Data Mining of Be stars in VO

Petr Škoda & Jaroslav Vážný

Astronomical Institute Academy of Sciences
Ondřejov
Czech Republic
&
Faculty of Theoretical Physics and Astrophysics
Masaryk University
Brno

IVOA Interoperability meeting KDD IG Session,
Naples, 16th May 2011

Motivations for Be Stars

The ~40 years of research in Ondřejov (data)

Stellar astronomers – non VO involved at all

- way to convince them to use VO (their topic)

Be stars are mysterious (after >100 years)

Challange for SSA postprocessing

Different time scales, quasi periodicity (not sure)

- emission epizodes, can look normal (20% of B = Be!)

Zoo of line profiles (winebottle, abs+em, high em)

Astro Data Mining

Its fancy – emerging – attracting informaticians

Photometric data mining (using color indices)

Spectral data mining

- specific profiles – emission in Halfa (other ?)
- parameters for characterization – Gaussian fit

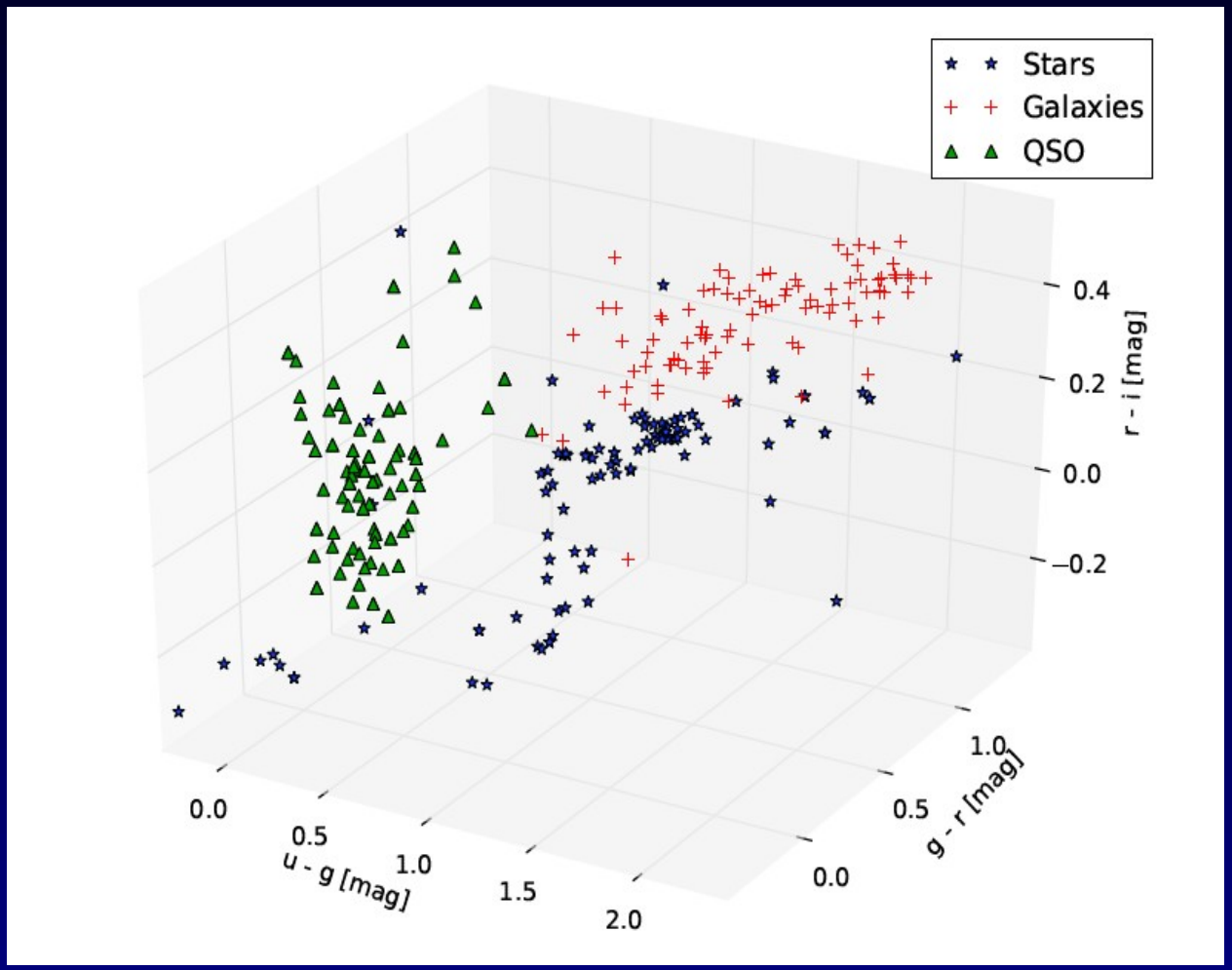
How the VO standards supports such DM ?

Classification of different objects

SDSS (types in CAS)

Feasibility study – J.V Bachelor + semestral in DM

s	g	q	<-- classified as
88	7	5	s
7	92	0	g
2	0	97	q



100 stars
99 galaxies
99 QSO

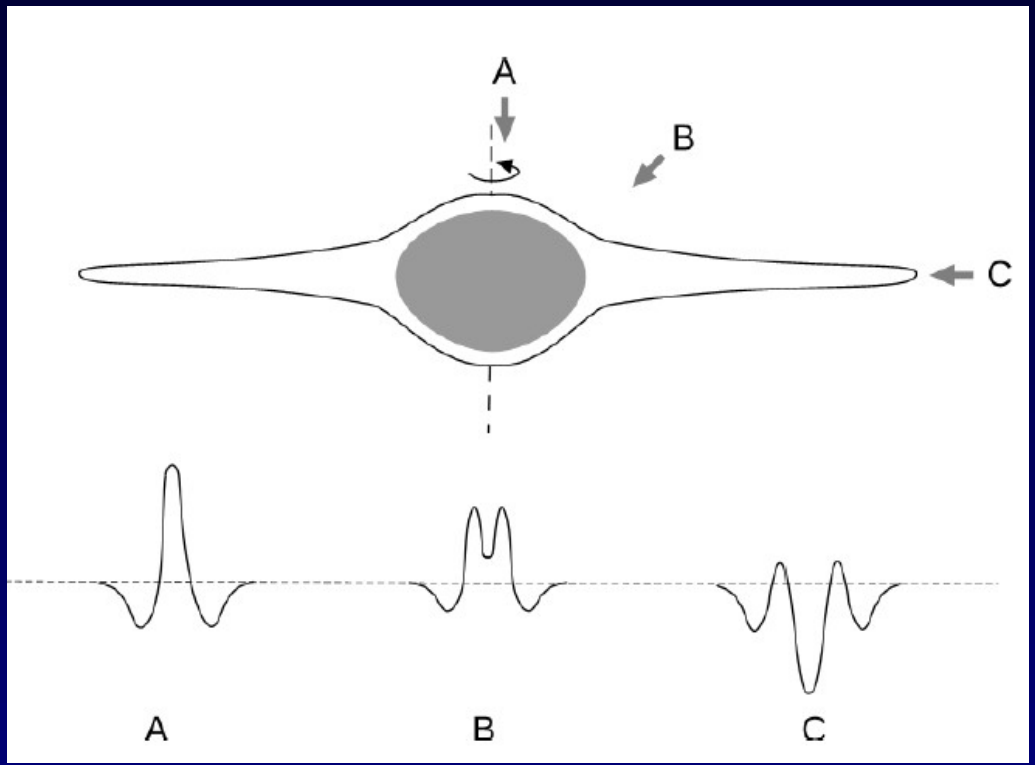
Be Stars - Introduction

Gamma Cas (Padre Angello Sechhi 1866)

Vatican obs predecessor – visual spectrograph

Some have or have had emission in Balmer lines

Slettebak 1988



Gamma Cas - 1st Be star

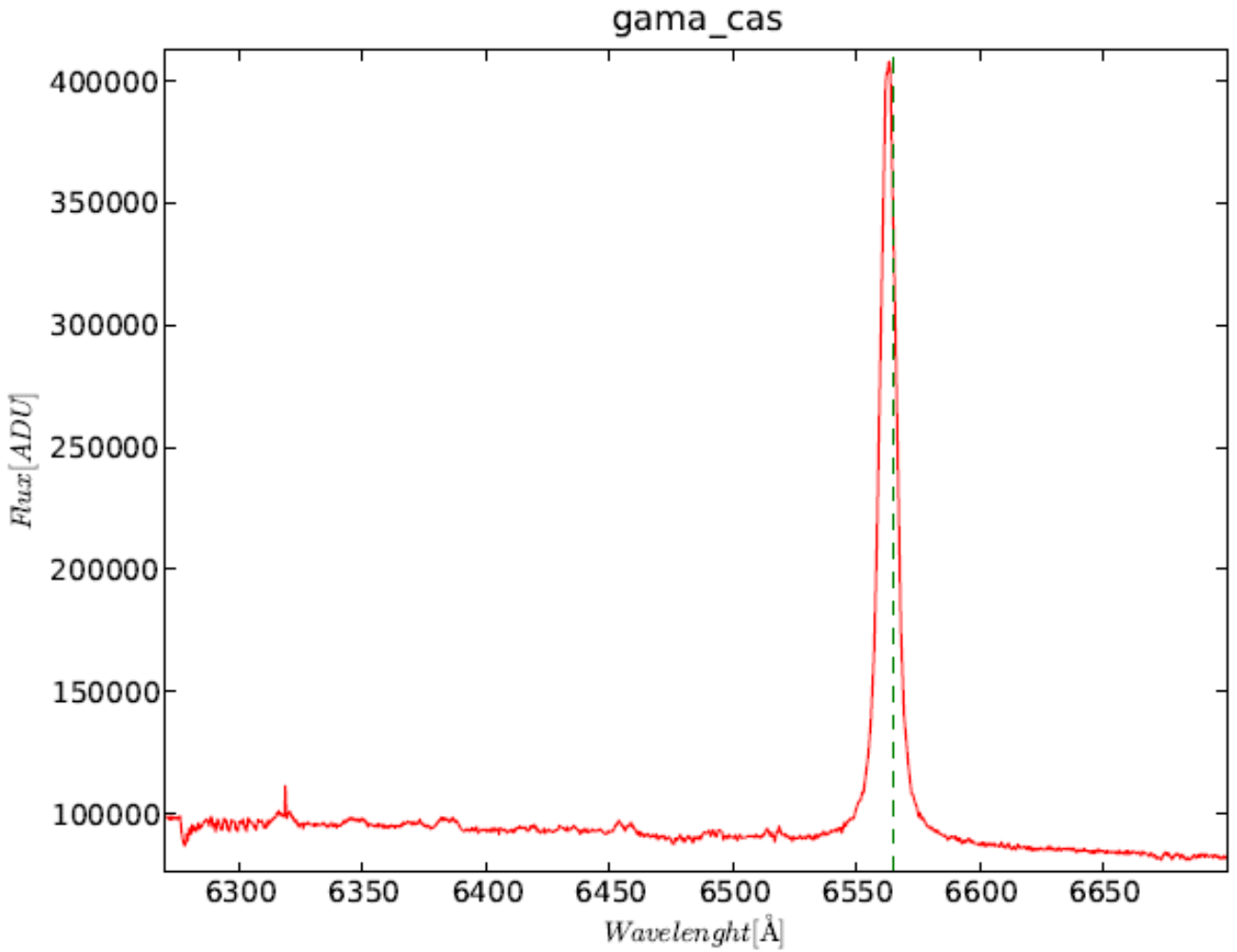
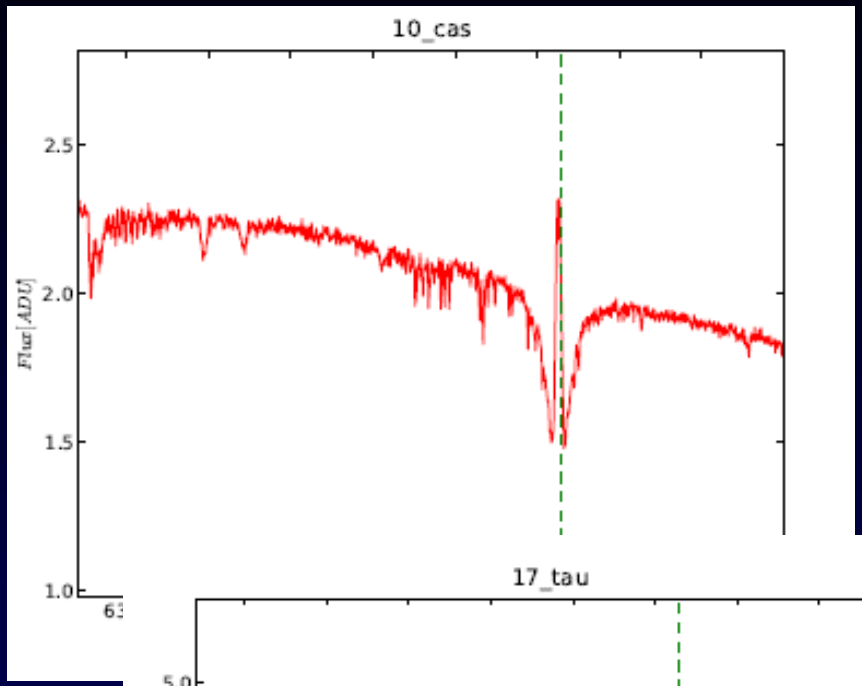
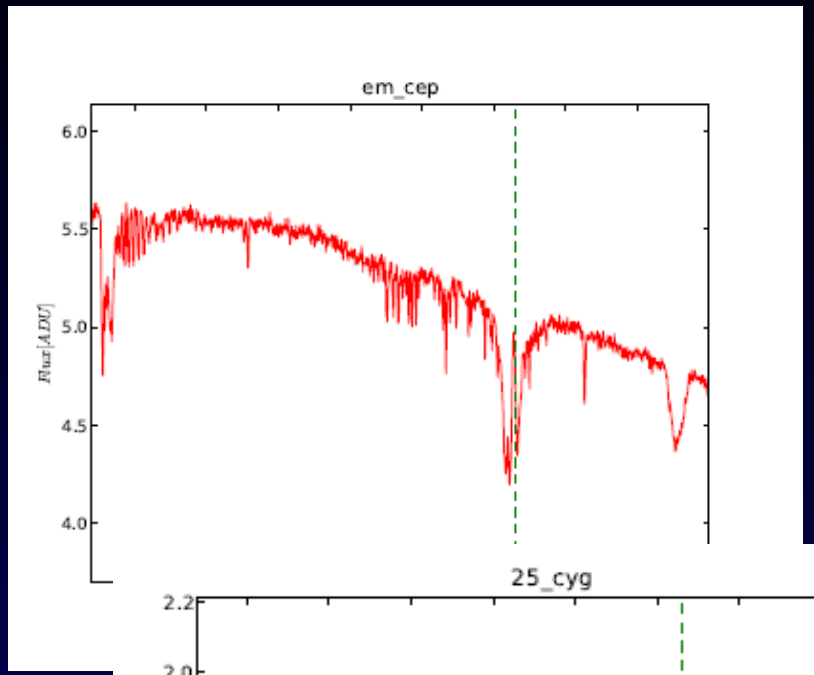
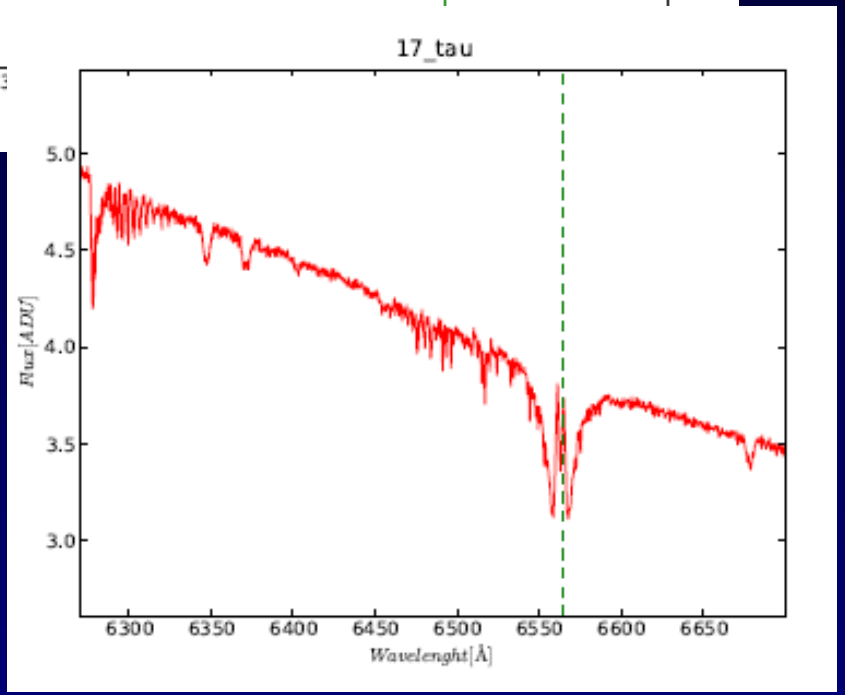


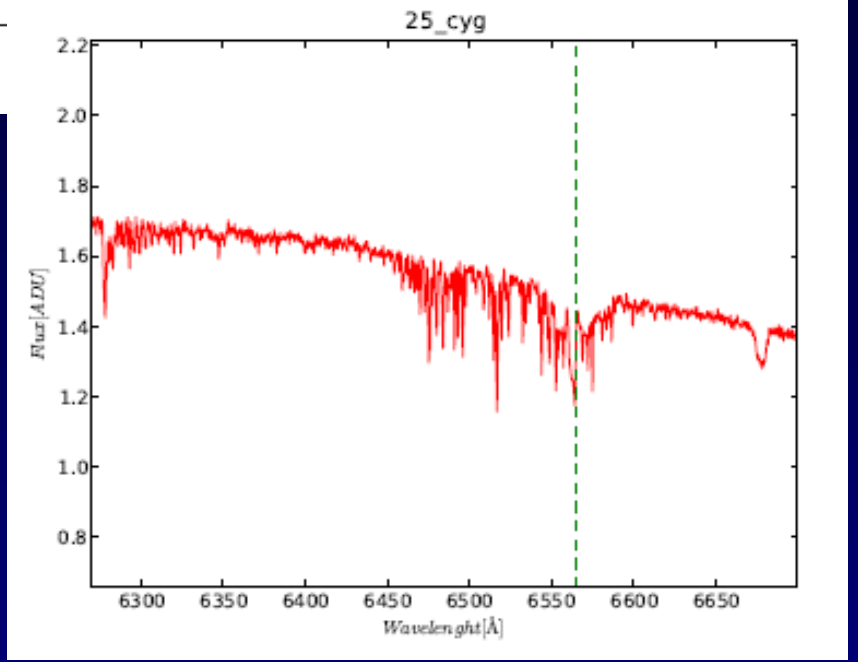
Figure 3.18: Spectrum of Gamma Cas. Be Star

Be Stars : Emission in absorption

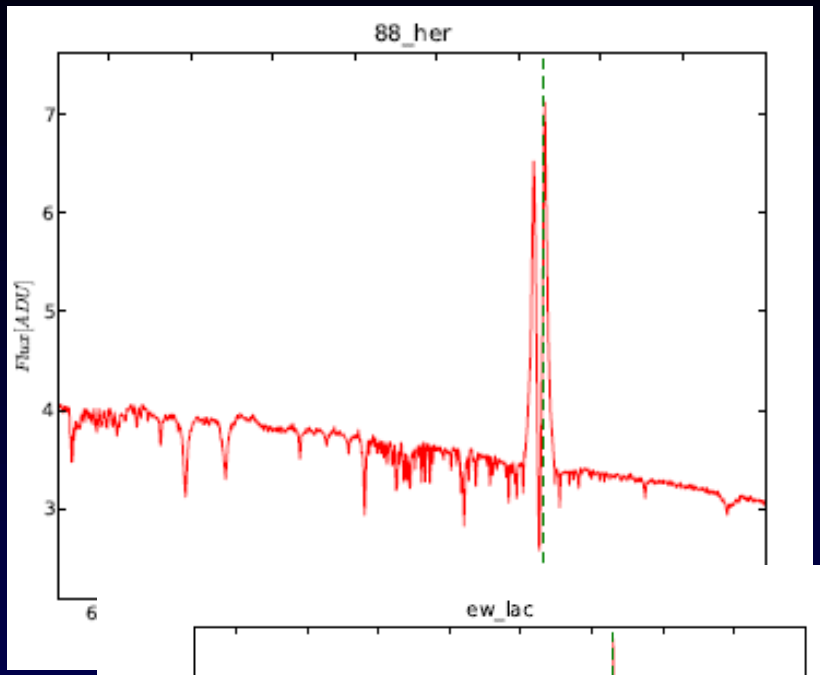


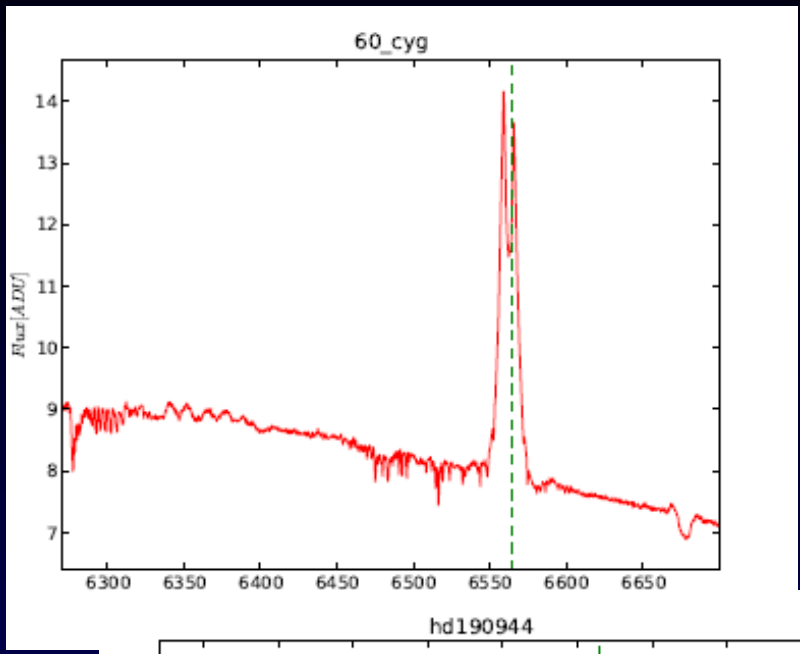


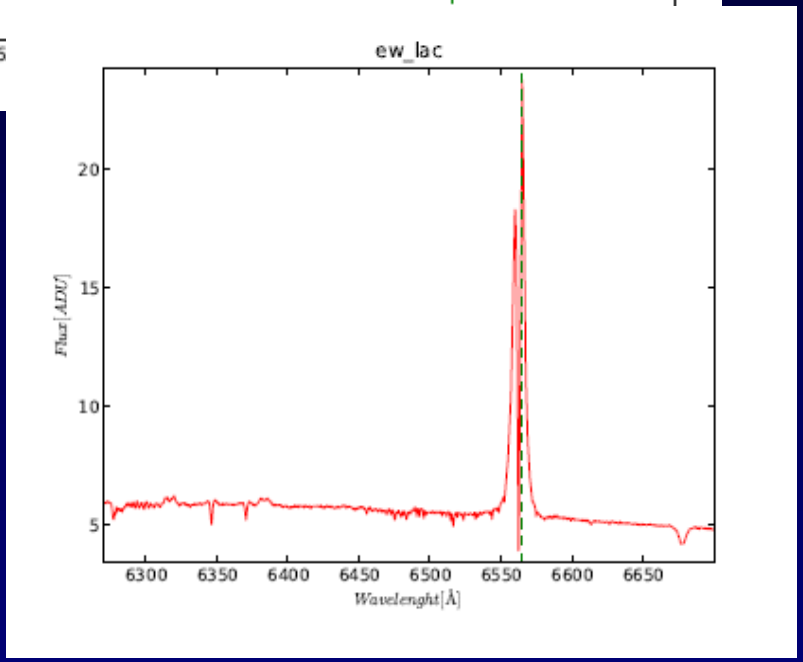


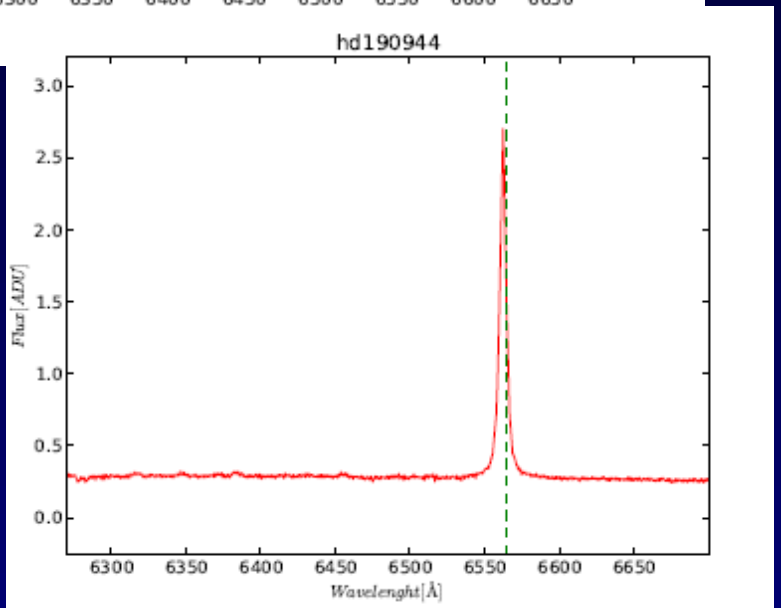


Be Stars : Shell lines vs. no absorption









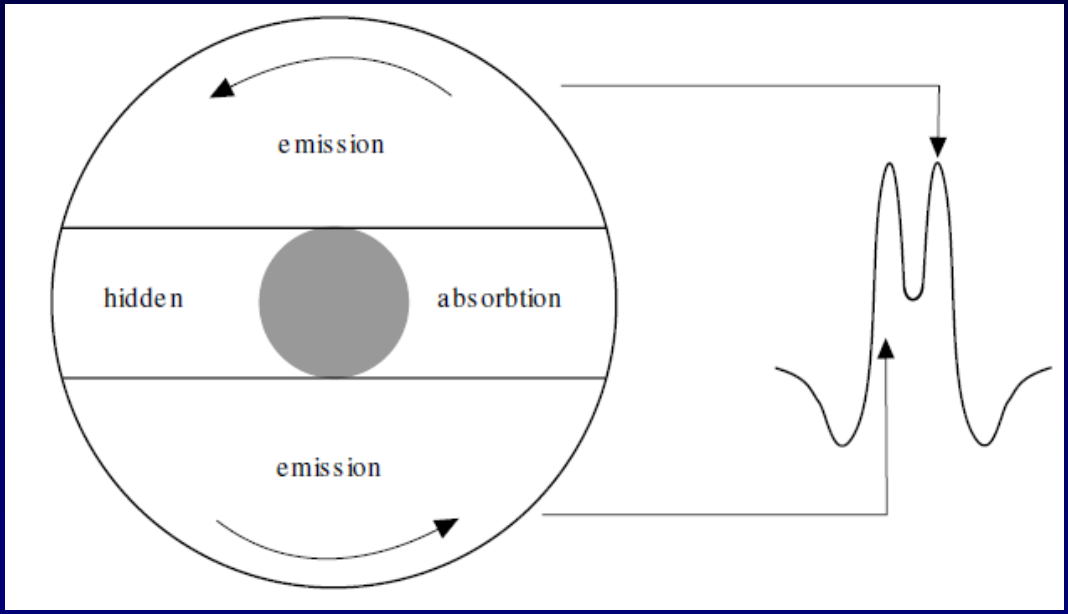
Be Stars

Circumstellar disk – decretion/accretion ???

Density waves (one-arm – ratio of peeks)

Some binary (multiple) stars

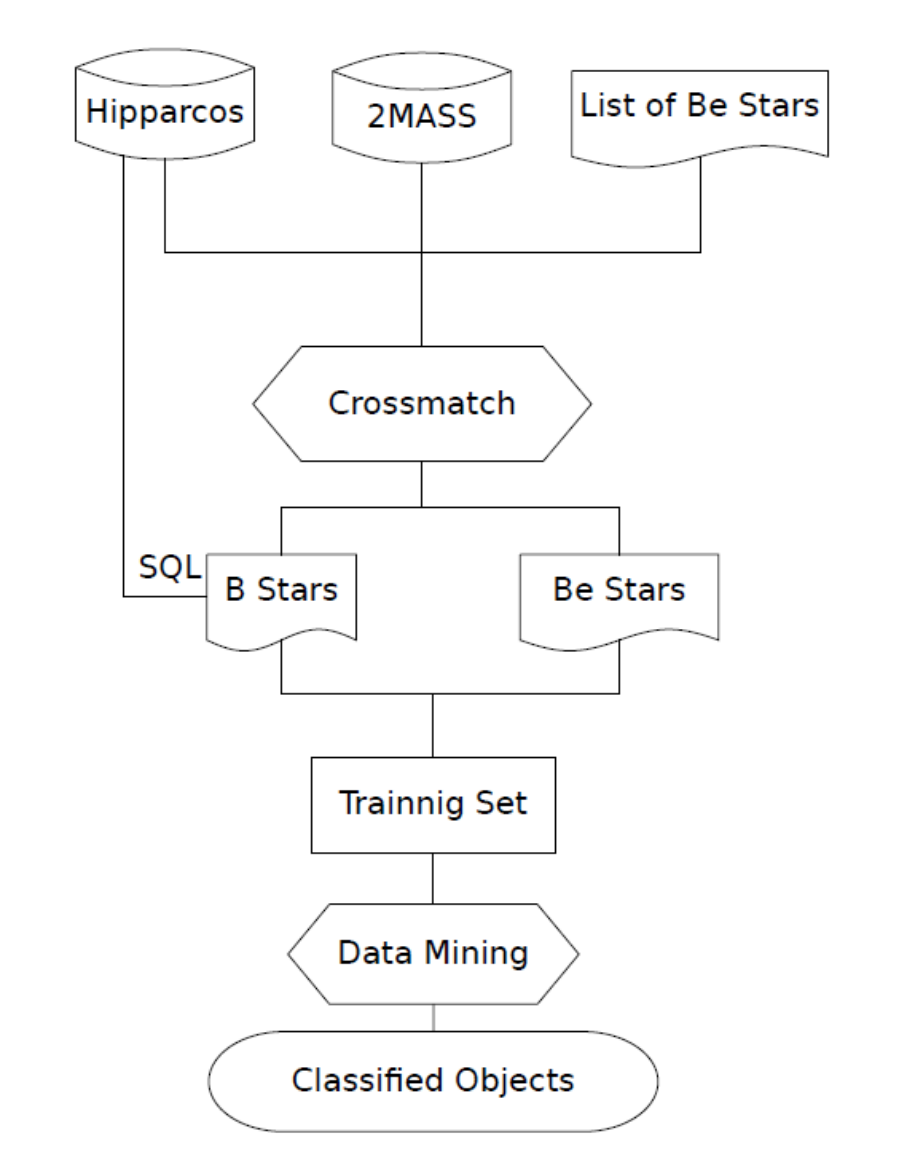
Pulsation components, DACs ...



Hirata &Kogure 1984

New models Siljaj 2010...

Photometry Data Mining



List of 645 Be stars HIP#

Selected from literature

But some have still indirect evidence (SED, IR ex)

Rest from HIP – Sptype B

excluded the above – (overlap in classification)

ConeSearch in 2MASS

to get JHK -> color indices

J48 Weka (SVM simillar)

Be Stars Data Availability

BAD (especially spectra)

The known ones too bright for surveys

BeSS – mostly amateurs (but important!)

Still the Data Jealousy + non-VO archives

Problem with other colours (transformation)

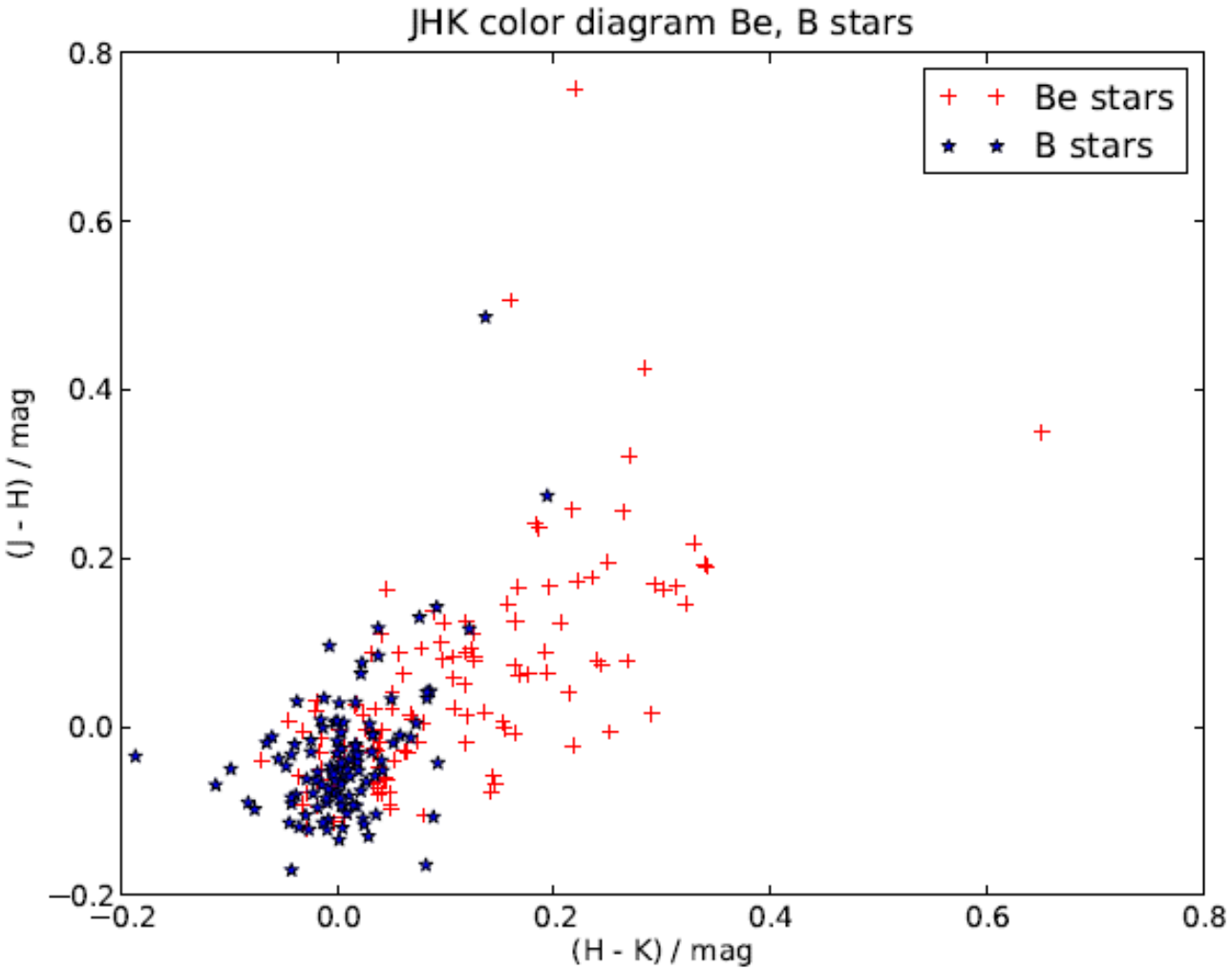
Different focus of surveys

IPHAS – failed (no bright stars)

Hipparcos + 2MASS

SEGUE in SDSS – galactic astronomy

Be vs other B stars (645+407)



B	Be	<-- classified as
304	126	B
157	465	Be

Beware of 20%

time evolution ?

Correctly Classified Instances	769	73.0989 %
Incorrectly Classified Instances	283	26.9011 %

Photometry Errors

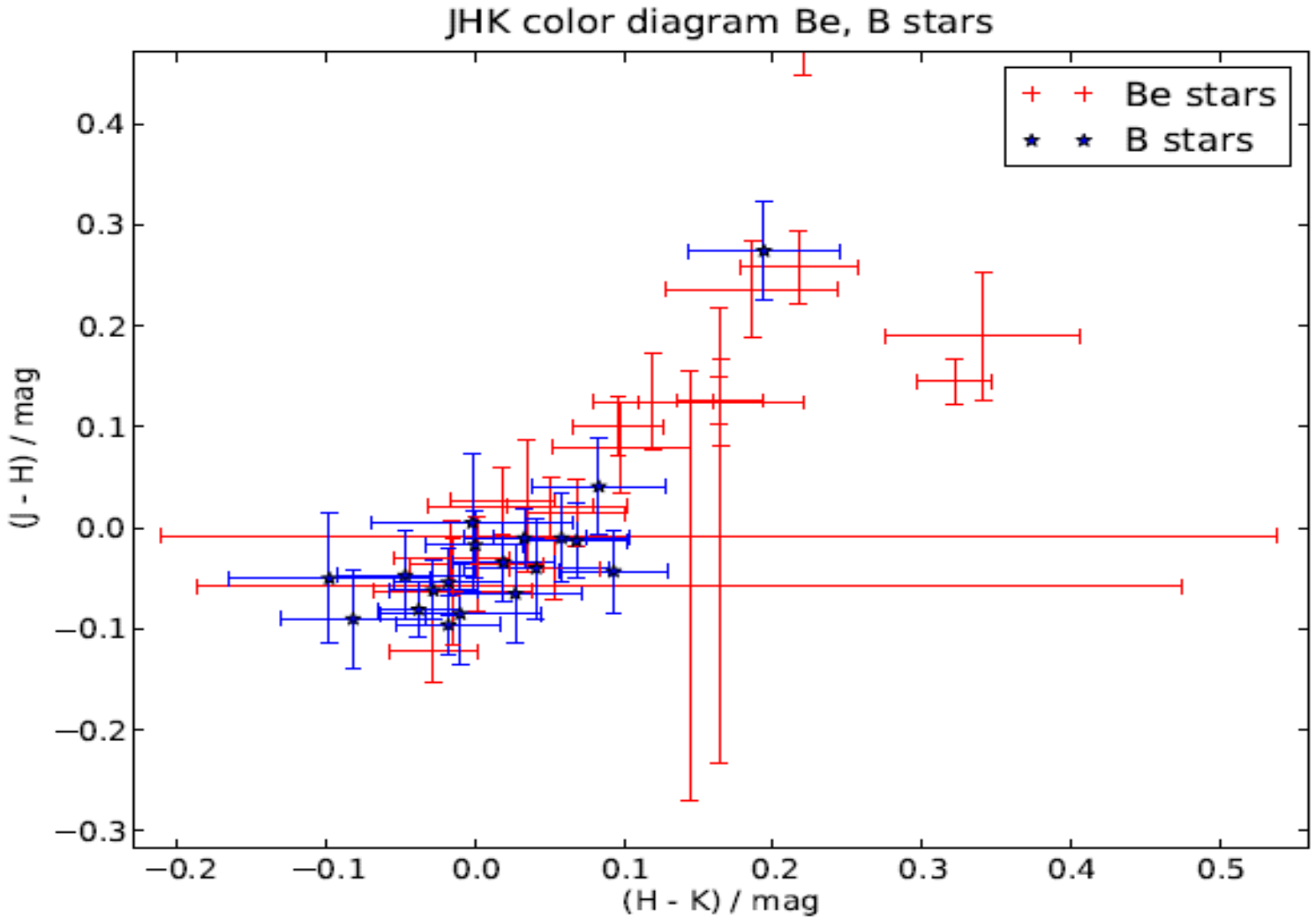
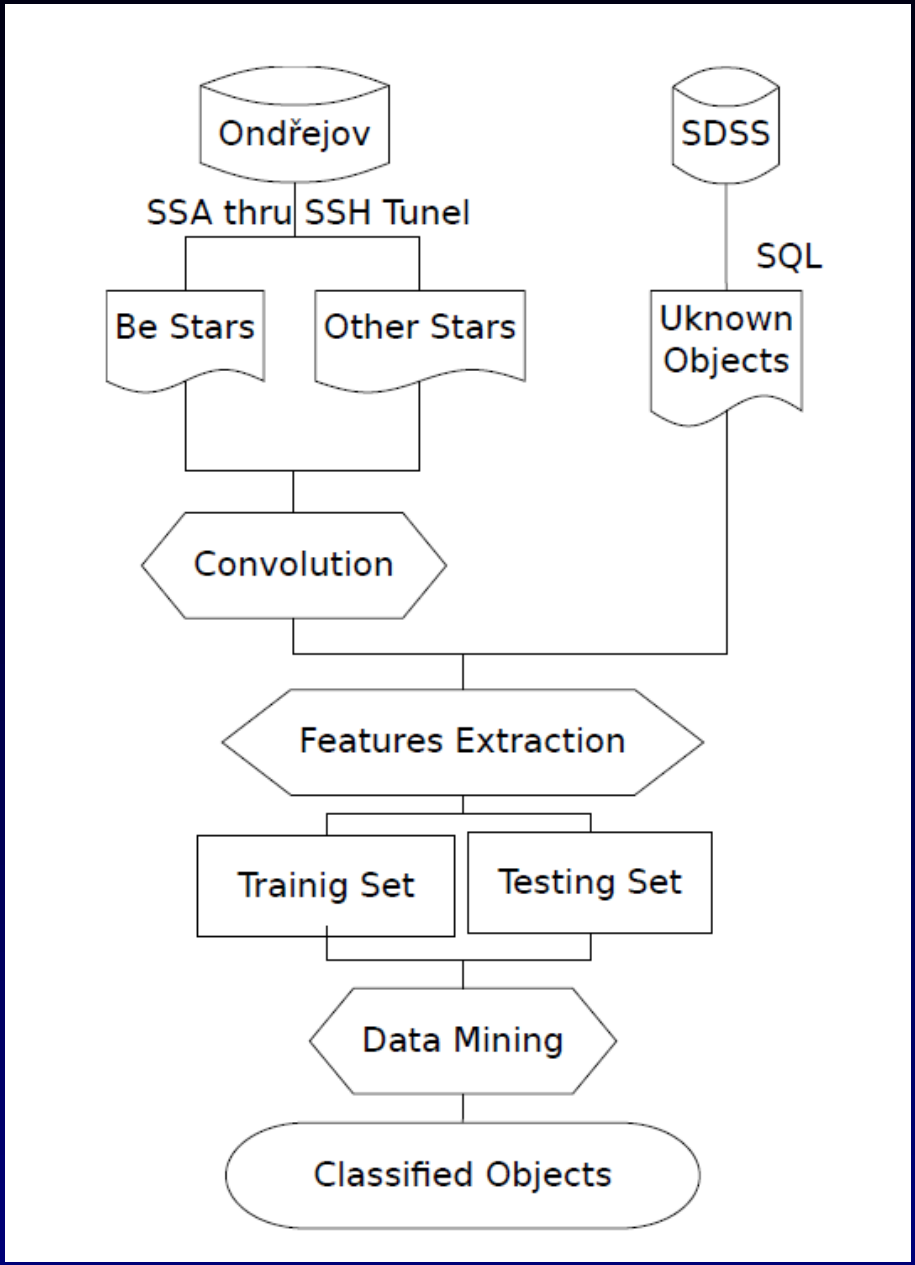


Figure 3.6: Color diagram of confirmed Be stars, B stars with errors

Spectral Data Mining



List of Be HIP# 645

OND spectra : 20 yr of digital spectra, same setup, ~1 representant

OND coverage ~400A

cutout of SDSS to OND

Only Halfa 50 A

but more ranges

(SSA BAND)

Continuum normalized

J48 in Weka (DT C45)

Example of OND CCD700 camera

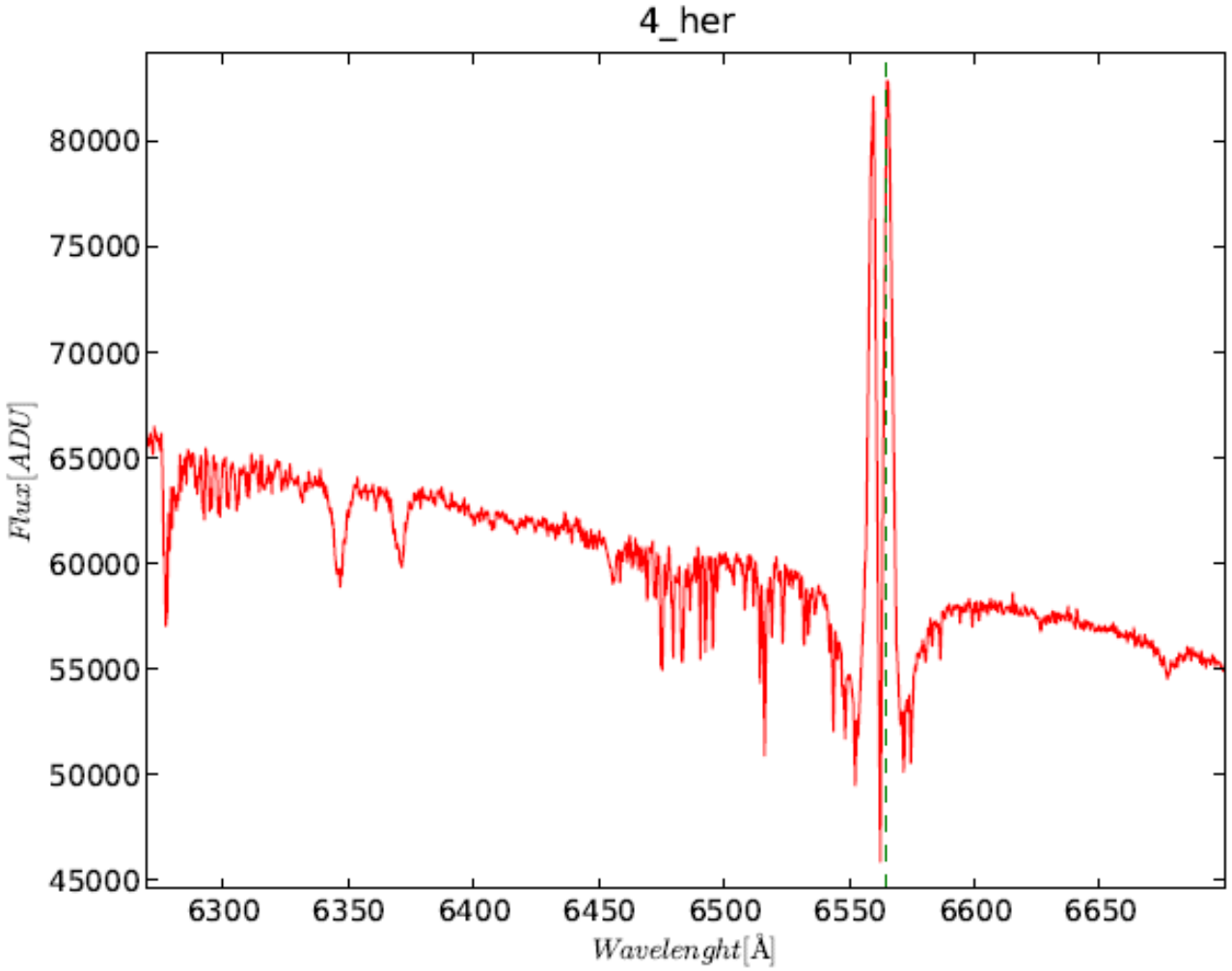


Figure 3.15: Spectrum of 4 Her. Be star

Resolution degradation + binning

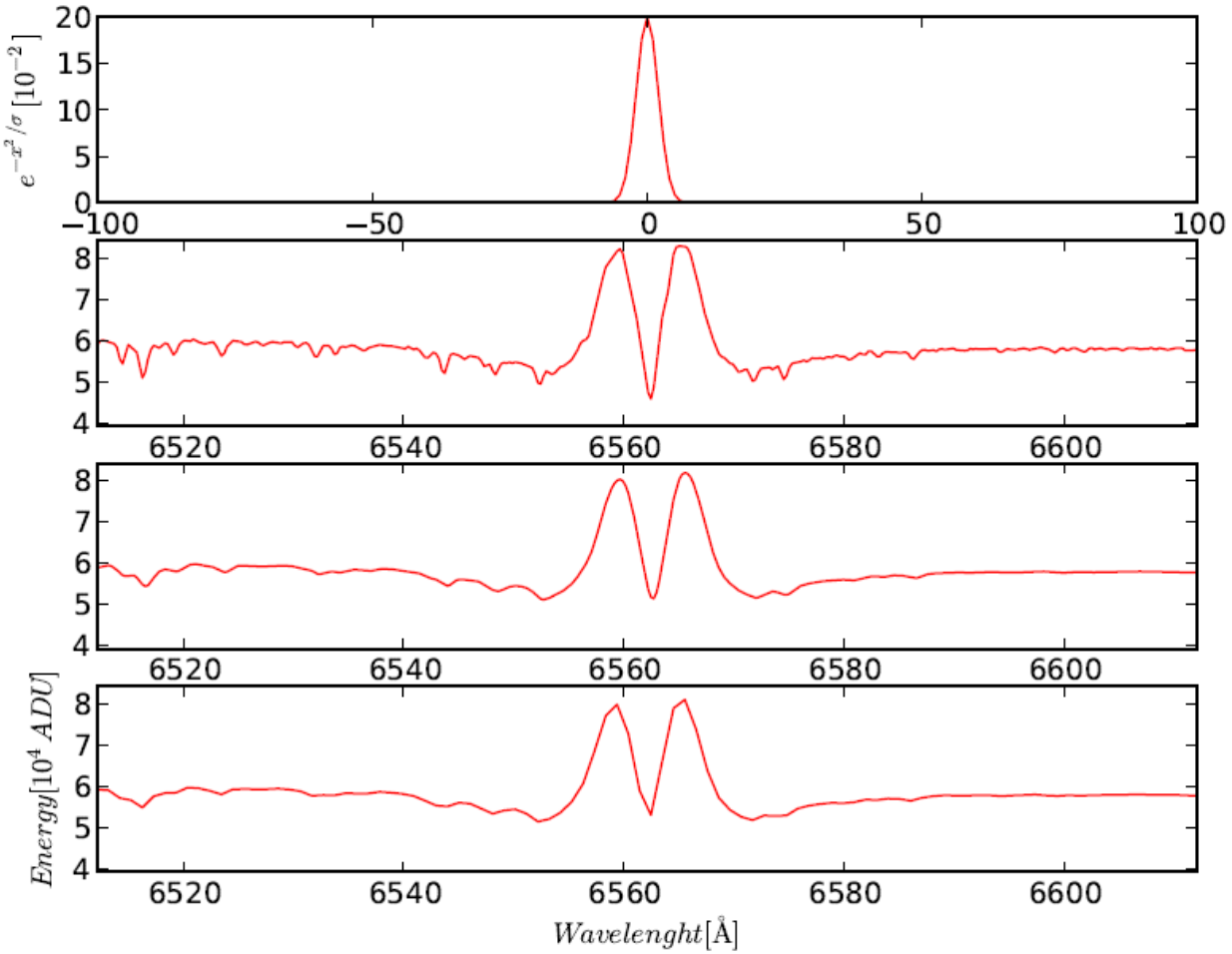


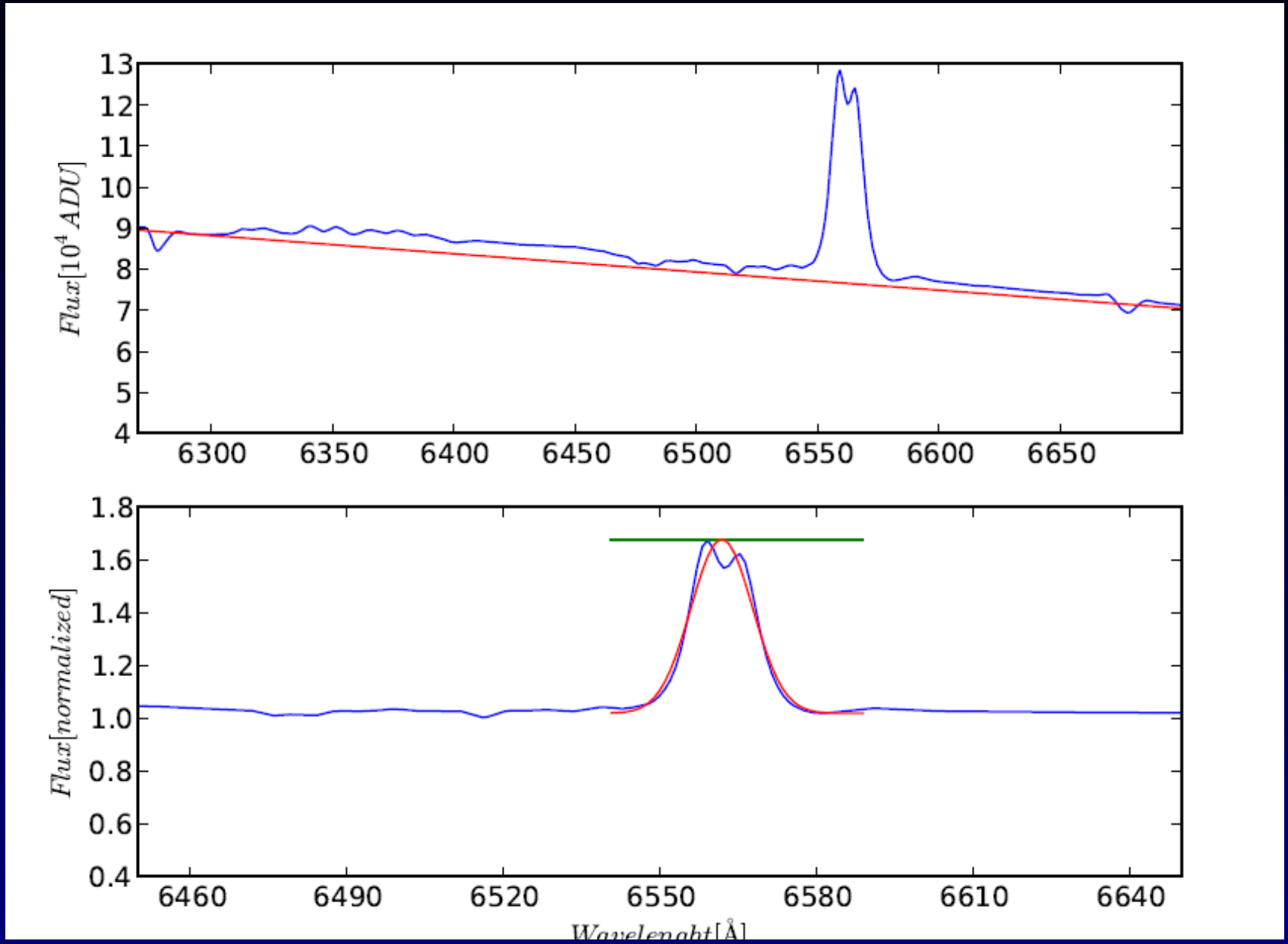
Figure 3.8: Reduction of spectral resolution of Ondřejov spectra of the Be star 4 Her. The top figure shows Gaussian function used for convolution with the spectrum, followed by the original spectrum then there is a spectrum after convolution with the Gaussian profile. The last is the final spectrum after binning

Diffent spectral resolution

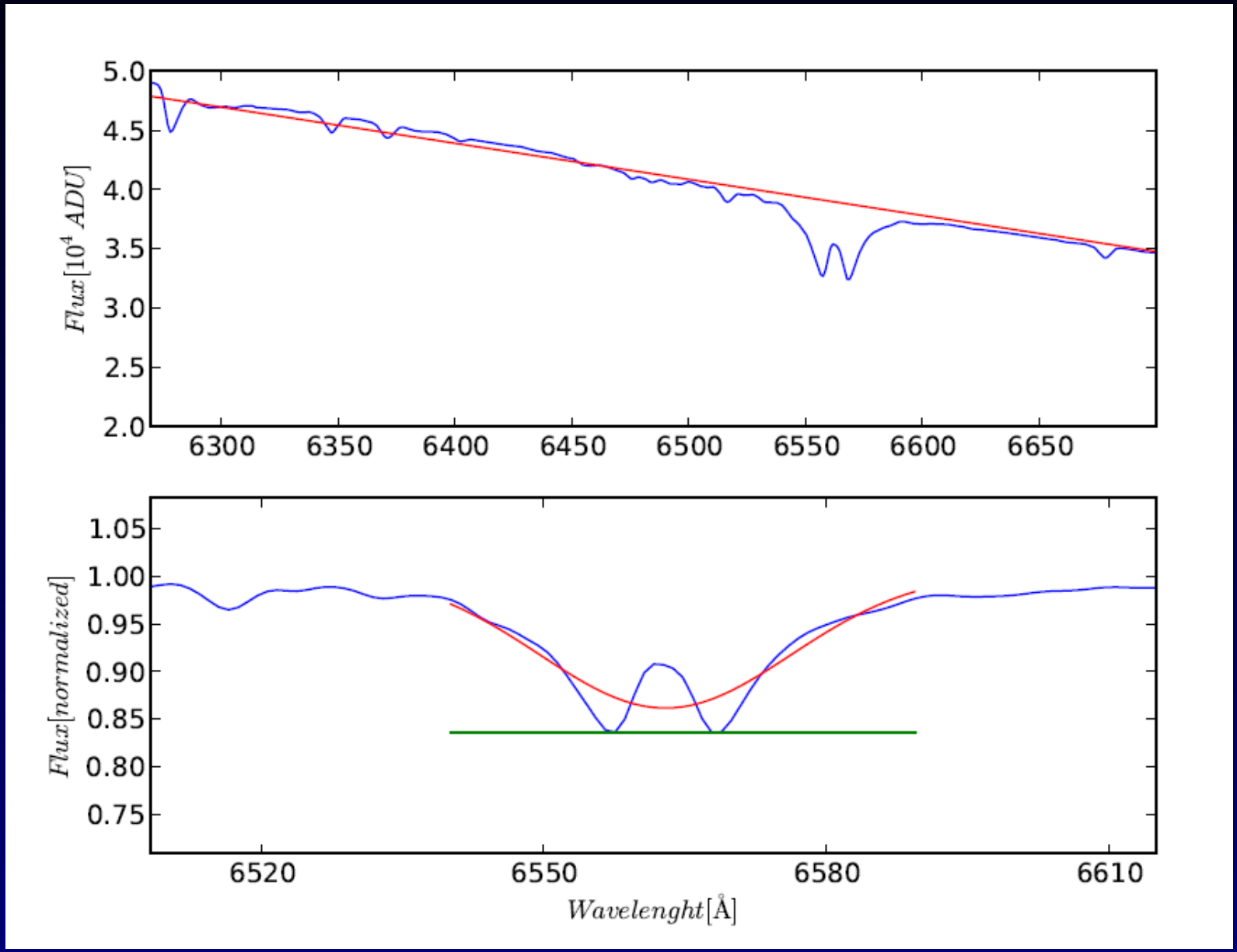
convolution of instrumental profile

binning
~4 pix in 1

Feature extraction (emission)



Feature extraction (>>abs+em)



Spectral Data Mining Results

```
=== Summary ===
Correctly Classified Instances         145               83.815 %
Incorrectly Classified Instances        28               16.185 %
Kappa statistic                          0.6529
Mean absolute error                      0.1849
Root mean squared error                  0.3652
Relative absolute error                 39.8819 %
Root relative squared error             75.8919 %
Total Number of Instances              173
```

```
=== Confusion Matrix ===
 Be Others  <-- classified as
 95 15   | Be
 13 50   | Others
```

If intensity+width + robust SNR used

```
=== Confusion Matrix ===
 Be     0   <-- classified as
102     6 |   Be
 35    40 |   Others
```

If only intensity of emission used

Be Star Candidate #1 in SEGUE

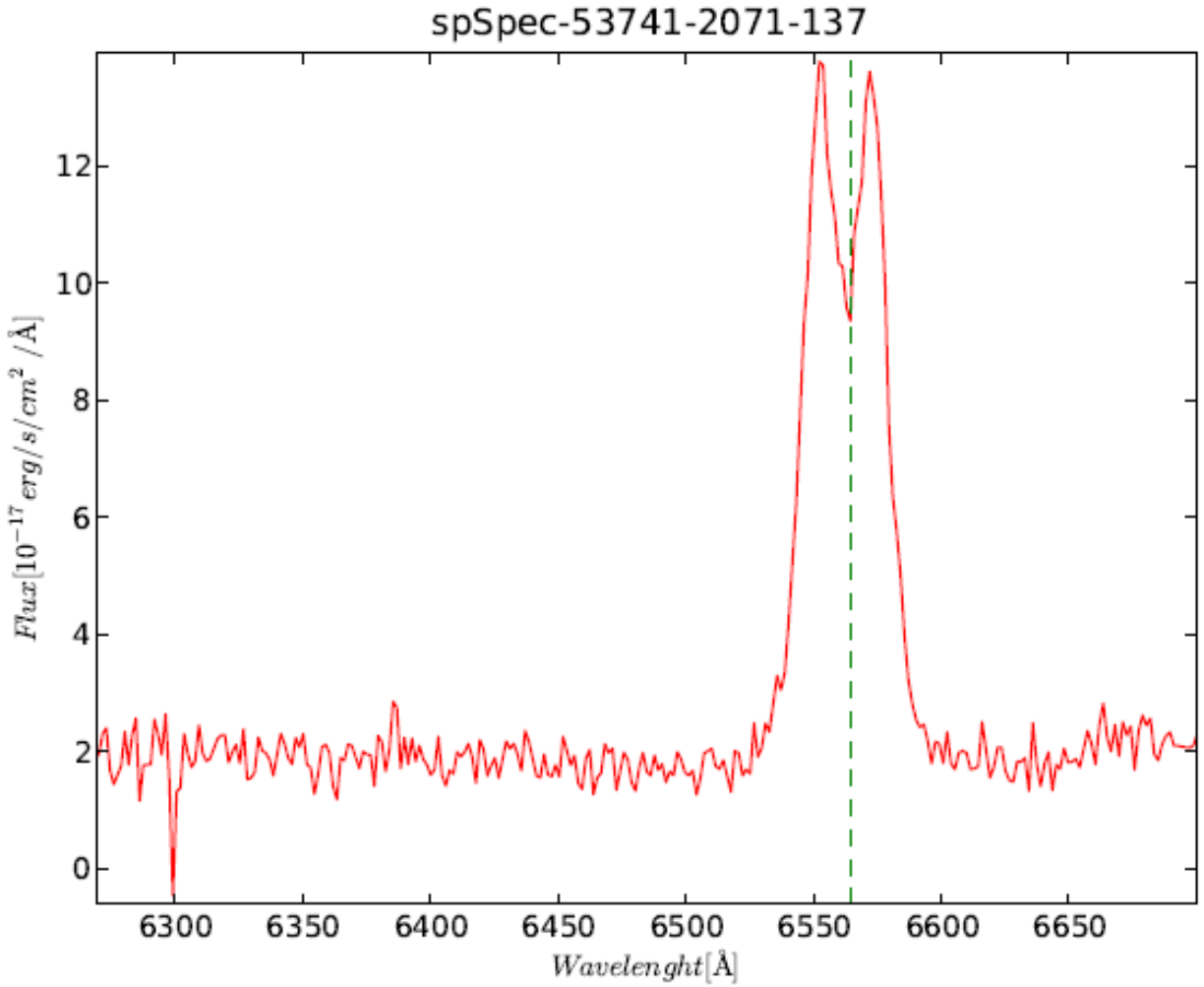


Figure 3.11: Example 1: SDSS J035747.16-063850.7

Be Star Candidate #2 in SEGUE

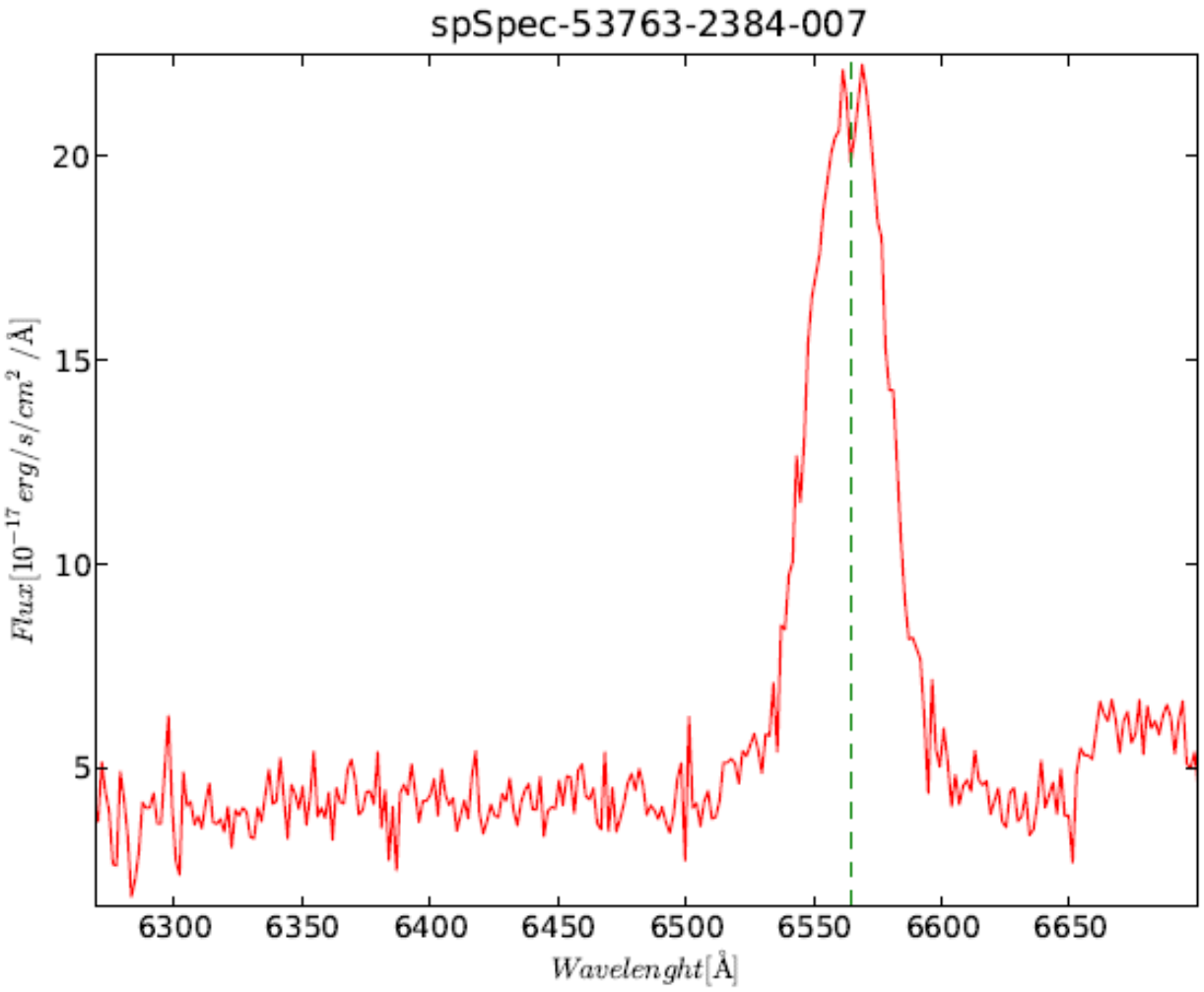


Figure 3.12: Example 2: SDSS J094325.89+520128.6

Be Star Candidate #3 in SEGUE

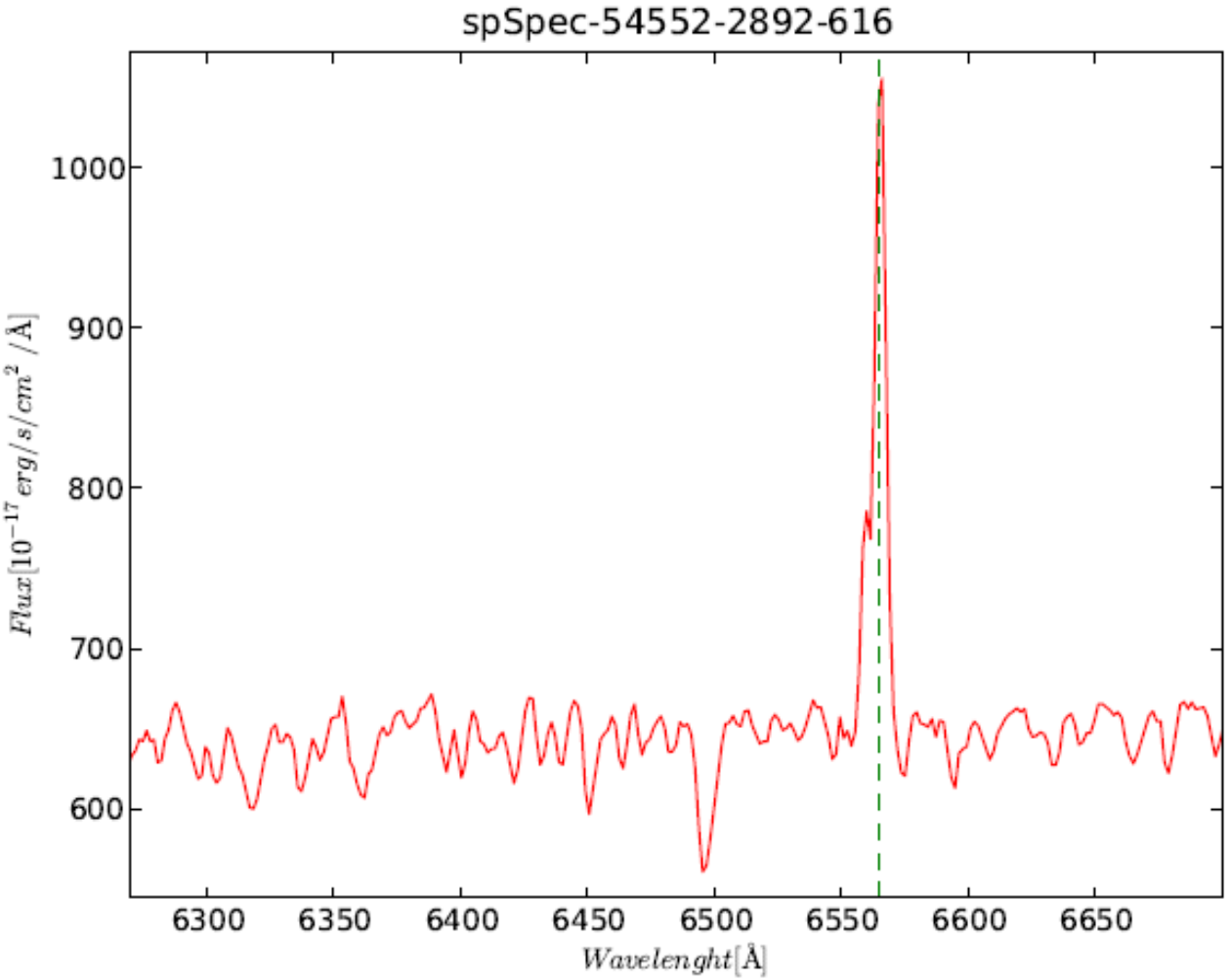
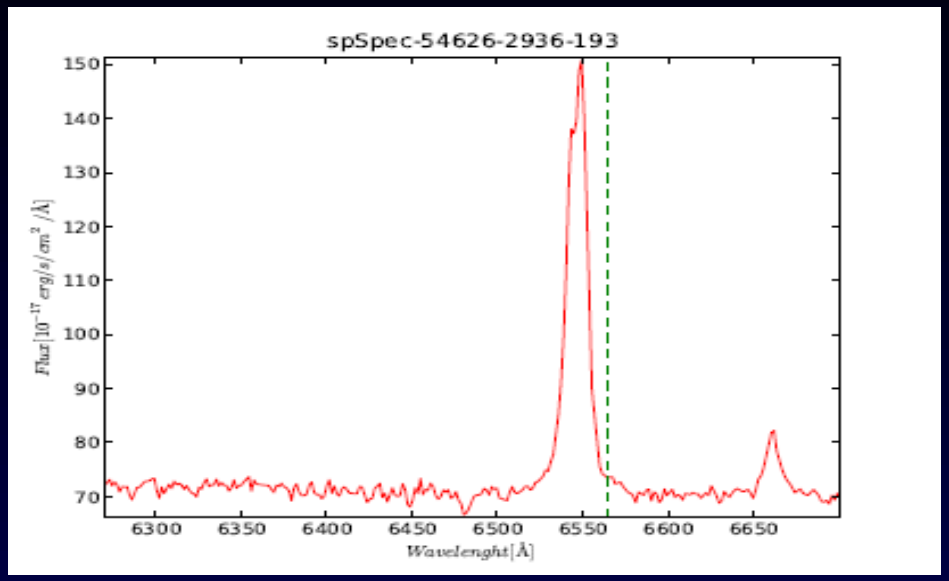
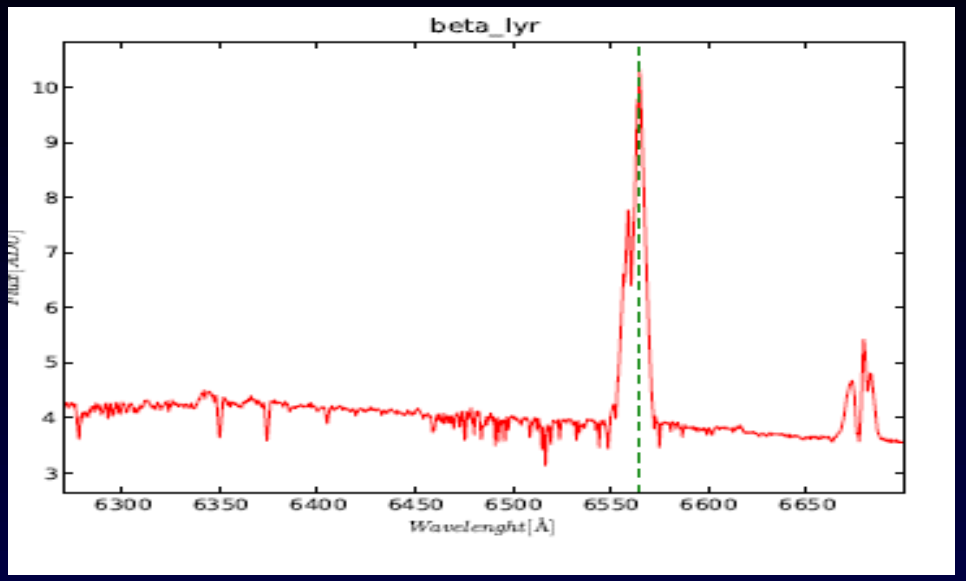
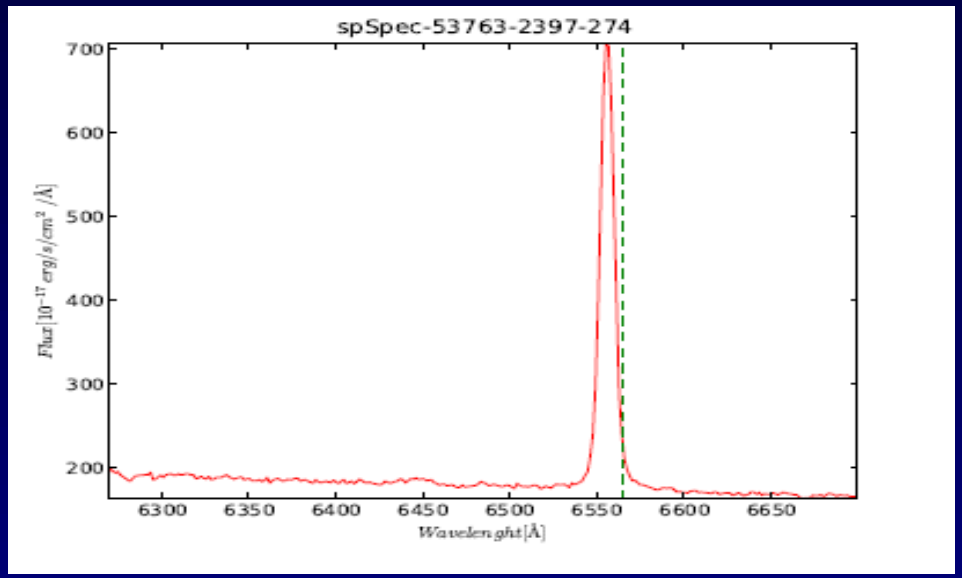


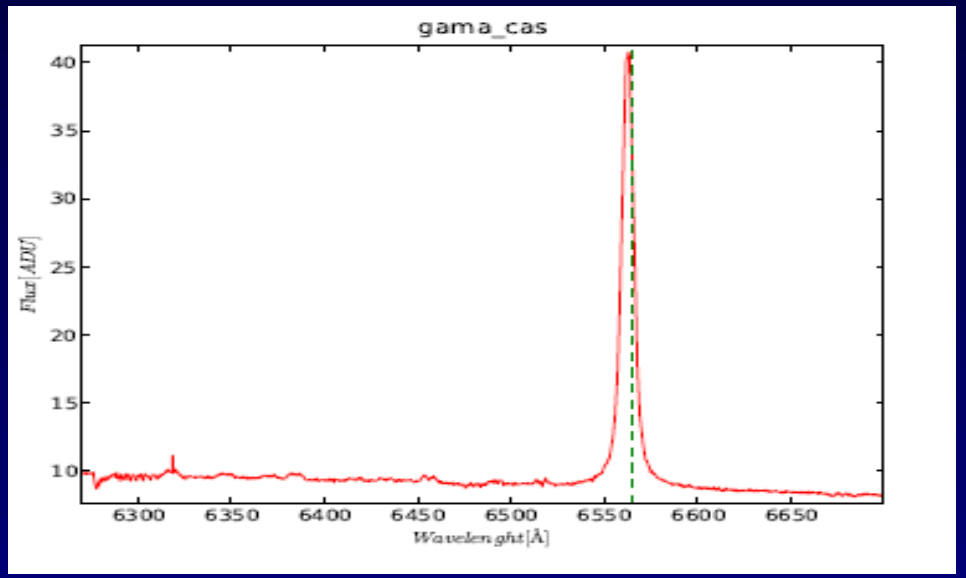
Figure 3.13: Example 3: SDSS J120729.12+003659.8

Candidates – Similarity !









Conclusions

Photometry not very convincing (more colours)

Verified candidates not in massive surveys !!!

VO discovery role

Spectral DM – support in VO standards needed

From 178314 SEGUE spectra 1110 identified as Be stars candidates (massive download of TB)

But lot of objects with emission,

Some probably novae or post-AGB phase

(e.g. V838 Mon – even more important !)

BaII lines at 6500 – see candidate #3

Conclusions – Future Plans

TIME evolution – LC + spectrum together

Specific shapes - Wavelet coefficients - features

- (quite new – feature extractions, licence plate recognition, pedestrian identification, human emotions, mental diseases in EEG, EKG, DNA features, not so far in astronomy)

The support in VO needed (SSA)

- Spectrum range cutout (BAND as service)
- Continuum normalization (FLUXCALIB=normalized)
- Resolution degradation (SRP service)
- The Obscore use case 2.3 – (em. width of Ha > 2000)
- VO mostly extragalactic support (stars = side effect)

Conclusions – Future Plans

The support in VO needed (Photometry)

- Work with time series as with spectra (SVO way)
- Period – metadata + folding (Period04, FROG)
- Plug-in algorithms – O-C trends, chaotic
- SED + Photometry DM + Filters (who provides curve)

The Uptake Committe influence ??

in stellar astronomy very low

- VO is not serious science, its extremely complicated way of storing data
- hoping someone will use them one day