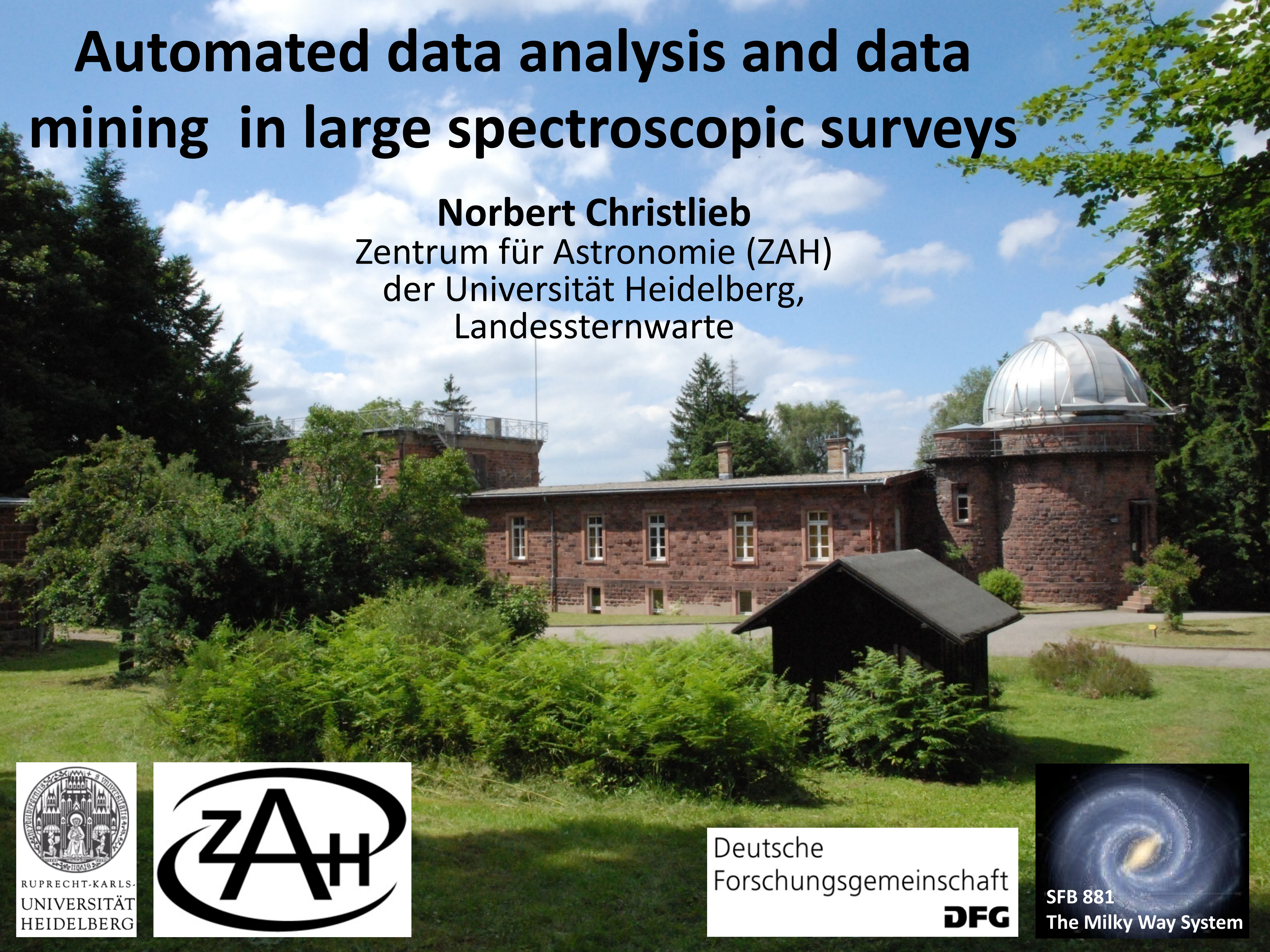
Automated data analysis and data mining in large spectroscopic surveys

Norbert Christlieb
Zentrum für Astronomie (ZAH)
der Universität Heidelberg,
Landessternwarte



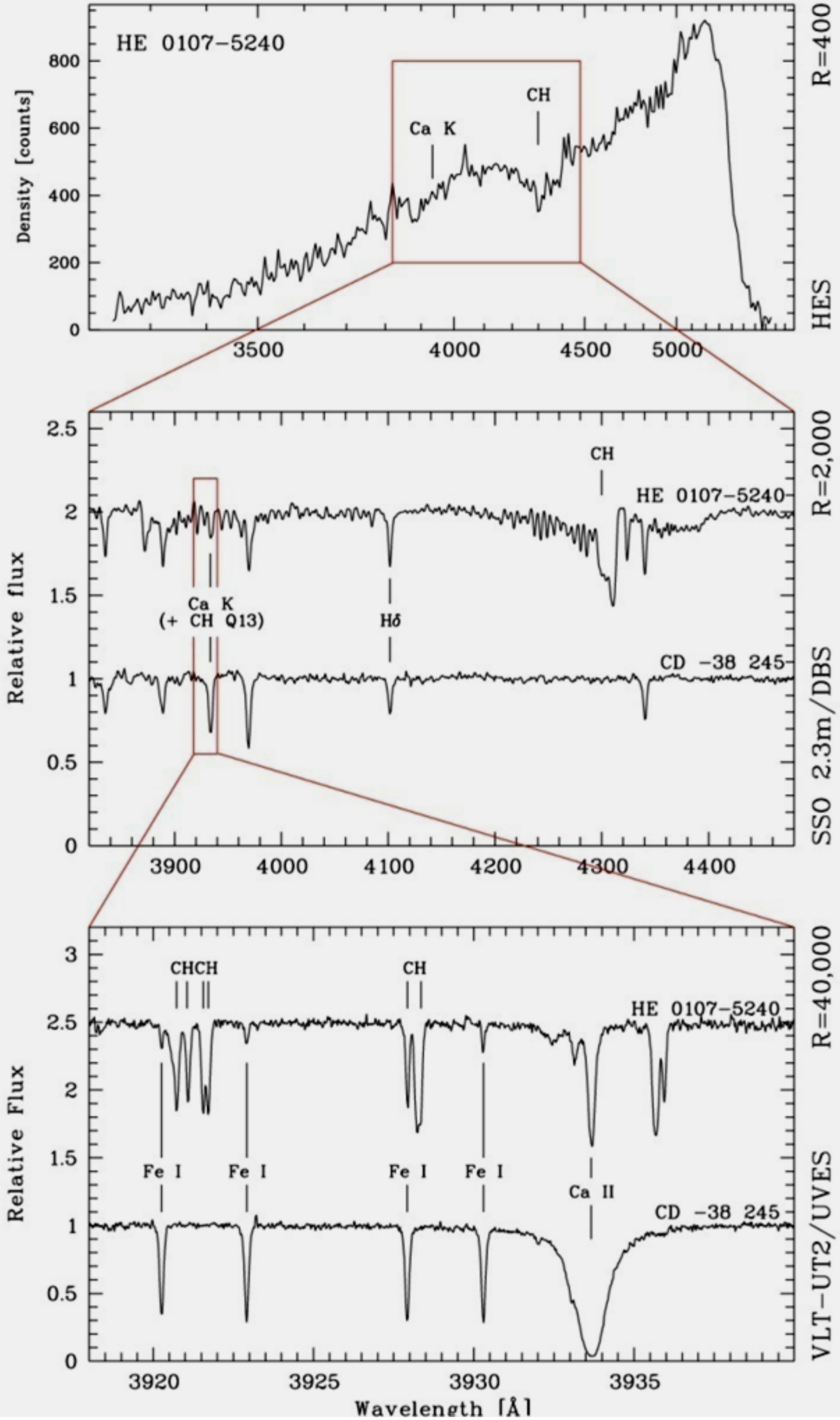
RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

ZAH

Deutsche
Forschungsgemeinschaft
DFG

SFB 881
The Milky Way System

Contents



Low resolution spectroscopy:

$R = \lambda/\Delta\lambda \approx 10-5{,}000$

Only very few, very strong lines and absorption features visible.

Medium resolution spectroscopy:

$R = \lambda/\Delta\lambda \approx 5{,}000-20{,}000$

Lines of a few elements visible (but depends strongly on *R*, *S/N,* elemental abundance, effective temperature of the star).

High resolution spectroscopy:

$R = \lambda/\Delta\lambda > 20{,}000$

Lines of many elements visible; “full” abundance analysis and determination of isotopic ratios of a few elements possible.

Selected upcoming spectroscopic sky surveys

Survey	**Targets**	***N* (approx.)**	λ	***R***	**Time**	**Remarks**
APOGEE	Stars	**100,000**	1.52-1.68 μm	30,000	2011-	
Gaia-ESO	Stars	**100,000**	Optical	20,000	2011-14	proposed
LAMOST	Stars, galaxies, quasars	**7,500,000** **5,700,000** **600,000**	370-900 nm 510-550 nm 830-890 nm	1700 10,000 10,000	2011-16	
HERMES	Stars	**1,000,000**	Optical	28,000	2012?-	
Gaia	No selection	**5,000,000**	847-874 nm	11,500	2013-18	
4MOST	Stars, galaxies, quasars	**25,000,000**	Optical	2000-20,000	2018-22	proposed

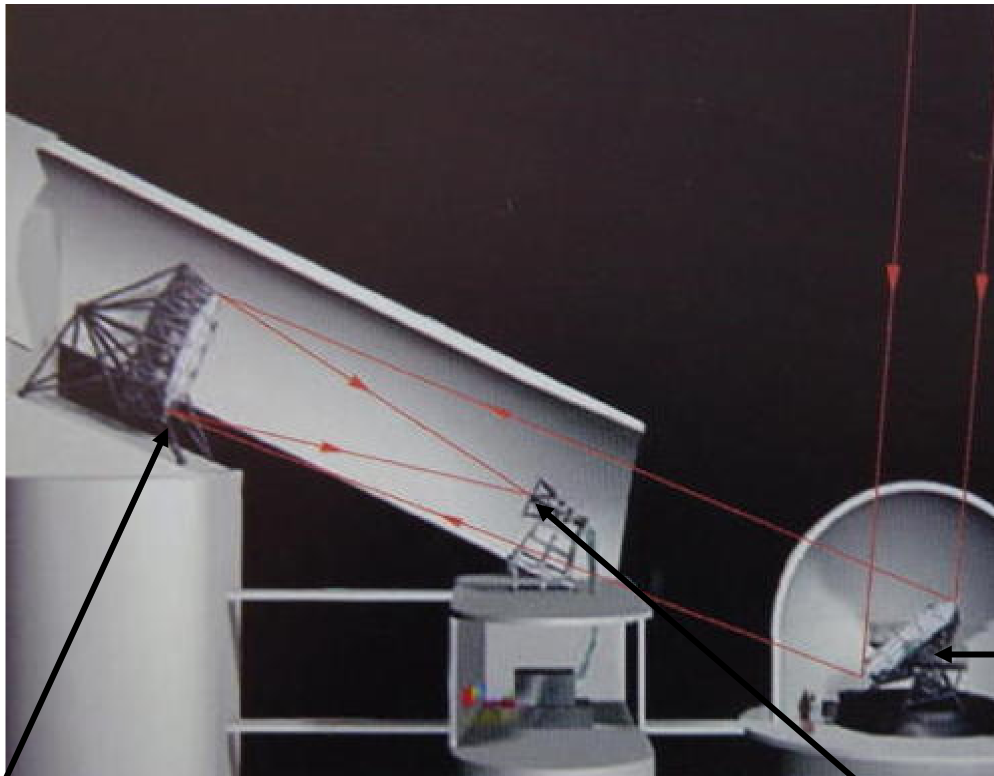
=> Automated data analysis methods are mandatory!

LAMOST at Xinglong Station

www.lamost.org

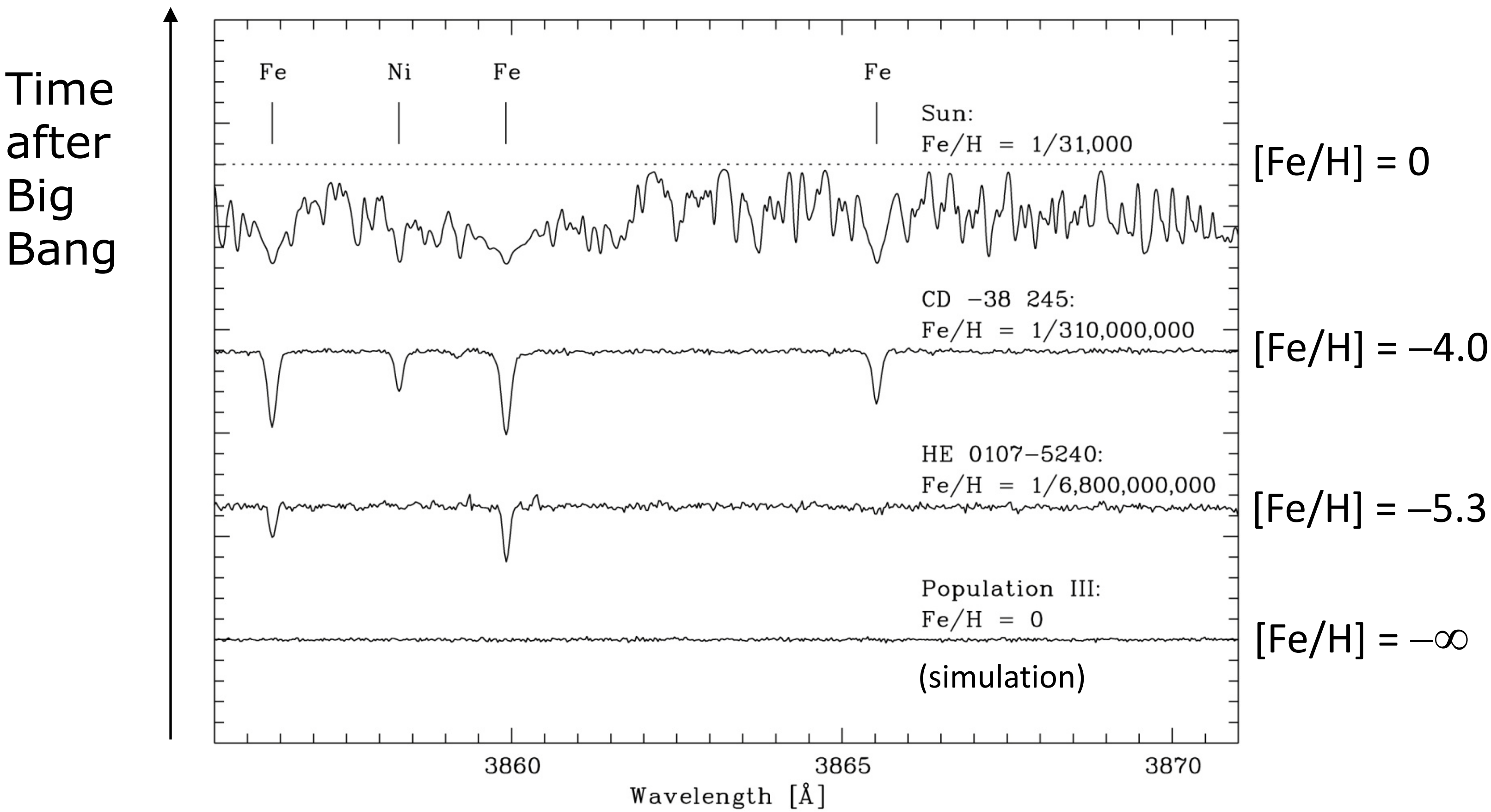
- Effective aperture: 4m
- 5° diameter field of view
- 4,000 fibers
- 16 fiber-fed two-arm spectrographs for low- to medium-resolution
- *S/N* = 10 at 1 nm resolution for a 20.5 mag star in 1.5 hours

LAMOST optical design



It is a meridian reflecting Schmidt telescope, laid down on the ground with its optical axis fixed in the meridian plane

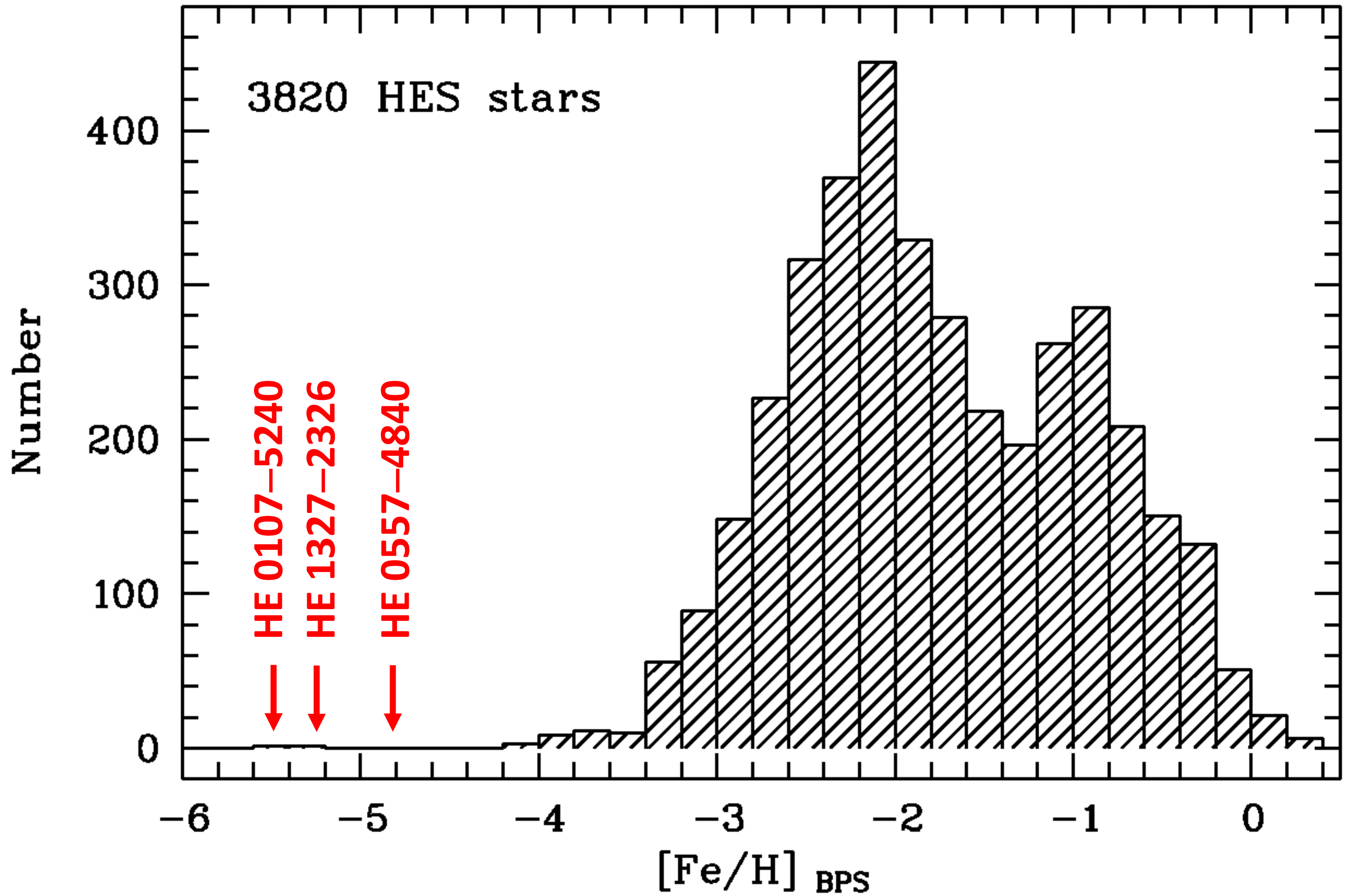
The chemical evolution of the Universe



$[X/Y] = \log_{10} (N_X/N_Y)_{star} - \log_{10} (N_X/N_Y)_{Sun}$ for elements X, Y

N = number density of atoms

The metallicity distribution function of the Galactic halo

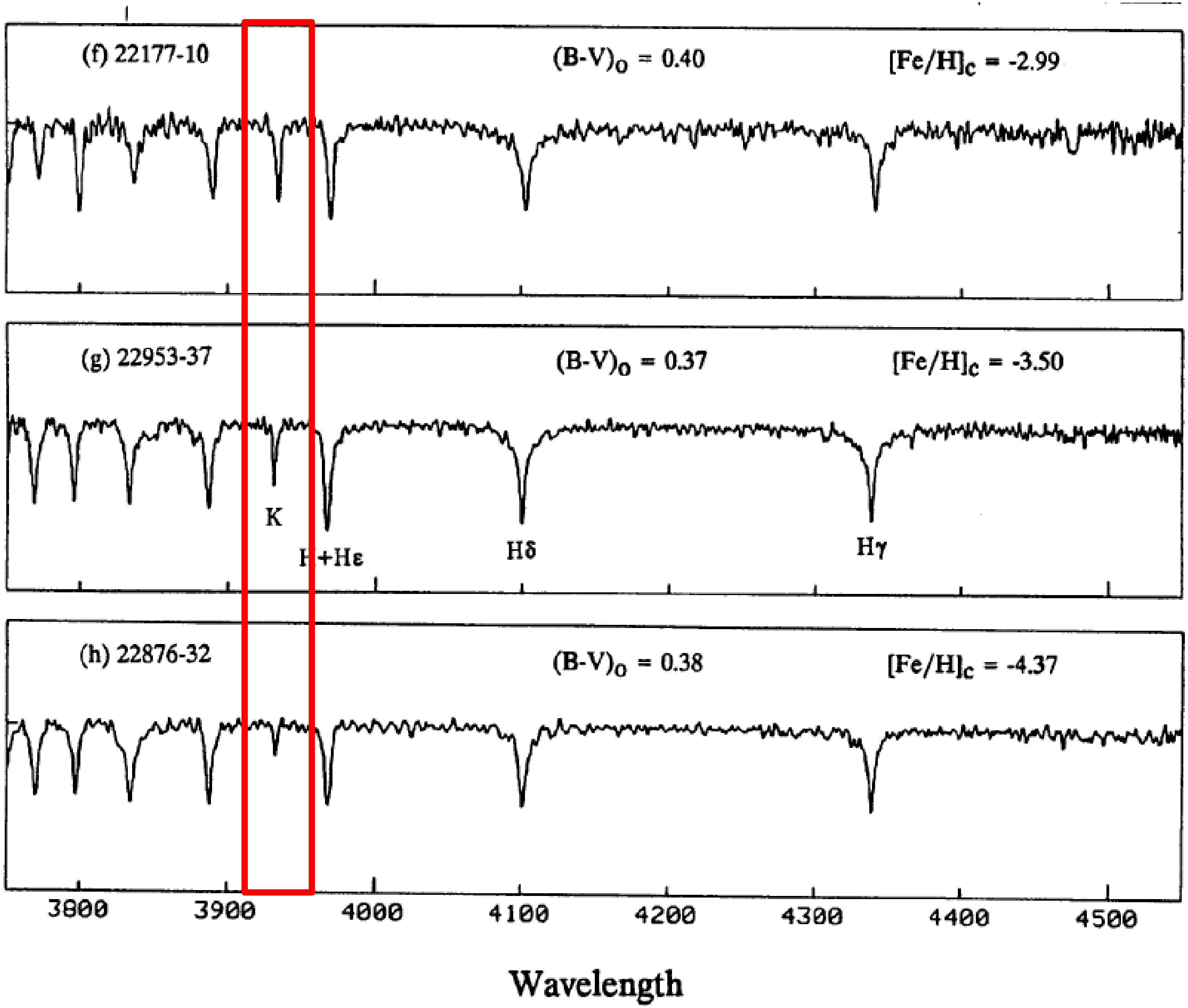


The LAMOST metal-poor star survey

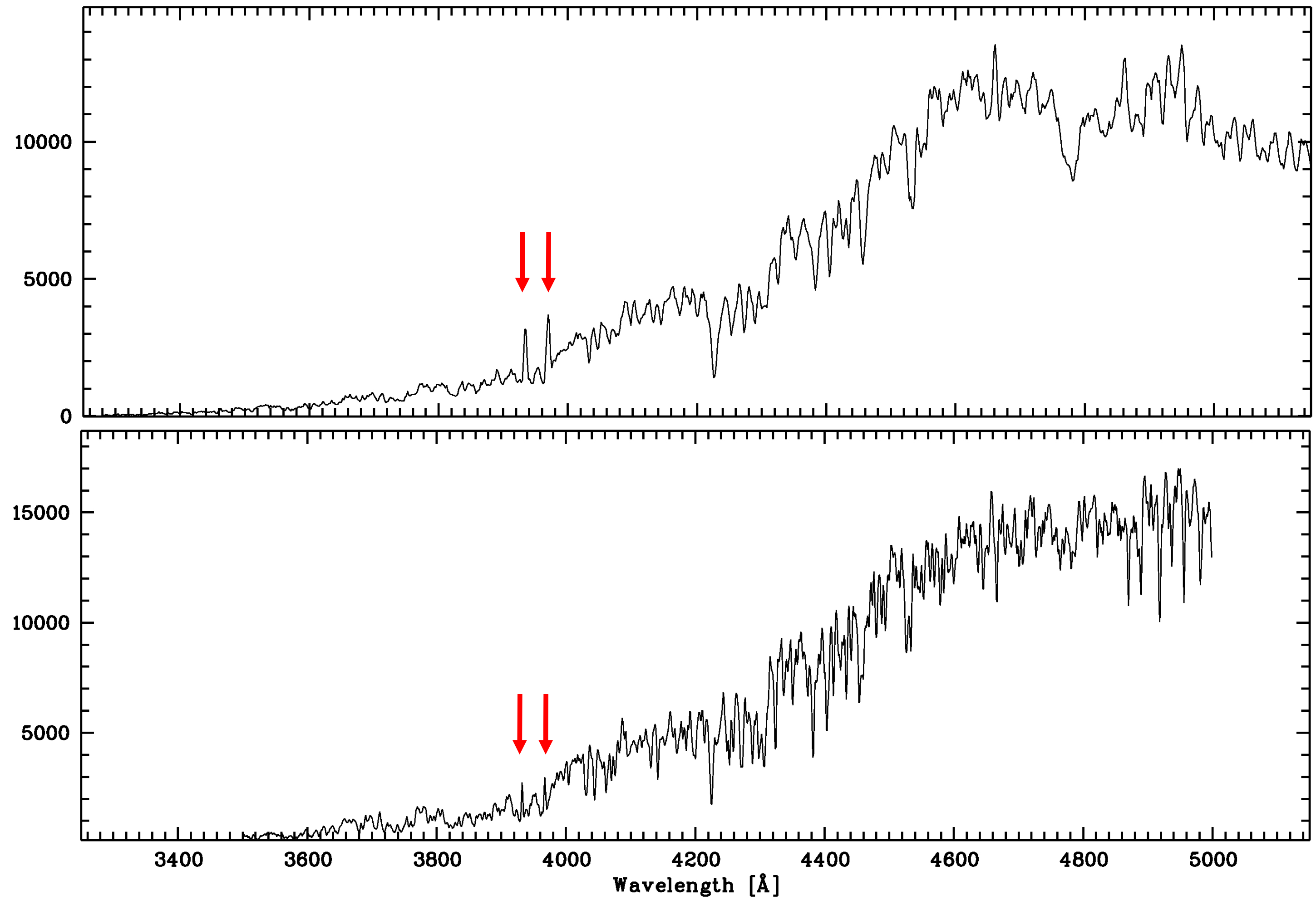
- Aims:
  - Identifying the most metal-poor stars for abundance analysis based on high resolution (R > 20,000) spectroscopy.
  - Statistical studies; e.g., metallicity distribution functions.
- Input catalog: 5.1 million stars selected from SDSS DR7.
- Medium-resolution (R = 1700) spectroscopy of 2.5 million stars => [Fe/H], [C/Fe].

Survey	Effective sky coverage	Effective mag limit	N < –3.0 (EMP)	N < –5.0 (HMP)
HES	6,400 deg^2	B < 16.5	200	2
SEGUE	1,000 deg^2	B < 19	1,000	10
LAMOST	12,200 deg^2	B < 18.0	3,000	30
SSS	20,000 deg^2	B < 17.5	2,500	25

Metallicity @ *R* = 2000

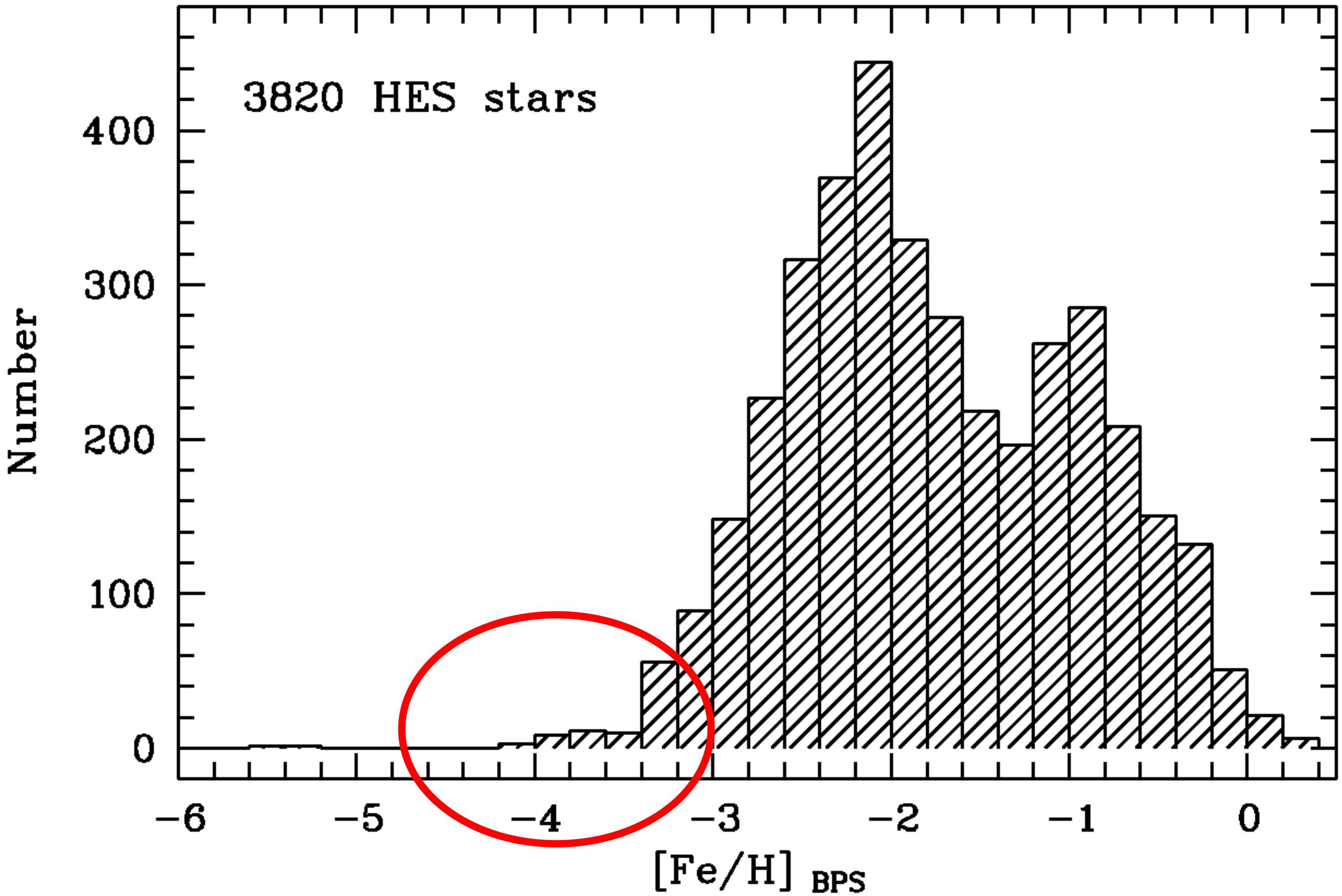


dMe stars

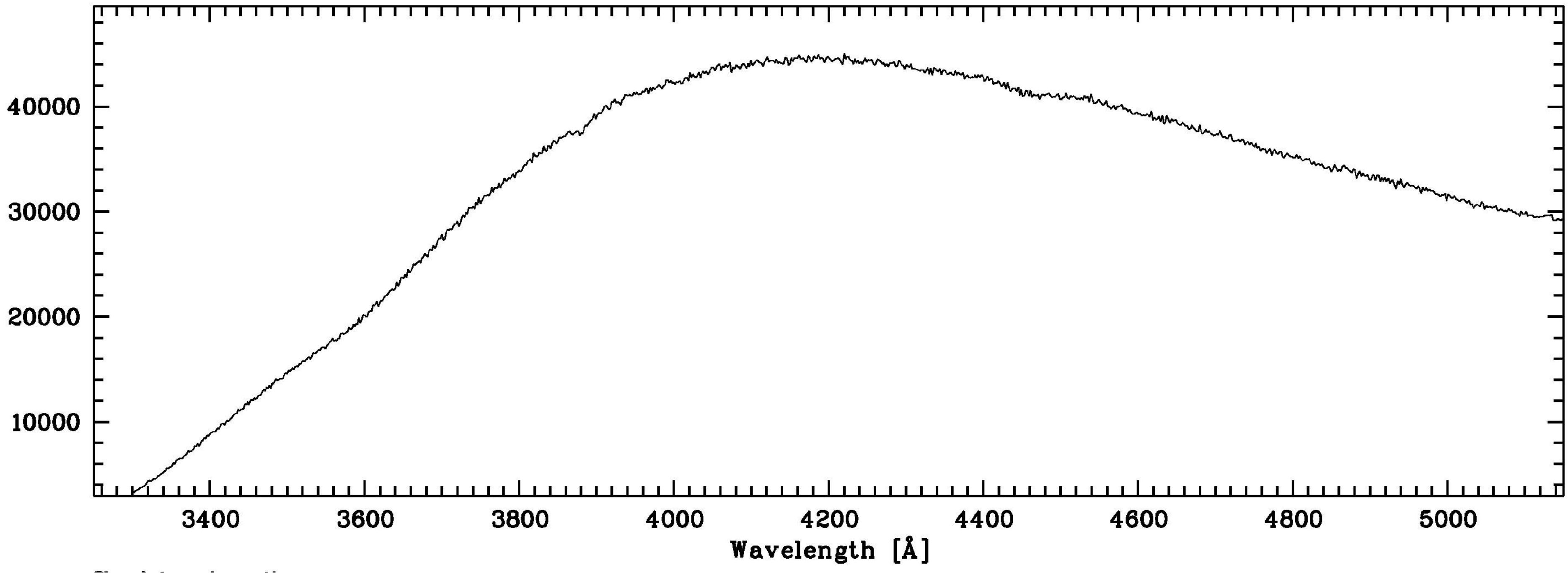


In samples of cool stars selected to be metal-poor by means of Ca K line strength, typically **a few percent** of the objects are dMe stars.

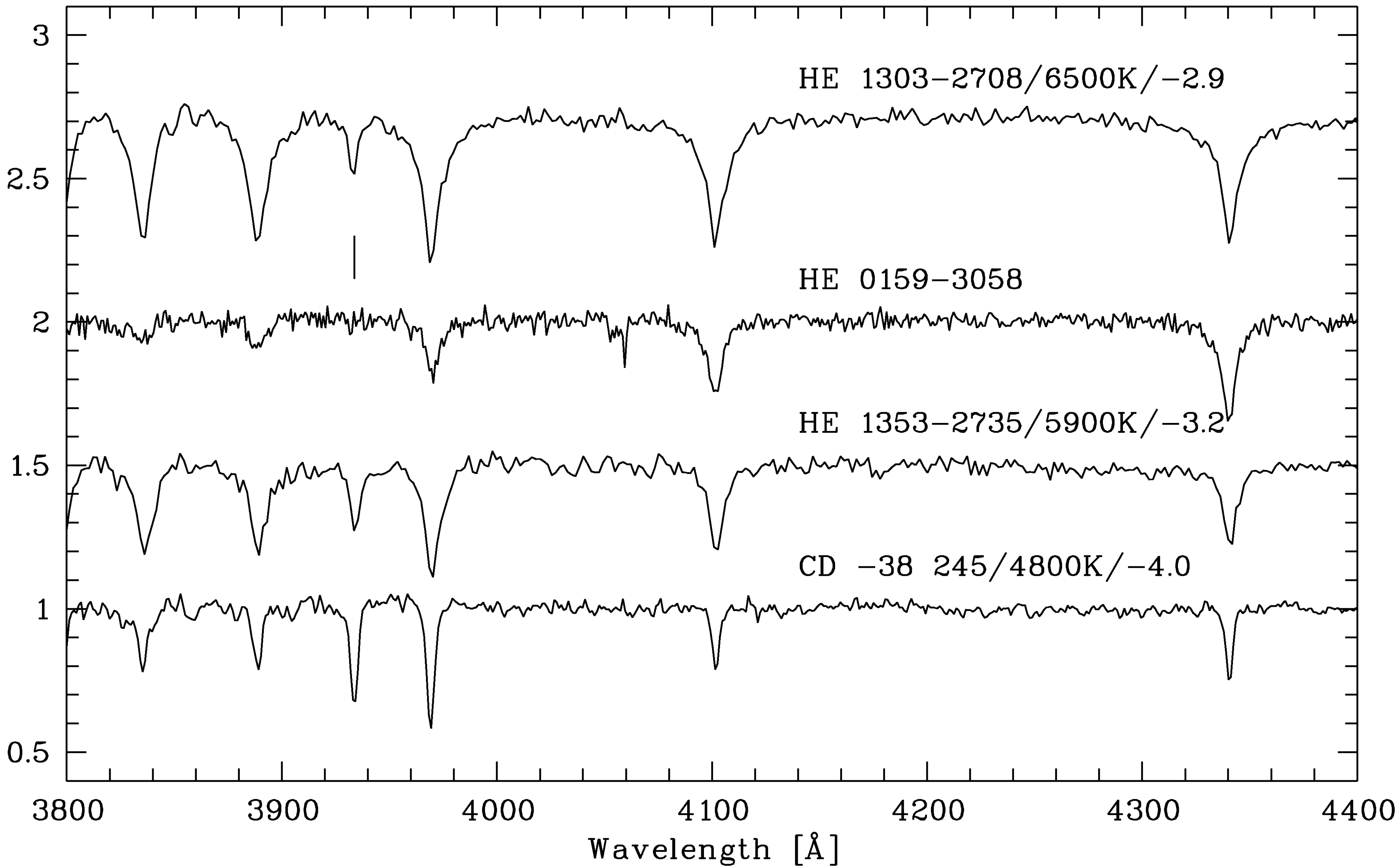
The metallicity distribution function of the Galactic halo



DC white dwarfs



Metal-poor stars vs. cool DA white dwarfs

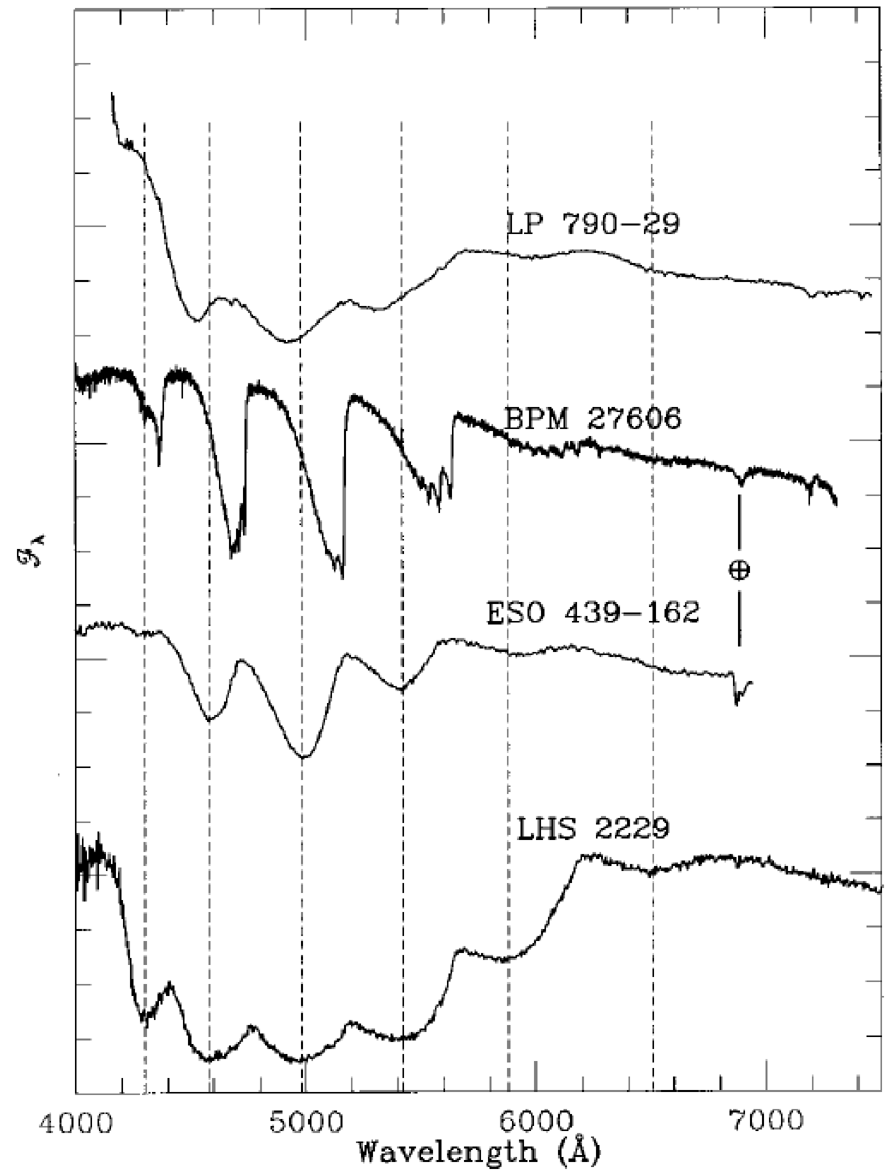


Strategy

- Automated detection of the relevant types of contaminating objects.
- Use "intelligent" features; e.g., Balmer line decrement.
- Apply various kinds of consistency checks.
  E.g., if no Ca K line detected, check if Balmer lines are there.

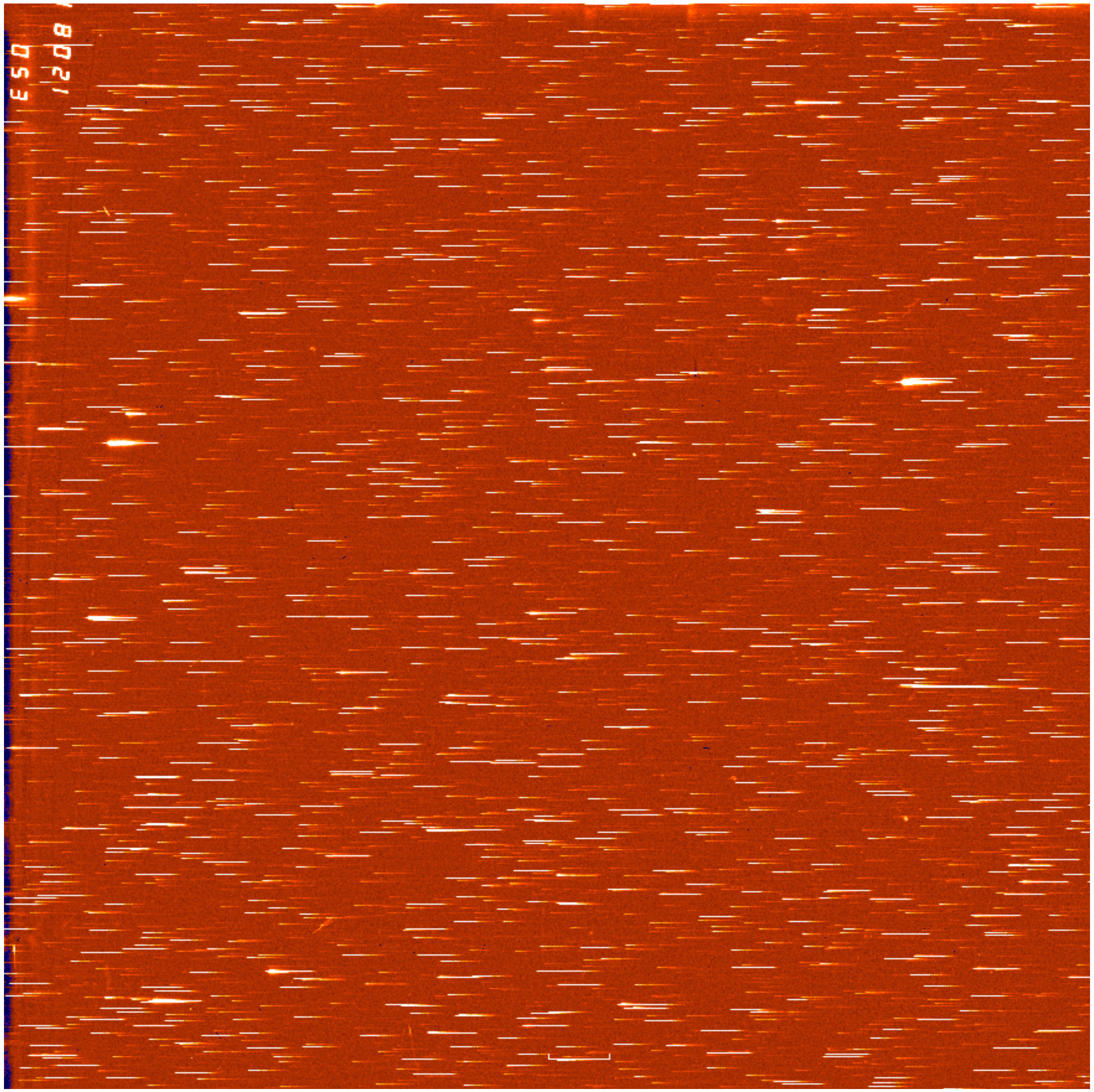
Search for "unusual" objects in the HES

- **Aim:** Identification of new, scientifically interesting classes of objects.
- **Method:** Search for outliers in projections of multi-dimensional parameter space.
- **Results:** None…



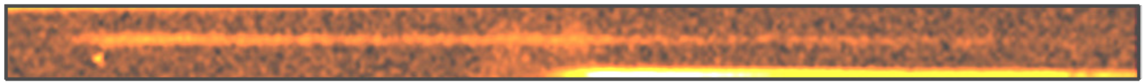
Schmidt et al. (1999, ApJ 512, 916)

The Hamburg/ESO Survey (HES)

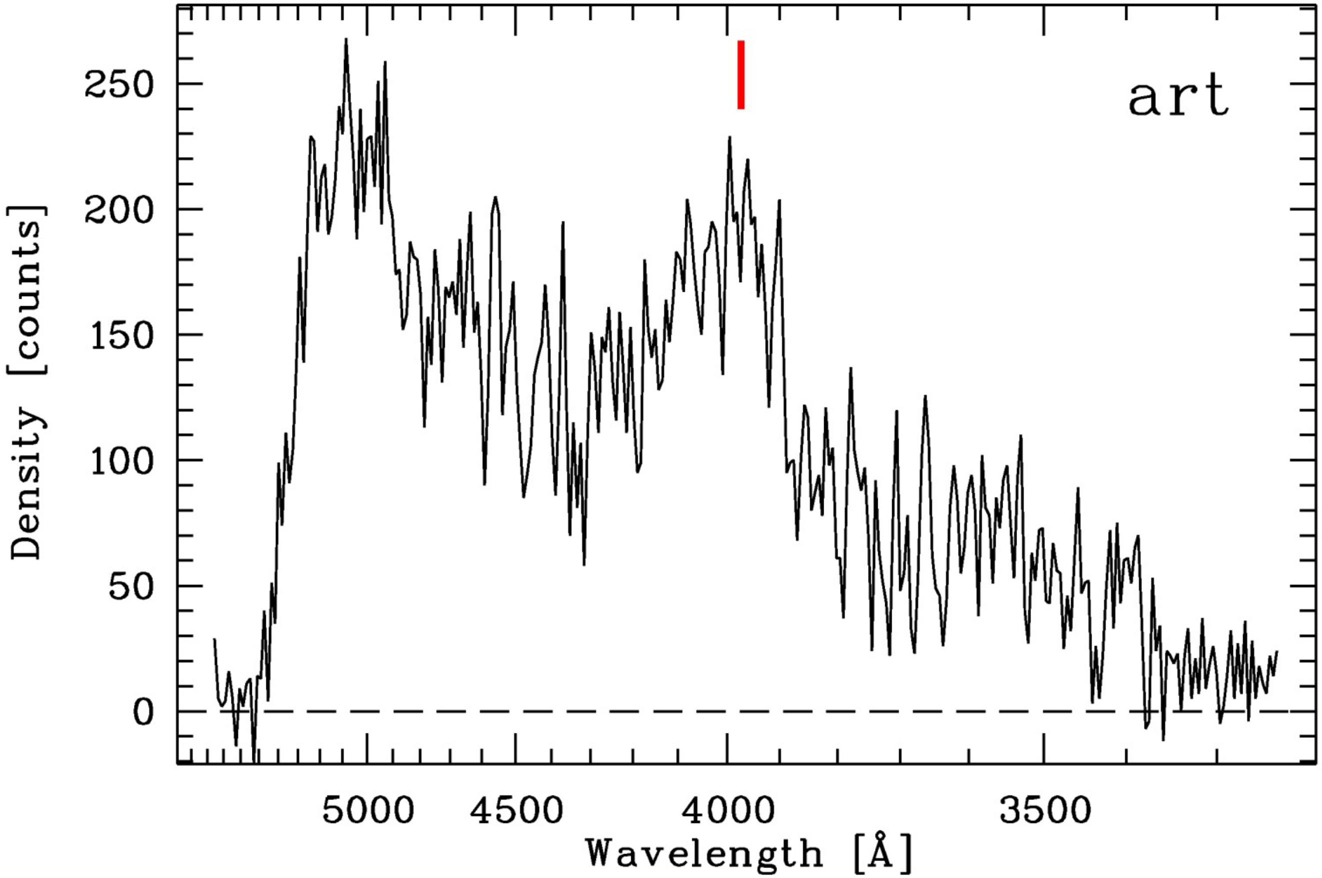


- Nominal sky coverage of one plate: 5° x 5°.
- $10 < B < 18$.
- $\Delta\lambda \approx 10$Å @ Ca II K.
- Typically ~30,000 spectra per plate.
- 379 plates => about 12 million spectra.
- On average, only 2-3 stars at [Fe/H] < –3.0 per plate.

HES outliers

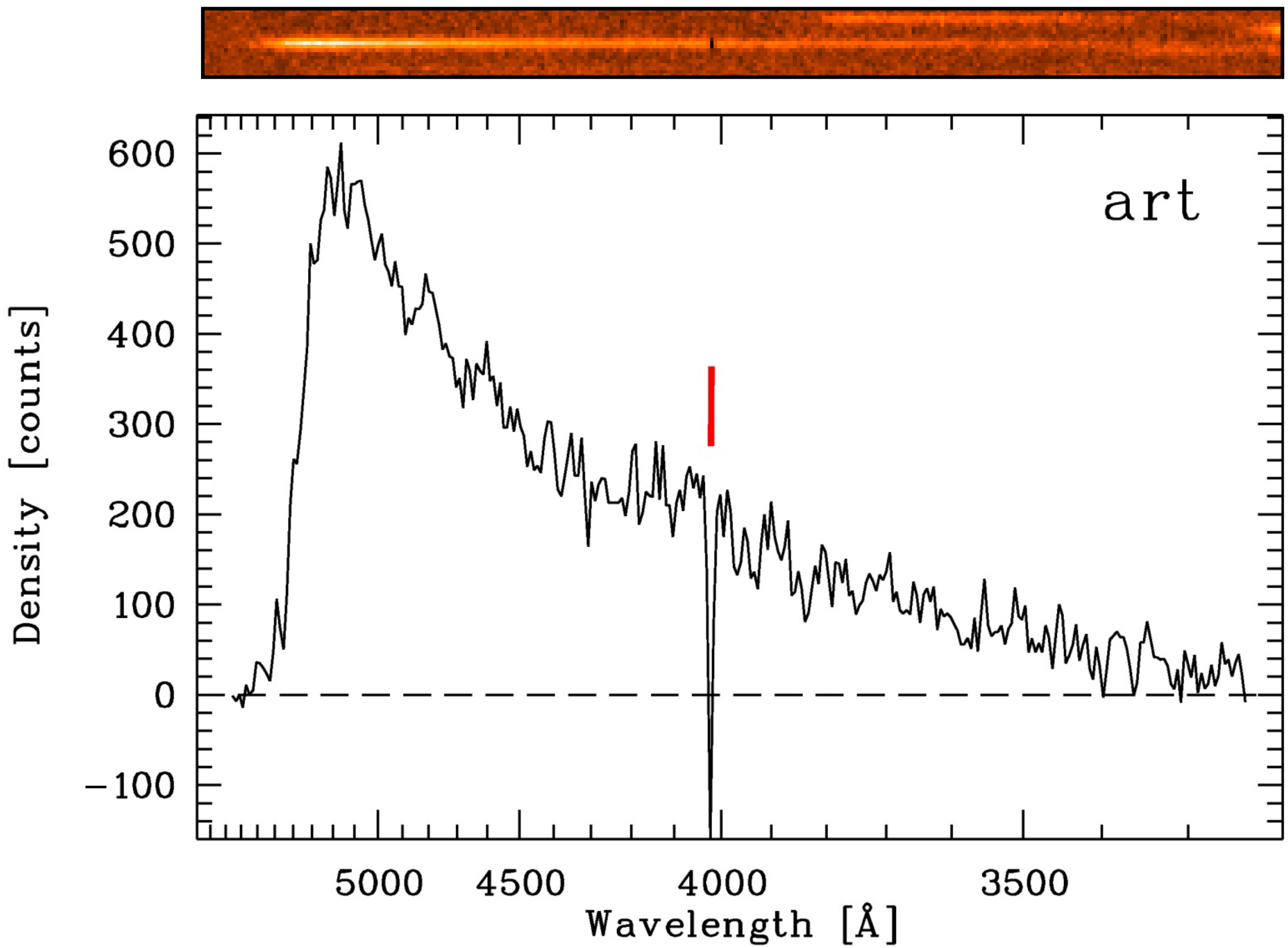


HES plate



Extracted HES spectrum

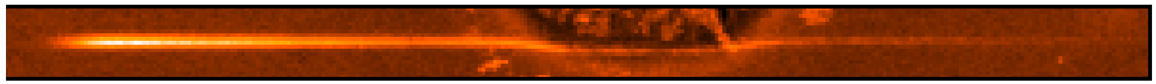
HES outliers



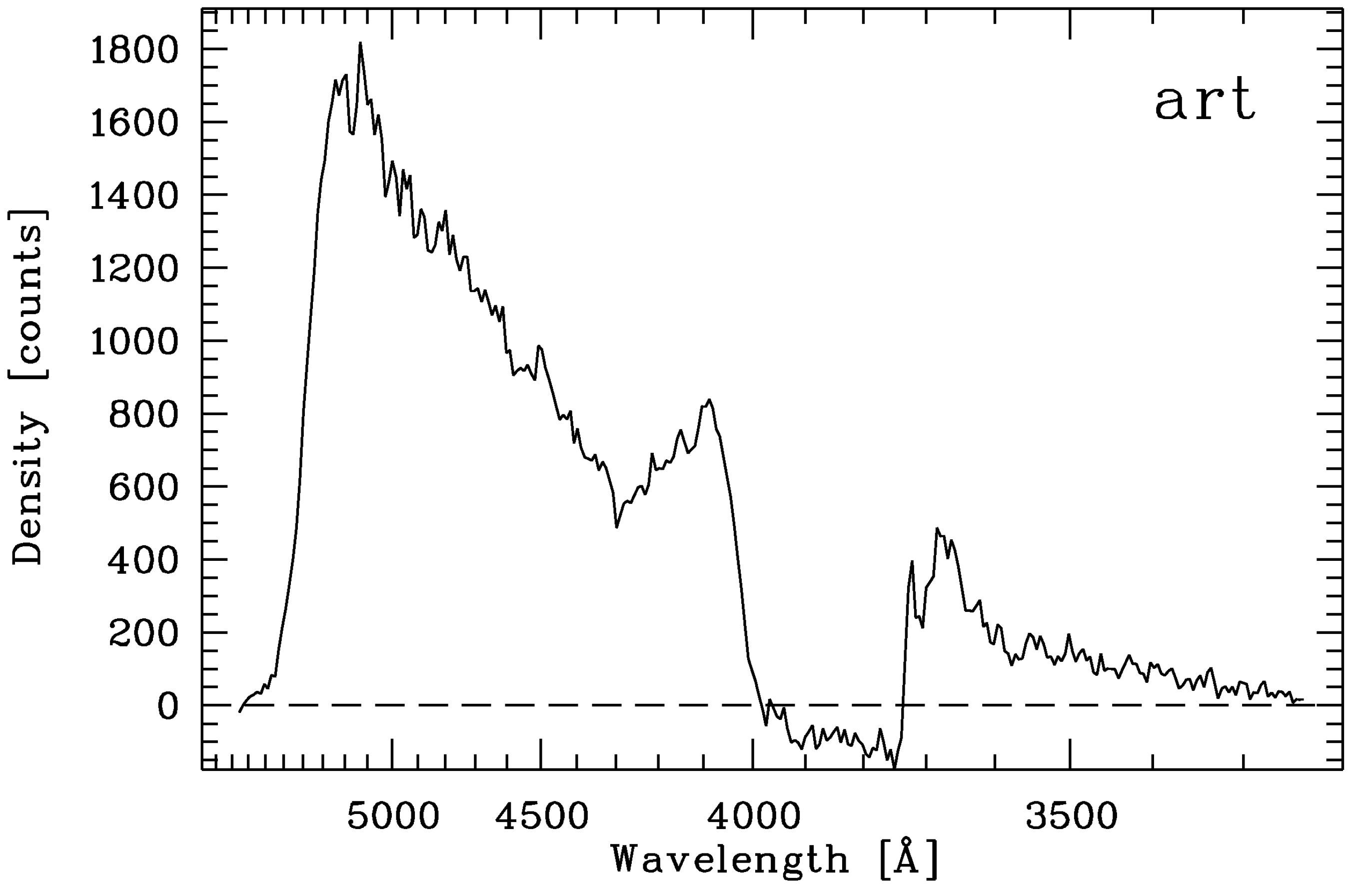
HES plate

Extracted HES spectrum

HES outliers



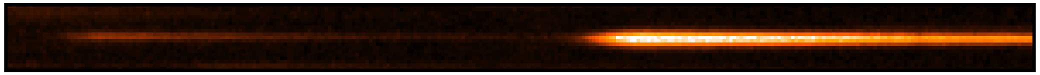
HES plate



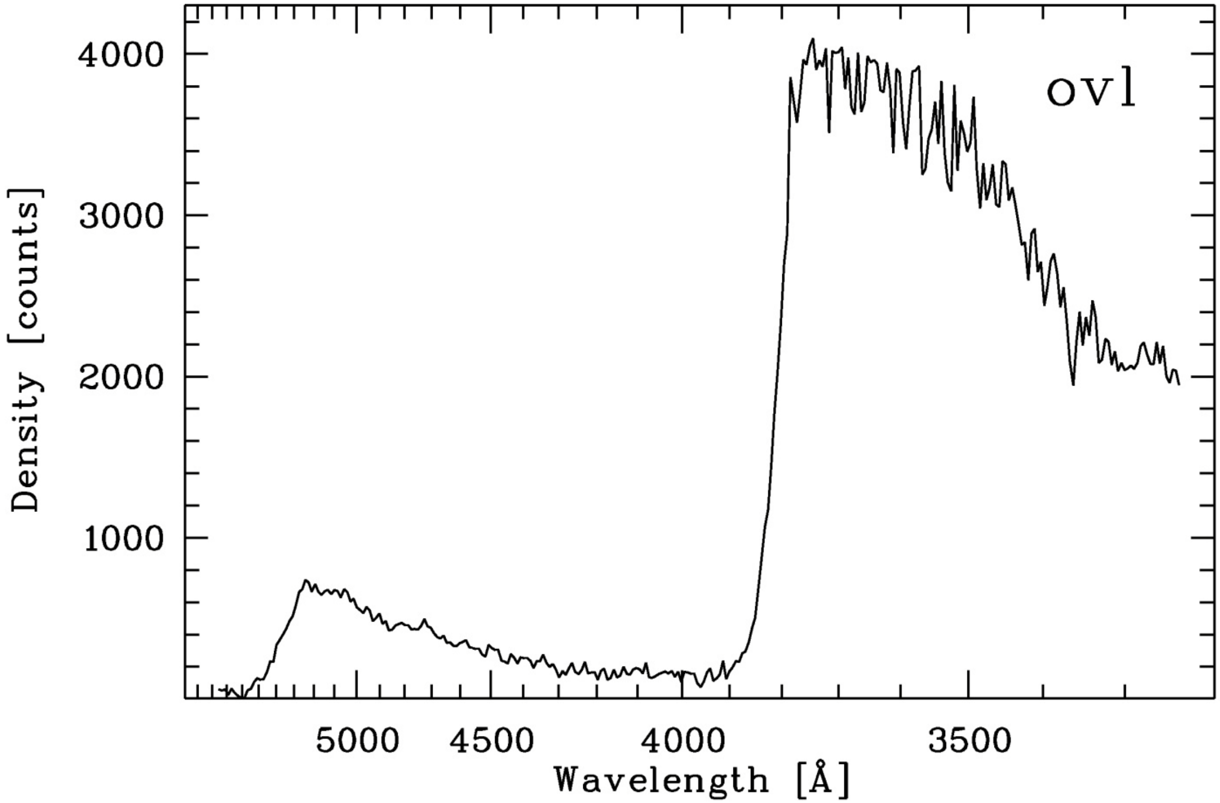
Extracted HES spectrum

HES outliers

DSS-I direct plate



HES plate



Extracted HES spectrum

Conclusions

- Detection of “unusual” objects is only feasible if instrumental artifacts have been reliably removed.
- For the exploitation of large spectroscopic sky surveys, we need to develop methods for automated detection of objects that contaminate the samples of the objects we want to study.