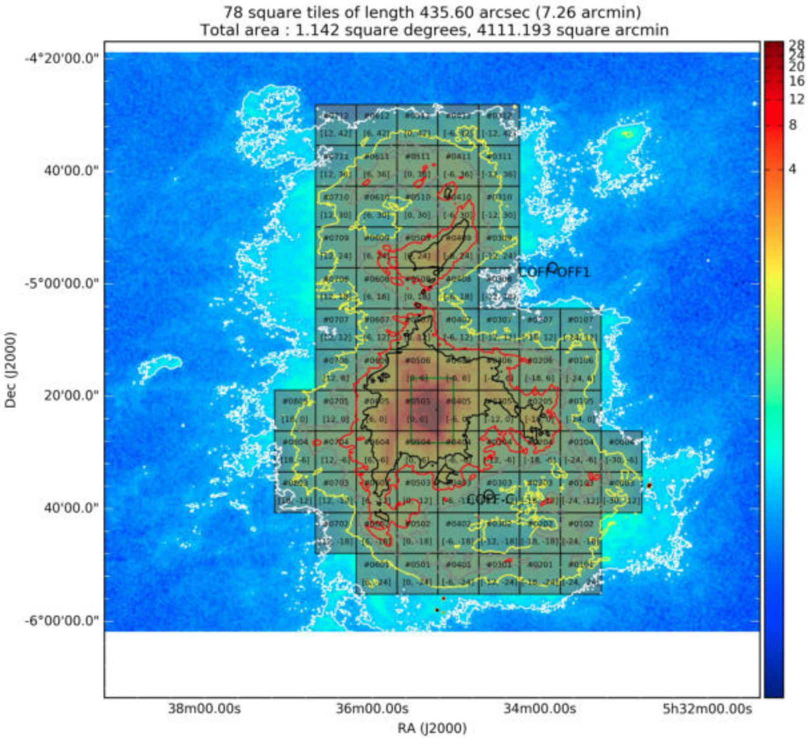
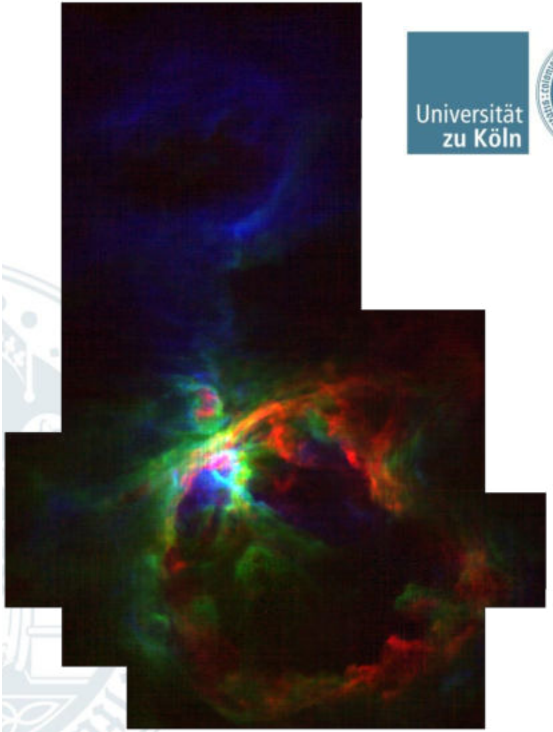




Data reduction of large maps with the upGREAT heterodyne instrument

SOFIA tele talks | Ronan Higgins | 26.01.2022

Talk outline

- Background to Orion Impact proposal
- Mapping strategy
- Data reduction process and data quality
- Overcoming data quality issues
- Looking to future
  - Data formats
  - Improving observing efficiency

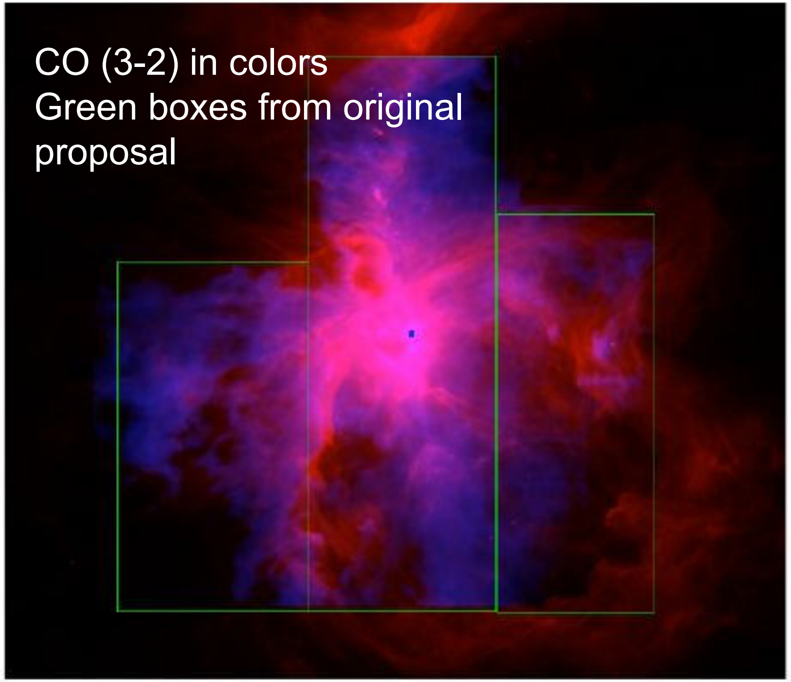
Impact proposal 04_0066

- Proposed in summer 2015 by a team led by Xander Tielens
- 42 hours observing time
- Observed over 13 flights
- Largest single map undertaken with SOFIA to date

SOFIA Proposal 04_0066

The large scale [CII] emission from the Orion molecular cloud

Principal Investigator: Prof. Alexander Tielens



Impact proposal 04_0066: publications/PhD theses

Observation and calibration strategies for large-scale multi-beam velocity-resolved mapping of the [CII] emission in the Orion molecular cloud*

R. Higgins[1], S. Kabanovic[1], C. Pabst[2], D. Teyssier[3], J. R. Goicoechea[4], O. Berne[5], E. Chambers[6], M. Wolfire[7], S. T. Suri[8], C. Buchbender[1], Y. Okada[1], M. Mertens[1], A. Parikka[6], R. Aladro[9], H. Richter[10], R. Güsten[9], J. Stutzki[1], and A. G. G. M. Tielens[2]

Disruption of the Orion molecular core 1 by wind from the massive star θ[1] Orionis C

C. Pabst[1], R. Higgins[2], J. R. Goicoechea[3], D. Teyssier[4], O. Berne[5], E. Chambers[6], M. Wolfire[7], S. T. Suri[2], R. Guesten[8], J. Stutzki[2], U. U. Graf[2], C. Risacher[8,9] & A. G. G. M. Tielens[1]*

Expanding bubbles in Orion A: [C II] observations of M 42, M 43, and NGC 1977

C. H. M. Pabst[1], J. R. Goicoechea[2], D. Teyssier[3], O. Berné[4], R. D. Higgins[5], E. T. Chambers[7], S. Kabanovic[5], R. Güsten[6], J. Stutzki[5], and A. G. G. M. Tielens[1]

[C II] 158 μm line emission from Orion A I. A template for extragalactic studies?

C. H. M. Pabst[1], A. Hacar[1,2], J. R. Goicoechea[3], D. Teyssier[4], O. Berné[5], M. G. Wolfire[6], R. D. Higgins[7], E. T. Chambers[8], S. Kabanovic[7], R. Güsten[9], J. Stutzki[7], C. Kramer[10], and A. G. G. M. Tielens[1,6]

[C II] 158 μm line emission from Orion A. II. Photodissociation region physics

C. H. M. Pabst[1], J. R. Goicoechea[2], A. Hacar[1,3], D. Teyssier[4], O. Berné[5], M. G. Wolfire[6], R. D. Higgins[7], E. T. Chambers[8], S. Kabanovic[7], R. Güsten[9], J. Stutzki[7], C. Kramer[10], and A. G. G. M. Tielens[1,6]

Molecular globules in the Veil bubble of Orion

IRAM 30 m ^{12}CO, ^{13}CO, and $C^{18}O$ (2–1) expanded maps of Orion A*

J. R. Goicoechea[1], C. H. M. Pabst[2], S. Kabanovic[3], M. G. Santa-Maria[1], N. Marcelino[1], A. G. G. M. Tielens[2], A. Hacar[2], O. Berné[4], C. Buchbender[3], S. Cuadrado[1], R. Higgins[3], C. Kramer[5], J. Stutzki[3], S. Suri[6], D. Teyssier[7], and M. Wolfire[8]

Carbon radio recombination lines from gigahertz to megahertz frequencies towards Orion A

P. Salas[1], J. B. R. Oonk[1,2], K. L. Emig[1], C. Pabst[1], M. C. Toribio[1], H. J. A. Röttgering[1], and A. G. G. M. Tielens[1]

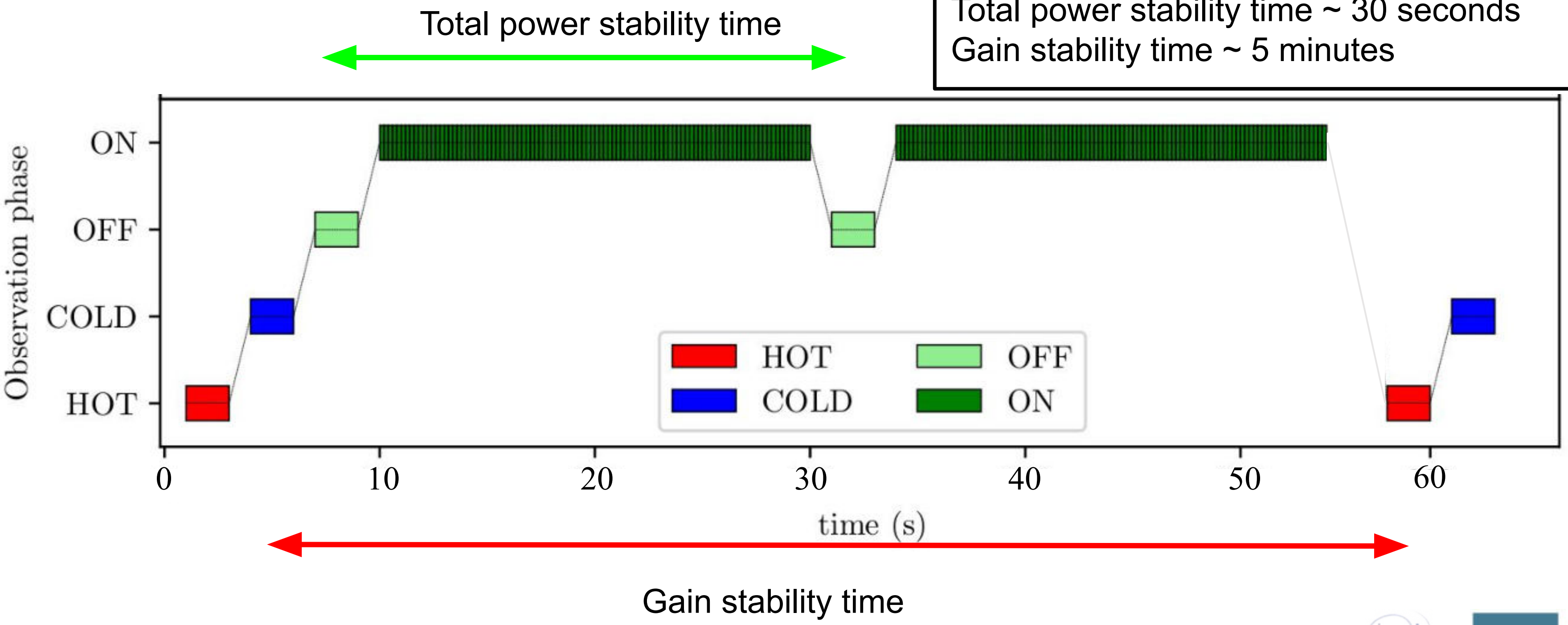
Phd theses:

Cornelia Pabst (Uni Leiden)
Sümeyye Suri (Uni Köln)
Slawa Kabanovic (Uni Köln Q4 2022)

Observing strategy

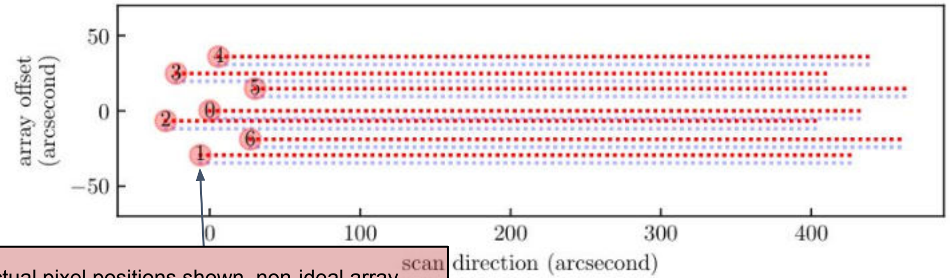
Observation details: Heterodyne mapping observation

upGREAT stability times:
Total power stability time ~ 30 seconds
Gain stability time ~ 5 minutes



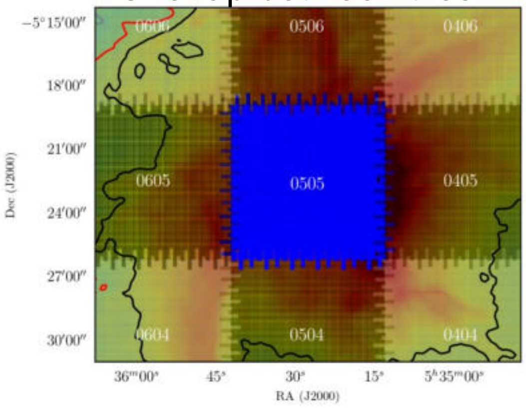
Observation details: array OTF mapping mode

84 OTF dumps of 0.3 second duration



Actual pixel positions shown, non-ideal array layout, note vertical distances between 2,0,5

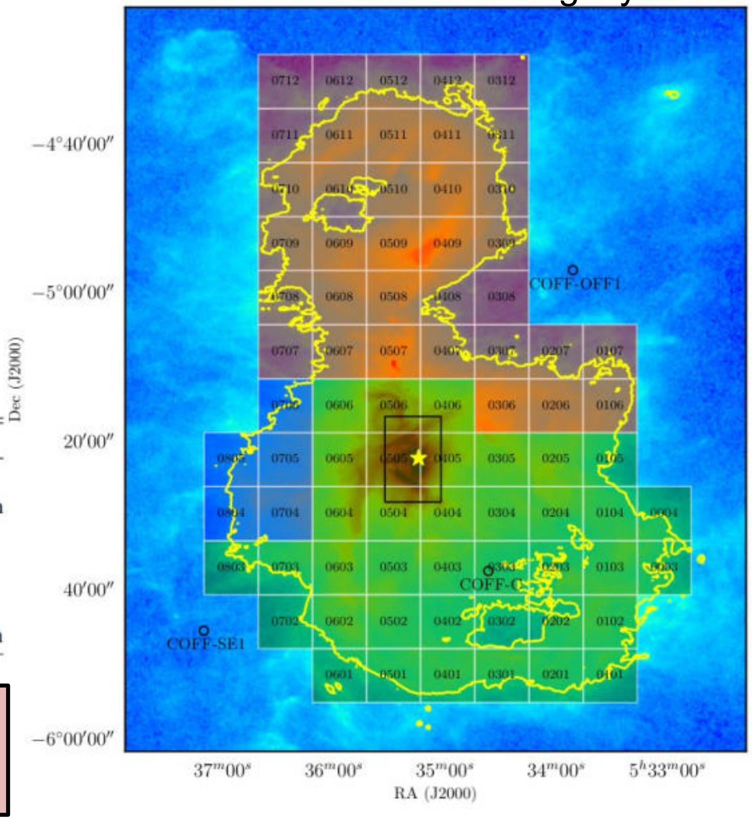
Overlap between tiles



Name	RA	Dec	Role
CENTER	5^{h}35^{m}27.6^s	−5°22′33.7″	Map center
COFF-C	5^{h}34^{m}36.5^s	−5°37′32.7″	Green OFF position
COFF-OFF1	5^{h}33^{m}51.0^s	−4°57′05.2″	Red OFF position
COFF-SE1	5^{h}37^{m}10.0^s	−5°45′33.7″	Blue OFF position
FOFF-E	5^{h}39^{m}21.6^s	−4°58′29.7″	Far OFF east
FOFF-W	5^{h}31^{m}15.5^s	−5°52′27.4″	Far OFF west
BAR_PEAK	5^{h}35^{m}20.9^s	−5°25′04.8″	Calibration position

Offset between map center and far-offs ~ 1 degree

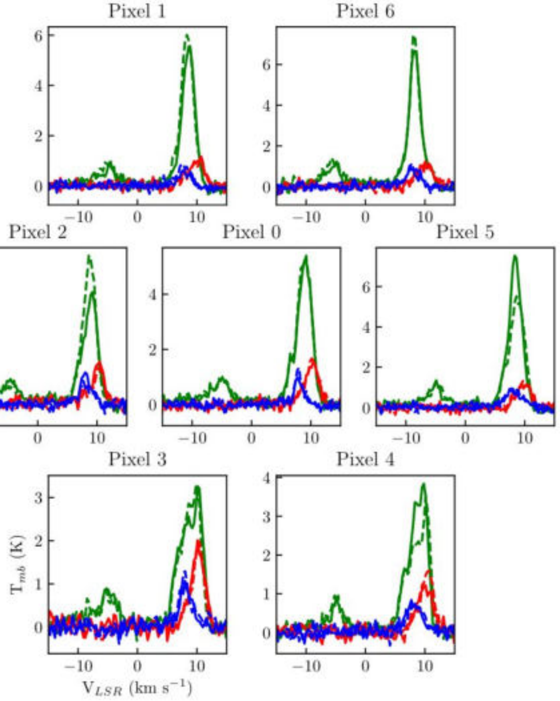
OFF selection and tiling layout

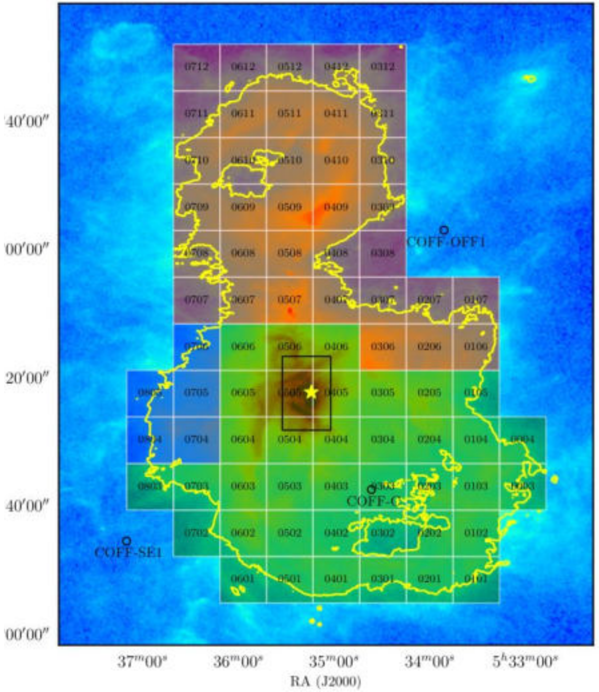


Slide 7

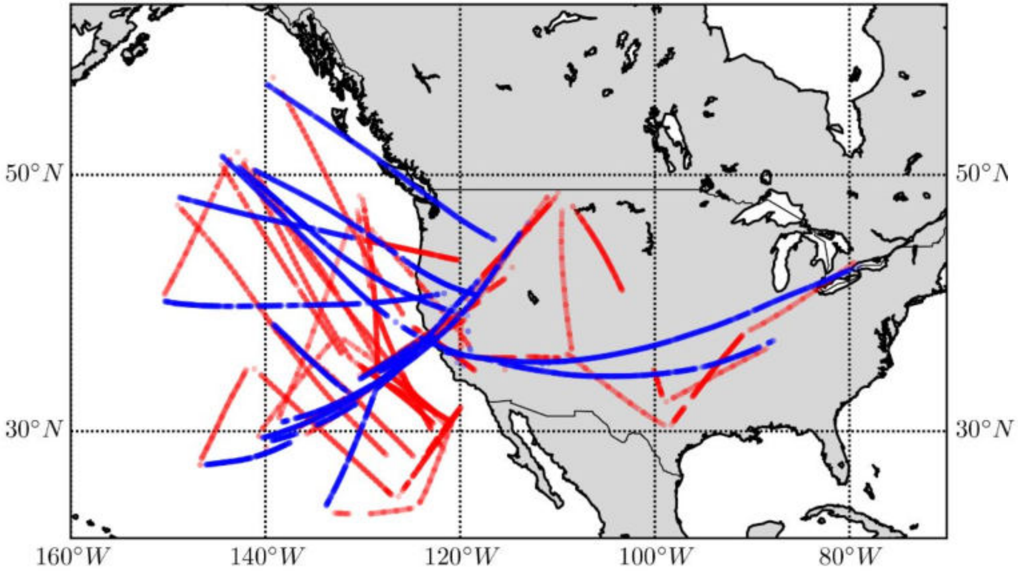
Observation details: OFF contamination

- OFF observation
- Concern about stability led to a minimisation of slew time
- 3 off positions had significant emission
- Concern about stability, OFF positions measured on each flight





Observation details: flight paths



Flight ID	Tiles	Tile IDs	Spectra count
2016-11-10_GR_F348	2	0505, 0605	34 272
2016-11-15_GR_F349	9	0105, 0205, 0304, 0305, 0306, 0307,0405, 0705, 0805	272 160
2016-11-16_GR_F350	8	0206, 0301, 0302, 0303, 0406, 0506, 0606, 0706	226 800
2016-11-17_GR_F351	8	0106, 0107, 0206, 0207, 0407, 0507, 0607, 0608	226 800
2016-11-18_GR_F352	8	0204, 0404, 0408, 0504, 0508, 0604, 0704, 0804	244 440
2017-02-08_GR_F371	8	0004, 0104, 0204, 0206, 0503, 0505, 0603, 0605	181 440
2017-02-09_GR_F372	10	0003, 0103, 0203, 0403, 0502, 0602, 0603, 0702, 0703, 0803	277 200
2017-02-10_GR_F373	5	0102, 0202, 0402, 0501, 0601	138 600
2017-02-14_GR_F374	8	0101, 0201, 0401, 0509, 0601, 0609, 0708, 0709	209 160
2017-02-15_GR_F375	8	0309, 0310, 0409, 0410, 0509, 0510, 0610, 0710	236 880
2017-02-16_GR_F376	7	0311, 0411, 0511, 0611, 0612, 0711, 0712	185 220
2017-02-17_GR_F377	9	0308, 0312, 0403, 0412, 0512, 0612, 0703, 0707, 0803	228 060
		Total Spectra	2 461 032

2.4 million spectra
140 Gb of raw data

SFB 956
Universität zu Köln

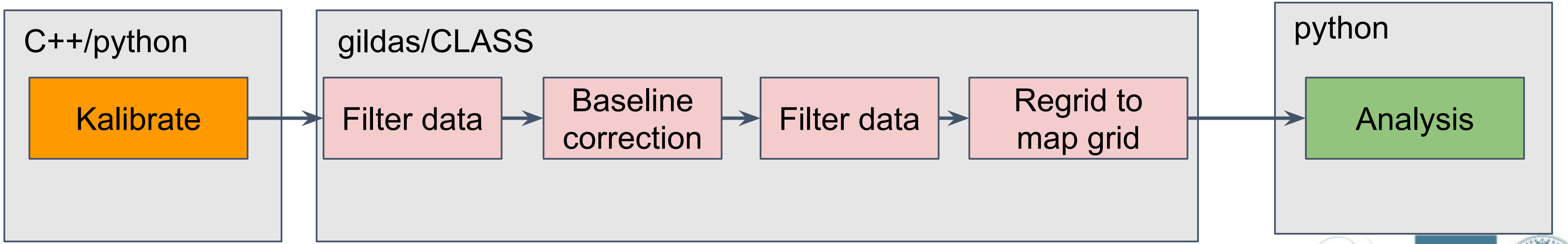
Data reduction and data quality

Data reduction process

GREAT/SOFIA atmospheric calibration★

X. Guan[1], J. Stutzki[1], U. U. Graf[1], R. Güsten[2], Y. Okada[1], M. A. Requena-Torres[2], R. Simon[1], and H. Wiesemeyer[2]

- Kalibrate code used to convert raw counts to rayleigh jeans corrected temperature scale and correct for atmospheric transmission
- Data written to gildas/CLASS data format
- Baseline correction and data filtered for outliers
- Velocity cube generated within class

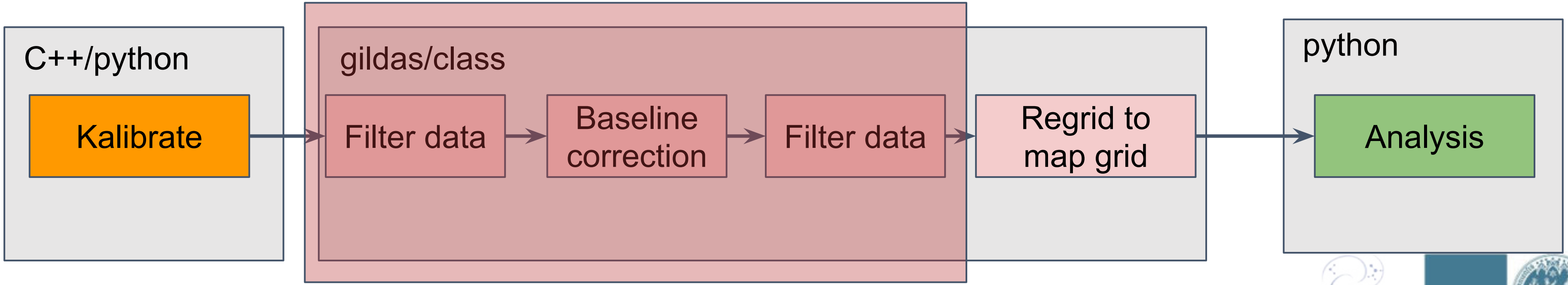


Data reduction process

GREAT/SOFIA atmospheric calibration*

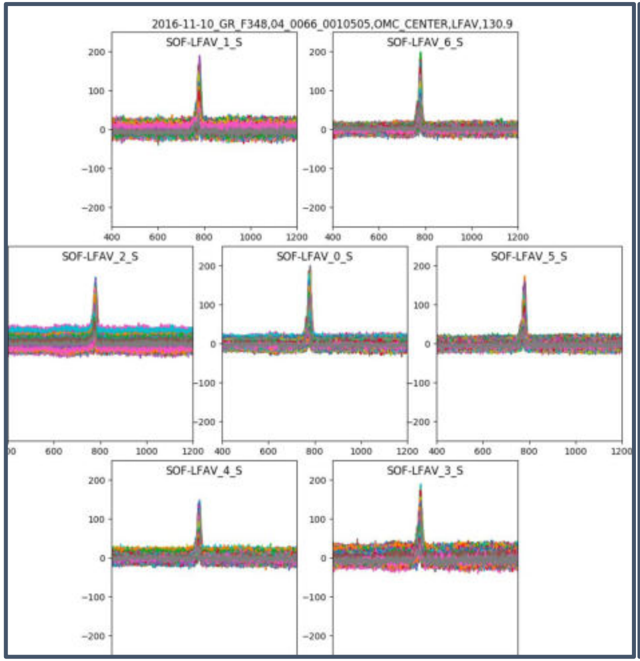
X. Guan[1], J. Stutzki[1], U. U. Graf[1], R. Güsten[2], Y. Okada[1], M. A. Requena-Torres[2], R. Simon[1], and H. Wiesemeyer[2]

- Kalibrate code used to convert raw counts to rayleigh jeans corrected temperature scale and correct for atmospheric transmission
- Data written to gildas/CLASS data format
- Baseline correction and data filtered for outliers
- Velocity cube generated within class

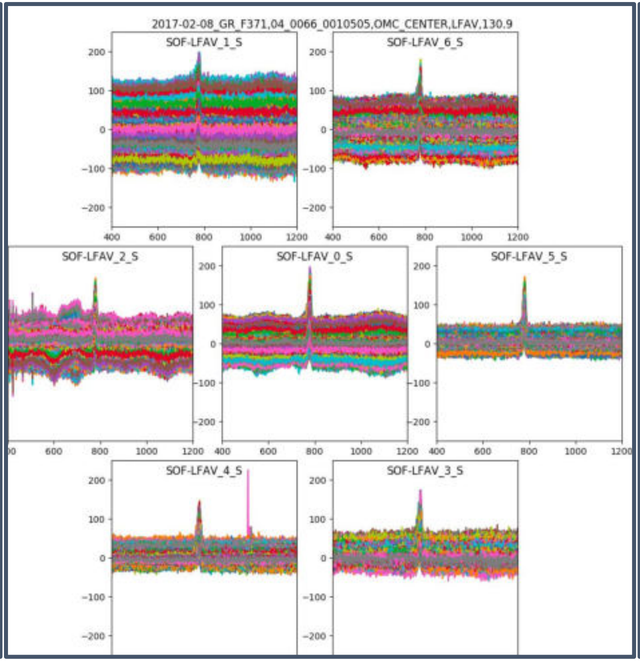


Data quality overview

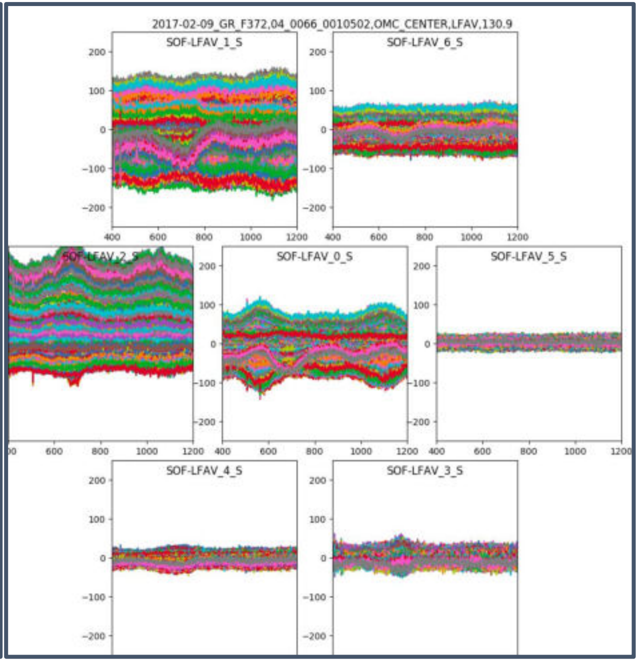
Good



Bad

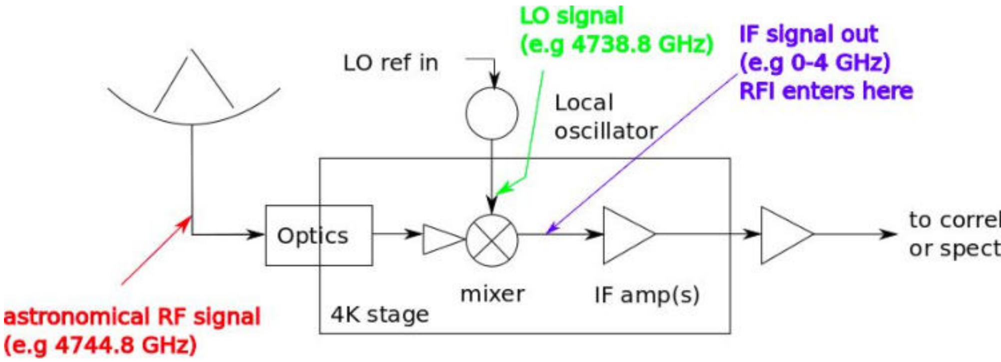


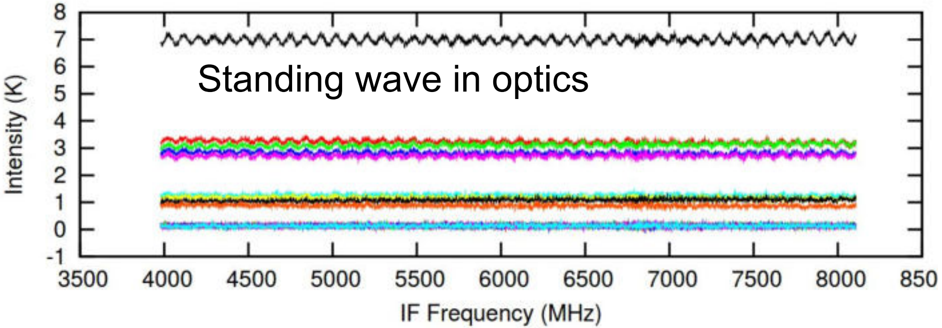
Worse

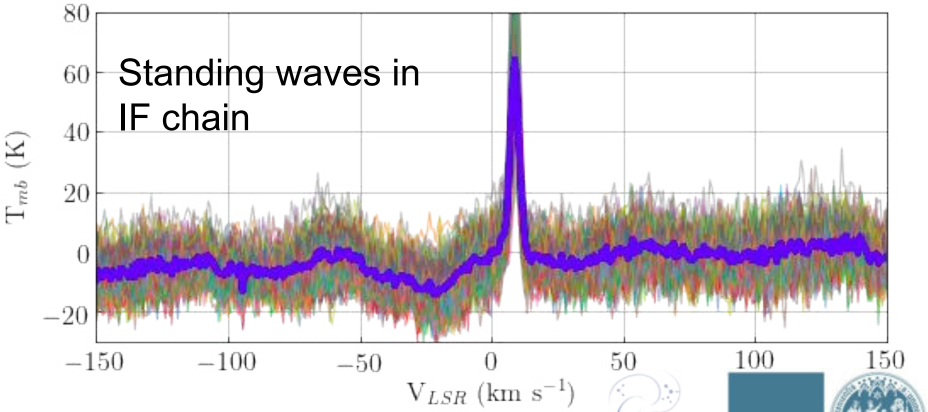


Baseline correction

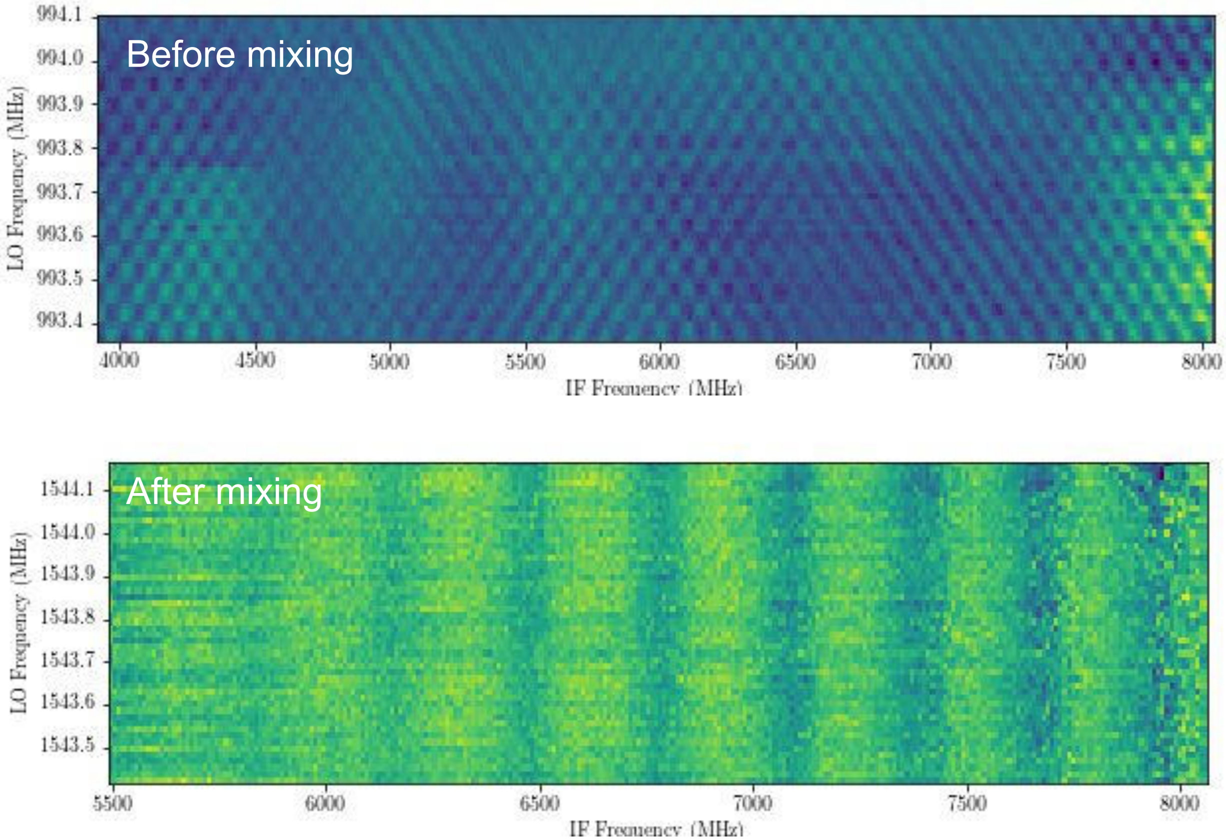
- Classic problem in heterodyne receivers
- Data can be corrupted by receiver systematics
  - Reflections from secondary mirror
  - Reflections from calibration loads
  - Reflections in IF chain (after mixing)







Baseline correction: standing waves physics



Standing waves in optics

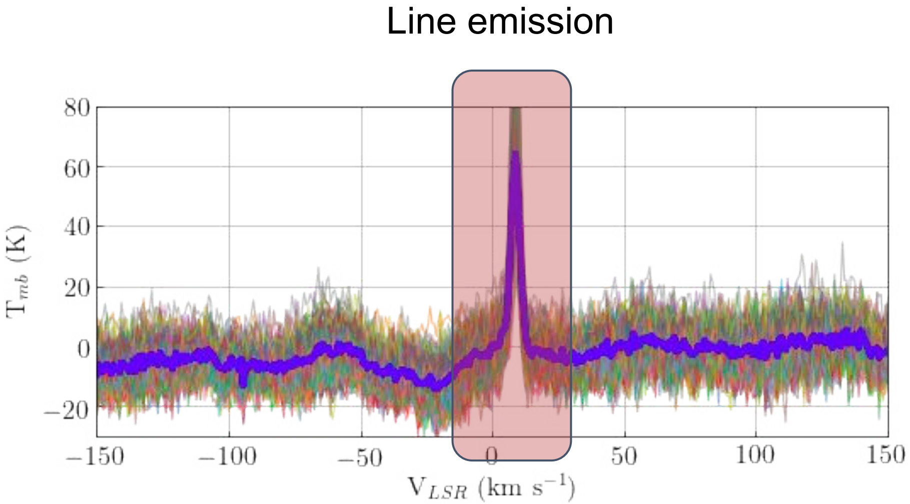
- Sinusoidal like but can have a modulated amplitude due to interference between sidebands
- Amplitude/phase is LO dependent due to sideband interference
- Period corresponds to cavity in optics

Standing waves in IF chain

- Irregular pattern across band, dependent on reflection properties of IF chain component (see here)
- Phase/amplitude LO independent
- Phase/amplitude dependent on electrical state of the calibration phases (stable system = no baseline effects)

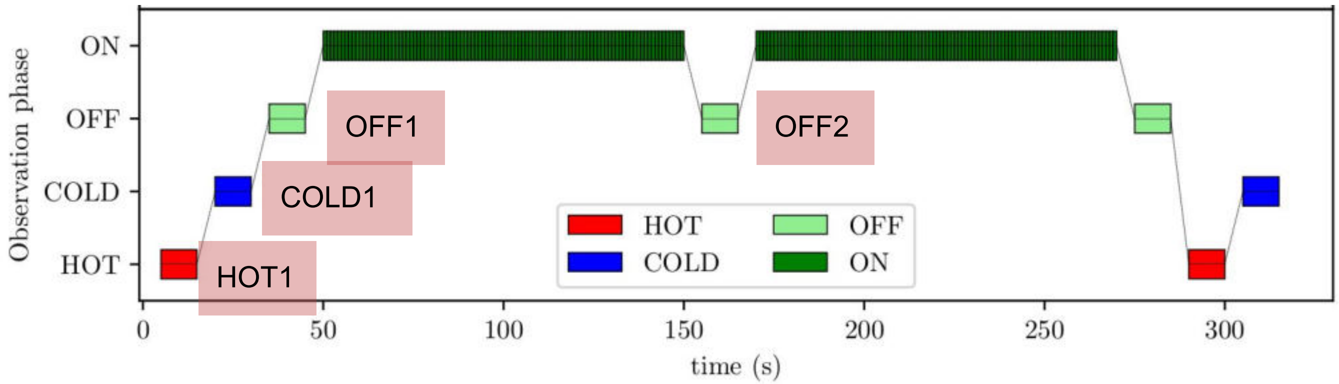
Baseline correction

- Typical step in heterodyne data processing processing
- Standard approach is to flag the line region and run a polynomial fit on the remaining baseline
- Where the baseline has local variations fitting a polynomial can corrupt the line emission
- Particular problem for regions with broad lines (e.g. M51 / galactic center)
- Alternative approach needed



Baseline correction: scaled spline

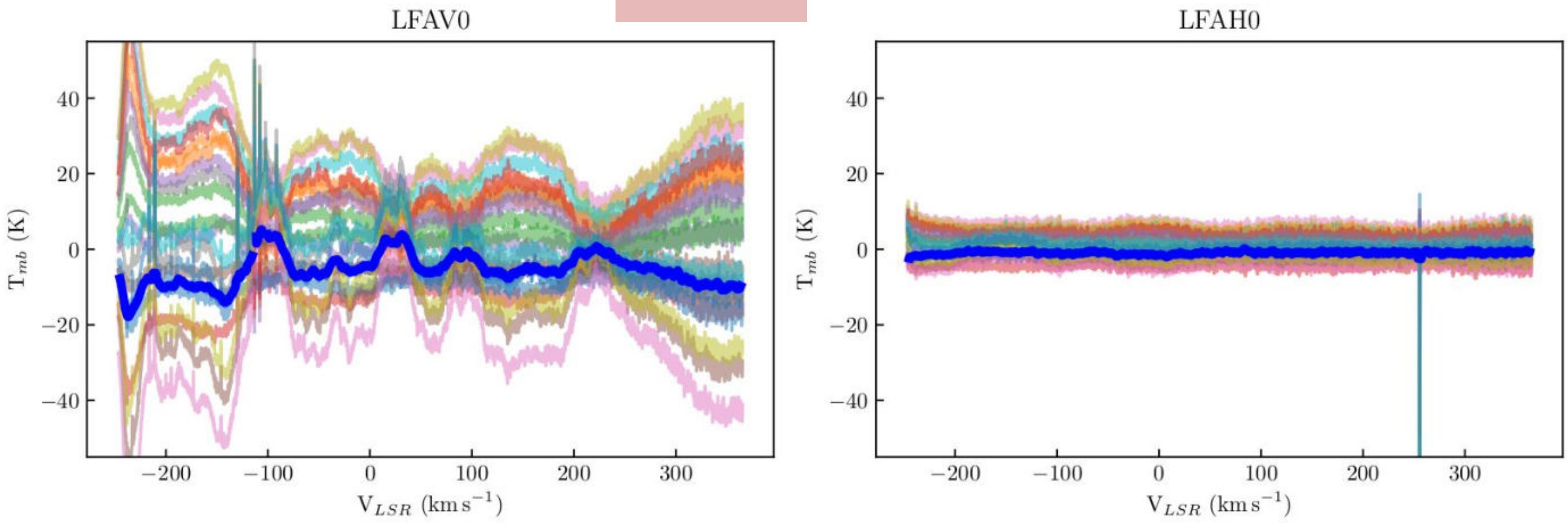
- From HIFI experience, a scaled splined based on off spectra observations worked well (see Do kester HEB standing waves paper)
- This approach was used for the Orion data set
- Residuals between OFF spectra used to generate a catalog of baseline shapes, under the assumption that OFF1-OFF2 spectra have similar baseline shapes to ON-OFF spectra



OFF1-OFF2
HOT1-HOT2
= BASELINE

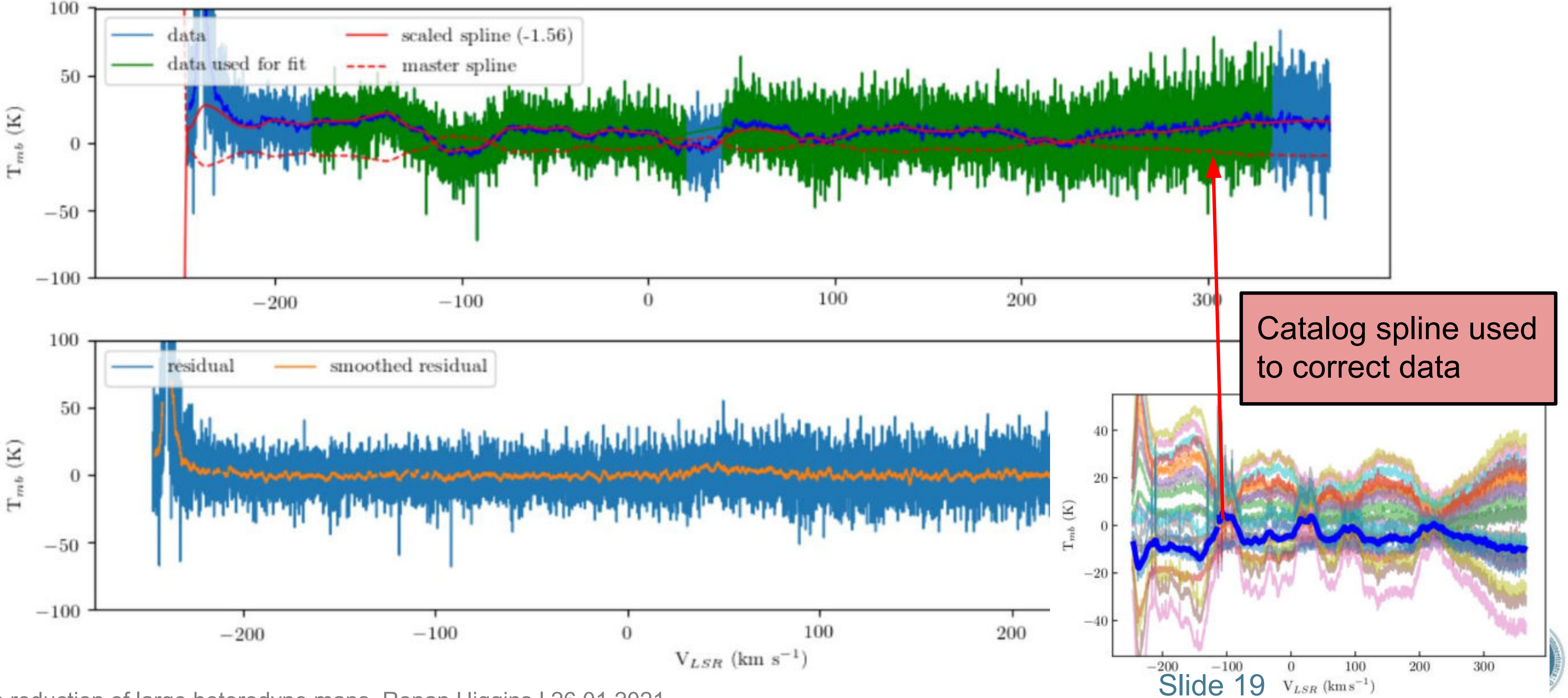
Baseline correction: OFF catalog example

<u>OFF(i)-OFF(j)</u>
HOT1-HOT2



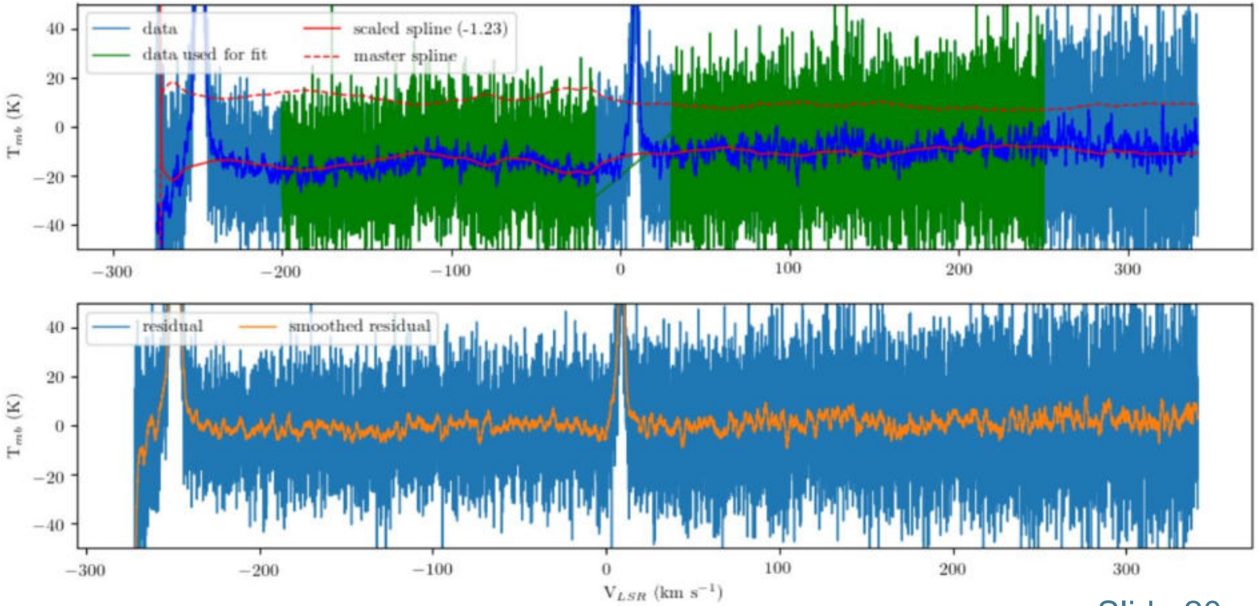
SFB 956
Universität zu Köln

Baseline correction: fitting spline to data

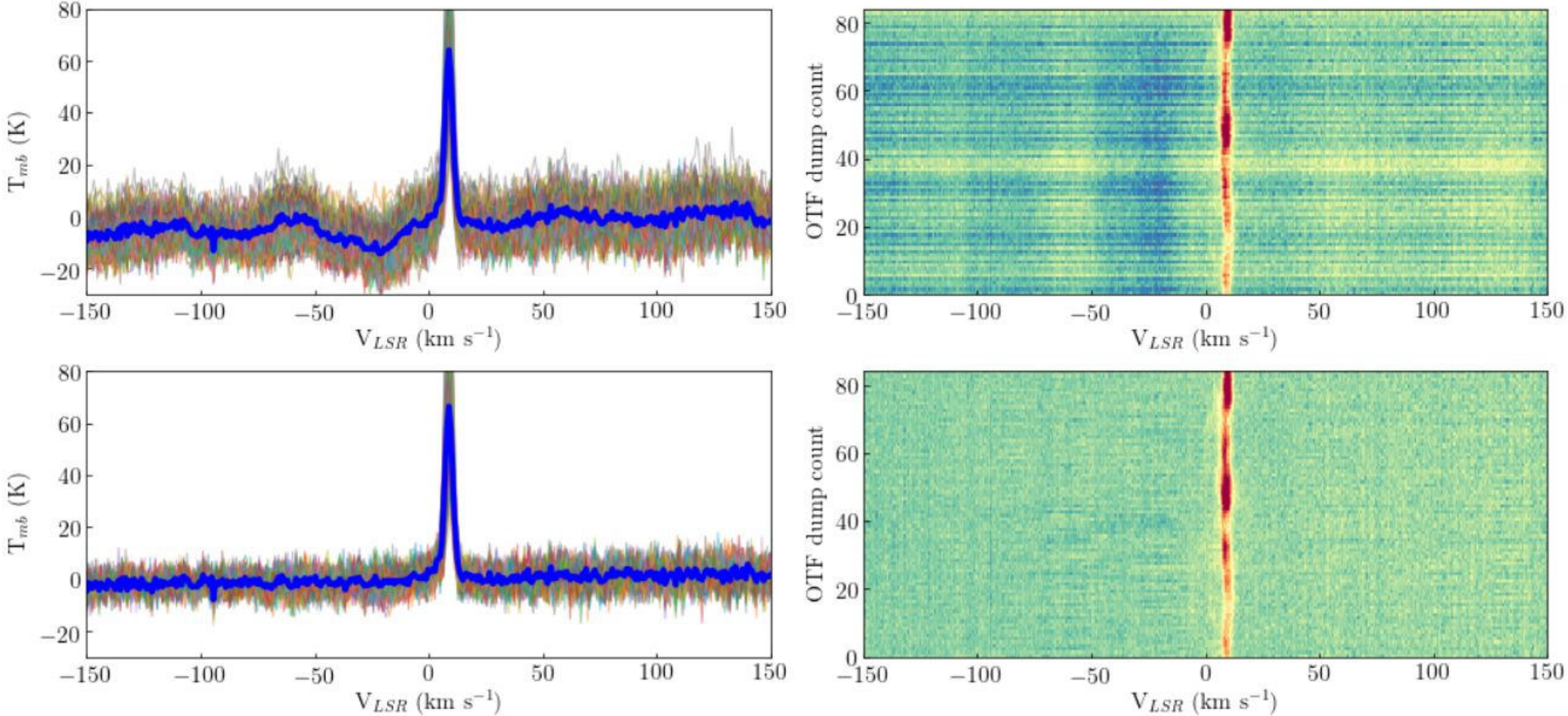


Baseline correction: spline correction process

Generate OFF catalog → Fit spline to each OFF residual spectra → Scale each catalog spline to find best fit to ON-OFF data → subtract scaled spline to correct spectra

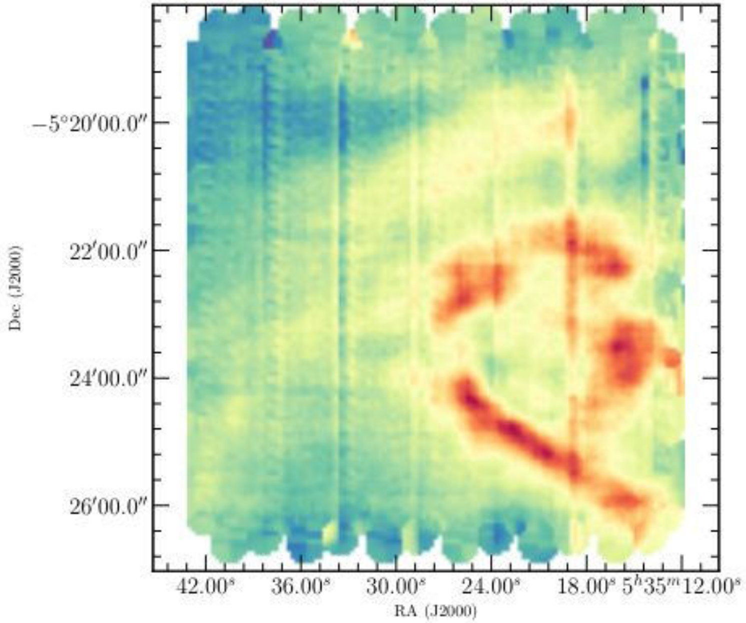


Baseline correction: spline correction outcome

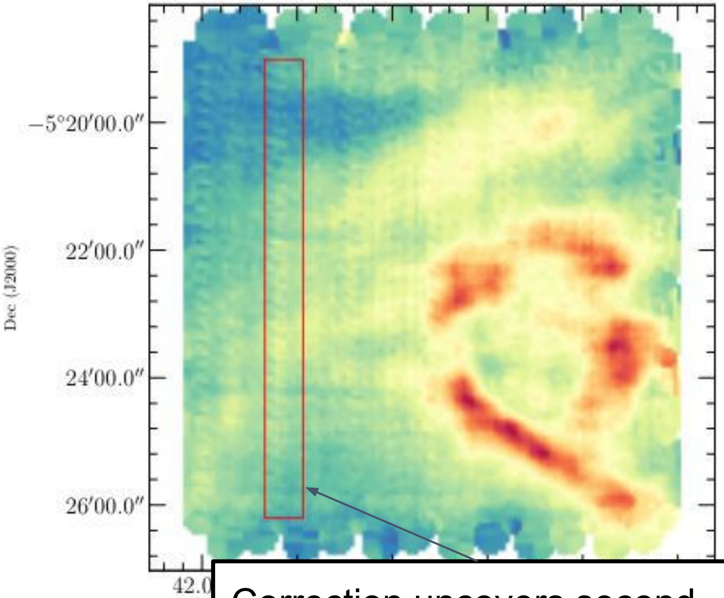


Baseline correction: spline correction outcome

Center tile with polynomial correction

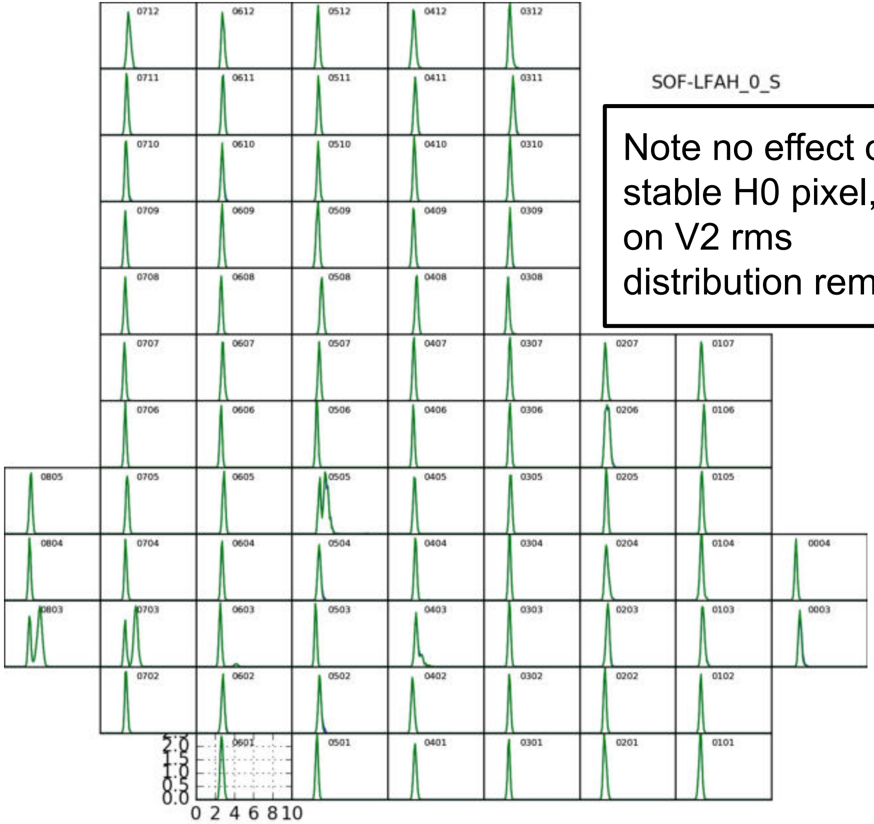


Center tile with spline correction

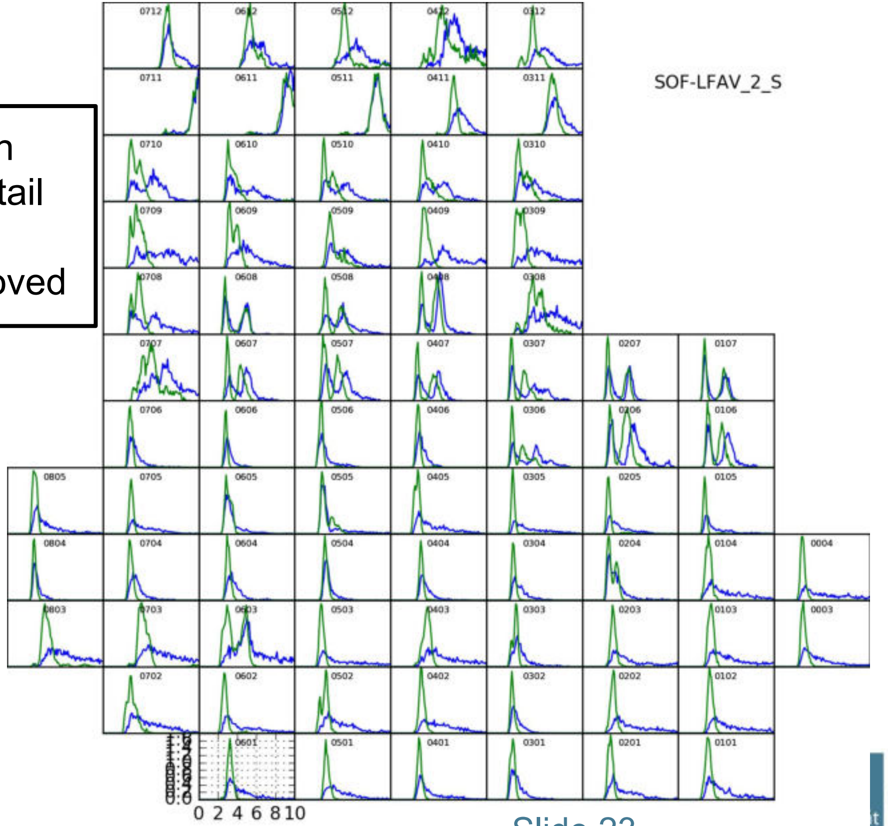


Correction uncovers second order effects in data: gain instability

Baseline correction: spline correction RMS before/after



Note no effect on stable H0 pixel, tail on V2 rms distribution removed

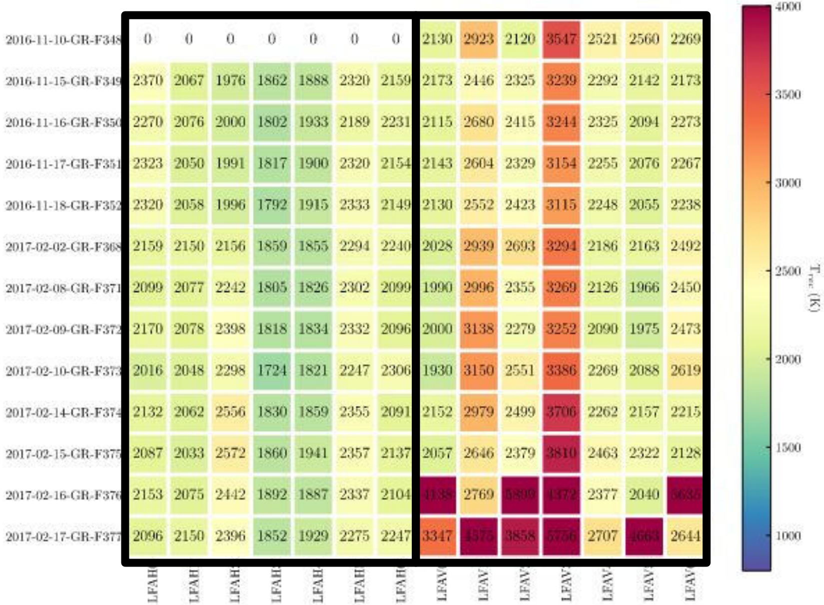


Improving data quality

- IF standing waves are a deficiency of receivers with HEB mixer
- Variable impedance of mixer leads to different reflection properties in mixer, small changes in **LO output power**, mechanical vibration, varying temperature can affect the pump state leaving baseline residuals in the data
- Problem first noted in the Herschel/HIFI instrument even when operating in a more stable thermal environment

LO dependent performance

LFAH LFAV



	LFAH	LFAH	LFAH	LFAH	LFAH	LFAH	LFAH	LFAV	LFAV	LFAV	LFAV	LFAV	LFAV	LFAV
2016-11-10-GR-F348	0	0	0	0	0	0	0	2130	2923	2120	3547	2521	2560	2269
2016-11-15-GR-F349	2370	2067	1976	1862	1888	2320	2159	2173	2446	2325	3239	2292	2142	2173
2016-11-16-GR-F350	2270	2076	2000	1802	1933	2189	2231	2115	2680	2415	3244	2325	2094	2273
2016-11-17-GR-F351	2323	2050	1991	1817	1900	2320	2154	2143	2604	2329	3154	2255	2076	2267
2016-11-18-GR-F352	2320	2058	1996	1792	1915	2333	2149	2130	2552	2423	3115	2248	2055	2238
2017-02-02-GR-F368	2159	2150	2156	1859	1855	2294	2240	2028	2939	2693	3294	2186	2163	2492
2017-02-08-GR-F371	2099	2077	2242	1805	1826	2302	2099	1990	2996	2355	3269	2126	1966	2450
2017-02-09-GR-F372	2170	2078	2398	1818	1834	2332	2096	2000	3138	2279	3252	2090	1975	2473
2017-02-10-GR-F373	2016	2048	2298	1724	1821	2247	2306	1930	3150	2551	3386	2269	2088	2619
2017-02-14-GR-F374	2132	2062	2556	1830	1859	2355	2091	2152	2979	2499	3706	2262	2157	2215
2017-02-15-GR-F375	2087	2033	2572	1860	1941	2357	2137	2057	2646	2379	3810	2463	2322	2128
2017-02-16-GR-F376	2153	2075	2442	1892	1887	2337	2104	4138	2769	[illegible]	4372	2377	2040	[illegible]
2017-02-17-GR-F377	2096	2150	2396	1852	1929	2275	2247	3347	4575	3858	5756	2707	4663	2644

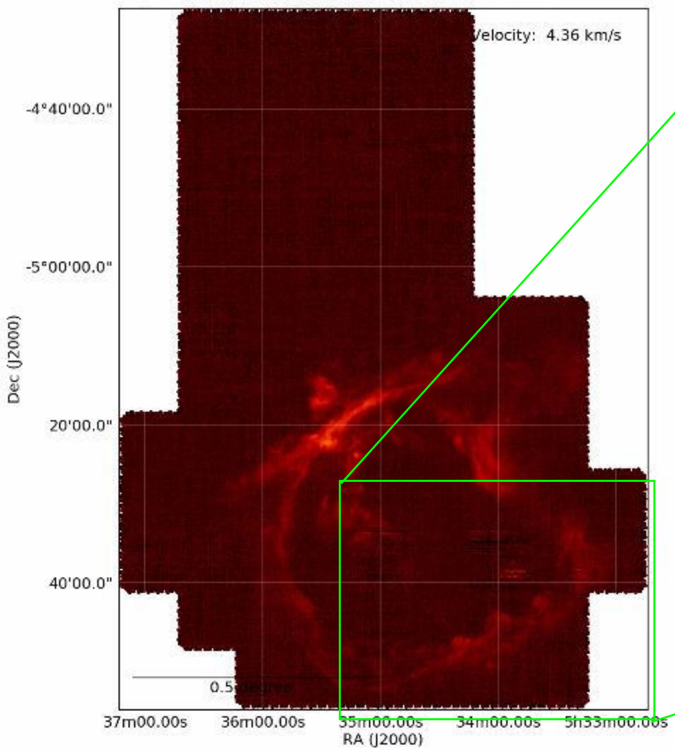
Note degradation in the LFAV LO towards later flights

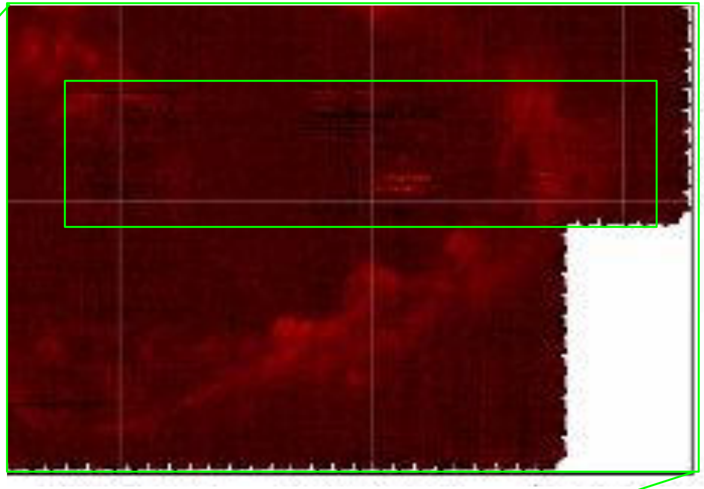
SFB 956

Universität zu Köln

Data quality: radio interference

Scratches seen in map at particular channels for particular pixels

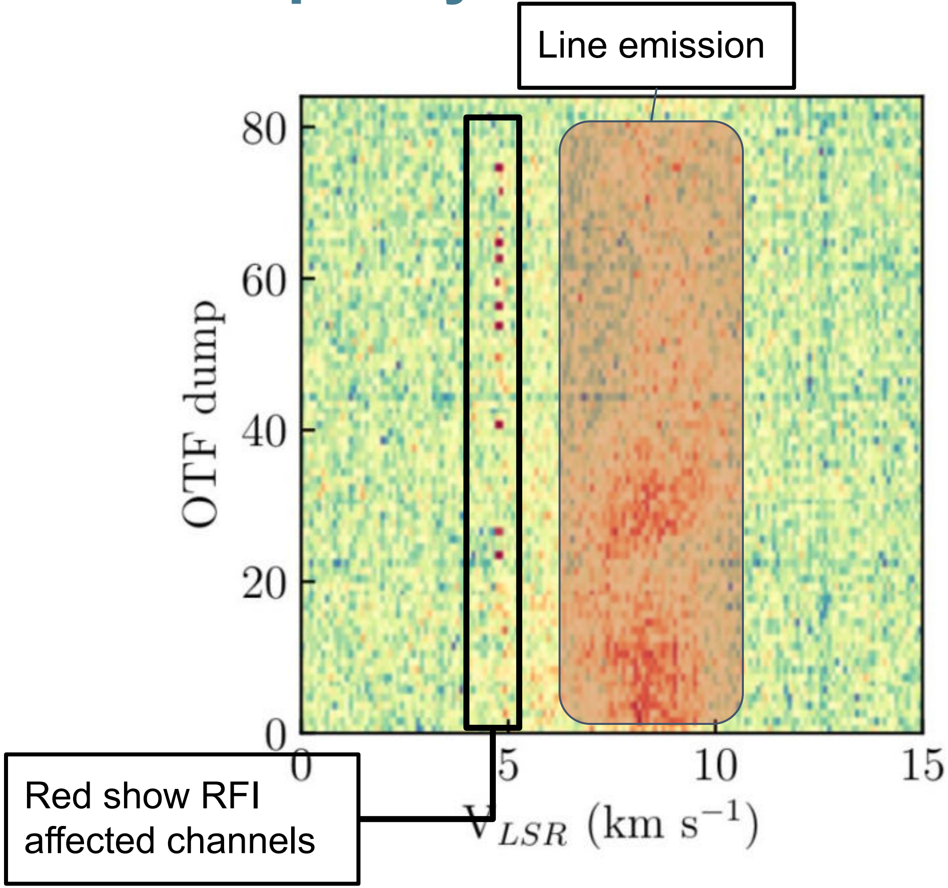


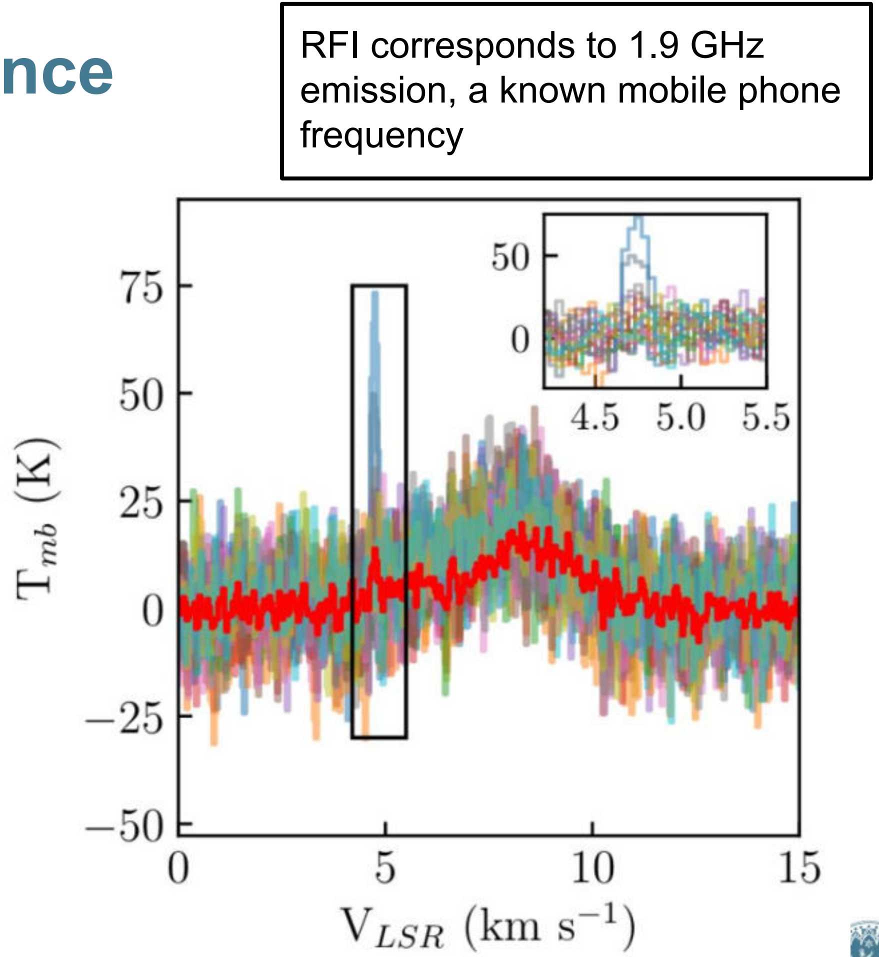


Universität zu Köln

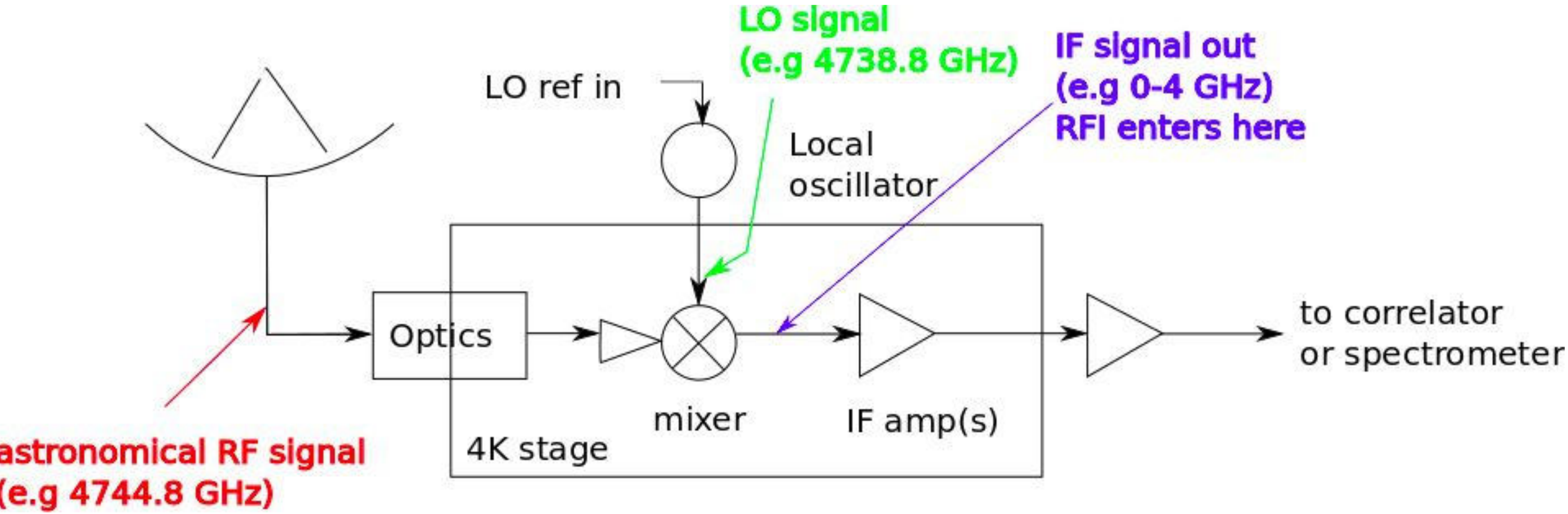
SFB 956

Data quality: radio interference



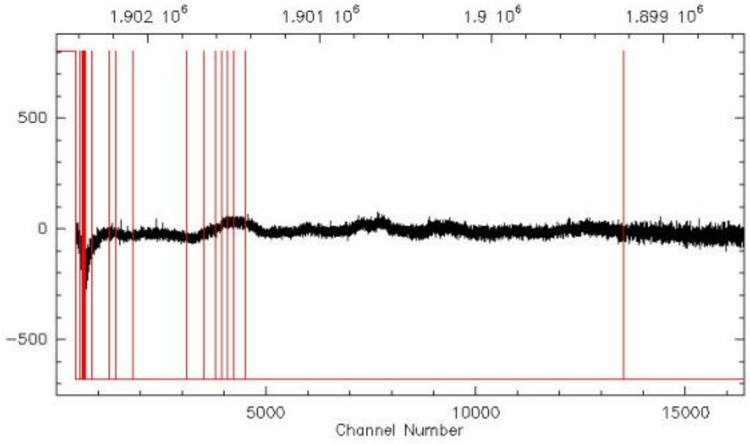


Data quality: radio interference



Data quality: overcoming radio interference

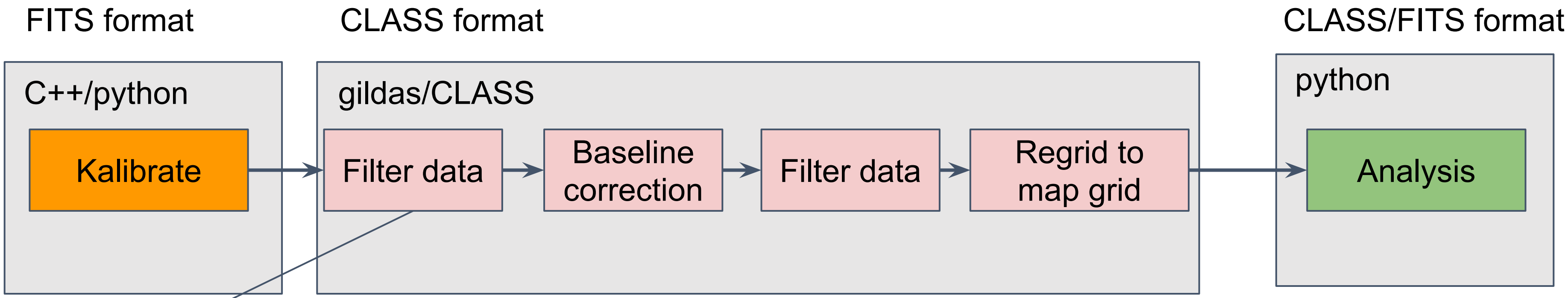
- 2-3 pixels in H array affects by RFI. Can't afford to drop whole pixel in cube generation, coverage too thin, leads to holes in map
- Weight RFI affected channels down in map making, required custom script in gildas (channel weighting not natively supported)
- Use associated array in gildas to track RFI affected spectra
- RFI affected channels are not recoverable!



Flagged spectra example, red shows the associated array for RFI

Flexible Data format and house keeping data synergies

upGREAT data processing: CLASS rigidity



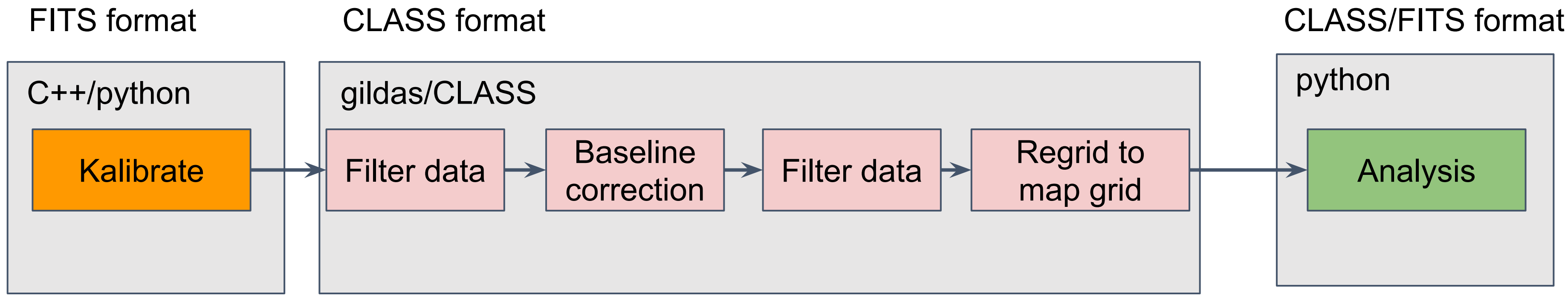
Quality measures: rms, radiometer rms, mean, median, FFT power at selected channels, RFI filters, water vapor, tsys

Stored in external csv file, not in class
Processed using python/pandas

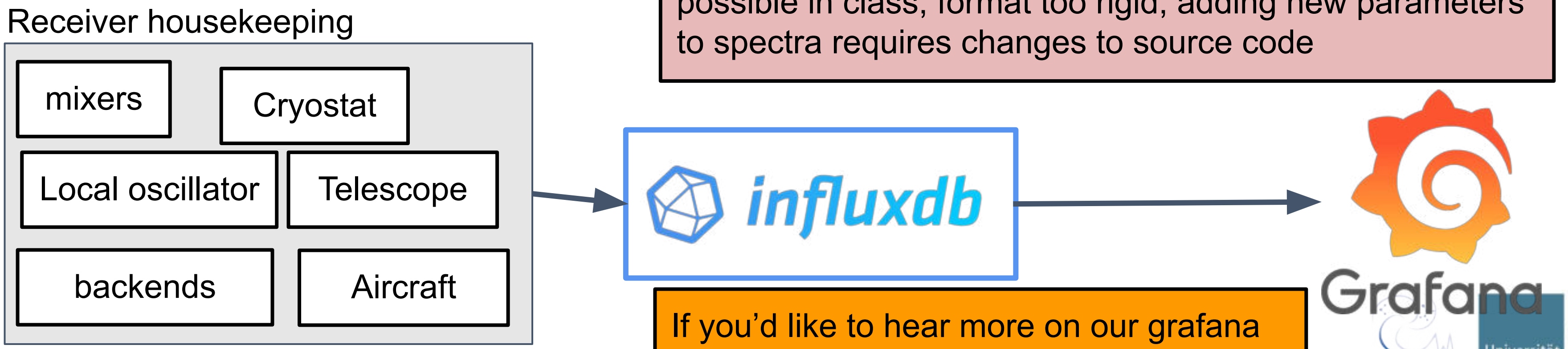
Ideally this is tagged to the data set, not possible in class right now

```
    !
    filter_index  /array_name to_be_ignored -
                        /table_filters "rms_radiometer_ratio>3" -
                        /table_filters "tsys == 0.0" -
                        /table_filters "tsys >= 3500 and telescope.str.contains('V')" -
                        /table_filters "rms.isna()" -
                        /table_filters "bad_channels>0"
    !
```

upGREAT data processing: class rigidity



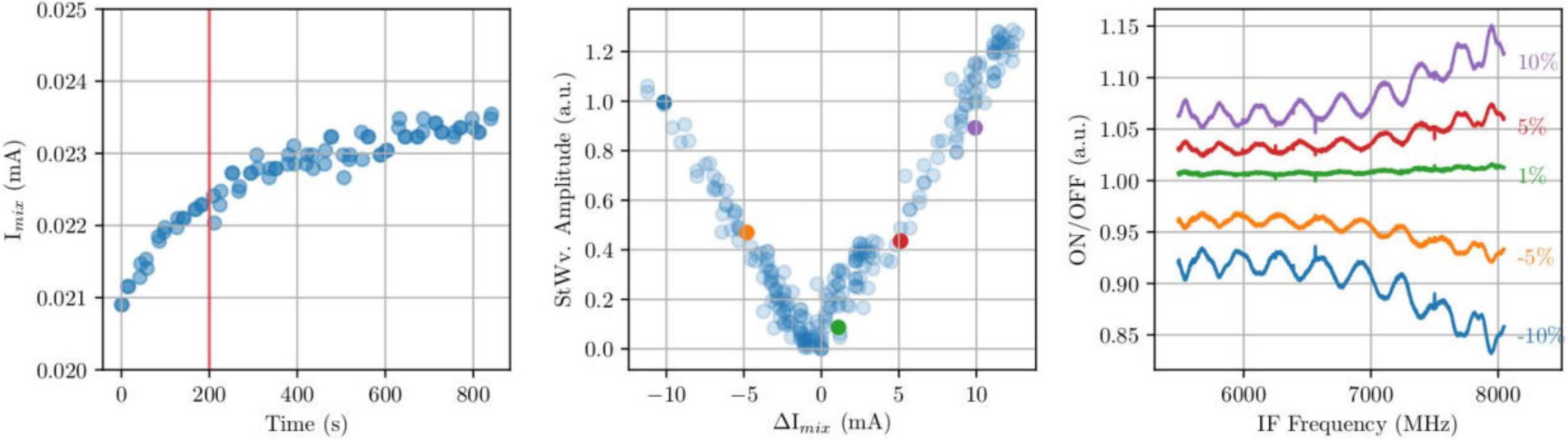
Ideally receiver data is contained in the data format, not possible in class, format too rigid, adding new parameters to spectra requires changes to source code



If you'd like to hear more on our grafana deployment check out the talk link here.

upGREAT data processing: using receiver data to detect anomalies

- IF Standing waves can be inferred direct from receiver housekeeping
- When you have the mixer current you can flag problem spectra, and identify source of baseline issue



Lessons learned from data processing

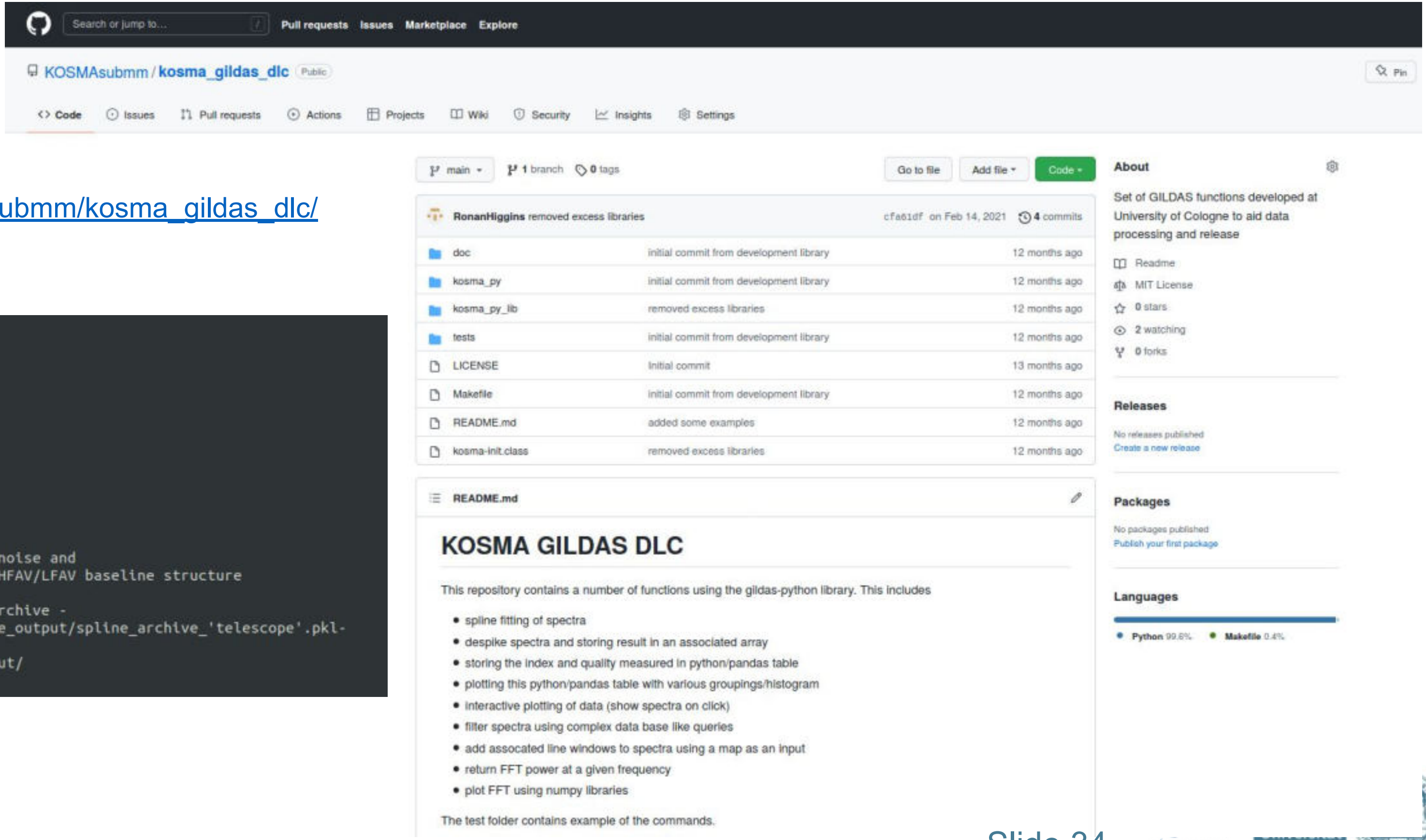
- Large part of data reduction was undertaken writing tools to get around short coming in CLASS data format and scripting environment
- CLASS is fast for core functionality, map making, baselining but struggles with more complex reductions
- Drawbacks:
  - Rigid data format
  - Limit time series support
  - Filtering of data limited
- To overcome these drawbacks we developed a python/pandas table add-on using the pyclass library, not a stable solution. Flexibility should be supported in data format
- Having all housekeeping data conveniently available in data reduction would help implementing an automatic reduction pipeline

Code used in data reduction

https://github.com/KOSMAsubmm/kosma_gildas_dlc/

Spline demo script

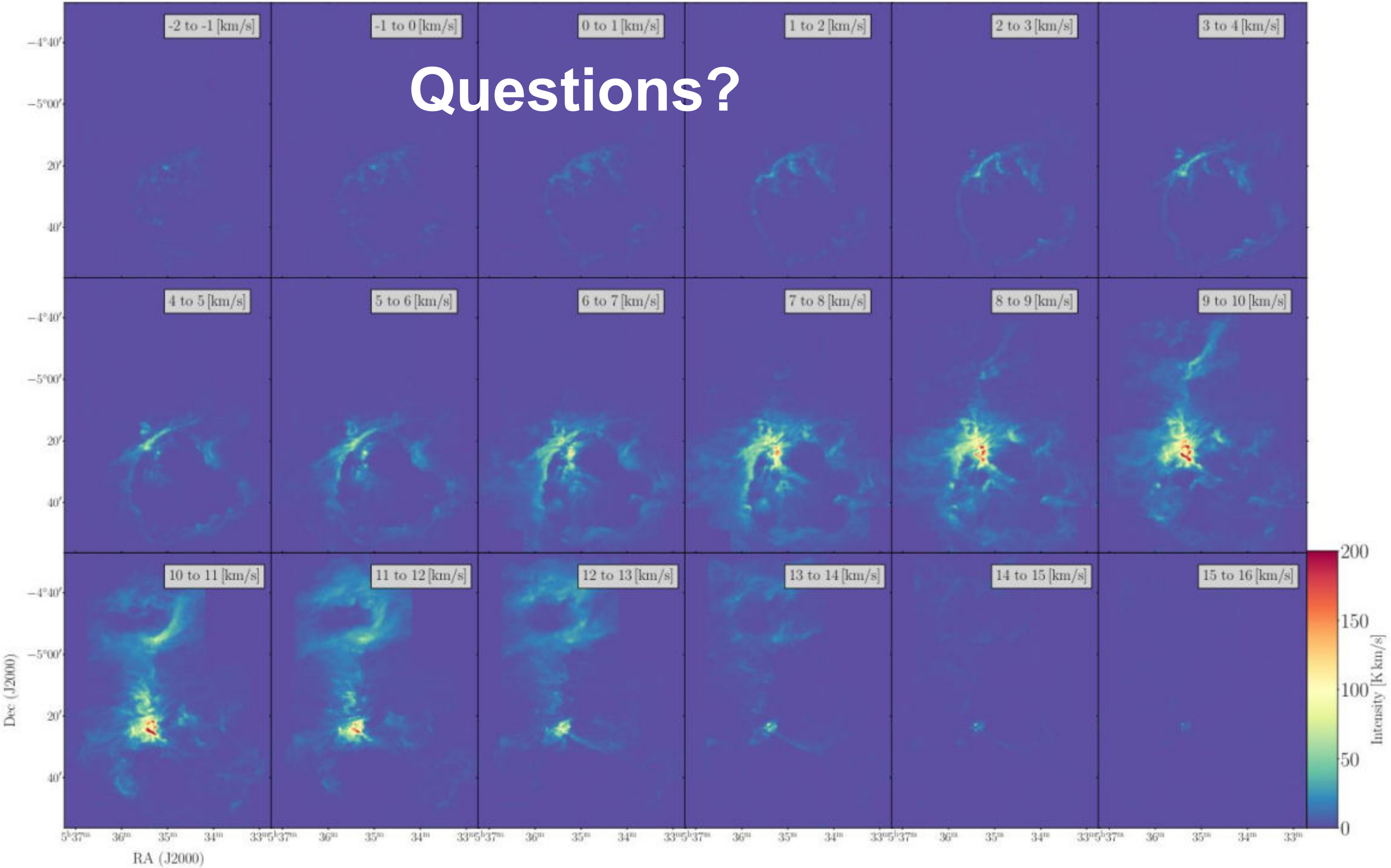
```
define structure kosma
file in data/spline_test_data.kosma
find
set unit v
set win -300 -180 20 40
!
! generate spline archive
!
set source *diff*
find
for i 1 to found
    get n
    ! smooth level is important
    ! baseline features can be below the noise and
    ! 40 is a good value for the typical HFAV/LFAV baseline structure
    smooth box 40
    spline /fit_spline /store_spline_in_archive -
           /spline_archive_filename spline_output/spline_archive_'telescope'.pkl-
           /logging_level info -
           /output_plots_path spline_output/
next
```



Lessons learned for future observations

- Angular distance to OFF position can be relaxed
  - Keeping to 20 arcminute limit added additional workload and lost observing time, recent GT Orion projects have gone directly to far off (~1 degree distance) without affecting data quality
- Consider relaxing total power stability time
  - Currently limited to 30 seconds for ON OTF scans
  - New baseline reduction techniques allows for longer scan times, consider testing longer scan times, increases efficiency
- Shorter dump times, more coverages
  - To improve redundancy consider going to shorter dump times but more coverages, recent tests at 0.1 seconds have worked well
  - The array OTF mapping mode requires uniform data quality across the array otherwise you get holes in your map due to poor pixel. Having multiple coverages with different pixels leads to better data quality (experience used in FEEDBACK legacy project)
- Don’t integrate single point observation into large single data blocks
  - OFF check observed with 15 second integration
  - Data quality was poor, 1 second instability corrupted entire spectra
  - Dump single point total power spectra in blocks of 0.5 seconds, average together afterwards
  - Recent GT tests using an adapted very slow OTF mode have worked well

Questions?



SFB 956

Universität zu Köln