Automated methods for abundance determination

Carlos Allende Prieto
Instituto de Astrofisica de Canarias

Pioneering work by Beatriz Barbuy in 1987

using the Lick 3.1m and the Robinson-Wampler image-dissector scanner

PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC
99:335-368, May 1987

CARBON AND NITROGEN ABUNDANCES IN METAL-POOR DWARFS OF THE SOLAR NEIGHBORHOOD*

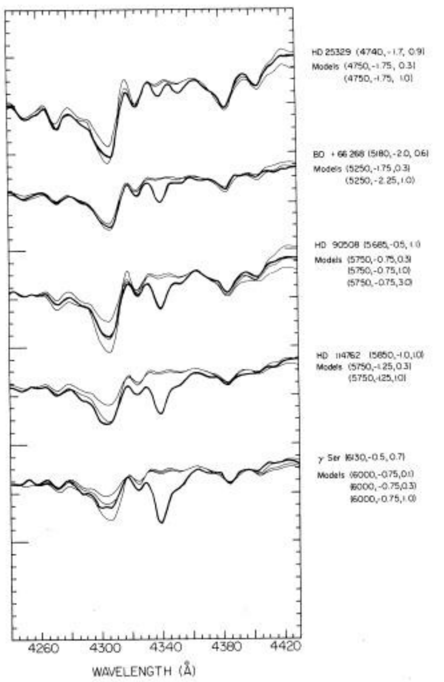
DUANE F. CARBON†
NASA-Ames Research Center, MS 245-6, Moffett Field, California 94035

BEATRICE BARBUY, ROBERT P. KRAFT, AND EILEEN D. FRIEL
Lick Observatory, Board of Studies in Astronomy and Astrophysics
University of California, Santa Cruz, California 95064

AND

NICHOLAS B. SUNTZEFF
Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatories,‡
Casilla 603, La Serena, Chile

Received 1986 October 24, revised 1987 February 2



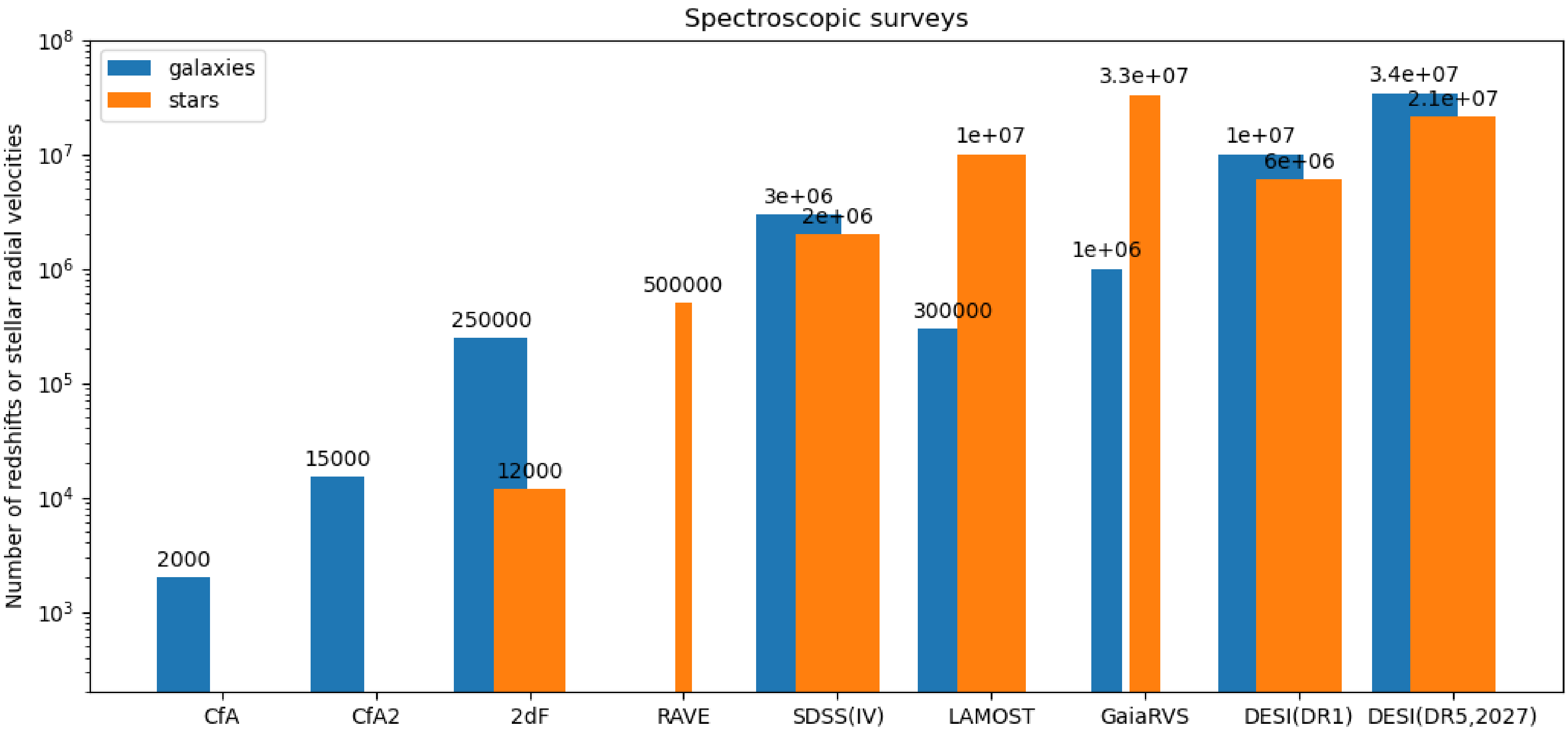
ABSTRACT

We have analyzed spectra of 83 subdwarfs obtained with the Lick Observatory 3.1-m Shane

The need for automation

- Make the spectroscopic analysis reproducible
- Make it optimal
- Make it faster
- Cope with the data acquisition rate

Scale of surveys



Machine Learning

- In fashion!
- Can be used (e.g. CNNs) to interpolate models
- Can be used to determine parameters by training on models
- Singular stars may not be properly identified by networks trained with limited data which do not include them
- Networks can learn from built-in relationships in the training sample

Data-driven methods

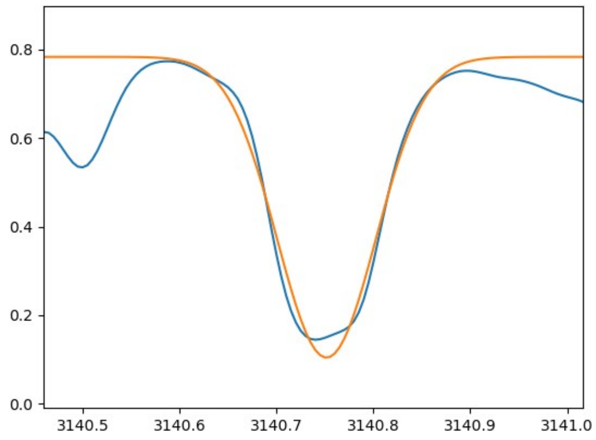
- Ultimately we need quantities we cannot measure directly: Teff, logg, [C/Fe], etc. so models are strictly necessary to assign tags to stars in the first place
- Data-driven applications, in many cases, are in fact model-driven, but without control of the quality/components of the models

Analysis of stellar spectra

- Reduce the data (e.g. using MIDAS)
- [measure equivalent widths]
- Determine atmospheric parameters
  get a model atmosphere
  compute a synthetic spectrum, compare with data
  iterate
- Determine chemical abundances
  compute a synthetic spectrum, compare with data

Equivalent widths vs. spectral synthesis

- EWs condense the information about a line in one number
- Independent of broadening such as rotation, macro-turbulence, instrumental profile
- Loss of information
- Cause blindness
- Errors tend to go in one direction



Model atmospheres

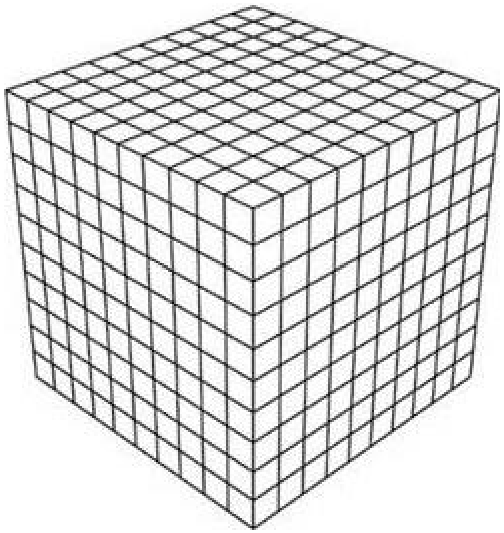
	purpose	name	main author(s)	features	comments
LTE 1-D	AS	ATLAS9	R. Kurucz	hs, pp, RE+MLT	opacity distribution functions
		ATLAS12			opacity sampling
		MARCS	B. Gustafsson, B. Plez	hs, pp/ss, RE+MLT	opacities interpolated from pre-computed tables
	DIA	SYNSPEC	I. Hubený, T. Lanz		used in combination with ATLAS, TLUSTY(LTE)
		SYNTHE	R. Kurucz		
		Turbospectrum	B. Plez		used in combination with MARCS, ATLAS, …
LTE 3-D	DIA	ASSET	L. Koesterke		SEDs from 3-D (RM)HD models
NLTE 1-D	AS+FI	CMFGEN	J. Hillier, L. Dessart	unif, ss, RE	complete CMF transfer[a]
		FASTWIND	J. Puls, E. Santolaya-Rey	unif, ss, EB	CMF transfer for individual elements[b]
		PHOENIX	P. Hauschildt, E. Baron	unif, ss, RE+MLT	complete CMF transfer[c]
		PoWR	W.-R. Hamann	unif, ss, RE	complete CMF transfer[d]
		WM-basic	A. Pauldrach, T. Hoffmann	hd (stat.), ss, EB	Sobolev line transfer[e]
	AS	TLUSTY	I. Hubený, T. Lanz	hs, pp, RE	used in combination with SYNSPEC
	DIA	SYNSPEC	I. Hubený, T. Lanz		used in combination with TLUSTY
		SYNPLE	C. Allende Prieto		python wrapper for SYNSPEC (LTE & NLTE)
		DETAIL/SURFACE	K. Butler, J. Giddings		used in combination with ATLAS[f] or TLUSTY
		Multi	M. Carlsson		used in combination with external atm. codes[g]
NLTE 3-D	AS+FI	PHOENIX	P. Hauschildt, E. Baron	unified, RE+MLT	cart/sph/cyl coordinate systems possible
	DIA	Multi3d	J. Leenaarts et al.[h]		one single atom in NLTE, others in LTE
		RH	H. Uitenbroek		multiple atoms and molecules[j]
RMHD	AS	CO5BOLD	B. Freytag, M. Steffen	hd (time-dep.), cart	see[m]
		Bifrost	B. Gudiksen et al.[k]	hd (time-dep.), cart	see[m]
		Stagger	Å. Nordlund et al.[l]	hd (time-dep.), cart	see[m]

Puls, Herrero CAP 2024

Model atmospheres

- For cool stars Kurucz, MARCS and Phoenix are the most heavily used ones, but MARCS and Phoenix codebase are not public
- Models can be taken from existing grids
- They can be interpolated from those grids
- They can be computed afresh with open source codes (Kurucz, Tlusty)

Existing grids



Kurucz: old, odfnew, Mészáros (APOGEE)

https://kurucz.harvard.edu

https://research.iac.es/proyecto/ATLAS-APOGEE/

https://data.sdss.org/sas/dr17/apogee/spectro/speclib/atmos/kurucz/

MARCS: web site, APOGEE grid

https://marcs.astro.uu.se/

https://data.sdss.org/sas/dr17/apogee/spectro/speclib/atmos/marcs/MARCS_v3_2016/

Phoenix grid from Husser et al. (2013)

https://www.astro.uni-jena.de/Users/theory/for2285-phoenix/grid.php

Interpolation of model atmospheres

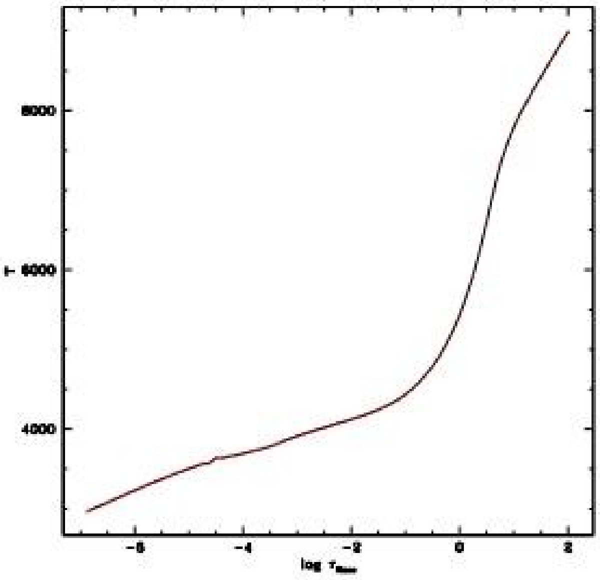
- Interpolation code for MARCS (T. Masseron)
  https://marcs.astro.uu.se/software.html

- Kmod
  https://www.as.utexas.edu/~hebe/stools/

- INNterpol
  https://github.com/cwestend/iNNterpol

Computing models afresh

- Kurucz codes are available
  https://kurucz.harvard.edu
- Sbordone port
  (Sbordone et al. 2004)
- mkk script
  https://github.com/callendeprieto/mkk-atlas9
- Mészáros (APOGEE) port
  on request to Sz. Mészáros



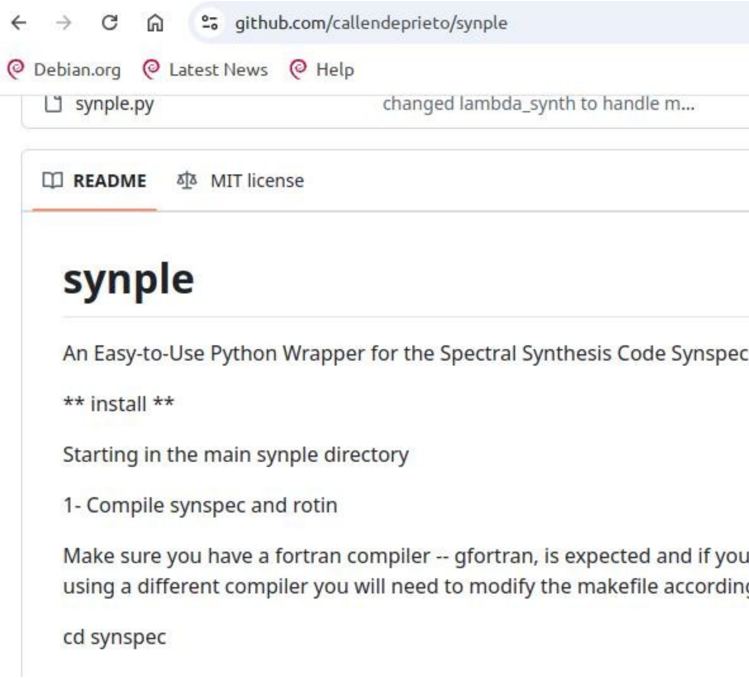
Spectral synthesis (public!)

- MOOG (Sneden)
- Turbospectrum (Plez)
- SPECTRUM (Gray)
- SYNTHE (Kurucz)
- Synspec (Hubeny)
- SME (Piskunov/Valenti)

iSpec (Blanco Cuaresma)

Synple (Allende Prieto)

synple



- Python3
- Minimal, command line interface
- Wraps synspec
- Works with MARCS, Kurucz, Phoenix and Tlusty models
- Requires only an input model atmosphere a the wavelength range
- Attemps to make *clever* decisions for you (line lists, sampling)
- Includes parallelization tools
- Includes tools for building grids of spectra

Comparing data to models

- SME (Valenti & Piskunov 1996, Piskunov & Valenti 2016)
- iSpec (Blanco Cuaresma et al. 2014; Blanco Cuaresma 2019)
- FERRE (Allende Prieto et al. 2006)
- BAS

See also (private): MATISSE – GAUGIN (Recio Blanco, Bijaui), q2 (Ramirez), BACCHUS (Masseron) ...

FERRE

- Fortran90, parallelized with OpenMP
- Intrinsically n-D, flexible to fit one, several, all dims.
- Stores pre-computed grids of synthetic spectra in RAM. Can PCA-compress the model database
- Interpolates in the grids with linear, quadratic or cubic polynomials
- Successfully used in APOGEE (also SDSS, DESI, WEAVE, etc.)
- Download from

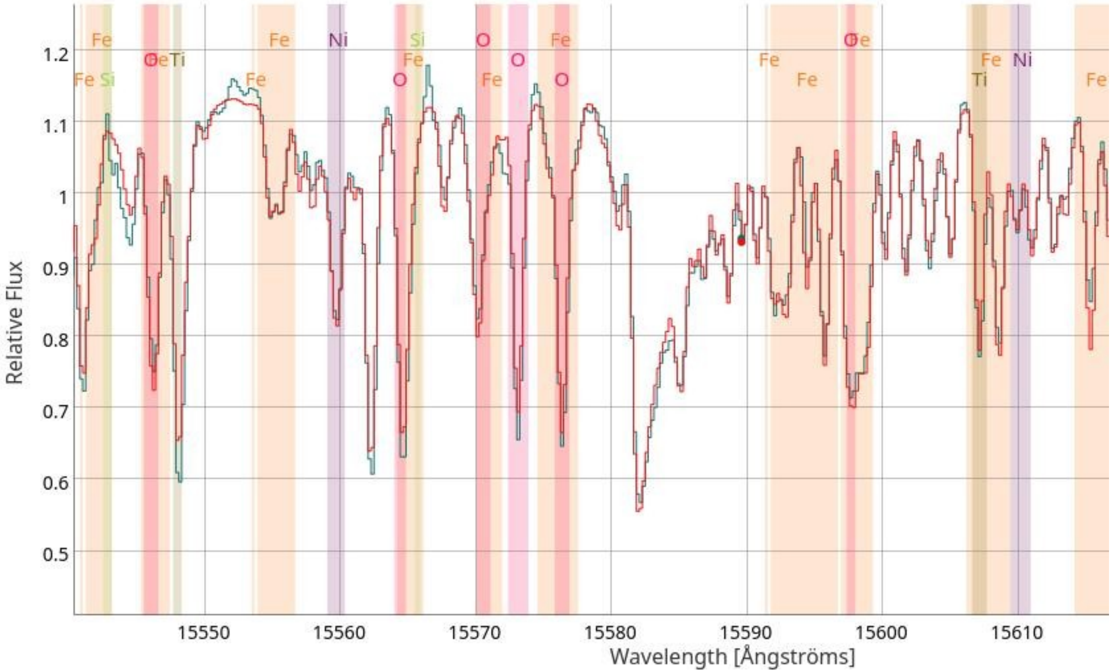
  https://github.com/callendeprieto/ferre

Optimization algorithms in FERRE

- Nelder-Mead algorithm (Neldel & Mead 1965)
- Boender-Timmer-Rinnoy Kan (Boender et al. 1982)
- UOBYQA (Powell 2000)
- MCMC with differential evolution (Vrugt et al. 2009)

FERRE in APOGEE

- 1st pass fitting the entire spectrum 1500-1700 nm to derive Teff,logg,[M/H],[alpha/M],[C/M],[N/M], vsini
- 2nd pass fitting abundances, one element at a time
- Increased complexity due to wavelength- and fiber-dependent LSF



Is FERRE good for all?

- A bit too heavy to deploy for 1 or few stars
- RVs have to be input
- Fitting [M/H] to derive abundances does not generally work at lower resolution
- 5-7D fits require multiple starting points for local algorithms, and global ones (MCMC) are expensive
- Regular grids are required by FERRE and increase complexity for handling large ranges in the parameters
- Size limitations already hit for APOGEE (1e4 frequencies, 7 parameters, grids with millions of models)
- Gridding effects still appear in some dimensions

BAS

- Interpolation reduces the model evaluation (s to ms) but the speed bottleneck is usually access to RAM
- Pre-computing the interpolations can save additional time
- Move to a Bayesian scheme brings a global algorithm, and more reliable error bar estimates
- Uses python for coding simplicity/libraries – it’s part of synple

Bayesian Algorithm in Synple

- Simplest possible Bayesian algorithm: flat priors

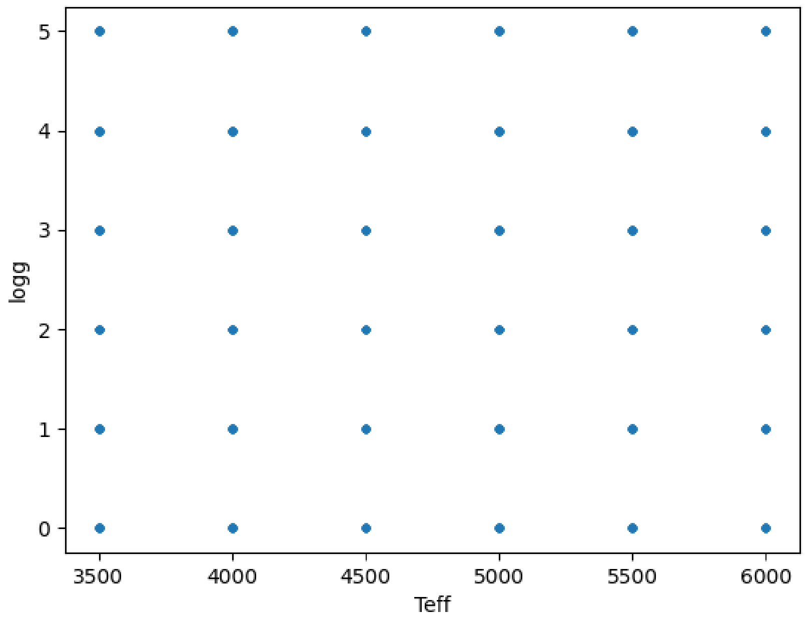
  $E(x) \sim \Sigma\, x * \exp(-chi/2)$

  $covar(x,y) \sim \Sigma\, (x-E(x)) * (y-E(y)) * \exp(-chi/2)$

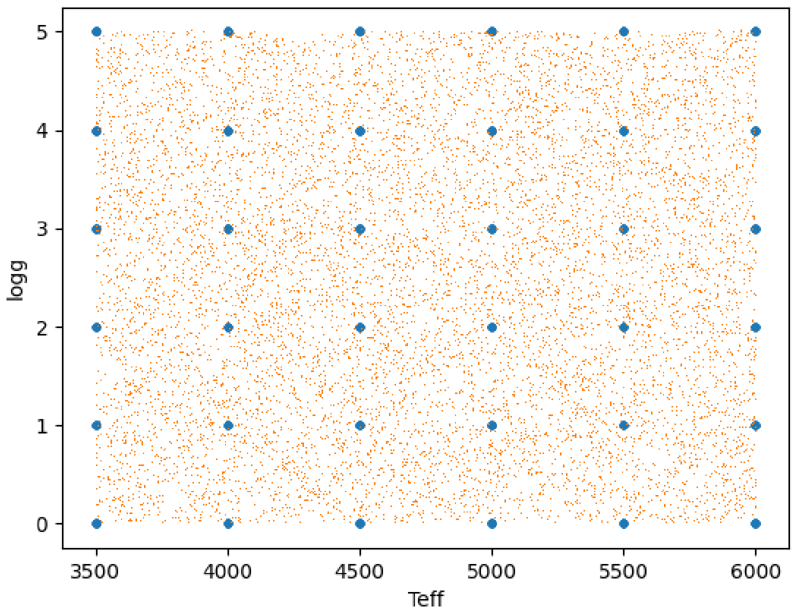
- Use all models in an interpolated irregular grid: may do a lot more work than with optimization algorithm, but always the same → new opportunities for parallelization

Irregular grids: RBF

Regular



Irregular

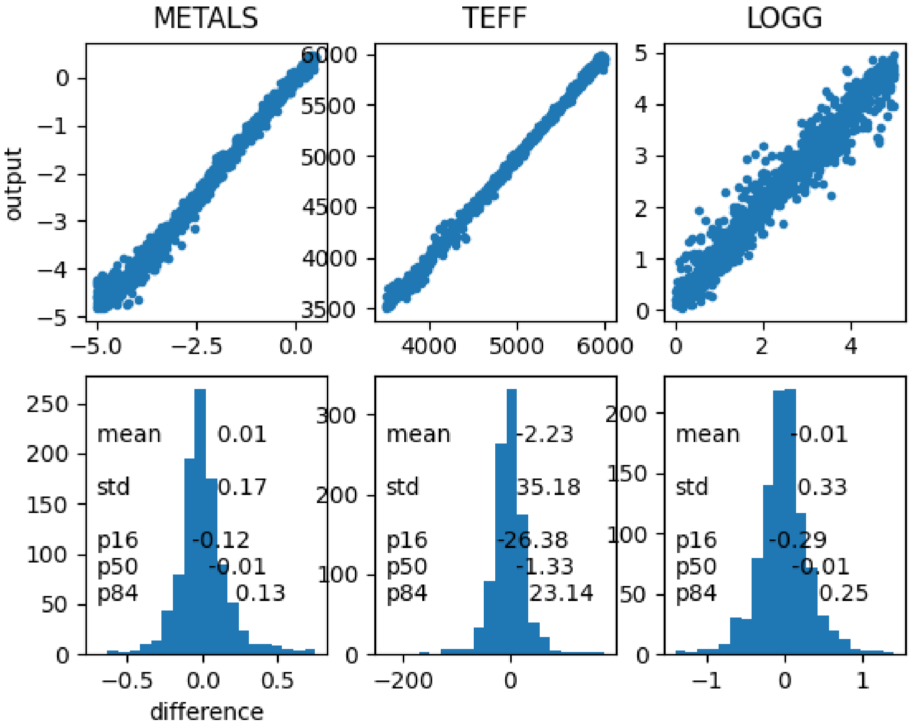


One step closer to full automation

- BAS recognizes the source of the data and them uses a specific grid
- Available grids for K-F-G-A type stars and white dwarfs
- Available for INT-IDS (MILES), HST-STIS/NICMOS, GTC-OSIRIS, Mayall-DESI, Gaia XP
- Incorporates cross-correlation and template matching for RV determination
- Incorporates on-the-fly reddening corrections
- And self-evaluation

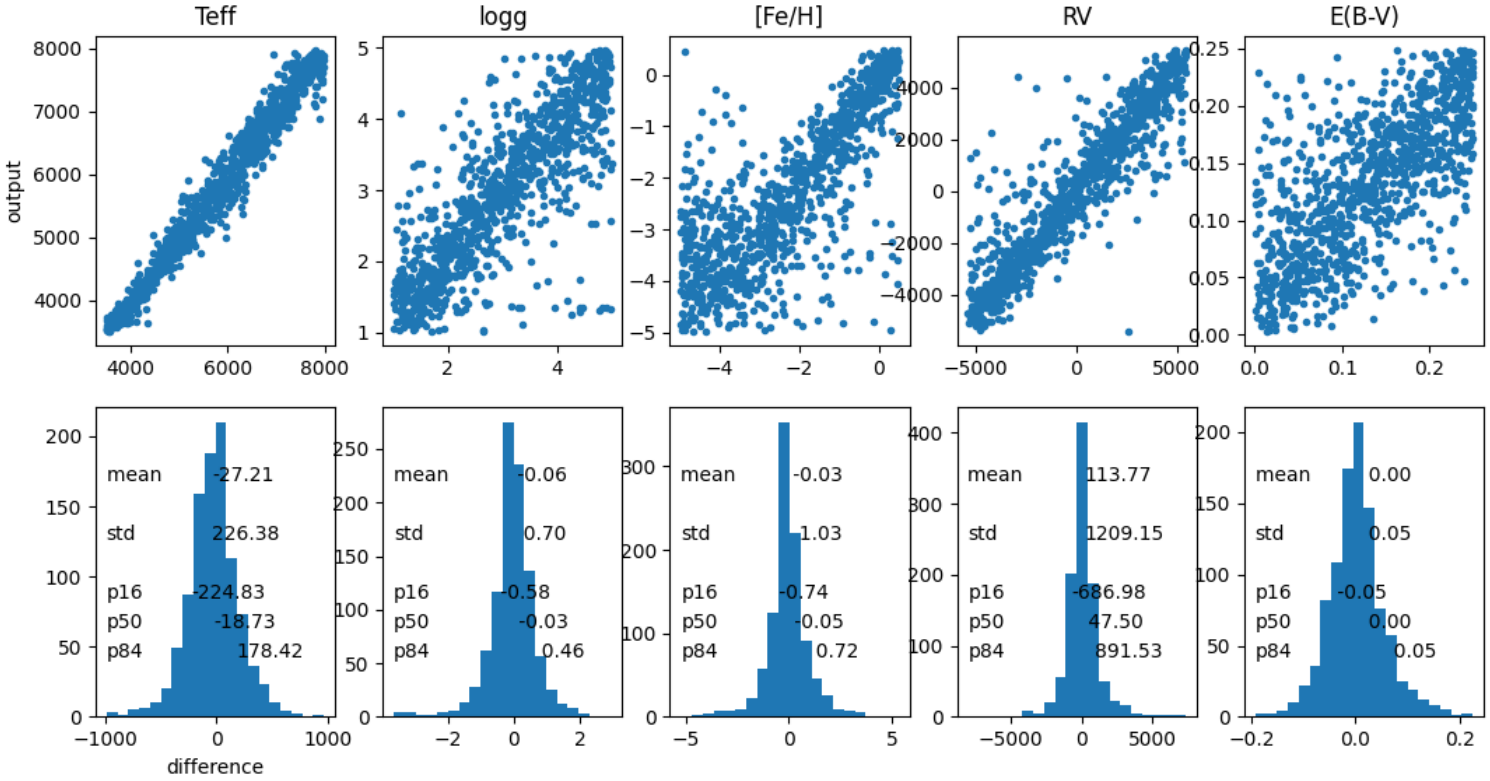
Test results

- R=500 (360-1000 nm)

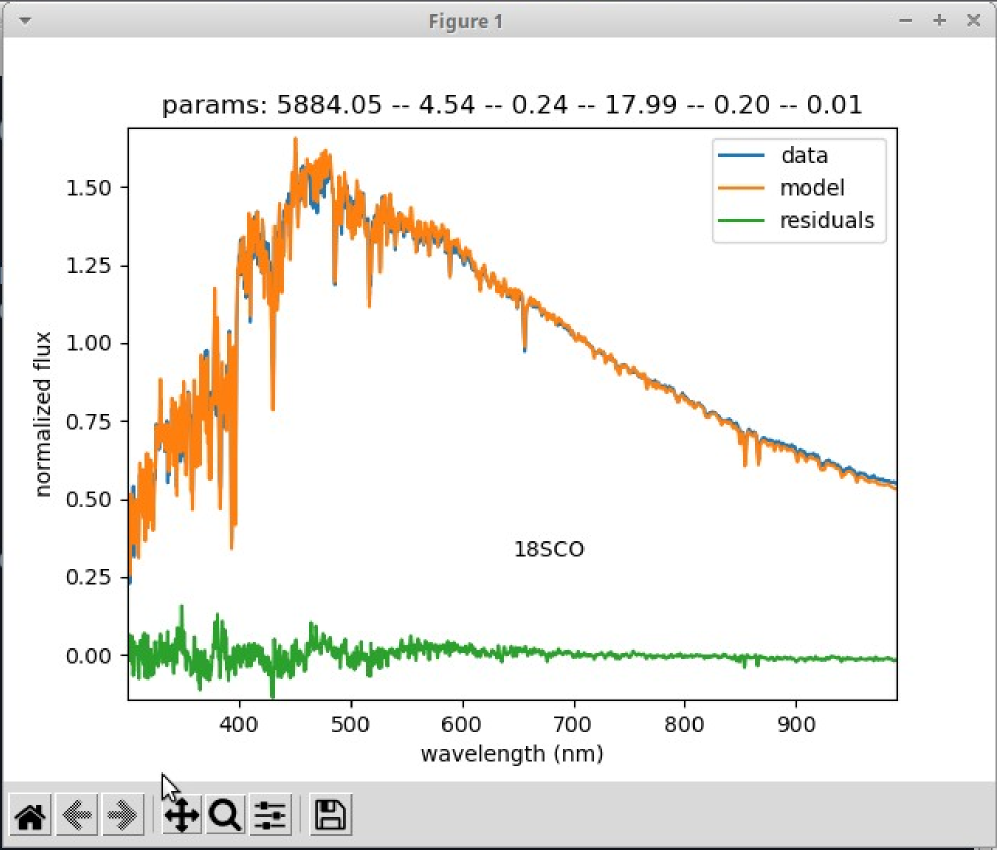


Test results

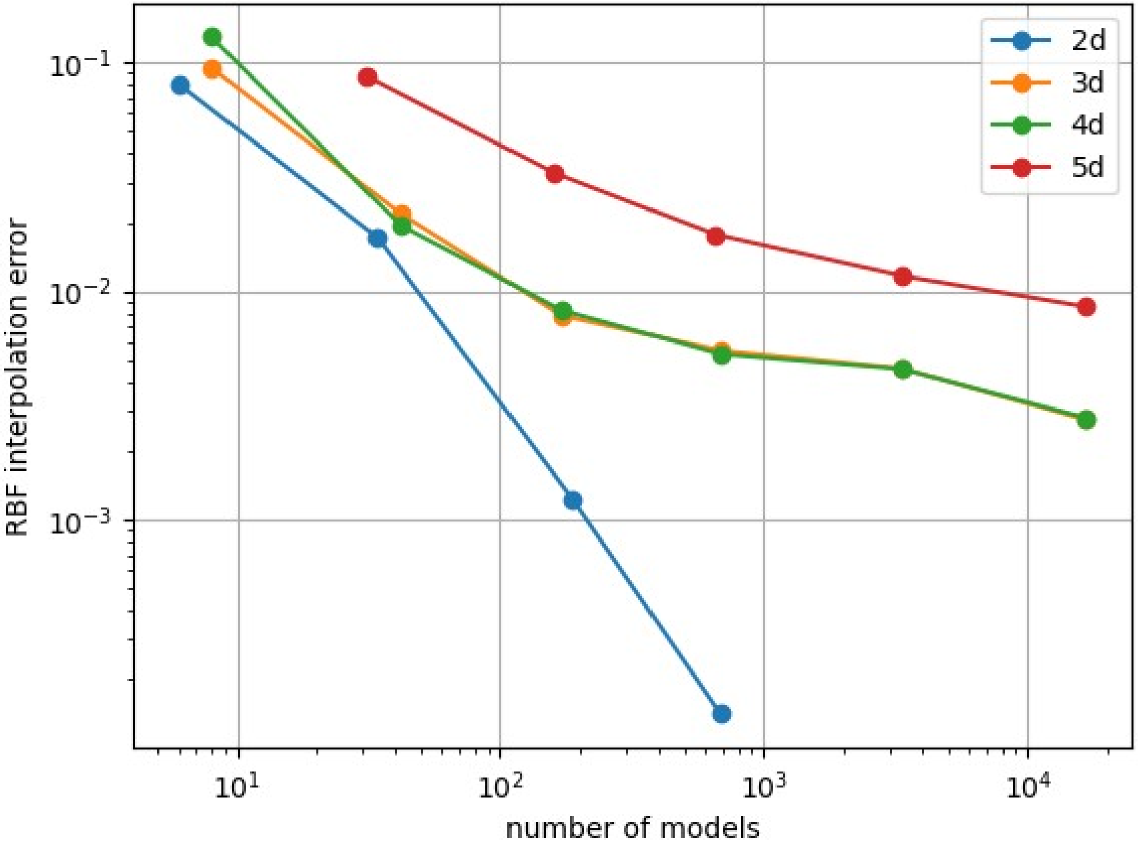
- R=55 (Gaia XP)



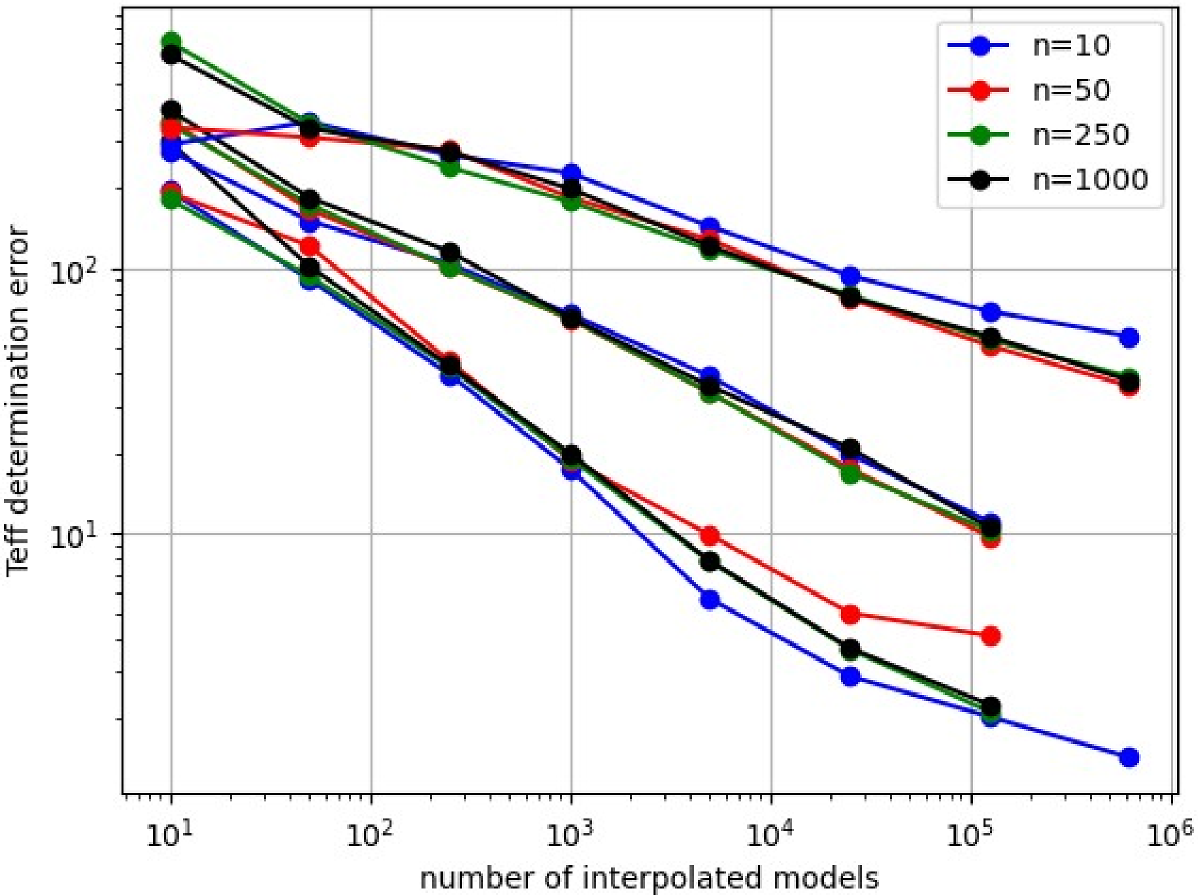
```
callende@pi:/work/callende/baltimore/calspec$ ls *fits
18sco_stis_004.fits          g191b2b_stiswfcnic_004.fits  kf
bd_17d4708_stisnic_007.fits  hd209458_stisnic_008.fits    si
callende@pi:/work/callende/baltimore/calspec$ ipython3
frPython 3.10.12 (main, Sep 11 2024, 15:47:36) [GCC 11.4.0]
Type 'copyright', 'credits' or 'license' for more informatio
IPython 7.31.1 -- An enhanced Interactive Python. Type '?' f

In [1]: from synple import bas, plot_spec

In [2]: bas('18sco_stis_004.fits')
data appear to be from  CALSPEC
adopting synthfile  n_sc2-STISrbf.pickle
reading grid /work/callende/synple/grids/n_sc2-STISrbf.pickl
reading header of the grid /work/callende/synple/grids/n_sc2
normalizing grid...
reading data from file 18sco_stis_004.fits...
nspec in bas: 1
spectrum  0  of  1  in  18sco_stis_004.fits
res= [5.88460622e+03 4.54337831e+00 2.44747185e-01] eres= [1
reduced lchi = 0.9828692796748426
RV =  -23.322775657220163  km/s
res= [5.88405287e+03 4.53666815e+00 2.44869648e-01] eres= [1.79874828e+01 1.95284456e-01 1.33902690e-02]
reduced lchi = 0.9623120652904065
closing opf file: 18sco_stis_004.fits.opf
Out[2]: ()

In [3]: plot_spec('18sco_stis_004.fits',res=True)
```



Interpolation errors



Uncertainty in the derived BAS parameters: Teff



New strategies with BAS

- 4-5 dimensions workable with ~ 1e5 data points randomly distributed
- Grids with +1 dimension are then used for abundance determination.
- These are exact (element X is actually changed), but may have limited wavelength coverage
- Parallelization for GPU exploits the nature of the problem: always the same models to evaluate and number known in advance

Grid building

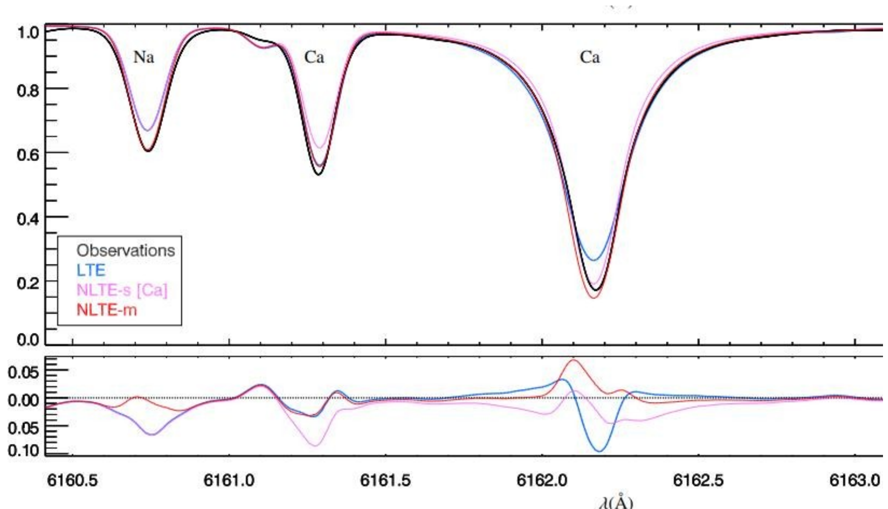
- Synple includes routines for building regular/irregular spectral grids from existing model atmospheres grids (MARCS, Kurucz), or computing them from scratch (Kurucz)
- It also includes RBF interpolation routines for creating irregular grids from existing regular/irregular grids

NLTE

- The overall strategy of using pre-computed spectral grids works the same regardless of the models
- NLTE already implemented in the DR7 APOGEE grids (Na, Mg, K, Ca) or in the SME analysis in GALAH
- Note that NLTE effects can propagate

  from one element to another!

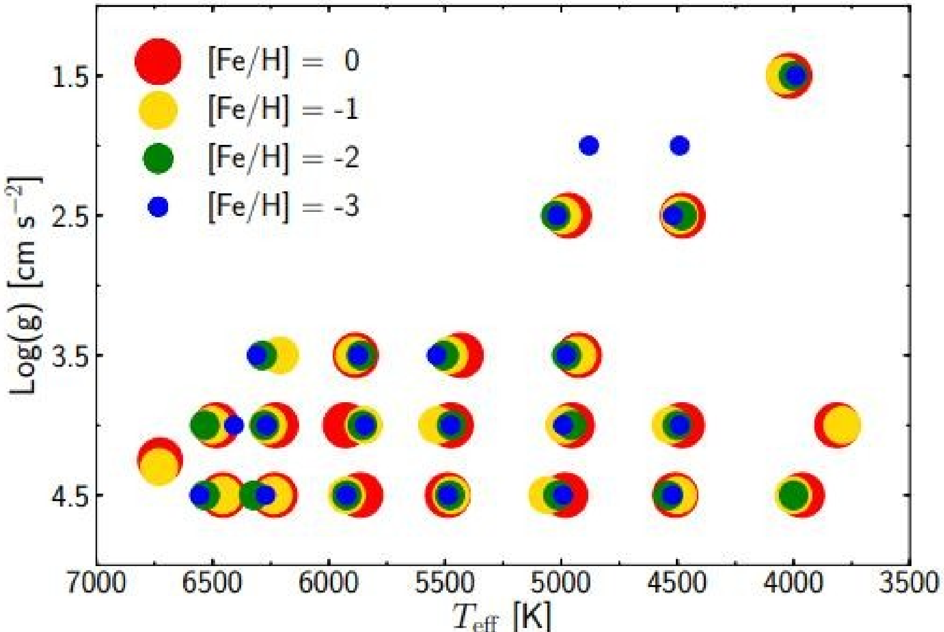
  (Osorio et al. 2022)



3D

- 3D models are scarce, so grid density is an issue, but possible to use a grid of flux ratios F(3D)/F(1D) on which interpolation can be used to correct high-density F(1D) grids (Bertran de Lis et al. 2023)

- Using 3D models requires doing NLTE in 3D, not <3D> = 1D



CIFIST MHD grid (Ludwig et al. 2009)

Summary

- Full automation means passing a reduced spectrum and getting all possible information back (parameters and abundances)
- We do not have such software yet,

  but we need it
- It will likely be available within 1-2 years,

  and we already have an approximation

  for some data sets
- Progress is hindered by keeping codes

  and models to ourselves, so please be

  open and share them!