Simulation-Based Inference for Detecting Blending in Spectra

Declan McNamara
Department of Statistics
University of Michigan
Ann Arbor, MI 48109
declan@umich.edu

Jeffrey Regier
Department of Statistics
University of Michigan
Ann Arbor, MI 48109
regier@umich.edu

Abstract

Many galaxies overlap visually from the vantage point of Earth; these galaxies are known as "blends". Undetected blends can lead to errors in the estimation of quantities of scientific interest, such as cosmological parameters and redshift. We propose a generative model based on a state-of-the-art simulator of galaxy spectra, and develop a simulation-based inference method to detect unrecognized blends. Our inference routine simulates both blended and unblended spectra with which it trains an inference network to solve the inverse problem, that is, to map spectra to a Bernoulli distribution indicating the presence or absence of blendedness. Our experiments demonstrate the potential of our method to detect unrecognized blends in high-resolution spectral data from the Dark Energy Spectroscopic Instrument (DESI).

1 Introduction

From the viewpoint of Earth, many galaxies overlap visually. These "blends" result in measurement contamination in astronomical data. In images, for example, a blend occurs when photons from multiple sources contribute to a pixel, while spectral blends occur when flux from multiple objects is registered on the same spectrograph fiber during an exposure. Blends are common due to the density of objects in the sky: about 62% of sources detected in images of the Legacy Survey of Space and Time (LSST) are expected to be blends [1]. Blends pose a significant difficulty to cosmology, as measured properties of blended objects do not reflect those of any individual constituent. These faulty measurements may propagate errors into further downstream cosmological estimation tasks [2, 3, 4, 5].

Detection of blends is thus critical to our understanding of cosmology. However, a significant portion of blends are undetected based on photometric data alone. Mock simulations for the LSST survey, for example, have estimated that approximately 12% of galaxies believed to be single objects are in fact unrecognized blends [6]. Detected blends in photometric data are often not targeted for spectroscopy [7], but undetected blends may still be, resulting in observed spectra that are blended.

We present a simulation-based inference approach to identifying blended astronomical spectra. While blending in spectra has been examined previously (e.g., [8]), it has not been the focus of most recent surveys, which tackle blending primarily by use of images and other photometry. However, recent advances in deep learning and astronomical simulation have sparked renewed interest in the potential to apply spectral data to the tasks of blend detection and separation [9]. Our work represents a step in this direction.

Astronomical spectra are high-dimensional, with many features potentially useful for detecting blends. However, spectra are often noisy and, in contrast to images, contain no spatial information to indicate

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

that another light source is nearby. Blends must be discerned from features present in the spectrum alone, such as emission and absorption lines [10]. A blended spectrum may exhibit i) larger or smaller emission or absorption lines than an unblended one, relative to overall flux magnitudes or ii) the "same" emission line in multiple wavelength locations if objects with different redshifts are blended.

While these spectral features may be informative for blend detection, they also pose a substantial roadblock to traditional Bayesian inference. The constituent spectra (one, two, or more) that form a noisy observation (a non-blend or blend) are unknown latent variables that must be marginalized out for inference. This task is computationally intractable in standard Bayesian methodologies due to the large dimension of latent spectra. The space to marginalize over is also trans-dimensional because the number of latent spectra is also unknown, complicating inference further. Our approach avoids these pitfalls by implicitly marginalizing out these nuisance variables through simulation, avoiding the need for explicit modeling of these complex latent quantities.

2 Generative Model

Detecting blends is an unsupervised task: though tens of thousands of spectra have been measured across various surveys, authoritative labels of the presence or absence of blending are unavailable. We thus posit a generative model of unblended and blended spectra. Our model makes use of the Probabilistic Value-Added Bright Galaxy Survey simulator (PROVABGS) [11] to generate both blended and unblended spectra (Section 2.1). As PROVABGS generates "noiseless" spectra, we must incorporate a realistic noise model for the synthetic spectra to resemble those measured in reality (Section 2.2).

2.1 Blended Spectra

The PROVABGS simulator is based on stellar population synthesis models of galaxy formation, and uses 12 parameters θ to produce synthetic spectra. We use the prior $\theta \sim p(\theta)$ on these 12 parameters suggested by [11], along with a Uniform$(0, 1)$ prior on redshift. Additionally, although the true proportion π of blends in observed spectra is unknown, we set $\pi = 0.5$ and set the prior on a blendedness indicator random variable z to be Bernoulli(π). This misspecified prior can be adjusted for posthoc (e.g., by case-controlled sampling). In the generative model, we draw z from the prior above, followed by θ_1, θ_2 and two redshifts r_1, r_2 independently from their respective priors. These yield two unblended spectra s_1', s_2' computed from the PROVABGS model (which are subsequently resampled onto a common wavelength grid).

Blending is performed by binwise addition of flux across all wavelengths; however, the priors above yield spectra on several different orders of magnitude of brightness. This can result in near-indiscernible blends. Therefore, we work with normalized and scaled spectra to construct blends that are detectable — normalized spectra have been used in related work to disentangle flux magnitudes from inference [12]. Denote the spectra after normalization by s_1, s_2, respectively. If $z = 0$, the distribution of the observed data x is modeled as a point mass at s_1, i.e. $x \sim \delta(s_1)$, whereas if $z = 1$ the observed spectrum is a blend of s_1 and s_2 constructed as $x \sim \delta(\text{Normalize}(m \cdot s_1 + s_2))$, where $m \sim \text{Uniform}(c, 1)$ and multiplication and addition are performed binwise across all wavelengths.

The scaling random variable m allows for blends where one source is significantly "brighter" than the other, as commonly occurs in reality. The source s_2 is permitted to be up to $\frac{1}{c}$ times brighter than s_1 by this construction. We take $c = 0.1$ in our experiments. The final normalization step after blending ensures that both blended and unblended spectra are on the same scale. If desired, one can rescale the data, i.e. generate $\tilde{x} = M \cdot x$ for some scalar M that multiplies the spectrum x binwise across the wavelength grid, to produce simulated spectra on the same scale as observed spectra. Together, the scalings M and m implicitly explore the threshold of detectability in spectral blends. The variable $m \in (c, 1)$ defines the *relative* brightness of the first source to the second, while M defines the *absolute* brightness of the spectrum. If either is too low, a blend may be undetectable: in the former case, because one source dominates, and in the latter because the noise model (Section 2.2) may overpower the signal.

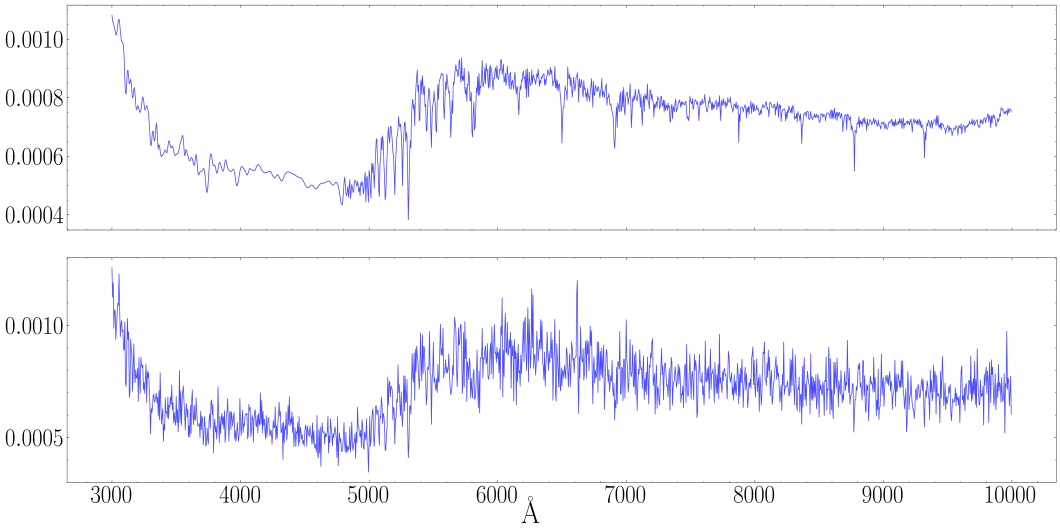

Figure 1: A noiseless blended spectrum produced from the model in Section 2.1 (top) has the constant SNR noise model applied (Section 2.2) with $\sigma = 0.1$ to obtain a more realistic spectrum (bottom).

2.2 Realistic Noise Models

We propose two noise models of different complexity. In the simpler case, we first draw samples (both blends and non-blends) from the model of spectra above, resampled onto a common wavelength grid of width 5 Angstrom (Å) between 3000 Å and 10000 Å. We work on the normalized scale (i.e. $M = 1$) throughout. We then add a simple noise model that imposes a fixed signal-to-noise ratio (SNR), constructed by multiplying the entire spectrum pointwise by a random variable $\epsilon \sim \mathcal{N}(1, \sigma^2 I_b)$ where $b = 1600$ is the dimension of the spectrum. This results in a signal-to-noise ratio (SNR) that is fixed across all wavelength bins for all observations. With μ denoting the flux at a particular wavelength for a given noiseless spectrum and $Z \sim \mathcal{N}(0, 1)$, the observed flux is modeled as

$$\mu \cdot (\sigma Z + 1) \sim \mathcal{N}(\mu, \mu^2 \sigma^2), \tag{1}$$

which has a SNR of $\frac{1}{\sigma}$. The choice of σ is a hyperparameter that we set to 0.1, which yields synthetic spectra that are similar to real spectra from SDSS. In Figure 1, we show a (resampled) blended spectrum both before and after applying the constant SNR noise model.

Our second noise model produces DESI-like spectra using the `specsim`[1] DESI noise model [13]. Given a normalized spectrum from the blending model above (resampled onto the DESI wavelength grid), we scale the signal by M and simulate the spectrograph response using Gaussian noise modeled after the atmosphere and instrumentation noise particular to DESI. As raw DESI spectra are observed on three spectrograph arms (blue (*b*), red (*r*), infrared (*z*)) [14], the simulated spectra are also observed in these three regions (cf. Figure 2).

Scaling by M is necessary because `specsim` expects signals of appropriate magnitude in units erg $\cdot$ cm^{-2}s^{-1}Å^{-1}, and the normalized spectra (both blended and unblended) are on a different scale. We experiment with a range of M values in Section 4, chosen to produce synthetic spectra of similar brightness to real DESI spectra (see Figure 2).

3 Variational Inference

We use a recently developed approach to Bayesian inference called forward amortized variational inference (FAVI) [16] to perform inference on the blendedness indicator random variable z. FAVI minimizes the forward KL divergence $\mathrm{KL}\big[p(z \mid x) || q_\phi(z \mid x)\big]$ between exact and approximate posteriors for all possible synthetic spectra x from the generative model [16, 17]. While the forward

[1] `https://github.com/desihub/specsim`

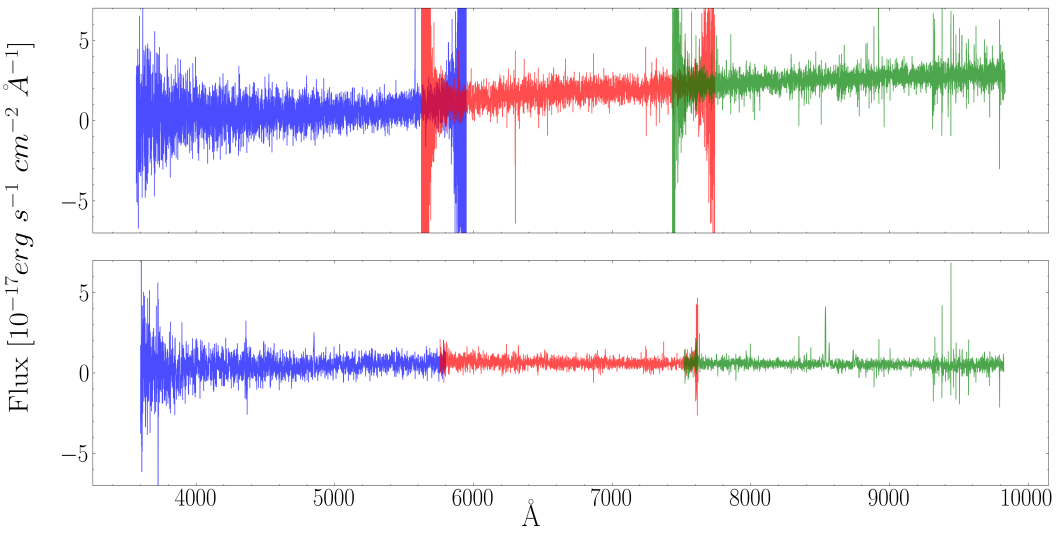

Figure 2: A simulated DESI-like blend with $M = 10^5$ (top), and a coadded spectrum from the DESI early data release (bottom) [15]. Coaddition smooths noise at camera boundaries.

KL divergence is generally intractable for any given spectrum x, FAVI averages over samples x from the generative model to enable unbiased estimation of gradients with respect to the parameters ϕ of the encoder network q_ϕ. More precisely, we have

$$\nabla_\phi \mathbb{E}_{p(x)} \mathrm{KL}\Big[p(z \mid x) || q_\phi(z \mid x)\Big] = \nabla_\phi \mathbb{E}_{p(x)} \mathbb{E}_{p(z|x)} \log \frac{p(z|x)}{q_\phi(z \mid x)} = -\mathbb{E}_{p(z,x)} \nabla_\phi \log q_\phi(z \mid x), \tag{2}$$

provided integration and differentiation may be interchanged. Accordingly, unbiased estimates of this gradient can be computed with only samples $z, x \sim p(z, x)$. Importantly, the likelihood is not required to be tractable or given in closed form, and so amortized inference by FAVI is a form of *likelihood-free inference* or *simulation-based inference* [18, 19]. FAVI is also known as "sleep-phase" training in the reweighted wake-sleep (RWS) algorithm [17, 20], and is an instance of neural posterior estimation (NPE) [21] with fixed prior proposal.

3.1 Encoder Architecture

We utilize a novel short-term Fourier transform (STFT)-based architecture for the encoder network. STFTs apply a Fourier transform to small, overlapping windows of the input signal, revealing features relevant for frequency decompositions [22]. Intuitively, STFT-based features may be able to capture features related to the frequency, number, or size of emission and absorption lines in the input spectra, that are subsequently passed through a deep neural network for detecting blends.

Our variational distribution is parameterized as a Bernoulli on the blendedness indicator z conditional on the observed spectrum x, i.e.

$$q_\phi(z \mid x) \sim \text{Bernoulli}(f_\phi(x)),$$

where $f_\phi(x)$ denotes the neural network with parameters ϕ applied to the STFT features of x. In this case, the FAVI objective (Equation 2) reduces to binary cross-entropy loss, averaged over simulated (z, x) pairs:

$$-\mathbb{E}_{p(z,x)} \log q_\phi(z \mid x) = -\mathbb{E}_{p(z,x)} \big(\mathbf{1}_{z=1} \log p + \mathbf{1}_{z=0} \log 1 - p\big). \tag{3}$$

where $p = f_\phi(x)$ for a given spectrum x, and $\mathbf{1}$ denotes an indicator random variable.

Accuracy	AUROC
0.9476 (0.0084)	0.9591 (0.0082)

Table 1: Results for the constant SNR noise model, blending performed with $c = 0.1$.

M	Accuracy	AUROC
10^4	0.5851 (0.0390)	0.6182 (0.0315)
10^5	0.7846 (0.0332)	0.8322 (0.0218)
10^6	0.8726 (0.0161)	0.9131 (0.0115)

Table 2: Results for DESI-like spectra from `specsim`, blending performed with $c = 0.1$ for differing values of M.

4 Experiments

We trained the encoder network and evaluated on held-out samples from our generative model[2]. On simulated test batches of spectra, we compute accuracy by performing blend prediction by choosing the most likely class specified by $q_\phi(z \mid x)$, i.e. with hard assignment using prediction threshold 0.5. We also compute a second metric, the area under the receiver operating curve (AUROC), that aggregates model prediction performance across multiple thresholds. For both metrics we report the average over simulated test batches, with standard errors in parentheses.

The trained encoder inferred blendedness to a high degree of accuracy, as evidenced by Table 1 and Table 2. For the constant SNR noise model, around 94% of predictions on unseen test simulations are correct. Detecting blends in the simulated DESI data is more challenging, reflected by generally lower degrees of accuracy on test simulations. The different values of M result in different levels of brightness, and thus of signal-to-noise levels, in the simulated spectra with the DESI noise model. At lower brightness, discriminating between blends and non-blends is difficult due to the noise level. For brighter instances, however, accuracy is around 87% on unseen test simulations.

5 Discussion & Future Work

For synthetic data sampled from two generative models of astronomical spectra, we inferred a blendedness indicator using simulation-based inference. The specific SBI methodology we use, FAVI, implicitly marginalizes out the complex spectral latent random variables, so no variational distributions need to be explicitly constructed for these nuisance latent variables. This is critical as the space over these nuisance latent variables is both high-dimensional and trans-dimensional: either one or two latent spectra need to be marginalized over in the case of a non-blend or blend, respectively.

Encoders based on STFT features performed well at this task, particularly for our simpler model of spectra. Ultimately, though, we aim to apply our model to detect blends in DESI spectroscopic observations. To this end, in future work we aim to further improve the quality of inference on simulated DESI data by inferring minor blends (e.g., $c < 0.1$) and blends of three or more objects. These changes can be easily incorporated into our existing generative model. We also aim to refine our model of synthetic DESI spectra and inference procedure to the observed spectra from the first DESI early data release [15]. In particular, we aim to generalize STFT features to wavelet transformations to enrich the feature space of spectra for inference. Our present work provides a baseline for spectral blend detection that future work may build upon.

[2]Code available at `https://github.com/declanmcnamara/spectral_blends`

Acknowledgements

The authors thank David Kirkby for initial discussions that led to this research project. This material is based upon work supported by the National Science Foundation under grant numbers 2209720 (OAC) and 1841052 (DGE).

References

[1] Javier Sanchez, Ismael Mendoza, David P. Kirkby, Patricia R. Burchat, and The LSST Dark Energy Science Collaboration. Effects of overlapping sources on cosmic shear estimation: statistical sensitivity and pixel-noise bias. *Journal of Cosmology and Astroparticle Physics*, 2021(07):043, 2021.

[2] Robert H. Lupton. SDSS image processing I : The deblender. 2005.

[3] David M. Reiman and Brett E. Göhre. Deblending galaxy superpositions with branched generative adversarial networks. *Monthly Notices of the Royal Astronomical Society*, 485(2):2617–2627, 2019.

[4] Derek Hansen, Ismael Mendoza, Runjing Liu, Ziteng Pang, Zhe Zhao, Camille Avestruz, and Jeffrey Regier. Scalable Bayesian inference for detection and deblending in astronomical images. In *ICML Workshop on Machine Learning for Astrophysics*, 2022.

[5] E. F. Schlafly, G. M. Green, D. Lang, T. Daylan, D. P. Finkbeiner, A. Lee, A. M. Meisner, D. Schlegel, and F. Valdes. The DECam plane survey: optical photometry of two billion objects in the southern galactic plane. *The Astrophysical Journal Supplement Series*, 234(2):39, 2018.

[6] Erfan Nourbakhsh, J. Anthony Tyson, Samuel J. Schmidt, and The LSST Dark Energy Science Collaboration. Galaxy blending effects in deep imaging cosmic shear probes of cosmology. *Monthly Notices of the Royal Astronomical Society*, 514(4):5905–5926, 2022.

[7] Beth Reid, Shirley Ho, Nikhil Padmanabhan, Will J. Percival, Jeremy Tinker, Rita Tojeiro, Martin White, Daniel J. Eisenstein, Claudia Maraston, Ashley J. Ross, Ariel G. Sánchez, David Schlegel, Erin Sheldon, Michael A. Strauss, Daniel Thomas, David Wake, Florian Beutler, Dmitry Bizyaev, Adam S. Bolton, Joel R. Brownstein, Chia-Hsun Chuang, Kyle Dawson, Paul Harding, Francisco-Shu Kitaura, Alexie Leauthaud, Karen Masters, Cameron K. McBride, Surhud More, Matthew D. Olmstead, Daniel Oravetz, Sebastián E. Nuza, Kaike Pan, John Parejko, Janine Pforr, Francisco Prada, Sergio Rodríguez-Torres, Salvador Salazar-Albornoz, Lado Samushia, Donald P. Schneider, Claudia G. Scóccola, Audrey Simmons, and Mariana Vargas-Magana. SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues. *Monthly Notices of the Royal Astronomical Society*, 455(2):1553–1573, 2015.

[8] Hynes, R. I. An optimal extraction of spatially blended spectra. *Astronomy & Astrophysics*, 382(2):752–757, 2002.

[9] Peter Melchior, Rémy Joseph, Javier Sanchez, Niall MacCrann, and Daniel Gruen. The challenge of blending in large sky surveys. *Nature Reviews Physics*, 3(10):712–718, 2021.

[10] Celia R. Fierro-Santillán, Jaime Klapp, Leonardo Di G. Sigalotti, Janos Zsargó, and Markus Hareter. Analysis of spectral lines in large databases of synthetic spectra for massive stars. *The Astronomical Journal*, 161(3):121, 2021.

[11] ChangHoon Hahn, K. J. Kwon, Rita Tojeiro, Malgorzata Siudek, Rebecca E. A. Canning, Mar Mezcua, Jeremy L. Tinker, David Brooks, Peter Doel, Kevin Fanning, Enrique Gaztañaga, Robert Kehoe, Martin Landriau, Aaron Meisner, John Moustakas, Claire Poppett, Gregory Tarle, Benjamin Weiner, and Hu Zou. The DESI PRObabilistic Value-added Bright Galaxy Survey (PROVABGS) Mock Challenge. *The Astrophysical Journal*, 945(1):16, 2023.

[12] Andrew Miller, Albert Wu, Jeffrey Regier, Jon McAuliffe, Dustin Lang, Mr. Prabhat, David Schlegel, and Ryan P Adams. A Gaussian process model of quasar spectral energy distributions. In *Advances in Neural Information Processing Systems*, 2015.

[13] David Kirkby, Thomas Robitaille, Julien Guy, Erik Tollerud, Brigitta Sipőcz, Michael Droettboom, E. M. Bray, Stephen Bailey, Larry Bradley, Benjamin Alan Weaver, Matt Craig, Christoph Deil, P. L. Lim, Kyle Barbary, Adam Ginsburg, Wolfgang Kerzendorf, Steve Crawford, Víctor Zabalza, and Hans Moritz Günther. desihub/specsim: August 2020 release.

[14] Sandrine Perruchot, Pierre-Éric Blanc, et al. Testing the 10 spectrograph units for DESI: approach and results. In *Ground-based and Airborne Instrumentation for Astronomy VIII*. SPIE, 2020.

[15] DESI Collaboration et al. The early data release of the dark energy spectroscopic instrument, 2023.

[16] Luca Ambrogioni, Umut Güçlü, Julia Berezutskaya, Eva van den Borne, Yağmur Güçlütürk, Max Hinne, Eric Maris, and Marcel van Gerven. Forward amortized inference for likelihood-free variational marginalization. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, 2019.

[17] Tuan Anh Le, Adam R. Kosiorek, N. Siddharth, Yee Whye Teh, and Frank Wood. Revisiting reweighted wake-sleep for models with stochastic control flow. In *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence*, 2019.

[18] Jan-Matthis Lueckmann, Jan Boelts, David Greenberg, Pedro Goncalves, and Jakob Macke. Benchmarking simulation-based inference. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, 2021.

[19] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

[20] Jörg Bornschein and Yoshua Bengio. Reweighted wake-sleep. In *3rd International Conference on Learning Representations*, 2015.

[21] George Papamakarios and Iain Murray. Fast ϵ-free inference of simulation models with Bayesian conditional density estimation. In *Advances in Neural Information Processing Systems*, 2016.

[22] L. Gelman and S. Braun. The optimal usage of the Fourier transform for pattern recognition. *Mechanical Systems and Signal Processing*, 15(3):641–645, 2001.