MNRAS **000**, 1–8 (2021) Preprint 19 December 2020 Compiled using MNRAS LATEX style file v3.0

You don't need high-resolution spectra to model galaxy growth

L.E. Abramson[1]★, D.D. Kelson[1], and A. Dressler[1]

[1] *Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We compare actual $R \sim$ **800** spectroscopy to predictions generated from galaxy star formation histories (SFHs) inferred from $R \sim 25$ rest-optical prism spectra and $ugrizJK_s$ photometry. Based on **XXX** systems with **characteristics** we find a median difference of $\leq$1% between all predicted and measured Lick absorption features except the Balmer lines—explainable by unmodeled emission—and Ca4227 in passive galaxies, which is up to 2.5% weaker than expected. These results hold using SFH models incorporating either five age bins, or over a hundred. Therefore, absent a Ca–age prior accurate to ∼2%, provided that a sufficient wavelength baseline has already been sampled, there is little utility in adding high resolution spectroscopy as a flux-related constraint in SFH modeling. Our results cast doubt on the extent to which spectra from the *James Webb Space Telescope* will enhance our understanding of galaxy growth, such that progress requires new tactics more than new data.

Key words: galaxies: spectroscopy — galaxies: evolution

1 INTRODUCTION

Understanding stellar mass growth is a central ambition of the study of galaxy evolution. Spectral energy distributions (SEDs) are the key empirical anchors in this work because they can be decomposed into combinations of distinct stellar populations of known ages. The resulting coefficients encode the stellar mass a galaxy formed at the lookback time corresponding to each population's age—i.e., its star formation history (SFH).

Different stellar populations have different but not orthogonal SEDs. Thus, SED decompositions and SFH inferences are degenerate (Cid Fernandes et al. 2005). The degeneracies are compounded by formally age-independent effects like metallicity and dust.

In theory, using high resolution spectra should ameliorate this issue: the absorption lines visible in such data should increase the contrast between stellar subpopulations, constrain metallicities, and yield better age/mass coefficients. Here, we perform an experiment to see how true this statement is in practice.

First, we infer SFHs from low resolution ($R \sim 25$) optical prism data and broadband UV–IR photometry. Next, we create high resolution ($R \sim 800$) model spectra from these SFH inferences and predict the strengths of absorption lines not visible in the original data. Finally, we compare those predictions to real $R \sim 800$ spectra obtained for the same objects. The residuals in these comparisons reveal how much information the high resolution spectra would have added to the SFH inference.

In a sample of **XXX** galaxies at $\langle z \rangle$ = **ZZZ**, using two independent SFH inference techniques, our model spectra predict all but one non-Balmer Lick absorption index to better than 1% in the median (Worthey 1994). This finding holds regardless of whether a galaxy is currently forming stars, and for features outside the original prism bandpass. The exception—Ca4227 in passive galaxies—is known to behave anomalously, and is only ∼2% weaker than expected. Meanwhile, the Balmer mismatches are clearly due to unmodeled nebular emission, not incorrectly modeled stellar ages.

Given the accuracy with which a model constrained by coarsely-but-broadly sampled data reproduces features only visible in finely-but-narrowly sampled data, we conclude that the additional pixels in high resolution spectra do not add meaningful flux-related SFH constraints for individual galaxies. As such, future surveys aimed at supporting SFH reconstruction should aim to maximize wavelength coverage as opposed to spectral resolution.

Below, Section 2 describes the data, 3 compares the spectral predictions to the high resolution data, and 3.2 presents our key results. Section 4 discusses implications. We use AB magnitudes and assume a Chabrier (2003) stellar initial mass function (IMF) with $(H_0, \Omega_M, \Omega_\Lambda) = (70 \text{ km s}^{-1} \text{ Mpc}^{-1}, 0.3, 0.7)$ throughout.

2 DATA

2.1 Master sample

We use data from the *Carnegie Spitzer IMACS Survey* (CSI; Kelson et al. 2014). CSI provides Magellan-IMACS Low- and Uniform-Dispersion Prism spectroscopy (Coil et al. 2011) for objects with *Spitzer* [3.5] ≤ 21 in **15.3 sq deg** from the **XMM-LSS, ELIAS S1,**

★ E-mail: labramson@carnegiescience.edu

and CDF South fields. Combined with supplemental $ugrizJK_s$ photometry from the NEWFIRM archive (Autry et al. 2003) and Canada-France-Hawai'i Telescope Legacy Survey (CFHTLS; Cuillandre & Bertin 2006), these data were used to derive flexible SFHs for each galaxy during redshift estimation.

The sample is unbiased to $\log M_*/\mathrm{M}_\odot \sim 10.3$ at $z \sim 0.7$. The spectral resolution of the prisms varies from $R \sim 50$ to $R \sim 15$ at $\lambda = 0.5$ and 1 μm, respectively—about 60× coarser than the Sloan Digital Sky Survey (York et al. 2000).

We derive SFHs from these data in two ways. Our results are quantitatively similar using either approach.

2.1.1 *Technique 1 for inferring SFHs from low resolution SEDs*

The CSI spectrophotometry was first modeled using 5 precomputed SEDs based on SFHs with constant star formation rates from:

- 0.0 to 0.2 Gyr prior to $t_{\rm obs}$;
- 0.2 to 0.5 Gyr;
- 0.5 to 1.0 Gyr;
- 1.0 to 2.0 Gyr;
- 2.0 Gyr prior to $t_{\rm obs}$ to $z = 5$;

where $t_{\rm obs}$ corresponds to the object's redshift and $z = 5$ is taken as the onset of star formation. If the data preferred, the oldest bin could also take the form of a 1 Gyr top hat starting at $z = 5$. Dressler et al. (2016, 2018) examine these SFHs in detail with the latter containing a thorough assessment of their quality in its Appendix.

In the fits, the above SED bases shared a metallicity but took independent extinctions (Calzetti et al. 2000 law). The latter were estimated by replicating the stellar template at four different $A_V \in \{0, 0.5, 1.0, 2.0\}$ values and finding their best-fit non-negative superposition. This process enabled each stellar population to be screened by potentially complex dust geometries. Global metallicities were inferred from templates spanning $-0.6 \leq \log Z/Z_\odot \leq 0.3$ in 0.1 dex steps with a prior peaked at $Z_\odot$. As such, the predicted spectra did not capture enrichment histories (cf. Pacifici et al. 2012; Morishita et al. 2019). The predictive accuracy of these models stands despite this shortcoming (Section 3).

The spectral templates were generated using Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009) assuming default abundances. When inferring the SFHs, these models—5 mass amplitudes × 4 A_Vs + 1 metallicity + 1 redshift + 4 emission line/blend amplitudes = 26 free parameters—were typically constrained by 7 photometric + **~150** spectral datapoints. Redshifts were gridded in $\Delta z = 0.005$ increments and inferred jointly with the other parameters.

Hereafter, we refer to this approach as "Technique 1."

2.1.2 *Technique 2 for inferring SFHs from low resolution SEDs*

Subsequent to Dressler et al. (2018), all CSI spectrophotometry was refit using a new SFH inference scheme based on a library of 500 $H = 1$ stochastic tracks (Kelson 2014; Kelson et al. 2016, 2020; Abramson & Kelson 2020, Abramson in preparation). The tracks comprised 200 independent increments but were agnostic to the age of the universe: they were simply stretched to span arbitrary $t \in [t_0, t_{\rm obs}]$ intervals with the absolute duration of each increment changing with a galaxy's redshift. The fits assumed a t_0 of $z = 10$, such that th star formation increment at the sample's mean redshift was 33.5 Myr. Model spectrophotometry was generated at $\Delta z = 0.01$ intervals for redshift estimation.

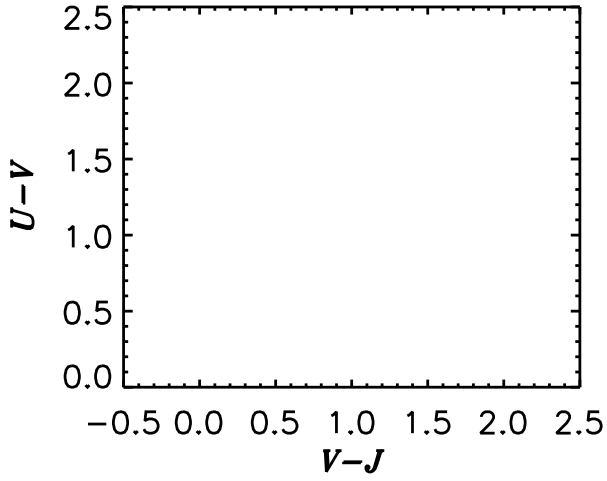

Figure 1. Some heuristic/contextual map of the sample data to orient the reader: UVJ plot, M_* histogram, half-mass time histogram, or all of the above. I don't want to clutter any of the subsequent plots with this stuff.

These models assumed monolithic values for both metallicity and extinction. Templates were generated by FSPS assuming default abundances at $-1.5 \leq \log Z/Z_\odot \leq 0.3$ in 0.3 dex steps **with a prior peaked at $Z_\odot$**.

Following Pacifici et al. 2012, the fitter found for each $H = 1$ track in the library the probability that it generated an observed SED given a redshift, metallicity, A_V, spectral fluxing function ($\propto \lambda^k$), and set of four emission line/blend amplitudes based on a chi-squared likelihood metric.

Hereafter, we refer to this approach as "Technique 2."

2.2 High resolution spectroscopy

We obtained high resolution—$R \sim 800$—Magellan-IMACS rest-optical spectra to study a class of galaxies inferred from the original CSI data to have formed at least half their stars within 1 to 2 Gyr of the epoch of observation—"late bloomers" (Dressler et al. 2016, 2018). The high-resolution observations covered late bloomers as well as objects outside of this class. We use these data here as the benchmarks against which to compare the predictions from the fits to the low-resolution CSI data discussed in the previous section.

We got the data in DATES. Here's how many objects, and a brief description of the data's S/N/quality/etc.

Figure 1 shows the *UVJ* **colour-colour** distribution of the high resolution sample objects as compared to the full CSI dataset.

3 RESULTS

We now compare the model spectra generated from our SFH inferences (Sections 2.1.1, 2.1.2) to actual high resolution spectra taken for the same objects. For all but one non-Balmer index, the median predicted line strength is within 1% of the measured value.

We study the model spectrum corresponding to the maximum-likelihood SFH fit. Uncertainties are tabulated for the SFHs and attendant parameters, but not propagated to the high resolution spectral models. As such, the following comparisons treat the predictions as more credible than they are. Combined with the models' monolithic, age-independent metallicities, our experiment is therefore biased towards revealing significant differences between the

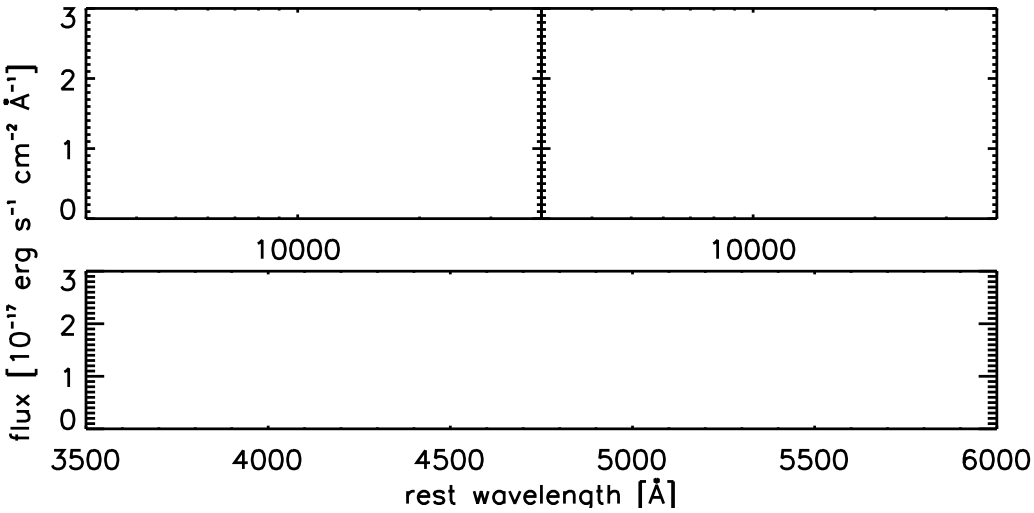

Figure 2. A 3-panel figure—2 top + 1 bottom. At top, show D18-like figures of the CSI data + SFH reconstructions for one object. I don't think we need to show the spectral sub-components for the chunky fit; just the best fit low-res SED and SFH. Other parameters at Dan's discretion, but they shouldn't clutter anything up; the key is the low-res SED. At bottom, show the full IMACS high-res spectrum for the same object and overlay the matched-resolution predictions from both SFH reconstruction techniques. Whatever colors we use here we'll stick with for the rest of the paper. The object probably should be like SA-type; nice absorption, some emission.

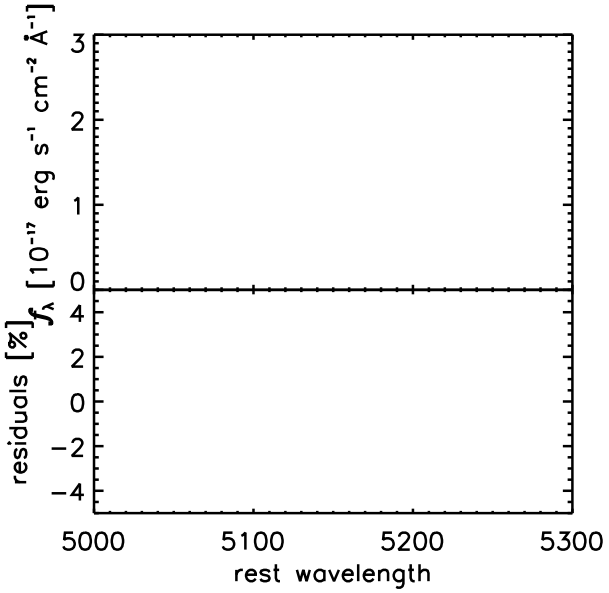

Figure 3. A graphic example of how we calculate whatever residuals we illustrate in Figure 4. Could be 2 panels, one showing the raw line + model, another showing the residuals. Could be 1 panel split in half. Could be for one object, or the full sample with the average models overlaid. Can just pick one SFH technique.

data and the predictions. Any offsets we find—in units of flux or standard deviations—are thus closer to upper limits.

LEA – some note here about whether we're going to focus on the full spectrum or just the "line strengths." The latter is justified because it's the real value-add of the high-res data, and also because I assume the same photometry was used to flux both the prism data and the high-res spectra. We just have to explain it.

3.1 Experimental setup

Figure 2 outlines the experiment. The **top** panels show the CSI spectrophotometry for a galaxy at redshift **ZZZ**, along with its inferred SFH using Technique 1 (left) and 2 (right). The CSI data have broad wavelength coverage and dense but coarse sampling over the rest optical. As such, they reflect a relatively complete accounting of a galaxy's stellar photospheric continuum emission. Except for the strongest/broadest absorption features—e.g., Hβ; the Mg λ5170 triplet—these data are insensitive to absorption lines and any historical information they might convey.

Figure 2, **bottom**, shows the complete high resolution IMACS spectrum of this object with the matched-resolution spectral predictions from the SFHs in the top panels overlaid. Results from Techniques 1 and 2 are in **color 1** and **color 2**, respectively.

The data in the two top panels are identical; only the models change from left to right. Likewise, both model predictions are compared to the same $R \sim 800$ IMACS data. We therefore perform two experiments in parallel identical except for the method used to generate the high resolution prediction. **We comment on the differences between the results from each technique in Sections XYZ, but they are ancillary to our main point.**

Figure 3 illustrates how we compare our spectral predictions to the high resolution data. Here, we re-present Figure 2, bottom, zoomed-in to the region surrounding **FEATURE**. LEA – If we need to divide the data by the model, we'll make this 2 panels. The central bandpass of the corresponding Lick index is shown in **however**. We define the mismatch between prediction and high resolution data as:

$$\Delta = \text{whatever Dan says.} \quad (1)$$

That is, a quoted difference of 5% in a given spectral feature corresponds to **whatever per-angstrom offset integrated over however many angstroms, or whatever Dan did.**

We then compute Δ for each Lick bandpass in each galaxy.

LEA – Here is where we talk about limiting ourselves to a

Table 1. Summary of experimental results. Mean equivalent widths are quoted for the full sample. **Asterisks denote feature outside the CSI prism wavelength coverage.**

Feature	Band center [Å]	Bandwidth [Å]	⟨EW⟩ [Å]	Full sample offset [%]		**Red** sample offset [%]		**Blue** sample offset [%]	
				Tech. 1	Tech. 2	Tech. 1	Tech. 2	Tech. 1	Tech. 2
CaHK	**LAMBDA**	**LAMBDA**							
CN_1	4159	35	foo	foo	foo	foo	foo	foo	foo
Ca4227	4228	12							
G4300	4298	35							
Fe4531	4536	45							
C_2 4668	4677	86							
Mg*b*	5176	32							
Fe5270	5265	40							
Fe5335	5332	40							
$H\delta_A$	4102	38							
$H\gamma_A$	4341	43							
Hβ	4862	28							

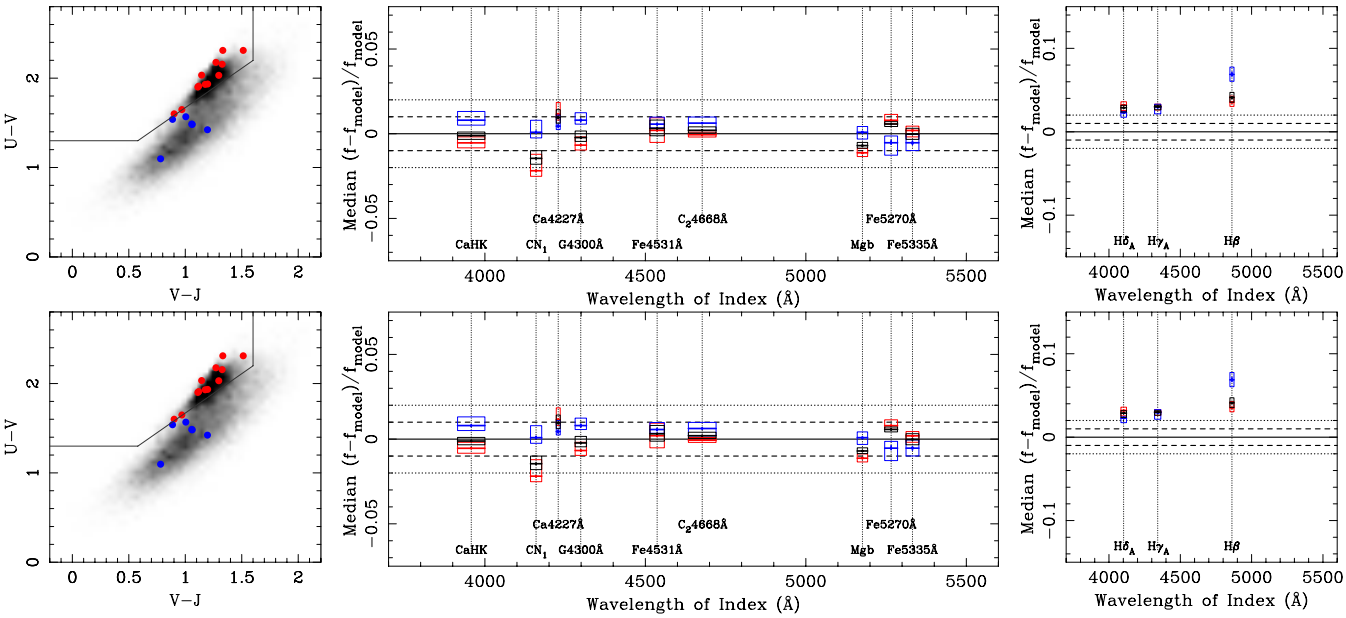

Figure 4. Main plot, only the absorption features (no Balmers), Technique 1 at top and 2 below. I think we should look at versions highlighting any ranges outside the prism bandpass (at the median or *N*th percentile redshift or whatever) and also see what happens when the individual residuals are lightly traced in grey or something. We also need to determine how we'll split the data (UVJ or (N)LB); if we go w/ UVJ perhaps we can add the LB/NLB version in an Appendix. I think we should also shrink the vertical scale to ±5%.

comparison in the lines or mention that we calculate global χ^2 values and point to where we'll discuss those.

3.2 Sample-wide outcomes

Figure 4 and Table 1 summarize the quality of our model predictions across the entire **XXX** object sample; i.e., the overall outcome of our experiment. In the figure, results from Techniques 1 and 2 are shown in the top and bottom panels, respectively. At the central wavelength of each feature, we show the median, 16th, and 84th percentile offsets in a given sample. Results based on the full sample are shown in black, with red and blue boxes denoting **SFR-defined** subsamples, respectively. The width of each box corresponds to the width of that feature's Lick bandpass. Regimes outside the original prism wavelength range are highlighted **in some color**. The dashed and dotted horizontal lines respectively benchmark mismatches at the ±1% and ±2% level.

Clearly, *almost all rest-optical galaxy absorption features visible at R ∼ 800 are predictable to better than 1% by models tuned to data with ∼30× lower resolution* but broader wavelength coverage. That is, to a very high degree of accuracy, one can infer the strengths of nearly all major optical stellar absorption features in integrated galaxy light without ever detecting any of them.

This statement holds for any feature in any subsample except one. **State mean total offsets for both types (Table 1).** It holds irrespective of the feature's width—**basic quantification**—and irrespective of the chemical species traced. It holds for strong features—e.g., the G band in passive objects **quote mean EW**—and weak ones—e.g., Mg*b* in starforming ones **quote mean EW**.

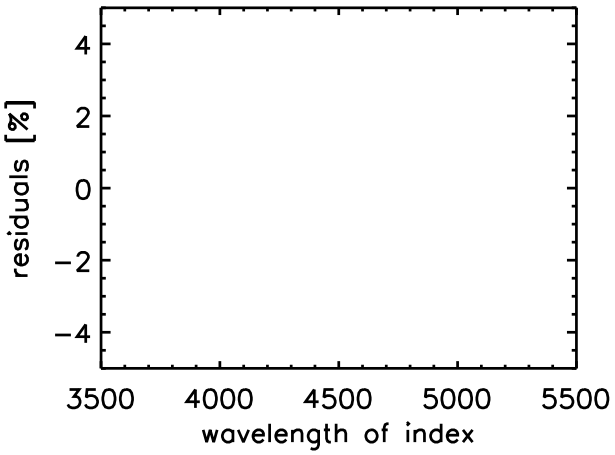

Figure 5. Show the offsets for the Balmer line predictions; same styling as Figure 4. Can potentially add another panel for, e.g., Hβ specifically showing the unmodelled emission. Can also put this up in Figure 3, but I think best to bury it down here.

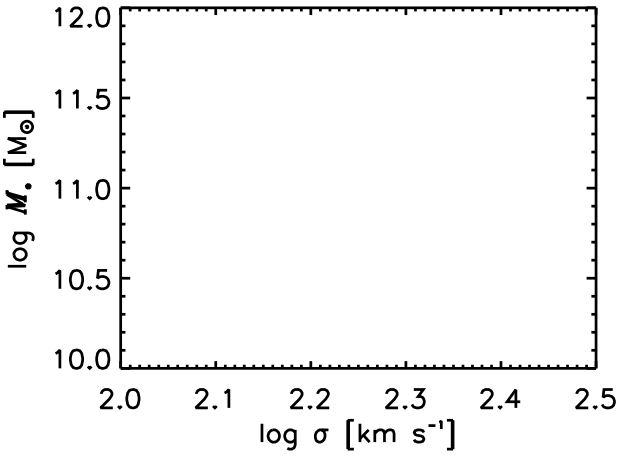

Figure 6. Show σ distributions or σ–M_* for LB/NLB objects.

Simply: we knew the fluxes of each Ångstrom of our high resolution spectra before we got to the telescope.

The lone exception is Ca4227 in **passive**, but not starforming galaxies. **Something about the history of this line being anomalous, Ca-rich supernovae, etc.** As CSI was designed to probe spatial overdensities, **and some of these objects are for real in dense places?**, we believe the ∼2.5% mismatch we see here may reflect the same phenomenon. Regardless, this finding suggests either that slightly different abundance patterns should be assumed when fitting passive objects (**CITE**), or that Ca4227 strengths might be used to enhance SFH estimates. We touch on the latter in Section 4, but note here that any prior would have to be accurate to better than 2% in the mean for this relationship to be useful.

3.3 The Balmer lines

Figure 5 shows the same offset summaries as Figure 4 but for the three strong Balmer lines in the prism data: Hδ, Hγ, and Hβ. LEA – we'll have to see what these look like. I don't think there's anything interesting to be said here, just to show that the only sizeable discrepancies are readily attributable to emission. If it holds up, we can also make a note about the enhanced performance of the $H = 1$ models here versus the chunky fits.

3.4 Systematic errors

The above comparisons are affected by **three** main systematics: redshift refinement, velocity broadening, and continuum matching.

Regarding redshift refinement, both SFH inference techniques adopted discrete redshift steps of $0.005 \leq \Delta z \leq 0.01$ (Section 2). This interval was sufficient given the low resolution of the CSI data, but the newer IMACS data allow for more precision. As such, there are typically mismatches of $\delta z \approx ZZZ$ between the predicted locations of spectral features and where they ultimately appear in the high resolution data.

Regarding velocity broadening, the CSI data are likewise too coarse to permit velocity dispersion estimates. As such, the high resolution spectral predictions must be broadened from the native FSPS resolution to whatever the $R \sim 800$ IMACS data imply for each object. Figure 6 shows the velocity dispersions inferred for our sample compared to their stellar masses.

We correct for both issues by **Dan to discuss how he shifts and broadens the models here**.

These offsets are important: they illustrate the only first-order orthogonal axes in the problem of SFH reconstruction: velocity and age. They also show us that higher resolution spectra add meaningful signal in velocity—if not age—space. It is not clear that independent Doppler information can be exploited to enhance SFH estimates (Section 4), but it is where high resolution data obviously help.

The last systematic is mismatches in the stellar continuum. **Dan to send methods.**

4 DISCUSSION

4.1 Implications for direct SFH inference

Our experiments imply that high resolution optical spectroscopy of any quality reasonable to expect for individual distant objects will not meaningfully enhance knowledge of their underlying stellar populations compared to much more coarsely and broadly sampled datasets (Section 4.1.4). That is, absorption line details—presented at galactic velocity dispersions—are sufficiently correlated to stellar continua that characterising the latter across a long enough wavelength baseline yields the same information as densely sampling the rest optical. There is therefore no need to obtain costly, high resolution spectroscopy for SFH inferences.[1] **In this, we reinforce findings by Pacifici et al. (2012), who show that—limited to rest optical data—$R \sim 1000$ spectra enhance SFH estimates less than other inferences compared $R \sim 100$ spectra (25% v. 45% drops in uncertainty; see their Table 2).** We would add, however, that, had that study incorporated UV–IR photometry, those gains would have been reduced. LEA – (Leja et al. 2019; Lower et al. 2020) show that once you have the kind of coverage

[1] A future paper will show that, at $z \sim 0.4$, the 26-band UltraVista filter set (Muzzin et al. 2013) performs similarly to CSI, meaning that even the prism data used here is likely oversampled.

we have changing the SFH parameters modulate the high resolution spectra at the 1%–2% level. Further insights into galaxy growth will not be gleaned by adding better spectra, but rather from the methods and results that reveal why this statement is true.

This lesson is not new. The fact that metallicity modulates continuum light on ∼100–1000 Å scales is apparent in the output of all stellar population synthesis codes. The phenomenon is also well recognized in, e.g., the correlation of Balmer line strengths with $U - B$ or D4000 (e.g., Kauffmann et al. 2003; Cid Fernandes et al. 2005). We have shown only that the same phenomenon holds for all (Lick) species and to a very great extent. The observational trick is to sample the SED so as to capture all of the relevant signal at minimal cost. We do not know the optimal sampling, only that CSI apparently meets the criteria that define it.

Nevertheless, these findings reinforce or suggest practical, informatic, and theoretical programs that we believe will be more informative than simply taking more or "better" data.

4.1.1 *Practical*

The practical upshot of the above is twofold. First, there is no need to expend telescope or computer time obtaining and modeling thousands of SED points for the purpose of inferring galaxy growth. Second, on the modeling side, the physical correlations between spectral elements should be accounted for mathematically to avoid drawing inferences of unreasonable certainty. Each pixel may be formally independent, but it does not add a new degree of freedom. LEA – this is the place to discuss previous work on or findings to this effect. The fact that S/N doesn't really matter has been noted in a bunch of places before, e.g., Leja et al. (2019). We disagree with those authors about the utility of adding bedder spectra—even though they see similar offsets to us (2%; their Figure 8)—but may agree with Ocvirk et al. (2006) about it?

The narrative here is something like: limited to the rest optical, Pacifici showed there's only marginal gains from going to R=1000 spectra for SFH inferences vs R=100. Leja showed that, using UV–IR photometry, S/N doesn't matter for SFH inferences. We do something like combine the two: if you have sufficiently broad SED sampling, the SFHs you infer are good enough to remove any reason for S/N or resolution to matter because all of the information those things could add has been accounted for.

Also, we're working at about 1/2 the age of the universe at least compared to Camilla, if not Leja, too.

4.1.2 *Informatic*

The informatic upshot of the above is that there exists a data compression scheme—a way to distill covariant pixels to independent degrees of freedom—that may be worth identifying if SFH inferences for large samples is to play a substantial role in extragalactic astronomy's future. Given the deluge of SEDs that PFS, LSST, and the Roman Space Telescope will deliver, modeling efficiency will be critical to characterising representative swaths of forthcoming datasets. It is likely that such a compression scheme will resemble what intuition and, e.g., CSI suggests: sparse UV and IR sampling with finer—but not fine—sampling over the rest optical, particularly from around the Balmer break to around the magnesium triplet, where intermediate age stars make the greatest impression (e.g., Dressler et al. 2016). However, it is also possible it will not resemble that configuration, so we encourage serious work on this issue (**or cite and amplify; Mosby hasn't published his work though**).

4.1.3 *Theoretical*

The theoretical upshot is that understanding this compression scheme—even if it is as basic as the above—represents understanding the extent to which the physics an object was subject to over its history (what we want) could possibly modulate its presentation at the one epoch we can observe it (what we get). It must therefore reveal the bedrock of the SFH ↦ SED mapping, and so something close to the epistemological ceiling for the study of galaxy evolution. Knowledge of such boundary conditions cannot but enhance the questions investigators ask and the investigations they pursue.

How mechanically to reverse engineer this compression scheme is a separate question. Machine learning could be used to identify via simulation the hypersurface characterising the precision to which $M_*(t)$ is knowable as a function of data—S/N, SED sampling—and galaxy properties—z, sSFR, Z, σ, A_V, environment, morphology. Alternately, one could simply invert the experiment we performed, fitting SFHs only to high resolution optical spectra and predicting broadband fluxes to the UV and IR. One could then repeat this exercise while downgrading the resolution or omitting portions of the starting spectrum and measuring how the extrapolations changed, adding outrigger photometry as soon as the predictions began to deviate significantly from the truth. Figures 12 and D2 of Abramson et al. (2020) show examples of this kind, and hint at the predictive quality as a function of SED type. Of course, this effort could also be coupled to machine learning.

Based on our results, however, similar exercises performed using only photometry (Abramson in preparation), and the physics at play, we suspect that full UV–IR wavelength coverage will be one of the foundations whatever map is finally produced: If all of the starlight is captured, the density of the sampling is secondary (provided it is sufficient to obtain accurate redshifts). Future surveys designed to support SFH reconstruction should therefore opt for increased wavelength coverage at the expense of SED resolution.

We are not asserting that there is never a reason to take high resolution spectra. If one is interested in the IMF (Conroy & van Dokkum 2012), specific abundance patterns, or kinematics, it is necessary. We only argue that it seems unlikely that those investigations will qualitatively enhance assessments of stellar mass growth.

The main counterpoint we find is the calcium line strengths of passive galaxies. These seem to add a relatively large amount of information to our predictions compared to the other Lick indices. However, while the prediction/data mismatches are twice the average offset, they are still only at the 2% level. As such, to enhance SFH predictions, priors linking Ca to age must be accurate to at least that level, and spectra must have $S/N \gtrsim 14\,\text{Å}^{-1}$ to see the signal at $1\,\sigma$. We believe attempting to construct such a prior or take such data are endeavours of diminishing returns: aperture and other observational effects surely distort one's ability to interpret such precision when present, and—as in the case of calcium—other, more easily observed properties like environment are probably equally good candidates for an age prior (**CITE**).

4.1.4 *What about stacking?*

If ∼1%–2% line strength precision is unobtainable for useful numbers of individual objects, it can be achieved through stacking (**CITE**). In principle, modeling the stacks could reveal aspects of the aggregate growth trajectories of the constituent objects. This statement is mathematically true, but may be physically unhelpful.

Figure 7 explores the $H = 1$ mass growth prior used in our Technique 2 SFH inferences (Section 2.1.2). These tracks are anal-

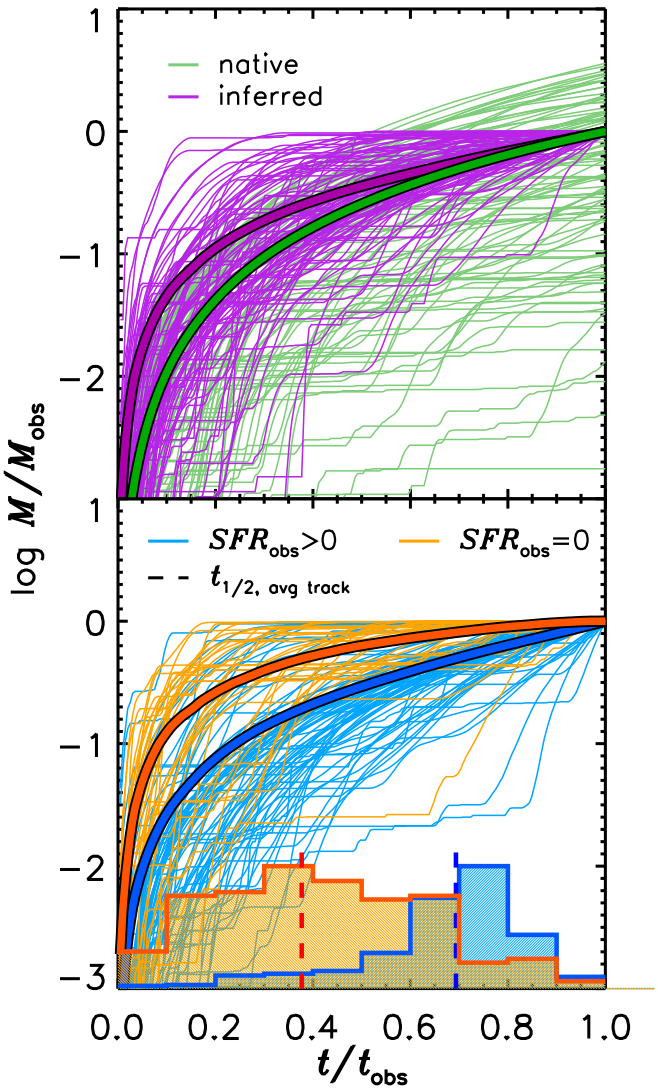

Figure 7. LEA – here's why stacking will not buy you much. Histograms are normalized to unity at peak.

ogous to the suite of semi-analytic model-derived histories used by Pacifici et al. (2012).

The top panel shows the native distribution of trajectories modeled forward from $t = 0$ to t_{obs} in green, with the ensemble's mean mass normalized to $\langle M \rangle = 1$ at t_{obs}. Overlaid in purple is the same suite of tracks, but with each independently normalized to $M(t_{obs}) = 1$. The purple ensemble is the result obtained by fitting: in the case of one object or a stack, the luminosity and color of the SED confines all tracks in the library to a small range of normalizations.

Although they are composed of exactly the same tracks, the mean of the inferred suite of histories is distinct from the mean of the native suite of histories. Stepping forward in time, as the galaxies do, results in a different outcome as stepping backwards—as the fitter must (see also Behroozi et al. 2013; Torrey et al. 2017). In this context, the physical question of interest is: "Does $H = 1$ stochasticity approximate the correct SFH reference set for real galaxies?" The answer to that question would be obtained by assessing whether the *difference* between the inferred history and the theoretical mean is as shown here (using a non $H = 1$ basis in the inference). Since this difference is large, high precision inferences are not needed; many lower-quality measurements will do fine.

The bottom panel shows a more practical example. We reproduce the purple tracks from the top panel split into those that lead to passive (orange) or starforming (blue) galaxies at t_{obs}. The half-mass time distributions for each subsample are shown at the bottom. From those distributions, it is clear that objects of the same stellar mass *and* SFR—intuitive parameters to stack on—may have half-mass times differing by 40% of a Hubble time. What the average of these tracks known to any precision encodes about the physics causing any one of them to behave as it did is hard to see.

The issue is that there are only two objects of physical interest: good assessments of (1) individual SFHs from which physics might be directly inferred, or (2) the typical SFH of a meaningful class of galaxies. Due to S/N requirements, (1) is not obtainable beyond what CSI-quality datasets support (Section 3). Meanwhile, (2) is an average over *histories*, and so is assessable by combining many low-fidelity inferences.[2] Both statements imply that the pursuit of data at the S/N needed to yield 1% line strengths—directly or via stacking—is unnecessary.

LEA – basically, the physics is in the histories (if anywhere), so you want to stack the histories (if anything), and for that you don't need good data at the single object level. Maybe there's room to compare the stack of histories to the history derived from a stack? What would that tell you? Something about loss-of-signal lookback horizons?

4.2 Prospects for exploiting Doppler information

High resolution spectroscopy does yield unique insight into velocity space. The obvious enhancement is to redshifts. This may be important to mitigate against, e.g., [O II] emission masquerading as an enhanced Balmer break at low resolution, biasing galaxy ages toward those of A stars. However, kinematics could also potentially aid in SFH reconstruction.

It's tempting to contemplate using galaxy velocity dispersions to place colour-independent total stellar mass priors on fits to SEDs of any resolution. This move is especially appealing because it could provide a handle on low-mass stars that imprint on the velocity dispersion but are undetectable in an SED. Unfortunately, a number of complications suggest that incorporating such data would not improve current methods.

The prior would most readily enter through a σ–M_* relation. The intrinsic scatter of this relation is ~0.2 dex (Figure 6 and **CITE**), much broader than the formal—or even systematic—errors on SED-inferred stellar masses (**quote what we find; s ystematics as cross-method or duplicate obs differences**). Further, that relation would be calibrated to stellar masses suffering the same systematic errors one wishes to correct by using the relation. Theoretically, a σ–Sérsic index–color relation—or some equivalent structure—could be used, but this entails other issues regarding spatial resolution, aperture matching, and dust effects. Finally, due, e.g., to inside-out growth (**CITE**) and the existence of thick disks, each stellar subpopulation may contribute its own velocity dispersion to the global profile. If so, modeling dispersions will reintroduce exactly that the same degeneracies in line decomposition

[2] It is not clear that (2) is obtainable, either. "Typical" in this context properly refers to the *mode* history; that which describes the largest fraction of class members. It may be that the *mean* history is our only proxy for this, but it is not in reality the object of interest.

as arose with SEDs. These effects would have to be marginalized over—almost certainly negating the precision gains—or the systematics from failing to do so understood.

UTILITY AFTER SELECTION BY SFH— Figure 6 illustrates the sample's velocity dispersion estimates. Here, we show **what we show** for galaxies split by their inferred growth histories. **LBs show behavior X while NLBs show behavior Y.** needed to say for certain, these trends suggest a link

We are bearish on the ultimate utility of kinematic stellar mass constraints, but we also do not know of any SFH modeling that has attempted to incorporate it. A study assessing its practicality and real-word effects would be edifying.

NIRSPEC no good for SFHs but ok for kinematics + environments. Cosmology.

5 SUMMARY

In the context of SFH reconstruction from full spectral fitting: high resolution doesn't buy you anything new.

- You can predict line strengths to 1% without measuring them.
- Wavelength coverage beats spectral resolution.
- Metal enrichment histories for at least our sample are consistent with flat to whatever corresponds to 1% differences in line strengths.
- The goal is not to fit SFHs to SEDs, but to understand the sampling/compression scheme that enabled us to find what we found. This is the SFH↦SED mapping and is equivalent to understanding the epistemological ceiling in the study of galaxy evolution.

Facilities: Magellan/IMACS
Software: Python (`CarPy`).

ACKNOWLEDGEMENTS

LEA thanks Dr. Lindsay Young for allowing him to find a place to write this text.

REFERENCES

Abramson L. E., Kelson D. D., 2020, arXiv e-prints, p. arXiv:2012.08050
Abramson L. E., Brammer G. B., Schmidt K. B., Treu T., Morishita T., Wang X., Vulcani B., Henry A., 2020, MNRAS,
Autry R. G., et al., 2003, in Iye M., Moorwood A. F. M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. pp 525–539, doi:10.1117/12.460419
Behroozi P. S., Marchesini D., Wechsler R. H., Muzzin A., Papovich C., Stefanon M., 2013, ApJ, 777, L10
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Chabrier G., 2003, PASP, 115, 763
Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, MNRAS, 358, 363
Coil A. L., et al., 2011, ApJ, 741, 8
Conroy C., van Dokkum P., 2012, ApJ, 747, 69
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Cuillandre J.-C., Bertin E., 2006, in Barret D., Casoli F., Lagache G., Lecavelier A., Pagani L., eds, SF2A-2006: Semaine de l'Astrophysique Francaise. p. 265
Dressler A., et al., 2016, ApJ, 833, 251
Dressler A., Kelson D. D., Abramson L. E., 2018, ApJ, 869, 152
Kauffmann G., et al., 2003, MNRAS, 341, 33
Kelson D. D., 2014, preprint, (arXiv:1406.5191)
Kelson D. D., et al., 2014, ApJ, 783, 110
Kelson D. D., Benson A. J., Abramson L. E., 2016, preprint, (arXiv:1610.06566)
Kelson D. D., et al., 2020, MNRAS, 494, 2628
Leja J., Carnall A. C., Johnson B. D., Conroy C., Speagle J. S., 2019, ApJ, 876, 3
Lower S., Narayanan D., Leja J., Johnson B. D., Conroy C., Davé R., 2020, arXiv e-prints, p. arXiv:2006.03599
Morishita T., et al., 2019, ApJ, 877, 141
Muzzin A., et al., 2013, ApJ, 777, 18
Ocvirk P., Pichon C., Lançon A., Thiébaut E., 2006, MNRAS, 365, 46
Pacifici C., Charlot S., Blaizot J., Brinchmann J., 2012, MNRAS, 421, 2002
Torrey P., Wellons S., Ma C.-P., Hopkins P. F., Vogelsberger M., 2017, MNRAS, 467, 4872
Worthey G., 1994, ApJS, 95, 107
York D. G., et al., 2000, AJ, 120, 1579

This paper has been typeset from a TEX/LATEX file prepared by the author.