

# A Method for the Spectroscopic Inference of Stellar Parameters

Ian Czekala, Sean M. Andrews, et al.

*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138*

## 1. Introduction

All astronomers recognize that spectroscopy offers a wealth of information that can help characterize the properties of the observing target. In the context of stellar astrophysics, spectroscopy plays many fundamental roles. The relative strengths and widths of stellar absorption lines provide access to key parameters like effective temperature ( $T_{\text{eff}}$ ) and surface gravity ( $\log g$ ), enabling model comparisons in the Hertzsprung-Russell diagram to estimate the masses and ages so crucial to understanding stellar evolution, as well as individual elemental (and molecular) abundances or the collective “metallicity” (typically parameterized as  $[\text{Fe}/\text{H}]$ ), facilitating study of the chemical hallmarks of different stellar populations. With sufficient spectral resolution, the velocity content of a spectrum can convey crucial information about stellar rotation ( $v \sin i$ ) and kinematics (e.g., association with a cluster or companion through the radial velocity,  $v_z$ ). While many fields benefit from the spectroscopic measurements of these stellar properties, there is acute interest in the exoplanet community. There, all estimates of the planet properties are made *relative* to the host properties (e.g., the planet-to-host mass or radius *ratio* is measured with the radial velocity or transit techniques, respectively). Moreover, essential clues to the planet formation process are encapsulated in the dependences of planet frequency on host mass (e.g., [Johnson et al. 2007](#); [Howard et al. 2010](#)) and metallicity (e.g., [Fischer & Valenti 2005](#); [Buchhave et al. 2014](#)).

That said, the robust and quantitative extraction of physical (or empirical) parameters from an observed spectrum can be an extraordinary challenge. Stellar models are ultimately required as a comparative benchmark to associate observed spectral features with the parameters of interest. Generating a synthetic model spectrum requires a complex numerical treatment of the stellar structure and radiative transfer through the atmosphere (e.g., [Kurucz 1993](#); [Castelli & Kurucz 2004](#); [Hauschildt et al. 1999](#); [Husser et al. 2013](#); [Paxton et al. 2011](#)). Detailed models calibrated to individual stars are important, but rare (e.g., the Sun, Vega); as such, these stellar models are relatively untested in large swaths of parameter-space. Moreover, they necessarily include simplifications to treat complicated physical processes (e.g., convection) or computational limitations (e.g., geometry, boundary conditions), and often must rely on incomplete or inaccurate atomic and molecular information (e.g., oscillator strengths, opacities). In principle, the models could be further improved with appropriate reference to spectroscopic datasets. Nevertheless, they achieve a remarkable level of success in reproducing many key diagnostic features in stellar spectra.

There are various well-tested techniques being used to compare these models with observed spectra and thereby infer basic stellar parameters; here we highlight three examples to illustrate some key ideas. First is a straightforward empirical approach that relies on distilling an information-rich subset of the spectral data, usually in the form of spectral line equivalent widths and/or local continuum shapes. A combined sequence of the ratios of these quantities are found to be especially sensitive to one key parameter in the models, and therefore can be used to quantify it (e.g., [Gray 1994](#); [Reid et al. 1995](#); [Rojas-Ayala et al. 2010, 2012](#)). The line index comparison approach has the major advantage of being trivially fast, but each index relationship is only informative over a limited range of parameter-space. A second technique exploits the cross-correlation of an observed spectrum with a suite of model templates to identify an optimized set of parameters, usually with some preferential weighting applied to specific spectral regions (e.g., [SPC](#); [Buchhave et al. 2012](#)). In this case, the speed advantage is maintained while taking advantage of more data content (particularly in the spectral dimension), thereby achieving high precision even for data with comparatively low sensitivity; the disadvantage is that the model quality and parameter inferences are assessed in an empirical, rather than probabilistic, framework. The third approach of interest prefers a direct, pixel-by-pixel comparison between model and observed spectra, and is usually associated with a spectral synthesis front-end (e.g., [MOOG](#); [Snedden 1973](#); [SME](#); [Valenti & Piskunov 1996](#)). This technique has the benefits of parametric flexibility (e.g., one can fit for atomic abundances or arbitrary structures) and a proper inference framework (usually a least-squares approach, although increasingly in a Bayesian format; [Shkedy et al. 2007](#); [Schönrich & Bergemann 2013](#)), but it is computationally expensive and can be biased by model systematics (e.g., [Mann et al. 2013](#)); in practice, this often requires an effort focused on a pre-selected subsample of the data.

In this article, we design a flexible forward-modeling approach to the general spectroscopic inference problem in a Bayesian framework, building on the best aspects of the latter two methods highlighted above. Employing a non-trivial covariance matrix parameterized by both global (stationary) and local (non-stationary) Gaussian process kernels, we are able to account for residual pixel-by-pixel correlations that arise when comparing observed spectra to intrinsically imperfect models. This approach minimizes the potential for bias, efficiently propagates systematic uncertainties into the parameter inferences, and ultimately can serve as tool for using observations to guide substantially improvements in the models themselves. An overview of the methodology of this approach is provided in [Section 2](#), including the covariance matrix formalism. Some relevant tests and example applications (specifically for a high resolution optical spectrum of an F star, and a medium-resolution near-infrared spectrum of a mid-M star) of the method are described in [Section 3](#). Finally, a discussion of the potential utility of the technique, and the possibility of extending it to develop data-driven spectral models, is provided in [Section 4](#).

## 2. Methodology

Here we describe a generative Bayesian modeling framework that confronts some of the key obstacles in the spectroscopic inference problem. The goal of this approach is to conservatively extract the maximal amount of information about a prescribed (and degenerate) parameter set by forward-modeling an observed spectrum, while also recognizing and explicitly accounting for the covariances and biases introduced by pathologically imperfect models or calibrations. The method is modular, and therefore can easily incorporate additional physical or nuisance parameters as desired without sacrificing an accurate reflection of the limitations in the data. Moreover, with a well-crafted observational sample, this data-driven approach should ultimately enable us to systematically learn how synthetic spectral models can be improved. The specific applications discussed here are related to the spectra of individual stars, but the methodology is generic (and could be used for composite spectra of unresolved stellar clusters, galaxies, etc.).

The remainder of this section describes the mechanics of this modeling framework. First, a model spectrum is generated for a given set of physical parameters (Section 2.1), and then post-processed to mimic reality using a set of observational and practical nuisance parameters (Section 2.2). Next, a direct, pixel-by-pixel comparison between the data and model spectra is made with a prescribed likelihood function and a parametric treatment of the covariances between pixel residuals (Section 2.3). That process is iterated under the guise of hierarchical Monte Carlo Markov Chain (MCMC) simulations to numerically explore the posterior probability density of the model conditioned on the data, and thereby to determine constraints on the parameters of interest (Section 2.4). Along the way, these procedures are illustrated with real observations of the high resolution optical spectrum from a nearby F star. That specific application, along with some alternative demonstrations of the method, are discussed in more detail in Section 3.

### 2.1. Generating a Model Spectrum

There are various approaches to synthesizing a spectrum,  $f_\lambda$ , for a specific set of model parameters,  $\theta_*$ . In an ideal case, a model stellar atmosphere is constructed and then subsequently processed through a radiative transfer code (e.g., Kurucz 1993; Hauschildt et al. 1999). However, in general this approach is still computationally prohibitive for any iterative method of probabilistic inference. One partial compromise is to interpolate over a library of atmosphere structures that were pre-computed for a discrete grid of parameter values,  $\{\theta_*\}^{\text{grid}}$ , for some arbitrary  $\theta_*$ , and then perform a radiative transfer calculation with that interpolated atmosphere to synthesize  $f_\lambda$  (e.g., as for SME; Valenti & Piskunov 1996). A more common variant is to instead rely on interpolation over a library of pre-synthesized model spectra,  $f_\lambda(\{\theta_*\}^{\text{grid}})$  (e.g., Castelli & Kurucz 2004; Allard et al. 2012; Husser et al. 2013). While technically the former approach is most similar to the ideal case, the computational cost of repeated spectral synthesis is sufficiently high to make a detailed exploration of parameter space (particularly for data with a large spectral range) considerably less

appealing. A related, but different approach is to eschew forward modeling entirely (and therefore repeated spectral syntheses and/or library interpolations), and instead evaluate the models only at the discrete grid points of the library. Then, these discretized samples of the posterior probability density can be interpolated to an arbitrary  $\theta_*$  to construct appropriate confidence intervals (similar to the method of SPC; Buchhave et al. 2012). The difficulty with this latter approach is that the parameter uncertainties can be smaller than the grid spacing; in that case, there is valid concern that this interpolation might not accurately recover intrinsic parameter degeneracies.

Here we opt to take the computationally expedient approach that employs a library of model spectra,  $f_\lambda(\{\theta_*\}^{\text{grid}})$ , where  $\theta_* = [T_{\text{eff}}, \log g, Z]$  (in practice, the metallicity  $Z$  is often parameterized by  $[\text{Fe}/\text{H}]$ ). However, it is worth noting that the techniques we will develop are applicable to *any* reasonable “front-end” that generates a model spectrum. In our adopted approach, the model spectrum for an arbitrary  $\theta_*$  must be interpolated from among the spectral library,

$$f_\lambda(\{\theta_*\}^{\text{grid}}) \rightsquigarrow f_\lambda(\theta_*), \quad (1)$$

where we assign the symbol  $\rightsquigarrow$  as an interpolation operator. The multi-dimensional interpolation in Eq. 1 needs to be performed many times, so computational efficiency is critical. In practice, a simple tri-linear interpolation is suitably fast, but introduces an undesirable level of inaccuracy (particularly in the  $Z$  dimension). The interpolation quality can be empirically estimated by performing the operation in Eq. 1 across a calculated location in  $\{\theta_*\}^{\text{grid}}$ , and then comparing the interpolated spectrum with the corresponding library spectrum. After an extensive exploration of such calculations (see also Husser 2012), we concluded that the best combination of speed and accuracy can be achieved by pre-computing a *refined* spectral library using a cubic spline interpolation with a  $\{\theta_*\}^{\text{grid}}$  spacing of [20 K, 0.1 dex, 0.1 dex], and then performing tri-linear interpolation over that refined grid. Overall, this interpolation technique is found to be accurate within a few percent per high resolution model pixel. Ideally, this pre-interpolation could be avoided if the spectral library was computed over a refined grid (with a substantial up-front computational investment); but for the time being, we can empirically propagate these interpolation uncertainties into the likelihood calculations (as will be described in Section 2.3). SA: I see nothing was done here regarding your updated approach; how should we handle that?

## 2.2. Post-Processing

Generally, the “raw” model spectrum  $f_\lambda(\theta_*)$  is highly over-sampled compared to a typical observed spectrum, and does not account for several additional observational and instrumental effects that become important in comparisons with real data. Therefore, a certain amount of post-processing is required before assessing the model quality. We treat that post-processing in two stages: the first deals with an additional set of “observational” parameters,  $\theta_{\text{obs}}$ , that incorporate dynamical effects, geometry, and the relative location of the target, while the second employs a suite of nuisance (hyper-)parameters,  $\Theta_{\text{n}}$ , designed to mitigate an imperfect data calibration.

We can further divide  $\boldsymbol{\theta}_{\text{obs}}$  into those parameters that impact the model primarily in the spectral or flux dimensions;  $\boldsymbol{\theta}_{\text{obs}} = [\boldsymbol{\theta}_{\text{obs},v}, \boldsymbol{\theta}_{\text{obs},f}]$ . For the former, we consider three kernels that contribute to the observed line-of-sight velocity distribution function,  $\varphi_v$ . The first,  $\mathcal{F}_v^{\text{inst}}$ , treats the instrumental spectral broadening. For illustrative purposes we assume  $\mathcal{F}_v^{\text{inst}}$  is a Gaussian with a mean at  $v = 0$  and a constant width  $\sigma_v$  at all  $\lambda$ , although more sophisticated forms could be adopted. The second,  $\mathcal{F}_v^{\text{rot}}$ , characterizes the broadening induced by (projected) stellar rotation, parameterized by  $v \sin i$  as described by Gray (2008, his Eq. 18.14). And the third,  $\mathcal{F}_v^{\text{dop}} = \delta(v - v_z)$ , incorporates the radial velocity through a Doppler shift. The model spectrum is modified by the parameters  $\boldsymbol{\theta}_{\text{obs},v} = [\sigma_v, v \sin i, v_z]$  through these kernels, using a convolution in velocity-space,<sup>1</sup>

$$f_\lambda(\boldsymbol{\theta}_*, \boldsymbol{\theta}_{\text{obs},v}) = f_\lambda(\boldsymbol{\theta}_*) \otimes \varphi_v = f_\lambda(\boldsymbol{\theta}_*) \otimes \mathcal{F}_v^{\text{inst}} \otimes \mathcal{F}_v^{\text{rot}} \otimes \mathcal{F}_v^{\text{dop}}, \quad (2)$$

and then re-sampled onto the discrete wavelengths corresponding to each data pixel,

$$f_\lambda(\boldsymbol{\theta}_*, \boldsymbol{\theta}_{\text{obs},v}) \mapsto \mathbf{M}(\boldsymbol{\theta}_*, \boldsymbol{\theta}_{\text{obs},v}), \quad (3)$$

where the  $\mapsto$  symbol denotes a re-sampling operator that maps the model spectrum onto the  $N_{\text{pix}}$ -dimensional array  $\mathbf{M}$  (and  $N_{\text{pix}}$  is the number of pixels in the spectrum). For reference, Figure 1 shows a graphical representation of these post-processing steps.

At this stage, the model is further modified in the flux dimension. A typical synthetic spectrum is computed as the flux that would be measured *at the stellar surface*, and so needs to be diluted by the subtended solid angle,  $\Omega = (R_*/d)^2$ , where  $R_*$  is the stellar radius and  $d$  is the distance. An additional wavelength-dependent scaling factor is applied to account for interstellar extinction, assuming some previously-derived extinction law  $A_\lambda$  (e.g., Cardelli et al. 1989) that is parameterized by  $A_V$ . The parameters  $\boldsymbol{\theta}_{\text{obs},f} = [\Omega, A_V]$  are applied as

$$\mathbf{M}(\boldsymbol{\theta}) \equiv \mathbf{M}(\boldsymbol{\theta}_*, \boldsymbol{\theta}_{\text{obs}}) = \mathbf{M}(\boldsymbol{\theta}_*, \boldsymbol{\theta}_{\text{obs},v}) \times \Omega \times 10^{-0.4 A_\lambda}, \quad (4)$$

where we have simplified the notation by writing  $\boldsymbol{\theta} \equiv [\boldsymbol{\theta}_*, \boldsymbol{\theta}_{\text{obs}}]$ .

So far, the procedure summarized in Eq. 1-4 is composed of relatively straightforward operations demanded by practical astronomical and computing issues. If the observed spectrum was *perfectly* calibrated, we could proceed to a likelihood calculation (Section 2.3) that makes a direct comparison with  $\mathbf{M}(\boldsymbol{\theta})$  at this point. However, that is unlikely to be the case. The primary concern is that an imperfect calibration produces (presumably low-level) mismatches in the underlying shape of the observed spectrum on relatively broad wavelength scales. When compared with the model, these mismatches might represent a non-trivial contribution to the likelihood, and thereby bias our estimates of the desired physical parameters. Often, this concern is treated externally to any modeling procedure, usually by dividing the observed spectrum (and model spectrum) by

---

<sup>1</sup>In practice, these convolutions are performed as multiplications in Fourier-space to better preserve spectral information (cf., Tonry & Davis 1979); the mathematical formalism is presented here for clarity.

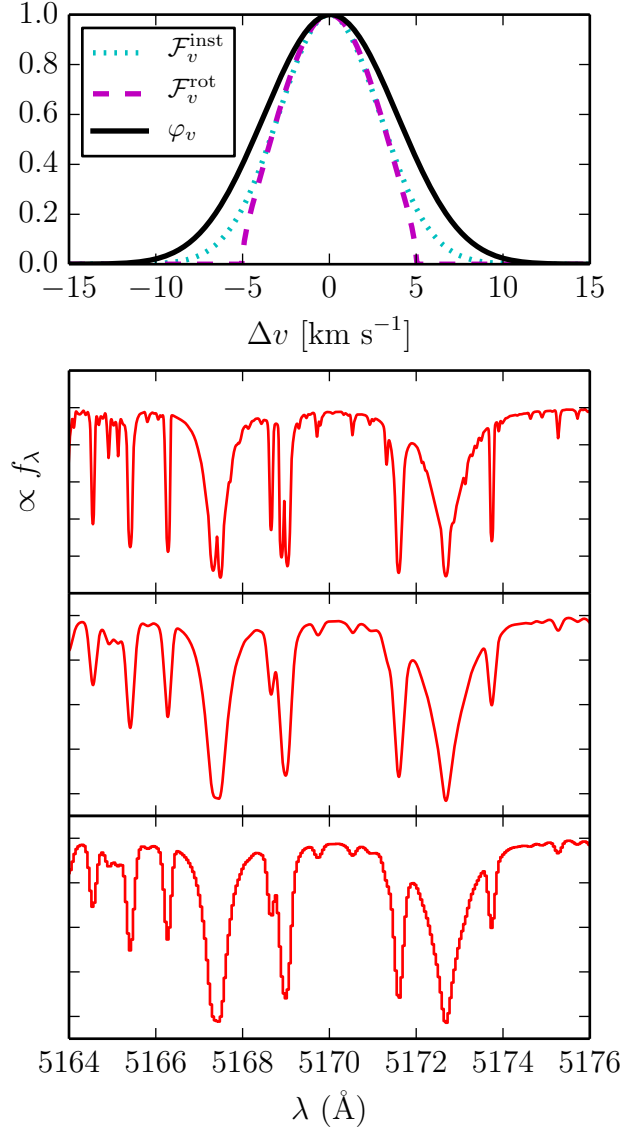


Fig. 1.— Panel 1: the instrumental ( $\text{FWHM} = 6.8 \text{ km s}^{-1}$ ) and rotational ( $v \sin i = 5 \text{ km s}^{-1}$ ) broadening kernels, and their composite. Panel 2: A section of the raw, high resolution synthetic model at  $R=500,000$ . Panel 3: The spectrum after applying the composite broadening kernel,  $\varphi_v$ . Panel 4: The broadened spectrum downsampled to the pixels of the detector.

an appropriate polynomial function. But that “normalization” procedure implicitly assumes that there is no relevant information content on those mismatched scales; if  $\theta$  also contributes to the broad spectral shape, then adopting this approach will corrupt the inferences of these parameters. Moreover, in practice this approach is limited, since defining an appropriate polynomial becomes

difficult in cases where the spectral line density is high (e.g., molecular bands for cool stars).

We adopt a somewhat analogous approach to deal with this issue, but cast it *internal* to the modeling framework to appropriately propagate the uncertainty introduced by additional degrees of freedom in the model. The residual calibration errors are treated as an explicit contribution to the model spectrum, enabling us to explore the distribution of possible calibrations (see Section 2.4) and then eventually marginalize out the associated nuisance parameters. In essence, the inferences for the relevant stellar parameters will properly account for the underlying uncertainty in the calibration process. This is achieved in practice by distorting the model spectrum with a (low-order) Chebyshev polynomial (e.g., Eisenstein et al. 2006; Koleva et al. 2009),

$$M(\boldsymbol{\theta}, \boldsymbol{\Theta}_{\text{cheb}}) = M(\boldsymbol{\theta}) \times \sum_n c_n T_n, \quad (5)$$

where  $T_n$  are the standard Chebyshev functions of order  $n$  and the coefficients are treated as a set of nuisance (hyper-)parameters,  $\boldsymbol{\Theta}_{\text{cheb}} = \{c_n\}^{\text{order}}$ , for each available spectral order in the dataset. With judicious priors on  $\boldsymbol{\Theta}_{\text{cheb}}$ , we can ensure that the unintended treatment of real spectral features (e.g., broad, deep molecular bands) as calibration artifacts is negligible (see Section 3 for examples). The lowest-order (scaling) coefficient,  $c_0$ , is by its nature degenerate with the solid angle parameter,  $\Omega$ . Therefore, we enforce an additional constraint by requiring that the mean of the polynomial is unity. For data with a single spectral order, this means setting  $c_0 = 1$ . If the goal is to model multiple spectral orders, we assign  $c_0 = 1$  in an arbitrary order as an anchor, but permit the  $c_0$  values in other orders to be different (as necessary). It is worth noting that this formalism can, in principle, be extended to develop models for completely uncalibrated spectra, rather than the residual calibration mismatches as described here (with a suitable relaxation of the priors on  $\boldsymbol{\Theta}_{\text{cheb}}$ ). Figure 2 offers a practical demonstration of how these nuisance parameters are applied.

### 2.3. Model Evaluation

The quality of the model spectrum is assessed by comparing to the data with a pixel-by-pixel likelihood calculation. If we denote the data spectrum as  $D$ , then a corresponding residual spectrum (of  $N_{\text{pix}}$  pixels) can be defined for any input parameter set,

$$R \equiv R(\boldsymbol{\theta}, \boldsymbol{\Theta}_{\text{cheb}}) \equiv D - M(\boldsymbol{\theta}, \boldsymbol{\Theta}_{\text{cheb}}). \quad (6)$$

To quantify the probability of the data conditioned on the model, we adopt a standard multi-dimensional Gaussian likelihood function,

$$p(D|M) = \frac{1}{[(2\pi)^{N_{\text{pix}}} \det(C)]^{1/2}} \exp\left(-\frac{1}{2} R^T C^{-1} R\right), \quad (7)$$

which penalizes models that yield larger residuals and explicitly allows for covariances in the residual spectrum through the  $N_{\text{pix}} \times N_{\text{pix}}$  matrix,  $C$ . For practical numerical reasons, we generally consider

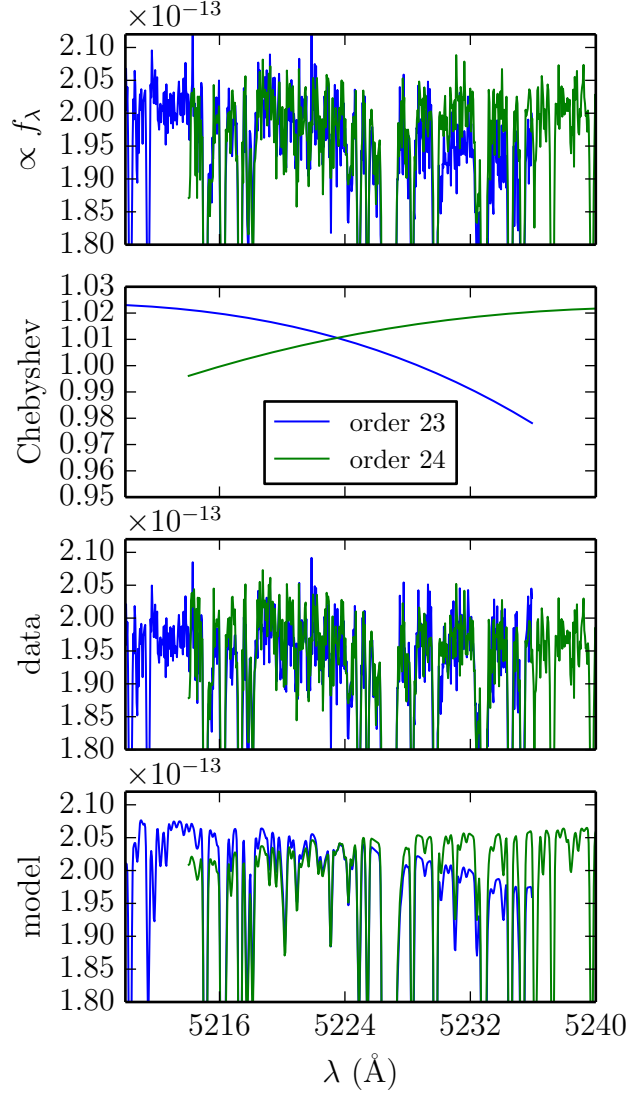


Fig. 2.— The spectrum at the overlap of two echelle orders (23 & 24). **Panel 1:** The flux-calibrated dataset shows a slight discrepancy of  $\lesssim 3\%$  between orders. **Panel 2:** To account for this residual error in the flux calibration, we multiply the model spectrum by a Chebyshev polynomial, whose coefficients are parameters that are part of the model. **Panels 3 & 4:** In principle, we could divide the data by these polynomials to recover what the true flux-calibrated data should be (panel 3), but instead we multiply the model by the polynomials to recover the original discrepancy between orders (panel 4). This has the advantage of preserving the original dataset as a fixed quantity and makes the model (Equation 5) linear in the Chebyshev polynomial coefficients.



the log-likelihood as the fit quality metric, which in this case is

$$\ln p(\mathbf{D}|\mathbf{M}) = -\frac{1}{2} \left( \mathbf{R}^\top \mathbf{C}^{-1} \mathbf{R} + \ln \det \mathbf{C} + N_{\text{pix}} \ln 2\pi \right). \quad (8)$$

The covariance matrix  $\mathbf{C}$  characterizes both the measurement uncertainty ( $\sigma$ ; “noise”) in each pixel and the intrinsic covariance between pixels. The special case where each pixel represents an independent measurement results in a diagonal covariance matrix,  $\mathbf{C}_{ij} = \delta_{ij} \sigma_i$  where  $\sigma_i$  is the uncertainty in pixel  $i$  and  $\delta_{ij}$  is the Kronecker delta function, and Eq. 8 reduces to the familiar

$$\ln p(\mathbf{D}|\mathbf{M}) = -\frac{1}{2} \sum_i^{N_{\text{pix}}} \frac{\mathbf{R}_i^2}{\sigma_i^2} \equiv -\frac{\chi^2}{2}, \quad (9)$$

the sum of the square of the residuals weighted by the inverse variances (squared uncertainties). However, the problem being addressed here necessitates the use of a more complex covariance matrix; additional off-diagonal terms that can explicitly characterize (1) pixel-to-pixel covariances imposed by the discrete over-sampling of the line-spread function, and (2) highly correlated residuals as manifestations of the still-imperfect model library are required to avoid biasing our inferences of the physically interesting parameters ( $\boldsymbol{\theta}$ ). The following subsections describe how these issues are addressed in the practical implementation of  $\mathbf{C}$ .

### 2.3.1. Global Covariance Structure

Astronomical spectrographs are designed so that the detector over-samples the instrumental line broadening function with at least a few pixels. Therefore, adjacent pixels are never completely independent samples of the observed spectrum. In that case, a difference between an observed and modeled spectral feature will create a residual that spans multiple pixels. This can be demonstrated in practice by examining the autocorrelation of the residual spectrum: a slight model mismatch will produce correlated residuals over a characteristic scale similar to the observational line-broadening kernel width. Figure 3 highlights a specific example of these correlated residuals in real data, where an imperfect model generates residuals which exhibit a significant autocorrelation signal on the scale of  $\sim 4$  pixels, which corresponds to the typical line width in this spectrum.

It seems important to distinguish here between “noise” and the fit residuals. Noise introduced to the spectrograph by astrophysical or instrumental effects is generally uncorrelated with wavelength. The arrival of each photon to the detector is an independent event; while these photons are scattered by the instrumental line-spread function, the magnitude and direction of that scatter is independent for each such event. In essence, the noise itself is uncorrelated, but the fit residuals likely are correlated. However, from a mathematical perspective the correlated residuals can be treated in the same way as correlated noise, by constructing a non-trivial covariance matrix with off-diagonal terms. In practice, this is achieved by parameterizing  $\mathbf{C}$  with a kernel that describes the

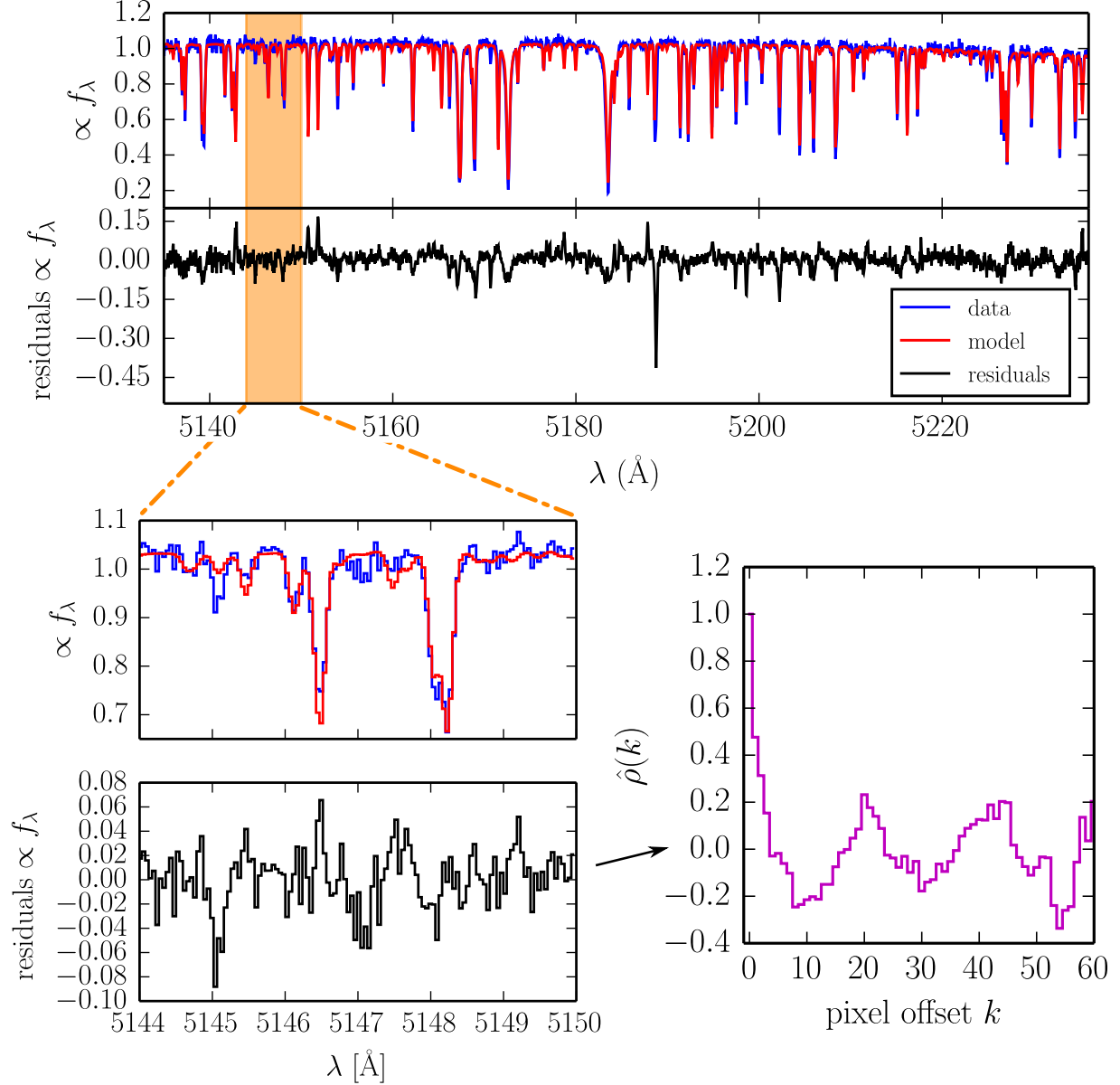


Fig. 3.— **Top**: the big spectrum. **Left** The same low-amplitude, mildly covariant residuals from Figure 5, panel 0, but enlarged to show the mildly covariant structure produced by slight mismatch between the data and model spectra. **Right** The autocorrelation of the residual sequence shown at left. Notice that there is significant correlation for offsets of  $\lesssim 4$  pixels.

covariance between any pair of pixels, representing wavelengths  $\lambda_i$  and  $\lambda_j$  (in many ways analogous to the standard two-point spatial correlation function used in cosmology).

For a well-designed spectrograph and sufficiently accurate model, this *global* (i.e., present throughout the spectrum) covariant structure should have a relatively low amplitude and small correlation length. To describe that structure, we use a stationary covariance kernel (or radial basis function) with an amplitude that depends only on the velocity distance between two pixels,

$$r_{ij} = r(\lambda_i, \lambda_j) = \Delta v = \frac{c}{2} \left| \frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j} \right|, \quad (10)$$

where  $c$  is the speed of light. The kernel is chosen to parametrically describe the covariance between pixel residuals, such that

$$\mathcal{K}_g(\lambda_i, \lambda_j) = \langle \mathbf{R}_i \mathbf{R}_j \rangle. \quad (11)$$

A variety of kernels have been used in the field of Gaussian processes to parameterize such a mildly covariant structure (e.g., [Rasmussen & Williams 2005](#)). We adopt the Matérn kernel with  $\nu = 3/2$ ,

$$\mathcal{K}_g(\lambda_i, \lambda_j | a_g, \ell) = a_g \left( 1 + \frac{\sqrt{3} r_{ij}}{\ell} \right) \exp \left( -\frac{\sqrt{3} r_{ij}}{\ell} \right), \quad (12)$$

which is parameterized by an amplitude  $a_g$  and scale  $\ell$  and makes for a smooth transition to negligible covariance at large  $r$ . To ensure that  $\mathbf{C}$  remains a relatively sparse matrix that enables computational expediency, we employ a Hann window function

$$w(\lambda_i, \lambda_j | r_0) = \begin{cases} \frac{1}{2} + \frac{1}{2} \cos \left( \frac{\pi r_{ij}}{r_0} \right) & r_{ij} \leq r_0 \\ 0 & r_{ij} > r_0 \end{cases} \quad (13)$$

to taper the kernel (so the effective kernel is  $w\mathcal{K}_g$ ). The truncation distance  $r_0$  can be fixed to some reasonable multiple of the scale parameter (here we set  $r_0 = 4\ell$ ). The examples in [Figure 4](#) demonstrate that such a kernel readily produces correlated structure in the residual spectrum that is similar to those seen in a typical data–model comparison.

### 2.3.2. Local Covariance Structure

Aside from the global covariance structure described above, there are likely also local regions of strong, highly correlated residuals that need to be treated in the modeling framework. These large amplitude residual regions are usually produced by imperfect spectral lines in the models (e.g., missing opacity sources, uncertain oscillator strengths, etc.); some representative examples are highlighted in [Figure 5](#). To parameterize such regions in the covariance matrix, we introduce a sequence of non-stationary kernels that explicitly depend on the actual wavelength values of a pair of pixels (on  $\lambda_i$  and  $\lambda_j$ ), and not simply the distance between them ( $r_{ij}$ ).

Assuming that these local residuals are produced primarily by pathological differences in the spectral line strength (rather than shape or center), a simple Gaussian is a reasonable residual

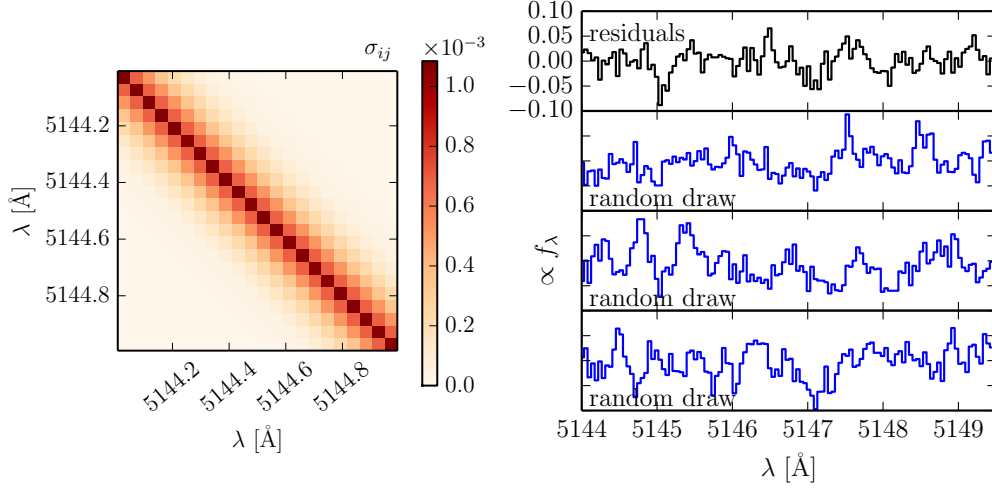


Fig. 4.— **Left** Inset zoomed to show a small region of a typical covariance matrix, generated using the kernel in Equation 12 and common values for the hyperparameters. There is a small degree of covariance in elements within a few pixels off the diagonal, which quickly tapers off so that the majority of the matrix remains sparse ( $\sigma_{ij} = 0$ ). This matrix is used to model a spectrum which is correlated on a  $\sim$ few pixel scale. **Right** To demonstrate that this matrix properly models the correlated structure of the residuals, we compare the residuals to random residuals generated from a multivariate normal distribution with this covariance matrix. The top panel shows the same residuals shown in Figure 3, and below are plotted three sets of simulated residuals. The amplitude and correlation length of the simulated residuals closely approximates the structure of the actual pixel residuals.

model. In that case, the  $k^{\text{th}}$  such local residual can be described as

$$R_\lambda(a_k, \mu_k, \sigma_k) = \frac{a_k}{\sqrt{2\pi}\sigma_k} \exp \left[ -\frac{r^2(\lambda, \mu_k)}{2\sigma_k^2} \right], \quad (14)$$

with an amplitude  $a_k$ , a mean wavelength  $\mu_k$ , and a width  $\sigma_k$ . Following Eq. 11, the kernel that describes the covariance of any two pixels related to the  $k^{\text{th}}$  residual region is

$$\mathcal{K}_k(\lambda_i, \lambda_j | a_k, \mu_k, \sigma_k) = \frac{1}{2\pi} \left( \frac{a_k}{\sigma_k} \right)^2 \exp \left[ -\frac{r^2(\lambda_i, \mu_k) + r^2(\lambda_j, \mu_k)}{2\sigma_k^2} \right]. \quad (15)$$

We again taper this kernel with a Hann window (Eq. 13) to ensure computational efficiency with a sparse covariance matrix; in this case, the truncation distance  $r_0$  can be set to some multiple of the Gaussian width (we set  $r_0 = 4\sigma_k$ ). Figure 6 demonstrates how this non-stationary Gaussian kernel generates a localized region of enhanced variance that successfully mimics the kind of residuals produced by an inaccurate spectral line model. These kernels down-weight the influence of such strong residuals in the likelihood calculation, mitigating any potential bias they might induce on

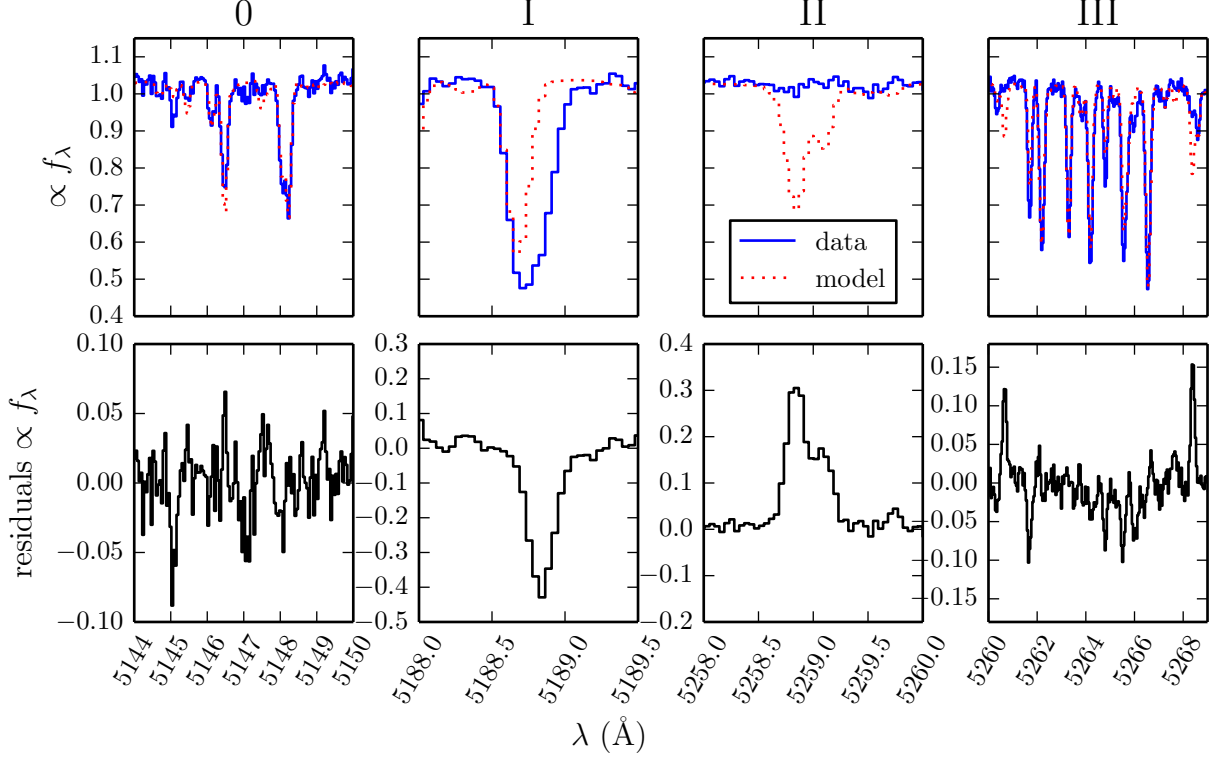


Fig. 5.— A collection of spectral lines which have imperfect model fits. From left to right: **Class 0** The majority of spectral lines ( $\gtrsim 60\%$ ) will have minor differences in strength between data and model spectrum, which produce low-amplitude correlations in the residuals on the length scale of the width of a typical spectral line. **Class I:** Sometimes ( $\lesssim 5\%$  of all lines), a missing opacity source in the model (in this case a line-blended Ca II) leaves a large, highly correlated patch of negative residuals. **Class II:** Sometimes ( $\lesssim 5\%$  of all lines), an extraneous line in the model leaves a large, highly correlated patch of positive residuals. **Class III:** If the line strengths are substantially discrepant ( $\lesssim 10\%$  of all lines), there will be many correlated residuals of moderate amplitude. The difficulty with class III lines is that for any specific line, there might exist a  $\theta_\star$  that will fit the line, but there does not exist a  $\theta_\star$  that will properly fit *all* the lines.

inferences of the interesting parameters ( $\theta$ ). In essence, this is a robust, flexible, and unbiased method for (correlated) outlier rejection that preserves the integrity of the probabilistic framework being developed (as opposed to manual or threshold-based clipping or masking).

**SA:** These local kernels can be further modified to account for more complex residual structures. For example, late-type stars with imperfectly modeled molecular bandheads may produce a complicated pattern of positive and negative residuals or a pronounced mismatch over a relatively large spectral scale. This phenomenologically different local covariance behavior can still be treated in this framework, if we design an appropriate covariance kernel. Given the behavior of molecular

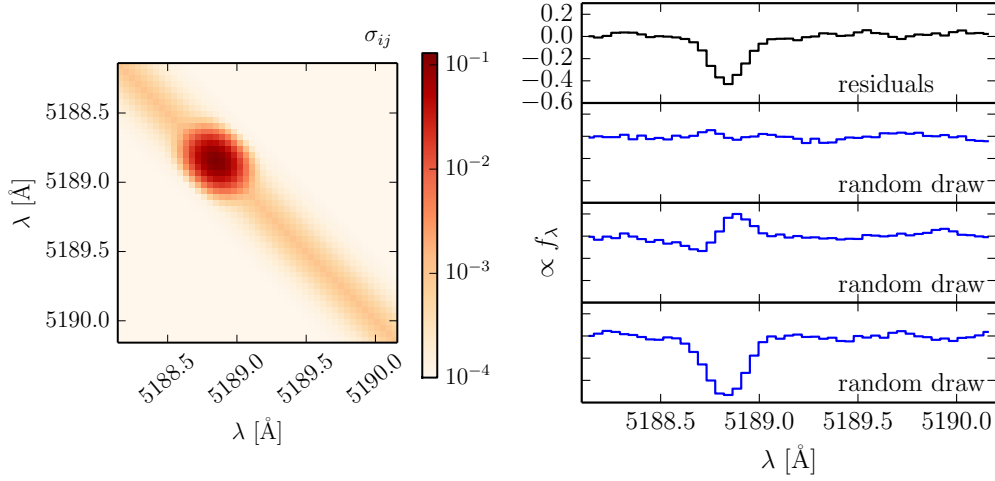


Fig. 6.— **Left** A typical covariance matrix including the Gaussian line kernel (Equation 15). The same global covariance shown in Figure 4 is still present with the same hyperparameters, however now there is an additional patch of high covariance corresponding to the large, Gaussian-shaped residuals. These larger elements in the covariance matrix effectively down-weight the contribution of a poorly modeled synthetic spectral line. **Right** The same spectroscopic residuals shown in Figure 5, class I, shown with three random draws from the covariance matrix. The random draws can take on a range of amplitudes—positive, negative, or even flat—because they are simply random draws that are described by the covariance matrix. The wide range of possible residual amplitudes match the structure and amplitude of the pixel residuals.

bandheads, a Matérn kernel tapered by the Gaussian kernel in Eq. 15 works well.  
 TODO: We'll need to round out the Gl 51 example in Sect 3.2 before this text is finished.

### 2.3.3. Composite Covariance Matrix

We can now compute the covariance matrix employed in the likelihood calculation (Eq. 8) as the linear combination of these kernels and the trivial pixel-by-pixel noise matrix,

$$\begin{aligned} \mathbf{C}_{ij}(\boldsymbol{\Theta}_{\text{cov}}) = & b \delta_{ij} \sigma_i + \\ & w(r_{ij}|r_0 = 4\ell) \mathcal{K}_g(\lambda_i, \lambda_j|a_g, \ell) + \\ & \sum_k w(r_{ij}|r_0 = 4\sigma_k) \mathcal{S}_k(\lambda_i, \lambda_j|h_k) \mathcal{K}_k(\lambda_i, \lambda_j|a_k, \mu_k, \sigma_k), \end{aligned} \quad (16)$$

where the covariance hyperparameters  $\boldsymbol{\Theta}_{\text{cov}} = [a_g, \ell, \{a_k, \mu_k, \sigma_k, h_k\}^{N_{\text{reg}}}]$ ,  $N_{\text{reg}}$  is the number of local residual regions (see below for details on how this is determined), and  $b$  is a fixed parameter that scales the pixel noise values to account for additional error (e.g., introduced by read noise, order

extraction, and the grid interpolation noted in Section 2.1; reasonable values are  $b \approx 1.02$ – $1.10$  for well-calibrated optical spectra, and examples are provided in Section 3).

## 2.4. Exploring the Posterior

To explore the posterior probability distribution and make inferences on the parameters of interest, we employ a hierarchical Monte Carlo Markov Chain (MCMC) approach that nests the standard Metropolis-Hastings algorithm (for sampling the parameters  $\theta$ ) inside the Gibbs sampler (which rotates through subsets of the hyper-parameters  $\Theta$ ).

The Gibbs sampler iteratively explores this hierarchy of parameters. At each level of the hierarchy, corresponding to a different subset of nuisance parameters, we use the Metropolis-Hastings algorithm to propose a new subset of parameters and then either accept or reject them. This process of proposal and acceptance/rejection is called “sampling.” The Gibbs sampler rotates between sampling in different subsets of parameters, eventually sampling all of the nuisance parameters at a certain cadence. To initialize the MCMC algorithm, we make a reasonable guess for the starting parameters. We use existing spectral types in the literature for the stellar parameters, flat Chebyshev polynomials (no flux-calibration correction), and no covariance structure in the residuals (the global covariance parameters are set to zero and no line kernels are instantiated). We use superscripts to denote iterations of the MCMC algorithm.  $i$  denotes the parameters from the current iteration and  $i - 1$  denotes the previous iteration. The Gibbs sampler rotates among subsets of the parameters as follows

1. Sample in the stellar parameters. For each proposal of the Metropolis-Hastings algorithm, generate a model spectrum following the steps in §2.2. When deciding whether or not to accept the parameters, the Gibbs sampler evaluates  $p(\theta_{\star}^i | \theta_{\text{orders}}^{i-1})$ .
2. For each order of the echelle spectrum
  - (a) Sample in the Chebyshev polynomial parameters. Adjust the spectrum as detailed in §2.2. Gibbs sampler evaluates  $p(\theta_{\text{Cheb}}^i | \theta_{\star}^i, \theta_{\text{global}}^{i-1}, \theta_{\text{lines}}^{i-1})$ .
  - (b) Sample in the global covariance parameters. Adjust the covariance matrix  $C$  as described in §2.3.1. Gibbs sampler evaluates  $p(\theta_{\text{global}}^i | \theta_{\star}^i, \theta_{\text{Cheb}}^i, \theta_{\text{lines}}^{i-1})$ .
  - (c) Check to see whether the algorithm should instantiate/delete new/old line kernels
  - (d) For each line kernel  $k$ 
    - i. Sample in the parameters for each line, conditional on the parameters for all the other lines. Adjust the covariance matrix  $C$  as described in §2.3.2. Gibbs sampler evaluates  $p(\theta_{\text{line}_k}^i | \theta_{\text{lines} \neq k}^{i-1})$ .

At each stage of the Gibbs sampler, the likelihood function (Equation 8) is evaluated for a proposal of a particular subset of parameters, conditional on the current values of all the other

parameters. The hierarchical structure of the parameters influences how the algorithm may be efficiently implemented in code. For a typical optical spectrum with  $\gtrsim 1,000$  pixels, the most computationally intensive step of the likelihood evaluation is usually the matrix product  $\mathbf{R}^T \mathbf{C}^{-1} \mathbf{R}$ . Since we have designed the covariance matrix to be sparse, we can use optimized sparse matrix algorithms which are much faster and memory efficient than dense matrix operations. Because we are not interested in the matrix inverse  $\mathbf{C}^{-1}$  by itself, but rather the product with the residual vectors, we can use efficient routines for solving linear systems to bypass the computationally difficult step of matrix inversion. Additionally, because the covariance matrix is positive semi-definite, we can use the Cholesky factorization of the matrix to optimize the evaluation of the matrix product. Once the covariance matrix is factorized, any subsequent evaluation of the matrix product for different residual vectors  $\mathbf{R}$  is extremely rapid. This makes the  $\theta_\star$  and  $\theta_{\text{Cheb}}$  steps of the Gibbs sampler extremely fast. When we sample in the nuisance parameters which affect the covariance matrix, we must redo the Cholesky factorization of  $\mathbf{C}$  for each update. However, because we designed the kernels to deliver a sparse matrix these operations are still moderately efficient. We use the high-performance SuiteSparse/CHOLMOD<sup>2</sup> library to implement the sparse matrix and Cholesky factorization operations (Chen et al. 2008; Davis & Hager 2009) and utilize the Metropolis-Hastings sampler included in the Python MCMC package `emcee` (Foreman-Mackey et al. 2012).

We run the MCMC Gibbs sampler for many iterations until the estimate of the posterior distribution has converged. To check that the chain is not stuck in a local maximum of the posterior, we redo the MCMC run many times with different starting parameters, to ensure that the algorithm converges to the same global maximum. A major advantage of using the MCMC algorithm to explore the multidimensional probability space is that it provides numerical samples in each dimension. Therefore, marginalizing out a parameter (i.e., numerically integrating over a dimension in probability space) is as simple as combining all of the samples in this dimension. This enables us to present a posterior of the stellar parameters  $\theta_\star$  (Figure ??) which has been marginalized over all of the nuisance hyperparameters. This posterior is the final estimate of the stellar parameters which *incorporates* any inherent uncertainty due to model mismatch (via the covariance hyperparameters) and flux-calibration (via the Chebyshev polynomials).

The benefit of including nuisance parameters and then marginalizing over them is that we can self-consistently model the uncertainty inherent to any spectrum while naturally capturing any degeneracy between the model parameters. If stellar parameters are estimated using a method which ignores these nuisance parameters (e.g., by-eye fitting) and the astronomer arbitrarily inflates the parameter uncertainties to reflect intuition, the degeneracy between parameters is artificially destroyed.

---

<sup>2</sup><http://www.cise.ufl.edu/research/sparse/cholmod/>



## 2.5. Applications

By cataloguing the covariance structure of the residuals, especially those generated from strong spectral line mismatch, we collect valuable information about the quality of the synthetic spectra. After fitting many stars, the accumulated knowledge of the data-model mismatch can be used when fitting a new star. The previous structure of the covariance matrix allows us to set priors on what the covariance in certain regions of the spectrum should be, which will speed convergence for this new star. Additionally, after fitting several stars, the average value of the covariance matrix will inform us about the quality of specific spectral lines in the synthetic models.

Linking the covariance matrices of stars could be done serially, where the aggregate covariance matrix of all previous fits is used as a prior for the current iteration. Or, the covariance matrices could instead be linked hierarchically, in that the parameters describing the depth and width of a line residual for a specific star  $a_1$  and  $\sigma_1$ , are modeled as coming from of a population of possible depths and widths for a given spectral type. Each stellar spectrum will have a slightly different realization of a spectral line, which will have some scatter about the average residual height. Linking the covariance matrix between spectra of similar stars allows us to grow more confident in our assessment that certain synthetic spectral lines are indeed outliers and should be appropriately down-weighted. In turn, as we become more certain of the weights, the stellar posteriors will become narrower and make our estimates of the stellar parameters more precise. This ability of the model to mutually inform sets of parameters is one of the major advantages of hierarchical Bayesian analysis (Kruschke 2010).

Once determined, this average covariance matrix could be delivered to the communities that created the synthetic libraries, which would enable them to rapidly pinpoint and correct non-physical lines in the synthetic models. Alternatively, we could correct the models ourselves by using the chain of logic and mathematical post-processing that we used to created the synthetic model spectrum to reverse-engineer what the behavior of the raw synthetic spectrum *should* be, at the raw  $R \gtrsim 100,000$  resolution. This fundamental application of machine learning would enable us to create our own library of data driven, semi-empirical stellar models. Rather than simply assembling an empirical spectral library using only real stellar spectra, this combined approach is more powerful because the synthetic stellar atmospheres provide an actual anchor point of fundamental stellar parameters tied down by the laws of stellar physics.

## 3. Examples

### 3.1. Fitting data

To demonstrate our technique, we fit spectra of two different stars with varying stellar properties and acquired with different instruments. The first object is WASP-14, an F5 star (Joshi et al. 2009). The spectrum was obtained with the Tillinghast Reflector Echelle Spectrograph (*TRES*)

(Mink 2011) on the FLWO 1.5m telescope on June 14th, 2009. *TRES* delivers a 51 echelle spectrum spanning the full optical wavelength spectrum  $3860 - 9100 \text{ \AA}$  at a resolution of  $R \sim 44,000$ . Torres et al. (2012) acquired the spectrum of WASP-14 to study the stellar properties of transiting exoplanet hosts. The raw spectrum was reduced using the well-tested *TRES* pipeline written by L. Buchhave. In the Mg b order ( $\sim 5100 \text{ \AA}$ ), the spectrum has a signal to noise of 150 per resolution element.

The SPC analysis within Torres et al. (2012) focused on a region of spectrum from 50005300  $\text{\AA}$  and used many methods. The other methods are not directly comparable because they use different underlying assumptions. Additionally, there are systematics introduced by metallicity assumptions and radiative transfer. Using this same spectral library as SPC, we recover the same parameters.

When we fit this spectrum using the same assumptions, (amounting to a  $\delta$ -function prior on  $\log g$ . global noise amplitude is so low, almost non-existent. This demonstrates sensitivity to a prior as well as the fact that surface gravity estimates from spectroscopy may be biased against fits derived from planetary transits.

The second object is Gl 51, an M5 star. The spectrum was obtained with the *SPEX* instrument at IRTF. The spectrum is at moderate resolution ( $R \sim 2,000$ ), mid-infrared spectrum from  $0.7 - 5.5 \mu\text{m}$ . Gl 51 is a M5 star from the IRTF spectral library (Cushing et al. 2005; Rayner et al. 2009) taken with the *SPEX* instrument (Rayner et al. 2003). (Rojas-Ayala et al. 2012) use the H20 -K index (defined over what wl range?).

A Bayesian approach allows you to explore sensitivity to prior assumptions.

### 3.2. Generic Tests

Designed to show off what the model framework can actually do.

Talk about giving more reasonable errors using regions, in the biased case. Show a figure of the regions determined, plotted with the increased variance envelope over the residuals.

- Spectral parameters can vary by a large margin depending on which spectral library you use (200  $K$  or more).
- Both spectral libraries have stars that they perform better and worse on.
- This will be reflected in the increased level of global noise, and number of “bad” regions that have been instantiated.

Summary end of this section is that we are

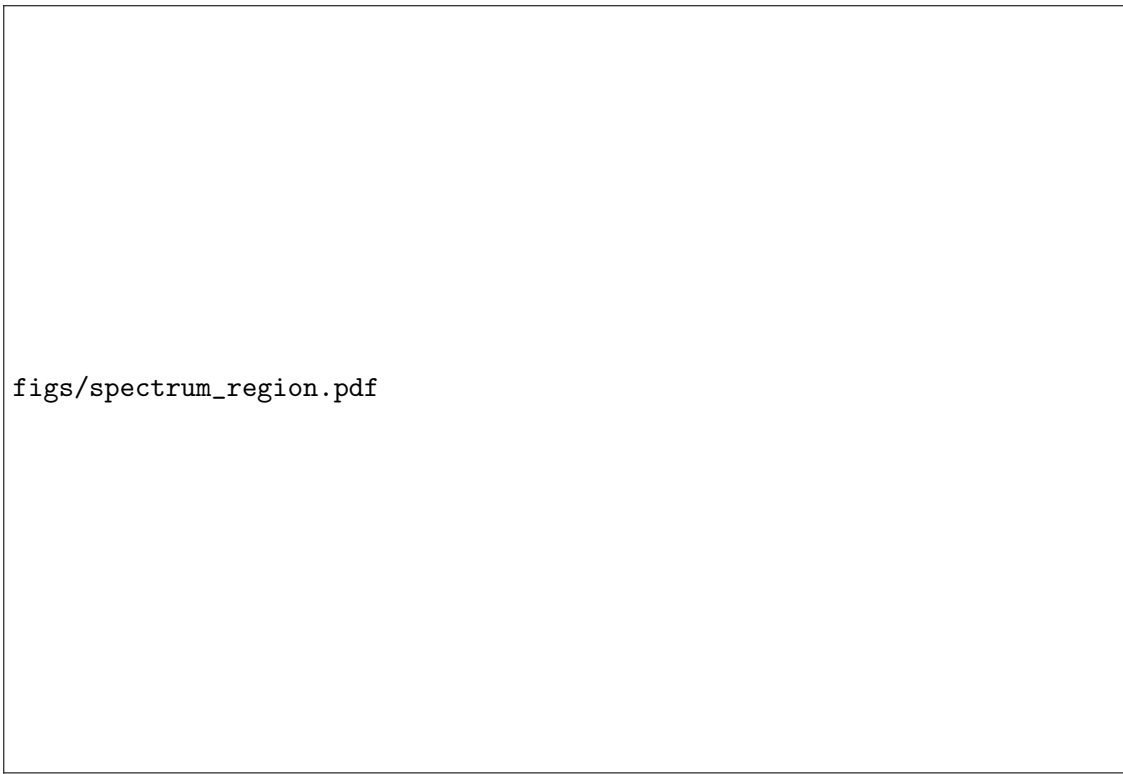


Fig. 7.— The spectrum residuals and many random draws from the final covariance matrix showing how the uncertainty in the “outlier” lines is identified by the local covariance kernels.

1) validating other techniques (WASP14, Kurucz, etc) 2) and two different synthetic libraries give different answers

this can feed into a discussion about tweaking models.

#### 4. Discussion

We can (should?) add an additional level to the hierarchy of hyperparameters and add parameters to describe the *population characteristics* of poorly modeled spectral lines (mostly the typical width, amplitude of these lines). This will tell us about the frequency and distribution of spectral modelling errors.

## 5. Conclusion

## REFERENCES

- Allard, F., Homeier, D., & Freytag, B. 2012, [Royal Society of London Philosophical Transactions Series A](#), **370**, 2765
- Buchhave, L. A., Latham, D. W., Johansen, A., et al. 2012, [Nature](#), **486**, 375
- Buchhave, L. A., Bizzarro, M., Latham, D. W., et al. 2014, [Nature](#), **509**, 593
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, [ApJ](#), **345**, 245
- Castelli, F., & Kurucz, R. L. 2004, ArXiv Astrophysics e-prints, [astro-ph/0405087](#)
- Chen, Y., Davis, T. A., Hager, W. W., & Rajamanickam, S. 2008, [ACM Trans. Math. Softw.](#), **35**, 22:1
- Cushing, M. C., Rayner, J. T., & Vacca, W. D. 2005, [ApJ](#), **623**, 1115
- Davis, T. A., & Hager, W. W. 2009, [ACM Trans. Math. Softw.](#), **35**, 27:1
- Eisenstein, D. J., Liebert, J., Harris, H. C., et al. 2006, [ApJS](#), **167**, 40
- Fischer, D. A., & Valenti, J. 2005, [ApJ](#), **622**, 1102
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2012, ArXiv e-prints, [arXiv:1202.3665 \[astro-ph.IM\]](#)
- Gray, D. F. 1994, [PASP](#), **106**, 1248
- . 2008, *The Observation and Analysis of Stellar Photospheres*
- Hauschildt, P. H., Allard, F., & Baron, E. 1999, [ApJ](#), **512**, 377
- Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, [Science](#), **330**, 653
- Husser, T.-O. 2012, *3D-Spectroscopy of Dense Stellar Populations* (Universitätsverlag Göttingen)
- Husser, T.-O., Wende-von Berg, S., Dreizler, S., et al. 2013, [A&A](#), **553**, A6
- Johnson, J. A., Butler, R. P., Marcy, G. W., et al. 2007, [ApJ](#), **670**, 833
- Joshi, Y. C., Pollacco, D., Collier Cameron, A., et al. 2009, [MNRAS](#), **392**, 1532
- Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, [A&A](#), **501**, 1269
- Kruschke, J. 2010, *Doing Bayesian Data Analysis: A Tutorial Introduction with R* (Academic Press)

- Kurucz, R. L. 1993, SYNTHE spectrum synthesis programs and line data
- Mann, A. W., Gaidos, E., & Ansdell, M. 2013, [ApJ](#), **779**, 188
- Mink, D. J. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 305
- Paxton, B., Bildsten, L., Dotter, A., et al. 2011, [ApJS](#), **192**, 3
- Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning) (The MIT Press)
- Rayner, J. T., Cushing, M. C., & Vacca, W. D. 2009, [ApJS](#), **185**, 289
- Rayner, J. T., Toomey, D. W., Onaka, P. M., et al. 2003, [PASP](#), **115**, 362
- Reid, I. N., Hawley, S. L., & Gizis, J. E. 1995, [AJ](#), **110**, 1838
- Rojas-Ayala, B., Covey, K. R., Muirhead, P. S., & Lloyd, J. P. 2010, [ApJ](#), **720**, L113
- . 2012, [ApJ](#), **748**, 93
- Schönrich, R., & Bergemann, M. 2013, ArXiv e-prints, [arXiv:1311.5558 \[astro-ph.SR\]](#)
- Shkedy, Z., Decin, L., Molenberghs, G., & Aerts, C. 2007, [MNRAS](#), **377**, 120
- Snedden, C. A. 1973, PhD thesis, The University of Texas at Austin.
- Tonry, J., & Davis, M. 1979, [AJ](#), **84**, 1511
- Torres, G., Fischer, D. A., Sozzetti, A., et al. 2012, [ApJ](#), **757**, 161
- Valenti, J. A., & Piskunov, N. 1996, *A&AS*, **118**, 595

## 6. Tables

This version of the paper was generated from a git repository available at <http://github.com/iancze/StellarSpectra/> with git hash ad85d14 (2014-07-06).

Table 1. Nomenclature used in this document

Symbol	Description
$i$	index specifying a pixel
$\lambda_i$	wavelength corresponding to a given pixel $i$
$\boldsymbol{\theta}_{\star,\text{grid}}$	fundamental stellar parameters, $T_{\text{eff}}, \log(g), [\text{Fe}/\text{H}], [\alpha/\text{Fe}]$ that parameterize a synthetic spectrum from the grid
$\boldsymbol{\theta}_{\star,\text{post}}$	stellar parameters $v \sin i, v_z, A_V$ , and $R^2/d^2$ that are applied during “post processing” of the synthetic spectrum
$\boldsymbol{\theta}_{\star}$	$\{\boldsymbol{\theta}_{\star,\text{grid}}, \boldsymbol{\theta}_{\star,\text{post}}\}$
$f_{\lambda,\text{inst}}(\lambda)$	data spectrum
$f_{\lambda,\text{synth}}(\lambda)$	synthetic spectrum
$\boldsymbol{\theta}_{\text{Cheb}}$	the set of Chebyshev polynomial coefficients $\{c_0, c_1, \dots, c_N\}$
$\boldsymbol{\theta}_{\text{line}_1}$	
$\boldsymbol{\theta}_{\text{lines}}$	
$\boldsymbol{\theta}_{\text{order}_1}$	
$\boldsymbol{\theta}_{\text{orders}}$	
$\boldsymbol{\theta}$	the parameters $\{\boldsymbol{\theta}_{\star,\text{grid}}, \boldsymbol{\theta}_{\star,\text{post}}, \boldsymbol{\theta}_{\text{N}}\}$ that completely describe a model spectrum
$D_i$	data flux for a given pixel, $D(\lambda_i)$
$\vec{D}$	data vector comprised of all $D_i, i = \{1, \dots, N\}$
$M_i$	model flux for a given pixel, $M(\lambda_i \boldsymbol{\theta})$
$\vec{M}$	model vector comprised of all $M_i, i = \{1, \dots, N\}$
$\sigma_i$	Poisson noise for a given pixel $i$
$R_i$	residuals $D_i - M_i$
$\mathbf{R}$	residual vector $\vec{D} - \vec{M}$
$C$	covariance matrix
$\sigma_{ij}$	element in the covariance matrix
$r(\lambda_i, \lambda_j)$	radial distance in wavelength space corresponding to $\Delta v$
$k_{\text{global}}$	global covariance kernel
$k_{\text{line}}$	regional covariance kernel

Table 2. Tests

Object	Orders	$\lambda$ range [ $\text{\AA}$ ]	library	$T_{\text{eff}} \pm \sigma$	$\log g \pm \sigma$	$[\text{Fe}/\text{H}] \pm \sigma$	comments
WASP-14	22-24	5060-5315	Kurucz	$6310 \pm 52$	$3.91 \pm 0.08$	$-0.38 \pm 0.03$	Poisson only
WASP-14	22-24	5060-5315	Kurucz	$6314 \pm 50$	$3.92 \pm 0.08$	$-0.39 \pm 0.03$	Matern only
WASP-14	22-24	5060-5315	Kurucz	$6512 \pm 32$	$\dots$	$-0.27 \pm 0.02$	Matern + fixed $\log g = 4$
WASP-14	23		PHOENIX	$6021 \pm 16$	$3.83 \pm 0.03$	$-0.50 \pm 0.01$	Poisson only
WASP-14	23		PHOENIX	$5965 \pm 70$	$3.78 \pm 0.09$	$-0.68 \pm 0.07$	Matern only, no regions (conv)
WASP-14	23		PHOENIX	$5883 \pm 55$	$3.71 \pm 0.07$	$-0.74 \pm 0.05$	Matern and regions (conv)
WASP-14	22-24	5060-5315	PHOENIX	$6117 \pm 30$	$3.73 \pm 0.06$	$-0.52 \pm 0.02$	Poisson only
WASP-14	22-24	5060-5315	PHOENIX	$5865 \pm 43$	$3.2 \pm 0.08$	$-0.85 \pm 0.03$	Matern only. Because of
WASP-14	22-24	5060-5315	PHOENIX	$\pm$	$\dots$	$\pm$	Matern + fixed $\log g = 4$
WASP-14	22-24	5060-5315	PHOENIX	$\pm$	$\pm$	$\pm$	Regions included