

Towards spectral fidelity

Reconstructing the spectroscopic instrumental profile using astrocombs

D. Milaković^{1,2} and P. Jethwa³

¹ Institute for Fundamental Physics of the Universe, Via Beirut, 2, 34151 Trieste, Italy

² INAF, Osservatorio Astronomico di Trieste, via Tiepolo 11, 34131, Trieste, Italy

e-mail: dinko@milakovic.net

³ University of Vienna, Department of Astrophysics, Türkenschanzstraße 17, A-1180 Vienna, Austria

Received September 15, 1996; accepted March 16, 1997

ABSTRACT

Context. Making accurate measurements of the wavelengths and the shapes of features observed in high resolution astronomical spectra requires the use of a spectrograph whose instrumental profile (IP) shape is well known and well characterised.

Aims. This work presents empirical IP models for the extremely stable HARPS spectrograph, installed at the ESO 3.6m telescope, and the improvements they bring for making accurate spectroscopic measurements.

Methods. The astronomical Laser Frequency Comb installed on HARPS produces ≈ 10000 bright emission lines covering 60% of the spectrograph wavelength range. These “astrocomb” lines are unresolved and are of known and stable wavelengths, making them excellent probes of the monochromatic IP over a large fraction of the detector. New methods, based on Gaussian Process regression, were developed to construct empirical IP models from astrocomb lines.

Results. We found that the HARPS IP is asymmetric over a large fraction of its detector and that this asymmetry changes smoothly. When empirical IPs are used to wavelength calibrate HARPS and to fit spectral features in the data, the accuracy of wavelength measurements is better than 3 m s^{-1} . When the Gaussian IP approximation is used instead, it is at least 14 m s^{-1} , almost five times worse. Applying the empirical HARPS IPs to existing AI-VPFIT models of the absorption system at $z = 1.15$ towards HE0515–4414 results in a systematic shift in the fine structure constant measurements (in this system) that is comparable to the statistical uncertainty.

Conclusions. Methods presented here will be crucial for obtaining objective and unbiased measurements of fundamental constants from high resolution spectra, as well as measurements of the redshift drift, isotopic abundances, and other science cases. They can also be applied to other instruments equipped with astrocombs, such as VLT/ESPRESSO and ELT/ANDES in the future. The empirical IP models are made freely available.

Key words. Instrumentation: spectrographs – methods: data analysis – techniques: spectroscopic – (cosmology:) cosmological parameters – (cosmology:) dark energy

1. Introduction

Astronomical studies using high resolution spectroscopy compare the positions and/or the shapes of spectral features seen in the spectra of astronomical objects, where the exact details of the comparison being made depends on the specific science case. Every kind of comparison imposes slightly different technical requirements on the spectrographs, which are commonly expressed in terms of instrument accuracy, precision, and stability. Recently, the term “spectral fidelity” came into use¹, referring to the spectrograph’s capability to produce spectra that have been minimally impacted by the spectrograph properties. High fidelity spectrographs aim to produce spectra that are the best possible representation of the true object spectrum such that the results derived from the analysis of the produced spectra do not depend on anything else except for the properties of the object being examined.

Knowing all relevant properties of the instrument impulse response function, i.e. the spectrograph’s instrument profile (IP), is crucial for achieving high spectral fidelity. Many laboratory tests

and on-sky experiments demonstrated that the best possible performance is achieved when the shape of the spectrograph IP is fully decoupled from tiny changes in the instrument, telescope, and observing conditions Baudrand & Walker (2001); Grupp (2003); Walker et al. (2003); Jovanovic et al. (2017); Oliva et al. (2019). Significant effort was therefore invested into ensuring that the IP shape is invariant under even the smallest changes in the illumination of the spectrograph entrance pupil.

Due to the nature of making spectroscopic measurements, instrument accuracy, precision, and stability (either for a single line or an ensemble of lines) are most commonly investigated by examining the properties of the instrument’s wavelength calibration. Because of their close similarities to an ensemble of ideal lines (unresolved and stable in time), astronomical Laser Frequency Comb calibration lamps (“astrocombs”) are especially suited for this purpose (Hänsch 2006; Murphy et al. 2007; Steinmetz et al. 2008; Wilken et al. 2010; Milaković et al. 2020).

Astrocombs have been used to demonstrate an ensemble wavelength calibration precision of $\approx 3 \text{ cm s}^{-1}$ (Wilken et al. 2012) by averaging over 10^5 lines observed by the HARPS spectrograph (Mayor et al. 2003). This was later improved to 1 cm s^{-1} with comparable instrument stability over a period of six hours (Milaković et al. 2020; Probst et al. 2020). Instrument precision

¹ For example, the “Spectral Fidelity” conference took place in Florence (Italy) in September 2023, <https://www.eso.org/sci/meetings/2023/fidelity.html>

of the same magnitude was demonstrated on the ESPRESSO spectrograph using an ensemble of 10^5 Fabry-Pérot etalon emission lines (Schmidt et al. 2021).

Wavelength calibration accuracy provided by the astrocombs was seldom investigated, also because they are superior to almost all other absolute wavelength calibration systems. A worrying result then emerged from a direct comparison between two independent astrocombs, operated simultaneously on HARPS. The comparison of the two astrocombs revealed an approximate 60 cm s^{-1} difference in the zero-points of their absolute wavelength calibrations for the instrument (Probst et al. 2020; Milaković et al. 2020). In other words, the ensemble instrument accuracy and stability depended on which astrocomb was used to measure it, with the difference between the two measurements being significantly larger than the uncertainty (approximately 1 cm s^{-1}). This has important implications for all scientific studies requiring long term (>1 year) ensemble instrument stability at the 1 cm s^{-1} level, in particular those planned with the future ANDES spectrograph (Marconi et al. 2022).

The exact causes for this 60 cm s^{-1} difference were never fully identified, although there were some indications on what they may be. Astrocomb lines were observed to collectively shift by as much as 15 m s^{-1} due to changes in the flux level of the astrocomb exposure (Milaković et al. 2020; Zhao et al. 2021). Interpreting these shifts as being caused by flux-dependent IP variations, and correcting for them using a simple model, the differences between the absolute calibrations of the two astrocomb were reduced to 40 cm s^{-1} (Milaković et al. 2020). This indicates that further improvements could be made if the IP shape is known, and also if its dependence on the relevant parameters, including flux, is measured. Ultimately, the best – and possibly the only – way to achieve high spectral fidelity is by knowing the IP shape.

The paper is organised as follows. **To be finished.**

2. The instrumental profile of a spectrograph

Most spectrographs have an approximately Gaussian instrumental profile, with strong departures from this shape generally caused by optical design choices or (rarely) defects in the spectrograph construction. The IP shape has received comparatively little attention so far, for two main reasons. Firstly, it was assumed that using a somewhat incorrect IP shape does not impact on the scientific results as long as that shape is applied consistently, especially when comparing multi-epoch observations in which spectral features appear at approximately the same locations on the detector (e.g. exoplanet studies). Secondly, the ideal way for IP shape reconstruction is by observing a set of unresolved, uniformly distributed and unblendend spectral features covering a large fraction of the instrument’s spectral range and such sources just recently became available.

The IP for several instruments was previously reconstructed from spectra in which an absorption cell containing molecular iodine (I_2) was inserted in the light path while observing astronomical objects (Marcy & Butler 1992; Valenti et al. 1995; Kambe et al. 2002). The I_2 cell methods has the benefit of using the light following the same optical path as the scientific observations so the reconstructed IP is exactly the one that appears in the data. However, the strong absorption from I_2 makes it unfeasible to use when observing faint targets. Another drawback comes from the fact that the lines in I_2 spectrum are blended even when observed at spectral resolution of $\sim 10^6$, complicating IP reconstruction. A comprehensive description on the use of I_2 cell for IP reconstruction can be found in Butler et al. (1996).

A different approach is offered by the astrocomb system, as the astrocomb lines have almost all necessary properties to make them suitable tools for IP reconstruction, as we explain in Sec. 2.1. Several other groups have recently used an astrocomb for this purpose Zhao et al. (2019); Hirano et al. (2020) and to characterise how the IP flux dependence impacts on the measured instrument stability (Zhao et al. 2021). A possible drawback associated to using astrocomb lines is that they appear in emission whereas the scientific analysis is most often performed on (comparably fainter) absorption lines. Therefore, the IP reconstructed from the astrocomb spectrum may not be the same as the IP appearing in absorption. A comparison between the IP reconstructed from data in emission and data in absorption is planned to be made for the ESPRESSO spectrograph, which will be reported on in a separate publication.

2.1. Astrocomb as a tool for IP reconstruction

The Laser Frequency Comb (LFC, Udem et al. 2002; Hänsch 2006) produces a train of emission lines (also known as modes) with frequencies determined by the equation

$$f_n = f_0 + n \times f_r, \quad (1)$$

where f_n is the frequency of the n^{th} mode and f_0 and f_r are the offset and repetition frequencies. Both f_0 and f_r are actively controlled frequencies in the MHz range, stabilised against a radio-frequency standard such as an atomic clock (see, e.g., Probst 2015, for more details). As such, the LFC line frequencies are known with a relative precision of $\approx 10^{-12}$ as long as a frequency lock with the reference is maintained. The accuracy and stability of the frequency standards can be transferred to the spectrograph, making LFCs the ideal wavelength calibration for astronomical spectrographs.

LFCs specifically adapted for astronomical spectrograph calibration are named “astrocombs” and have f_n and f_r in the GHz range to match the spectrograph resolving power. First astrocomb prototypes were tested on the German Vacuum Tower Telescope (Steinmetz et al. 2008) and the HARPS (Wilken et al. 2010) spectrographs. The left panel of Fig. 1 shows a small part of the astrocomb spectrum observed by HARPS.

In the context of IP reconstruction considered here, astrocombs show another useful property: the comb lines are intrinsically extremely narrow, 300 kHz to 500 kHz (Tilo Steinmetz, private communication). This corresponds to a relative uncertainty of 5×10^{-10} to 1×10^{-9} (or $\lesssim 1 \text{ m s}^{-1}$, when scaled by the speed of light). For comparison, the resolution element width of a spectrograph with spectral resolving power $R = \lambda/\delta\lambda = f/\delta f = 120\,000$ (where $\delta\lambda$ and δf are the full width at half maximum, FWHM, of the resolution element in wavelength and frequency space, respectively) at 550 nm is $\delta f = 4.54 \text{ GHz}$. This is a factor of 10 000 broader than the intrinsic astrocomb line width. The astrocomb lines are therefore completely unresolved by even the highest resolution astronomical spectrographs and, as such, individual modes can be considered monochromatic impulse inputs to the optical system. Their image on the detector is therefore a direct digitized representation of the 2d IP. Correspondingly, the extracted astrocomb spectrum then represents the 1d IP. For example, the middle panel of Fig. 1 shows a stack of flux normalised astrocomb lines observed by HARPS on top of their best-fit Gaussian centres.

Lastly, the astrocomb lines are high signal-to-noise (S/N), as seen in the right panel of Fig. 1. They also have very uniform flux levels: the dynamic range of the HARPS astrocomb

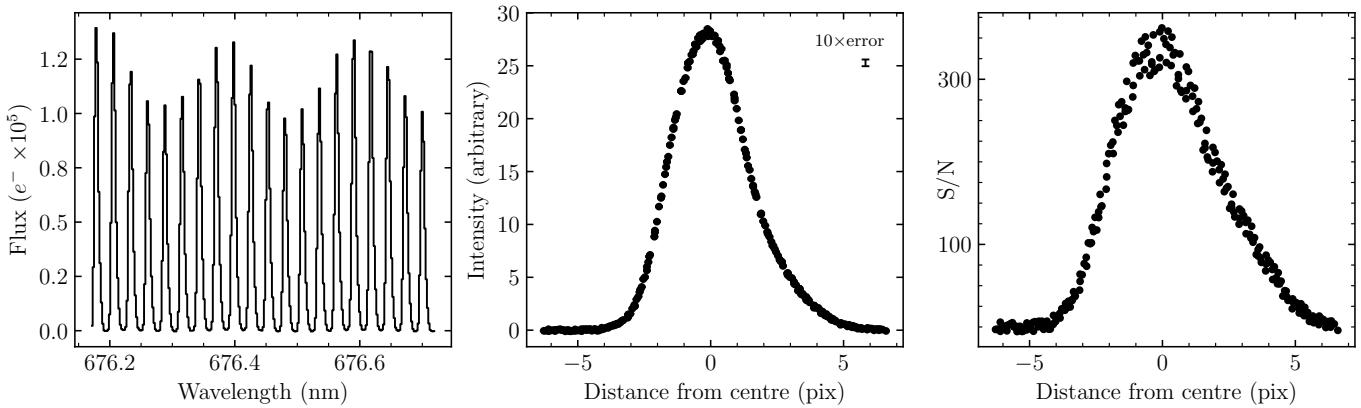


Fig. 1. *Left panel:* The extracted HARPS astrocomb spectrum observed by HARPS in the wavelength range $676.2 \text{ nm} \leq \lambda \leq 676.7 \text{ nm}$ containing 20 astrocomb modes, blaze-corrected and background-subtracted. *Middle panel:* The same lines as in the left panel have been flux normalised and stacked on their best-fit Gaussian centres (see Sec. ??). The result is a remarkably well sampled representation of the 1d IP in this wavelength range with 256 points. The sub-pixel sampling is achieved due to the IP falling slightly differently with respect to pixel centre for each line. The right-skewed asymmetry in the IP shape is immediately noticeable. The error-bar in the top right corner shows 10 times the median error on individual points (enlarged for visibility). *Right panel:* Individual points reach S/N of several hundred across a large fraction of the HARPS wavelength range.

is 0.000 27 dB from 500 nm to 690 nm. We are therefore able to use a single astrocomb calibration frame to reconstruct the IP without needing to worry about its variation with recorded light intensity within a single frame. In fact, exposures taken with different flux levels can be used to determine the relation between light intensity and IP shape.

2.2. The digitised IP

This work was heavily influenced by the paper by Anderson & King (2000), henceforth referred to as A&K. We were surprised to realise that, while they were interested in performing accurate astrometric measurements and hence the 2-dimensional IP and we are interested in accurate wavelength measurements and hence the 1-dimensional IP, most of the considerations made by the authors apply here unchanged.

Similarly to Valenti et al. (1995) previously, A&K considered the fact that a pixelated detector does not directly record information on the *true* instrumental profile (their “iPSF”). Instead, what is recorded is the integrated product of the iPSF with the 2-dimensional pixel sensitivity profile (\mathcal{R}), where the integration limits are determined by the pixel’s boundaries (see their section 3 and formulae (1)-(5), specifically). This led them to define the *effective* IP (their “ePSF”) as the convolution of the iPSF with \mathcal{R} . The digitized image of a point source on the detector then provides samples of ePSF, with the values in individual pixels determined precisely by the distance of the centre of the pixel from the point source, multiplied by some flux factor intrinsic to the source.

Importantly, A&K showed that one does not need to determine the iPSF to measure the positions of features with sub-pixel accuracy. Instead, determining ePSF is sufficient because the recorded pixel values – the only observable – are already integrated over the pixel area with appropriate weights determined by the pixel response function. Going further, A&K demonstrated how repeated observations of the same point sources, each offset from the other by $\lesssim 1$ pix, can be used to reconstruct ePSF on a subpixel scale and its variation across the detector.

Their approach has three major advantages, as noted by the authors themselves. Firstly, it avoids integrating the true IP over the pixel area when fitting for the position of the feature (the cen-

tre of a star image in their case, and the centre of an astrocomb line for us), as the integration over the area of the pixel is already included into the definition of the effective IP. Secondly, fewer computational resources are needed to solve for the effective IP than the true IP itself. Finally, because of the definition of the ePSF as the convolution of iPSF and \mathcal{R} , the latter is automatically incorporated into the modelling process and needs not to be untangled from the recorded point source image.

We agreed with the reasoning of A&K above, and have applied it to 1d astrocomb spectrum in order to reconstruct the 1-dimensional IP. All references to the IP in the rest of the paper therefore refer to the effective IP, unless stated otherwise.

2.3. The 1-dimensional effective IP

The flux in an extracted spectral pixel is given by:

$$F_i = f_* \int_{-\infty}^{+\infty} \psi_I(x - x_*) \times Q(x - i) dx + B_i, \quad (2)$$

where the index i identifies the pixel centred at $x = i$ (in a single echelle order). The values f_* and x_* are the total integrated flux (brightness) and the position of a monochromatic light source, respectively, and B_i is the background level. $\psi_I(\Delta x)$ is the 1-dimensional true IP, i.e. it quantifies the fraction of light (per 1d pixel unit area) that falls onto the point offset by Δx from x_* . Equation (2) is analogous to A&K’s equation (1) and our ψ_I is a 1d analogue of their iPSF.

Q replaces A&K’s \mathcal{R} in our equations. In addition to the 2d pixel response function contained in the latter, Q carries the information on all other operations applied to the data by the data acquisition and spectral extraction processes: e.g. bias subtraction, flat-fielding, weighted summation of pixels in the spatial direction, etc. Untangling contributions of each of these operations in 1d spectra is almost impossible in practice, and doing so would require a full 2d approach. However, like with \mathcal{R} , it is not necessary to know Q ’s value – as long as it is (nearly) unchanged between the data used for IP reconstruction and the scientific data to which the IP is applied, which is the case for observations considered here.

Applying the same transformations as in A&K, appropriately modified for 1d, we obtain:

$$F_i = f_* \psi_E(i - x_*) + B_i \quad (3)$$

where:

$$\psi_E(\Delta x) = \int_{-\infty}^{+\infty} Q'(x) \psi_I(\Delta x - x) dx, \quad (4)$$

and $Q'(x) = Q(-x)$. Equation (4) is the 1d analogue of A&K's ePSF.

2.4. A non-parametric functional form for the IP

Previous studies modelled the non-Gaussian shapes using different modifications to the Gaussian, including modifying the wings (Cardelli et al. 1990) or adding satellite Gaussian profiles (Valenti et al. 1995; Zhao et al. 2019). While we tried out some of these approaches, we were unable to obtain satisfying results. Ultimately, we decided to use Gaussian Process (GP) regression to obtain a smooth function that is most likely representation of the HARPS instrumental profile. This is similar to what Hirano et al. (2020) have done for the InfraRed Doppler spectrograph on the Subaru telescope.

2.4.1. Gaussian Process regression

GP regression is a technique for probabilistic, non-parametric regression (Rasmussen & Williams 2006). Here we provide a basic overview; see Aigrain & Foreman-Mackey (2022) for a detailed review in an astronomical context. Given pairs of data-points (x_i, y_i) with noise ϵ_i , the goal of regression is to find a function f such that $y_i = f(x_i) + \epsilon_i$. A GP is a prior over this space of functions. The defining property of a GP is that for any finite sample of points \mathbf{x} , the sampled function values $\mathbf{f} = \{f(\mathbf{x})\}$ have a multivariate normal probability distribution

$$p(\mathbf{f}) = \mathcal{N}(\mathbf{m}, \mathbf{K}), \quad (5)$$

which is parameterised by a mean vector \mathbf{m} and covariance matrix \mathbf{K} . These are, in turn, specified by a mean function m and covariance function k , where

$$m_i = m(x_i; \theta), \quad (6)$$

$$K_{ij} = k(x_i, x_j; \phi), \quad (7)$$

which depend on hyperparameters θ and ϕ respectively. Assuming that the noise ϵ_i is normally distributed with zero mean and scale σ_i , i.e.

$$p(\epsilon_i) = \mathcal{N}(0, \sigma_i), \quad (8)$$

then the likelihood of the data \mathbf{y} as a function of the hyperparameters is given by

$$\mathcal{L}(\theta, \psi) = \mathcal{N}(\mathbf{y}; \mathbf{m}, \mathbf{K} + \text{diag}(\{\sigma_i\})). \quad (9)$$

Given observed data (x_i, y_i) , training the GP means to find the hyperparameters (θ, ϕ) which maximise this likelihood. Once trained, the GP encodes a probability distribution over suitable regression functions f . Ensembles of functions f can be sampled from the trained GP (see Aigrain & Foreman-Mackey (2022) for details), and the variance amongst these functions encodes the uncertainty in the regression. The specific choices of mean functions and covariance functions that we adopt in this work are introduced in later sections: in section 3.2 we describe a GP model for the line profile, and in section 4.6, a GP to describe the observational errors.

3. Methods

The instrumental profile can be estimated from the astrocomb spectra by inverting Eq. (3):

$$\hat{\psi}(\Delta x) = \frac{F_i - B_i}{f_*}. \quad (10)$$

where we have now dropped the subscript “E” on ψ_E for the ease of notation. The hat above ψ indicates estimation from observations. In the equation above, $\Delta x = i - x_*$ is the distance between the pixel and astrocomb line centres and f_* is the integrated astrocomb line flux. Knowing x_* and f_* allows for determining at which offset Δx has the pixel centered at $x = i$ sampled ψ and how the sampling has been scaled. Stacking the observations on top of each other provides subpixel sampling of ψ , as shown in the middle panel of Fig. 1. The implicit assumption in performing the stacking is that ψ does not change across the range covered by the lines being stacked, but we relax this assumption later by allowing ψ to be different at the location of each astrocomb line.

When performing the GP regression, we considered also the variance on $\hat{\psi}$. It is, however, not simple to calculate this quantity analytically because it means calculating the variance of the ratio of two random Poisson variables (the first one being $F_i - B_i$ and second one being f_*), one of which that is conditional on the other, i.e. $f_* = f(F_i - B_i)$. Instead, we use the following simple approximation that preserves S/N in the data:

$$\sigma_{\hat{\psi}}^2 = \frac{1}{f_*^2} [\sigma_{F_i}^2 + \sigma_{B_i}^2]. \quad (11)$$

Simple MCMC calculations of the variance on Eq. (10) confirm that this approximation works very well, with differences between values determined through Eq. (11) and the MCMC variance differing by 5% at most, always in lowest flux pixels. We later describe how we used the data themselves to modify $\sigma_{\hat{\psi}}^2$ where necessary.

3.1. Obtaining line position and brightness

Accurate determination of x_* and f_* that go into calculating $\hat{\psi}$ requires that ψ is already known and an accurate estimation of ψ , in turn, requires x_* and f_* to also be known. This motivated the use of an iterative approach.

Initial guesses for x_* and f_* were determined by approximating ψ with a Gaussian profile and fitting it to the data with three free parameters: the amplitude A , the mean μ , and the standard deviation τ of the Gaussian, i.e.

$$I(x; A, \mu, \tau) = \frac{A}{\tau \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\tau}\right)^2\right]. \quad (12)$$

Since we are interested in the effective IP, integration under pixel area when optimizing for the Gaussian parameters was not performed. Instead, we simply evaluated the Gaussian profile at pixel centres. The impact this has on estimating x_* is minimal (the average change in line centres approximately 1×10^{-5} pix, which is negligible). We identify x_* and f_* with the mean and the integrated area under the Gaussian profile, i.e. $x_* = \mu$ and $f_* = A\tau\sqrt{2\pi}$.

In subsequent iterations, once ψ has been modelled using our GP regression methods, the Gaussian approximation is replaced by the formula:

$$I(x; A_*, x_*, \omega) = A_* \psi(\omega(x - x_*)). \quad (13)$$

Here, ψ is the IP normalised to unit area, and A_* , x_* , and ω are free parameters. A_* and x_* are the line amplitude and position and ω is parameter whose value is ≈ 1 allowing for small stretching or compressing of the pixel scale when performing the fit. The line brightness is $f_* = \sum_{i=1}^N I(i)$, where the sum goes over the N pixels of the line.

3.1.1. Fitting an IP model to the data

For all IP models considered in this work, astrocomb line model parameters were determined using non-linear least-squares minimisation of the normalised model residuals of Equations (12) and (13).

Let $I(\zeta)$ be the predicted model flux, where ζ is the set of all free model parameters. Parameter optimization finds ζ which minimises the normalised residuals of the model to the data, i.e.

$$\zeta = \arg \min f(\zeta), \quad (14)$$

where

$$f(\zeta) = \sum_{j=0}^{j=N} \left(\frac{I(\zeta)_j - F_j}{\sigma_j} \times w_j \right)^2. \quad (15)$$

The summation in Eq. (15) goes over the N pixels of the astrocomb line, where the $j = 0$ and $j = N$ correspond to the locations of the minima either side of the line and w is the weight applied to each pixel.

The weighting scheme was implemented to ensure the best possible fit at astrocomb line centres, following the discussion and recommendations made in A&K. In this scheme, a unity weight is given to the central pixels, zero weight to the pixels far from the line centre, and linear weights to pixels in between. The weight of a pixel centred at $x = j$ and at a distance $\Delta x = j - x_*$ from the astrocomb line, is given by

$$w_j = \begin{cases} 1 & \text{for } |\Delta x| \leq x_1 \\ \frac{x_2 - |\Delta x|}{x_2 - x_1} & \text{for } x_1 < |\Delta x| \leq x_2 \\ 0 & \text{for } |\Delta x| > x_2, \end{cases} \quad (16)$$

where $x_1 = 2.5$ pix and $x_2 = 5$ pix. The weighting scheme is implemented when fitting both empirical IP models and the Gaussian IP approximation.

The reported best-fit parameters correspond to the values of ζ that satisfy Eq. (14). Parameter errors were calculated from the diagonal of the covariance matrix at the best-fit solution.

3.2. A Gaussian Process for the instrumental profile

The empirical ψ was modelled using GP regression, conditioning the GP on the observed $\hat{\psi}(\Delta x)$ samples. Because the shapes of astrocomb lines are approximately Gaussian, the mean function was chosen to be

$$m(x; \theta) = A \mathcal{N}(x; \mu, \tau) + y_0 \quad (17)$$

where \mathcal{N} represents the standard normal probability density function. This function takes parameters $\theta = (A, \mu, \tau, y_0)$ corresponding to the normalisation, the mean, the standard deviation, and the zero-offset, in that order.

The choice of covariance function affects the smoothness of the regression curve. The squared exponential covariance function was chosen, i.e.

$$k_{\text{SE}}(x_i, x_j; \phi) = a^2 \exp \left(-\frac{|x_i - x_j|^2}{2l^2} \right), \quad (18)$$

which depends on the hyperparameters $\phi = (a, l)$ i.e. an amplitude and a length-scale. The amplitude controls how far the regression curve can stray from the mean function, while the length-scale determines typical length of wiggles in the curve.

An additional diagonal term was added to the covariance matrix in order to account for the uncertainties on the data:

$$k(x_i, x_j; \phi) = k_{\text{SE}}(x_i, x_j; \phi) + \sigma_i \sigma_j \delta_{ij}, \quad (19)$$

where $\sigma_i = \sqrt{\sigma_{\hat{\psi}}^2 + \sigma_0^2}$, and σ_0^2 is an additional error term that is common to all $\hat{\psi}$ samples and is also a free hyperparameter. Finally, δ_{ij} is Kronecker delta.

3.3. Defining the IP centre

Defining the location of the IP centre, i.e. deciding where to set $\Delta x = 0$, is crucial for line centre x_* measurements and, by extension, the construction of the IP models themselves. For an unimodal and symmetric IP profile, any of the mean, mode, or median can be meaningfully used as the centre. While these quantities are also well defined for asymmetric profiles, they can be very different from each other and also be sensitive to small changes in the shape of the profile.

We wanted our centre estimator be robust against small changes in hyperparameter values. We thus set to identify the optimal centre estimator for use in this work by comparing the properties the mean, the mode, the median, and the centre definition from Anderson & King (2000). In this last case, the IP centre is asserted to be at the location which results in equal fluxes for the two brightest pixels: $\psi(\Delta x = 0.5) = \psi(\Delta x = -0.5)$.

To find the most robust estimator, a single IP model was perturbed 1000 times and the four centre estimators were calculated for each perturbed model. The details of the comparison can be found in Appendix A. The results show that A&K's centre estimate is the most robust of the four examined, shifting by $\leq 5 \times 10^{-8}$ pix across the 1000 perturbed profiles. For comparison, the mode was the least stable estimator with shifts $\geq 1 \times 10^{-3}$ pix.

We also followed A&K's suggestion to apply a rigid shift to the $\hat{\psi}$ samples in Δx to ensure that the IP models are properly centred before modelling. The transformation applied was:

$$\Delta x \rightarrow \Delta x + \frac{\psi(0.5) - \psi(-0.5)}{\frac{\partial \psi}{\partial x}(-0.5) + \frac{\partial \psi}{\partial x}(0.5)}. \quad (20)$$

Subsequently to the data shifting, GP hyperparameters were optimised once more and the GP was retrained on the shifted data. The data shifting and hyperparameter optimization procedure was repeated until the difference between consecutive shifts was smaller than 1×10^{-3} pixor for a maximum of twenty times (the stopping criterion was typically satisfied after three to four repetitions). This ψ at the end of this procedure was the final model of a single iteration.

4. Reconstructing the HARPS IP

4.1. Data

The methods described in the previous Sections were applied the High Accuracy Radial-velocity Planet Searcher (HARPS) spectrograph, and its instrumental profile was reconstructed from a single astrocomb calibration frame taken as a part of an observing programme to measure the fine structure constant (α) at $z = 1.15$ towards the quasar HE0515–4414 (ESO observing

programme 0102.A-0697, PI Milaković). The primary reason for choosing HARPS is that the IP models could be applied to the science spectra in order to explore their impact on measurements of α (Sec. 6).

While ESPRESSO observations of the same quasar are also available and astrocomb calibrated, those observations required significant manual interventions in the data (Murphy et al. 2022) for reasons not yet fully understood (ESPRESSO consortium, private communication). Additionally, the wavelength calibration of ESPRESSO contains systematic effects which are also not yet understood. The most obvious example is the high frequency correlation with wavelength in the residuals of its wavelength calibration. The residuals also are not normally distributed nor do they have properties expected from photon noise statistics (Schmidt et al. 2021). These problems have not been observed in HARPS.

Finally, the HARPS dataset was modelled using AI-VPFIT (Lee et al. 2021a,b), whereas the ESPRESSO spectrum was not. This is important in light of recent evidence AI-VPFIT models do not suffer from biases present in models obtained by other means (Webb et al. 2022; Lee et al. 2023).

4.1.1. HARPS

HARPS is a fibre-fed echelle spectrograph designed for extremely precise spectroscopic measurements (Mayor et al. 2003), installed at the ESO 3.6m telescope at the La Silla Observatory. HARPS's primary scattering element is an R4 grism which disperses the incoming light into 72 echelle orders, 89 through 161. Optical fibres transport the light from the telescope to the instrument in two channels. One channel (the primary, fibre A) generally carries the astronomical target light. The other channel (the secondary, fibre B) can be used to simultaneously record the sky light or the light from a wavelength calibration lamp. HARPS is equipped with three calibration lamps: a Thorium-Argon (ThAr) arc lamp, a Fabry-Pérot etalon, and an astrocomb, which can be observed in either fibre. The nominal spectral resolution is $R = \lambda/\delta\lambda = 115000$, where $\delta\lambda$ is the FWHM of the instrument resolution element. The entire instrument is enclosed in a temperature and pressure controlled vacuum vessel in order to minimise environmental disturbances and their impact on the scientific observations.

4.1.2. Data description

The astrocomb spectrum used for the modelling was recorded on 2018-12-05 at 08:12:52.040 UT in both fibres. This spectrum was used for wavelength calibration of quasar observations in Milaković et al. (2021) and was thus the obvious choice. The offset and repetition frequencies of the astrocomb were $f_0 = 4.58$ GHz and $f_r = 18$ GHz. The reduced astrocomb spectrum (HARPS pipeline version 3.8) was downloaded from the ESO archive. The pipeline extracts the 1d spectra of each echelle order using optimal extraction (Horne 1986; Robertson 1986), producing separate files for the two channels. We corrected the flux for the detector gain and for the measured spectrograph blaze function².

The spectral variance array was not saved by the pipeline so we estimated it from the flux array assuming Poissonian statistics, i.e. $\sigma_{F_i}^2 = F_i$, where F_i is the flux in i^{th} pixel in electrons. Given the high flux values in the astrocomb spectra, the con-

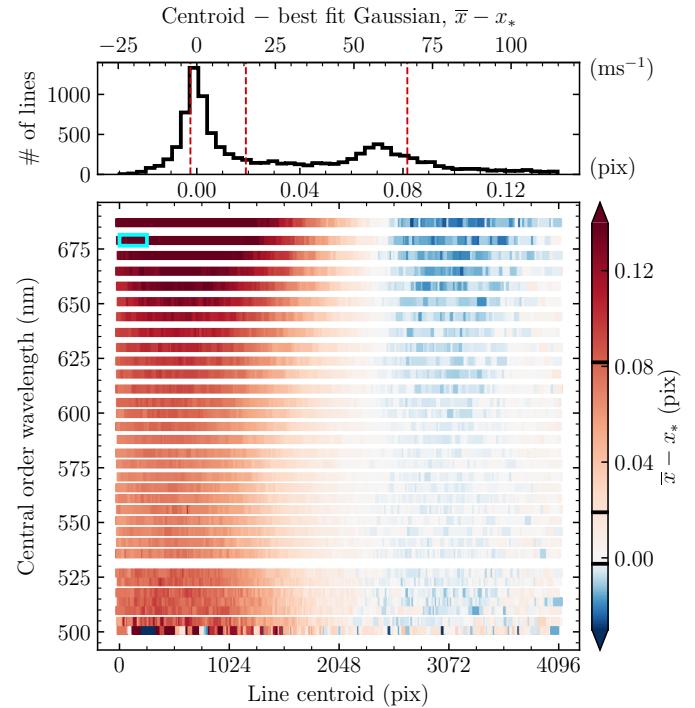


Fig. 2. *Main panel:* Each coloured square shows one of the 10576 detected astrocomb lines in the HARPS LFC spectrum described in the text. The position of the square in the (x, y) plane approximately matches its position on the HARPS detector frame. The gap seen at ≈ 525 nm is due one echelle order falling between the two HARPS detector CCD chips so is not recorded. Wavelength within a single echelle order increases towards right. The colour of the square is proportional to the difference (in units pixels) between the measured line centroid and the line position determined by fitting a Gaussian IP to the data. The zero point of the colour bar to the right is set to have a difference of zero, such that the blue and the red colours denote negative and positive shifts, respectively. Smooth variations across the detector most likely correspond to IP shape variations. The red rectangle indicates the location of the 10th segment in optical order 110, for which we show an empirical IP model in Fig. 3. *Top panel:* Histogram of the values shown in the main panel. The secondary x -axis on the top shows the line centre differences in units m s^{-1} ($1 \text{ pix} = 820 \text{ m s}^{-1}$). The vertical dashed red lines show the positions of the distribution median (0.0219 pix) and the central 50% distribution limits (-0.002 pix and 0.082 pix). The same quantities are shown as thick black lines in the colour bar.

tributions from the dark current and the read-out process were assumed to be negligible.

4.2. Background estimation

The spectral background contains contributions from scattered light ($\approx 3\%$, Rodler & Lo Curto 2019) and from background associated to the astrocomb system itself. The astrocomb background is produced by the high-power amplifier of the system (Probst et al. 2015) and contributes as much as 30% of the total flux below 500 nm (Milaković et al. 2020).

We determined the full spectral background in each echelle order independently. We started by identifying the minima as the locations at which the first derivative of the spectral flux array is closest to zero and, simultaneously, second derivative of the flux array is positive. Before calculating the derivatives, the flux array was upsampled by a factor of 10 and smoothed using a Nuttall window function (Nuttall 1981) to more precisely determine the location of the minima and to avoid falsely detecting minima in

² The relevant blaze function is saved in the observations taken on 2018-12-04 at 20:14:42.379 UT

the noise. The most appropriate width for the window function was determined by calculating the power spectrum of the spectral flux array and identifying the frequency carrying the most power. In practice, the window widths were between 11 and 15 pixels, depending on the wavelength.

The background was then determined by connecting the minima between the lines using straight lines and smoothing the resulting array with a Nuttall window function with a window size of 51 pixels. We assumed that the background follows Poisson statistics, i.e. its variance is $\sigma_{B_i}^2 = B_i$, where B_i refers to the background in the i^{th} pixel.

4.3. Line detection and line properties

Boundaries between individual astrocomb lines were identified by applying the minima detection algorithm to the background-subtracted spectra. 10576 astrocomb lines were detected, with mode numbers between 24087 and 33384 (1440 modes falling in the spectral region covered by two adjacent orders are seen twice). The separation between lines ranges between 10 pix at 500 nm and 18 pix at 690 nm, whereas the FWHM of the best fitting Gaussian for each line was < 3.7 pix everywhere. Lines were therefore sufficiently separated such that line blending is expected to be negligible. This provided $> 100\,000$ samples of $\hat{\psi}(\Delta x)$ across a large fraction of the detector in a single exposure.

Visual examination of lines reveals asymmetry in their shapes, with most lines being skewed to the right, and this asymmetry also appears to vary across the detector. We confirmed this by examining differences between two line centre estimates: the centroid (\bar{x} , a flux-weighted average position of the line) and the centre of the best fit Gaussian IP. This comparison served to examine line asymmetry, as lines which follow an approximately Gaussian IP should show zero difference between these two centre estimates and positive or negative values imply that the line is skewed to the right or to the left, respectively.

The main panel of Fig. 2 shows values of this quantity for all detected lines as coloured squares. Lines appearing in a single order are aligned along the x -axis and echelle orders are separated along the y -axis, so that the location of the squares in the panel approximately corresponds to their location on the detector. Wavelength within each order increases with increasing pixel number (left to right), whereas the central wavelength of the order increases (order number decreases) from the bottom to the top of the panel.

Correlated departures from a symmetrical shape across the detector are immediately noticeable. The most obvious is the division of the detector area into two halves: left, with positive values (coloured red) and right, with negative values (coloured blue). Lines appearing in the middle of each order have values closest to zero (coloured white). Maximal positive departures from the symmetrical shape are seen in the top left of the panel, with values as large as 0.14 pix, and maximal negative departures are seen in the top right of the panel, with values of approximately -0.03 pix. In between the two regions the values vary quickly. Negative values seem to appear around 3000 pix in every order, but in some orders they recover to zero at the very right edge. The width of regions with negative values seem to increase as one moves vertically from the middle of the detector towards the top and the bottom.

The top panel of Fig. 2 shows the histogram of the values from the main panel. The mode of the distribution is around zero pixels, and its median is 0.020 pix. The central 68% of the distribution lies within -0.002 pix and 0.083 pix. A single HARPS

pixel covers a wavelength range of approximately 820 m s^{-1} , so the same quantities can be expressed as velocity shifts. Converting the median and the central 68% distribution limits into velocities, we obtained 16.3 m s^{-1} , -1.7 m s^{-1} , and 67.7 m s^{-1} (respectively). Out of the 10576 lines, 4266 of them (40%) have values within ± 0.01 pix (8.2 m s^{-1}). We conclude from this that the HARPS IP is not symmetric and that its shape varies strongly across the detector in a correlated way.

4.4. Determining the IP locally

To reduce the impact from positional IP variability on our reconstructed models as much as possible, the detector was artificially segmented in both the main and the cross dispersion directions and the IP models was then determined for each segment independently. Echelle orders provide a natural division in the cross dispersion direction and we further divided individual orders into 16 segments, each spanning 256 pix or $\approx 0.5 \text{ nm}$, in the main dispersion direction. The 256 pixel width was chosen to avoid a known “stitching” issue of the HARPS detector which appears as small changes in the detector pixel size every 512 pix in the main dispersion direction and arises from imperfections during CCD construction (Wilken et al. 2010; Bauer et al. 2015; Coffinet et al. 2019; Milaković et al. 2020). One such segment is shown as a red rectangle in Fig. 2.

While we explicitly assumed that IP does not change within the segment when performing the modelling, we allowed for smooth variation of ψ when fitting IP models to the data, as follows. Let the sets $\{\psi_1, \dots, \psi_{16}\}$ and $\{X_1, \dots, X_{16}\}$ denote the 16 IP models calculated within single echelle order and the central segment positions, respectively. We asserted that the models are correct at those locations, i.e. $\psi|_{x=X_i} \equiv \psi_i$, where i is the ordinal number of the segment and x is now a coordinate within the echelle order and not Δx . The IP at an arbitrary x was then interpolated from the two closest ψ models by applying weights inversely proportional to the distance between x and the two nearest segment centres:

$$\psi(\Delta x)|_{X_i < x < X_{i+1}} = \frac{d_2}{D} \psi(\Delta x)|_{X_i} + \frac{d_1}{D} \psi(\Delta x)|_{X_{i+1}}. \quad (21)$$

Here, $d_1 = |x - X_i|$, $d_2 = |x - X_{i+1}|$, and $D = |X_{i+1} - X_i|$. If $x < X_1$ or $x > X_{16}$, no interpolation was done and only the single closest IP was used (ψ_1 or ψ_{16}).

4.5. Hyperparameter determination

To model the IP for one segment, we optimize \mathcal{L} from Eq. (9) using $\hat{\psi}$ samples falling within the segment boundaries to find the maximum likelihood values of the hyperparameters θ and ϕ . The mathematical framework for hyperparameter optimisation was implemented using the GP library `tinygp` (Foreman-Mackey 2023), which was performed using the limited memory Broyden - Fletcher - Goldfarb - Shanno algorithm with boxed constraints (also known as L-BFGS-B, Byrd et al. 1995) as implemented in the Python optimization library `jaxopt` (Blondel et al. 2021).

Parameters (A, μ, τ, y_0) were initialised at the values determined from fitting Eq. (17) to $\hat{\psi}(\Delta x)$ using non-linear least-squares. These parameters were constrained not to move further than ± 5 standard deviations, which were calculated from the covariance matrix diagonal at the best fit solution. We required hyperparameters a , l , and σ_0 to always be positive so they were expressed as logarithms. Their initial values were $\log a = 1$,

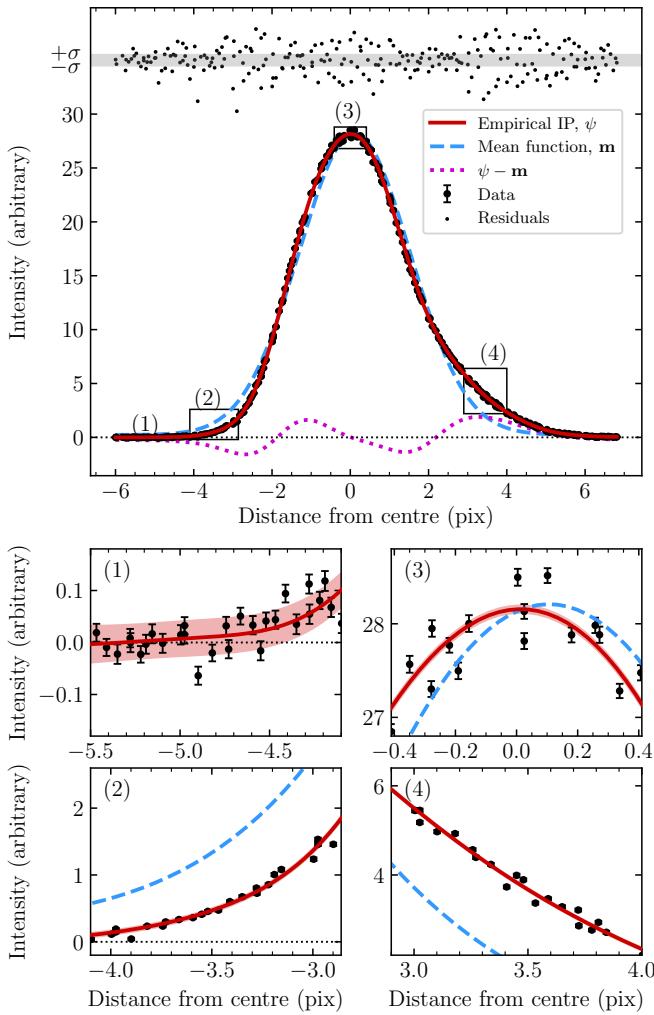


Fig. 3. Gaussian Process fits to the IP. *Top panel:* The large black points are the 245 $\hat{\psi}$ samples with corresponding error bars in the 1st segment of order 90 (average wavelength $\lambda = 676.5$ nm), also indicated as a rectangle in Fig. 2. The thick red line is our reconstruction of ψ using GP regression (Sec. 2.4.1) after recentering, in the first iteration. The secondary GP for variance estimation was not used, resulting in a $\chi^2_v = 7.15$ for this fit. The red shaded regions show the 1σ ranges of the model. The dashed blue line is the mean function of the GP, $\mathbf{m}(\mathbf{x}; \theta)$, Eq. (17). The small black points above the model show the residuals normalised with respect to the error on the data, with the grey shaded regions indicating region $\pm 1\sigma$. Parameters θ (controlling \mathbf{m}) and ϕ (controlling the correlation matrix \mathbf{K}_{ij} , Eq. (18)) were fitted simultaneously. The purple dotted line is the difference between the GP model and the mean function, i.e. it is the function controlled by \mathbf{K}_{ij} . Departures from the Gaussian shape are observed almost everywhere. The thin dotted black line indicates zero flux. Rectangles indicate regions plotted in the bottom panels. *Panels with numbers (1)-(4):* Zoom-ins of the main panel. The number in the top left corner identifies one rectangle in the main panel. The shaded red bands around the GP model show the 1σ uncertainty regions.

$\log l = 0$, and $\log \sigma_0 = -5$. These parameters were constrained to have the following values: $-4 \leq \log a \leq 4$, $-1 \leq \log l \leq 2$, and $-15 \leq \log \sigma_0 \leq 1.5$. Only σ_0 ever reaches its imposed boundary (on the lower side).

After this step, known as training, we can draw sample functions f from the GP evaluated at arbitrary locations, and the vari-

ation of these samples give us an estimate of the uncertainty in the regression function f . The mean f is the IP model, ψ .

Note that we also explored a fully Bayesian treatment, whereby we introduce hyperpriors on θ and ϕ and perform Markov Chain Monte Carlo (MCMC) sampling of the posterior distribution, thus propagating additional uncertainty to the regression function. We found MCMC sampling prohibitively slow for general use (several hours per segment using MCMC compared to one minute per segment using L-BFGS-B). However, our tests suggest that our maximum likelihood uncertainty estimates are very similar to the fully Bayesian approach and the IP models derived through L-BFGS-B are indistinguishable from the MCMC one. More details are provided in Appendix B.

4.6. Empirical variance estimation

We already commented that our $\sigma_{\hat{\psi}}$ estimates, i.e. Eq (11), are formally incorrect. Fig. 3 shows the model derived from $\hat{\psi}$ samples in the 1st segment of order 90 (wavelength range 676.1 nm $\leq \lambda \leq 676.7$ nm) in detail. While the fit is generally good, the residuals scatter beyond what is expected from errors on individual data points. This is most obviously seen in the peak region, where the residuals show correlated scatter outside of the expected $\pm 1\sigma$ range indicated by the gray shaded area in the top of the Figure's main panel. A detailed view, shown in panels (1)-(4) of the same Figure, reveals that the scatter between nearby $\hat{\psi}$ samples is inconsistent with the scatter expected from their estimated errors and that this inconsistency changes with Δx . For example, panel (3) shows that the errors in the region close to the profile peak are underestimated.

The increased scatter also affects the goodness of fit statistic, reduced χ^2 : $\chi^2_v = \chi^2/v$, where v is the number of degrees of freedom in the fit. χ^2 is the weighted sum of squared deviations from the expected value

$$\chi^2 = \sum_{i=1}^N \left(\frac{O_i - E_i}{\sigma_i} \right)^2, \quad (22)$$

where O_i and E_i are the observed and the expected values of the i^{th} observation, and σ_i is its standard error. In this case, $O_i = \hat{\psi}_i$ and $E_i = \psi(\Delta x_i)$ and the summation goes over the N samples of a single segment. For the model shown in Fig. 3, $\chi^2_v = 12.44$, significantly higher than the desirable value of ≈ 1 . Models for other segments had similar χ^2_v values, most likely due to incorrect errors on $\hat{\psi}$. Even though we also obtained models with $\chi^2_v \approx 1$, strong departures of the normalised residuals from the $\pm 1\sigma$ range, such as those seen in Fig. 3, were observed also in those models.

4.6.1. A secondary GP for variances

Since our model currently underestimates the observational variance, our resulting uncertainty estimates on line positions were also likely to be underestimated. We therefore implemented additional methods to deal with this issue by extending our modelling procedures to more correctly capture the observed variance. A second GP was introduced to modify σ_i from Eq. (19), of the form

$$g(\Delta x) \sim \mathcal{N}(C_g, \mathbf{G}_{i,j}), \quad (23)$$

where C_g is a constant hyperparameter and \mathbf{G} is a squared exponential kernel with the amplitude a_g and the length-scale l_g as

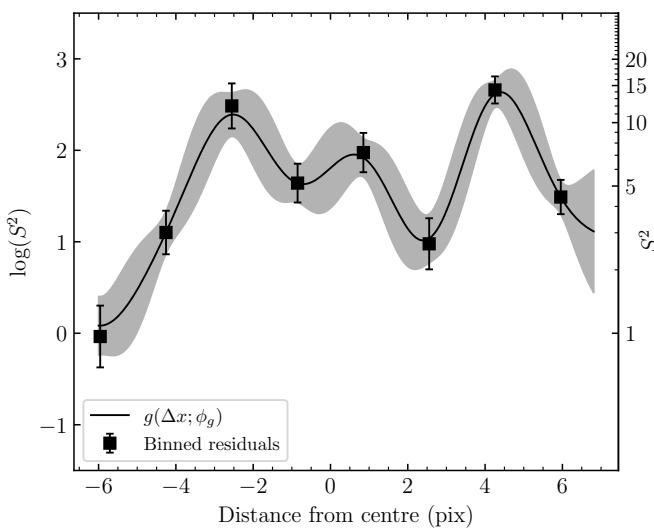


Fig. 4. $g(\Delta x)$ multiplies $\sigma_{\hat{\psi}}$ to capture the observational variances more correctly. Black squares with error bars are the logarithms of (S^2/σ^2) , where S^2 is calculated as explained in the text. The solid black line is the mean prediction of $g(\Delta x)$ from GP regression and the grey shaded are the 1σ ranges. The y-axis label on the right shows S^2 and $g(\Delta x)$ on a linear scale.

hyperparameters. The mean prediction of $g(\Delta x)$ acts as a multiplicative factor modifying the observed variance of $\hat{\psi}_i$ such that σ_i going into Eq. (19) becomes:

$$\sigma_i = \sqrt{g_i \sigma_{\hat{\psi}_i}^2 + \sigma_0^2}, \quad (24)$$

where $\sigma_{\hat{\psi}_i}^2$ is the unmodified error on $\hat{\psi}_i$ and $g_i = g(\Delta x_i)$.

The training data for g was derived from the normalised model residuals, $(\hat{\psi} - \psi)/\sigma_{\hat{\psi}}$, as follows. In the first step, the N residuals were divided into equally spaced bins in Δx such that there were at least 15 points in every bin. Sample variance, S^2 , was then calculated together with the variance on S^2 for each bin:

$$\text{Var}(S^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} S^4 \right). \quad (25)$$

Here, n is the number of residuals in the bin and μ_4 is their fourth central moment.

Hyperparameters $\phi_g = \{C_g, a_g, l_g\}$ were then optimised using equations in Sec. 2.4.1 on pairs of points $(x_b, \log(S^2))$, where x_b are bin centres. Limits on ϕ_g were: $-10 \leq \log C_g \leq 5$, $-3 \leq \log a_g \leq 3$, and $-1 \leq \log l_g \leq 3$. Logarithms were used to ensure $g(x)$ and hyperparameters ϕ_g are always positive. The value $\log(\text{Var}(S^2))$ was added to the diagonal of \mathbf{G} to account for the uncertainty on the variances analogously to Eq. (19).

An example of $g(\Delta x)$ determined in this way is shown in Fig. 4, using the data and the model previously shown in Fig. 3. This determination of $g(\Delta x)$ is also in excellent agreement with its MCMC determination (Appendix B).

Finally, hyperparameters of the GP describing the instrumental profile were recalculated using the empirical variances. Appendix C shows a detailed plot for the same data as in Fig. 3 obtained in this way. The multiplication function applied to the variances on $\hat{\psi}$ is the same as shown in Fig. 4. The improvement from applying this method is obvious: the residuals have improved significantly and show no dependence on Δx and the value of χ^2_ν is much closer to unity. Similar improvements are seen for all segments.

4.7. The final models

We produced ten iterations of IP models in pixel space (later used for wavelength calibration) and in velocity space (later used on scientific observations) for each of the 16 segments of the 32 echelle orders illuminated by the astrocomb. In each iteration, all astrocomb lines were refitted using the latest ψ models, where the models were interpolated to best represent the IP at a specific location within the echelle order.

The data and the procedures used for creating IP models in velocity space were exactly the same as described above (including the secondary GP for variance estimation), except that Δx was expressed in terms of velocity instead of pixel. The velocity of the i^{th} pixel is:

$$\frac{v_i}{c} = \frac{\lambda_i - \lambda_*}{\lambda_*}, \quad (26)$$

where c is the speed of light, and λ_i is the wavelength at pixel centre. In Eq. (26), λ_* is the wavelength at the astrocomb line centre determined by fitting the velocity space IP to the data in exactly the same way as was previously done in pixel space, except that limits for the weights in Eq. (16) have been changed to be $x_1 = 2.5 \text{ km s}^{-1}$ and $x_2 = 5 \text{ km s}^{-1}$.

Wavelengths for all pixels were obtained from the wavelength calibration, which was done independently in each iteration using the latest λ_* estimates.

The reason for determining λ_* from the data and not, for example, setting it to the wavelength predicted by the LFC equation is that we do not *a priori* know where that wavelength appears with respect to the line centre. Similarly, if our wavelength calibration is not perfect, we will not know the exact wavelength at λ_* nor will we know whether it is accurate. This point is expanded upon in Sec. 5.

Surprisingly, we found that the uncertainty on astrocomb line centre does not always decrease with each performed iteration nor does the likelihood of the model always increase. We thus also separately saved a set of 512 “most likely” ψ models from this set of 5120 by choosing the model with the largest likelihood from the set of 10 models produced for one segment.

IP models in pixel space for all 32 orders are shown in Fig. 5, where one order is shown per panel. Looking at any of the panels, the strong variation of the IP shape within the order is immediately noticeable, with the bluest line (corresponding to the segment covering the bluest wavelength range) generally being the most sharply peaked and the reddest line (corresponding to the reddest wavelength covered by that order) being more broad. The IPs in the reddest echelle orders exhibit a tail on their right side that is the strongest for blue lines and that almost completely disappears for the reddest line, changing smoothly with wavelength. Inter-order variations are best appreciated by comparing the bluest lines in each panel, revealing that the aforementioned tail grows with increasing wavelength.

Fig. 6 shows the IP models in velocity space for all investigated orders. Similarly to models in pixel space, the IP in velocity space is obviously asymmetric and the bluest segments in each order exhibit a strong wing at positive velocities that decreases in strength with increasing wavelength. Unlike before, however, the asymmetric wing is prominent also in the bluest echelle orders examined, around 500 nm. Profile width also changes more quickly within a single order than was the case for IP models in pixel space.

All of this is consistent with what is seen in Fig. 2, where the lines in top left corner of the detector showed the strongest signs of asymmetry in their shapes and lines in the right detector

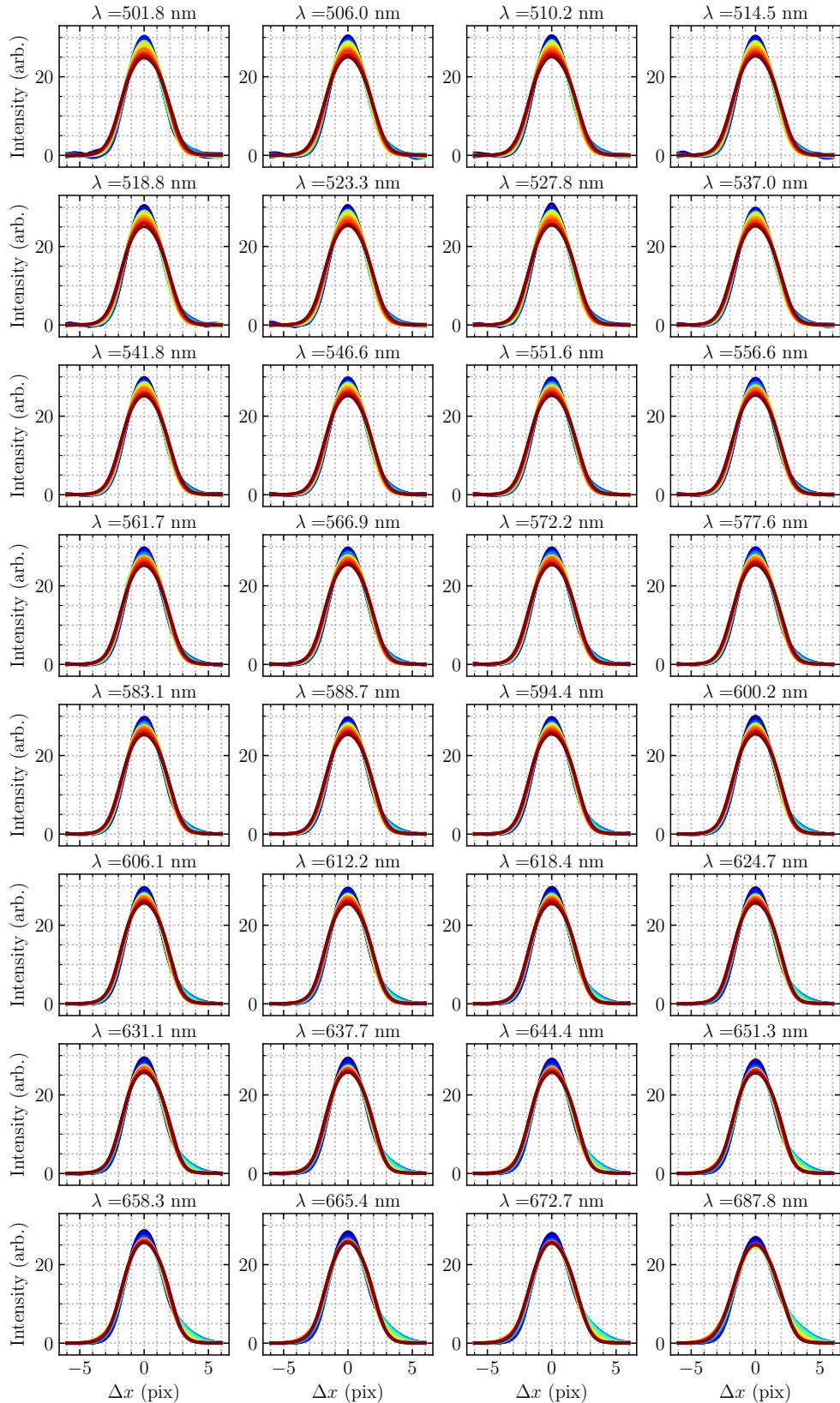


Fig. 5. HARPS IP models in pixel space for orders 89 through 122. Each panel shows the IP models for a single echelle order, whose central wavelength is printed above it. The 16 coloured lines inside each panel show how the IP changes with wavelength (line colour changes from blue to red with increasing wavelength).

edge were more symmetric, albeit not fully. Similar behaviour is seen in all echelle orders, once more consistent with the general picture on line shapes provided by Fig. 2.

5. Improving spectral fidelity

5.1. Fit quality of astrocomb lines and uncertainties on line position measurements

All astrocomb lines were independently fitted using Gaussian IPs and our empirical IPs, yielding two sets of line parameters. Fig. 7 shows the models for one randomly chosen astrocomb line fitted with three different IPs: (1) the empirical IP at the line location (i.e. ψ was linearly interpolated from the nearest two segment centres), left panel; (2) the empirical IP from the nearest segment centre (no ψ interpolation was done), middle panel; and (3) a Gaussian IP, right panel. The best fit quality, in terms of χ^2_v , is provided by the interpolated empirical IP, followed by the empirical IP without interpolation, and finally by the Gaussian IP. Note the large difference in χ^2_v between the two empirical IP models and the Gaussian IP model, demonstrating once more that the Gaussian IP is inadequate. Similar results are obtained for all other examined lines.

The χ^2_v for all 10576 detected lines fitted using the most likely IP models in pixel space (where the IP was interpolated to line position) are shown in Fig. 8 as a function of line position on the detector. Unlike during fitting models to the data, no weights were applied in calculating χ^2_v . We note several things. Firstly, the most commonly observed χ^2_v values are in the range $0.9 \leq \chi^2_v \leq 1.1$, indicating excellent overall agreement between our models and the data. Secondly, there is a cluster of lines with $\chi^2_v \gtrsim 20$ in the middle of the detector. The increased χ^2_v does not correlate with line brightness nor any other examined parameter, leaving its cause a mystery. However, we found a peculiarity while examining model residuals for several randomly selected lines from that cluster, which we report in [Appendix TBD](#). Thirdly, lines at the order's red wavelength edges have systematically larger χ^2_v values than lines in order centres.

The median χ^2_v over the entire ensemble of lines is 1.77, with 70% (95%) of the lines having $\chi^2_v < 2.62$ (7.42). For comparison, the median χ^2_v for the Gaussian fits is 375.32, and 70% (95%) of the lines have $\chi^2_v < 589.56$ (900.83). Overall, the fit quality is significantly higher when using our ψ models than when using a Gaussian IP approximation. The improvement is also appreciated by comparing our Fig. 7 and 8 to figures 4 and 5 in Milaković et al. (2020).

Fig. 9 shows the uncertainties on line centres in pixel space, σ_{x_*} , determined from the covariance matrix at the best-fit solution. As expected, larger uncertainties are associated to lines with higher χ^2_v values (e.g. in the central region of the detector) and in regions with low S/N in the data. A radial dependence can also be seen in which lines close to detector edges also have larger uncertainties compared to lines close to the detector centre (excluding the cluster of lines with high χ^2_v). The median uncertainty is 1.9 mpix and the central 68% distribution limits are 1.2 mpix and 3.2 mpix. 95% of all lines have $\sigma_{x_*} \leq 5.8$ mpix. Approximating σ_{x_*} in units m s^{-1} (1 pix = 820 m s^{-1}), the median of the σ_{x_*} distribution corresponds to 1.56 m s^{-1} and the 16th, 68th, and 95th percentiles correspond to 0.97 m s^{-1} , 2.65 m s^{-1} , and 4.78 m s^{-1} , respectively.

5.2. Differences between Gaussian and empirical IP centres and the impact on wavelength calibration

First we report on the comparison between x_* values measured by fitting empirical IPs to their Gaussian IP counterparts. Fig. 10 shows the differences between the two measurements, $\Delta x_* = x_{*,G} - x_{*,E}$, where the subscripts on x_* indicate which IP was used (E for our empirical IPs and G for the Gaussian IP). The Figure reveals striking differences between the two sets of centre estimates. Firstly, zero differences are seldom observed. Secondly, the differences vary smoothly both within and across orders. For most orders, the maximum difference occurs at their blue wavelength edges (left half of the Figure). Moving towards the red wavelength edge, Δx_* approaches zero approximately in the middle of the order before increasing again at its red edge. The maximum observed Δx_* , 0.18 pix (corresponding to a velocity shift of 147.6 m s^{-1}), appears at the red wavelength edges of the two reddest orders (top right corner). Negative values are observed only in the very low S/N part of the spectrum (bottom left corner). 95% of the lines have $|\Delta x_*| < 0.127$ pix (104 m s^{-1}) and 50% have $|\Delta x_*| < 0.079$ pix (65.0 m s^{-1}).

Next, the two sets of x_* measurements (Gaussian and empirical IP) were used to wavelength calibrate HARPS and the resulting wavelength calibrations were then compared to each other. In producing the calibration, a single echelle order was divided into eight 512 pixel wide segments and a seventh order polynomial was fitted in each segment individually (Milaković et al. 2020). No continuity conditions were imposed, meaning that the wavelength calibration is allowed to jump at each segment boundary. Both calibrations cover the wavelength range $499 \text{ nm} \leq \lambda \leq 680 \text{ nm}$.

Expressing the differences between the two wavelength calibrations as velocity shifts, we found the same general pattern already seen in Fig. 10 so it is not shown. The only difference with respect to that pattern is that this pattern is of the opposite sign, which can be understood by considering that the empirical IP astrocomb positions have lower numerical pixel values than their Gaussian counterparts (as inferred from Fig. 10). The calibration based on the empirical IP line positions thus assigns the same wavelength (determined from Eq. (1)) to a smaller pixel number. Equivalently, the same pixel is assigned a shorter wavelength. This results in a wavelength difference that has an opposite sign with respect to the difference between pixel positions. The median velocity shift between the two calibrations is -46.62 m s^{-1} (Gaussian minus empirical IP calibration). The maximum difference in absolute is approximately -150 m s^{-1} , appearing at the same locations where maximum positive Δx_* is observed in Fig. 10, and is consistent in amplitude with the simple conversion based on the velocity content of HARPS pixels.

Examining the histogram of the wavelength calibration residuals (not shown) for the two calibrations reveals that they both follow a normal distribution with a zero mean and that their distributions have similar widths. The root-mean-square (RMS) of the residuals when empirical IP line centres are used is 4.27 m s^{-1} whereas it is 4.28 m s^{-1} when Gaussian IP centres are used. Both calibrations are therefore “correct” in a statistical sense.

5.3. Differences between wavelength measurements made using Gaussian and empirical IP models

For the purpose of quantifying the impact of our empirical IP models on wavelength measurements, the astrocomb spectrum was next treated as if it were a science spectrum, i.e. it was cali-

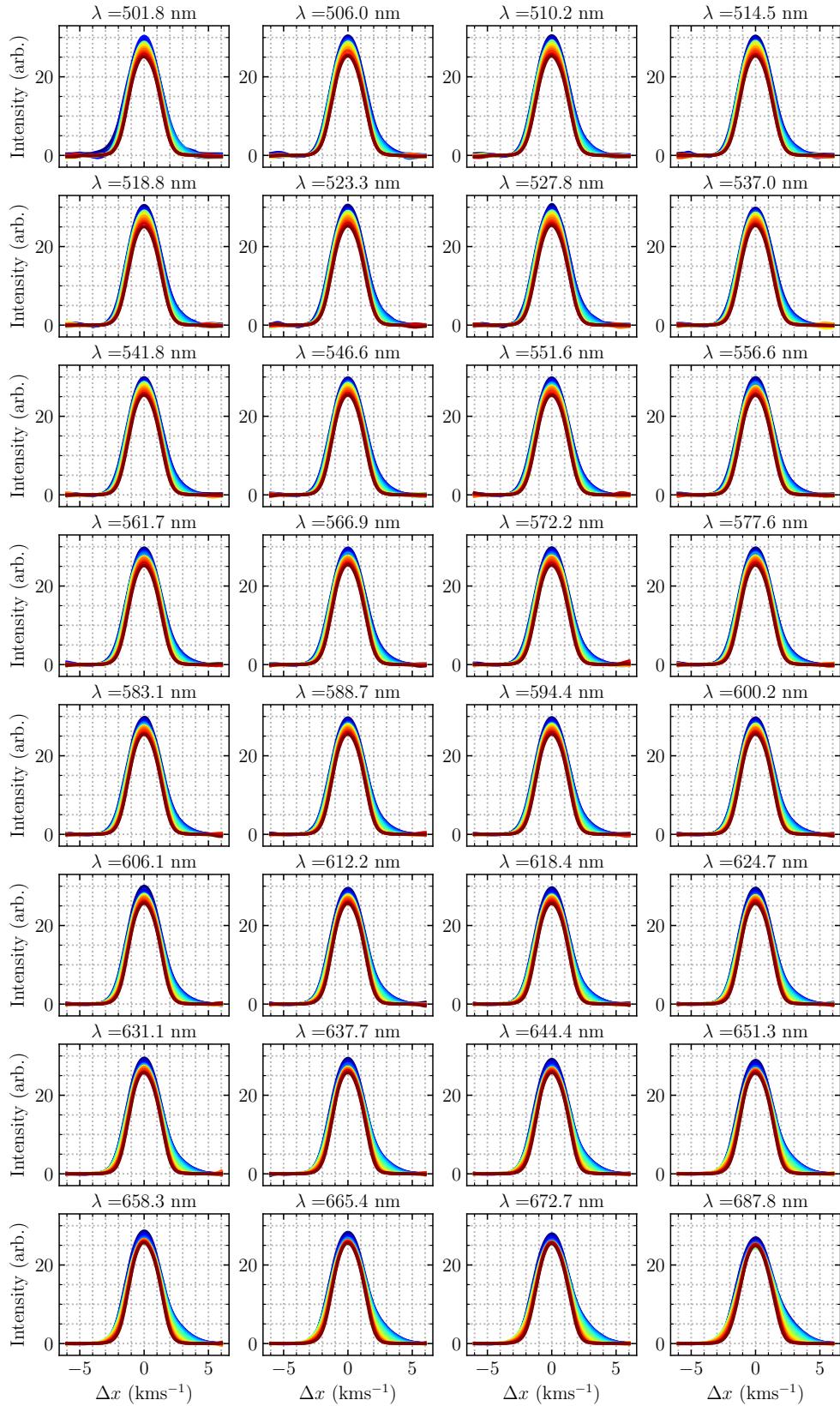


Fig. 6. HARPS IP models in velocity space for orders 89 through 122. Each panel shows the IP in 16 segments of one order, whose central wavelength is printed above it. Line colour within the panel changes from blue to red with increasing wavelength.

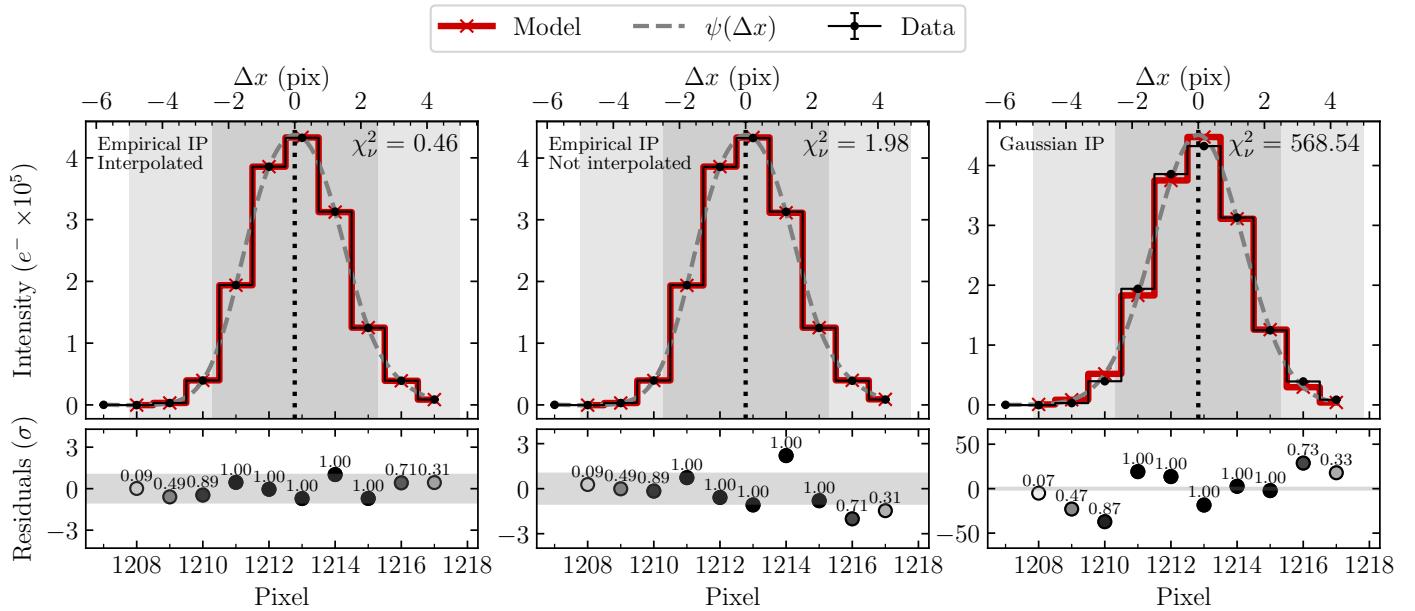


Fig. 7. The figure shows the results of fitting the same astrocomb line using three different IPs. They are, in order from left to right, (i) the most likely empirical IP, when it is allowed to vary locally within the echelle order (interpolated), (ii) when the empirical IP from the nearest segment is used instead (not interpolated), and (iii) the Gaussian IP. *Top panels:* In each top panel, the solid black histogram with error bars shows the astrocomb line flux and the solid red line with crosses shows the best fit model of that data. The line model is described by Eq. (13) when fitting the empirical IP and by Eq. (12) when fitting a Gaussian IP. The location of the best-fit line centre, x_* , is indicated by the vertical dotted black line in each panel, with the ticks and labels on the top indicating Δx . The two vertical grey shaded bands illustrate the weighting scheme, Eq (16). χ^2_{ν} for each fit is printed in the top right corner of the panel. Finally, the continuous dashed grey line shows the IP model, ψ , shifted and rescaled to match the data (for visual comparison purposes only). *Bottom panels:* Dots in each panel are the normalised fit residuals and the grey shaded band indicates the $\pm 1\sigma$ region. Dot opacity is proportional to the pixel weight in the fit, which is also printed above each dot. Dots falling within the darker grey band in the top panel all had a weight of unity, whereas the dots falling outside of both bands had zero weight (so are not plotted). Dots falling in between, i.e. inside the lighter grey band of the top panel, had weights that change linearly between zero and unity. Note that the y-axis scale differs between panels.

bilated using the wavelength calibration produced in Sec. 5.2 and line wavelengths, λ_* , were measured from this spectrum without any prior knowledge from Eq. (1).

Like before, the wavelength of each line was measured by fitting a specific IP model and two sets of λ_* measurements were produced: the first one using our empirical IP models and the second one using Gaussian IP models. To allow a more straightforward interpretation of the results, the same kind of IPs was used both to wavelength calibrate the data and to measure astrocomb line wavelengths. In both cases, the fitting followed the methods explained in Sec. 3.1.1, with λ_* taking the place of x_* . Gaussian IPs were fitted in wavelength space but empirical IP models were fitted in velocity space because of the way they were originally constructed. Zero velocity corresponded to the location of the line centroid in wavelength space, $\bar{\lambda}$. Following this, the data was fitted using the appropriate IP model and λ_* was extracted from the best-fit solution.

The uncertainties on λ_* associated to empirical IPs resemble those on x_* both in the distribution on the detector and in their distribution (so are not shown). The distribution of σ_{λ_*} has a median value of 1.55 m s^{-1} , with the 16th, 68th, and 95th percentiles being 0.95 m s^{-1} , 2.66 m s^{-1} , and 4.77 m s^{-1} (respectively). This is in excellent agreement with the uncertainties on x_* expressed in velocity units (c.f. with numbers in Sec. 5.1) and consistent with theoretical predictions based on the photon counting statistics, σ_{pn} (Bouchy et al. 2001). The median ratio of $\sigma_{\lambda_*}/\sigma_{pn}$ (calculated for each line individually) is 1.04 and the central 68% of the distribution falls in the range between 0.65 and 1.62.

Fig. 11 shows the differences between the two sets of λ_* measurements expressed in velocity units:

$$\frac{\Delta v_*}{c} = \frac{\lambda_{*,G} - \lambda_{*,E}}{\lambda_{LFC}}. \quad (27)$$

The subscripts G and E on λ_* again indicate which IP was fitted to the data and $\lambda_{*,LFC}$ is the expected wavelength determined through Eq. (1). **to be rewritten** This Figure shows that Δv_* varies smoothly across the detector in a way that is similar to Fig. 10. The largest positive departures from zero can be found in the blue wavelength edge of the reddest echelle order, reaching values as large 40 m s^{-1} , which is also the largest absolute departure. The mode of the histogram in the top panel of the same Figure is $\approx 4 \text{ m s}^{-1}$ and its median is 5.90 m s^{-1} . Values below zero occur for 9% of the lines. Interpreting the histogram as a probability distribution, the probability of finding $|\Delta v_*| \leq 5 \text{ m s}^{-1}$ is 42%. In absolute terms, the 50%, 70%, and 95% of the lines have $|\Delta v_*| \leq 5.98 \text{ m s}^{-1}$, 9.64 m s^{-1} , and 20.63 m s^{-1} , respectively.

5.4. Examination of astrocomb modes appearing twice on the detector

This intriguing result was explored further by examining the consistency between λ_* measurements for astrocomb modes appearing twice on the detector due to the overlap between two adjacent spectral orders. If the IP models used to fit those lines are correct and if the wavelength calibration is accurate, the measured λ_* for the two instances of the same mode should be identical. We identified 1472 astrocomb modes appearing twice in our

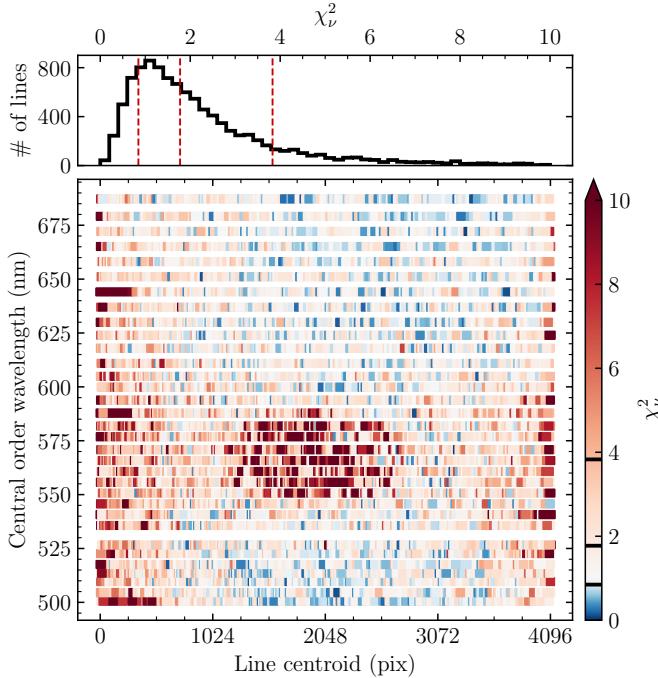


Fig. 8. *Main panel:* The reduced χ^2 for all detected astrocomb lines as a function of line position on the detector. The colour bar's zero point corresponds to $\chi^2_\nu = 1$, such that blue and red colours show lines whose χ^2_ν is smaller and larger than 1 (respectively). Most notable is a cluster of lines with $\chi^2_\nu \gtrsim 20$ in the middle of the panel. The increased χ^2 in this region does not correlate with line amplitude or any other examined parameter. *Top panel:* Histogram of χ^2_ν values plotted in the main panel. The vertical dashed red lines show the locations of the median and the central 68% distribution limits. The same quantities are shown as thick black horizontal lines in the colour bar.

astrocomb spectrum, for which we calculated the relative wavelength difference between the pair:

$$\Delta v_{\text{pair}} = \frac{\lambda_{*,2} - \lambda_{*,1}}{c}. \quad (28)$$

In the formula above, $\lambda_{*,1}$ and $\lambda_{*,2}$ are the two measured wavelengths for the same mode, and c is the speed of light.

The distribution of Δv_{pair} associated to using our empirical IP models is plotted as the red histogram in Fig. 12. Reassuringly, it peaks at zero difference. The median of the distribution is -0.31 m s^{-1} , with central 68% of the pairs falling in the range $-12.45 \text{ m s}^{-1} \leq \Delta v_{\text{pair}} \leq 13.54 \text{ m s}^{-1}$. Examining absolute values, 50% of the pairs have $|\Delta v_{\text{pair}}| < 7.96 \text{ m s}^{-1}$, 70% have $|\Delta v_{\text{pair}}| < 13.51 \text{ m s}^{-1}$ and 95% have $|\Delta v_{\text{pair}}| < 34.58 \text{ m s}^{-1}$.

In contrast, the distribution of Δv_{pair} values associated to using Gaussian IPs peaks at -12.5 m s^{-1} (black histogram in the same Figure), with the median value of -13.59 m s^{-1} . The central 68% of the distribution is contained within the range $-19.14 \text{ m s}^{-1} \leq \Delta v_{\text{pair}} \leq -7.95 \text{ m s}^{-1}$. In this case, 50% of the pairs have $|\Delta v_{\text{pair}}| < 13.66 \text{ m s}^{-1}$, and 70% (95%) of the pairs have $|\Delta v_{\text{pair}}| < 16.33 \text{ m s}^{-1}$ (25.41 m s^{-1}).

5.5. Examination of wavelength scale distortions

Another important question is whether choosing a particular IP produces correlated distortions in the wavelength scale of the instrument. Avoiding such distortions is important whenever relative wavelength shifts between different spectral features are

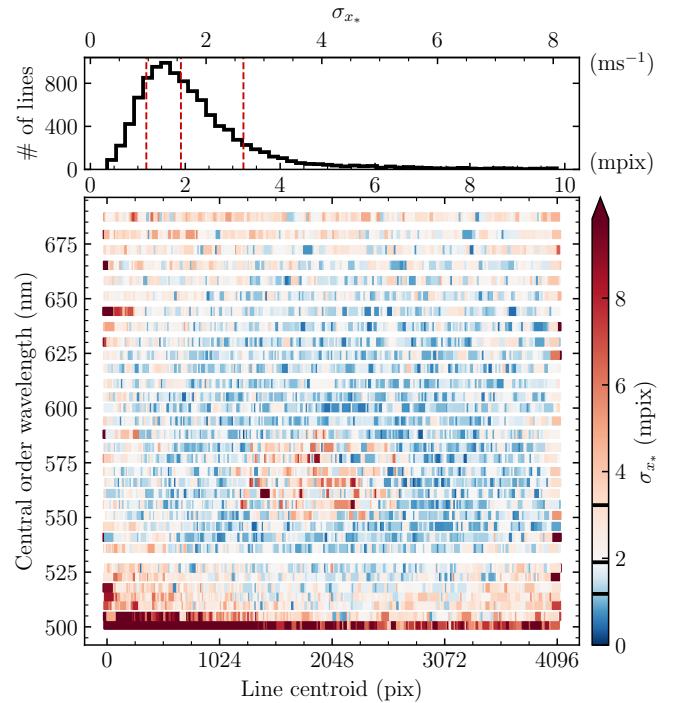


Fig. 9. *Main panel:* The colour encodes the uncertainty on astrocomb line centres, σ_{x_*} . The colour bar zero point is set to the sample median, such that the red and the blue colours correspond to values larger and smaller than the median, respectively. Lines with larger uncertainties generally have larger χ^2_ν (c.f. Fig. 8) or low S/N (at 500 nm). *Top panel:* Histogram of the values plotted in the main panel. The vertical dashed red lines show the median and the central 68% distribution limits, which are also shown as thick black horizontal lines in the colour bar. The x -axis on the top shows σ_{x_*} in units m s^{-1} (1 pix = 820 m s^{-1}).

measured, as is always done in fundamental constant studies and sometimes in studies of isotopic abundances of chemical elements.

The existence of wavelength scale distortions in the UVES (Dekker et al. 2000) and HIRES (Vogt et al. 1994) spectrographs is well documented and their impact on fundamental constant measurements is well understood (Rahmani et al. 2013; Whitmore & Murphy 2015; Dumont & Webb 2017; Milaković et al. 2020). In fact, removing wavelength scale distortions is one of the major expected benefits from using astrocombs for instrument wavelength calibration and a major reason for their deployment on numerous current and future spectrographs. Yet, some level of distortions were found to remain even when astrocombs are used (Milaković et al. 2020; Schmidt et al. 2021).

First, we examined how Δv_{pair} distributes across the detector. 13 shows this distribution for the empirical IP calibration and 14 shows it for the Gaussian IP calibration. In order to more easily detect any correlations with line position on the detector, the colour scheme in the two Figures has been modified such that the white colour corresponds to the distribution median (which is different in each Figure).

In both Figures, the values and the patterns seen on the left hand side (i.e. appearing in the blue wavelength edge of an echelle order) are copied on the right hand side, one order down (i.e. in the red wavelength edge of the following echelle order). This is a plotting choice: we only had one Δv_{pair} value for two lines yet we wanted to see where all of them fall on the detector. The apparent spreading out of the pattern is a consequence of instrument optics (lines appearing on the blue wavelength edge

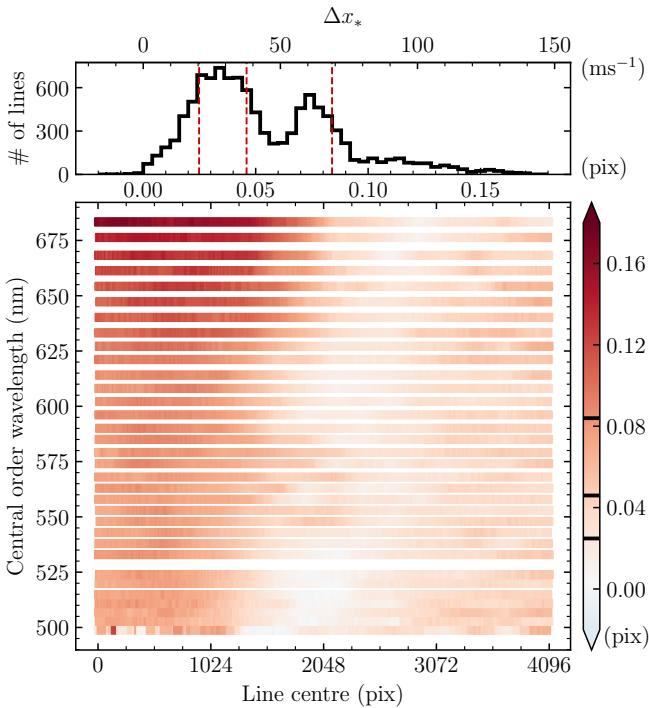


Fig. 10. *Main panel:* The difference between astrocomb line position measurements, x_* , derived from fitting two different IP models to the data. The colour bar shows the quantity $\Delta x_* = x_{*,G} - x_{*,E}$, where the subscripts indicate which IP model was used to determine the centre during line fitting (G for the Gaussian and E for our empirical IP models). The zero point of the colour bar is set to $\Delta x_* = 0$ pix, such that the red and blue colours correspond to positive and negative shifts. *Top panel:* The histogram of the values plotted in the main panel. The bottom x-axis is in units pixel and the top axis in units m s^{-1} ($1 \text{ pix} = 820 \text{ m s}^{-1}$). The vertical dashed red lines show the median (0.064 pix or 51.7 m s^{-1}) and the central 68% region limits (0.034 pix and 0.090 pix , corresponding to 28.1 m s^{-1} and 74.0 m s^{-1}). The same quantities are shown as horizontal thick black lines in the colour bar to the right of the main panel.

of the order are more densely packed than the lines appearing at its right edge) and of data plotting (line position on the detector is only approximate). Neither pattern copying nor its changing width have an impact on the following discussion or on the conclusions.

Looking at the empirical IP case first (Fig. 13), Δv_{pair} values are seemingly almost randomly distributed with no obvious correlations with line position. There is possibly one exception to this, seen as a blue pattern in order with a central wavelength of 580 nm, which may indicate the presence of a short-range wavelength distortion with an amplitude of -20 m s^{-1} (corresponding to $\lesssim 3\%$ of the HARPS pixel) within 0.2 nm range (the width of the pattern). We have not explored this further. The largest absolute departures of Δv_{pair} from zero are seen in the lowest S/N data (at 500 nm), which is also not very large in amplitude, $\approx \pm 20 \text{ m s}^{-1}$.

The values used to produce the histogram in the top panel of the Figure are the same ones previously used to construct the red histogram in Fig. 12, but the former focuses on a narrower range and has more bins. Overall, there is no evidence that using the empirical IP models to wavelength calibrate HARPS introduces wavelength scale distortions larger than several m s^{-1} over the wavelength range 500 nm to 680 nm.

Changing focus onto Fig. 14, which shows the results for the Gaussian IP case, evidence for both short-range and long-range

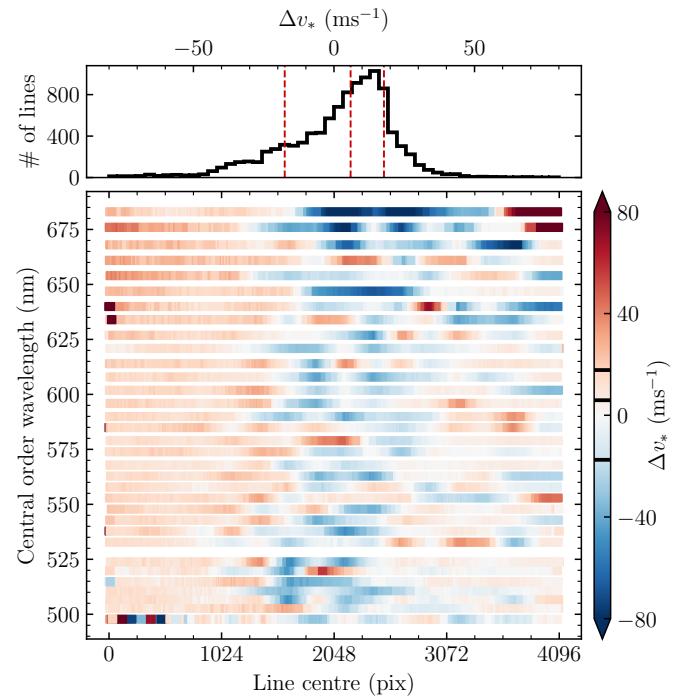


Fig. 11. *Main panel:* The plot shows the relative difference between the measured astrocomb lines wavelengths in velocity units, Δv_* , for all 10576 lines. The zero point of the colour bar is set to $\Delta v_* = 0 \text{ m s}^{-1}$, such that the red and blue colours correspond to positive and negative shifts. *Top panel:* The histogram of the values plotted in the main panel. The vertical dashed red lines show the median 5.90 m s^{-1} and the central 68% region limits (1.39 m s^{-1} and 13.82 m s^{-1}). The same quantities are shown as horizontal thick black lines in the colour bar to the right of the main panel.

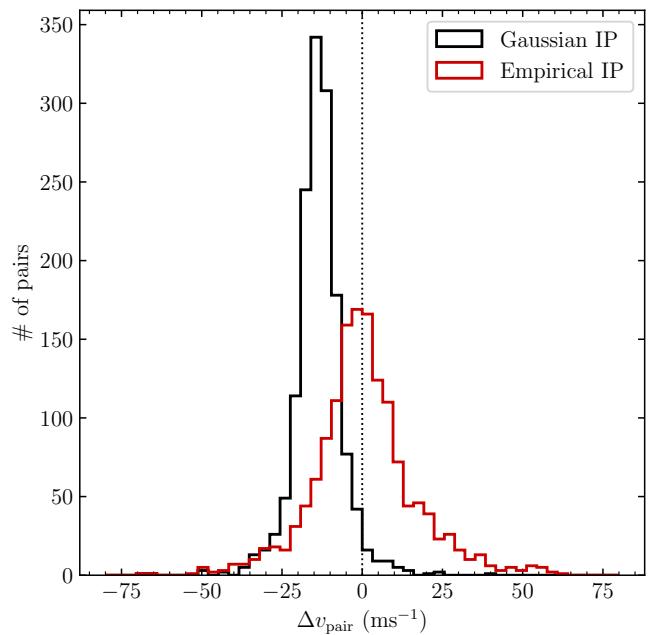


Fig. 12. Histogram of measured velocity shifts between 1472 astrocomb modes appearing twice in the spectrum. Line wavelengths were measured either by fitting the most likely empirical IP model (red histogram) or a Gaussian IP model (black histogram). Zero shift means that the same wavelength was measured for both lines of the pair.

distortions is immediately noticeable. Firstly, the amplitudes and the range of Δv_{pair} are significantly larger than was the case when empirical IP was used, as first noted from the comparison of the two histograms in Fig. 12. Secondly, Δv_{pair} varies smoothly within the order, changing abruptly every 512 pix. These discontinuities are a consequence of using segmented polynomials for wavelength calibration and appear at the edges between two independent polynomials. Segmented polynomials were used, not only to avoid CCD manufacturing errors, but also because using a single polynomial to wavelength calibrate the full length of the echelle order created correlated calibration residuals even when CCD errors were compensated for and high order polynomials were used (Milaković et al. 2020).

It is therefore interesting to note the complete absence of the discontinuities when empirical IP were used (Fig. 13). This gives another indication that the accuracy derived from using the empirical IPs is significantly superior to what was so far been possible to achieve using Gaussian IPs.

We make an additional note about the discontinuities every 512 pix. Because the same pattern is repeated for visual purposes on the right hand side of the plots, this discontinuity is artificially transferred (copied) also there, where it appears at a location which is not a multiple of 512 pix. Collectively, the copied values create a slanted pattern seen on the bottom right corner of the plot. This is a purely artificial effect introduced by the way the data was plotted and should not be interpreted as being real.

However, discontinuities at 3072 pix and 3584 pix seen in the bottom right corner of the plot, in echelle orders with the central wavelengths in the 500 nm to 530 nm range, are real because they appear at multiples of 512 pix. Analogously to before, because the values are copied and plotted twice, an artificial slanted pattern can also be seen in the bottom left corner of the plot.

We can now answer the original question, i.e. does using a particular IP introduce distortions into the wavelength scale of the instrument. The answer is a clear yes.

Overall, Fig. 14 tells us that, should one infer the distance between two spectral features, one appearing

6. Impact on fine structure constant measurements

VPFIT (Carswell & Webb 2014) is the most commonly used tool for making fine structure constant (α) measurements from astronomical absorption spectra. Provided with a set of model parameters that describe the system being examined, including a relative change in α from its terrestrial laboratory value ($\Delta\alpha/\alpha$), VPFIT searches for parameter values that minimise $\chi^2 = (\text{data} - \text{model})/\text{error}$. This process involves a convolution of the absorption system model with an user-provided IP which is compared against the data, i.e. used to calculate χ^2 . In the case when the exact IP is unknown, VPFIT approximates it by a Gaussian whose width corresponds to the FWHM of the resolution element, also provided by the user. All previous $\Delta\alpha/\alpha$ measurements in quasar absorption systems used the Gaussian IP approximation when performing parameter optimisation. We thus set out to quantify the impact of replacing the Gaussian IP approximation with known empirical models by using them to re-measure $\Delta\alpha/\alpha$ from existing HARPS observations of the quasar HE0515–4414.

6.1. Quasar data and absorption system models

Observations, data reduction, and data combination procedures are described in Milaković et al. (2021). Quasar exposures were

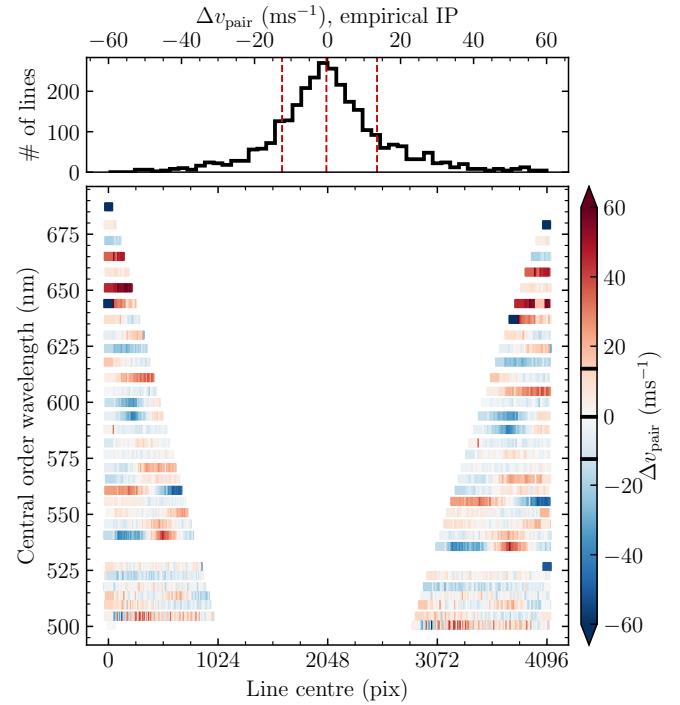


Fig. 13. *Main panel:* The distribution of measured velocity shifts between 2880 astrocomb lines (1440 pairs) appearing twice in the spectrum when the empirical IP is used for line centre measurements and for the wavelength calibration. No spatial correlations are observed. The unfilled detector area contains no line pairs. The zero point of the colour bar is set to the median observed shift (-4.0 m s^{-1}) such that the red and blue colours represent positive and negative shifts with respect to that value. *Top panel:* The histogram of values shown in the main panel (the same as the red histogram in Fig. 12). The median and the central 68% distribution limits are shown as vertical dashed red lines. The same quantities are shown as thick black lines in the colour bar to the right.

bracketed by astrocomb frames for wavelength calibration and monitoring spectroscopic drifts. Figure 2 in Milaković et al. (2021) shows that astrocomb lines shifted by $\approx 0.5 \text{ m s}^{-1}$ over the 8 nights over which the observations took place, demonstrating that instrumental drifts were negligible. The wavelength calibration for the data is based on Gaussian IP centres for astrocomb lines.

The absorption system at $z = 1.15$ is $\approx 600 \text{ km s}^{-1}$ wide and has a complicated structure with a large number of components. In Milaković et al. (2021), the system was divided into five regions (labeled with roman numerals I–V) to simplify calculations. Transitions appearing in all five regions are Mg I $\lambda 2852$, Mg II $\lambda 2796, 2803$, and Fe II $\lambda 2344, 2374, 2383, 2586, 2600$. Region V additionally contains Mn II $\lambda 2576, 2594, 2606$.

Models for all five regions were produced using AI-VPFIT (Lee et al. 2021a) and can be found in Milaković et al. (2021). In making the models, both thermal and turbulent broadening mechanisms were allowed to contribute to the final Doppler b -parameters and $\Delta\alpha/\alpha$ was a free fit parameter as soon as more than one atomic transition were included in the fit. Further details and plots of the system can be found in Milaković et al. (2021).

6.2. Matching the IP in the data

To obtain the correct results, the IP provided to VPFIT should match the IP in the combined quasar spectrum. The best approach to do this would be to calculate a weighted sum of the IPs

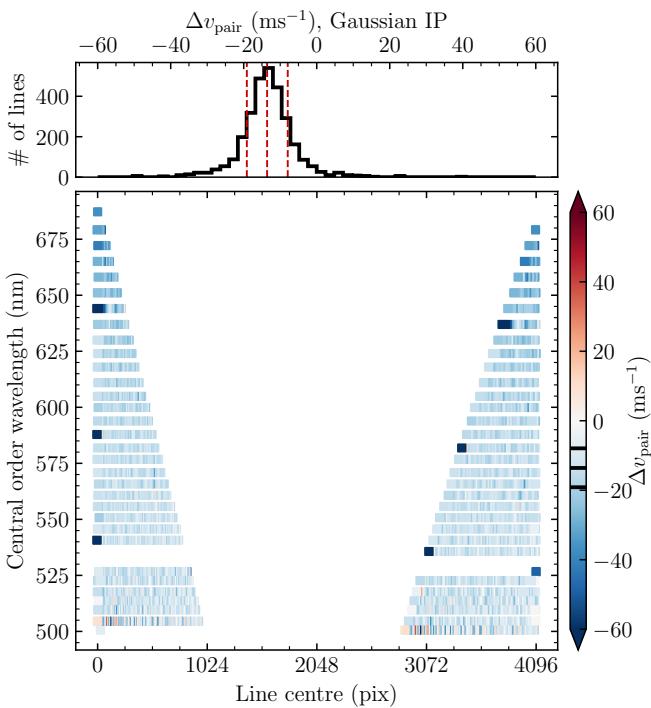


Fig. 14. *Main panel:* The distribution of measured velocity shifts between 2880 astrocomb lines (1440 pairs) appearing twice in the spectrum when the Gaussian IP is used for line centre measurements and for the wavelength calibration. Strong spatial correlations are observed, as are sudden jumps every 512 pixels which occur at the locations of HARPS detector stitching. The unfilled detector area contains no line pairs. *Top panel:* The histogram of values shown in the main panel (the same as the black histogram in Fig. 12). The median and the central 68% distribution limits are shown as vertical dashed red lines. The same quantities are shown as thick black lines in the colour bar to the right.

derived from astrocomb frames closest in time to the astronomical observations, with the weights on the IPs exactly the same as those applied to the data during summation. Additionally, the astrocomb frames should be taken such that the observed flux counts in the astrocomb lines matches that of quasar observations.

The flux in the astrocomb spectra associated to the quasar observations are several thousand times higher than the flux in the quasar observations. While some changes in the IP shape are expected with changing flux levels due to the effects of charge transfer inefficiency, we had no suitable data to assess what the suitable IP for our data would look like. We therefore proceeded with the use of IPs reconstructed from high flux astrocomb observations, leaving the investigation of the impact of changing flux levels on the IP shape for future work.

We also did not attempt to reconstruct the IP for all astrocomb spectra associated to these observations, for the following reasons. Firstly, the necessary calculations were computationally prohibitive and secondly, given that instrument stability was $\approx 0.5 \text{ m s}^{-1}$ over the duration of observations, any changes in the IP with time are likely to be negligible. Instead, we reconstructed the IP using the astrocomb spectrum used for wavelength calibration of the quasar data, i.e. the spectrum recorded on 2018-12-05 at 08:12:52.040 UT in fibre A.

Pixels provided with the data are the only ones VPFIT knows about and the user-provided IP must match them exactly. In our quasar data, individual exposures that went into the final co-added quasar spectrum were redispersed onto a common

wavelength grid with pixel steps linear in logarithmic space, corresponding to a constant step in velocity space of $\Delta v = 0.83 \text{ km s}^{-1}$.

6.3. Producing IPs covering a specific wavelength range

The IP in each spectral region of interest, i.e. a spectral chunk bracketed by wavelengths λ_{start} and λ_{end} , was calculated as a weighted sum of IPs in detector regions falling in that wavelength range. The weights were determined by considering the number of pixels contributing to the final spectrum and the spectral variance array of each exposure that went into creating the final quasar spectrum.

First, the wavelength calibration for a single exposure was considered. The calibration was shifted to the Solar-system barycentre restframe and pixels in that exposure with wavelengths $\lambda_{\text{start}} \leq \lambda \leq \lambda_{\text{end}}$ were identified. Next, we determined how these pixels are distributed with respect to our segmentation of echelle orders and we identified all segments with contributing pixels. In each of these segments, we summed the inverse flux variances of those pixels, $W = \sum_i (1/\sigma_i^2)$, where i is pixel number. This way, a segment with many contributing pixels and/or smaller variances weights more than a segment which contributes either fewer pixels or the same number of pixels with larger uncertainties on the flux.

We repeated this procedure for every quasar exposure and obtained a list of (echelle order, segment, W) values, where the sum of inverse variances now includes contributions from each exposure and W s were normalised such that they sum up to unity. Finally, the IP for a single spectral chunk was:

$$\psi(\lambda_{\text{start}} \leq \lambda \leq \lambda_{\text{end}}) = \sum_{o,i} W(o,i) \psi_{o,i}, \quad (29)$$

where the indices $\{o, i\}$ identify the segment i of the echelle order o which contributes pixels to this chunk.

6.4. VPFIT results

The IPs derived in the previous subsection were saved into files readable by VPFIT. The sampling used was 25 locations per 0.83 km s^{-1} wide data pixel and the total range covered was $\pm 5 \text{ km s}^{-1}$. This sampling and range corresponds to 350 points separated by 33.2 m s^{-1} . For comparison, in the models for the $z = 1.15$ system published in Milaković et al. (2021), the Doppler b -parameter of the most narrow heavy element absorption component is 0.9003 km s^{-1} and the smallest uncertainty on the redshift of a single heavy element absorption line is 3×10^{-7} (41.8 m s^{-1}).

The saved IP models were provided to VPFIT in the way described in the documentation, together with the AI-VPFIT models and the quasar and atomic data. VPFIT version used was 11.1, the same as in Milaković et al. (2021). The atomic data and the relevant VPFIT settings including stopping criteria were kept unchanged, with one exception: the two subpixelisation parameters contained in the “vpfit_setup.dat” file were set to match the subpixelisation of the IP, “nsubmin”=“nsubmax”=25.

The best fit value for $\Delta\alpha/\alpha$ and the corresponding statistical uncertainty (σ_{stat}) from the VPFIT calculations are given in Tab. 1 for all five regions, together with the final model χ^2_v . Values for the same quantities from Milaković et al. (2021) are also given for comparison with our results. In Milaković et al. (2021), VPFIT used a Gaussian IP for the numerical model convolution. Looking at the numbers in the first column of the two tables, we

see that best fit $\Delta\alpha/\alpha$ value has shifted by $\approx -2 \times 10^{-6}$ in regions I, II, and III; by $\approx -4 \times 10^{-6}$ in region IV, and remained basically unchanged in region V. Looking at the next column, we see that the statistical uncertainty has increased slightly in regions I and V when the empirical IP was used but remained unchanged in regions II, III, and IV. Finally, the χ^2_v in all regions has also not changed much.

Looking at other parameter values, we found that the smallest uncertainty on the redshift of any heavy element absorption line in the model remained unchanged and the smallest Doppler b -parameter increased to 0.9341 km s^{-1} (+3%). Taking the ratio between the b -parameters in the new and the old VPFIT models on a component by component basis, we find that b has changed by <1% in 62% of the components and by <0.1% in 9% of the components. The median ratio is 1.0.

6.5. Interpreting VPFIT results

There are two major limitations to the models used above. The first one is that they were derived from data that was calibrated using the Gaussian approximation for the HARPS IP. Wavelength calibration in the data is therefore systematically biased with respect to a calibration that would have been produced by using empirical IPs, as shown in Sec. 5. The second limitation is that a Gaussian IP was used for the numerical model convolution inside VPFIT. A simple exchange of the IP used for the convolution, without also changing the wavelength calibration in the data and without also producing new models using that exact IP, is therefore unlikely to produce trustworthy $\Delta\alpha/\alpha$ measurements.

With this in mind, the redefinition of the IP centre (i.e. changing from a symmetric Gaussian IP to an asymmetric empirical IP shape) is the simplest explanation for the systematic shift in $\Delta\alpha/\alpha$ values in Tab. 1. To demonstrate this, we compared the redshift values for individual absorption components in the new (using the empirical IP) and the old (using the Gaussian IP) models. Expressing redshift differences as velocity shifts, $\Delta v_z/c = z_{\text{old}} - z_{\text{new}}/(1+z_{\text{old}})$, and calculating the weighted mean of those shifts using their inverse errors as weights, we found that the new redshifts are systematically larger by 19.2 m s^{-1} and that the central 68% of the distribution of Δv_z fall within $[-13.9 \text{ m s}^{-1}, 41.8 \text{ m s}^{-1}]$. Shifts between the new and the old component redshift are thus similar to the shifts seen between astrocomb line centres measured using our empirical IP models and those measured using the Gaussian IP.

Following this, we calculated the theoretically expected shifts in the line centres due to $\Delta\alpha/\alpha \neq 0$:

$$\frac{\Delta v_\alpha}{c} = \frac{q}{2\omega} \frac{\Delta\alpha}{\alpha}. \quad (30)$$

In the formula above, Δv_α is the expected shift of a specific atomic transition for a given $\Delta\alpha/\alpha$ value, q is its theoretical sensitivity to $\Delta\alpha/\alpha$ change (Dzuba et al. 1999; Webb et al. 1999), and ω is its wavenumber.

Taking $\Delta\alpha/\alpha = -1.74 \times 10^{-6}$ (the average difference between $\Delta\alpha/\alpha$ values in Tab. 1 over all five regions), we found that the modelled transitions should have Δv_α in the range $[-50.8 \text{ m s}^{-1}, -6.8 \text{ m s}^{-1}]$. Due to their high S/N in the data and due to the large difference between their q factors, two transitions ($\text{Mg II } \lambda 2796$ and $\text{Fe II } \lambda 2382$) provide most of the constraint on $\Delta\alpha/\alpha$ in this system. The predicted shifts for these two transitions are: $\Delta v_\alpha = -6.2 \text{ m s}^{-1}$ ($\text{Mg II } \lambda 2796$) and $\Delta v_\alpha = -37.4 \text{ m s}^{-1}$ ($\text{Fe II } \lambda 2382$). The difference between these two numbers (-31.2 m s^{-1}) is close to the negative average change

Table 1. The two tables give the best fit $\Delta\alpha/\alpha$ values returned by VPFIT calculations performed on HARPS observations of HE0515–4414 and using AI–VPFIT models from Milaković et al. (2021). The table at the top shows the results when VPFIT was provided with empirical IP models produced in this work to perform the numerical model convolution. The bottom table shows the results when the IP was approximated with a Gaussian with FWHM=1.11 km s⁻¹. The first column in both tables identifies the region being fitted using roman numerals I–V. The second column gives the best fit $\Delta\alpha/\alpha$ value in units 1×10^{-6} . The third column, σ_{stat} , gives the statistical uncertainty on the values in the second column, derived from the covariance matrix evaluated at best fit solution. The units on σ_{stat} are 1×10^{-6} . The fourth column gives χ^2_v for the model. The final row in both tables gives the weighted average across all five regions and the associated uncertainty on that quantity.

Empirical IP models

Region	$\Delta\alpha/\alpha$	σ_{stat}	χ^2_v
I	-6.40	4.51	0.9910
II	14.95	4.31	0.9858
III	16.92	15.51	0.9847
IV	-7.92	4.16	0.8616
V	-2.93	3.64	0.9822
All	-0.49	2.03	

Gaussian IP approximation^a

Region	$\Delta\alpha/\alpha$	σ_{stat}	χ^2_v
I	-3.90	4.42	0.9892
II	17.47	4.30	0.9859
III	18.45	15.67	0.9836
IV	-6.39	4.15	0.8595
V	-2.59	3.12	0.9860
All	-0.94	1.97	

^a The numbers in this table are from table 2 in Milaković et al. (2020).

in component redshifts (19.2 m s^{-1}), suggesting that the change in $\Delta\alpha/\alpha$ simply compensates for the higher redshifts such that the observed wavelengths at which the components appear in the spectrum remain unchanged.

7. Summary of results

We conducted a detailed study of the HARPS instrumental profile of HARPS using its astronomical Laser Frequency Comb calibration system. In Sec. ??, we introduced a new method for the reconstruction of the 1-d instrumental profile for spectroscopic instruments equipped with astrocombs. The method, based on Gaussian Process regression and iterative procedures, allows for the determination the most likely non-parametric function what describes the empirical IP. Applying these methods to HARPS data, we reconstructed its IP from the astrocomb spectrum taken on UTC 2018-12-04T08:12:52, producing IP models for HARPS. Models were produced both in pixel space (for the purpose of wavelength calibration) and in velocity space (for the purpose of fitting spectral lines in a wavelength calibrated spectrum).

The main results of this work are To be rewritten:

1. In the entire range covered by the astrocomb, the HARPS IP significantly differs from a Gaussian shape and the IP asym-

- metry varies significantly across the detector. The IP is more asymmetric in velocity space than in pixel space.
2. The IP shape strongly correlates with wavelength both within and across orders. In every examined order, the IP at the blue wavelength edge exhibits a rightwards tail that gradually diminishes with increasing wavelength. The maximal tail amplitude also changes with wavelength, with bluest wavelength edge of the reddest orders having the strongest tails.
 3. The most robust centre estimator for an asymmetric IP proved to be the estimator introduced by Anderson & King (2000). The second most robust estimator was the median, followed by the mean and the mode. The differences between these four centre estimators for our models are ≈ 0.1 pix.
 4. Fitting astrocomb lines using our empirical IPs models produced high quality fits using only three free parameters, including line centre. The distribution of χ^2_v for the fits peaks at approximately unity and 70% of the lines have $\chi^2_v < 2.38$. Uncertainties on the fitted astrocomb line centres are consistent with the theoretical predictions from photon counting statistics.
 5. We found significant differences between astrocomb centres measured using our empirical IP models and centres measured using the Gaussian IP approximation. Additionally, we found that those differences vary across the detector in a correlated way. Within a single order, the maximum difference between the two centre measurements can reach 0.22 pix. The average difference in the wavelength range covered by the astrocomb is 0.06 pix.
 6. An identical correlated pattern emerges when comparing wavelength calibrations derived from those two sets of measured line centres (empirical and Gaussian IP). The average shift between the two calibrations (Gaussian minus empirical IP calibration) in the wavelength range covered by the astrocomb is -51 m s^{-1} , and the most commonly observed value is in the range $[-25 \text{ m s}^{-1}, -20 \text{ m s}^{-1}]$. Maximal difference, -180 m s^{-1} , was observed in the blue edge of the reddest echelle order $\lambda \approx 684.8 \text{ nm}$.
 7. Assuming that the empirical IPs derived here are correct and that line centres derived by fitting those IPs to the astrocomb lines are trustworthy, we conclude that using the Gaussian IP for wavelength calibration introduces distortions with amplitudes as large as 180 m s^{-1} between 500 nm and 680 nm. The distortions within single orders are similar in magnitude ($\approx 100 \text{ m s}^{-1}$), with the reddest orders showing having the most extreme distortions. Left uncorrected, such distortions can spoil radial velocity and fundamental constants measurements.
 8. We set an upper limit on the accuracy of spectroscopic measurements by fitting for the empirical wavelengths of 1440 astrocomb modes (2880 lines) appearing twice in the spectrum. Calculating the distribution of wavelength differences (shifts) for all line pairs, we estimate that the accuracy of our empirical IPs is 5 m s^{-1} at most, whereas it is at least ten times worse for the Gaussian IP. Worryingly, when a Gaussian IP is used, wavelength shifts correlate with detector position. No such correlations are seen for our empirical IPs.
 9. Applying our empirical IPs to HARPS observations of the quasar HE0515–4414 and using existing AI-VPFIT models to measure $\Delta\alpha/\alpha$, we find that its value changes by $\approx -2 \times 10^{-6}$ with respect to the values reported in Milaković et al. (2021). This is a consequence of the redefinition of a line centre.

8. Discussion

An additional benefit of having IP models is to use them to remove instrumental signatures from the data analysis, providing also the ability to directly compare and even combine spectra obtained using different spectrographs. Currently, the spectra of the same objects obtained using different instruments are non-reproducible, complicating result interpretation. The most obvious example relates to the studies of quasar absorption systems in the context of fundamental coupling constants' variability (Webb et al. 1999; King et al. 2012). For example: UVES, HARPS, and ESPRESSO observations produced different physical properties of the absorption system at $z = 1.15$ towards the quasar HE0515–4414 (Kotuš et al. 2017; Milaković et al. 2021; Murphy et al. 2022). While some differences can be attributed to the astrophysical objects, improving the spectral fidelity is crucial for performing tests of fundamental physics using astronomical spectra. Characterising the spectrograph's IP is the first step towards in this direction.

Knowledge of the IP

As mentioned in the Introduction, the ANDES instrument aims to maintain wavelength calibration stability. Ensuring wavelength calibration stability at the cm s^{-1} level over periods of several years or decades will probably require significant advancements in accurate absolute wavelength measurements. Here, we define accuracy as bias with respect to the true value, precision as the variance of N repeated measurements of the same quantity, and stability as the evolution of the bias with time.

We stress once more that, strictly speaking, the values in Sec. ?? do not quantify the accuracy of wavelength measurements because we cannot assess the bias in the astrocomb, if there is any. They do, however, give

determine the accuracy of our spectral wavelength measurements because the wavelength calibration was performed using the same astrocomb under examination and may therefore be biased. Instead, we simply showed that the bias in such measurements, whatever it may be, is smaller when empirical IPs are used.

8.0.1. Advantages of effective IP

We note further advantages and some limitations of this approach. For example, the definition of ePSF makes no assumptions regarding the physical sizes of pixels. This way, information on the impact of small variations in the pixel area is contained in \mathcal{R} . The drawback of this is that interpolating between IPs constructed in different areas of the detector is more difficult as we have no information on \mathcal{R} there.

8.0.2. A comparison with previous work

Zhao et al do not give models.

8.0.3. A comparison with an IP derived from an iodine absorption cell

A comparison with the IP derived from an iodine (I_2) absorption cell would be interesting. I_2 cells have previously been used to determine the IP shape of optical echelle spectrographs directly from the science observations. The absorption of light by the molecular iodine in the cell imprints the shapes of hundreds of sharp, intrinsically narrow molecular lines in the optical range (5000–6300 Å). HARPS was equipped with an I_2 absorption cell but, unfortunately, it has been decommissioned in 2004 (Rodler

& Lo Curto 2019) so a direct comparison of the IP derived from astrocomb calibrations and I₂ cell calibrations is not possible on HARPS.

A project to compare the IP derived from astrocomb (in emission, following only partly the full optical path of an astronomical target) with the IP derived from I₂ absorption cell is planned for ESPRESSO as a part of preparations for the ELT.

Acknowledgements

D. Milaković thanks Luca Pasquini and John Webb on discussions which improved this work. P. Jethwa acknowledges funding by the Austrian Science Fund (FWF): F6811-N36, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 724857 (Consolidator Grant ArcheoDyn).

References

- Aigrain, S. & Foreman-Mackey, D. 2022, arXiv e-prints, arXiv:2209.08940
- Anderson, J. & King, I. R. 2000, PASP, 112, 1360
- Baudrand, J. & Walker, G. A. H. 2001, PASP, 113, 851
- Bauer, F. F., Zechmeister, M., & Reiners, A. 2015, Astronomy and Astrophysics, 581, A117
- Bingham, E., Chen, J. P., Jankowiak, M., et al. 2019, J. Mach. Learn. Res., 20, 28:1
- Blondel, M., Berthet, Q., Cuturi, M., et al. 2021, arXiv preprint arXiv:2105.15183
- Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733
- Butler, R. P., Marcy, G. W., Williams, E., et al. 1996, PASP, 108, 500
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. 1995, SIAM Journal on Scientific Computing, 16, 1190
- Cardelli, J. A., Ebbets, D. C., & Savage, B. D. 1990, ApJ, 365, 789
- Carswell, R. F. & Webb, J. K. 2014, VPFIT: Voigt profile fitting program, Astrophysics Source Code Library, record ascl:1408.015
- Coffinet, A., Lovis, C., Dumusque, X., & Pepe, F. 2019, arXiv.org, arXiv:1901.03294
- Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4008, Optical and IR Telescope Instrumentation and Detectors, ed. M. Iye & A. F. Moorwood, 534–545
- Dumont, V. & Webb, J. K. 2017, MNRAS, 468, 1568
- Dzuba, V. A., Flambaum, V. V., & Webb, J. K. 1999, Phys. Rev. Lett., 82, 888
- Foreman-Mackey, D. 2023, tinygp: The tiniest of Gaussian Process libraries
- Geyer, C. 2011, Introduction to Markov Chain Monte Carlo (CRC Press), 3–48
- Grupp, F. 2003, A&A, 412, 897
- Hänsch, T. W. 2006, Reviews of Modern Physics, 78, 1297
- Hirano, T., Kuzuhara, M., Kotani, T., et al. 2020, PASJ, 72, 93
- Hoffman, M. D. & Gelman, A. 2014, Journal of Machine Learning Research, 15, 1593
- Horne, K. 1986, PASP, 98, 609
- Jovanović, N., Schwab, C., Guyon, O., et al. 2017, A&A, 604, A122
- Kambe, E., Sato, B., Takeda, Y., et al. 2002, PASJ, 54, 865
- King, J. A., Webb, J. K., Murphy, M. T., et al. 2012, MNRAS, 422, 3370
- Kotuš, S. M., Murphy, M. T., & Carswell, R. F. 2017, MNRAS, 464, 3679
- Lee, C.-C., Webb, J. K., Carswell, R. F., et al. 2023, MNRAS[arXiv:2212.00791]
- Lee, C.-C., Webb, J. K., Carswell, R. F., & Milaković, D. 2021a, MNRAS, 504, 1787
- Lee, C.-C., Webb, J. K., Milaković, D., & Carswell, R. F. 2021b, MNRAS, 507, 27
- Marconi, A., Abreu, M., Adibekyan, V., et al. 2022, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 12184, Ground-based and Airborne Instrumentation for Astronomy IX, ed. C. J. Evans, J. J. Bryant, & K. Motohara, 1218424
- Marcy, G. W. & Butler, R. P. 1992, PASP, 104, 270
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, The Messenger (ISSN0722-6691), 114, 20
- Milaković, D., Lee, C.-C., Carswell, R. F., et al. 2021, MNRAS, 500, 1
- Milaković, D., Pasquini, L., Webb, J. K., & Lo Curto, G. 2020, MNRAS, 493, 3997
- Murphy, M. T., Molaro, P., Leite, A. C. O., et al. 2022, A&A, 658, A123
- Murphy, M. T., Udem, T., Holzwarth, R., et al. 2007, Monthly Notices of the Royal Astronomical Society, 380, 839
- Nuttall, A. H. 1981, IEEE Transactions on Acoustics Speech and Signal Processing, 29, 84
- Oliva, E., Rainer, M., Tozzi, A., et al. 2019, A&A, 632, A21
- Phan, D., Pradhan, N., & Jankowiak, M. 2019, arXiv preprint arXiv:1912.11554
- Probst, R. A. 2015, PhD thesis, Ludwig Maximilian Universität, <http://hdl.handle.net/11858/00-001M-0000-002A-31A0-E>
- Probst, R. A., Milaković, D., Toledo-Padrón, B., et al. 2020, Nature Astronomy, 4, 603
- Probst, R. A., Wang, L., Doerr, H.-P., et al. 2015, arXiv.org, 023048
- Rahmani, H., Wendt, M., Srianand, R., et al. 2013, MNRAS, 435, 861
- Rasmussen, C. E. & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning
- Robertson, J. G. 1986, PASP, 98, 1220
- Rodler, F. & Lo Curto, G. 2019, HARPS User Manual
- Schmidt, T. M., Molaro, P., Murphy, M. T., et al. 2021, A&A, 646, A144
- Steinmetz, T., Wilken, T., Araujo-Hauck, C., et al. 2008, Science, 321, 1335
- Udem, T., Holzwarth, R., & Hänsch, T. W. 2002, Nature, 416, 233
- Valenti, J. A., Butler, R. P., & Marcy, G. W. 1995, PASP, 107, 966
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. 2019, arXiv e-prints, arXiv:1903.08008
- Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 2198, Instrumentation in Astronomy VIII, ed. D. L. Crawford & E. R. Craine, 362
- Walker, G. A. H., Shkolnik, E., Bohlander, D. A., & Yang, S. 2003, PASP, 115, 700
- Webb, J. K., Flambaum, V. V., Churchill, C. W., Drinkwater, M. J., & Barrow, J. D. 1999, Phys. Rev. Lett., 82, 884
- Webb, J. K., Lee, C.-C., & Milaković, D. 2022, Universe, 8, 266
- Whitmore, J. B. & Murphy, M. T. 2015, MNRAS, 447, 446
- Wilken, T., Curto, G. L., Probst, R. A., et al. 2012, Nature, 485, 611
- Wilken, T., Lovis, C., Manescu, A., et al. 2010, MNRAS, 405, L16
- Zhao, F., Lo Curto, G., Pasquini, L., et al. 2021, A&A, 645, A23
- Zhao, F., Zhao, G., Liu, Y., et al. 2019, MNRAS, 482, 1406

Appendix A: Choosing a centre for an asymmetric profile

Our method requires that *some* measure of an IP centre is provided, so that line centres can be determined through fitting (Eq. 7-9). For asymmetric profiles, there is no meaningful definition of the line centre and, looking through the relevant literature, we found scarce information about this particular point.

A&K introduced their own centre estimator based on the considerations for the distribution of flux on the detector produced by an unresolved source. They posited that the IP is properly centered when an unresolved source located at the boundary between the two brightest pixels results in equal fluxes in the boundary pixels, i.e. $\psi(\Delta x = -0.5) = \psi(\Delta x = 0.5)$. As a reminder, Δx is the distance of the pixel centre from the unresolved spectral feature. This definition is of the centre depends, of course, on the IP shape in the central region but its only assumption is that the IP is a smooth function peaking somewhere in between the two pixel centres. It was developed specifically to deal with asymmetric IP shapes but its performance with respect to other centre estimators has never before been quantified so we were unsure whether it was the most appropriate for our purposes.

We therefore performed a numerical exercise to determine what is the most appropriate centre estimator for asymmetric IPs. The criterion chosen for their ranking was the estimator's invariability under small perturbations of the IP shape (robustness), where the best estimator would be the least variable. Four different estimators were considered: the mean, the median, the mode, and the definition of line centre from Anderson & King (2000). The numerical exercise made use of the MCMC models from Appendix B and posterior distributions for hyperparameter values in particular. The posterior distribution for relevant hyperparameters was sampled 1000 values to create 1000 IP models shown in Fig. A.1. Centres of these models were subsequently determined using each of the four estimators and their distributions were examined.

The mean values for all four estimators are different and they are not consistent with each other based on the widths of their distributions (see the inset in Fig. A.1). The largest difference between any two estimators is ≈ 0.12 pix (between the median and A&K's estimator) and the smallest difference is ≈ 0.01 pix (between the median and the mean). Distributions for all four estimators show a single peak and are approximately symmetric around their central value (see Fig. A.2). Concentrating on the distribution width, we see significant differences across the estimators. The text in each of the four panels of the same Figure gives the central 68% distribution limits, when the distribution median is set to zero. Ranking the four estimators according to their distribution widths (from the most narrow to the most broad), we obtain: the A&K's estimator ($\psi(\Delta x = -0.5) = \psi(\Delta x = 0.5)$), the median, the mean, and the mode.

We therefore chose to centre all our empirical IP profiles ψ at the location which results in $\psi(\Delta x = -0.5) = \psi(\Delta x = 0.5)$. This is also reflected in Fig. A.1, which is plotted with its centre at that location.

Appendix B: A full Markov Chain Monte Carlo approach

All of the analysis presented in the main body of this work optimises the GP hyperparameters. While our results show that this is sufficient to describe the data, we wanted to check whether

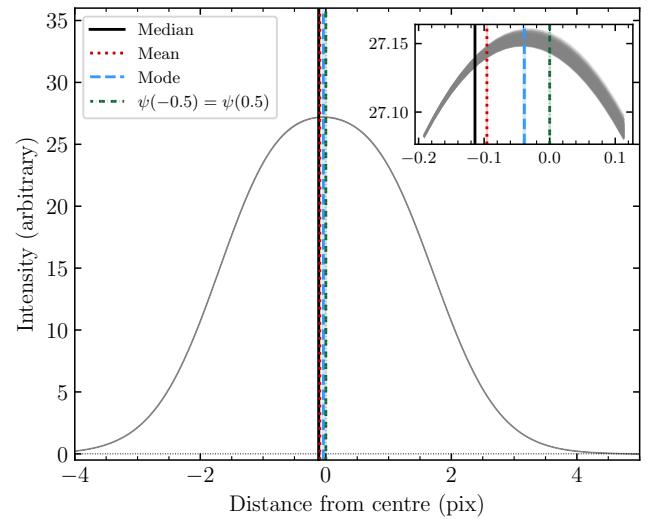


Fig. A.1. A set of 1000 MCMC HARPS instrumental profiles, trained on the data and shown on a fine grid. We estimated the centres of the 1000 profiles using the median, the mean, the mode, and the location satisfying $\psi(-0.5) = \psi(0.5)$ for each of the 1000 profiles, producing distributions for each of the four quantities. The means of the four distributions is shown as a vertical line of different colour and stroke, as explained in the figure legend. The x-axis zero point is set to the mean of the distribution for $\psi(-0.5) = \psi(0.5)$ centres. The inset on the top right shows a zoom-in of the region around the profile peak. Units in the inset are the same as in the main panel.

uncertainty in the hyperparameter values may significantly affect our conclusions. To test this, for one (?) line, we sampled the posterior distribution of the hyperparameters via Markov Chain Monte Carlo (MCMC) algorithm. Specifically, we used the Hamiltonian Monte Carlo/No-U-Turn sampler (Hoffman & Gelman 2014) as implemented in the `numpyro` (Bingham et al. 2019; Phan et al. 2019) software. The posterior comprises a likelihood identical to that used in section Sec. 4.6 (i.e. with one GP for the line profile and another for the excess variance) along with prior distributions on the hyperparameters. For these *hyperpriors*, we used normal distributions, whose mean and variance were set to some reasonable number (see Fig. B.1). We ran four chains using 200 and 450 sampling steps. We assessed convergence of the chains via two diagnostics: the split Gelman-Rubin R (Vehtari et al. 2019) and effective sample size N_{eff} (Geyer 2011). The values of these diagnostics for all fitted parameters are shown in Tab. B.1.

The median IP model derived through MCMC is indistinguishable from the IP model derived through the L-BFGS-B method and the differences between the two models are consistent with the noise (Fig. B.2). The posterior probability distributions for hyperparameters $A, \mu, \sigma, y_0, a, l$ for the instrumental profile, and (a_g, l_g) for the excess variance are shown in Fig. B.4. The hyperparameter values returned by the L-BFGS-B method for the same data are also shown as red lines and squares, and printed out in the top right corner of the figure. Hyperparameter values associated to the IP shape returned by the L-BFGS-B method all fall within the central 68% of the posterior distribution determined using MCMC.

On the other hand, hyperparameters controlling the variance modification all fall outside of the central 68% of their respective MCMC posterior distributions. However, overplotting the $g(\Delta x)$ from both methods in Fig. B.3, we see that the two functions

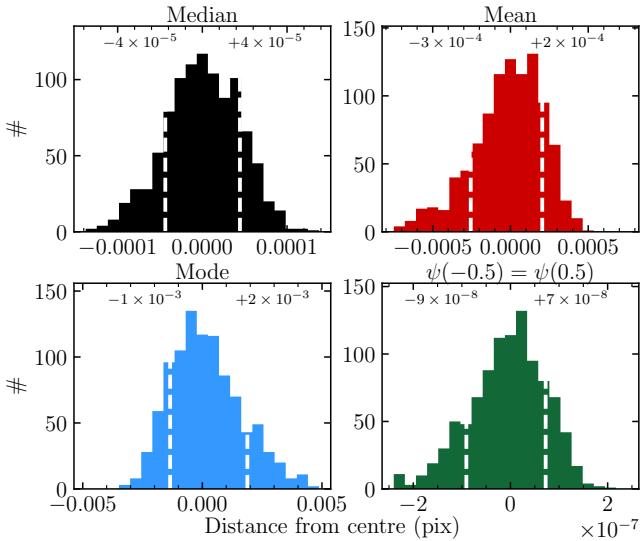


Fig. A.2. Each panel shows the distribution of “centre” measurements for the 1000 instrumental profiles shown in Fig. A.1, where the centre estimator used is indicated at the top of the panel. Since we were interested in quantifying the stability (distribution width) of each centre estimator under small perturbations of the IP, the distributions plotted were shifted to have zero median. The dashed white vertical lines show the limits of the central 68% of the distribution and the text above the lines gives their values in units pixel. Ranking centre estimators from the most stable to the least stable, we get: $\psi(-0.5) = \psi(0.5)$, the median, the mean, and the mode.

Table B.1. The convergence of the MCMC chains was confirmed using two statistics: \hat{R} (should be near 1) and N_{eff} (should be over 50).

Quantity	\hat{R}	N_{eff}
A	1.000	1262.332
μ	1.001	913.540
$\log(\sigma)$	1.001	1587.845
y_0	0.999	1070.650
$\log(a)$	0.999	959.304
$\log(l)$	1.001	810.947
$\log(\sigma_0)$	1.014	93.589
$\log(a_g)$	1.006	198.493
$\log(l_g)$	0.999	54.589
$\log(C_g)$	0.999	179.070
$\psi(\Delta x)$	1.005*	711.299*
$g(\Delta x)$	1.091*	82.120*

* Average over the data points.

agree very well so this disagreement in hyperparameter values is assumed not to be problematic.

Appendix C: Details of a single example model

Appendix D: Examples of astrocomb line fits

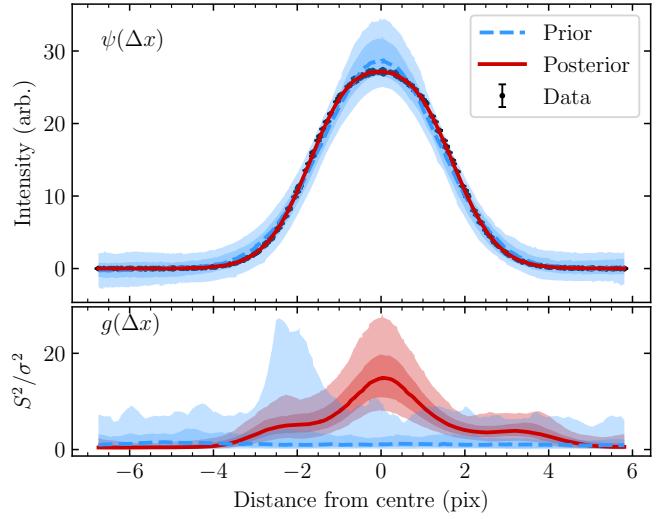


Fig. B.1. *Top panel:* The data are the same as in Fig. 3. The blue dashed line is the median of the prior distribution for the shape of the IP. The blue shaded bands around the blue line show the central 68% and 95% of the prior distribution. The red solid line and the corresponding red bands show the same quantities for the posterior distribution. *Bottom panel:* The lines and the bands have the same meaning as in the top panel but are for function $g(\Delta x)$ which modifies the variance array as per Eq. (24).

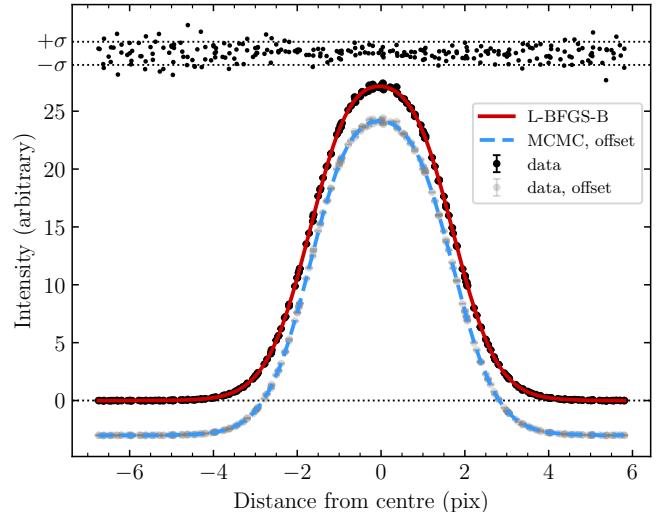


Fig. B.2. The IP models derived using the L-BFGS-B (solid red line overlaid on top of black points) and the MCMC (dashed blue line overlaid on top of grey points) methods are indistinguishable from one another. The two sets of lines and data points have been offset by 3 in y direction for clarity. The normalised residuals between the two models, i.e. the MCMC model minus the L-BFGS-B model divided by the error on the data, are shown as a black dots at the top of the panel. The two dotted horizontal lines bracketing the residuals indicate $\pm 1\sigma$, where σ is the error on the data.

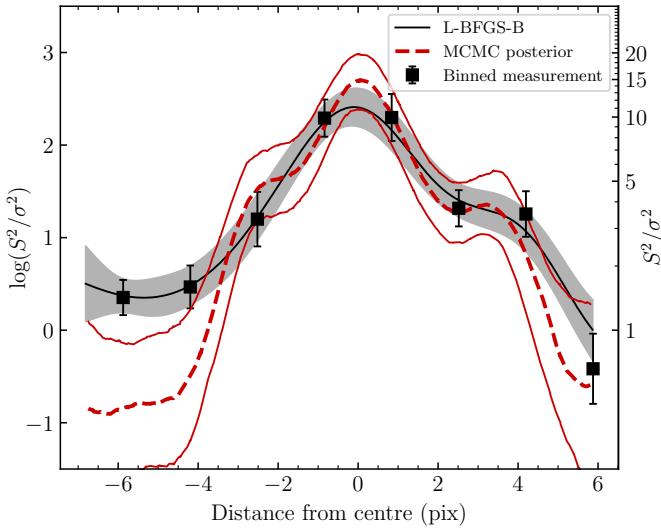
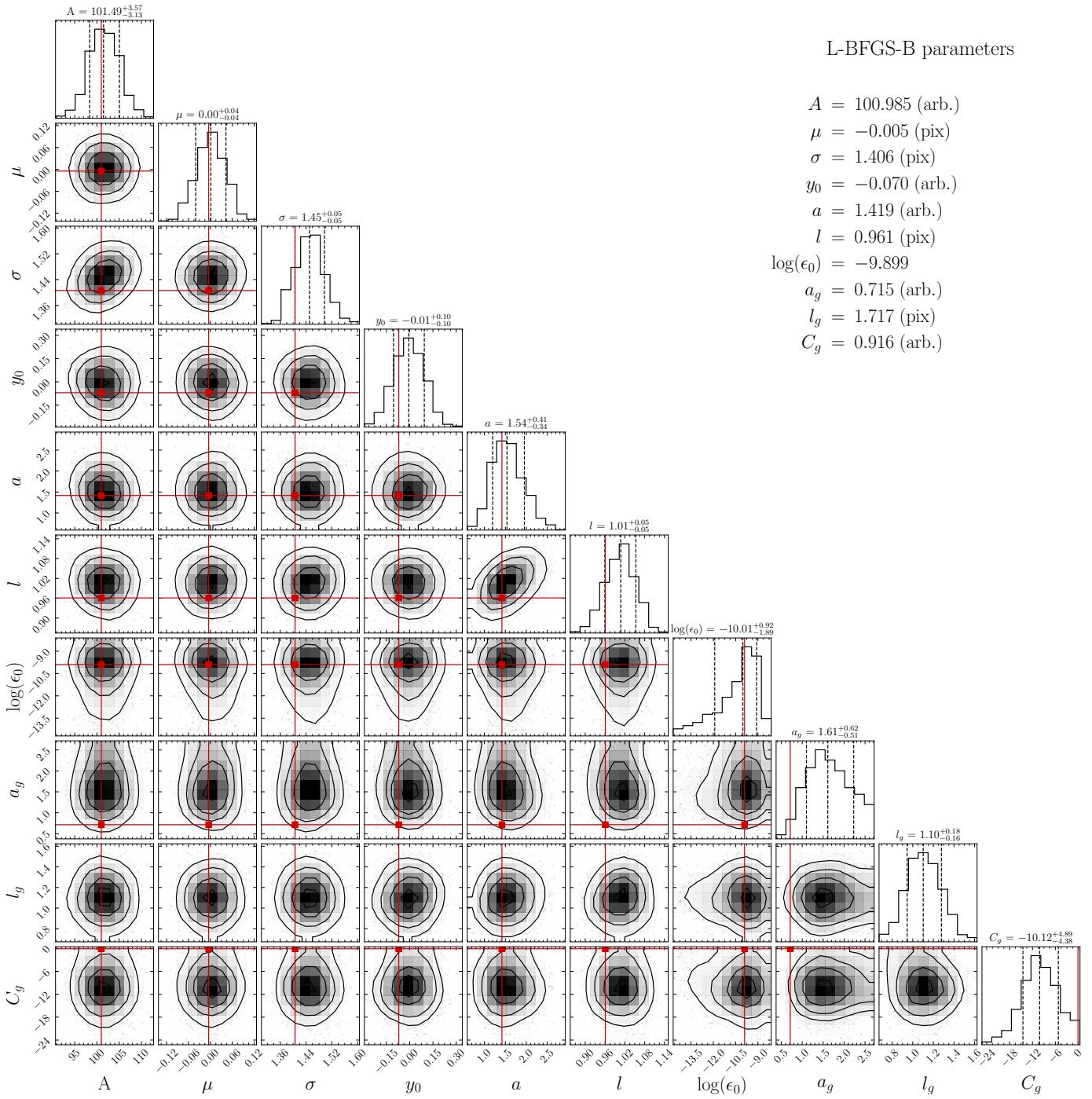


Fig. B.3. The same as in Fig. 4 but with MCMC posterior distribution for $g(x)$ overplotted. The black squares and the solid black line show the method for variance modification based on L-BFGS-B presented in the main body of the paper and implemented in this work. The dashed red line shows the median of the MCMC posterior distribution for $g(\Delta x)$ and the thin solid red lines show the central 68% limits. The two models are in excellent agreement.

**Fig. B.4.**

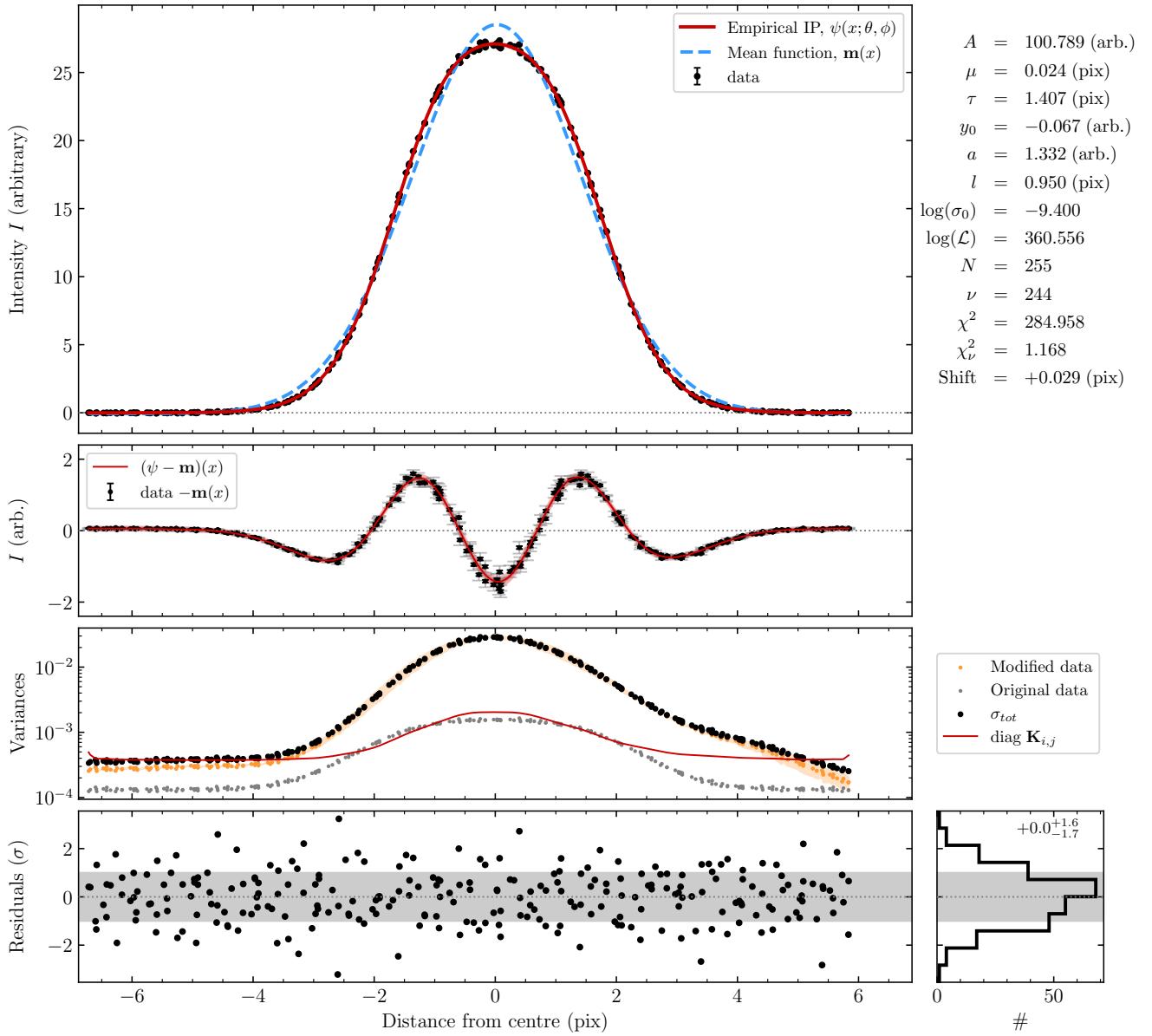


Fig. C.1. *First panel from the top:* The black points are the 255 samples of the IP, $\hat{\psi}$, in the 10th segment of order 110 derived from flux normalised and stacked astrocomb lines (Eq. (10)). The thick red line is the empirical IP model, $\psi(x)$ in the first iteration. The model was derived by performing a Gaussian Process regression on the black points (Sec. 2.4.1) with a secondary GP for variance estimation. The dashed blue line is the mean function of the GP, Eq. (17). The numbers to the right of the panel give additional information. The first seven numbers are the values of the parameters (θ, ϕ, σ_0) that minimise the loss function, $-\log \mathcal{L}$, where \mathcal{L} is the model likelihood (also quoted). N and ν are the number of data points and the number of degrees of freedom in the fit, respectively. χ^2 is given by Equation (22) and $\chi^2_\nu = \chi^2/\nu$. The large χ^2_ν value indicates that data variance has been underestimated. This can also be seen from the scatter of black points at $x \approx 0$ in the first and second panels from the top, which scatter beyond expectation based only on the formal error on each data point. Large χ^2 values of points at $x \approx 0$ also reflect this. We assess the empirical variance in Sec. ?? and show the resulting improvement in Fig. C.1. The final number in the list on the right gives shift applied to Δx to ensure proper centroiding, Eq. (20). *Second panel:* The black points show the difference between the data and the GP mean function (a Gaussian with parameters $\theta = \{A, \mu, \sigma, y_0\}$). The red line shows $\psi(x) - \mathbf{m}(x)$, i.e. departures from the Gaussian shape. *Third panel:* Grey dots are the variances on the $\hat{\psi}$, i.e. Equation (11). Orange points and bands are the variances inferred from the data (as per Eq. (24)) and the corresponding uncertainty. The inferred variance is substantially higher than the formal one everywhere, especially in the central region $-1 \leq x \leq 1$. The larger black dots are the sum of modified ϵ_{data} and σ_0 . The red line is the variance of the GP model, i.e. the diagonal of $\mathbf{K}_{i,j}$, Equation (18). *Fourth panel:* The grey dots show the residuals, model minus data, normalised by ϵ_{data} . The grey shaded area shows the $\pm 1\sigma$ range and the black dashed lines show the 5th and 95th percentiles. Note that the residuals blow up in the range $-4 \leq x \leq 4$, indicating that ϵ_{data} in that region is underestimated. The converse is true at $x \leq -5$. The small panel to the right shows the histogram of the residuals. The shaded area and the dashed lines are the same as in the panel immediately to the left. The number in the top right corner of the panel is the median, and the upper and lower limits correspond to the 5th and 95th percentiles.