Selection and verification of non-seismic data

Sofia Feltzing
Lund Observatory

What is non-seismic data?

- Stellar properties derived
  - from spectra
  - from photometry
  - from interferometric measurements
  - from astrometric measurements
  - … other?

- This talk will touch on selection of non-seismic measurements and the derivation of the non-seismic data

- Mainly focus on validation – *the action of checking or proving the validity or accuracy of something*

Selecting measurements and non-seismic data

Selecting non-seismic measurements

- I have one issue here
  - first need to decide which non-seismic measurements to use and how to curate them - not all PLATO objects may have the same measurements available
- Would be good to leave the workshop with some resolutions about these two issues

Selecting non-seismic data

- I have a few thoughts here
  - not all data needs to be derived, some already exist
  - need to decide which non-seismic data to derive "in house" and which ones to curate from other sources
- Would be good to leave the workshop with some way forward on this matter

Validation

Validation

- There are few things to consider here
  - choice of data to validate against: ***data***
  - choice of how the validation is carried out: ***process***
  - choice of when we decide that something is good and when something is bad and should be rejected: ***decision***

data – process – decision

Validation

2

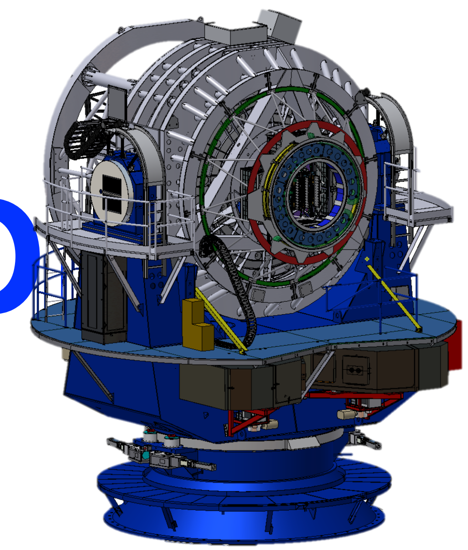
- We need to develop a process that provides the probability distribution function for each non-seismic datum
  - this needs to take into account that there ***might*** be more than on datum available for each object
  - rather than just lumping all of these together it might be better to use our knowledge and understanding and have a decision tree to decide which data to use and which to combine

decision tree – PDFs

Precision vs. accuracy

- Precision and accuracy to be treated separately
- It may not always be possible to assess accuracy for all non-seismic data – e.g., [Fe/H] – instead we can refer to reference values

4MOST as an examp



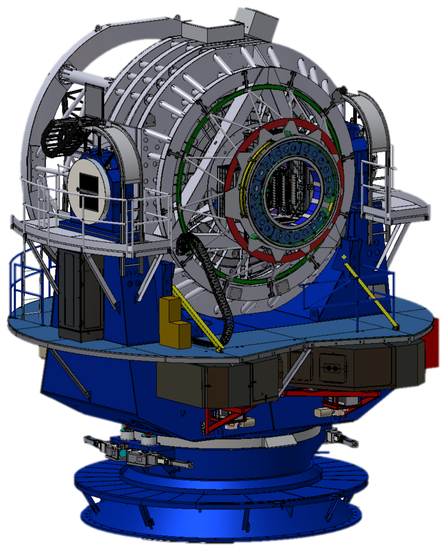
- 4MOST is a facility to go on the ESO VISTA telescope
- 2400 fibres
- Stellar surveys for up to 20 million stars with low (~6000) and high (~18 000) resolution spectra covering all or most of the visible spectrum
- The analysis is of course different from what we do in PLATO but it is a good example of a survey with bulk analysis facing similar problems with validation
- 4MOST clocks about 10 000 individual spectra per night
- Starts observing late 2022

4MOST example

- L2 data-products from the pipeline will include radial velocities and stellar parameters (effective temperature, log g etc) as well as abundances of selected elements.
- The requirements (and goals) are split into precision and accuracy
- Precision is method dependent and as the pipeline will use more than one method to derive the data products the way precision is assessed will be documented for each data release
- The accuracy will be assessed against extant data off well understood stars. The data used will be documented for each data release

4MOST example

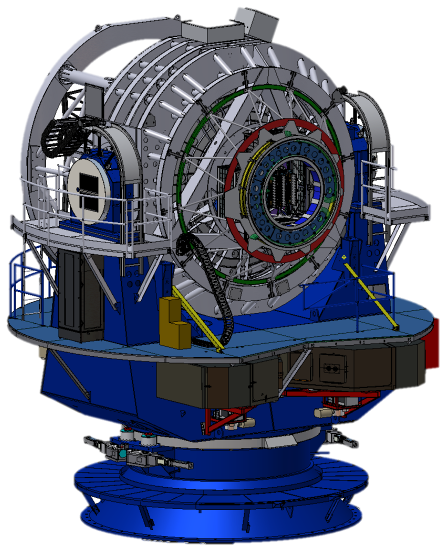
- This approach gives flexibility and possibilities for future developments - if it is built in from start we have much to gain
- The decision is further to for every data release have
  - a frozen pipeline version
  - publish the list of reference values



4MOST example

- This approach gives flexibility and possibilities for future developments - if it is built in from start we have much to gain
- The decision is further to for every data release have
  - a frozen pipeline version
  - publish the list of reference values

Call for LoI expected 25 November



Validation objects

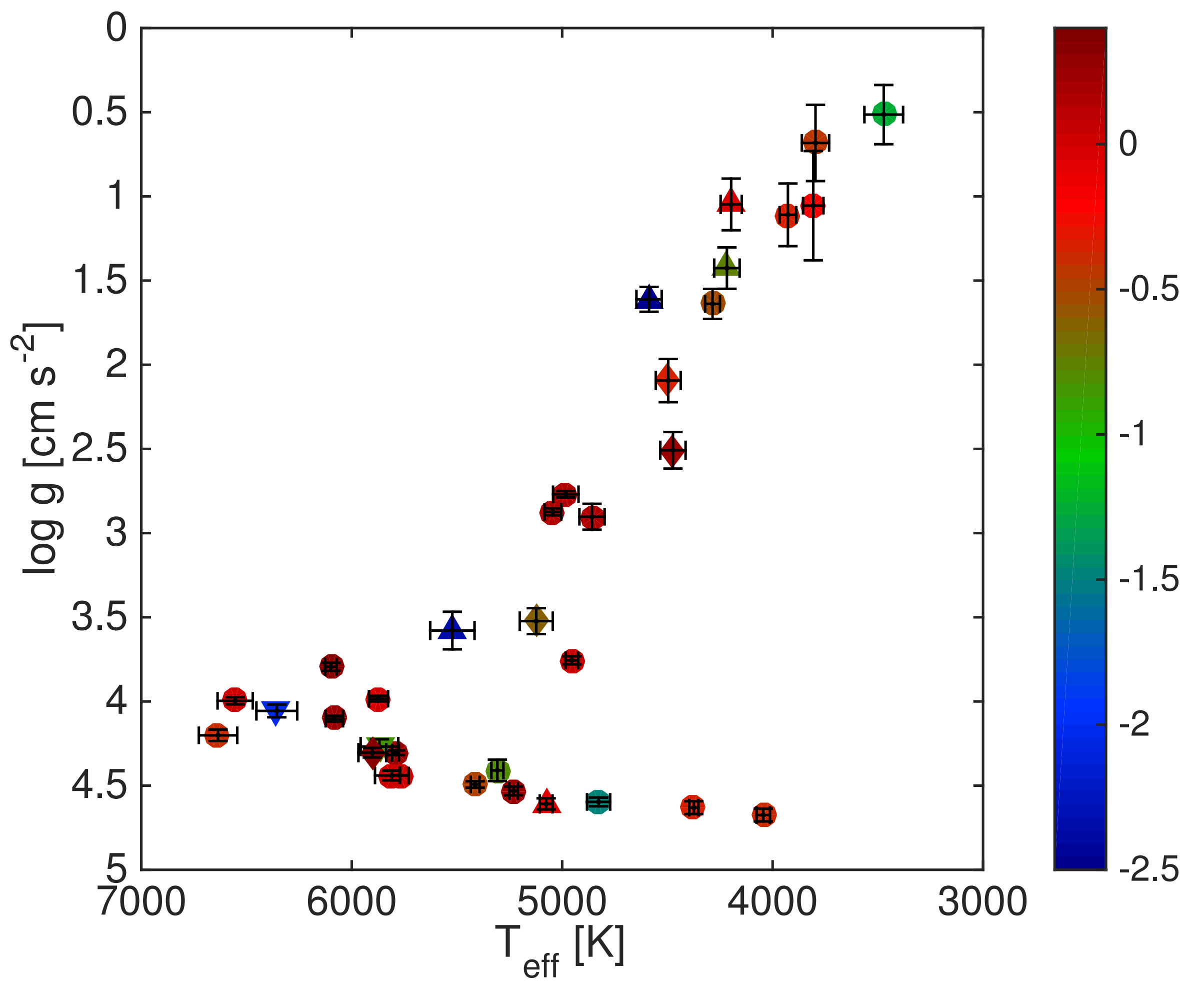
- Common choices of reference objects for verification and validation of data and methods include
  - Gaia benchmark stars
  - eclipsing binaries
  - (open) stellar clusters

- But there could be others, such as
  - all Gaia objects with a certain datum available
  - …

- But if they are used for validation are pipelines then allowed to use them to, e.g., train their methods?

Gaia benchmark stars

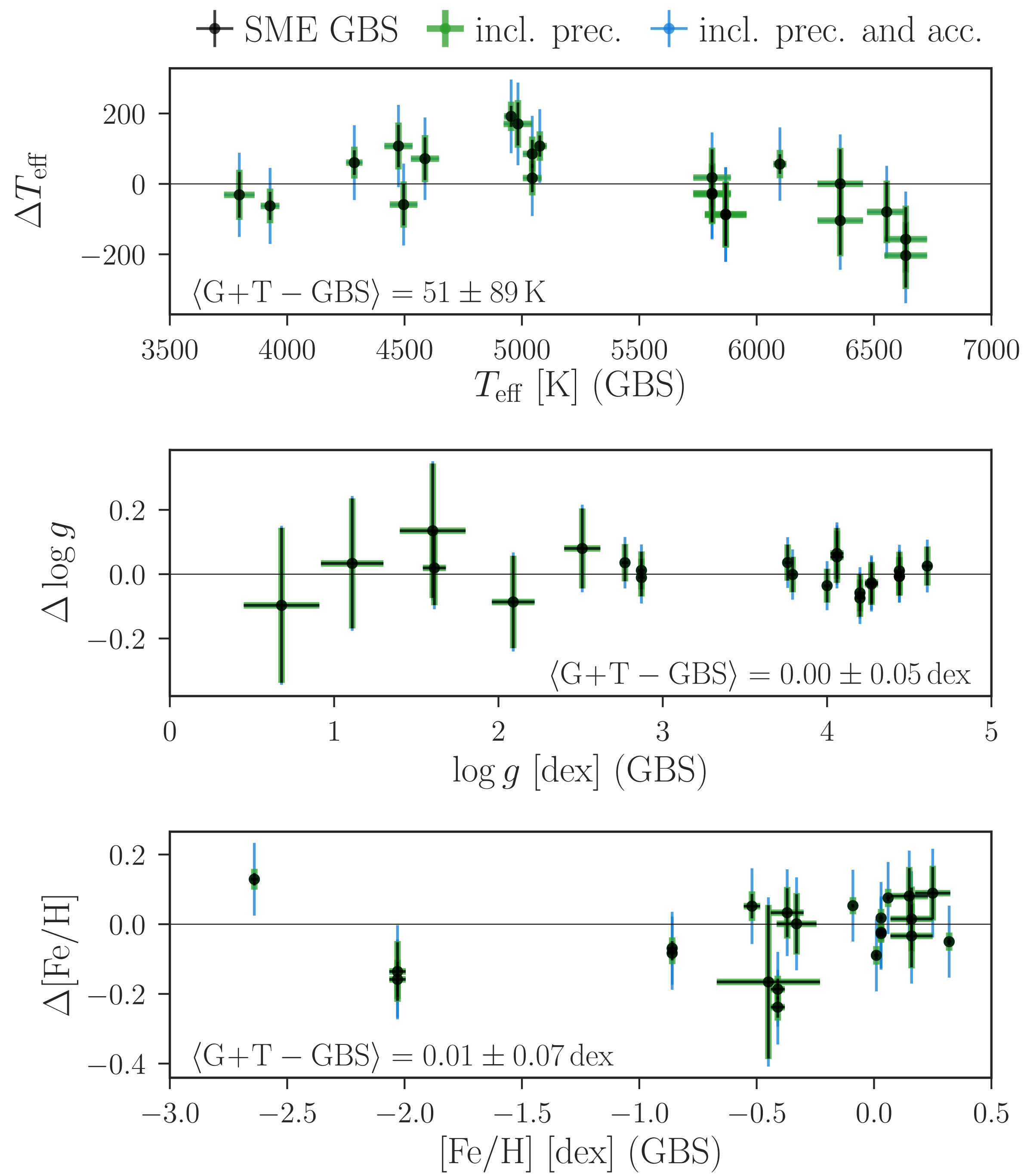
- Sample designed to anchor Gaia astrophysical parameters (Apsis, Bailer-Jones et al. 2013)
- Accurate knowledge of radii and flux —> T_{eff} and log *g*
- Reference values for [Fe/H] and various elements are defined



Heiter et al. (2015) A&A 582 A49
Jofré et al. (2015) A&A 582 A81

Assessing accuracy

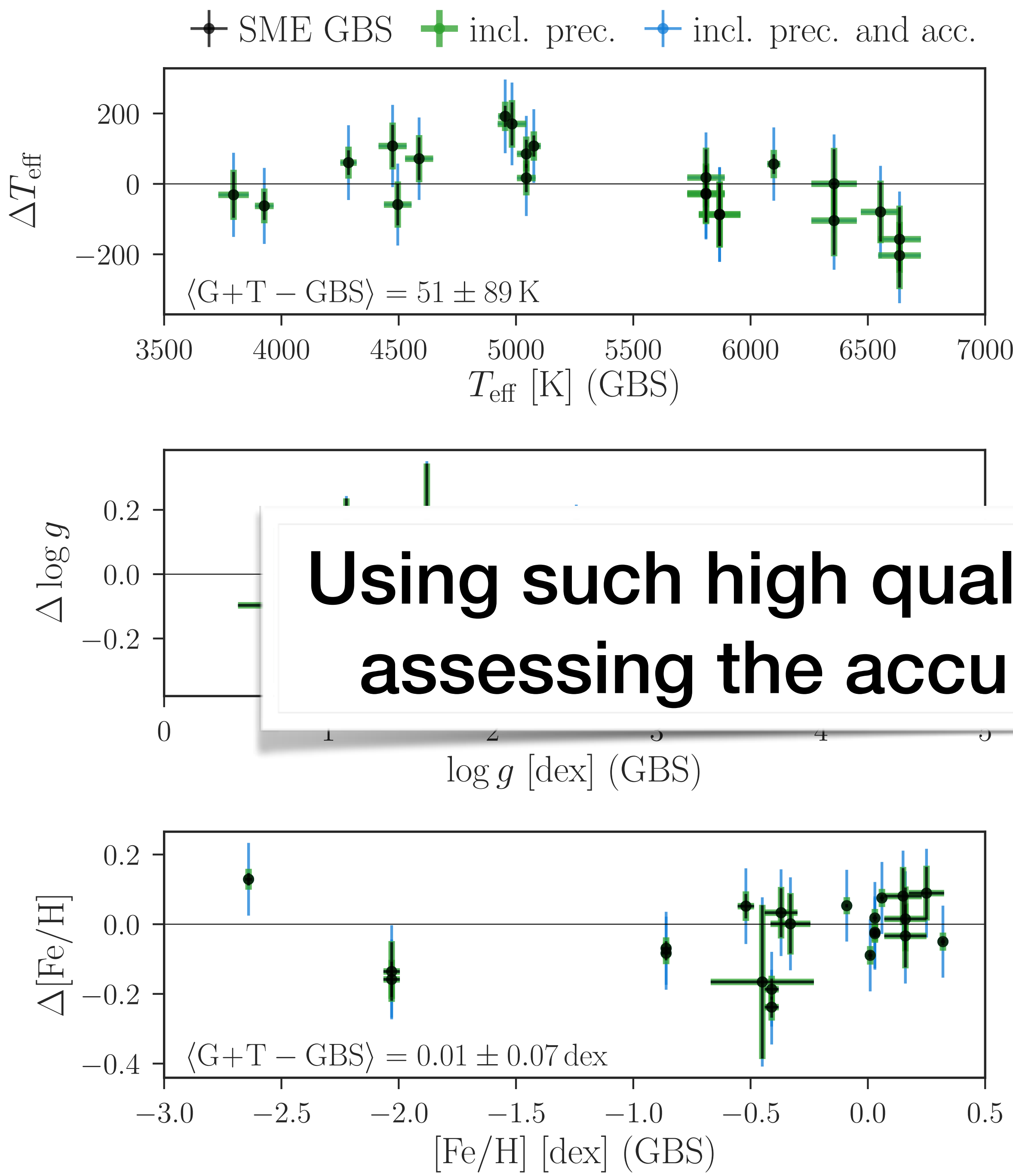
GALAH example



- Use Gaia benchmark stars do decide when and if the *analysis method* provides reliable values
- Assessed against reference values
- They decided this was good enough for their purposes

Buder et al. (2019) A&A 624 A19

GALAH example



- Use Gaia benchmark stars do decide when and if the *analysis method* provides reliable values
- ...nce
- They decided this was good enough for their purposes

Using such high quality data sets is one way of assessing the accuracy of the analysis itself

Buder et al. (2019) A&A 624 A19

Assessing accuracy

Gaia data

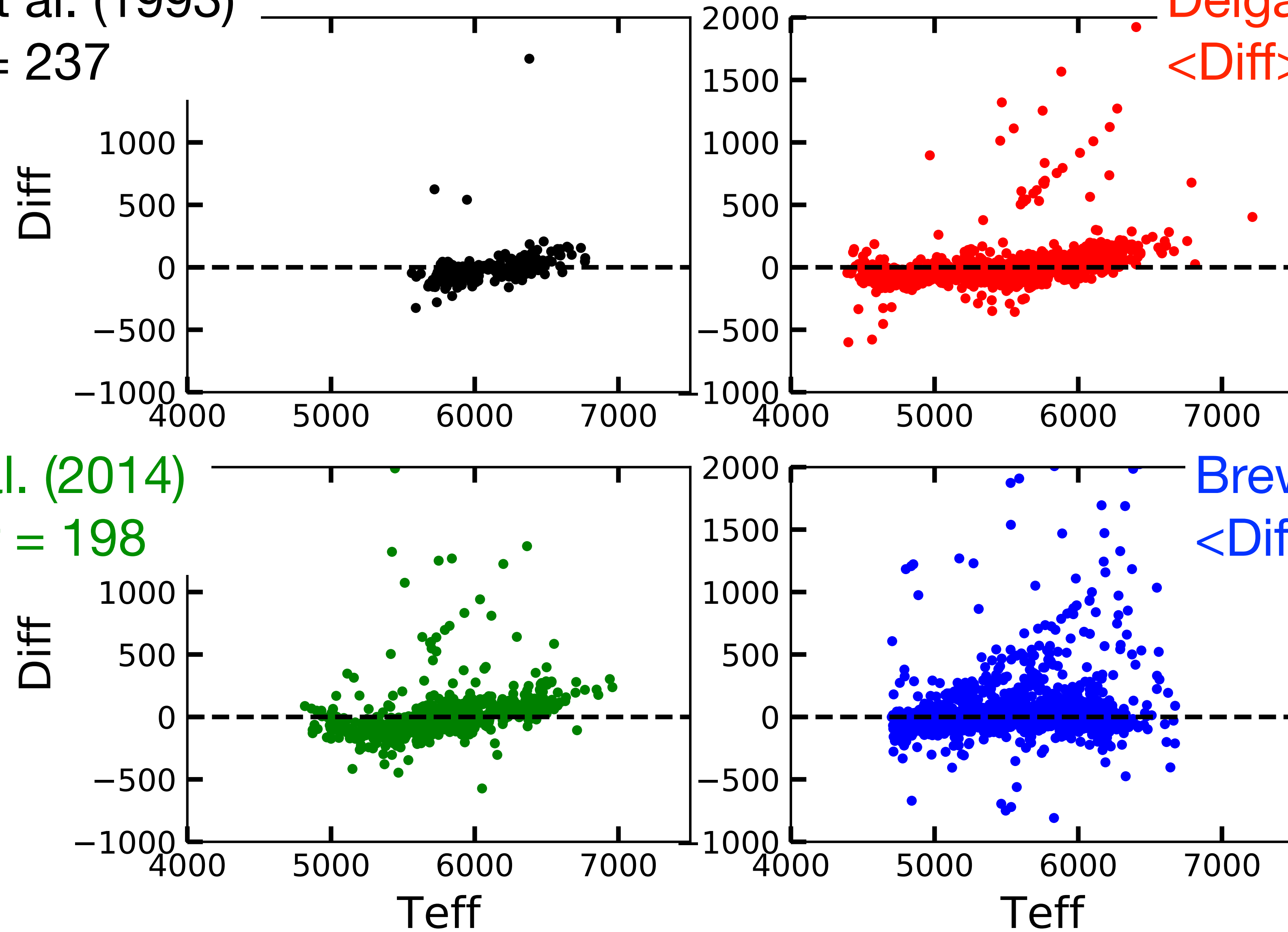
- Gaia allows you for essentially the full sample to assess the accuracy of the final result (but not same as w. BSs)

Edvardsson et al. (1993)
<Diff>=7.3 σ = 237

Delgado-Mena et al. (2014)
<Diff>=17 σ = 198

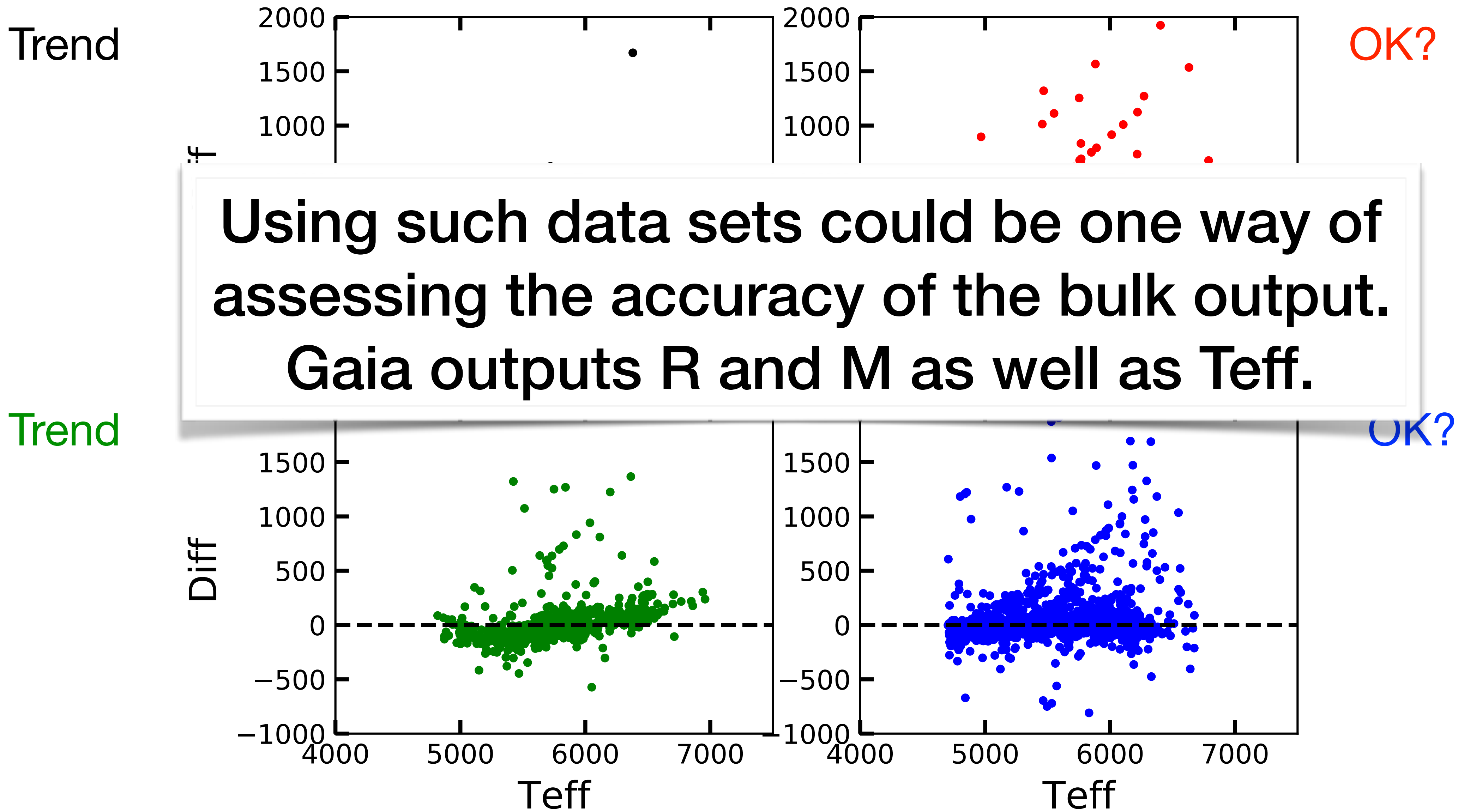
Bensby et al. (2014)
<Diff>=17 σ = 198

Brewer et al. (2016)
<Diff>=49 σ = 278



Gaia data

- Other examples could be in house IRFM Teff



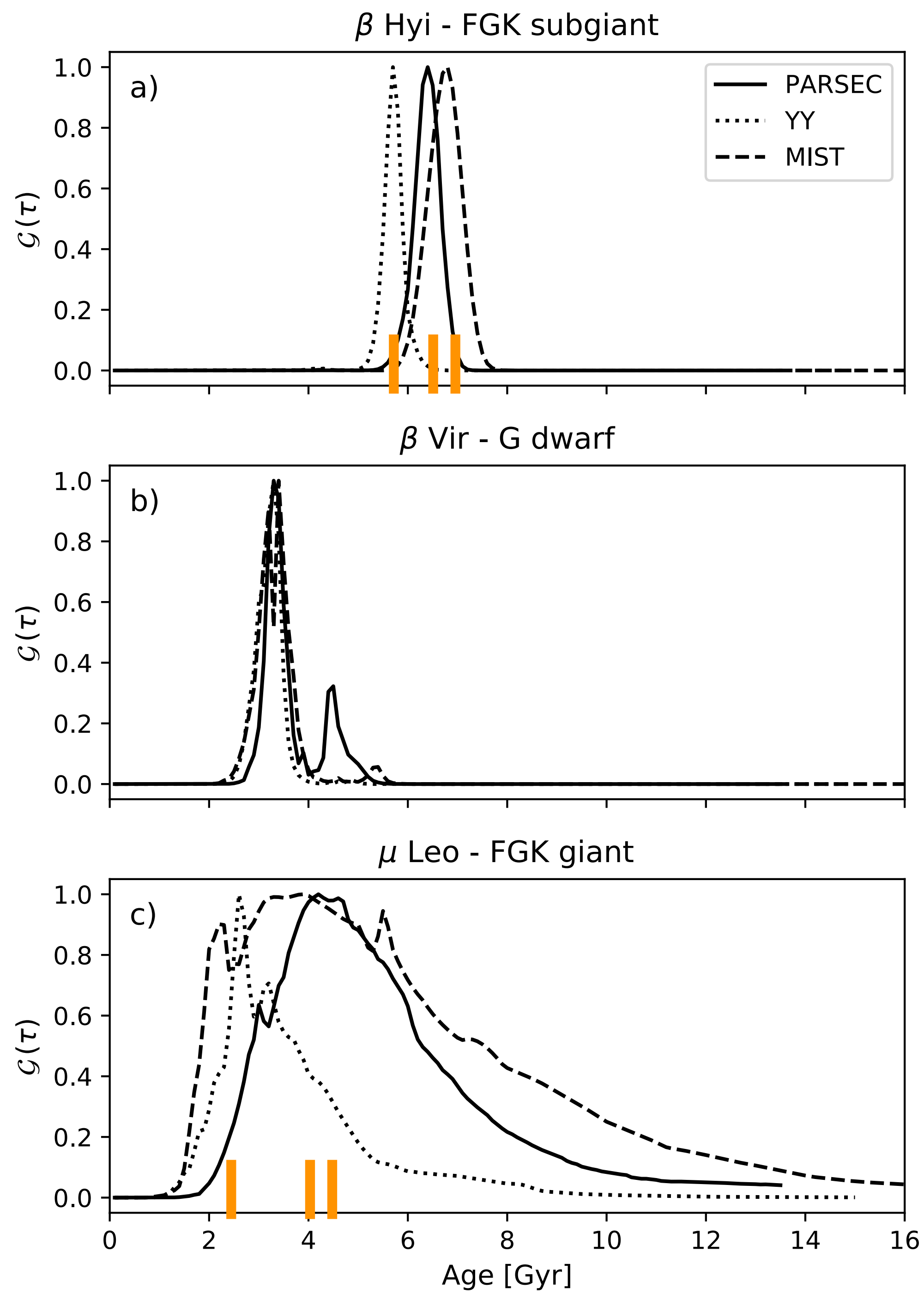
Using such data sets could be one way of assessing the accuracy of the bulk output. Gaia outputs R and M as well as Teff.

Combining results

When and if we have more than one datum

Combining data

Example I



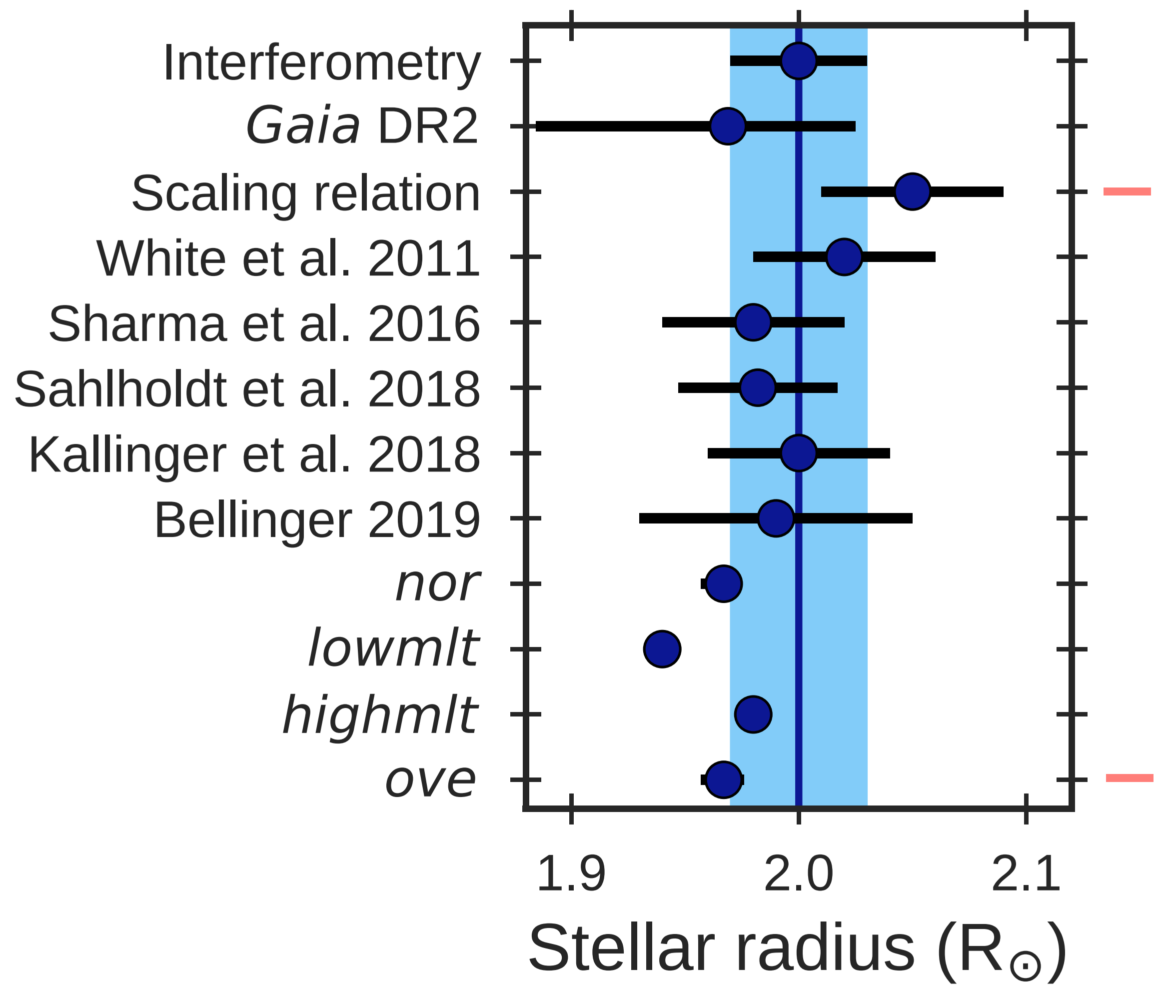
Peak values

- A combined PDF weighted with the "1σ" error might seem natural if we have no better understanding
- But, I guess, most in this room would put more weight on the MIST isochrones (?)
- Perhaps we would even decide just to use the MIST results?

Sahlholdt et al. (2019) MNRAS 482 895

Combining data

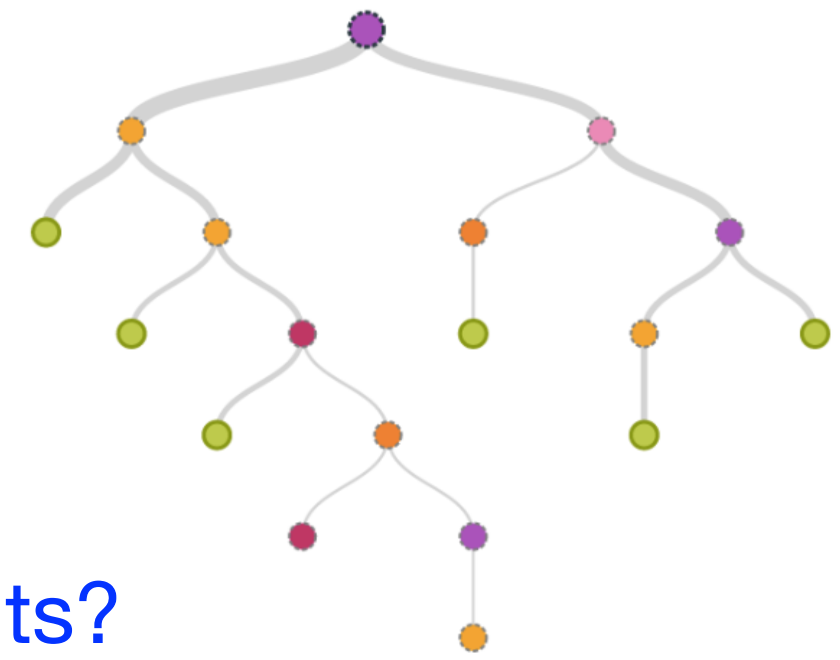
Example II



- A combined PDF weighted with the "1σ" error might seem natural if we have no more better understanding
- But, I guess, most in this room would say that interferometry is the best way to derive R?
- Should we then only use that datum and its associated error distribution?

Stockholm et al. (2019) MNRAS 489 928

Some real ??



- How do we best combine method results?
- Which ones do we keep and which ones do we throw out? Keeping everything may not be the best (even with very low weights)
- How do we assess which one is closest to the truth?
- Or should we just lump them all together as more is always better?

Many methods=good?

- Deriving data from measurements is a task that can result in many different values even when one would think they should give the same answers.

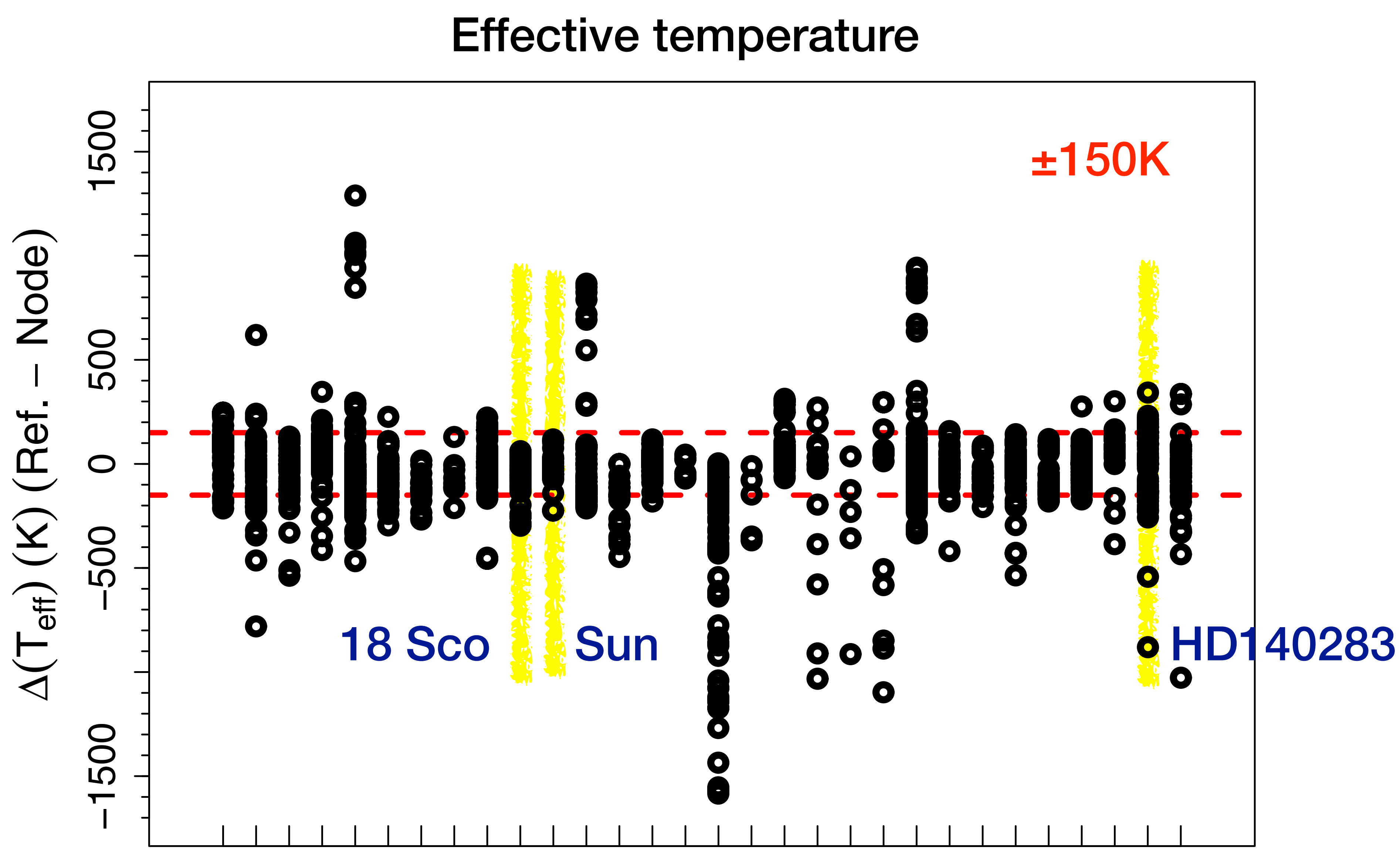
- Here I will give one example
  - Gaia-ESO Survey individual analysis

- Lets look at everyones best effort to analyse the Gaia Benchmark stars

- Same model atmospheres, same atomic data - different methods or different implementations of the same method

Spectroscopic analysis only

Gaia-ESO Survey



- It is difficult to even with the same input physics achieve a precise result amongst 10+ analysis methods

log *g* - T_{eff} iterations

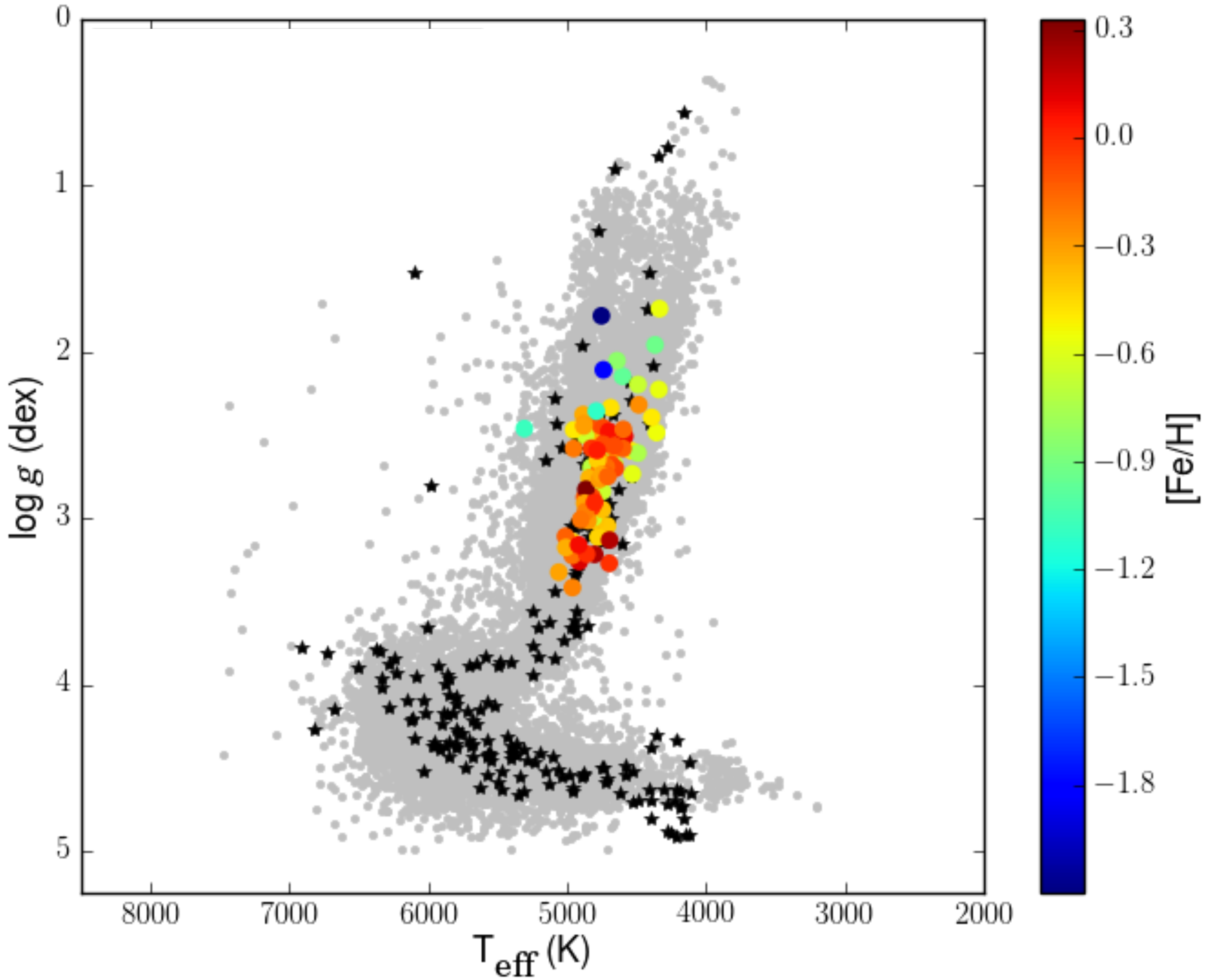
➡ But we can fix this by adding a log *g* making use of asteroseismology - is that true? Let's take a look how it works in practise in spectroscopic survey mode

$$\log g = \log g_\odot + \log(\nu_{\rm max}/\nu_{\rm max,\odot}) + \frac{1}{2}\log(T_{\rm eff}/T_{\rm eff,\odot})$$

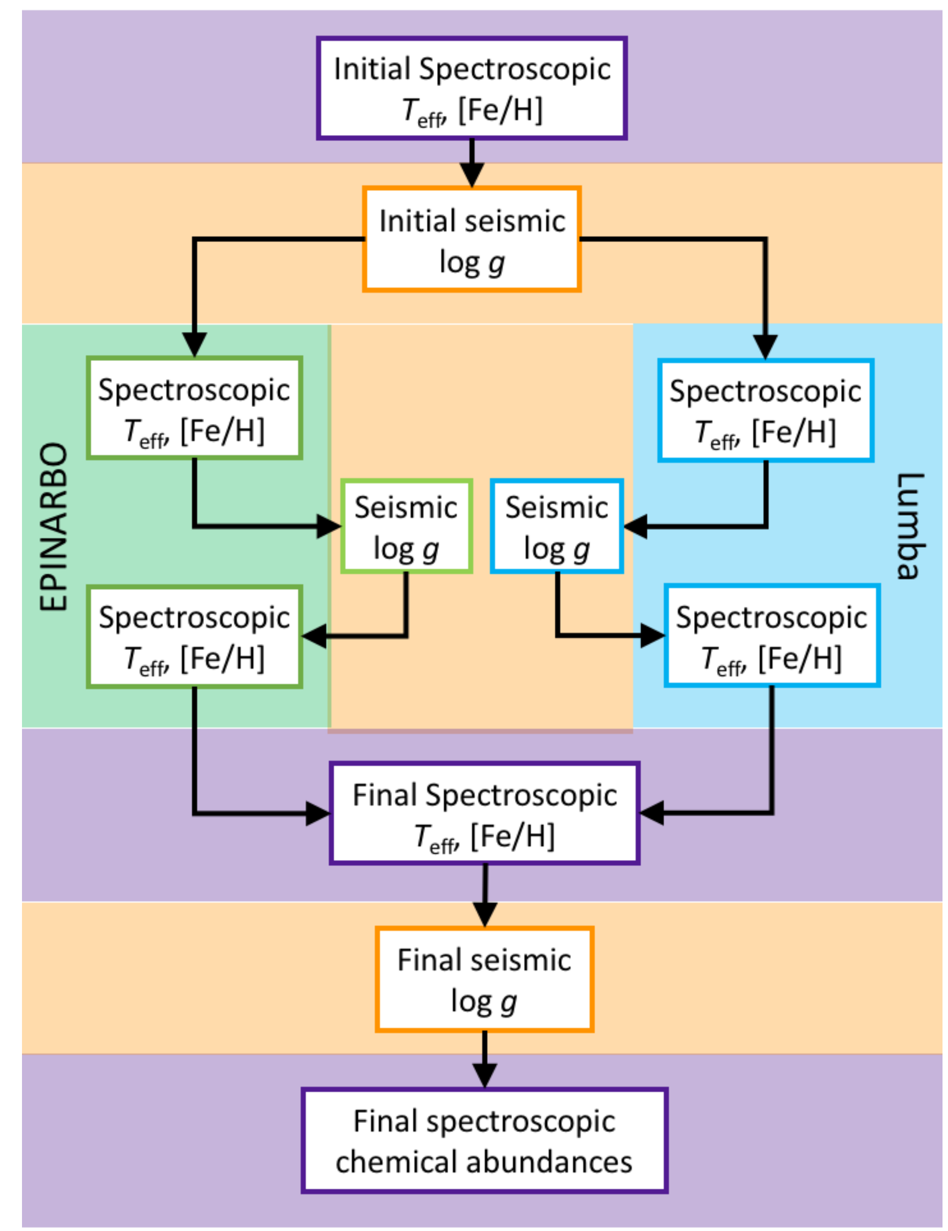
- This means that the determination of log *g* is very weakly dependent on T_{eff}, which in turn means that not much will happen when the analysis tries to iterate to a better estimation of the data
- Here I will give one example
  - K2 + *Gaia*-ESO Survey - stellar parameters

Gaia-ESO & K2

Stars analysed (giants only)



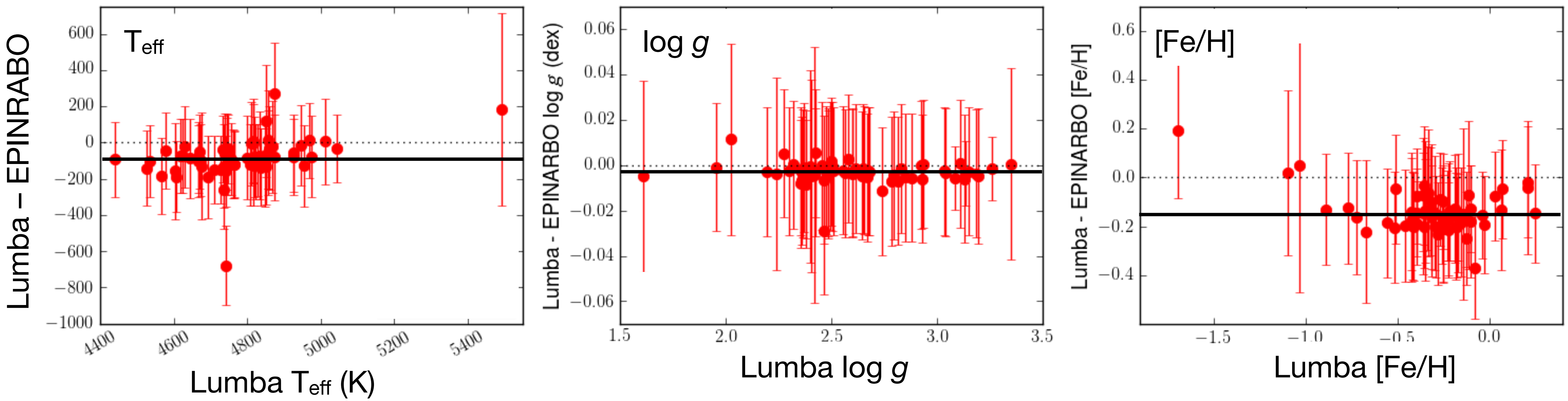
Process



Worley et al. (submitted)

Summary of results

- Two teams analysed the spectra, one using EWs, the other using spectral synthesis (plots are synt-EWs)



–82 ± 63 K **–0.003 ± 0.003** **–0.147 ± 0.070**

Median ± MAD

- Should PLATO have one or several derivations (internally?) of the stellar parameters?

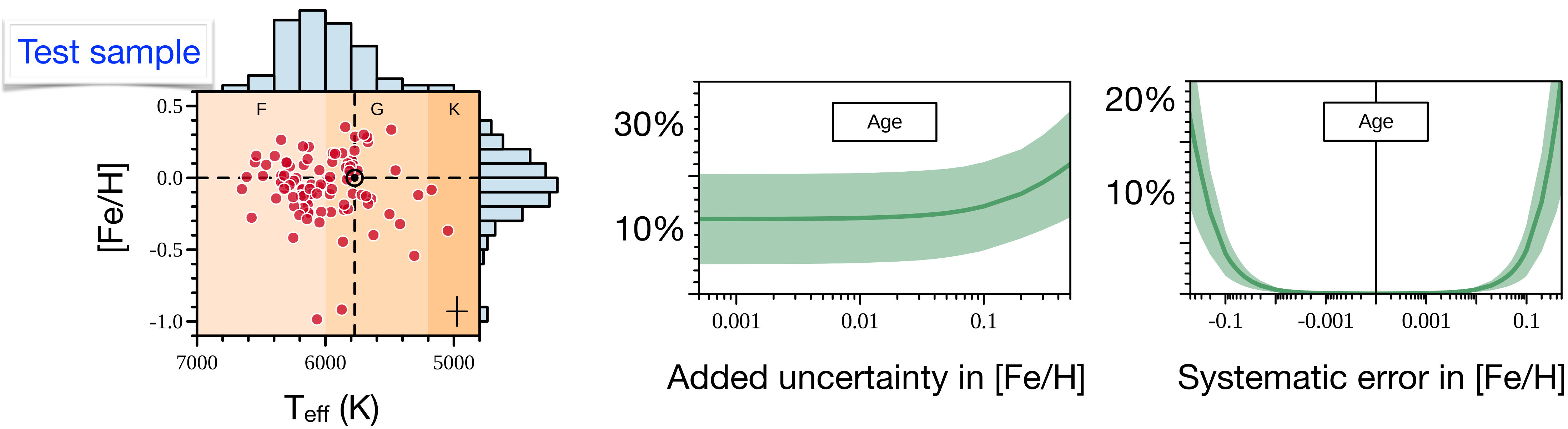
Worley et al. (submitted)

An unrelated question

Do we need [Fe/H]?

precise

- Does [Fe/H] need to be precis and/or accurate?
- Example with SPI (Bellinger et al. 2016)



- When you look at R, M and ρ (for this sample) the answer appears to be no
- But if we want an age with 10% error [Fe/H] will (at least) need to be <u>precise</u>

Bellinger et al. (2019) A&A <u>622</u> A130
Bellinger et al. (2016) ApJ <u>830</u> 31

Validation

Validation 0.1

- Essential that benchmark stars are observed by PLATO and analysed <u>in the same way</u> as the main sample(s)

- The analysis should preferably be blind such that it is not possible to identify what is a benchmark star and what is not for the analysers

- The accuracy of the results can then be validated against the benchmark reference values straightforwardly

- This should result in assigning a global error PDF to the results taking into account precision as defined by the pipeline(s)

Validation 0.2

- Verification of bulk products against extant large data-sets or against large data-sets created “in house”

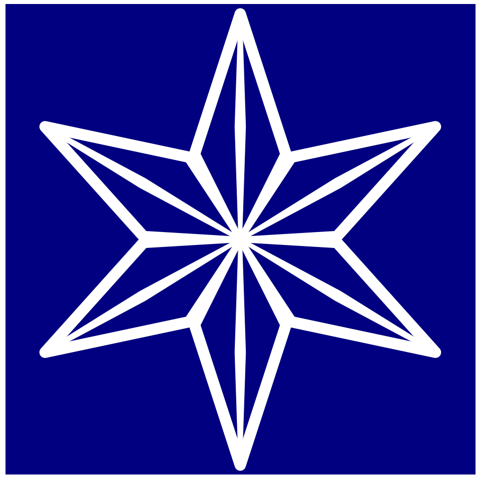
- Examples include
  - data from Gaia
  - Teff IRFM
  - stellar parameters derived using only classical methods (allows to understand how the PLATO results relate to eg those from large spectroscopic surveys)s

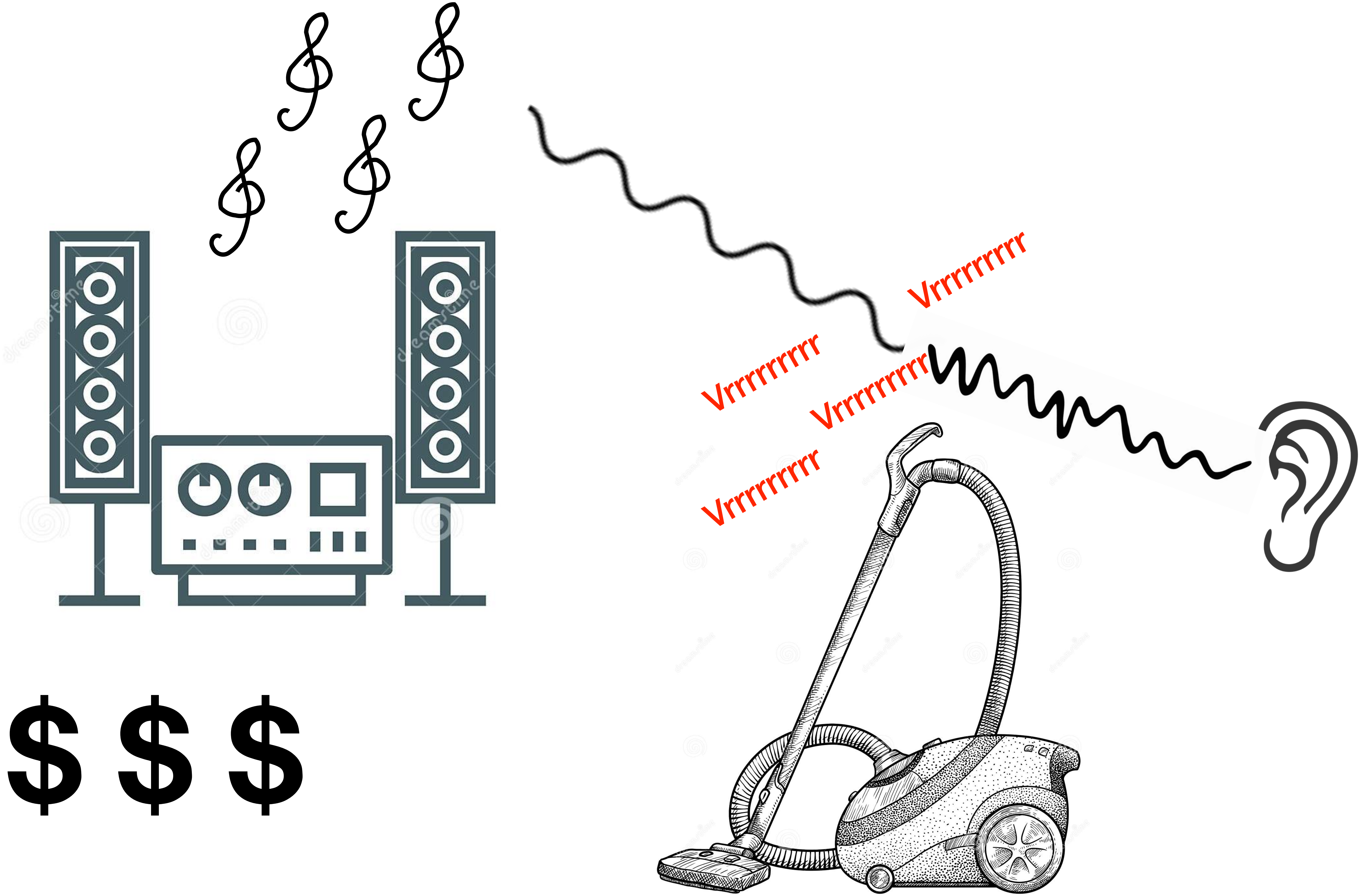
Next steps

1. *Reference objects* - defined per data release. Work on more reference objects on-going
2. Propagation of PDFs between different types of analysis needs be defined - would suggest a small team with broad interests and skill sets is given this task (it could live in WP125 200) - for example are pipelines able to deal with non-symmetric PDF (and is it necessary to be able to do so)?
3. Process of validation needs to be defined and documented – would suggest some over-arching working group for this as it spans many WP12X - but led from WP125
4. Definition of what data should be flagged and/or rejected needs to be specified – many may have an opinion on this; how do we move forward?

List of things to do after meeting

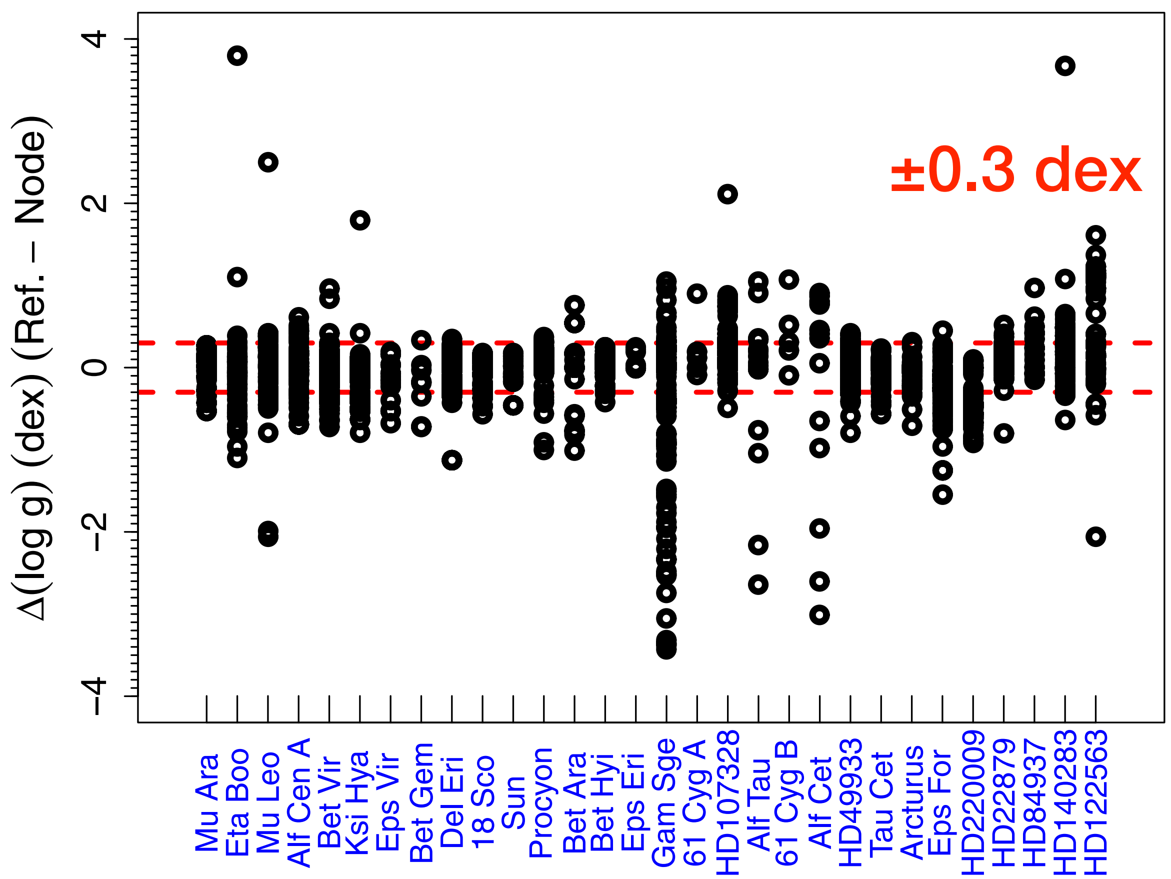
- Decide which non-seismic measurements to use and how to curate them - not all PLATO objects may have the same measurements available
- Need to decide which non-seismic data to derive “in house” and which ones to curate from other sources
- Precision and accuracy to be treated separately
- Bulk validation with “in house” or extant data - OK?
- How do we best combine method results? Which ones do we keep and which ones do we throw out? Or should we just lump them all together as more is always better?
- How to define Teff if you use scaling relation to define log g?
- PLATO analysis should be blind
- Should PLATO have one or several derivations (internally?) of the stellar parameters?
- How to use validation results to determine an “error” - would like input on this
- Propagation of PDFs between different types of analysis needs be defined - would suggest a small team with broad interests and skill sets is given this task (it could live in WP125 200) - for example are pipelines able to deal with non-symmetric PDF (and is it necessary to be able to do so)?
- Process of validation needs to be defined and documented – would suggest some over-arching working group for this as it spans many WP12X - but led from WP125
- Definition of what data should be flagged and/or rejected needs to be specified – many may have an opinion on this; how do we move forward?



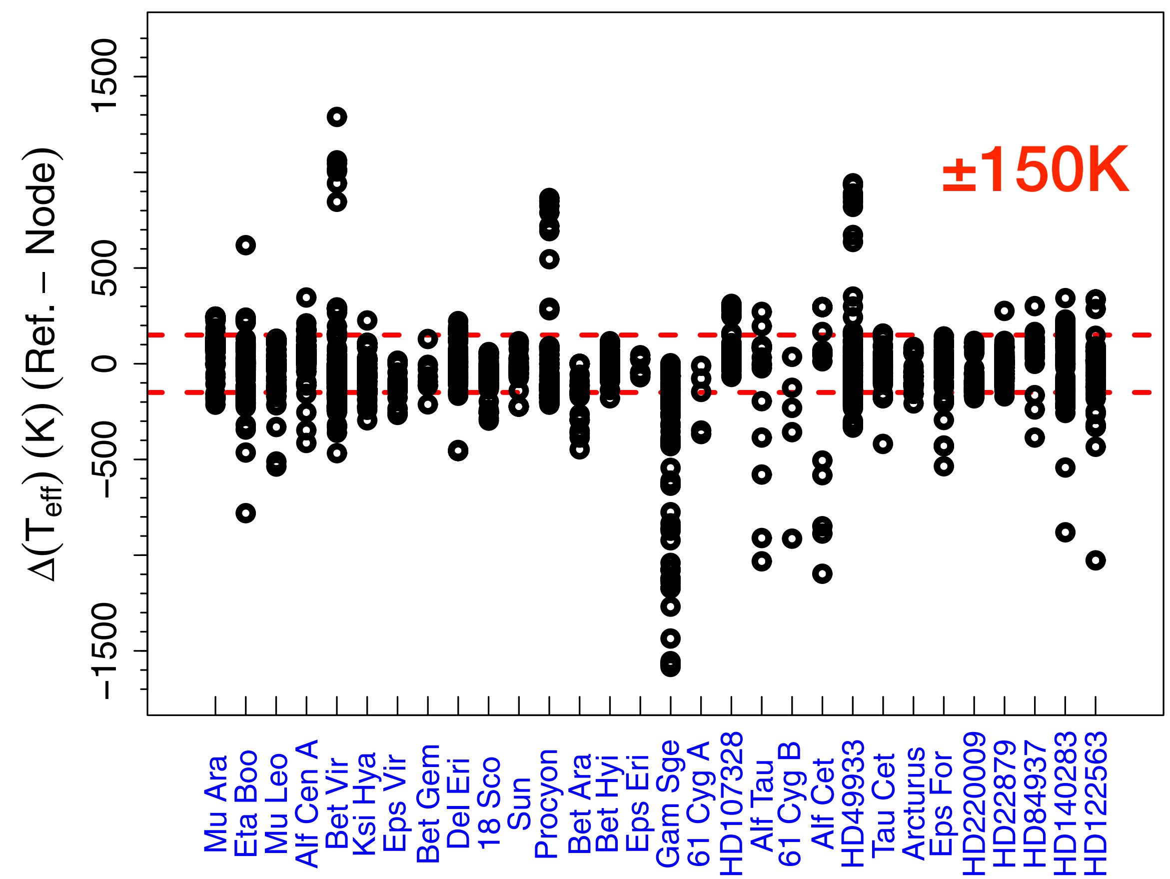

Vrrrrrrrr
Vrrrrrrrr
Vrrrrrrrrrr
Vrrrrrrrr
$ $ $

Gaia-ESO Survey

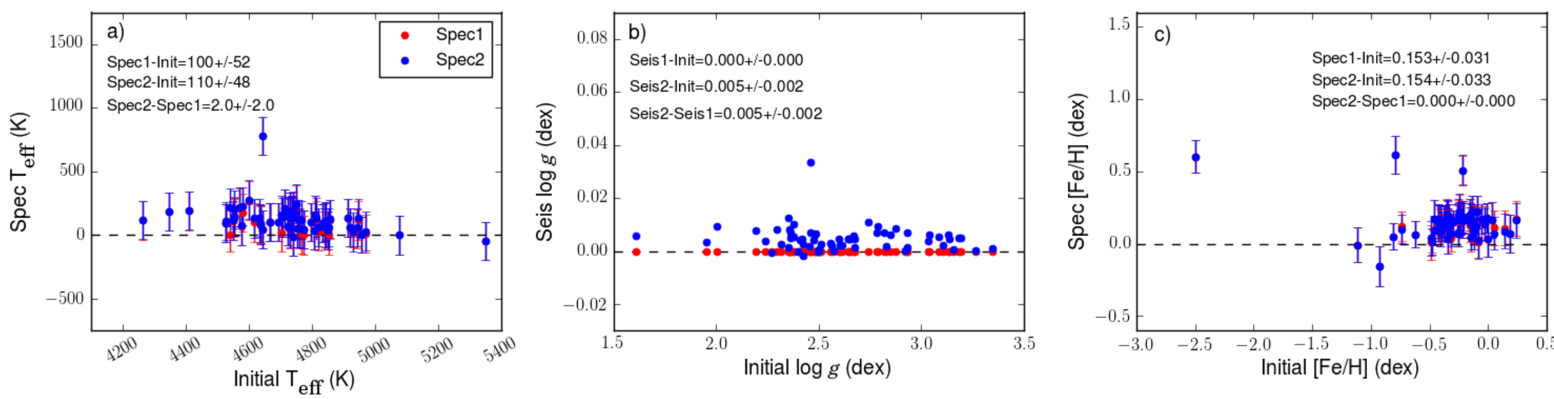
Surface gravity

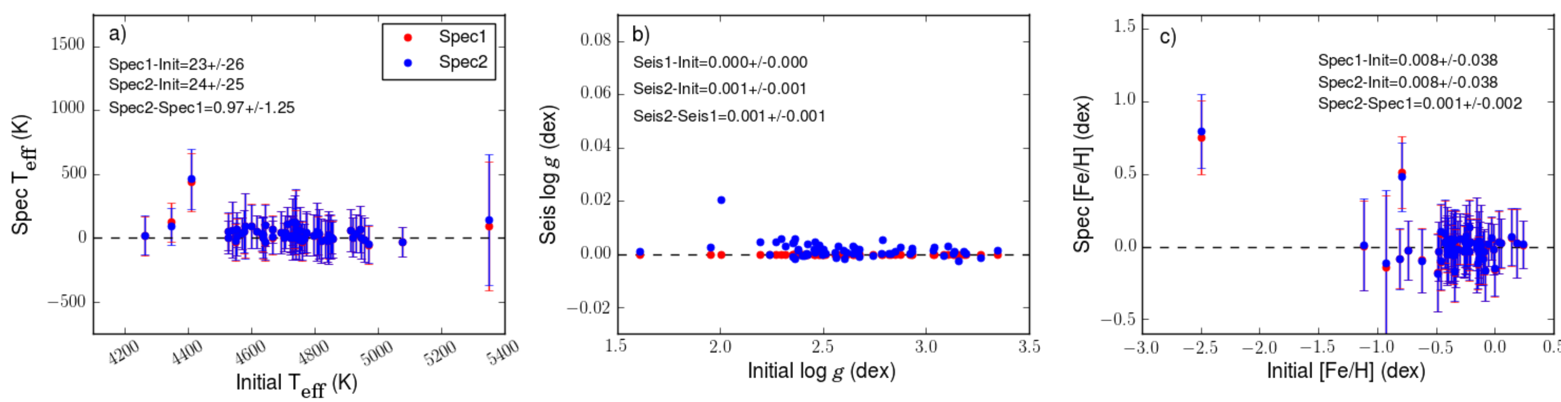


Effective temperature



- Lets start with everyones best effort to analyse the Gaia Benchmark stars
- Same model atmospheres, same atomic data - different methods or different implementations of the same method





We point to the discussion of Morel et al. (2014) that a change of 100 K in $T_{\rm eff}$ only affects $\log g$ by about 0.005 dex. Therefore, significant improvement in $\log g$ by, say, a change of the order of 0.1 dex, requires a change in $T_{\rm eff}$ that is much larger than typical uncertainties of $T_{\rm eff}$. Nonetheless, by fixing $\log g$ by

Worley et al. (submitted)