

Methods for quasar absorption system measurements of the fine structure constant in the 2020s and beyond

D. Milaković^{1,2,3}, C.C. Lee⁴, P. Molaro^{2,1}, and J.K. Webb⁴

¹ Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34532 Trieste, Italy

² INAF – Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy

³ INFN, Sezione di Trieste, Via Valerio 2, I-34127 Trieste, Italy

⁴ Clare Hall, University of Cambridge, Herschel Rd, Cambridge CB3 9AL, United Kingdom

e-mail: dinko@milakovic.net, dmilakov@sissa.it

Received: 18-11-22; Accepted: 21-11-22

Abstract. This article reviews the two major recent developments that significantly improved cosmological measurements of fundamental constants derived from high resolution quasar spectroscopy. The first one is the deployment of astronomical Laser Frequency Combs on high resolution spectrographs and the second one is the development of spectral analysis tools based on Artificial Intelligence methods. The former all but eliminated the previously dominant source of instrumental uncertainty whereas the latter established optimal methods for measuring the fine structure constant (α) in quasar absorption systems. The methods can be used on data collected by the new ESPRESSO spectrograph and the future ANDES spectrograph on the Extremely Large Telescope to produce unbiased $\Delta\alpha/\alpha$ measurements with unprecedented precision.

Key words. cosmological parameters; methods: data analysis; instrumentation: spectrographs, quasars: absorption lines

1. Introduction

The fine structure constant, $\alpha \equiv e^2/\hbar c$ is a dimensionless parameter of the Standard Model of Particle Physics whose value determines the strength of the electromagnetic interactions. As such, it is considered one of the fundamental constants of nature. Variations of fundamental constants with time are possible, and even required, in higher-dimensional theories (e.g. string and Kaluza-Klein) and in theories in which a scalar field ϕ provides late-time

acceleration of the universal expansion rate (e.g. quintessence and Bekenstein type models). Detecting any such variations would disprove the Weak Equivalence Principle, a postulate of General Relativity. Uzan (2011) and Martins (2017) give an overview of theoretical models associated with fundamental constant variations.

Any deviation in α would cause a small, systematic shift in the energy levels of atoms and thus their wavelengths. This shift can

equivalently be expressed as a small spectral velocity shift, Δv :

$$\frac{\Delta v}{c} = -\frac{\Delta\alpha}{\alpha} \frac{2q}{\omega}. \quad (1)$$

Above, ω is the wavenumber of the transition, q quantifies its sensitivity to α variation, c is the speed of light, and

$$\frac{\Delta\alpha}{\alpha} = \frac{\alpha_{obs} - \alpha_{lab}}{\alpha_{lab}}, \quad (2)$$

where α_{obs} and α_{lab} refer to the values in the observed astronomical object and the laboratory, respectively.

q -coefficients have been determined through many-body quantum mechanical calculations (e.g. Dzuba et al. 1999). Measuring relative velocity shifts between a set of multiplets, covering a wide range of q -coefficient values, is the basis of the Many Multiplet (MM) method (Webb et al. 1999; Dzuba et al. 1999). To illustrate the required spectroscopic wavelength accuracy for such measurements, a 1 part-per-million (ppm) change in $\Delta\alpha/\alpha$ produces a relative velocity shift of $\approx 20 \text{ m s}^{-1}$ between Fe II $\lambda 2344$ and Mg II $\lambda 2796$ transitions, both commonly observed in quasar absorption systems.

The MM method was applied to quasar absorption system (Webb et al. 1999, 2001), white dwarf (Berengut et al. 2013; Bainbridge & Webb 2017; Hu et al. 2021), and stellar spectra (Hees et al. 2020; Murphy et al. 2022a). The application of the MM method to 293 quasar absorption systems (observed by VLT/UVES and Keck/HIRES) found evidence for a spatial dipole with an amplitude of ≈ 10 ppm and a 4.2σ statistical significance (King et al. 2012).

Later studies revealed the existence of short- and long-range wavelength-scale distortions in both UVES and HIRES (Rahmani et al. 2013; Whitmore & Murphy 2015). Left uncorrected, such distortions spoil $\Delta\alpha/\alpha$ measurements. While incorporating the knowledge on the distortions and the associated uncertainty is possible (Dumont & Webb 2017), it is clear that significant advancements to wavelength calibration of astronomical spectrographs is necessary to advance the field.

Quasar absorption systems commonly consist of clustered gas clouds which appear as a complex structure of blended Voigt profiles in the spectrum. It is possible to produce statistically acceptable models (in terms of χ^2 statistics) of the system with different – and discrepant – $\Delta\alpha/\alpha$ measurements (Wilczynska et al. 2015).

Information criteria (see Section 4.1) can discriminate between competing models, but modelling complex absorption systems is often time consuming for humans. Additionally, a human may unconsciously bias the results towards some preferred value. Unbiased and automated spectral fitting tools are necessary to explore these issues and ensure robust measurements.

In this paper, we review recent advancements in instrumentation delivered by the Laser Frequency Comb technology (Sec. 2), by the ESPRESSO spectrograph (Sec. 3), and data analysis tools based on Artificial Intelligence, AI-VPFIT (Sec. 4). We then present some lessons learned by applying AI-VPFIT to simulated and real quasar spectra in Sec. 5, with conclusions in Sec. 6.

2. Laser Frequency Combs

The Laser Frequency Comb (LFC, Udem et al. 2002; Hänsch 2006) in an ideal wavelength calibrator for high-resolution spectrographs. It produces a set of unresolved emission lines equidistant in frequency space. The frequency of an n^{th} line, f_n , is given by:

$$f_n = f_0 + n \times f_r, \quad (3)$$

The “offset” and “repetition” frequencies (f_0 and f_r , respectively) are stabilised to an external radio frequency standard with an accuracy of $\Delta f/f \sim 10^{-12}$ or, equivalently, 2 mm s^{-1} (Probst et al. 2020).

Astronomical LFCs have $f_r \approx 20 \text{ GHz}$, approximately a factor of three above their instrumental resolution element width. Imaging the LFC spectrum provides thousands of densely spaced lines on the detector. The measured line positions, together with their known frequencies and thus wavelengths, can be used to

wavelength calibrate the spectrograph and remove distortions to better than $\Delta\lambda/\lambda \simeq 10^{-8}$ or, equivalently, 3 m s^{-1} (Milaković et al. 2020). Whilst there are other potential sources of uncertainty, reducing the wavelength calibration errors to this level can reduce its contribution to the overall error budget on $\Delta\alpha/\alpha$ to $\leq 0.1 \text{ ppm}$.

2.1. Lessons from HARPS

Following a development period (Wilken et al. 2010, 2012), in 2015 HARPS (Mayor et al. 2003) became the first high resolution spectrograph to be equipped with an LFC for regular operations (Probst et al. 2014). The LFC performance was characterised in a series of tests described here.

The use of both static and dynamic mode scrambling methods in optical fibres carrying LFC light was found to be crucial for obtaining a uniform and stable line spread function and therefore for obtaining wavelength calibration stability at the photon noise limit (Probst et al. 2020). When fitting Gaussians to LFC lines, it was found that line centres shift by $\approx 1 \text{ m s}^{-1}$ between consecutive exposures when dynamic scrambling is not used (Probst et al. 2020). Furthermore, the measured Gaussian line centres were found to depend on the flux level, due to charge transfer inefficiency (CTI) of the detector, with shifts as large as 3 m s^{-1} for sufficiently low fluxes (Milaković et al. 2020; Zhao et al. 2021). Using static and dynamic scrambling and applying a simple CTI correction, the HARPS LFC demonstrated a 3 m s^{-1} wavelength calibration accuracy (the difference between the wavelength calibration polynomial and the known LFC line wavelength, averaged over $\sim 10 \text{ k}$ lines) in a single exposure and $< 1 \text{ cm s}^{-1}$ wavelength calibration stability in a series of 194 exposures taken over period of several hours (Milaković et al. 2020).

A disturbing result emerged from the simultaneous use of two LFCs on HARPS (Probst et al. 2020; Milaković et al. 2020). The two LFCs shared the same basic design but had different f_0 and f_r and were operated independently. The comparison of the wavelength calibrations from the two LFCs revealed an unexpected $60 \pm 1 \text{ cm s}^{-1}$ discrepancy in their zero-

points. Applying a simple CTI correction decreased the discrepancy to 45 m s^{-1} (c.f. the thin and thick black and red lines in figure 15 of Milaković et al. 2020). Further improvements could be achieved using a more accurate CTI model and possibly by employing the empirical instrumental profile instead of its Gaussian approximation (see Sec. 3.1 regarding the latter).

Although the tests discussed above have not identified any systematic effects associated with the use LFC itself, the origin of these shifts is not yet fully understood. A tunable LFC (i.e. one for which f_0 or f_r can be precisely controlled) may be useful in this respect. Such a system can be used to change the LFC line positions at will and, in this way, scan the detector in sub-pixel steps. In addition to producing more detailed CTI models and reconstructing the instrumental profile on very small wavelength scales, this could also be used to quantify the impact of intra-pixel sensitivity variations on line centre measurements.

3. ESPRESSO

ESPRESSO is a high resolution ($R = \lambda/\delta\lambda = 145\text{k}$, where $\delta\lambda$ is the full-width at half-maximum of the resolution element) spectrograph designed for measurements at the photon noise limit, with α measurements being one of its main scientific goals (Pepe et al. 2021). Wavelength calibration errors were reduced to a negligible level in ESPRESSO using an LFC (Schmidt et al. 2021b; Murphy et al. 2022b). In addition, ESPRESSO is also equipped with a thorium-argon (ThAr) lamp which is observed simultaneously with a Fabry-Pérot etalon (FP) for high precision calibration.

3.1. Instrumental profile reconstruction

Another significant advantage of the LFC is the narrow intrinsic width of its lines, $\approx 1 \text{ MHz}$ (c.f. full-width at half maximum of the instrument resolution element $\approx 5 \text{ GHz}$), before passing through the instrument optics. As such, images of LFC lines on the detector directly trace the instrumental profile (IP). Reconstruction of ESPRESSO's IP (Fig. 1)

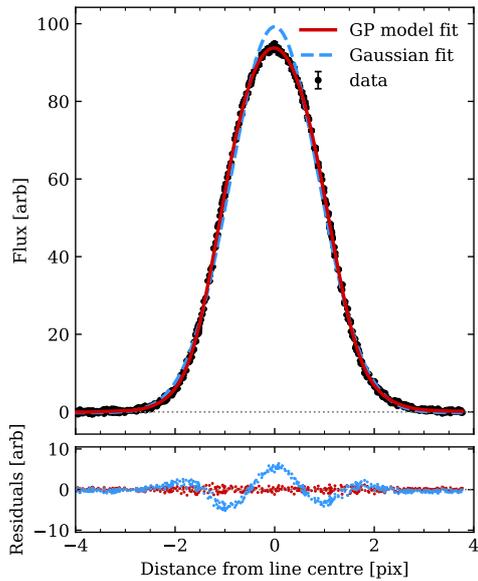


Fig. 1. *Top panel:* LFC lines in the region 500 ± 0.25 nm, observed in one slice of ESPRESSO, were flux normalised and stacked on their Gaussian profile centres (black points, $N = 520$). The reconstructed, non-parametric, non-Gaussian IP is shown as a solid red line. The reconstruction was performed using Gaussian process (GP) regression on the black data points (Milaković et al., in prep.). A Gaussian IP fitted to the same data is shown as a dashed blue line for comparison. The difference in the modes of the two fitted profiles is 0.01 pix, corresponding to ≈ 4 m s $^{-1}$. *Bottom panel:* Residuals to the two IPs fitted above. Residuals to the non-parametric IP are shown in red. Asymmetry of the empirical IP is seen as a wavy pattern in the residuals to the Gaussian IP approximation, shown in blue.

was done in a non-parametric way, i.e. using Gaussian process regression (Rasmussen & Williams 2006). Preliminary results revealed that the IP is asymmetric in pixel space and that the asymmetry varies as a function of position on the detector (Fig. 2). Fitting the LFC lines using the empirical IP results in differences in the line mode as large as 80 m s $^{-1}$ with respect to when a Gaussian IP is assumed (Milaković et al., in prep.). For comparison, the desired accuracy in measuring line centroids in $\Delta\alpha/\alpha$ studies is 5 m s $^{-1}$.

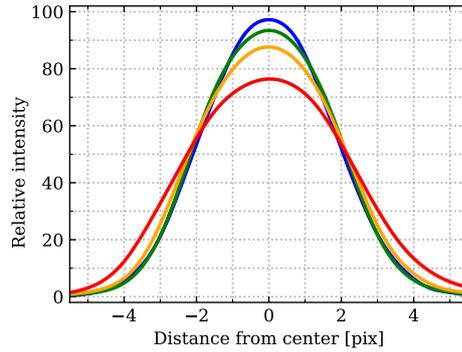


Fig. 2. The coloured lines show the empirical IP of ESPRESSO in a single slice of the optical order 122 (central wavelength 500 nm). Different colours correspond to different sections of the order, changing from blue to red with increasing wavelength.

Knowledge of the empirical IP can be used to improve wavelength calibration accuracy, better understand instrumental effects, and in spectral modelling. For example, some of the problems with ESPRESSO wavelength calibration (discussed in Section 3.2) may be resolved by better knowledge of the line centres. Importantly, they may also explain the observed difference in the zero-points of two independent LFCs on HARPS (Sec. 2.1). Lastly, Voigt profile fitting tools such as VPFIT (Carswell & Webb 2014) and AI-VPFIT (see Sec. 4) already allow for the use of an empirical IP so, as soon as the IP models are validated, they will be used in future analyses.

3.2. Limitations of the ESPRESSO LFC

Since its installation in 2018, ESPRESSO's LFC demonstrated a series of issues preventing its use for regular instrument calibrations and monitoring. Hardware issues caused the LFC's spectral coverage and line intensities to vary on timescales of minutes, preventing a reliable wavelength calibration for the instrument. Additionally, a detailed study of the ESPRESSO's LFC identified other unexpected problems preventing its performance at theoretical limits (Schmidt et al. 2021b).

Firstly, the ThAr+FP and LFC wavelength calibrations differ up to 30 m s^{-1} within a single echelle order, higher than expected considering the $2\text{--}4 \text{ m s}^{-1}$ uncertainties on the positions of LFC and ThAr lines (cf. figure 13 of Schmidt et al. 2021b). This issue probably points to inaccuracies of ThAr atlases but has not been investigated further. Secondly, the residuals to the LFC (but also ThAr+FP) wavelength calibration solution show peculiar properties. The histogram of the residuals is double-peaked and the standard deviation of the residuals around the best fit wavelength calibration solution show excess scatter: approximately 6 m s^{-1} (68% c.l.) instead of the expected 2 m s^{-1} (cf. figure 10 of Schmidt et al. 2021b). Thirdly, residuals of consecutive LFC lines are correlated on short pixel-scales (cf. figure 9 of Schmidt et al. 2021b). The mentioned effects, while unexpected, are unlikely to spoil $\Delta\alpha/\alpha$ measurements.

Finally, technological limitations constrain ESPRESSO LFC’s spectral range to the region between 490 and 720 nm, approximately 43% of the wavelength range of ESPRESSO. Important ultraviolet transitions, e.g. Fe II $\lambda 1608$ and $\lambda 1611$, are therefore not covered at $z < 2$. Additionally, the lack of LFC coverage limits the wavelength range in which IP can be determined. This is an additional motivation for the development of a new generation of astronomical LFCs (Obrzud et al. 2019, 2021; Schmidt et al. 2021a). Instrument servicing in May 2022 and in October 2022 aimed to resolve the technical issues. Preliminary results show a broader wavelength coverage and more stable flux levels.

3.3. A digitization issue in ESPRESSO before May 2022

All exposures taken between ESPRESSO’s commissioning in November 2018 and 19 May 2022 were affected by a digitization issue caused by a malfunction in the Next Generation Controller boards (Baade et al. 2009). The problem was first reported in Murphy et al. (2022c) and resolved quickly after being detected in late 2021, as explained shortly. The malfunction caused individual

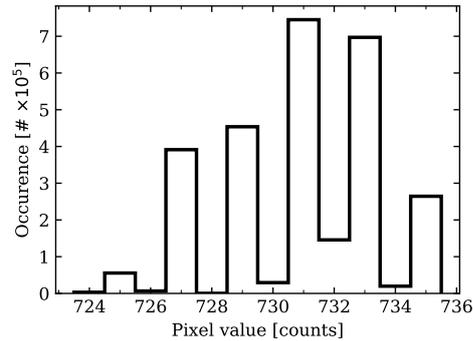


Fig. 3. A histogram of raw pixel values in a single read-out region of ESPRESSO blue detector’s bias frame taken on 3rd November 2018 at 10.11 UTC. Approximately 93% (7%) of the values are odd (even) as opposed to the expected 50/50% occurrence. The issue affects all read-out regions and appears in all ESPRESSO raw frames. See text in Sec. 3.3 for more details.

segments of the detector, each corresponding exactly to a single read-out region, to preferentially output odd or even numbers. The issue was tracked to a small variability in the reference voltage used in the ADC (Antonio Manescau, priv. comm.) In the worst affected segments, such as the one shown in Fig. 3, 7% (93%) of pixels had even (odd) values.

Raw detector frames thus contained pixels for which the read-out value is off by one with respect to their true values. The affected pixels seem to vary randomly from one exposure to the next. The amplitude of the effect does not depend on the recorded flux levels in the exposure. The impact of this effect on science measurements derived from quasar spectra has not yet been assessed. The boards were since replaced and observations taken after 19 May 2022 do not suffer from this problem.

4. Spectral modelling using Artificial Intelligence

Recent years saw the development of novel spectral modelling tools designed to produce objective, robust and unbiased $\Delta\alpha/\alpha$ measurements: GVPFIT (Bainbridge & Webb 2017) and AI-VPFIT (Lee et al. 2021a).

AI-VPFIT combines non-linear least squares minimisation, genetic algorithms, Monte Carlo (MC), and Artificial Intelligence (AI) methods to automatise the process of absorption system modelling. Its genetic algorithm relies on an information criterion to choose the best-fit model in each generation and to decide when to stop the modelling process. The use of high performance computing increased the efficiency by a factor of > 100 over a human modeller: a model of an arbitrarily complex absorption system can be produced in several hours or days using supercomputers¹. This allowed a series of checks on the assumptions made in previous studies, revealing biases and sources of systematic uncertainty, as discussed in this section. Importantly, it also identified ways to avoid them in future measurements.

4.1. Model building based on information theory

The optimal model should strike a balance between simplicity (i.e. the number of model parameters) and parameter bias and variance, as discussed in Bozdogan (1987); Webb et al. (2021). An information criterion (IC) penalises the introduction of unnecessary model parameters by adding a penalty term (\mathcal{P}) to the χ^2 statistic, $IC = \chi^2 + \mathcal{P}$. When comparing several competing models, each of which gives $\chi^2/\nu \simeq 1$ (where ν is the number of degrees of freedom), the preferred model is the one that minimizes the IC. This way, the IC allows for objective model selection based on information theory (see, e.g., Liddle 2007).

AI-VPFIT allows the user a choice between three IC:

1. the corrected Akaike Information Criterion (AICc), $\mathcal{P} = 2kN/(N - k - 1)$ (Akaike 1974; Hurvich & Tsai 1989).
2. the Bayesian Information Criterion (BIC), $\mathcal{P} = k \ln(N)$ (Schwarz 1978);
3. the Spectroscopic Information Criterion (SpIC), for which the exact \mathcal{P} formula is in Webb et al. (2021).

¹ A human would require several weeks or months for completing the same task.

Above, k and N are the total number of free model parameters and the number of data points, respectively.

In a typical quasar absorption system modelled by AI-VPFIT, AICc tends to overfit the data and BIC tends to underfit it, while SpIC strikes a balance between the two (c.f. fig 2 in Webb et al. 2021). The greatest benefit of SpIC is that its formulation considers the “local” impact of model parameters by taking into account the number of pixels they affect and the strength of the spectral feature being fitted. This is crucial in order to avoid systematic effects in $\Delta\alpha/\alpha$ measurements, as discussed in Section 5.

4.2. The importance of the Monte Carlo approach

The introduction of Monte Carlo approaches in conjunction with AI-VPFIT offer several vital advantages. However, first it is helpful to clarify what exactly the Monte Carlo aspects are. Inherent to the AI-VPFIT itself is the randomised way in which trial components are added to the absorption complex, during model construction. There is no “intelligent” placement of components, which may at first sight seem inconsistent with an interactive human approach. Interactively, a human may be more likely to target the strongest visible component, or if not, perhaps the “least blended” component. Clearly, throwing down components randomly does not necessarily achieve this human selection procedure, although equally clearly, *some* of the trial positions in a Monte Carlo sense *will* replicate the interactive approach.

But there is second Monte Carlo aspect to AI-VPFIT usage; one can form a large number of models of any given absorption system, each time using a different random number seed for trial component placement. In this way, a large number of independently constructed models has the capacity to incorporate all possible human modellers.

Together, these two Monte Carlo procedures not only permit a completely unbiased approach to model construction, but also enable an exploration of parameter space so that

multiple minima in χ^2 space can be mapped out. At the time of writing this paper, relatively few absorption systems have been modelled using AI-VPFIT. Therefore, how common multiple minima are is not yet known. Nevertheless, on the basis of what has been done, it seems that the phenomenon is likely to be common (Lee et al. 2021b; Webb et al. 2022).

The discussion above naturally raises the question as to what *causes* multiple minima in χ^2 space in the context of $\Delta\alpha/\alpha$ measurements. For example, would some completely different statistical procedure (e.g. machine learning) identify the same local minima as does non-linear least squares? The answer to this is not known. How sensitive is the shape of χ^2 space to spectral S/N and resolution? Again, this has not yet been quantified. Nevertheless, it is clear that individual error bars (e.g. see the blue points in figure 6 of Webb et al. 2022) will grow with decreasing S/N and lower resolution, such that local minima will be less pronounced or indistinguishable.

5. Insights from AI-VPFIT

Several thousand AI-VPFIT models were constructed for both simulated and real quasar spectra and the impact of particular choices for the model building process were investigated in a series of papers by Lee et al. (2021a,b); Webb et al. (2021, 2022). These include assumptions on the gas broadening mechanism, the choice of the IC, the decision to hold the $\Delta\alpha/\alpha$ fixed throughout the model building process. The following section outlines the main results of the four publications mentioned in this paragraph.

5.1. On the line broadening mechanism

Assumptions on absorbing gas motion proved to be crucial in obtaining robust $\Delta\alpha/\alpha$ measurements. Thermal and turbulent motions in the absorbing gas cloud impact on the width of its observed Voigt profile (b -parameter) such

that the total, compound, b -parameter is given by:

$$b_{\text{tot}}^2 = b_{\text{turb}}^2 + b_{\text{therm}}^2, \quad (4)$$

where b_{turb} arises due to turbulence and $b_{\text{therm}}^2 = 2k_B T/m$ depends on gas temperature T and atomic mass m , with k_B the Boltzmann's constant.

Simulations showed that superfluous components are required to obtain a good fit to the data when compound broadening is not permitted, and that the extra components often emulate compound broadening. Furthermore, because of degeneracies introduced by extra components' parameters, the statistical uncertainty² on $\Delta\alpha/\alpha$ increases. Allowing for compound broadening was the most reliable way to retrieve the input parameters. The converse was also true: assuming an incorrect line broadening mechanism may generate false local minima and hence produce a spurious result (Lee et al. 2021a; Webb et al. 2022).

5.2. On model non-uniqueness

Model non-uniqueness was further explored by Lee et al. (2021b) using HARPS observations of the quasar HE0515–4414 ($R = 115k$, data described in Milaković et al. 2021) and UVES observations of the quasar Q0528–2505 ($R = 70k$, data described in Murphy et al. 2019)). Several hundred models were made for each of the two quasars using different combinations of the line broadening mechanism (turbulent, thermal, or compound) and the IC (AICc or SpIC) and their resulting $\Delta\alpha/\alpha$ measurements compared.

The main results of that analysis can be seen in their figure 1. For both quasars, non-uniqueness was present for certain combinations of the broadening mechanism and IC. When non-uniqueness was present, the scatter in $\Delta\alpha/\alpha$ was at least comparable to the average statistical uncertainty (see, e.g., panels (a) and (d) of the aforementioned figure). Non-uniqueness was the most severe when AICc

² Calculated from the diagonal of the covariance matrix at the best fit solution.

and turbulent broadening were used (c.f. panels (a) and (g) of the same figure). Yet, when SpIC and compound broadening were used, HE0515–4414 showed no evidence for non-uniqueness (c.f. panels (a) and (e)). The benefit of SpIC and compound broadening was less evident for Q0528–2505 (c.f. panels (g) through (l)), possibly because of the lower resolution of the UVES observations.

The combination of using SpIC and compound broadening significantly decreased the probability of ending up in a local minimum, therefore lowering the systematic uncertainty associated with non-uniqueness in the analysed HARPS observations of HE0515–4414. The opposite was true for all other combinations of IC and gas broadening (as seen from the large scatter of points in the corresponding panels of figure 1 in Lee et al. 2021b).

5.3. On bias

One of the conclusions from AI-VPFIT is that modelling procedures matter. Assumptions on gas broadening and the choice of IC are directly (but partially) related to the probability of ending up in a local minimum and hence the level of bias. This raises an important question regarding avoid bias in future measurements.

Prior to the AI-VPFIT studies that have revealed that substantial bias may actually be introduced, one approach attempting to avoid biasing $\Delta\alpha/\alpha$ measurements has been to fix $\Delta\alpha/\alpha = 0$ during model building. Only once the model has been obtained in this way was $\Delta\alpha/\alpha$ included as a model parameter. The assumptions here are that the procedure just described does not cause the model to get stuck in a local minimum, i.e. that an unbiased value of $\Delta\alpha/\alpha$ is always recovered. This approach was followed in almost all published analyses (e.g. Webb et al. 1999, 2001; King et al. 2012; Kotuš et al. 2017; Murphy et al. 2022b).

Going a step further, some authors applied additional methods to their data in order to avoid biasing results (e.g. Evans et al. 2014; Kotuš et al. 2017; Murphy & Cooksey 2017; Murphy et al. 2022b). Weak long-range and intra-order distortions were introduced into the wavelength-scales of individual

quasar exposures before their combination, effectively offsetting $\Delta\alpha/\alpha$ from its true value by $\lesssim 5$ ppm (Murphy et al. 2022b). The offset magnitude is random and unknown outside of the software used to introduce the distortions (UVES_pop1er, Murphy 2018). The modelling is done on the distorted spectrum in the way described above. The only exception happens at the very end, when the final optimisation of model parameters is done on the original, undistorted data instead.

However, Webb et al. (2022) have shown that the combination of fixing $\Delta\alpha/\alpha = 0$ and applying distortions is particularly vulnerable to bias on $\Delta\alpha/\alpha$ when turbulent broadening is used to simultaneously model transitions arising in atoms of different masses. The reason for that is twofold. Firstly, if (for example) the absorption components being modelled are in reality thermally broadened but are fitted using turbulently broadened profiles, the final model will be forced to contain more components to achieve a good fit to the data. The presence of superfluous absorbing components creates more degeneracy i.e. the individual parameters are less well constrained. Secondly, if $\Delta\alpha/\alpha$ is fixed to be zero during model construction, the parameters describing those superfluous absorbing components can more readily adjust to give a good fit to the data. The net result is that the model obtained in this way is likely to end up in a false local minimum. It has previously been assumed that the final step of releasing $\Delta\alpha/\alpha$ as a free parameter can “release” the model from its false local minimum and iterate to the true value of $\Delta\alpha/\alpha$, whatever that may be. We now know this is wrong and in fact, once *in* a local minimum, it is likely that the model stays there, ending up with a spurious result.

The bias caused by fixing $\Delta\alpha/\alpha = 0$ and applying distortions was investigated using a simulated ESPRESSO spectrum of HE0515–4414 (Webb et al. 2022). The simulated spectrum was created with $\Delta\alpha/\alpha = 8.08$ ppm (a value previously recovered from the real data). One hundred AI-VPFIT models were created, each with $\Delta\alpha/\alpha$ set to zero, for each of the four combinations of turbulent/compound broadening and AICc/SpIC.

$\Delta\alpha/\alpha$ was only allowed to vary (together with other model parameters) in the final optimisation step.

Considering compound broadening only, and referring to figure 6 in Webb et al. (2022) (i.e. the red points), we can see good consistency between the individual measurements. There seems to be no bias (the measurements reflect the input value of $\Delta\alpha/\alpha$) and there are no multiple minima. Conversely, considering only turbulent broadening (i.e. the blue points), we see enhanced scatter as well as clear multiple minima. In fact, a $\Delta\alpha/\alpha$ value consistent with the input value was recovered 100% of the time only when compound broadening and SpIC were used, with compound broadening plus AICc almost as good. A measurement consistent with the input $\Delta\alpha/\alpha$ was recovered only $\approx 10\%$ of the time when turbulent broadening and AICc was used. Instead, a measurement consistent with zero was found $\approx 50\%$ of the time. Given that the average statistical uncertainty in those models was $\langle\sigma_{\text{stat}}\rangle \approx 2.5$ ppm, a no-variation result is $\sim 3\sigma$ away from the simulated input.

6. Discussion & conclusions

This article reviews a series of publications that have advanced measurements of the fine structure constant over the last several years. The most significant of them are the deployment of astronomical Laser Frequency Combs for wavelength calibration and the development of AI-VPFIT for spectral analysis. The former led to more accurate wavelength calibration of the quasar spectra from which the measurements are derived and the latter led to significant improvements in measurement methods.

The application of LFC wavelength calibration methods on HARPS and ESPRESSO removed the previously dominant source of instrumental uncertainty in $\Delta\alpha/\alpha$ measurements (Sec. 3). The development of AI-VPFIT (Sec. 4) has been equally groundbreaking. Not only did it increase the speed at which measurements can be made, but it identified methods to avoid non-uniqueness and reduce measurement bias (Sec. 5).

ESPRESSO is expected to observe dozens of quasars during its operational lifetime, producing many more high quality, high S/N spectra suitable for varying constant studies. The developments described here will be useful for the analysis of that data, but also of the data collected using the ANDES spectrograph (Marconi et al. 2022), to be installed on the Extremely Large Telescope in the future.

The conclusions, summarised as enumerated points below, are derived mostly from published work referenced in the text. The references to the original work from which conclusions numbered 1–7 are derived can be found in the sections therein stated. Conclusions numbered 8, 9, and 10 are derived jointly from the four publications referenced at the beginning of Section 5. For this reason, claims mentioned in conclusions 8–10, but originally presented elsewhere, are given their exact references.

The main conclusions are:

1. Instruments calibrated using LFCs are free from wavelength-scale distortions at the level of $\Delta\lambda/\lambda \simeq 10^{-8}$ (Sec. 2). This translates into a maximum systematic uncertainty on $\Delta\alpha/\alpha$ of approximately < 0.1 ppm due to wavelength calibration errors. Even in the absence of an LFC, distortions in the ThAr+FP wavelength calibration on ESPRESSO are only marginally worse (Sec. 3).
2. LFCs are also excellent tools for characterising instrumental effects (e.g. CTI, Sec. 2.1) and reconstructing the empirical IP (Sec. 3.1). Tunable LFCs may increase their utility as tools for instrument characterisation.
3. Empirical models of the ESPRESSO IP will further improve wavelength calibration accuracy and hence reduce associated systematics (Sec. 3.1). When provided to software such as AI-VPFIT, the IP models will lead to more accurate models of absorption systems.
4. Developments in LFC technology are expected to address outstanding issues like LFC reliability and wavelength coverage (Sec. 3.2).

5. The zero-points of two independent LFCs observed on HARPS differ by 45 cm s^{-1} (Sec. 2.1). This has significant implications on science cases requiring $\sim 1 \text{ cm s}^{-1}$ instrument stability over periods > 1 year. Future instruments which will rely on LFC to track instrument drifts, such as NIRPS (Bouchy et al. 2017) and ANDES (Marconi et al. 2022), could cross-check instrument stability by using redundant calibration sources with similar precision (e.g. a secondary LFC or an absorption cell).
6. An unexpected digitisation issue introduced weak flux artefacts into ESPRESSO spectra of faint targets (such as quasars) taken before 19 May 2022. Repeating some faint target observations may be useful in understanding the effect so it can be efficiently and automatically removed from the affected datasets. Currently, that data must be cleaned manually (Sec. 3.3).
7. The experience from using HARPS and ESPRESSO for high precision spectroscopic measurements highlights the importance of understanding the full data production and data analysis chains. Further work is required before these instruments can be used optimally.
8. When fitting multiple species simultaneously to measure $\Delta\alpha/\alpha$, only compound broadening and SpIC (or AICc) should be used for model building (Webb et al. 2022). Turbulent broadening should not be used for simultaneous fits to multiple species as doing so will likely bias the measurement (Webb et al. 2022).
9. Uncertainty on $\Delta\alpha/\alpha$ due to model non-uniqueness is sometimes comparable to or even larger than the statistical uncertainty (Lee et al. 2021a,b; Webb et al. 2021, 2022). For any particular absorption system, multiple models should be produced to quantify this uncertainty (Webb et al. 2022). Interestingly, using compound broadening and SpIC can eliminate non-uniqueness in some observations (Lee et al. 2021b). A future analysis using archival (e.g. UVES) and novel ESPRESSO observations may be used to explore the role of spectral resolution and S/N.
10. Non-uniqueness may be endemic in quasar absorption systems. Conclusions on the variability of fundamental constants will then have to be made on a large statistical sample instead of on individual quasar absorption systems (Webb et al. 2022).

Acknowledgements. DM is also supported by the INFN PD51 INDARK grant.

References

- Akaike, H. 1974, IEEE Transactions on Automatic Control, 19, 716
- Baade, D., Balestra, A., Cumani, C., et al. 2009, The Messenger, 136, 20
- Bainbridge, M. B. & Webb, J. K. 2017, MNRAS, 468, 1639
- Berengut, J. C., Flambaum, V. V., Ong, A., et al. 2013, Phys. Rev. Lett., 111, 010801
- Bouchy, F., Doyon, R., Artigau, É., et al. 2017, The Messenger, 169, 21
- Bozdogan, H. 1987, Psychometrika, 52, 345
- Carswell, R. F. & Webb, J. K. 2014, VPFIT: Voigt profile fitting program, Astrophysics Source Code Library, record ascl:1408.015
- Dumont, V. & Webb, J. K. 2017, MNRAS, 468, 1568
- Dzuba, V. A., Flambaum, V. V., & Webb, J. K. 1999, Phys. Rev. Lett., 82, 888
- Evans, T. M., Murphy, M. T., Whitmore, J. B., et al. 2014, MNRAS, 445, 128
- Hänsch, T. W. 2006, Reviews of Modern Physics, 78, 1297
- Hees, A., Do, T., Roberts, B. M., et al. 2020, Phys. Rev. Lett., 124, 081101
- Hu, J., Webb, J. K., Ayres, T. R., et al. 2021, MNRAS, 500, 1466
- Hurvich, C. M. & Tsai, C.-L. 1989, Biometrika, 76, 297
- 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., & Milaković, D. 2021a, MNRAS, 504, 1787
- Lee, C.-C., Webb, J. K., Milaković, D., & Carswell, R. F. 2021b, MNRAS, 507, 27
- Liddle, A. R. 2007, MNRAS, 377, L74
- 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
- Martins, C. J. A. P. 2017, *Reports on Progress in Physics*, 80, 126902
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, *The Messenger*, 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. 2018, *Mtmurphy77/Uves_Popler: Uves_Popler: Post-Pipeline Echelle Reduction Software*, Zenodo
- Murphy, M. T., Berke, D. A., Liu, F., et al. 2022a, arXiv e-prints, arXiv:2211.05150
- Murphy, M. T. & Cooksey, K. L. 2017, *MNRAS*, 471, 4930
- Murphy, M. T., Kacprzak, G. G., Savorgnan, G. A. D., & Carswell, R. F. 2019, *MNRAS*, 482, 3458
- Murphy, M. T., Molaro, P., Leite, A. C. O., et al. 2022b, *A&A*, 658, A123
- Murphy, M. T., Molaro, P., Schmidt, T. M., et al. 2022c, *The Messenger*, 188, 15
- Obrzud, E., Brasch, V., Voumard, T., et al. 2019, *Optics Letters*, 44, 5290
- Obrzud, E., Denis, S., Sattari, H., et al. 2021, *APL Photonics*, 6, 121303
- Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, *A&A*, 645, A96
- Probst, R. A., Lo Curto, G., Avila, G., et al. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, ed. S. K. Ramsay, I. S. McLean, & H. Takami, 91471C
- Probst, R. A., Milaković, D., Toledo-Padrón, B., et al. 2020, *Nature Astronomy*, 4, 603
- 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*
- Schmidt, T., Bouchy, F., Brasch, V., et al. 2021a, in *European Planetary Science Congress*, EPSC2021–323
- Schmidt, T. M., Molaro, P., Murphy, M. T., et al. 2021b, *A&A*, 646, A144
- Schwarz, G. 1978, *Annals of Statistics*, 6, 461
- Udem, T., Holzwarth, R., & Hänsch, T. W. 2002, *Nature*, 416, 233
- Uzan, J.-P. 2011, *Living Reviews in Relativity*, 14, 2
- 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., Carswell, R. F., & Milaković, D. 2021, *MNRAS*, 501, 2268
- Webb, J. K., Lee, C.-C., & Milaković, D. 2022, *Universe*, 8, 266
- Webb, J. K., Murphy, M. T., Flambaum, V. V., et al. 2001, *Phys. Rev. Lett.*, 87, 091301
- Whitmore, J. B. & Murphy, M. T. 2015, *MNRAS*, 447, 446
- Wilczynska, M. R., Webb, J. K., King, J. A., et al. 2015, *MNRAS*, 454, 3082
- Wilken, T., Curto, G. L., Probst, R. A., et al. 2012, *Nature*, 485, 611
- Wilken, T., Lovis, C., Manescau, A., et al. 2010, *MNRAS*, 405, L16
- Zhao, F., Lo Curto, G., Pasquini, L., et al. 2021, *A&A*, 645, A23