SFH MEASURES FROM BAYESIAN FULL SPECTRUM FITTING

BENJAMIN D. JOHNSON, ET AL.

Draft version January 24, 2014

ABSTRACT

Testing

1. INTRODUCTION

?. We've been at this for over forty years. Spectral models improving (in resolution, accuracy, and characterization of uncertainties), spectral observations rapidly improving.

SFH constraints (in rough order of preference depending on S/N or depth: resolved stellar CMDs, integrated spectroscopy, line indices, integrated broadband SED)

Existing SFH methods for spectra (**?**). SSP equivalent (indices), various least-squares minimization, maximum likelihood, and matrix inversion schemes, ICA/PCA, NMF. Various amounts of regularisation required. `STARLIGHT, STECKMAP, ULYSS, pPXF, MOPED/VESPA, K_CORRECT`. ICA can be considered an implementation of Bayesian BSS with several assumptions (strong priors).

Why is this better? Open source, for one (STARLIGHT is closed source, VESPA/MOPED is not even publically available as a binary), not in IDL (ULySS, STECKMAP, pPXF), much more flexible, and fully Bayesian (i.e. not maximum likelihood, allows prior information in a principled way, allows marginalization)

2. METHODOLOGY

Fully Bayesian SFH reconstruction using MCMC. Advantages are no linearity constraint, full marginalized posterior PDFs for model parameters, incorporation of prior information (this is key!), an extensible generative model, full samples of the likelihood for use in heirarchical models of galaxy evolution (i.e. to infer hyperparameters of the SFH distribution from largish galaxy samples)

3. THE MODEL

The model consists of a number of SSPs that can be linearly combined, including metallicity variations and velocity dispersion. In principle, emission lines can be added, as well as dust attenuation of stars of different ages, and uncertainties in SPS models can be propogated. Complexity of the model is only limited by the expense of the likelihood call and the ability of MCMC routines to efficiently explore the parameter space.

The general model is effectively specified by the classical population synthesis equation (though we expand on the typical dust specification):

$$\begin{aligned} L_\lambda = & K_\lambda(\vec{b}) + I_\lambda(\vec{a}) \times \\ & \Big\{ \big[\sum_i \mathrm{A}_i S_\lambda(t_i, Z_i, \xi) e^{-\tau_{V,i} k_\lambda(\tau_V, DF, R_v, \Theta)} \big] * G(\sigma_{v,stars,i}) + \\ & \mathrm{Q} \cdot S_{\lambda,neb}(\bar{U}, Z_{gas}) e^{-\tau_{\lambda,neb}} * G(\sigma_{v,gas}) + \\ & \mathrm{C} \cdot S_{\lambda,AGN} + \mathrm{D} \cdot S_{\lambda,dust}(U_{dust}, R_v) \Big\} \end{aligned} \quad (1)$$

where t_i is the age, Z_i is the metallicity, ξ are parameters related to uncertain ingredients of the SSP models (e.g. f_{BHB}, binary fraction, IMF, etc..), $\tau_{V,i}$ is the characteristic V band optical depth toward stars of that component. k_λ relates the effective attenuation at λ to the characteristic attenuation at V and depends on: τ_V itself (e.g. **?**); DF, the distribution of attenuation values (e.g., a delta-function at τ_V, log-normal, or uniform up to some value); R_V, the shape of the extinction curve, which may be dependent on dust composition; and Θ, a parameter that encapsulates the effects of relative star/dust and global geometry and scattering on the shape of the effective attenuation curve. Note that k_V is not necessarily 1 if the DF is complex. $G(\sigma_v)$ is the gaussian broadening function, where σ_v is the velocity dispersion. I_λ is a multiplicative calibration factor depending on parameters a and $K_\lambda(b)$ is an additive calibration factor (perhaps sky emission) depending on parameters b .The amplitudes A_i describe the SFH (i.e. the total stellar masses of each component).

We will first be concerned with a more restricted model, in which we ignore dust emission, nebular emission, AGN, the various SSP parameters ξ and many of the more complicated terms in the attenuation curve. We also assume a single velocity dispersion for all stars. We write this model as

$$L_\lambda = K_\lambda(\vec{b}) + I_\lambda(\vec{a}) \sum_i \mathrm{A}_i S_\lambda(t_i, Z_i, \xi) e^{-k_\lambda \tau_{V,i}} * G(\sigma_{v,stars}) \quad (2)$$

The parameter list for this model is

$$\theta = \{\mathrm{A}_i \ldots \mathrm{A}_N, \tau_{V,i} \ldots . \tau_{V,N}, \sigma_{v,stars}, \vec{a}, \vec{b}\} \quad (3)$$

where there are $N = N_{age} \times N_Z$ separate amplitudes. In practice, the number of dust attenuations should be smaller, perhaps only 2 or 3, and certainly less than N_{age} but we keep the full set for generality.

3.1. *Likelihood and Likelihood Gradient*

We write the likelihood of the model as

$$\mathcal{L} = \prod_\lambda \frac{1}{\sigma_\lambda \sqrt{2\pi}} e^{-\frac{(L_\lambda - F_\lambda)^2}{2\sigma_\lambda^2}} \quad (4)$$

where F_λ is the observed spectum and σ_λ is the true noise at each wavelength. If we consider an additional term from uncertainty on the noise such that $\sigma_\lambda = \tilde{\sigma}_\lambda + JF_\lambda$ (jitter) then the natural logarithm of the likelihood, after expanding L_λ of the restricted model, is given by

$$\ln\mathcal{L} = -\frac{1}{2} \sum_\lambda \frac{(I_\lambda(\vec{a}) \sum_i A_i S_\lambda(t_i, Z_i) e^{-k_\lambda \tau_i} + K(\vec{b}) - F_\lambda)^2}{(\tilde{\sigma}_\lambda + JF_\lambda)^2} + \sum_\lambda \ln[2\pi(\tilde{\sigma}_\lambda + JF_\lambda)^2] \tag{5}$$

where the second term now depends on the parameter J instead of being constant.

Now, in highly covariant spaces it can be useful to have a measure of the gradient of $\ln\mathcal{L}$. In the amplitude directions this is

$$\frac{\partial \ln\mathcal{L}}{\partial A_i} = -\frac{1}{2}\sum_\lambda 2\frac{(L_\lambda - F_\lambda)}{\sigma_\lambda^2} S_\lambda(t_i, Z_i) e^{-k_\lambda \tau_i} I_\lambda(\vec{a}) \tag{6}$$

while in the τ_V directions this is

$$\frac{\partial \ln\mathcal{L}}{\partial \tau_{V,i}} = \frac{1}{2}\sum_\lambda 2\frac{(L_\lambda - F_\lambda)}{\sigma_\lambda^2} I_\lambda(\vec{a}) A_i S_\lambda(t_i, Z_i) e^{-k_\lambda \tau_i} k_\lambda \tag{7}$$

We note that, for sets of SSPs or components i where the dust attenuation is required to be the same (i.e. for all components of a similar age), then the partial derivative of each of these components should be summed over those components.

In the $\sigma_{V,stars}$ direction we have

$$\frac{\partial \ln\mathcal{L}}{\partial \sigma_{V,stars}} = \tag{8}$$

In the $\vec{a}$ and $\vec{b}$ directions we have

$$\begin{aligned} \frac{\partial \ln\mathcal{L}}{\partial a_i} &= -\frac{1}{2}\sum_\lambda 2\frac{(L_\lambda - F_\lambda)}{\sigma_\lambda^2} \frac{L_\lambda - K_\lambda(\vec{b})}{I_\lambda(\vec{a})} \frac{\partial I_\lambda(\vec{a})}{\partial a_i} \\ \frac{\partial \ln\mathcal{L}}{\partial b_i} &= -\frac{1}{2}\sum_\lambda 2\frac{(L_\lambda - F_\lambda)}{\sigma_\lambda^2} \frac{\partial K_\lambda(\vec{b})}{\partial b_i} \end{aligned} \tag{9}$$

Finally, in the J direction we have

$$\frac{\partial \ln\mathcal{L}}{\partial J} = -\frac{1}{2}\sum_\lambda -2\frac{(L_\lambda - F_\lambda)^2}{\sigma_\lambda^3}(F_\lambda) + \frac{1}{2\pi\sigma_\lambda} F_\lambda \tag{10}$$

It is difficult to write down expressions for the partial derivative with respect to the parameters ξ. In practice these will likely be determined via finite differnceing, or the MCMC can be run on each point of a grid of these parameters. Alternatively, we note that the SSP essentially specifies the distribution of stellar parameters for a given age and metallicity $p(L,T,g|t,Z)$ (according to the IMF and isochrones) and the ξ result in perturbations to this distribution. One could therefore imagine attempting to infer small deviations around a strong prior for this function through a model of linear combinations of individual stellar spectra, though this dos not then include uncertainties in the spectra themselves for a given L,T,g,Z. Anyway, such work is beyond the scope!

3.2. *Number of components*

In principle, can be determined from the data (e.g. find the binning which minimizes covariance), or left to be very large. In practice, this takes a long time to reach autocorrelation. Biases due to considering wide bins? can mock spectra using high temporal resolution and solve with low temporal resolution, see if you get biases. This does indeed result in biases, need to quantify/explore a bit more.

While at infinite S/N there is no covariance between different components and they can be recovered exactly as long as they are linearly idependent, there is significant covariance in the different components at moderate signal to noise. If we consider three time bins whose spectra are nearly indistinguishable, then the likelihood surface for the amplitude of these components, assuming the other coomponents to be fixed, will describe the surface of an ellipsoid in the positive octant (or in a 2-d slice resembling something like a banana distribution). Generalizing to higher order covariances the likelihood function may be expected to approximately describe the surface of a hyper-ellipsoid in the positive closed orthant, but could get significantly more complex.

Possibilities for dealing with these complicated likelihood surfaces

1. Hamiltonian MC - this technique explicitly makes use of gradient information (loosely analagous to covariance information) to explore the parameter space efficiently. It is not clear that HMC will be more efficient than emcee in this respect though. And of course it means writing down expressions for the gradient of the likelihood with respect to every parameter, but this may not be too difficult if the model is constructed carefully.

2. reversible-jump MCMC. allows the number of components to vary. requires model-comparison and all the issues therein.

3. **?** has developed some MCMC routines that use covariance information of accepted proposals to specify new proposals. Not clear how this maintains detailed balance. Applied to exoplanet parameter inference.

4. making the basis spectra orthogonal before calculating likelihoods. A potential problem with this is dust. Also, the spectral matrix will have to be diagonalized separately for each combination of observed wavelengths and velocity dispersions.

5. solve with a lower time -resolution and use the likelihood samples as intial guesses for the amplitudes of sub-bins when solving at higher resolution. This at least keeps you near a likelihood maximum as the dimensionality increases.

6. something more formally and strictly heirarchical than the last method?

3.3. *The Use of Photometry*

Photometry can be added in cases where the the spectroscopic normalization is unsure. If there is strong confidence in the spectroscopic normalization, then the photometric information does provide much power for inference, unless the spectroscopic S/N is much lower than the photometric uncertainty (by a factor of approximately the number of independent spectroscopic elements within the filter bandpass).

In everything that follows w consider linear photomtric units, maggies, defined as

$$f_b(L_\lambda) = \frac{1}{K_b} \int_0^\infty \lambda L_\lambda R_b(\lambda) d\lambda \quad (11)$$

where $R_B(\lambda)$ is the bandpass B response (detector signal/photon) and K_B is a constant related to the zeropoint of the magnitude scale.

The addition of photometry to the model is made simply by the addition of a photometric term to $\ln \mathcal{L}$ from spectroscopy

$$\ln \mathcal{L}_{tot} = \ln \mathcal{L}_{spec} + \ln \mathcal{L}_{phot} \quad (12)$$

We also write

$$\ln \mathcal{L}_{phot} = \sum_b \frac{(f_b - 10^{-0.4 M_b})^2}{\sigma_b^2} \quad (13)$$

where $\sigma_b = 1.086 \delta M_b 10^{-0.4 M_b}$.

When does the addition of photometry change the total likelihood and thus provide additional constraints on the model? Consider the likelihood ratio of two models, where the normalization of the spectroscopic data is considered unknown and allowed to vary, but the normalization of the photometric data is fixed. Consider the lieklihood ratio of two models where both normalizations are fixed and 1) the photometric data is within the spectroscopic bandpass or 2) the photometric data point is outside the range of the spectroscopy.

3.3.1. *Gradients of the photmetric likelihood*

$$\frac{\partial \ln \mathcal{L}_{phot}}{\partial a_i} = -\frac{1}{2} \sum_b 2 \frac{(f_b((L - K(\vec{b})/I(\vec{a}))) - m_b)}{\sigma_b^2} f_b(S_{\lambda,i} e^{-k_\lambda \tau_{V,i}})$$
$$\frac{\partial \ln \mathcal{L}_{phot}}{\partial \tau_{V,i}} = -\frac{1}{2} \sum_\lambda 2 \frac{((f_b((L - K(\vec{b})/I(\vec{a}))) - m_b))}{\sigma_b^2} f_b(A_i S_{\lambda_i} e^{-k_\lambda \tau_{V,i}} k_\lambda) \quad (14)$$

3.3.2. *Aperture Correction*

4. TESTS: RECOVERY OF SFHS

4.1. *SSPs and constant SFR*

4.2. *Parameterized SFHs - τ models with bursts*

4.3. *Realistic SFHs from ANGST*

For testing we consider the SFHs from the angst project.

4.4. *SFHs from hydro-simulations and/or SAMs*

5. TESTS: DEPENDENCE OF RESULTS ON OBSERVATIONAL PARAMETERS

1. S/N ratio
2. wavelength range ($\lambda_{min}, \lambda_{max}$)
3. spectral or velocity resolution

show contours of $\delta\theta$ as a fn of these instrument characteristics for a number of the test SFHs. compare to the uncertainties on the CMD based SFHs?

6. CAVEATS AND LIMITATIONS

subject to uncertainties in the SPS models (AGB lifetimes and SEDs, IMFs, isochrones or tracks). In principle these aspects can be modeled and marginalized over (**?**) but the likelihood calls become very expensive.