Avoiding Bias in Measurements of Fundamental Constants from High Resolution Quasar Spectra †

John K. Webb [1,*] , **Chung-Chi Lee** [1] **and Dinko Milaković** [2,3,4]

1   Clare Hall, University of Cambridge, Herschel Rd., Cambridge CB3 9AL, UK
2   Institute for Fundamental Physics of the Universe, Via Beirut 2, 34151 Grignano, Italy
3   National Institute for Astrophysics—Osservatorio Astronomico di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
4   Italian National Institute for Nuclear Physics (INFN)—Sezione di Trieste, Via Bonomea 265, 34136 Trieste, Italy
*   Correspondence: jkw.phys@gmail.com
†   This article is a contribution for the AlteCosmoFun'21 conference volume, Szczecin, Poland, 6–10 September 2022, organised by the Szczecin Cosmology Group.

Abstract: Recent advances in spectroscopic instrumentation and calibration methods dramatically improve the quality of quasar spectra. Supercomputer calculations show that, at high spectral resolution, procedures used in some previous analyses of spacetime variations of fundamental constants are likely to generate spurious measurements, biased systematically towards a null result. Developments in analysis methods are also summarised and a prescription given for the analysis of new and forthcoming data.

Keywords: cosmological parameters; quasar absorption systems; methods: observational; techniques: spectroscopic

1. Introduction

High resolution spectra of distant quasars reveal numerous narrow absorption lines caused by gaseous components of galaxies intersecting the Earth–quasar sightline. The large number of atomic species and transitions detected allow precise measurements of the fine structure constant $\alpha_{SI} = e^2/4\pi\epsilon_0\hbar c$ over cosmological distances, where e is the electron charge, ϵ_0 the vacuum permittivity, $\hbar$ the reduced Planck constant, and c the speed of light. The dimensionless α is the ratio of the speed of an electron in the lowest energy orbit of the Bohr-Sommerfeld atom to the speed of light, and hence connects quantum mechanics (through $\hbar$) with electromagnetism (through the remaining quantities).

The 1999 invention of the *Many Multiplet method* (MM) [1,2] created an order of magnitude precision gain over previous methods in searches for spacetime variations of α. The previously used method had been the *Alkali Doublet method* (AD), in which measurements were made relative to the same ground state, by-passing the invaluable sensitivity of the ground state to any change in α. Further, the excited levels in the $^2P_{3/2}$–$^2P_{1/2}$ AD structure itself generally have lower sensitivity compared to the commonly observed singly ionised multiplets of Fe, Zn, Cr, and others. In contrast, the MM method takes into account ground-state shifts. Also, excited state relativistic corrections in multiplets can be large; s–p and s–d transitions for example may even be of the opposite sign. Any real change in α therefore generate a unique pattern of observed wavelength shifts that is not degenerate with a simple cosmological redshift. The MM method thus produces sensitive results when applied to multiplets of the same atomic species, or to species having widely differing atomic masses.

New and forthcoming scientific facilities will intensify searches for spacetime variations of fundamental constants. In particular, the Echelle SPectrograph for Rocky Exoplanets and Stable Spectroscopic Observations (ESPRESSO) on the European Southern Observatory's Very Large Telescope (VLT) [3] and the ArmazoNes high Dispersion Echelle

Spectrograph (ANDES, formally named HIRES) on the forthcoming Extremely Large Telescope (ELT) e.g., [4,5] will play important roles in this field of research. As the data quality and quantity increase, it becomes all-important to ensure that analysis techniques produce fully unbiased and optimal estimates. Here we summarise recent methodological advances facilitating these things and to scrutinise assumptions made and procedures used in previously published analyses that can produce bias. The remainder of this paper is composed as follows: Sections 2 and 3 summarise several recent advances in the varying α field. Section 4 presents some new calculations showing that "blinding" methods, as used in some published measurements, generate measurement bias. Section 5 gives a précis of the do's and don'ts when analysing high quality absorption spectra.

2. Wavelength Calibration

Measuring α requires the wavelength scale of the astronomical spectrum to be established with high fidelity. For example, a change of $\Delta\alpha/\alpha \equiv (\alpha_z - \alpha_0)/\alpha_0 = 1 \times 10^{-6}$ gives rise to a relative shift in the wavelength of Fe II 2383 of approximately $\Delta\lambda/\lambda = 1 \times 10^{-7}$. For echelle spectrographs such as UVES on the VLT, the standard wavelength reference is established by imaging the spectrum of a ThAr arc lamp. A comparison between ThAr wavelength calibrated data and the solar spectrum (as first done by [6]), revealed the presence of long-range wavelength scale distortions with amplitudes as large as $\Delta\lambda/\lambda = 1 \times 10^{-5}$. It was initially thought that distortions this large could potentially spoil fundamental constant measurements [7,8], although subsequent analyses showed that such distortions can be modelled and associated uncertainties on $\Delta\alpha/\alpha$ allowed for (see [9] and Appendix B9 of [10]).

The new generation of astronomical spectrographs aims to avoid the long-range distortions seen in some ThAr calibrated spectra altogether. Laser Frequency Comb technology (LFC, [11–13]) or Fabry-Perót etalons (FP) combined with ThAr can provide wavelength calibration with accuracy of the order $\Delta\lambda/\lambda = 1 \times 10^{-8}$ [14–16]. Both of these advanced calibration sources are installed on the High Accuracy Radial-velocity Planet Searcher (HARPS, [17]) and ESPRESSO. One current difficulty is that none of the currently available astronomical LFCs provide wavelength calibration below 5000 Å. Wavelengths below this cut-off are currently calibrated by simultaneously imaging both ThAr and FP spectra. The ThAr + FP combination provides lower accuracy than LFCs. Since the spectral region below 5000 Å is generally very important for $\Delta\alpha/\alpha$ measurements and since, this important problem is yet to be solved, raising significant problems not only for varying constant measurements but also for redshift drift projects.

Another important concern that must be addressed in future high-precision spectroscopy is that using an LFC or ThAr+FP may not be sufficient to remove all systematic effects related to wavelength calibration. A comparison of two independent LFCs used simultaneously on HARPS revealed an unexpected offset in the zero-points of their wavelength calibrations of $\Delta\lambda/\lambda = 1.5 \times 10^{-9}$ [14,15]. Although this is a very small effect, combining observations calibrated using either two different LFCs or a single LFC that has been modified in some way between two observations should be carefully performed.

3. Absorption Profile Modelling

3.1. VPFIT and AI-VPFIT

VPFIT [18] is a non-linear least squares code for modelling high resolution absorption spectra that has been developed over a number of years [18,19] and it forms the core of our procedures. The theoretical background on which VPFIT is based, plus some recent enhancements, are described in [10,20]. Throughout VPFIT's development, considerable effort has gone into ensuring high precision internal calculations. It has a comprehensive online user guide, updated frequently [21]. The ever-increasing quality of high resolution spectroscopic data requires extremely careful treatment of every aspect of profile calculation and fitting. Some of the studies carried out in this context are:

1. The Voigt profile 2-level atom approximation is good enough, even at very high spectral signal to noise, for non-damped column densities i.e., $\log N \leq 20$ [22].
2. Voigt function $H(a, u)$ look-up tables must be sufficiently high resolution in the relevant parameter (u) to render non-linear effects negligible. Interpolation within those tables must also be sufficiently precise [10].
3. It is most important to allow for non-linearity in the Voigt profile shape by computing model Voigt profiles in extremely fine bins [19].
4. As far as possible, any blends/interlopers must be allowed for, whether they arise in identified species or not. Failure to do can significantly increase the measurement error and bias individual measurements [23].
5. Fitting region selection is important. If line wings/continuum regions are truncated, best-fit models for $\Delta\alpha/\alpha$ exhibit an unnecessarily large scatter [24].
6. Voigt function derivatives must be accurate [10,20]. The Hessian is derived from the derivatives of $H(a, u)$, the inverse of which provides parameter uncertainties. The Hessian, of course, determines parameter search directions.
7. How to select a "final" absorption system model, given non-uniqueness and alternative information criteria? [25].
8. Absorption line broadening: models must include temperature as a free parameter and should not be assumed to be *turbulent* [26,27].
9. Using the correct instrumental profile for model calculations is important [14].
10. All spectral regions used in the measurement must be carefully checked for potential contaminating atmospheric features [28].

Using VPFIT involves human decision making and the final results obtained can depend on that human input [25]. Whilst this has negligible consequences for many applications, measurements of fundamental constants at high redshift push the limits of the data and are more susceptible to small systematic errors or biases than other less challenging measurements. Measurements of $\Delta\alpha/\alpha$ thus motivate full automation in which all human input is avoided and all potential bias removed. Artificial Intelligence methods were first applied to this problem in [29,30] (GVPFIT) and significantly extended in [23] (AI-VPFIT). The advantages in avoiding human decision making are achieving objectivity, reproducibility, and the ability to explore multiple models to the same absorption system. These advantages turn out to be crucial for measurements of fundamental constants. The calculations reported in the following sections make use of both VPFIT and AI-VPFIT.

3.2. Information Criterion or χ_ν^2 to Select Models?

In selecting a single best-fit model for an absorption complex, many previous measurements have made use of a simple normalised χ^2 approach, accepting a model once χ_ν^2 is "sufficiently" close to unity,

$$\chi_\nu^2 = \frac{1}{\nu} \sum_{i=1}^{n_d} \left(\frac{d_i - f_i}{\sigma_i^2} \right)^2 \tag{1}$$

where d_i is the spectral data array, f_i is the model, σ_i is the spectral error array, n_d is the number of data points, and the number of degrees of freedom $\nu = n_d - n_p$, where n_p is the total number of free parameters in the model. Alternatively, one can use an *information criterion* (IC) to select models, the general form of which is

$$\text{IC} = \chi^2 + \mathcal{P}(n_p, n_d) \tag{2}$$

where $\chi^2 = \nu \chi_\nu^2$ and $\mathcal{P}(n_p, n_d)$ is a penalty factor that increases with increasing number of model parameters. An IC applies a "principle of parsimony", balancing parameter variance with model bias. The application of ICs in astrophysics has been discussed by [31,32] and a comprehensive treatise is given in [33]. Problems associated with noise characteristics in

calculating ICs are discussed in [34]. Employing an IC allows an optimal number of model parameters to be identified in an objective and reproducible way [35].

Using an asymptotic χ_ν^2 as a means of deciding how many model parameters to use requires the user to choose a maximum acceptable value for χ_ν^2. This in itself is not particularly disadvantageous because one can formalise the problem through the relationship between χ_ν^2 and its probability distribution, hence defining an acceptance probability rather than a numerical value of χ_ν^2 for model selection. However, in practice it is generally difficult to obtain σ_i accurately, so χ^2 itself is only approximate. The difficulties in accurately estimating the spectral error array are well known and have been discussed in the section titled *Modifying the error arrays* in the RDGEN user guide [18].

For the reasons outlined above, in the context of $\Delta\alpha/\alpha$ at least, it is preferable to use an IC for model selection rather than χ_ν^2. However, using an IC raises a question: what is the optimal form of the penalty term $\mathcal{P}(n_p, n_d)$ in Equation (2)? This point has been studied in [35], where three ICs are compared : the corrected Akaike Information Criterion AICc, the Bayesian Information Criterion BIC, and the Spectral Information Criterion SpIC. The latter is a new IC designed specifically for spectroscopy. All three ICs (there are others) perform in slightly different ways; AICc tends to over-fit the data, allowing too many free parameters, whilst the converse is true for BIC. SpIC appears to fall in the Goldilocks zone.

3.3. Contributions to the $\Delta\alpha/\alpha$ Error Budget

Fully understanding all potential contributions to the error budget, random and systematic, is of course a crucial aspect of assessing the reality of any deviation of $\Delta\alpha/\alpha$ from zero. Uncertainties associated with varying α measurements have been discussed in numerous papers, e.g., [36,37]. Since our understanding of uncertainties has significantly improved since those studies, we list here possible sources of error and their attribution:

Inherent uncertainties:

1. Statistical error i.e., VPFIT covariance matrix error.
2. If turbulent broadening is used to model the system and if the true intrinsic broadening is *compound* (see Equation (3)), a systematic error is introduced.
3. Absorption system model non-uniqueness error.
4. Continuum estimate error.

Errors from astrophysical factors:

5. Isotopic relative abundances.

Errors associated with theoretical and experimental uncertainties:

6. Q coefficient uncertainties (these bias towards $\Delta\alpha/\alpha = 0$).
7. Oscillator strength uncertainties.
8. Laboratory wavelength uncertainties.

Errors associated with data extraction or instrumental factors:

9. Wavelength calibration error (for pre-LFC/FP data).
10. Bad pixels.
11. Flat-fielding errors.
12. Weak cosmic rays removal.
13. Significant point-spread function variations across the detector.

3.4. Future Measurements Require Monte Carlo AI

The recent application of AI methods (AI-VPFIT) provides full automation of modelling quasar absorption systems [23,30]. Automation has allowed us to explore in detail how final best-fit models depend on the construction sequence and hence the extent to which model non-uniqueness contributes to the total $\Delta\alpha/\alpha$ error budget. Preliminary studies of this sort have been made recently and so far indicate that indeed $\Delta\alpha/\alpha$ measurements are impacted by the sequence in which models are developed. The inference is that, for any particular absorption system, it is difficult to properly assess the overall uncertainty on

$\Delta\alpha/\alpha$ *unless* multiple models are produced, each one constructed differently (emulating the different approaches that would be taken by different human modellers). The contribution of non-uniqueness to the overall $\Delta\alpha/\alpha$ uncertainty budget can only be determined on a case by case basis because we now know that the degree of non-uniqueness differs from one absorption system to another [25]. In other words, it is necessary to form many models of each particular absorption system in order to quantify the intrinsic non-uniqueness behaviour and hence determine the overall uncertainty on $\Delta\alpha/\alpha$.

4. Spectral Simulations and Distortion-Blinding

The methods used to measure $\Delta\alpha/\alpha$ must be entirely free of any bias. To that end, a "blinding" method has been used in several published measurements [38–41], employing the following steps. During the initial data reduction stages, long-range and intra-order distortions of the wavelength scale are applied to individual exposures. The individual distortions are designed to leave a non-zero $\Delta\alpha/\alpha$ in the final co-added spectrum. VPFIT is then used on that distorted spectrum, fixing $\Delta\alpha/\alpha$ to be zero. Once the final model has been obtained in this way, one final tweak of the model is carried out by fitting a non-distorted version of the spectrum, allowing the existing model parameters plus $\Delta\alpha/\alpha$ to vary freely. However, fixing $\Delta\alpha/\alpha = 0$ corresponds to requiring that all rest-frame wavelengths involved in the initial absorption system modelling are precisely those measured in terrestrial laboratories. If the true value of $\Delta\alpha/\alpha$ is zero, the method just described should indeed be unbiased. However, if the true $\Delta\alpha/\alpha$ is *not* zero (and it is explicitly non-zero in the "blinded" data of [38–41]), forcing $\Delta\alpha/\alpha$ to be initially zero necessarily produces a flawed model. The final step of "switching on" $\Delta\alpha/\alpha$ as a free parameter may or may not subsequently be able to correct that flawed model. If it does not, the result is a systematically biased measurement. Notwithstanding the previous comments, it should be noted that the spectral resolution of the simulations in the analyses in this paper are far higher than in e.g., [38–40]. At lower spectral resolution, fewer components are detected. This could mean that the level of bias is diluted. Further calculations are needed to answer this point. Irrespective of this, what is instead needed is a method that is fully unbiased, *no matter what* the true value of $\Delta\alpha/\alpha$ is. We next describe some preliminary simple spectral simulations to illustrate the concerns just expressed. To distinguish the sort of blinding described above from other potential blinding methods, we will refer to the method described above as "distortion-blinding".

4.1. Preliminary Illustration of How Distortion-Blinding + Turbulent Line Broadening Creates Bias

A simulated spectrum of a single component absorption system is created, with seven transitions: Mg II 2796, 2803, Fe II 2344, 2374, 2383, 2586, 2600 Å. A spectral resolution of 2.10 km/s FWHM is used for the three transitions having $\lambda_{obs} < 5120$ Å and 2.05 km/s FWHM for the four transitions having $\lambda_{obs} > 5540$ Å. The pixel size and signal to noise per pixel were 0.4 km/s and 75. The spectral parameters correspond to those of the ESPRESSO spectrum in [41]. The absorption line parameter values used are similar to one component from the $z_{abs} = 1.15$ system towards the quasar HE0515–4414. The upper panel of Table 1 (labelled "Generating model") shows the actual absorption line parameters used to generate the simulated spectrum. For the purposes of this preliminary illustration, we use an extreme value of $\Delta\alpha/\alpha = 10^{-4}$, so that line shifts caused by the non-zero $\Delta\alpha/\alpha$ are easily visible in the plotted data. A subset of the simulated spectral lines is illustrated in Figure 1. As Table 1 shows, the simulated spectrum is generated using compound broadening i.e., line broadening comes from both turbulent and thermal contributions,

$$b_{obs}^2 = b_{\text{turb}}^2 + \frac{2kT}{m} \qquad (3)$$

Having created the data, for the purposes of this illustration we first fit all lines simultaneously with VPFIT, using compound line broadening. $\Delta\alpha/\alpha$ is a free parameter throughout. No distortion-blinding is used. The second panel of Table 1 (labelled "Fitting

(compound)") illustrates the measured parameter values and their uncertainties, showing, as expected, that only a single absorption component is required and that all parameters returned are consistent with the input values.
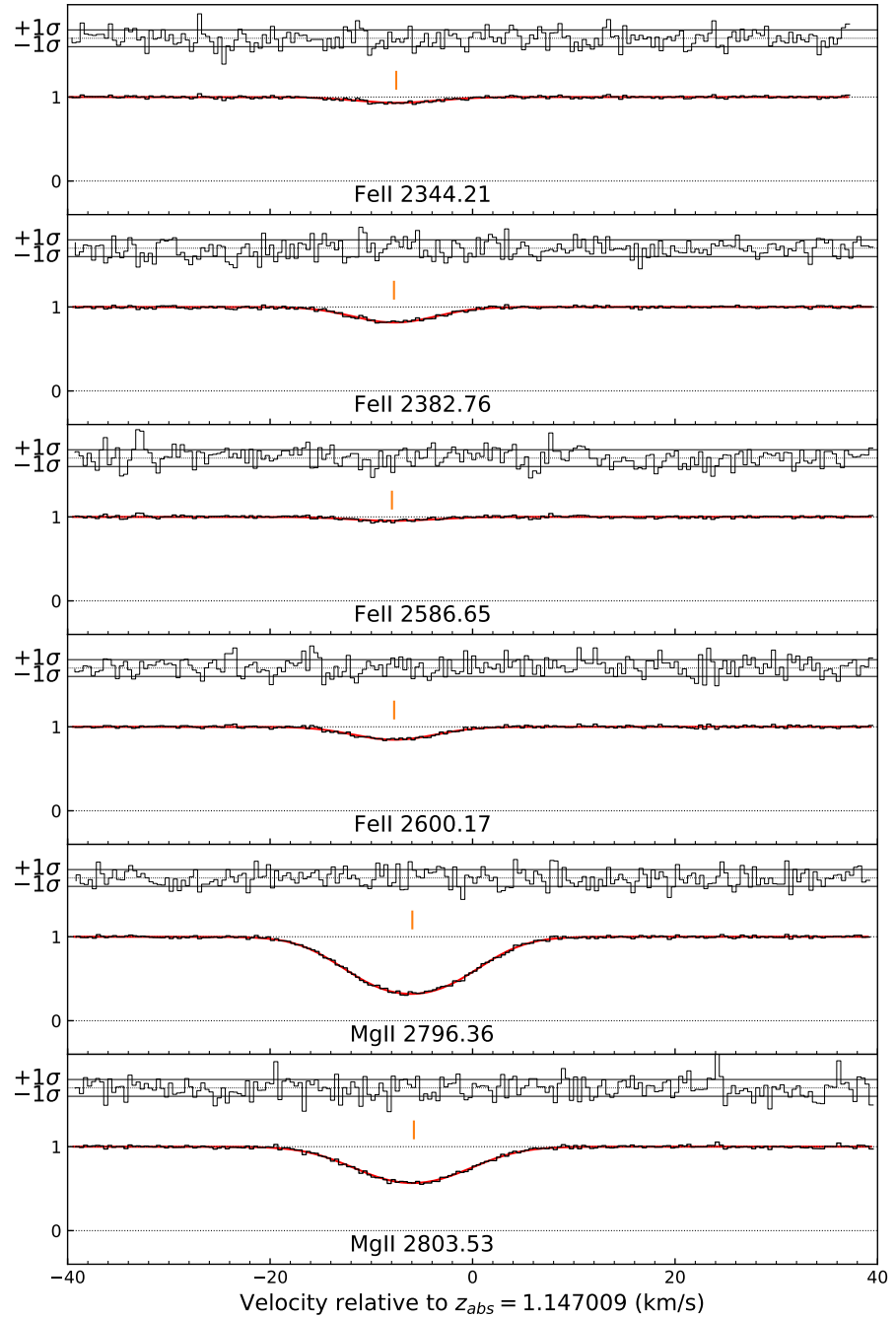
Figure 1. The data (black histogram) in this plot corresponds to the upper panel labelled "Generating model" in Table 1. The model (continuous red curve) corresponds to the middle panel labelled "Fitting (compound)". See Section 4.1.

Table 1. Results obtained using VPFIT on a single component simulation. See Section 4.1 for details. $v = (z - z_0)c/(1 + z_0)$, where $z_0 = 1.1469691$ and c is the speed of light. In the lower panel, one of the Fe II components fell below the VPFIT detection limit, indicated in the log N column by "-". The negative velocities (v) for the turbulent case are a consequence of forcing $\Delta\alpha/\alpha = 0$ (the true value is non-zero).

	Species	log N	b_{turb} (km/s)	T (10^4 K)	v (km/s)	$\Delta\alpha/\alpha$ (10^{-5})
Generating model	Mg II	12.500	3.87	5.31	0.000	10.0
	Fe II	12.000				
Fitting (compound)	Mg II	12.499 (0.002)	3.70	5.53	0.000	10.1
	Fe II	12.000 (0.007)	(0.27)	(0.34)	(0.028)	(0.4)
Fitting (turbulent)	Mg II	12.493 (0.017)	7.02	-	−0.279	
	Fe II	-	(0.05)		(0.084)	-
	Mg II	10.682 (1.099)	5.44	-	−2.152	
	Fe II	12.001 (0.007)	(0.10)		(0.070)	

We now reach the point of this illustration. The seven transitions are again fitted simultaneously using VPFIT, but this time with two important differences: the data are fitted using a fixed $\Delta\alpha/\alpha = 0$ and using turbulent line broadening, i.e., taking the limiting case of $T = 0$. The numerical results are given in the third (lower) panel of Table 1 (labelled "Fitting (turbulent)") and the resulting model is illustrated in Figure 2. This fit now requires *two* absorption components (one of the Fe II components is below the column density threshold so tick marks are not shown). Figure 2 and Table 1 reveal that adopting a turbulent model creates the necessity for an additional (fake) velocity component *and* because of the distortion-blinding, that fake velocity component is required to correspond perfectly with (redshifted) terrestrial laboratory wavelengths.

The purpose of this simple simulation is purely to demonstrate how easy it is to derive a strongly biased result if distortion-blinding + turbulent broadening are used; the normalised residuals are plotted above each transition in Figure 2. The residuals show that the turbulent distortion-blinded fit forcing $\Delta\alpha/\alpha = 0$ provides a good fit, even though it is far from zero in the data. The normalised χ^2 for both fits quantify that: the non-blinded compound model produces an overall fit of $\chi^2_\nu = 1.0250$ for 1394 degrees of freedom and AICc = 1440.882. The distortion-blinded $\Delta\alpha/\alpha = 0$ turbulent model produces $\chi^2_\nu = 1.0291$ for 1393 degrees of freedom and AICc = 1447.581. Both normalised χ^2 values are acceptable. In Section 4.2 we examine this problem in more detail.
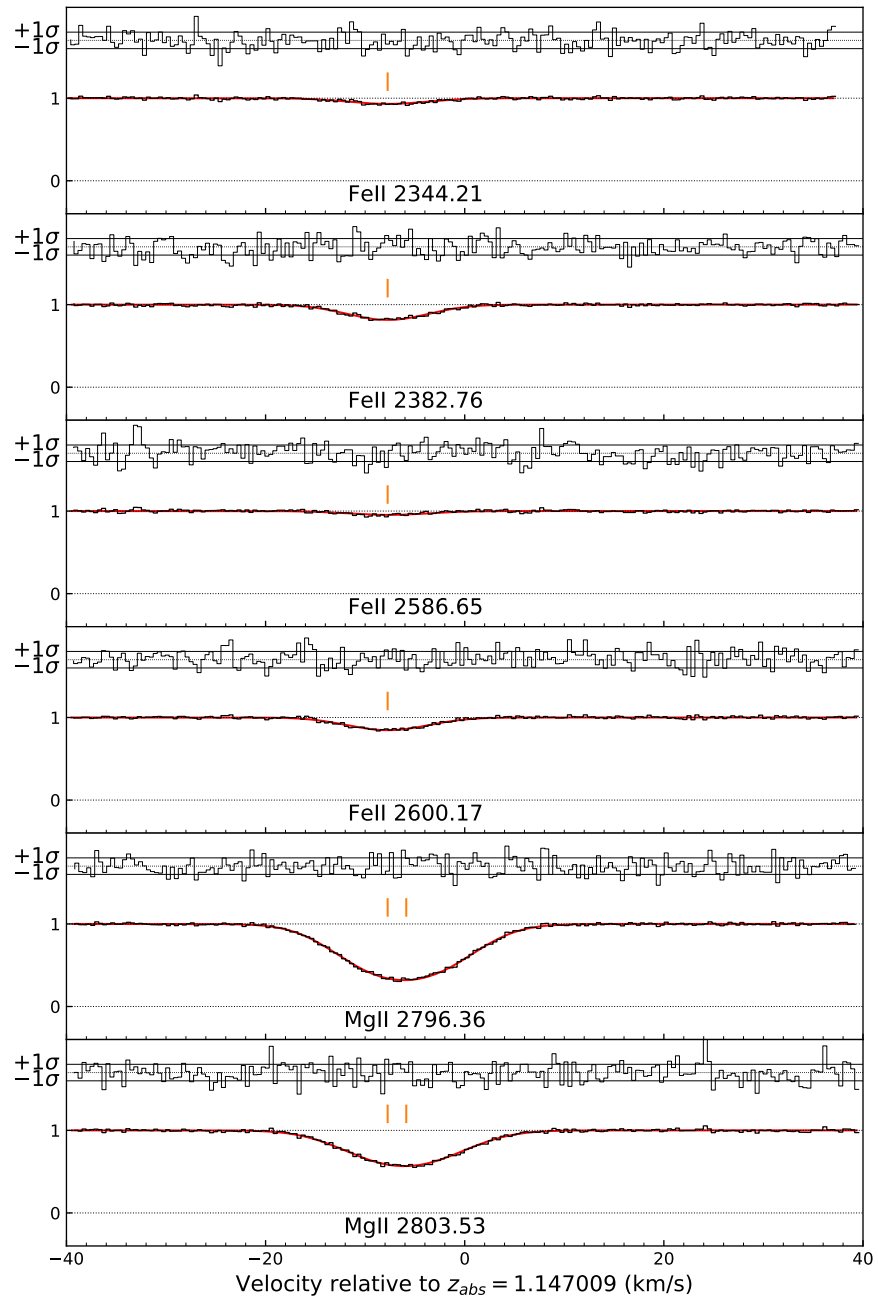
Figure 2. The data (black histogram) in this plot corresponds to the upper panel labelled "Generating model" in Table 1. The model (continuous red curve) corresponds to the middle panel labelled "Fitting (turbulent)". See Section 4.1.

4.2. Detailed Calculations Using AI-VPFIT

Section 4.1 illustrates that distortion-blinding + turbulent modelling has the capacity to significantly bias $\Delta\alpha/\alpha$ measurements towards zero. In this section we show that distortion-blinding + turbulent modelling does not merely have the capacity to bias, but that it inevitably does so.

We first generate a simple synthetic spectrum. The absorption line parameters used were extracted from measurements (by us) of a real absorption system at $z_{abs} = 1.15$ towards the publicly available ESPRESSO spectrum of the quasar HE0515–4414 [41]. A small section

of the system was used, rather than the entire system, to make calculating time shorter and because a small section serves the required purpose. The spectral characteristics are the same as the simulated spectrum described in Section 4.1. The absorption line parameters used to generate the synthetic spectrum are given in Table 2. The synthetic spectrum was created using $\Delta\alpha/\alpha = 8.08 \times 10^{-6}$, a value measured by us, without applying any kind of blinding, using the full absorption system (i.e., not a small section). The same seven transitions were generated as above.

Table 2. Absorption line parameters used to generate the spectrum illustrated in Figure 3. See Section 4.2.

log N(Mg II)	log N(Fe II)	b_{turb} (km/s)	T (10^4 K)	z_{abs}
12.66	12.20	1.49	1.64	1.1469691
12.61	12.16	7.19	4.06	1.1469952

As Table 2 shows, the line broadening used to generate the absorption simulation is compound i.e., each redshift has its own unique thermal and turbulent contribution. Many previous detailed analyses show that individual quasar absorption components exhibit compound and not turbulent broadening e.g., [27], so our synthetic spectrum emulates real data in this respect. The next stage of our analysis is to model the synthetic data using the same procedures that have been applied to real data in e.g., [41]. Models are fitted using turbulent line broadening and $\Delta\alpha/\alpha$ is fixed to zero i.e., we carry out the same distortion-blinding approach used in some previous works. We then apply AI-VPFIT [23] multiple times, each time deriving an independently constructed best fit. AICc is used to select the best-fit model during each AI-VPFIT calculation. In each case, after the best-fit model is established, one final iteration using VPFIT is done, adding $\Delta\alpha/\alpha$ as an additional free parameter.

Figure 4 gives histograms for four sets of 100 AI-VPFIT calculations (one for each combination of turbulent, compound, AICc, and SpIC models), revealing remarkable effects. The panel on the left demonstrates that using turbulent broadening with AICc generates a distribution that is clearly non-Gaussian, with severe bias towards $\Delta\alpha/\alpha = 0$. No bias is seen for AICc/compound and the distribution looks well-behaved. The panel on the right shows results obtained using SpIC. In this case, significant bias is again seen if turbulent broadening is used, but not for compound broadening. The SpIC/compound $\Delta\alpha/\alpha$ results are more reliably determined than those for AICc/compound (compare the spreads of the compound points for AICc and SpIC results). The bin size in the histograms above is 1.25×10^{-6} and the 1σ statistical uncertainty for a typical compound broadened model with 2 components is 2.13×10^{-6}. Figure 4 shows that the combination of distortion-blinding + turbulent broadening is particularly damaging to the final result.

Figure 5 is a visual breakdown of the number of redshift components needed to fit the spectral lines illustrated in Figure 3, for the four different combinations of AICc, SpIC, turbulent, and compound. Light blue bars correspond to metals (i.e., Mg II and Fe II in this case). Remembering that the true number of components is 2, we see that the only combination with 100% success is SpIC/compound, with AICc/compound almost as good. The turbulent results badly overfit the data, both AICc and SpIC giving most probable numbers of 4. The yellow bars correspond to interlopers i.e., additional absorption components required by AI-VPFIT to achieve an acceptable fit, but which were not identified as metals. No interlopers were present in the simulated data being fitted so these are weak features caused by chance correlated noise patterns that have no detectable impact (in terms of systematic bias) on the $\Delta\alpha/\alpha$ estimate.
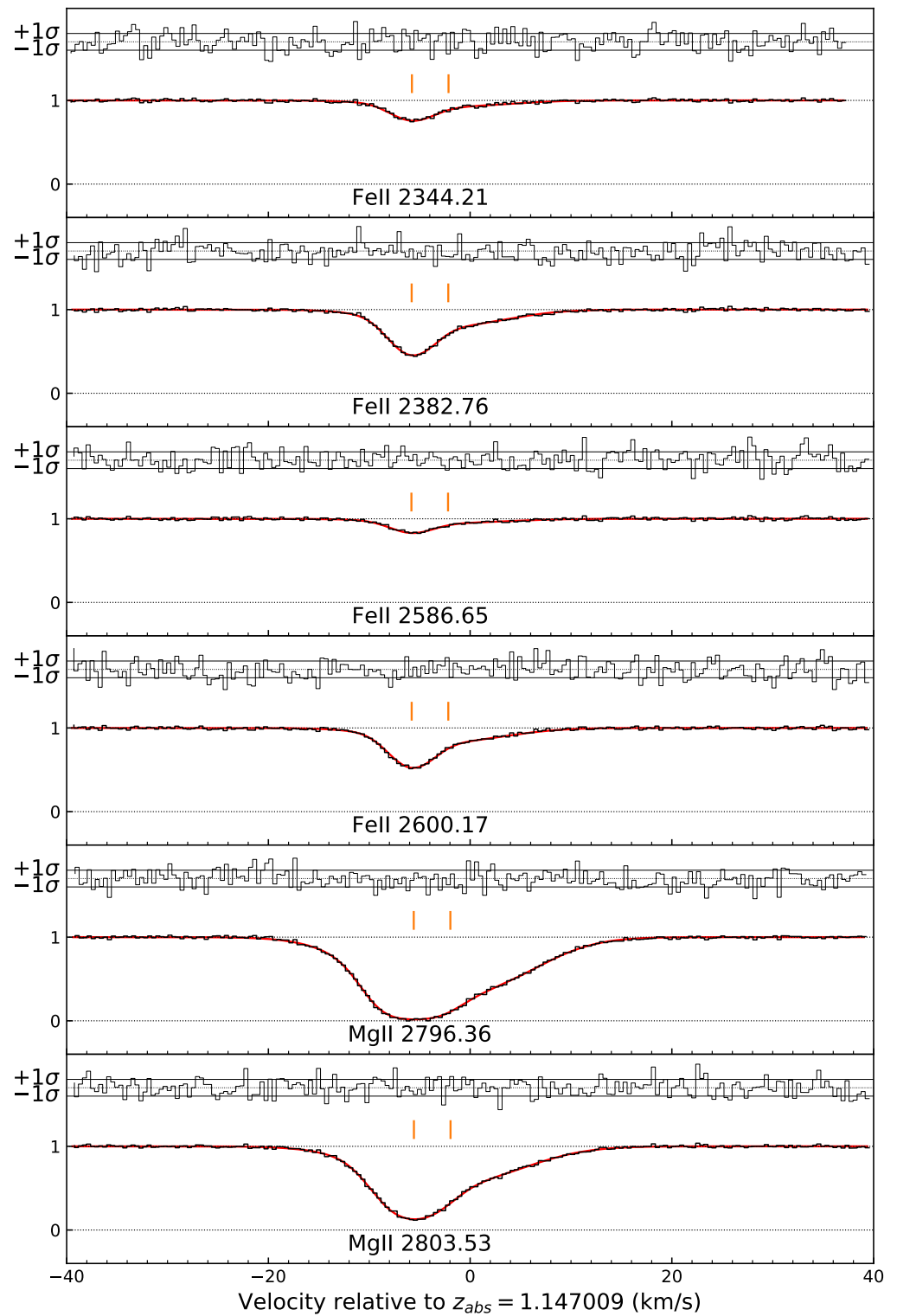
Figure 3. Synthetic spectrum fitted with 4 sets of 100 AI-VPFIT models, as described in Section 4.2. Only 6 transitions are shown (the excluded transition, Fe II 2374 Å, is very weak). The parameter details are given in Table 2. The spectral characteristics are described in Section 4.1. The model (continuous red line) is a compound broadened model, randomly selected from our set of 100 fits. Vertical tick marks illustrate the best-fit component positions.
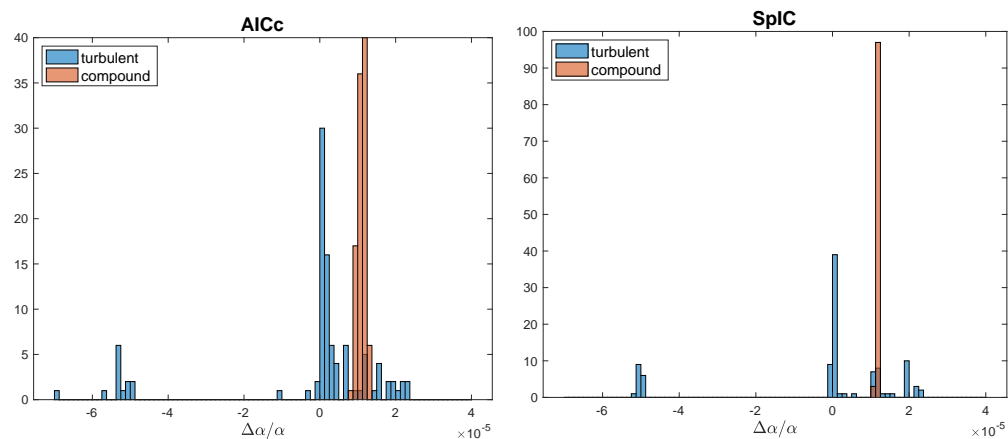
Figure 4. Results from calculations simulating the impact of "blinding" on measurements of $\Delta\alpha/\alpha$. The red bars show the results from fitting compound broadened models without using distortion blinding. Those results are correct and unbiased. The blue bars show the results from fitting turbulent broadened models and using distortion blinding. The results are stongly biased and the correct $\Delta\alpha/\alpha$ is not recovered. See Section 4.2.
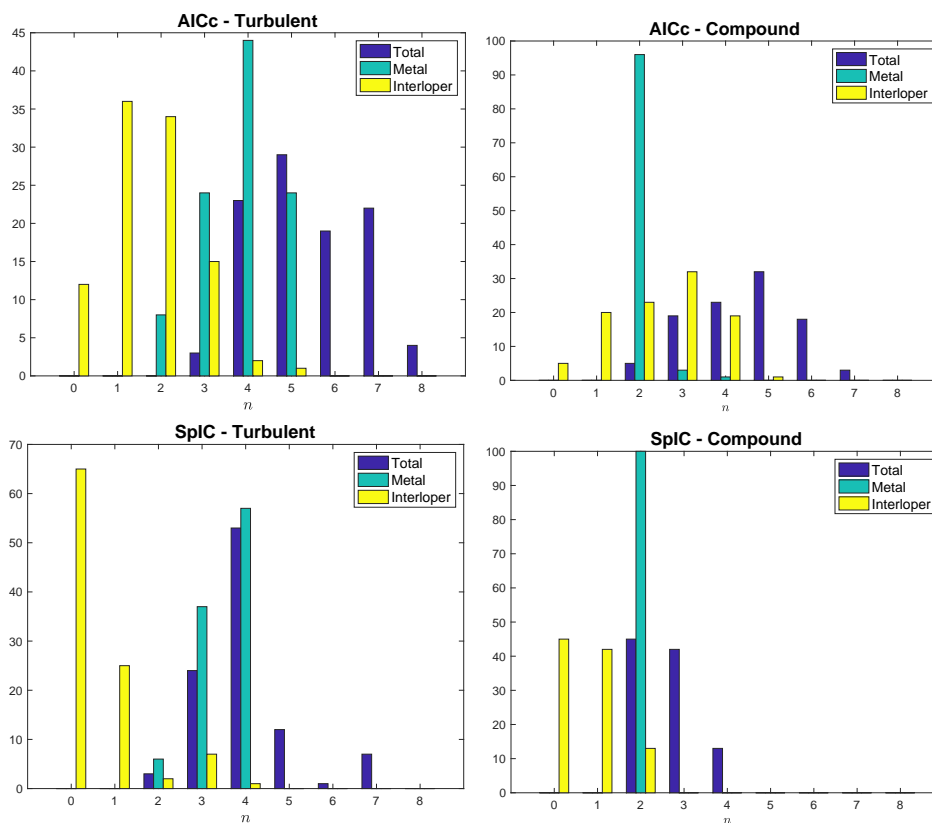
Figure 5. Number of absorption components for the results illustrated in Figure 4. The spectrum being modelled comprised 2 components. **Top left**: AICc/turbulent results in severely over-fitted data (light blue). **Lower left**: SpIC/turbulent also over-fits the data but not quite as severely as AICc/turbulent. The turbulent model results have been obtained using distortion blinding, as described in Section 4.2. **Top right**: AICc/compound produces very slight over-fitting. **Lower right**: SpIC/compound produces no over-fitting. Distortion blinding has *not* been used for the compound model results. The yellow bars indicate additional interloper (i.e., components not identified as heavy element) components automatically included in the fits by AI-VPFIT. These are purely a consequence of chance noise correlations in the input spectrum and appear to have no impact on the final value of $\Delta\alpha/\alpha$ measured.

Figure 6 plots the 100 individual AI-VPFIT $\Delta\alpha/\alpha$ measurements. The distortion-blinded turbulent points (blue) are clearly heavily biased towards zero. Further, the majority of those points have far larger error bars (because of the overfitting and hence artificially increased blending) than a small subset closer to the correct value of $\Delta\alpha/\alpha = 8.08 \times 10^{-6}$. The compound points (not distortion-blinded, red) have substantially smaller error bars (no overfitting) and do not suffer bias. Interestingly, the blue points clearly exhibit model non-uniqueness whereas the red ones do not.
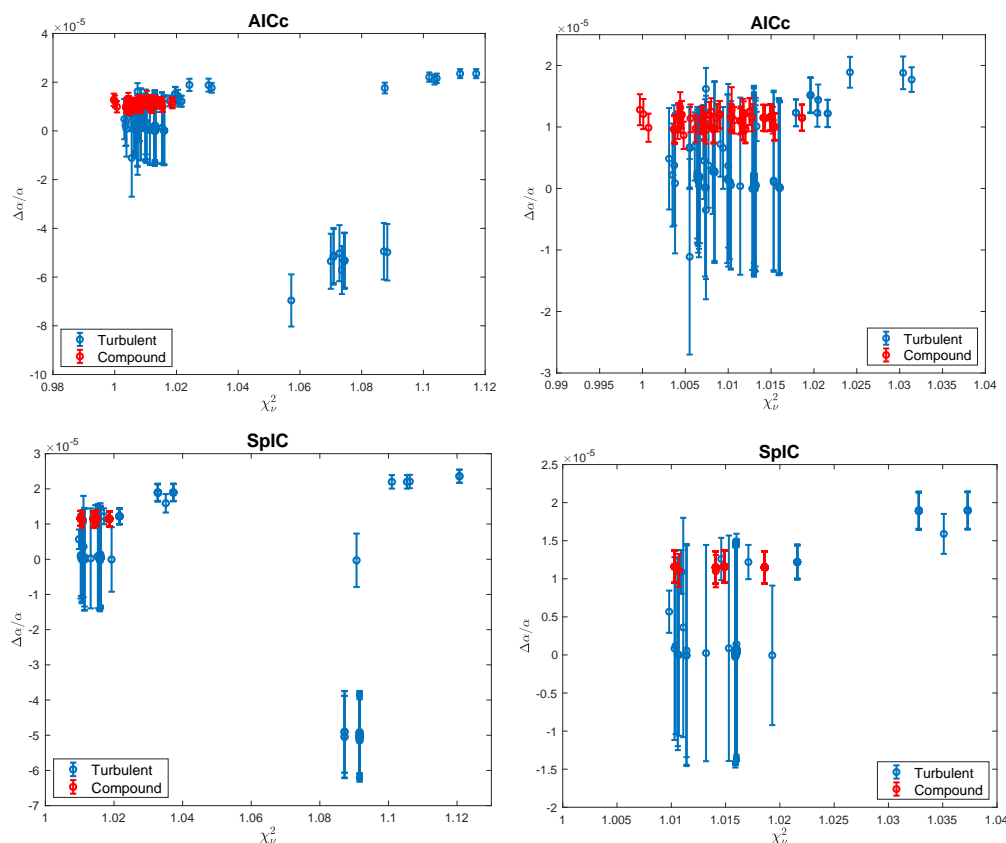
Figure 6. Plots of $\Delta\alpha/\alpha$ vs. χ_ν^2 for the 4 combinations of AICc, SpIC, turbulent, and compound. Distortion-blinding was applied to the turbulent (blue) points but not the compound (red) points. See Section 4.2. The two right hand panels are zoom-ins of the two left hand panels.

4.3. A Retrospective on Distortion-Blinding

Distortion-blinding was an attempt at taking a conservative approach to the subject. It is only because we have developed AI Monte-Carlo methods that we can now identify the adverse impact of interactive modellers (each probably taking a slightly different approach) using distortion-blinding + turbulent broadening. AI-VPFIT shows us how to avoid bias in measurements using high quality data from facilities like ESPRESSO on the VLT and, in the future, HIRES on the ELT. In fact, in the absence of distortion-blinding, any reasonable analysis is effectively inherently blind anyway. If the fitting is done "manually" (i.e., using VPFIT), the process contains so many steps that it is inconceivable for a human to subconsciously construct a biased model unless the final measured value of $\Delta\alpha/\alpha$ is influenced directly by iteratively adjusting the model or adding new absorption components until a "preferred" value of $\Delta\alpha/\alpha$ is obtained. Given that analyses involving distortion-blinding have been carried out by different people at different times, the degree to which this has or has not happened in previously published measurements cannot be recovered.

Looking to future measurements, no tampering with the wavelength scale needs to be or should be done. If a fit is done using VPFIT, a far simpler (and *unbiased*) method of

"blinding" would be to switch off all output to the user containing any information about $\Delta\alpha/\alpha$ during the entire model construction process. $\Delta\alpha/\alpha$ must be a free parameter from the outset i.e., as soon as two or more transitions or species with differing sensitivities to α are incorporated—see [23] for a discussion on this important point. Switching off $\Delta\alpha/\alpha$ output naturally renders any conscious or subconscious steering by the user impossible.

Analyses using the AI-VPFIT method, as described in [23], are automatically blinded. There is no interactive input during the modelling procedure.

5. Discussion

The calculations described in this work and accompanying discussions allow us to define some general requirements for spacetime measurements of varying fundamental constants:

1. When modelling/solving for $\Delta\alpha/\alpha$, if fitting multiple species simultaneously, turbulent broadening should not be used. Instead only compound broadening should be used. If compound broadening is problematic because T is not sufficiently well constrained, either the measurement should be discarded, or possibly one could adopt a representative T as a fixed parameter, although in this case the uncertainty on $\Delta\alpha/\alpha$ may be artificially lowered. We have not investigated such an approach in this paper.
2. Distortion-blinding biases results and should be avoided. Deliberately distorting a spectrum and then solving for its velocity structure with $\Delta\alpha/\alpha$ forced to be zero is most likely to create a result that is biased towards $\Delta\alpha/\alpha = 0$. Previous $\Delta\alpha/\alpha$ measurements carried out in this way should be repeated.
3. Some absorption systems suffer from model non-uniqueness. Therefore it is desirable to model each absorption system multiple times to quantify this.
4. In obtaining the observational data, wavelength coverage should be done using LFCs or FPs. If observations from different observing runs at different epochs are to be combined/jointly analysed, there should be calibration redundancy i.e., ideally calibrations always done with two LFCs or FPs. Accurate calibration reaching as blue as the atmospheric cutoff is essential to pick up lower rest-wavelength transitions.

An aside on the above is an interesting consequence of turbulent broadening with important implications for abundance measurements. Turbulent modelling generates spurious absorption components and these appear with meaningless relative abundances. The result is that the relative abundances from one absorption component to another appear to be highly (artificially) scattered. This is seen in Section 4.1 and Figures 1 and 2. Let $\Delta = \log N(\mathrm{Mg\,II}) - \log N(\mathrm{Fe\,II})$. Table 1 gives $\Delta(\mathrm{true}) = 0.5$, $\Delta(\mathrm{compound}) = 0.5$, $\Delta(\mathrm{turbulent}, v = -2.152) = -1.3$, $\Delta(\mathrm{turbulent}, v = -0.279) \sim 2.5$, where, for the purposes of this illustration, the latter assumes a column density detection threshold of 10.0, so is an order of magnitude estimate. This means that claims for significant relative abundance fluctuations across an absorption complex may be incorrect if turbulent broadening has been assumed. If, instead, *summed* column densities are used for a complex, no spurious effects would be seen.

Author Contributions: Conceptualization: J.K.W.; Formal analysis: C.-C.L.; Investigation: D.M. All authors have read and agreed to the published version of the manuscript.

Funding: We are grateful to the John Templeton Foundation for support.

Data Availability Statement: We thank the ESPRESSO consortium for making public the reduced ESPRESSO spectrum of HE0515–4414 that we used for some of the calculations in this paper.

Acknowledgments: The supercomputer calculations described in this paper were carried out using OzSTAR at the Centre for Astrophysics and Supercomputing at Swinburne University of Technology. We also thank Mariusz P. Dąbrowski and Vincenzo Salzano for organising the Alternative Gravities and Fundamental Cosmology ALTECOSMOFUN'21 conference, Szczecin, Poland, 6–10 September 2021, and Bob Carswell and Paolo Molaro for comments on a draft of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dzuba, V.A.; Flambaum, V.V.; Webb, J.K. Space-Time Variation of Physical Constants and Relativistic Corrections in Atoms. *Phys. Rev. Lett.* **1999**, *82*, 888–891. [CrossRef]
2. Webb, J.K.; Flambaum, V.V.; Churchill, C.W.; Drinkwater, M.J.; Barrow, J.D. Search for Time Variation of the Fine Structure Constant. *Phys. Rev. Lett.* **1999**, *82*, 884–887. [CrossRef]
3. Pepe, F.; Cristiani, S.; Rebolo, R.; Santos, N.C.; Dekker, H.; Cabral, A.; Di Marcantonio, P.; Figueira, P.; Lo Curto, G.; Lovis, C.; et al. ESPRESSO at VLT. On-sky performance and first results. *Astron. Astrophys.* **2021**, *645*, A96. [CrossRef]
4. Marconi, A.; Marcantonio, P.D.; D'Odorico, V.; Cristiani, S.; Maiolino, R.; Oliva, E.; Origlia, L.; Riva, M.; Valenziano, L.; Zerbi, F.M.; et al. EELT-HIRES the high-resolution spectrograph for the E-ELT. In Proceedings of Ground-Based and Airborne Instrumentation for Astronomy VI, Edinburgh, UK, 26 June–1 July 2016.
5. Tamai, R.; Koehler, B.; Cirasuolo, M.; Biancat-Marchet, F.; Tuti, M.; Gonzáles Herrera, J.C. The ESO's ELT construction status. In Proceedings of Ground-Based and Airborne Telescopes VII, Austin, TX, USA, 10–15 June 2018.
6. Molaro, P.; Levshakov, S.A.; Monai, S.; Centurión, M.; Bonifacio, P.; D'Odorico, S.; Monaco, L. UVES radial velocity accuracy from asteroid observations. I. Implications for fine structure constant variability. *Astron. Astrophys.* **2008**, *481*, 559–569. [CrossRef]
7. Rahmani, H.; Wendt, M.; Srianand, R.; Noterdaeme, P.; Petitjean, P.; Molaro, P.; Whitmore, J.B.; Murphy, M.T.; Centurion, M.; Fathivavsari, H.; et al. The UVES large program for testing fundamental physics—II. Constraints on a change in μ towards quasar HE 0027-1836. *Mon. Not. R. Astron. Soc.* **2013**, *435*, 861–878. [CrossRef]
8. Whitmore, J.B.; Murphy, M.T. Impact of instrumental systematic errors on fine-structure constant measurements with quasar spectra. *Mon. Not. R. Astron. Soc.* **2015**, *447*, 446–462. [CrossRef]
9. Dumont, V.; Webb, J.K. Modelling long-range wavelength distortions in quasar absorption echelle spectra. *Mon. Not. R. Astron. Soc.* **2017**, *468*, 1568–1574. [CrossRef]
10. Webb, J.K.; Carswell, R.F.; Lee, C.C. Precision in high resolution absorption line modelling, analytic Voigt derivatives, and optimization methods. *Mon. Not. R. Astron. Soc.* **2021**, *508*, 3620–3633. [CrossRef]
11. Hänsch, T.W. Nobel Lecture: Passion for precision. *Rev. Mod. Phys.* **2006**, *78*, 1297–1309. 1297. [CrossRef]
12. Steinmetz, T.; Wilken, T.; Araujo-Hauck, C.; Holzwarth, R.; Hänsch, T.W.; Pasquini, L.; Manescau, A.; D'Odorico, S.; Murphy, M.T.; Kentischer, T.; et al. Laser Frequency Combs for Astronomical Observations. *Science* **2008**, *321*, 1335. [CrossRef]
13. Hänsch, T.W.; Picqué, N. Laser Spectroscopy and Frequency Combs. *J. Phys. Conf. Ser.* **2013**, *467*, 012001. [CrossRef]
14. Milaković, D.; Pasquini, L.; Webb, J.K.; Lo Curto, G. Precision and consistency of astrocombs. *Mon. Not. R. Astron. Soc.* **2020**, *493*, 3997–4011. [CrossRef]
15. Probst, R.A.; Milaković, D.; Toledo-Padrón, B.; Lo Curto, G.; Avila, G.; Brucalassi, A.; Canto Martins, B.L.; de Castro Leão, I.; Esposito, M.; González Hernández, J.I.; et al. A crucial test for astronomical spectrograph calibration with frequency combs. *Nat. Astron.* **2020**, *4*, 603–608. [CrossRef]
16. Schmidt, T.M.; Molaro, P.; Murphy, M.T.; Lovis, C.; Cupani, G.; Cristiani, S.; Pepe, F.A.; Rebolo, R.; Santos, N.C.; Abreu, M.; et al. Fundamental physics with ESPRESSO: Towards an accurate wavelength calibration for a precision test of the fine-structure constant. *Astron. Astrophys.* **2021**, *646*, A144. [CrossRef]
17. Mayor, M.; Pepe, F.; Queloz, D.; Bouchy, F.; Rupprecht, G.; Lo Curto, G.; Avila, G.; Benz, W.; Bertaux, J.L.; Bonfils, X.; et al. Setting New Standards with HARPS. *Messenger* **2003**, *114*, 20–24.
18. Carswell, R.F. Bob Carswell's Homepage. Available online: https://people.ast.cam.ac.uk/~rfc/ (accessed on 27 January 2022).
19. Carswell, R.F.; Webb, J.K. VPFIT: Voigt Profile Fitting Program. Available online: http://ascl.net/1408.015(accessed on 27 January 2022)
20. Lee, C.C.; Webb, J.K.; Carswell, R.F. Addendum: Precision in high resolution absorption line modelling, analytic Voigt derivatives, and optimisation methods. *arXiv* **2021**, arXiv:2112.14490.
21. Carswell, R.F.; Webb, J.K. VPFIT User Guide. Available online: https://www.overleaf.com/read/vbxkcfnfgksr (accessed on 27 January 2022).
22. Lee, C.C.; Webb, J.K.; Carswell, R.F. Quantum mechanics at high redshift - modelling damped Lyman-α absorption systems. *Mon. Not. R. Astron. Soc.* **2020**, *491*, 5555–5571. [CrossRef]
23. Lee, C.C.; Webb, J.K.; Carswell, R.F.; Milaković, D. Artificial intelligence and quasar absorption system modelling; application to fundamental constants at high redshift. *Mon. Not. R. Astron. Soc.* **2021**, *504*, 1787–1800. [CrossRef]
24. Wilczynska, M.R.; Webb, J.K.; King, J.A.; Murphy, M.T.; Bainbridge, M.B.; Flambaum, V.V. A new analysis of fine-structure constant measurements and modelling errors from quasar absorption lines. *Mon. Not. R. Astron. Soc.* **2015**, *454*, 3082–3093. [CrossRef]
25. Lee, C.C.; Webb, J.K.; Milaković, D.; Carswell, R.F. Non-uniqueness in quasar absorption models and implications for measurements of the fine structure constant. *Mon. Not. R. Astron. Soc.* **2021**, *507*, 27–42. [CrossRef]
26. Milaković, D.; Lee, C.C.; Carswell, R.F.; Webb, J.K.; Molaro, P.; Pasquini, L. A new era of fine structure constant measurements at high redshift. *Mon. Not. R. Astron. Soc.* **2021**, *500*, 1–21. [CrossRef]
27. Noterdaeme, P.; Balashev, S.; Ledoux, C.; Duchoquet, G.; López, S.; Telikova, K.; Boissé, P.; Krogager, J.K.; De Cia, A.; Bergeron, J. Sharpening quasar absorption lines with ESPRESSO. Temperature of warm gas at z $\sim$ 2, constraints on the Mg isotopic ratio, and structure of cold gas at z $\sim$ 0.5. *Astron. Astrophys.* **2021**, *651*, A78. [CrossRef]

28. European Southern Observatory. SKYCALC Sky Model Calculator. Available online: http://www.eso.org/observing/etc/bin/gen/form?INS.MODE=swspectr+INS.NAME=SKYCALC (accessed on 27 January 2022).

29. Bainbridge, M.B.; Webb, J.K. Evaluating the New Automatic Method for the Analysis of Absorption Spectra Using Synthetic Spectra. *Universe* **2017**, *3*, 34. [CrossRef]

30. Bainbridge, M.B.; Webb, J.K. Artificial intelligence applied to the automatic analysis of absorption spectra. Objective measurement of the fine structure constant. *MNRAS* **2017**, *468*, 1639–1670. [CrossRef]

31. Liddle, A.R. How many cosmological parameters? *Mon. Not. R. Astron. Soc.* **2004**, *351*, L49–L53. [CrossRef]

32. Liddle, A.R. Information criteria for astrophysical model selection. *Mon. Not. R. Astron. Soc.* **2007**, *377*, L74–L78. [CrossRef]

33. Burnham, K.; Anderson, D. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; Springer: New York, NY, USA, 2002.

34. Rossi, R.; Murari, A.; Gaudio, P.; Gelfusa, M. Upgrading Model Selection Criteria with Goodness of Fit Tests for Practical Applications. *Entropy* **2020**, *22*, 447. [CrossRef]

35. Webb, J.K.; Lee, C.C.; Carswell, R.F.; Milaković, D. Getting the model right: An information criterion for spectroscopy. *Mon. Not. R. Astron. Soc.* **2021**, *501*, 2268–2278. [CrossRef]

36. Murphy, M.; Webb, J.; Flambaum, V.; Churchill, C.; Prochaska, J. Possible evidence for a variable fine-structure constant from QSO absorption lines: Systematic errors. *Mon. Not. R. Astron. Soc.* **2001**, *327*, 1223–1236. [CrossRef]

37. King, J.A.; Webb, J.K.; Murphy, M.T.; Flambaum, V.V.; Carswell, R.F.; Bainbridge, M.B.; Wilczynska, M.R.; Koch, F.E. Spatial variation in the fine-structure constant—New results from VLT/UVES. *Mon. Not. R. Astron. Soc.* **2012**, *422*, 3370–3414. [CrossRef]

38. Evans, T.M.; Murphy, M.T.; Whitmore, J.B.; Misawa, T.; Centurion, M.; D'Odorico, S.; Lopez, S.; Martins, C.J.A.P.; Molaro, P.; Petitjean, P.; et al. The UVES Large Program for testing fundamental physics—III. Constraints on the fine-structure constant from three telescopes. *Mon. Not. R. Astron. Soc.* **2014**, *445*, 128–150. [CrossRef]

39. Murphy, M.T.; Cooksey, K.L. Subaru Telescope limits on cosmological variations in the fine-structure constant. *Mon. Not. R. Astron. Soc.* **2017**, *471*, 4930–4945. [CrossRef]

40. Kotuš, S.M.; Murphy, M.T.; Carswell, R.F. High-precision limit on variation in the fine-structure constant from a single quasar absorption system. *Mon. Not. R. Astron. Soc.* **2017**, *464*, 3679–3703. [CrossRef]

41. Murphy, M.T.; Molaro, P.; Leite, A.C.O.; Cupani, G.; Cristiani, S.; D'Odorico, V.; Génova Santos, R.; Martins, C.J.A.P.; Milaković, D.; Nunes, N.J.; et al. Fundamental physics with ESPRESSO: Precise limit on variations in the fine-structure constant towards the bright quasar HE 0515–4414. *arXiv* **2021**, arXiv:2112.05819.