Uncertainty of Line-of-sight Velocity Measurements of Faint Stars from Low- and Medium-resolution Optical Spectra

László Dobos[1] , Alexander S. Szalay[1] , Tamás Budavári[1,2,3] , Evan N. Kirby[4] , Robert H. Lupton[5] , and
Rosemary F. G. Wyse[1]
[1] Department of Physics & Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA; dobos@jhu.edu
[2] Department of Applied Mathematics & Statistics, Johns Hopkins University, Baltimore, MD 21218, USA
[3] Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA
[4] Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA
[5] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA

Abstract

Massively multiplexed spectrographs will soon gather large statistical samples of stellar spectra. The accurate estimation of uncertainties on derived parameters, such as the line-of-sight velocity v_{los}, especially for spectra with low signal-to-noise ratios (S/Ns), is paramount. We generated an ensemble of simulated optical spectra of stars as if they were observed with low- and medium-resolution fiber-fed instruments on an 8 m class telescope, similar to the Subaru Prime Focus Spectrograph, and determined v_{los} by fitting stellar templates to the simulated spectra. We compared the empirical errors of the derived parameters—calculated from an ensemble of simulations—to the asymptotic errors determined from the Fisher matrix, as well as from Monte Carlo sampling of the posterior probability. We confirm that the uncertainty of v_{los} scales with the inverse square root of the S/N, but also show how this scaling breaks down at low S/N and analyze the error and bias caused by template mismatch. We outline a computationally optimized algorithm to fit multiexposure data and provide a mathematical model of stellar spectrum fitting that maximizes the so called significance, which allows for calculating the error from the Fisher matrix analytically. We also introduce the effective line count, and provide a scaling relation to estimate the errors of v_{los} measurements based on stellar type. Our analysis covers a range of stellar types with parameters that are typical of the Galactic outer disk and halo, together with analogs of stars in M31 and in satellite dwarf spheroidal galaxies around the Milky Way.

Unified Astronomy Thesaurus concepts: Galactic archaeology (2178); Radial velocity (1332); Doppler shift (401); Spectroscopy (1558); Late-type stars (909); Late-type giant stars (908)

1. Introduction

To decipher the assembly history of the Milky Way and neighboring galaxies, Galactic archeology relies on measuring the dynamical and chemical properties of individual stars throughout the outer disk and the halo of the Galaxy, as well as stars in satellite galaxies and M31. The low number density of target stars, the requirement of sufficiently large sample sizes for statistical analysis, and the large celestial area to be covered make wide field of view 4 and 8 m class telescopes with low- and medium-resolution optical fiber-fed spectrographs—such as the Subaru Prime Focus Spectrograph (PFS; Sugai et al. 2014; Takada et al. 2014) or the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2022; Dey et al. 2023)—are the best available tools for observations. Even with these advanced instruments, constraints on observational time limit the achievable signal-to-noise ratio (S/N) to a range of S/N = 5–100 per resolution element, with the large majority of the targets observed at S/N < 15. Typically, a constraint on the maximally acceptable uncertainties for measured velocities $\sigma(v_{los})$ and metallicities $\sigma([Fe/H])$ drive the requirements on S/N for faint stars. Measuring the abundances of most individual elements is typically reserved for brighter stars with higher S/Ns and higher-resolution instruments.

In this study, our primary focus is on characterizing $\sigma(v_{los})$, the uncertainty in the line-of-sight velocity v_{los}, based on many repeated realizations of simulated stellar spectra. We characterize the error as a function of spectral type and a wide range of flux S/Ns. Understanding the sources of uncertainty in line-of-sight velocity determination and precise estimation of the error are essential to investigate the velocity distributions of dynamically cold systems such as stellar streams and ultrafaint dwarf spheroidal (dSph) satellite galaxies and tackle problems such as the radial mass profile of dark matter halos in dSph galaxies, as overestimating the uncertainties can directly lead to underestimating the velocity dispersion and vice versa (Koposov et al. 2011). Pessimistic error estimates lead to underestimation of the velocity dispersion, especially when the velocity dispersion is comparable to the measurement error. To compile a statistically significant sample, the low potential target density of nearby dSph galaxies might demand a fainter magnitude limit to the survey because targeting a large number of faint stars with less precise line-of-sight velocities—but with well understood errors—might have better constraining power on the radial dark matter profile than brighter but smaller samples.

It has been established that the close binary fraction in the stellar populations of the Milky Way increases with decreasing metallicity (Badenes et al. 2018; Moe et al. 2019), and a similar dependence is inferred for stellar populations in dSph galaxies, from the frequency of blue stragglers (Wyse et al. 2020). The majority of targets in stellar spectroscopic surveys of Local

Group galaxies are expected to be in multiple systems. When the brightnesses of the two stars are comparable, spectroscopic binaries have a significant effect on the uncertainty of v_{los} measurements when the multiplicity is not accounted for. In the case of evolved (giant) primaries, the contribution of the companion (dwarf) star to the spectrum is usually negligible. However, at the distance of extragalactic targets, such as red giants in M31 and M33, the stellar surface density can be high enough that the probability that two stars are aligned along the line of sight becomes significant (Kounkel et al. 2021). Even though these are not gravitationally bound systems, the effect of this multiplicity on the v_{los} measurement is similar to that of binaries. The effect of unresolved binaries on the v_{los} measurement is an important topic, but it is beyond the scope of the present paper and we defer further discussion to a future study. For a review of currently available template-based and model-independent methods for disentangling spectroscopic binaries from multiepoch, low-resolution spectra, we refer the reader to the recent study by Seeburger et al. (2024).

We simulate spectroscopic observations of single stars with a wide range of flux S/Ns, for a broad range of stellar types. We model both the random and systematic observational effects with software based on the spectroscopic exposure time calculator (ETC) originally developed by Hirata et al. (2012) for the Roman Space Telescope[6] and adapted to the Subaru PFS by Yabe et al. (2017).

We determine v_{los} and its measurement error with several methods. We compare the empirical uncertainty of v_{los}, determined from an ensemble of simulations, to those that one can derive from individual spectra via mathematical methods. These methods include the evaluation of the Fisher matrix of a maximum likelihood or maximum significance fit using a single model spectrum as a template. In addition to fitting a single template to determine v_{los}, we also optimize for the atmospheric parameters by interpolating the templates on a synthetic stellar spectrum grid. Although this allows for inferring the stellar parameters, including effective temperature, metallicity, and surface gravity, in the present study, we restrict our focus to the uncertainties of the line-of-sight velocity. Nevertheless, we also compare the asymptotic errors derived from the Fisher matrix to those determined from the full Monte Carlo sampling of the posterior distribution of v_{los} and the stellar atmospheric parameters.

The accuracy of the line-of-sight velocity measurement is known to depend on the proper choice of template, and template mismatch introduces an additional systematic error (David et al. 2014). In the case of fitting observations with synthetic spectra, template mismatch can mean the following: (i) the template is inaccurate because the synthetic spectra do not match real observations due to shortcomings of the stellar models, and (ii) the stellar models are accurate but a template with the wrong atmospheric parameters is used. Although the use of templates derived from inaccurate stellar models to fit v_{los} is considered a significant source of systematic error, we do not investigate this topic in this work. In fact, we fit the simulated observations using the same synthetic spectrum grid that we used to generate the simulated observations. We investigate, however, the magnitude of the bias and systematics caused by mismatched atmospheric parameters and characterize them as a function of S/N.

We also omit an analysis of sources of additional systematic error terms such as convolving stellar templates with a mismatched line-spread function (LSF), systematic errors in the wavelength solution due to an inaccurate wavelength calibration, or time-dependent drifts of the wavelength solution over the course of the night.

Software libraries for direct pixel template fitting use a range of approaches. The approaches vary as to (i) whether fitting is done using individual exposures on a per pixel basis or resampled, stacked spectra; (ii) how the fluxing errors (or template inaccuracies) are corrected for; (iii) whether errors in the wavelength solution are corrected for or not; (iv) whether a full Bayesian method is used including priors on the parameters, as opposed to calculating the likelihood function only; and (v) whether only a multivariate maximum finder is used or a full Monte Carlo sampling of the posterior (likelihood) distribution is done to estimate the parameters and their uncertainties. In addition to these, many template-fitting v_{los} finders are designed to correct for the peculiarities of a specific instrument, which can include fluxing systematics, wavelength shifts from slit mis-centering (Sohn et al. 2007), etc.

We developed our of line-of-sight velocity measurement algorithm that also works by fitting stellar templates to observed or simulated spectra on a per pixel basis and supports fitting multiple exposures with potentially different LSFs. The code allows for a multiplicative flux correction to the observation in the form of a wavelength-dependent polynomial to account for fluxing systematics, as well as imprecisions of the templates. Our approach is very similar to Koposov et al. (2011), with some additions: we calculate the full covariance matrix of the parameter errors analytically, also taking the covariances of coefficients of the flux correction polynomial into account, and consider the case of nonuniform priors on v_{los} and the template parameters when calculating the full error covariance matrix. Our software implementation is immensely optimized to efficiently perform grid interpolation and convolutions at high resolution with wavelength-dependent kernels. The optimizations were necessary to fit a large number of simulations in a reasonable time.

It is generally useful to be able to estimate the expected uncertainty of a parameter measurement before observations. Scaling relations provide a simple way to approximate $\sigma(v_{\mathrm{los}})$ from the spectral resolution, S/N, bandwidth, and spectral type of the target. Based on our empirical results on the uncertainty of the line-of-sight velocity, we propose a modified version of the scaling relation of Hatzes & Cochran (1992) that also takes the spectral type into account.

The paper is structured as follows. In Section 2, we review former work related to uncertainty estimation of v_{los} and then move over to describing the simulated observations in Section 3. In Section 4, we introduce the significance function, which provides a convenient mathematical formalism to calculate $\sigma(v_{\mathrm{los}})$ analytically, but we relegate the detailed calculation to Appendix A. We explain our performance optimized template-fitting algorithm in Section 5, and in Section 6, we present the outcome of the simulations. Results are discussed and compared to earlier work in Section 7, and the paper is concluded in Section 8.

2. Former Work

Inferring physical parameters, including the line-of-sight velocity from spectra is fundamental and has been in the focus of algorithm development for decades. Although computationally

---

more intensive, direct pixel fitting of model templates has largely replaced Fourier methods, originally to extract the line-of-sight velocity distribution from galaxy spectra (e.g., Rix & White 1992; Newman et al. 2013), but later more complex and refined spectrum processing pipelines have been developed for single star $v_{\rm los}$ measurement as well (Koleva et al. 2009; Koposov et al. 2011; Cunningham et al. 2019; Jenkins et al. 2021).

Among the many other spectrum fitting codes, Koleva et al. (2009) introduced the ULySS software library to fit single stellar spectra as well as to analyze the star formation history and chemical evolution of composite stellar populations. They emphasized the importance of using an LSF that closely matches that of the instrument to achieve good results. More recently, Koposov et al. (2011) analyzed Very Large Telescope (VLT)/GIRAFFE fiber-fed spectra on a per pixel basis highlighting the advantages of using native detector pixels and likelihood stacking to fit the data. Koposov et al. (2011), based on the earlier work of Koleva et al. (2008), developed a pixel-fitting method that allows for a flux correction to the template in the form of a wavelength-dependent polynomial to account for fluxing systematics in the observations as well as imprecisions of the templates. Koposov et al. (2011) consider the coefficients of the flux correction polynomial nuisance parameters and marginalize them from the joint likelihood of $v_{\rm los}$ and the template parameters.

Koposov et al. (2011) and later Jenkins et al. (2021) realized that the sky lines in the raw VLT/GIRAFFE spectra had systematic shifts and developed an algorithm to recalibrate the wavelength solutions of 1D spectra based on the positions of sky emission lines. Walker et al. (2006, 2015) developed a Bayesian technique to fit spectra of dSph members at medium resolution and found that the posterior probability distribution of $v_{\rm los}$ and the atmospheric parameters often become strongly non-Gaussian at low S/Ns.

To evaluate the effect of template mismatch on measuring $v_{\rm los}$, North et al. (2002) suggested fitting spectra with several templates and found that template mismatch is not a significant source of systematic errors in case of G, K, and M stars. David et al. (2014) quantified the effect of template mismatch as a function of the atmospheric parameters and found deviations of $\Delta v_{\rm los} \pm 0.1$ km s^{-1} in the temperature range $4000 \leqslant T_{\rm eff} \leqslant 6000$ K based on fitting mismatched templates to noiseless spectra. They also showed that the bias can be significantly higher for spectra with relatively broad Balmer lines, Paschen lines, or the Ca II triplet. Walker et al. (2015) showed—using Monte Carlo sampling of the posterior distribution of $v_{\rm los}$, the atmospheric parameters, and the flux correction coefficients—that $v_{\rm los}$ is essentially uncorrelated with the other parameters.

In order to estimate the uncertainty of $v_{\rm los}$ measurements without detailed simulations, Hatzes & Cochran (1992) provided a scaling formula to calculate $\sigma(v_{\rm los})$ from the S/N, spectral resolving power, and instrument bandwidth and proved the validity of the formula on simulations of high-resolution spectroscopic observations. Bouchy et al. (2001) derived theoretical constraints on $\sigma(v_{\rm los})$ and emphasized that the spectral type plays an important role in the precision of $v_{\rm los}$ measurements. They determined that the local slope of the spectrum—degraded to the resolving power of the instrument and evaluated at the center of each detector pixel—determines how much information is carried by the spectrum about Doppler shift.

In the present work, we generalize the formalism proposed by Kaiser (2004), originally to detect faint point sources in noisy pixelated images. Kaiser (2004) showed how to quantify the significance of the detections and calculate unbiased error estimates of the parameters such as the centroid positions. When the method is adopted to spectroscopy of faint objects, the significance function becomes independent of the normalization of the observed spectrum and the template, and the location of its maximum will coincide with the maximum of the likelihood function. The formalism allows for calculating the Fisher information matrix at the best-fit parameters analytically and deriving expressions for the uncertainties of the model parameter as well as the Doppler shift, as we will show in Section 4.

Table 1
Assumed Parameters of the Optical Spectrograph Arms, Which are Very Similar to the Parameters of the Subaru PFS Instrument

	B	R	MR
λ coverage [nm]	380–650	630–970	710–885
pixel dispersion [Å]	0.7	0.9	0.4
$\Delta\lambda$ spectral resolution [Å]	2.1	2.7	1.6
Δv velocity resolution [km s^{-1}]	130	100	60
R resolving power	2300	3000	5000

3. Simulated Observations

To assess the uncertainty of Doppler shift estimates from synthetic template fitting, we rely on detailed simulations of spectroscopic observations with known $v_{\rm los}$, as well as atmospheric and observational parameters. For this study, we generated 1000 simulated observations, in three different spectrograph arm configurations, of every stellar type listed in Table 2. The distribution of the observational parameters was the same for all stellar types, representing a realistic dark night similar to the environment on Maunakea. Our goal was to sample a broad range of S/Ns instead of generating a synthetic catalog. Hence, magnitudes were sampled uniformly from an interval that will typically be accessible by Subaru PFS (Sugai et al. 2014). The three spectrograph arm configurations of PFS that we investigated are listed in Table 1. Simulations were executed for the blue and red arms, with the two different resolution configuration of the latter: low and medium resolution. We denote these with B, R and MR, respectively. The resolution (FWHM of the LSF) is approximately constant within each of these arms. A single pointing of PFS consists of an observation in B plus either R or MR. PFS also has a near-infrared channel, but we do not include it in our simulations.

3.1. Simulation Software

Our simulations are based on the ETC code version 5, originally written by Hirata et al. (2012) and modified by the Subaru PFS team (Yabe et al. 2017) for the fiber-fed PFS instrument. The ETC implements an optical model of the telescope, the wide field corrector, and the spectrograph to calculate the LSF, the fiber trace width, the effective aperture, and vignetting. It also applies the combined transmission function of the atmosphere, the optical elements, as well as the quantum efficiency curve of the detector and a model of the detector noise characteristics to calculate the noise terms. To calculate the photon counts and the noise, object photons, sky photon (continuum and lines separately), photons of light

Table 2
Summary of the Stellar Models Used as the Basis for the Simulated Observations

No.	[Fe/H]	T_{eff} (K)	log g	Phase	Exp. Count
1.	−2.0	6250	4.5	MS	12
2.	−2.0	5500	4.5	MS	12
3.	−2.0	4750	4.5	MS	12
4.	−0.5	6250	4.5	MS	12
5.	−0.5	5500	4.5	MS	12
6.	−0.5	4750	4.5	MS	12
7.	−1.5	5500	3.5	RGB	12
8.	−1.5	4750	2.0	RGB	12
9.	−1.5	4000	0.5	RGB	12
10.	−2.0	8000	3.0	HB	12
11.	−2.0	7000	3.0	HB	12
12.	−2.0	6000	2.5	HB	12
13.	−0.7	5500	3.5	RGB	20
14.	−0.7	4750	2.0	RGB	20
15.	−0.7	4000	0.5	RGB	20
16.	0.0	4750	3.0	RGB	20
17.	0.0	4000	1.5	RGB	20
18.	0.0	3250	0.0	RGB	20

Note. We list the synthetic spectrum atmospheric parameters, stellar evolutionary phase, and the number of simulated 15 minute exposures for each model.

Figure 1. Location of the simulated observations in the plane of atmospheric parameters (left) and the Subaru Hyper-Suprime Cam color–magnitude diagram (right). We plot a selection of PARSEC isochrones for reference, as well as the locations of the models we simulated. Note that the models were chosen to be aligned with the synthetic spectrum grid, thus may not lie on the isochrones. See Table 2 for a detailed list of parameters.

scattered by the Moon, and stray light from the spectrograph are taken into account.

Since the aforementioned computations can take a significant time, we modified the ETC to write out the LSF, the transmission function, and the contributions of the various noise terms to the final flux variance. Then we ran the ETC on a grid of observational parameters including seeing, the angle of the target and the Moon with respect to the zenith and each other, and the field angle of the target with respect to the optical axis to calculate vignetting. By interpolating these pretabulated ETC outputs, simulating a single observation takes only a few hundred milliseconds. In order to generate tens of thousands of simulations, we implemented a parallel spectroscopic simulation library in Python that can handle very fast synthetic spectrum grid interpolation, LSF convolution, and resampling. The highly customizable software suite was originally developed to generate large training sets for machine learning spectral analysis methods and its details will be published elsewhere.

Our simulations are based on the high-resolution PHOENIX library of synthetic stellar spectra (Husser et al. 2013), which are available at a resolving power of $R = 500,000$ in the optical.

Whenever S/Ns are quoted, we refer to the 95th percentile S/N per resolution element in the MR arm. If the simulation was performed for the low-resolution R arm, the equivalent S/N for the MR arm is quoted.

3.2. Model Parameters

We selected 18 stellar types that span the range of metallicity, effective temperature, and surface gravity that optical Galactic archeology surveys typically cover in the Milky Way, dSph galaxies, and nearby galaxies with resolvable stars such as in M31. Stellar types are summarized in Table 2 and indicated in Figure 1 on top of PARSEC isochrones (Bressan et al. 2012), both in terms of the physical parameters and the colors and absolute magnitudes in the Subaru Hyper-Suprime Cam
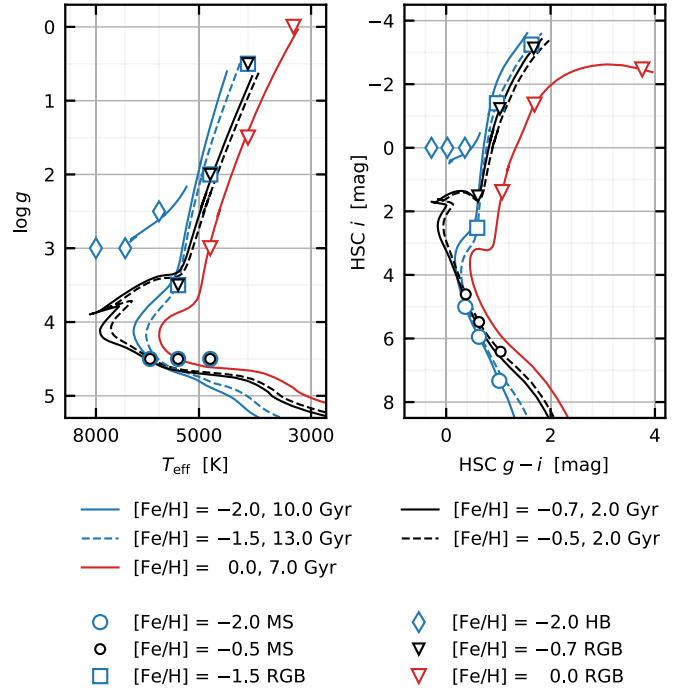
broadband filter set. The actual parameters slightly differ from the isochrones because the synthetic grid in Table 2 is not exactly aligned with the isochrones. For the sake of simplicity, all our models have solar [α/M] and, when fitting templates to simulations, we fit the three fundamental parameters only.

3.3. Simulation Details

We start the simulations from fluxed synthetic models, which have much higher resolution than the instrument, to be able to evaluate the ability of the template-fitting algorithm to detect subpixel Doppler shifts. As mentioned in Section 3.2, the atmospheric parameters were chosen to match the grid so no model interpolation was necessary to generate the simulated spectra. When fitting the simulated observations with stellar templates, the atmospheric parameters are treated as free parameters, and we linearly interpolate the grid to calculate the flux of models in between grid points.

We choose the Doppler shift randomly from a uniform distribution and shift the wavelengths of the spectral elements of the high-resolution model before smoothing it with the LSF and interpolating it to the pixels of the detector. The LSF calculated by the ETC already takes the pixelization into account. Hence, we use linear interpolation, since flux-conserving resampling would duplicate the effect of pixelization. We convolve the spectrum, still at high resolution, with the instrumental LSF, expressed as a Gaussian kernel with a wavelength-dependent width. The width of the Gaussian is taken from the output of the ETC, and the kernel is evaluated at high resolution, still on the wavelength grid of the input models.

Choosing the Doppler shift for each simulated spectrum randomly from a broad interval, instead of just measuring $\sigma(v_{\mathrm{los}})$ at $v_{\mathrm{los}} = 0$, allows us to test two important things: (i) We

Table 3
Parameters of the Simulated Spectroscopic Observations

Parameter	Range
seeing	0.6–1″0
extinction $E(B-V)$	0.0–0.5 mag
target zenith angle	0°–60°
field angle	0°0–0°5
single exposure time	15 minutes
exposure count	12 or 20
object apparent magnitude	see text
line-of-sight velocity	−500 to 500 km s^{-1}

can investigate the sensitivity of the template-fitting algorithm to subpixel shifts and whether there are any systematics depending on how the strong absorption features are shifted with respect to the pixel edges. We ran these tests and found no systematics. (ii) We can better test the part of the template-fitting algorithm that provides a coarse initial estimate of v_{los} before starting the maximum likelihood optimization.

We assume observational parameters similar to those of at the Maunakea Observatories on dark nights, as summarized in Table 3. Each of these parameters is sampled uniformly between the given limits. The 15 minute observation time was chosen according to the design parameters of the PFS instrument and we assumed 12 exposures (a total of 3 hr) per field for the MW and dSph targets (model Nos. 1–12) and 20 exposures (a total of 5 hr) for the M31 targets (model Nos. 13–18).

Instead of using S/N as the input parameter, we normalize the synthetic flux of the model spectrum to a magnitude drawn from a uniform distribution that covers the typical magnitude range of potential targets of PFS. Although this method will not result in uniform sampling in S/N, it will make sure that the flux errors from sky subtraction, scattered light, and the noise contribution of the detector are taken into account at the right ratio. At faint magnitudes, the sky photons are the most significant source of flux error, hence, rescaling the noise level of a simulation to a given S/N would change the ratio of the noise components. We use the pretabulated output from the ETC to obtain the combined transmission curve of the atmosphere and the instrument, as well as the continuum and line emission of the sky. We only simulate dark time observations here but the software is also capable of calculating the scattered light from the Moon. To calculate the resulting S/N, we approximate the photon noise with a Gaussian distribution and sum up the detector noise terms in quadrature to calculate the variance in each pixel.

We sample the value of the $E(B-V)$ extinction randomly between 0.0 and 0.5 mag and apply the average Milky Way extinction curve of Cardelli et al. (1989). When fitting the simulated spectra, we do not correct for extinction, but rather let the flux correction polynomial (see Section 4.4) account for it.

We apply a multiplicative function to the flux in the form of a slowly changing, but high-order function with a maximum amplitude of 2%. The purpose of this function is to simulate the effects of improper fluxing as well as differences between the continua of the synthetic and real spectra.

Even though we do not simulate sky subtraction systematics or residuals, we mask out pixels where the strong sky lines are, including the wings of those lines. We also mask out pixels where the S/N would be extremely low due to low instrumental transmission.

The actual noise realization is only added to the simulated spectra during template fitting to allow generating multiple noise realizations to imitate multiple exposures. When fitting templates to the simulated observations, we assume 15 minute exposures of the target that add up to a 3 hr or 5 hr total observation time. We assume the same instrumental noise variance and observational parameters for each exposure. As explained in Section 4, the exposures are not stacked, but the templates are fitted on a per pixel basis.

We will also demonstrate how the zero-point errors of the wavelength calibration affect the uncertainties of the v_{los} measurements. To simulate the zero-point error, we execute the simulations as explained above, then we apply a small shift to the pixel central wavelengths, which corresponds to an affine transformation in λ and translation in $\log \lambda$. Simulating the zero-point error this way, instead of before the step when we interpolate the spectrum to the detector pixels, is a simple shortcut that makes fitting templates to multiple exposures with different wavelength solutions easier. Note, that the majority of the simulations were generated without a wavelength calibration error, hence the results for $\sigma(v_{\mathrm{los}})$ represent the theoretical lower limit that does not take the systematic error floor of the instrument into account. The instrumental velocity error floor is generally thought to be around or slightly above a hundredth of a pixel, or approximately 1.3, 1.0, and 0.6 km s^{-1} in the spectrograph arms B, R, and MR, respectively.

The steps of the simulation procedure are as follows.

1. Load the high-resolution input model from the synthetic spectrum grid.
2. Shift the model to the desired Doppler shift between −500 and 500 km s^{-1}.
3. Apply the extinction model.
4. Given a randomly chosen observed magnitude, calculate the flux in physical units.
5. Convolve the synthetic spectrum with the instrumental LSF on the original wavelength grid of the high-resolution model.
6. Interpolate the spectrum to the grid of detector pixels.
7. Calculate the photon count in each pixel from the source, as well as the different components of the sky and Moon.
8. Calculate the S/N in each pixel.
9. Apply the model for simulated flux calibration error.
10. Generate a pixel mask of strong sky lines and very noisy pixels.
11. Optionally, shift the central wavelengths of each pixel to mimic the wavelength calibration zero-point error.

4. Template Fitting

In this section, we will follow Kaiser (2004) who introduced a formalism for the optimal detection of faint, point-like objects in noisy pixelated images. Here we generalize his approach in the context of spectroscopy, where we consider the optimal extraction of a signal from noisy observations by cross-correlating them with a stellar template. While the objective function derived from these calculations coincides with the quantity that spectrum template-fitting codes typically aim to maximize, we provide formulae for the full covariance matrix and an analytic expression for the uncertainty of v_{los}.

In the most general case that we consider, targets are observed using multiple spectrograph arms and many detector readouts. We assume that repeated observations can happen at

different times during different atmospheric conditions and with different fiber configurations, where the latter can completely change both the LSF and the pixelization of the spectrum when a different fiber is assigned to the same target. As usual, we do likelihood stacking instead of attempting to resample the spectra to the same wavelength grid and stack the flux. Resampling and stacking, in addition to mixing the LSF from different observations, would result in strong correlations in the flux error. Likelihood stacking, on the other hand, simplifies handling the flux error but has consequences on computational performance that have to be addressed.

4.1. Analytic Calculation of the Significance Function

Kaiser (2004) introduced a formalism to detect faint point sources in images. Here we generalize this formalism for spectroscopy and show how it allows the Fisher information matrix and elements of its inverse to characterize the error of $v_{\rm los}$ measurements. While Kaiser (2004) assumed a point source convolved with a spatially extended point-spread function (PSF) to cross-correlate with the observed image, we consider a template spectrum convolved with the LSF and its cross-correlation with the observed spectrum. The template can be fixed or the atmospheric parameters of the template can also be subject to optimization, along with $v_{\rm los}$.

While the amplitude of the image of a point source can be described by a single number, in the case of spectra, we also have to discuss corrections to incorrect fluxing of the observations or, not quite equivalently, inaccurate model continua. We defer discussing the error arising from inaccurate wavelength calibration and the methods of wavelength correction to a follow-up paper. The most generic model we consider here allows for finding $v_{\rm los}$, the atmospheric parameters of the template and a multiplicative correction of the observed flux in the form of a linear combination of wavelength-dependent basis functions where the coefficients are unknown. When fitting the template parameters, or including a correction for the fluxing systematics, the calculations quickly become elaborate, hence we moved them to Appendix A, and only quote the most important results here.

With p representing the pixel label that indexes each pixel of each exposure, f_p is the sky-subtracted photon count of the observed spectrum, and σ_p^2 is the total noise variance of the observed spectrum. The variance σ_p^2 is calculated from the total photon count n_p, which includes the object photons and sky photons, as well as additional error terms such as the readout noise and dark current. The model $m_p(z)$ is the stellar spectrum template, Doppler shifted by $z = v_{\rm los}/c$, convolved with the LSF at high resolution, and resampled to the detector pixels. The orthogonal basis functions $q_j(\lambda)$ are defined across the full wavelength coverage and evaluated at the wavelengths of pixel centers λ_p to get the matrix $q_{jp} = q_j(\lambda_p)$.

Assuming Gaussian errors of the flux, the likelihood function that includes the template parameters, as well as the flux correction, takes the form of

$$L(A_j, z, \theta)$$
$$= \prod_p \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left[-\frac{1}{2}\frac{\left[f_p - \sum_j A_j q_{jp} \cdot m_p(z, \theta)\right]^2}{\sigma_p^2}\right], \quad (1)$$

where the product goes over the p pixels of each spectrograph arm and each exposure. Evaluated at each pixel

center, m_p is the flux of the template, which depends on the Doppler shift z and the atmospheric parameters θ, whereas the unknown A_j coefficients form a linear combination with the q_{jp} orthogonal functions. Here we also assume the independence of the flux errors, measured in each pixel. This is a good first approximation if the spectrum is not rebinned, however, the flux error in neighboring pixels is usually covariant due to sky and scattered light subtraction.

If we only consider a single scalar amplitude A instead of a wavelength-dependent flux correction, the likelihood becomes

$$L(A, z) = \prod_p \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left[-\frac{1}{2}\frac{[f_p - A\,m_p(z)]^2}{\sigma_p^2}\right]. \quad (2)$$

The log likelihood, with the constant terms omitted, can be written as the sum of two functions in the form of

$$\mathcal{L}(A, z) = A\varphi(z) - \frac{1}{2}A^2\chi(z), \quad (3)$$

where

$$\varphi(z) = \sum_p \frac{f_p m_p(z)}{\sigma_p^2}, \quad \text{and} \quad \chi(z) = \sum_p \frac{m_p^2(z)}{\sigma_p^2}. \quad (4)$$

In the original problem, Kaiser (2004) considers point sources, so the equivalent of m_p in his case is the PSF, which means $\varphi(z)$ is the double convolution of the ideal image, i.e., the PSF-convolved observed image. In our case, $m_p(z)$ is the spectrum template convolved with the instrumental LSF so $\varphi(z)$ is the cross-correlation function of the observation and the LSF-convolved template, i.e., a matched filter, weighted by the inverse variance. It is interesting to see that $\varphi(z)$ becomes independent of the observation when the object is so bright that the noise is dominated by the object photons and σ_p^2 becomes approximately equal to f_p. When observing faint objects, however, the noise is dominated by the sky photons and $\varphi(z)$ becomes independent of the template. Kaiser (2004) compares $\chi(z)$ to an exposure map whereas in our case $\chi(z)$ can be related to the information content of the spectral pixels.

At the maximum of the likelihood, the partial derivatives of $\mathcal{L}$ should vanish, which yields the equations

$$\frac{\partial \mathcal{L}}{\partial A} = \varphi(z) - A\chi(z) = 0, \quad (5)$$

$$\frac{\partial \mathcal{L}}{\partial z} = A\varphi'(z) - \frac{1}{2}A^2\chi'(z) = 0, \quad (6)$$

where the prime denotes differentiation by z. The first equation can be solved for A and gives

$$A = \frac{\varphi(z)}{\chi(z)}, \quad (7)$$

which can, in turn, be substituted into the second one to get

$$\frac{\varphi'(z)}{\varphi(z)} = \frac{1}{2}\frac{\chi'(z)}{\chi(z)}. \quad (8)$$

This latter can be rewritten as

$$\frac{d}{dz}\left(\frac{\varphi(z)}{\sqrt{\chi(z)}}\right) = 0. \quad (9)$$

Kaiser (2004) called

$$\nu(z) = \frac{\varphi(z)}{\sqrt{\chi(z)}}, \tag{10}$$

the significance function. This is the function that needs to be maximized to get the best-fit Doppler shift. The significance is only a function of z and the lack of explicit dependence on A expresses the fact that the normalization of the template carries no information on the Doppler shift.

Let us denote the quantities evaluated at the maximum of $\nu(z)$ with the subscript 0. Since at the maximum of the significance function $\varphi_0 = A_0 \chi_0$ holds and $\mathcal{L}_0 = A_0 \varphi_0 - A_0^2 \chi_0$, the maximum of the log likelihood is at the same z as the maximum of $\nu(z)$ and

$$\mathcal{L}_0 = \frac{\varphi_0^2}{\chi_0} - \frac{1}{2}\frac{\varphi_0^2}{\chi_0} = \frac{1}{2}\nu_0^2. \tag{11}$$

4.2. The Fisher Information Matrix

We are now at a point to calculate the Fisher information matrix for measuring the line-of-sight velocity from a noisy observation, defined as

$$F_{Az} = -\left\langle \frac{\partial^2 \mathcal{L}}{\partial A \partial z} \right\rangle_0, \tag{12}$$

where the 0 subscript indicates that this should be evaluated at the parameter values corresponding to the maximum point z_0 of the significance function $\nu(z)$, and the associated A_0 that can be calculated from Equation (7). The averaging in Equation (12) is done over all possible realizations of the noise.

Let us first write down the curvature matrix of a single noise realization, which consists of the second derivatives of the log likelihood

$$C = \begin{pmatrix} -\dfrac{\partial^2 \mathcal{L}}{\partial A^2} & -\dfrac{\partial^2 \mathcal{L}}{\partial A \partial z} \\ -\dfrac{\partial^2 \mathcal{L}}{\partial A \partial z} & -\dfrac{\partial^2 \mathcal{L}}{\partial z^2} \end{pmatrix} = \begin{pmatrix} -\chi & -\varphi' + A\chi' \\ -\varphi' + A\chi' & -A\varphi'' + \dfrac{1}{2}A^2\chi'' \end{pmatrix}. \tag{13}$$

By using Equations (5) and (6) at the maximum point, we can simplify this into the Fisher matrix in the form of

$$F = \begin{pmatrix} \chi_0 & \varphi_0' \\ \varphi_0' & -\dfrac{\varphi_0\varphi_0''}{\chi_0} + \dfrac{1}{2}\varphi_0^2\dfrac{\chi_0''}{\chi_0^2} \end{pmatrix}, \tag{14}$$

in agreement with Equation (22) of Kaiser (2004). We can calculate the second derivative of the significance $\nu(z)$ as

$$\frac{\nu''}{\nu} = \frac{\varphi''}{\varphi} - \frac{1}{2}\frac{\chi''}{\chi} + \frac{3}{4}\left(\frac{\chi'}{\chi}\right)^2 - \left(\frac{\varphi'}{\varphi}\right)\left(\frac{\chi'}{\chi}\right). \tag{15}$$

At the maximum point this becomes

$$\frac{\nu_0''}{\nu_0} = \frac{\varphi_0''}{\varphi_0} - \frac{1}{2}\frac{\chi_0''}{\chi_0} + \left(\frac{\varphi_0'}{\varphi_0}\right)^2. \tag{16}$$

Multiplying this with $\nu_0^2 = \varphi_0^2/\chi_0$, we get

$$\frac{\varphi_0''\varphi_0}{\chi_0} - \frac{1}{2}\varphi_0^2\frac{\chi_0''}{\chi_0^2} = \nu_0\nu_0'' - \frac{\varphi_0'^2}{\chi_0}, \tag{17}$$

with which we can rewrite the Fisher matrix in the simpler form of

$$F = \begin{pmatrix} \chi_0 & \varphi_0' \\ \varphi_0' & -\nu_0\nu_0'' + \dfrac{\varphi_0'^2}{\chi_0} \end{pmatrix}. \tag{18}$$

The determinant of F is particularly simple and takes the form of

$$|F| = -\chi_0\nu_0\nu_0''. \tag{19}$$

The determinant is always positive as ν has a maximum so ν'' must be negative. Now we can easily see that the Cramér–Rao lower bound of the Doppler shift error is

$$\sigma_z^2 = -\frac{1}{\nu_0\nu_0''}. \tag{20}$$

Equation (20) has no explicit dependence on A, the normalization of the template, which means that the lower bound on the uncertainty of the Doppler shift is independent of the normalization (but the errors of z and A are not uncorrelated). The scaling of σ_z^2 is very intuitive: $1/\nu_0\nu_0''$ scales with the variance of the flux σ_p, thus so will σ_z^2, and the dimensions of σ_z are kilometers per second. In other words, the Doppler shift error is proportional to the inverse of the square root of the typical S/N of the flux measurements.

4.3. Including the Atmospheric Parameters

When fitting synthetic spectrum templates to observations, often not only the Doppler shift, but the atmospheric parameters are also unknown. In this more generic case, in addition to z, the template model $m_p(z, \theta)$, and consequently φ and χ, depend nonlinearly on the θ parameters, and the derivatives of the significance function and ϕ, with respect to z and θ, will appear in the Fisher matrix in the block form of

$$F = \begin{pmatrix} \chi & \varphi_z & \varphi_\beta \\ \varphi_z & -\nu\nu_{zz} + \dfrac{\varphi_z^2}{\chi} & -\nu\nu_{z\beta} + \dfrac{\varphi_z\varphi_\beta}{\chi} \\ \varphi_\alpha & -\nu\nu_{\alpha z} + \dfrac{\varphi_\alpha\varphi_z}{\chi} & -\nu\nu_{\alpha\beta} + \dfrac{\varphi_\alpha\varphi_\beta}{\chi} \end{pmatrix}, \tag{21}$$

where z in the index of φ, ν means partial differentiation by z, and the Greek indices indicate differentiation with respect to the θ template parameters. All functions in Equation (21) have to be evaluated at the maximum point of ν. In Appendix A.2, we show that even in this case, the Doppler shift error can be calculated analytically and the result is

$$\sigma_z^2 = \left(\frac{1}{-\nu\nu_{zz}}\right)\frac{1}{1 + \sum_\alpha \frac{s_\alpha^2}{\lambda_\alpha}}, \tag{22}$$

where the sum of always positive numbers in the denominator is related to the total error of the template fit caused by the

uncertainties of the template parameters. This result is also very intuitive as it expresses that optimizing the template parameters —in comparison to optimizing for the Doppler shift only with a fixed template—always decreases the error of the Doppler shift. In general, the more free parameters the stellar template has, the smaller uncertainty the Doppler shift can be estimated with.

4.4. Linear Model for Continuum and Flux Correction

Since the flux calibration of observed spectra is often affected by unknown systematics, a wavelength-dependent, nonlinear correction function is typically used to correct for fluxing errors, incorrect dereddening, discrepancies between the templates and real continua, etc. The most practical, generic flux correction function is a low-order polynomial in the form of $A(\lambda) = \sum A_n q_n(\lambda)$. The correction function does not depend on the Doppler shift, which means that we always correct the observed flux, even when the difference between the templates and the observations comes from the incorrect templates. This is because correcting the flux of the template would require Doppler shifting the correction function, which we want avoid to keep the math simple.

Introducing a wavelength-dependent flux correction results in the log likelihood of

$$\mathcal{L}(\boldsymbol{A}, z) = \boldsymbol{A}^T \boldsymbol{\varphi} - \frac{1}{2} \boldsymbol{A}^T \boldsymbol{\chi} \boldsymbol{A}, \qquad (23)$$

where we again omitted the additive constants and $\boldsymbol{A}$, $\boldsymbol{\varphi}$, and $\boldsymbol{\chi}$ are now vectors and matrices, which take the form

$$\varphi_k(z, \theta) = \sum_p q_{kp} \frac{f_p m_p(z, \theta)}{\sigma_p^2}, \qquad (24)$$

$$\chi_{kn}(z, \theta) = \sum_p q_{kp} \, q_{np} \, \frac{m_p^2(z, \theta)}{\sigma_p^2}, \qquad (25)$$

where q_{kp} are the linearly independent set of vectors introduced earlier. The linear independence of the q_{kp} vectors will ensure that the matrix χ is invertible and $\boldsymbol{A}$ can be calculated as

$$\boldsymbol{A} = \boldsymbol{\chi}^{-1} \boldsymbol{\varphi}. \qquad (26)$$

We can now write down the curvature matrix in the block form of

$$
C
$$
$$
= \begin{pmatrix}
-\boldsymbol{\chi} & \boldsymbol{\varphi}_z - \boldsymbol{\chi}_z \boldsymbol{A} & \boldsymbol{\varphi}_\beta - \boldsymbol{\chi}_\beta \boldsymbol{A} \\
\boldsymbol{\varphi}_z^\top - \boldsymbol{A}^\top \boldsymbol{\chi}_z & \boldsymbol{A}^\top \boldsymbol{\varphi}_{zz} - \frac{1}{2} \boldsymbol{A}^\top \boldsymbol{\chi}_{zz} \boldsymbol{A} & \boldsymbol{A}^\top \boldsymbol{\varphi}_{z\beta} - \frac{1}{2} \boldsymbol{A}^\top \boldsymbol{\chi}_{z\beta} \boldsymbol{A} \\
\boldsymbol{\varphi}_\alpha^\top - \boldsymbol{A}^\top \boldsymbol{\chi}_\alpha & \boldsymbol{A}^\top \boldsymbol{\varphi}_{\alpha z} - \frac{1}{2} \boldsymbol{A}^\top \boldsymbol{\chi}_{z\alpha} \boldsymbol{A} & \boldsymbol{A}^\top \boldsymbol{\varphi}_{\alpha\beta} - \frac{1}{2} \boldsymbol{A}^\top \boldsymbol{\chi}_{\alpha\beta} \boldsymbol{A}
\end{pmatrix},
$$
$$(27)$$

where z in the indices of $\boldsymbol{\varphi}$ and $\boldsymbol{\chi}$ means element-wise differentiation with respect to the Doppler shift and the Greek letters in the index mean differentiation by each of the template parameters θ.

The corresponding significance function depends only on the Doppler shift and the template parameters, and can be defined as

$$\frac{\nu^2(z, \theta)}{2} = \boldsymbol{\varphi}^T \boldsymbol{\chi}^{-1} \boldsymbol{\varphi}. \qquad (28)$$

This is still a scalar function with no explicit dependence on the flux correction coefficients. It has the maximum at the same values of z and θ where the likelihood function does. We can now write down the Fisher matrix in a block matrix form as

$$
F = \begin{pmatrix}
\boldsymbol{\chi} & -\boldsymbol{\varphi}_z & -\boldsymbol{\varphi}_\beta \\
-\boldsymbol{\varphi}_z^T & -\nu \nu_{zz} + \boldsymbol{\varphi}_z^\top \boldsymbol{\chi}^{-1} \boldsymbol{\varphi}_z & -\nu \nu_{z\beta} + \boldsymbol{\varphi}_z^\top \boldsymbol{\chi}^{-1} \boldsymbol{\varphi}_\beta \\
-\boldsymbol{\varphi}_\alpha^T & -\nu \nu_{\alpha z} + \boldsymbol{\varphi}_\alpha^\top \boldsymbol{\chi}^{-1} \boldsymbol{\varphi}_z & -\nu \nu_{\alpha\beta} + \boldsymbol{\varphi}_\alpha^\top \boldsymbol{\chi}^{-1} \boldsymbol{\varphi}_\beta
\end{pmatrix}, \qquad (29)
$$

where all expressions have to be evaluated at the parameters that maximize the significance function. According to a calculation similar to the previous case described in Section 4.3, the uncertainty of the Doppler shift will be the same as in Equation (22).

The Fisher matrix in Equation (29) can be evaluated numerically by calculating the Hessian of the significance function around the maximum, as well as the element-wise first partial derivatives of the vector $\boldsymbol{\varphi}$. We must point out that the numerical differentiation happens with respect to z and the template parameters only and not with respect to the flux correction coefficients. This is advantageous, since algorithms for numerical evaluation of the Hessian typically scale as $\mathcal{O}(D^2)$, where D is the number of parameters, but adaptive algorithms that attempt to reduce the errors of the differentials can scale even worse. Calculating the Fisher matrix using our formalism efficiently reduces the number of necessary function evaluations as there is no differentiation with respect to the flux correction coefficients. Yet, the inverse of Equation (29) yields the errors and covariances of the flux correction coefficients, as well as the template parameters θ and z. When twice differentiating the significance functions with respect to v_{los} and the three fundamental template parameters, computation of the Hessian still requires several hundred function evaluations using the numdifftools Python package.

4.5. Bayesian Formalism with Priors

When fitting stellar spectra with templates with unknown parameters, one can often benefit from a priori knowledge about the spectral type of the observed object. For example, photometric data can constrain the effective temperature and provide some limits on the surface gravity. To incorporate this into a probabilistic model, most often the Bayesian posterior probability is considered, which, in the generic case of optimizing for the template parameters and the flux correction coefficients, can be written as

$$p(\boldsymbol{A}, z, \theta \mid D) = \frac{L(\boldsymbol{A}, z, \theta) p(\boldsymbol{A}) p(z) p(\theta)}{\int L(\boldsymbol{A}, z, \theta) p(\boldsymbol{A}) p(z) p(\theta) \, d\boldsymbol{A} \, dz \, d\theta}, \qquad (30)$$

where D is a symbol for the observations and $p(z)$ and $p(\theta)$ are the prior probability distributions defined on the Doppler shift and the atmospheric parameters, respectively. We will soon see that the prior $p(\boldsymbol{A})$ on the flux correction coefficients must be flat in order to follow the same formalism as in Section 4.4. The integral in the denominator of Equation (30) is the usual normalization constant that appears in Bayesian posteriors. It is not necessary to calculate it, as the location of the maximum of the posterior does not depend on it, and Monte Carlo sampling of $p(z, \theta \mid D)$ requires a function that is only proportional to the posterior.

The location of the maximum of Equation (30) coincides with the location of the maximum of its logarithm, hence, after dropping the additive constants that come from the denominator or elsewhere, the log posterior will be the sum of the log likelihood and the logarithm of the priors as

$$\mathcal{P}(\boldsymbol{A}, z, \theta \mid D) = \mathcal{L}(\boldsymbol{A}, z, \theta) + \pi(z) + \pi(\theta), \qquad (31)$$

where we denoted the logarithm of the priors with $\pi(z)$ and $\pi(\theta)$ and already assumed that $p(\boldsymbol{A})$ is flat, hence its partial derivatives are 0. The location of the maximum of Equation (31) can be found by equating its partial derivatives to 0, as before, in the case of the maximum likelihood method. Since $\pi(\boldsymbol{A}) = 0$, we can write down the same system of equations for A as in Equations (5) and (6) but φ and χ, as well as the significance function ν, as defined in Equation (28), will also depend on the θ model parameters. Similarly to Equation (11), at the optimum of the amplitude, the log posterior can be rewritten as

$$\mathcal{P}(\boldsymbol{A}_0, z, \theta \mid D) = \frac{1}{2} \nu^2(z, \theta) + \pi(z) + \pi(\theta), \qquad (32)$$

where the 0 index expresses that the posterior and the significance function are to be taken at the optimum with respect to the flux correction coefficients $\boldsymbol{A}$. The Doppler shift and template parameters maximizing the posterior can be found using the same nonlinear maximization methods as in the case of Equation (11). When the priors on z and each of the θ parameters are independent, no mixed second derivatives of them will appear in the Fisher matrix, which takes the form of

$$
F = \begin{pmatrix}
\chi & -\varphi_z & -\varphi_\beta \\
-\varphi_z^T & -\nu\nu_{zz} + \varphi_z^\top \chi^{-1}\varphi_z - \pi_{zz} & -\nu\nu_{z\beta} + \varphi_z^\top \chi^{-1}\varphi_\beta \\
-\varphi_\alpha^T & -\nu\nu_{\alpha z} + \varphi_\alpha^\top \chi^{-1}\varphi_z & -\nu\nu_{\alpha\beta} + \varphi_\alpha^\top \chi^{-1}\varphi_\beta - \pi_{\alpha\beta}
\end{pmatrix}. \qquad (33)
$$

In practice, we find that the variances of the flux correction coefficients are several orders of magnitude smaller than the variances of the Doppler shift and the template parameters. The same is true for the covariances between the flux correction coefficients and the template parameters, including the Doppler shift. For example, Walker et al. (2015) solved a stellar template-fitting problem by Monte Carlo sampling the posterior of a hierarchical Bayesian model very similar to Equation (30). They treated the flux correction coefficients as random variables and found no correlations between the variances of the coefficients and the rest of the model parameters including z and the atmospheric parameters. It is general practice, therefore, to determine the covariance matrix of the uncertainty of the line-of-sight velocity and the template parameters by taking the flux correction coefficients at the location of the maximum significance and calculating the covariance matrix from the inverse of the Hessian of the likelihood function with respect to z and θ only.

5. Algorithms

We have developed our own spectral template-fitting software library that integrates with the observation simulation code and implements the significance maximization formalism outlined in Section 4.1, and, in more detail, in Appendix A.3.

Our template-fitting algorithm is straightforward but we have two important goals: precision and performance. In order to fit v_{los} with subpixel precision, we decided to use templates with as high resolution as possible and perform the LSF convolution at the original resolution of the synthetic spectra. High-performance implementations of both the simulation code and the fitting were necessary to be able to execute the large number of simulations and fit the simulated spectra in many configurations, including full Monte Carlo sampling of the posterior distribution. In particular, we focused on high-performance convolution and synthetic spectrum grid interpolation.

To evaluate the significance function, the following procedure is followed.

1. The synthetic spectrum grid is interpolated to the given values of the atmospheric parameters.
2. The interpolated template is shifted to the given value of z.
3. The template is convolved with the LSF.
4. The template is interpolated to the pixels of the observed spectrum, for each exposure.
5. The quantities χ, φ, and the logarithm of the priors are evaluated.
6. The value of $\chi^{-1}\phi$ is calculated.

Step 3 of the algorithm can be performed in conjunction with Step 1 to optimize performance, as we will explain in the next section.

5.1. Convolution with the Line-spread Function

When simulating spectra, as well as when processing the templates while fitting the simulated observations, we convolve the synthetic spectra with the LSF at the original resolution of the synthetic spectrum grid, followed by resampling to the detector pixels. It is important to point out that the LSF we use already accounts for the pixelization; consequently, resampling consists of a simple linear interpolation of the LSF-convolved high-resolution template to the center of the pixels instead of flux-conserving rebinning, which would take pixelization into account twice. As a consequence, convolution, and interpolation are not interchangeable and the convolution must be done first, in the high-resolution representation.

The convolution of templates with a wavelength-dependent LSF is implemented using a data compression method based on principal component analysis (PCA). This algorithm is applicable when the template spectrum uses regular binning, either linear or logarithmic. To calculate the convolution of a pixelated spectrum f_p with a kernel $k(\lambda)$ that has a nonlinear dependence on wavelength, one wants to evaluate

$$\hat{f}_p = [f \star k(\lambda)]_p = \sum_{n=-N}^{N} f_p [k(\lambda_p)]_{p-n}, \qquad (34)$$

where $[k(\lambda_p)]_i$ is the kernel evaluated at the wavelength λ_p of pixel p and discretized on the pixel grid around the central pixel. The value N is the half kernel size necessary to evaluate the convolution with some prescribed precision, i.e., the kernel either has a finite support or its value is negligible outside the $[-N, N]$ interval. The kernel $k(\lambda)$ is usually a high-order function and calculating the convolution requires at least $P \times (2N + 1)$ function evaluations, where P is the number of pixels. As a consequence, when convolving many spectra with the same kernel, pretabulating the kernel around each pixel is

an obvious way of optimization. We denote the pretabulated kernel by $[k(\lambda_p)]_i$, around pixel p where the i index runs over the neighboring pixels. A downside of pretabulated kernels, especially when the synthetic spectra are high resolution, is that convolution algorithms require shuffling around very large arrays in memory. Here we show how to leverage the linearity of the convolution operation to reduce the variable-kernel convolution into the sum of a few fixed-kernel convolutions, which has very fast implementations available.

Let us evaluate $[k(\lambda_p)]_i$ for each pixel p and shift $i \in [-N, N]$ and express the variable kernel as a linear combination of some basis kernels K_{pi} in the form of

$$K_{pi} = [k(\lambda_p)]_i = \sum_j a_{pj} E_{ji}, \tag{35}$$

where a_{pj} are wavelength-dependent coefficients and E_{ji} is a matrix consisting of some orthogonal basis vectors. To find the most optimal basis, we apply PCA to K_{pi} and factorize its product with its transpose using singular value decomposition, which yields

$$K_{ip} K_{pj}^\top = \sum_l E_{il} \Sigma_l^2 E_{lj}, \tag{36}$$

where the E_{il} matrix consists of the eigenvectors of $\boldsymbol{KK}^\top$ and $\boldsymbol{\Sigma}$ is a diagonal matrix of its eigenvalues. We can then determine the T_{pi} principal components of the kernel at each pixel as $T_{pi} = \sum_j K_{pj} E_{ji}$ and reconstruct the kernel from its principal components as

$$K_{pi} = \sum_j T_{pj} E_{ji}. \tag{37}$$

The sum over j in Equation (37) goes over $2N + 1$ values, the width of the tabulated kernel. Data compression in PCA is achieved by limiting the sum in Equation (37) to only a few eigenvectors belonging to the largest eigenvalues of $\boldsymbol{KK}^\top$. This data compression technique is lossy, but in practice, varying width Gaussian kernels can be compressed into the first 5–10 principal components with an accuracy of 10^{-5}. Compare this to the typical kernel width of $N > 50$ when convolving down the high-resolution templates to the low-resolution PFS instrument.

Most importantly, the first few T_{pj} principal components can be interpolated to any intermediate wavelength or fitted with low-degree polynomials as a function of wavelength and evaluated at any wavelength. Hence, the kernel can be reconstructed with high accuracy at any wavelength—as long as the spectral bins remain regular.

To evaluate the convolution of Equation (34) in this compressed representation, one has to compute

$$\hat{f}_p = \sum_j \sum_{n=-N}^{N} T_{pj} f_p E_{nj}. \tag{38}$$

When the kernel is compressed with PCA, the sum over j in Equation (38) goes over only a few values, hence we replaced the expensive array reshuffling necessary to evaluate the convolution using a pretabulated kernel with a few fast, optimized fixed-kernel convolutions and a summation.

5.2. Synthetic Grid Interpolation Combined with Convolution and Caching

We chose our input models of the simulations to have atmospheric parameters that are on the grid points of the synthetic spectrum grid, but when fitting the simulated observations with free template parameters, interpolation of the synthetic stellar grid is inevitable. At the same time, when working with high-resolution spectra, one needs to make flux interpolation as fast as possible. By ruling out higher-order methods for performance reasons, we opted for a simple linear interpolator, which only takes the 2^D spectra in the corners of a grid cell into account, where D is the number of grid parameters.

The simple linear interpolations of the synthetic spectrum grid offers several optimization opportunities, especially when the convolution step can be combined with the interpolation step. One has to recognize that iterative optimization algorithms, Monte Carlo samplers, and numerical Hessian computation methods take very small steps in the direction of each parameter around the most optimal parameters. Consequently, once convergence has been reached, interpolation will only happen within the very same grid cell. (Unless, of course, the best-fit parameters fall onto a grid cell boundary.) Precomputing and caching the LSF-convolved synthetic spectra at the grid points surrounding the cell and interpolating the convolved templates can save a significant amount of processing. On the other hand, convolution with the LSF is normally done on the template only *after* it is shifted to some nonzero, but typically small, line-of-sight velocity during template fitting. Caching and reusing the LSF-convolved templates for interpolation is only possible when the wavelength dependence of the LSF is weak compared to the wavelength shifts and the order of the convolution and wavelength shift operations can be exchanged.

5.3. Maximizing the Significance Function

We find the maximum of the significance function using the classical Nelder–Mead method. When applying a polynomial flux correction to the observed spectra, one has to evaluate the significance function according to Equation (28) which involves a matrix inversion. Instead of computing the inverse matrix χ^{-1}, it is numerically much more stable to solve the equation $\chi \boldsymbol{x} = \varphi$ for the vector $\boldsymbol{x}$, which will yield $\boldsymbol{x} = \chi^{-1} \varphi$. Numerical stability of this operation also dictates that the flux values of the template spectrum have to be scaled by a factor to match the interval of typical flux values of the observation.

When template fitting is done in a Bayesian setting and priors are defined on the template parameters, the posterior probability distribution of the parameters is given by Equation (32). The maximum a posteriori solution can be found the same way as the maximum significance solution.

5.4. Uncertainty Estimators of v_{los}

In Section 4.2, we derived an analytic formula for the uncertainty of the line-of-sight velocity. The error estimator based on the Fisher matrix, also called the asymptotic error, gives only a lower bound on $\sigma(v_{los})$ and this limit is only achievable in practice when the likelihood function is strictly Gaussian about the maximum and has no local maxima in the few σ vicinity of the global maximum. This is rarely the case when dealing with low-S/N spectra, as noise "smears" the likelihood function so that its maximum will be smaller and the curvature at the maximum larger, increasing the influence of

nearby local maxima. Hence, precise characterization of $\sigma(v_{los})$ in low-S/N situations requires Monte Carlo sampling of the significance function. We are going to demonstrate this in Section 6.

In addition to the asymptotic uncertainty estimator derived from the Fisher matrix, we also determine the full covariance matrix of the parameters from full Monte Carlo sampling of the Bayesian posterior probability distribution when v_{los}, as well as the template parameters, are treated as unknown.

When we present our results in Section 6, we are going to compare the uncertainty estimates from the Fisher matrix and Monte Carlo sampling to those that we measure from an ensemble of repeated simulations.

5.5. Monte Carlo Sampling of the Posterior

We implemented an adaptive Monte Carlo algorithm to generate samples of the significance function or the Bayesian posterior probability distribution. The sampler works by alternating between v_{los} and the rest of the parameters. In case of high-S/N spectra when v_{los} is well constrained, this Gibbs-like sampling resulted in better mixing of the Monte Carlo chains than sampling all parameters together. The proposal distributions of both semisteps are initialized from the covariance matrix of the parameters, as determined from the maximum a posteriori template fitting. The initial states were generated from the maximum a posteriori best-fit parameters by perturbing them with a random vector drawn from a multivariate Gaussian distribution with the same covariance matrix as the step proposal. Initializing the state and the step proposal distributions this way allowed us to minimize the length of the burn-in phase and reach a high acceptance rate sooner.

6. Results

We generated 1000 realizations of the simulated observed spectra for each stellar type listed in Table 2 and fitted them using different combinations of the spectrograph arms and template spectra, as well as various different configurations of the template-fitting algorithm. In the case of maximum likelihood and maximum a posteriori fitting, we calculated Δv_{los} as the difference between the input line-of-sight velocity and the best velocity estimate obtained by fitting the templates to the simulations.

The results we present here do not account for some instrumental systematics, such as the wavelength calibration. As a result, velocity uncertainties below the typical value equivalent to a hundredth of a detector pixel (approximately 0.6 km s^{-1} in the MR arm) are unrealistic. Nevertheless, we plot these results as they establish the theoretical lower limit on the uncertainty of v_{los}. The instrumental error floor is considered to be due to systematics and wavelength calibration errors, neither of which were taken into account when we ran the simulations. We are going to discuss this issue in some detail in Section 6.3.

When simulating the observations, we took the typical fluxing errors into account by multiplying the flux with a random, slowly changing function that alters the flux by about 2%, see Section 3. In addition to this, we applied the reddening law of Cardelli et al. (1989) with a randomly chosen value of A_V between 0 and 0.5 and but we do not explicitly correct for this reddening when fitting the templates. Instead, we use a fifth-order Chebyshev polynomial over the entire wavelength

Table 4
Parameters of the Priors Used to Fit the Simulations When the Template Parameters Are Treated as Unknowns

Parameter	Distribution	Parameters	Limits	Unit
[Fe/H]	truncated normal	$\sigma = 0.5$	± 1.0	dex
T_{eff}	truncated normal	$\sigma = 50$	± 250	K
$\log g$	truncated normal	$\sigma = 1.0$	± 2.0	⋯

Note. The probability distributions are always centered on the input parameters of the simulations and the limits are calculated with respect to the input parameter values.

range to correct both the fluxing error and reddening, as described in Section 4.

We fitted the simulated spectra with templates using various configurations. In the most optimistic scenario, the atmospheric parameters of the template were chosen to match the input parameters. In case the atmospheric parameters of the template are also optimized for, instead of solving the maximum significance problem, we performed a full Bayesian maximum a posteriori fit to the simulations. The priors used in the Bayesian model are normally distributed in a finite interval centered on the input parameters, as summarized in Table 4.

6.1. Signal-to-noise Ratio

Throughout this section, we adopt the following definition of S/N, denoted with S_p/N_p, when talking about the per pixel S/N of a simulated spectrum. The signal is taken as the flux in a pixel of the noiseless spectrum, convolved down to the resolution of the instrument and resampled to the pixels of the detector. The noise is taken as the standard deviation of the Gaussian noise calculated from the observational parameters, the sky model, and the instrumental model, as described in Section 3.

To characterize the S/N of an entire spectrum by a single S/N number, we take the 95th percentile of the per pixel S_p/N_p values. The advantage of using a high percentile to calculate the overall S/N as opposed to the mean or the median is that the 95th percentile is less sensitive to the noisy pixels of the deep absorption lines and more sensitive to the continuum pixels.

Characterizing a spectrum with a single S/N value has its caveats, however. In Figure 2, we plot S_p/N_p for three stellar spectra that show different levels of absorption in the MR arm. In the plots, S_p/N_p is normalized to have a 95th percentile of 1 pixel^{-1}, indicated by the red line, and the median of S_p/N_p is plotted in blue. While the 95th percentile S/N traces the highest-S/N pixels of the continuum remarkably well in the case of the higher-temperature models, it is no longer a good measure when there is no well-defined continuum, as in the case of the cool M giant in the top panel of Figure 2. On the other hand, using the maximum of S_p/N_p in place of the 95th percentile to characterize a spectrum would not be robust enough.

Stoehr et al. (2008) suggested measuring the S/N from a single realization of a noisy spectrum in the form of

$$S/N = \frac{\sqrt{6}}{1.482602} \cdot \frac{\mu(f_p)}{\mu(|2f_p - f_{p-2} - f_{p+2}|)}, \quad (39)$$

where μ denotes the median and f_p is the flux in pixel p. This is a robust way of estimating the S/N from spectra when the signal and the variance are not known individually, only the
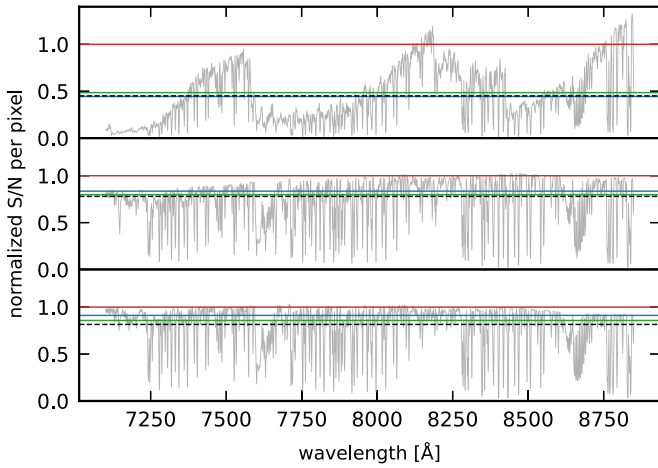
Figure 2. S/N per pixel for three different stellar types (from top to bottom: solar metallicity M giant, metal-poor M dwarf, and metal-poor K giant). The 95th percentile S/N is plotted in red and the median S/N in blue, whereas the dashed black line indicates the S/N as defined by Stoehr et al. (2008). See the text for discussion.

noisy flux. For reference, we plot this quantity in Figure 2 as well, noting that it is subject to the same dependence on spectral type as the mean or the median, and hence cannot characterize well the typical S/N of continuum pixels.

For easy comparison, we will refer to the 95th percentile S/N as measured in the MR arm, even when plotting the results for different spectrograph arms, or a combination of arms. Whenever the S/N is expressed per resolution element, as opposed to per pixel, we calculate it by multiplying the 95th percentile value of the per pixel S/N by the square root of the typical number of pixels per resolution element. Based on Table 1, the multipliers are $\sqrt{3}$, $\sqrt{3}$, and $\sqrt{4}$ for the B, R, and MR arms, respectively.

6.2. Empirical Error of v$_{los}$ from Repeated Simulations

Figure 3 shows the primary result of this work: the standard deviation of Δv_{los}, calculated in logarithmic bins of S/N per resolution element. S/N is quoted for the single arm MR configuration, even when the curves correspond to the R arm or a combination of two arms. We will refer to the standard deviation of Δv_{los} as the empirical error and simply denote it with $\sigma(v_{los})$. We determined the empirical errors for fitting each spectrograph arm separately (thin curves) and in the blue and red arms in combination (thick curves). The colors represent the arms used for fitting B (blue curves), R (red curves), and MR (gray curves). We verified that the slope and y-intercept of these curves are insensitive to the binning in S/N: using logarithmic bins in S/N yields very similar curves regardless of the number of bins, see Figure 8.

The curves in Figure 3 show the general trend of $\sigma(v_{los})$ decreasing with the square root of S/N, which is expected based on what we calculated in Section 4.2 from the Fisher matrix. Deviations are visible from this trend at low S/N in most cases and at high S/N in case of the M giant. We discuss these deviations in Section 6.3.

6.3. Error Estimators of v$_{los}$ for Single Spectra

In the previous section, we looked at the empirical error of v_{los} as a function of S/N, which was calculated from a large number of simulations as the standard deviation of Δv_{los} in

logarithmic bins of S/N. Let us now compare the empirical error to uncertainty estimators that can be determined for a single simulation or a nonrepeated observation, such as the one we calculated in Section 4.2 from the Fisher information matrix or from Monte Carlo sampling of the likelihood function or the Bayesian posterior.

To compare the various estimators of uncertainty, we fitted simulated spectra of a low-metallicity G subgiant, typical of stars found in dSph galaxies. We performed the fitting in two configurations, in order to test the effect of an uncertain wavelength calibration. First, we assumed perfectly matching wavelength solutions for each of the 12 exposures in both the B and MR arms. Second, we kept the flux in each pixel, but multiplied the wavelength values of each exposure with a random Doppler shift sampled from a normal distribution with a standard deviation of 25 km s^{-1}. Since the wavelength errors simulated this way are independent random variables, they are expected to average out according to $\sqrt{N_{exp} \cdot N_{arm}}$. Consequently, the amplitude of the error was chosen to be very large to emphasize the effect. In both cases, with and without wavelength errors, we optimized for the three atmospheric parameters, and a fifth-order Chebyshev polynomial was used to correct for the simulated systematics of the fluxing. The template was fitted using a maximum finder algorithm, as well as Monte Carlo sampling of the posterior probability distribution.

In panels (a) and (b) of Figure 4, we plot the empirical error (solid red line) calculated from an ensemble of simulations, in logarithmic bins of S/N, the asymptotic error derived from the Fisher matrix (blue dots) for each simulation and the standard deviation of Monte Carlo samples drawn from the posterior probability distribution (red dots), also for each simulation, as functions of the S/N for the case of perfect wavelength calibration (left) and with simulated wavelength errors (right). In the case of perfect wavelength calibration, only at low S/N, and in the case of simulated wavelength error, both at low and high S/Ns, obvious deviations of the empirical error from the asymptotic error are visible.

At low S/N, the empirical error diverges sharply from the asymptotic error, which could indicate either that the optimization algorithm did not find the real maximum of posterior, or that the logarithm of the posterior is no longer approximated well by a quadratic function around its maximum. However, the error estimates determined from the covariance matrix of the Monte Carlo samples are consistent with the empirical error at low S/N, which indicates that the maximum finder worked as expected but instead, the posterior becomes non-Gaussian around the mode. To investigate the Gaussianity of the posterior, we plot the skewness and kurtosis of the Monte Carlo samples drawn from the Bayesian posterior in panels (c)–(f) of Figure 4. These indicate that the posterior—or more precisely its marginalization over all parameters except v_{los}—is indeed Gaussian at higher S/N, but can become non-Gaussian at low-S/N values.

At high S/N, when perfect wavelength calibration is assumed, the empirical error and the asymptotic error remain consistent and follow the same $(S/N)^{-1}$ scaling as at intermediate S/N. This result is unrealistic as the practical error floor of fiber-fed spectrographs is thought to be around a hundredth of a pixel, which is equivalent to 0.6 km s^{-1} for the MR arm. On the other hand, the result shows that, despite the pixelization of the spectrum, Doppler shift measurements could theoretically be better than the 1/100 pixel limit if there was a way to more precisely calibrate the instruments. When we
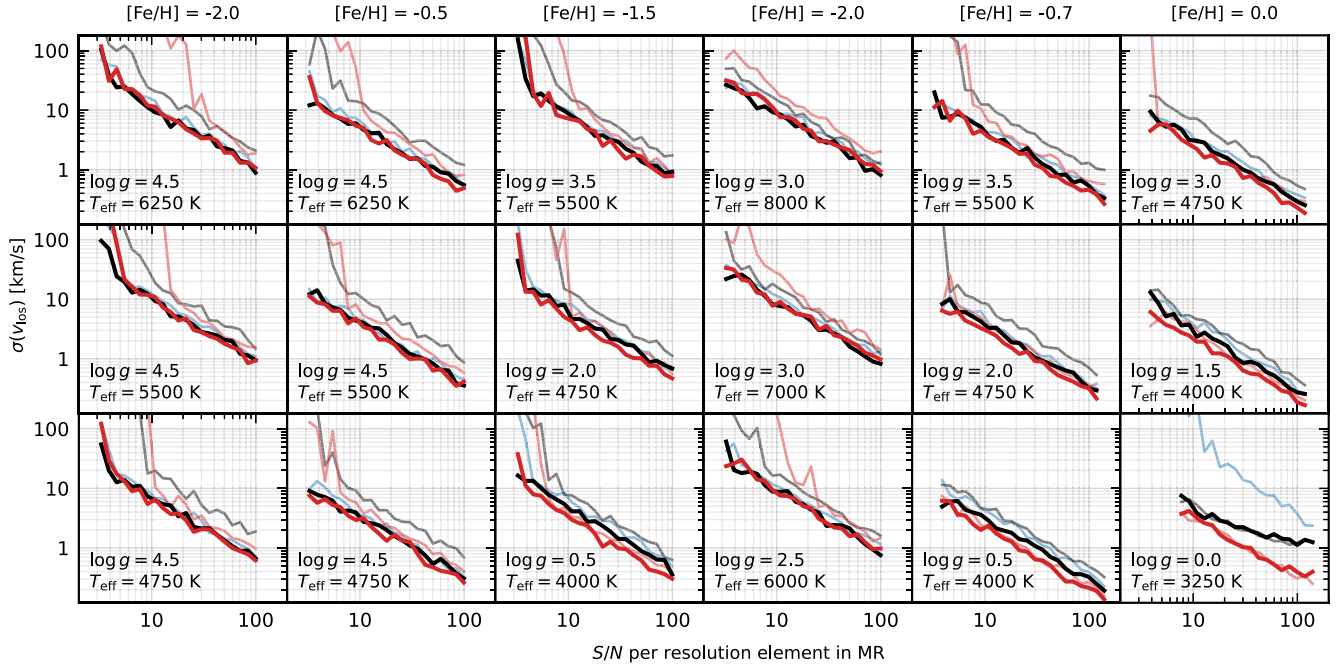
Figure 3. Comparison of the typical empirical line-of-sight velocity error, without any instrumental systematics, for the models listed in Table 2 using the Subaru PFS B (thin blue curves), R (thin red curves), and MR (thin gray curves) arms, as well as the combination of B + R (thick black curves) and B + MR (thick red curves). The curves show the standard deviation of the line-of-sight velocity error $\Delta v_{\rm los}$, with respect to the model input, in logarithmic bins of S/N. For purposes of easy comparison, the values of S/N on the horizontal axes are always the 95th percentile S/N in the MR arm only, even when the fitting was done using different arms or a combination of arms.

include the effect of incorrect wavelength calibration, the empirical errors calculated from multiple simulations show the expected error floor, as visible in the top right panel of Figure 4. Neither the asymptotic error, nor the error from Monte Carlo sampling, is capable of accounting for the error floor, however. This is the result of the simplification of the likelihood function, which only accounts for the uncertainty of the observed flux but has no additional error terms to account for the imprecise wavelengths. We will address this issue in an upcoming paper.

6.4. Effect of Mismatched Template Parameters

So far we have seen how the uncertainty of line-of-sight velocity measurement depends on the S/N of the observed spectrum when fitting templates with atmospheric parameters that match the input parameters of the simulated spectra. Let us now study the case when the template parameters are far off the real values. This is a realistic scenario because synthetic spectra are precomputed on grids of the parameters and the atmospheric parameters of observed stars seldom coincide with the grid points. On the other hand, we do not take into account the imperfections of the models because we generated the simulated observations from the same synthetic spectrum grid from which we take our templates.

In Figure 5, we plot the uncertainty and bias of $v_{\rm los}$ as a function of S/N determined using template spectra with different atmospheric parameters. The curves show the standard deviation (top panels) and the mean (bottom panels) of $\Delta v_{\rm los}$ for many repeated simulations. The observations were simulated with artificially introduced fluxing errors, as described in Section 3.3 and we repeated fitting $v_{\rm los}$ with and without correcting for fluxing errors. The tests were performed on simulated observations of a model with atmospheric parameters [M/H] = −1.5, $T_{\rm eff}$ = 4750 K, and log g = 2.0, whereas the templates were

chosen to be one or two grid points away in the direction of each parameter (but only one parameter at a time). Step sizes of $\Delta T_{\rm eff} = \pm 200$ K, $\Delta [{\rm M/H}] = \pm 1.0$ dex, and $\Delta \log g = \pm 2.0$ were used in effective temperature, metallicity, and surface gravity, respectively.

The two top panels of Figure 5 show $\sigma v_{\rm los}$ from fitting the different templates with and without the polynomial flux correction. The curves overlap remarkably well, which indicates that using mismatched templates has very little effect on the uncertainty of $v_{\rm los}$. In general, fitting the simulated spectra with a template that matches the input atmospheric parameters, or letting the parameters vary freely, yields the best results in terms of bias and error variance while nonmatching templates tend to cause a bias of 1–2 km s^{-1}, even at higher S/N, which is consistent with earlier results (Walker et al. 2015). Templates with mismatched metallicity have the largest detrimental effect on the precision of $v_{\rm los}$ measurements.

Slight deviations from the exactly matching template (blue curves) are visible only at S/N > 50 where using certain templates resulted in an overestimate of the uncertainty. However, as we showed in Section 6.3 and illustrated in Figure 4, at such high S/N, errors of the wavelength calibration already start to dominate the uncertainty, which we did not take into account in the simulations.

A stronger effect on the bias of $v_{\rm los}$ is visible in the two bottom panels of Figure 4 when using mismatched templates to fit the line-of-sight velocity. Using a polynomial for flux correction tends to eliminate the bias, as it can be seen in the bottom right panel of Figure 4. It is important to point out, however, that even the bias from fitting with a perfectly matching template (blue curves) shows visible fluctuation at lower S/Ns, which we attribute to the small number of simulations.
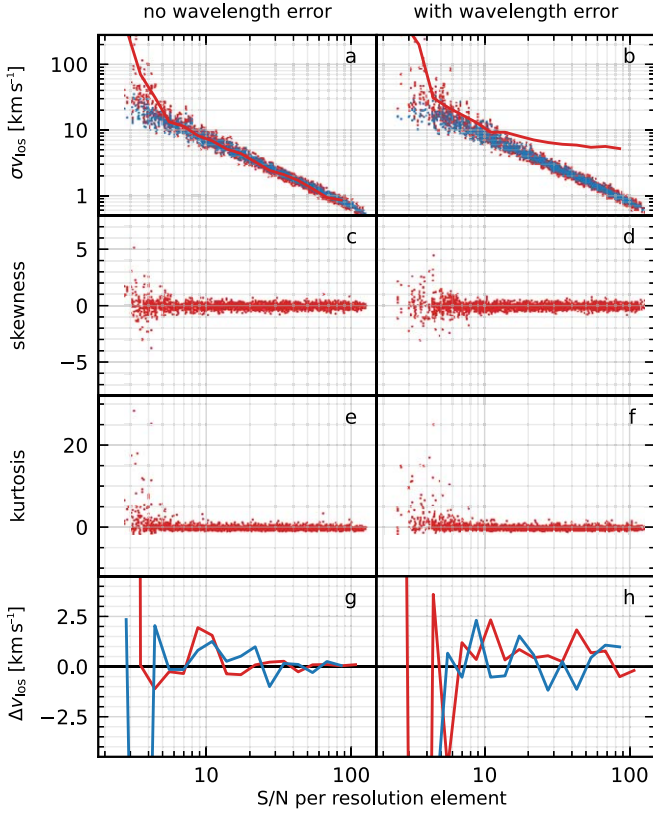
no wavelength error   with wavelength error

Figure 4. Comparison of the error estimates and bias for a model with atmospheric parameters [M/H] = −1.5, T_{eff} = 5500 K, and log g = 3.5, for the case of fitting v_{los} plus the atmospheric parameters, both with and without simulated wavelength calibration error when fitting of arms B + MR in combination. In panels (a) and (b), the blue dots indicate the asymptotic error calculated from the Fisher matrix at the maximum value of the significance function, the red dots show the standard deviation of Monte Carlo samples drawn from the Bayesian posterior, and the red curve shows the empirical error calculated from an ensemble of simulations, all in logarithmic bins of S/N. For this particular stellar type, different error estimates start to diverge below S/N < 6 and the error estimator from the Fisher matrix is no longer a good measure of the uncertainty. Panels (c)–(f) show the skewness and the kurtosis of the Monte Carlo samples drawn from the Bayesian posterior, to demonstrate that the posterior marginalized for the atmospheric parameters is no longer Gaussian at low S/N. In panels (g) and (h), the red curve indicates the expectation value of the bias Δv_{los}, as determined from a Monte Carlo sample of the likelihood function, while the blue curve shows the same bias determined from the maximum a posteriori method. Both bias curves are calculated in logarithmic bins of S/N. The bias shows no significant trend over this S/N range, but its large scatter at low S/N is obvious. These plots are qualitatively the same for lower temperature, but more luminous, metal-poor giant stars.

6.5. Correlations of the Template Parameter Errors with σ(v$_{los}$)

In order to assess the correlations between the measurement error of v_{los} and the variance of atmospheric parameters of the best-matching template spectrum, we fitted the simulated spectra with full Monte Carlo sampling of the posterior probability. To illustrate the results of a Monte Carlo run, we used scatterplots of Δv_{los} as a function of the difference of atmospheric parameters from their input values in Figure 6 for a single simulated observation of the model with atmospheric parameters [M/H] = −1.5, T_{eff} = 4750 K, and log g = 2.0. The various colors show three different Monte Carlo chains, with 1000 samples each, which are plotted without any thinning. The good mixing of the Monte Carlo chains is apparent, as well as that practically no correlation between the
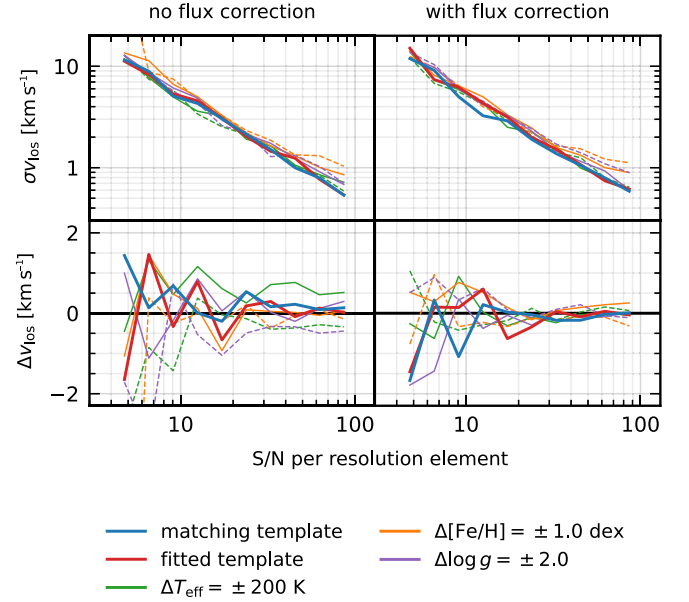

no flux correction   with flux correction

matching template   Δ[Fe/H] = ± 1.0 dex
fitted template   Δlog g = ± 2.0
ΔT_{eff} = ± 200 K

Figure 5. The variance and bias of line-of-sight velocity fits to simulations of a star with atmospheric parameters [M/H] = −1.5, T_{eff} = 4750 K, and log g = 2.0, with and without flux correction. The statistics are calculated from an ensemble of simulations in logarithmic bins of S/N. The blue curves indicate the results from fitting a matching template with the same parameters as the input model, whereas the red curves show the result from fits where we optimized for the atmospheric parameters as well. The rest of the curves, as indicated by the color given in the legend, show the results when we used templates that are one or two grid points away from the input model, on the synthetic stellar spectrum grid. Solid (dashed) lines indicate a deviation in the positive (negative) direction from the input parameters for [M/H], T_{eff}, and log g. See the text for detailed discussion.

atmospheric parameters and v_{los} exists. For reference, the Pearson correlation coefficients are printed in the top left corner of each panel.

In Figure 7, we plot the histograms of correlation coefficients between v_{los} and the error of the template parameters for all realizations of all models in Table 2 to demonstrate that the correlations are indeed very small, independently of stellar type and S/N. We also plot the distribution of the coefficient of multiple correlations, which expresses how well v_{los} can be expressed as a linear combination of the atmospheric parameters within a Monte Carlo sample drawn from the posterior. Similar to the bivariate correlation, a value much smaller than 1 means linear independence. The bivariate correlations are normally distributed around 0 with a standard deviation of about 0.1, which indicates the lack of any significant correlations. This reinforces the results presented in Section 6.4, where we observed that fitting the simulated spectra with templates that did not precisely matched the input parameters did not result in increased uncertainties of the line-of-sight velocity. This is true in a broad range of S/Ns except for high S/Ns where the effect of using a mismatched template is more clearly detectable. On the other hand, at high S/N, our simulations are inaccurate because we do not take any kind of wavelength calibration error or systematics into account.

7. Discussion

7.1. Scaling Relations of σ(v$_{los}$) and Signal-to-noise Ratio

To estimate the uncertainty of line-of-sight velocity measurements from the S/Ns of observations and the primary parameters of the spectrograph arms, Hatzes & Cochran (1992)
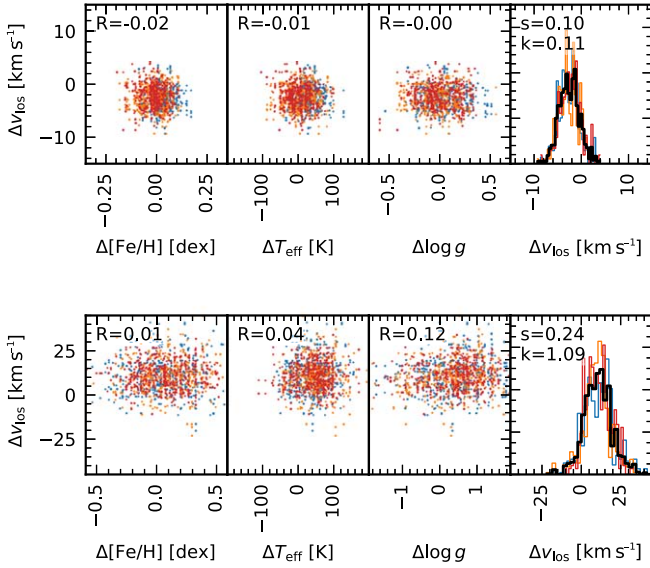
Figure 6. Results from Monte Carlo sampling of the posterior for two different realizations of the model with atmospheric parameters $[M/H] = -1.5$, $T_{eff} = 4750$ K, and $\log g = 2.0$, and S/Ns (measured per resolution element in the MR arm) of $S/N = 11.76$ (top) and $S/N = 2.33$ (bottom). The first three panels of each row show scatterplots of the deviations of the atmospheric parameters from the model input vs. Δv_{los}. The colors indicate samples from three different Monte Carlo chains. The rightmost panel of each row shows histograms of Δv_{los} for each Monte Carlo chain in color, whereas the black histogram is the combined histogram of all samples. In the top left of each scatterplot, the correlation coefficient of the two variables is printed. The small values of the Pearson correlation coefficient indicate that v_{los} does not correlate with the atmospheric parameters. In the corner of the histograms, the values of the skewness s and the kurtosis k (0 for a Gaussian distribution) of the Monte Carlo samples are indicated. Note the different scales on the axes.

introduced the formula

$$\sigma^{HC}(v_{los}) = 1.45 \times 10^6 \cdot (S/N)^{-1} \cdot R^{-1} \cdot B^{-1/2}, \quad (40)$$

where velocity is measured in units of km s^{-1}, S/N is the S/N of the flux measurement per pixel,[7] R is the dimensionless spectral resolution, and B is the total wavelength coverage in units of Å. While the formula is certainly an approximation—since it does not take important effects into account, such as the varying S/N over the wavelength coverage, nor the contrast, width, and density of the absorption lines—it captures the primary relation between the uncertainty of the line-of-sight velocity measurements and the square root of S/N.

Bouchy et al. (2001) pointed out that the information a spectral pixel carries about the Doppler shift depends on the slope of the original spectrum (after instrumental broadening) in the pixel. They provide the formula

$$\sigma^{BPQ}(v_{los}) = c \cdot \frac{1}{\sqrt{\sum W_p}}, \quad \text{where} \quad W_p = \frac{\lambda_p^2 [\partial f_p/\partial \lambda]^2}{\sigma^2(f_p)}, \quad (41)$$

to predict the measurement error of v_{los}. On the other hand, when the $[\partial f_p/\partial \lambda]^2$ slope is to be calculated numerically at pixel boundaries from neighboring spectral pixels, one has to

Figure 7. The distribution of the correlation coefficient of v_{los} and the atmospheric parameters (left three panels), and the coefficient of multiple correlation (rightmost panel), calculated from the Monte Carlo sampling of all realizations of all spectral types. According to these, the covariance of v_{los} with the template parameters is minimal on average.
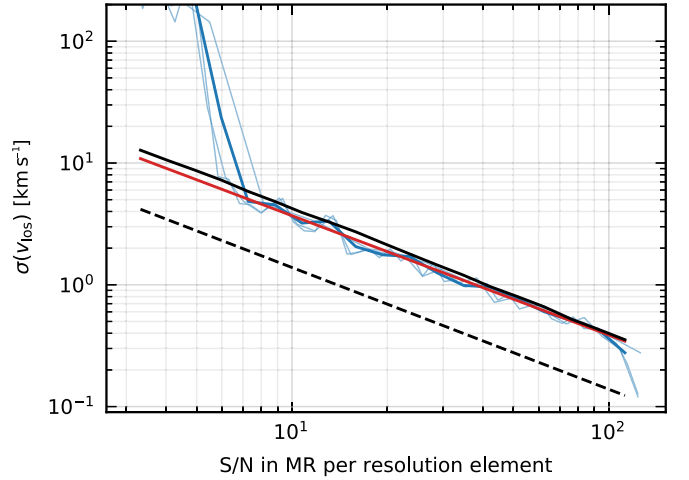
Figure 8. Comparison of $\sigma(v_{los})$ predictors of Hatzes & Cochran (1992, dashed black line, see Equation (40)) and Bouchy et al. (2001, solid black line, see Equation (41)) compared to our results from repeated simulations (blue lines) for a low-metallicity K giant with $T_{eff} = 4000$ K. The multiple blue lines indicate different logarithmic binning of Δv_{los} to calculate $\sigma(v_{los})$. The red line is a linear fit to the logarithm of the data with a slope of -0.95, very close to the theoretical value of -1.

take the covariances into account, so the symbol $\sum W_p$ becomes

$$\sum W_p = \text{Tr } \boldsymbol{w}^\top \boldsymbol{C}^{-1} \boldsymbol{w}, \quad (42)$$

where the $\boldsymbol{w}$ vector is one element shorter than f_p and takes the form of

$$w_i = \frac{1}{2}(\lambda_i + \lambda_{i+1})\left[\frac{f_{i+1} - f_i}{\lambda_{i+1} - \lambda_i}\right]. \quad (43)$$

The covariance matrix is tridiagonal and can be written as

$$C_{ij} = \begin{cases} \sigma^2(f_i) + \sigma^2(f_{i+1}) & \text{if } i = j, \\ \sigma^2(f_i) & \text{if } j = i + 1 \text{ or } i = j + 1. \end{cases} \quad (44)$$

In Figure 8, we plot the uncertainty estimates of v_{los} calculated from the formulae of Hatzes & Cochran (1992) and Bouchy et al. (2001), comparing their results to the empirical errors determined from repeated simulation of the spectrum of a low-metallicity K giant with $T_{eff} = 4000$ K.

Here we introduce a new quantity, the effective line number, which can be calculated from synthetic spectra without identifying the individual lines. We are going to use the

---

[7] The exact definition of S/N is unspecified by Hatzes & Cochran (1992). Hence, we assume it was set to the same in every pixel of their simulations.
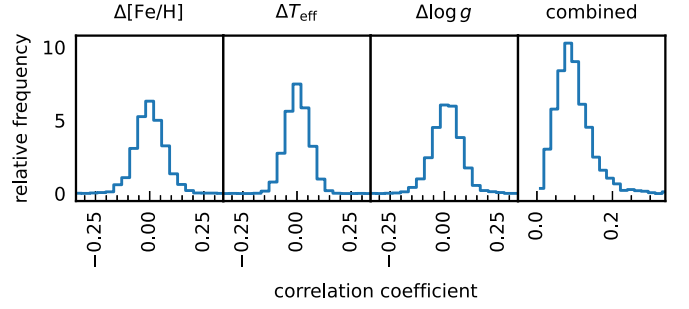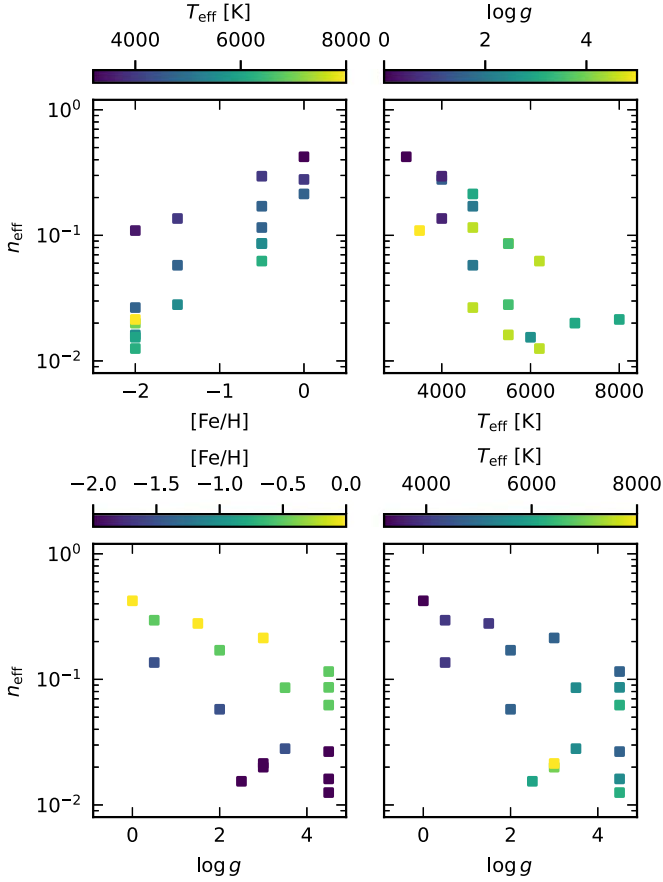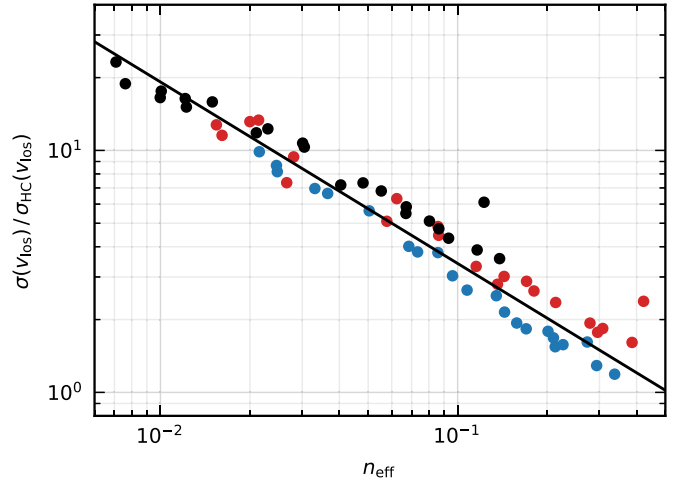
Figure 10. The dependence of the ratio of Equation (40) of Hatzes & Cochran (1992) to empirical measurements of $\sigma(v_{\rm los})$ based on our simulations, plotted as a function of the effective line number per resolution element as defined in Equation (45) for the arm configurations B (blue dots), R (red dots), and MR (black dots). Correcting the formula in Equation (40) using $n_{\rm eff}$ can account for a factor of 10 change in the uncertainty of the line-of-sight velocity as a function of spectral type. The black line is not a fit but rather the function $0.6 \cdot n_{\rm eff}^{-3/4}$. The outlier point is an M giant, see Section 7.2 for discussion.

Figure 9. Dependence of the effective line number $N_{\rm eff}$ on the atmospheric parameters in four different projections, calculated for the MR arm. In this wavelength range, the effective number of lines depends primarily on metallicity and temperature and only secondarily on surface gravity, as it can be seen in the top two panels and and bottom left one, respectively. While the depth of the hydrogen lines grows with temperature in the range we tested, the roughly inverse relation between $T_{\rm eff}$ and $N_{\rm eff}$ suggests that the calcium and iron lines play a more significant role.

effective line number to derive a correction to Equation (40) of Hatzes & Cochran (1992). Let us define the effective line number as

$$N_{\rm eff} = \sum \left| \frac{df_i}{d\ln(\lambda_i/1\ {\rm \AA})} \right| \Big/ \sum f_i, \qquad (45)$$

where f_i is the flux in the ith pixel and λ_i is the wavelength measured at the pixel center. We use the derivative by $\ln\lambda$ since the broadening of spectral lines scales with the wavelength. This definition is additive in the sense that the effective number of lines scales with the wavelength coverage and the effective number of lines is always the sum of parts of the spectrum. We also introduce the quantity $n_{\rm eff}$, which is the effective number of lines per resolution element.

Figure 9 illustrates the behavior of $N_{\rm eff}$ as a function of the fundamental atmospheric parameters for the models in Table 2, where the effective line number is determined for the MR arm configuration. The strong dependence of the effective line number on metallicity is obvious, as well as that $N_{\rm eff}$ decreases sharply with increasing temperature, at least in the parameter range of our models. This suggests that $N_{\rm eff}$ is indeed sensitive to metal lines instead of the hydrogen line series. $N_{\rm eff}$ also depends strongly on surface gravity, having a much larger

value for giant stars than dwarf stars of the same metallicity. This is in accordance with our intuition that the line-of-sight velocity of giant stars with narrow lines can be measured with lower uncertainty at the same S/N of the flux than dwarf stars with pressure-broadened lines.

To adopt Equation (40) to different stellar types, we calculated the ratio of the empirical $\sigma(v_{\rm los})$, as measured from our simulations, and $\sigma_{\rm HC}(v_{\rm los})$, as calculated from Equation (40), as a function of S/N for the arm configurations B and MR. The 95th percentile of the per pixel S/N turned out to be a much better characterization of noise in the continuum pixels than the median of the per pixel S/N when the overall S/N of spectra from different spectral types as compared. In Figure 10, we plot the ratio $\sigma(v_{\rm los})/\sigma_{\rm HC}(v_{\rm los})$ as a function of the effective line count per resolution element. We find it remarkable how the points distribute along a single line in the log–log plot even though we did the analysis on fluxed spectra covering a broad range of atmospheric parameters.

The black line in Figure 10 is not a fit to the data but indicates the function $0.6 \cdot n_{\rm eff}^{-3/4}$ that we can use to adapt Equation (40) to different spectral types in the form

$$\sigma(v_{\rm los}) = 0.87 \times 10^6 \cdot n_{\rm eff}^{-3/4} \cdot ({\rm S/N})^{-1} \cdot R^{-1} \cdot B^{-1/2}, \quad (46)$$

where the unit of velocity is in units of km s^{-1}.

7.2. Different Scaling of σ(v$_{los}$) for M giants

In the bottom right panel of Figure 3, we plotted the uncertainty of $v_{\rm los}$ as a function of the 95th percentile S/N per resolution element for various combinations of the spectrograph arms. The scaling of $\sigma(v_{\rm los})$ is significantly different for the MR arm relative to the other two and also differs from the scaling of $\sigma(v_{\rm los})$ of any other models. The flattening of the black curve is most likely due to the adopted definition of an overall S/N, which characterizes the rest of the studied stellar types well.

In Figure 2, we plotted the per pixel S/N for three different stellar types in the MR arm, as a function of wavelength, where the top panel is for an M giant, compared to the normalized

S/N of a metal-poor M dwarf and a metal-poor K giant. The horizontal red line indicates the 95th percentile of the S/Ns, which we used on the x-axis of Figure 3. The 95th percentile was chosen to define a measure of noise, which is mostly sensitive to the continuum pixels where the S/N is the best—as in the case of absorption spectra. Figure 2 clearly shows that the mean and median S/N cannot capture the continuum pixels while the 95th percentile works for most stellar types except when the absorption features are so strong that we see no continuum pixels anymore. As a consequence, the M giant is an outlier on the rest of the plots as well.

7.3. Uncertainties at Low Signal-to-noise Ratios from Markov Chain Monte Carlo

We have seen in Section 6.3 that the uncertainty estimate of $v_{\rm los}$ calculated from the Fisher matrix becomes unreliable at low S/N and significantly underestimates the empirical error determined from repeated simulations. On the other hand, Monte Carlo sampling of the same objective function yields uncertainties consistent with the empirical errors. Even when flux correction coefficients are set aside, a full Monte Carlo sampling of the posterior of $v_{\rm los}$ and the template parameters has a high computational cost compared to evaluating the Fisher matrix at the best-fit parameters.

Nevertheless, some applications might benefit from measuring $v_{\rm los}$ from a large number of stellar spectra observed at very low S/N. For example, the surface density of relatively bright stars of faint satellite galaxies of the Milky Way might be too low to fill the fibers of an instrument similar to the Subaru PFS. However, measuring the line-of-sight velocity of a large number of fainter stars—even if only with high uncertainties—might carry useful information when the uncertainties are estimated correctly and the data are treated in a statistically robust way.

8. Conclusions

We have investigated the dependence of the uncertainty of line-of-sight velocity measurements on the S/N of the flux in the case of a wide selection of stellar types that are relevant to Galactic archeological observations. We found that at intermediate S/N, the uncertainty of $v_{\rm los}$ scales with the square root of S/N as expected from theory, but significant deviations arise at low S/N. We have shown that asymptotic estimators of the uncertainty, such as the Cramér–Rao lower bound, calculated from the Fisher matrix, is unable to correctly characterize $\sigma(v_{\rm los})$ when measured from noisy spectra, but a full Monte Carlo sampling of the posterior probability can account for the empirical variance that we see in repeated simulations of the template fitting.

We have calculated the elements of the Fisher matrix for the parameters of the best-fitting template spectrum and also given analytic results for the uncertainty of the line-of-sight velocity. We have shown how the full covariance matrix of the flux correction function coefficients and the template parameters can be calculated efficiently when the systematics of the flux calibration of the observation or the differences between the template spectrum and the observation require correcting the stellar continuum with some wavelength-dependent correction function in the form of a linear combination of base functions, such as polynomials.

We have defined a new quantity, the effective absorption line density, which can be used to estimate the expected uncertainty of line-of-sight velocity measurements from a spectrum with a certain stellar type, as a function of the continuum S/N. The method works as well as the uncertainties calculated by Bouchy et al. (2001) and gives more realistic results than the formula of Hatzes & Cochran (1992).

Our work informs best practices for estimating uncertainties in $v_{\rm los}$. Our conclusions are relevant to the new era of massively multiplexed spectroscopy, featuring spectrographs such as Subaru PFS, DESI, WEAVE, 4MOST, and VLT/MOONS. These spectrographs will inevitably observe many—perhaps a majority—stars at low S/N where we have found that simple estimators of uncertainty are inaccurate. We advocate for the computation of full posterior $v_{\rm los}$ distributions for science cases where accurate estimates of uncertainty are important, such as the computation of dark matter density profiles.

Acknowledgments

Appendix A
Properties of the Fisher Matrix

In Sections 4.2–4.5, we calculated the Fisher matrix of the significance function to estimate the uncertainty of $v_{\rm los}$ measurements. Here we outline the calculations necessary to evaluate the formulate numerically, as well as show that the uncertainty of velocity measurements decreases when the template parameters are allowed to vary. We also prove that the formulae introduced in Sections 4.3 and 4.4 for simultaneously fitting the Doppler shift, the template parameters and the coefficients of the envelope correction are correct.

A.1. Numerical Evaluation Scheme for the Fisher Matrix

While Equation (18) is an intuitive analytic result, this is not the most convenient way to compute the Fisher matrix numerically. In case of high S/N, values of the significance function ν become large, and in the second derivative, one needs to take small differences of large numbers, leading to fairly large numerical errors. We can minimize the errors by calculating the relevant second derivatives as pixel-wise sums over the numerical derivatives of the template spectrum. In order to facilitate this, we will introduce the following set of new functions containing the derivatives of the models

$$\psi_{00} = \sum_p \frac{m_p(z_0)m_p(z_0)}{\sigma_p^2}, \tag{A1}$$

$$\psi_{01} = \sum_p \frac{m_p(z_0)m_p'(z_0)}{\sigma_p^2}, \tag{A2}$$

$$\psi_{11} = \sum_p \frac{m_p'(z_0)m_p'(z_0)}{\sigma_p^2}, \tag{A3}$$

$$\psi_{02} = \sum_p \frac{m_p(z_0)m_p''(z_0)}{\sigma_p^2}, \tag{A4}$$

$$\phi_{00} = \sum_p \frac{f_p m_p(z_0)}{\sigma_p^2}, \tag{A5}$$

$$\phi_{01} = \sum_p \frac{f_p m_p'(z_0)}{\sigma_p^2}, \tag{A6}$$

$$\phi_{02} = \sum_p \frac{f_p m_p''(z_0)}{\sigma_p^2}. \tag{A7}$$

It turns out that the function ψ_{02} will never appear in the Fisher matrix. We can also define the matrix Ψ as

$$\Psi = \begin{pmatrix} \psi_{00} & \psi_{01} \\ \psi_{01} & \psi_{11} \end{pmatrix}. \tag{A8}$$

Also note that $\psi_{00} = \chi(z_0) = \chi_0$. The ensemble averages of the different terms taken at the maximum of the significance function are the following

$$\langle \varphi \rangle_0 = A_0 \phi_{00}, \quad \langle \varphi' \rangle_0 = A_0 \phi_{01}, \quad \langle \varphi'' \rangle_0 = A_0 \phi_{02},$$
$$\langle \chi \rangle_0 = \psi_{00}, \quad \langle \chi' \rangle_0 = 2\psi_{01}, \quad \langle \chi'' \rangle_0 = 2(\psi_{11} + \psi_{02}). \tag{A9}$$

At the maximum of the significance function $\phi_{00} = A_0 \psi_{00}$ and $2\phi_{01}/\phi_{00} = \psi_{01}/\psi_{00}$ hold. We can use this and substitute the ensemble averages of the derivatives into the curvature matrix at the true position of the maximum, giving us the Fisher matrix in the form of

$$F = \begin{pmatrix} \psi_{00} & A_0 \psi_{01} \\ A_0 \psi_{01} & A_0^2(\psi_{11} + \psi_{02}) - A_0 \phi_{02} \end{pmatrix}. \tag{A10}$$

We can write the determinant of F as

$$|F| = A_0^2 |\Psi| + A_0 \psi_{00}(A_0 \psi_{02} - \varphi_{02}). \tag{A11}$$

The last term in the determinant should go to 0 as we approach the optimal template. Now we can compute the σ_z^2 element, which gives the Cramér–Rao lower bound on the error of line-of-sight velocity measurement as

$$\sigma_z^2 = \frac{\psi_{00}}{|F|}. \tag{A12}$$

We can use $A_0 = \varphi_0/\chi_0$ to reduce this to quantities that are already calculated during the likelihood evaluation. In any case, A_0 is dimensionless, χ_0 scales with the inverse variance, and Ψ scales as the square of the inverse variance. Overall, σ_z^2 should approximately scale with the variance of the noise, and its units should be km s^{-1} from $|\Psi|$.

A.2. Including the Template Parameters

Including the template parameters into the optimization—as opposed to only optimize for the line-of-sight velocity with a fixed template—involves taking the partial derivatives of the significance function with respect to the atmospheric parameters θ_α. The curvature matrix of the likelihood surface will be larger, augmented by a row and column for each parameter, but the expressions for the matrix elements will be very similar. Here we will use the subscript z to denote differentiation with respect to z, and α and β to denote differentiation with respect to the template parameters. Omitting the zero subscript for the

maximum, the augmented Fisher matrix can be written as

$$|F| = \begin{pmatrix} \chi & \varphi_1 & \varphi_\beta \\ \varphi_1 & -\nu\nu_{11} + \frac{\varphi_1^2}{\chi} & -\nu\nu_{1\beta} + \frac{\varphi_1 \varphi_\beta}{\chi} \\ \varphi_\alpha & -\nu\nu_{\alpha 1} + \frac{\varphi_\alpha \varphi_1}{\chi} & -\nu\nu_{\alpha\beta} + \frac{\varphi_\alpha \varphi_\beta}{\chi} \end{pmatrix}. \tag{A13}$$

In order to calculate the Cramér–Rao bound for the Doppler shift, we need to calculate the zz matrix element of the inverse matrix. This requires calculating the determinant of the matrix

$$S = \begin{pmatrix} \chi & \varphi_\beta \\ \varphi_\alpha & -\nu\nu_{\alpha\beta} + \frac{\varphi_\alpha \varphi_\beta}{\chi} \end{pmatrix}, \tag{A14}$$

where we eliminated the column and row of F corresponding to the zz element. The velocity error bound σ_z^2 is given by

$$\sigma_z^2 = \frac{|S|}{|F|}. \tag{A15}$$

We can get a relatively simple expression if we rewrite these matrices in a block form as

$$F = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} a & b \\ b^T & d \end{pmatrix}, \tag{A16}$$

where B and b are row vectors. The determinants of F and S can be expressed as

$$|F| = |A||D - B^T A^{-1} B| = |A||G|, \tag{A17}$$

$$|S| = |a||d - b^T a^{-1} b| = |a||g|, \tag{A18}$$

where we introduced $G = D - B^T A^{-1} B$ and $g = d - b^T a^{-1} b$.

Let us now consider our actual problem where, in the optimum of the likelihood function, the matrix blocks will be $A = \chi$, $B = (\varphi_z, \varphi_\alpha)$, and $A^{-1} = 1/\chi$, thus

$$G_{ij} = -\nu\nu_{ij}, \tag{A19}$$

where $i, j = (z, \alpha...)$ go over z and the atmospheric parameters of the template. Similarly

$$g_{ij} = -\nu\nu_{\alpha\beta}, \tag{A20}$$

but the indices only run over the templates parameters α. Here g is the Hessian of the log likelihood at the maximum point, with respect to the template parameters, and its determinant is the Gaussian curvature. With these transformations, the variance of the Doppler shift becomes much simpler. With $|A| = |a| = \chi$, it can be written as

$$\sigma_z^2 = \frac{|g|}{|G|}. \tag{A21}$$

Let us separate off the first row and column of G and write it as

$$G = \begin{pmatrix} -\nu\nu_{zz} & -\nu\nu_{z\beta} \\ -\nu\nu_{\alpha z} & g_{\alpha\beta} \end{pmatrix}. \tag{A22}$$

Knowing that $G_{11} = -\nu\nu_{zz}$, the determinant of G can be expressed as

$$|G| = -\nu\nu_{zz} \left| g + \frac{\nu\nu_\alpha \nu_\beta}{\nu_{zz}} \right| = -\nu\nu_{zz}|\hat{g}|, \tag{A23}$$

where

$$\hat{g}_{\alpha\beta} = g_{\alpha\beta} + \left(\frac{\nu}{\nu_{zz}}\right)\nu_\alpha\nu_\beta, \qquad (A24)$$

with which the variance becomes

$$\sigma_z^2 = \left(\frac{1}{-\nu\nu_{11}}\right)\frac{|g|}{|\hat{g}|}. \qquad (A25)$$

One can see how this is different from Equation (20), where the template parameters were not considered. We will now show that $|g|/|\hat{g}| \leqslant 1$, which indicates that the error of the line-of-sight velocity measurement decreases when the template parameters are optimized as well.

Let us factorize the symmetric, positive definite matrix g as $g = U^T\Sigma U$, where Σ is diagonal and rotate the ν_γ partial derivatives of the significance with U as

$$s_\beta = \nu_\gamma U_{\gamma\beta}, \qquad s_\alpha^T = U_{\alpha\gamma}^T\nu_\gamma. \qquad (A26)$$

Since the determinant is invariant under the rotation, we can write $|\hat{g}|$ as

$$|\hat{g}| = \left| U^T\Sigma U + \left(\frac{\nu}{\nu_{zz}}\right)U^T s^T s U \right| = \left| \Sigma + \left(\frac{\nu}{\nu_{zz}}\right)s^T s \right|. \qquad (A27)$$

According to the determinant lemma, see Appendix B.3, the determinant of a matrix of this special form (diagonal plus an outer product of a vector with itself) can be calculated analytically as

$$|\hat{g}| = |g|\left(1 + \sum_\alpha \frac{s_\alpha^2}{\lambda_\alpha}\right), \qquad (A28)$$

where λ_α are the eigenvalues (diagonal elements) of Σ. This yields the very intuitive result of

$$\sigma_z^2 = \left(\frac{1}{-\nu\nu_{zz}}\right)\frac{1}{1 + \sum_\alpha \frac{s_\alpha^2}{\lambda_\alpha}}, \qquad (A29)$$

which relates to the simple case where we only maximized the likelihood to get the amplitude and the line-of-sight velocity, but not the atmospheric parameters. Furthermore, it shows that the more parameters we optimize for, the lower the uncertainty of the inferred line-of-sight velocity will be.

A.3. Linear Model for Envelope Correction

We made the assumptions in Sections 4.3 and Appendix A.2 that the observed spectrum is noisy and the template differs from the observation only by a linear amplitude; this assumption is obviously too simple. It is reasonable to expect that the models need to be corrected with a smooth multiplicative function of wavelength to account for small errors in the spectrophotometric calibration, or slight differences in the continuum of the template and the real spectrum. As introduced in Section 4.4, we model the flux correction with a function that is a linear combination of functions $q_n(\lambda)$, such as Chebyshev or Legendre polynomials, which only depend on the

wavelength in the form of

$$A(\lambda) = \sum_{n=0}^{N} A_n q_n(\lambda). \qquad (A30)$$

The orthogonality of the functions is not necessary but the $q_n(\lambda_p)$ vectors must be linearly independent to ensure that we can always solve for the best-fit A_n coefficients when all template parameters, including z, are given.

We can now rewrite the log likelihood using a vector notation, where the quantities A, φ, and χ are vectors and matrices of dimension of $N+1$, as

$$\mathcal{L}(A, z) = A^T\varphi - \frac{1}{2}A^T\chi A, \qquad (A31)$$

where $\varphi(z)$ is a vector-valued function and $\chi(z)$ is matrix-valued function of the Doppler shift, defined as

$$\varphi_k(z) = \sum_p q_k(\lambda_p)\frac{f_p m_p(z)}{\sigma_p^2}, \qquad (A32)$$

$$\chi_{kn}(z) = \sum_p q_k(\lambda_p) q_n(\lambda_p)\frac{m_p^2(z)}{\sigma_p^2}. \qquad (A33)$$

These are the natural extensions of Equation (4) from Section 4.1, with the addition of the array of multiplicative functions q_k. Evaluating the partial derivatives of $\mathcal{L}$ with respect to A_k and z, an equating them to 0, we get

$$\frac{\partial\mathcal{L}}{\partial A} = \varphi(z) - \chi(z)A = 0, \qquad (A34)$$

$$\frac{\partial\mathcal{L}}{\partial z} = A^T\varphi'(z) - \frac{1}{2}A^T\chi'(z)A = 0, \qquad (A35)$$

where we denoted the differentiation with respect to z by the prime and continued to use vector notation. Since χ is a symmetric, positive definite matrix, it has an inverse, and from the first equation we can solve for A at the maximum point of $\mathcal{L}$, denoted by the zero subscript. The solution of Equation (A34) is simply

$$A_0 = \chi_0^{-1}\varphi_0, \qquad A_0^T = \varphi_0^T\chi_0^{-1}. \qquad (A36)$$

The solution of Equation (A35) gives us the line-of-sight velocity. At the maximum point with respect to the A_n coefficients, it becomes

$$A_0^T\varphi_0' - \frac{1}{2}A_0^T\chi_0'A_0 = A_0^T\left[\varphi_0' - \frac{1}{2}\chi_0'A_0\right] = 0, \qquad (A37)$$

which means that the quantity in the square brackets, composed of the first derivatives of φ and χ, is orthogonal to A_0 at the maximum. Substituting the values of A_0 from Equation (A36), we can rewrite this in the form of

$$\frac{\partial\mathcal{L}}{\partial z}\bigg|_0 = \varphi_0^T\chi_0^{-1}\varphi_0' - \frac{1}{2}\varphi_0^T\chi_0^{-1}\chi_0'\chi_0^{-1}\varphi_0. \qquad (A38)$$

Here and later we will be using the identity

$$K' = -K\chi_0'K, \text{ and } K'' = -K\chi_0''K + 2K\chi_0'K\chi_0'K, \qquad (A39)$$

for the derivative of the inverse of a symmetric invertible matrix, where $K = \chi_0^{-1}$. With this, we can rewrite

Equation (A38) in the more suggestive form of

$$\frac{\partial \mathcal{L}}{\partial z}\bigg|_0 = \boldsymbol{\varphi}_0^{\mathrm{T}} \boldsymbol{\chi}_0^{-1} \boldsymbol{\varphi}_0' + \frac{1}{2}\boldsymbol{\varphi}_0^{\mathrm{T}}(\boldsymbol{\chi}_0^{-1})' \boldsymbol{\varphi}_0$$

$$= \frac{\partial}{\partial z}\left(\frac{1}{2}\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\chi} - 1\boldsymbol{\varphi}\right)\bigg|_0. \quad (A40)$$

The expression in the parentheses is identical in function to the significance function ν, defined in Equation (10) of Section 4.1 as $\nu^2/2 = \varphi^2/\chi$. The significance function in vector form can be written as

$$\frac{\nu^2}{2} = \boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\chi}^{-1}\boldsymbol{\varphi}, \quad (A41)$$

which can be evaluated at any Doppler shift z, and whose maximum coincides with the maximum of $\mathcal{L}$.

A.4. The Fisher Matrix in Vector Form

Similarly to Section 4.1, we can calculate the elements of the Fisher information matrix. First, let us write down the second derivatives of the log likelihood, which are

$$\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{A}^2} = -\boldsymbol{\chi}, \quad (A42)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{A}\partial z} = \boldsymbol{\varphi}' - \boldsymbol{\chi}'\boldsymbol{A}, \quad (A43)$$

$$\frac{\partial^2 \mathcal{L}}{\partial z^2} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\varphi}'' - \frac{1}{2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\chi}''\boldsymbol{A}. \quad (A44)$$

The Fisher matrix is the negative of this, evaluated at the maximum point as

$$F = \begin{pmatrix} \boldsymbol{\chi}_0 & -\boldsymbol{\varphi}_0' + \boldsymbol{\chi}_0'\boldsymbol{A}_0 \\ -\boldsymbol{\varphi}_0'^{\mathrm{T}} + \boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0' & -\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\varphi}_0'' + \frac{1}{2}\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0''\boldsymbol{A}_0 \end{pmatrix}. \quad (A45)$$

This is a symmetric block matrix of size $N + 2$, composed of an $N + 1$ matrix $\boldsymbol{\chi}_0$ augmented by a single row and column that corresponds to the Doppler shift. We can use the so called bordering formula, see Equation (B1), to evaluate the elements of the inverse of the Fisher matrix $S = F^{-1}$, block by block, starting with S_{22} as

$$\frac{1}{S_{22}} = -\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\varphi}_0'' + \frac{1}{2}\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0''\boldsymbol{A}_0$$
$$- (\boldsymbol{\varphi}_0'^{\mathrm{T}} - \boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0')\boldsymbol{\chi}_0^{-1}(\boldsymbol{\varphi}_0' - \boldsymbol{\chi}_0'\boldsymbol{A}_0), \quad (A46)$$

$$= -\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\varphi}_0'' + \frac{1}{2}\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0''\boldsymbol{A}_0 - \boldsymbol{\varphi}_0'^{\mathrm{T}}\boldsymbol{\chi}_0^{-1}\boldsymbol{\varphi}_0'$$
$$- \boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0'\boldsymbol{\chi}_0^{-1}\boldsymbol{\chi}_0'\boldsymbol{A}_0 + \boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0'\boldsymbol{\chi}_0^{-1}\boldsymbol{\varphi}_0', \quad (A47)$$

where we already aggregated the symmetric terms. Let us now show that this term is exactly $-\nu_0 \nu_0''$ by using the expressions from Equations (B5), (B6), and (A36)

$$\frac{\partial^2}{\partial z^2}\left(\frac{1}{2}\nu^2\right)\bigg|_0 = \frac{1}{2}(\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\chi}^{-1}\boldsymbol{\varphi})''\bigg|_0, \quad (A48)$$

$$= \boldsymbol{\varphi}_0^{\mathrm{T}}\boldsymbol{K}_0\boldsymbol{\varphi}_0'' + \frac{1}{2}\boldsymbol{\varphi}_0^{\mathrm{T}}\boldsymbol{K}_0''\boldsymbol{\varphi}_0 + \boldsymbol{\varphi}_0'^{\mathrm{T}}\boldsymbol{K}_0\boldsymbol{\varphi}_0' + 2\boldsymbol{\varphi}_0^{\mathrm{T}}\boldsymbol{K}_0'\boldsymbol{\varphi}_0', \quad (A49)$$

$$= \boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0'' - \frac{1}{2}\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0''\boldsymbol{A}_0 + \boldsymbol{\varphi}_0'^{\mathrm{T}}\boldsymbol{\chi}_0^{-1}\boldsymbol{\varphi}_0'$$
$$+ \boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0'\boldsymbol{\chi}_0^{-1}\boldsymbol{\chi}_0'\boldsymbol{A}_0 - 2\boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0'\boldsymbol{\chi}_0^{-1}\boldsymbol{\varphi}_0'. \quad (A50)$$

We can thus express S_{22} through $\nu_0 \nu_0''$ in a much simpler form as

$$\mu = \frac{1}{S_{22}} = -\nu_0 \nu_0'', \quad (A51)$$

and the rest of the matrix elements are

$$S_{11} = \boldsymbol{\chi}_0^{-1} + \frac{1}{\mu}\boldsymbol{\chi}_0^{-1}(-\boldsymbol{\varphi}_0' + \boldsymbol{\chi}_0'\boldsymbol{A}_0)(-\boldsymbol{\varphi}_0'^{\mathrm{T}} + \boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0')\boldsymbol{\chi}_0^{-1},$$

$$(A52)$$

$$S_{12} = -\boldsymbol{\chi}_0^{-1}(-\boldsymbol{\varphi}_0' + \boldsymbol{\chi}_0'\boldsymbol{A}_0)/\mu, \quad (A53)$$

$$S_{21} = -(-\boldsymbol{\varphi}_0'^{\mathrm{T}} + \boldsymbol{A}_0^{\mathrm{T}}\boldsymbol{\chi}_0')\boldsymbol{\chi}_0^{-1}/\mu. \quad (A54)$$

Next, we define the vector $\boldsymbol{B}$ for the first derivatives as

$$\boldsymbol{B}_0 = \boldsymbol{\chi}_0^{-1}(\boldsymbol{\varphi}_0' - \boldsymbol{\chi}_0'\boldsymbol{A}_0). \quad (A55)$$

With this, the expression of inverse of the Fisher matrix simplifies considerably to

$$S = \begin{pmatrix} \boldsymbol{\chi}_0^{-1} - \dfrac{\boldsymbol{B}_0\boldsymbol{B}_0^{\mathrm{T}}}{\nu_0\nu_0''} & \dfrac{\boldsymbol{B}_0}{\nu_0\nu_0''} \\ \dfrac{\boldsymbol{B}_0^{\mathrm{T}}}{\nu_0\nu_0''} & -\dfrac{1}{\nu_0\nu_0''} \end{pmatrix}. \quad (A56)$$

To us the most important is the S_{22} element. This is in exact agreement with the previous, scalar result, except that ν^2 is calculated as a quadratic form of φ and χ^{-1}. The rest of the inverse matrix can be calculated numerically using the formulae provided above.

Appendix B
Matrix Identities

In the paper, we use several identities to aid in the inversion of matrices and element-wise derivatives of matrices. Here we summarize these identities along with some explanation.

B.1. The Bordering Formula

To analytically calculate the inverse of the Fisher matrix, we used a matrix identity called the bordering formula. Here F is a symmetric matrix of size $N + 1$, built by padding an existing invertible square matrix M of size N with an extra row and column in a symmetric fashion, using the vector d and the scalar c as

$$F^{-1} = \begin{pmatrix} \boldsymbol{M} & \boldsymbol{d} \\ \boldsymbol{d}^T & c \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{M}^{-1} + \boldsymbol{M}^{-1}\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{M}^{-1}/\mu & -\boldsymbol{M}^{-1}\boldsymbol{d}/\mu \\ -\boldsymbol{d}^T\boldsymbol{M}^{-1}/\mu & 1/\mu \end{pmatrix}, \quad (B1)$$

where

$$\mu = c - \boldsymbol{d}^T\boldsymbol{M}^{-1}\boldsymbol{d}. \quad (B2)$$

B.2. Derivative of an Inverse Matrix

This formula deals with computing the element-wise derivative of a matrix inverse. Let us take a symmetric invertible matrix χ, and denote its inverse as $\boldsymbol{K} = \boldsymbol{\chi}^{-1}$. We can

show that the derivative of K can be calculated as $K' = -K\chi'K$. Let us start from the equality

$$1 = \chi\chi^{-1} = \chi K. \tag{B3}$$

By differentiating the above equation we arrive at

$$0 = \chi'K + \chi K'. \tag{B4}$$

Rearranging this gives

$$K' = -K\chi'K, \tag{B5}$$

from which we can calculate the second derivative as

$$K'' = -2K'\chi'K - K\chi''K = -K\chi''K + 2K\chi'K\chi'K. \tag{B6}$$

B.3. The Determinant Lemma

The determinant lemma states that the determinant of a matrix in a special form (diagonal plus an outer product of a vector with itself) can be calculated as

$$|A u v^{\mathrm{T}}| = |A|(1 + v^{\mathrm{T}}A^{-1}u), \tag{B7}$$

where v and u are arbitrary column vectors.

ORCID iDs

László Dobos ⓘ https://orcid.org/0000-0001-7679-9478
Alexander S. Szalay ⓘ https://orcid.org/0000-0002-4108-3282
Tamás Budavári ⓘ https://orcid.org/0000-0002-7034-4621
Evan N. Kirby ⓘ https://orcid.org/0000-0001-6196-5162
Robert H. Lupton ⓘ https://orcid.org/0000-0003-1666-0962
Rosemary F. G. Wyse ⓘ https://orcid.org/0000-0002-4013-1799

References

Badenes, C., Mazzola, C., Thompson, T. A., et al. 2018, ApJ, 854, 147
Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Cunningham, E. C., Deason, A. J., Rockosi, C. M., et al. 2019, ApJ, 876, 124
David, M., Blomme, R., Frémat, Y., et al. 2014, A&A, 562, A97
DESI Collaboration, Abareshi, B., Aguilar, J., et al. 2022, AJ, 164, 207
Dey, A., Najita, J. R., Koposov, S. E., et al. 2023, ApJ, 944, 1
Hatzes, A. P., & Cochran, W. D. 1992, in ESO Workshop on High Resolution Spectroscopy with the VLT, ed. M.-H. Ulrich (Garching: ESO), 275
Hirata, C. M., Gehrels, N., Kneib, J.-P., et al. 2012, arXiv:1204.5151
Husser, T. O., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6
Jenkins, S. A., Li, T. S., Pace, A. B., et al. 2021, ApJ, 920, 92
Kaiser, N. 2004, PanSTARRS Internal Report, PSDC-002-010-00
Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, A&A, 501, 1269
Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2008, MNRAS, 385, 1998
Koposov, S. E., Gilmore, G., Walker, M. G., et al. 2011, ApJ, 736, 146
Kounkel, M., Covey, K. R., Stassun, K. G., et al. 2021, AJ, 162, 184
Moe, M., Kratter, K. M., & Badenes, C. 2019, ApJ, 875, 61
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
North, R. C., Marsh, T. R., Kolb, U., Dhillon, V. S., & Moran, C. K. J. 2002, MNRAS, 337, 1215
Rix, H.-W., & White, S. D. M. 1992, MNRAS, 254, 389
Seeburger, R., Rix, H.-W., El-Badry, K., Xiang, M., & Fouesneau, M. 2024, MNRAS, 530, 1935
Sohn, S. T., Majewski, S. R., Muñoz, R. R., et al. 2007, ApJ, 663, 960
Stoehr, F., White, R., Smith, M., et al. 2008, ASP Conf. Ser. 394, Astronomical Data Analysis Software and Systems XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis (San Francisco, CA: ASP), 505
Sugai, H., Tamura, N., Karoji, H., et al. 2014, Proc. SPIE, 9147, 91470T
Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1
Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2006, AJ, 131, 2114
Walker, M. G., Olszewski, E. W., & Mateo, M. 2015, MNRAS, 448, 2717
Wyse, R. F. G., Moe, M., & Kratter, K. M. 2020, MNRAS, 493, 6109
Yabe, K., Moritani, Y., Shimono, A., & Lupton, R. 2017, PFS Exposure Time Calculator and Spectrum Simulator, https://github.com/Subaru-PFS/spt_ExposureTimeCalculator