DOCUMENT

Spectral Line Catalogue from Herschel-HIFI Spectral Scans:

Explanatory Supplement

Prepared by	David Teyssier (ESAC), Lisa Benamati (IRAP)
Reference	HERSCHEL-HSC-DOC-2196
Issue	1
Revision	Release 1.0
Date of Issue	2017-12-21
Status	For release
Document Type	Release Note
Distribution	HSC

Document approval

Prepared by:	D. Teyssier, L. Benamati	Date: 2017-12-21
Approved by:		Date:

Change log

Version	Date	Change	Author
0.1	2017-06-16	First draft of document	David Teyssier, Lisa Benamati
1.0	2017-12-21	Final document issue	David Teyssier, Lisa Benamati

1 Introduction

The Heterodyne Instrument for the Far-Infrared (HIFI, (7)) was the high-resolution spectrometer on-board *Herschel*, offering a continuous spectral coverage in the ranges [480–1270] and [1430–1910] GHz with a resolving power in excess of 10^7. One of the richest output of the HIFI scientific database lies in the hundreds of spectral line surveys it conducted using its Spectral Scan observing mode. Overall, HIFI collected around 500 observations in this mode, over about a hundred of different lines-of-sight, in partial or full spectral coverage.

A dedicated effort was conducted in order to extract and identify as many spectral lines as possible from these observations, using baseline-corrected versions of the Spectral Scan products specially curated as Highly-Processed Data Products (HPDP, (17)). The approach is based on an automatic spectral feature detection algorithm, coupled to a spectroscopy data-base used to assign the most likely species and transition to any detection. The process used a combination of tools specially coded in HIPE[1], and functionalities offered in CASSIS[2].

This document describes the method used to generate and validate the Spectral Line Catalogue obtained from these observations. The Catalogue is made available through the Herschel Science Archive as a Highly-Processed Data Product, and we provide here further information about the content of this data-set.

2 Cautionary notes

The primary goal of this work is to provide a list of spectral features present in the HIFI spectral scans, and offer a best-guess of which line and transition that each correspond to. The completeness of the catalogue is limited by the data noise (as is explained in the following section, a SNR threshold of 5 was applied to the selection of spectral line feature), as well as the complexity of the line profiles (intrinsic to the source, or altered by possible OFF position contamination). On top of that, not all spectral lines could be assigned to a specific transition (so-called U-lines), and some could actually correspond to multiple solutions (so-called blended lines). The U-lines are not provided as output of this study (see also Sect. 8).

Although all integrity checks show a high level of reliability in both line detection and transition assignment (see Sect. 7), special care should be taken when using lines with low intensity (typically below some hundreds of mK) and/or signal-to-noise ratio, in particular when those lines are stamped as blends. Finally, the fitted line intensities should be used with precaution when the line profiles deviate from a simple Gaussian, or when the spectral line density is particularly high (which will manifest also in the line blends) – see also Sect. 5.2 for additional warning flags provided with the catalogue.

---

[1] `https://www.cosmos.esa.int/web/herschel/hipe-download`

[2] `http://cassis.irap.omp.eu`

3 The HIFI Spectral Scans

3.1 Archive content and science readiness

We consider here the ~500 Spectral Scan observations obtained in standard observing mode (in contrast with those performed in the so-called engineering mode) with one of the three possible referencing schemes for this mode (namely Double Beam Switching, Load Chop and Frequency Switching). Of those, we ignored the 18 Spectral Scan observations stamped as non-public, as they usually made use of non-standard observing modes or instrument configuration settings. The standard product generation pipeline provides calibrated deconvolved spectra in each polarisation for each observation, and takes care of masking all spurious spectral features still present in the Level 2 products (see e.g. the HIFI Product Explained document, or the Spectral Scan cookbook). What the automatic pipeline does not cover, though, is the treatment of any residual baseline distortion resulting from an imperfect bandpass calibration (e.g. standing waves, see Section 5.3 of the HIFI Handbook for further details).

3.2 Baseline-corrected Spectral Scans and considered sample

In order to circumvent those residual artefacts, baseline-corrected spectra have been generated by the Herschel Science Centre (HSC) in the form of Highly Processed Data Products[3] (17). Due to the semi-automatic approach used in this process, there are some limitations to the type of observations that can be accurately treated by this method. In particular, its result is relatively poor in case of relatively line-rich spectra, or complex, broad, line profiles. This is particularly true for observations towards sources such as Orion KL (e.g. (5)), the SgB2 lines-of-sight (e.g. (4), (18)), η Carina (e.g.(10)) or line rich evolved stars like IRC+10216 (e.g. (3)), VY CMa (e.g. (1)) or OH231.8+4.2 (e.g. (2)). When applied to these complicated sources, the baseline correction usually results in erroneous line intensities and/or profiles, which we decided would not qualify as HPDP.

As a consequence, about 1/3 of all observed Spectral Scans was discarded from this exercise, and it was therefore decided that the same crtierion should apply to the sample eventually used to generate the Spectral Line Catalogue described in this document. Table 7 gives an inventory of the omitted observations, together with references to past or future work expected to provide the information missing from the baseline-corrected data-set provided by the HSC. The expectation is that those products (both baseline corrected and Spectral Line Catalogue) should eventually become available as User-Provided Data Products[4].

Exceptionally, however, Spectral Line Catalogues were generated for sources where the baseline correction was not considered accurate enough to release the corrected data themselves, but where, however, the line detection and species identification would make sense. For those, no information about line intensity or line width is provided. They are listed in Table 8. This approach is also followed for some particular cases of the catalogue generation where input data have been smoothed for dedicated purpose (see Section 4.2.2).

---

[3] `http://archives.esac.esa.int/hsa/legacy/HPDP/HIFI/HIFI_spectral_scans/`

[4] `https://www.cosmos.esa.int/web/herschel/user-provided-data-products`

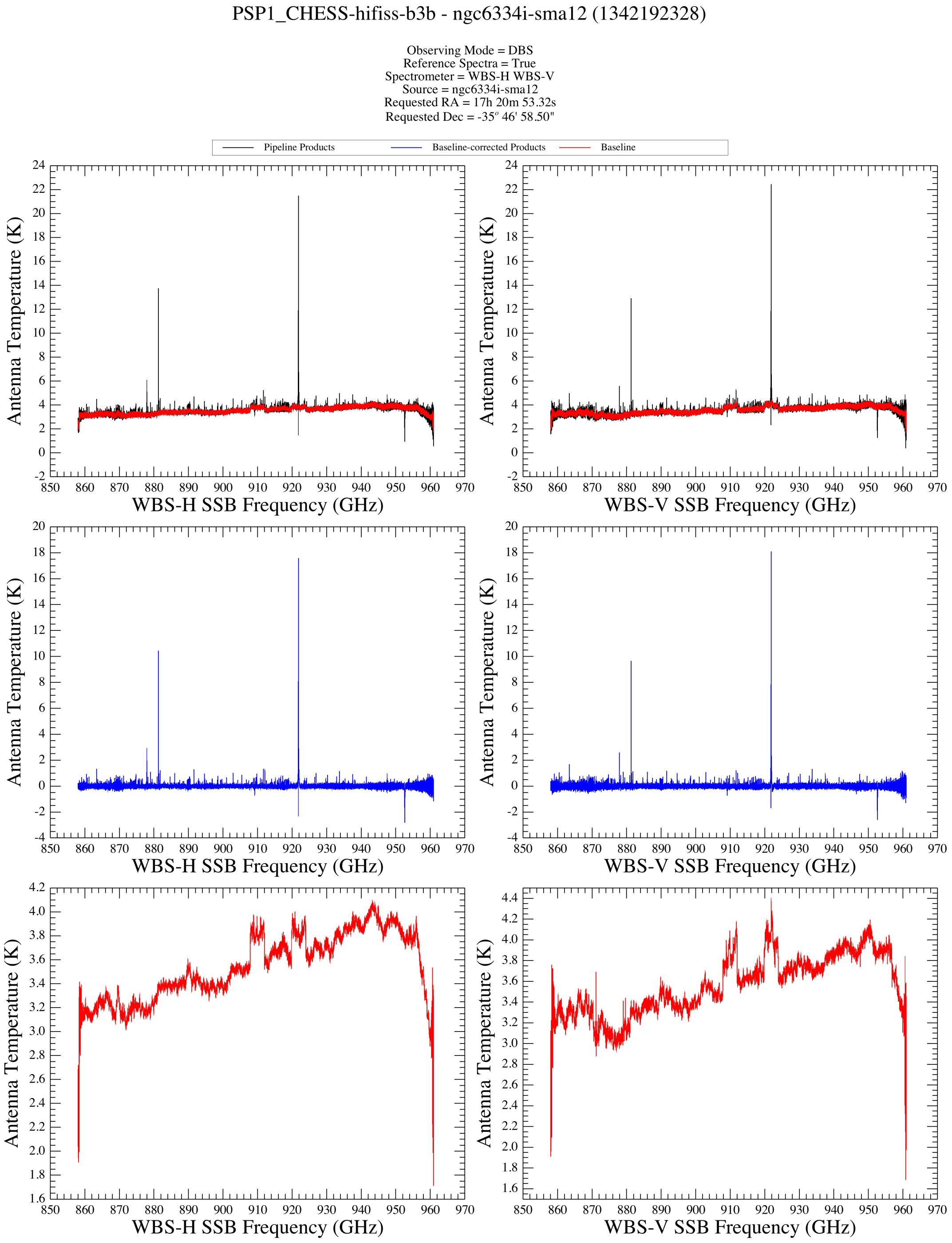

Figure 1: Illustration of a baseline-corrected HPDP for obsid 1342192328. The upper panel shows the products from the standard pipeline generation (black) with the removed baseline in red. The middle panel (blue spectra) shows the resulting baseline-corrected spectra. Finally, the lower panel shows the removed baseline. Note that the removed baseline contains both the intrinsic source continuum and the residual baseline artefacts.

In total, the catalogue generation considered an input of 339 baseline-corrected spectral scans. Table 6 summarises the sources they correspond to, and the spectral coverages offered by the associated observations.

4 Generation of the Spectral Line Catalogue

4.1 Method

The input to the catalogue generation is in fact based on the merging of the two polarisation spectra provided for each of the baseline-corrected Spectral Scans. In all but two cases (obsid 1342181163, where the H spectrum is plagued with artefacts that we could not clean effectively, and obsid 1342253144 where only the V channel was selected by the observer), both H and V polarisations are averaged. Note that there are known reasons for the H and V spectra to display slightly discrepant information, both on line and continuum (see e.g. Section 5.8.4 of the HIFI handbook).

Based on these input files, the catalogue generation essentially relies on a dedicated task developed within HIPE and called `identifyLines` – a general description of the task and its parameters is given in the correspond HIPE manual page[5]. The task works in two steps:

- First, the spectral feature detection is performed, based on an automatic line masking functionality provided in the HIPE software (so-called smoothBaseline, see the task description in HIPE manual[6]). The main parameters for this stage of the processing are (i) the number of segments to split the input spectrum into in order to speed up the process, (ii) the spectral smoothing factor, usually used to lower the noise rms, (iii) the best-guess line width for emission and absorption lines respectively, (iv) the signal-to-noise (SNR) threshold for a feature to qualify as a robust detection. Detected lines are then fit with a Gaussian line profile and the outcome is stored in the table. For absorption lines, the resulting line will appear with negative intensity because the input spectrum has had its baseline subtracted (Section 3.2).
- The second step consists of assigning the most likely species and transition to each of the detections resulting from the first step. For this a line list template is used. Section 4.3 describes how this template was built. Because this line list template provides line frequencies in the Local Standard of Rest (LSR), a fundamental step in the assignment is the assessment of the systemic velocity of the studied target in the LSR, called thereafter v_{LSR}. We explain in Section 4.3.2 how this latter was figured out for each observation.

4.2 Line detection

The HIPE task `identifyLines` was used systematically on all 339 baseline-corrected polarisation-averaged spectra described in Section 3.2. We used the following parameters as

---

[5] `http://herschel.esac.esa.int/hcss-doc-15.0/load/hifi_um/html/hdrg_lineid.html`

[6] `http://herschel.esac.esa.int/hcss-doc-15.0/load/hcss_urm/html/herschel.ia.toolbox.spectrum.standingwaves.SmoothBaselineTask.html`

default:

- nbOfSplit = 4 (i.e. split the spectrum in four equal ranges to speed up the process)
- smoothBox = 7 (i.e. smooth the spectra with a 7 channel box to lower the rms and increase the line detectability)
- fwhmAbsorption = 1 km/s (best-guess line width for absorption lines – note that this number had to be increased for some obsids exhibiting broader absorption lines).
- fwhmEmission = 4 km/s (best-guess line width for emission lines – note that this number had to be increased for some obsids exhibiting broader emission lines).
- snrType = ”Intensity” (parameter used to compute the signal to noise ratio – the alternative is Flux)
- minSNR = 5 (value of the minimum signal to noise ratio – all lines detected with a value smaller than that are rejected)

The decision on the SNR threshold is based on an exploratory study of the outcome of the line identification for various values of the SNR. The threshold of 5 was found to be a good trade-off between limiting to the maximum the number of false positives (which would show up at too low SNR) while still allowing recognition of lines appearing as an unequivocal detection. There were, however, still some false positive resulting from the chosen threshold, and a dedicated cleaning exercise was then performed as a subsequent step.

4.2.1 Line parameters

Each detected spectral feature is then fitted by a model consisting of a Gaussian sitting on top of a first order polynomial continuum. The fitted parameters and their uncertainties are then tabulated in the line identification table (see Section 5.1). It should be noted that Gaussian fits are performed irrespective of the exact line profile, and as such there is no additional information about the existence of e.g. outflows or peculiar line profiles (e.g. top hat in AGB stars). This also means that the derived line parameters (width, centroid, integrated intensities) might not be fully representative of the values applying to the more complex line profiles.

4.2.2 Smoothed and un-smoothed input data

Because the initial spectral feature detection step relies on an automatic line mask generation, there can be particular cases where the line profile of a given feature makes it difficult for the algorithm to yield a detection. This will generally happen with relatively broad lines, lines in absorption, or complex profile such as P-Cygni ones, or their look-alike that arise in case of severe OFF position contamination (see also (16)). In these cases, the task was shown to do a better job when the input spectra was smoothed in frequency.

The positive consequence of this alternative process is that not only complex profiles would end up being detected and then fed to the line identification step, but also weaker lines initially buried into the noise at native resolution would be picked up by the detection step, and also make it into the catalogue table. On the downside, however, the smoothing needed to achieve such result is sometimes relatively large compared to the intrinsic line widths and, as a consequence, the fitted line parameters of the severely broadened line profiles would no longer be representative of the real line characteristics. Also, for very crowded spectra, significant smoothing will blend lines located close to each other, potentially precluding to detect them as they would no longer be spectrally resolved.

Because there is still a significant added value in terms of detected species, we are providing the line catalogue associated to these smoothed spectra as a complementary information to those achieved with native resolution data. The line parameters such as line peak intensity or line width are, however, omitted from the corresponding catalogue files. In crowded line regions, those inputs should be interpreted with particular caution, as the large smoothing will tend to further blend lines already very close in frequency. An example of such case is illustrated in Fig. 3. In this particular case, lines such as the ^{13}CO 10–9 around 1101.3 GHz or C^{18}O 10–9 around 1097.2 GHz are clearly detected, but their complex line profile (here due to severe OFF position line contamination) made it impossible to the task to detect and therefore report the line identification. The lower plot shows the same result using a smoothed input spectrum, which allows circumvention of that problem and allows additional line detections, albeit at the expense of getting non-representative line profile fitting parameters.

4.3 Line identification

4.3.1 Line list template

The line identification is part of the HIPE `identifyLines` task used in the process. It relies on a line list template against which the detected and fitted line parameters are compared.

In order to build an initial line list template representative of the species and transition typically achievable by HIFI, we analysed three prototypical sources (IRAS 16293-2422, NGC6334I and IRC+10216) with the help of the CASSIS line analysis tool package. CASSIS harbours two of the most commonly used spectroscopic databases (JPL[7] (13) and CDMS[8] (11)) and allows easy display of all possible identifications of a line in a spectrum.

For each source, the task generally found a lot of unidentified lines that were actually known lines because the initial line list template was not complete enough. Possible candidates for those unidentified lines were assessed using CASSIS, and the final line assignment was further secured by checking the presence of other transitions of the same species at similar excitation in other spectral ranges observed for the same source. All confirmed identifications were then added to the line list template to be used for the next source. All non-confirmed identifications were kept in a table to build the so-called Unidentified Line List. The consolidated line list template used for the catalogue generation is also provided together with the other catalogue deliverables (Section 5).

In some cases, the same source was observed in the same band with different obsids, but the outcome of the line identification is not necessarily strictly the same. This can be due to different noise levels, and non-fully repeatable line intensity that could affect detections just marginally above the SNR threshold of 5.

---

[7] Developed by the Jet Propulsion Laboratory, see `https://spec.jpl.nasa.gov`

[8] Developed by the Cologne University, see `https://www.astro.uni-koeln.de/cdms`

4.3.2 Source velocity

The HIFI pipeline provided spectral with a frequency scale calibrated in the LSR. As such, all lines appear shifted by the source systemic velocities with respect to the LSR. This velocity hence needs to be provided to the line identification task in order to related directly the fitted line centroid to that of the LSR line frequencies tabulated in the spectroscopic database, and therefore in the line list template.

In order to determine the best guess value of the source systemic velocity, we relied primarily on the detection of the ^{12}CO (5-4) line (or the ^{13}CO (5-4) line if the profile of the ^{12}CO line was too complex) in band 1a. When those were not observed or non-detected, our first backup was to try and recognise the $H^{12}CO^+$ (7-6) or $H^{12}CN$ (7-6) lines present in band 1b. When even this was not possible, we then tried to rely on higher transitions of the same molecules which could have been detected in other bands. When none of this was available, we finally relied on velocities reported in the literature. This initial centroid value is tabulated as v_0 in Table 9.

Once the line catalogues were generated for each observation, we computed for each source the mean v_{LSR} value and its uncertainty as the respective median and the standard deviation of all v_{LSR} derived from an identified line in this source. The presence of high values of the v_{LSR} uncertainty for some sources is due to their complex nature and structure. Indeed, some sources are known to harbour maser lines emission, outflows, or broad emission lines. Table 9 summarises all these numbers, together with other statistical figures applying to the sources covered by the catalogue (e.g. line width). When only one line was identified, the v_{LSR} is given without any uncertainty.

4.3.3 Blended Entries

When several entries from the line list template match the fitted position of a given spectral feature, multiple entries are populated into the catalogue table. Those entries will share all of the line fitting parameters and the only difference will be in the species/transition names, as well as the rest frequency. Those multiple entries should be interpreted as alternative line identification and the final decision on which of those it is most likely is left to the user.

4.3.4 Unidentified Lines

It is sometimes the case that a spectral feature considered as a genuine detection does not result in any line identification, most likely due to the incompleteness of the template line list. Such features are then considered Unidentified Lines, or U-lines. The tables of U-lines are also one of the outputs of the whole process, however, they are not delivered in the present data-set. This is essentially due to a lack of resources in order to perform a dedicated validation of the corresponding entries, in particular to discard any false positive left-over. Those may be considered in a subsequent release of the Spectral Line Catalogue (Section 8).

5 Spectral Line Catalogue Content

The delivered Spectral Line Catalogue corresponds to the successful line detection and line identification achieved on a total of **278** observations. As explained in Section 4.2.2, however, some line identifications could only be performed once the data were spectrally smoothed. As such, the full catalogue will contain multiple inputs for a given observation, with the following breakdown:

- There are **270** observations that lead to a Spectral Line Catalogue of at least one entry when using the native resolution Spectral Scan data. In those, **13** observations correspond to Spectral Scan data where only sub-optimal baseline correction could be achieved (Table 8, see also Section 3.2).
- There are **76** observations for which complementary catalogues are provided based on smoothed Spectral Scan data. Only **8** of those observations have no counterpart in the list of the 270 observations above, because they lead to at least one line detection only when the input data were smoothed.

The above material is provided in the form of tables of the detected and identified lines, together with postcards aimed at illustrating the content of the tables. For each observation ID, the following files are provided:

- at most two line lists are provided in the form of FITS files: one resulting from the spectrum at native resolution, and, when applicable, one resulting from the spectrum smoothed to a certain resolution (see Section 4.2.2). In the latter case, no information is provided about the line flux and width, as well as the fitted continuum level. The naming convention used for those tables is `<obsid>_HIFI_LineCatalogue_<method>.fits.gz`, where `obsid` is the observation ID and `method` is either "native" or "smoothed" depending on the spectral resolution used. Note that no particular naming convention is used to discriminate the 13 observations listed in Table 8. However, as for the smoothed spectra catalogues, no information is provided about the line flux, width and continuum level due to the sub-optimal baseline correction used for those data.
- one, so-called *prime* postcard, is provided for both native and smoothed resolution catalogues where applicable. On top of this, for crowded and/or large spectral coverage observations, *zoomed* postcards are provided in narrower chunks of spectral range. The respective file names for those are `<obsid>_HIFI_LineCatalogue_postcard_PRIME_<method>.png` and `<obsid>_HIFI_LineCatalogue_postcard_ZOOM_<nb>_<method>.png` where `nb` is just a running number for the various zoomed postcards generated for a given obsid.
- the line list template used by the `identifyLines` task to assign transitions to detection, and described in Section 4.3.1 is also made available[9]. Note that this list differs from the two default lists provided in the HIPE software distribution[10].

---

[9] `http://archives.esac.esa.int/hsa/legacy/HPDP/HIFI/HIFI_line_catalogue/LineListTemplate.txt`

[10] `http://herschel.esac.esa.int/hcss-doc-15.0/load/hifi_um/html/Hifi.Um.Sec.LinesCatalog.html`

- finally, the full catalogues for the respective native and smoothed resolutions are provided as a concatenation of all entries from all considered obsids. Those full catalogues are named `HIFI_Spectral_Line_Catalogue_Native_Full.fits.gz` and `HIFI_Spectral_Line_Catalogue_Smoothed_Full.fits.gz`.

Although the above files will be provided by the Herschel Science Archive[11], they can also be fetched from a dedicated legacy repository: `http://archives.esac.esa.int/hsa/legacy/HPDP/HIFI/HIFI_spectral_line_catalogue/`. In this repository, the above files will be organised in separate folders `TablesNative`, `TablesSmoothed`, `PostcardsNative`, `PostcardsSmoothed` and `FullCatalogues`. Within each postcard directories, sub-directories separate the PRIME ones from those dedicated to zoomed ranges. On top of that, bundles of all files applying to one given obsid can be retrieved in the directory `CataloguesPerObsid` under the names `<obsid>_HIFI_LineCatalogue.tar.gz`. The total size of the HPDP is 56 Mb.

5.1 Description of the Table Datasets

The catalogues come in the form of table data-sets saved as FITS files. Their header content is described in Table 1. Apart from some general meta-data about the observation itself, the header also features some specific flags providing information of interest to the user. Those are described in the next section. The table content in itself consists of one row per identified spectral feature, and provides a mixture of information on the detected species and transition, as well as on the line characteristics. The latter are based on a Gaussian fit of the detected spectral feature. Table 2 gives the details about the columns present in the data-set.

Header Parameter	**Description**
obsid	Observation ID
sourceId	Name of the source
longitude	Right Ascension of the source in degrees
latitude	Declination of the source in degrees
coordinateSystem	Name of reference frame for ephemeris data
smoothingWidth	Size of smoothing box (in channels) when applicable
manualMaskMin_n	Lower limit of manual line mask #n (optional)
manualMaskMax_n	Upper limit of manual line mask #n (optional)
localStandardOfRest	Local standard of rest
type	Herschel product type
creator	Name of the S/W that produced the product
creationDate	Creation date of this product
description	Product description (here 'Herschel Spectral Line List Product')
instrument	Instrument attached to this product
modelName	Model name attached to this product
startDate	Start date of this product
endDate	End date of this product
formatVersion	Version of product format
author	Author of the product

---

[11] `http://archives.esac.esa.int/hsa/whsa/`

profile	Line extraction method
fluxUnit	Unit of the fluxes
backgroundType	Type of background determination
references	References, e.g. Herschel observations/products
explanatoryTest	Additional comments
ctype1	Wavelength type
cunit1	Unit of wavelength axis
fwhmEmission	Full Width at Half Maximum used to fit emission lines (km/s)
fwhmAbsorption	Full Width at Half Maximum used to fit absorption lines (km/s)

Table 1: List of Line Catalogue FITS header parameters.

Column Name	**Description**
obsid	Observation ID
name	Source Name (as given by proposer)
ra	Right Ascension (Eq. 2000.0) of the target
dec	Declination (Eq. 2000.0) of the target
observingMode	HIFI Observing mode name (here 'HifiSScan')
spectrograph	Name of spectrometer backend. Save for a handful of exceptions, it will always be "WBS-HV" as both polarisations should have been averaged to form the input spectrum
spectralResolution	Spectral resolution of input spectrum (MHz). For the default Line Catalogue, this corresponds to the native resolution (around 1 MHz), while for smoothed tables, it indicates the resulting spectral bin size after smoothing
database	Data-base from which the line spectroscopy parameters are taken (JPL or CDMS)
species	Name of the molecular or atomic species
transition	Name of the transition
restFrequency	Frequency of the line transition in the LSR (GHz)
eup	Upper energy level E_{up} of the line (K)
aij	Einstein coefficient $A_{\mathrm{i,j}}$ (1/s)
position	Fitted position of the line centroid (GHz)
stdPosition	Standard deviation error on the above fitted position (GHz)
sideBand	Name of the sideband frequency scale (here "SSB", i.e. Single Sideband)
v0	Position of the velocity centroid (km/s)
stdV0	Standard deviation error on the fitted centroid (km/s)
width[†]	Line width of the Gaussian line fit (km/s).
stdWidth[†]	Standard deviation error on the fitted line width (km/s).
flux[†]	Integrated intensity of the fitted line (K.km/s). Note that this number is negative in case of absorption lines because the spectra used to perform the fit have been baseline-subtracted.
stdFlux[†]	Standard deviation error on the integrated intensity (K.km/s).

background[†]	Continuum level estimated within the frequency interval used for the fit (K). Because the input spectra have had their baseline subtracted, this level is essentially zero.
stdBackground[†]	Standard deviation error on the above continuum level (K).
rms	RMS noise level of the spectra estimated from the spectrum on two windows of 50 MHz on either side of the line profile (K)
signalToNoiseRatio	Fitted line peak signal-to-noise ratio
evidence	Provides a quality measure of the Gaussian fit (based on a Bayesian statistics fitter)
peakPosition	Frequency position of the fitted line peak (GHz)
lowerFrequencyFitWindow	Lower frequency of the interval used to perform the Gaussian fit (GHz)
upperFrequencyFitWindow	Higher frequency of the interval used to perform the Gaussian fit (GHz)

Table 2: Description of the Line Catalogue table columns. [†]Parameter not provided in the case of smoothed tables (Section 4.2.2) or set to zero for obsids listed in Table 8 (see also Section 3.2).

5.2 Flags

A collection of flags have been used in order to pass on additional information to the users of the Spectral Line Catalogue. They are most often warnings about peculiarities either in the baseline-corrected data used as input, or in the table content. Table 3 lists all possible flags in the data-set, and their meaning.

Flag name	**Description**
resolutionFlag	Set to either ”native” or ”smoothed”, it indicates whether the catalogue was generated with the native spectral resolution data, or with smoothed input data. When smoothing was applied, the value of the smoothing width is given given in the header as smoothingWidth
highLineDensityFlag	Warns against spectra were the average line density is above 1 line/GHz. This can be indicative of less accurate line fit parameters due to sub-optimal baseline correction
baselineCorrectionFlag	Warns against possible sub-optimal baseline correction that could lead to less accurate line fit parameters
absLineFlag	Indicates that absorption lines are reported in catalogue. Line fit parameters for those should be treated with care as the automatic line masking on such lines can sometimes be sub-optimal
manualBaselineMasking	When set to true is indicates that additional line masks where added by the instrument experts in order to generate the baseline-corrected input data. In those cases, the masks are given in the header. Otherwise set to false (fully automated baseline correction)
blendedFlag	Indicates if the catalogue contains blended lines

v0Flag	This flag will provide warning and information about table entries for which the fitted line centroid deviate noticeably from the mean centroid computed over the whole table. When such entries were not removed from the table, it is most likely because the line identification is considered reliable, but the line centroid fit got altered for a particular reason. The flag provides further hints as to why it could have been the case.
fswFlag	Indicates that the input data were obtained in Frequency Switching mode. Such spectra may suffer from residual spectral ghosts
freqGroupingFlag	Indicates that the input data were obtained with the frequency grouping option. Such spectra may suffer from residual baseline artefacts, potentially affecting the line detection
brightLineFlag	Indicates that spectral ghosts due to bright lines are present in the standard pipeline products, and that dedicated bright line flags were involved in the generation of the baseline-corrected spectra used to build the catalogue
offContaminationFlag	Indicate that some lines are affected by OFF position contamination. This will lead to either missed line detection, or inaccurate line flux computation from the line fitting

Table 3: List of possible flags added in the FITS file headers.

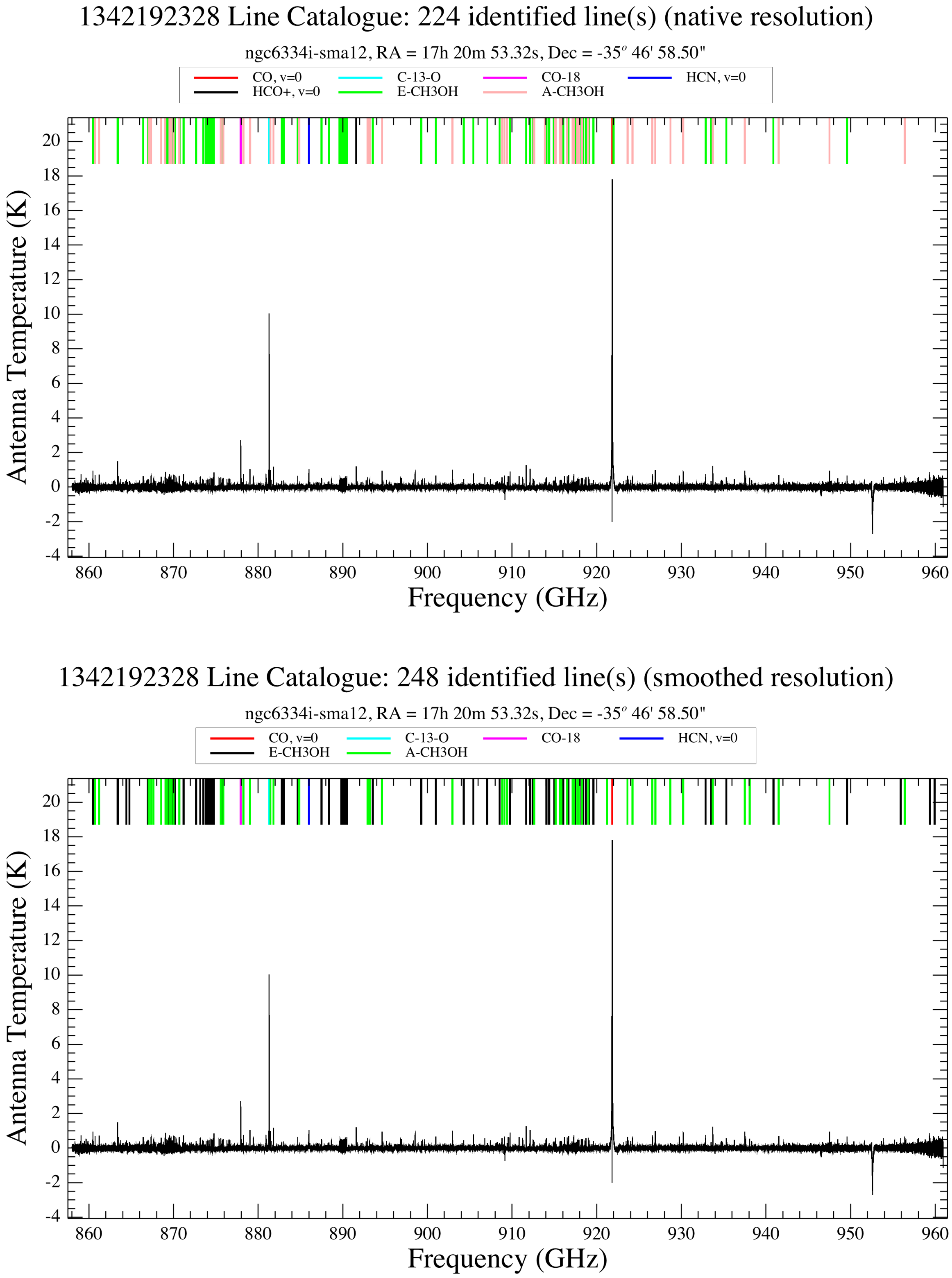

Figure 2: Illustration of a postcard for the Line Catalogue tables associated with obsid 1342192328. In this particular case, both a catalogue originating from native and smoothed resolution spectra have been generated, and respective postcards are being produced. As in most cases, there are slightly more line being detected/identified in catalogue formed from the smoothed data.

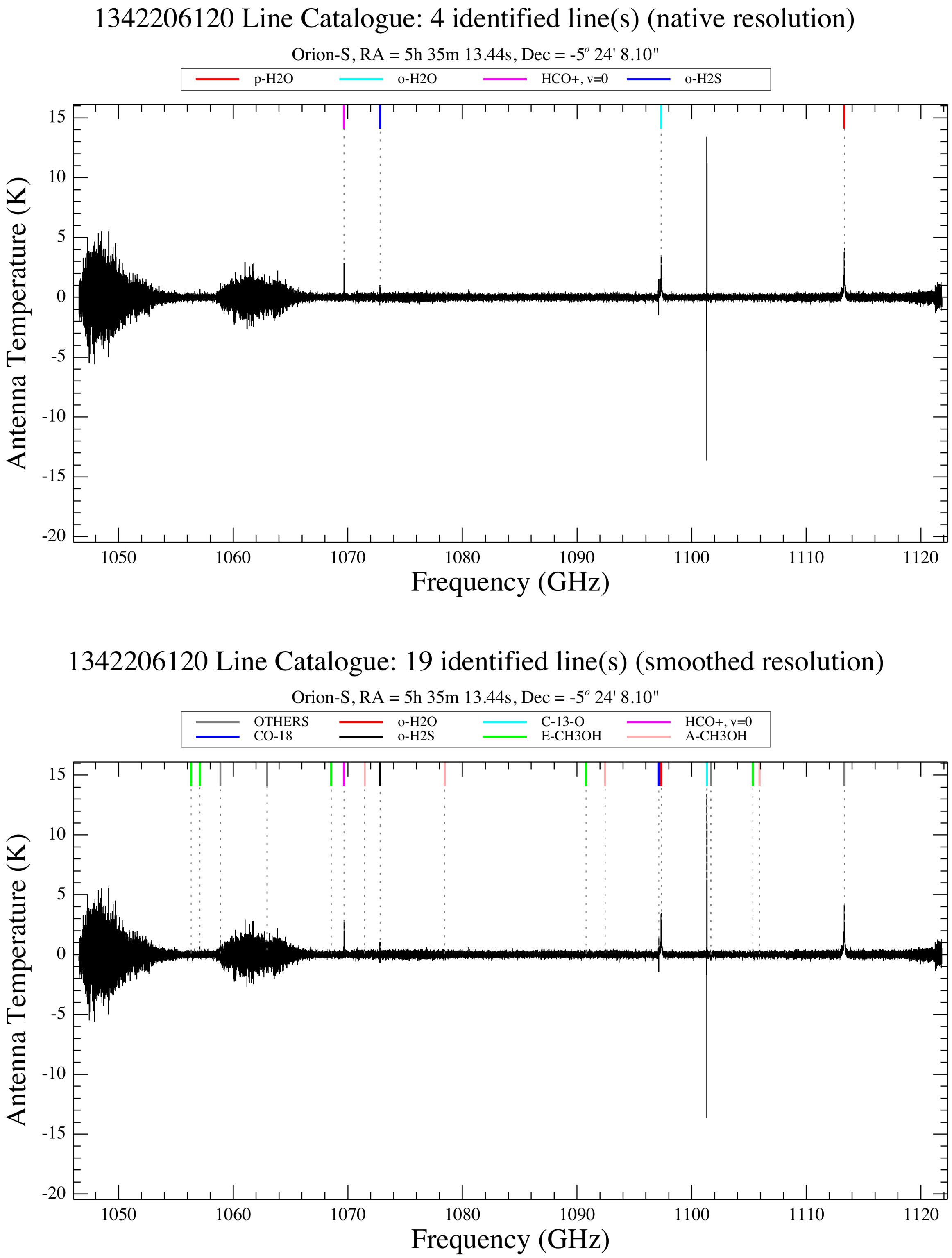

Figure 3: Illustration of a postcard for the Line Catalogue tables associated with obsid 1342206120 (Orion-South), where complex line profiles preclude the detection of certain (high SNR) species at native resolution (up), while they are picked up in the smoothed data. See text for further details.

6 Line Catalogue Statistics

We provide here some top-level statistics and associated plots about the content of the HIFI Spectral Line Catalogue. The numbers below correspond to the native resolution catalogue. The numbers for the smoothed catalogue are smaller, simply because fewer observations have been considered in this sub-sample.

- the catalogue contemplates 74 individual sources
- when combining those, a total of 101 unique species have been identified
- the combined catalogue contains 7594 entries in total, spanning 3065 unique transitions.
- there are, however, 1751 entries corresponding to blended identification, i.e. entries that share the same fitted parameters as another entry in the same source.

Figure 4 illustrates the distribution of detection per species for both the native and smoothed resolution catalogues. This simply indicates the number of occurrence of a given species detection, when accumulating all the relevant transitions. Note surprisingly, the most frequently reported molecules (methanol, $C^{17}O$) are those with a very large number of transition degeneracy, or having hyperfine structure. One should bear in mind that a large fraction of the latter are often tabulated as blends. A more detailed statistical analysis of the catalogue will be presented in Caux et al. (in preparation).

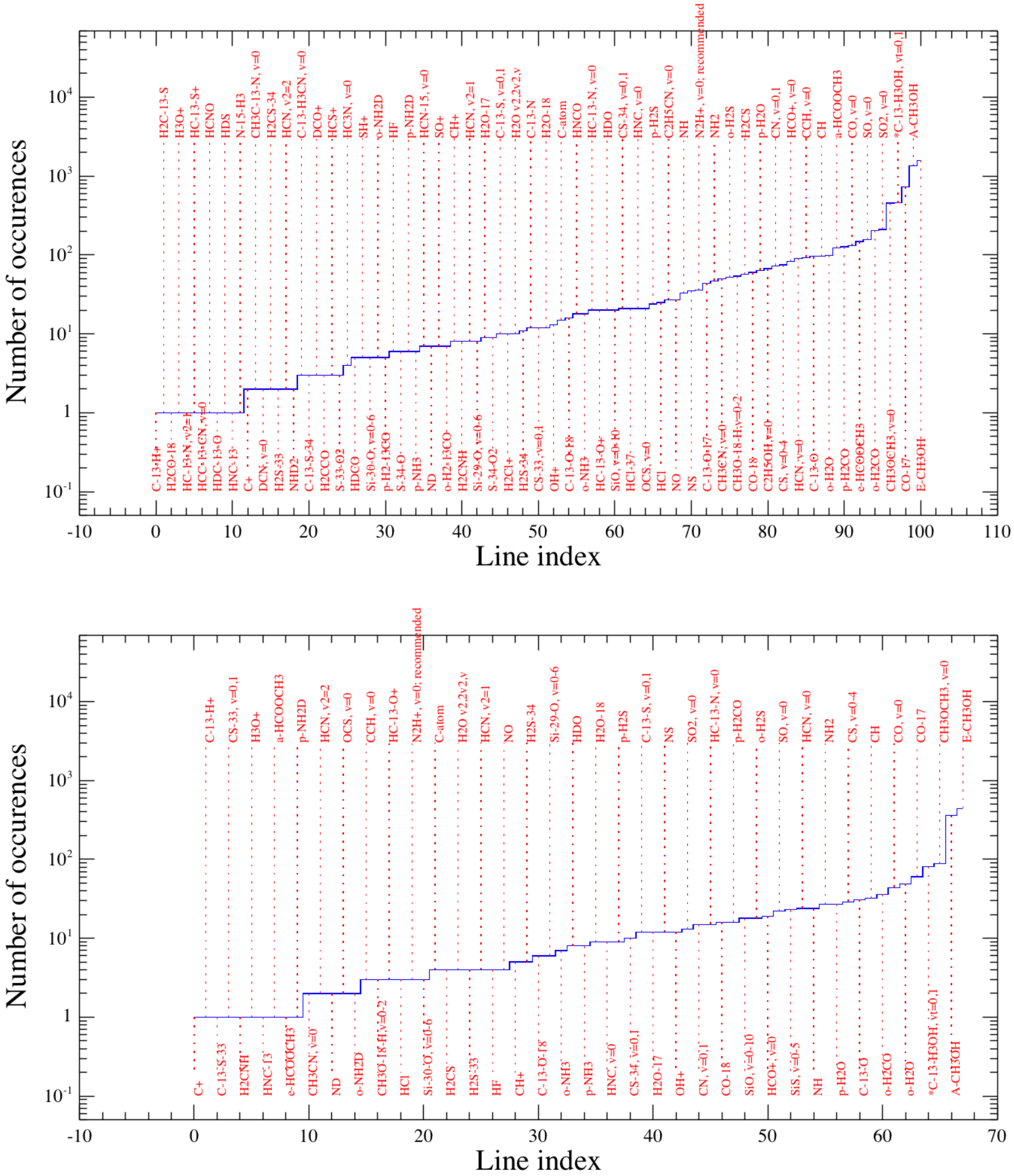

Figure 4: Cumulative number of detection for a given species in the combined catalogues at native (upper plot) and smoothed (lower plot) resolution. The line index is just an arbitrary coding for each species present in the catalogue (here ordered in increasing number of occurrence.

7 Validation of the catalogue

7.1 Consistency checks and removal of false positive

In order to check the validity of a given line identification, we considered all the observations available for a given source. We first checked the likeliness of a given molecular transition by confirming the presence of other transitions in the same source at other frequencies. Using CASSIS we then checked the likeliness of a given transition by comparing our findings with that of an LTE model for that given molecule and source. We had a fairly conservative approach and decided to discard line transitions that would not fit the LTE model properly. This was particularly important for lines exhibiting intensities just marginally above the SNR threshold of 5. In addition, we eyeballed all absorption line identifications and checked in particular those with upper energy levels in excess of 200 K. Finally, we checked for any significant outliers in fitted systemic velocities. We compared each velocity to the mean velocity and its standard deviation within a given obsid, and flagged those deviating from the mean by more than five times the standard deviation. When the discrepancy could be explained by peculiar line profiles (which in practice imply an improper fit of the real source velocity through the Gaussian fit), a flag was added to the data FITS header (see also Section 5.2).

The above checks lead to the removal of several dozens of false positives. We also noted the presence of obvious false negatives, where strong lines would be missed by the algorithm. The majority of those cases correspond to lines exhibiting complex profiles (see e.g. Fig. 3), which could be circumvented to some extent by generating an alternative catalogue, making use of smoothed data.

7.2 Comparison to published data

We also compared the output of our Line Catalogue to that published in the literature for the observations towards Orion South. The HIFI spectral survey and a line catalogue was published by Tahani et al. 2016 (15). This line-of-sight has the advantage of providing a relatively line rich spectrum, which is very adequate to check the ability of our catalogue to detect a large variety of species and transitions. The line lists reported by Tahani were kindly provided to us by the author in electronic form and a direct comparison could be made. Their line identification allows transitions up to E_{up}=1500 K, and let the line velocity be at ± 1.5 km/s from an assumed systemic velocity of 7 km/s. Tahani et al. reported two line lists:

- a list of line detections with SNR above 3 (thereafter called ”3-sigma” line list): this list is limited to the species name (i.e. no transition is given), the frequency (which we interpreted as the rest frequency of the transition) and the antenna temperature. The comparison is here based matches between the species name, **and** the proximity of rest frequencies (within 10 MHz)
- a list of line detections with SNR above 5 (thereafter called ”5-sigma” line list): this list provides both species and detailed transition names, as well line peak intensity, integrated intensity, and line width. The comparison is here based on the matches between the unique combination of species and transition name

No. of entries	No. of unique species	No. of unique transitions	No. of blended entries
Native resolution HIFI spectral line catalogue (14 obsids)			
551	47	506	189
Smoothed resolution HIFI spectral line catalogue (4 obsids)			
27	13	27	2
Tahani et al. 5-sigma line list			
404	40	388	16
Tahani et al. 3-sigma line list			
724	53	707	17

Table 4: Statistics of the native and smoothed resolution catalogue entries for Orion South. The fact that the total number of entries is larger than the number of unique transitions is related to the fact that the HIFI Spectral Scans of different bands can overlap in sky frequency, leading to the possible register of the same line multiple times. Smoothed catalogue tables were produced for four Orion South observations, summing up 13 uniques species, 12 of which were already detected in the native resolution catalogue. The only new species introduced by the smoothed catalogue is HF.

Comparison catalogue	Species unique to HIFI line catalogue	Species unique to Tahani's 5-sigma	Species unique to Tahani's 3-sigma
HIFI line catalogue	–	C^+, ^{13}CS	SiO, SH^+, CO^+, NH_2 $C^{13}S$, $HC^{18}O^+$, C^+ $H_2^{33}S$, p-$H_2^{18}O$, p-NH_3
Tahani's 5-sigma	HF, $C^{33}S$, OCS SO^+, $CH_3OCH_3{}^b$ H_2CNH, $C_2H_5OH^b$ $^{13}CH_3OH$, $HC^{15}N$, HCS^{+b}	–	CH_3OCH_3, $HC^{18}O^+$ HCS^+, $HC^{15}N$ SiO, SH^+, o-$H_2^{18}O$ CO^+, NH_2, $H_2^{33}S$ p-NH_3, p-$H_2^{18}O$, HF
Tahani's 3-sigma	$C^{33}S$, OCS SO^+, H_2CNH $C_2H_5OH^b$, o-$H_2^{18}O$ $^{13}CH_3OH$	NA	–

Table 5: List of species unique to a given catalogue compared to others. For species uniquely found in the HIFI line catalogue, we indicate when they correspond to a blended assignment with superscript ^{b}.

Table 4 summarises the statistics of the HIFI catalogue content for the Orion South observations, as well as those corresponding to the comparison catalogues from Tahani. Overall, some species appear to be missing in both catalogues (Table 5) – a noticeable no-match in the HIFI catalogue is the C^+ line, which is clearly detected, but was not picked by the automatic algorithm due to a complex line profile created by strong OFF position contamination (Tahani's observations were corrected from this artefact as part of their data processing). Not surprisingly, the number of matches increases when comparing the HIFI catalogue to Tahani's 3-sigma list (Sect. 7.2.1) instead of the 5-sigma one (Sect. 7.2.2) – see also Fig. 7. It is also evident from a study of the no-matches that they lie mostly at the faint intensity end (Fig. 6), where the SNR is probably just marginally above 3, and not necessarily the same in both studies. In terms of line parameters, those derived for the matched transitions show good agreement between the two catalogues (Fig. 5).

In summary, the comparison of our catalogue content with that published by Tahani confirms the large completeness level achieved by the automatic line detection and assignment engine, but also that the technique reaches its limitation when approaching SNR levels in the range 3 to 5, or line intensities under 0.1 K, as is evidenced by the match scores obtained for the respective line lists reported by Tahani. Finally, the vast majority of no-matches do apply to lines attributed to blends, which therefore correspond to alternative line assignments of a same spectral feature. By definition those must be used with special care and the choice of the most likely species and transition is left to the appreciation of the user.

The following sections provide further details about how our catalogue for HIFI observations made on Orion South match Tahani's two line lists.

7.2.1 Comparison with the 5-sigma line list

As described in Table 4, the total number of unique species when combining the native and smoothed catalogues amount to 48, while this number is 40 for the 5-sigma catalogue of Tahani. There are 38 species in common, meaning that 2 species are uniquely found in Tahani's 5-sgima line list, and 10 uniquely found in the HIFI line catalogue (see Table 5 for details).

HIFI catalogue vs Tahani 5-sigma

Out of the 506 unique transitions reported in the HIFI line catalogue, 373 have at least one match with the 5-sigma table (352 transitions have a unique match in the 5-sigma line list), of which 47 are actually reported as blends. Regarding the non-matched transitions, we note that 3/4 of those actually correspond to blended lines which share their fitted frequencies with other transitions featured as a match. As such, in the end a total of 56 transitions from the HIFI catalogue does not have any counterpart in Tahani's 5-sigma line list. It is, however, interesting to note that a match for almost each of those is found when extending the comparison to the 3-sigma line list (Fig. 7, and Section 7.2.2).

Fig. 5 illustrates how the line parameters of matched transitions compare between the two catalogues. The agreement is relatively good, with some outliers, especially in the fitted line width, which are mostly due to sub-optimal Gaussian fitting of non-Gaussian profiles. Lines identified as blend also show larger discrepancies. It is also interesting to note that the tabulated rest frequencies shows a bi-modal distribution, with the strongest component exhibiting a shift of about +0.5 km/s in Tahani's tables. The exact reason for this is unclear. Finally, a comparison between the fitted line centroid frequency and the rest frequency of the associated transition consistently peaks around 7 km/s, corresponding to the systemic velocity of the Orion South line-of-sight.

Tahani 5-sigma vs HIFI catalogue

When checking Tahani's 5-sigma line list with the HIFI catalogue, we note that 66 out of the 388 unique transitions from the Tahani 5-sigma line list has no match within the HIFI catalogue. Fig. 6 illustrates the line intensity distribution of the respective matches and no-matches transitions for this cross-check. The typical noise level in the Orion South spectral scans is of order 0.05 K, implying a 5-sigma detection level around 0.25 K. This is consistent with the location of the bulk of the matches for the respective line lists, as well as with the line peak intensity locus of the transition between matches and no-matches. As such, most non-matched lines lie at the faint end, where the SNR might be just too low for an automatic detection algorithm to work as efficiently as the more dedicated effort conducted in Tahani's study. There are some isolated exceptions at high intensity, but these are typically bright lines affected by strong OFF contamination (e.g. C^+) where the line profile prevented achievement of a proper line detection.

7.2.2 Comparison with the 3-sigma line list

Again, the total number of unique species when combining the native and smoothed catalogues amount to 48, while this number is 53 for the 3-sigma line list of Tahani. There are 42 species in common, meaning that 11 species are uniquely found in Tahani's 5-sigma catalogue, and 6 uniquely found in the HIFI line catalogue (see Table 5 for details).

HIFI catalogue vs Tahani 3-sigma

As indicated above, the comparison with this second line list cannot make use of the transition information, as this latter is not provided in Tahani's list. To circumvent that, matches were considered positive in case the rest frequencies were within 10 MHz of each others, **and** the species names would match. This time, only 9 transitions have no counter-part in the 3-sigma catalogue, while all blended lines having no direct match actually share their fitted frequency with that of a matched transition. This means that the vast majority of the non-blended no-matches with the 5-sigma line list could be matched in the 3-sigma line list. Fig. 7 illustrates the (HIFI catalogue) peak temperature locus of those complementary matches (cyan histograms on the right panel), compared with those that were still missing in the 5-sigma list (red histograms on the left panel). Not surprisingly, those lie at the faint intensity end, also confirming that most of the remaining no-matches with the 3-sigma line list belong to the weakest intensity bin of these diagrams.

Tahani 3-sigma vs HIFI catalogue

This time, the number of entries from the 3-sigma line list with no counterpart in the HIFI catalogue amounts to 330 transitions, which is not surprising given the different SNR threshold used in the respective studies. Fig 7.2.1 (right panel) confirms that a large fraction of those are indeed below the detection threshold ($5\sigma \simeq 0.25$ K) applied to the HIFI catalogue.

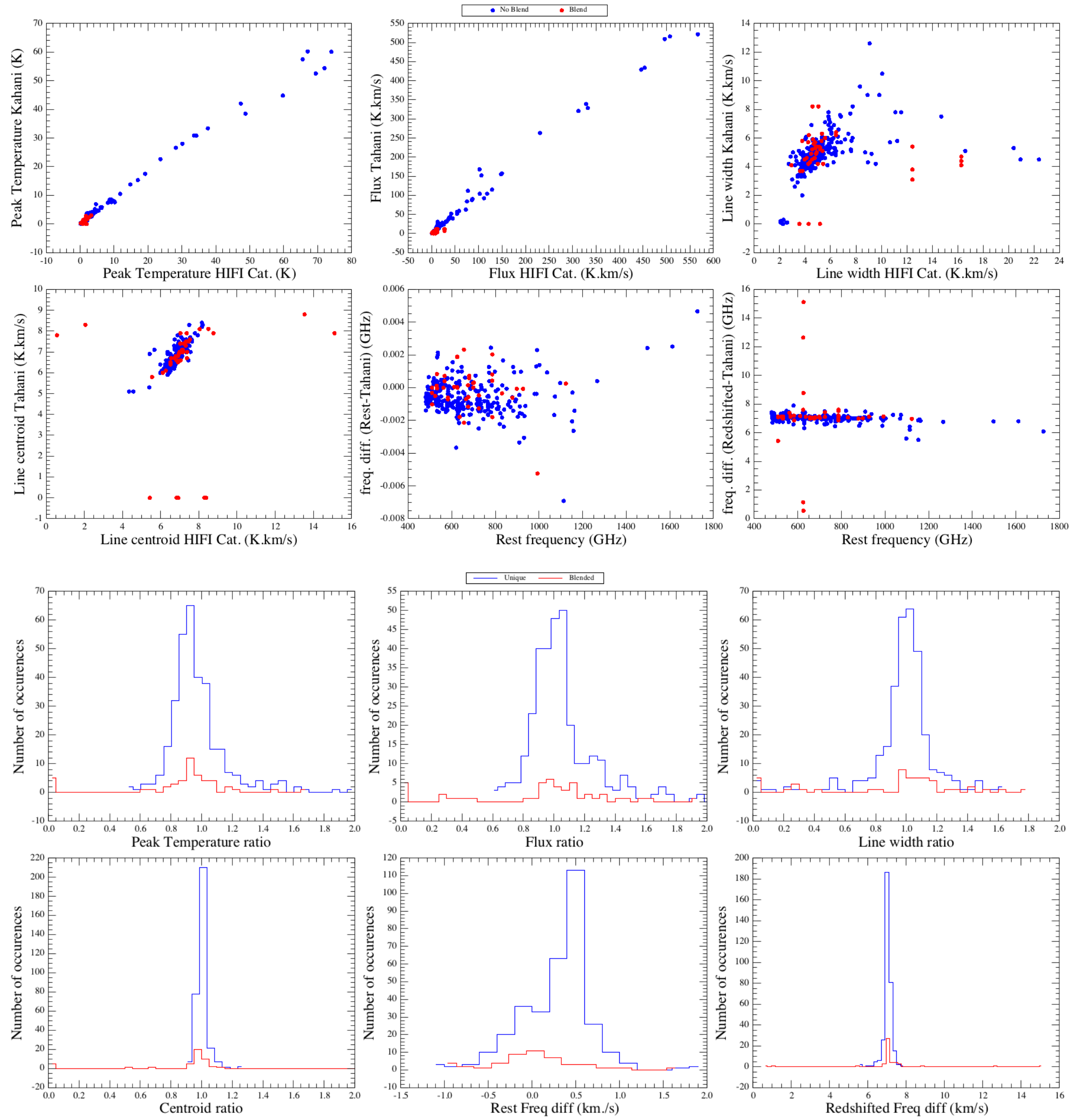

Figure 5: Comparison of the fitted line parameters for matches between the HIFI catalogue and Tahani's 5-sigma line list. *Upper panels*: from left to right, top to bottom: scatter plots of the line peak intensities, line integrated fluxes, line Full Width at Half Maximum, Line centroid (in the LSR), differences between rest frequencies, and differences between rest and fitted frequencies. *Lower panels*: same as above displayed as histograms of the line parameter ratios.

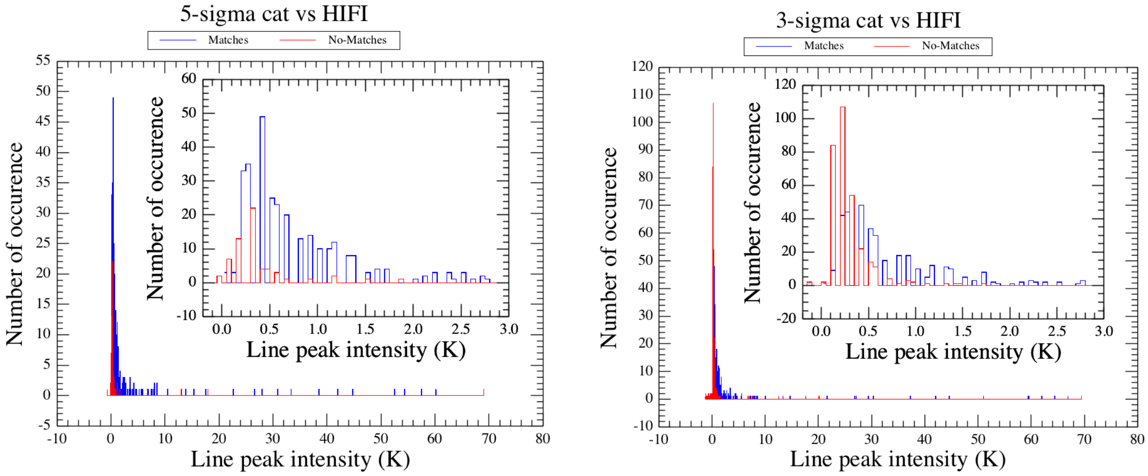

Figure 6: Line intensity distribution of the transitions from the Tahani line lists (5-sigma on the left, 3-sigma on the right) that have a match in the HIFI catalogue (blue) and those that do not have a match in the HIFI catalogue (red).

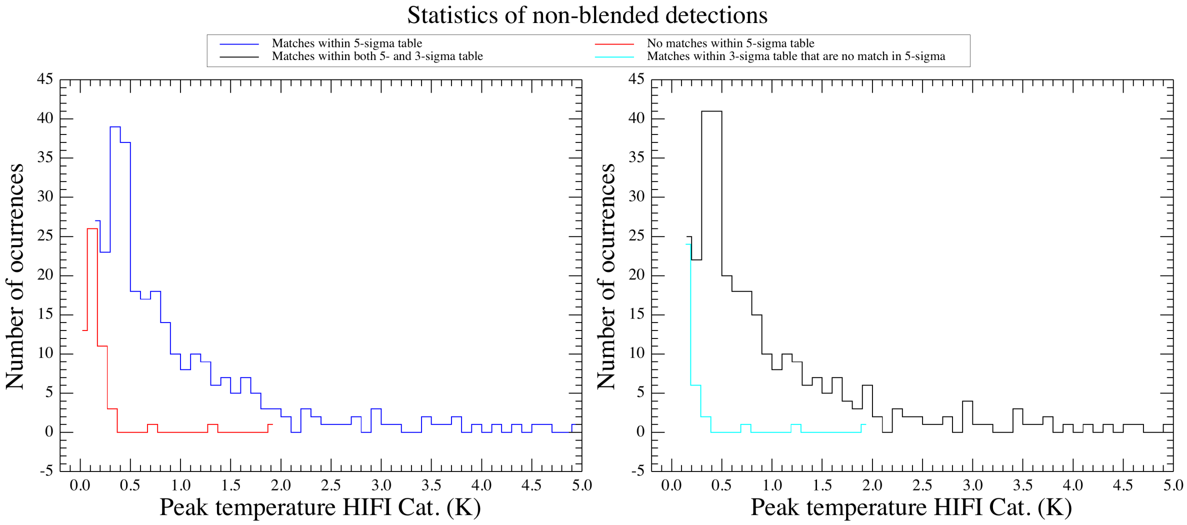

Figure 7: Histograms of the HIFI catalogue fitted peak temperatures of non-blended transitions. *Left*: For lines matched (blue) and un-matched (red) with the 5-sigma line list (blue). *Right*: For lines matched in both the 5-sigma and 3-sigma line lists (black) and for line matched in the 3-sigma line list, but un-matched in the 5-sigma line list (cyan). All histograms use a 0.1K bin.

8 Possible improvements

As indicated in Section 4.3.4, the list of U-lines resulting from the line identification exercise will not be part of the present release. Once the necessary checks of the corresponding entries have been performed, the idea is to make those tables available as an additional component of the delivery. On top of this, additional work is needed in order to assign line identification to high signal-to-noise spectral features clearly missed by the algorithm. This concerns essentially lines with a complex profile (broad lines, P-Cygni profiles, strong contamination by OFF position, absorption lines, etc) where a proper detection could be not performed. Finally, the combination of the two previous items should lead to the enhancement of the line template list, to increase the list of possible transitions present in the HIFI spectral scans.

References

[1] Alcolea, J., Bujarrabal, V., Planesas, P. et al., 2013, A&A 559, A93

[2] Sanchez-Contreras, C., Velilla Prieto, L., Agúndez, M. et al., 2015, A&A 577, A52

[3] Cernicharo, J.; Waters R., Decin L. et al., 2010, A&A, 521, L8

[4] Comito C., Schilke, P., Rolffs, R. et al., 2010, A&A 521, L38

[5] Crockett, N.; Bergin, E., Neill, J. et al., 2014, ApJ, 787, 112

[6] Crockett, N.; Bergin, E., Neill, J. et al., 2015, ApJ, 806, 239

[7] de Graauw, T., Helmich, F. P., Phillips, T. G et al., 2010, A&A, 518, L6

[8] Goldsmith P., Liseau R., Bell T., et al., 2011, ApJ 737, 96

[9] Hartogh, P., Jarchow, C., Lellouch, E. et al., 2010, A&A 521, L49

[10] Morris P., Gull, T., Hillier, D. et al., 2017, ApJ, 842, 79

[11] Müller, H., Thorwirth, S., Roth, D. et al., 2001, A&A 370, L49

[12] Neill J., Bergin E., Lis, D. et al., 2014, ApJ, 789, 8

[13] Pickett, H., 1991, J. Molec. Spectroscopy, 148, 371

[14] Shipman R., Beaulieu S., Teyssier, D. et al., 2017, A&A 608, A49

[15] Tahani K., Plume R., Bergin E.A etl al., 2016, ApJ 232, 12

[16] Teyssier, D., 2016, *HIFI Reference Position Spectra Data Products: Release notes*, HERSCHEL-HSC-DOC-2111.

[17] Teyssier, D., 2017, *HIFI Spectral Scans Highly-Processed Data Products: Release Notes*, HERSCHEL-HSC-DOC-2198.

[18] Zernickel, A., Schilke, P., Schmiedeke, A. et al., 2012, A&A 546, A87

9 Appendix

9.1 List of source considered in the Line Catalogue Generation, and the total frequency coverage of the corresponding observations

Source Name	Covered frequency ranges (GHz)
IC1795-1	[555.4, 631.2] [958.8, 1050.2]
W3(H2O)	[647.8, 676.3] [1139.9, 1168.3]
L1448MM-1	[647.8, 676.3] [1139.9, 1168.3]
irc+50096	[525.8, 551.9] [602.8, 629.9] [690.8, 720] [778.8, 821.1] [866.8, 898.2] [949.9, 987.2] [1053.9, 1089.1] [1141.9, 1179.2]
NGC1333 IRAS4A	[626, 800.9] [1139.9, 1168.3]
NML Tau	[479.5, 628.5] [641.7, 719.5] [746.4, 775.3] [799.1, 843.3] [858.1, 933.8] [966.2, 1042.2] [1058.7, 1220.2] [1434.6, 1535] [1608.5, 1695] [1713.2, 1797.4] [1832.3, 1906.8]
R Dor	[479.5, 560.2] [641.7, 719.5]
Orion S	[479.5, 1280] [1426.5, 1532.7] [1573.3, 1906.7]
OMC-1 (Peak 1)	[484.7, 501.9] [520.8, 545.1] [771.3, 788.5]
Orion Bar CO+ peak	[479.5, 1279.7] [1427.2, 1562.7] [1573.3, 1906.7]
OMC-2 FIR4-1	[479.5, 1243.9] [1489.1, 1508.9] [1535.4, 1556.6] [1592.9, 1638.8] [1650.5, 1680.3] [1707.2, 1734.9] [1756.8, 1777.6]
CIT 6	[525.8, 551.9] [602.8, 629.9] [690.8, 720] [778.8, 821.1] [866.8, 898.2] [949.9, 988.1] [1053.9, 1089.1] [1141.9, 1179.2]
CarinaN-point-I F	[958.8, 1050.2]
Garradd (C/2008 Q3)	[541.9, 559.5]
IRAS 12326-6245	[514.8, 547.1] [969.9, 1001.2] [1020.4, 1039.9] [1146.8, 1181.1] [1205.8, 1226.5]
Y CVn	[525.8, 551.9] [602.8, 629.9] [690.8, 720] [778.8, 820.2] [866.8, 898.2] [949.9, 988.1] [1053.9, 1089.1] [1141.9, 1179.2]
G316.81-0.06	[512.9, 538.2]
II Lup	[479.5, 1121.9]
G323.74-0.26	[512.9, 538.2]
G327.3-0.6	[514.8, 547.1] [969.9, 1001.2] [1020.4, 1040] [1146.8, 1181.1] [1205.9, 1226.5]
G331.28-0.19	[512.9, 538.2]
rho Oph A	[479.5, 508] [771.1, 776.5] [783.1, 788.5]
iras16293-2422A B-1	[479.5, 1238.4] [1481, 1510.9] [1573.2, 1798.6]
IRAS16293.2422	[554.5, 628.5]
16293E	[479.5, 636.5]
NGC6334i	[750.9, 771.4] [1223.9, 1241.5] [1840.1, 1857.1]
ngc6334i-sma12	[479.5, 1279.8] [1575.1, 1906.7]
NGC6334I(N)-SMA 1-1	[626.1, 801.9]
IRAS 17233-3606	[514.8, 547] [969.9, 1001.2] [1020.4, 1039.9] [1146.8, 1181.1] [1205.8, 1226.5]
GCM+0.693-0.027	[647.8, 676.3] [1140, 1168.3]
G0.55-0.85	[512.8, 538.2]
G5.90-0.43	[512.9, 538.2]
G5.90-0.44	[512.8, 538.2]

G8.14+0.23	[512.8, 538.2]
G9.62+0.19	[512.8, 538.2]
G8.67-0.36	[512.8, 538.2]
G10.47+0.03	[514.8, 547] [969.9, 1001.2] [1020.4, 1039.9] [1146.8, 1181.1] [1205.8, 1226.5]
G10.30-0.15	[512.8, 538.2]
G10.34-0.14	[512.8, 538.2]
G10.32-0.16	[512.8, 538.2]
G10.6-0.4 (W31 C)	[750.9, 771.4] [877.9, 956.1] [1108, 1232.3] [1232.5, 1238.2] [1645.1, 1678.9] [1840.1, 1857.1]
GAL 12.21-0.10	[647.7, 676.3] [1139.8, 1168.3]
GAL 012.91-00.2 6	[647.7, 676.3] [1139.8, 1168.3]
GAL 19.61-0.23	[647.8, 676.3] [1139.8, 1168.3]
G23.44-0.18	[512.8, 538.2]
G24.79+0.08	[512.8, 538.2]
G25.83-0.18	[512.8, 538.2]
GAL 31.41+0.31	[647.8, 676.3] [1139.8, 1168.3]
GAL 034.3+00.2	[647.7, 676.3] [1139.8, 1168.3]
W49N	[968.7, 986.9]
w51-e1/e2	[1573.3, 1702.7]
AFGL2591	[479.5, 1238.4]
GAL79.29+00.46	[555.4, 636.0]
W75N	[647.8, 676.3] [1139.9, 1168.3]
DR21	[968.7, 986.9] [1059.9, 1120.8] [1823.1, 1844.9]
LDN1157-B1	[479.5, 1178.1] [1191.9, 1228.1] [1595.1, 1674.9]
V Cyg	[525.8, 551.9] [602.8, 629.9] [690.8, 720] [778.8, 821.1] [866.8, 898.2] [949.9, 988.1] [1053.9, 1089.1] [1141.9, 1179.2]
NGC7023	[521.8, 568.4] [572.2, 580.4] [1050.6, 1121.2]
NGC7027	[509.5, 545.1] [556.4, 591.2] [602.8, 635.7] [670.3, 722.2] [765.5, 800.9] [807, 850.1] [869.4, 900.1] [913.1, 944.1] [961.5, 996.4] [1022.8, 1056.5] [1227.1, 1279.9] [1459.6, 1510.9] [1577, 1605] [1633.1, 1662.9] [1713.1, 1739]
S CEP	[525.8, 551.9] [602.8, 629.9] [690.8, 720] [778.8, 821.1] [866.8, 898.2] [949.9, 988.1] [1053.9, 1089.1] [1141.9, 1179.2]
NGC7538 IRS1	[1058.7, 1116]
CRL3068	[525.8, 551.9] [602.8, 629.9] [690.8, 720] [778.8, 821.1] [821.1, 866.8] [866.8, 898.2] [949.9, 987.4] [1141.9, 1179.2]

Table 6: Frequency range of available HIFI Spectral Surveys.

9.2 List of observations discarded for the generation of baseline-corrected Spectral Scan and Spectral Line Catalogues

Source	Observation ID	Reference for supplementary material
Orion KL	1342190871, 1342190872, 1342191504, 1342191592, 1342191601, 1342191649, 1342191725, 1342191727, 1342191728, 1342191755, 1342192220, 1342192329, 1342192562, 1342192563, 1342194176, 1342194178, 1342194540, 1342194732, 1342205334, 1342216387, 1342266895, 1342192673, 1342192674, 1342194733	(5), (6), HEXOS UPDP[12]
SgrB2(M)	1342191482, 1342191565, 1342191680, 1342192546, 1342204723, 1342204739, 1342205848, 1342206455, 1342206640, 1342215935, 1342216702, 1342218200, 1342243701, 1342243702, 1342251112, 1342192656, 1342206501, 1342266904	(4), HEXOS UPDP[12]
SgrB2(N)	1342204692, 1342204703, 1342204731, 1342204812, 1342204829, 1342205491, 1342205855, 1342206364, 1342206370, 1342218198, 1342266903, 1342206498, 1342206643, 1342215934, 1342216701	(12), HEXOS UPDP[12]
SgrB2(S)	1342190897, 1342191483, 1342191740, 1342190899, 1342190900, 1342191684	HEXOS UPDP[12]
SgrA*	1342230279, 1342230394, 1342239594, 1342239609, 1342243685, 1342243700, 1342243707, 1342251185, 1342230396, 1342243697, 1342243705, 1342251446, 1342252173, 1342253143, 1342253145, 1342266608	Goicoechea et al. in prep.
IRC+10216	1342196414, 1342196423, 1342196473, 1342196475, 1342196483, 1342196514, 1342196516, 1342196518, 1342196541, 1342196543, 1342196566, 1342196574, 1342196590, 1342210102, 1342210742, 1342210754, 1342221429	(3) and upcoming UPDP

---

[12] `http://archives.esac.esa.int/hsa/legacy/UPDP/HEXOS_HIFI/ReleaseNote/hexos_release_note.pdf` and `http://archives.esac.esa.int/hsa/legacy/UPDP/HEXOS_HIFI/Data/`

VY CMa	1342228611, 1342230402, 1342231504, 1342244486, 1342244491, 1342244512, 1342244631, 1342244789, 1342244791, 1342244945, 1342244960, 1342244962, 1342231467, 1342244537, 1342244610	Quintana-Lacaci et al. in prep. and upcoming UPDP
OH231.8+4.2	1342231503, 1342231526, 1342231532, 1342244632, 1342244942, 1342244944, 1342244964, 1342245270, 1342245371	Sanchez-Contreras et al. in prep., and upcoming UPDP
η Carina	1342181171, 1342180817, 1342180818, 1342180819, 1342181165, 1342181170, 1342232978, 1342232982, 1342235769, 1342235809, 1342235831	(10) and upcoming UPDP
Mars	1342194492, 1342194496, 1342194545, 1342194685, 1342194693, 1342194742, 1342194744, 1342194746, 1342194748, 1342194751, 1342194753, 1342235092	Upcoming UPDP from HssO (9)

Table 7: List of sources and observations discarded for the generation of Spectral Line Catalogues (see text for details).

Source	**Observation ID**	**Reference for supplementary material**
NGC6334I	1342191481, 1342191561, 1342206085, 1342206087, 1342251671	(18)
OMC-1	1342191503, 1342191754, 1342192215	(8), HOP UPDP[13]
W49N	1342229905	Unknown
W3IC1795	1342190881	Unknown
G327.3	1342238588	Unknown
IRAS17233-3606	1342239610	Unknown
G10.47	1342242817	Unknown

Table 8: List of sources and observations that were only considered for line identification, but for which no line fit parameter was provided due to imperfect baseline correction (see text for details).

9.3 Line parameters derived for all individual sources in the Line Catalogue

Table 9: Various line parameters derived for all individual sources in the Line Catalogue. ”N/A” indicates either that no standard deviation value could be derived as only one transition was reported in the source, or that the source corresponds to one of those identified as not providing accurate fitted line flux and width (Table 8).

HSA Source Name	RA hh:mm:ss	Dec deg:mm:ss	v_0 km/s	Median v_{LSR} km/s	σ v_{LSR} km/s	Median FWHM km/s	σ FWHM km/s	Min FWHM km/s	Max FWHM km/s	Nb Species observed	Nb Transitions observed	Total GHz covered
IC1795-1[a]	02 25 43.72	62 06 11.48	-38.90	-38.44	1.77	4.43	3.81	4.19	18.64	9	35	167
W3(H2O)	02 27 4.39	61 52 23.8	-46.90	-46.57	1.31	5.84	1.95	3.42	19.48	18	97	57
L1448MM-1[a]	03 25 38.83	30 44 5.11	6.00	5.25	1.54	1.68	6.40	1.40	12.63	3	3	57
irc+50096	03 26 29.45	47 31 49.17	-15.60	-15.29	N/A	15.78	N/A	N/A	N/A	1	1	266
NGC1333 IRAS4A[a]	03 29 10.36	31 13 31.98	7.10	6.95	3.12	2.34	12.56	1.85	27.24	5	5	203
NML Tau[b]	03 53 28.84	11 24 22.33	35.80	34.91	0.94	21.01	6.70	1.91	26.55	10	34	959
R Dor	04 36 45.58	- 62 04 37.35	6.00	7.66	0.56	6.64	1.02	5.01	9.22	10	26	158
Orion S	05 35 13.44	- 05 24 6.81	6.90	6.80	0.86	4.64	2.39	1.76	22.39	47	506	1240
OMC-1 (Peak 1)[a]	05 35 13.78	- 05 22 8.36	9.60	8.91	5.32	N/A	N/A	N/A	N/A	39	301	58
Orion Bar CO+ peak	05 35 20.73	- 05 25 12.92	10.50	10.58	0.33	2.15	0.92	1.29	10.01	29	171	1269
MWZ90OMC-2 FIR 4-1	05 35 27.13	- 05 09 51.38	11.90	11.92	1.13	4.65	2.96	1.45	20.62	32	430	930
CIT 6[a]	10 16 2.28	30 34 19.01	-1.30	-1.22	2.25	18.62	1.50	17.58	21.91	5	10	267
CarinaN-point-IF	10 43 35.08	-59 34 5.75	-12.00	-11.53	N/A	4.89	N/A	N/A	N/A	1	1	91
Garradd (C/2008 Q3)	12 33 48.37	-05 53 36.46	0.00	0.03	N/A	1.22	N/A	N/A	N/A	1	1	18
IRAS 12326-6245	12 35 35.0	-63 02 32.16	-39.60	-39.58	2.14	5.84	3.71	3.05	29.73	33	181	138
Y CVn[a]	12 45 7.84	45 26 25.11	22.30	22.42	2.11	12.56	1.90	7.89	12.56	2	8	266
G316.81-0.06	14 45 26.77	-59 49 14.86	-39.30	-39.38	0.57	4.59	0.72	3.65	5.56	4	9	25
II Lup	15 23 5.12	-51 25 58.81	-14.70	-14.34	3.11	25.44	4.20	18.96	43.16	12	33	642
G323.74-0.26	15 31 45.77	-56 30 49.63	-49.60	-50.37	0.37	5.48	1.29	4.63	8.21	5	7	25
G327.3-0.6[b]	15 53 7.85	-54 37 6.58	-44.20	-44.29	2.62	N/A	N/A	N/A	N/A	50	307	138
G331.28-0.19	16 11 26.91	-51 41 56.77	-87.70	-87.74	0.95	6.57	2.91	4.43	16.12	8	15	25
rho Oph A	16 26 27.75	-24 23 55.86	3.80	3.61	0.20	0.96	0.68	0.96	3.12	4	18	39
iras16293-2422AB-1	16 32 22.65	-24 28 33.12	3.90	3.91	0.72	4.81	2.44	0.81	18.29	44	378	1014
IRAS16293-2422	16 32 22.73	-24 28 33.56	4.10	3.89	1.70	4.50	3.48	2.49	16.27	19	56	74
16293E	16 32 28.47	-24 29 3.14	3.70	4.02	0.90	1.29	0.71	1.27	4.35	10	24	157
NGC 6334I	17 20 53.23	-35 46 59.14	-7.80	-8.23	1.47	N/A	N/A	N/A	N/A	10	59	55
ngc6334i-sma12[a]	17 20 53.47	-35 46 59.6	-7.80	-7.94	1.85	N/A	N/A	N/A	N/A	71	2838	1132
NGC6334I(N)-SMA1-1[a]	17 20 55.18	-35 45 3.9	-4.00	-3.95	0.67	5.06	3.08	2.49	22.20	16	108	176
IRAS 17233-3606[a]	17 26 42.51	-36 09 17.91	-2.90	-3.29	2.12	N/A	N/A	N/A	N/A	43	239	138
GCM+0.693-0.027	17 47 21.87	-28 21 26.69	69.80	69.29	1.89	18.20	3.90	15.18	24.97	7	8	57
G0.55-0.85	17 50 14.53	-28 54 30.34	16.80	16.91	1.18	6.27	1.20	3.19	9.85	13	51	25
G5.90-0.43	18 00 40.92	-24 04 20.15	6.50	6.64	0.52	4.55	0.84	3.88	6.42	6	10	25
G5.90-0.44	18 00 43.9	-24 04 46.67	9.00	8.85	0.78	7.04	2.93	2.73	10.42	3	6	25
G8.14+0.23	18 03 0.82	-21 48 9.69	20.00	20.31	1.33	4.96	2.11	4.83	10.63	3	7	25
G9.62+0.19	18 06 14.81	-20 31 36.77	4.42	4.23	1.48	5.92	1.90	4.29	12.95	14	39	25
G8.67-0.36	18 06 18.97	-21 37 32.17	34.85	34.91	0.32	4.10	1.97	3.18	9.62	8	10	25
G10.47+0.03	18 08 38.23	-19 51 50.19	66.50	66.43	1.71	N/A	N/A	N/A	N/A	21	112	138
G10.30-0.15	18 08 55.51	-20 05 57.81	13.40	13.21	0.64	5.87	1.22	3.50	7.44	5	10	25
G10.34-0.14	18 08 60.0	-20 03 35.24	12.50	12.81	0.64	4.29	4.61	2.85	12.63	3	4	25
G10.32-0.16	18 09 1.5	-20 05 7.51	12.40	12.56	0.69	4.41	1.13	3.47	5.84	3	5	25
G10.6-0.4[a]	18 10 28.71	-19 55 50.79	-2.40	-3.07	1.89	6.95	2.68	5.55	13.29	8	17	279
GAL 12.21-0.10	18 12 39.76	-18 24 21.19	23.90	24.49	0.70	7.10	1.06	5.76	11.35	9	23	57
GAL 012.91-00.26	18 14 39.0	-17 52 3.18	38.00	37.63	0.80	5.41	1.54	3.71	12.53	10	38	57
GAL 19.61-0.23	18 27 38.03	-11 56 42.47	41.10	41.66	1.83	9.14	3.39	4.81	20.07	14	46	57
G23.44-0.18	18 34 39.1	-08 31 32.7	101.40	101.84	0.46	9.93	3.84	7.21	12.65	2	2	25
G24.79+0.08	18 36 12.33	-07 12 10.78	111.07	110.97	1.11	7.02	2.06	4.14	12.64	10	24	25
G25.83-0.18	18 39 3.57	-06 24 10.35	93.80	93.70	0.33	11.09	5.39	7.28	14.90	2	2	25
GAL 31.41+0.31	18 47 34.58	-01 12 43.07	96.90	96.88	0.73	6.58	1.22	4.07	11.04	15	59	57
GAL 034.3+00.2	18 53 18.5	01 14 56.76	58.50	58.56	1.28	6.27	2.10	3.58	15.47	28	170	57
W49N[a]	19 10 13.22	09 06 12.04	8.00	10.20	11.10	N/A	N/A	N/A	N/A	6	10	18
w51-e1/e2[b]	19 23 43.76	14 30 28.44	58.44	58.74	N/A	20.13	N/A	N/A	N/A	1	1	129
AFGL2591	20 29 24.82	40 11 20.06	-5.50	-5.53	0.90	3.56	2.16	2.28	16.85	28	166	759
Gal 79.29+00.46	20 31 42.04	40 21 59.12	0.80	0.90	N/A	1.98	N/A	N/A	N/A	1	1	81
W75N	20 38 35.92	42 37 22.58	9.40	9.69	0.88	6.40	2.12	4.03	15.68	13	61	57
DR21[a]	20 39 1.09	42 22 49.12	-2.60	-2.77	0.36	10.01	3.46	5.57	15.05	4	4	101
LDN1157-B1[a]	20 39 10.18	68 01 9.68	-0.05	0.27	1.61	5.66	2.78	3.20	14.65	14	44	815
V CYg	20 41 18.23	48 08 28.7	14.00	14.32	0.44	13.61	2.20	10.34	18.13	4	9	267
NGC 7023 (H2)[a]	21 01 32.29	68 10 27.14	2.30	2.36	0.27	1.14	0.49	0.86	2.90	7	27	125
NGC7027[a]	21 07 1.6	42 14 10.41	30.20	27.71	4.41	26.18	6.06	19.63	36.87	3	7	552
S CEP	21 35 13.24	78 37 28.78	-13.20	-13.94	0.53	23.58	0.56	23.18	23.98	1	2	267
NGC7538 IRS1	23 14 16.51	61 28 8.78	-58.20	-57.67	1.03	4.30	1.11	2.99	7.71	7	7	57
CRL3068	23 19 12.38	17 11 35.75	-30.60	-30.68	0.60	15.27	1.39	13.84	17.84	3	6	277

---

[a] Source with outflows

[b] Source with masers