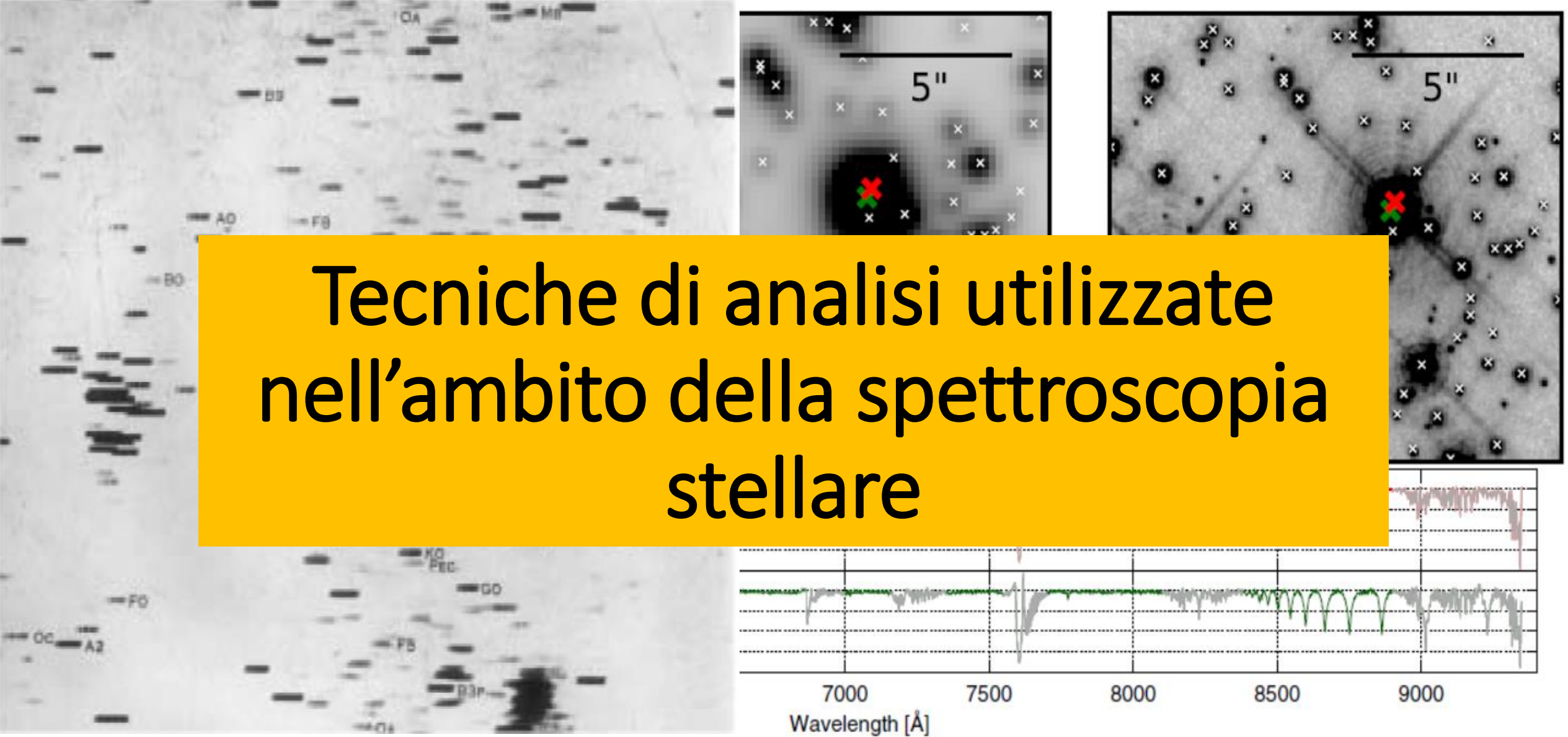


# Tecniche di analisi utilizzate nell'ambito della spettroscopia stellare



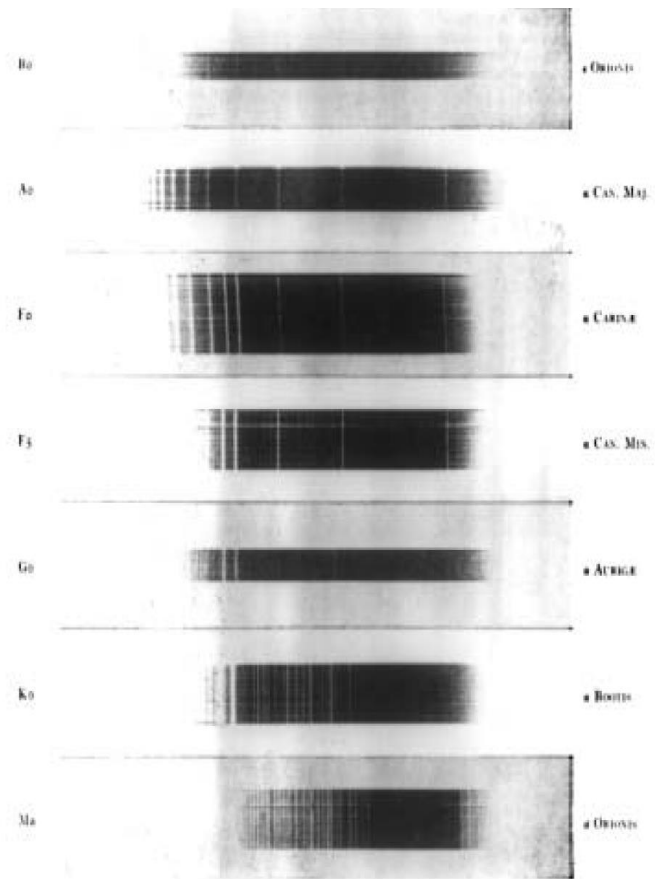
Raffaele Gratton, Sara Lucatello, Valentina D'Orazi

INAF – Osservatorio Astronomico di Padova

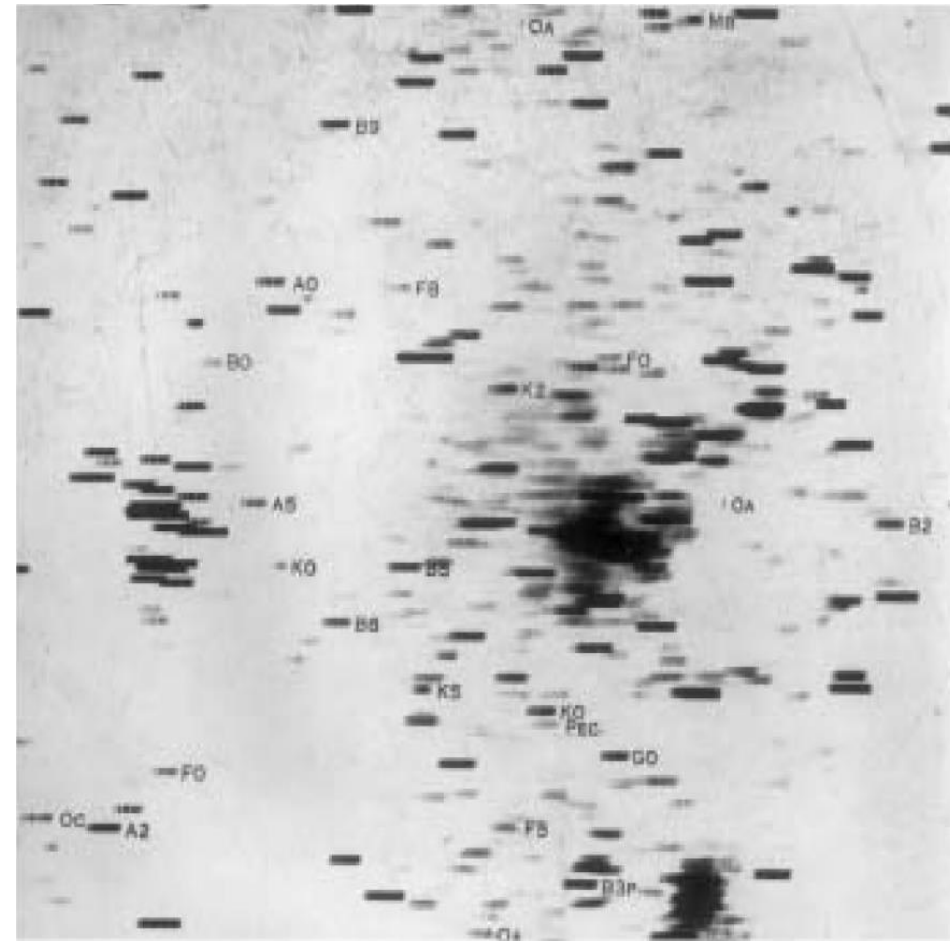
# L'astronomia moderna (?) e' basata su grandi quantita' di dati → Survey spettroscopiche stellari

Survey	Year	N. Objects (x10 <sup>6</sup> )	Resolution	Lim mag
RAVE	2004-2013	0.48	7,500	9<I<13
ESO – GAIA	2012-	0.1	20,000	R<19
SDSS-Apogee-2	2014-2020	0.26	22,500	J<16
Hermes-GALAH	2013-	1	28,000	V<14
LAMOST-Legue	2012-	7.5	1,800	R<19
GAIA DR3	2021	5	11,500	G<12
MOONS	2021-2026	>2	5,000 20,000	I<19.5 I<17.5
Weave	2020-2024	~6	5,000 20,000	V<21 V<16
4-MOST	2022-2027	20 3	6,000 20,000	G<20.5 G<18

# HENRY DRAPER CATALOGUE



TYPICAL STELLAR SPECTRA



- Pubblicato tra il 1918 e il 1924: 225300 stelle (half of RAVE and twice ESO-GAIA)
- Estensione nel 1949: 359083 stelle
- Posizioni e tipi spettrali

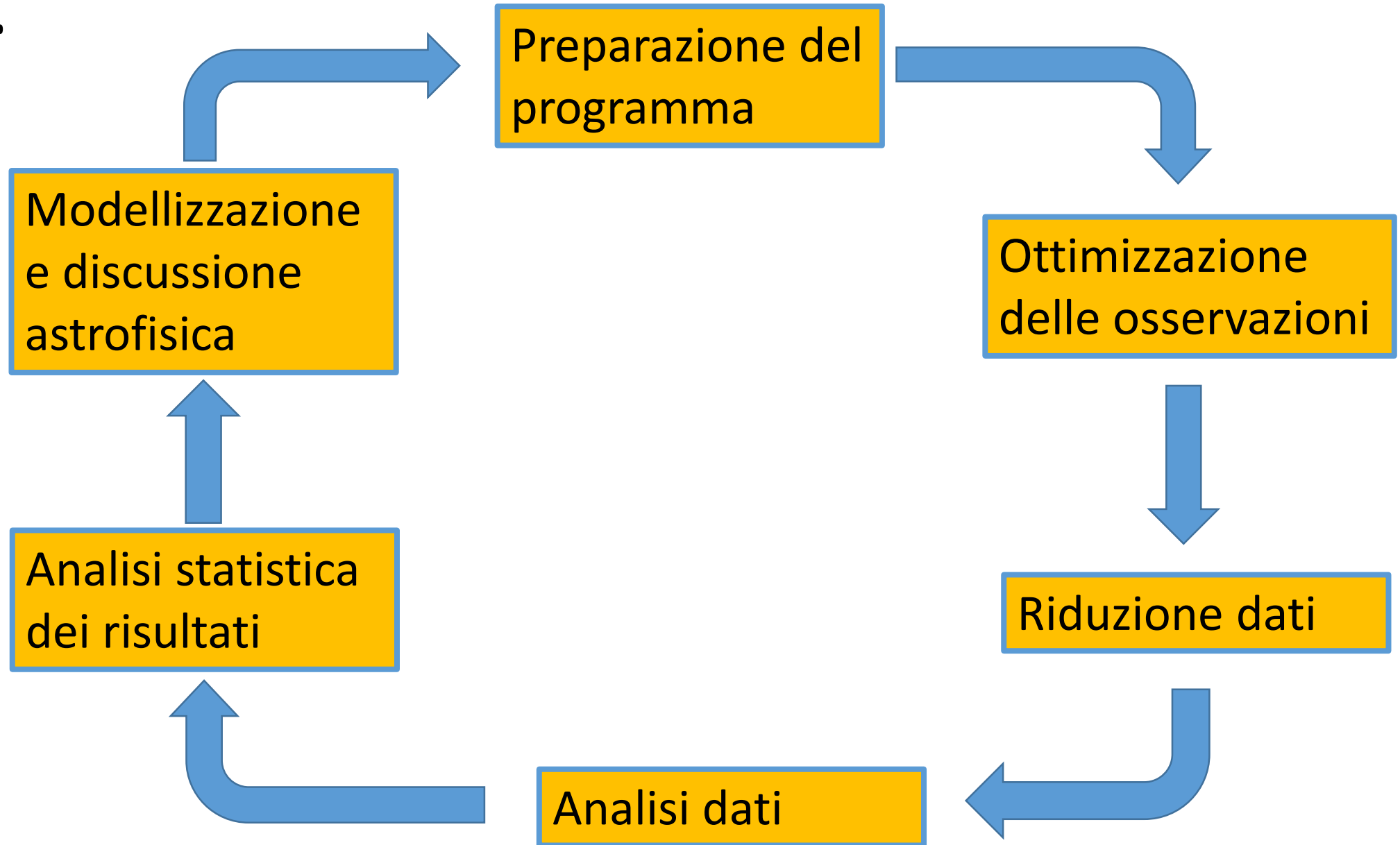
Annie Cannon

# How Miss Cannon was able to realize the HD catalogue?

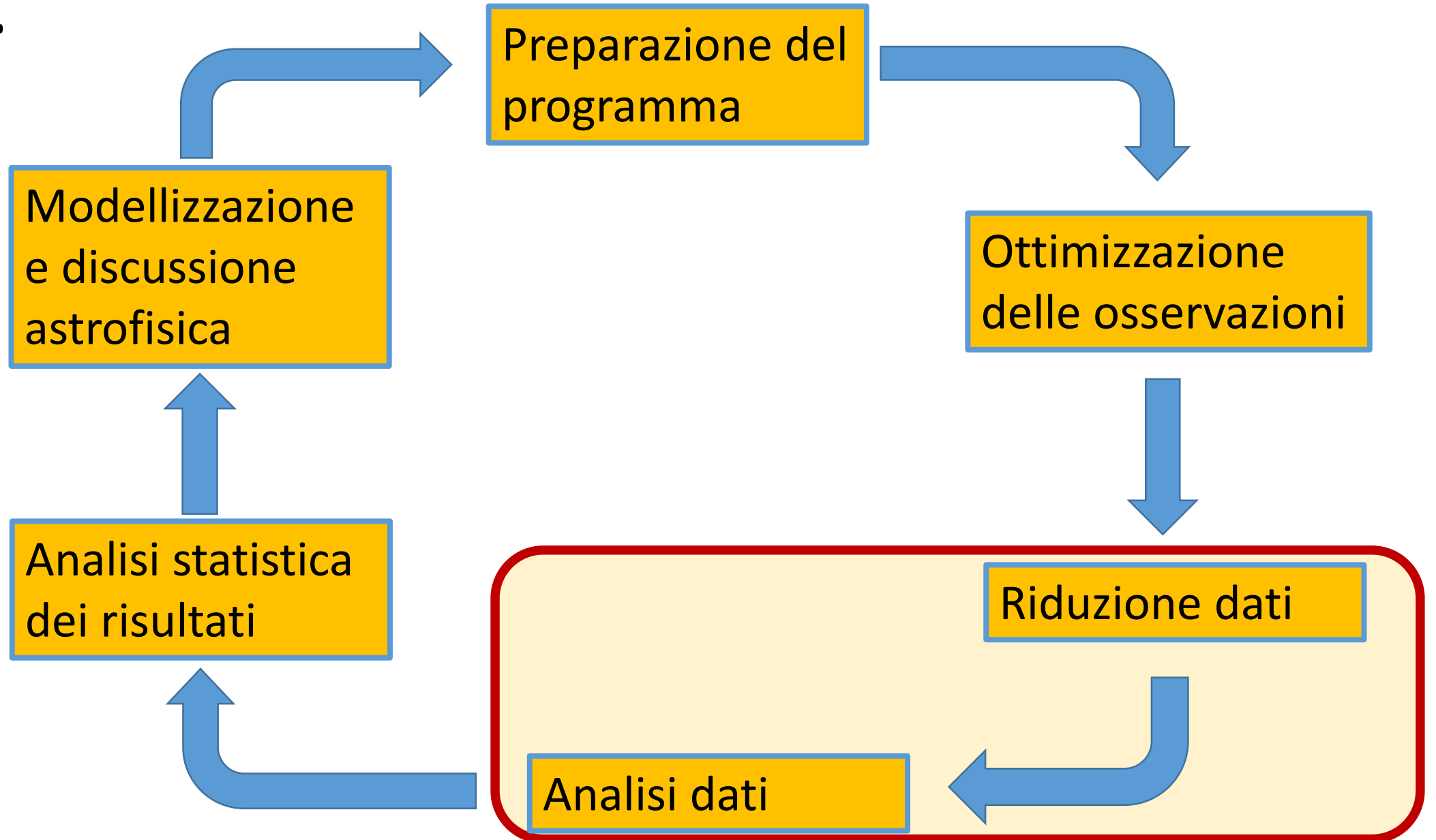
Cecilia Payne-Gaposchkin quoted Henry Norris Russell as saying that someone should ask Miss Cannon exactly how she classified the spectra. The descriptions she repeated in each of the HD volumes were derived from careful examination of large-dispersion spectra. Few of those details could be recognized on the very small-dispersion spectra of the much fainter stars that make up the bulk of the HD Catalogue. Payne-Gaposchkin's reply to Russell was that **Cannon would not be able to tell exactly how she classified: "She was like a person with a phenomenal memory for faces. . . . She had amazing visual recall, but it was not based on reasoning . . .she simply recognized them."** Cannon no longer compared individual spectra with standards; she just looked at the spectra with a magnifying glass and **classified them at the astonishing rate of about 300 stars an hour, or one every 12 seconds!** She first numbered the stars to be classified on the plates. Then, while she examined their spectra, she dictated to an assistant the number and spectral class to be entered into a notebook. For the fainter stars, for which no lines were discernable, she relied on the distribution of light in the continuum

(from E. Dorrit Hoffleit, Phys. perspect. 4 (2002) 370–398)

# Strumenti software necessari ad ogni passo...



# Strumenti software necessari ad ogni passo...



# Riduzione dati

- Dipende dalla natura dei dati → di solito inserita nella pipeline di data reduction, ma non sempre in modo completo!
- Focus qui su due modi in particolare:
  - Strumenti multi-fibre
    - Es. FLAMES, SDSS, WEAVE, 4MOST
  - Strumenti a campo integrale
    - Es. SINFONI, MUSE, HARMONI
  - Esistono anche casi intermedi
    - es. KMOS



# MUSE spectrum extraction (Kamann+2013)

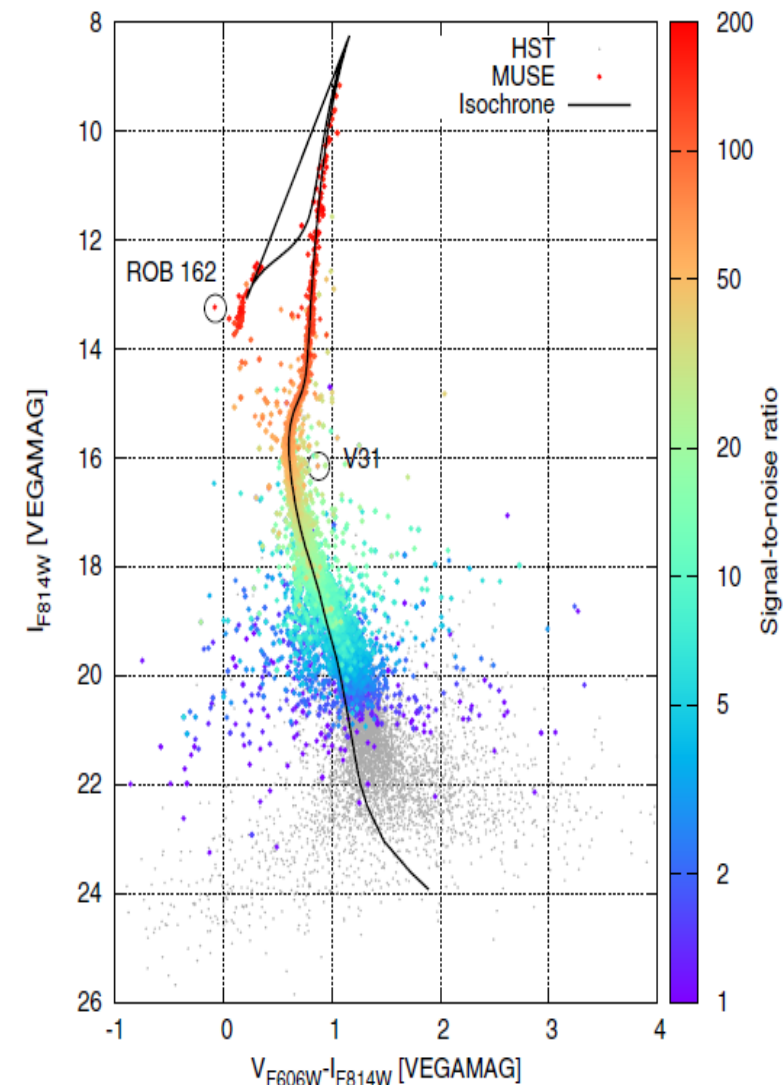
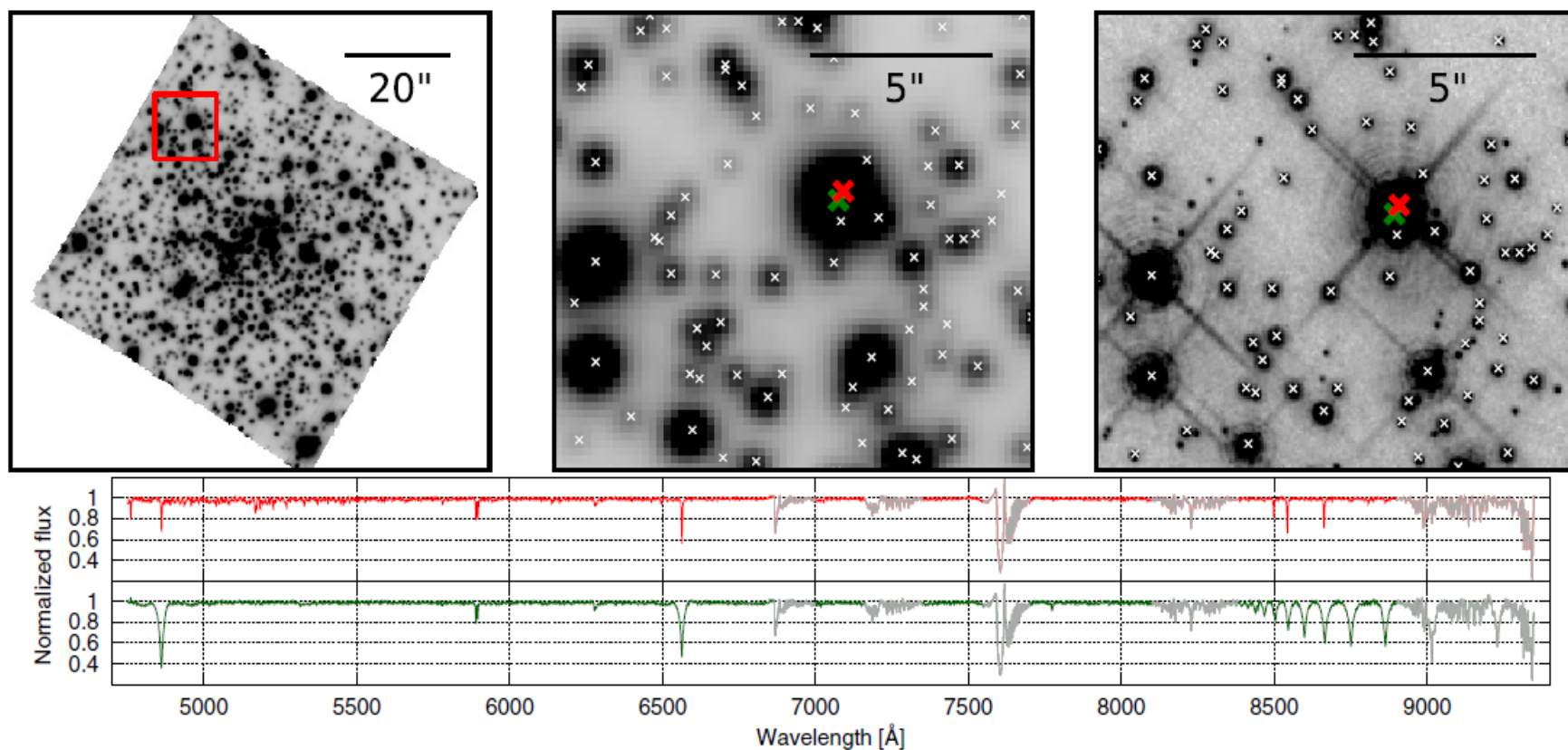
## Source deblending via PSF fitting

- Based on a photometric reference catalogue
- Initial guess of the PSF in the MUSE data
- The PSF model is used to create a mock MUSE image from the reference catalogue
- By cross-correlating this image against the MUSE cube, an initial guess for the coordinate transformation is obtained
- The sources for which meaningful spectra can be extracted are identified
- This set of sources is then used in the actual extraction process, which works iteratively on every layer, i.e. every wavelength slice, of the datacube
- In each layer, the analysis starts with a simultaneous fit to the fluxes of all sources, using the current estimate of the PSF and source coordinates.
- Afterwards, all sources except those identified as isolated enough to model the PSF are subtracted and the parameters of the PSF and coordinate transformation are refined
- The new estimates are then used in another simultaneous flux fit. This process is iterated until convergence



# Estrazione di spettri dai dati MUSE

T.-O. Husser et al.: MUSE crowded field 3D spectroscopy of over 12 000 stars in the globular cluster NGC 6397. I.



# MUSE + ottica adattiva



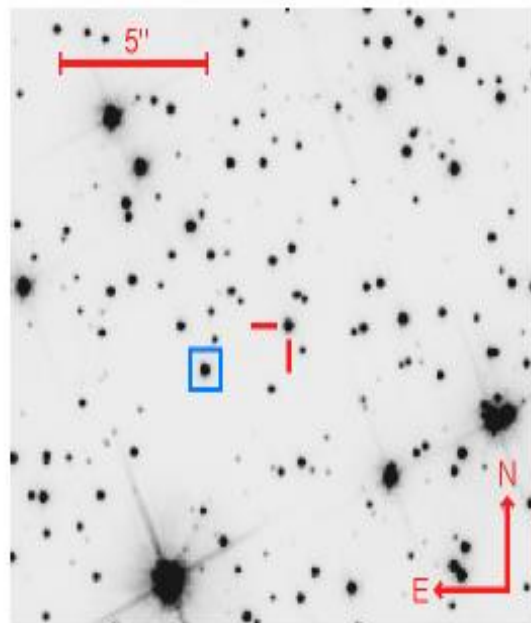
VLT+MUSE Wide Field Mode  
without Adaptive Optics

VLT+MUSE Narrow Field Mode  
with Adaptive Optics

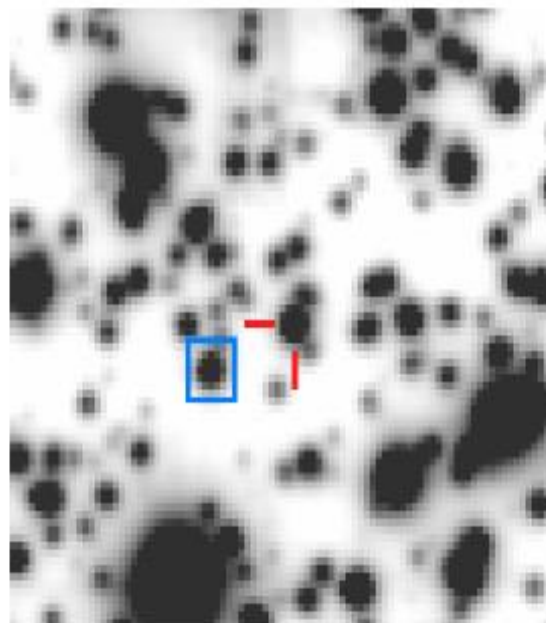
# Potenzialita' di MUSE

A detached stellar-mass black hole candidate in the globular cluster NGC 3201

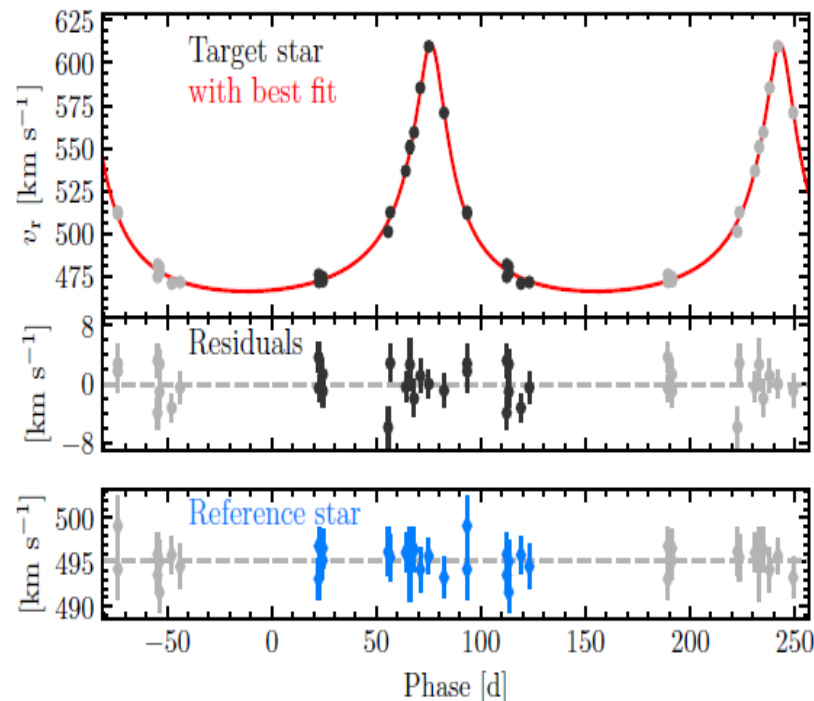
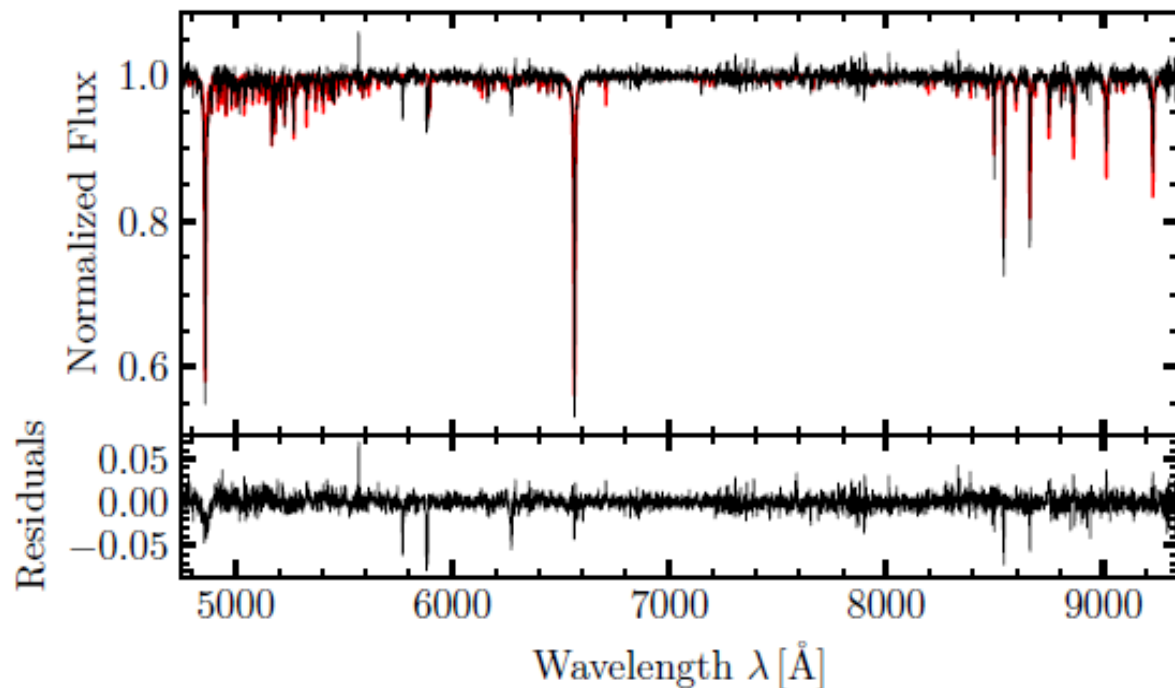
Benjamin Giesers,<sup>1\*</sup> Stefan Dreizler,<sup>1†</sup> Tim-Oliver Husser,<sup>1</sup> Sebastian Kamann,<sup>1,2</sup>  
Guillem Anglada Escudé,<sup>3</sup> Jarle Brinchmann,<sup>4,5</sup> C. Marcella Carollo,<sup>6</sup>  
Martin M. Roth,<sup>7</sup> Peter M. Weilbacher,<sup>7</sup> Lutz Wisotzki<sup>7</sup>



(a) HST

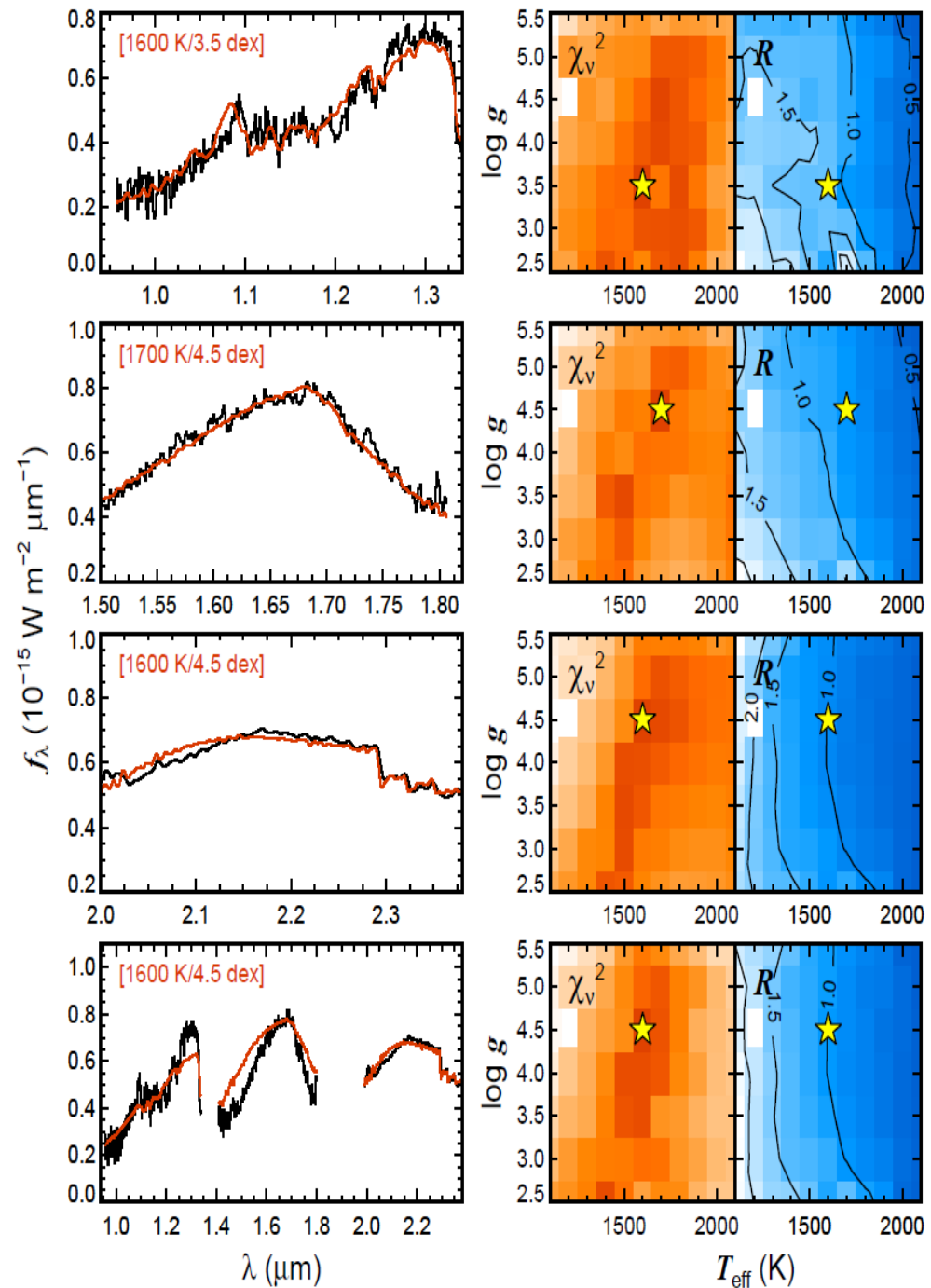
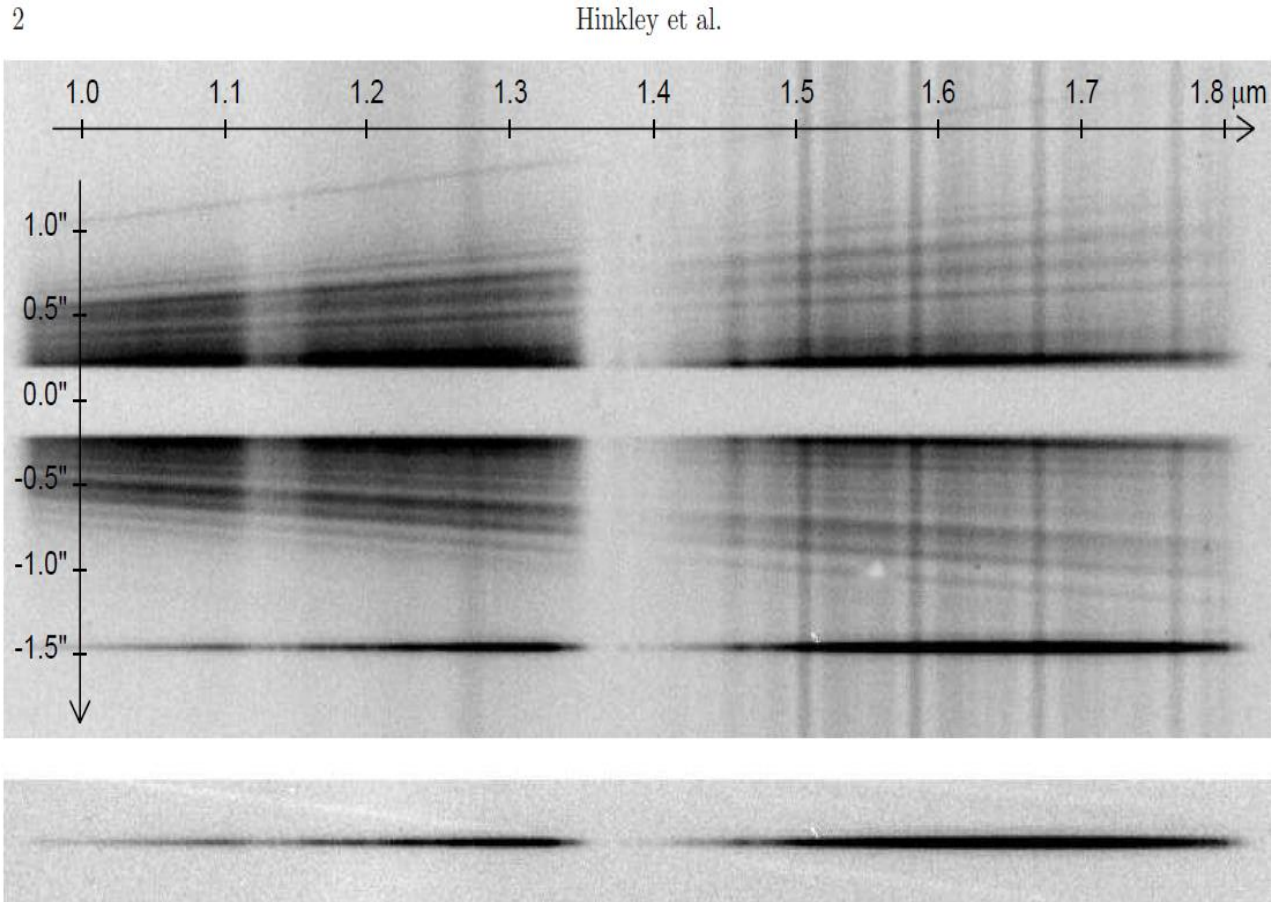


(b) MUSE





# Casi speciali: Immagini ad alto contrasto e tecniche di sottrazione delle speckles



# Casi speciali: tecniche di derivazione di velocita' radiale di alta precisione

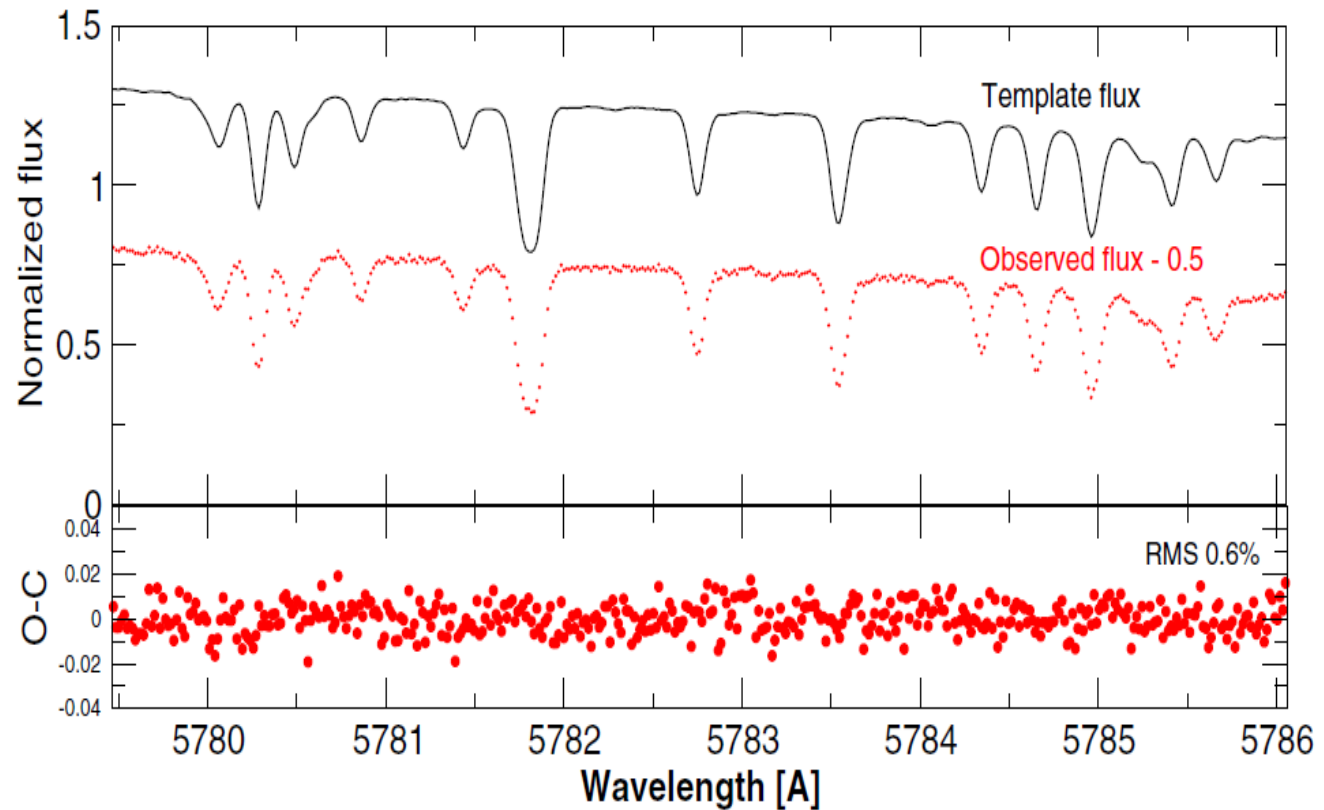
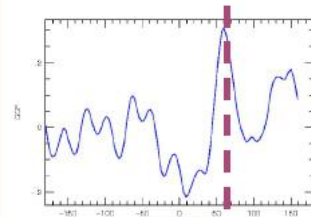
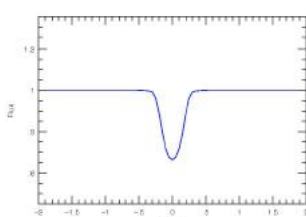
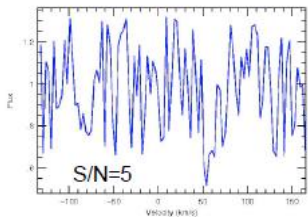
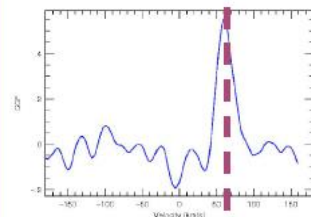
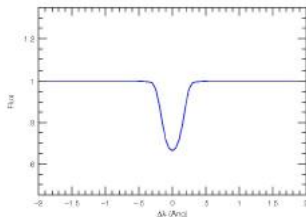
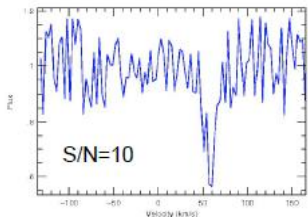
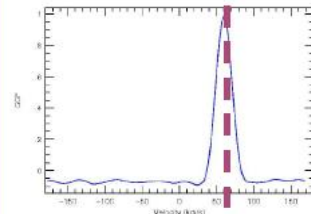
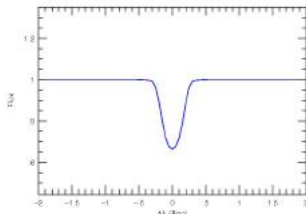
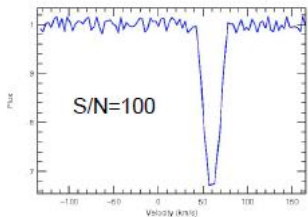
CCF: e.g. HARPS pipeline

TERRA: Anglada-Escude'+Butler 2012

Observation

Standard

CCF



Maybe replaced by a mask

Tecniche di analisi dei dati spettroscopici: estratto uno spettro (segnale in funzione della lunghezza d'onda) estrarre un dato quantitativo sul target

Due metodologie principali

Confronto con modelli e ottimizzazione di parametri (physically oriented)

Machine Learning

- Supervised machine learning → libreria di templates
- Unsupervised machine learning

# Esempio: estrazione dei parametri principali di una stella ( $T_{\text{eff}}$ , $\log g$ , composizione chimica)

- Confronto con modelli di atmosfere e ottimizzazione di parametri
- Es. MOOG, FAMA, Spectroscopy Made Easy (SME):
  - Misura di EWs oppure cfr con sintesi
  - $T_{\text{eff}}$  da potenziali di eccitazione
  - $\log g$  da equilibrio di ionizzazione
- ➔ Abbondanze



# Varianti

- Possibilita' di stabilire dei priors:
  - Temperatura da colori
  - Gravita' da confronto con isocrone – se nota la distanza
- ➔ Possibilita' di un'analisi Bayesiana (anche se finora scarsamente applicata)

# Casi speciali: misure di velocità radiale di alta precisione

- Modellizzazione del noise astrofisico (attività)
- Possibilità di usare metodi statistici (es. Processi Gaussiani e un'analisi Bayesiana dei parametri relativi)

# Pro-Con della modellizzazione fisica

## Pro's:

- Si tratta di una metodologia “assoluta” (non dipende da altre tecniche)
- Il senso fisico delle quantità ricavate è chiaro
- Diretta connessione tra osservabili e parametri derivati → relativamente facile determinazione delle incertezze nei parametri derivati e comprensione degli outliers

## Con's:

- Procedura può richiedere molto tempo di calcolo → eventualmente troppo per la mole dei dati fornite dalle survey (e comunque il rapporto costi/benefici non è probabilmente ottimizzato: nessuna economia di scala)
- I risultati sono model-dependent (es. trattazione dello scattering, Non-LTE, 3-D) e modifiche importanti richiedono cambiamenti nella pipeline

# Supervised machine learning

- Determinazione dei parametri fisici fatta per confronto con librerie di templates definiti da delle labels (es. Parametri atmosferici, composizione chimica)
- In effetti, si tratta di una classificazione, eventualmente con possibilita' di interpolazione
- Ruolo dei templates e' fondamentale
  - In questo approccio, risultati affidabili sono ottenuti solo per interpolazione all'interno della libreria
  - La libreria deve essere ottenuta con le stesse tecniche usate per le osservazioni (es. Deve avere la stessa risoluzione spettrale)
  - Problemi se la libreria non e' sufficientemente estesa e fitta (es. problemi con rotazione/binarieta')
- Pros's:
  - E' una tecnica molto veloce, adatta alla trattazioni di enormi quantita' di dati
- Con's:
  - Non chiaro il legame fisico tra osservabili e labels → gli errori nelle labels possono essere ottenuti solo dal cfr con tecniche basate sul modelling fisico
  - Difficolta' di capire la natura di outliers

# The Cannon ([Ness et al. 2015](#), [Casey et al. 2016](#))

Parameterizes the spectral fluxes as a function of a set of independently-determined stellar parameters and abundances

In principle, these could be any physical or parameterized quantities, so they are generically referred to as labels

The Cannon method uses a training set of stellar spectra to determine the values of function parameters that best match the training set spectra

This function is then applied to a larger data set to derive their unknown labels

The method has the advantage of exploiting all of the information that may be present in the stellar spectra, even if a complete physical understanding of it is lacking. For more details, refer to the papers listed above

# The Payne (Ting et al. 2018)

A general method for the precise and simultaneous determination of numerous stellar labels from observed spectra, based on fitting physical spectral models

- exploits the information from much of the available spectral range
- fits all labels (stellar parameters and element abundances) simultaneously
- uses spectral models, where the atmosphere structure and the radiative transport are consistently calculated to reflect the stellar labels
- At its core The Payne has an approach to accurate and precise interpolation and prediction of the spectrum in high-dimensional label-space, which is flexible and robust, yet based on only a moderate number of ab initio models ( $O(1000)$  for 25 labels)
- Uses a simple neural-net-like functional form and a suitable choice of training labels

# Confronto tra diversi metodi

- Da metodi empirici (es. The Cannon)
  - Libreria di templates osservati con lo stesso strumento in cui le labels sono definite da una precedente analisi basata su un modelling fisico
  - Pro's:
    - le distorsioni strumentali sono sotto controllo
  - Con's:
    - La libreria di templates deve essere sufficientemente estesa
    - Dipendenza dalla qualita' del modelling fisico dei templates
- Da metodi teorici (es. Payne)
  - Libreria di templates ottenute ad es. mediante spettri sintetici
  - Pro's:
    - Possibilita' di avere una libreria estesa ad libitum
  - Con's:
    - Necessario un modelling accurato degli effetti strumentali
    - Dipendenza dal modelling adottato



# Unsupervised machine learning

- E' un metodo di classificazione "oggettivo" basato sulla cluster analysis
- Es. K-means (es. Garcia-Dias+ 2018 per Apogee DR12)
- Il programma trova da solo le labels che ottimizzano la classificazione (divisione in gruppi) dei dati osservativi (in K-means questo e' fatto usando una Principal Component Analysis)
- Pro's:
  - Molto veloce e "unbiased" (non proprio: il risultato dipende da parametri scelti a priori, ad es. il numero di gruppi)
- Con's:
  - Il significato fisico delle labels usate nella classificazione puo' non essere chiaro
  - → difficolta' a dare una chiara interpretazione dei risultati rispetto a modelli/scenari interpretativi

# Cosa implica avere a disposizione del software

- La pianificazione e lo sfruttamento di grandi survey richiede molto software
  - Solo avere un “repository” accessibile a tutti?
  - Il software deve essere:
    - Affidabile
    - Verificato
    - Documentato
    - Mantenuto
    - Supportato
  - Eventualmente:
    - Portabile
    - Sviluppabile
    - Interfacciabile
- ➔ Chi fa questo lavoro?
- ➔ Chi decide qual e' il software che dobbiamo avere e le prioritá'?
- ➔ Solo software o anche archivi?