The IWG7 4MOST Galactic Pipeline (4GP): Development Report

Dominic Ford, Lund Observatory

21 July 2017

The 4MOST Galactic Pipeline (4GP) is a collection of python modules which wrap spectral synthesis and processing tools including Turbospectrum, 4FS, radial-velocity estimating codes, and the Cannon[i]. A common data format is used to store spectra and their associated metadata, including the ability to search for spectra by arbitrary metadata constraints. This allows spectra to be easily passed between any combination of tools. A simple web interface allows the contents of spectrum libraries to be searched and viewed quickly for diagnostic purposes.

In due course we will use this framework to develop a fully automated pipeline for extracting radial velocities, stellar parameters and elemental abundances from observed spectra. Presently, we are using it to characterise the accuracy and speed with which the Cannon can operate when trained on synthetic spectra and tested on noisy synthetic spectra. We have also developed a working pipeline for determining radial velocities, but in the present report all of the spectra are in the object rest frame.

The code and associated documentation is available on GitHub:

https://github.com/dcf21/4most-4gp
https://github.com/dcf21/4most-4gp-scripts

The Cannon is a data-driven code which uses machine-learning algorithms to learn how each pixel within a spectrum correlates with each parameter which is to be derived. *Parameter* in this context may refer to any numerical quantity – e.g. $T_{\rm eff}$, log(g) and [Fe/H], as well as the abundances of particular elements.

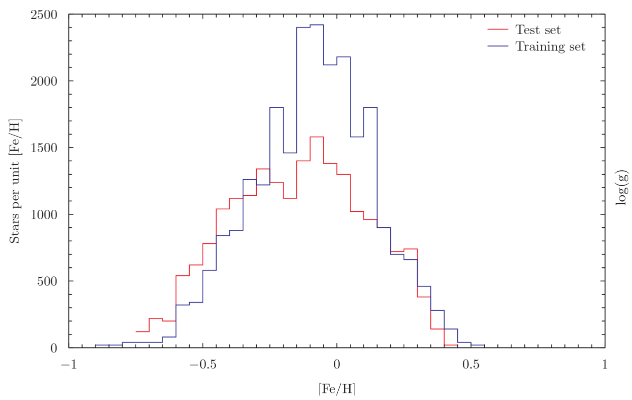
The Cannon requires a training set of stars with known parameters in addition to the test stars whose parameters are to be determined.

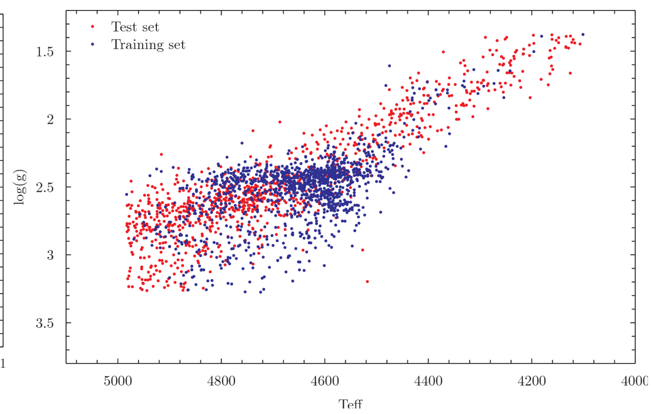
Input Spectra

For the purposes of our tests, we use synthetic spectra for a sample of 1,232 training stars and 1,001 test stars. To make them as realistic as possible, the parameters used to synthesise each star are taken from real giant stars with $T_{\rm eff} < 5000$ K catalogued by APOGEE. This sample is identical to the stars used by Casey & Hawkins in their preliminary tests in 2016.[ii]

The synthetic spectra were generated using Turbospectrum and span the wavelength range 3700 Å to 9500 Å at a spectral resolution of 50,000. This is significantly higher resolution than 4MOST HRS, and is selected so that the output can be reused for future tests at a range of wavelengths. Each spectrum takes ~ 7 minutes to synthesise on a single CPU core, so synthesising the complete sample takes roughly 10 CPU days.

The parameter-space coverage of the test and training sets are shown below:





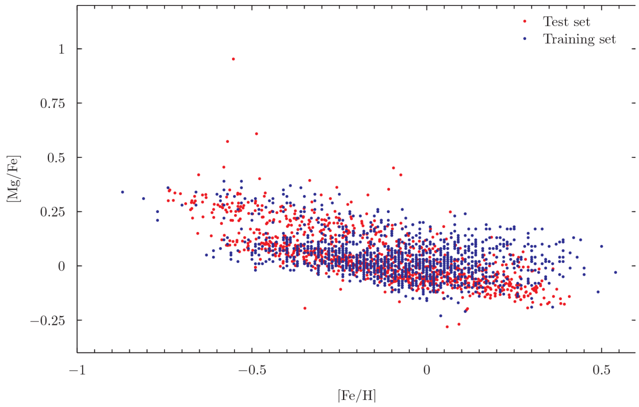


Figure 1: The parameter space coverage of the test and training sets.

To add noise to the synthetic spectra, we use the 4FS Exposure Time Calculator (ETC). This can calculate the length of observation needed to observe each object to any given signal to noise ratio, including a model of the sky transmission and emission at Paranal. It also down-samples the synthetic spectra to the resolution expected from 4MOST, and accounts for the wavelength-dependent sensitivity of each of 4MOST's detector chips.

The process is illustrated in Figure 2 for one of the test stars in our sample. Panel (a) shows the synthesised flux for the test object, as returned by Turbospectrum. Panel (b) shows the shape of the continuum, as provided by Turbospectrum (see below). Panels (c) and (d) show the degraded spectra returned by 4FS for 4MOST's HRS and LRS modes respectively. The three wavelength arms within each mode are shown in different colours.

For each object, the Cannon requires a single spectrum, sampled on a common raster of wavelengths shared with all the input spectra. To achieve this, we stitch the spectra from the three wavelength arms together, as shown in panels (e) and (f). In the LRS mode, the wavelength arms overlap, and we made the arbitrary decision to splice them at wavelengths of 5327.7 Å and 7031.7 Å. These splicing points are identical to those used by Casey & Hawkins (2016).

Defining SNR

The nominal SNR of a spectrum needs to be precisely defined, because the SNR is wavelength dependent. The SNR of the spectrum discussed above is shown in panels (g) and (h) on the next page.

Following Casey & Hawkins (2016), we define the nominal SNR to be the median SNR per pixel in the range 6180 Å to 6680 Å. This region was chosen because it is common to both LRS and HRS, and it does not fall near the edges of any chips, where the SNR is atypically low. The 4FS configuration settings allow us to define SNR either per pixel or per Å; we chose the former in order to reproduce the earlier tests as closely as possible.

The nominal SNR of the spectrum shown here is 250 per pixel. In our tests, we degraded each test spectrum to seven SNRs per pixel: 5, 10, 15, 20, 50, 100 and 250. The training spectra were only degraded to an SNR per pixel of 250; we never train the Cannon on spectra with a SNR lower than this.

This process takes ~ 5 seconds per spectrum on a single core, including both LRS and HRS modes at all seven SNRs. Degrading the full sample of 2,233 stars takes around three CPU hours in total.

Continuum normalisation

The Cannon requires its input spectra to be continuum normalised. Applying continuum normalisation to the output of 4FS is not straightforward, especially at the edges of the chips, where the flux is low.

Turbospectrum produces the input spectra in both continuum-normalised and non-continuum-normalised form, and so as a work-around, we extract the continuum (with no lines) by dividing the latter by the former. This is shown in panel (b) of Figure 2. Passing this through a separate run of 4FS gives a continuum-only output spectra. Dividing the spectrum with lines by that without lines gives a perfectly continuum-normalised output from 4FS.

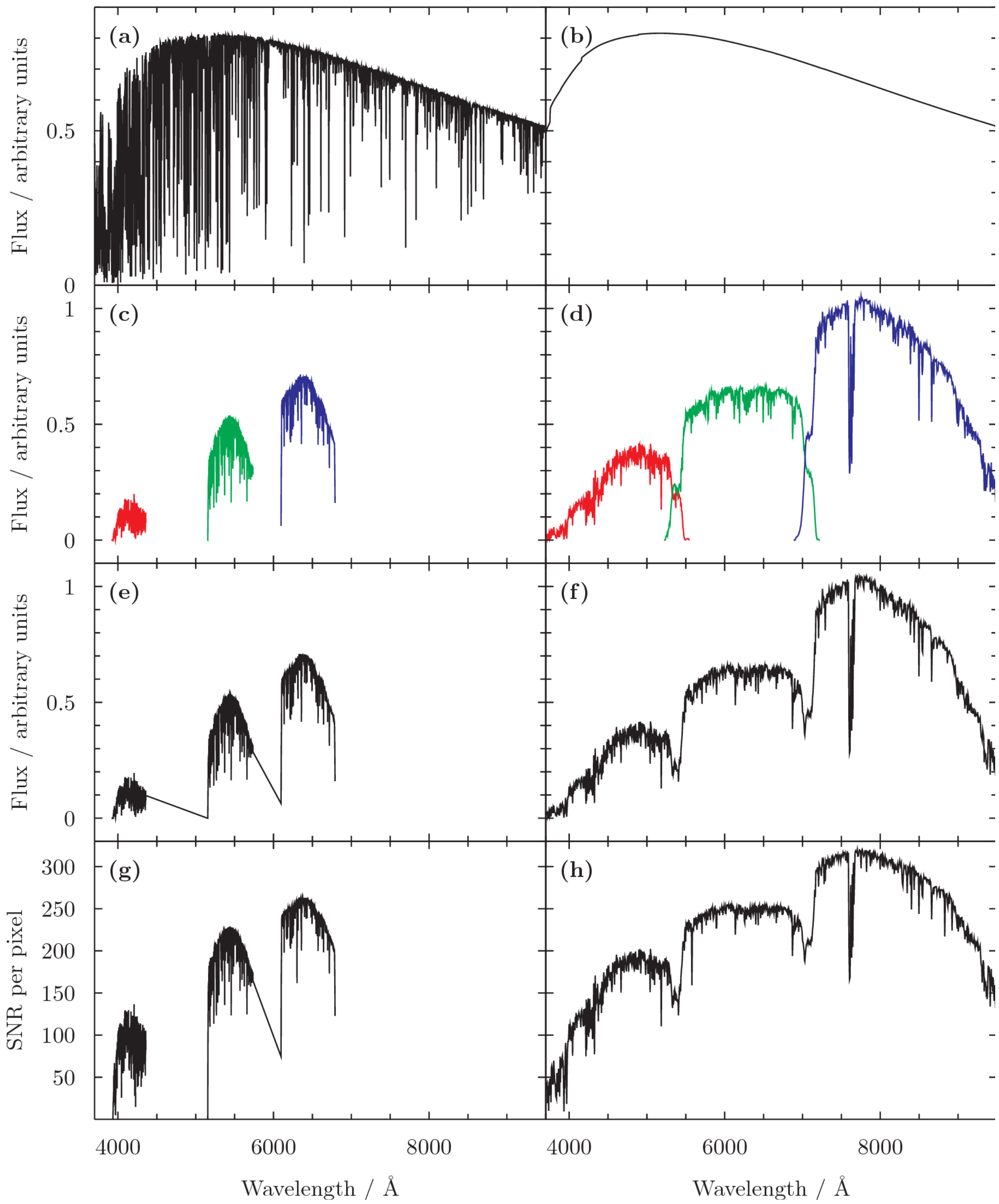


Figure 2: The process of putting an example synthetic spectrum through 4FS. See the text for details. The star shown here has the APOGEE ID: 2M22552267+5331345.

Accuracy of the Cannon's parameter determination

The Cannon takes roughly 12 hours on a quad-core computer to estimate parameters from 7,007 spectra (i.e. 1,001 stars degraded to seven SNRs). Of this time, two hours is spent training the Cannon, and the remainder is spent working through the test spectra.

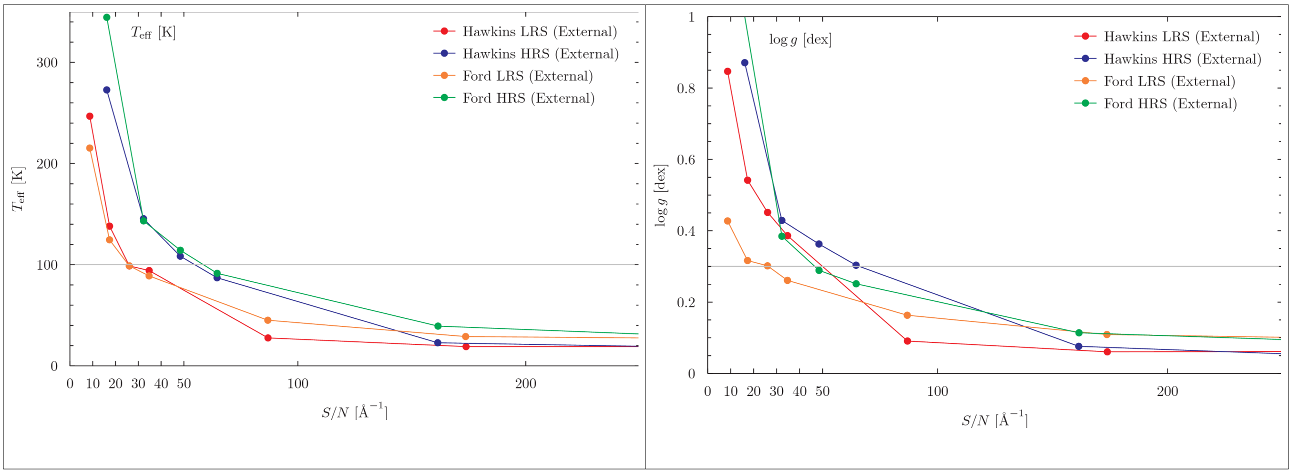
Since the training step is a one-off expense, this indicates that to meet the eventual computational needs of 4MOST, roughly one quad-core server will be required to run the Cannon per every 15,000 spectra that are to be analysed each day.

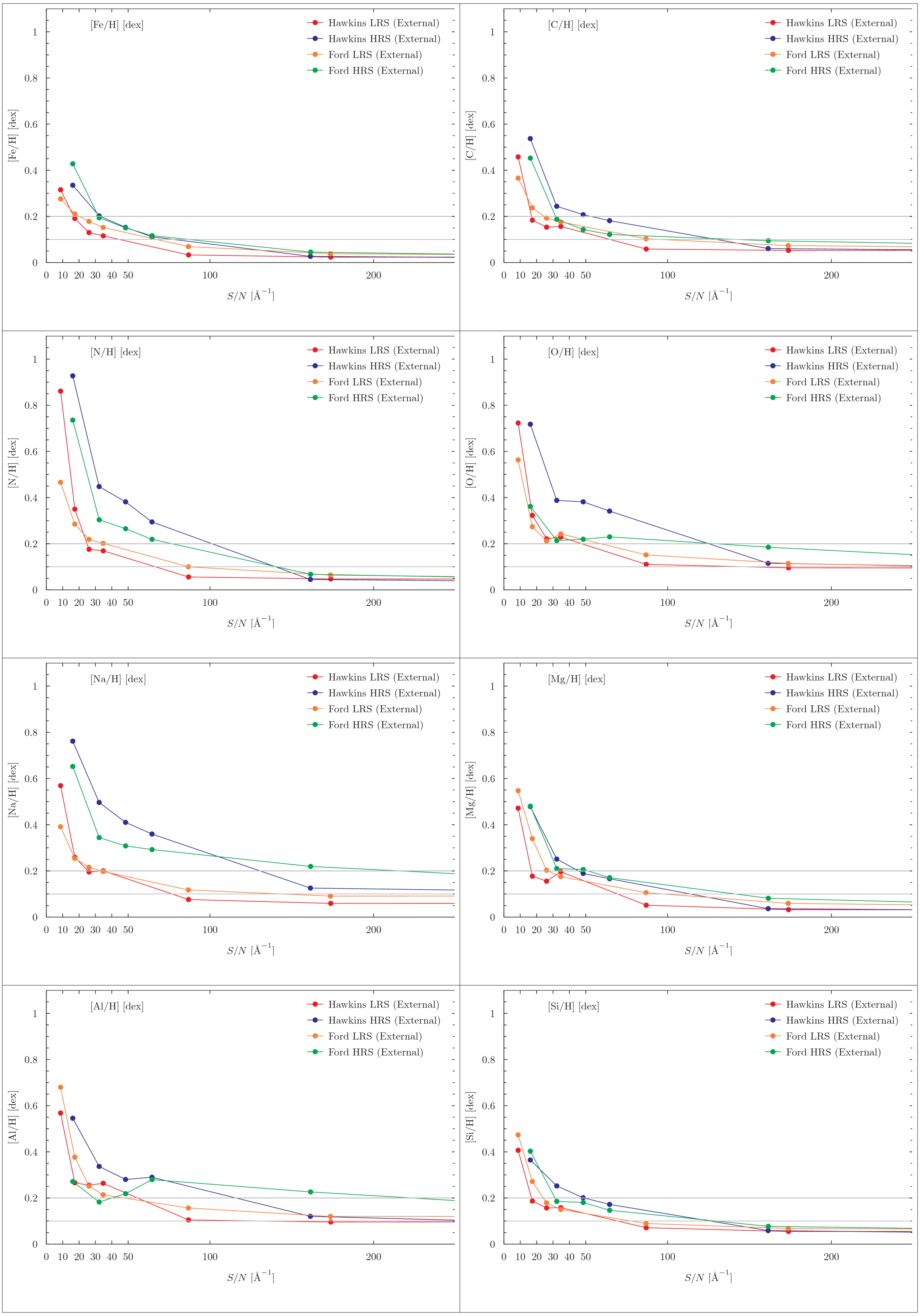
The plots below show the mean absolute offset of the parameter values estimated by the Cannon, from those used to synthesise the spectra. This is plotted against the nominal SNR per Angstrom to which the input spectra were degraded. Each data point is an average of all 1,001 test objects. The conversion from SNR/pixel to SNR/Å is achieved by dividing by the square root of the median number of pixels per Å across the spectrum.

Grey horizontal lines show the target accuracies for each parameter, as quoted by Thomas Bensby at the DFDR meeting. These are 0.1-0.2 dex for individual elemental abundances, 100 K for T_{eff}, and 0.3 dex for log(g).

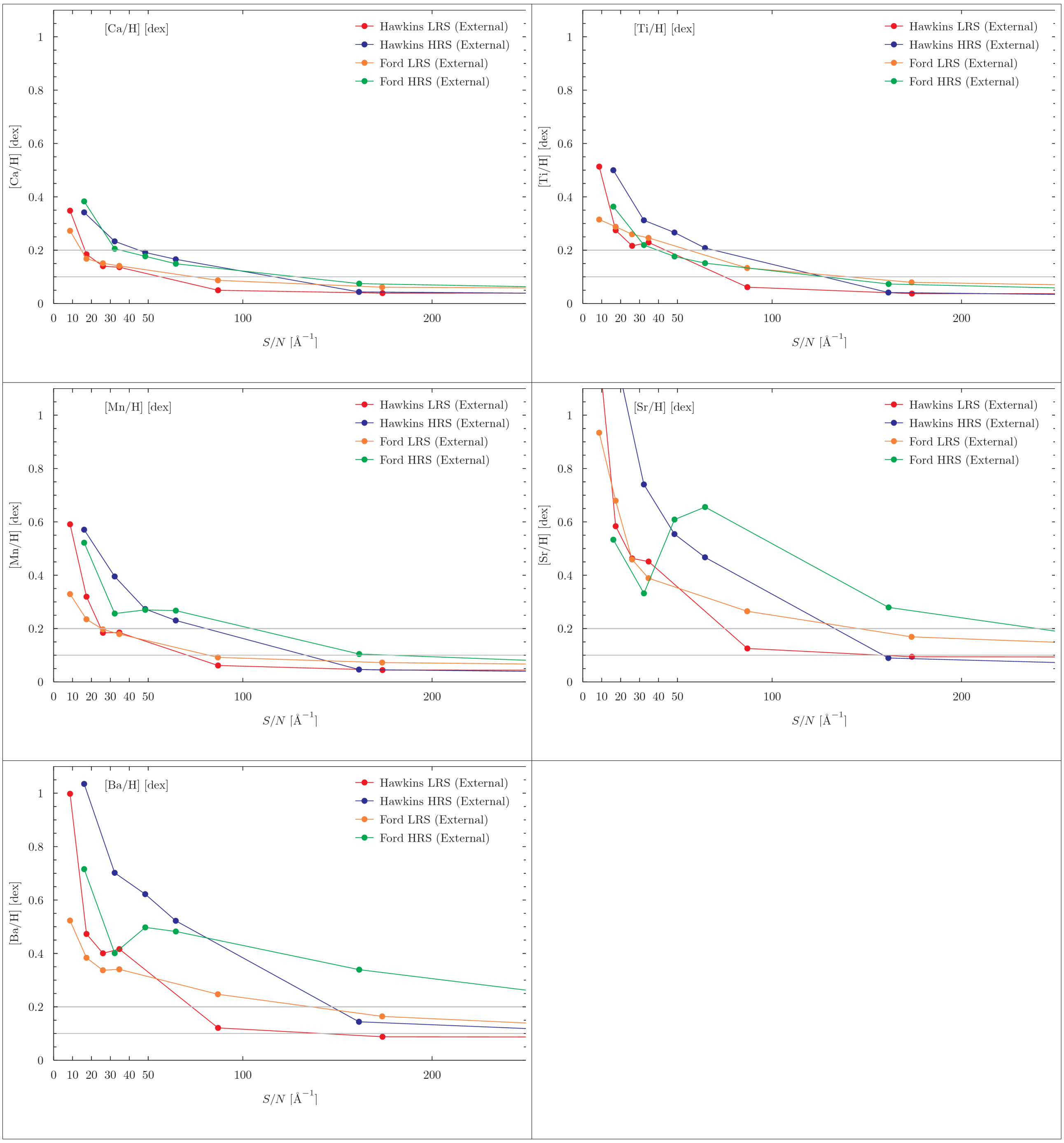
The lines labelled "Ford (External)" are based on spectra which were synthesised and degraded by Dominic Ford as described above. The lines labelled "Hawkins (External)" were constructed by exactly the same method, and using the same sample of stars, but we used the degraded spectra generated by Casey & Hawkins (2016). These plots may slightly differ from their original report, since we have repeated the step of parameter determination by training the Cannon on their spectra.

Although there is some scatter between the Ford and Hawkins data sets, both are broadly consistent that LRS spectra require a SNR/Å > 50 to achieve the target parameter accuracy for all of the elements listed below. For HRS spectra, this increases to around SNR/Å > 100. This difference can probably be explained because the HRS spectra cover roughly half the wavelength span of LRS spectra, and the Cannon is good at combining statistical evidence from many lines across the whole observed wavelength range. Also, more flux is needed at any given stellar magnitude to achieve the same SNR at higher resolution.





[Fe/H] [dex]
[C/H] [dex]
[N/H] [dex]
[O/H] [dex]
[Na/H] [dex]
[Mg/H] [dex]
[Al/H] [dex]
[Si/H] [dex]
Hawkins LRS (External)
Hawkins HRS (External)
Ford LRS (External)
Ford HRS (External)
S/N [Å^{-1}]



Next steps

We plan to generalise the tests to dwarf stars in addition to the sample of giants considered here. We also intend to investigate more thoroughly the performance of the code on different stellar types and metallicity regimes.

In reality, 4MOST spectra will be subject to many kinds of noise whose properties are quite different from that modelled here. In coming weeks we will test the sensitivity of the Cannon to these other sources of noise by repeating the simulations while varying the kinds of degradation applied to the spectra (as performed on individual lines in the requirements document[iii]). This may include, for example:

- Replacing some small fraction of the light in each test spectrum with a solar spectrum, to simulate stray light.

- Enabling 4FS's facility to add sky emission into the test spectra, which is currently disabled. This will simulate poor sky subtraction.
- Adding additional weak lines into the test spectra at random positions.
- Applying small radial velocity offsets to the test spectra, to simulate bad RV measurements, or bad wavelength calibration.

References

i Ness, M. et al. 2015, ApJ, 808, 16 – http://adsabs.harvard.edu/abs/2015ApJ...808...16N

ii Available on the 4MOST Docushare:
https://ds-web.aip.de/docushare/dsweb/Get/Document-5284/20160715_First%20Results.pdf

iii Document MST-SPE-PSC-20307-9237-0001 on the Docushare:
https://ds-web.aip.de/docushare/dsweb/Get/Document-3091/MST-SPE-PSC-20307-9237-0001_1_00.pdf