

ON THE ESTIMATION OF SYSTEMATIC UNCERTAINTIES OF STAR FORMATION HISTORIES

ANDREW E. DOLPHIN

Raytheon Company, Tucson, AZ 85734, USA; adolphin@raytheon.com
 Received 2012 February 18; accepted 2012 March 20; published 2012 May 4

ABSTRACT

In most star formation history (SFH) measurements, the reported uncertainties are those due to effects whose sizes can be readily measured: Poisson noise, adopted distance and extinction, and binning choices in the solution itself. However, the largest source of error, systematics in the adopted isochrones, is usually ignored and very rarely explicitly incorporated into the uncertainties. I propose a process by which estimates of the uncertainties due to evolutionary models can be incorporated into the SFH uncertainties. This process relies on application of shifts in temperature and luminosity, the sizes of which must be calibrated for the data being analyzed. While there are inherent limitations, the ability to estimate the effect of systematic errors and include them in the overall uncertainty is significant. The effects of this are most notable in the case of shallow photometry, with which SFH measurements rely on evolved stars.

Key words: methods: data analysis – galaxies: stellar content

1. INTRODUCTION

Resolved stellar populations provide a wealth of information from which a galaxy’s star formation history (SFH) can be inferred. The various features visible in a color–magnitude diagram (CMD) provide indicators of eras of star formation. For example, the presence of upper main-sequence stars indicates recent star formation, while horizontal branch stars indicate much older periods of star formation. Since most epochs of star formation can be tied to specific populations, consideration of all age-sensitive populations permits a qualitative measurement of a galaxy’s SFH, as summarized by Hodge (1989).

More recently, quantitative approaches to this problem have been developed, in which the exact distribution of stars in a CMD is analyzed in order to more precisely determine the SFH (e.g., Gallart et al. 1996; Tolstoy & Saha 1996; Dolphin 1997; Hernandez et al. 1999; Holtzman et al. 1999; Harris & Zaritsky 2001). These techniques rely on the use of synthetic CMDs that are generated by making use of theoretical isochrones, a stellar initial mass function, a model of observational effects, and often other factors such as reddening, foreground stars, and unresolved binaries. These synthetic CMDs can be generated for an unlimited number of possible SFHs, and the history producing the synthetic CMD most resembling the observed data is reported as the measured SFH.

To make this comparison, the CMD is generally divided into a number of bins. Within each bin, the number of observed stars is compared with the synthetic model using a goodness of fit statistic such as χ^2 , which can be written in the case of Poisson statistics as

$$\chi^2 = \sum \frac{(n_i - m_i)^2}{m_i}, \quad (1)$$

where m_i is the number of model points in the bin and n_i is the number of observed points. A comparable statistic based on a Poisson noise model (Dolphin 2002) is given by

$$-2 \ln P = 2 \sum m_i - n_i + n_i \ln \frac{n_i}{m_i}. \quad (2)$$

With either statistic, the resulting figure of merit from the CMD comparison is related to the probability that the observed data were drawn from the synthetic model. Assuming a uniform

prior on all potential SFHs ($P(\text{SFH}) = 1$), Bayes’ theorem can be invoked, and this value can be interpreted as the probability that the SFH used to generate that synthetic model is correct:

$$P(\text{SFH}|\text{CMD}) = \frac{P(\text{CMD}|\text{SFH}) P(\text{SFH})}{P(\text{CMD})}. \quad (3)$$

Uncertainties of the measured SFH are generally calculated with some form of Monte Carlo analysis. This involves the generation of a large number of simulated observed CMDs, either from random realizations of the best-fitting model or from random resamplings of the original photometry. The SFH is measured for each of these simulated CMDs, and the variation in these SFHs is interpreted as the uncertainty in the SFH of the original data. Effects of uncertainties from distance, extinction, or the solution method itself can be incorporated as well.

A factor excluded from SFH analysis as described above is the effect of systematic uncertainties in the adopted isochrone set. This raises several concerns about the method. First, the possibility that observed data cannot be *exactly* modeled by any combination of isochrones could potentially invalidate the use of a probabilistic technique for CMD comparison. Second, if the probabilistic technique is capable of finding a well-constrained best match, the errors induced in that fit need to be understood and quantified. The present study attempts to address these issues.

2. EFFECTS OF SYSTEMATICS ON THE SOLUTION SPACE

To examine the effect of systematic errors on the ability to arrive at a solution, I have chosen the stressing case of extremely deep photometry (below the main-sequence turnoff) of a simple stellar population. (The term “simple” is used instead of “single” because the adaptations of isochrones used in this analysis contain spreads of 0.05 dex in age and 0.1 dex in metallicity.) The depth of photometry requires that all parts of the CMD are adequately fit in order to achieve a good solution, while the simple stellar population assumption requires that all observed stars are modeled using a single isochrone (rather than allowing multiple isochrones to cover the entire observed data). The latter requirement is particularly stressing on a probabilistic approach,

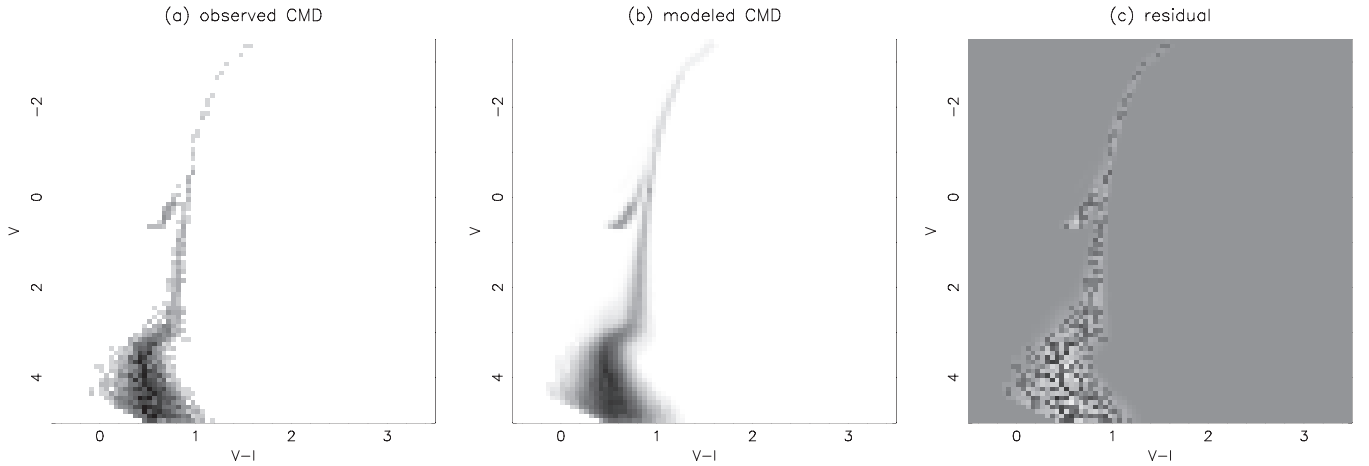


Figure 1. Best-fit model CMD for a simple stellar population. The left panel shows the simulated observed data, generated using the Padua isochrones (Marigo et al. 2008; Girardi et al. 2010). The middle panel shows the best-fit model, also using the Padua isochrones. The right panel shows the residual from subtracting the best-fit model from the simulated observations.

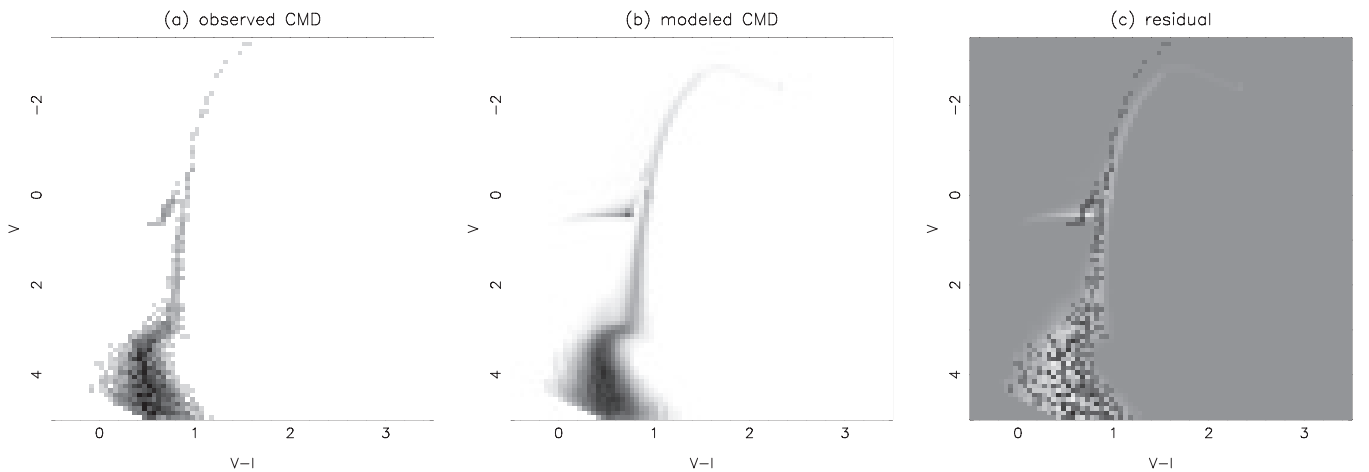


Figure 2. Same as Figure 1, but using the BaSTI isochrones (Pietrinferni et al. 2004) to fit the data.

since either formulation of the likelihood (Equations (1) and (2)) becomes infinite if a CMD bin contains observed data but no model data.

The analysis was done by creating a synthetic population of stars with ages from 10.0 to 11.2 Gyr ($\log(\text{age})$ of 10.00 to 10.05), metallicities from -1.9 to -2.0 in $[\text{Fe}/\text{H}]$, and the Padua isochrones (Marigo et al. 2008) with updated asymptotic giant branch (AGB) models (Girardi et al. 2010). Distance modulus and extinction were set to zero, and the V , I filter combination was used. The SFH was then measured using the MATCH package (Dolphin 2002). Three different isochrone sets were used for the solutions: the identical Padua models used to create the data, solar-scaled BaSTI models (Pietrinferni et al. 2004) with $\eta = 0.4$ and overshoot, and solar-scaled Dartmouth models (Dotter et al. 2008). Some amount of variability in red giant branch (RGB) mass loss has been modeled in the BaSTI and Dartmouth synthetic CMDs, but not in those computed from the Padua models. The solution was a four-parameter fit for age, metallicity, distance, and extinction. The age and metallicity parameters are the ones of primary scientific interest; the variations in distance and extinction allow for slight differences in the photometric calibrations adopted when computing the isochrones.

The best solution obtained using the Padua isochrones is shown in Figure 1. There are no statistically significant

differences between this solution and the synthetic population, and the best-fitting synthetic CMD was from the same population used to generate the data. This is naturally the best-case scenario.

The best solution obtained from the BaSTI isochrones is shown in Figure 2. As expected, the fit is not nearly as clean as the previous solution. Quantitatively, the synthetic photometry is 10^{150} times more likely to have been produced by the Padua models than by the BaSTI models. Because no single BaSTI isochrone exactly matches the Padua isochrone used to generate the data, the algorithm must prioritize the CMD regions. The choice is made to model the most strongly populated parts of the CMD (the main sequence and subgiants) as well as possible, while allowing the upper RGB to deviate very strongly. The horizontal branch is also fit as well as possible given the differences in the models. The best fit is obtained with a slightly younger age (7.9–8.9 Gyr), higher metallicity (-1.3 to -1.4), and slight errors in both distance and extinction.

Finally, the best solution obtained from the Dartmouth isochrones is shown in Figure 3. As with the BaSTI solution, there are significant errors in the fit, as again no single isochrone exactly matched the Padua models used to generate the simulated observations. In this case, the synthetic photometry is 10^{106} times more likely to have been drawn from the Padua models than from the Dartmouth models—an improved fit relative

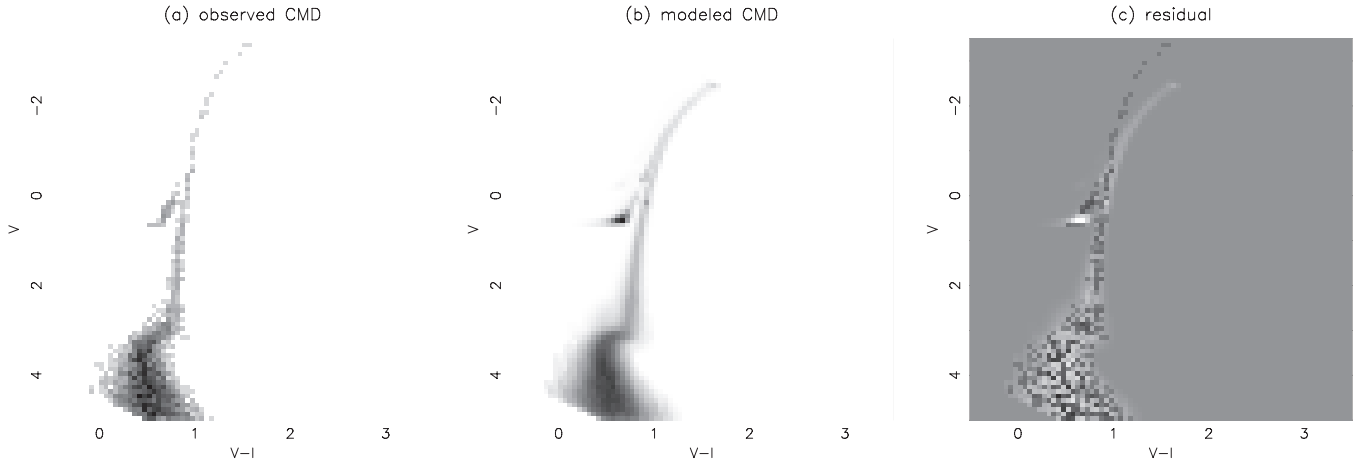


Figure 3. Same as Figure 1, but using the Dartmouth isochrones (Dotter et al. 2008) to fit the data.

Table 1
Measured Parameters of a Simple Stellar Population

Data Set	Age (Gyr)	Metallicity ([Fe/H])	Distance ($m - M$)	Extinction (A_V)
Input population	10.59	-1.95	0.000	0.000
Padua solution	10.73 ± 0.07	-1.948 ± 0.015	0.003 ± 0.004	-0.005 ± 0.003
BaSTI solution	8.90 ± 0.10	-1.318 ± 0.033	0.082 ± 0.003	-0.090 ± 0.002
Dartmouth solution	9.46 ± 0.09	-1.080 ± 0.013	0.011 ± 0.004	-0.161 ± 0.003

Note. The data were simulated using the Padua models; three different model sets were used to measure the SFHs.

to the BaSTI models but nevertheless not a good match. And, as with the BaSTI solution, the model producing the best fit did not match the input population; the stars were younger (8.9–10 Gyr) and significantly higher metallicity (−1.0 to −1.1). The distance was nearly identical, but the extinction was 0.16 magnitudes less in A_V .

In addition to using the best solution to estimate the values of the four parameters, the dependencies of the goodness-of-fit on the four solution parameters can be used to estimate how well those parameters are constrained. As the fit parameter from Equation (2) is derived from the Poisson probability distribution, it can be converted directly into a four-dimensional probability density map. Each parameter’s probability density function can be obtained by marginalizing the other three. Applying this procedure, the best fits and uncertainties from the three solutions above are listed in Table 1.

While there are clear systematic errors in the results shown in this table, it is encouraging that, even under stressing conditions, a probabilistic routine is able to measure a most probable set of parameters. More significantly, the minimum surrounding the best solution is similarly well defined in the presence of systematic errors as it is without them. As seen in Table 1, the uncertainties in age and metallicity measured when using the BaSTI models are about twice those measured using the Padua models; the uncertainties in the Dartmouth-based solution are comparable to the Padua-based solution.

It should be noted that the behavior of a CMD-fitting routine in the face of such low probabilities likely depends strongly on the robustness of the algorithm. However, at least in the case of the SFH measurement package being used in this analysis, the ability to find a solution is not significantly impeded by the presence of significant systematic errors.

3. SIMPLE STELLAR POPULATIONS

With confidence that precise (though inaccurate) fits can be obtained despite the presence of systematic errors, a logical step is to quantify the size of the errors in measured stellar populations induced by these errors in the models used to measure them. To accomplish this, the experiment described in the previous section was expanded. All three isochrone sets were used to generate the synthetic photometry, and five populations of different ages and metallicities were simulated.

Results from this experiment are shown in Table 2. Looking at the summary lines, it is clear that the effects of systematic differences between the three isochrone sets is not constant between the populations. For example, there were no errors in the youngest population’s age measurement, but the metallicity was in error by 0.2 dex in both cases. On the other hand, the ~2.5 Gyr population had a smaller error in metallicity but a larger error in age.

Because of the non-uniformity of age and metallicity systematic uncertainties, it is tempting to draw the conclusion that uncertainties should be estimated by measuring the SFH with as many isochrone sets as is possible, and setting the uncertainty equal to the standard deviation of the solutions. While this is possible in some cases, there are two significant limitations. First, not all isochrone sets will cover the necessary ranges in age, metallicity, and evolutionary phases. For example, of the models used in this analysis, only the Padua models provide isochrones younger than 25 Myr. Thus, systematic errors could not be measured for systems containing ongoing star formation.

Even in the case of older populations for which all isochrone sets used here are available, the use of three SFH measurements to estimate the uncertainty created by systematic errors is

Table 2
Best-fitting Models for Various Simple Stellar Populations

Input			Solution		
Isochrones	Log Age	[Fe/H]	Isochrones	Log Age	[Fe/H]
Padua	7.825	−0.75	BaSTI	7.825	−0.95
BaSTI	7.825	−0.75	Padua	7.825	−0.55
Standard Dev				0.00	0.25
Padua	8.425	−0.75	BaSTI	8.425	−0.55
BaSTI	8.425	−0.75	Padua	8.375	−0.85
Standard Dev				0.03	0.19
Padua	9.425	−0.85	BaSTI	9.375	−0.65
Padua	9.425	−0.85	Dartmouth	9.525	−0.65
BaSTI	9.425	−0.85	Padua	9.475	−1.05
BaSTI	9.425	−0.85	Dartmouth	9.475	−0.85
Dartmouth	9.425	−0.85	Padua	9.425	−1.15
Dartmouth	9.425	−0.85	BaSTI	9.375	−0.85
Standard Dev				0.06	0.15
Padua	9.725	−1.05	BaSTI	9.725	−1.05
Padua	9.725	−1.05	Dartmouth	9.725	−0.95
BaSTI	9.725	−1.05	Padua	9.725	−1.05
BaSTI	9.725	−1.05	Dartmouth	9.775	−1.15
Dartmouth	9.725	−1.05	Padua	9.675	−0.85
Dartmouth	9.725	−1.05	BaSTI	9.675	−0.85
Standard Dev				0.03	0.13
Padua	10.025	−1.35	BaSTI	9.975	−0.95
Padua	10.025	−1.35	Dartmouth	10.025	−0.85
BaSTI	10.025	−1.35	Padua	10.025	−1.65
BaSTI	10.025	−1.35	Dartmouth	10.075	−1.45
Dartmouth	10.025	−1.35	Padua	10.075	−1.65
Dartmouth	10.025	−1.35	BaSTI	9.975	−1.45
Standard Dev				0.04	0.36

Notes. The left columns indicate the isochrone set and stellar population used to generate the CMD. The right columns indicate the isochrone set used to solve for the best fit, as well as the mean age and mean metallicity of that best fit. Simple stellar populations were used in both CMD generation and SFH measurement. Note that standard deviations were estimated using the mean absolute value of the differences, which is an unbiased estimator of standard deviation. For the case of a Gaussian distribution, the standard deviation equals the mean error multiplied by $\sqrt{\pi/2}$. Note also that Dartmouth isochrones were not used at ages younger than 1 Gyr, as the younger isochrones are not implemented in MATCH.

analogous to using two Monte Carlo runs to estimate random errors.

4. ESTIMATION OF SYSTEMATIC UNCERTAINTIES

In order to estimate systematic uncertainties, one must be able to generate nearly unlimited representations of reasonable systematic errors, each of which can be analyzed. This section outlines one such method, although other viable alternatives certainly exist. It should be emphasized that the primary motivation of this study is to understand and quantify effects of model differences on the measured SFH, not to examine or critique the underlying modeling choices and methods that created these differences.

The adopted approach is to model the isochrone differences as shifts in M_{bol} and $\log T_{\text{eff}}$. The rationale behind the choice of $\log T_{\text{eff}}$ rather than color is that a simple color shift could result in unphysical temperatures (e.g., $V - I = -1$). It also produces a system that is more likely to be portable between different filter sets with minimal modifications.

It should be emphasized that this is not to be confused with shifts in distance modulus and extinction that are commonly

applied during the determination of the best-fit SFH (including the solutions shown in Table 1). The distance modulus and extinction shifts are used to compensate for uncertainties such as those in the photometric zero points, photometric calibration of the isochrone sets, and foreground extinction. The shifts described in this section are used to intentionally introduce systematic errors between the observed data (which are not shifted) and the synthetic models (which are). It is important to note that, for solutions in which these shifts are applied, no solution in distance modulus or extinction is allowed, since this would mitigate the inserted error. Specifically, any error induced in the bolometric magnitudes could exactly be offset by an identical shift in distance modulus, thereby resulting in exactly the same SFH as would have been observed had no shifts been applied to either variable.

For this to work successfully, it is necessary that a set of shifts in the M_{bol} , $\log T_{\text{eff}}$ space can adequately reproduce the errors in age and metallicity for all five populations shown in Table 2. In order to make this transformation, the change in recovered age and metallicity can be measured as a function of the shifts in luminosity and temperature. For the youngest population in Table 2, this can be described with the following equation:

$$\begin{pmatrix} \Delta \log(\text{age}) \\ \Delta [\text{Fe}/\text{H}] \end{pmatrix} = \begin{pmatrix} -0.25 & -0.05 \\ 13.5 & 1.2 \end{pmatrix} \begin{pmatrix} \Delta \log T_{\text{eff}} \\ \Delta M_{\text{bol}} \end{pmatrix}. \quad (4)$$

This indicates that the accurate recovery of the age of the population (even in the presence of systematic errors) is because it would require a very large error in temperature or luminosity to create a noticeable change in the measured age. However, a relatively small difference in either temperature or luminosity could create a significant error in the recovered metallicity (a result of the relative insensitivity of upper main-sequence color to metallicity).

One can invert the matrix in Equation (4) to solve for the temperature and luminosity shifts that would result in the observed errors in age and metallicity. In this example, there was zero age error but -0.2 dex of metallicity error when generating data with the Padua models and solving with the BaSTI models. Put into the above equation, this translates to temperature and luminosity shifts of

$$\begin{pmatrix} \Delta \log T_{\text{eff}} \\ \Delta M_{\text{bol}} \end{pmatrix} = \begin{pmatrix} 3.2 & 0.13 \\ -36 & -0.67 \end{pmatrix} \begin{pmatrix} 0.0 \\ -0.25 \end{pmatrix} = \begin{pmatrix} -0.033 \\ 0.17 \end{pmatrix}. \quad (5)$$

Indeed, when applying shifts of -0.027 in $\log T_{\text{eff}}$ and 0.133 in M_{bol} while solving the SFH using the Padua models on data generated with the Padua models, one induces the expected error in metallicity, with zero age error. Expanding this result, many solutions could be made with shifts comparable in size to these. In doing so, a fuller sampling of the effects of systematics can be obtained.

Recalling that it is necessary to find shifts that adequately model all of the populations in Table 2, similar math is performed on the other four. The most extreme case is the oldest population, in which the matrix is nearly singular and thus very large variations of $\sigma_{\log T_{\text{eff}}} = 0.142$ and $\sigma_{M_{\text{bol}}} = 1.14$ are obtained. However, the uncertainties on those values are also very high. When combining the results of all five populations, values of $\sigma_{\log T_{\text{eff}}} = 0.012$ and $\sigma_{M_{\text{bol}}} = 0.18$ are found.

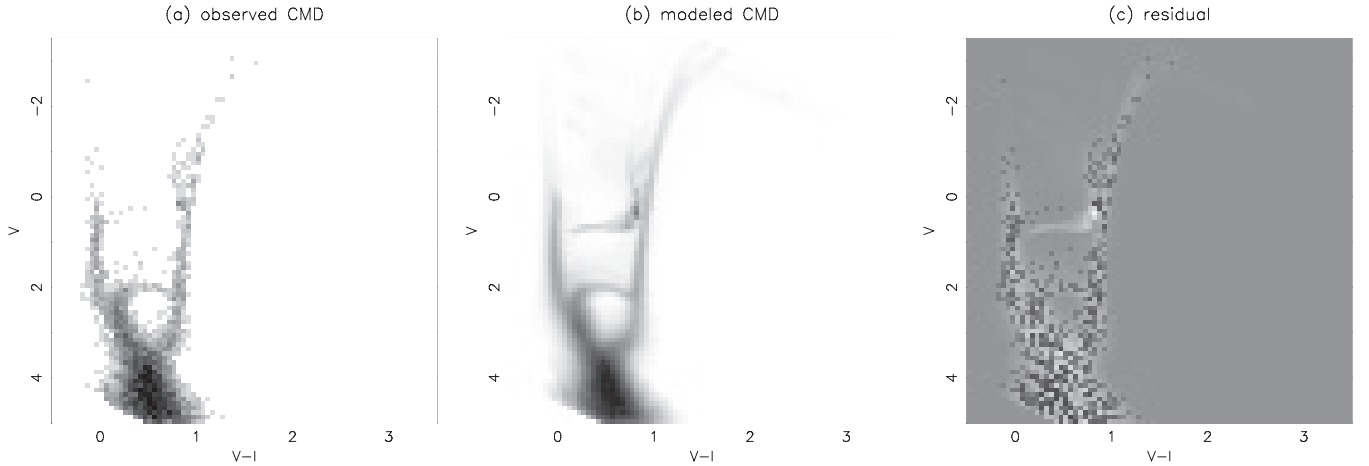


Figure 4. Best-fit model CMD for a multi-burst stellar population. The BaSTI models were used to generate the simulated observations, and the Padua models were used to fit the data. Panels are the same as in Figure 1.

Table 3

Comparison between Measured and Estimated Errors due to Systematic Errors

Input		Measured		Calculated	
Log Age	[Fe/H]	$\sigma_{\log \text{ age}}$	$\sigma_{[\text{Fe}/\text{H}]}$	$\sigma_{\log \text{ age}}$	$\sigma_{[\text{Fe}/\text{H}]}$
7.825	-0.75	0.00	0.25	0.01	0.27
8.425	-0.75	0.03	0.19	0.00	0.25
9.425	-0.85	0.06	0.15	0.08	0.18
9.725	-1.05	0.03	0.13	0.04	0.18
10.025	-1.35	0.04	0.36	0.03	0.21

Notes. The first two columns show the age and metallicity of the population used to generate the data. The third and fourth columns show the measured standard deviation in age and metallicity errors and are repeated from Table 2. The final two columns show the errors in age and metallicity obtained by applying random shifts with standard deviations of $\sigma_{\log T_{\text{eff}}} = 0.012$ and $\sigma_{M_{\text{bol}}} = 0.18$.

Table 3 shows the resulting age and metallicity errors obtained by applying shifts randomly selected from a Gaussian distribution of mean zero and standard deviation as specified above. While these errors do not exactly match those measured from the simple populations, they are well within a factor of two in all cases, and within 30% for all but one. To put into context, MATCH's maximum resolutions in age and metallicity are 0.05 and 0.1 dex, respectively. More importantly, in cases in which only two isochrones were used, the uncertainty of the standard deviation is $\sim 75\%$ of the measured standard deviation. Even in the cases for which three isochrone sets were used, the uncertainty is about half the measured standard deviation. Thus, in the presence of at minimum 50% uncertainties in the measured error distributions, the results from Table 3 show the temperature and luminosity shifts to adequately reproduce the observed systematic errors in all five populations.

5. MULTIPLE STELLAR POPULATIONS

While the above section demonstrated that consistent shifts in temperature and luminosity can induce appropriately sized errors in recovered simple populations across a wide spectrum of age, it needs to be seen whether or not the process would work for more complex systems such as field populations of resolved galaxies. In this section, this topic is addressed by analysis of the two extreme cases: a population whose entire SFH is comprised of a small number of short bursts, and a population with constant star formation rate for its entire history.

For the first case, a four-burst stellar population was created using the BaSTI isochrones. The four bursts were defined as follows:

1. 50–56 Myr, mean [Fe/H] = -0.45 ;
2. 500–562 Myr, mean [Fe/H] = -0.75 ;
3. 2.5–2.8 Gyr, mean [Fe/H] = -1.15 ; and
4. 12.6–14.1 Gyr, mean [Fe/H] = -1.95 .

The SFH of this population was then measured using the Padua isochrones, with no shifts in temperature or luminosity. In addition, 50 solutions were made with randomly chosen shifts (using Gaussian distributions with $\sigma_{\log T_{\text{eff}}} = 0.012$ and $\sigma_{M_{\text{bol}}} = 0.18$). The variation between these solutions was used to measure the systematic uncertainties. Finally, to understand what the measurement would have been in the absence of systematic errors, the history was measured using the BaSTI isochrones.

The best fit using the Padua models is shown in Figure 4, and is not nearly as bad as the fits shown in Section 2. The reason for this is that MATCH is allowed to mix populations, allowing more portions of the CMD to be fit. In this case, only the horizontal branch was fit poorly.

The measured SFHs are shown in Figure 5. The solid line and shaded region show the best fit using the Padua models and the uncertainty estimated from the fits with the Padua models and shifts in luminosity and temperature. The dashed line shows the best fit obtained using the BaSTI models. The error-free measurement generally sits near the upper error bars, though at some times (for example, from 0.5 to 2.5 Gyr ago) falls significantly outside the bounds. In this case, the ratio of star formation in the middle two bursts was systematically misestimated in all of the runs with the Padua models. Most likely this is a result of the two isochrone sets having different lifetimes of certain evolutionary phases, an error source not accounted for in this analysis. On the other hand, the estimated uncertainties on the measured burst ages were correct.

The second test was carried out in the other extreme, a constant star formation rate from 14.1 Gyr ago to the present. As before, the synthetic data were generated using the BaSTI models and solved using the Padua models, including 50 runs with temperature and luminosity shifts to estimate the systematic uncertainties.

Results of this test are shown in Figures 6 and 7. The quality of the fit was not nearly as good as in the multiple-burst test, as

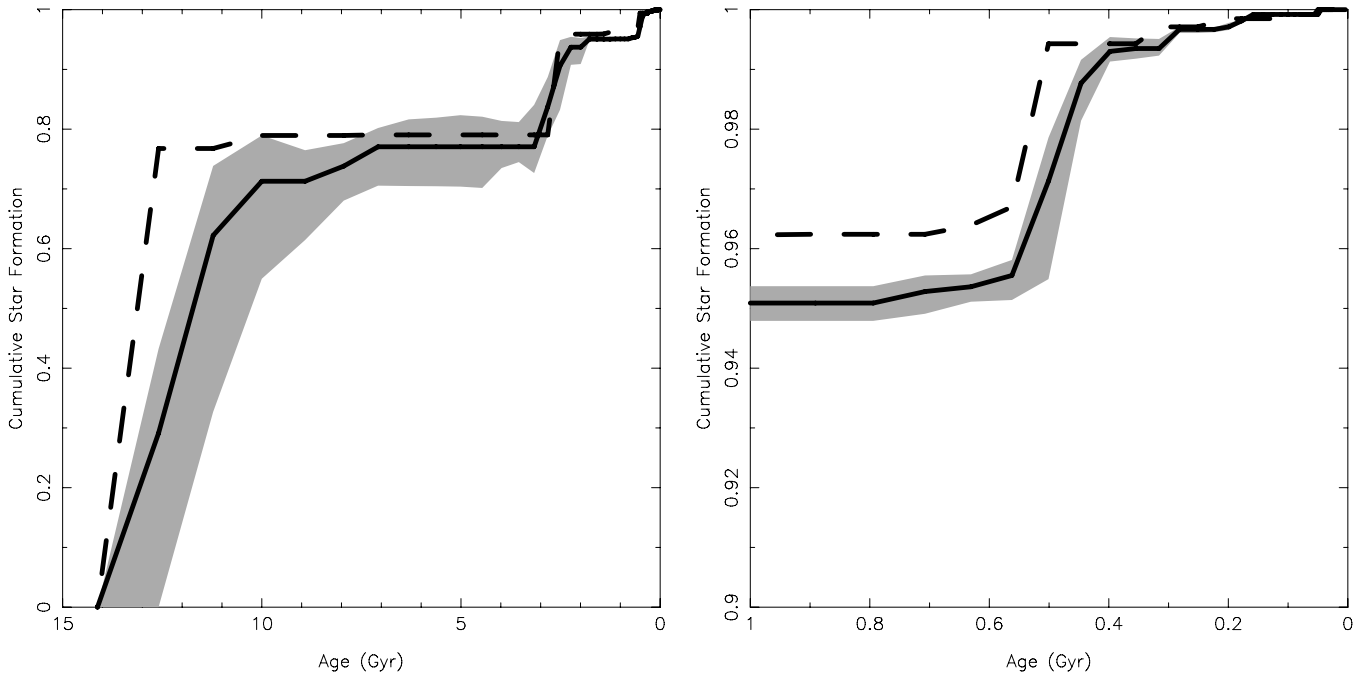


Figure 5. Measured SFH of a stellar population with four bursts of star formation, in the presence of systematic errors. The solid line represents the best fit to the SFH, and the shaded region shows the 1σ uncertainties. Finally, the history as measured with no systematic errors is shown by the dashed line. The left panel shows the entire history of the galaxy; the right panel zooms into the past Gyr.

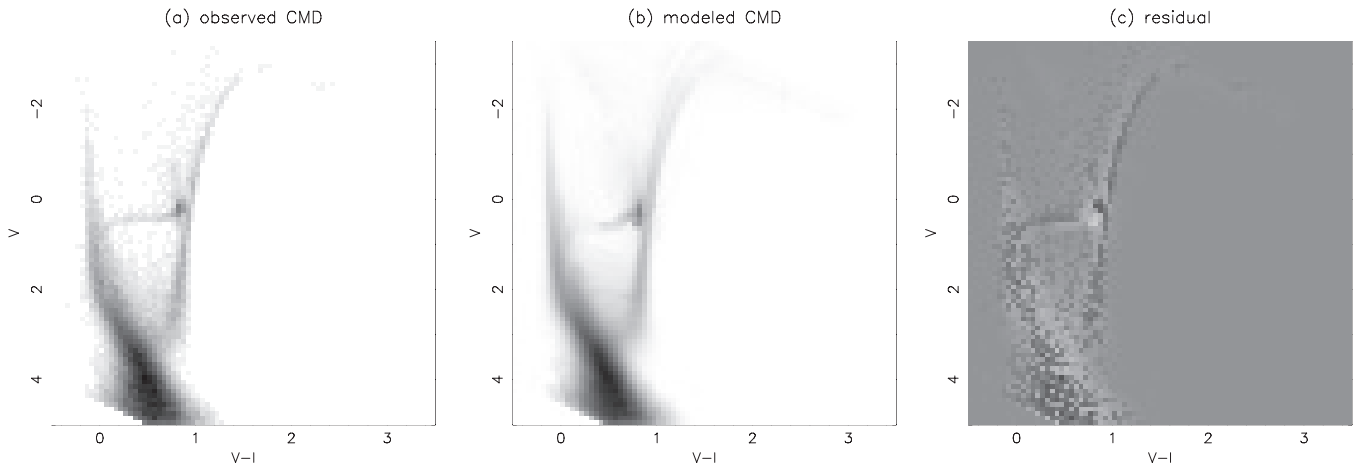


Figure 6. Best-fit model CMD for a stellar population with continuous star formation. The BaSTI models were used to generate the simulated observations, and the Padua models were used to fit the data. Panels are the same as in Figure 1.

the presence of all ages required all CMD features to be fit. The estimation of the uncertainties appears to have been successful; the measurement with no systematic errors fell generally within the error bars and never outside them by much. The largest deviation, seen at 8.9 Gyr, is equivalent to a 2σ error.

6. DISCUSSION

Ideally, there would be a nearly limitless number of suitable isochrone sets covering the entire space of reasonable choices of the modeling parameters. The reality is that one must make the best possible use of a handful of models, not all of which may cover the entire range of initial mass or evolutionary phases that are needed in an analysis. Thus, the direct measurement of systematic errors by analysis of multiple isochrone sets is inaccurate at best, and impossible at worst.

The approach I have outlined for estimating the uncertainties due to systematic errors has relied on finding some simple proxy for the isochrone differences that can be varied in a large number of Monte Carlo runs in order to observe its effects. This provides the advantage that it can be applied in populations for which only one isochrone set provides coverage (for the three isochrone sets analyzed here, this would be for ages younger than 25 Myr), and also that rather than relying on one or two independent pairs of isochrone sets to sample the distribution of errors produced in the SFH measurement, a nearly unlimited sample can be generated by random draws.

The choice made was to use shifts in luminosity and effective temperature, which can be applied relatively simply during the CMD fitting process. This ignores systematic differences in measured histories that could result from different lengths of evolutionary phases. In cases in which the SFH is dominated by

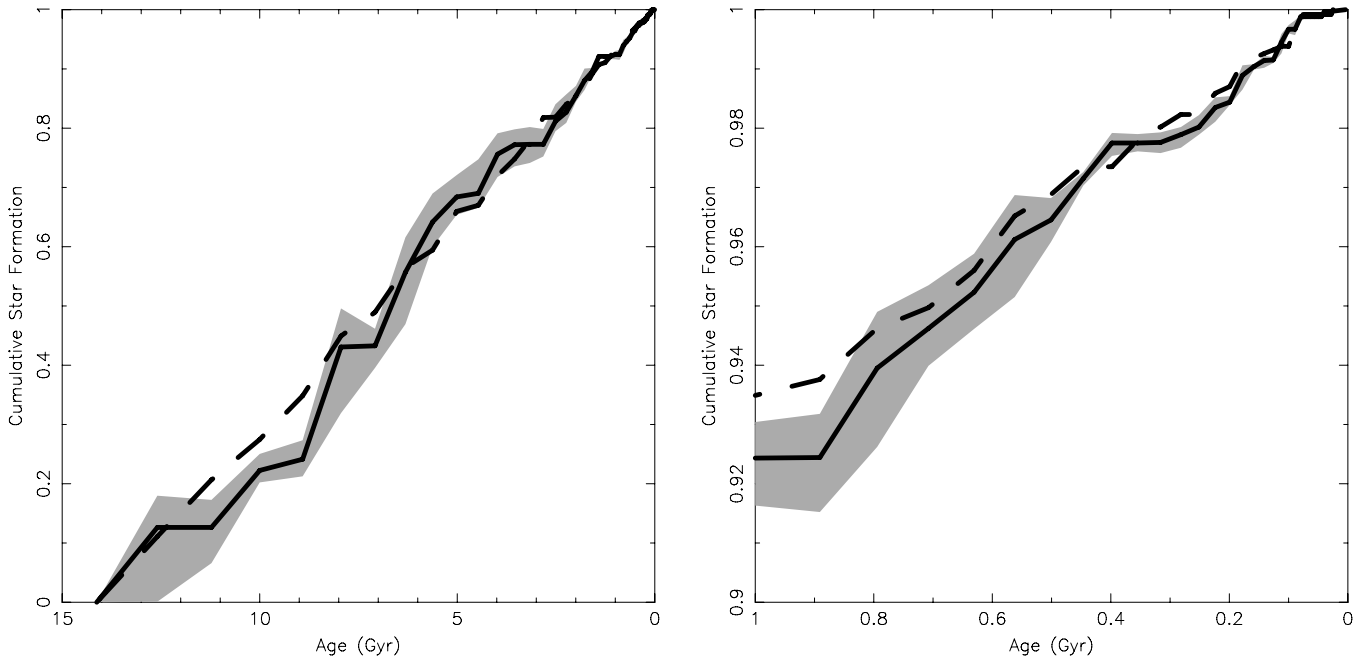


Figure 7. Measured SFH of a stellar population with continuous star formation, in the presence of systematic errors. Panels and lines are identical to those in Figure 5.

Table 4

Variation of Temperature and Luminosity Shifts as a Result of Variations in Photometric Depth

M_V Limit	M_I Limit	$\sigma_{\log T_{\text{eff}}}$	$\sigma_{M_{\text{bol}}}$
-2.0	-2.5	0.032	0.40
-1.0	-1.5	0.024	0.35
0.0	-0.5	0.026	0.35
1.0	0.5	0.019	0.18
2.0	1.5	0.018	0.18
3.0	2.5	0.020	0.31
4.0	3.5	0.013	0.19

Notes. The left two columns show the 50% completeness limit; the right two columns show the shifts needed to adequately model effects of systematic errors.

a small number of short bursts, this can cause the uncertainties in the ratio of star formation in each burst to be underestimated. For field populations with continuous star formation, this is less of a problem.

The measurement of appropriate shift sizes is thus key to the success. Because the systematic error model is not physics based, the shifts should be measured for every data set. For example, photometric depth will affect which CMD features dominate the measurement of the SFH, causing the best overall shift size to change somewhat based on how sensitive those features are to modeling choices. Effects of this on the shift sizes are shown in Table 4. Data shown in this table are computed using the averages of five measurements of the shifts at each photometric depth. While the naive expectation might have been a reduction in shifts as the photometry is deeper, this is not true in every case. What is true is that the resulting errors in age and metallicity become progressively smaller as the photometry becomes deeper.

In addition to photometric depth, the filter choice and stellar population being examined can affect the ideal shifts to apply to a data set. Table 5 shows the shifts applied to deep photometry (beyond the ancient main-sequence turnoff) for three different

Table 5

Size of Temperature and Luminosity Shifts as a Function of Filter Combination

Blue Limit	Red Limit	$\sigma_{\log T_{\text{eff}}}$	$\sigma_{M_{\text{bol}}}$
$M_B = 5.0$	$M_V = 4.5$	0.011	0.17
$M_V = 4.0$	$M_I = 3.5$	0.013	0.19
$M_J = 3.0$	$M_K = 2.5$	0.016	0.26

Notes. In all three cases, the photometry reaches below the ancient main-sequence turnoff. As with Table 4, the 50% completeness limits are given in the left two columns, while the required shifts are given in the right columns.

Table 6

Variation of Temperature and Luminosity Shifts as a Result of Variations in Photometric Depth for the B , V Filter Combination

M_B Limit	M_V Limit	$\sigma_{\log T_{\text{eff}}}$	$\sigma_{M_{\text{bol}}}$
-1.0	-1.5	0.038	0.47
0.0	-0.5	0.023	0.39
1.0	0.5	0.027	0.38
2.0	1.5	0.015	0.24
3.0	2.5	0.015	0.34
4.0	3.5	0.009	0.25
5.0	4.5	0.011	0.17

Note. Columns are the same as in Table 4.

filter combinations. While the B , V filter combination behaves similarly to V , I , much larger luminosity variations are needed in J , K . Population effects can also be present. As seen in Section 3, the same shifts are not obtained for every age or metallicity. Thus, the appropriate shifts must be measured for the specific conditions seen in the data being analyzed. For the sake of completeness, shifts as a function of photometric depth for the B , V filter combination is given in Table 6, and in Table 7 for the J , K combination.

The goal of this analysis is a reliable method for estimating systematic uncertainties that can be incorporated into SFH results. In some cases, this significantly increases the uncertainties and resolves issues in which SFH algorithms report error bars

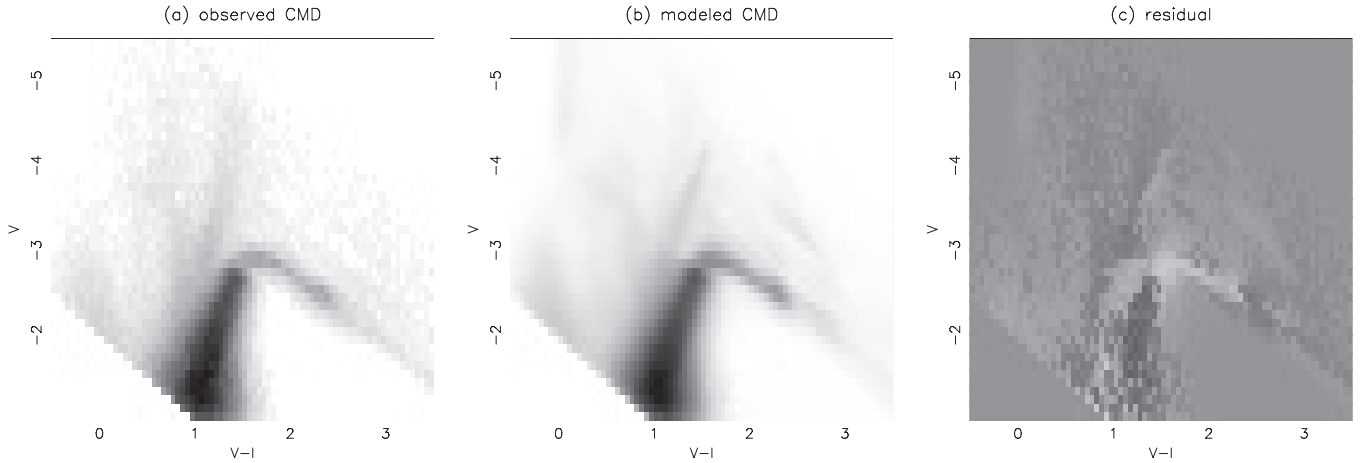


Figure 8. Best-fit model CMD for a stellar population with continuous star formation and shallow photometry. The BaSTI models were used to generate the simulated observations, and the Padua models were used to fit the data. Panels are the same as in Figure 1.

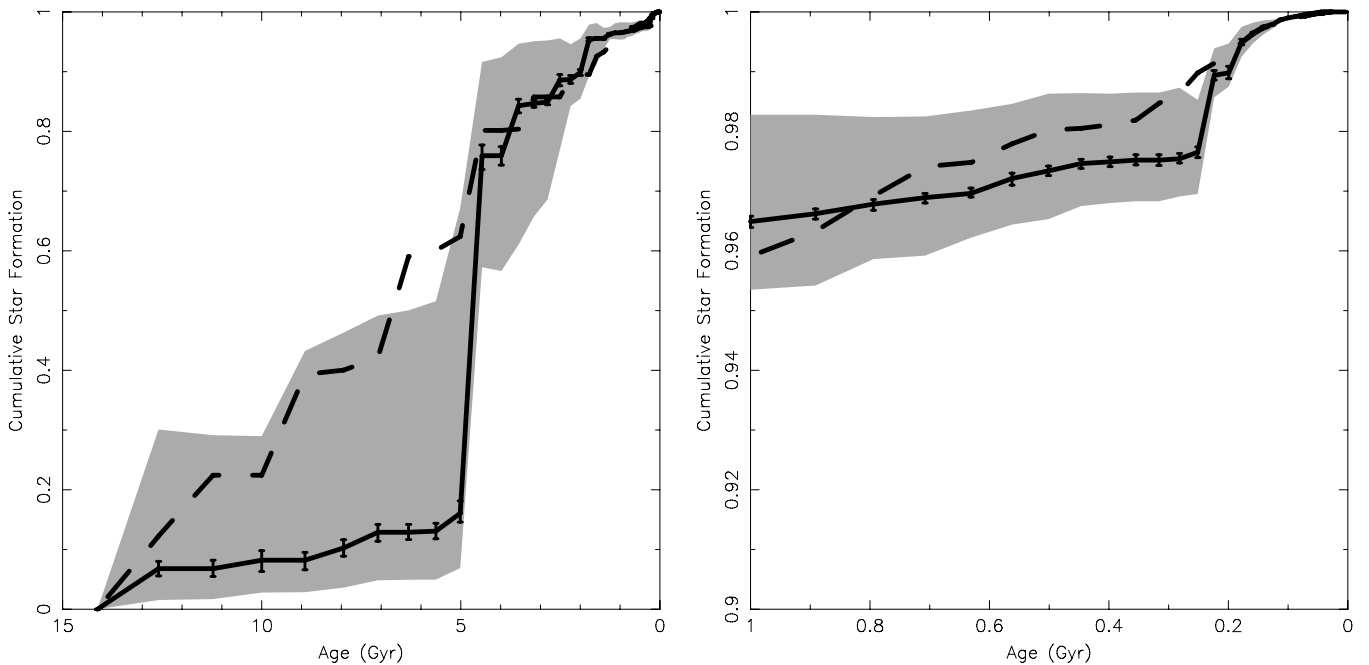


Figure 9. Measured SFH of a stellar population with continuous star formation, in the presence of systematic errors. Panels and lines are identical to those in Figure 5, except that uncertainties due to random errors (shown as the error bars) have been added.

Table 7

Variation of Temperature and Luminosity Shifts as a Result of Variations in Photometric Depth for the J , K Filter Combination

M_J Limit	M_K Limit	$\sigma_{\log T_{\text{eff}}}$	$\sigma_{M_{\text{bol}}}$
-3.0	-3.5	0.021	0.41
-2.0	-2.5	0.020	0.51
-1.0	-1.5	0.019	0.19
0.0	-0.5	0.019	0.15
1.0	0.5	0.018	0.13
2.0	1.5	0.018	0.23
3.0	2.5	0.016	0.26

Note. Columns are the same as in Table 4.

that appear unrealistically small. For example, the CMD shown in Figure 8 contains very little information for older populations: the bright red giants and AGB stars, for which not only are significant age/metallicity degeneracies present, but also

the evolutionary modeling is less certain. The measured SFHs from this field are shown in Figure 9. The systematic error estimate (gray region) shows the expected large uncertainty, while random errors only (error bars) are extremely small, and incorrectly show that approximately 70% of the star formation to have occurred in a sharp burst 5 Gyr ago. When incorporating systematic errors, the lack of age resolution is apparent and no statistically significant burst is measured. The solution with no systematic errors is thus in error at more than a 10σ level if using random errors only, while it is within the estimated systematic uncertainties at most ages.

7. SUMMARY

Systematic uncertainties, in the form of uncertainties in the adopted isochrones, are a significant (and frequently dominant) source of uncertainty in the measurement of SFHs. Despite this, SFH measurements are generally reported with minimal (if any)

analysis indicating the potential effects of systematic errors on the results.

I have outlined a process for estimating the effects of systematic errors on the measured SFH. This process is based on shifts in luminosity and temperature that can be applied. The shifts can be varied within some probability distribution, allowing the estimation of a large number of SFHs with reasonably sized systematic errors.

Being an entirely empirical technique, there are limitations. First, the sizes of the shifts to apply are dependent on the stellar population being observed, the filter set in use, and the depth of the photometry. The implication is that the shifts will be data set dependent, and thus should be recalculated for each data set. Second, while the method was calibrated to measure errors in inferred age and metallicity to reasonable accuracy, effects due to lifetimes of certain evolved populations are not an explicit part of the calibration. For field populations, this is not seen to cause significant issues. However, uncertainties in the histories of populations formed by a small number of short bursts can be underestimated.

Limitations notwithstanding, the analysis presented here indicates a significant improvement in the estimation of uncertainties in SFHs. The ability to obtain extremely small random errors in shallow, wide fields has long been known (e.g.,

Dolphin 2002), simply due to the effects of sample size on random errors. Inclusion of systematic errors in the uncertainty analysis prevents this, and results in error bars that more accurately represent the degree of confidence in the measured SFH.

Support for program number HST-GO-11986.07 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

REFERENCES

- Dolphin, A. 1997, *New Astron.*, **2**, 397
 Dolphin, A. E. 2002, *MNRAS*, **332**, 91
 Dotter, A., Chaboyer, B., Jevremovic, D., et al. 2008, *ApJS*, **178**, 89
 Gallart, C., Aparicio, A., Bertelli, G., & Chiosi, C. 1996, *AJ*, **112**, 1950
 Girardi, L., Williams, B. F., Gilbert, K. M., et al. 2010, *ApJ*, **724**, 1030
 Harris, J., & Zaritsky, D. 2001, *ApJS*, **136**, 25
 Hernandez, X., Valls-Gabaud, D., & Gilmore, G. 1999, *MNRAS*, **304**, 705
 Hodge, P. 1989, *ARA&A*, **27**, 139
 Holtzman, J. A., Gallagher, J. S., III, Cole, A. A., et al. 1999, *AJ*, **118**, 2262
 Marigo, P., Girardi, L., Bressan, A., et al. 2008, *A&A*, **482**, 883
 Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, *ApJ*, **612**, 168
 Tolstoy, E., & Saha, A. 1996, *ApJ*, **462**, 672