Survey-to-survey comparison and homogenization

Elena Pancino

Osservatorio Astrofisico di Arcetri (INAF), Space Science Data Center (ASI-INAF-INFN)

INAF
ISTITUTO NAZIONALE
DI ASTROFISICA

The concept of standard

to an accuracy of <1mm ($<10^{-3}$) and a precision of 10^{-5} m

- How long is a meter?
- We should have one template
- All other meters should be like that
- All quantities expressed in meters agree

to be fair: even for the meter there are different methods, more accurate and precise than the platinum-iridium meter

Musée des arts et Métiers. Paris. France

INAF ISTITUTO NAZIONALE DI ASTROFISICA

SSDC SPACE SCIENCE DATA CENTER

In astrophysics

There are only four direct observables

1. Flux of photons (neutrinos, etc.)
2. Arrival time of photons
3. Sky position of photons
4. Wavength/frequency of photons

standardized units, but in physics!

- Plus any combination of the above
  - Light curve, spectrum, proper motion …

The Whirlpool Galaxy, Wikipedia

Example: the case of Vega

Vega is one of the few stars with a directly measured flux, used to define several photometric systems

However, it is a:

1. pole-on rotator (Peterson+06)
2. variable to ±0.03 mags (Fernie 81)
3. with a dust/debris belt (Su+05)

Vega, artist impression, NASA/JPL

What is a meter ±3 cm? Not a good standard!

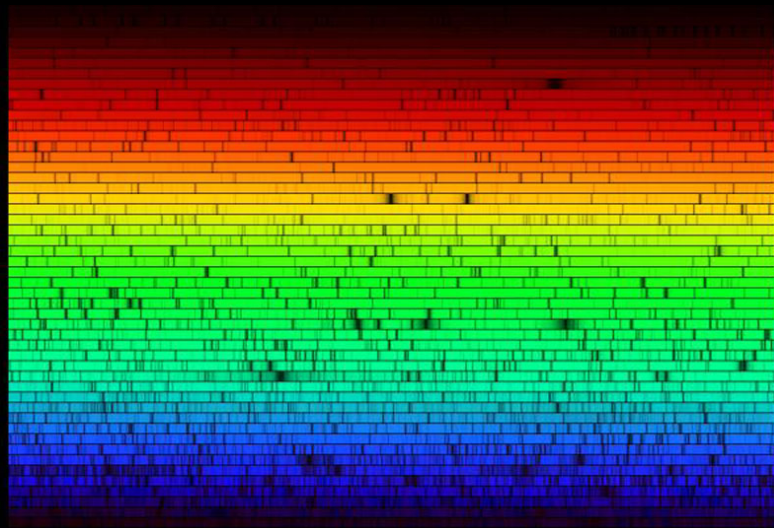
In spectroscopy

In stellar spectroscopy we derive, e.g.:

1. Radial velocity (Doppler effect)
2. Stellar parameters (from models)
3. Abundance ratios (from models)

These rely on:

line parameters (X_{ex}, loggf, λ, damping), solar abundances, atmospheric models, model atoms … and more



https://apod.nasa.gov/apod/ap070624.html

Standardized spectroscopy?

Not completely impossible, but certainly very challenging …
… nobody really can do that for the moment.

Therefore it is safe to say that:

there are no standard objects
(in the iridium-platinum meter sense)
in stellar spectroscopy

e.g., we cannot place
a thermometer into
a star to get Teff

we can build reference
(consensus) objects

What to do if we cannot count on standards?

1. ***Compare*** with different «measurements»
2. Possibly obtained with different/independent methods
3. Possibly obtained with different instruments
4. Possibly of a much higher quality
5. Or at least look for ***self consistency***, like in the EW method

compare, re-compare, compare again, and compare once more

Now that we know

1. That there are no proper/true spectroscopic «standards»
2. That we are not measuring, but inferring from data+models
3. That we cannot calibrate, but just compare and choose

We can go on and freely use the words «standard», «measurement», and «calibration», being aware that they mean indeed «reference object», «estimate/infer/derive», and «compare», respectively

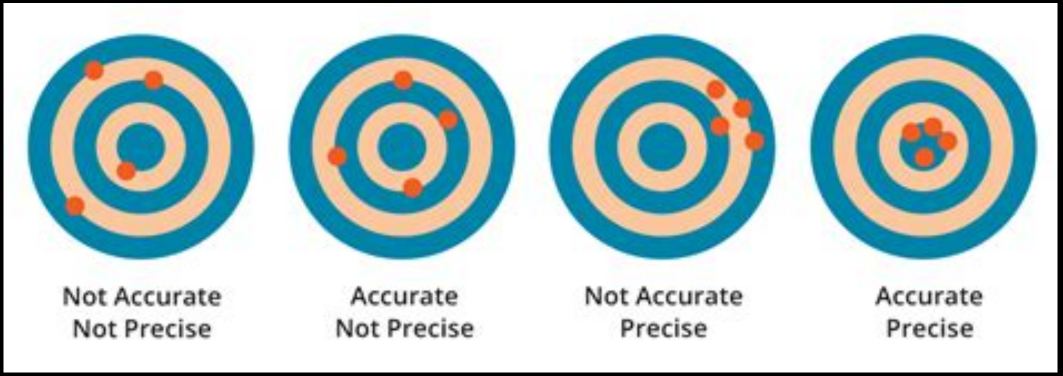
Internal and external uncertainties

Self consistency → homogeneity → precision → internal → scatter → random

Comparison → calibration → accuracy → external → bias → systematic

The terminology will vary, and things can be more complicated, e.g.:

- Continuum placement
- Choice of parameters



Enter the main parameters

In the following, I will focus on the following parameters

- Effective temperature, or Teff (K)
- Surface gravity, or logg (dex)
- Iron abundance, or [Fe/H] (dex) – alternatively: metallicity
- Radial velocities (RVs)

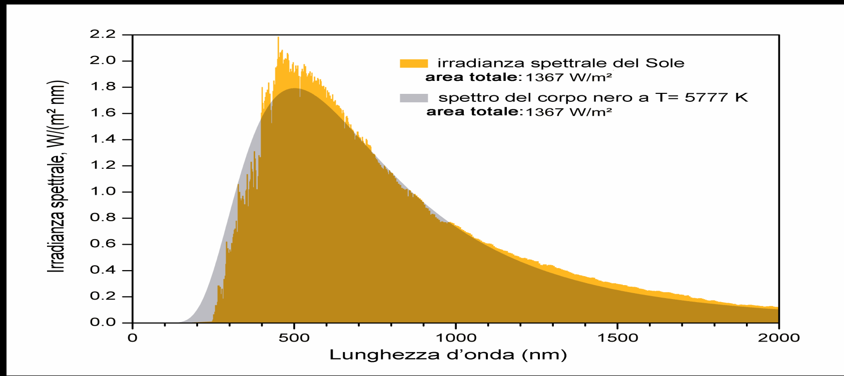
But some considerations will remain valid for other parameters as well, such as v_{micro}, v_{macro}, [El/Fe], vsini, etc.

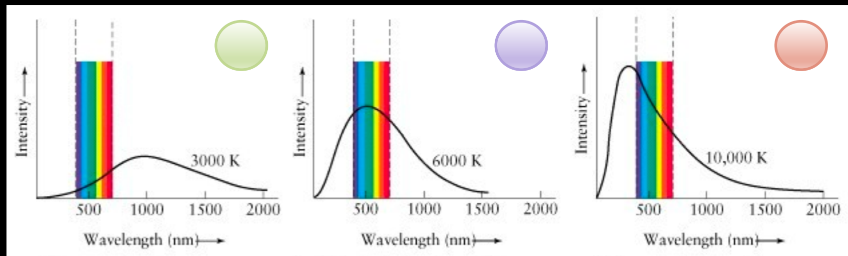
Effective temperature

Teff at the «bottom» of the atmosphere (optical depth of $\tau = 2/3$)

- Wien's law: $\lambda_{peak} \propto 1/T$

But stars are not perfect blackbodies and very few stars have their flux (F_{bol}) measured across the whole spectrum





Methods for effective temperature

- From the lines/EWs – excitation equilibrium, line depth ratios
- From spectral synthesis – find the most T_{eff}-sensitive features
- From bolometric fluxes and interferometric radii
  - Link T_{eff} and F with the Stefan-Boltzmann equation: $F \propto \sigma T^4$
  - Link L, r, and F with stellar luminosity eq.: $L \propto 4\pi\, r^2\, F$
  - Thus: $T_{eff} \propto \theta^{-1/2}\, F_{bol}^{1/4}$ and from the IR you can get F_{bol}
- From the IR flux method (simple relations for F_{bol})
- From color-T_{eff} relations (within 50—100 K)

see Chris' lecture

need stellar diameters from interferometry

and/or IR observations to get the F_{bol}

need good photometry and a reddening estimate

might need distances as well

Surface gravity

The gravitation law in astronomical notation:

see also
Chris' talk

$$\log \frac{g}{g_\odot} = \log \frac{M}{M_\odot} - 2\log \frac{R}{R_\odot}$$

$F \propto \sigma T^4$; $L \propto 4\pi\, r^2 F$

First basic consideration:

The cross-match with photometric & astrometric surveys and gas/dust maps is mandatory

You either need T_{eff} and F_{bol} or you need the radius

In any case there is no way to get logg *without a mass estimate*

unless you use spectroscopy (ionization equilibrium or synthesis)

Astroseismology

Periodograms obtained by space missions like CoRoT, Kepler, or TESS provide quantities that correlate with the mass and radius of stars

$$R \approx R_\odot \left(\frac{\Delta\nu_\odot}{\Delta\nu}\right)^2 \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right) \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{1/2},$$

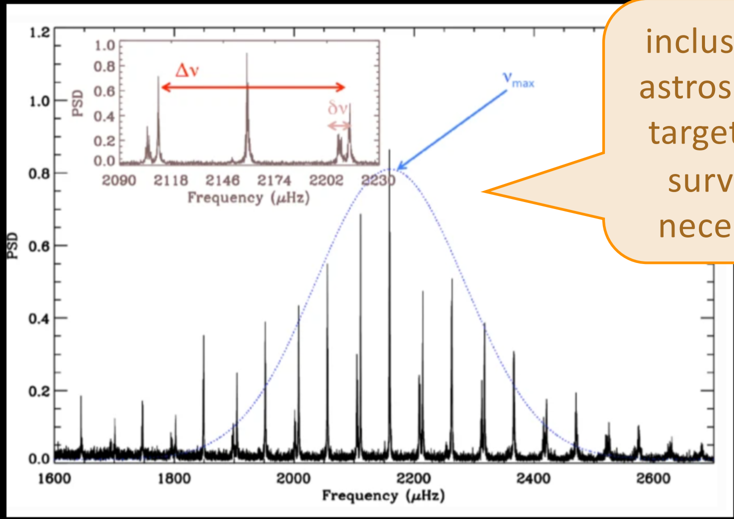
$$M \approx M_\odot \left(\frac{\Delta\nu_\odot}{\Delta\nu}\right)^4 \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right)^3 \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{3/2}$$

$$\nu_{max} \propto {}^{g}/_{\sqrt{T_{eff}}}$$

thus, with T_{eff} in hand, one can have a good handle on logg, with uncertainties 5-10 times smaller than in spectroscopy (δlogg < 0.1 dex)

Second consideration:

inclusion of astroseismic targets in a survey is necessary



Power spectrum density of 16 Cyg A, Garcia & Ballot (2019)

Ways to homogenize and calibrate

Focusing on the three parameters Teff, logg, and [Fe/H], we now review some possible ways of comparing and combining data

They can be used to compare and combine data obtained with different methods within a survey (e.g., Gaia-ESO survey, Gilmore et al. 2012)

e.g., Smiljianic et al. 2014

Or they can be used to compare and combine different surveys (e.g., the Survey of Surveys, Tsantaki et al., submitted to A&A)

e.g., Spina et al. 2014

Every spectroscopic work benefits from comparisons, of course, including yours

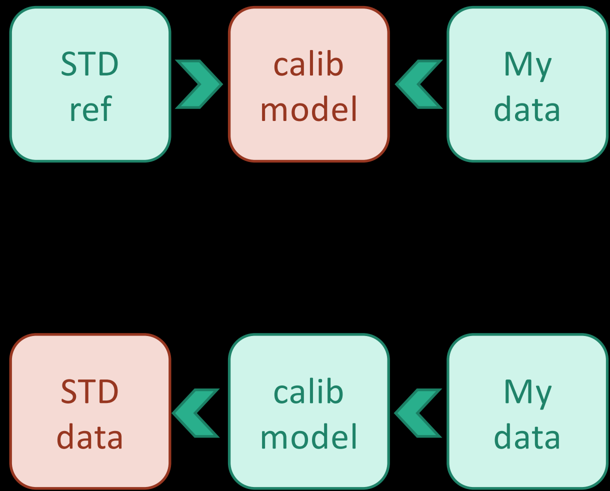
or scientific

Method zero: the calibration concept

Fundamental concept, applied at least as a check in most if not all the surveys

If you have standards

1. You observe them and create your calibration model
2. With the model, you can then calibrate all your observations



Method 1: look for astrophysical insight

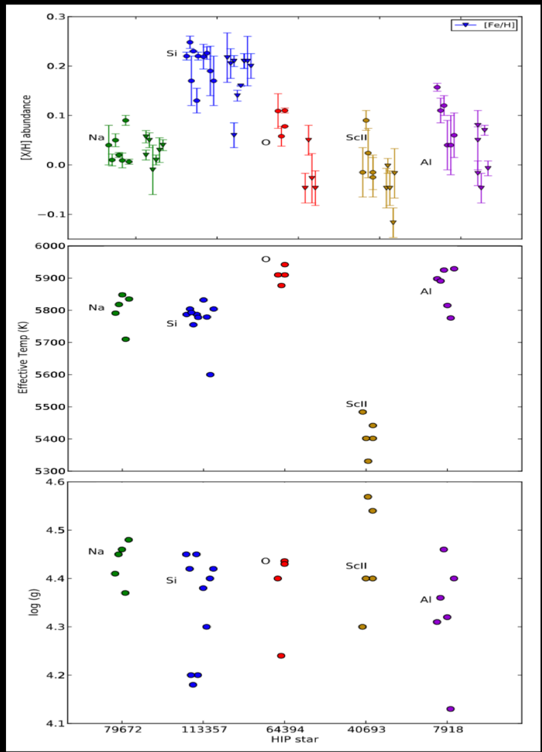
Generally used for relatively small, high quality datasets, or *new independent measurements*:

- the Gaia benchmark stars (Jofre et al. 2015)
- the Hypatia catalogue (Hinkel et al. 2014, 2018)
- the PASTEL database (Soubiran et al. 2016)

All surveys use method 1 and you should, too

The method does:

- Try to make sense of spreads and trends
- NOT recalibrate unless there is a very clear reason for doing so (e.g., solar reference abundance, different Teff reference scale)



Soubiran et al. 2016

Method 2: go straight to the end

Some surveys like APOGEE use method 2

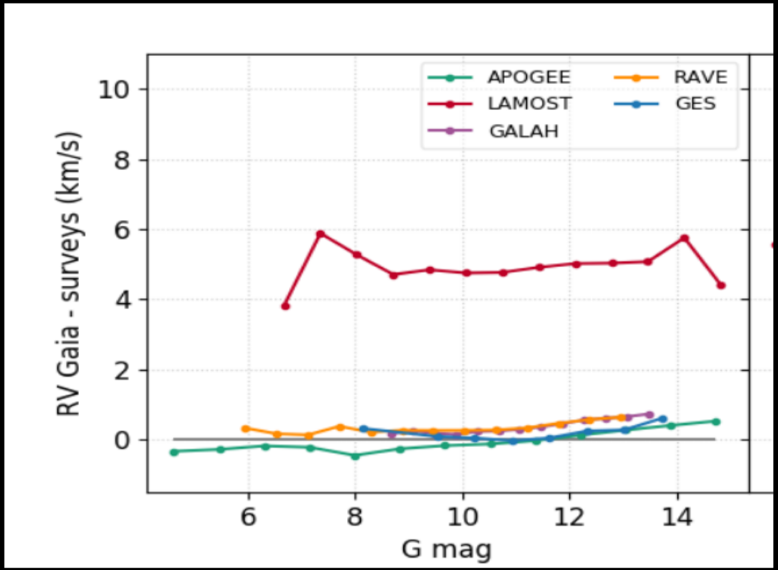
This method is *phenomenological*

- Compare different datasets (at least three)
- Use the RV/Teff/logg/[Fe/H] differences
- Reconstruct the «true» spreads

$$\sigma_{12}^2 = \sigma_1^2 + \sigma_2^2, \quad \sigma_{13}^2 = \sigma_1^2 + \sigma_3^2, \quad \sigma_{23}^2 = \sigma_2^2 + \sigma_3^2$$

Three cornered-hat method

- Similarly with the ZP and trends:
  - If one survey only shows offsets from all others
  - Then correct the ZP of that survey only
- Once errors and trends are recalibrated: merge



Survey of Surveys, Tsantaki et al., in press (see also Katz et al., 2018)

Method 3: standardize even before beginning

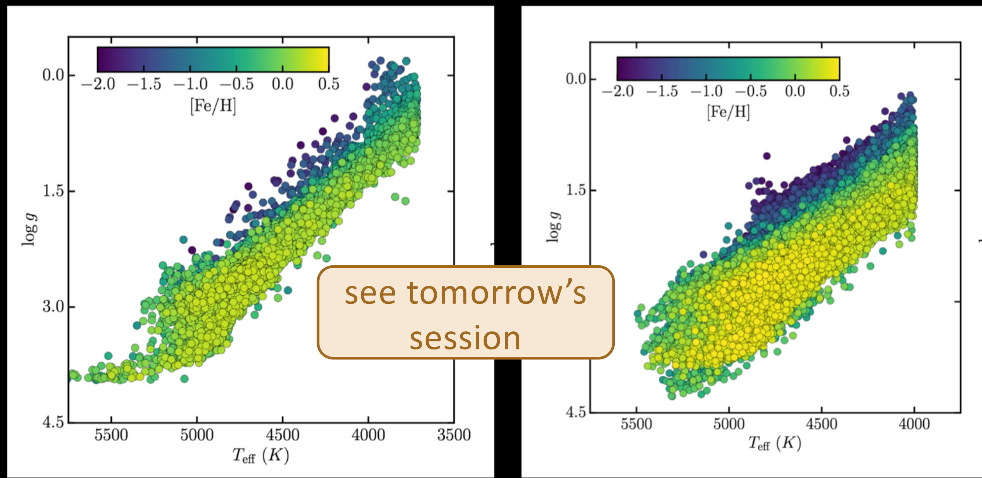
Enter the realm of machine learning:

- Use high(er) quality measurements
- Train your algorithm on them
- Apply to the whole survey set

no independent results

Examples:

- Cannon (APOGEE, Ness et al. 2015)
- Payne (LAMOST, Tiang et al. 2017)



see tomorrow's session

The ≃12000 stars in the Cannon training set for APOGEE (left) and the results for the ≃90000 stars in the APOGEE DR14 release (from Casey et al. 2016)

train set > ML < Spec-tra

out-put < ML < My data

I have one survey: I observe calibrators

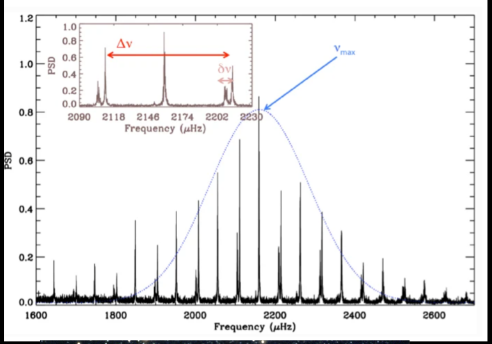
Imagine you are designing your own survey,

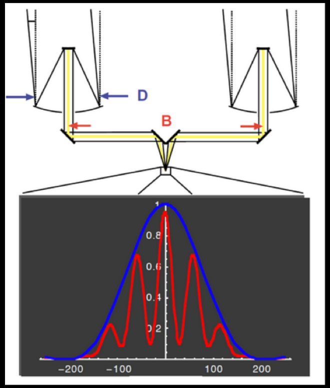
then do plan for observations of «calibrators»

- Stars with higher quality data in the literature
- Stars with independently determined parameters
- Stars in common with other large surveys
- Objects studied by many others

You have two broad aims

- Ensure your data are homogeneous
- Ensure your data compare well with others





INAF
ISTITUTO NAZIONALE DI ASTROFISICA

SSDC
SPACE SCIENCE DATA CENTER

Gaia benchmark stars

Every classical abundance analysis is tested on the Sun or on Arcturus, by tradition

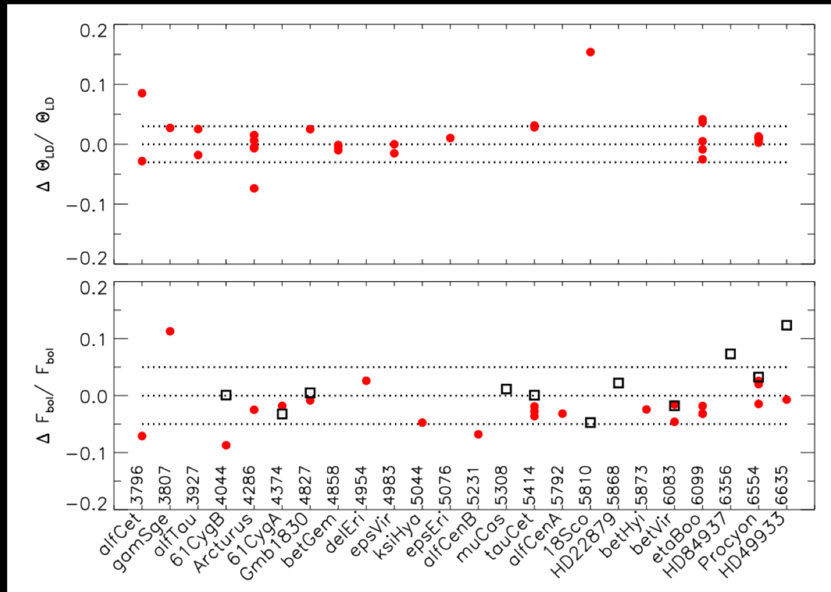
The Gaia benchmark stars are a compilation of stars with (multiple) determinations of:

- Bolometric fluxes
- Interferometric radii

too bright for most surveys
2 < V < 9 mag

For them, Teff and logg can be derived to 30-80 K and 0.02-0.20 dex respectively

fully independent from typical methods



Gaia benchmark stars, Heiter et al. 2015
Used in Gaia-ESO Survey (Jofre et al. 2015)

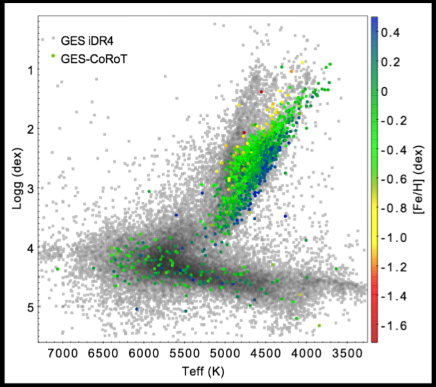
Asteroseismology

Samples of stars with astrosismology help in improving the *logg estimates*

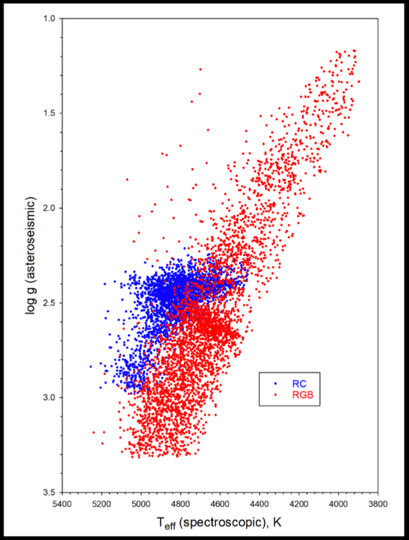
- CoRoT
- Kepler
- TESS

This helps also with the evolutionary phase

using the scaling relations one can check and adjust the survey data analysis strategy, identify biases, and/or recalibrate parameters



The GES-CoRoT sample (Pancino et al. 2017)



The APOKASC sample (Pinsonneault et al. 2018)

Star clusters

Test your homogeneity

- Same metallicity (mostly)
- Same distance
- *Same mass above the TO*
- *Trends of [Fe/H] vs. Teff and logg*

«solve» biggest problem for logg

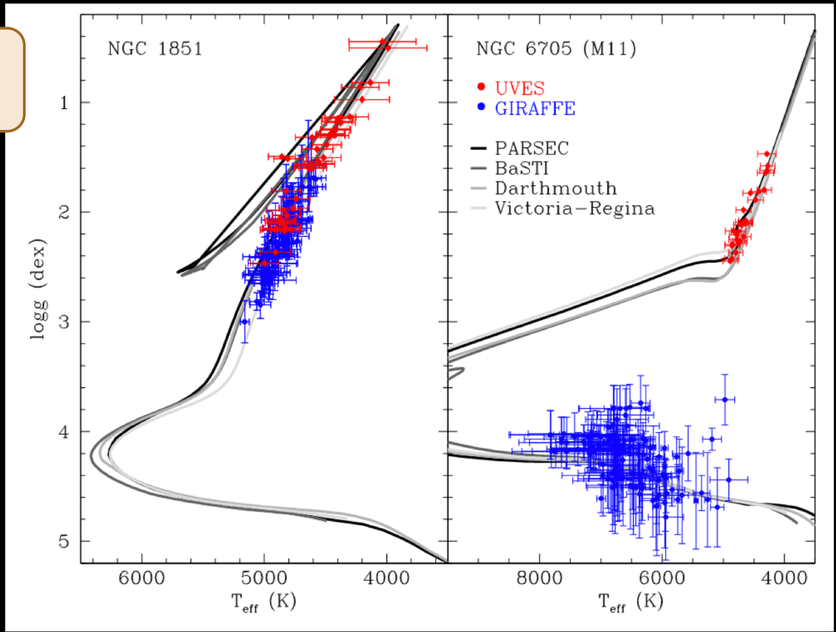
species not affected by evolutionary effects

Test your metallicity scale

- On well studied clusters

Test your parameters

- With models and with photometry



Two GES calibrating clusters (Pancino et al. 2017)

INAF ISTITUTO NAZIONALE DI ASTROFISICA

SSDC SPACE SCIENCE DATA CENTER

Stars in common with other surveys/sets

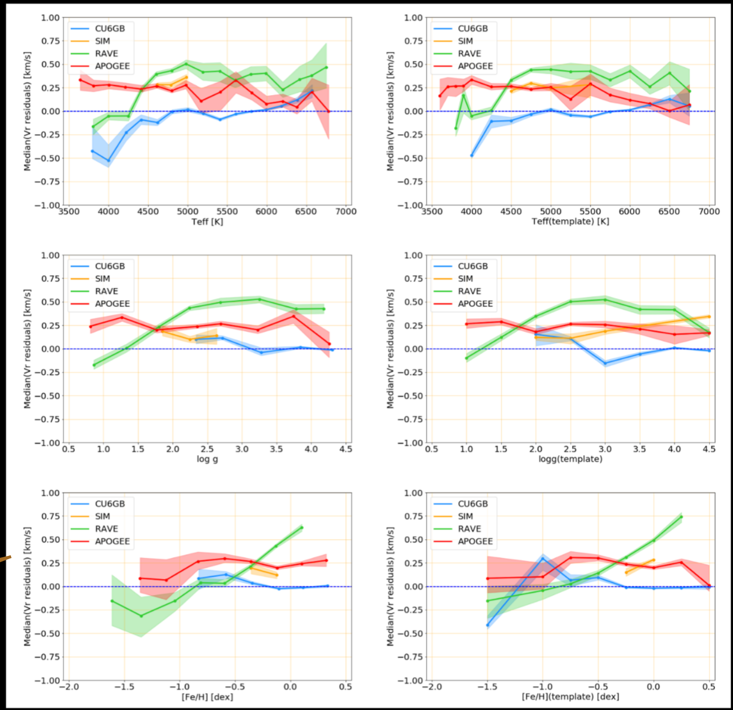
If presenting new, independent estimates, comparison with others has the goals

- understanding the reference system
- presenting the data to the community

But any «judgement» is suspended and the data are not corrected or changed in any way

This is method 1

Different surveys differ in RV ZP by about 0.5 km/s



The Gaia DR2 RVs (Katz et al 2018)

I have many surveys/datasets: homogenization

See Antonella and Angela's talks for current surveys

If you need to combine different datasets, you cannot avoid recalibrating the data for a meaningful interpretation, to report:

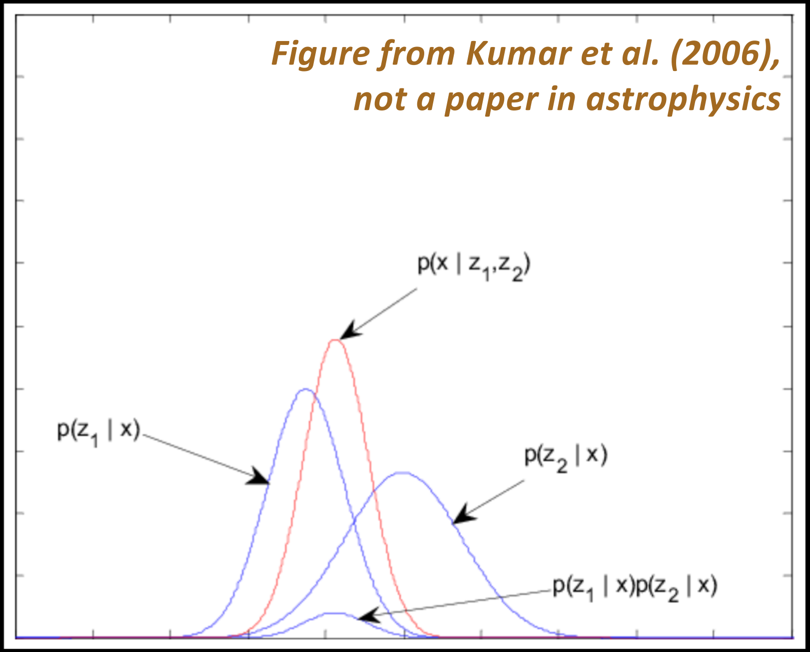
- measurements on compatible systems
- errors on compatible systems

before combining them with, e.g., a weighted mean or median

Later, you can rigidly calibrate the ZP on higher quality measurements

(or do both simultaneously)

This is method 2



The Survey of Surveys

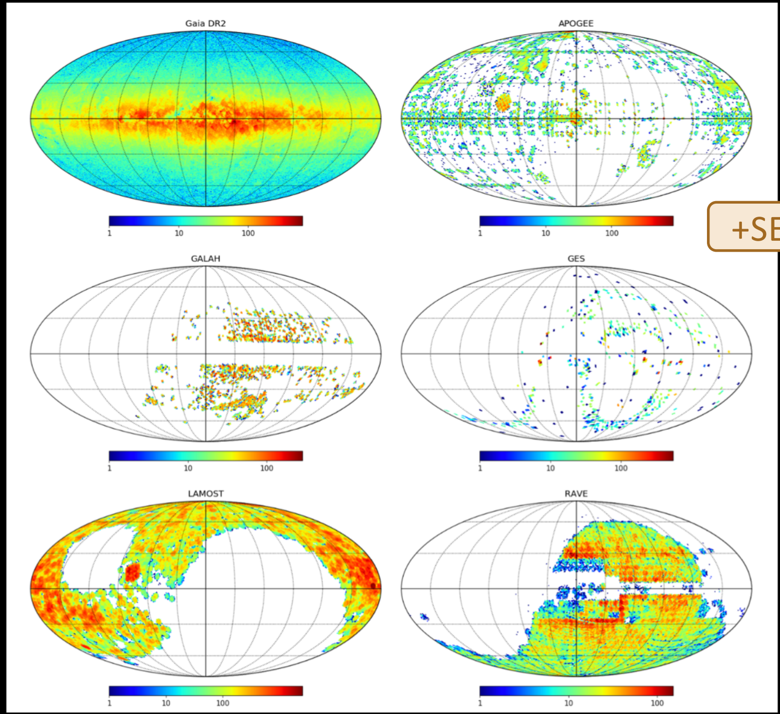
A project to meaningfully combine spectroscopic surveys into one catalogue

- Uses (internal) homogeneization
- Uses (external) recalibration

A pilot paper has been submitted to prove the concept, presenting the largest ever catalogue of 11 million RVs (ZP≃300 m/s, errors ≃1 km/s) and received a favorable referee report

Advertisement: we are looking for collaborators

I will use SoS as a test case in the following



+SEGUE

Survey of Surveys, Tsantaki et al., in press

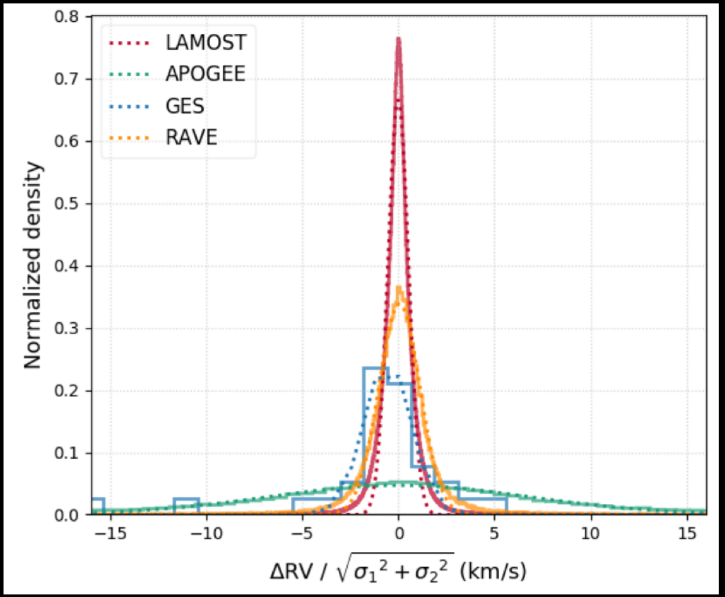
Error recalibration

As we have seen, the error computation is quite complex and thus we could say that

- Uncertainties from different datasets are often not compatible with each other

Methods:

- If there are duplicated measurements, the spreads must be compatible with errors
- If there are no duplicates, one could use the three cornered hat method



Survey of Surveys, Tsantaki et al., in press

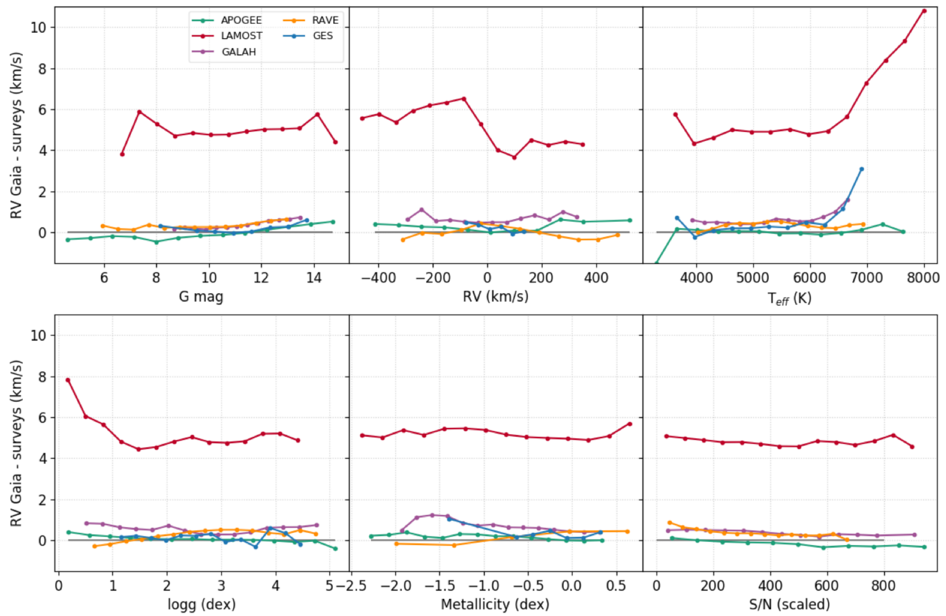
Detrending

Model the differences as a function of relevant parameters, either:

- All possible paired differences
- Differences with the largest or more reliable dataset

The goal is to reduce the number of equations and of terms in those equations

$$\Delta RV = \alpha T_{\rm eff}^2 + \beta T_{\rm eff} + \gamma \log g + \delta [Fe/H] + \epsilon S/N + \zeta RV + \eta$$



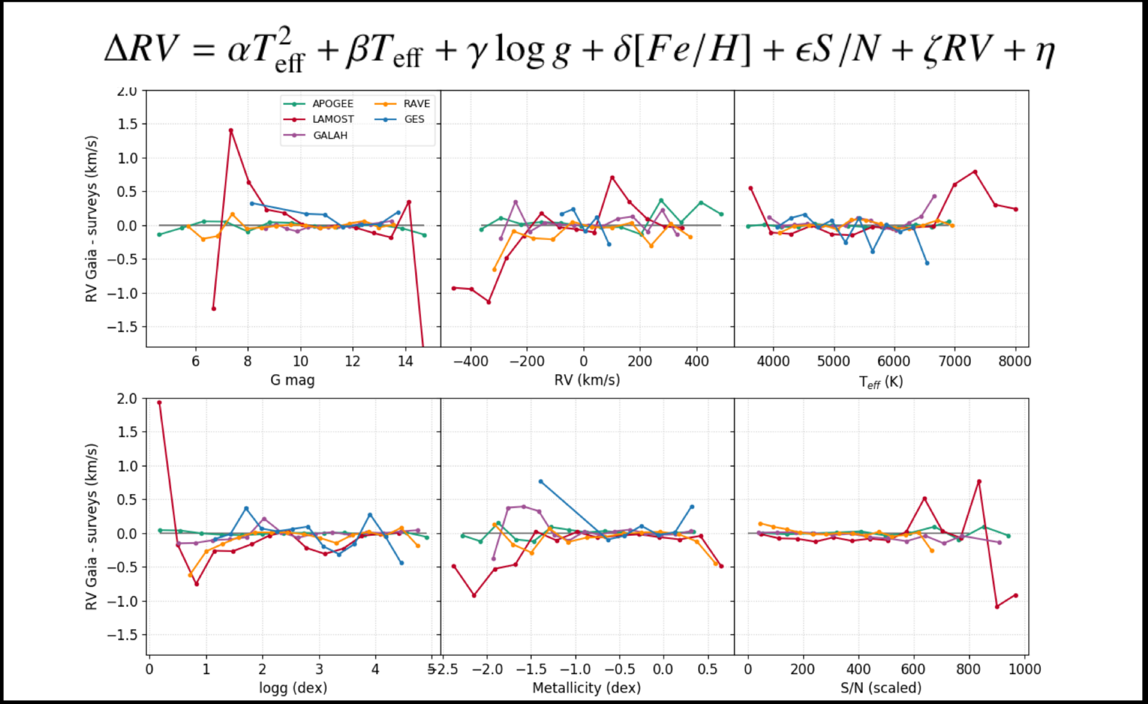
Survey of Surveys, Tsantaki et al., in press

Detrending

Once the trends are eliminated (or reduced), merge all catalogues:

- weighted mean
- weighted median
- choose highest quality
- …

Note how in the figure the scale has changed



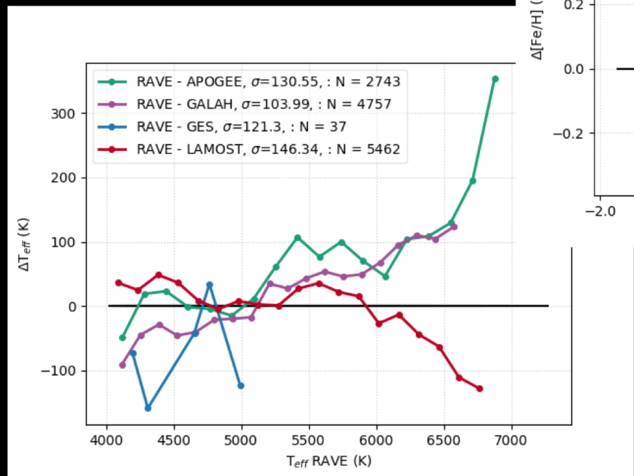
Survey of Surveys, Tsantaki et al., in press

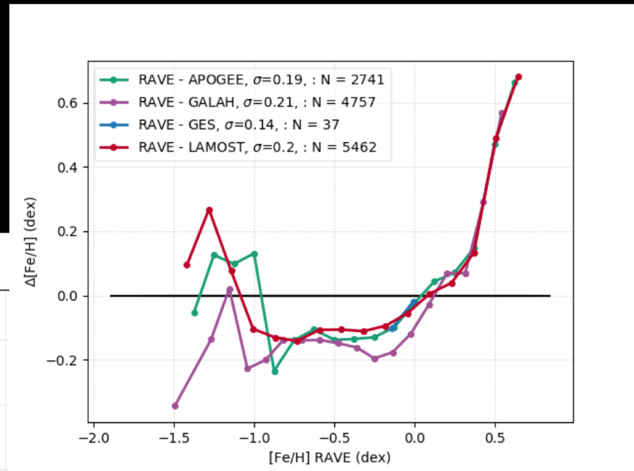
How do major surveys compare?

Based on five major surveys

Before homogenization ZP (and spread):
- RV: ±0.5 km/s (±1-5 km/s)
- Teff: ±50-100 K (±150 K)
- [Fe/H]: ±0.05 dex (±0.2 dex)
- logg: ±0.2 dex (±0.5 dex)

After homogenization ZP:
- RV: ±0.3 km/s (±0.5-2 km/s)





Survey of Surveys, Tsantaki et al., in press

External recalibration

Now all data are in a somewhat arbitrary but homogeneous system

- Does the system compare well with Benchmarks, astroseismology, stellar clusters, other surveys, and higher quality datasets?
- Shall we apply method 1 (understand/explain without correcting)?
- Shall we apply method 2 (recalibrate at least the ZP)?

It depends on the goals – in the Survey of Surveys we recalibrated, but it really depends on the datasets and on the scientific goals

INAF
ISTITUTO NAZIONALE
DI ASTROFISICA

SSDC
SPACE SCIENCE DATA CENTER

Gaia as a reference framework

Experiments such as the Survey of Surveys not possible without Gaia:

- Overlaps with all surveys
- Exquisite measurements
- Sophisticated cross-match software (Marrese et al. 2017, 2019)

Gaia is becoming a reference catalog, anyone observing checks whether their targets are in Gaia