

Stellar population synthesis diagnostics

Y.K. Ng

Padova Astronomical Observatory, Vicolo dell'Osservatorio 5, I-35122 Padua, Italy

Received November 4, 1997; accepted March 27, 1998

Abstract. A quantitative method is presented to compare observed and synthetic colour-magnitude diagrams (CMDs). The method is based on a χ^2 merit function for a point (c_i, m_i) in the observed CMD, which has a corresponding point in the simulated CMD within $n\sigma(c_i, m_i)$ of the error ellipse. The χ^2 merit function is then combined with the Poisson merit function of the points for which no corresponding point was found within the $n\sigma(c_i, m_i)$ error ellipse boundary.

Monte-Carlo simulations are presented to demonstrate the diagnostics obtained from the combined (χ^2 , Poisson) merit function through variation of different parameters in the stellar population synthesis tool. The simulations indicate that the merit function can potentially be used to reveal information about the initial mass function. Information about the star formation history of single stellar aggregates, such as open or globular clusters and possibly dwarf galaxies with a dominating stellar population, might not be reliable if one is dealing with a relatively small age range.

Key words: methods: data analysis, numerical — stars: HR-diagram, statistics

1. Introduction

In the last decade the simulation of synthetic Hertzsprung-Russell and colour-magnitude diagrams (hereafter respectively referred to as HRDs and CMDs) has advanced at a rapid pace. It has been applied successfully in various studies of young clusters in the Magellanic Clouds (Chiosi et al. 1989; Bertelli et al. 1992; Han et al. 1996; Mould et al. 1997; Vallenari et al. 1990, 1992, and 1994a,b), dwarf galaxies in the Local Group (Aparicio & Gallart 1995, 1996; Aparicio et al. 1996, 1997a,b; Ferraro et al. 1989; Gallart et al. 1996a–c; Han et al. 1997; Tolstoy 1995, 1996; Tosi et al. 1991), open clusters in our Galaxy

(Aparicio et al. 1990; Carraro et al. 1993, 1994; Gozzoli et al. 1996), and the structure of our Galaxy (Bertelli et al. 1994, 1995, 1996; Ng 1994, 1997a,b; Ng & Bertelli 1996a,b; Ng et al. 1995, 1996a,b, 1997).

Generally all studies focus in the first place on matching the morphological structures at different regions in the CMDs (Gallart 1998). In the recent years a good similarity is obtained between the observed and the simulated CMDs. Unfortunately, the best fit is in some cases distinguished by eye. The morphological differences are large enough to do this and the eye is actually guided by a detailed knowledge of stellar evolutionary tracks. However, the stellar population technique has improved considerably and an objective evaluation tool is needed, to distinguish quantitatively one model from another.

Bertelli et al. (1992, 1995) and Gallart et al. (1996c) defined ratios to distinguish the contribution from different groups of stars. The ratios are defined so that they are sensitive to the age, the strength of the star formation burst and/or the slope of the initial mass function. Vallenari et al. (1996a,b) demonstrated that this is a good method to map the spatial progression of the star formation in the Large Magellanic Cloud. Robin et al. (1996), Han et al. (1997) and Mould et al. (1997) use a maximum likelihood method to find the best model parameter(s), while Chen (1996a,b) adopted a multivariate analysis technique. Different models can be quantitatively sampled through Bayesian inference (Tolstoy 1995; Tolstoy & Saha 1996) or a chi-squared test (Dolphin 1997).

In principle one aims with stellar population synthesis to generate a CMD which is identical to the observed one. The input parameters reveal the evolutionary status of the stellar aggregate under study. To obtain a good similarity between the observed and the simulated CMD one needs to implement in the model in the first place adequately extinction, photometric errors and crowding. It goes without saying that the synthetic population ought to be comparable with the age and metallicity (spread) of the stellar aggregate. Only with a proper choice of these parameters, one can start to study in more detail the

stellar initial mass function and the star formation history for an aggregate.

This paper describes a quantitative evaluation method based on the combination of a chi-squared and Poisson merit function¹. It allows one to select the best model from a series of models. In the next section this method is explained and Monte-Carlo simulations are made, to display how the diagnostic diagrams of the residual points can be employed. It is demonstrated that the non-fitting, residual points provide a hint about the parameter that needs adjustment in order to improve the model. This paper ends with a discussion about the method and the diagnostic diagrams together with their limitations.

It is emphasized that this paper deals with a description of a quantitative evaluation method for CMDs. Aspects related with the implementation of an automated CMD fitting program and comparison with simulated or real data sets will not be considered, because they are not relevant for the general validity of the method described in the following section.

2. Method

2.1. The chi-squared merit function

The method is based on minimizing a chi-squared merit function between all the observed points in a CMD which have a corresponding point in the synthetic CMD within a $3 - 5\sigma(c_i, m_i)$ error ellipse. In this range one can assume that the errors follow a Normal distribution. The χ^2 is a measure for the goodness of the fit for N observed points in a CMD and is defined as:

$$\chi^2(O, S) = \sum_{i=1}^N \frac{(c_{i,O} - c_{i,S})^2}{\sigma_i^2(c_{i,O})} + \frac{(m_{i,O} - m_{i,S})^2}{\sigma_i^2(m_{i,O})} \quad (1)$$

$$= \sum_{i=1}^N \left(\frac{\delta m(c_i, m_i)}{\sigma_i(c_{i,O}, m_{i,O})} \right)^2, \quad (2)$$

where the subscripts (O, S) refer respectively to the observed and synthetic CMD, (c_i, m_i) respectively to the colour and magnitude for each point i in the CMD, $\sigma_i(c_{i,O}, m_{i,O})$ is the error ellipse around each point i , and $\delta m(c_i, m_i)$ is the (colour, magnitude) difference between the observed and the synthetic star. In addition, the reduced merit function F_χ is defined as:

$$F_\chi = \overline{\chi^2} = \sqrt{\chi^2(O, S)/N_{\text{match}}}, \quad (3)$$

where N_{match} refers to the number of points found within $3 - 5\sigma(c_i, m_i)$ of the error ellipse for each point between the observed and the synthetic CMD. Depending on the selection criteria $\overline{\chi^2}$ is defined to be smaller than or equal to 5. In general this is one of the functions that should be

¹ The method introduced in Sect. 2 actually mimics closely the procedure used in the ‘‘fit by eye’’ method.

minimized. Acceptable models are those with $F_\chi \lesssim 1$, i.e. models for which the difference in the (colour, magnitude) of the matched points between the observed and synthetic CMDs is on average less than 1σ .

2.2. The Poisson merit function

There are some points which do not have counterparts in the observed or the synthetic CMD, due to the limits imposed in the comparison. For a good fit the number of unmatched points, observed and simulated (respectively $N_{O,\text{not}}$ and $N_{S,\text{not}}$), should in the ideal case be smaller than the Poisson uncertainty for the total number of N_O observed and N_S synthetic points:

$$N_{O,\text{not}} + N_{S,\text{not}} \lesssim \sqrt{N_O} + \sqrt{N_S}, \quad (4)$$

or written as the Poisson merit function F_P

$$F_P = \frac{N_{O,\text{not}} + N_{S,\text{not}}}{\sqrt{N_O} + \sqrt{N_S}} \lesssim 1. \quad (5)$$

All the residual points can be placed in a CMD. This diagram contains indications about parameters that need to be optimized. In practice F_P will not be smaller than 1, due to simplifications adopted for the model CMD, to a not optimum representation of some evolutionary phases or even to limitations in the transformation from the theoretical to the observational plane. In Sect. 3 the diagnostics derived from the CMD filled with the residuals are explained in more detail. The CMDs filled with the residuals are hereafter also referred to as the diagnostic diagrams.

2.3. The global merit function

Both F_χ and F_P span a two-dimensional plane, see for example Fig. 2, which displays the merit for the various models (see Sect. 3). An acceptable model is obtained when both F_χ and F_P are about 1 or smaller. The best fit can therefore be obtained by minimizing the global merit function F , which is defined as

$$F = F_\chi^2 + F_P^2. \quad (6)$$

Both F_P and F_χ are in units of σ , say σ_P and σ_χ respectively. The difference between the observed and synthetic points is therefore on average about $\sqrt{F}\sigma$, where $\sigma = \sigma(F_P, F_\chi)$ is a function of the average chi-squared difference of the matched points combined with the Poisson uncertainty of the unmatched points. In general the minimization of the global merit function is mainly due to minimization of the Poisson merit function, i.e. the reduction of the number of unmatched points. An acceptable fit of the data is obtained when $F \simeq 2$ or smaller.

Note that this procedure is not limited to finding the best fit for a single-colour, magnitude diagram. It can easily be adjusted to fit multi-colour, magnitude diagrams.

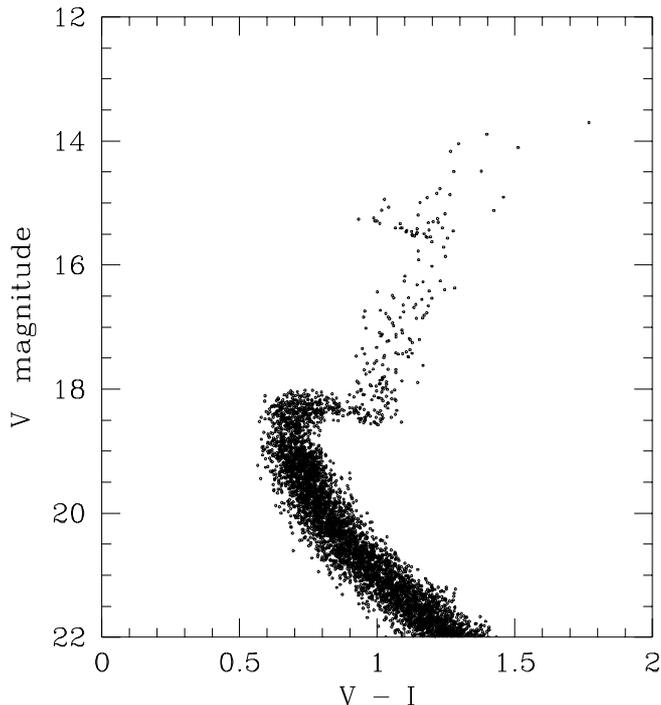


Fig. 1. The $(V, V-I)$ colour-magnitude diagram of the original stellar population ($Z = 0.005 - 0.030$, $t = 8 - 9$ Gyr) placed at 8 kpc, see text for additional details. Note that this simulation reflects a reddening free case and that the shape of the red horizontal branch is due to a large spread in metallicity

2.4. Speeding up

A comparison between the observed and the synthetic CMD on a point by (nearest) point basis can slow down the fitting procedure considerably, especially when a CMD consists of many data points. To speed up the whole procedure one can make a concession in accuracy by binning the N data points in M average, colour-magnitude boxes (\bar{c}_j, \bar{m}_j) , each with its own average error ellipse. Equation (1) can then be re-written to

$$\chi^2(O, S) = \sum_{j=1}^M k_j \left(\frac{(\bar{c}_{j,O} - \bar{c}_{j,S})^2}{\bar{\sigma}_j^2(\bar{c}_{j,O})} + \frac{(\bar{m}_{j,O} - \bar{m}_{j,S})^2}{\bar{\sigma}_j^2(\bar{m}_{j,O})} \right) \quad (7)$$

$$= \sum_{j=1}^M k_j \left(\frac{\delta \bar{m}(\bar{c}_j, \bar{m}_j)}{\bar{\sigma}_j(\bar{c}_{j,O}, \bar{m}_{j,O})} \right)^2, \quad (8)$$

where M is the resolution of the binned CMD, i.e. the number of colour-magnitude bins, and k_j is the number of points in each bin. This method has not been applied here, because the gain is small for the low number of objects used for the examples in Sect. 3.

2.5. Optimizing with a genetic algorithm

This paper does not deal with the actual implementation of an automated search program, which will be the subject of a forthcoming paper (Ng et al., in preparation).

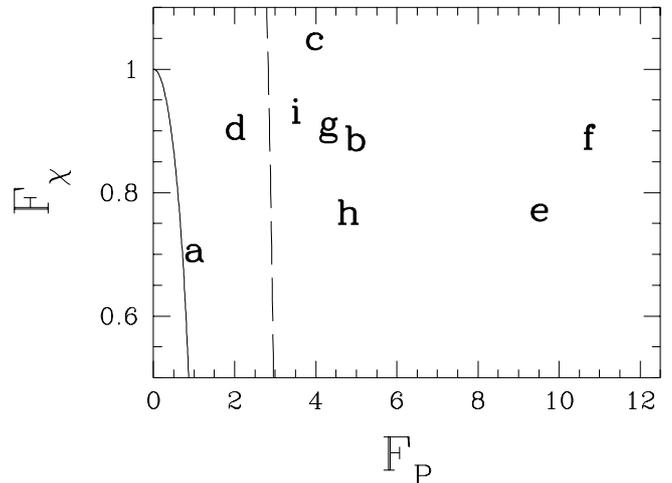


Fig. 2. The global merit for various simulations. The letters a–bf i refer to a model for which the result are given in Table 1. The solid and long dashed lines denote respectively the 1σ and 3σ area, where acceptable simulations for the stellar populations ought to be located. Model d could be acceptable. However, Fig. 3d hints that there is still a systematic discrepancy between the observations and simulations

The following part has been included for completeness as an example for a possible approach.

Genetic algorithms are a class of heuristic search techniques, capable of finding in a robust way the optimum setting for a problem (Charbonneau 1995; Charbonneau & Knapp 1996 and references cited therein). The optimum setting is searched with a so-called fitness parameter f , which ranges from zero (worst) to one (best). The fitness parameter f can be expressed as follows in terms of the global merit function:

$$f = \frac{1}{1 + F}. \quad (9)$$

Acceptable solutions yield $F \lesssim 2$ or $f \gtrsim \frac{1}{3}$. An estimate of the uncertainty of the input parameters can be obtained by doing multiple, time consuming simulations. However, another way to estimate the uncertainty is to vary for the fittest solution one parameter at a time. The fitness $f_{\sigma,k}$ for each parameter k is defined in such a way that the global merit F changes with $1\sigma(F_P, F_\chi)$ when this parameter is varied:

$$f_{\sigma,k} = \frac{1}{1 + |\sqrt{F_k} - \sqrt{F_{\min}} - 1|}, \quad (10)$$

where F_k is the merit for parameter k and F_{\min} is the merit obtained for the fittest population. This procedure corresponds in Fig. 2 with a jump in the (F_P, F_χ) -plane from the contour of the optimum solution to a contour displaced by exactly $1\sigma(F_P, F_\chi)$.

Note that the contour of the fittest population has the value $f_{\sigma,k} = \frac{1}{2}$ for $F_k = F_{\min}$, while the contour for the estimated uncertainty has the value $f_{\sigma,k} = 1$ for $\sqrt{F_k} = \sqrt{F_{\min}} + 1$.

Table 1. Description of the diagnostic statistics for the various models displayed in Figs. 2–4, see text for additional details

model	$N_{O,not}$	$N_{S,not}$	N_{match}	F_χ	F_P	F	f	comment
a	73	70	4927	0.70	1.01	1.52	0.398	no variation, see Sect. 3.1s
b	353	351	4647	0.89	4.98	25.6	0.038	distance, see Sect. 3.2
c	282	280	4718	1.04	3.97	16.9	0.056	extinction, see Sect. 3.3
d	143	143	4857	0.90	2.02	4.91	0.169	photometric errors, see Sect. 3.4
e	675	671	4325	0.77	9.52	91.2	0.011	crowding, see Sect. 3.5
f	760	759	4240	0.89	10.7	116.	0.009	age, see Sect. 3.6
g	306	306	4694	0.90	4.33	19.5	0.049	metallicity, see Sect. 3.7
h	340	340	4660	0.77	4.81	23.7	0.040	initial mass function, see Sect. 3.8
i	249	249	4751	0.93	3.52	13.3	0.070	star formation rate, see Sect. 3.9

3. Diagnostics

Some examples are given with a synthetic population to elucidate the method and its associated merit function as described in Sect. 2. It is further shown that the residual points of the data not matched can provide an indication about the parameter to be improved. A diagnostic diagram is much easier to interpret due to the relative small number of residual points than a CMD with the simulation of the best fit which may have 5000 or more points. It might be advantageous for other methods such as ratios, maximum likelihood, multivariate analysis or Bayesian inference to display the residuals from the best fit obtained. In this way one can obtain an independent visual impression of the performance of different methods and even obtain hints about possible improvements.

The test population used as reference in the simulations has the following specifications:

- a metallicity range, spanning $Z = 0.005 - 0.030$;
- an age range from 8 – 9 Gyr;
- an initial mass function with a Salpeter slope;
- an exponentially decreasing star formation rate with a characteristic time scale of 1 Gyr.

This population, displayed in Fig. 1, was found to contribute significantly to the CMD of Baade’s Window (Ng et al. 1996a). In the simulation the test population is placed at 8 kpc distance. The observational limits are $V_{lim} = 22^m$ and $I_{lim} = 21^m$. In each simulation $N = 5000$ stars are considered. For simplicity an extinction and crowding free case is considered with Gaussian distributed photometric errors amounting to $\sigma \lesssim 0^m05$ per passband. This is then compared with models, in which one of the specified parameters is varied. For the χ^2 -merit function a 3σ limit around each point (c_i, m_i) is adopted for the range of the error ellipse. These variations are then followed by comparison with different age, metallicity, initial mass function and star formation history.

Table 1 shows the results of the various simulations performed and discussed below, while Fig. 2 displays the global merit function of these simulations. Figures 3a–i display the diagnostics diagrams, resulting from the comparison between the “observed” and synthetic CMDs.

3.1. No variation

The first step is to demonstrate what values one obtains for F_χ , F_P and F with an acceptable model population. Such a population is obtained from a different realization of the test population by modifying the seed of the random number generator.

The values of the merit functions for this simulation (model a) are listed in Table 1 and are consistent with the expectation that for 5000 points about 71 points will not be matched in each of the observed and simulated dataset. Figure 2 shows further that this population has in the (F_P, F_χ) -plane an average σ close to 1. Figure 3a shows that the residual points (open circles for “observations” and crosses for simulations) are randomly distributed over the original (shaded) population. This is indicative that the model a simulation is an acceptable replacement for the original population.

Table 2 gives the formal 1σ uncertainties determined with Eq. (10). The estimated error in the distance gives an uncertainty in the distance modulus of about 0^m06 . This is a realistic value, because it is in close agreement with the $0^m07 - 0^m08$ uncertainty in the distance modulus obtained, for example, by Gratton et al. (1997) for globular clusters. The uncertainty in the extinction is about $dA_V \simeq 0^m06$, which is also in good agreement with the best estimates for the reddening of the globular clusters mentioned above. They yield $E(B - V) = 0^m02$, which is equivalent with an extinction of about 0^m06 with a standard reddening law. The 5 – 10% uncertainties in the age limits indicate that at a 1σ level the lower and upper age limit are likely not the same, but the whole population might on the other hand be almost indistinguishable from a population with a single age of about 8.7 Gyr. In addition, a 10% uncertainty is not very different from the 12% random errors estimated for the ages of globular clusters (Gratton et al. 1997).

The uncertainties in the metallicity range hint that an unambiguous determination of the presence of a metallicity range can be established. The estimated errors range from 0.1 – 0.2 dex, which is once more a realistic value if one

Table 2. Formal 1σ uncertainties determined with Eq. (10) for **model a** (see Table 1 and Sect. 3.1), where $[Z] = \log Z/Z_{\odot}$, α is the index of the power-law initial mass function and β is the index of the exponential star formation rate, specifying its characteristic time scale

parameter	value
$\log d$ (pc)	3.906 ± 0.012
A_V	$0^m00 \pm 0^m06$
$\log t_{\text{low}}$ (yr)	9.903 ± 0.043
$\log t_{\text{high}}$ (yr)	9.954 ± 0.023
$[Z]_{\text{low}}$	-0.60 ± 0.18
$[Z]_{\text{high}}$	$+0.18 \pm 0.08$
α	2.35 ± 0.03
β	1.0 ± 1.4

considers that the uncertainty in the metallicity scale is of the order of 0.1 dex. Most remarkable is the very small error obtained for the index of the slope of the power-law IMF. This is mainly due to the fact that small differences in the slope tend to result in large number of residuals of main sequence stars for the Poisson merit function. It strongly suggests that the method described in this paper potentially could reveal crucial information about the IMF. This is in high contrast with the estimated uncertainty for the index of an exponentially decreasing star formation rate. Due to the relative small age range the uncertainty is quite large. As a consequence, a population with a constant star formation rate cannot be discarded. On the other hand, Dolphin (1997) has demonstrated that in case of a considerably larger age range stronger constraints on the star formation rate could be obtained.

3.2. Distance

The distance of a stellar aggregate should be determined as accurately as possible, because uncertainties in the distance modulus can result in a different age for the stellar population, see Gratton et al. (1997) for a detailed discussion. For the next simulation (**model b**) the distance of the synthetic population is modified to 8.5 kpc. Figure 3b displays the diagnostic diagram with the residuals.

A fraction of the residuals, located on the main sequence or the sub-giant branch, are fainter than the original input population (shaded area). This feature provides a strong hint that the adopted population is not located at the proper distance. It results in a considerably larger value for the Poisson merit function. This leads to a rather large value for the global merit and gives an indication that the parameters for the matching stellar population are not yet optimally tuned. This furthermore appears in Fig. 2, which shows that the simulation for **model b** is not in the $1\sigma - 3\sigma$ range of acceptable solutions.

3.3. Extinction

A synthetic population (**model c**) is made with an average extinction of $A_V = 0^m10$, to demonstrate the effects of differences in the extinction. In addition, a random scatter is added to the extinction, amounting to 0^m02 . Figure 3c shows the corresponding diagnostic diagram with two residual sequences (original versus extinction modified), each located respectively near the blue and red edge of the original input population (shaded area). The sequences are also shifted from each other along the reddening line. A comparison between Figs. 3b and 3c further shows that there are distinct differences between the distribution of the residuals when the distance or extinction are not optimally tuned.

3.4. Photometric errors

In this simulation the effect of overestimating the photometric errors (**model d**) in the simulations is demonstrated. The photometric errors as a function of magnitude are increased by 50%. The results are displayed in Fig. 3d. The figure shows that the residuals for the modified population are mainly located outside the shaded region of the original population, while the residuals from the original population are almost centered on the shaded area.

Figure 2 indicates that model d might be an acceptable solution, but Fig. 3d hints that the parameters are not yet optimally tuned. Figure 3d further displays that the distribution of the residuals is significantly different from the residuals shown in any of the other panels of Fig. 3. It is therefore possible to identify an inadequate description of the photometric errors from the residuals and improve with this information the description of the simulated errors.

3.5. Crowding

In crowded fields single stars can lump together and thus form an apparent (not always resolved) binary on an image (Ng 1994; Ng et al. 1995; Aparicio et al. 1996). This results in the “disappearance” of some of the fainter stars. One can mimic this effect easily in the Monte-Carlo simulations of a synthetic CMD. One should avoid to include in these simulations the completeness factors as obtained from artificial star experiments. In crowding limited fields one obtains from the simulations of apparent binaries the completeness as a function of magnitude. This apparent, binary induced completeness function should be compared with the completeness factors obtained from artificial stars experiments.

The crowding simulations (**model e**) are performed under the assumption that 5% of the observed stars in the CMD are (apparent) binaries. Note that $\sim 10\%$ of the total number of simulated stars are involved in the crowding

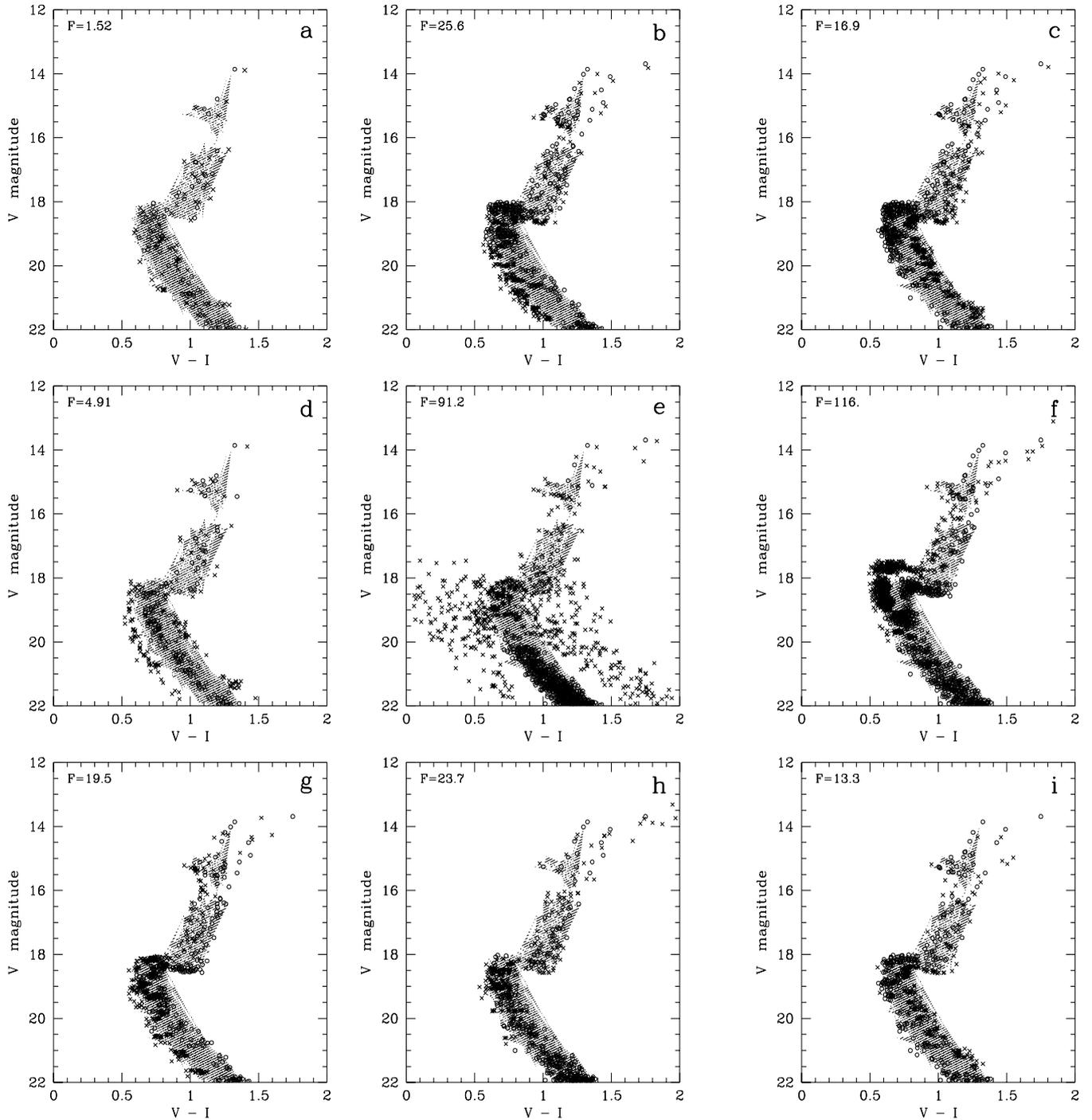


Fig. 3. Diagnostics diagrams, resulting from the various simulations given in Table 1. Frame **a**) shows the diagram for a different realization, while non of the major input parameters are varied; in frame **b**) the distance is varied; frame **c**) shows the diagram when some extinction is added; frame **d**) displays the residuals when different photometric errors are adopted; frame **e**) represents the case where the amount of crowding is overestimated; frame **f**) when a younger age is adopted for the stellar population; frame **g**) when the upper metallicity limit is underestimated; frame **h**) shows the effect for a different slope of the power-law initial mass function; and frame **i**) shows the case when the star formation rate is increasing towards younger age, instead of decreasing. The residual “observed” points which are not fitted are indicated by open circles, while the residual synthetic points are indicated by crosses. In each frame the shaded area shows approximately the part of the CMD covered by the original population. In addition the global merit F is reported in each frame

simulations. The actual number is slightly lower, because the simulations allow for the small possibility of the formation of multiple star lumps. The results are displayed in Fig. 3e. The diagnostics show, except for a significantly larger scatter, some similarity with those from Fig. 3d. However, the difference is that one part of the residuals from Fig. 3d forms a thinner sequence on top of the lower main sequence band, while the other part of the residuals is located close to the lower main sequence band. In the case of crowding the central band is broader and the remaining residuals are not located near the lower main sequence band.

3.6. Age

The galactic bulge might contain a rather young stellar population (Ng et al. 1995; Kiraga et al. 1997), however the results by Bertelli et al. (1995) and Ng et al. (1996) contradict such a suggestion. The results by Ng et al. (1995) might be induced by the limited metallicity range used in the simulations, while the results from Kiraga et al. (1997) do not appear to allow for a different interpretation. The next simulation (*model f*) is therefore made with a population which has an age in the range 4 – 5 Gyr and Fig. 3f displays the diagnostic diagram. At first sight this appears to be similar to the diagram of Fig. 3c, but there is a pronounced difference: the residual sequences are not parallel and furthermore one of the two sequences is slightly brighter and bluer than the main sequence turn-off of the original population.

In Sect. 3.1 it is noted that ages can be determined to an accuracy of at least 10%. An analysis of a deep bulge CMD with the method discussed in this paper might well constrain the actual age of the major stellar populations in the galactic bulge.

3.7. Metallicity

In the following simulation (*model g*) the upper metallicity limit of the synthetic population is decreased to $Z = 0.020$. Figure 3g displays the diagnostic CMD and shows two almost parallel sequences as in Fig. 3c, where the extinction is varied. The similarity is due to the fact that extinction and metallicity differences show a similar behaviour in $(V, V - I)$ CMDs. Photometry in other passbands should be explored to avoid this degeneracy. Ng & Bertelli (1996) demonstrate that near-infrared $(J, J - K)$ photometry would resolve unambiguously the degeneracy between extinction and age-metallicity.

3.8. Initial mass function

The sensitivity of the slope of a power-law initial mass function is demonstrated by changing the slope of the original population from $\alpha = 2.35$ to 1.35 (*model h*). Figure 3h

displays the corresponding diagnostic CMD. The residuals for the simulation with a shallower slope are dominating the upper part of the diagnostic diagram, while the residuals from the original population are concentrated near the faint detection limit. The large number of residuals give rise to an increase of the value obtained for the Poisson merit function. As mentioned in Sect. 3.1 this behaviour provides a strong constraint in the determination of the slope of the power-law IMF. Note in addition that the distribution of the residuals is quite different from the other panels in Fig. 3. One should realize however that no strong constraint for the power-law IMF can be obtained when main sequence stars below the turn-off are not available for analysis.

3.9. Star formation history

The final demonstration (*model i*) invokes a synthetic population generated for an increasing star formation rate with a characteristic time scale of 1 Gyr. The uncertainties given in Table 2 (see also Sect. 3.1) already suggest that a study of the star formation rate cannot be done reliably when a small age range is considered. However, the differences in the index of the exponential star formation rate are for this simulation suitably chosen, such that some differences will show up. The residuals of the simulation with an increasing star formation rate are displayed in Fig. 3i. The diagnostic CMD shows two parallel sequences comparable to the sequence in Figs. 3c and 3g. This partly provides an indication of the difficulties involved in studies of the star formation rate. It further shows once more that $(V, V - I)$ CMDs are not an optimal choice to study differences in star formation histories, because it will be difficult to distinguish differences in the extinction, metallicity, star formation history and even small age differences from one another. Additional photometry in other passbands ought to be used instead.

4. Discussion

4.1. The RGB and other caveats

The examples shown thus far are highly idealized cases. The colours of the stars from a synthetic populations are as reliable as the colours of the isochrones. Recent analysis of globular clusters (see Reid 1997 or Gratton et al. 1997 and references cited in those papers) indicate that a good fit for the main sequence stars does not necessarily imply a good agreement with the stars on the red giant branch. The discrepancy partly originates from the uncertainties in the colour transformations for the late type stars from the theoretical to the observational plane. Another cause is likely related to the mixing length parameter used in the calculations of the evolutionary tracks. One therefore

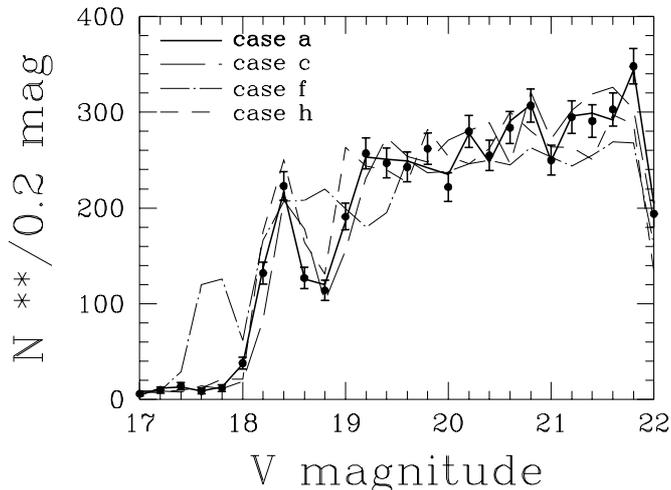


Fig. 4. Luminosity function for the cases a, c, f and h (Table 1). The open circles are obtained from the original input data set. The uncertainty per $0^m.2$ bin displayed are Poisson error bars

has to be cautious in the analysis of selected regions in the CMDs with old stellar populations.

A nice aspect of the diagnostic diagram is that the residuals clearly indicate how large the deviations are. In addition, uncertainties in the treatment of particular evolutionary phases might show up in the diagnostics. However, this can only be properly evaluated through a massive study, where one searches for systematic clumping of the residuals in particular regions of the CMDs. The identification of systematic clumps can then be used to improve the description of a particular evolutionary phase in the computation of new evolutionary tracks. In contrast to Dolphin (1997) it is argued that one should avoid the introduction of factors to reduce the weight of these particular phases in the fitting procedure.

All these caveats are however not related to the general validity of the method presented in Sect. 2. They will depend on the actual implementation of an automated fitting method and they will become important when a comparison with real data is made. However, a thorough discussion of problems associated with the implementation of an automated fitting method or a comparison with real data is beyond the scope of this paper.

4.2. Combining the diagnostics

It will be quite rare that one is going to deal with one of the idealized diagnostic diagrams from Figs. 3a-i. More likely the resulting diagnostic diagram is a combination of these diagrams, indicating that a number of parameters ought to be modified. One has to remain careful, because some effects might partly cancel each other out, like for example age and distance. A different distance for the stellar aggregate induces a change of the best-fitting age of the stellar population (Gratton et al. 1997). However, in a CMD the

distribution of the stars is not exactly the same for populations with a different age. The subtle differences might not cancel out through variation of the distance. The resulting diagnostic diagram might therefore indicate that yet another parameter ought to be optimized - such as the star formation rate - and in the end indicate that an acceptable fit has been obtained, while in reality one is dealing with an artifact. However, one of the major problems in $(V, V - I)$ CMDs remains the similarity in the behaviour of the extinction, small age differences, metallicity and the star formation rate. The results therefore might not always be as reliable as they are presented. A heuristic search for the optimum fit obtained with a genetic algorithm might properly disentangle the information for these parameters, but it would be more convenient to avoid this degeneracy and use photometry from additional passbands in which this degeneracy does not occur.

4.3. Unmatched evolutionary phases

In general, one will not always find for large amplitude variable stars a synthetic counterpart within the error ellipse. Those stars give rise to a small bias in the Poisson merit function. But, the total number of large amplitude, variable stars in any field is expected to be considerably smaller than \sqrt{N} . Therefore, one can ignore in first approximation any bias in the results due to variable stars.

Some fast evolutionary phases are not necessarily well described by theory or even not well covered by the small number of stars observed. This may lead to the presence of systematic clumps in the CMDs of the residuals from a massive study. The information obtained from these clumps can be used to compute new tracks and isochrones. In many cases however the number of stars present in these clumps is expected to be smaller than \sqrt{N} . It is therefore expected that unmatched evolutionary phases in general will not affect significantly the search for an optimum fit to the data.

As an aside, Gallart (1998) demonstrates that models are quite capable to predict subtle details in the observations, despite the fact that some evolutionary phases are not fully understood.

4.4. Galactic structure

In galactic structure studies the stars are distributed along the line of sight. The diagnostics procedure outlined here is also useful for these type of studies, in which the observed stars are a complex mixture of different stellar populations. Instead of applying a tedious scheme to deconvolve this mixture in its individual components, it is more liable to construct a synthetic mixture and compare this directly with the observations. The diagnostics will provide in the first place information about the galactic structure along

the line of sight. Once this has been established, one can explore in more detail the initial mass function and the star formation history of the different stellar populations. It should be possible to obtain some feedback for the input stellar library with an improved calibration of the galactic model, and to obtain on the long run in a self-consistent way indications about the adopted solar abundance partition or enrichment law $\Delta Y/\Delta Z$ (see Chiosi 1996 and references cited therein).

4.5. Open & globular clusters and the Local Group

In the studies of open clusters, metal-rich globular cluster and galaxies with resolved stellar populations from the Local Group a considerable amount of fore- and background stars can be present. It is not easy to take this contribution into account, because it is sometimes not clear if a particular feature is due to stellar evolution and intrinsic to the aggregate². One can clean in a statistical sense the galactic contribution from a neighbouring field. But this is only possible if the extinction and photometric errors are comparable. Ng et al. (1996c,d) used a galactic model to account for the contribution of the fore- and background stars. An unambiguous determination of the age was hampered by the large metallicity range and partly by the estimated amount of differential extinction. In this or other cases the diagnostics scheme as provided in this paper might contribute to a significantly deeper analysis.

4.6. The stellar luminosity function

Figure 4 displays the luminosity function of the original population, together with another realization of this population (**case a**), a population with a modification in the extinction (**case c**) and age (**case f**), and one for which the IMF slope was modified (**case h**). One can easily verify in Fig. 4 that the differences between the various populations for the majority of the magnitude bins are relatively small with respect to the generally adopted Poisson error bars. Only **case f** is significantly different, due to the large age difference adopted.

For a large number of bins (**models c** and **h**) the number of stars is not exactly within the 1σ uncertainty in case of Poisson errors, but they roughly are within 2σ . This is a first indication that one did not yet obtain an optimum solution. The diagnostic diagrams, Figs. 3c and 3h, indicate clearly that this is indeed the case. A comparison with **model a** further indicates that the uncertainty in

the number of stars in each bin is slightly over-estimated with Poisson error bars. This is mainly because the bins are not independent from one another. An acceptable solution - like **model a** - should go through almost every observed point in Fig. 4. About $\sqrt{N_{\text{bins}}}$ are expected to deviate from this expectation. In case of **model a** about 5 points might not show a close match in Fig. 4. A close inspection shows that this is indeed the case.

The results indicate that the luminosity functions of various realizations - such as **cases a**, **c**, and **h** - might apparently not be so significantly different from one another and might therefore all be acceptable solutions for the test population. The values for the merit function and the diagnostic diagrams (Figs. 3a, 3c and 3h) however strongly indicate that only **model a** is acceptable. The method presented in Sect. 2, together with the diagnostic diagram, is more powerful than an analysis in which different model luminosity functions are compared. The global merit function, together with the diagnostic diagrams, gives a better discrimination between different models. The crucial point lies in the application of the Poisson statistics. As outlined in Sect. 2, the only independent quantity to which the Poisson statistics can be applied is the total number of stars brighter than a specific limiting magnitude³. Suffice to mention that Table 1 and Figs. 3a, 3c, and 3h show that significant differences are present between **cases c** and **h** with respect to **case a**.

4.7. Future work

Firstly, an automatic procedure should be developed based on the merit functions and the diagnostic diagrams. A search with a genetic algorithm appears to be a promising approach. As a first test one should apply this program to a synthetic dataset, such as the test population used in this paper. The use of real datasets should be avoided initially, because unforeseen problems - which are not associated with the validity of the method - might arise with real data sets. In particular, problems related with relative fast phases of stellar evolution or the colour transformation from the theoretical to the observational plane, see for example Sects. 4.1 and 4.3.

Secondly, a comparison should be made between the results from an automated search program and the results obtained from the isochrone fitting technique. A study of old open clusters in which these techniques are applied

² See for example the suspected intervening stellar population to the LMC suggested by Zaritsky & Lin (1997) from a feature in the CMD. This interpretation is questioned and Beaulieu & Sackett (1998) argue that it is actually intrinsic to the LMC. Note further that another feature identified as the RGB-bump is most likely the AGB-bump (Gallart 1998).

³ For example: the number counts of field SSA 22 (see Fig. 2a from Reid et al. 1997) is within the Poisson error bars in good agreement with the model predictions; their Fig. 2b however shows that too many stars are predicted by their model, i.e. 61 predicted versus 31 observed for $I_{\text{lim}} = 21^{\text{m}}5$; this gives $F_P > 5$, indicating that the model does not have an optimal parameter setting. It is however beyond the objective of this paper to determine which parameter(s) in their model has the culprit.

is underway (Carraro et al., in preparation). The purpose is to determine if the age of the oldest open cluster Berkeley 17 (Phelps 1997) is as old as the globular clusters or if it has an age comparable to or slightly older than the old clusters in the sample defined by Carraro et al. (1998). The next step is to apply this method to the resolved, multiple stellar populations of dwarf galaxies. However, with respect to the clusters the results might not be as reliable.

Finally, it is intended to improve through (self-) calibration from studies of essentially single stellar populations, like open & globular clusters, the library of stellar evolutionary tracks and isochrones.

5. Conclusion

A method based on the χ^2 merit function is presented to compare “observed” with synthetic single stellar populations. Monte-Carlo simulations have been performed to display the diagnostic power from a CMD containing the points for which no corresponding synthetic point was found within a reasonable error ellipse. The simulations indicate that the CMDs of residual points might provide hints about model parameters to be improved. The simulations further indicate that one ought to be cautious with the analysis of stellar luminosity functions and that strong hints can be obtained about the shape of the initial mass function.

Acknowledgements. I. Saviane and L. Portinari are thanked for reading of the manuscript. G. Bertelli and C. Chiosi are thanked for their stimulating discussions. In particular, G. Bertelli is acknowledged for making available the synthetic stellar population generator (HRD-ZVAR), which produced the stellar populations used for this paper. The research of Ng is supported by TMR grant ERBFMRX-CT96-0086 from the European Community.

References

- Aparicio A., Gallart C., 1995, *AJ* 110, 2105
Aparicio A., Gallart C., 1996, in Proceedings “From Stars to Galaxies”, 9–13 October, 1995, Elounda (Crete, Greece), Leitherer C., von Alvensleben U.F. and Huchra J. (eds.), ASP Conf. Ser. 98, 31
Aparicio A., Bertelli G., Chiosi C., Garcia-Pelayo J.M., 1990, *A&A* 240, 262
Aparicio A., Gallart C., Chiosi C., Bertelli G., 1996, *ApJ* 469, L97
Aparicio A., Gallart C., Bertelli G., 1997a, *AJ* 114, 669
Aparicio A., Gallart C., Bertelli G., 1997b, *AJ* 114, 680
Beaulieu J.-P., Sackett P.D., 1998, *AJ* (in press) (astro-ph/9710156)
Bertelli G., Mateo M., Chiosi C., Bressan A., 1992, *ApJ* 388, 400
Bertelli G., Bressan A., Chiosi C., Ng Y.K., Ortolani S., 1994, *Mem. S. A. It.* 65, 689
Bertelli G., Bressan A., Chiosi C., Ng Y.K., Ortolani S., 1995, *A&A* 301, 381
Bertelli G., Bressan A., Chiosi C., Ng Y.K., 1996, *A&A* 310, 115
Bienaymé O., Robin A.C., Crézé M., 1987, *A&A* 180, 94
Carraro G., Bertelli G., Bressan A., Chiosi C., 1993, *A&AS* 101, 381
Carraro G., Chiosi C., Bressan A., Bertelli G., 1994, *A&AS* 103, 375
Carraro G., Ng Y.K., Portinari L., 1998, *MNRAS* 296, 1045
Carraro G., Girardi L., Ng Y.K., 1998, *MNRAS* (in preparation)
Charbonneau P., 1995, *ApJS* 101, 309
Charbonneau P., Knapp B., 1996, “A User’s Guide to PIKAIA 1.0”, NCAR Technical Note 418+IA (Boulder: National Center for Atmospheric Research)
Chen B., 1996a, *A&A* 306, 733
Chen B., 1996b, *A&AS* 118, 181
Chiosi C., 1996, in Proceedings “From Stars to Galaxies”, 9–13 October, 1995, Elounda (Crete, Greece), Leitherer C., von Alvensleben U.F. and Huchra J. (eds.), ASP Conf. Ser. 98, 181
Chiosi C., Bertelli G., Meylan G., Ortolani S., 1989, *A&A* 219, 167
Dolphin A., 1997, *New Astron.* 2, 397
Ferraro F.R., Fusi Pecci F., Tosi M., Buonanno R., 1989, *MNRAS* 241, 433
Gallart C., 1998, *ApJ* 495, L43
Gallart C., Aparicio A., Bertelli G., 1996a, in Proceedings “From Stars to Galaxies”, 9–13 October, 1995, Elounda (Crete, Greece), Leitherer C., von Alvensleben U.F. and Huchra J. (eds.), ASP Conf. Ser. 98, 339
Gallart C., Aparicio A., Bertelli G., Chiosi C., 1996b, *AJ* 112, 1950
Gallart C., Aparicio A., Bertelli G., Chiosi C., 1996c, *AJ* 112, 2596
Gozzoli E., Tosi M., Marconi G., Bragaglia A., 1996, *MNRAS* 283, 66
Gratton R.G., Fusi Pecci F., Carretta E., et al., 1997, *ApJ* 491, 749
Han M., Hoessel J.G., Gallagher J.S. III, Holtzman J., Stetson P.B., 1997, *AJ* 113, 1001
Mould J.R., Han M., Stetson P.B., et al., 1997, *ApJ* 483, L41
Kiraga M., Paczyński B., Stanek K.Z., 1997, *ApJ* 485, 611
Ng Y.K., 1994, Ph.D. thesis, Leiden University, The Netherlands
Ng Y.K., 1997a, in Proceedings “The impact of large-scale near-IR sky surveys”, 24–26 April, 1996, Puerto de la Cruz (Tenerife, Spain), Garzon-Lopez et al. (eds.), p. 91
Ng Y.K., 1997b, in Proceedings 12th IAP Colloquium “Variable stars and the astrophysical returns from microlensing surveys”, 8–12 July, 1996, Paris (France), Ferlet R., Maillard J.-P. and Raban B. (eds.), p. 113
Ng Y.K., Bertelli G., 1996a, *A&A* 315, 116
Ng Y.K., Bertelli G., 1996b, in Proceedings “From Stars to Galaxies”, 9–13 October, 1995, Elounda (Crete, Greece), Leitherer C., von Alvensleben U.F. and Huchra J. (eds.), ASP Conf. Ser. 98, 547
Ng Y.K., Bertelli G., Bressan A., Chiosi C., Lub J., 1995, *A&A* 295, 655 (erratum *A&A* 301, 318)
Ng Y.K., Bertelli G., Chiosi C., Bressan A., 1996a, *A&A* 310, 771

- Ng Y.K., Bertelli G., Chiosi C., Bressan A., 1996b, in Proceedings ESO/MPA workshop on "Spiral Galaxies in the Near-IR", 7–9 June 1995, Garching bei München (Germany), Minniti D. and Rix H.-W. (eds.), p. 110
- Ng Y.K., Bertelli G., Chiosi C., 1996c, A&A (submitted)
- Ng Y.K., Bertelli G., Chiosi C., 1996d, A&A (submitted)
- Ng Y.K., Bertelli G., Chiosi C., Bressan A., 1997, A&A 324, 65
- Ng Y.K., Chiosi C., Bertelli G., Bressan A., 1998, A&A (in preparation)
- Phelps R.L., 1997, ApJ 483, 826
- Reid I.N., 1997, AJ 114, 161
- Reid I.N., Gizis J.E., Cohen J.G., et al., 1997, PASP 109, 559
- Robin A.C., Haywood M., Cr ez e M., Ojha D.K., Bienaym e O., 1996, A&A 305, 125
- Santos J.F.C. Jr., Frogel J.A., 1997, ApJ 479, 764
- Tolstoy E., 1995, Ph.D. thesis, Groningen University, The Netherlands
- Tolstoy E., 1996, ApJ 462, 684
- Tolstoy E., Saha A., 1996, ApJ 462, 672
- Tosi M., Greggio L., Marconi G., Focardi P., 1991, AJ 102, 951
- Vallenari A., Chiosi C., Bertelli G., Meylan G., Ortolani S., 1990, A&AS 87, 517
- Vallenari A., Chiosi C., Bertelli G., Meylan G., Ortolani S., 1992, AJ 104, 1100
- Vallenari A., Aparicio A., Fagotto F., Chiosi C., 1994a, A&A 284, 424
- Vallenari A., Aparicio A., Fagotto F., et al., 1994b, A&A 284, 447
- Vallenari A., Chiosi C., Bertelli G., Ortolani S., 1996a, A&A 309, 358
- Vallenari A., Chiosi C., Bertelli G., Aparicio A., Ortolani S., 1996b, A&A 309, 367
- Zaritsky D., Lin D.N.C., 1997, AJ 114, 2545