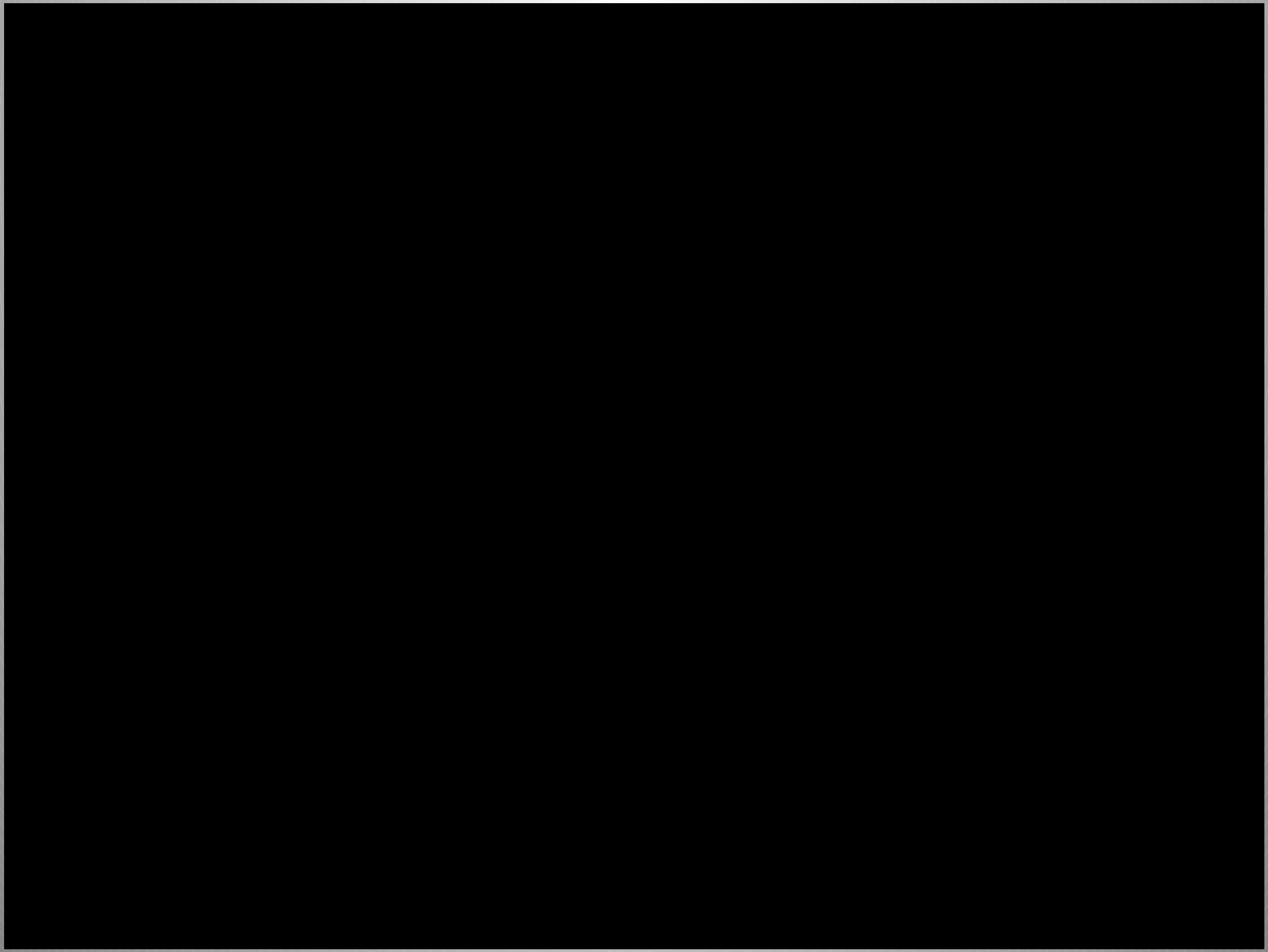
Automated Pipelines for Spectroscopic Analysis

Carlos Allende Prieto
Instituto de Astrofísica de Canarias

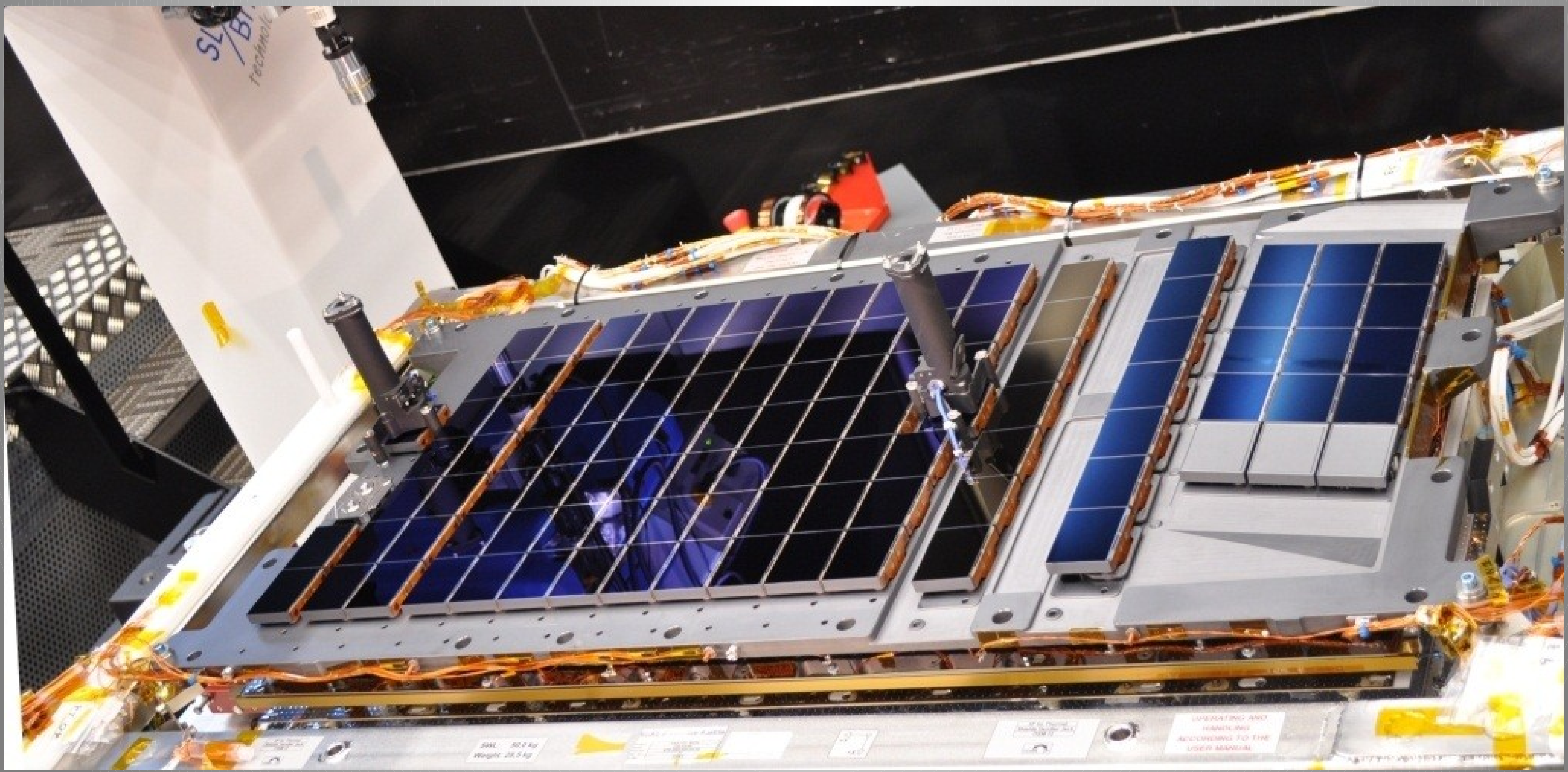
Gaia



Gaia: basics

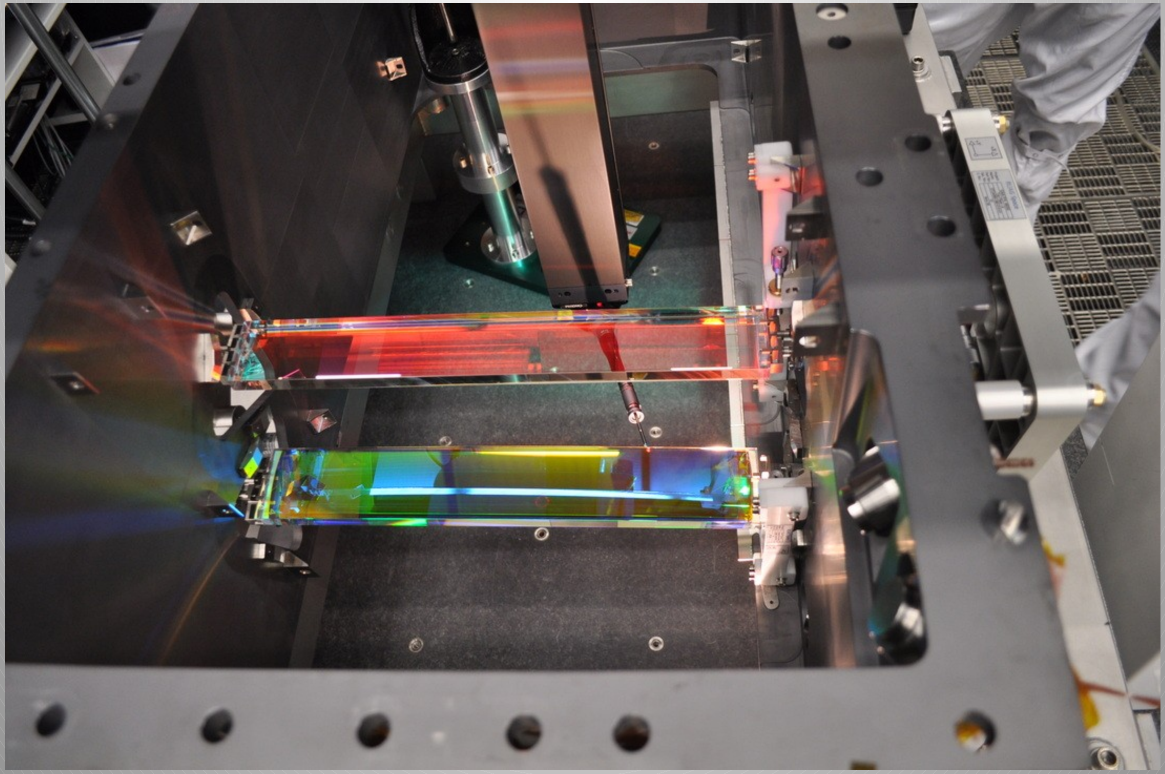
- All sky
- Point source detection onboard: No selection effects (other than reddening and brightness)
- Astrometry and spectrophotometry for 1e9 sources down to V~20
- High-resolution spectroscopy (847-874 nm) for 1e8 sources down to V~16

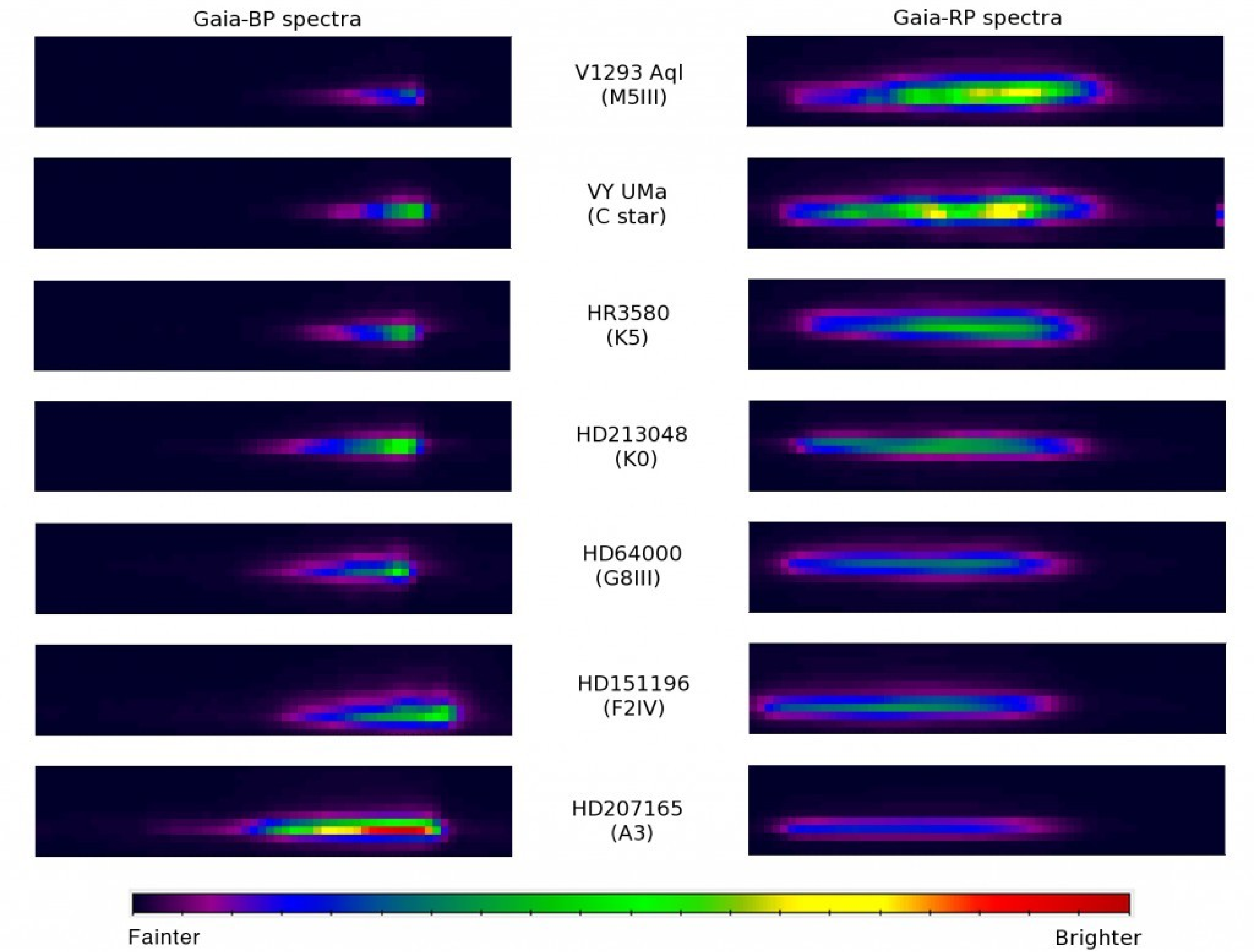
Gaia: focal plane assembly



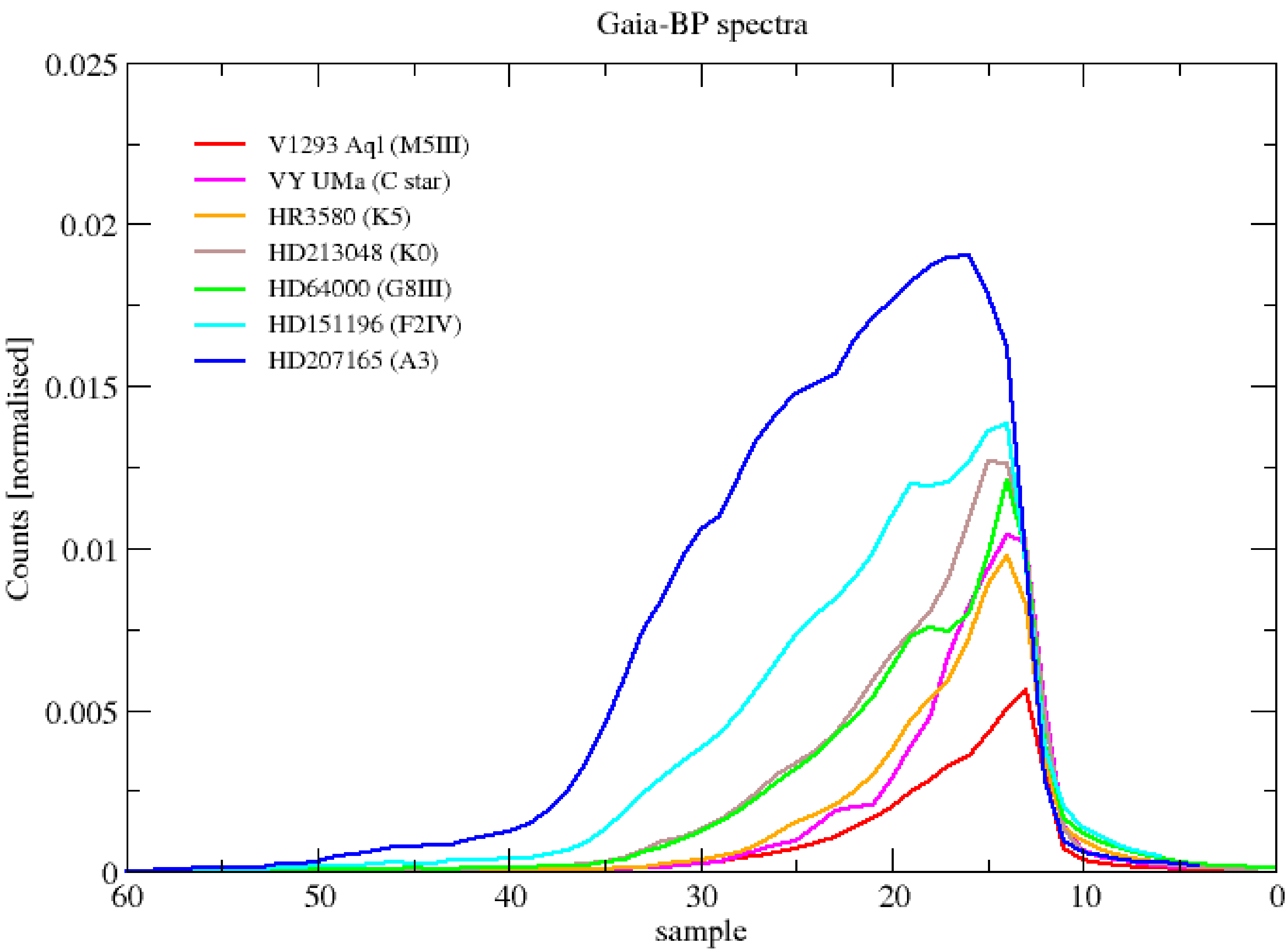
Gaia: basics

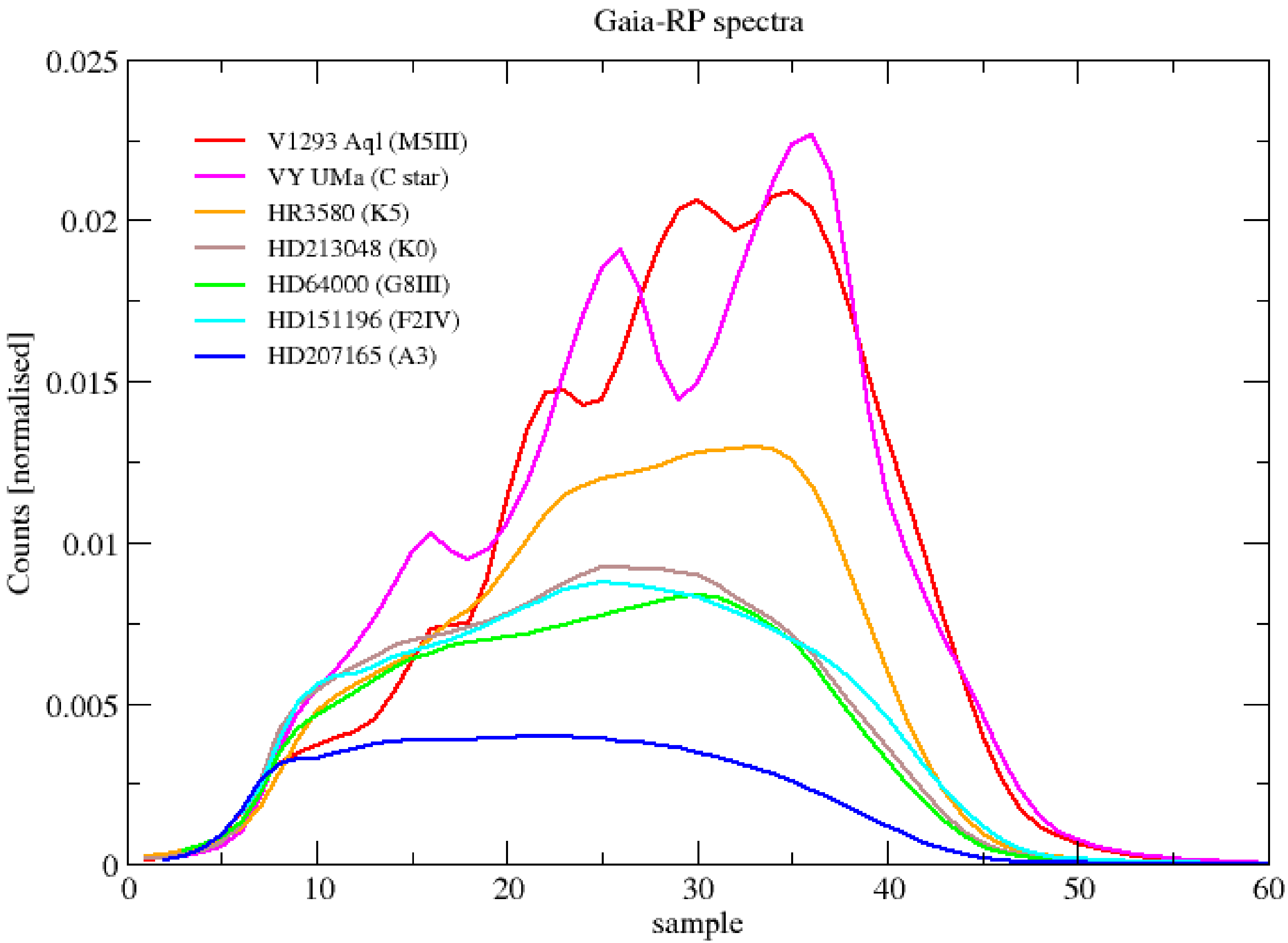
- Spectrophotometry: BP/RP





Gaia-BP spectra
Gaia-RP spectra
V1293 Aql
(M5III)
VY UMa
(C star)
HR3580
(K5)
HD213048
(K0)
HD64000
(G8III)
HD151196
(F2IV)
HD207165
(A3)
Fainter
Brighter

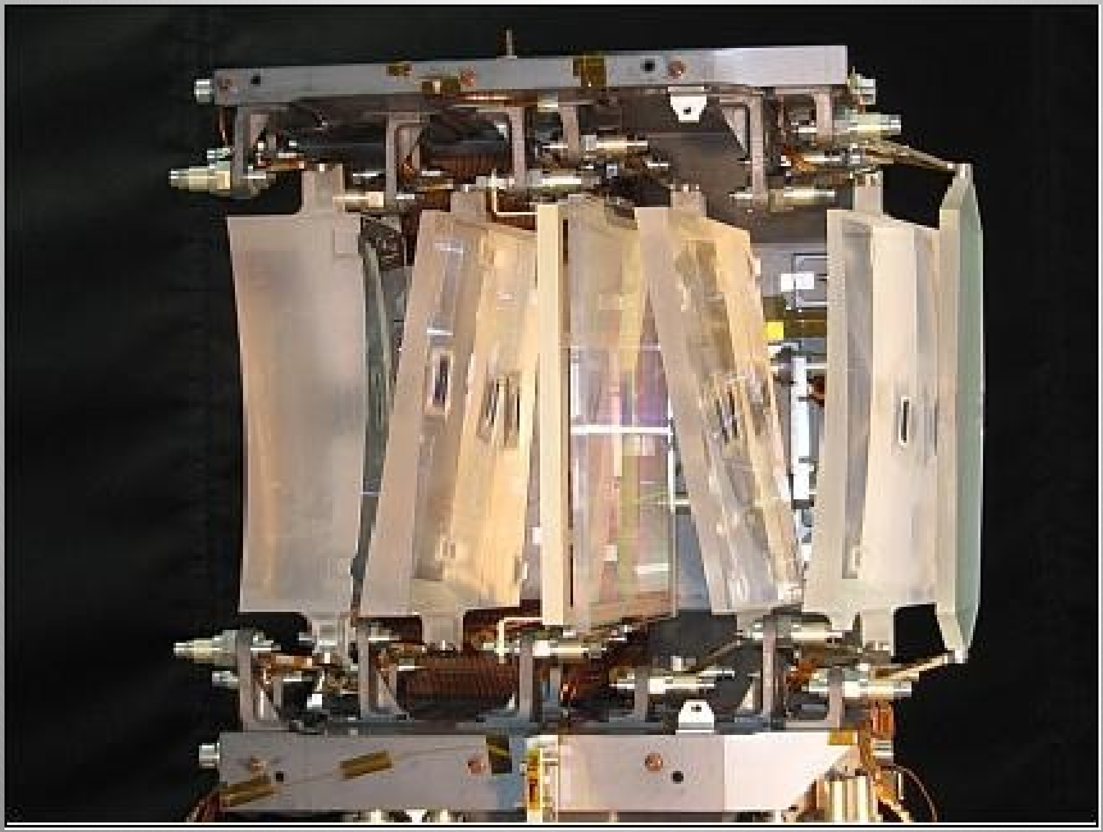

Gaia-BP spectra
Counts [normalised]
0.025
0.02
0.015
0.01
0.005
0
60
50
40
30
20
10
0
sample
V1293 Aql (M5III)
VY UMa (C star)
HR3580 (K5)
HD213048 (K0)
HD64000 (G8III)
HD151196 (F2IV)
HD207165 (A3)

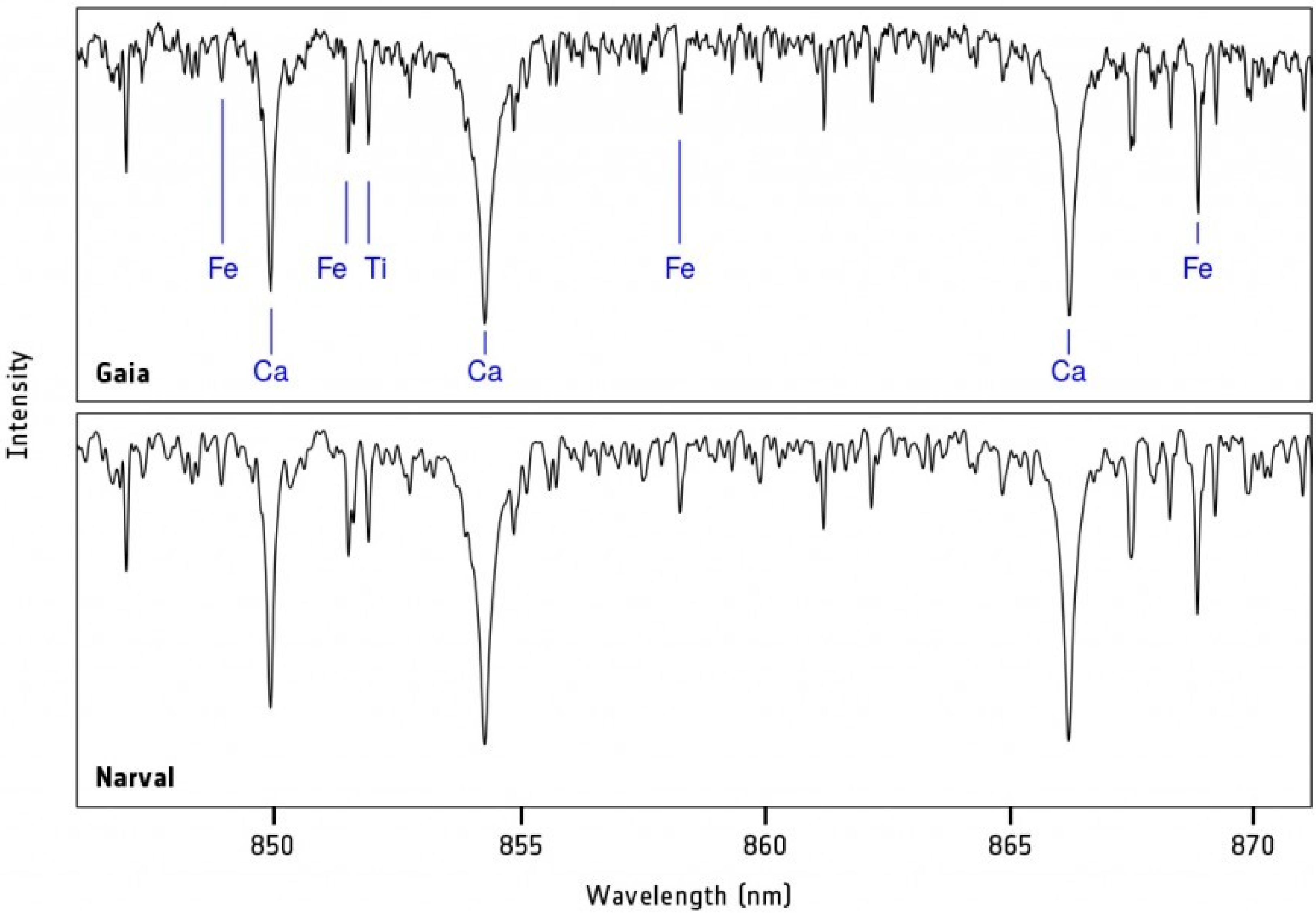


Gaia-RP spectra
V1293 Aql (M5III)
VY UMa (C star)
HR3580 (K5)
HD213048 (K0)
HD64000 (G8III)
HD151196 (F2IV)
HD207165 (A3)
Counts [normalised]
sample
0.025
0.02
0.015
0.01
0.005
0
0
10
20
30
40
50
60

Gaia: basics

- RVS




Intensity
Fe
Fe
Ti
Fe
Fe
Ca
Ca
Ca
Gaia
Narval
850
855
860
865
870
Wavelength (nm)

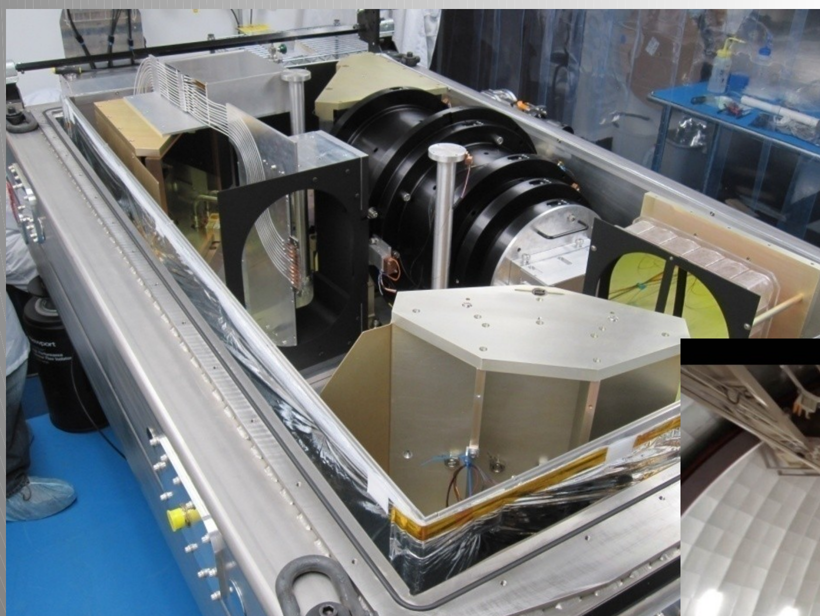
What’s Gaia missing?

- The vast majority of RVS spectra will have a very low S/N ratio: no chemistry information for stars fainter than V~12
- No RVS data for V>16

The answer

- Community organized complementary projects to carry out spectroscpy from the ground

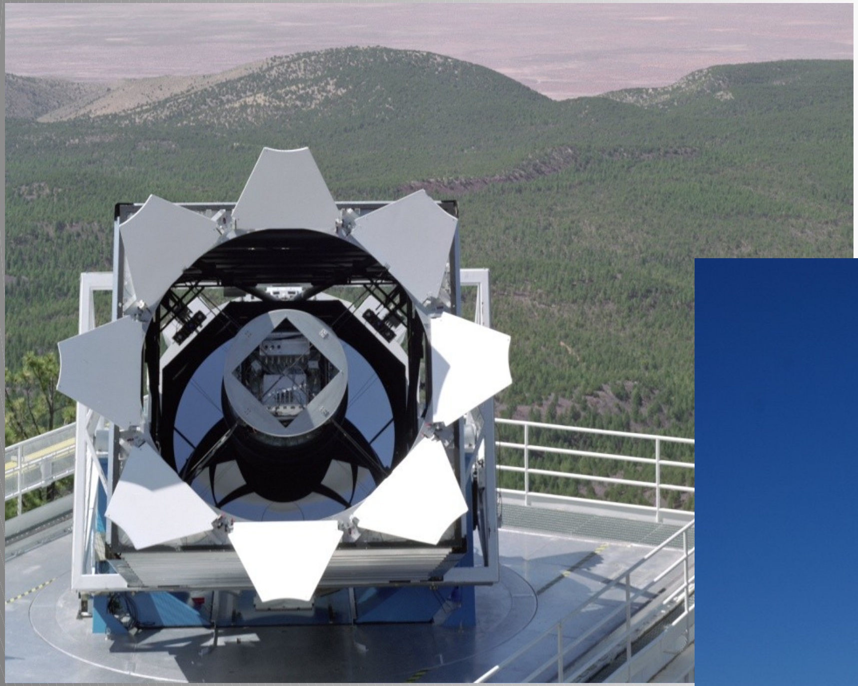
Ongoing high-resolution surveys



APOGEE

HERMES

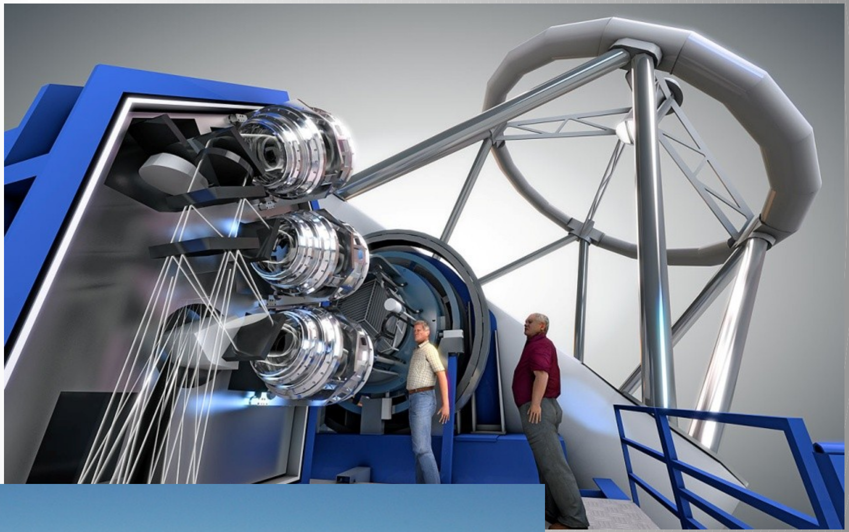
And others at low-resolution…



LAMOST

SDSS

In the future



4MOST
MOONS

DESI

WEAVE

Automation

- Data acquisition
- Reduction
- Analysis

Analysis

- Classification
- Parameterization

1. empirical
2. theoretical

Modeling spectra

- Model atmospheres
- Radiative transfer
- Line formation

Algorithms

1. Projection (ANN, MATISSE …)
2. Local optimization (Nelder-Mead, Newton, conjugate gradient…)
3. Global optimization (genetic algorithms, annealing …)

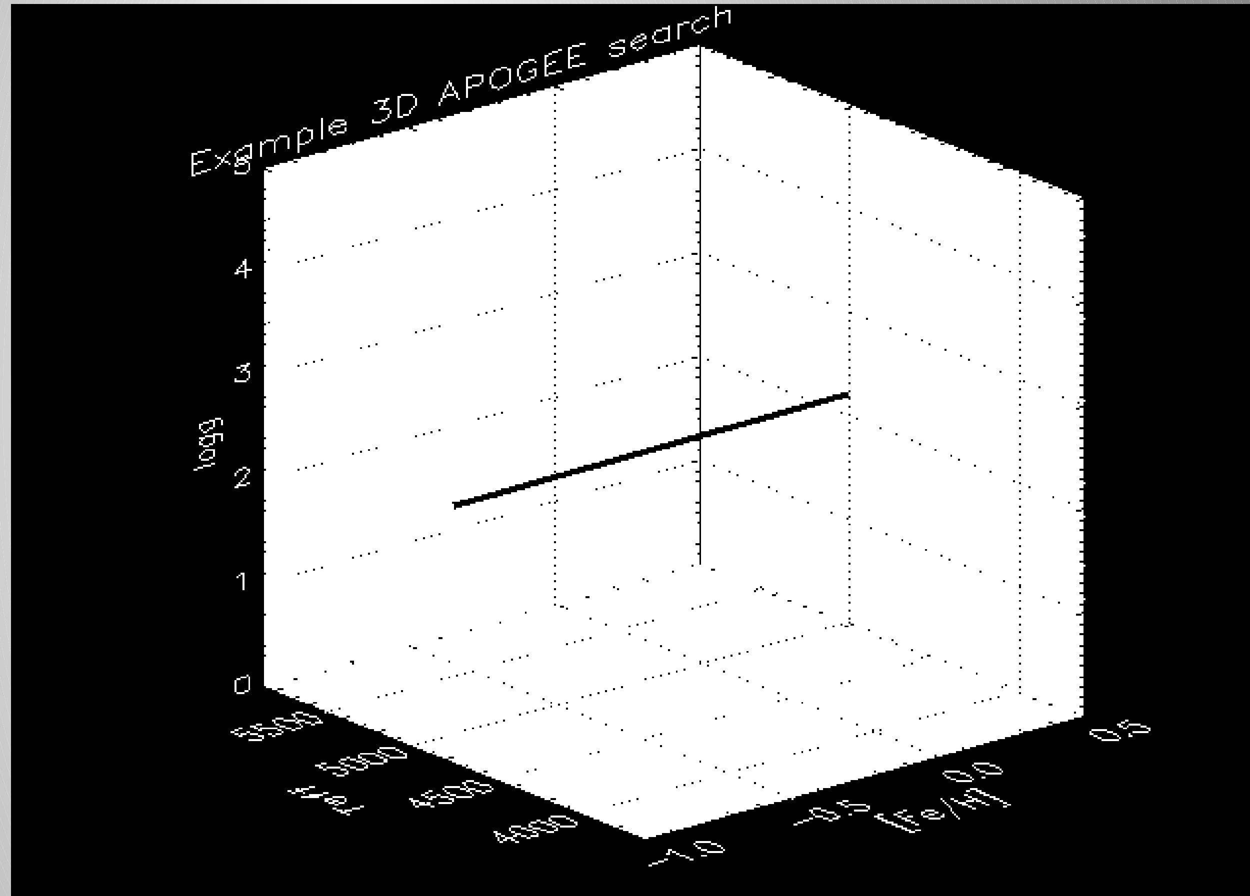
Searches for the best solution in a chi-squared sense

FERRE

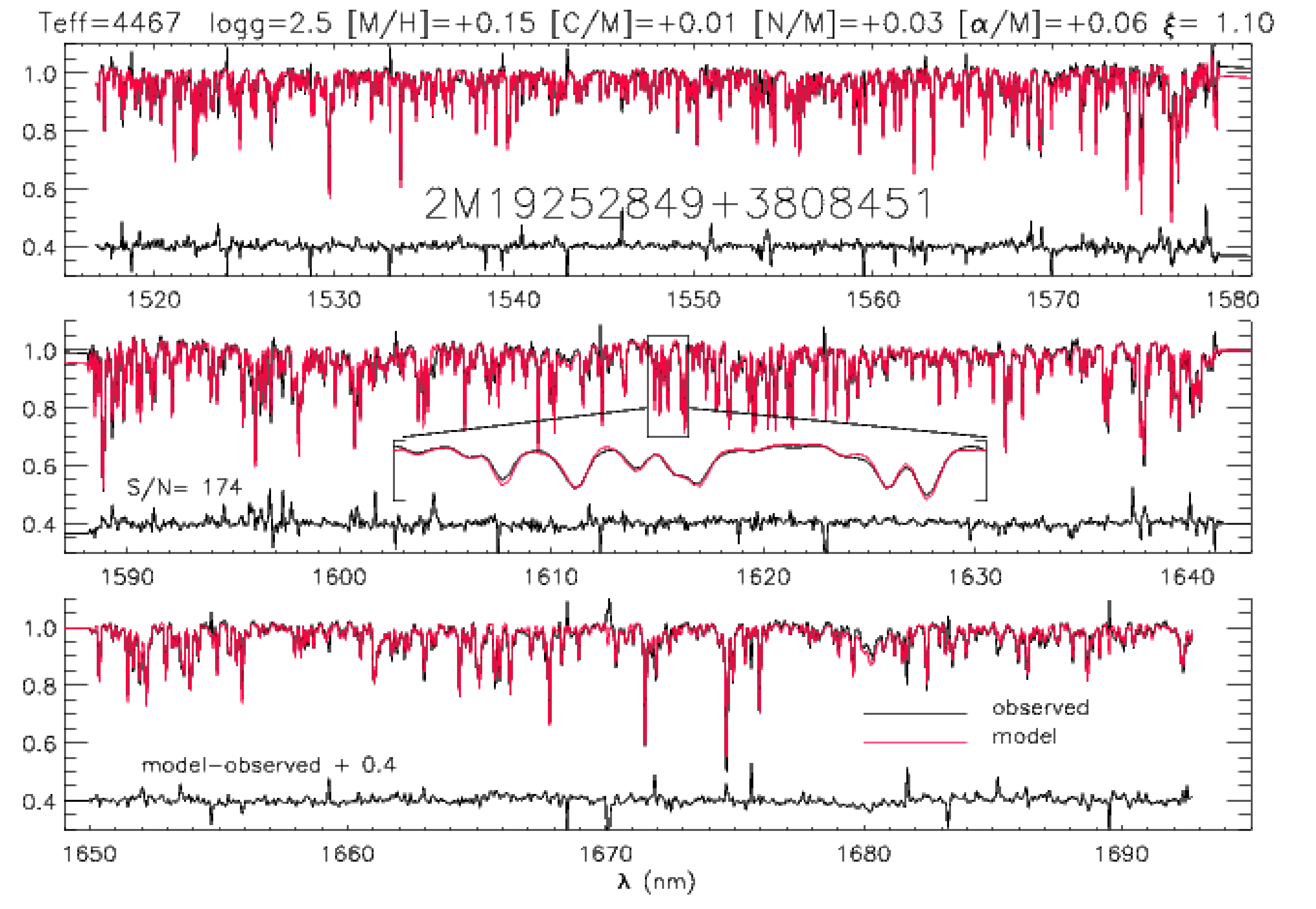
- FORTRAN90
- Highly portable
- Models held in RAM or a database
- Linear, quadratic/cubic bezier, cubic splines interpolation
- PCA compression can be used
- openMP parallelization
- Multiple algorithms (Nelder-Mead, Newton …)
- Highly flexible: fit one, two … all model parameters
- Internal calculation of covariance matrix
- Successfully used on SDSS/SEGUE, APOGEE, VLT, MMT data
- Fast: seconds for a typical search with 7 parameters, millions of models and 1e4 frequencies

Publicly available at **http://hebe.as.utexas.edu/ferre**

3D projection (Teff-logg-[Fe/H]) of a 7D APOGEE search



Example: APOGEE

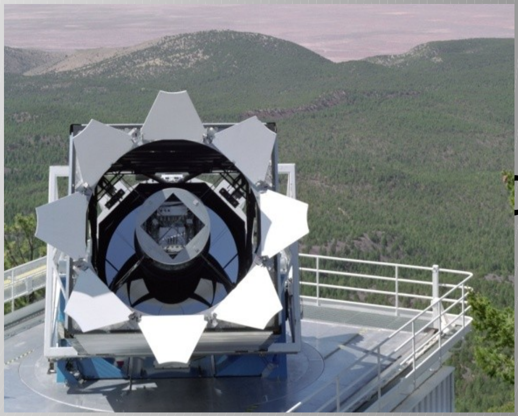


Cross-comparison

- SDSS-SEGUE: SEGUE Stellar Parameter Pipeline (SSPP)
- Gaia-ESO: *nodes* system
- APOGEE: the APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP)

Examples: SDSS/SEGUE

- 380/360-960/1000 nm coverage, R~2000
- SEGUE Stellar Parameters Pipeline (SSPP)
- ‘try all you can’
- Many algorithms adopted d implemented on a single pipeline



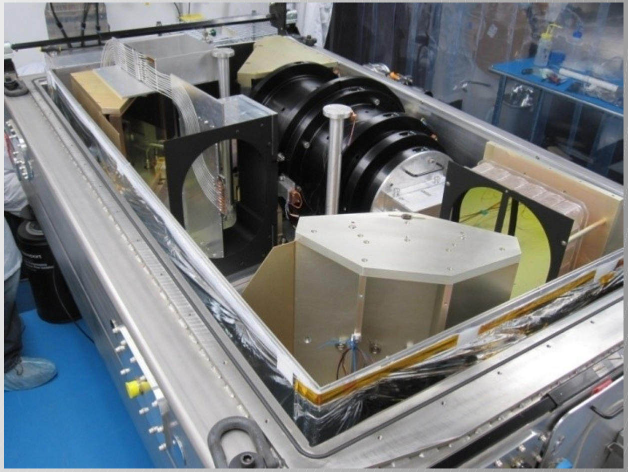
- Some algorithms provide estimates for a single parameter, others for multiple parameters
- Results are combined based on a decision tree

Examples: Gaia-ESO

- GIRAFFE (R~18,000) spectra (HR10-HR21-HR15N…) covering several tens of nm
- Also UVES spectra (R~50,000)
- ‘subcontracting the job’
- Independent nodes download nd analyze the data ‘at home’, return results
- Results are finally homogenize by a third party based on ‘benchmark’ stars

Examples: APOGEE

- R~22,500 in the H band (1500-1700 nm)
- Highly uniform sampling
- ‘single algorithm’
- Efforts concentrate on implementing and testing a single algorithm



- Single pipeline ran on a dedicated machine under strict version control

Cross comparison

- Ease of implementation
- Computational demand
- Error calculation
- Human resource demands
- Repeatability
- Handle on systematics
- Clarity/traceability

Conclusions

- Importance of developing software under version control that runs and it is maintained at a given location. Otherwise repeatability and traceability are compromised, and huge human efforts are required to run
- Better to focus on one or few algorithms, implemented afresh and thoroughly tested, rather than ‘as many as you can get’
- Multiple algorithms can only provide estimates of systematic errors if truly independent, i.e. independent atomic/molecular data, model atmospheres, synthesis codes, etc.