).

The threshold of 120 sec is empirical, almost all calibration-type SIMCALs are shorter or equal.

This schema works well to save the science-type SIMCAL products and avoid calibration-type products containing only noise.

In Table 1 we have collected the most typical patterns we have seen in GIRAFFE Medusa OBs:

Table 1: Possible OzPoz/SimCal combinations and their data product setup

within the same OB: DPR.TYPE ...	resulting in ...
OZPOZ (only)	classical set of products, no SIMCAL attached
OZPOZ | OZPOZ	two sets of products, no stacking
SIMCAL (only)	if `exptime` is longer than 120 sec: accepted as SCIENCE if central wavelength shorter than about 700nm (with a few exceptions), otherwise not
SIMCAL | OZPOZ	OZPOZ products, with 1 SIMCAL attached
SIMCAL | OZPOZ | SIMCAL	OZPOZ products, with 2 SIMCALs attached
OZPOZ | OZPOZ | SIMCAL	both OZPOZ get their own product dataset, with the SIMCAL file attached
SIMCAL60 | SIMCAL30 | SIMCAL10 | OZPOZ	multiple SIMCALs with short exposure times: all attached to OZPOZ (with SIMCAL10 meaning 10 sec exposure time)

There are very rarely false suppressions, with a SIMCAL file intended by the PI as science file. This could have happened for a SIMCAL short exposure (less than 120 sec) taken in the settings beyond 700nm. In the few cases which we discovered and checked these were "exotic" observations, like 1 science fibre with lots of sky fibres. These likely are flux or RV standard star observations.

There are very rare cases when an OZPOZ file did not contain a single SCIENCE fibre (only SKY fibres). In that case no product was created.

General processing pattern. We have processed only those Medusa science data for which certified master calibrations exist in the archive. These master calibrations exist normally at daily frequency. They were all processed by the Quality Control Group, close to the date of acquisition, in order to provide quality feedback to the Observatory. All master calibrations used here were certified, meaning checked for quality and proper registration of instrument effects.

In general, the master calibrations were processed with different (earlier) pipeline versions than the science data in this release. The most significant change in the GIRAFFE pipeline was related to the replacement of the old CCD by the new CCD in 2008-05. While data from the old CCD often have a glow induced by a diode, the new CCD does not have this artefact. The old CCD also had two bad columns while the new CCD has no such imperfections. The reduction strategy required the subtraction of a scaled master dark for the old CCD, and the masking of the bad columns with a bad pixel mask. Both calibration steps are not required for the new CCD. See "Master calibrations" below for a discussion of the main differences.

Science data with the `OBS.PROG.ID` starting with 60. or 060. have been de-selected, considering

them as test data. Data taken at daytime (with obviously wrong 'SCIENCE' tag) have been ignored. Otherwise the header tag 'SCIENCE' has been blindly accepted from the raw data (originally defined by the PI), thus including sometimes standard stars intended by the observer for use as calibrators for flux or radial velocity. Please check the `OBJECT` header key for such cases. It is likely that most of the products with just one science fibre fall into that regime. Also, there are rare cases when test observations were executed under the SCIENCE label. Some very short exposures with no signal fall into that category but have not been suppressed. For a discussion of the science vs calibration property of the SIMCAL data see the above section.

GIRAFFE does not foresee flux calibration for the Medusa mode. The data come extracted and wavelength-calibrated but with no flux calibration.

There is no raw science data selection based on quality. Likewise, we have not selected by OB grades[6]: the observations might have any grade between A and D (if taken in Service Mode, SM), or X (in Visitor Mode, VM). The availability, or non-availability, of a particular file in this release does not infer a claim about the data quality. OB grade and comments are stored in the product FITS header.

Release Notes

Pipeline Description

The data reduction uses the standard GIRAFFE pipeline recipe `giscience`. Find its description in the FLAMES/GIRAFFE Pipeline User Manual[7].

Find the pipeline version used for processing in the header of the product file, under "HIERARCH ESO PRO REC1 PIPE ID". With the reprocessing of the single exposures in 2019-02 with the cosmic-ray cleaning, the pipeline version for all data processed until 2019-03 was `giraf/2.16.3`. It will evolve in the future. Very few GIRAFFE spectra could not be reprocessed with the pre-processing cleaning step because of algorithmic failures. Those data still carry a pipeline version earlier than 2.16.3.

For the recipe parameters, almost any are set to the pipeline defaults. The only exceptions are:

* `bsremove-method=PROFILE+CURVE` (in order to take into account the structure of   the master bias which has some curvature);
* `bsremove-yorder=5` (the polynomial coefficient for the curvature fit);
* `flat-apply=TRUE`;
* the technical parameters required to make a product compliant with the ESO Data Product Standard[8].

Furthermore, for those single SIMCAL spectra that qualify for the stacking (see part II), we have added:

* `scal-zmax=-1`.

This parameter value disables the (very small) simultaneous wavelength correction which might fail in case of saturated spectra. This would be an issue for the stacking, since some of the input files for the stack could have this correction applied, others not (if it failed due to saturation). If the SIMCAL spectrum is the final product, we have used the default value (which enables the little shift).

Find more details in the Pipeline User Manual[7], section 9.6 (for issue 6).

---

[6] As given by the observatory staff to assess the match with user-defined constraints, for SM data.
[7] Under the GIRAFFE link in http://www.eso.org/sci/software/pipelines/.
[8] Namely: `generate-SDP-format=TRUE` (to generate the output format of the spectra), and `dummy-association-keys=4` for the technical ASSOC/ASSON keys.

Information about the GIRAFFE pipeline (including downloads and manual) can be found under the URL http://www.eso.org/sci/software/pipelines/. The QC pages[9] contain further information about the GIRAFFE data, their reduction and the pipeline recipes.

Data Reduction and Calibration

Reduction steps.

Pre-processing: As a pre-processing step, the raw file is cleaned for cosmics using the `PyCosmic` algorithm.[10] This python code is called separately from the pipeline as the first step. It is a robust detection algorithm for cosmics developed specifically for fibre data, combining an edge-detection algorithm with a point-spread function convolution scheme. The master bias is subtracted from the raw file, it is converted to electrons, then `PyCosmic` is applied, the cleaned data are converted back to ADU and the master bias is added again, to be handed over to the GIRAFFE pipeline. This raw file is cleaned for cosmics but otherwise identical to the original raw file. The pixels for which a cosmic has been found are replaced by an average of the neighbouring pixels. (Unfortunately there is no scheme for quality flags in place, to flag those pixels.)

The `PyCosmic` code is the same as the one implemented in the Reflex workflow for GIRAFFE[11]. Find there a detailed description of the algorithm.

With this pre-processing step in place, the single-exposure data products are effectively of the same cosmetic quality as the OB-stacked data products. All fibres are cleaned, including the SKY fibres. We have decided to pre-process all qualifying raw files independent of their exposure time. Even spectra with EXPTIME=100 sec or less have been exposed to `PyCosmic`, without noticeable (positive or negative) effect then.

The pixels affected by cosmics are replaced by an average of the surrounding pixels. This is a difference to the strategy, and quality, of the cleaning in case of stacks, see Part II. The impact of the preprocessing is quite impressive (Figure 1). There is no direct way for checking if a particular wavelength bin is affected by cleaned pixels. If in doubt, compare the product with the earlier version 1. This option exists only for data until 2019-01-14 (acquisition date).

Pipeline processing. The main reduction steps after pre-processing are the following:

- The input spectrum is bias-corrected, using a polynomial fit to the provided master bias.
- For data earlier taken earlier than 2008-05 (with the old CCD), the closest-in-time master dark is scaled for exposure time and subtracted. This helps removing the glow in the upper right part of the CCD. The master darks have been taken about monthly and are associated closest-in-time to the science frame. No such correction is needed for data taken after that date (new CCD).
- For the same date range, the bad-pixel mask is applied, to suppress the wrong and meaningless signal in the two bad columns. No such correction is needed for data taken after that date (new CCD).
- Using the localization and fibre-width tables (coming together with the master flat as solution from the `gimasterflat` recipe), the science spectrum is extracted. We have used the default (SUM) extraction. The optimal extraction has issues with high-frequency oscillations, and would also have required the creation of new master flats.
- The master flat is used to correct for fringing, pixel-to-pixel gain variations and fibre-to-fibre transmission.
- Finally the dispersion solution is used to apply the wavelength calibration, and the resampling to a wavelength grid.
- For SIMCAL data, the provided line-mask is used to cross-correlate with the obtained

[9] http://www.eso.org/qc/GIRAFFE/pipeline/pipe_gen.html
[10] Husemann B. et al., 2012, A&A 545, 137 [2012A&A...545A.137H]
[11] Find the latest version of the Reflex tutorial under http://www.eso.org/sci/software/pipelines/.

wavelength solution, in order to derive a residual wavelength drift. The resulting differential shifts are applied to the SIMCAL data and are written into the binary output table of the SIMCAL observation. (But see the remark about `scal-zmax` in the previous section.)

- The individual spectra have been shifted to the heliocentric RV system, using the barycentric, heliocentric and geocentric corrections. *The ancillary 2D products are not RV corrected* (since the RV shift has subtle differences from fibre to fibre).

Note:
- In the cases where the observer has chosen to have separate OZPOZ and SIMCAL observations, the shift as derived in the SIMCAL observation(s) has *not been applied to the OZPOZ data*.
- No SKY subtraction has been applied. Instead, all fibres with a sky signal (as defined by the observer) are collected in the associated MOSSKY file (with name GI_SSKY). It is generally not possible for an automatic reduction process to select the appropriate sky signals. Many observers configure their sky fibres across the entire field of view. The user may want to use a mean, a 2D fit, or fine-tuned strategies depending on fibre location. If you need SKY subtraction and want to use the signal registered in the SKY fibres, please keep in mind that you may have to *correct back* the correction to the heliocentric rest frame that has been applied to the science spectrum (header key HELICORR).
- (For the older version 1 data, please also note the historical bug in the MOSSKY file for SIMCAL observations, see section "Known features and issues".)

Master Calibrations used for data reduction[12].
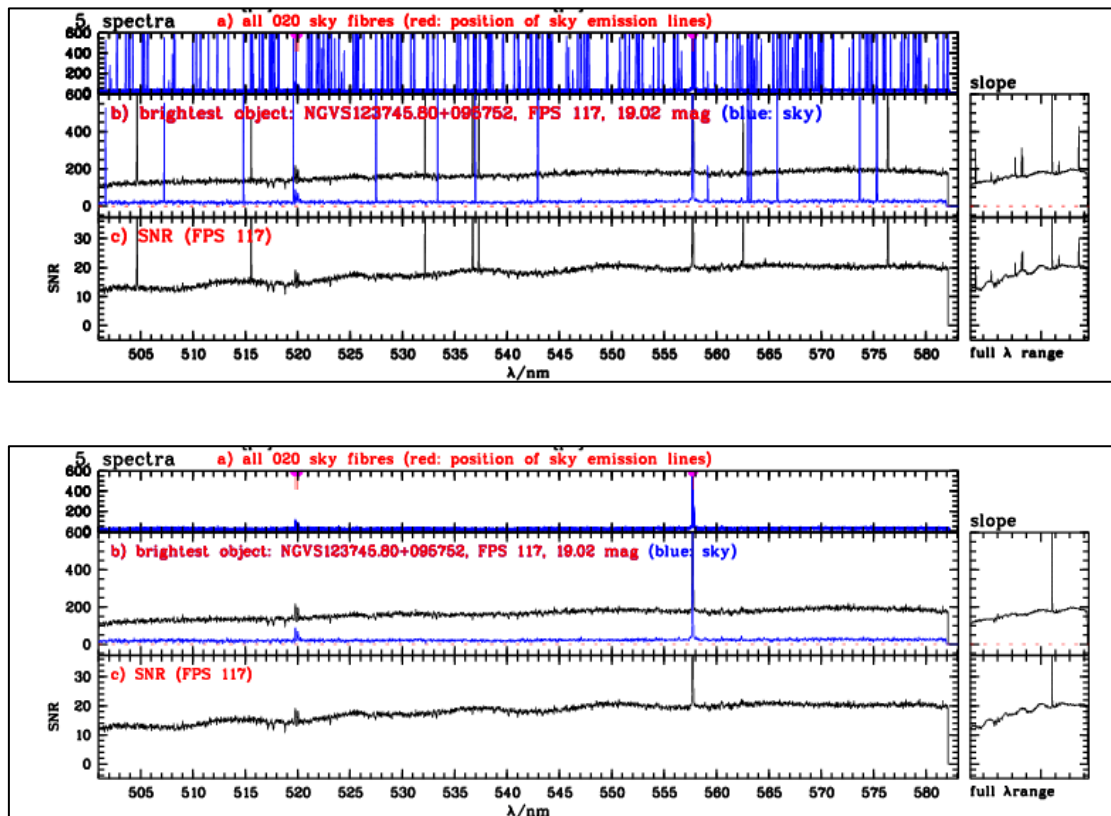All data products from 2005-06-01 on have been reduced using the historical master calibrations,

Figure 1. Effect of the pre-processing cleaning algorithm `PyCosmic`. On top, we display one uncleaned spectrum (the old version 1), at bottom the cleaned version 2. The panels b) display the respective spectrum, the panels c) the respective SNR. In addition, the panels a) display in blue the signals from all 20 sky fibres which consist of a forest of cosmics in the upper case, while they are nicely cleaned in the bottom case, with only two real sky lines left.

---

[12] Please check Sects. 8 and 9 of the Pipeline User Manual for the description of calibration data.

as available in the archive. These master calibrations have been processed close to the time of the data acquisition, to control the quality of the instrument. They have been processed with the respective pipeline version available at that time, which is in general different from the current one used for the science processing. We have carefully checked that the quality of these "historical" master calibrations is comparable to the ones as produced by the current pipeline. More about this in the section "Data quality/master calibrations" below.

The master calibrations from the start of operations (2003-04) until just before 2005-06-01 have been reprocessed using the current pipeline version. This was done for two reasons:
- in those early years, the processing scheme for GIRAFFE master calibrations was not complete but focussed on Service Mode data only; for Visitor Mode data no master calibrations existed;
- for the Medusa1 fibre system, there was the peculiar effect that the "slit geometry setup" tables, in their versions used for later data, were not listing the fibre #27 which was however giving a signal in the first half year until 2003-08-27. Later that fibre was disabled for poor performance. This situation could not be handled properly by the pipeline and hence only Medusa2 master calibrations were available.

Both deficiencies were considered serious enough to be remedied by a complete (re)processing of the master calibrations. Therefore, the initial set of data products has been reduced using a fully consistent set of master calibrations processed with the same pipeline version.

Wavelength scale. The GIRAFFE Medusa products are wavelength calibrated. The wavelength scale is *heliocentric*. The correction values that have been applied are stored in the header (HELICORR, BARICORR, GEOCORR, in km/s).

Telluric absorption. No correction for telluric absorption lines has been applied. Generally telluric correction is not a major concern for GIRAFFE observations. No dedicated telluric standard star observations exist. Users may want to visit the ESO *skytools* web page[13] for appropriate tools.

Table 2. Set of master calibrations used for data reduction

Type (pro.catg)	name (first part)	Mandatory*/ optional	content
MASTER_BIAS	GI_MBIA	optional	created from 5 raw bias frames; removes bias level and bias structure.
MASTER_DARK	GI_MDRK	optional (but always provided for data before 2008-04)	created from 3 raw dark frames; removes or reduces diode glow (dark current is negligible).
FF_LOCWIDTH, FF_LOCCENTROID, FF_EXTSPECTRA, FF_EXTERRORS	GI_PLOC, GI_PLOW, GI_PFEX, GI_PFEE	mandatory	tables derived from flat field (3 input flat frames), with localization information for each fibre, and with extracted signal for pixel-to-pixel gain correction (incl. error)
DISPERSION_ SOLUTION	GI_PDIS	mandatory	Dispersion table derived from single arc frame
BAD_PIXEL_MAP	GI_PBPX	Optional (but always provided for data before 2008-04)	Static bad pixel map, corrects for bad columns
LINE_MASK	GI_GMSK	Optional	Static line mask, required for cross-correlation of SIMCAL data
GRATING_TABLE	GI_GRAT	mandatory	Static grating table
SLIT_GEOMETRY_ SETUP	GI_GSGS	mandatory	Static slit geometry table, describing

---

[13] http://www.eso.org/sci/software/pipelines/skytools/

Type (pro.catg)	name (first part)	Mandatory*/ optional	content
			fibre arrangement in the slit (Medusa1: 2003-08-28 and later; Medusa2: since start of operations)
OR			
SLIT_GEOMETRY	GI_GSGT	mandatory	Static slit geometry table, used for Medusa1 until 2003-08-27

* Mandatory: if missing, pipeline would fail; optional: not strictly required but always provided for science data products

Bad pixel map. We have used the same bad pixel map throughout the epoch of the old CCD. This is generally a reasonable strategy but we cannot exclude that in certain cases a custom bad pixel map would mask the affected bad columns in a better way. We have marked the potentially affected data products with a quality flag (see below).

Master darks. We have reduced the spectra with closest-in-time master darks (intended to correct for the diode glow on the old CCD) selected from a pool of master darks taken at about monthly intervals. In individual cases they still do not match the prevailing conditions for the science data very well, since the glow level was variable. We have marked the potentially affected spectra with a quality flag (see below).

Master calibration names and recipe parameters used for reduction. The product header contains a list of all used master calibrations, look for keys "HIERARCH ESO PRO REC1 CAL<n> NAME" and "... CATG", with the index n. The used pipeline parameters and their values are listed as "HIERARCH ESO PRO REC1 PARAM<n> NAME" and "... VALUE".

Products. The final GIRAFFE science data product in the binary spectroscopic data format combines information from the following 2D pipeline products:

- 2D extracted spectrum (de-biassed and if applicable dark-corrected, flat-fielded, extracted, wavelength-calibrated, rebinned): all fibres combined (SCIENCE, SKY, SIMCAL if used).
- Corresponding 2D error file.

The spectra are extracted by the pipeline from the 2D pipeline product, one spectrum per SCIENCE fibre. Each spectrum is a binary FITS table file, with the wavelength values as first column, then the extracted signal for the corresponding fibre, the corresponding error, and the SNR as ratio of the signal and the error (provided for convenience, it does not contain new information).

The following additional files are delivered as ancillary FITS files:
- The entire 2D extracted spectrum is delivered for quality checks.
- If SKY fibres exist, there is an extraction from the 2D file of all SKY fibres, with the product category 'ANCILLARY_MOSSKY'.
- If existing, the attached SIMCAL observations are added (see 'File Structure' below).

Finally, there are two overview plots added as ancillary non-FITS files to each product spectrum. One is a QC report, applicable to the whole set of products from the parent raw file, and one is a preview plot of the spectrum.

The spectra contain some header keywords that have been added by the pipeline. They are listed in the following table:

Table 3. FITS keywords added by the pipeline

parameter	values	meaning
OB related information:		
SM_VM	SM or VM	Data taken in Service Mode or Visitor Mode; VM data are less constrained in terms of OB properties; they have no user constraints defined and therefore no OB grades
OB_GRADE	A/B/C/D; X	Immediate grade given by night astronomer, considering

parameter	values	meaning
		ambient conditions checked against user constraints (VM data are formally graded X meaning 'unknown')
OB_COMM*n*	Free text	Any optional comments added by the night astronomer, together with the *approximate* UT hh:mm (truncated after 200 characters). The comments are put into *n* keywords with maximum length of 68 chars for the comment. [Note that OBs might have been executed several times during the night, with or without comments. In those cases the user should always carefully check that the listed comment applies to the data product with the closest *previous* timestamp.]
QC related information:		
QCFLAG	e.g. 000000100	QC flag composed of 9 bits, see Table 4.
QC_COMM*n*	Free text	Automatically added comment if a quality issue is discovered by the processing system (human-provided comments also exist but are very rare): • SCIENCE SIMCAL (a stand-alone SIMCAL observation processed as SCIENCE) • No SKY fibres defined • More than 1000 pixels saturated The comments are put into *n* keywords with maximum length of 68 chars for the comment.

Rejected or failed processing. The GIRAFFE pipeline evaluates the binary OZPOZ table (FITS extension of the raw file), column TYPE, value M (for "Medusa science"). The TYPE values (could also be S for "sky") are user-designed. Products are created only for fibre type M. If a user has not assigned a science fibre (by mistake or by intent), there will be no product.

If no certified master calibrations exist, no processing can be done. This was a very rare event.

In cases of quality issues (high number of saturated pixels, time or temperature mismatch of calibrations) we have flagged those data in the QC flag but leave the final assessment to the user (see Section 'Data Quality').

Data Quality

Master calibrations. All used master calibrations have been quality-reviewed and certified at the time of acquisition, as part of the closed QC loop with the Observatory which also includes trending[14]. The set of master calibrations for the initial period 2003-04 until 2005-06 has been completely reprocessed. As part of the certification process there is always a scoring to bring non-compliant behavior of the calibrations to the attention of the QC scientist. All these cases have been handled as part of the certification procedure. Hence there is reasonable evidence that the master calibrations catch all instrument properties, as relevant for the reduction, correctly and completely.

The most important parameters for the quality of the products are the SNR of the master flats, and the *rms* of the dispersion solution. The SNR of the master flats was always high enough (both for the old and the new CCD) to be dominated by the fixed-pattern (gain) noise, which is important to not compromise the SNR of the science data.

---

[14] Check out for more under the GIRAFFE link of http://www.eso.org/qc/ALL/daily_qc1.html.

The *rms* of the wavelength dispersion (Figure 2) was slightly higher in the first few years than later, due to a slight degradation of the focus over the years. This effect is entirely instrumental and not due to evolving quality of the pipeline. *Actually the quality of the master calibrations has been very stable over the years[15].* The red bullets represent data from the reprocessing of early master calibrations (2003-2005) and demonstrate their homogeneous quality despite using different pipeline versions.

 SNR. There is a column "SNR" in the spectral product that is calculated from the signal and the corresponding error. It has no independent information but is calculated from the two columns FLUX_REDUCED and ERROR_REDUCED and is provided for convenience. Its mean value across the spectrum is written into the header as key SNR.
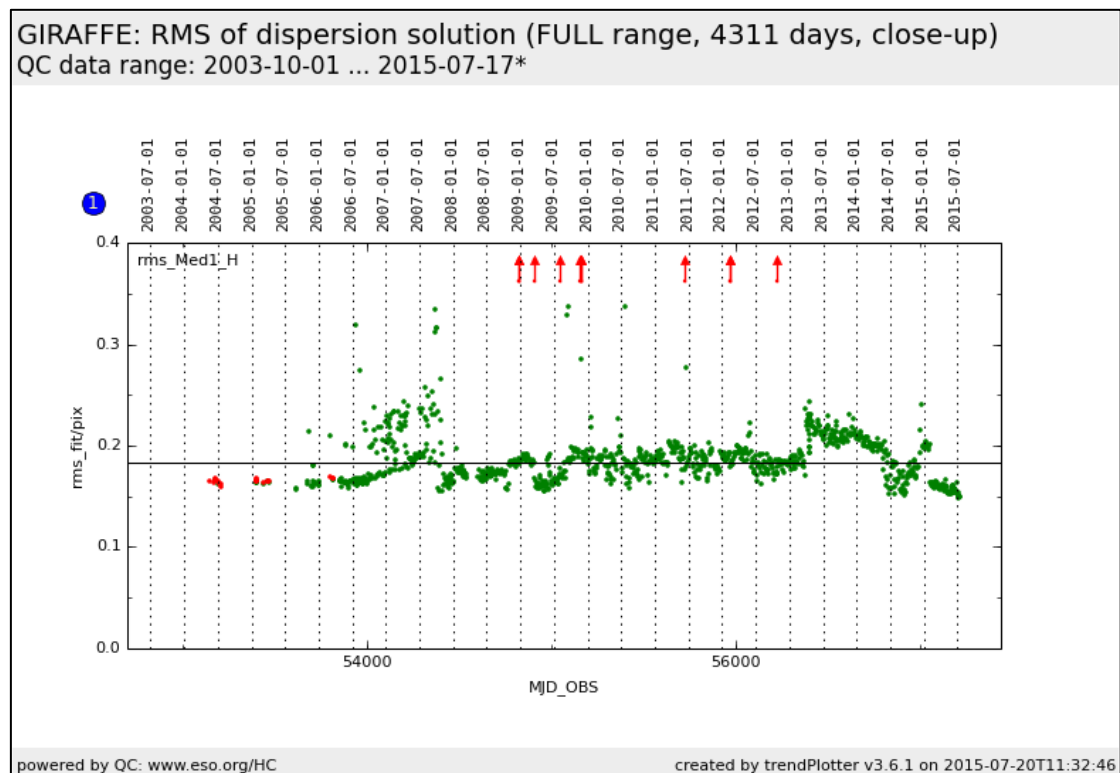
Figure 2. Performance of the RMS of the dispersion solution for the reference setting Medusa1 H525.8B. This Health Check plot is based on ARCLAMP calibrations taken every 3rd day.

---

[15] See http://www.eso.org/qc/GIRAFFE/reports/HEALTH/trend_report_ARC_RMS_HC.html for the trending plots; click "FULL" for the entire history.
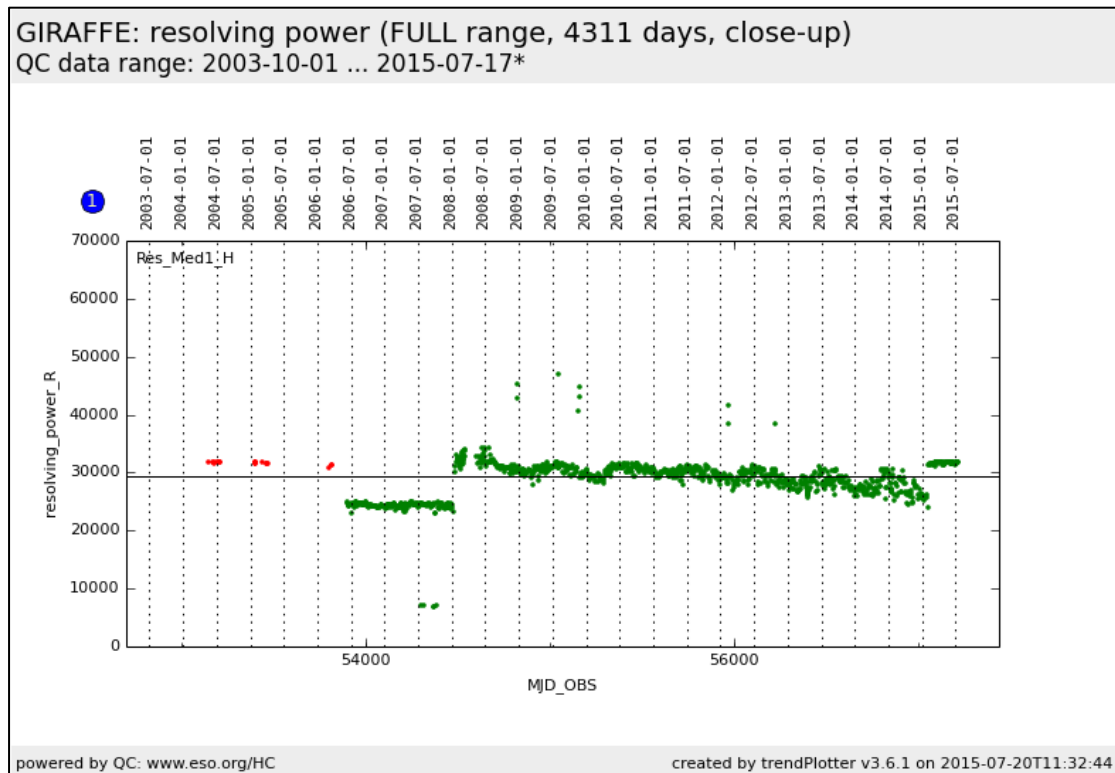
Figure 3. Resolving power measured from ARCLAMP frames for the reference setting Medusa1 H525.8B.

SPEC_RES. The resolving power is derived from the arclamp calibrations for the same setting as the science data. The actual resolution of the science data has not been measured directly but very likely is the same as for the arclamp calibration, since all spatial information (position, FWHM of the source due to seeing) is lost due to scrambling within the fibre. The header key SPEC_RES has been filled from a look-up table containing mean values for each setting, derived from the HC data. The stability of SPEC_RES is illustrated in Figure 3 for a reference setting. As in the previous figure, red bullets mark the initial 2003-2005 dataset from the reprocessing.

QC flag. The header key "QCFLAG" in the GIRAFFE products contains a quality flag. It is composed of nine binary bits which are listed in Table 4. For each bit, the value 0 means "no concern".

The final QC flag is composed of all components. For instance, 000000000 is a pristine product. 010000000 or 100000000 or 110000000 describe products that have been processed with an ARCLAMP calibration product that deviates in ambient temperature, or time, by more than a nominal value (1.5 degrees or 1 day). The motivation for this comes from the temperature sensitivity of the grating which moves slightly in both cross-dispersion and dispersion direction with temperature[16]. These thermal drifts are minimized by the daily calibration plan which foresees these calibrations to be taken right after the science data, in daytime. Occasionally this rule cannot be satisfied, and the score #2 (deltaTime) flags this. This occurs only rarely, and is likely to be uncritical, but in principle the wavelength scale could be less accurate in those cases.

The QC flag 001000000 marks an issue with the number of saturated pixels. In practise, saturated pixels are not intrinsic to the spectrum but arise from cosmics, or sky emission lines. Flag #4 is for process control only. Flag #5 indicates if no SKY fibre is available (or not marked as such by the observer). In practice this happens rarely.

Flag #6, if 0, marks an OZPOZ file that has attached SIMCAL file(s). Although it is very normal for an OZPOZ file to have *no* SIMCAL file attached, it was decided to encode flag #6 in this way since with an attached SIMCAL, the observation can have in principle the highest possible wavelength accuracy. It follows that a value 1 *does not imply any issue with the data*. The value 0 is only possible

---

[16] Check out http://www.eso.org/qc/GIRAFFE/reports/HEALTH/trend_report_STABILITY_HC.html for more.

for an OZPOZ file.

Bit	Content (if YES, value is 0, otherwise 1)	Motivation
#1 –deltaTemp	abs(deltaTemp) below 1.5 degree? (deltaTemp=difference between ambient temperature for science and for arclamp calibration)	increased probability of calibration mismatches between science and arclamp calibrations (rarely violated)
#2 - deltaTime	abs(deltaTime) below 1 day? (deltaTime=difference between acquisition time of science and arclamp calibration)	usually daytime calibrations come within 0.5 days; if more than a day difference, probability is higher that shifts have occurred for arclamp data (impacts wavelength scale) (rarely violated)
#3 – saturated pixels	number of saturated pixels below 1000?	not critical if due to cosmics; never due to saturation of larger spectral portions
#4 – number of science fibres	≥ 10?	relevant for process control only
#5 – number of sky fibres	≥ 1	flag cases where no sky fibre is available
#6 – flag for simcal attached	attached SIMCAL file(s) exist(s)?	0 is possible for OZPOZ files only; if attached SIMCAL file exists, accuracy of wavelength scale can be checked and corrected to highest possible level
#7 – OBJECT type	OBJECT,OzPoz?	if yes, that product does not suffer from contamination by SIMCAL fibre signal; if no (value=1), be aware of possible contamination
#8 – bad column flag	product unaffected by bad column (old CCD)?	if no (value=1), be careful with the signal, it could have wrong offsets
#9 – glow flag	product unaffected by CCD glow (old CCD)?	if no (value=1), be careful with the signal; the master_dark could have over/under-corrected the glow

Table 4. QC flags. Flags #1-#7 apply to all products from the same raw file in the same way. Flags #8 and #9 apply to individual fibres (could be 1 for some products and 0 for others), depending on their position on the CCD. These flags have a meaning only for data before 2008-05 (old CCD).

Flag #7, if 0, marks that the file itself is from an OZPOZ observation. The SIMCAL fibre signal in the neighbourhood of emission lines could in principle cause some cross-talk, or even saturation in extreme cases, for the SCIENCE fibres. Therefore a SIMCAL product is always marked by the value 1, to make the user aware of potential issues[17]. In combination, the flags #6 and #7 can have the values 000000000 (OZPOZ file, with SIMCAL file attached); 000001000 (OZPOZ file, no SIMCAL attached); or 000011000 (SIMCAL science file). Of course all other flags could be 0 or 1.

While flags #1-#7 apply to all products from the same parent raw file in the same way, flags #8 and #9 may vary among them. They are applicable to data from the old CCD. Flag #8 marks products from a fibre that is affected by the bad columns, and flag #9 is for fibres that are affected by the CCD glow.

The cleaning of cosmics is not indicated in the QC flags since it has been applied to all products of single exposures.

Science products. The pipeline processing of the science data is done automatically, with some internal quality control, monitoring:
- the quality of the associations of calibration data (checking that the master calibrations

---

[17] Therefore some observers have chosen a long OZPOZ exposure and one or more short SIMCAL exposures in the same setting and same OB.

used are applicable to the science data, flagged in the QCFLAG);
- flags for the number of saturated pixels, and for potential issues from the data reduction (see above, Table 4);
- on-demand QC reports and quick-look overviews.

This information has largely been used to improve and fine-tune the reduction process. An individual one-by-one inspection of the products has not been done. The QC reports and spectral overviews are delivered along with the spectral products.

Previews. The preview plots have been originally developed as quick-look plots for process quality control. It was felt that they might also be useful to the archive user. They are delivered as ancillary files along with the products. There are two plots:

1. the main QC plot (Figure 4), one per dataset (the same for all products from the same raw file);
2. the preview plot (Figure 5), one per product.

In the upper panel of box 5, the sky emission lines are marked.[18]

Process quality control. The quality of the data reduction is monitored with quality control (QC) parameters, which are stored in a database. The database is publicly accessible through a browser and a plotter interface[19].

QC parameters are used to monitor the reduction quality. The most important check is the "SNR versus signal" control plot (the signal being expressed as the mean across the entire reduced spectrum, and the SNR being the mean over the entire spectral range).

Figure 6 shows data points from the reference setting H665.0 which has been selected because it has a rich selection of products. Only data for the brightest fibre are displayed, to avoid confusion. Two data sets are displayed: one (in red) is for the data from 2014 only, the other (in blue) is for products from 2009-09 on. (From that date on, the pipeline version `giraf/2.8` was used for the calibration data processing. Previous versions had a bug with the computation of the error and hence with the SNR.)

---

[18] Taken from http://www.eso.org/qc/UVES/pipeline/sky_spectrum.html (based on Hanuschik R.W., A&A 407, 1157 [2003]).

[19] Browser: http://archive.eso.org/qc1/qc1_cgi?action=qc1_browse_table&table=giraffe_science_public
    Plotter: http://archive.eso.org/qc1/qc1_cgi?action=qc1_plot_table&table=giraffe_science_public
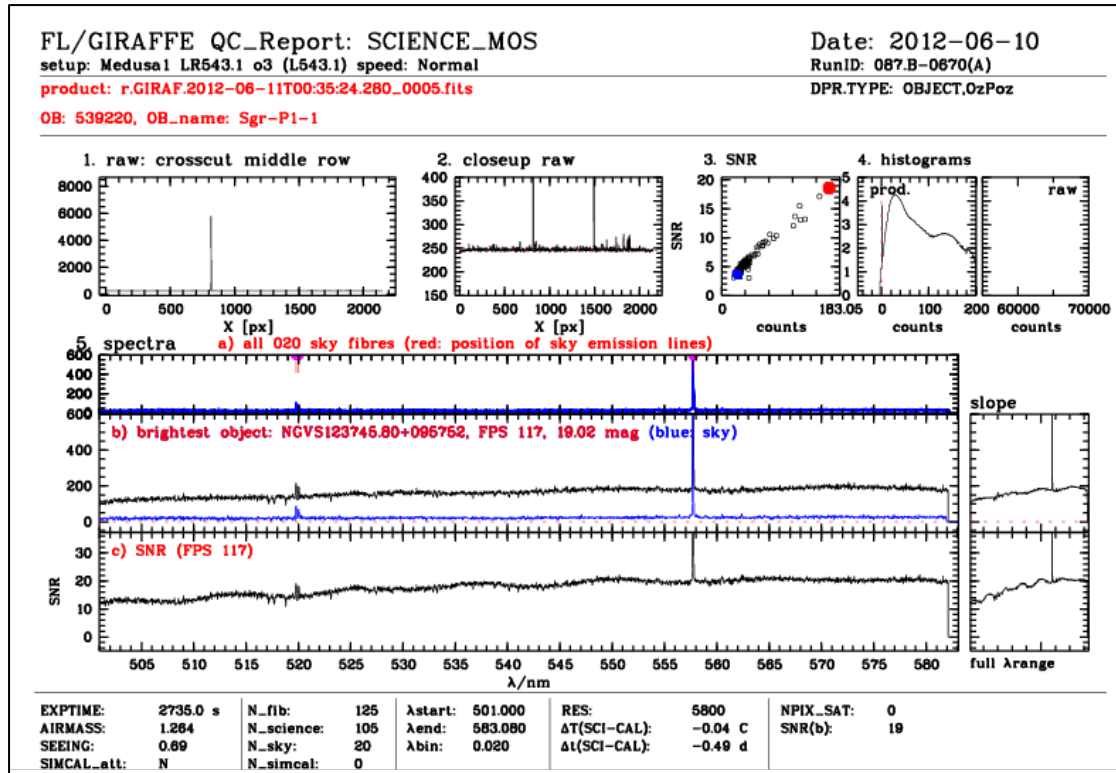
Figure 4. Main QC plot for the whole dataset. Description:

- on top: acquisition date, setup information, run ID; product ID, DPR.TYPE, OB ID, OB name;
- boxes 1 and 2: crosscuts from the raw file;
- box 3: SNR vs. counts for all fibres (SCIENCE in black, SKY in blue, brightest SCIENCE in red);
- box 4: histograms for 2D product; *4a*: product file, the histogram focussing on the region around the background, counts are on a logarithmic scale; *4b*: histogram for the parent raw file, focussing on the region around saturation, 65000 ADUs;
- boxes 5: spectral overviews; *top*: all available sky spectra plotted in blue; *middle*: signal from *brightest fibre*, in blue the last sky fibre; *bottom*: SNR plot for *brightest fibre*; sky lines are plotted from http://www.eso.org/qc/UVES/pipeline/sky_spectrum.html;
- at bottom: a set of QC parameters applicable to the parent raw file and the 2D products, including the flag for attached SIMCAL (Y/N), number of fibres (total, science, sky, simcal), spectral coordinates in nm, resolving power, deltaTime and deltaTemp values, number of saturated pixels, and SNR of brightest spectrum.
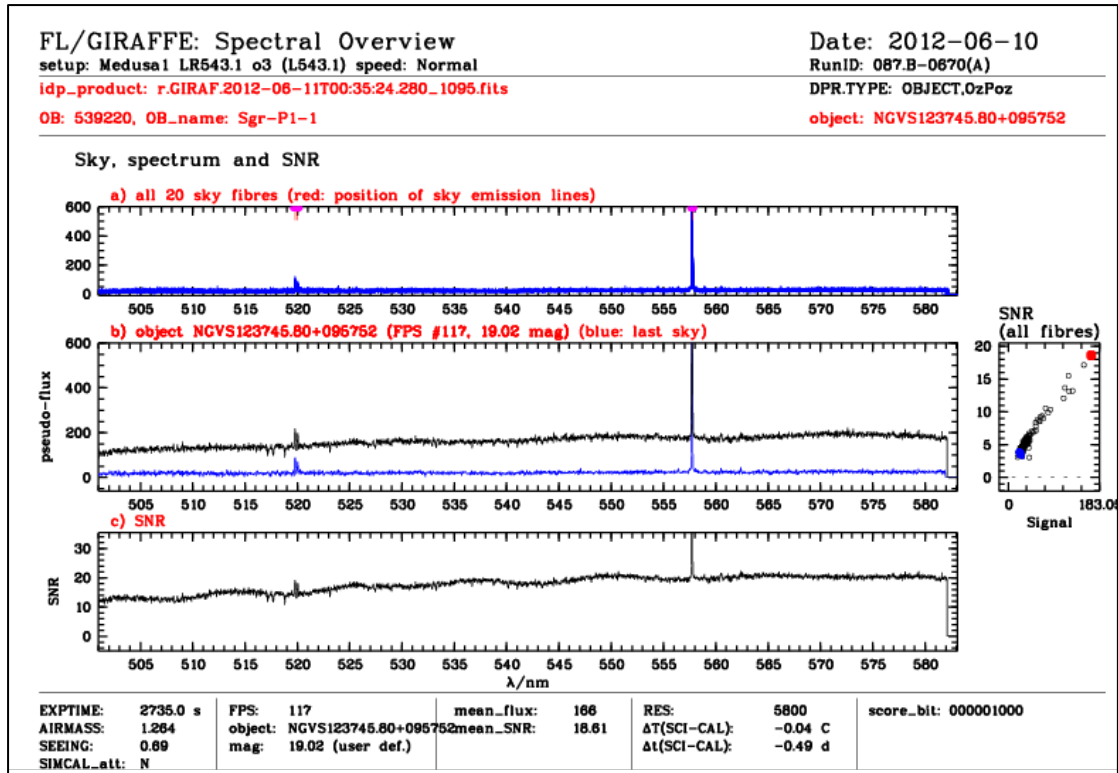
FL/GIRAFFE: Spectral Overview
setup: Medusa1 LR543.1 o3 (L543.1) speed: Normal
idp_product: r.GIRAF.2012-06-11T00:35:24.280_1095.fits
OB: 539220, OB_name: Sgr-P1-1

Date: 2012-06-10
RunID: 087.B-0670(A)
DPR.TYPE: OBJECT,OzPoz
object: NGVS123745.80+095752

Sky, spectrum and SNR

a) all 20 sky fibres (red: position of sky emission lines)

b) object NGVS123745.80+095752 (FPS #117, 19.02 mag) (blue: last sky)

SNR (all fibres)

c) SNR

EXPTIME:	2735.0 s	FPS:	117	mean_flux:	166	RES:	5800	score_bit: 000001000
AIRMASS:	1.264	object:	NGVS123745.80+095752	mean_SNR:	18.61	ΔT(SCI-CAL):	-0.04 C	
SEEING:	0.69	mag:	19.02 (user def.)			Δt(SCI-CAL):	-0.49 d	
SIMCAL_att:	N							

Figure 5. Preview plot, one per spectral product. Description:
- on top: acquisition date, setup information, run ID; product ID, DPR.TYPE, OB ID, OB name, target name;
- panel 1a: all sky fibres (same plot as #5 a) of the previous figure);
- panel 1b: the spectrum, plus the sky signal (last fibre) for comparison;
- panel 1c: the SNR of the spectrum;
- also displayed: the SNR plot for all spectra of the same raw file, with the current one highlighted in red.
- At bottom: a set of QC parameters applicable to this spectrum, including the score bits (QC flag), the flag for attached SIMCAL (Y/N), mean flux and SNR, resolving power, deltaTime and deltaTemp values, and object properties like FPS (fibre index), object name, and user defined magnitude (if available).
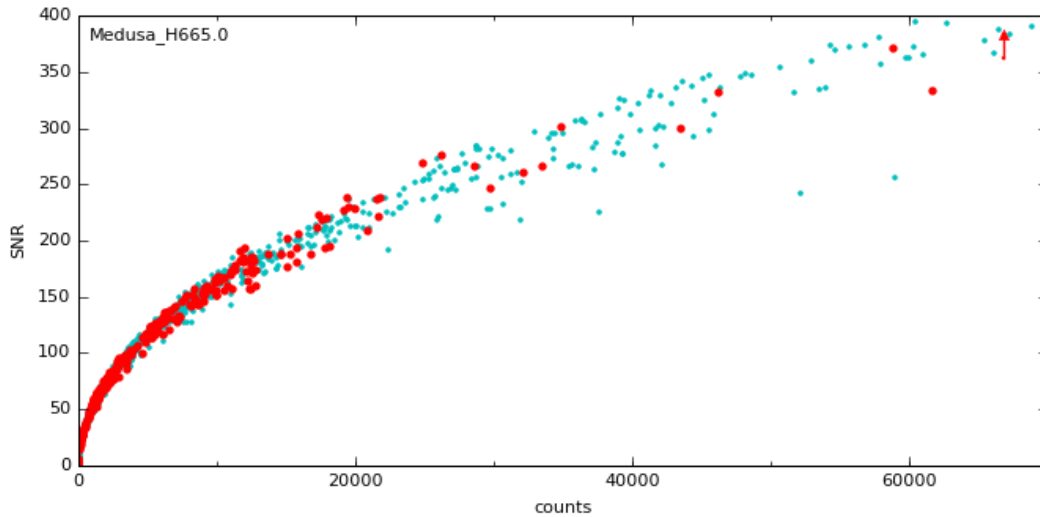
Figure 6. SNR vs signal. All data points are from the same setting, H665.0, both Medusa fibre systems. Only the signal from the brightest fibre is displayed here. Data points in red refer to the year 2014. The other data points cover the range 2009-09 and later.

This figure shows the square-root law expected if the data quality is limited by photon shot noise. Out to SNR values as high as 400 there is no flattening observed, as one might otherwise expect if the SNR is limited by imperfect data reduction, e.g. by imperfect flat-fielding. Another effect is visible in this figure, namely the stability of the reduction scheme and of the instrument. There is no significant difference between the entire dataset displayed in this figure (blue points) and the data from 2014 only (red points).

The highest values for SNR found are in excess of 700 (Figure 7).
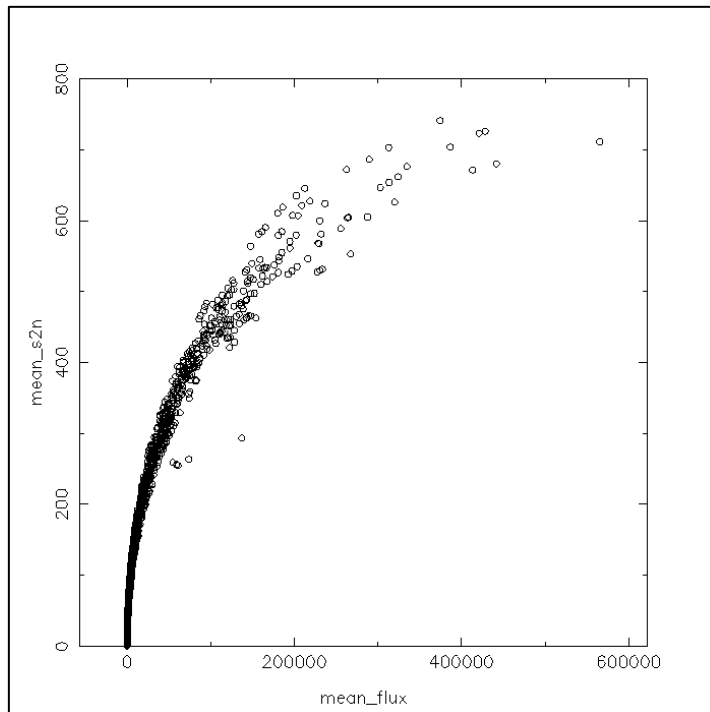
Figure 7. All Medusa1, HR data (brightest fibre), data from 2010 until end of 2014. Maximum SNR found is in excess of 700.

Known features and issues

Issues.

Old CCD (data acquired until 2008-04), dispersion solution. All care has been applied to process the data with the proper master calibrations. Since visual checks were unaffordable, we cannot entirely exclude that there are minor differential shifts across the fibre dispersion solution. We recommend to inspect the GI_SRBS file with an image browser. Any sky emission or absorption lines should be perfectly straight across all fibres.

Old CCD, bad columns. The old CCD has two bad columns that cover part of the CCD and affect mostly the signal in fibre FPS24. Depending on the spectral format, the impact may be strong or negligible. A bad pixel table has been used to mask out these two bad columns. But the extent and strength of the artefact (non-linear or negative counts) may have been variable and hence the masking imperfect. We have marked those spectra with the bad column flag #8. Please check those spectra carefully, their count level might be off (with the shape of spectral features being still correct).

Old CCD, glow correction. With the monthly master darks, the strength of the glow may have been over/under-subtracted occasionally. The affected area is the upper right part of the CCD which translates into a fibre index between approximately FPS109 and FPS124 and affects about the last quarter of the spectrum at most. Such data are marked in the previews and scores, check your preview. If the slope is flat, all is ok. If there is an upwards or downwards slope, the spectrum is affected by the improper glow correction. The glow has low spatial frequencies only, hence neither the shape of spectral lines nor their RV are affected. If a proper slope is important, correcting with a low-frequency fit to the slope is likely to be sufficient. In more severe cases a full reprocessing is required, with appropriate scaling of master dark.

Old CCD, fringing in red setups. The old CCD has a significant fringing in the red settings. The correction by the flat-fielding is very good, as is visible in Figure 8.

New CCD (since 2008-05). For the period 2011-11-20 until 2011-12-22, and the fibre system Medusa1, one fibre (FPS27) was enabled that was never enabled before or after. There is no precise configuration available for the pipeline to process that fibre, and it was decided to not process those datasets into data products. A total of 13 input raw GIRAFFE files is affected. Medusa2 data are not affected by this issue, and their products have been processed.

General (old or new CCD).
SKY spectra to correct back. If you need SKY subtraction and want to use the signal registered in the SKY fibres, please keep in mind that you may have to correct back the correction to the heliocentric rest frame that has been applied to the science spectrum (header key HELICORR).

[The following issue is listed for historical reasons. It has been solved by now and is not applicable to any of the currently visible spectra.
2D SKY file. That ancillary product file is intended to contain the signal from all SKY fibres (if any), with the goal to make the selection and handling of the SKY signals easier for the user. However, due to a pipeline bug, this file contains the signal from the SKY fibres and from the SIMCAL fibres if the
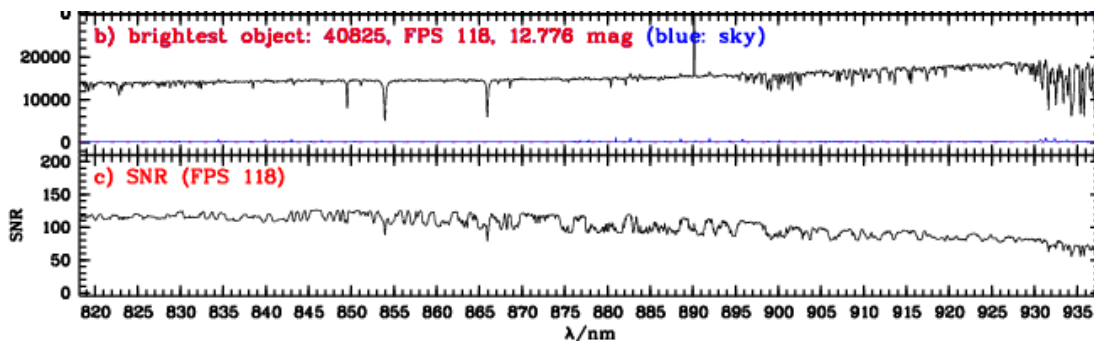

Figure 8. The lower panel (the SNR plot from the spectral preview plot) shows the fringing pattern typical of the red settings observed with the old CCD. The master flat shows exactly the same pattern which then cancels out in the extracted signal (top panel).

raw file contains SIMCAL fibres (OBJECT,SimCal). The five SIMCAL fibres can easily be identified by their typical emission lines. Of course they should be ignored for sky background fits and sky removal. This bug has been fixed with pipeline version giraf-2.14.2, for data observed on 2015-08-01 and later.]

For cases with no observer-defined SKY fibres, this file contains SIMCAL fibres only and should be ignored. Naturally, for OBJECT,OzPoz files (the typical case) there is no such issue.

Last fibre(s). The pipeline extracts only fibre signals falling entirely onto the chip, meaning out to FPS135 at most. Any signal in incomplete fibres is ignored.

Sky/science type in FPOSS. The type is configured by the observer. The pipeline extracts spectra for fibres assigned type "M" (science), and collects fibres assigned type "S" into the ANCILLARY_MOSSKY file. If the observer has confused sky and science in one or more fibres, the pipeline cannot correct for this. See Figure 9 for a possible example.

Possibly wrong OB grades in FITS keywords. If in Service Mode an OB is executed twice (or in general more than once) during the same night, this is typically done because during the first attempt the ambient conditions unexpectedly degraded, resulting in a C grade. A second attempt might have been undertaken, often resulting in a B or A grade. In such case, the grade (as recorded in the OB_GRADE FITS keyword) is always taken from the last execution. This is wrong in the typical case of a grade C-A pair and due to a bug. The grade as retrieved from the nightlog tool NLT is correct. For repetitions in different nights this bug does not appear, and the quoted grades are correct.

Sky emission lines. Under unfavourable conditions, sky lines are sometimes confused with cosmics by the cleaning process and are removed by mistake, resulting in a negative signal if a sky fibre signal is subtracted.

Features
Attached flats. Occasionally observers have acquired single flat field exposures during the night, attached to the science template ("attached flats"). These have been taken with the Nasmyth flat screen. Their purpose is very user specific and their quality is lower than that of the daytime flats provided by the calibration plan. Their main purpose is likely quality control, to have localization information very close in time with the science data. They have been ignored for this release. Users interested in these calibrations should query the archive.

Calculated SNR too low before 2009-09-01. Pipeline versions earlier than `giraf/2.8` created master flats with an error computation that was too high. This error propagates to the computed SNR of the products which is too low, by roughly 20-30%. The error does not affect the quality of the data but only the values listed in the ERROR and the SNR columns of the data products. The header key SNR is also affected.

QC preview plots. For fibres with indexes between FPS109 and FPS124 there is the message "Check for residual glow" in the preview plot. By mistake, this note is displayed not only for spectra taken with the old CCD (where it is correct) but also for many spectra taken with the new CCD (where it is a false alert and should be ignored).

RV standard star observations in stand-alone SIMCALs. With our strategy for the suppression of stand-alone SIMCAL observations with short exposure time, it might have happened (rarely) that such data represent observations of RV standard stars. We could not find a strategy to identify them safely. In those cases users should search the raw data archive for the same dates as the data products, download and process them. Such cases are confined to VM nights, to our best knowledge.

Parallel FLAMES/UVES observations. The FLAMES spectroscopic facility on UT2 is able to feed up to 8 science fibres into the UVES spectrograph, in parallel to the FLAMES/GIRAFFE observations, as part of the same OB and observing field. Those UVES data are not processed and are not part of this data release.

Find more documentation about features and issues on-line:
http://www.eso.org/qc/PHOENIX/GIRAFFE/processing.html.

Data Format

File Types

The primary GIRAFFE Echelle product is the extracted fibre spectrum, in binary spectroscopic data format:

ORIGFILE names starting with	product category HIERARCH.ESO. PRO.CATG	format	how many?	which fibres	Description
GI_SIDP	SCIENCE_RBNSPEC_IDP	binary table, signal plus error plus SNR	1	SCIENCE	main data product; extracted, rebinned, heliocentric scale, identified by FPS index

There are *N* products per parent raw file, if the raw file contains *N* science fibres (marked as type "M" in the OzPoz table attached to the raw file).

Each product has one or more ancillary FITS files delivered with it:

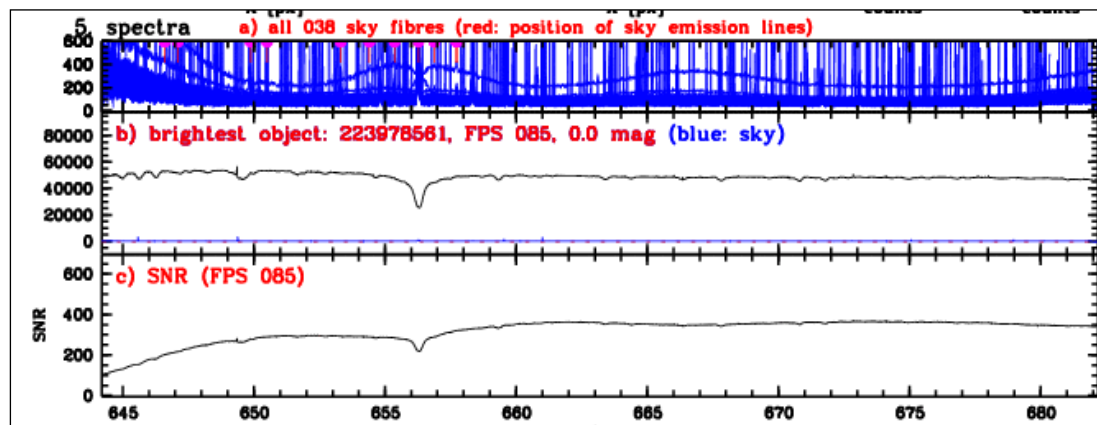| ORIGFILE	product category	format	how many?	which fibres	Description

Figure 9. The upper row in the preview plots displays all sky signals. Occasionally it shows a sky signal with signature from a point source. This might indicate a target confusion.

names start-ing with	HIER-ARCH.ESO.PRO.CATG				
GI_SRBS	SCIENCE_RBNSPECTRA	2D image, plus FIBRE_SETUP table	1	all: SKY, SCIENCE, SIMCAL	extracted, rebinned
GI_SSKY	ANCILLARY_MOSSKY	2D image, plus FIBRE_SETUP table; SKY fibres extracted for convenience from GI_SRBS	1 (in rare cases: none)	SKY only (due to a bug, it also contains the SIMCAL fibres for OBJECT,SimCal files, until 2015-07-31)	extracted, re-binned; useful for determination of SKY to be sub-tracted from 1D data product
GI_SSIM	SIMCAL_RBNSPECTRA	2D image, plus FIBRE_SETUP table, including residual shifts WLRES; origin: attached exposures, same OB	0, 1 or more	all : SKY, [SCIENCE*], SIMCAL	

*Formally, the fibres in this product also have the type 'M' but due to the short exposure time contain only noise.

The list of ancillary FITS files is identical for all products from the same raw file.

Furthermore the following non-FITS files are delivered with each spectrum:

ORIGFILE names starting with	PRODCATG	format	how many?	Description
r.GIRAF…gif	ANCILLARY.PREVIEW	gif file	1	see Figure 4; labelled "QC report"
r.GIRAF…gif	ANCILLARY.PREVIEW	gif file	1	see Figure 5; labelled "Spectral over-view"

The following naming convention applies to the ORIGFILE product: e.g. the name
`GI_SIDP_964735_2013-05-07T01:58:20.737_F122_Med1_H504.8_o11.fits`
has the components:

GI	SIDP	964735	2013-05-07T01:58:20.737	F122	Med1_H504.8_o11.fits
GIRAFFE	product type (S stands for science)	OB ID	timestamp of first raw file archival	fibre with index FPS122	setup string: Med1 for Medusa1; H504.8_o11 for central wavelength and grating order

The ancillary files have the following ORIGFILE names:

Table 5. Naming conventions of ANCILLARY files

type	example	rule
ANCILLARY.MOSSPECTRA	GI_SRBS_964735_2013-05-07T01:58:20.737_Med1_H504.8_o11.fits	Same as primary product except for fibre index, name starting with GI_SRBS
ANCILLARY.MOSSPECTRA.SKY	GI_SSKY_964735_2013-05-07T01:58:20.737_Med1_H504.8_o11.fits	Same as GI_SRBS, name starting with GI_SSKY
ANCILLARY.MOSSPECTRA.CALSIM	GI_SSIM_964735_2013-05-06T23:37:37.133_Med1_H504.8_011.fits	Same as GI_SRBS, starting with GI_SSIM
ANCILLARY.MOSSPECTRA.PREVIEW	r.GIRAF.2013-05-07T01:58:20.737_0005.fits.gif	Technical filename of the GI_SRBS file, with 0005.fits.gif appended
ANCILLARY.PREVIEW	r.GIRAF.2013-05-07T01:58:20.737_0005.fits_f122.gif	same name, plus fibre index

The user may want to read the ORIGFILE header key and rename the archive-delivered FITS files.

File relations

Pipeline view: the following figures illustrate the relation between the 1D data products and the ancillary files, as seen from the product creation (pipeline) process.
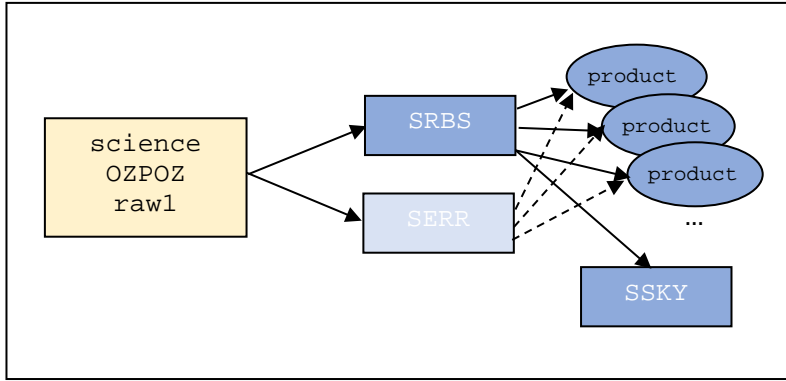
Figure 10: Pipeline processing scheme for GIRAFFE products, OZPOZ case. The main pipeline product of the first step is the SRBS file, with the errors in the SERR file. In the second step, the individual products are extracted fibre by fibre from the science fibres in SRBS; the corresponding errors are extracted from SERR. All sky fibres are extracted from SRBS into the 2D file SSKY. All of these products, except for the SERR file which is considered obsolete, are delivered with the selected products.
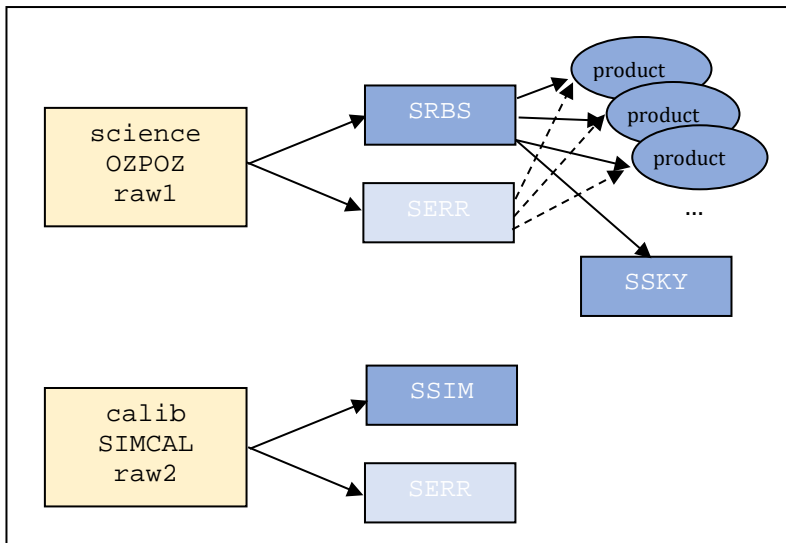

Figure 11. Same, for the OZPOZ plus SIMCAL case. The SSIM file(s), if measured within the same OB and night, are delivered with the products.
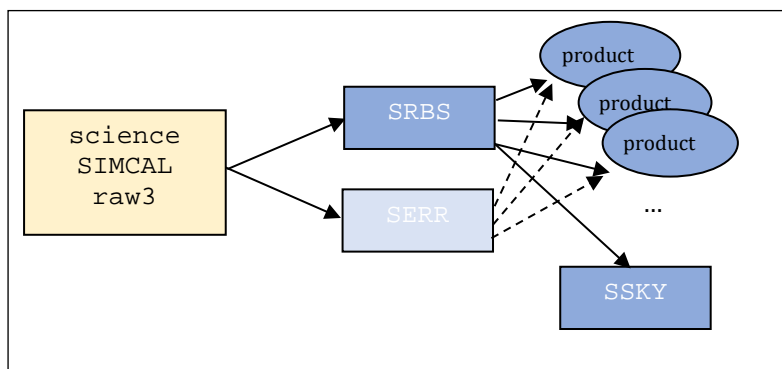
Figure 12. Same, for the SimCal SCIENCE case.

Science data product view (as seen from the delivered dataset):
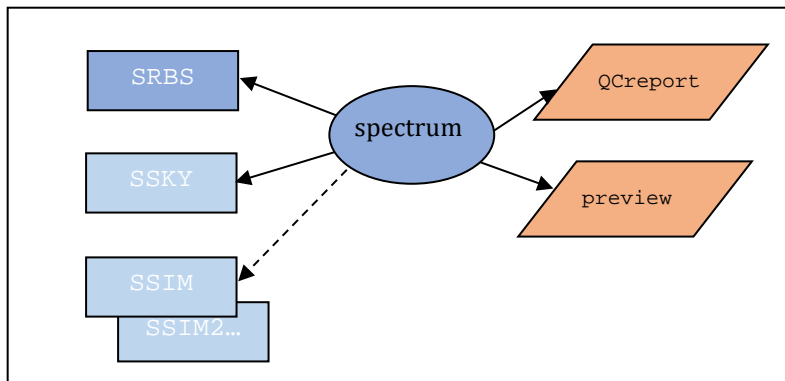

Figure 13. Product scheme as seen from the delivered spectrum product.

File structure

The primary GIRAFFE product GI_SIDP comes as binary FITS table in multi-column format. The columns are labeled as follows:

Table 6. Internal structure of the GIRAFFE 1D spectra.

column	label	content
#1	WAVE	wavelength in nm, corrected to heliocentric system
#2	FLUX_REDUCED	extracted, wavelength-calibrated but not fluxed SCIENCE signal in counts
#3	ERR_REDUCED	Error of FLUX_REDUCED (same units)
#4	SNR	Signal-to-noise ratio = FLUX_REDUCED/ERR_REDUCED

The SNR column is provided for convenience.

The ancillary FITS files (no matter which type) have a 2D image format, with the fibre index as horizontal axis and the heliocentric wavelength as vertical axis. Since the correction from topocentric to heliocentric is slightly different for each fibre (it depends on target coordinates), it is *not* applied to any of the 2D products, but to each 1D spectrum individually. The applied correction is stored in the product header (as keys HELIOCOR and GEOCORR; listed is also the barycentric correction BARYCORR). Each 2D ancillary file also has the *fibre setup* table, merged by the pipeline from the two binary raw file tables *ozpoz table* (user-provided FPOSS target parameters) and *fibre table* (technical fibre parameters). Here the user finds information connecting fibre properties and target properties, like the fibre types and indexes, target coordinates, names and estimated magnitudes etc.

fibre type	content	
S	sky fibre	assigned by observer; contained in GI_SRBS and GI_SSKY
M	science	assigned by observer; contained in GI_SRBS and GI_SIDP
<none>	CALSIM	simultaneous calibration fibre

File size

The GIRAFFE science data products have about 0.2 MB size. The ancillary 2D files come with several MB size (depending on number of fibres). Files are always uncompressed.

Part II: OB-stacked data products

Data Selection

The OB-stacked data products are created from single-exposure data products. All selection rules for single data products apply (see Part I). The following additional rules apply for the OB-stacks:

- There must be more than one exposure per OB execution[20].
- Type (DPR.TYPE) can be OBJECT,OzPoz or OBJECT,SimCal but not mixed; the frequently used pattern OzPoz | SimCal | OzPoz (with a short SimCal exposure and longer OzPoz exposures) is stacked into OzPoz spectra, with the SimCal exposure ignored.
- Only identical setups are combined.
- Only SCIENCE fibres are stacked (no SKY, no SIMCAL).

We have filtered out those files in an otherwise valid stack that deviate by more than 40% in exposure time from the average value. This cleans up stacks of the type "10|20|20 sec", or stacks with aborted exposures. We have not combined multiple executions of the same OB.

In Table 7 we have collected some typical patterns.

Table 7: Possible OzPoz/SimCal combinations in an OB and their data products

Within the same OB: DPR.TYPE …	Single products	Stacked products?
OZPOZ (single exposure)	one product file	No
OZPOZ | OZPOZ	N product files	Yes
SIMCAL | SIMCAL	N product files (if `exptime` is longer than 120 sec and central wavelength shorter than about 700nm, with a few exceptions)	Yes
OZPOZ | SIMCAL | OZPOZ	two OZPOZ products, with the SIMCAL file attached	OZPOZ yes

Pipeline Description

The stacking has been done using the High-level Data Reduction Library (HDRL) recipe `esotk-spectrum1d-combine`[21]. Find the pipeline version used for processing in the header of the product file, under "HIERARCH ESO PRO REC2 PIPE ID". The version for the initial datasets is `esotk-0.8.1`.

For N=2 input spectra, we have used a preprocessing step for the cleaning of cosmics. The algorithm first collapses the stack of input spectra by using the median to generate a master spectrum. Then it subtracts the master spectrum from each individual spectrum and derives the bad pixels on the residual spectra by thresholding, rejecting all pixels exceeding the thresholds. We have set the lower threshold to such a high value (1e6) that it has no effect, so we filter only for cosmics (positive outliers). While this works very well for N=2, we found that the preprocessing step rejects too many pixels for N>2, with a negative effect on the final SNR. Also we found that for N>2 the preprocessing is unnecessary because the main sigma clipping does a very good job for the suppression of cosmics then. Therefore we have skipped the preprocessing for N>2.

For the stacking we have used the sigma clipping method (`--collapse.method=SIGCLIP`). Input spectra are normalized to the flux level of the first input file (`--rescale-spectra=TRUE`), then the clipping is performed, using a kappa value of 7 (`--collapse.sigclip.kappa-low=7; --collapse.sigclip.kappa-high=7`).

---

[20] Strictly speaking we always refer to a single OB execution, with all candidate files having the same OBS.START key. If an OB has been executed several times, no matter whether in sequence or not, it has created several datasets which are not stacked together.

[21] Currently an internal procedure, not publicly available.

This stacking strategy provides the best results for the relatively small number of input files used by the observers for GIRAFFE stacking (often 2, rarely beyond 4). The main benefits of the stacking are the improved SNR and the suppression of cosmics. [Note that cosmics suppression is now also available for single-exposure data products, through a pre-processing cleaning algorithm based on *PyCosmic*, see Part I.]

We have applied a filtering rule for those cases where the exposure time of the stack candidates was different. This typically happens when the last exposure in a template is aborted. Accepting such candidates for a stack would result in adding more noise than signal. We have rejected those candidates when they differed by more than 40% from the average exposure times of the input files. Such cases were rare.

There are cases when – despite formally the same exposure time - the signal level in the input files differed strongly, e.g. because of degrading extinction. We have not rejected those cases automatically, but we have marked them by a score called mean_dev which is 1 if the mean fluxes differ by more than 20%.

Data Quality

SNR. The stacked products have a column "SNR_REDUCED" that is calculated from the signal "FLUX_REDUCED" and the corresponding error "ERR_REDUCED". Its mean value across the spectrum is written into the header as key SNR. There is also the column "CONTRIB_REDUCED" that lists the number of contributing pixels in every wavelength bin. If lower than the number of input spectra, the signal of one or more input files has been rejected, e.g. because of a cosmic or a noise outlier. The average number of contributing spectra is written into the header of the stacked product as key "CONTRIB". This number, normalized by the number of input spectra, is also scored as contrib which gets the value 1 if the relative contribution is less than 0.95 (5% loss).

QC flag. Like for the single spectra, the header key "QCFLAG" in the GIRAFFE stack products contains a quality flag. It is composed of nine binary bits inherited from the single spectra (Table 4), and of two added bits, mean_dev and contrib. The value 0 means "no concern". Inheritance of the single spectra bits means that they are set to 0 if they are 0 for all input files, and to 1 if at least one input file has a value 1. The bit mean_dev is based on the mean flux differences of all spectra of the OB stack, since it is reasonable to assume that it is due to extinction variations and hence applicable to all spectra of the OB stack in the same way. The bit contrib is based on the average of the CONTRIB_REDUCED column of the individual stacked spectrum. It is a property of the reduction process and might differ for each spectrum in the OB stack.

Bit	Content (if YES, value is 0, otherwise 1)	Motivation
#1-#9 propagated from the bits of the single spectra, as defined in Table 4; set to 0 if they are 0 *for all input files*, otherwise set to 1.		
#10 – mean_dev	Mean deviation of fluxes below 20%	All input spectra are normalized to the first one; if they differ too much in their flux levels, this might indicate a low-quality stack since too much noise has been added.
#11 - contrib	Maximum of 5% of the pixels lost in stacking	Pixels are lost if cosmics outside the 7sigma range are clipped; higher values might indicate an issue with one of the input files, SNR might get compromised.

Table 8. QC flags for stacked spectra.

Science products. Like for the single exposures, the pipeline processing of the stacked data is done automatically, with some internal quality control monitoring:
- the values of mean_dev and contrib in the above table;

- on-demand QC reports and quick-look overviews.

This information has largely been used to improve and fine-tune the reduction process. An individual one-by-one inspection of the products has not been done. The spectral overviews are delivered with the spectral products.

Previews. As for the single-exposure products, there are two preview plots, delivered as ancillary files along with the products:

1. one main QC plot which is the same for all spectra from the stack one per dataset (the same for all products from the same raw file), see Figure 14; this plot is essentially the same as for the single-exposure spectra (Figure 4), except for some added labels: the total exposure time TEXPTIME for the stack (instead of EXPTIME for the single exposure); the number of combined spectra; the maximum deviation between exposure levels, max_dev, for the brightest fibre.
2. the preview plot (Figure 15), one per product. It is equivalent to the preview for the single spectra (Figure 5).
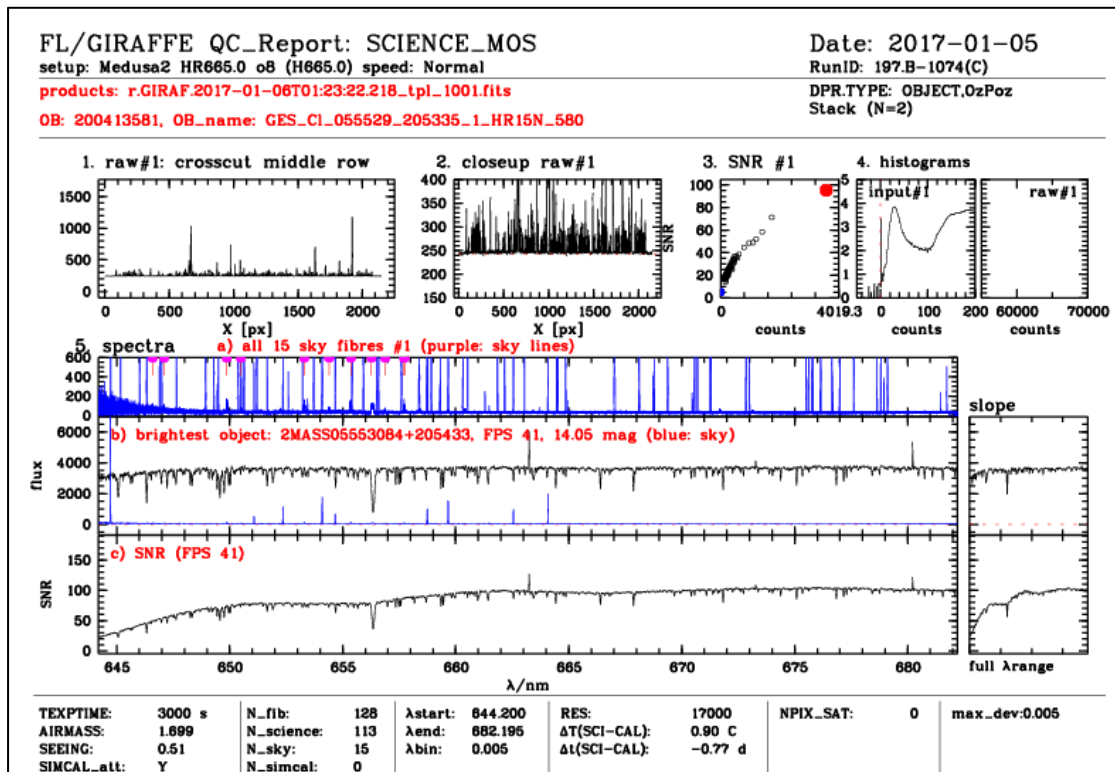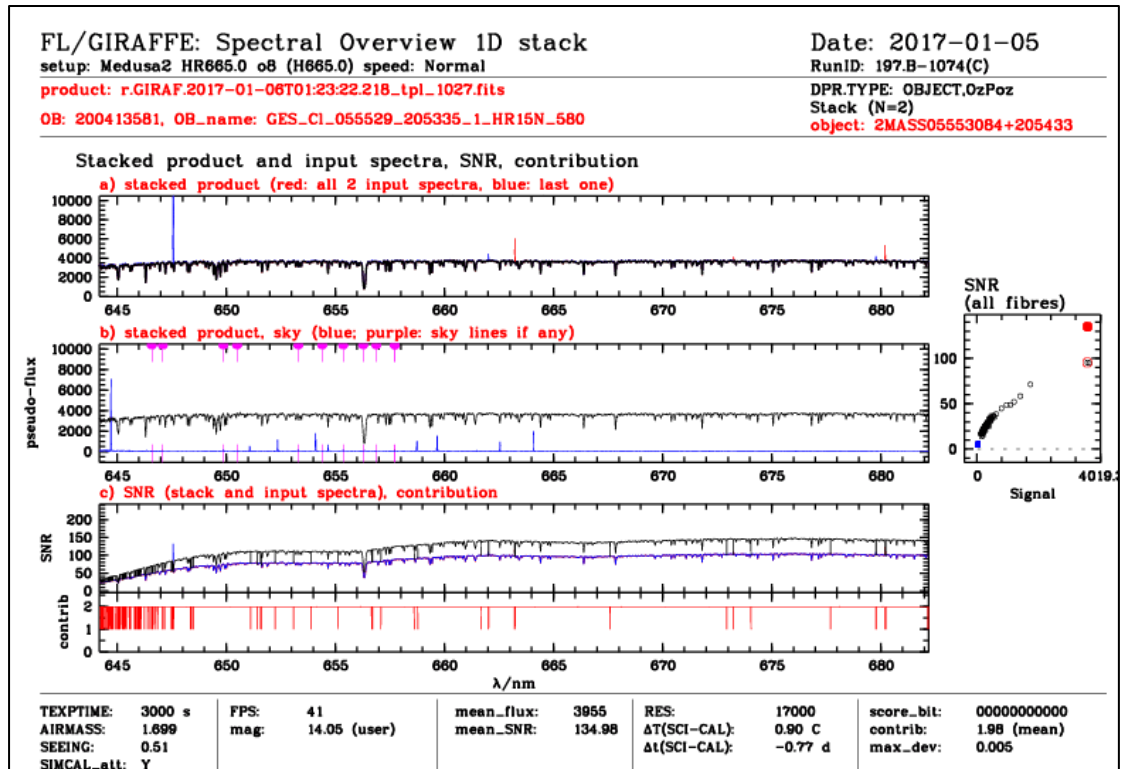

Figure 14. Main QC plot for the whole dataset. Description:
- On top: acquisition date, setup information, run ID; product ID, DPR.TYPE, OB ID, OB name, and number of combined spectra for the stacks.
- Boxes 1 and 2: crosscuts from the first raw file (#1).
- Box 3: SNR vs. counts for all fibres (SCIENCE in black, SKY in blue, brightest SCIENCE in red), from the first single exposure (#1).
- Box 4: histograms for 2D product of the first single exposure (#1); *4a*: product file, the histogram focussing on the region around the background, counts are on a logarithmic scale; *4b*: histogram for the first parent raw file, focussing on the region around saturation, 65,000 ADUs.
- Boxes 5: spectral overviews; *top*: all available *sky* spectra for the first exposure, plotted in blue; *middle*: signal from the *brightest fibre*, first exposure; in blue: the last sky fibre for this exposure; *bottom*: SNR plot for the *brightest fibre*, first exposure; sky emission lines, if any, are plotted from http://www.eso.org/qc/UVES/pipeline/sky_spectrum.html.
- At bottom: a set of QC parameters applicable to the first raw file and the 2D products, including the flag for attached SIMCAL (Y/N), number of fibres (total, science, sky, simcal), spectral coordinates in nm, resolving power, deltaTime and deltaTemp values, number of saturated pixels, and the

maximum deviation between the flux levels of the contributing exposures, for the brightest spectrum.

Process quality control. The quality of the data reduction is monitored with quality control (QC) parameters, which are stored in a database. The database is publicly accessible through a browser and a plotter interface[22].

QC parameters are used to monitor the reduction quality. The most important check is the "SNR versus signal" control plot (both parameters being averages across the entire spectrum). Figure 16 displays this plot for stacked data from 2016 and 2017.

Figure 15. Preview plot, one per stacked product. Description:
- On top: acquisition date, setup information, run ID; product ID, DPR.TYPE, OB ID, OB name, object name, and number of combined spectra for the stack.
- Panel a: stacked product (black), all N input spectra (red), last input spectrum (blue); this to visualize potential flux variations; with no flux variations, only blue or red spikes due to cosmics are visible.
- Panel b: the stacked product, plus the sky signal (last sky fibre) for comparison; sky emission lines are marked if any[23].
- Panel c: the SNR of the stacked spectrum (in black), of all input spectra (in red), and of the last input spectrum (in blue).
- The plot of the column CONTRIB_REDUCED with the number of contributing spectra per wavelength bin.
- Also displayed at right: the SNR of this stacked product (full red dot), the SNR of the first input spectrum (open red circle), and the SNR values for all other fibres of the first exposure of this OB stack, establishing the SNR curve for a single exposure and visualizing the SNR gain of this stack.
- At bottom: Score bits (QC flag), averaged contribution, maximum deviation of flux levels from the mean; other parameters as applicable for all input spectra, see previous Figure 14.
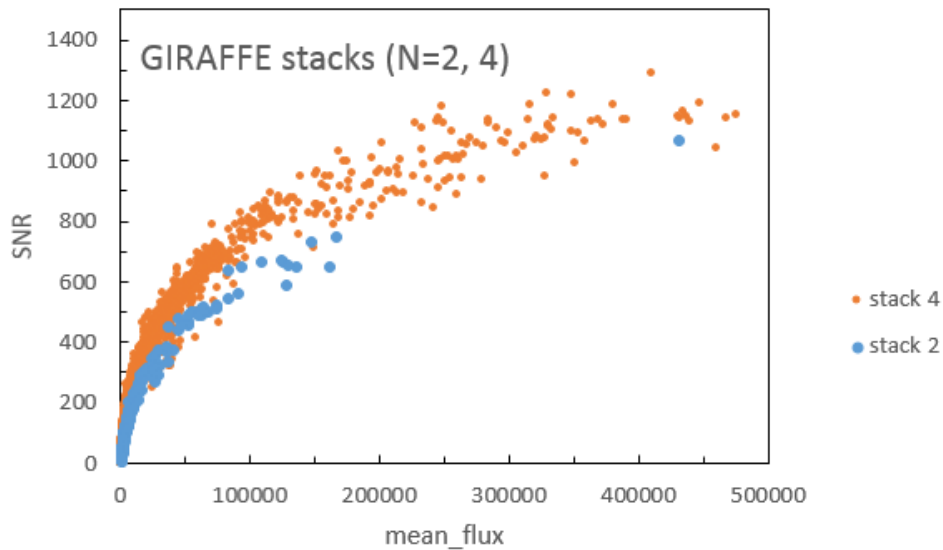

Figure 16. SNR vs. signal for N=2 and N=4 GIRAFFE stacks. 'mean_flux' is the average pseudo flux of a spectrum, SNR its average signal-to-noise. By combining multiple spectra of the same mean_flux, the SNR is expected to increase by a factor sqrt(N). This figure is based on the complete stacked data from 2016 and 2017, all settings. The N=2 data points refer to the brightest fibre in the OB stack (to avoid crowding). For the N=4 data (which are much rarer) we display all data points.

In Figure 17 and Figure 18 we display the main panels of the spectral preview plots for two examples of N=2 and N=4 stacks. They demonstrate the efficient suppression of cosmics and the improvement of SNR in the stacked products.
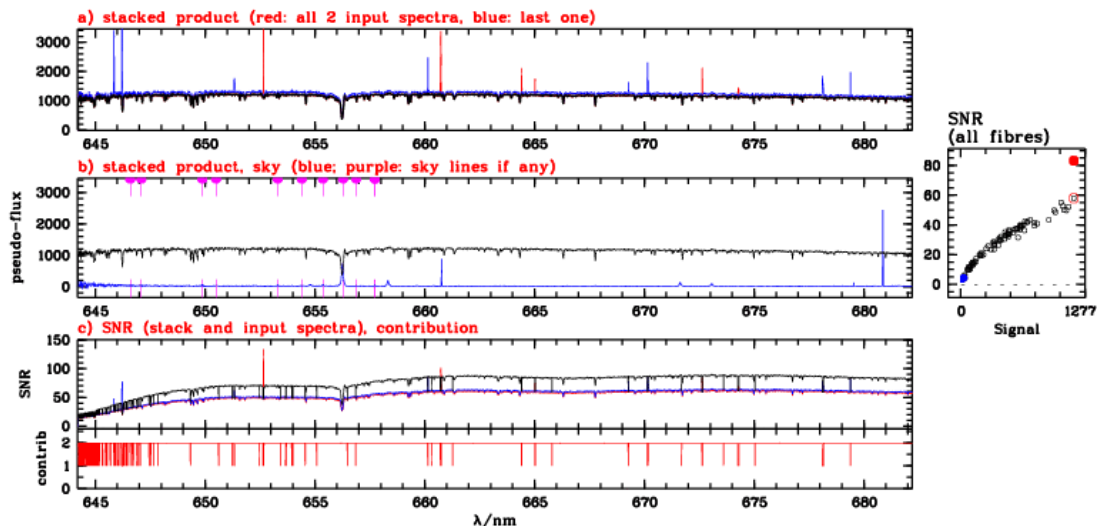

Figure 17. Cosmics suppression and SNR improvement for N=2 stacks, from an example preview plot. From top to bottom: a) two single spectra (in red and blue) plus the stacked product; since flux levels differ by less than 2%, the only noticeable differences occur in the spikes (red or blue) due to cosmics; b) the stacked product only, where the cleaning effect of the N=2 pre-processing is visible (no spikes left); c) the SNR plot for the two single spectra (red and blue) plus for the stacked product which is higher by about sqrt(2); the contribution plot has N=2 almost everywhere (contrib=1.97). Finally, the rightmost box shows the SNR vs. signal control plot (all fibres) for one of the input files, and the comparison of the single SNR (open red circle) and the stack SNR (full red circle) for this spectrum.
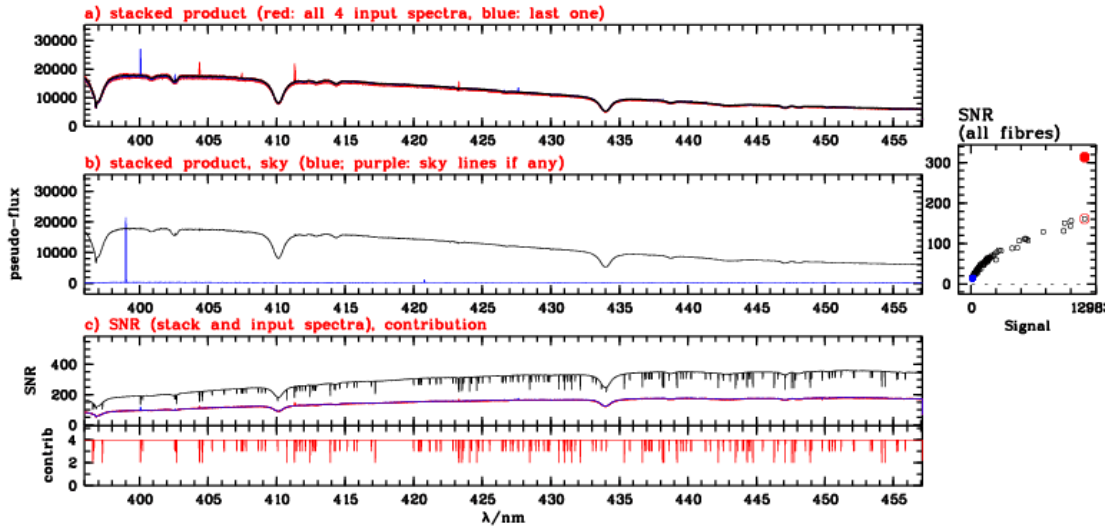
28

Figure 18. Cosmics suppression and SNR improvement for N=4 stacks, from an example preview plot. The improvement in SNR is about a factor 2. The cosmics cleaning is provided by the clipping routine. The SNR curve in panel c) reflects the contribution function which oscillates between 2 and 4. With an average contribution of 3.93 it looses less than 2% of the bins in the input spectra.

Known features and issues

Issues.

In rare cases, observers apply small telescope offsets between consecutive exposures in an OB. These offsets can be recognized only by comparison of the raw files belonging to the OB. The information in the OzPoz table (in particular target names and coordinates) is not reliable anymore and should not be trusted. It relies on the assumption that the telescope is NOT moved between individual exposures. Nevertheless, these offsets have been applied by observers to support two different scenarios: a) mapping background nebular emission, or b) getting an accurate sky background in exposure #2 for all stellar signals in exposure #1 by applying an offset by a few arcseconds. When recognized, such OBs have been rejected for stacking, and the products from the single exposures are available.

Table 9: List of known programmes that have used telescope offsets between single exposures in an OB

Programme ID	OB IDs	Dates	Comment
072.C-0348(A)	200134720, 200134730	2003-12-02	mapping pattern applied, target names and coordinates in the products are correct in the first exposure but then the telescope offsets are not properly registered in the products; check the raw files if this is an issue
074.D-0518(A)	200146024	2004-12-05	mapping around standard star
	200146036	2004-12-06	
092.C-0820(A)	1014429, 1014431	2014-01-31	#1: objects, correct target names and coordinates; #2: corresponding sky for these targets, coordinates slightly wrong
	1014433	2014-02-04	
	1014424	2014-02-07	
	1014427	2014-03-01	

There might be more cases which escaped proper recognition. If data from a stack look suspicious:

- compare the SNR of the stacked product (SOBS) with the SNR of the first single product (ancillary file SSXP), for a given fibre: if the stack SNR is lower than the single SNR, this indicates the stacking of an object exposure with a sky exposure, with the sky exposure not marked as such and with tiny offsets; use the first ancillary SSXP file for scientific analysis instead of the delivered SOBS file.
- the mapping pattern can only be recognized from the raw files; in case of concern, compare line profiles in the individual ancillary files; if they look different, check RA and DEC in the raw file headers, to check for potential offsets; if there are offsets, use the SSXP files instead of the SOBS file.

Please inform us about further cases where you suspect stacking of wrong candidates.

Features.

Pseudo-flux. The "FLUX_REDUCED" scale in the stacked products is not flux-calibrated in physical units, but is in flat-fielded counts per sec. There is no response curve for the MEDUSA mode. This is also true for single exposures.

Ambient parameters. Ambient parameters like airmass and seeing in general have evolved in the OB stack. They are always inherited from the first input file.

SKY spectra. We have not stacked the SKY spectra, for technical reasons. We have not subtracted the sky signal since in general we do not know which one, or which subtraction strategy, is the best suited for a given object. If you need to subtract the sky:
- Stack the 2D sky files (which come with the single spectra), or (better) create a smoothed version.
- If sky subtraction is critical, check also for variability of the sky.
- Select the best (closest) sky fibre (if it contains background) and subtract it from the stacked spectrum.

SIMCAL corrections. The binary tables in the SSIM files attached to the stacked spectra have information about the SIMCAL shift.

For those single SIMCAL spectra that qualify for the stacking we have disabled the (very small) simultaneous wavelength correction which might fail in case of saturated spectra. This would be an issue for the stacking, since some of the input files for the stack could have this correction applied, others not (if it failed due to saturation).

Single exposure spectra. The spectra from all single exposures are added as ancillary files to the stacked spectra. We recommend to download them in order to obtain additional quality information, and also to judge the significance of spectral features if in doubt.

Data Format

File Types

The primary GIRAFFE stack product is the stacked fibre spectrum in binary spectroscopic data format:

ORIGFILE names starting with	product category HIERARCH.ESO. PRO.CATG	format	how many?	which fibres	Description
GI_SOBS	SCIENCE_RBNSPEC_STACK	binary table, signal, error, contribution, and SNR	1	one SCIENCE	main data product; stacked from *N* single GIRAFFE spectra of the same OB

(SOBS stands for "science OB stacked").

There are *n* products per OB stack, if each of the parent raw file contains *n* science fibres (marked as type "M" in the OzPoz table attached to the raw files).

Each stack product has several ancillary FITS files delivered with it:

ORIGFILE names starting with	product category HIERARCH.ESO. PRO.CATG	format	how many?	which fibres	Description
GI_SSXP ("science single exposure")	SCIENCE_RBNSPEC_IDP	binary table, signal, error, and SNR	*N*	One SCIENCE	*N* corresponds to the number of input spectra for the stack; these spectra have the same format and content as the single-exposure GI_SIDP products
GI_SRBS	SCIENCE_RBNSPECTRA	2D image, plus FIBRE_SETUP table	*N*	all: SKY, SCIENCE, SIMCAL	extracted, re-binned; one for each GI_SSXP (single exposure)
GI_SSKY	ANCILLARY_MOSSKY	2D image, plus FIBRE_SETUP table; SKY fibres extracted for convenience from GI_SRBS	*N or 0* (in rare cases: none)	SKY only	extracted, re-binned; useful for determination of SKY to be subtracted from 1D data product; one for each GI_SSXP (single exposure)
GI_SSIM	SIMCAL_RBNSPECTRA	2D image, plus FIBRE_SETUP table, including residual shifts WLRES; origin: attached exposures, same OB	0, *N*, or more	all : SKY, [SCIENCE*], SIMCAL	extracted and re-binned

*Formally, the fibres in this product also have the type 'M' but due to the short exposure time contain only noise.

Furthermore the following non-FITS files are delivered with each spectrum:

ORIGFILE names starting with	PRODCATG	format	how many?	Description
r.GIRAF...gif	ANCILLARY.PREVIEW	gif file	1	see Figure 14Figure 4, labelled "QC report"
r.GIRAF...gif	ANCILLARY.MOSSPECTRA.PREVIEW	gif file	1	see Figure 15; labelled "Spectral overview"

The following naming convention applies to the ORIGFILE product: e.g. the name
`GI_SOBS_200413581_2017-01-06T01:23:22.218_F89_Med2_H665._o8.fits`
has the components:

GI	SOBS	200413 581	2017-01-06T01:23:22.218	F89	Med2_H665._o8.fits
GI-RAFFE	product type (SOBS stands for "science OB stacked")	OB ID	timestamp of first raw file	fibre with index FPS89	setup string: Med2 for Medusa2; H665._o8 for central wavelength (665.0nm) and grating order (8)

The ancillary files have the following ORIGFILE names:

Table 10. Naming conventions of ANCILLARY files

PRODCATG	example	rule
ANCILLARY.SPECTRUM	GI_SSXP_964735_2013-05-07T01:58:20.737_F89_ Med1_H504.8_o11.fits	Timestamp from raw file name; other name fields from GI_SOBS file
ANCILLARY.MOSSPECTRA	GI_SRBS_964735_2013-05-07T01:58:20.737_ Med1_H504.8_o11.fits	Same as corresponding SSXP file except for fibre index; name starting with GI_SRBS
ANCILLARY.MOSSPECTRA. SKY	GI_SSKY_964735_2013-05-07T01:58:20.737_ Med1_H504.8_o11.fits	Same as GI_SRBS, name starting with GI_SSKY
ANCILLARY. MOSSPECTRA.CALSIM	GI_SSIM_964735_2013-05-06T23:37:37.133_Med1_H504.8_011.fits	Same as GI_SRBS, starting with GI_SSIM
ANCILLARY.MOSSPECTRA. PREVIEW	r.GIRAF.2013-05-07T01:58:20.737_tpl.fits.gif	Technical filename of the first GI_SRBS file, with 'tpl.fits.gif' appended
ANCILLARY.PREVIEW	r.GIRAF.2013-05-07T01:58:20.737_tpl.fits_f122.gif	same name, with fibre index included

The user may want to read the ORIGFILE header key of the archive-delivered FITS file and rename it to the ORIGFILE name.

File Structure

The stacked spectrum GI_SOBS comes as binary FITS table in multi-column format. The columns are labeled as follows:

Table 11. Internal structure of the GIRAFFE stacked spectra binary table.

column	label	content
#1	WAVE	wavelength in nm, corrected to heliocentric system
#2	FLUX_REDUCED	extracted, wavelength-calibrated but not fluxed SCIENCE signal in counts
#3	ERR_REDUCED	Error of FLUX_REDUCED (same units)
#4	QUAL_REDUCED	Not significant for GIRAFFE stacks (always 0)
#5	CONTRIB_RE-DUCED	Contribution per WAVE bin (number of actually used input files, either N or less)
#6	SNR_REDUCED	Signal-to-noise ratio = FLUX_REDUCED/ERR_REDUCED

Acknowledgment text

According to the Data Access Policy for ESO Data held in the ESO Science Archive Facility, all users are thus required to acknowledge the source of the data with an appropriate citation in their publications. Since processed data downloaded from the ESO Archive are assigned a Digital Object Identifier (DOI), the following statement must be included in any publications making use of them:
Based on data obtained from the ESO Science Archive Facility with DOI(s) : https://doi.eso.org/10.18727/archive/27.