Machine Learning *techniques* *for* Stellar Spectroscopy

Part 2

Lorenzo Spina

INAF - Astronomical Observatory of Padua (Italy)

spina.astro@gmail.com

@lorenzospina

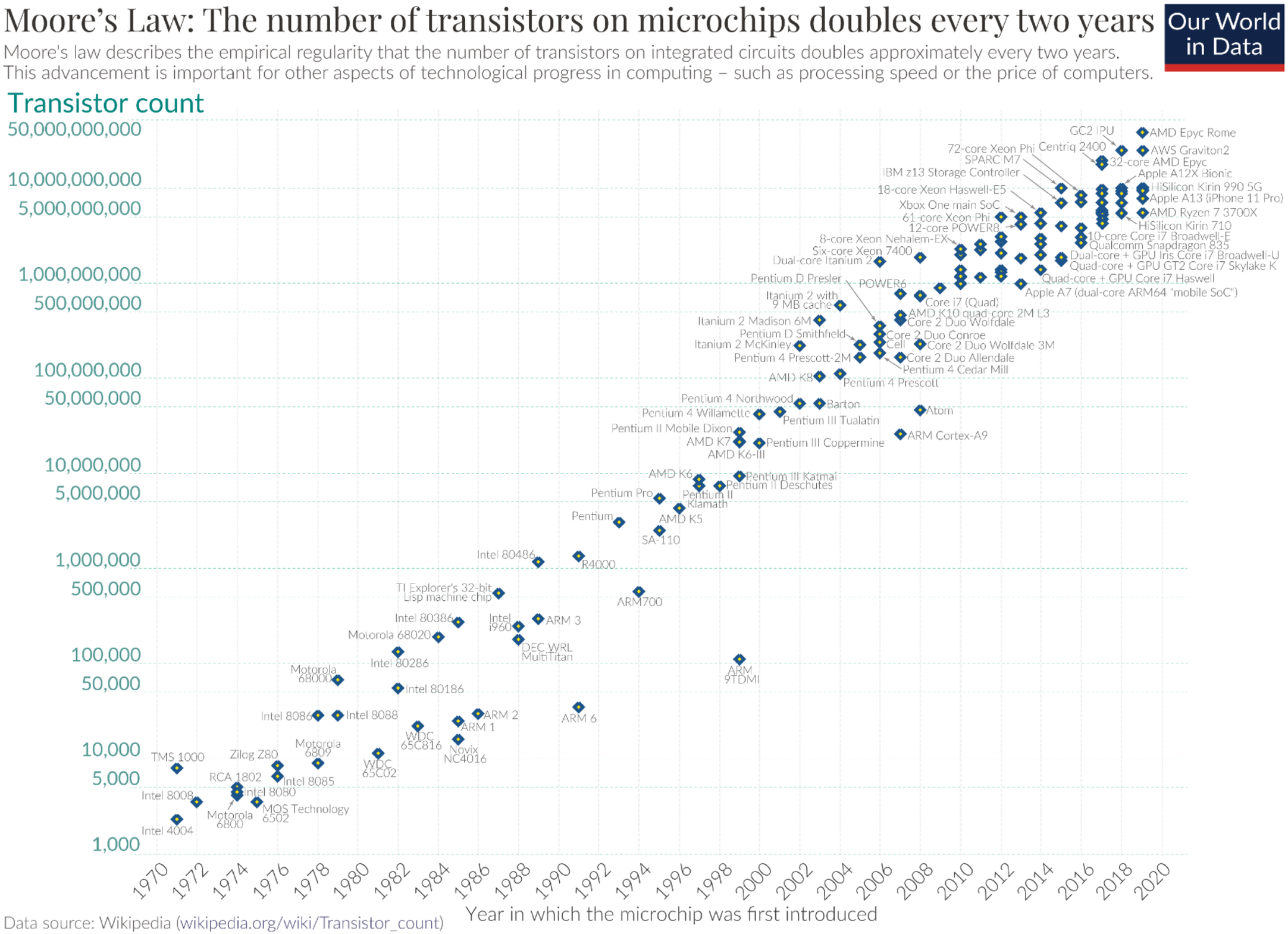
Part 1

- What is Machine Learning?
- How does it work?
- Brief introduction to Artificial Neuronal Networks.

Part 2

- The Cannon
- The Payne
- Issues of using ML for stellar spectroscopy and possible solutions

Machine Learning: why should we care?



Machine Learning: why should we care?

Tsunami of data (especially with the advent of new telescopes and missions).

Machine Learning: why should we care?

Tsunami of data (especially with the advent of new telescopes and missions).

Growing progress in available algorithms and research.

Machine Learning: why should we care?

Tsunami of data (especially with the advent of new telescopes and missions).

Growing progress in available algorithms and research.

Increasing support from industries.

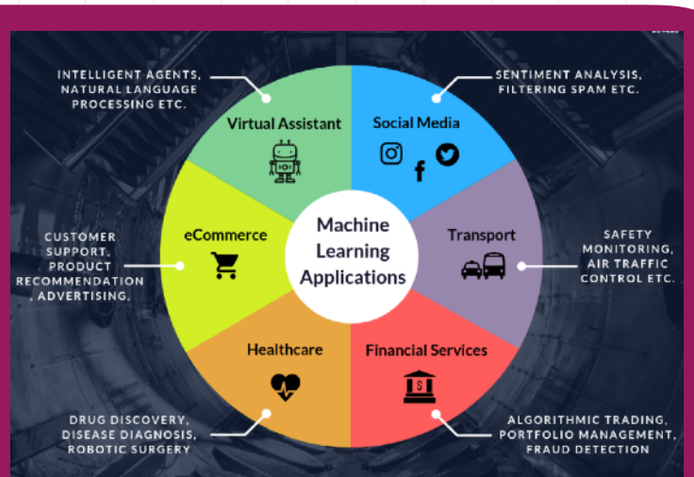
Machine Learning: why should we care?

Tsunami of data (especially with the advent of new telescopes and missions).

Growing progress in available algorithms and research.

Increasing support from industries.

Machine learning is fully transferable to a broad range of fields.



In our dreams...

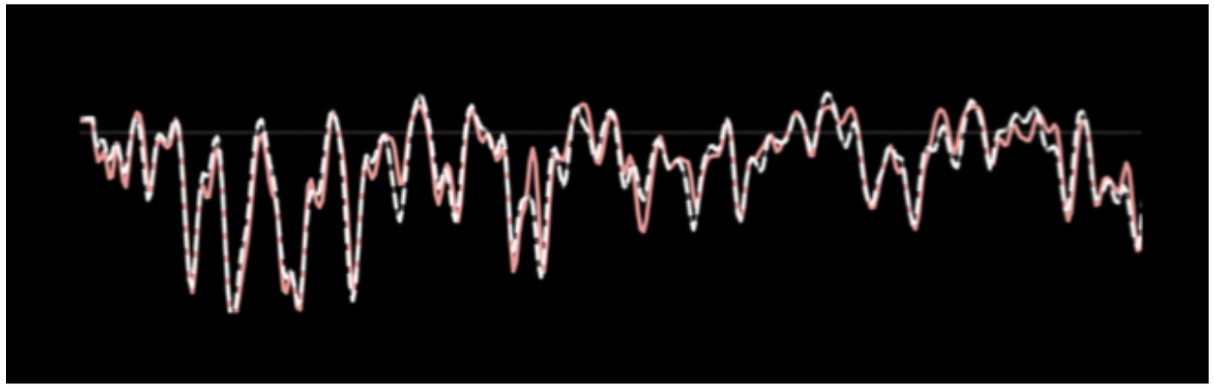
Stellar labels
(T_{eff}, log g, [Fe/H], [X/Fe])

↓

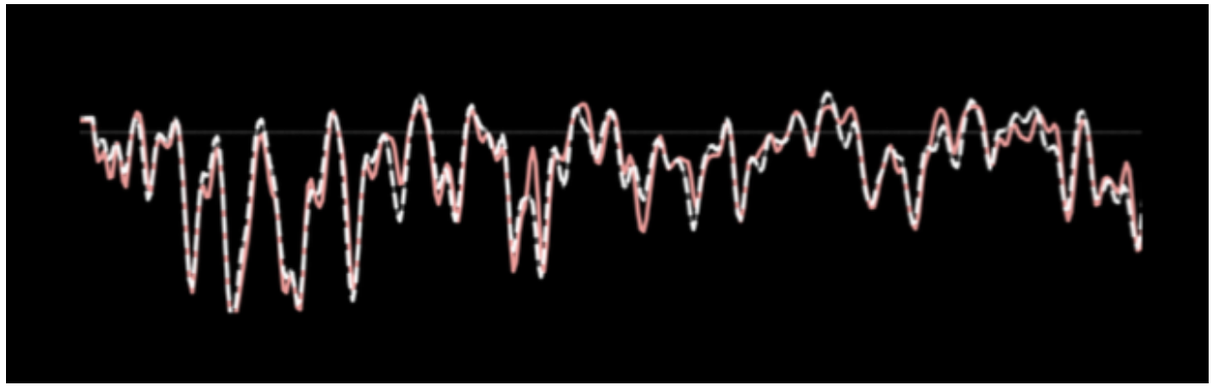
ML algorithm

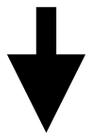
(magics happen)

↓



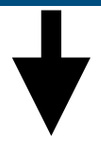
In our dreams...





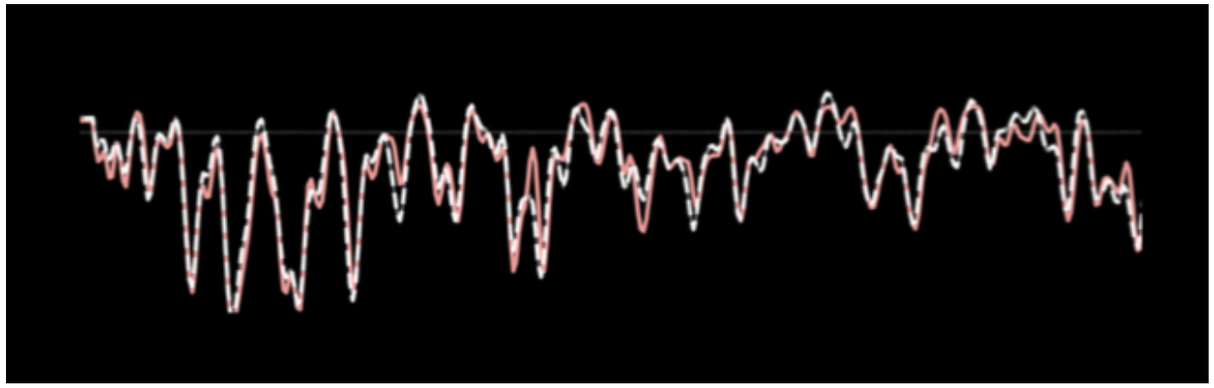
ML algorithm

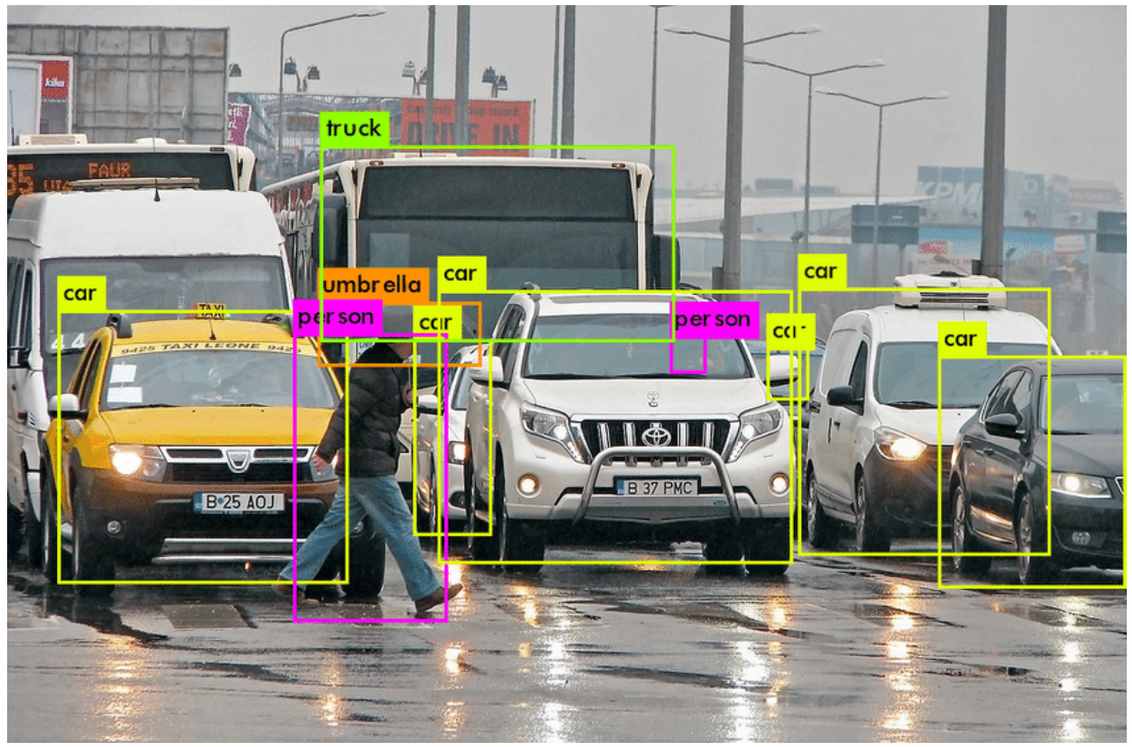
(magics happen)



Stellar labels
(T_{eff}, log g, [Fe/H], [X/Fe])

In our dreams...





ML algorithm

(magics happen)

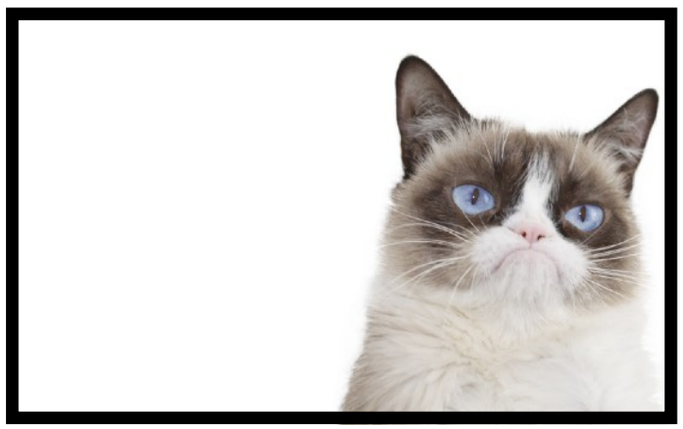
Stellar labels
(T_{eff}, log g, [Fe/H], [X/Fe])

Computer vision VS stellar spectroscopy

a cat

Computer vision VS stellar spectroscopy

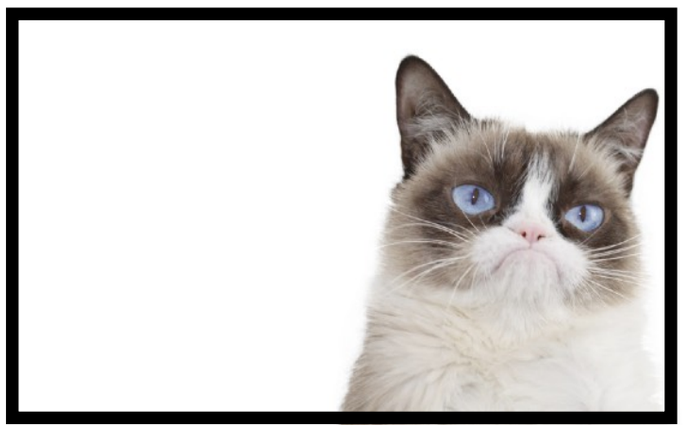
a cat



a translated cat:
the same cat

Computer vision VS stellar spectroscopy

a cat

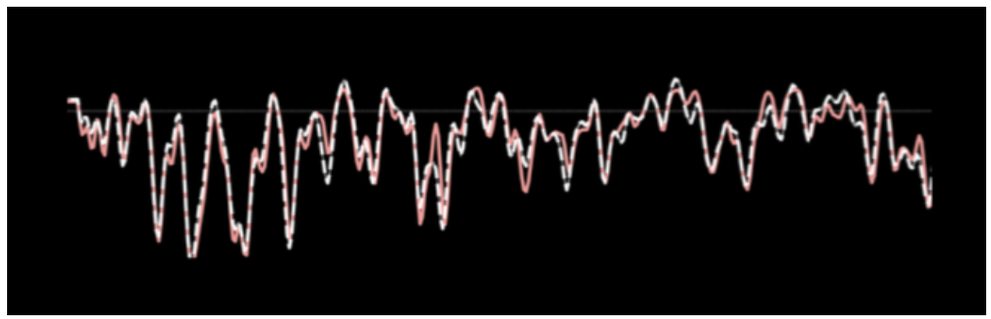


a translated cat:
the same cat

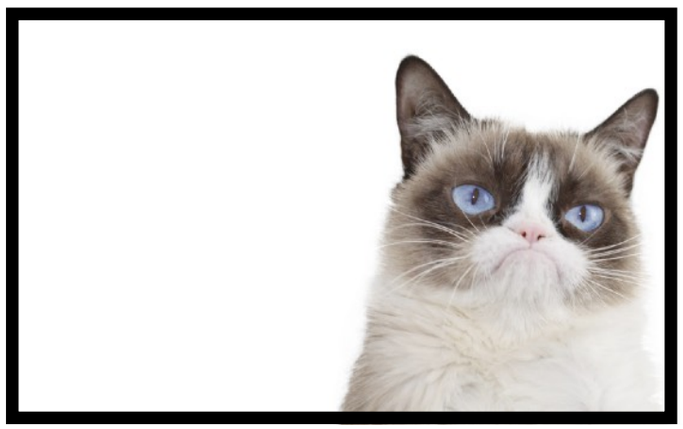
a distorted cat:
the same cat

Computer vision VS stellar spectroscopy

a cat



a spectrum

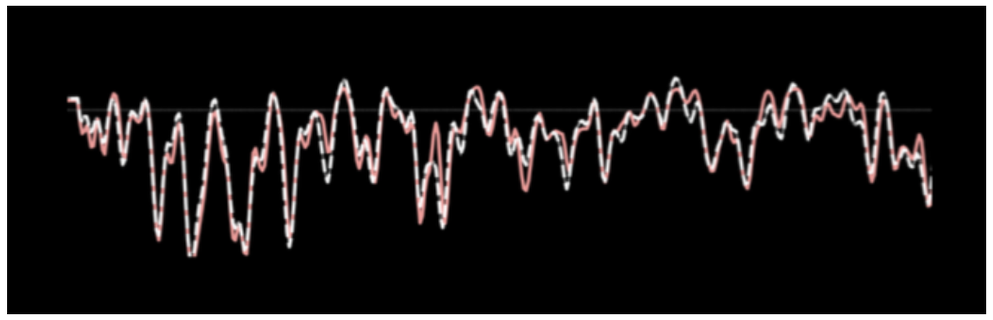


a translated cat:
the same cat

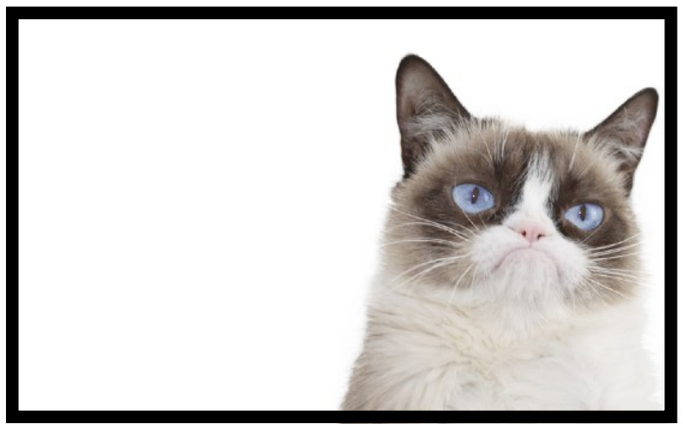
a distorted cat:
the same cat

Computer vision VS stellar spectroscopy

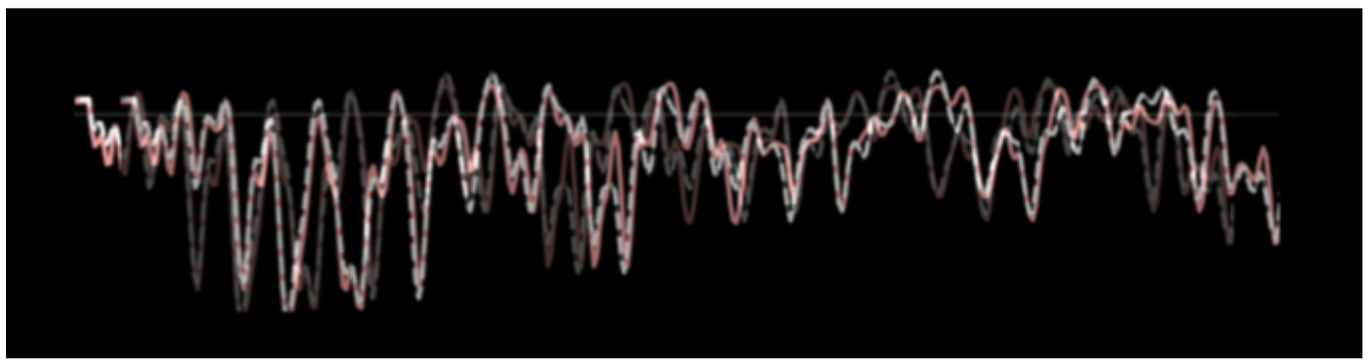
a cat



a spectrum



a translated cat:
the same cat

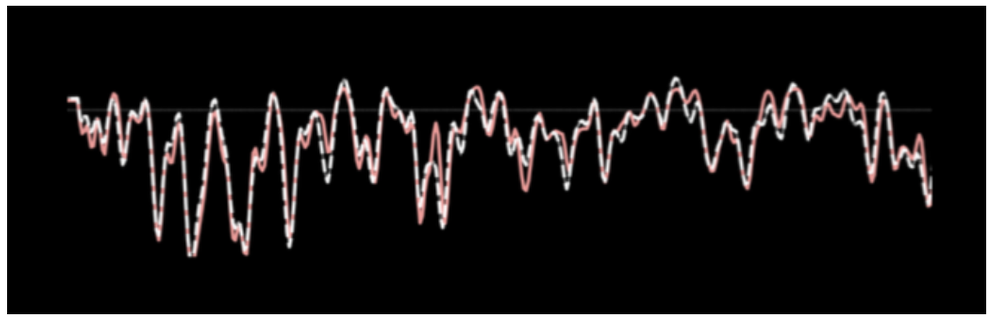


a translated spectrum:
not the same spectrum!

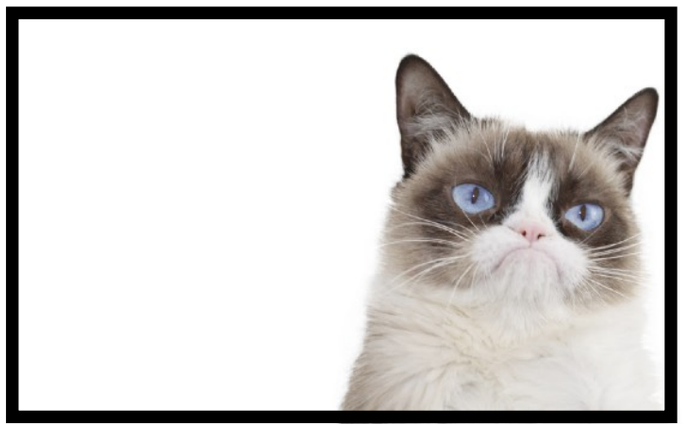
a distorted cat:
the same cat

Computer vision VS stellar spectroscopy

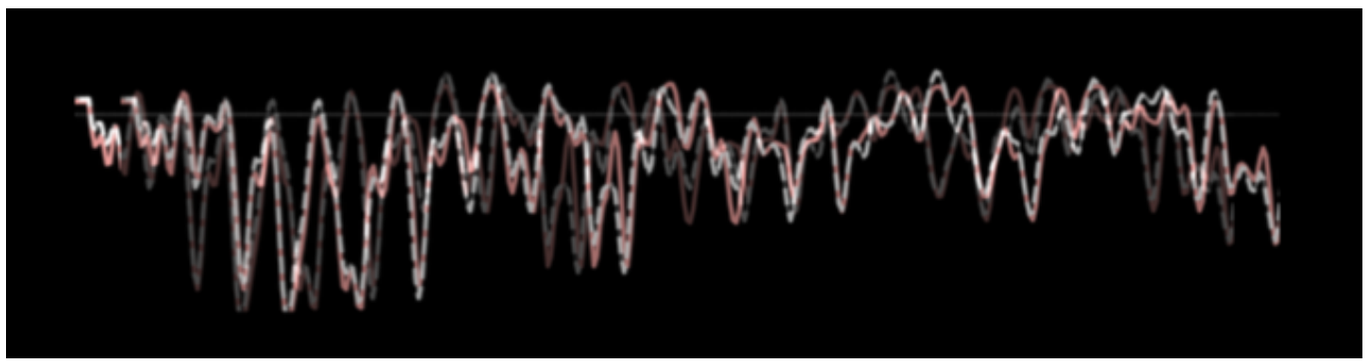
a cat



a spectrum

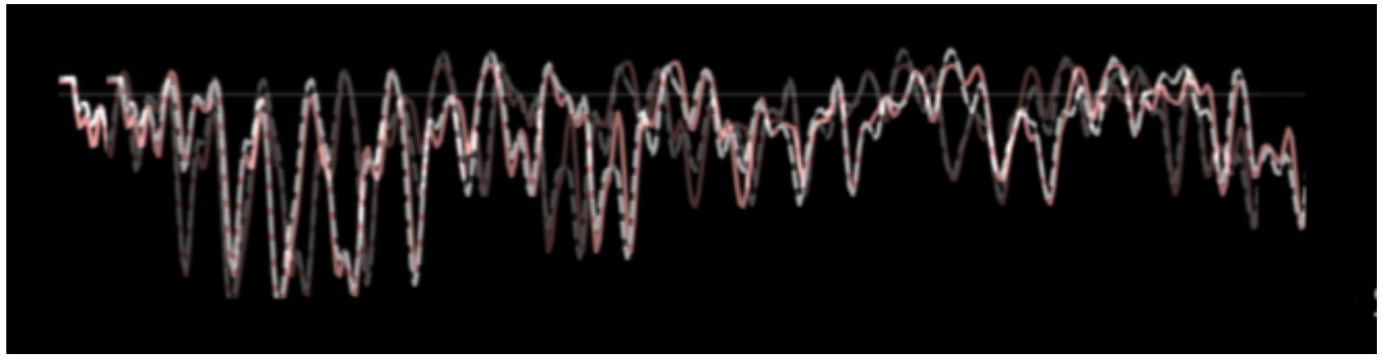


a translated cat:
the same cat



a translated spectrum:
not the same spectrum!

a distorted cat:
the same cat



a distorted spectrum:
not the same spectrum!

https://annayqho.github.io/TheCannon/

The Cannon Ness et al. (2015)

A pure data-driven approach to Spectroscopic Analysis.

Flux at every pixel for the nth-star

$$f_{n\lambda} = g(\mathbf{l_n} | \theta_\lambda) + \mathbf{noise}$$

Stellar parameters and abundances of the nth-star. (known labels)

Coefficients. One per pixel. (unknown, trainable)

The **coefficients** and the **noise** can be "learnt" from a training set of spectra whose stellar parameters are already know.

Once the model is ready, you can apply it to other stars with unknown labels.

The Cannon relies on the following to be true:

- stellar flux varies smoothly with stellar labels
- stars with the same labels have the same spectra

https://annayqho.github.io/TheCannon/

The Cannon Ness et al. (2015)

A pure data-driven approach to Spectroscopic Analysis.

Flux at every pixel for the nth-star

$$f_{n\lambda} = g(\mathbf{l_n} | \theta_\lambda) + \mathbf{noise}$$

Stellar parameters and abundances of the nth-star. (known labels)

Coefficients. One per pixel. (unknown, trainable)

The g function is a polynomium, whose order is set by the user.

First-order polynomium. The most simple model.

$$f_{n\lambda} = a_\lambda + b_\lambda(\mathrm{Teff})_n + c_\lambda(\mathrm{logg})_n + d_\lambda([\mathrm{Fe/H}])_n + \text{noise}$$

$$\theta_\lambda = (a_\lambda, b_\lambda, c_\lambda, d_\lambda)$$

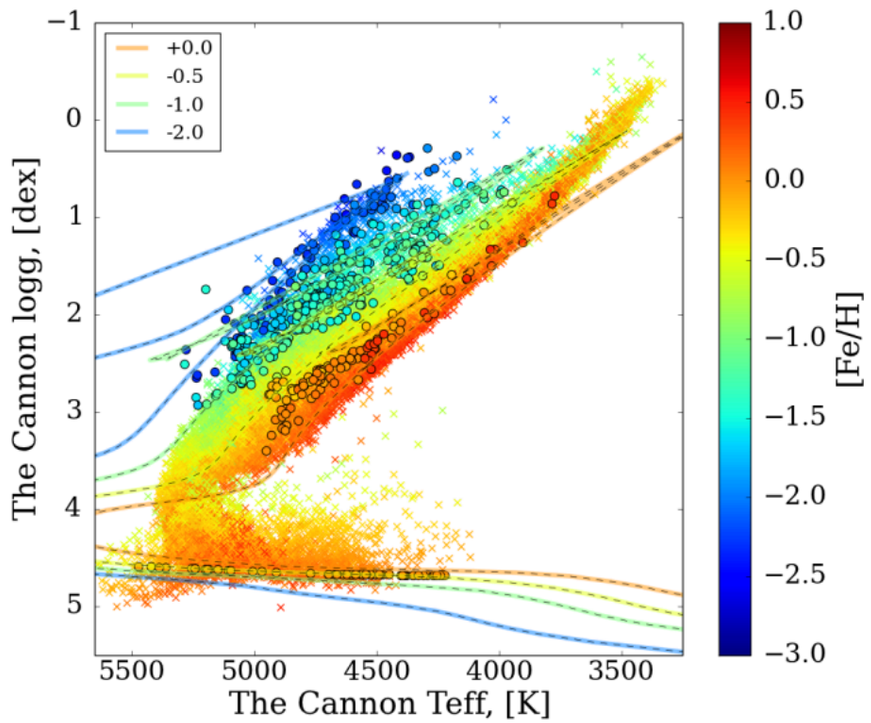
Authors suggest to use a second-order polynomium (quadratic-in-label case).

$$\begin{aligned}
F_{n\lambda} = {} & \theta_0 && \text{(constant term)} \\
& + \theta_{T_{\rm eff}} T_{\rm eff} + \cdots + \theta_{X_N}[X_N/\mathrm{Fe}] && \text{(linear terms)} \\
& + \theta_{T^2_{\rm eff}} T^2_{\rm eff} + \cdots + \theta_{X^2_N}([X_N/\mathrm{Fe}])^2 && \text{(squared terms)} \\
& + \theta_{T_{\rm eff}\,\log(g)} T_{\rm eff}\log(g) + \ldots \\
& \quad + \theta_{X_N X_{N-1}}[X_N/\mathrm{Fe}][X_{N-1}/\mathrm{Fe}] && \text{(cross-terms)} \\
& + \text{error}
\end{aligned}$$

https://annayqho.github.io/TheCannon/

The Cannon Ness et al. (2015)

A pure data-driven approach to Spectroscopic Analysis.

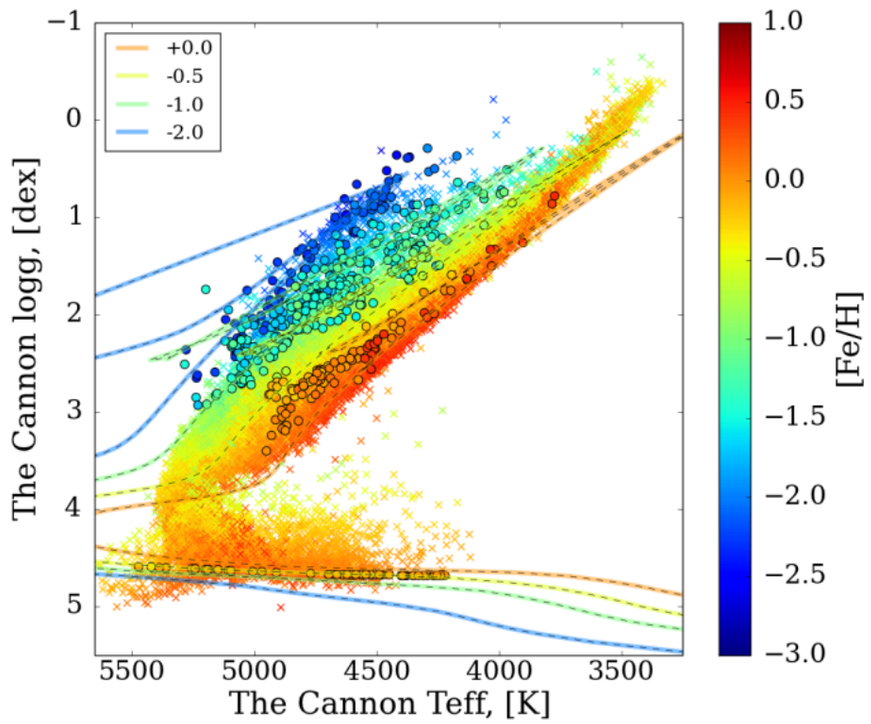


Ness et al. (2015)

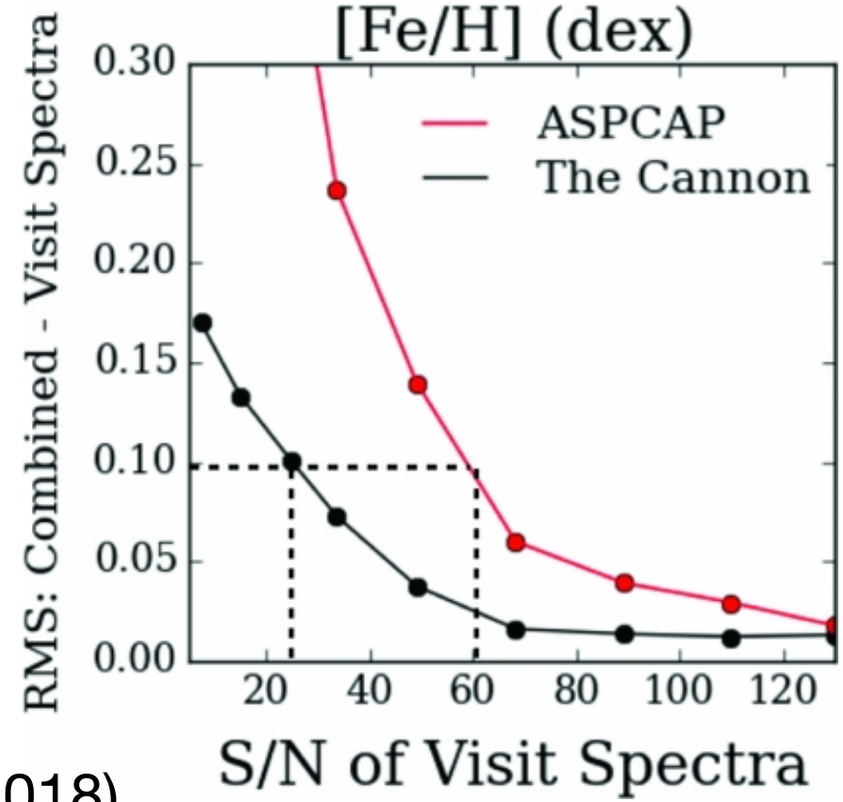
https://annayqho.github.io/TheCannon/

The Cannon Ness et al. (2015)

A pure data-driven approach to Spectroscopic Analysis.



Ness et al. (2015)

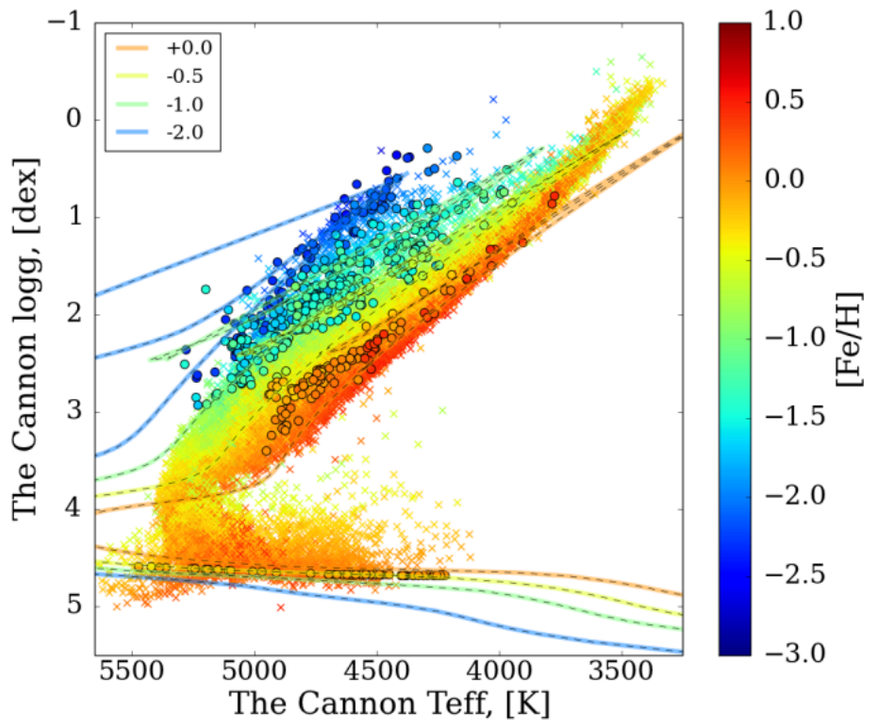


Ness (2018)

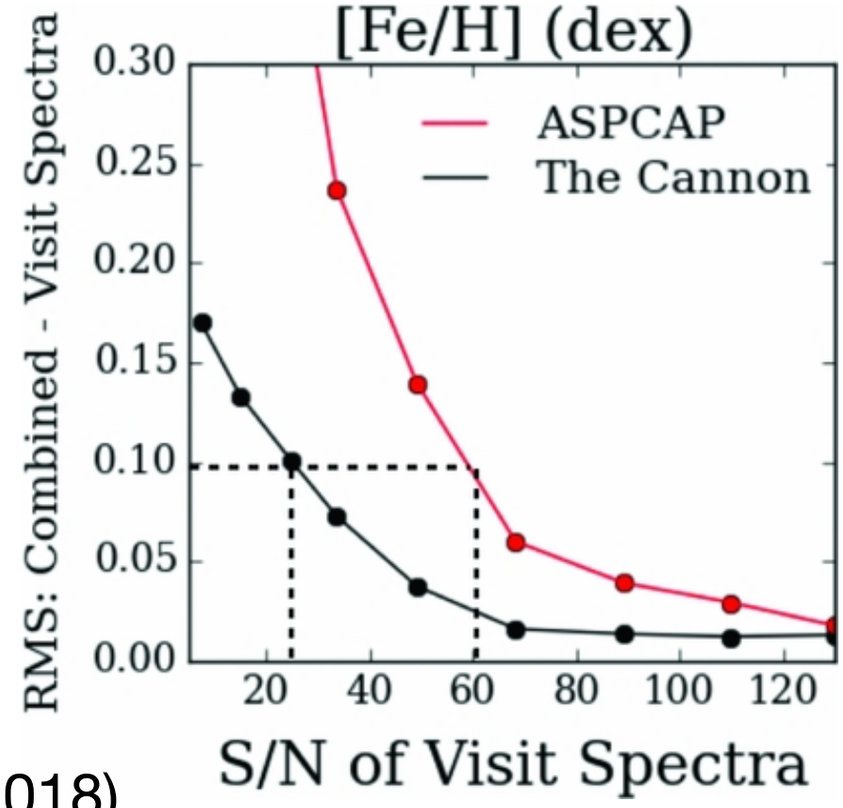
https://annayqho.github.io/TheCannon/

The Cannon Ness et al. (2015)

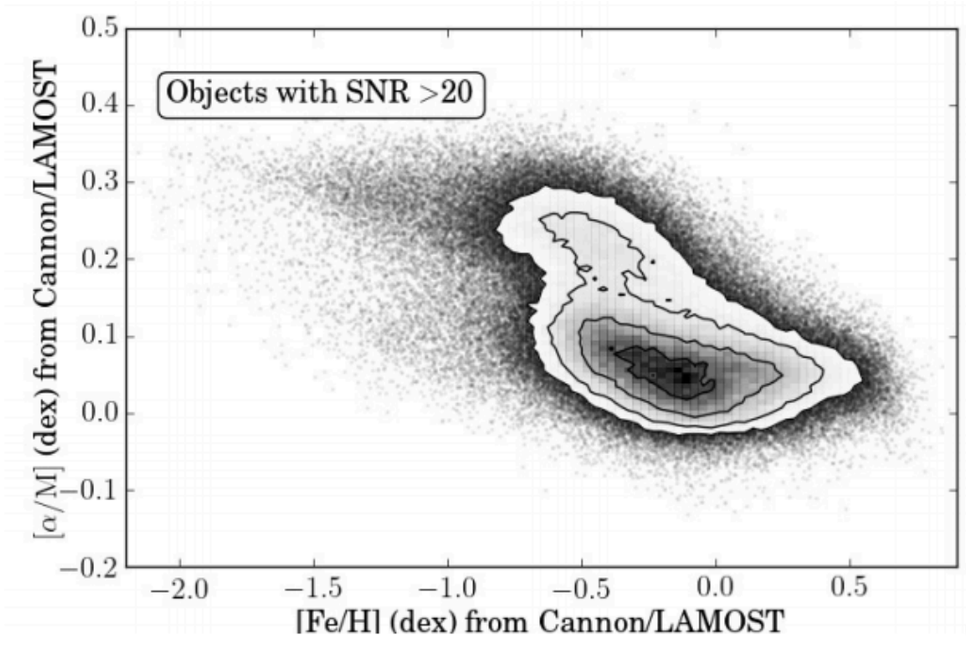
A pure data-driven approach to Spectroscopic Analysis.



Ness et al. (2015)



Ness (2018)

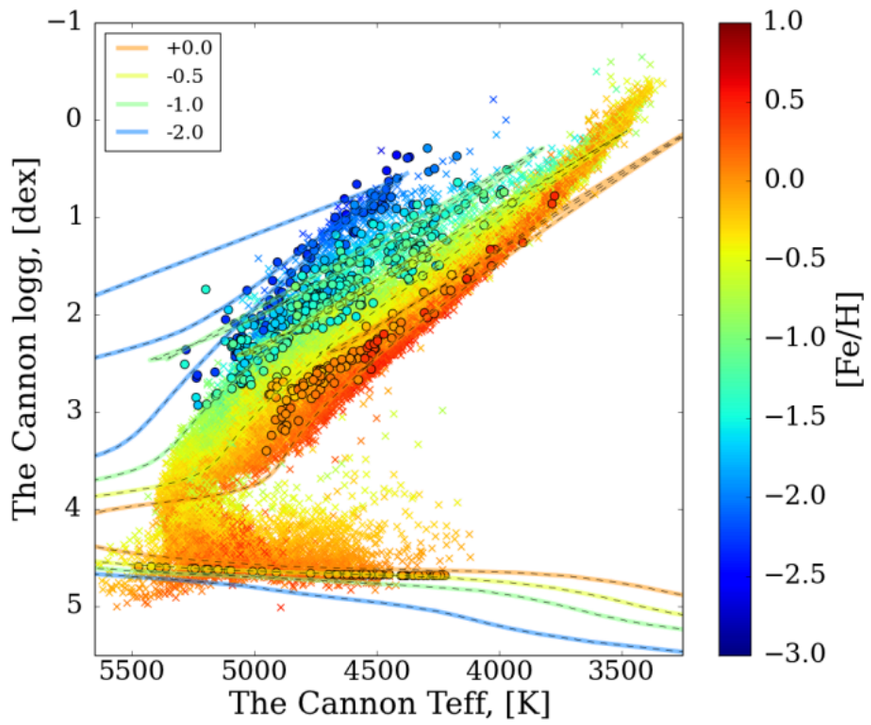


Oh et al. (2017)

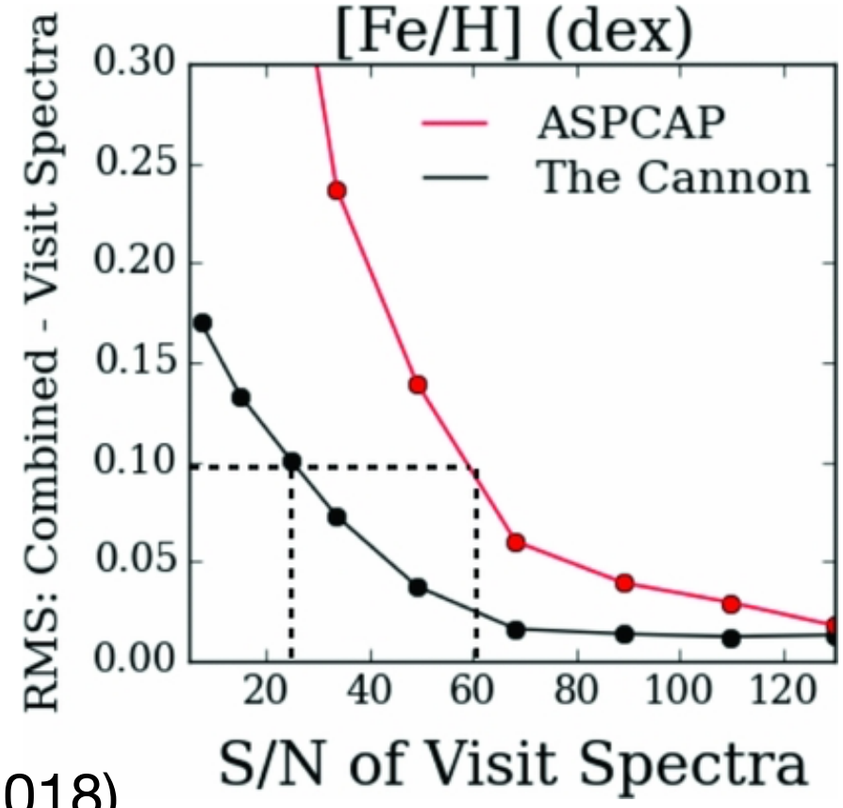
https://annayqho.github.io/TheCannon/

The Cannon Ness et al. (2015)

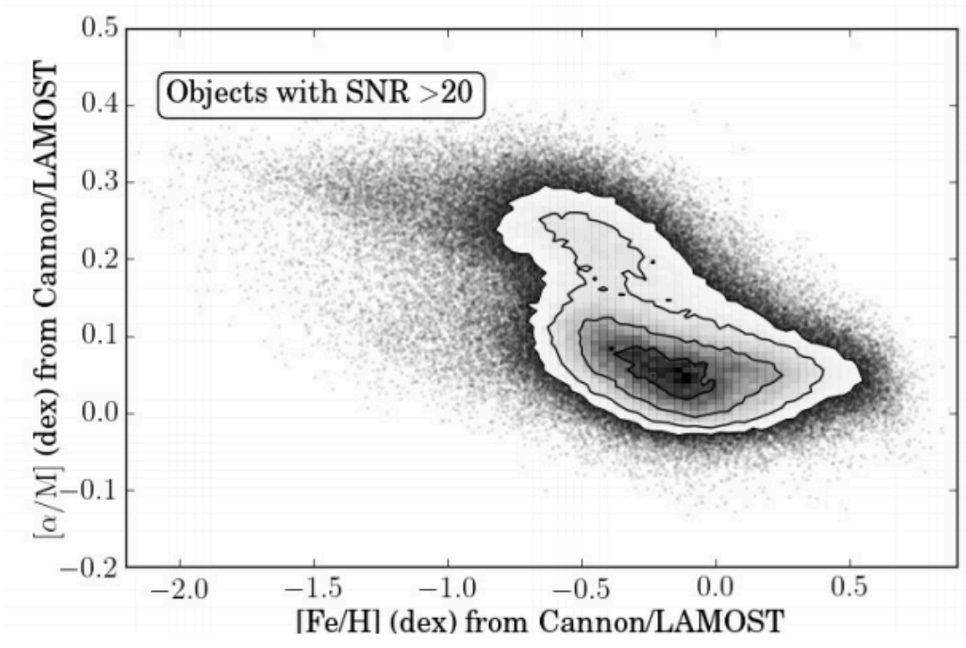
A pure data-driven approach to Spectroscopic Analysis.



Ness et al. (2015)



Ness (2018)



Oh et al. (2017)

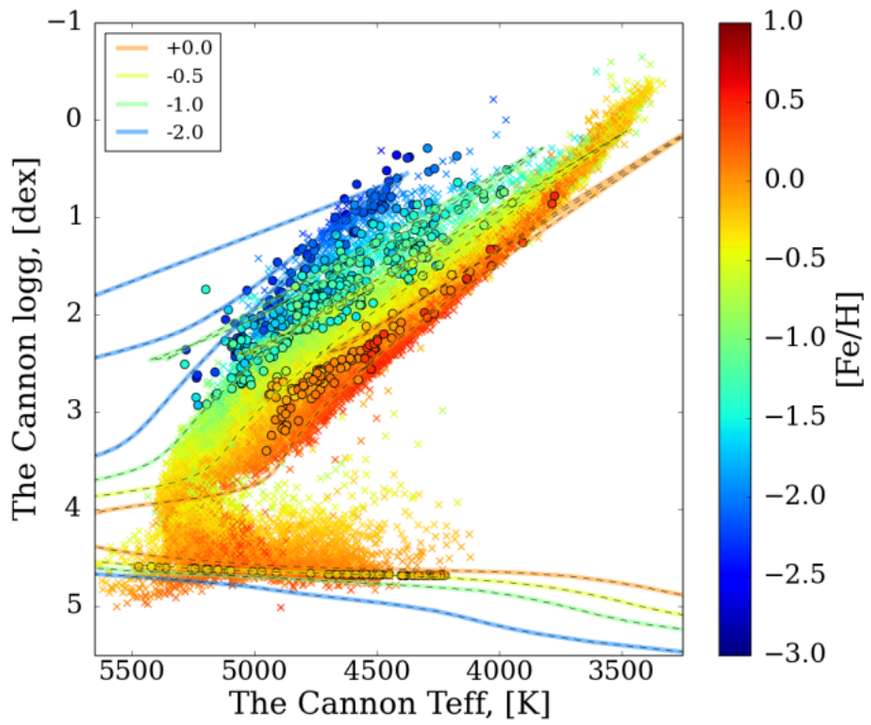
Advantage

The method is extremely symple and fast.

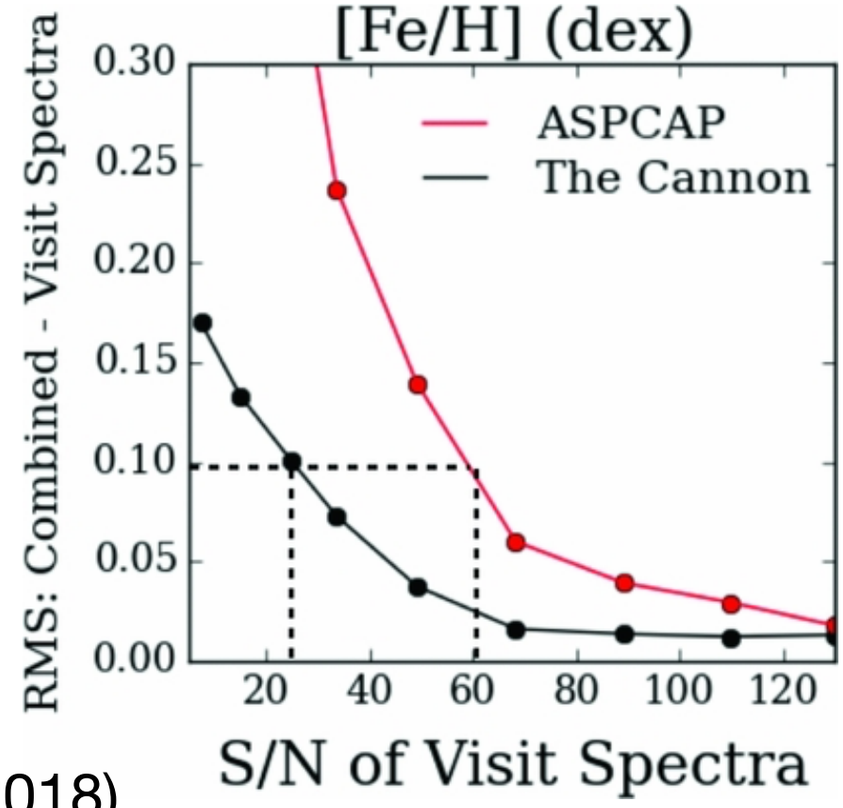
https://annayqho.github.io/TheCannon/

The Cannon Ness et al. (2015)

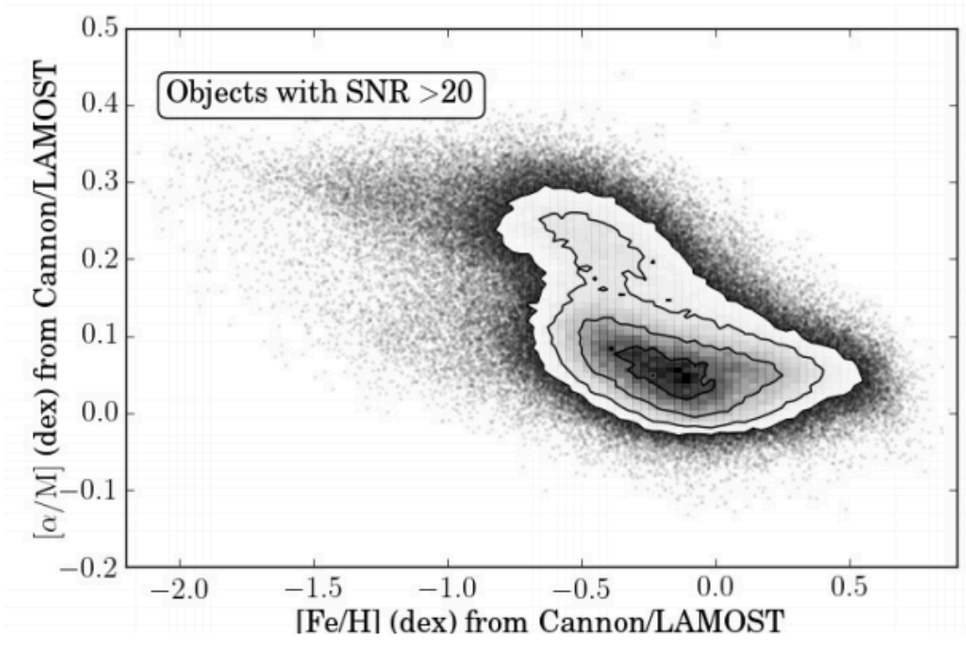
A pure data-driven approach to Spectroscopic Analysis.



Ness et al. (2015)



Ness (2018)



Oh et al. (2017)

Advantage

The method is extremely symple and fast.

Danger

The method uses no physical stellar models.

The pitfalls of a data-driven approach

Beware of hidden **bias** and **correlations**!

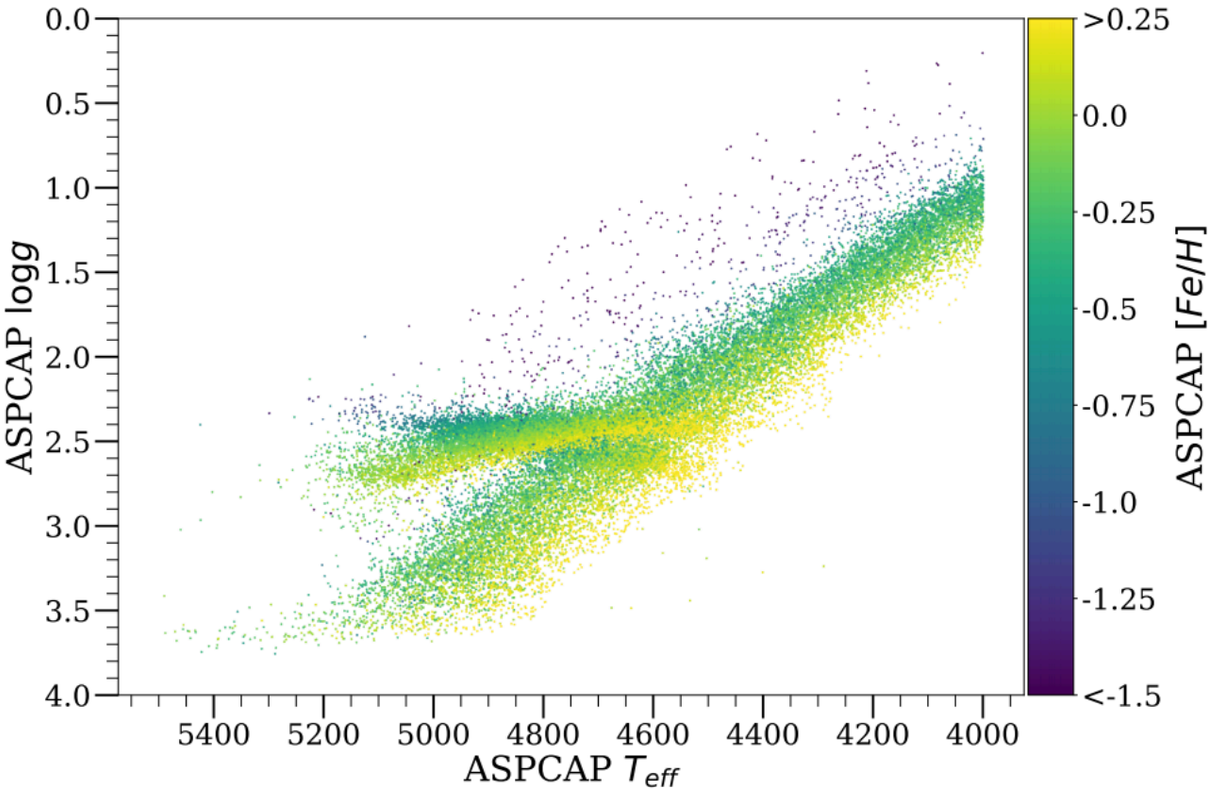
The danger of using data-driven models.

We need to be extra cautious of what the ML algorithms are "learning" and what training sets we build.

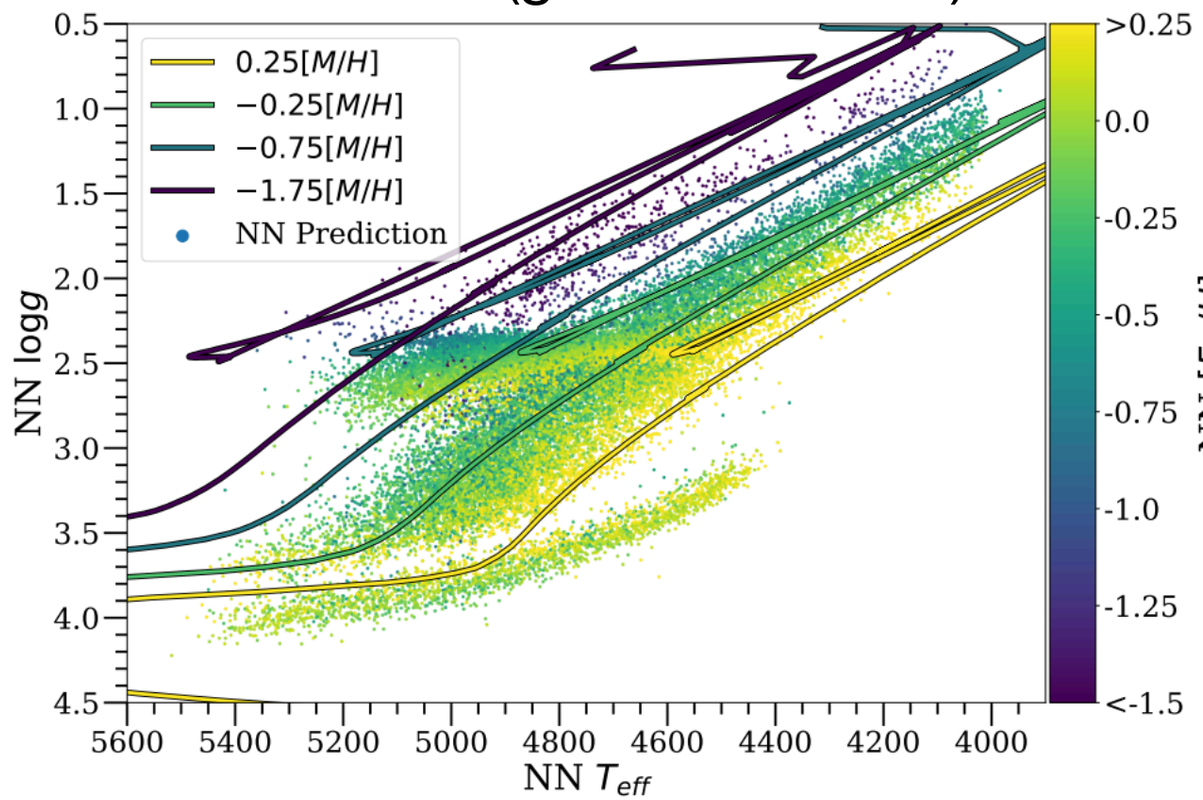
The pitfalls of a data-driven approach

The effect of **biases** in the training set...

Training set (only giants)



Test set (giants + dwarfs)



Leug & Bovy (2018)

Take-home messages:

The training set must be representative of the specific problem we want to address.

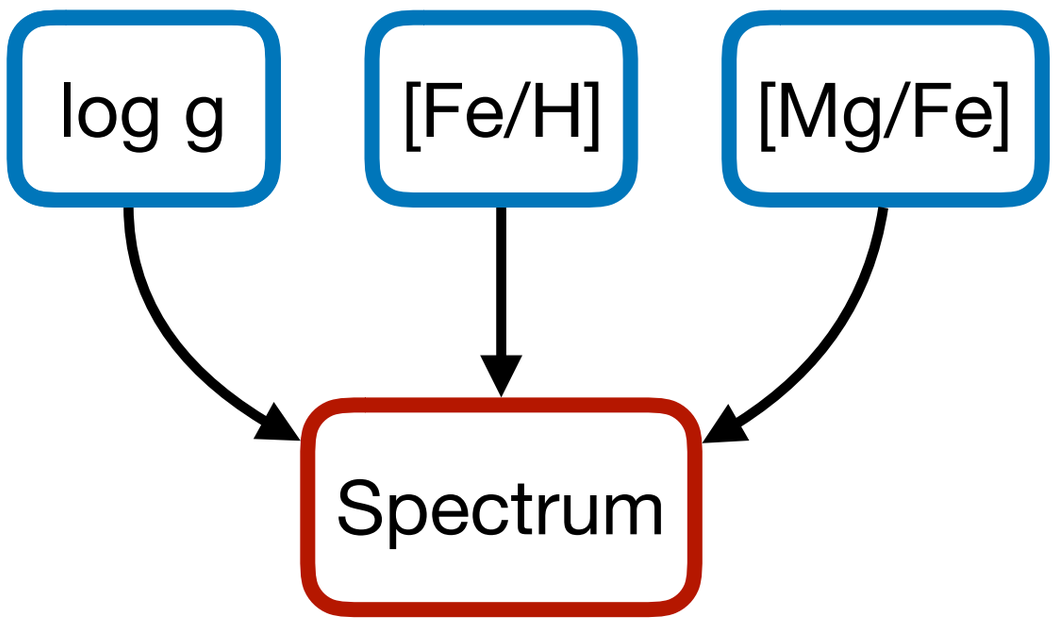
Any bias that exists in the training set will be folded into the test set.

The pitfalls of a data-driven approach

The effect of **correlations** in the training set.

Is the model
"learning"
measurements?

...or is it "learning"
hidden correlations?

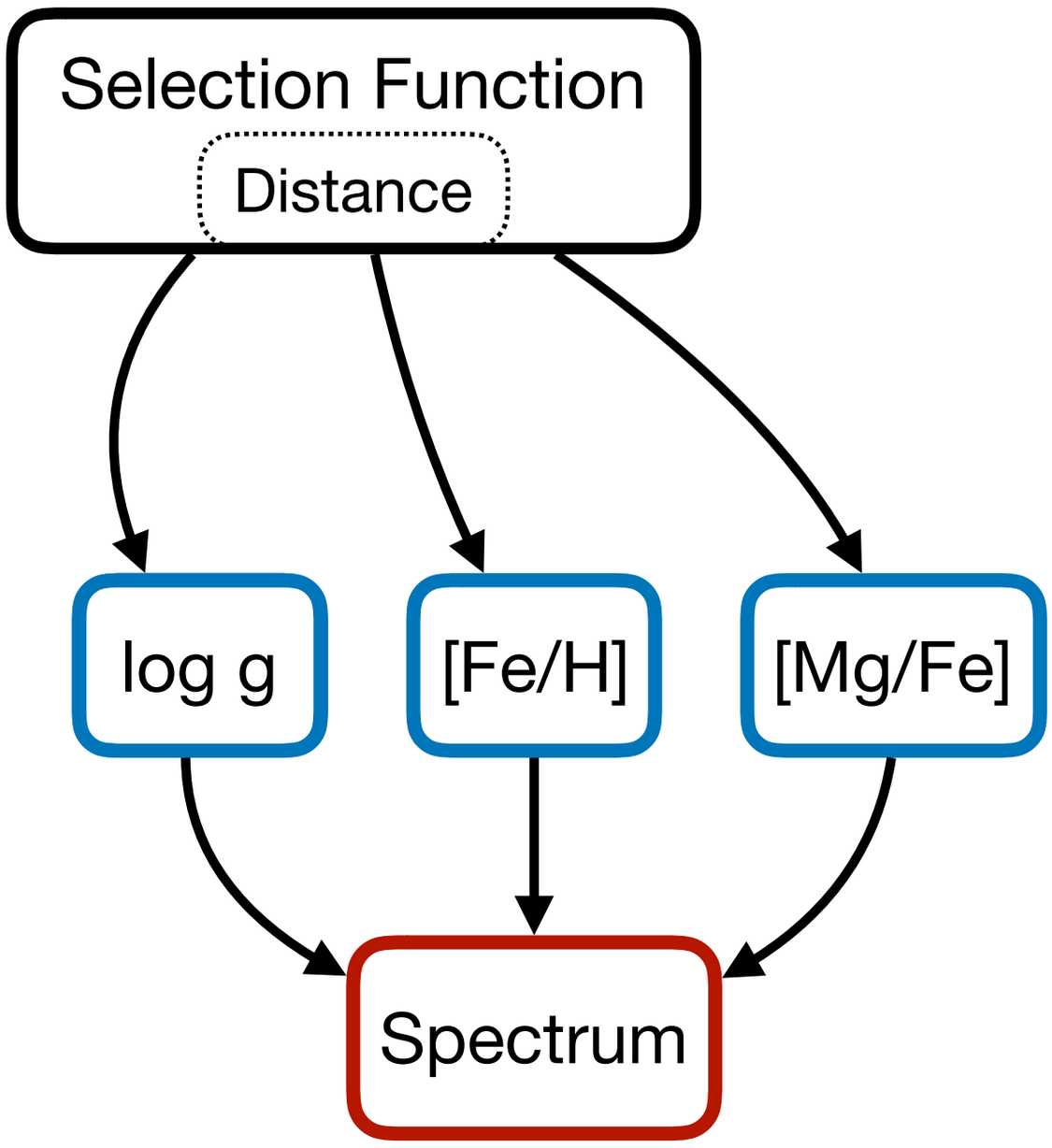


The pitfalls of a data-driven approach

The effect of **correlations** in the training set.

Is the model "learning" **measurements**?

...or is it "learning" **hidden correlations**?

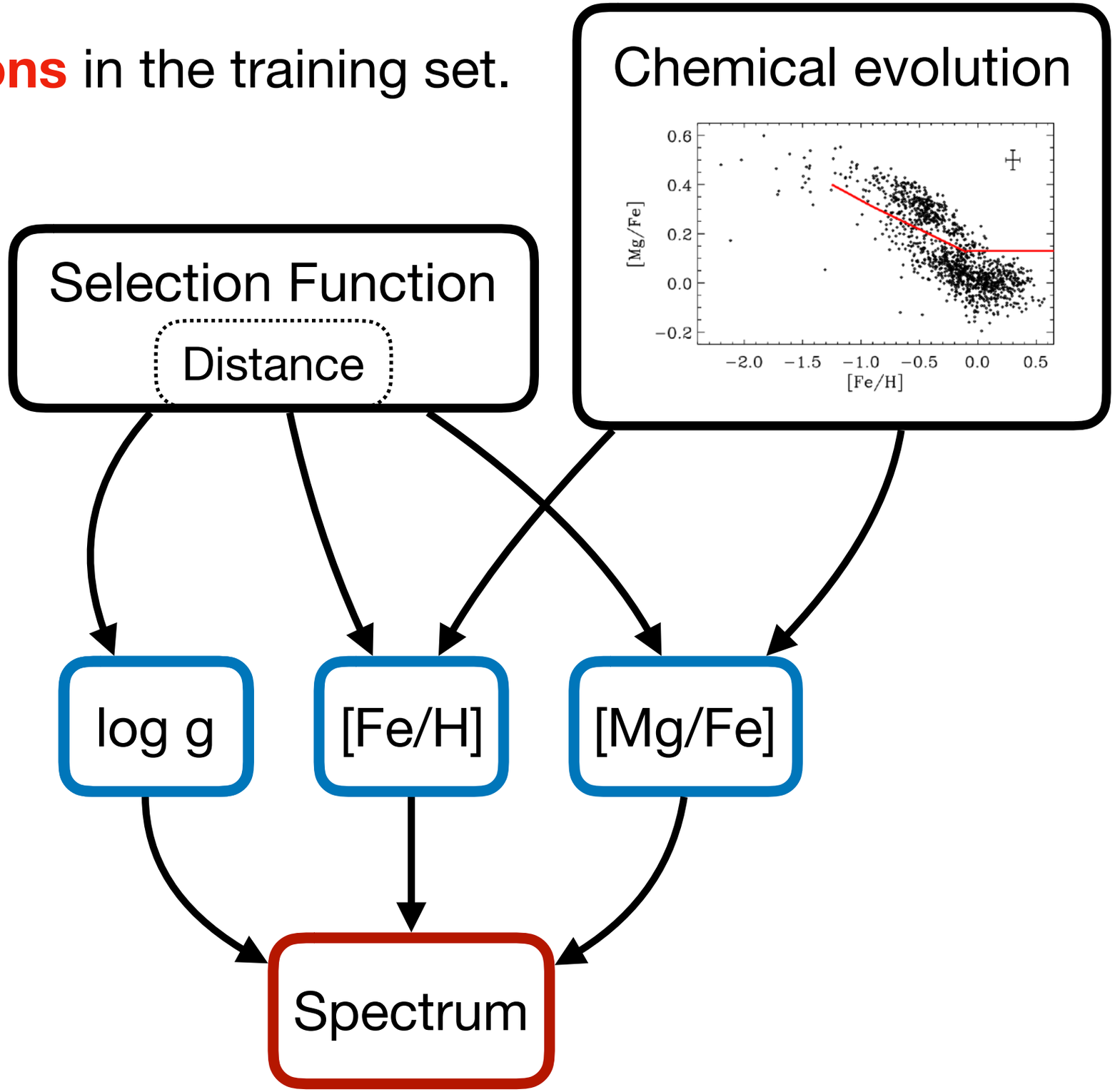


The pitfalls of a data-driven approach

The effect of **correlations** in the training set.

Is the model "learning" **measurements**?

...or is it "learning" **hidden correlations**?

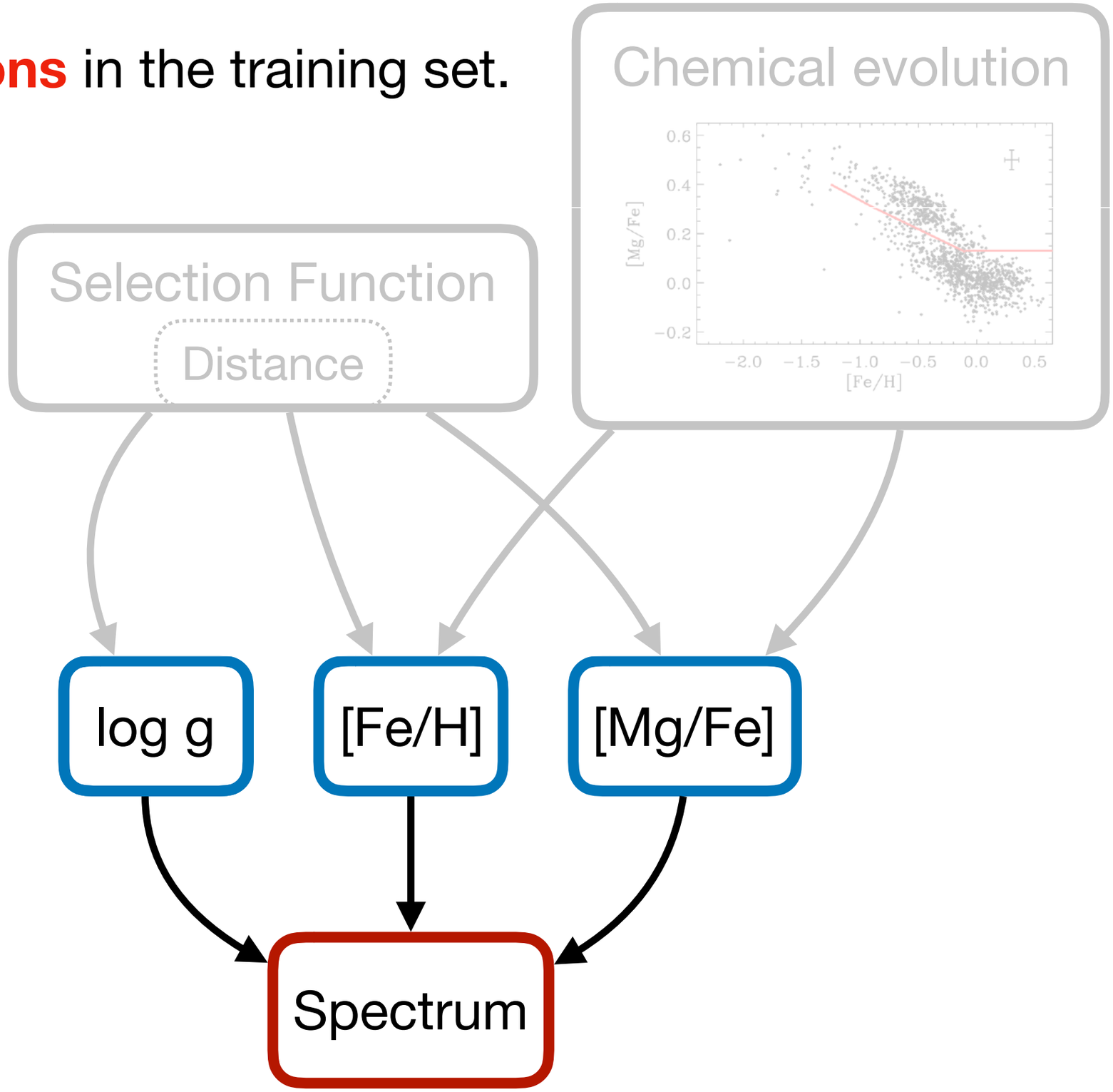


The pitfalls of a data-driven approach

The effect of **correlations** in the training set.

Is the model "learning" **measurements**?

...or is it "learning" **hidden correlations**?



The pitfalls of a data-driven approach

Beware of hidden **bias** and **correlations**!

The danger of using data-driven models.

We need to be extra cautious of what the ML algorithms are "learning" and what training sets we build.

The pitfalls of a data-driven approach

Beware of hidden **bias** and **correlations**!

The danger of using data-driven models.

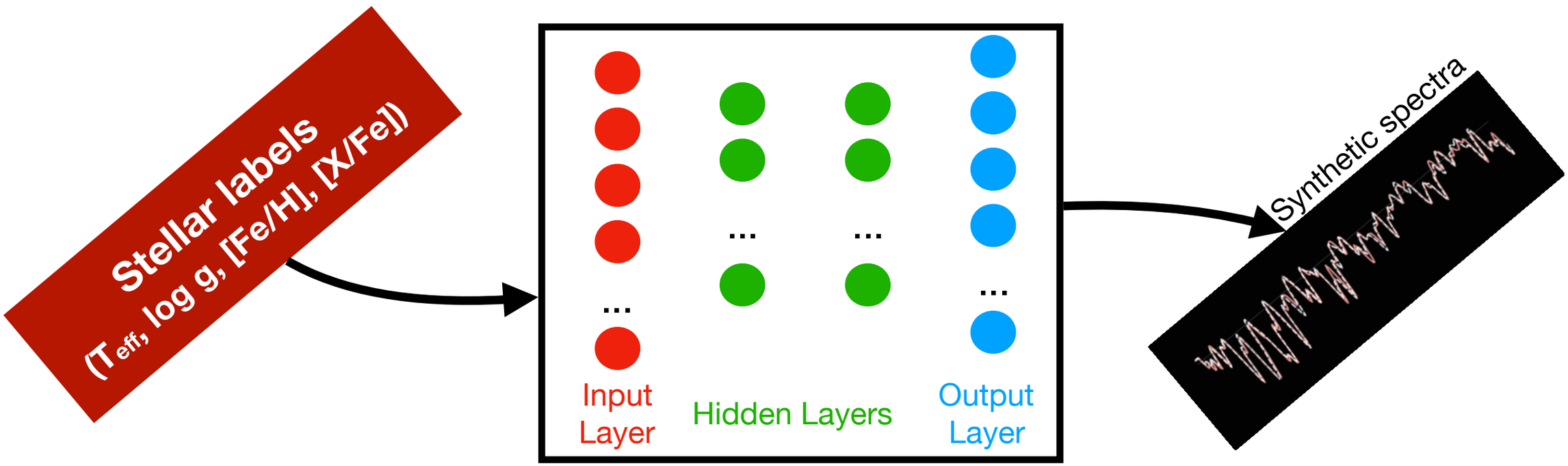
We need to be extra cautious of what the ML algorithms are "learning" and what training sets we build.

Unfortunately bias and correlations are unavoidable in stellar spectroscopy...

https://github.com/tingyuansen/The_Payne

The Payne Ting et al. (2019)

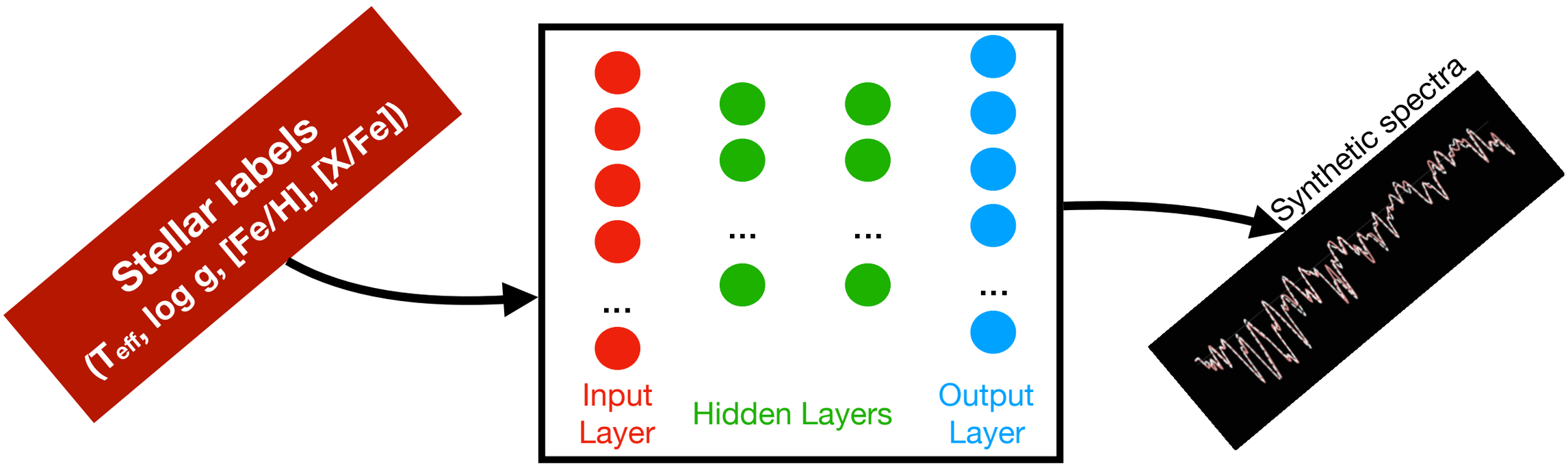
Physical generalization: the training sample is composed by synthetic spectra.

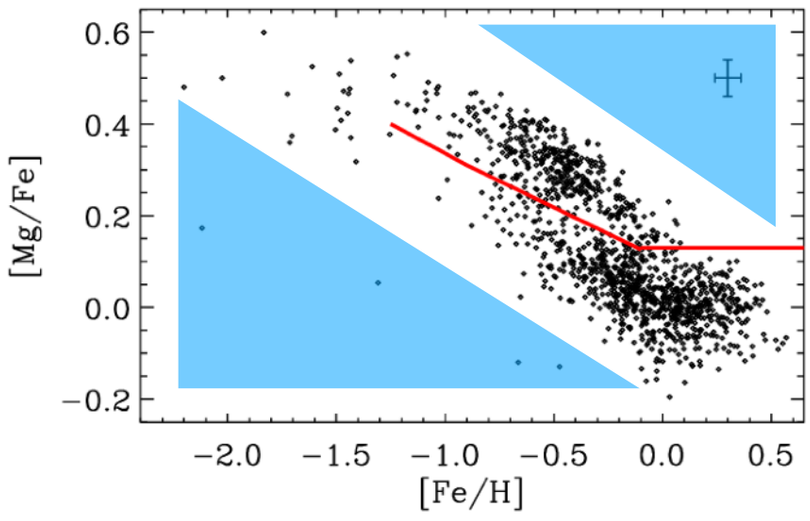


https://github.com/tingyuansen/The_Payne

The Payne Ting et al. (2019)

Physical generalization: the training sample is composed by synthetic spectra.



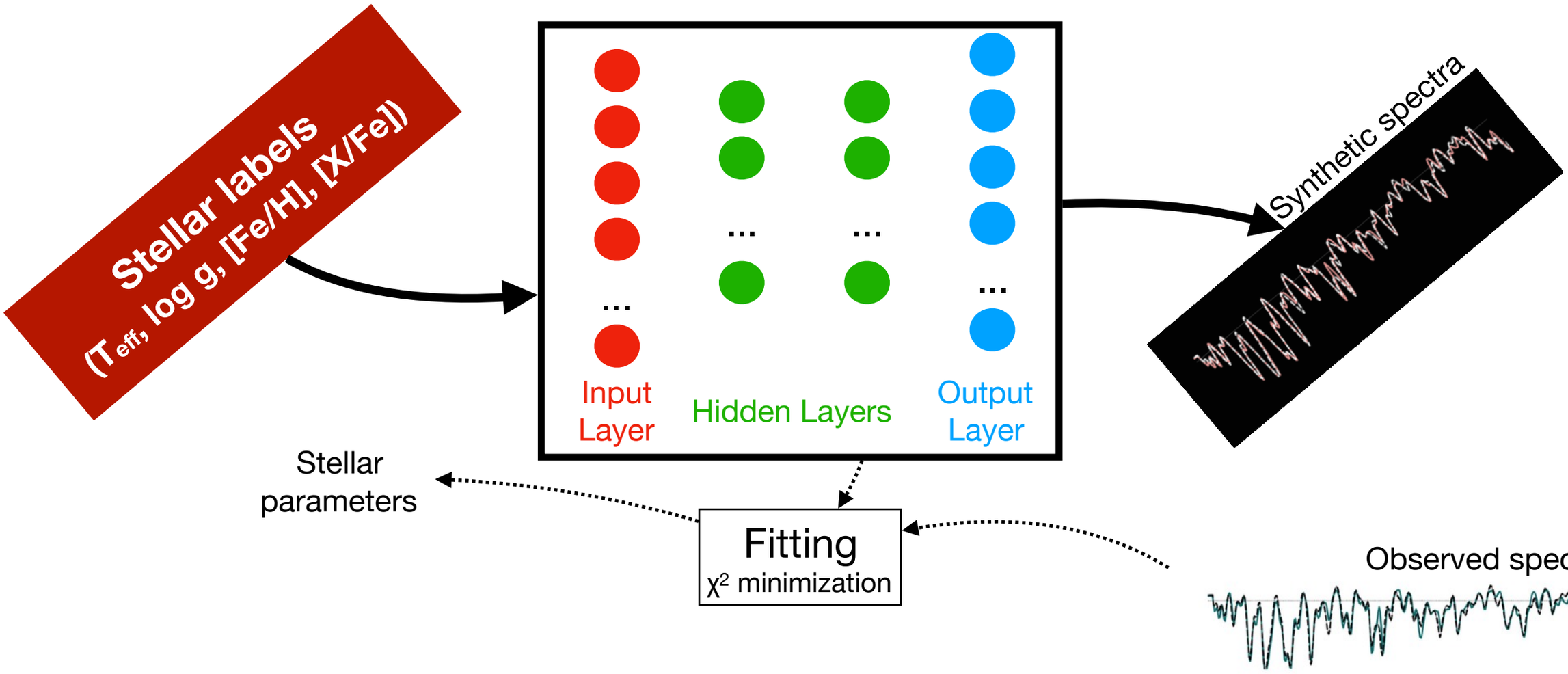


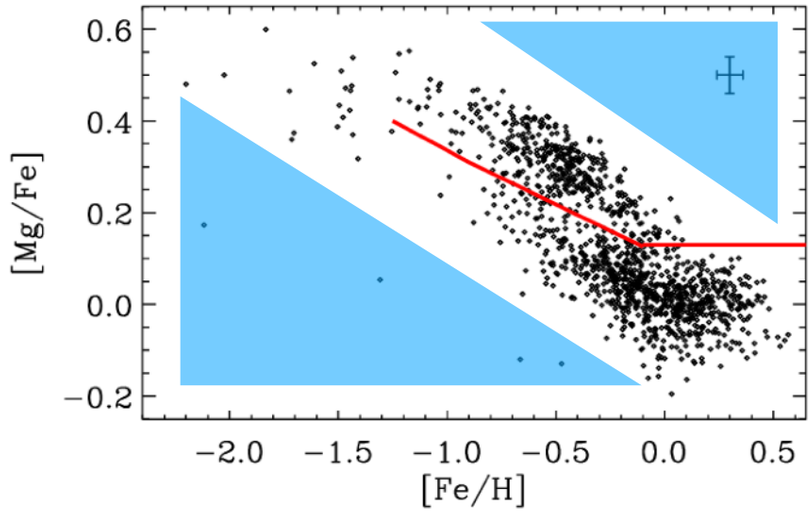
Full control on the training sample
You can sample the entire space of parameters.

https://github.com/tingyuansen/The_Payne

The Payne Ting et al. (2019)

Physical generalization: the training sample is composed by synthetic spectra.





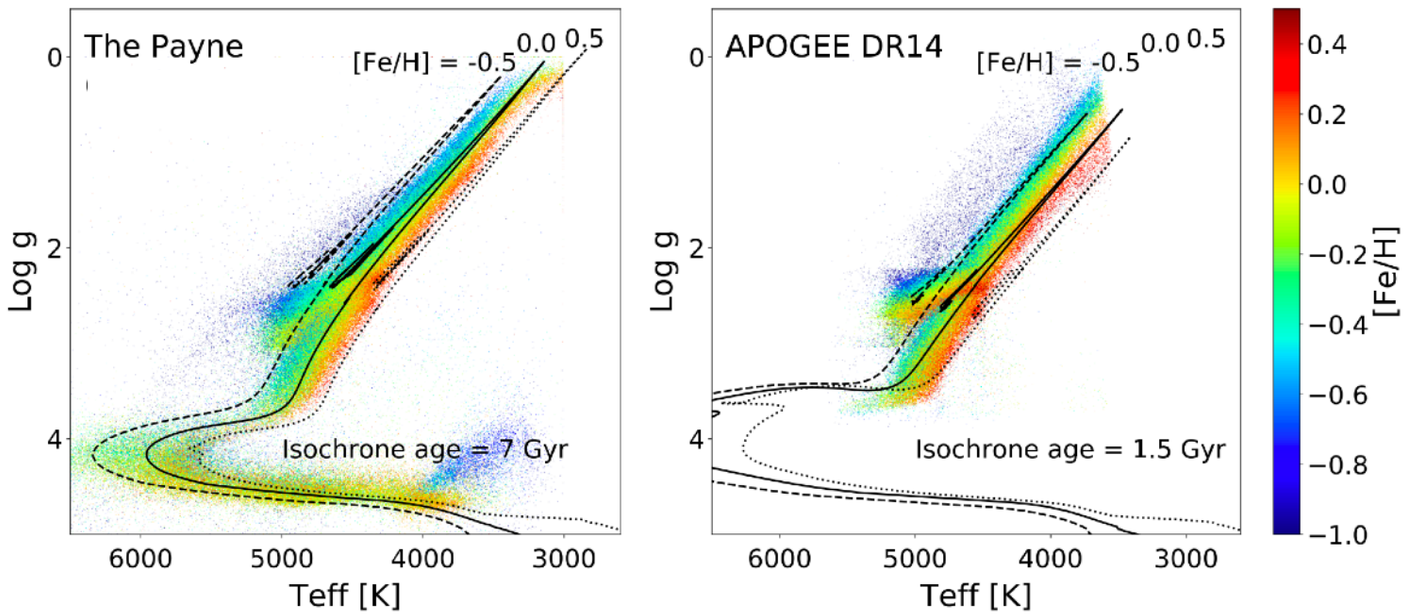
Full control on the training sample
You can sample the entire space of parameters.

Thanks to this fast interpolator of atmospheric models you can simply fit your spectra in order to calculate the atmospheric parameters.

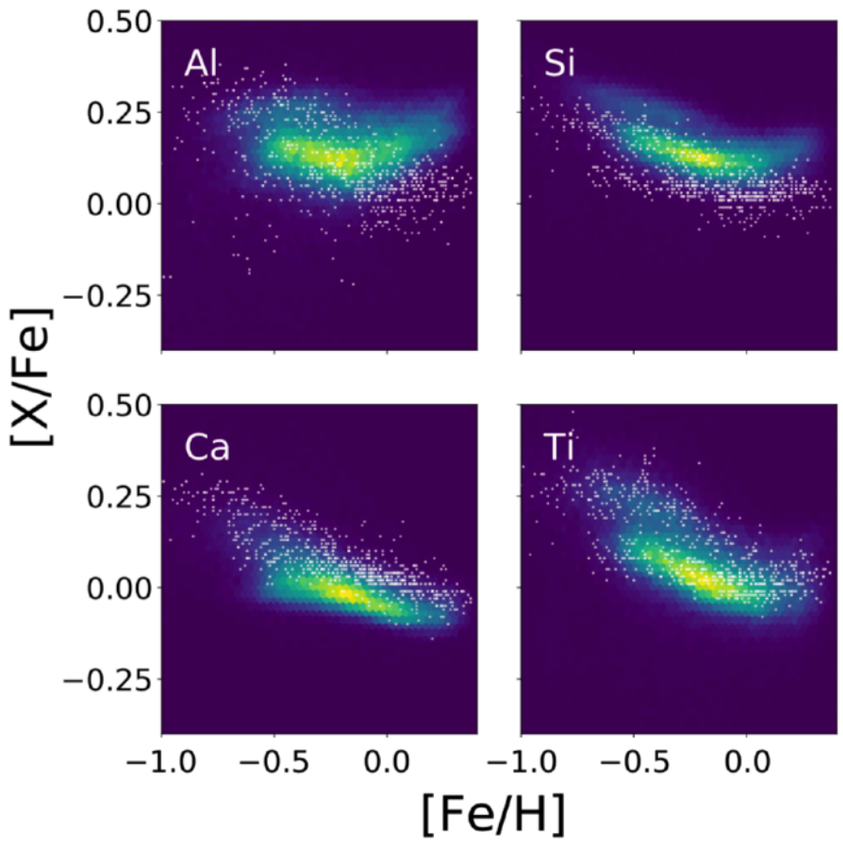
https://github.com/tingyuansen/The_Payne

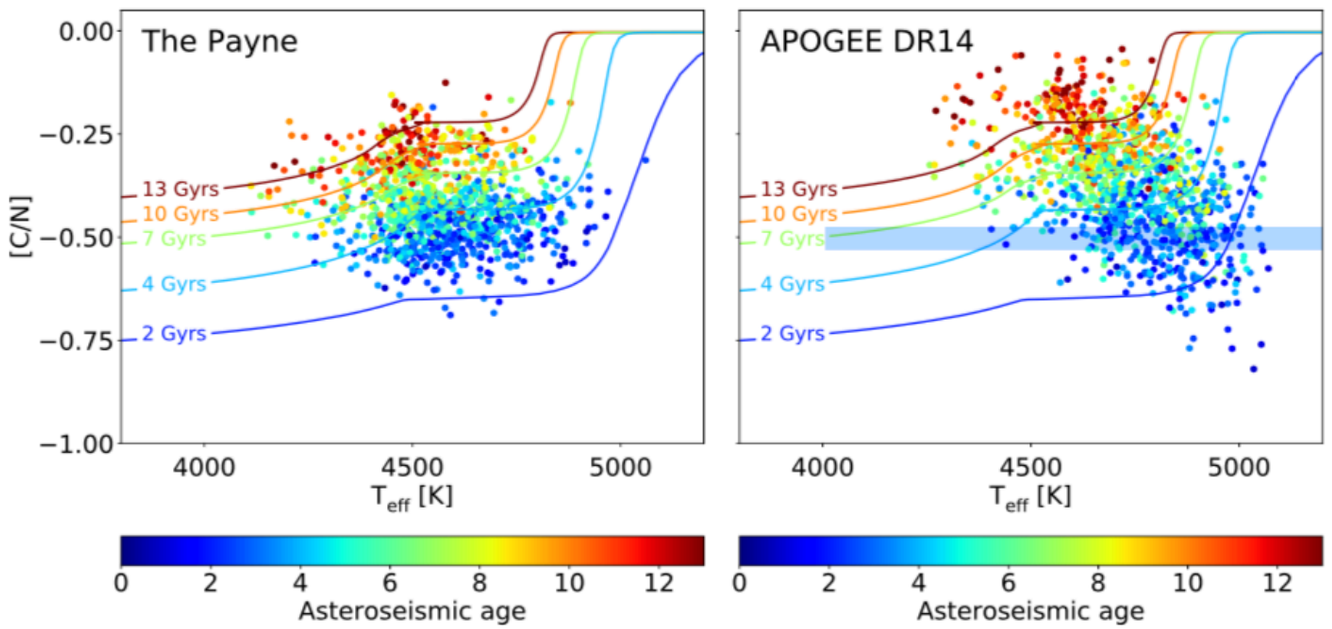
The Payne Ting et al. (2019)

Physical generalization: the training sample is composed by synthetic spectra.



Analysis of the APOGEE DR14 dataset with The Payne (Ting et al 2019).





The pitfall of physical generalisation: the synthetic gap

Models are good approximations of reality, but they are not perfect.
Synthetic gap: differences in feature distributions between synthetic and observed spectra.

Tips to reduce the synthetic gap:

- pre-processing the spectra: matching the resolution and sampling of the spectra to a common wavelength scale and removing the continuum.
- argumenting the spectra data set: adding Gaussian noise, rotational and radial velocities, masking telluric regions, and zeroing flux values to mimic bad pixels

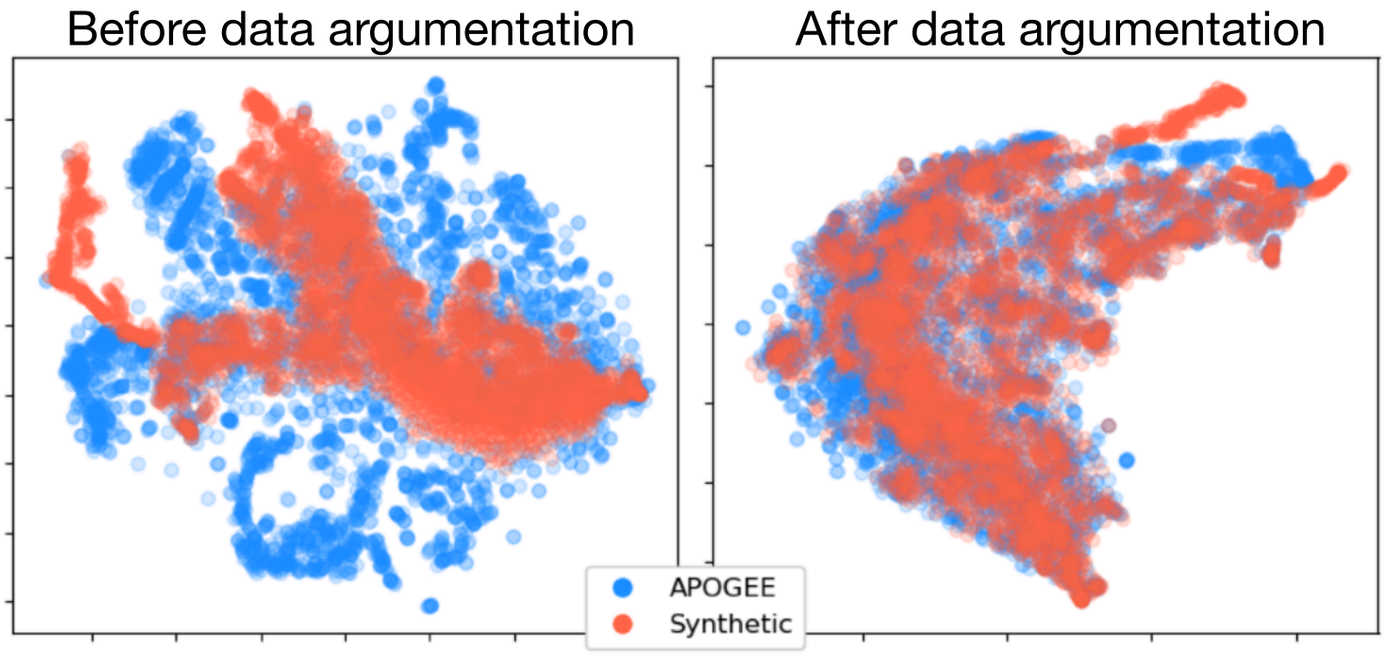
The pitfall of physical generalisation: the synthetic gap

Models are good approximations of reality, but they are not perfect.
Synthetic gap: differences in feature distributions between synthetic and observed spectra.

Tips to reduce the synthetic gap:

- pre-processing the spectra: matching the resolution and sampling of the spectra to a common wavelength scale and removing the continuum.
- argumenting the spectra data set: adding Gaussian noise, rotational and radial velocities, masking telluric regions, and zeroing flux values to mimic bad pixels

t-SNE visualization of the synthetic and APOGEE spectra



Fabbro et al. (2017)

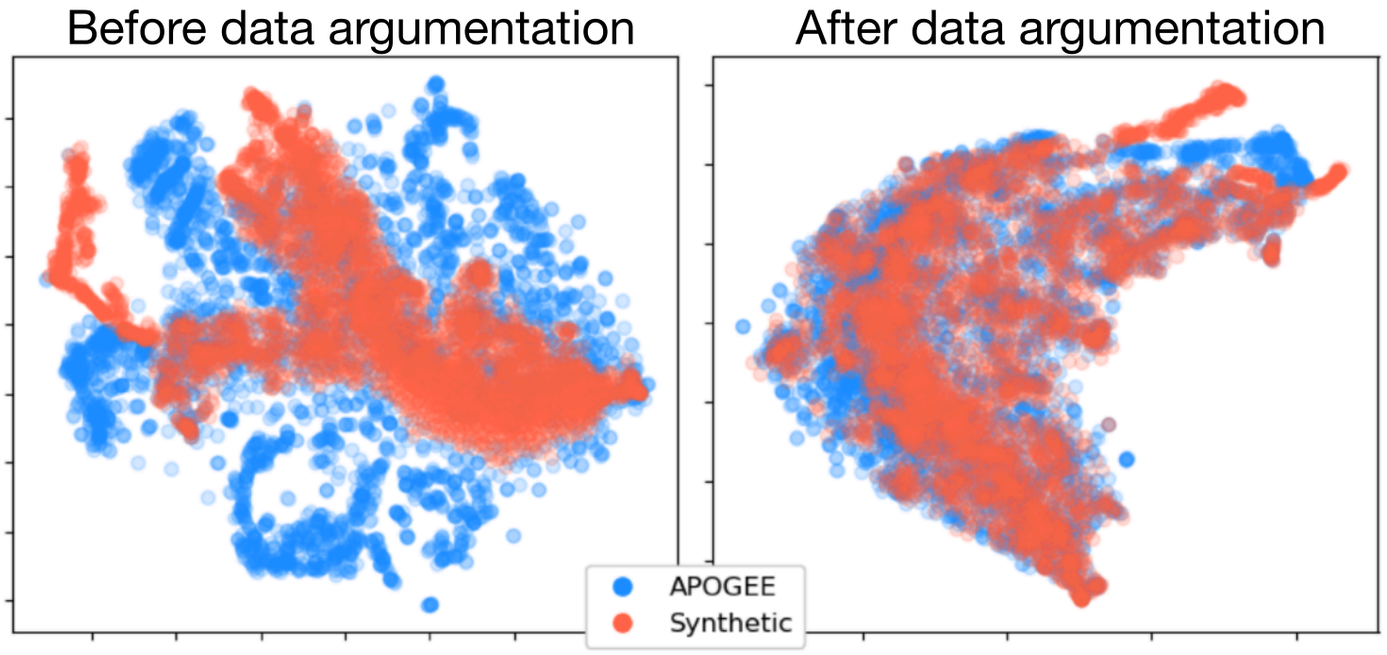
The pitfall of physical generalisation: the synthetic gap

Models are good approximations of reality, but they are not perfect.
Synthetic gap: differences in feature distributions between synthetic and observed spectra.

Tips to reduce the synthetic gap:

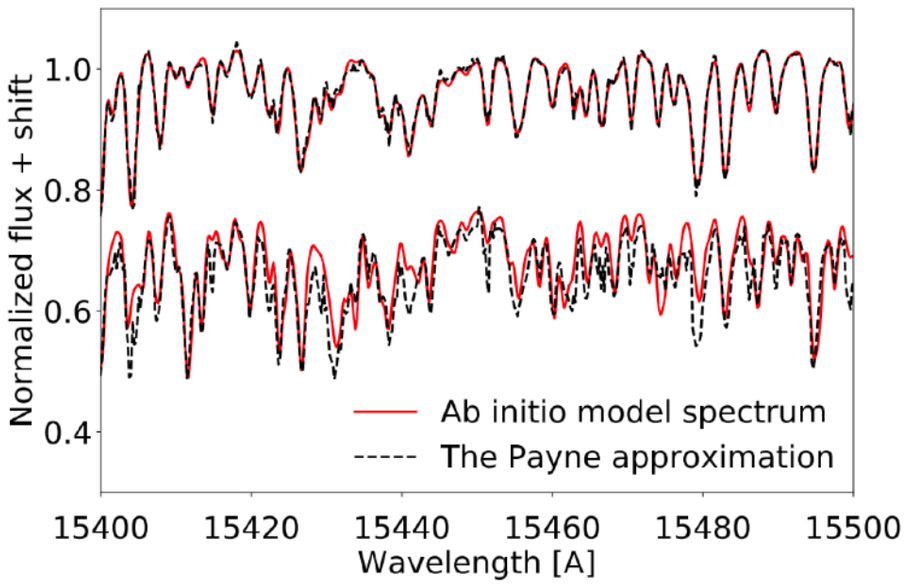
- pre-processing the spectra: matching the resolution and sampling of the spectra to a common wavelength scale and removing the continuum.
- argumenting the spectra data set: adding Gaussian noise, rotational and radial velocities, masking telluric regions, and zeroing flux values to mimic bad pixels

t-SNE visualization of the synthetic and APOGEE spectra



Fabbro et al. (2017)

Spectra generalisation



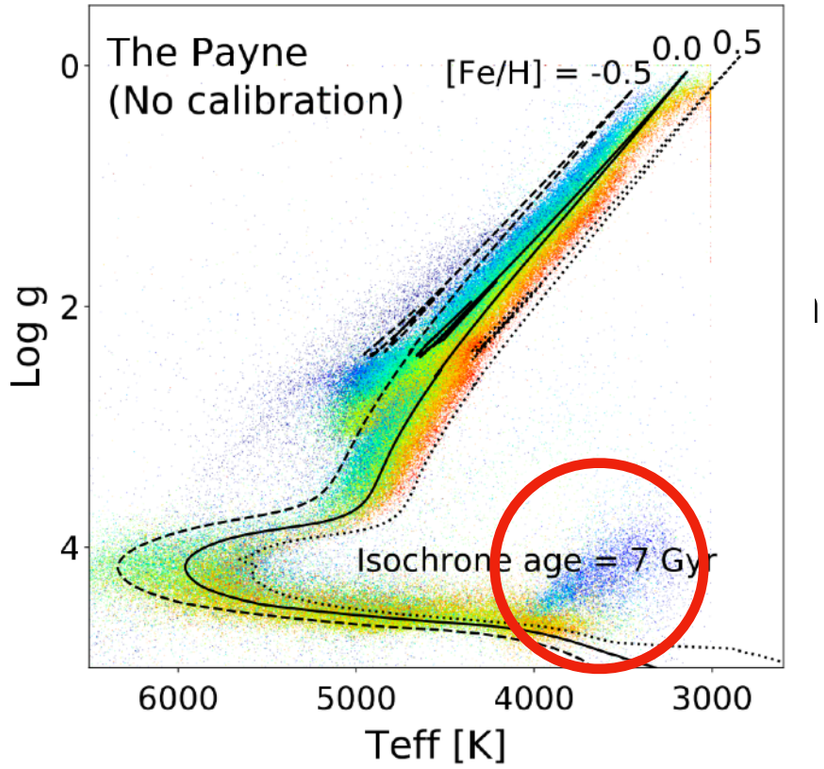
Ting et al. (2019)

The pitfall of physical generalisation: the synthetic gap

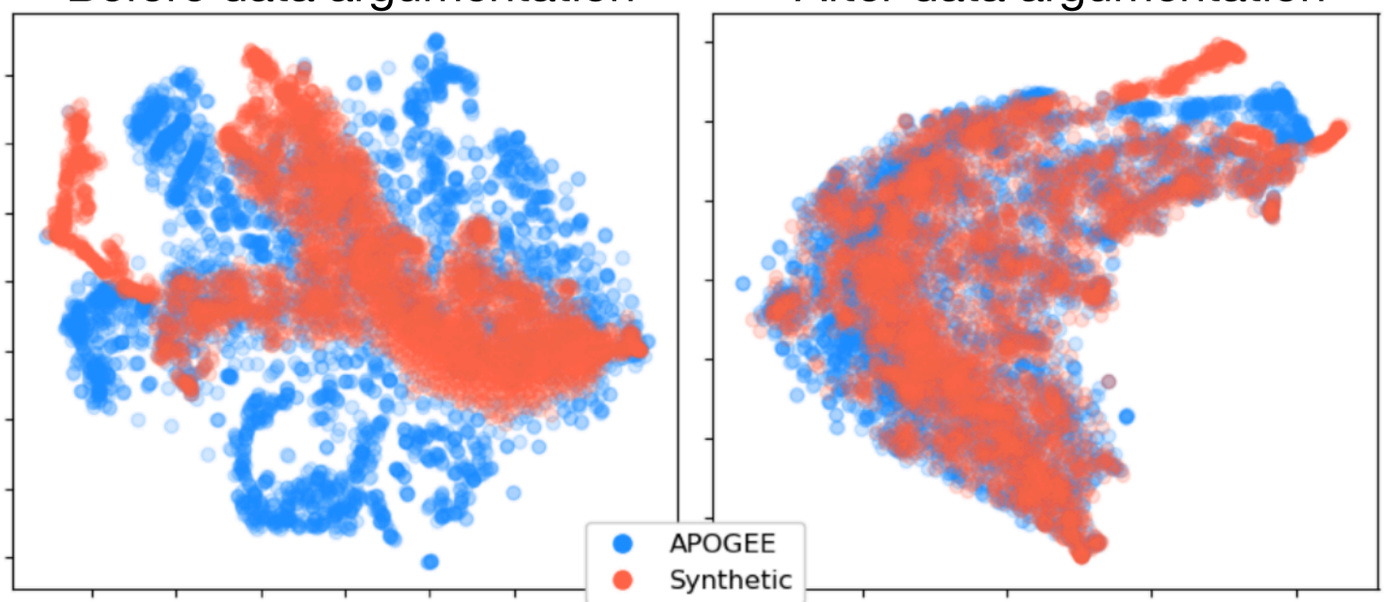
Models are good approximations of reality, but they are not perfect.
Synthetic gap: differences in feature distributions between synthetic ar

Tips to reduce the synthetic gap:

- pre-processing the spectra: matching the resolution and sampling of scale and removing the continuum.
- argumenting the spectra data set: adding Gaussian noise, rotational a regions, and zeroing flux values to mimic bad pixels

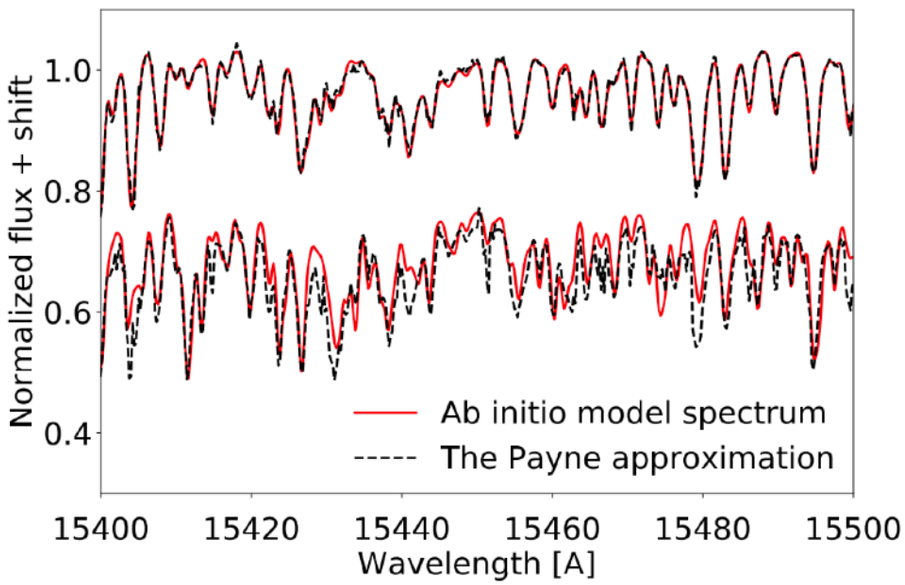


t-SNE visualization of the synthetic and APOGEE spectra



Fabbro et al. (2017)

Spectra generalisation



Ting et al. (2019)

The two approaches with machine learning

Pro. Better causual inference.
Cons. Models suffer from systematics and neglect part of physics.

Pro. Real data are a better representation of the real world.
Cons. Surveys always suffer of selection functions and hiddden biases

The two approaches with machine learning

Physical generalization

Data-driven

The Payne

Sweet spot?

The Cannon

Pro. Better causual inference.
Cons. Models suffer from systematics and neglect part of physics.

Pro. Real data are a better representation of the real world.
Cons. Surveys always suffer of selection functions and hiddden biases

Feature transfer

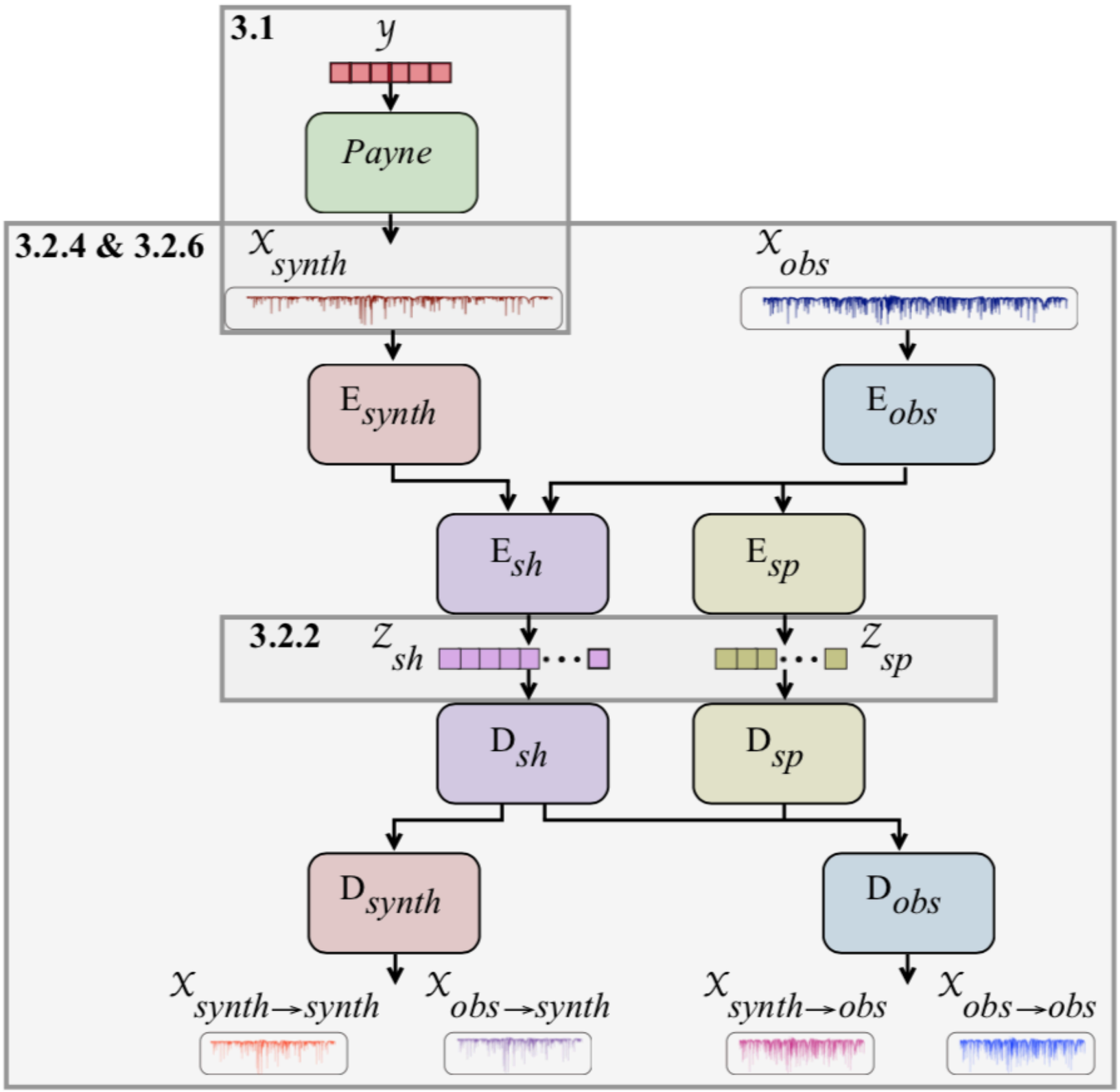
Feature transfer

Feature transfer

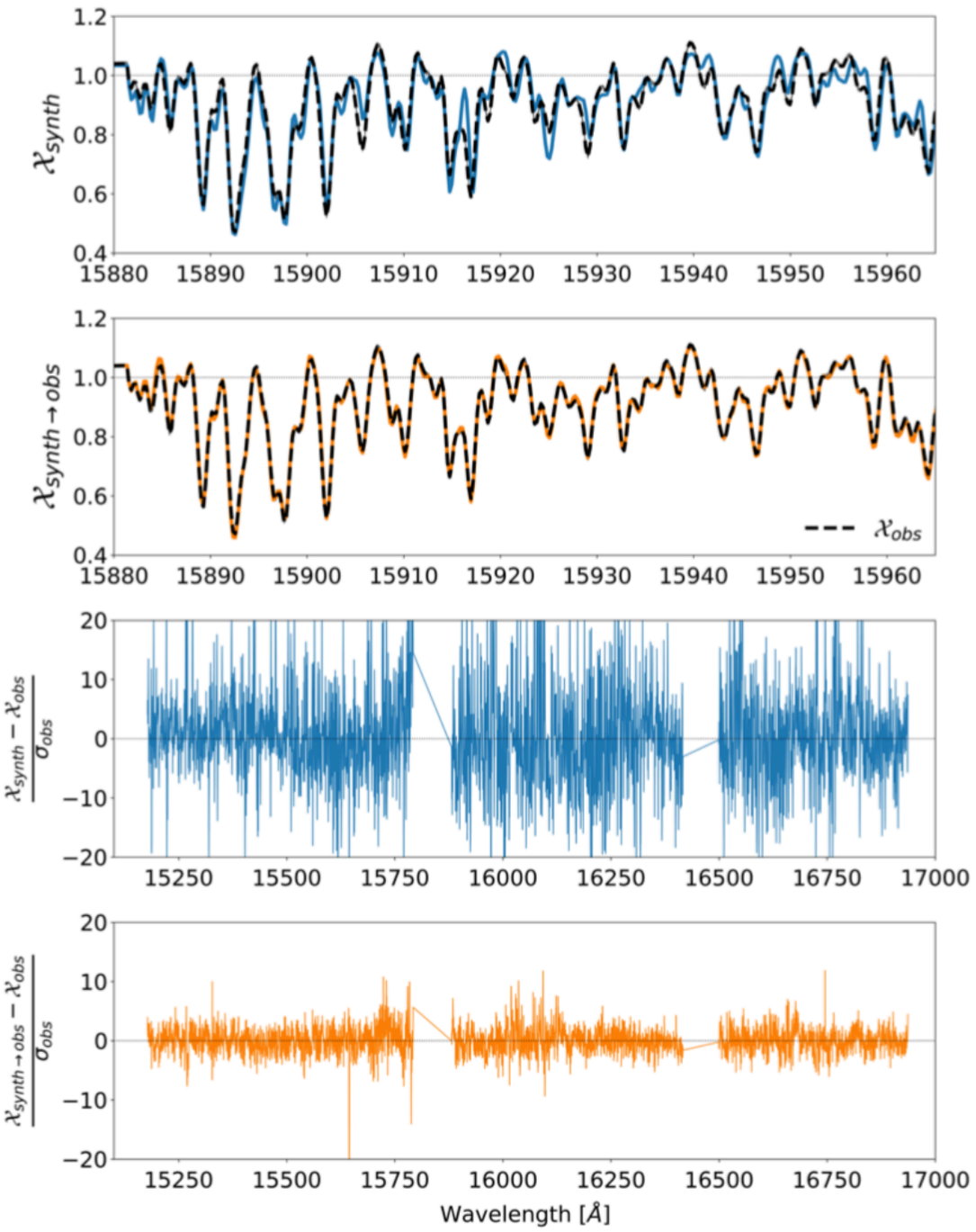
Feature transfer

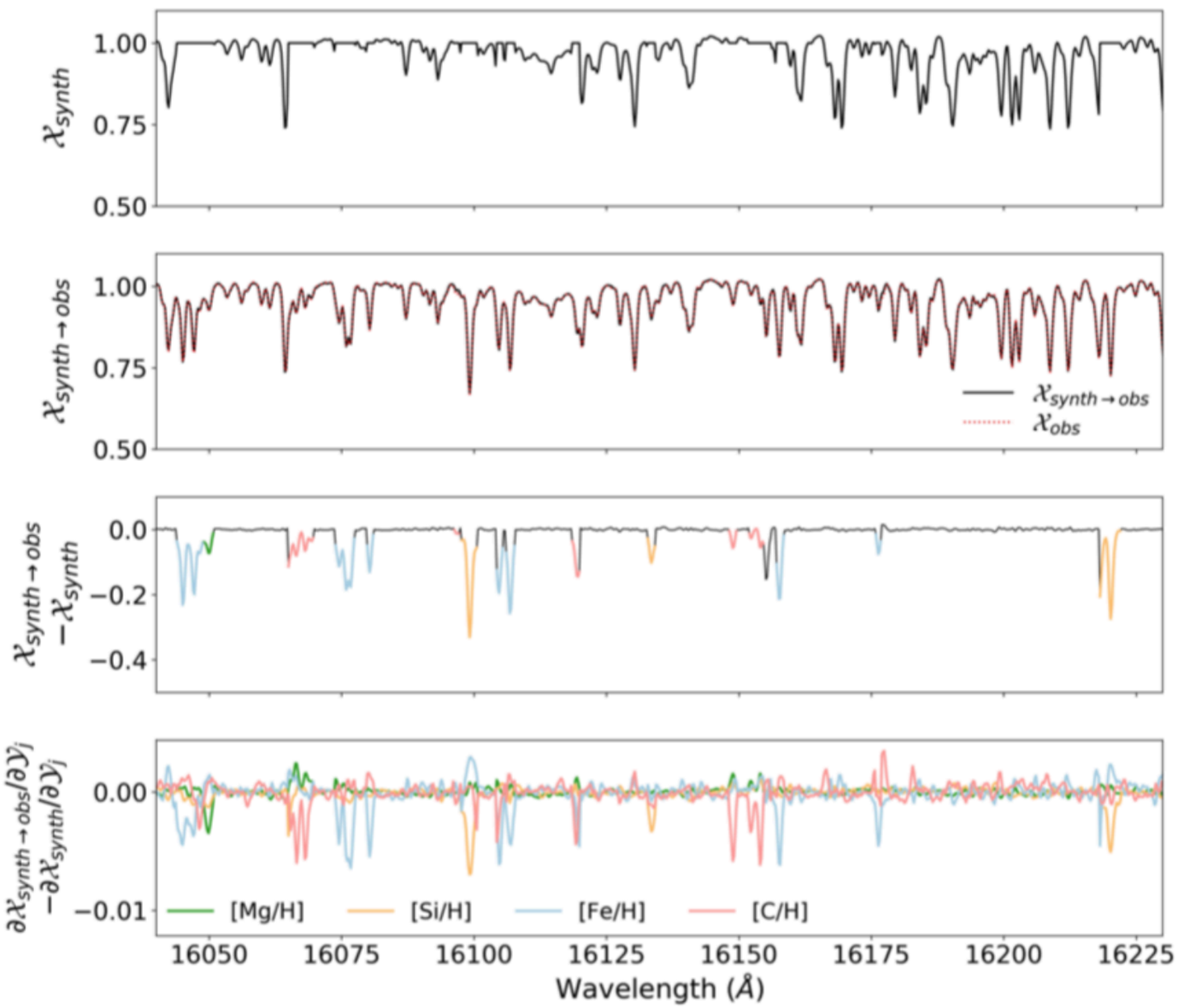
Cycle-Starnet

O'Briain et al. (2021)



Cycle-Starnet O'Briain et al. (2021)





Cycle-Starnet

Limitations:

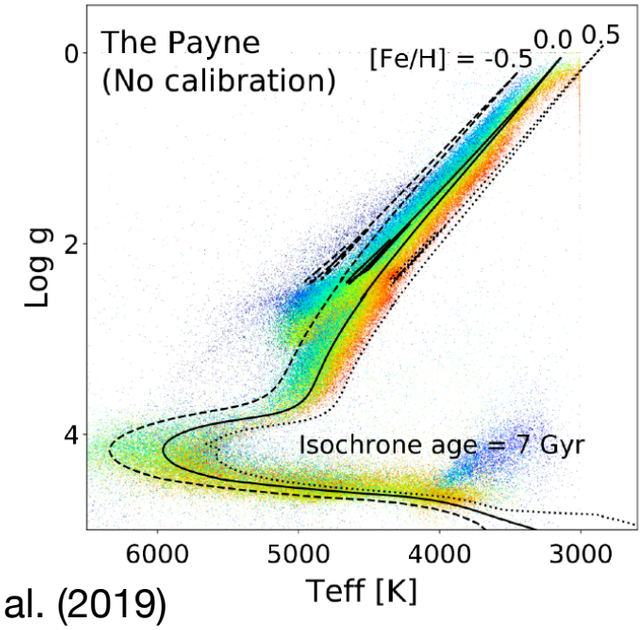
- It is not always straightforward to make the algorithm converge during training.
- It relies on the assumption that the uncertainties in the observed spectra are perfectly known. How a mis-characterisation of the noise affect the final results is yet to be studied.
- Labels from the synthetic and observad domains must span the same range. However, for a new survey we may not know the range of stellar labels.
- Cycle-Starnet may not extrapolate well for outliers and exotic stellar objects.
- Cycle-Starnet can correct for inaccurate atomic parameters in the linelist. However, this adds much more complexity in the analysis.
- It is impossible to correct for all the possible unknowns. The synthetic gap is reduced but not completely eliminated.

However, the study of Tranfer Features algorithms is currently proceeding at high speed. It is very likely that in the next few years we will have better algorithms that could alleviate all these limitations.

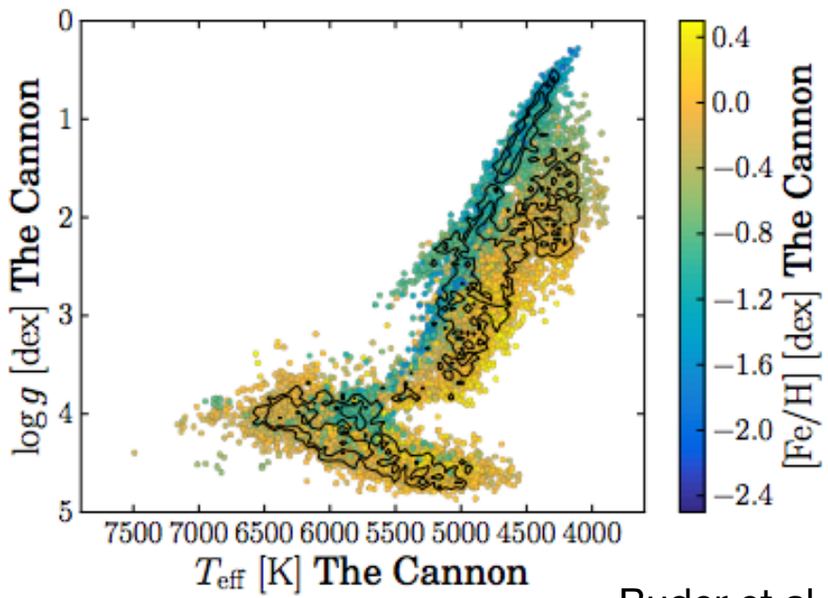
Conclusions

- Machine learning algorithms are a powerful tool.
  Significant computational power, huge amount of data, new algorithms: we can embrace new methods in our research, passing from a model-drive approach to a data-driven approach.

- Machine learning algorithms applied to stellar spectroscopy have some pitfalls and limitations.
  Data-driven models: bias and correlations.
  Physical generalisation models: synthetic gap.
  These are unavoidable pitfalls. However there are codes that achieve reasonably good results.



Ting et al. (2019)



Buder et al. (2018)

...also see Nandakumar et al. (2020) for a nice example of data homogeneisation (GALAH-APOGEE) with the Cannon.

- There is a huge effort in researching new machine learning algorithms. The "state-of-the-art" is fluid and can evolve very rapidly. Stay tuned!