

Using the Agile Software Development Lifecycle to develop a standalone application for generating Colour Magnitude Diagrams

K. Fitzgerald^{a,b,1}, L-M. Browne^{a,2}, R.F. Butler^{a,3,*}

^aCentre for Astronomy, School of Physics, National University of Ireland - Galway, University Road, Galway, H91 CF50, Ireland

^bPermanent address: Athlone Institute of Technology, Dublin Road, Bunnally, Athlone, Co. Westmeath, N37 HD68, Ireland

^cDeclarations of Interest: none for any author

Abstract

Virtual observatories allow the means by which an astronomer is able to discover, access, and process data seamlessly, regardless of its physical location. However, steep learning curves are often required to become proficient in the software employed to access, analyse and visualise this trove of data. It would be desirable, for both research and educational purposes, to have applications which allow users to visualise data at the click of a button. Therefore, we have developed a standalone application (written in Python) for plotting photometric Colour Magnitude Diagrams (CMDs) - one of the most widely used tools for studying and teaching about astronomical populations. The *CMD Plot Tool* application functions “out of the box” without the need for the user to install code interpreters, additional libraries and modules, or to modify system paths; and it is available on multiple platforms. Interacting via a graphical user interface (GUI), users can quickly and easily generate high quality plots, annotated and labelled as desired, from various data sources. This paper describes how *CMD Plot Tool* was developed using Object Orientated Programming and a formal software design lifecycle (SDLC). We highlight the need for the astronomical software development culture to identify appropriate programming paradigms and SDLCs. We outline the functionality and uses of *CMD Plot Tool*, with examples of star cluster photometry. All results plots were created using *CMD Plot Tool* on data readily available from various online virtual observatories, or acquired from observations and reduced with IRAF/PyRAF.

Keywords: Agile software development; Object oriented development; Scientific visualisation; Hertzsprung-Russell diagram; Globular clusters; General

1. Introduction

In recent years, online archives and virtual observatories or VOs (Hanisch et al., 2015) have increased the availability and quantity of digitised, multi-wavelength astronomical data accessible to research scientists, academics, students and hobbyist astronomers. As the data access problem diminishes, more focus is now required on effectively and efficiently analysing, interpreting and visualising results from the data. To one major institution active in the VO sphere, the European Southern Observatory, the term “Virtual Observatory Tools” only means tools to query and retrieve raw or processed data from its archives⁴; not to manipulate or visualise it. Popular examples of existing tools which do address some forms of manipulation and visualisation include SAO ds9 (available as a desktop application), which is restricted to pixel data and has limited options for overplotting of markers and regions; CDS Aladin (available in both full-functionality desktop and “Lite”-functionality HTML5 browser versions), which can perform many operations on VO images, catalogs and overlays - but lacks the ability

to plot an “x vs y” graph such as a CMD; and TOPCAT and VOPlot, both of which operate on VO catalog (tabular) data. EURO-VO maintains an online list⁵ of these and other tools for VO data. But what this list makes clear is that no one piece of user-friendly software is readily available to manage, analyse and visualise this diversity of data. As a result of this, a growing number of researchers, students and hobbyists are developing their own tools for their own specific purposes.

Furthermore, new users also experience steep learning curves in installing, understanding and utilising specialist software. One example is where a user must create a colour magnitude diagram (CMD) to visualise and interpret photometric data. A CMD is an observational representation of the Hertzsprung-Russell diagram (Russell, 1913), not in physical units (absolute luminosity and temperature) but in observational units (stellar magnitudes and magnitude differences, in particular filter band-passes). The flowchart shown in Figure 1 outlines the steps involved in generating a CMD using IRAF⁶ (ascl:9911.002) (Tody, 1993) or PyRAF⁷ (ascl:1207.011) (Greenfield and White, 2000). The sophistication, flexibility and breadth of

*Corresponding author

¹Email: Mr. Karol Fitzgerald - kfitzgerald@ait.ie

²Email: Ms. Lisa-Marie Bowne - l.browne1@nuigalway.ie

³Email: Dr. Ray Butler - ray.butler@nuigalway.ie

⁴<https://archive.eso.org/cms/virtual-observatory-tools.html>

⁵<http://www.euro-vo.org/?q=science/software>

⁶IRAF: Image Reduction and Analysis Facility.

⁷PyRAF is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

IRAF have made it one of the “industry standard” platforms in observational astronomy research - but is it necessary or reasonable to expect mastery in IRAF (or its ilk) as a pre-requisite for CMD generation? Even tools like VOPlot and TOPCAT, which take photometric catalog data as their starting point, may intimidate some users: VOPlot has a 64-page online user manual; TOPCAT’s PDF manual runs to 527 pages! This paper introduces *CMD Plot Tool*: a standalone application designed to read various sources of data and to easily produce consistent and high quality plots of CMDs - allowing novices to bypass the usual learning curve, and experts to get results more readily. The executables and source code are publically available on Sourceforge at <https://cmd-plot-tool.sourceforge.io>

The paper outline is as follows: Section 2 discusses the development of specialised software in astronomy and recent research highlighting its importance and current trends. The software development lifecycle (SDLC) used in developing *CMD Plot Tool*, and the process of software freezing to create an “out of the box” application, are outlined. These techniques have wider utility for other application developers. This section concludes by documenting how this application can be deployed on various platforms without the need for additional development environments, code interpreters, or users having to modify system paths. Section 3 describes the functionality of the *CMD Plot Tool*, how the various data sources are integrated, and the operations needed to produce CMDs. Section 4 illustrates how CMDs can be used to interpret, visualise and analyse celestial populations, with focus on sample datasets to produce CMDs of globular clusters (GCs). Section 5 presents our conclusions.

2. Software Development in Astronomy

A survey carried out by [Momcheva and Tollerud \(2015\)](#) focused on the use of software and the software skills of the participants in the worldwide astronomy community, between December 2014 and February 2015. Participants consisted of 1142 astronomers: 380 graduate students, 340 postdocs, 385 research scientists and faculty, and the remaining 37 consisted of undergraduate students, observatory scientists, etc. All survey participants responded “yes” when asked if they used software as part of their research; 11% of responders said they used software developed by others; 57% used software developed by themselves and others; while 33% said that they used software they developed themselves for specific purposes, as there was no software readily available. This research also revealed the open source language, Python, to be the programming language of choice. However, there was no reference in this research to the software development life cycle (SDLC) employed by developers in the astronomy domain.

Prior to our development of *CMD Plot Tool*, we investigated which (if any) SDLCs are utilised in the development of astronomy-specific software. Examples of promising software publications we checked include [Collins et al. \(2017\)](#), [Konstantopoulos \(2015\)](#), [Sybilski et al. \(2014\)](#), [VanderPlas et al. \(2012\)](#), [Economou et al. \(2014\)](#), [Allen and Schmidt \(2015\)](#), [Bray \(2014\)](#), [Davies et al. \(2013\)](#), [Kent \(2013\)](#), [Goodman](#)

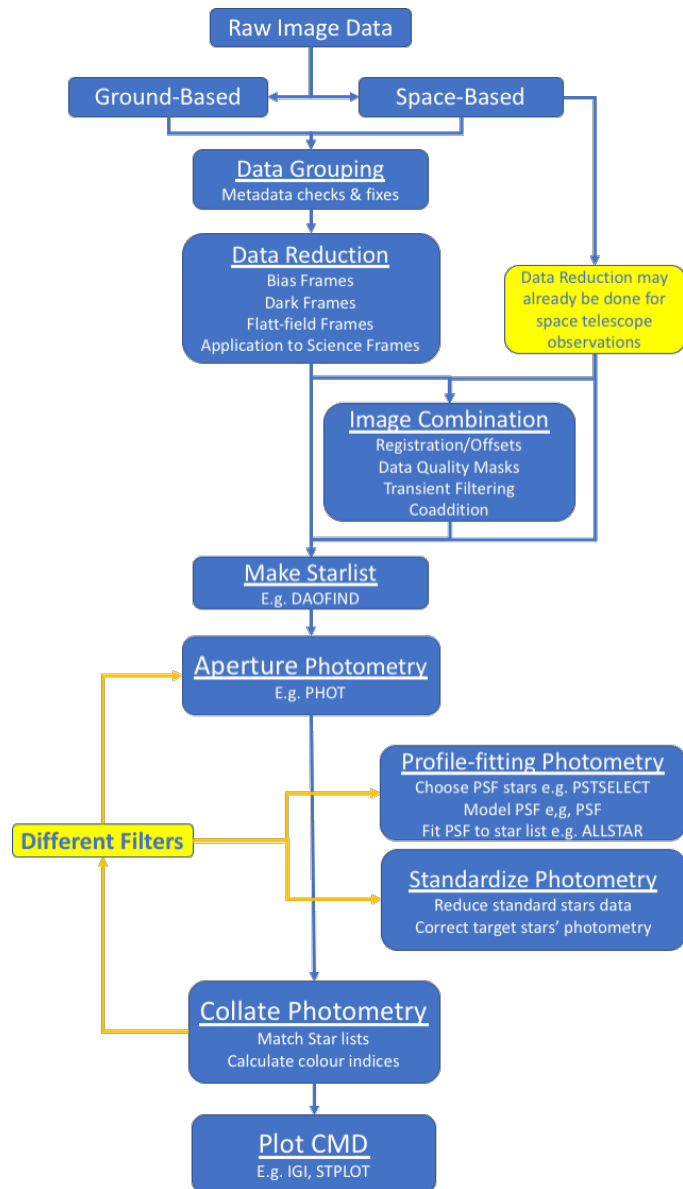


Figure 1: CMD Generation Flowchart. This illustrates the steps and learning curve required to generate a CMD of stellar objects, starting with raw image data. Task names in IRAF/PyRAF/DAOPHOT, which would achieve the photometry steps, are given for illustration.

(2012), and [McMullin et al. \(2007\)](#). We found no information regarding the process (programming paradigms) by which the published astronomy-specific software was written, i.e. the developers do not state whether they employed any particular SDLC or test driven development (TDD: [Beck \(2002\)](#); [Astels \(2003\)](#)) approach. In order to preclude our own selection bias, we then checked every paper in the most recent year’s volume of freely available Astronomical Data Analysis Software and Systems (ADASS) conference e-proceedings ([Taylor and Rosolowsky, 2015](#)) - choosing ADASS because of its pre-eminence as a forum for communicating best practice in this field. Not all of these 128 ADASS papers involve writing code - in fact, over half are concerned with other topics (databases, data standards and models, metadata and archive

management, VO interfacing, creative use of archival data, collaborative workspaces, cloud computing, source code libraries). However, our trawl revealed that in nearly all cases where a code or pipeline was developed (typically using Python, Java, C or C++; sometimes in conjunction with MPI or Perl), it is described in terms of its algorithms and functionality, layers and internal architecture, external connectivity, inputs and outputs; but not in terms of its development approach. The few exceptions to this rule were Ballester et al. (2015), Vallejo et al. (2015), Surace et al. (2015) and Mulumba et al. (2015). The first of these papers outlines a process of iterating with users on their requirements, but does not explicitly identify this approach by its name. The other three, in contrast, give comprehensive treatment to their Agile and User-Centred design methodologies. This investigation indicates that only the order of 5% of astronomical coding papers provide such information, and presumably this reflects the proportion of projects which take a SDLC approach.

The advantages of employing and adhering to a SDLC include: clarity in project objectives, requirements and estimates; more stable systems where missing functionality can be easily identified; developing a valuable relationship between users and developers. A lack of awareness of these benefits is an unfortunate aspect of the astronomical software development culture, and if unchecked can potentially lead to software/systems being over budget, delivered late, or missing functionality; and in worst cases, to complete project failure. So in an attempt to steer this culture, we will describe programming paradigms and the SDLC/TDD approach, including the advantages and disadvantages of various SDLCs.

2.1. Software Development Life Cycle (SDLC) and Programming Paradigms

A SDLC defines a structured sequence of stages in software engineering to develop the intended software product. A TDD approach relies on a shorter development cycle, where requirements become specific test cases that the software must pass. Another factor to consider is determining which programming paradigm to use. In developing *CMD Plot Tool*, much emphasis was put on determining an appropriate SDLC/TDD approach and programming paradigm, while also setting out a strict set of requirements beforehand. This section will discuss the options available and justify the choices made.

Programming Paradigms. Different paradigms allow for alternative approaches when developing software applications. However, it is important to note that while programming languages are usually classified by one paradigm, there are some languages - such as Python - that can handle multiple paradigms. The two most common programming paradigms are the *procedural paradigm* and the *object-oriented paradigm*. Procedural programming relies on the premise that the coder utilises procedures (routines or subroutines) to operate and manipulate data. This type of programming is sequential in nature, and so not particularly complicated. Object-orientated programming (OOP) amalgamates procedures and data into *objects*, allowing for more complicated functionality, while min-

imising the amount of code required. Objects can either be independent or associated with other objects, and they interact by passing information to each other. In the instance where an object interacts with another object, regardless of their similarity or differences, then the object contains information about itself (*encapsulation*) and the objects it can interact with (*inheritance*). Inheritance enables new objects to inherit the properties and methods of existing objects. OOP also utilises *classes*, which are user-defined prototypes for an object. They define a set of attributes that characterise any object of the class. Classes allow for the generation of multiple objects of the same type that can be used anywhere in the code, allowing for a significant reduction in coding but also more flexibility and functionality when dealing with multiple objects of the same type. Inheritance may be exploited for further coding efficiencies: rather than creating a new object from scratch, developers can reference a pre-existing object or superclass and create a subclass based on the previous one, allowing them to reuse code and functionality more effectively.

Using OOP, developers can manage and break software projects down into smaller, more manageable modular problems, one object at a time. The modularity of objects makes trouble-shooting easier; encapsulation ensures that objects are self-contained and functionality contained within methods/functions specific to that class. When errors do occur, developers know where to look without having to navigate through large amounts of code. Another advantage of OOP is in software maintenance; applications may evolve by additional functionality or improvements to the user experience, and upgrades might be required to allow software to work on newer computer systems. While an OOP-based application requires a great deal of planning pre-release, less work is needed to maintain it over time. Given these advantages, *CMD Plot Tool* was developed using OOP, written in Python.

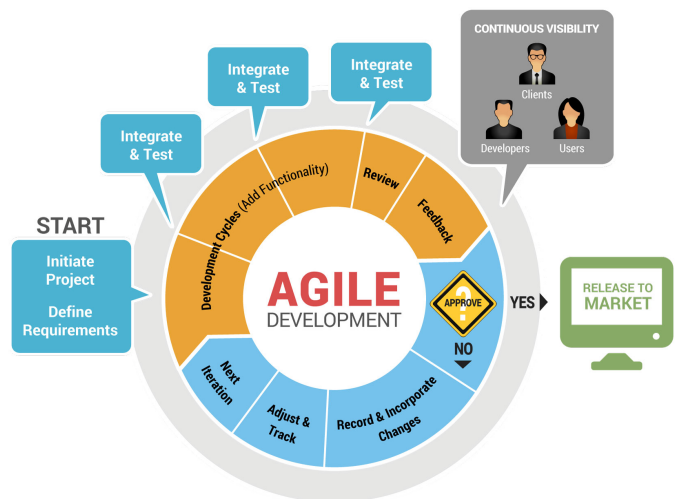


Figure 2: The Agile Software Development Lifecycle. Image reproduced from <http://cssmith.co/wp-content/uploads/2017/10/scrum-diagram-agile-application-and-cloud-computing.jpg>.

Agile Software Development. The Agile Software Development Lifecycle (ASDL) is a set of methodologies (four values and twelve principles) defined by the Agile Manifesto (Beck et al., 2001). It values:

1. **Individuals and interactions** over processes and tools
2. **Working software** over comprehensive documentation
3. **Customer collaboration** over contract negotiation
4. **Responding to change** over following a plan

The twelve principles expand on these values in more detail, giving more emphasis to the items highlighted in bold. The ASDL is illustrated in Figure 2. It combines iterative and incremental process models, and focuses on adaptability and satisfaction by rapid delivery of working software products before the product is reviewed. During each development lifecycle, additional functionality is incorporated and tested. If the product meets the specification and satisfaction of the project owner, then the product is released to the market or end users; if not, then another iteration of the development phase begins after all incremental changes have been noted. Each development phase focuses on particular functionality.

A comparison of the ASDL and traditional software development life cycles is presented in Table 1 (Leau et al., 2012). This explains our main reasoning for choosing the ASDL over traditional SDL: the project (1) was relatively low scale, (2) had high customer involvement, and (3) was characterised by a changeable development direction and evolving user requirements. Table 2 (Sami, 2012) outlines the rubric used in deciding on the ASDL over the alternatives for this project; ASDL again scores highest or joint-highest in factors such as user requirement and “visibility of stakeholders” (another name for customer involvement), but low cost and short time schedule were also important in a busy, non-commercial, academic development team.

Our utilisation of the ASDL to develop of *CMD Plot Tool* proved extremely beneficial. *User stories* - concise, written descriptions of specific functionality that is valued to the user or owner of the software - were generated by two co-authors (R.B., L.-M.B.) prior to the commencement of coding by K.F. A description of a user story follows the following template (Rehkopf, 2017):

As a [user role], I want to [goal], so i can [reason]

For example: *As an astronomer, I want to be able to select two magnitude files output by IRAF/DAOPHOT; calculate the colour index for each star; format the axis, title, plotting colour; and add annotation; so I can generate a CMD plot for the classroom or for publication.*

User stories must be detailed enough to start work; however, further details can be established and clarified as the project progresses. The relatively small scale of this project benefited the authors by allowing them to discuss the overall goal and functionality of the software prior to and during development. Regular communication and continuous inputs from the customers (in this instance the co-authors) left little room for guesswork regarding functionality. Requirements specified by

the customers were developed into 13 user stories (often with some elements in common).

These user stories focused the development cycle for each iteration of *CMD Plot Tool* and saved time: by allowing the developer to concentrate on breaking requirements down into functionality, this breaks a large project into smaller, more manageable tasks (in total 14 iterations were required, with 1-2 weeks allocated for each depending on the complexity of the functionality). For example, the above user story is accomplished by the following tasks:

1. Read in DAOPHOT-format photometry file for each filter/waveband
2. Extract necessary fields from each file.
3. Merge fields from each file, removing errors & duplicates.
4. Calculate the colour indices from pair(s) of magnitudes, and calculate their errors.
5. Retrieve user specified parameters regarding axis scale, annotation, and colour options from the GUI.
6. Generate and save the plot.
7. Display the plot.
8. Delete temporary files.

Another benefit of having well defined user stories is the ability to anticipate the functionality of each software class and how they interact with each other. Upon completion, *CMD Plot Tool* consisted of 6 classes, illustrated in the UML diagram of Figure 3, with approximately 1000 lines of code; an analysis of each class is given in Table 3.

Software Testing. Software testing is essential in evaluating the quality, performance and accuracy of code by detecting differences between actual and expected outputs, for given inputs. The ASDL (Figure 2) recommends that this process be carried out during each development phase, to ensure *verification* and *validation* of the product. Verification ensures that the product satisfies the conditions imposed at the start of the development phase, to confirm that the application behaves as it should. Validation ensures the application satisfies the requirements at the end of the development phase to certify that it is built as per the customer requirements. There are numerous types of software testing, which can be categorised as being either functional, non-functional or maintenance. TDD, unit testing and GUI testing were considered as possible testing techniques for testing the *CMD Plot Tool*.

TDD is a technique developed by Kent Beck (Beck, 2002), (Figure 4) for constructing software. It follows three simple steps repeatedly:

- Construct a test for the next functionality to be coded
- Write functional code until the test is passed
- Refactor new and old code, making it well structured

This particular software development technique relies on the repetition of a very short development cycle where requirements are turned into very specific test cases, and software is then improved to pass only the new tests. However, implementing a TTD approach in this project proved difficult due to lack

	Agile	Traditional
User requirement	Iterative acquisition	Detailed user requirements are well defined before coding
Development direction	Readily changeable	Fixed
Testing	On every iteration	After coding phase completed
Customer involvement	High	Low
Extra quality required for developers	Interpersonal skills & basic business knowledge	Nothing in particular
Suitable project scale	Low to medium scaled	Large-scaled

Table 1: Agile SDLC versus Traditional SDLC models (Leau et al., 2012)

Factors	Waterfall	V-Shaped	Evolutionary Prototyping	Spiral	Iterative & Incremental	Agile
Unclear User Requirement	Poor	Poor	Good	Excellent	Good	Excellent
Unfamiliar Technology	Poor	Poor	Excellent	Excellent	Good	Poor
Complex System	Good	Good	Excellent	Excellent	Good	Poor
Reliable system	Good	Good	Poor	Excellent	Good	Good
Short Time Schedule	Poor	Poor	Good	Poor	Excellent	Excellent
Strong Project Management	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
Cost limitation	Poor	Poor	Poor	Poor	Excellent	Excellent
Visibility of Stakeholders	Good	Good	Excellent	Excellent	Good	Excellent
Skills limitation	Good	Good	Poor	Poor	Good	Poor
Documentation	Excellent	Excellent	Good	Good	Excellent	Poor
Component reusability	Excellent	Excellent	Poor	Poor	Excellent	Poor

Table 2: Rubric used in deciding on the SDL for this project (Sami, 2012)

of experience in designing test cases prior to coding. As a consequence of this steep learning curve, it was decided to use unit testing.

The *CMD Plot Tool* was developed and tested using the PyCharm IDE (integrated development environment) 2017 Community Edition, Python 2.7 with later iterations using Python 3.7. The free community edition of PyCharm offers usage of both testing frameworks⁸ and code analysis tools⁹. Code inspections detect - and suggest corrections for - compiling errors, code inefficiencies including unreachable code, unused code, non-localised string, unresolved method, memory leaks and even spelling mistakes. We found the code inspection feature extremely useful in identifying and eliminating potential errors prior to running any unit testing. In addition to this automated testing, the co-authors behind the user stories used their expertise in astronomical photometry to test the GUI, based on a range of inputs and expected outputs - sampling the 3 types of input file format (see 3.1), and input catalogs of sizes which varied up to approximately 10^5 stars. Testing covered all of the functionality provided in the GUI (see 3), including error bars, plot annotation, axis specification, and side-by-side plotting. Figures 6, 7, 8, and 9 illustrate some of these features. Feedback was provided to co-author K.F., who then made the appropriate changes, and the code was tested again. This process was done at the end of each iteration.

⁸PyCharm Testing Frameworks: <https://www.jetbrains.com/help/pycharm/testing-frameworks.html>

⁹PyCharm Editions Comparison: https://www.jetbrains.com/pycharm/features/editions_comparison_matrix.html

2.2. Freezing Python Code

Installation of astronomical software packages can be non-trivial, and a source of frustration. It is often the case that the end-user installs programming interpreters to execute code, only then to be informed that the software package requires the presence of another to work correctly, and that system paths must be set for referencing additional libraries. Other issues can arise when there are multiple versions of a particular programming language on a system (e.g. *Python 2* and *Python 3*).

Freezing of code allows for the creation of a single executable file that can be distributed to users. This application/executable contains all the code and any additional resources required to run the application, and includes the Python interpreter that it was developed on. The major advantage for distributing applications in this manner is that it will work immediately, without the need for the user to have the required version of Python (or any additional necessary libraries) installed on their system. A disadvantage of generating a single file is that it will be larger, as all necessary libraries are incorporated. The increase in file size is acceptable when considering other issues, for example ease of installation, running, and portability to other platforms. Python freezing tools and platforms supported are listed in Table 4. *CMD Plot Tool* was developed on OSX using Python 3.6, and the following libraries were used and frozen with it: *numpy*, *pandas*, *tkinter*, *PIL*, *csv*, *math*, *matplotlib*, *glob*, *os* and *datetime*. The size of the frozen output file is 152 MB.

<i>Class Name (.py)</i>	<i>Lines of Code (Including Comments)</i>	<i>Description</i>
SplashScreen	41	Creates and positions a splash screen in the centre of the monitor when the application starts.
CMD_Plot_Tool	554	Constructs and manages operations via the GUI. Parameters and operational requests based on user interaction are passed to appropriate classes.
Plot_DAOPHOT	189	Selected DAOPHOT-format photometry (magnitude) files are passed to this class, which constructs and saves a CMD plot based on user-specified parameters.
Plotting	250	Selected .zpt and .txt photometry (magnitude) files are passed to this class, which constructs and saves a CMD plot based on the user-specified parameters and requirements.
Plot_Text	21	Manages every instance of an annotation on the plot. A reference to the annotation text, with its corresponding x,y coordinate and its pointer x,y coordinate, allows for multiple instances within a single plot.
ThePlot	41	Updates and displays the plot on the screen. Additional functionality also allows for this window to be moved, based on mouse clicks and movement.

Table 3: Characteristics of each class within *CMD Plot Tool*. Note that the "lines of code" column includes comments, i.e. non-executable code.

Tool	Apple OSX	Linux	Microsoft Windows
py2app	Yes	No	No
py2exe	No	No	Yes
pyInstaller	Yes	Yes	Yes
cx_Freeze	Yes	Yes	Yes

Table 4: Python freezing tools and the platforms they support. As this software was developed on OSX, *py2app* was used to successfully freeze the code into a single file. Further single-file applications/executables were created using *cx_Freeze* for the platforms listed above. These are available with the source code.

3. Functionality of *CMD Plot Tool*

The Graphical User Interface (GUI) of *CMD Plot Tool*, shown in Figure 5a, allows the user to select files and format plots from a single window. Users generating plots follow the flowchart outlined in Figure 5b. Current functionality includes:

- Set working directory: Select the directory where all plots will be saved.
- Save option: Depending on the scenario, users may generate multiple plots. Each time a plot is generated, it is saved using a timestamp. This aids efficiency, while also allowing the user to keep a detailed track record of plots generated. Users can overwrite this option.
- Select input file type: Current functionality allows the user to work on standard text files, *.mag* files generated from IRAF/DAOPHOT, and *.zpt* files from the ACS Survey of Galactic Globular Clusters (Sarajedini et al., 2007). Each file type is handled differently due to their unique format.
- Title and Axis options: format plot titles; select which columns in the data are to be plotted on X and Y axes; set axis limit values and units of scale on each axis independently from each other. This option enables zooming

into particular areas of the CMD plot. Examples of this are illustrated in Figure 7b and Figure 8b.

- Annotate options: Allows users to position text and format the colour of text and pointer labels.
- Error bars: Selecting this option displays error bars on each data point, while also allowing the user to format the colour and markers used. An example of this is illustrated in Figure 7b.
- Star count: The user will be informed as to the number of stars plotted within a region.
- Plot comparison: as an educational tool, it is extremely useful to illustrate to students and non-professionals the differences between data acquired from various sources. Figure 6 is an example of this.
- Delete plots: During the course of plotting CMDs, users may wish to test variations of different parameters. As a result, and over time, the number of plots generated will increase. If selected, all plots within the working directory are deleted.
- Reset parameters: Allows the user to reset all parameters within the GUI to their defaults.

3.1. Supported Input File Types

Text Files. The *CMD Plot Tool* offers users the option to plot data contained in standard text files. The format of the input data depends on the type of plot, i.e. with or without error bars. Plotting without error bars requires data to be presented in two column format, with each line containing *X Y* paired data, and the first line containing the heading name for each column. Plotting data with error bars follows the same format, with additional fields: *X Y X_error Y_error*.

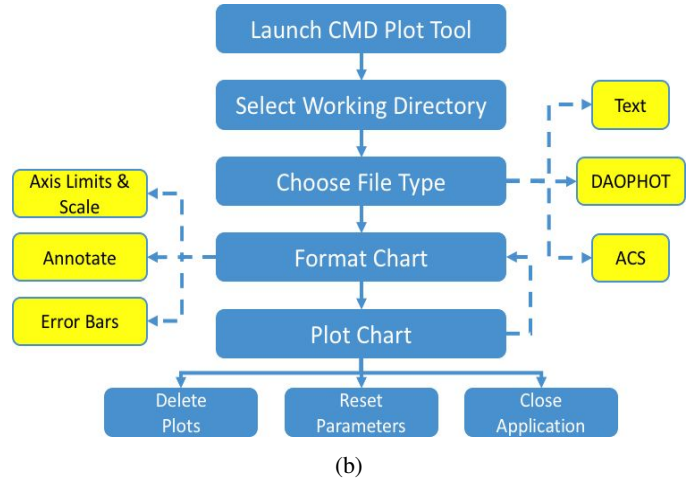
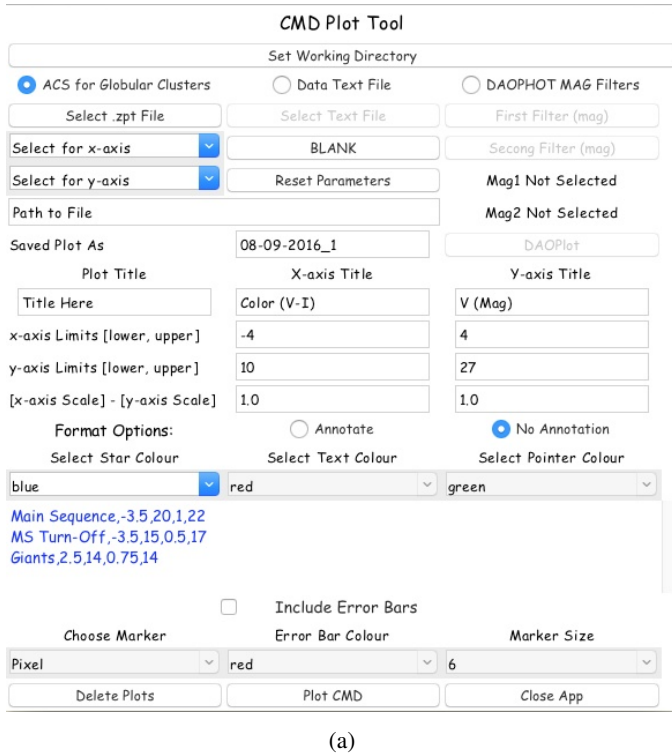


Figure 5: *CMD Plot Tool*: Panel (a) shows its Graphical User Interface (GUI). Panel (b) outlines its operational flowchart.

<i>Column Name</i>	<i>Description</i>
id	Identification of the star
x,y	Position of the star on the master image
Vvega, err, Vvega, err, Ivega, err	Magnitudes/colour and errors on the VEGAmag photometric system - stars with one measurement have errors based on root(n) noise.
Vground, Iground	Magnitudes on the ground VI system.
Nv, Ni	Number of times the star was measured in each filter.
wV, wI	Flags for the nature of image in V and I - short or deep.
xsig, ysig	RMS of x and y positions (large scatter indicates poor photometry)
othv, othi	Fraction of light from other stars (measure of crowding).
qfitV, qfitI	Quality of the fits - smaller is better
RA, Dec	The right ascension and declination (celestial coordinates) of the star at the J2000 equinox, from the HST image headers.

Table 5: ACS *.zpt* file format. Adapted from https://archive.stsci.edu/pub/hlsp/acsggct/hlsp_acsggct_photometry_readme.txt

ACS .zpt Files. The ACS Survey of Galactic Globular Clusters (Sarajedini et al., 2007) is an imaging survey of Galactic GCs using the Advanced Camera for Surveys instrument on board the Hubble Space Telescope (HST). This HST Treasury project¹⁰ was designed to obtain photometry with S/N $\gtrsim 10$ for main sequence stars with masses $\gtrsim 0.2M_{\odot}$ in the central 2 arcminutes of a sample of 66 globulars. Filters used included F606W (“Wide V” band; yellow light) and F814W (~I band; near-infrared light).

The column format of the *.zpt* file is described in Table 5. Users of *CMD Plot Tool* can select which magnitude/colour columns they wish to plot via a drop down menu that is made available when this file type is selected; normally the VEGA-mag columns are used. Corresponding error bar data is included if the user opts to display error bars on the plot.

DAOPHOT .mag files. DAOPHOT (ascl:1104.011) (Stetson, 1987) was developed for performing point-source photometry, particularly in crowded stellar fields; initially as a standalone package, and subsequently ported into IRAF. The typical DAOPHOT *.mag* file (star magnitude file) contains 33 columns of data relating to detected stars, in addition to meta-data about the observation and relevant DAOPHOT parameter settings. Not all of this is necessary for visualising a CMD plot. *CMD Plot Tool* selects a subset of fields within each *.mag* file (two *.mag* files must be selected for plotting). Corresponding fields are then merged and sorted based on each star’s unique ID within the file. Calculations are performed based on their

¹⁰Data available online: <https://archive.stsci.edu/prepds/acsggct/>

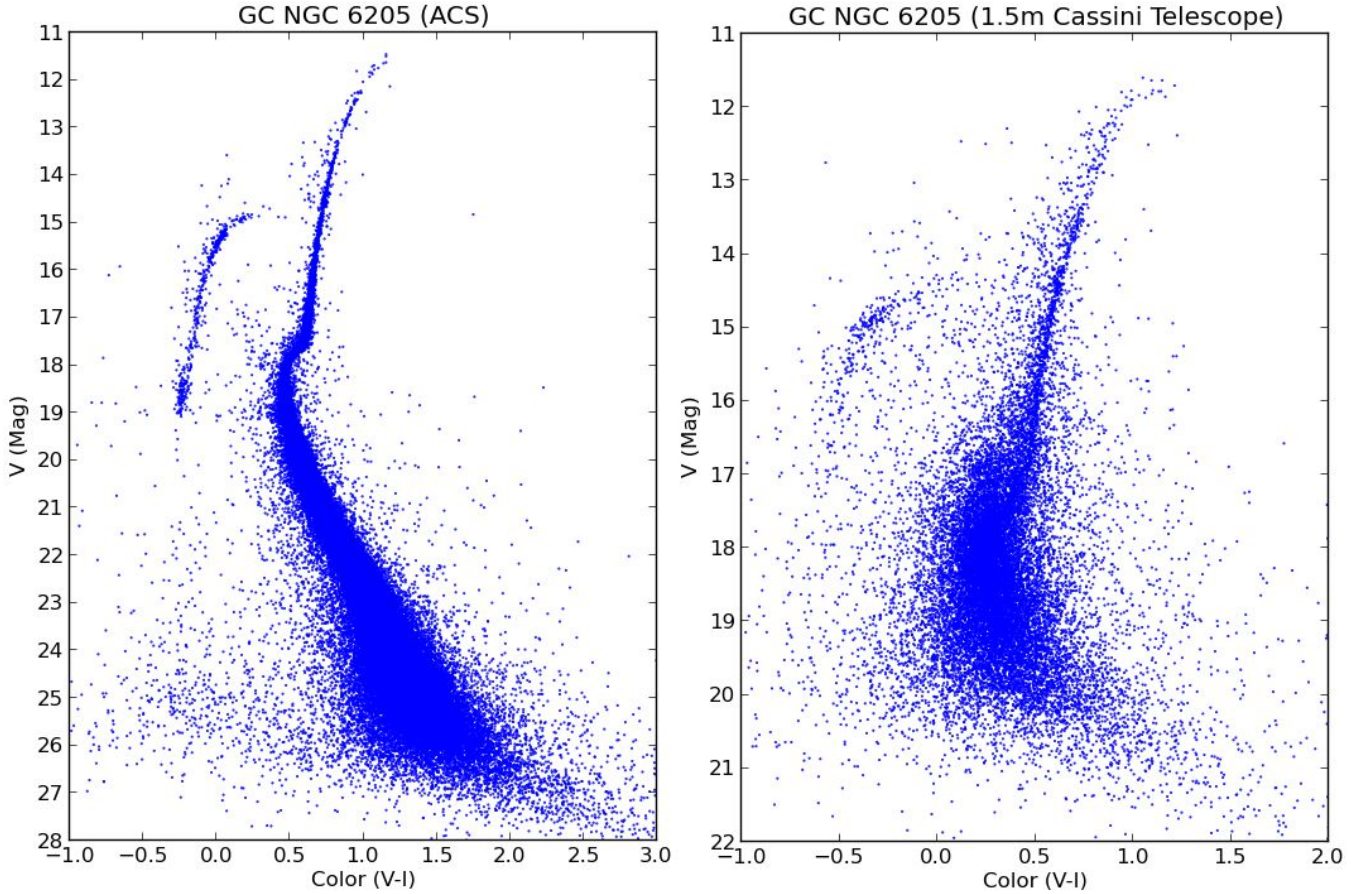


Figure 6: Comparison of CMDs of globular cluster NGC 6205 (M13), generated as a single output using the Plot Comparison functionality (see 3) of *CMD Plot Tool*. The plot on the left was generated from reduced Hubble Space Telescope photometry available from the ACS Survey of Galactic Globular Clusters (Sarajedini et al., 2007). The plot on the right was generated from data acquired by the authors with the 1.52m Cassini Telescope of the Osservatorio Astronomico di Bologna at Loiano and reduced using IRAF/DAOPHOT.

unique IDs, magnitudes and magnitude errors. The results are then saved in temporary storage and plotted. Once the application is terminated, all temporary files are removed.

It is important to note that users are not required to have IRAF/PyRAF installed on their system in order to plot CMDs based on DAOPHOT *.mag* files. The algorithm written to accomplish this utilises the Python Data Analysis Library¹¹ (*pandas*) and the Python CSV module. The *pandas* library is an open source library which provides high performance, easy to use data structures (data frames) and other data analytics tools for Python. Incorporating this with the CSV module allows for fast, reliable and efficient merging and calculating of large scale data. Performance is excellent even when the dataset being worked on contains tens of thousands of stars.

3.2. Annotation

The colour options available for text, markers, pointers and error bars, and shape options available for markers, are outlined in Table 6. Adding annotation to the plot is done via an algorithm that parses the text input by the user via the text area

Colour Options	Marker Options
Blue	Circle (○)
Red	Triangle (△)
Green	Diamond (◇)
Black	Point (●)
Cyan	Pixel (·)
Yellow	Star (★)
Magenta	X (×)

Table 6: Colour options available for text, markers, pointers and error bars, and shape options available for markers.

available on the GUI. Annotation is entered in the following format: *text*, *x1*, *y1*, *x2*, *y2*. *Text* represents the string that the user wishes to display, *x1*, *y1* represents the start position of the text, and *x2*, *y2* represents where the labelling arrow will point to. Data entered is parsed using a comma as a separator, removing unnecessary spaces. For example, the annotation in Figure 7b was entered as follows:

Main Sequence, 0.85, 17.92, 0.75, 17.75
MS-Turn Off, 0.85, 17.70, 0.75, 17.45

¹¹<http://pandas.pydata.org>

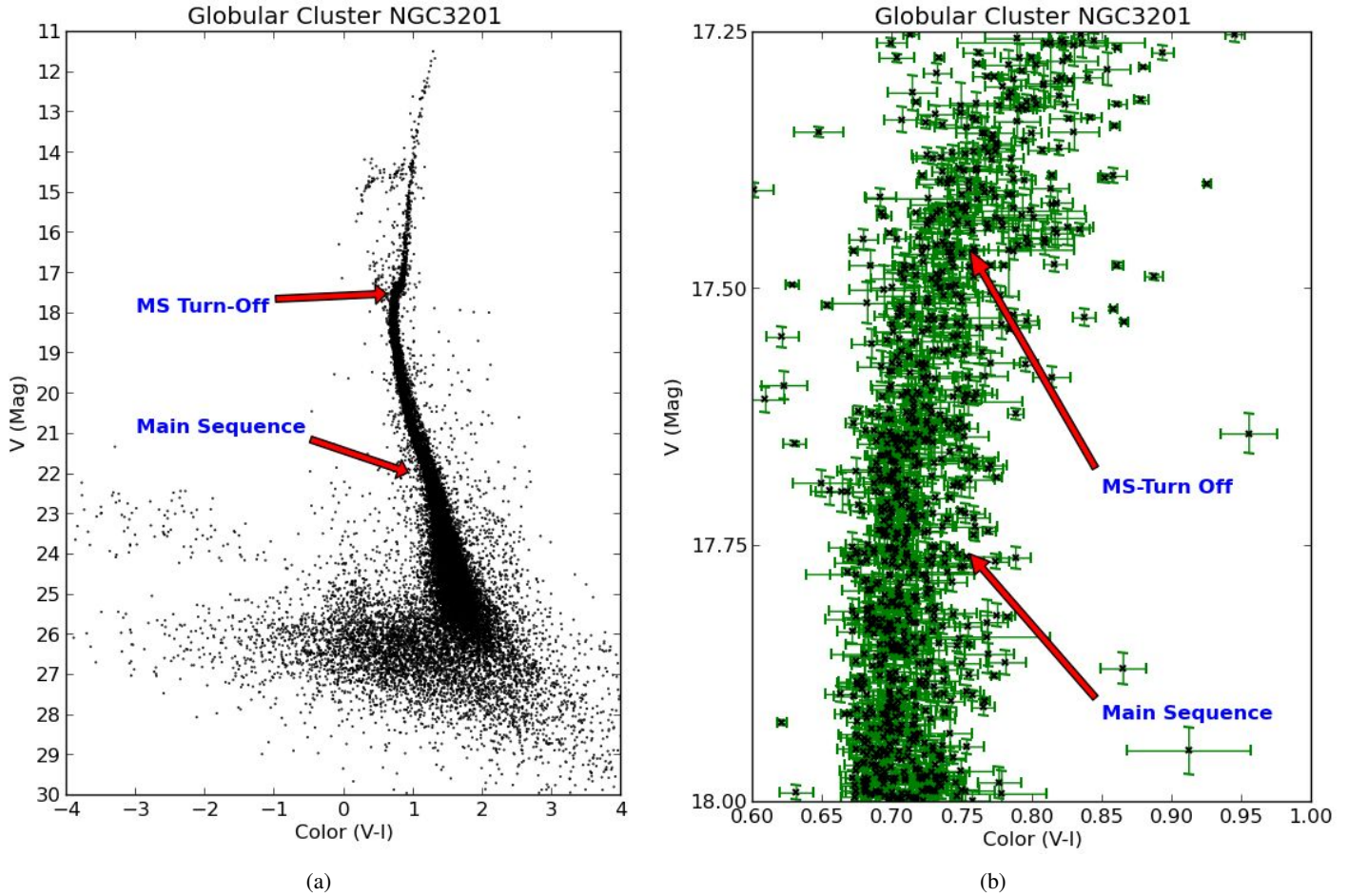


Figure 7: CMDs for GC NGC 3201, generated from *ACS Survey of Galactic Globular Clusters* data (Sarajedini et al., 2007). Panel (a) shows the full CMD without error bars, while Panel (b) shows error bars, and is zoomed in and annotated to highlight stars at the top of the main sequence and main sequence turn-off.

There is no restriction to the amount of annotation on each plot. For example, in Figure 8a where multiple arrows are projecting from text, the following annotation was applied:

```
MS, -2.75, 23, 0.4, 22
MS, -2.75, 23, 0.75, 24.5
```

If the user wishes to just add text with no pointer, then the $x1$, $y1$ and $x2$, $y2$ coordinates must be the same.

4. Usage Case Study for *CMD Plot Tool*: Interpreting the CMDs of Globular Clusters

Star Clusters. Stars form in clustered environments which are encapsulated in a giant molecular cloud (Walker et al., 2015). They are subjected to the gravitational forces of their stellar brethren, consume fuel to sustain themselves, and transform as they age. The energy created and exchanged between stars causes gas to heat to a point where it is blown away; during this time some stars may escape a newly forming cluster to become “runaways”. Those that remain become gravitationally bound. As a star approaches its end, it will swell to become a red giant; and depending on its mass, it will either eject its bloated atmosphere forming a planetary nebula and exposing its core

as white dwarf, or go supernova with a shockwave powerful enough to instigate star formation in other interstellar clouds. The remains of the core of the exploded star will form a neutron star or black hole; which is created will depend on the mass of the original star. This lifecycle can take anywhere from a few million to billions of years, again depending on the star. Star clusters are divided into two principal categories: open (Galactic) clusters and globular clusters.

Globular Clusters. Globular Clusters (GCs) are among the oldest radiant objects in the universe, with estimated ages between 10 Gyr – 16 Gyr (Krauss and Chaboyer, 2003). There are approximately 150 known GCs in the Milky Way galaxy, most residing in its spherical halo. Their stellar populations are extremely dense reaching up to 10^6 stars pc^{-3} . Due to their density, it is highly probable that collisions occur between stars within the cluster, such collisions would lead to a variety of exotic stellar objects such as pulsars, Blue Straggler stars, close interacting binaries, and X-ray sources (Díaz-Sánchez et al., 2012). GCs are also readily observable in external galaxies due to their compact sizes¹² and masses ranging

¹²half-light radii of a few parsecs

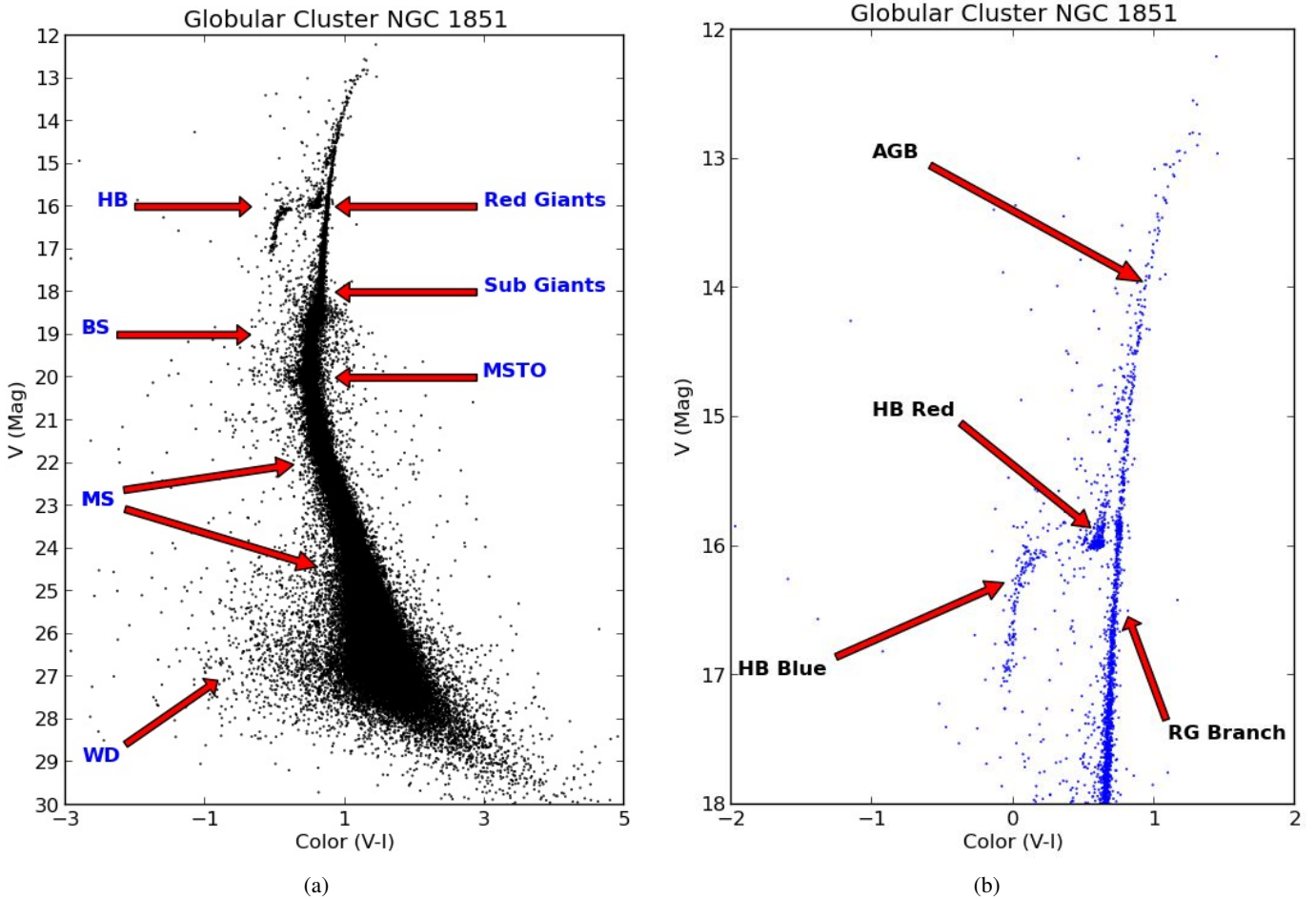


Figure 8: CMDs of Globular Cluster NGC 1851. Data from the ACS Survey of Galactic Globular Clusters (Sarajedini et al., 2007). Panel (a) shows the full CMD, while Panel (b) zooms in on the brighter sequences. Explanation of labels: MS: Main Sequence, MSTO: Main Sequence Turn Off, HB: Horizontal Branch, BS: Blue Stragglers, WD: White Dwarfs, RGB: Red Giant Branch, HB-Blue: blue population of the Horizontal Branch, HB-Red: red population of the Horizontal Branch, AGB: asymptotic giant branch.

from $10^4 M_{\odot} - 10^6 M_{\odot}$ (Brodie and Strader, 2006), although it is harder to resolve extragalactic GCs into individual stars.

GCs are of interest because of their age, assumed relatively homogeneous populations, relative isolation in their parent galaxies, and abundance of unusual objects, such as blue stragglers, X-ray binaries, radio pulsars, and intermediate-mass black holes (Portegies Zwart et al., 2010). Krauss and Chaboyer (2003) describe three independent ways to reliably infer the ages of the oldest star's in our galaxy: (i) Radioactive Dating, (ii) White Dwarf Cooling and (iii) Main Sequence Turnoff Time Scale: a robust prediction of the theoretical models for determining absolute GC age is the time it takes a star to exhaust the supply of hydrogen in its core and leaves the main sequence. However, uncertainties in determining distances to GCs are a source of doubt when estimating ages.

4.1. CMDs for Globular Clusters

The CMD is an extremely useful tool in aiding the understanding of the stellar populations within a cluster, and their evolution. Since a CMD plots apparent magnitudes, stars of the same inherent luminosity, but different distances from the ob-

server, are displaced vertically on a CMD. But since the depth of a star cluster is much less than its distance, all its stars can be taken to be at the same distance. The colours of the stars, calculated as the ratio of their fluxes (difference in logarithmic magnitudes) in two filter bandpasses, reflect their surface temperature. Age and chemical composition differences between stars of a given mass are (mainly) evidenced in temperature (colour) differences, displacing stars horizontally on the CMD. But since the stars in a cluster essentially formed from the same gas nebula at the same time, the assumption of equal age and chemical composition also generally holds. A typical cluster CMD therefore shows rather tightly defined sequences of stars, with the mass of a star determining where it lands in these sequences. As indicated in Figure 8a and Figure 8b, a typical CMD for a GC will include a main sequence (MS), main sequence turnoff point (MSTO), subgiant branch (SGB), red giant branch (RGB), horizontal branch (HB), and asymptotic giant branch (AGB). Blue stragglers (BS) may also be present between the MSTO and HB, and if the photometry reaches deep enough, white dwarfs (WDs) may be detected as well.

These CMD sequences and branches trace the evolutionary

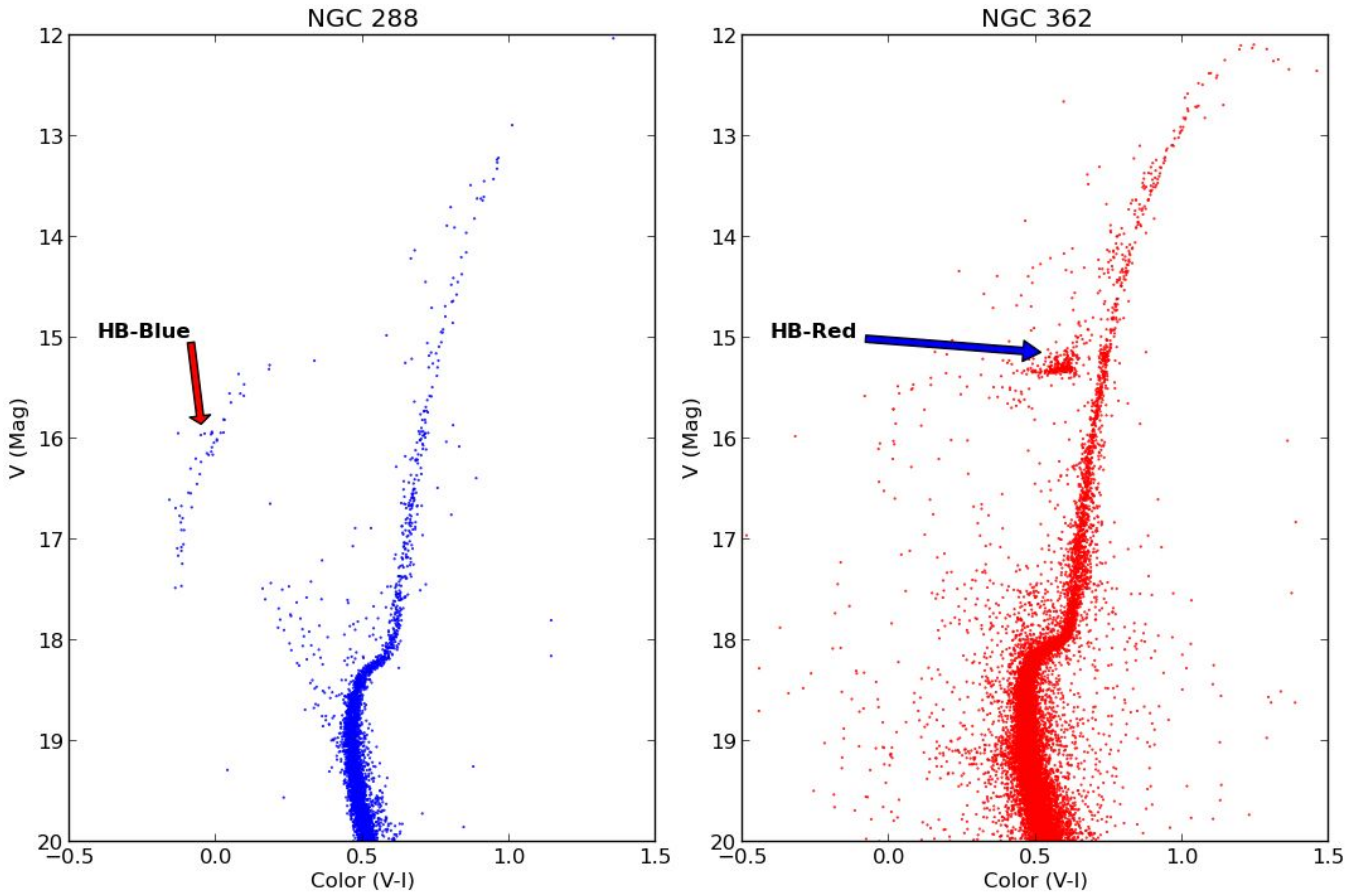


Figure 9: A single output file in Plot Comparison mode. Left side: CMD for GC NGC 288, where HB-Blue is the blue population of the horizontal branch. Right side: CMD for GC NGC 362, where HB-Red is the red population of the horizontal branch. Data from the ACS Survey of Galactic Globular Clusters (Sarajedini et al., 2007).

processes in the cluster’s stars. The relative numbers of stars in each branch also quantify the relative duration of each evolutionary phase - few stars present in a branch means that they quickly move on to the next phase.

The main sequence (MS). Most of a star’s life is spent on the main sequence (MS): a phase when hydrogen in the star’s core is fused into helium, with temperatures reaching ~ 15 million Kelvin. Energy is transported to the star’s surface through radiation and convection. There is difference between the CMD of a globular cluster and that of solar neighbourhood: in the latter, main sequence stars do not reach the horizontal branch. This occurs as such young stars are uncommon in GCs. The age of the cluster is estimated by the main sequence turnoff (MSTO).

The subgiant branch (SGB). This is a narrow section of the CMD, which joins the main sequence at the turnoff point. A star moves onto the turnoff point when the hydrogen in its core has been depleted and an inert helium core created. Hydrogen will continue burning (fusing) in a shell as the core contracts. There is a small increase in the luminosity of the star as the cooler envelope beyond the shell expands. This is known as the subgiant phase. The position of the SGB moves considerably with age, so the narrowness of this feature on a GC CMD limits

the stellar formation period to no more than $\sim 2\%$ of the cluster’s age (Stetson, 1993). The metallicity (relative abundance of elements heavier than helium) of the cluster governs the position of the MS and SGB on the CMD. Metal-poor clusters tend to have a much flatter SGB than that of metal-rich clusters.

The red giant branch (RGB). As the envelope of the star expands, the effective temperature decreases and it reaches an adiabatic temperature gradient. This is due to convection within the star’s interior and the efficiency of energy transportation to the surface. At this point the star will rise quickly up the RGB. When the star reaches the top of the RGB, its core temperature rises and helium fusion begins, producing carbon and oxygen. This is known as the helium flash.

The horizontal branch (HB). After the helium flash, energy is being released from helium burning in the star’s core plus hydrogen burning continues in the thin shell surrounding the core; as a result the star shrinks, becomes hotter and bluer, and jumps to the horizontal branch. CMDs reveal how the metallicity of the cluster influences the composition of the HB. Metal poor clusters have bluer stars than metal rich clusters. RR Lyrae pulsating variable stars can be found on the HB in the instability strip, which is located between red and blue HB stars. The

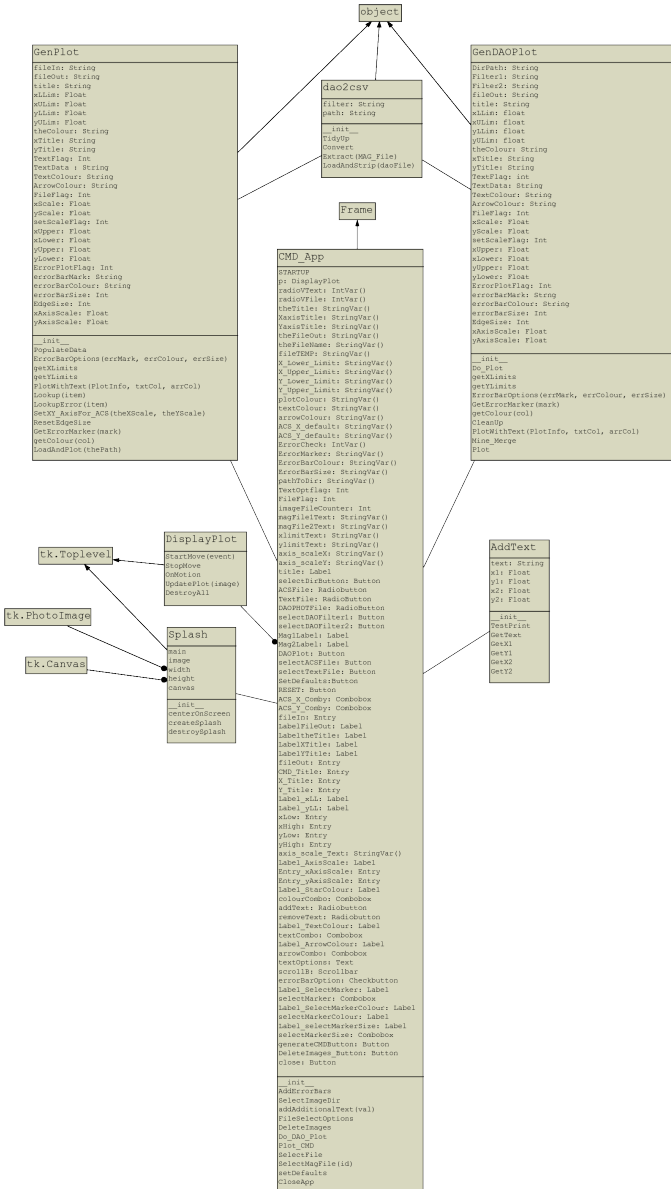


Figure 3: UML diagram for classes outlined in the Python files in Table 3

blue portion of the branch consists of helium burning similar to the hydrogen burning of the MS, at shorter time scales. As the star moves to the red portion of the branch, the helium in its core is depleted and has been completely converted into carbon and oxygen. This evolution occurs rapidly and forms a carbon-oxygen inert core. The populations of blue and red stars on the HB can be seen in Figure 8b. Although there is a correlation between HB colour and metallicity, there can also be a spread of HB star colours at any given metallicity. For example NGC 288 and NGC 362 have identical metallicities; yet NGC 288 has a very blue HB and NGC 362 had a very red HB, as can be seen in Figure 9.

The asymptotic giant branch (AGB). The AGB derives its name from the manner in which the evolutionary track approaches the line of the RGB from the left (see Figure 8b).

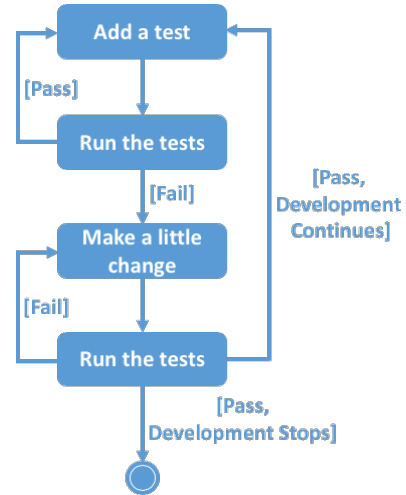


Figure 4: Test Driven Development Image reproduced from <http://agiledata.org/essays/tdd.html>. © Scott W Ambler

The inert carbon-oxygen core shrinks and heats up, with both helium and hydrogen burning shells around it. The star then moves up the AGB. Here the star has a higher temperature than when it was on the RGB. Stars on this branch will again cool and grow quickly. The core of such a star can be unstable, and its outer envelope can be ejected slowly while the core contracts. This leads to the creation of planetary nebulae with white dwarfs at their centres.

White dwarfs (WD). Because of their faintness, the most elusive sequence in most CMDs is the white dwarf cooling sequence. The older WDs have had more time to cool and fade, and the rate of cooling can be modelled, so the faintest detectable WD can be dated relative to the brightest WD.

CMDs and GC age. The oldest GCs have been of particular interest to researchers as they tried to determine their age. In doing so, they provided an excellent constraint on determining the age of the universe and subsequently on cosmological models used to describe it (VandenBerg et al., 2002). Isochrones are synthetic CMD sequences computed for a grid of different ages and metallicities. Determining the best isochrone fit for stars in the region of the turnoff (MSTO) best estimates the age of the GC. Analysing the variances of these isochrones between the TO and the start of the subgiant branch is shown to be nearly independent of age and chemical abundances. More recently, VandenBerg et al. (2013) determined the ages of 55 GCs using CMDs and an improved ΔV_{TO}^{HB} method. In that study, isochrones were fitted to just the TO portion of the observed CMD where morphology is predicted to be nearly independent of age and metallicity; furthermore inferred ages were based on the location of the beginning of the subgiant branch. This shows how several features on a CMD can be used for chronology.

5. Conclusions

We have presented a new application, *CMD Plot Tool*, for calculating and plotting CMDs. It can handle multiple file for-

mats to generate professional and customised plots, without the usual steep learning curve. Development using Python, OOP and a formal software development lifecycle model (Agile) allowed for the creation of an application that can be deployed on multiple systems. It works “out of the box” and does not require any installation of development environments, additional libraries or resetting of system paths. The tool is available as a single application/executable file, with the source code, on the public open-source repository, Sourceforge. Sample data is also bundled, to demonstrate its complete functionality to users. Other functionality within this application is the ability to convert DAOPHOT magnitude files to CSV format. We have illustrated one usage of the application: exploring the CMDs of globular star clusters.

Funding

This work was supported by Athlone Institute of Technology (K.F.) and National University of Ireland - Galway (L.-M.B.)

References

- Allen, A., Schmidt, J., 2015. Looking before leaping: Creating a software registry. *Journal of Open Research Software*, 3(1), p.e15, doi:<http://doi.org/10.5334/jors.vb>.
- Astels, D., 2003. *Test-Driven Development: A Practical Guide*. 1 ed., Prentice Hall.
- Ballester, P., Gabasch, A., Jung, Y., Modigliani, A., Taylor, J., Coccato, L., Freudling, W., Neeser, M., Marchetti, E., 2015. The High Level Data Reduction Library, in: Taylor, A.R., Rosolowsky, E. (Eds.), *Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV)*, p. 383.
- Beck, K., 2002. *Test Driven Development. By Example (Addison-Wesley Signature)*. Addison-Wesley Longman, Amsterdam.
- Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J., Thomas, D., 2001. *Manifesto for Agile Software Development*. <https://agilemanifesto.org/iso/en/manifesto.html>. Accessed: 2016-09-20.
- Bray, E.M., 2014. Astropy: Building Blocks for Astronomy Software, in: Manset, N., Forshay, P. (Eds.), *Astronomical Data Analysis Software and Systems XXIII*, p. 343.
- Brodie, J.P., Strader, J., 2006. Extragalactic Globular Clusters and Galaxy Formation. *Annu. Rev. Astron. Astrophys* 44, 193–267. doi:[10.1146/annurev.astro.44.051905.092441](https://doi.org/10.1146/annurev.astro.44.051905.092441), arXiv:[astro-ph/0602601](https://arxiv.org/abs/astro-ph/0602601).
- Collins, K.A., Kielkopf, J.F., Stassun, K.G., Hessman, F.V., 2017. AstroImageJ: Image processing and photometric extraction for ultra-precise astronomical light curves. *The Astronomical Journal* 153, 77. doi:[10.3847/1538-3881/153/2/77](https://doi.org/10.3847/1538-3881/153/2/77), arXiv:[1601.02622](https://arxiv.org/abs/1601.02622).
- Davies, R.I., Agudo Berbel, A., Wiezorrek, E., Cirasuolo, M., Förster Schreiber, N.M., Jung, Y., Muschiello, B., Ott, T., Ramsay, S., Schlichter, J., Sharples, R., Wegner, M., 2013. The Software Package for Astronomical Reductions with KMOS: SPARK. *A&A* 558, A56. doi:[10.1051/0004-6361/201322282](https://doi.org/10.1051/0004-6361/201322282), arXiv:[1308.6679](https://arxiv.org/abs/1308.6679).
- Díaz-Sánchez, A., Pérez-Garrido, A., Villó, I., Rebolo, R., Pérez-Prieto, J.A., Oscoz, A., Hildebrandt, S.R., López, R., Rodríguez, L.F., 2012. High-resolution optical imaging of the core of the globular cluster M15 with FastCam. *Monthly Notices of the Royal Astronomical Society* 423, 2260–2269. doi:[10.1111/j.1365-2966.2012.21034.x](https://doi.org/10.1111/j.1365-2966.2012.21034.x), arXiv:[1204.5340](https://arxiv.org/abs/1204.5340).
- Economou, F., Hoblitt, J.C., Norris, P., 2014. Your data is your dogfood: DevOps in the astronomical observatory. arXiv e-prints arXiv:[1407.6463](https://arxiv.org/abs/1407.6463).
- Goodman, A.A., 2012. Principles of high-dimensional data visualization in astronomy. *Astronomische Nachrichten* 333, 505. doi:[10.1002/asna.201211705](https://doi.org/10.1002/asna.201211705), arXiv:[1205.4747](https://arxiv.org/abs/1205.4747).
- Greenfield, P., White, R.L., 2000. A New CL for IRAF Based On Python, in: Manset, N., Veillet, C., Crabtree, D. (Eds.), *Astronomical Data Analysis Software and Systems IX*, p. 59.
- Hanisch, R.J., Berriman, G.B., Lazio, T.J.W., Emery Bunn, S., Evans, J., Mcglynn, T.A., Plante, R., 2015. The Virtual Astronomical Observatory: Re-engineering access to astronomical data. *Astronomy and Computing* 11, 190–209. doi:[10.1016/j.ascom.2015.03.007](https://doi.org/10.1016/j.ascom.2015.03.007), arXiv:[1504.02133](https://arxiv.org/abs/1504.02133).
- Kent, B.R., 2013. Visualizing Astronomical Data with Blender. *Publ. Astron. Soc. Pac* 125, 731–748. doi:[10.1086/671412](https://doi.org/10.1086/671412), arXiv:[1306.3481](https://arxiv.org/abs/1306.3481).
- Konstantopoulos, I.S., 2015. The starfish diagram: Visualising data within the context of survey samples. *Astronomy and Computing* 10, 116–120. doi:[10.1016/j.ascom.2015.01.007](https://doi.org/10.1016/j.ascom.2015.01.007), arXiv:[1407.5619](https://arxiv.org/abs/1407.5619).
- Krauss, L.M., Chaboyer, B., 2003. Age estimates of globular clusters in the milky way: Constraints on cosmology. *Science* 299, 65–69. doi:[10.1126/science.1075631](https://doi.org/10.1126/science.1075631).
- Leau, Y.B., Loo, W.K., Tham, W.Y., Tan, S.F. (Eds.), 2012. *Software Development Life Cycle AGILE vs Traditional Approaches*. volume 37, International Conference on Information and Network Technology (ICINT), Singapore.
- McMullin, J.P., Waters, B., Schiebel, D., Young, W., Golap, K., 2007. CASA Architecture and Applications, in: Shaw, R.A., Hill, F., Bell, D.J. (Eds.), *Astronomical Data Analysis Software and Systems XVI*, p. 127.
- Momcheva, I., Tollerud, E., 2015. Software Use in Astronomy: an Informal Survey. ArXiv e-prints arXiv:[1507.03989](https://arxiv.org/abs/1507.03989).
- Mulumba, P., Gain, J., Marais, P., Woudt, P., 2015. Scientific Visualization of Radio Astronomy Data using Gesture Interaction, in: Taylor, A.R., Rosolowsky, E. (Eds.), *Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV)*, p. 145.
- Portegies Zwart, S.F., McMillan, S.L.W., Gieles, M., 2010. Young Massive Star Clusters. *Annu. Rev. Astron. Astrophys* 48, 431–493. doi:[10.1146/annurev-astro-081309-130834](https://doi.org/10.1146/annurev-astro-081309-130834), arXiv:[1002.1961](https://arxiv.org/abs/1002.1961).
- Rehkopf, M., 2017. Atlassian Agile Coach - User Stories. <https://www.atlassian.com/agile/project-management/user-stories>. Accessed: 2018-09-13.
- Russell, H.N., 1913. “Giant” and “dwarf” stars. *The Observatory* 36, 324–329.
- Sami, M., 2012. Choosing the right Software Development Life Cycle Model. <https://melsatar.blog/2012/03/21/choosing-the-right-software-development-life-cycle-model/>. Accessed: 2017-09-20.
- Sarajedini, A., Bedin, L., Chaboyer, B., Dotter, A., Siegel, M., 2007. The ACS Survey of Galactic Globular Clusters. I. Overview and Clusters without Previous Hubble Space Telescope Photometry. *The Astronomical Journal* 133, 1658–1672. doi:[10.1086/511979](https://doi.org/10.1086/511979), arXiv:[astro-ph/0612598](https://arxiv.org/abs/astro-ph/0612598).
- Stetson, P.B., 1987. DAOPHOT - A computer program for crowded-field stellar photometry. *Publ. Astron. Soc. Pac* 99, 191–222. doi:[10.1086/131977](https://doi.org/10.1086/131977).
- Surace, J., Laher, R., Masci, F., Grillmair, C., Helou, G., 2015. The Palomar Transient Factory: High Quality Realtime Data Processing in a Cost-Constrained Environment, in: Taylor, A.R., Rosolowsky, E. (Eds.), *Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV)*, p. 197. arXiv:[1501.06007](https://arxiv.org/abs/1501.06007).
- Sybilski, P.W., Pawłaszek, R., Kozłowski, S.K., Konacki, M., Ratajczak, M., Helminiak, K.G., 2014. Software for autonomous astronomical observatories: challenges and opportunities in the age of big data, in: *Software and Cyberinfrastructure for Astronomy III*, p. 91521C. doi:[10.1117/12.2055836](https://doi.org/10.1117/12.2055836).
- Taylor, A.R., Rosolowsky, E. (Eds.), 2015. *Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV)*. volume 495 of *Astronomical Society of the Pacific Conference Series*.
- Tody, D., 1993. IRAF in the Nineties, in: Hanisch, R.J., Brissenden, R.J.V., Barnes, J. (Eds.), *Astronomical Data Analysis Software and Systems II*, p. 173.
- Vallejo, J.C., Lopez, F.P., Ortiz, I., Macfarlane, A., Osuna, P., Gill, R., Casale, M., 2015. Flexible and Modular Design for the BepiColombo Science Operations Control System, in: Taylor, A.R., Rosolowsky, E. (Eds.), *Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV)*, p. 277.
- VandenBerg, D.A., Brogaard, K., Leaman, R., Casagrande, L., 2013. The Ages of 55 Globular Clusters as Determined Using an Improved ΔV HB TO Method along with Color-Magnitude Diagram Constraints, and Their Implications for Broader Issues. *The Astrophysical Journal* 775, 134. doi:[10.1088/0004-637X/775/2/134](https://doi.org/10.1088/0004-637X/775/2/134), arXiv:[1308.2257](https://arxiv.org/abs/1308.2257).
- VandenBerg, D.A., Richard, O., Michaud, G., Richer, J., 2002. Models of metal-poor stars with gravitational settling and radiative accelerations. ii. the age of the oldest stars. *The Astrophysical Journal* 571, 487. URL: <http://stacks.iop.org/0004-637X/571/i=1/a=487>.

VanderPlas, J., Connolly, A.J., Ivezić, Z., Gray, A., 2012. Introduction to astroML: Machine learning for astrophysics. doi:[10.1109/CIDU.2012.6382200](https://doi.org/10.1109/CIDU.2012.6382200), [arXiv:1411.5039](https://arxiv.org/abs/1411.5039).

Walker, D.L., Longmore, S.N., Bastian, N., Kruijssen, J.M.D., Rathborne, J.M., Jackson, J.M., Foster, J.B., Contreras, Y., 2015. Tracing the Conversion of Gas into Stars in Young Massive Cluster Progenitors. ArXiv e-prints [arXiv:1502.03822](https://arxiv.org/abs/1502.03822).