Title

Astrostatistical Analysis in Solar and Stellar Physics

Author

Stenning, David Craig

Publication Date

2015

UNIVERSITY OF CALIFORNIA,
IRVINE

Astrostatistical Analysis in Solar and Stellar Physics

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

David Craig Stenning

Dissertation Committee:
Professor David A. van Dyk, Co-Chair
Professor Yaming Yu, Co-Chair
Professor Babak Shahbaba

2015

DEDICATION

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to offer my sincere gratitude to my advisor David van Dyk. This dissertation would not exist had it not been for your guidance, patience, and encouragement over the years. I cannot express how thankful I am for all the opportunities you have provided me throughout this journey.

I would also like to thank Yaming Yu, Vinay Kashyap, Thomas Lee, C. Alex Young, Nathan Stein, and everyone else at the International CHASC Astro-Statistics Collaboration.

I am also grateful to Ted von Hippel for his continued support. I have greatly enjoyed our work on Bayesian analysis of stellar evolution. And I cannot forget to thank Rachel Wagner-Kaiser and Elliot Robinson. Rachel went above and beyond in answering all my naive astrophysics questions, especially during the last few weeks of writing this dissertation. And I will forever be indebted to Elliot for his programming assistance.

I would like to thank all my professors at UCI for their instruction and support over the years. I am also grateful to both Rosemary Busta and Lisa Stieler for all the assistance they provided during my tenure at UCI. In addition, I would like to thank my fellow graduate students for being a source of intellectual stimulation, motivation, and (at times necessary) distraction. In particular, I would like to acknowledge Andrew Gelfand: I sorely missed our daily walk to Peet's during this final year.

I am thankful to Imperial College London for hosting me as a visiting graduate student. My time in London was more remarkable than I could have hoped.

Last, and most importantly, I would like to thank my parents for their unconditional support.

I am grateful for the financial assistance I have received from NSF grant DMS-09-07522, NSF grant DMS-12-09232, NASA grant 10-ADAP10-0076, and a Dean's Fellowship from the Donald Bren School of Information and Computer Sciences.

I am grateful to Wiley Periodicals, Inc. for allowing me to use material from the article "Morphological feature extraction for statistical learning with applications to solar image data" published in the journal *Statistical Analysis and Data Mining*. Chapter 2 of this dissertation is adapted from that article, and I am grateful to my co-authors—Thomas C. M. Lee, David A. van Dyk, Vinay Kashyap, Julia Sandell, and C. Alex Young—for directing and supervising the research which forms the basis for that chapter.

I am grateful to Taylor and Francis Group LLC Books for allowing me to republish the chapter "A Bayesian Analysis of the Solar Cycle Using Multiple Proxy Variables" from the edited volume *Current Trends in Bayesian Methodology with Applications* (eds: Satyanshu K. Upadhyay, Umesh Singh, Dipak K. Dey, and Appaia Loganathan). Chapter 3 of this dissertation is adapted from that book chapter, and I am grateful to my co-authors—David A. van Dyk, Yaming Yu, and Vinay Kashyap—for directing and supervising the research which forms the basis for that chapter.

CURRICULUM VITAE

David Craig Stenning

EDUCATION

Doctor of Philosophy in Statistics **2015**
University of California, Irvine *Irvine, CA*

Master of Science in Statistics **2013**
University of California, Irvine *Irvine, CA*

Bachelor of Science in Applied Mathematics **2007**
Columbia University *New York, NY*

RESEARCH EXPERIENCE

Graduate Research Assistant **2009–2015**
University of California, Irvine *Irvine, California*

Laboratory Research Assistant **2007–2009**
Columbia Astrophysics Laboratory *New York, NY*

TEACHING EXPERIENCE

Teaching Assistant **2009, 2011–2012, 2014**
University of California, Irvine *Irvine, CA*

SELECTED HONORS AND AWARDS

Dean's Fellowship, Information and Computer Sciences **2009–2013**
University of California, Irvine

REFEREED JOURNAL PUBLICATIONS

Morphological Feature Extraction for Statistical Learning with Applications to Solar Image Data **Aug 2013**
Statistical Analysis and Data Mining

REFEREED BOOK CHAPTERS

A Bayesian Analysis of the Solar Cycle Using Multiple Proxy Variables **2015**
Current Trends in Bayesian Methodology with Applications, Editors: S. Upadhyay, D.K. Dey, U. Singh and A. Loganathan

Morphological Image Analysis and its Application to Sunspot Classification **2012**
Statistical Challenges in Modern Astronomy V, Editors: G.J. Babu and E.D. Feigelson

UNREFEREED CONFERENCE PUBLICATIONS

The Power of Principled Bayesian Methods in the Study of Stellar Evolution **Nov 2014**
EES 2013: The Age of Stars

Deriving the Ages of Field White Dwarfs **Jan 2015**
Proceedings of the 19th European Workshop on White Dwarfs

Deriving Precise Ages of Field White Dwarfs using Bayesian Techniques **Jan 2015**
Proceedings of the 19th European Workshop on White Dwarfs

TECHNICAL DOCUMENTATION

Bayesian Analysis for Stellar Evolution with Nine Parameters (BASE-9): User's Manual **Nov 2014**
`http://arxiv.org/abs/1411.3786`

INVITED PRESENTATIONS

Bayesian Analysis of the Solar Cycle Using Multiple Proxy Variables **Nov 2014**
U.S. Naval Research Laboratory

An Introduction to Bayesian Statistical Methods for Astronomy **Nov 2014**
Embry-Riddle Aeronautical University

Science-Driven Feature Extraction with Solar Image Data **Dec 2013**
Royal Observatory of Belgium

Commentary on "Algorithms for Solar Active Region Identification and Tracking." **Feb 2012**
Solar Statistics 2012, Harvard-Smithsonian Center for Astrophysics

Automatic Classification and Tracking of Solar Features **Dec 2011**
Workshop on Current Challenges in Statistical Learning, Banff International Research Station

CONTRIBUTED POSTERS

Bayesian Analysis of the Solar Cycle Using Multiple Proxy Variables **Aug 2014**
Solar Information Processing Workshop VII, La Roche-en-Ardenne, Belgium

Automatic Classification and Tracking of Solar Features **Aug 2012**
Solar Information Processing Workshop VI, Montana State University

Automatic Classification and Tracking of Solar Features **Feb 2012**
Solar Statistics 2012, Harvard-Smithsonian Center for Astrophysics

ABSTRACT OF THE DISSERTATION

Astrostatistical Analysis in Solar and Stellar Physics

By

David Craig Stenning

Doctor of Philosophy in Statistics

University of California, Irvine, 2015

Professors David A. van Dyk, Yaming Yu, Co-Chairs

This dissertation focuses on developing statistical models and methods to address data-analytic challenges in astrostatistics—a growing interdisciplinary field fostering collaborations between statisticians and astrophysicists. The astrostatistics projects we tackle can be divided into two main categories: modeling solar activity and Bayesian analysis of stellar evolution. These categories from Part I and Part II of this dissertation, respectively.

The first line of research we pursue involves classification and modeling of evolving solar features. Advances in space-based observatories are increasing both the quality and quantity of solar data, primarily in the form of high-resolution images. To analyze massive streams of solar image data, we develop a science-driven dimension reduction methodology to extract scientifically meaningful features from images. This methodology utilizes mathematical morphology to produce a concise numerical summary of the magnetic flux distribution in solar "active regions" that (i) is far easier to work with than the source images, (ii) encapsulates scientifically relevant information in a more informative manner than existing schemes (i.e., manual classification schemes), and (iii) is amenable to sophisticated statistical analyses.

In a related line of research, we perform a Bayesian analysis of the solar cycle using multiple proxy variables, such as sunspot numbers. We take advantage of patterns and correlations among the proxy variables to model solar activity using data from proxies that have become available more recently, while also taking advantage of the long history of observations of sunspot numbers. This model is an extension of the Yu *et al.* (2012) Bayesian hierarchical model for the solar cycle that used the sunspot numbers alone. Since proxies have different temporal coverage, we devise a multiple imputation scheme to account for missing data. We find that incorporating multiple proxies reveals important features of the solar cycle that are missed when the model is fit using only the sunspot numbers.

In Part II of this dissertation we focus on two related lines of research involving Bayesian analysis of stellar evolution. We first focus on modeling multiple stellar populations in star clusters. It has long been assumed that all star clusters are comprised of single stellar populations—stars that formed at roughly the same time from a common molecular cloud. However, recent studies have produced evidence that some clusters host multiple populations, which has far-reaching scientific implications. We develop a Bayesian hierarchical model for multiple-population star clusters, extending earlier statistical models of stellar evolution (e.g., van Dyk *et al.*, 2009; Stein *et al.*, 2013). We also devise an adaptive Markov chain Monte Carlo algorithm to explore the complex posterior distribution. We use numerical studies to demonstrate that our method can recover parameters of multiple-population clusters, and also show how model misspecification can be diagnosed. Our model and computational tools are incorporated into an open-source software suite known as BASE-9. We also explore statistical properties of the estimators and determine that the influence of the prior distribution does not diminish with larger sample sizes, leading to non-standard asymptotics.

In a final line of research, we present the first-ever attempt to estimate the carbon fraction of white dwarfs. This quantity has important implications for both astrophysics and fundamental nuclear physics, but is currently unknown. We use a numerical study to demonstrate that assuming an incorrect value for the carbon fraction leads to incorrect white-dwarf ages of star clusters. Finally, we present our attempt to estimate the carbon fraction of the white dwarfs in the well-studied star cluster 47 Tucanae.

Part I

Modeling Solar Activity

Chapter 1

Background

1.1 The Sun

The Sun is the closest star to the Earth; without it life on our planet could not exist. The core of the Sun is hot and dense, providing conditions for the thermonuclear fusion of hydrogen into helium that fuels the Sun's energy output. The Sun is a ball of *plasma*—an ionized gas with charged particles (ions) that is electrically conductive. Since the Sun is plasma, as opposed to gas, internal motion in the Sun generates a magnetic field, or *magnetosphere*. The physical process that generates the Sun's magnetic field is known as the *solar dynamo*.

Solar physicists monitor the Sun nearly continuously using spaced-based observatories such as the *Solar and Heliospheric Observatory* (SoHO), the *Transition Region and Coronal Explorer* (TRACE), and the *Solar Dynamics Observatory* (SDO). These observatories carry high-resolution and high-cadence instruments operating at various wavelengths. For instance, SDO obtains images of the solar *corona*—the Sun's "atmosphere"—in seven different filters ranging from extreme ultraviolet to soft X-ray wavelengths at a rate of roughly one 4096 × 4096 pixel image per second. An image of the Sun captured by SDO on August 31, 2013 in the 171 angstrom extreme ultraviolet (EUV) wavelength band is presented in Figure 1.1. The bright filaments that can be seen are plasma-filled structures that trace the Sun's magnetic field, known as *coronal loops*. The loop structures viewed in Figure 1.1 can be used to represent the complexity of the Sun's magnetic field structure, and complex magnetic field configurations are related to volatile events originating with the release of magnetic energy in the corona.

1.2 Solar Activity and Space Weather

A major concern of current solar physics, and a stated mission for current solar observatories, is to improve understanding of the Sun's influence on the Earth and on Near-Earth space (NASA, 2009). Activity in the solar corona can result in extreme space weather events and have a damaging impact on Earth. In particular, highly energetic events such as *solar flares*—sudden bursts of radiation following the release of magnetic energy—and *coronal mass ejections* (CMEs)—massive bursts of coronal material—eject charged particles into space, which have the potential to damage technological infrastructure (NRC, 2008). For instance, a geomagnetic storm in 1989 was responsible for the collapse of the Hydro-Québec power grid and left millions of people without power for nine hours (Bolduc, 2002). In addition, charged particles pose a danger to astronauts on the International Space Station, or even passengers flying in aircraft at high altitude both through exposure to radiation and through the potential damage to aircraft computer systems.

Solar flares and CMEs are known to be related to various observed solar features, in particular *sunspots* and their corresponding magnetic *active regions*. Sunspots are dark areas on the Sun's *photosphere*—the region that emits the light that we see—that form when convection is inhibited by intense magnetic fields. Sunspots are classified based on the complexity of the associated magnetic flux distribution as viewed in *magnetograms*—images of the spatially-

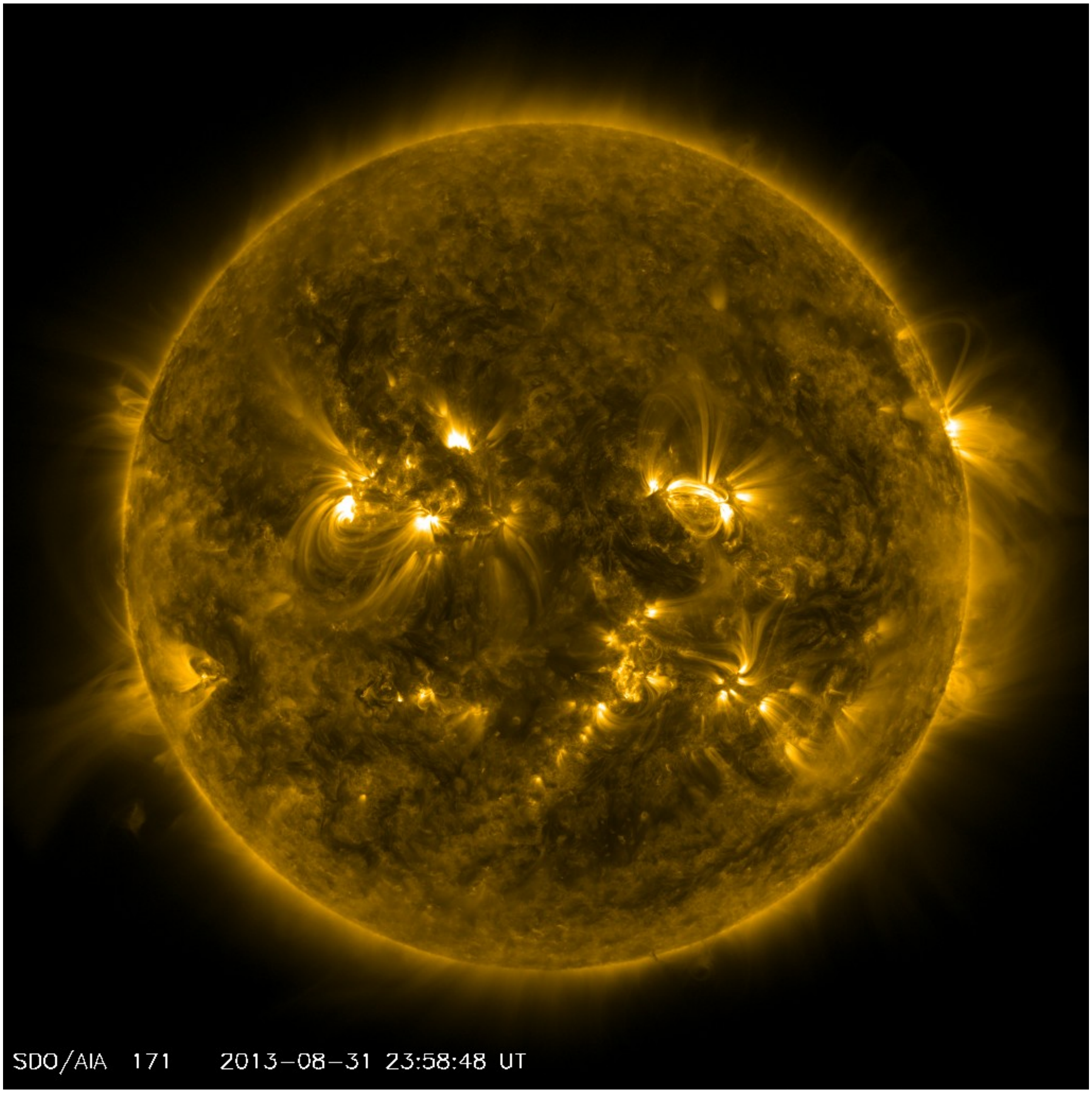


Figure 1.1: The Sun. This image was taken by an instrument onboard the Solar Dynamics Observatory on August 31, 2013. We are viewing the Sun as it appears in a wavelength of extreme ultraviolet light, specifically at a wavelength of 171 angstroms. At this wavelength we observe plasma at temperatures of approximately 6.3×10^5 K, which corresponds to the Sun's corona. Regions in the image that appear brighter correspond to higher temperature plasma. Furthermore, the filaments that appear connected to the brightest regions are coronal loops, which trace the Sun's magnetic field. (Image courtesy of NASA/SDO.)

resolved line-of-sight magnetic field in the photosphere. A *white-light* image of the Sun's photosphere taken by SDO on August 31, 2013 is presented in the left panel of Figure 1.2.

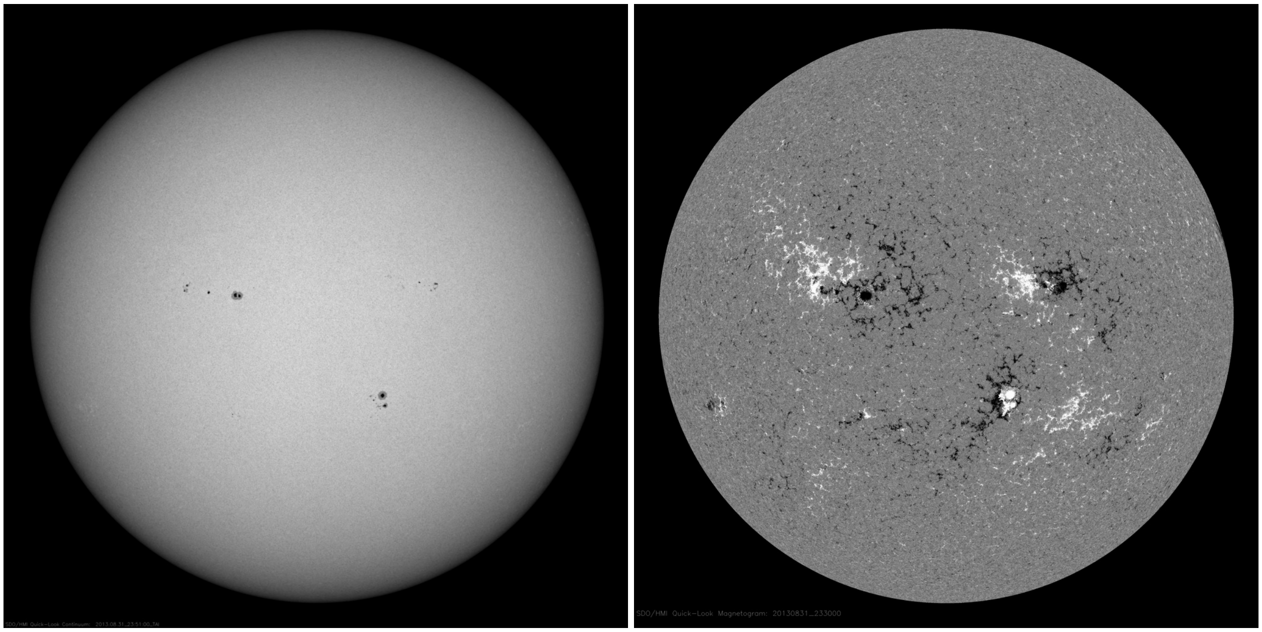


Figure 1.2: The Photosphere and Magnetosphere. *Left:* An image of the Sun's photosphere captured by an instrument onboard SDO on August 31, 2013. The dark spots visible on the solar disk are sunspots. *Right:* A line-of-sight projection of the Sun's magnetosphere, represented by a magnetogram obtained by instruments onboard SDO on August 31, 2013. White and black areas correspond to areas of high magnetic flux, where white corresponds to positive magnetic polarity and black corresponds to negative magnetic polarity (with grey indicating a lack of magnetic flux). We observe that areas of high magnetic flux correspond to the sunspots visible in the left panel. (Image courtesy of NASA/SDO.)

The dark spots visible on the face of the Sun are sunspots. A line-of-sight magnetogram obtained by SDO at nearly the same time is presented in the right panel of Figure 1.2. The black and white colors represent regions of strong magnetic activity, while grey represents regions where no magnetic field is present. Black regions represent "south" or "negative" magnetic polarity (i.e. the magnetic field is directed inwards toward the sun), while white regions represent "north" or "positive" magnetic polarity (i.e. the magnetic field is directed outwards, towards SDO/Earth).

It can be seen in Figure 1.2 that areas of high magnetic activity observed in the magnetogram correspond to regions where we find sunspots on the photosphere in the white-light image. Hence, sunspots are associated with areas of high magnetic activity. Furthermore, coronal loops trace the magnetic fields that erupt from sunspots. All these structures—sunspots, active regions, and coronal loops—are therefore scientifically meaningful. Stenning *et al.* (2012, 2013) develop a method for extracting and quantifying scientifically meaningful numerical features from magnetogram, white-light, and EUV images; this is an important first step in a proposed plan to describe the evolution of the Sun's active regions and predict which will erupt in massive solar flares and/or CMEs. This work is described in Chapter 2.

1.3 The Solar Cycle

Highly energetic solar eruptions such as solar flares and CMEs do not occur at a constant rate; rather, space weather events like these are more common during periods of high *solar activity*—a loose term that is defined only by observable proxy variables. Variations in the level of solar activity follow a roughly 11-year cyclic pattern, which is known as the *solar cycle*. This pattern is readily apparent in the time series of monthly average sunspot numbers that is presented in Figure 1.3. Since solar flares and CMEs are more common near the *solar maximum*—the peak in solar activity during the 11-year cycle—there is considerable interest in predicting the timing and amplitude of future solar maxima, which has practical value in the planning of space missions. Nevertheless, predicting solar maxima remains a difficult task, with different methods yielding substantially different predictions. See Pesnell (2012) for an analysis of the various predictions made for the current solar cycle.

Yu *et al.* (2012) proposes a statistical model of the solar cycle that was constructed using the monthly average sunspot numbers. This model summarizes the evolution of the solar cycle in a set of relations that yield the amplitude, duration, and time of rise to maximum for a cycle, given the values from the previous cycle. Stenning *et al.* (2015) extends this model to incorporate additional proxies besides the monthly average sunspot numbers. The use of additional proxies reveals important characteristics of the solar cycle that are missed by the sunspot numbers alone. This work is described in Chapter 3.

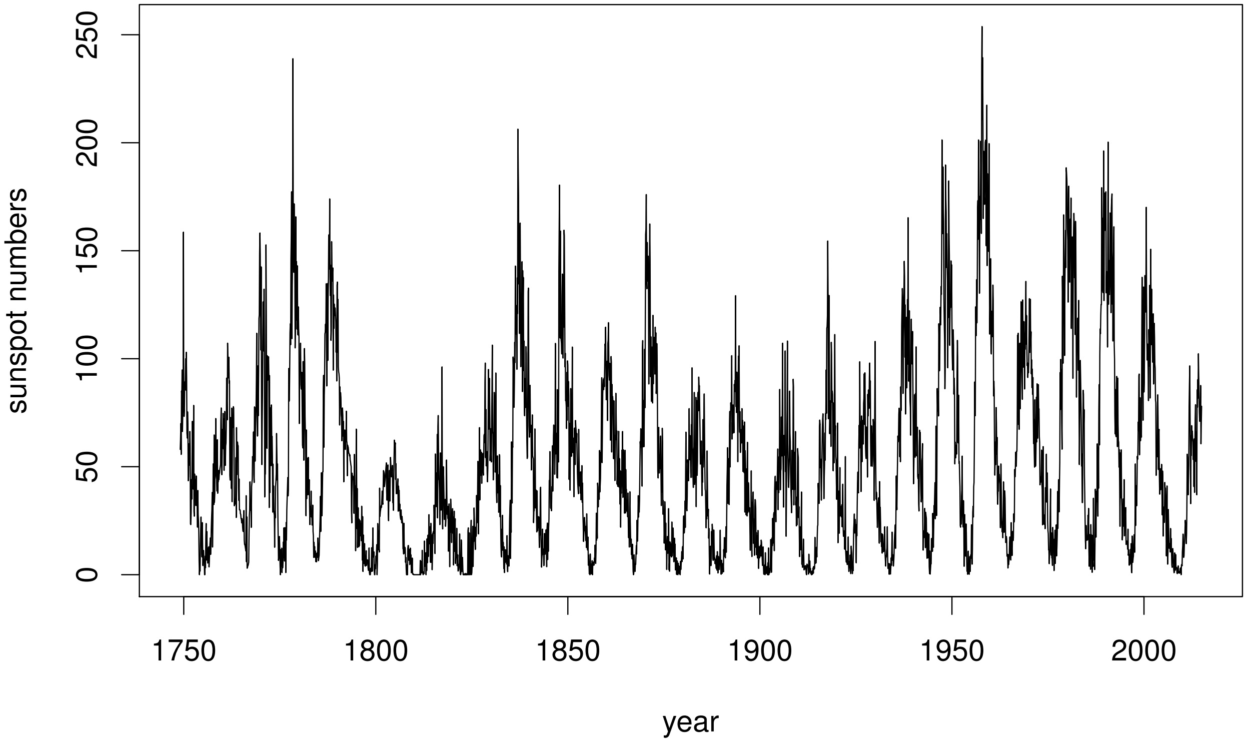


Figure 1.3: The Solar Cycle. The time series is the monthly average sunspot number from January, 1749 through December, 2014. To obtain the sunspot number for a particular day, solar physicists first count both the number of individual sunspots and the number of "sunspot groups" visible on the solar disk. A sunspot group is a physically co-located collection of spots that share a common active region. The sunspot number is then the sum of the number of individual spots and ten times the number of sunspot groups, after applying a correction for systematic differences between instruments and observatories. More details on calculating sunspot numbers can be found in Section 3.1. The solar cycle was discovered by noticing the 11-year cyclic pattern in the time series of sunspot numbers.

Chapter 2

Morphological Feature Extraction in Statistical Image Analysis

2.1 Introduction

The ability to extract meaningful information from large amounts of image data has far-reaching applications in such diverse fields as medicine, computer vision, and astronomy (Gonzalez and Woods, 2002). Advances in imaging technology are yielding massive data sets that are increasingly laborious to process manually. Studying complex images "by eye" also limits the types of analyses that can be performed since interesting features must be extracted and propagated in machine-readable form before they can be utilized in sophisticated statistical procedures. The need for automated methods is particularly apparent in the field of solar physics, with space-based observatories such as SoHO, TRACE, and SDO producing unprecedented amounts of high-quality image data. Experts typically detect and analyze features in SoHO and TRACE images manually, but newer observatories such as SDO—with its continuous science data downlink rate of 130 Megabits per second—render common labor-intensive techniques impractical. With routines arising from *mathematical morphology* (Section 2.2), we develop general techniques for extracting scientifically meaningful numerical quantities from complex high-throughput images and use them as covariates in statistical learning methods for classification and ultimately tracking and prediction of solar features. *Our overriding goal is to extract scientifically meaningful and interpretable numerical features from solar images.* The numerical features can be carried forward into secondary analyses that will also be interpretable in terms of meaningful scientific quantities.

Our goal of extracting numerical features from images for use in secondary statistical analysis is similar in spirit to the use of functional data as predictor variables in regression. This is typically accomplished using a set of independent basis functions that represent the functional data. Although this is a mathematically attractive strategy, it does not generally lead to scientifically meaningful summaries. One notable exception involves the use of a dependent library of generating functions to represent the functional predictors (Woodard *et al.*, 2012). This allows quantities such as the frequency, locations, and size of dips, bumps, and plateaus to be captured and passed on to the secondary analysis. Like this, we also aim to preserve scientifically meaningful summaries, but of very different predictors: images of complex solar features.

Mathematical morphology (MM) is a valuable tool for extracting shape characteristics from image data, and is well suited to the task of analyzing complex solar features. It is a non-linear process that is highly effective in extracting useful numerical summaries from image data, as we show below. Using appropriate morphological operations, images can be simplified by preserving the essential shape of geometric structures and eliminating noise. Therefore, MM is an excellent imaging tool for filtering, segmentation, and taking measurements such as feature areas from an image.

Our general approach to solving practical solar imaging problems is to break the original problem into a sequence of sub-problems each of which can be solved in a relatively simple manner. For example, one may decompose an image classification problem into: (i) clean the image, (ii) perform segmentation to delineate the features of interest, (iii) extract various measurements from the image, and (iv) feed these measurements to a classifier. In this example MM can naturally be applied to solve sub-problems (i) to (iii). In the remaining of this section we describe a specific solar imaging problem for which MM can be employed to solve some sub-problems.

As discussed in the previous chapter, energetic solar eruptions such as solar flares and CMEs eject charged particles into space and can damage technological infrastructure; these events are known to be related to various observed solar features, in particular sunspots and their corresponding magnetic active regions. Experts classify sunspots based on the complexity of associated magnetic flux distribution as viewed in magnetograms. One sunspot classification scheme in particular, the *Mount Wilson scheme*, has some power to predict solar flares and CMEs when combined with other space weather data (Ireland *et al.*, 2008). However, this classification is carried out manually and as a result is both laborious and prone to inconsistencies stemming from human observer bias (Ireland *et al.*, 2008). That is, manual classification results in non-reproducible catalogues since two experts looking at the same set of images will not always agree. Automated sunspot classification procedures based on statistical learning methods will result in reproducible catalogues, but require numerical covariates as inputs. Using our general numerical feature extraction techniques, we produce summaries of sunspots/active regions from SoHO images that are relevant to the sunspot's classification. The scientific relevance of these numerical summaries is demonstrated by their successful use as input covariates to a supervised learning algorithm that can reproduce manual classifications with an acceptable level of agreement. As we will discuss in further detail in Section 2.3, it is not necessary or desirable to have the automatic classifier exactly mimic the manual class assignments. Insofar as the Mt. Wilson classification scheme contains relevant information regarding activity around a sunspot (Ireland *et al.*, 2008), by constructing numerical summaries guided by the Mt. Wilson classification rules we aim to capture the same useful scientific information. The key is that the information is obtained in a self-consistent manner, leading to more objective and reproducible data analysis. The scientific information will also be encoded in numerical feature vectors instead of images, opening increased opportunities for downstream analyses.

The increase in quantity and quality of solar image data has spurred interest in developing automated techniques for processing such data. A general review of existing image processing techniques—including MM—useful for automated feature recognition with solar data is given in Aschwanden (2010). Simple MM is used by Curto *et al.* (2008) in procedures for automatically detecting and grouping sunspots. While this method is broadly similar to the initial step of our approach, it focuses on *identifying* sunspot groups whereas we are interested both in classifying sunspot groups and in obtaining numerical summaries of sunspot groups and active regions that can be used for statistical learning. Identification of the sunspot groups is a necessary precursor to both of these tasks. Colak and Qahwaji (2009) presents a system for automatically detecting and classifying sunspot groups according to the McIntosh classification scheme (McIntosh, 1990). While we develop our methodology to match the Mt. Wilson scheme, which is more useful as a measure of the complexity of the magnetic field structure, our results and reclassifications will be applicable in either case.

This chapter is divided into four sections. We begin in Section 2.2 with a brief introduction to MM and standard image analysis tools, and describe our general approach for extracting scientifically meaningful numerical features from images that can be used for statistical learning. In Section 2.3 we show how our general techniques can be used to extract numerical features from complex SoHO magnetogram and white light images that can be used in an automatic sunspot classification algorithm. Finally, in Section 2.4 we discuss our results and directions for future work. Throughout this chapter we use the word *feature* to describe interesting aspects of an image, such as sunspots or active regions. This is not to be confused with the numerical summaries that are typically referred to as *features* in the machine learning literature. We refer to the latter as *numerical features* to avoid potential confusion.

2.2 Science-Driven Image Analysis

The goal of science-driven image analysis is to derive scientifically meaningful quantities and machine-readable representations of images that can be used for statistical learning. When combined with standard image analysis techniques, MM is a powerful tool for capturing the essential scientific information in a simple "sketch" representation—a segmented image that resembles the drawing an expert would make in copying the raw image by hand. For example, magnetograms can be segmented into simplified "trinary" images in which regions of negative magnetic polarity, positive magnetic polarity, and background are assigned values of two, one and zero, respectively. The trinary images sketch the solar feature of interest, the magnetic active region, so that numerical summaries capturing important scientific information can be calculated.

2.2.1 Feature Recognition

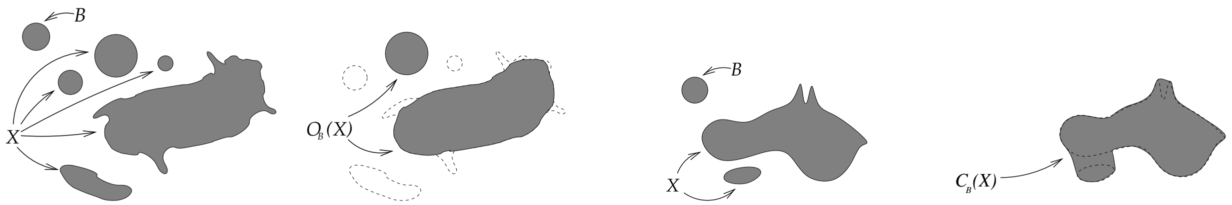
The first step in science-driven image analysis is to recognize scientifically meaningful features. For example, we need to be able to detect sunspots and active regions in solar image data since those features provide rich information about solar processes. By looking at an intensity histogram of an image, we can often determine whether the interesting features are best identified by *thresholding* the histogram at some particular value. Typical strategies to determine the threshold value include using the standard-deviation in the histogram, using a global or local median filter, etc. However, this method is not universally applicable because the features of interest may not be the brightest, or may exhibit variation in intensity. There is also no justification to choose one type of thresholding over another. Thus, care must be used to ensure that the adopted threshold is not destructive to the feature we wish to study.

2.2.2 Mathematical Morphology

MM is a powerful tool for extracting and processing scientific information from image data since morphological operations relate directly to the shape of observed features. Here we introduce some morphological operations that are useful in extracting scientifically meaningful numerical features from images. A more detailed introduction to morphological analysis is given in the Appendix. More in depth coverage can be found in Serra (1982) and Soille (2003).

Dilation and Erosion—Dilation and erosion are the two fundamental operations in MM. They form a duality and they both use a structuring element (SE) Y to probe and alter the shapes of geometric structures inside an image I. The dilation of I by Y is the set of points z such that Y *hits* I when the origin of Y is placed at z. Therefore the dilation of I always enlarges I. The erosion of I by Y is defined as the set of points z such that Y *fits* wholly inside I when the origin of Y is at z. In contrary to dilation, erosion always shrinks I. For real-valued/greyscale images, the SE smoothes the three-dimensional image surface, with the height of the image surface at each pixel being equal to its intensity value.

Morphological Opening—A morphological opening operation involves, first, an erosion of the image with an SE, followed by a dilation with the same SE. Since after an erosion, only those features in the image that are morphologically similar to the SE are still present, this effectively enhances such features in the image and smoothes them from the interior. Opening also has a filtering effect: image structures that cannot completely contain the SE are removed from the image. A simple example of a morphological opening operation on a binary image is given in Figure 2.1(a).



(a) Opening of a set X by a disk B.

(b) Closing of a set X by a disk B.

Figure 2.1: Illustration of morphological opening and closing on binary images.

Morphological Closing—The opposite operation to opening is morphological closing, which smoothes features from their exterior. A closing operation is a dilation, followed by an erosion, which essentially smoothes out the image and fills in gaps without degrading or distorting the salient features, as would occur with normal boxcar or Gaussian smoothing. A simple example of a morphological closing operation on a binary image is given in Figure 2.1(b). In practice, choosing between morphological opening and closing depends on the features to be enhanced or type of noise to be removed.

Morphological Skeletonization—Skeletonization extracts the interior "skeletons" in extended regions; the locus of the points that form the skeleton traces out the spine of the region, yielding a sketch representation of the original features. They are the innermost possible pixels in the region, and are ideally suited to capture, for example, a simplified representation of coronal loops that can then be used to extract location/shape information.

Morphological Pruning—Morphological pruning removes the small offshoots that may exist in a morphological skeleton due to irregularities in the boundaries of the region. Such offshoots can be eliminated by first identifying the locations where the offshoots exist, then finding the lengths of such regions, and then eliminating all structures that are a few pixels long or smaller, to produce a cleaned skeleton that better represents the feature of interest.

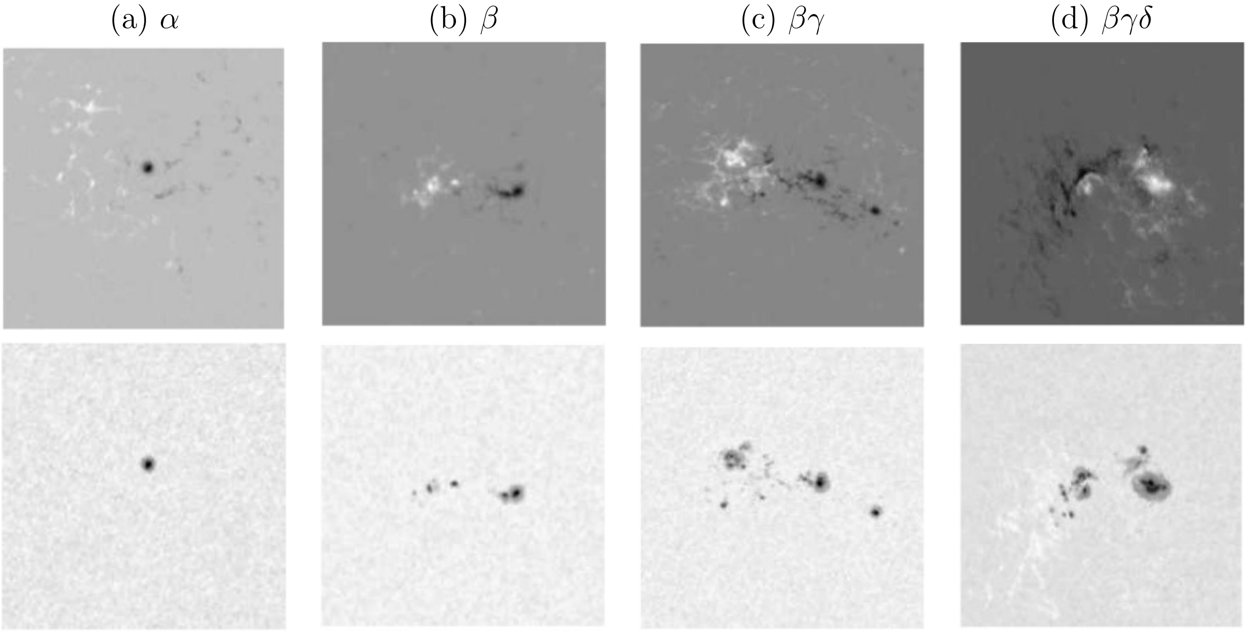


Figure 2.2: Examples of the four classes of sunspot groups used in the Mt. Wilson scheme, with magnetograms in the top row and white-light images in the bottom row. The α class (a) is dominated by a single unipolar sunspot that appears white or black in the magnetogram, depending on the polarity (positive or negative). The β class (b) has spots of both positive and negative polarity that can be separated by a single North-South polarity inversion line. The $\beta\gamma$ class (c) exhibits a complex distribution of polarities, and a single North-South polarity inversion line cannot cleanly divide the positive and negative regions of magnetic flux. In the $\beta\gamma\delta$ class (d), examination of the white-light image in conjunction with the magnetogram reveals umbrae of different polarity within a single enclosed penumbra.

2.3 Sunspot Classification

2.3.1 Mount Wilson Classification

The Mt. Wilson classification scheme groups sunspots into four broad classes based on the morphology of magnetically active regions as viewed in magnetogram images. Examples of the four classes appear in Figure 2.2. The simplest class morphologically is the α class, defined as a single *unipolar* sunspot—a single spot of either positive or negative polarity, which is often linked to a *plage* of opposite polarity. Plage is a diffuse network of magnetic fluxtube footpoints formed when magnetic field lines shooting outward from the photosphere scatter down over a wide area. For *bipolar* sunspot groups, spots of opposite magnetic polarity are visible in magnetogram images and multiple sunspots tend to be present in the white-light images, forming a *sunspot group.* The simplest bipolar class morphologically is the β class, which is a pair of sunspots of opposite magnetic polarity with a single *North-South polarity inversion line*—a simple and distinct linear spatial division oriented in the solar North-South direction—between the polarities. If a bipolar group is sufficiently complex that a single North-South polarity inversion line cannot divide the two polarities, then it is a $\beta\gamma$ sunspot group. If a $\beta\gamma$ group also contains *umbrae* of different polarity inside a single *penumbra*, which is known as a *delta spot*, then it is a $\beta\gamma\delta$ sunspot group. The umbra is the dark, inner part of the sunspot, and is surrounded by the slightly lighter penumbra as can be clearly seen in the white-light image (bottom row) of Figure 2.2(d).

Classification of sunspots is commonly performed through visual inspection by experts, and publicly available sunspot lists are manually determined. The Mt. Wilson scheme is popular because it is based on a simple and interpretable set of rules (as described above) and has some power to predict flares when combined with other solar data (Ireland *et al.*, 2008). However, while the classification rules are simple, the morphology of active regions is better described by a continuum rather than a discrete clustering. For example, the morphology of a particular active region may exist somewhere between a β group and a $\beta\gamma$ group and experts may disagree as to the "correct" classification. As a result, manual classification in general suffers from human observer bias stemming from the subjective and often ambiguous morphologies of active regions (Ireland *et al.*, 2008). A catalogue of sunspot identifications and classifications constructed manually is non-reproducible, which partly motivates our automated procedure.

2.3.2 Generating Numerical Summaries of Solar Active Regions

In this section we describe how MM can be used to extract scientifically meaningful and statistically useful numerical features. In particular, we detail our step-by-step procedure for generating numerical summaries of active region morphology using SoHO magnetogram and white-light images. In particular, we use the white-light images to obtain the general location of active regions in magnetograms to better differentiate between active region and plage network. We also calculate additional numerical summaries that characterize active region complexity that are of scientific interest in addition to serving as input covariates to statistical learning algorithms aimed at sunspot classification. Our general strategy is to obtain simple sketches (in the form of trinary images) of sunspot groups in white-light images and magnetically active regions in magnetograms. Then, we calculate numerical summaries from the sketches that summarize the morphology of magnetic flux distribution that are relevant to a sunspot group's classification and can therefore be used for statistical learning. Since these numerical summaries are based on the Mt. Wilson classification rules, they have a scientific basis and are interpretable to a solar physicist. In this way, we reduce complex images to real-valued numerical feature vectors that summarize the morphological characteristics of sunspot groups and associated active regions. Our general methodology is illustrated and summarized in a schematic form through Figures 2.3-2.5.

2.3.2.1 Sunspot and Active Region Identification

In the first part of our procedure, we use the white-light images to identify sunspots. This provides us with a general location of the active regions in magnetogram images, and helps distinguish the active regions from plage network. To do this, we first clean (i.e. smooth) the inverted white-light image (the image obtained by multiplying each pixel value by -1) with a morphological opening operation using a spherical SE with radius five. We use a round

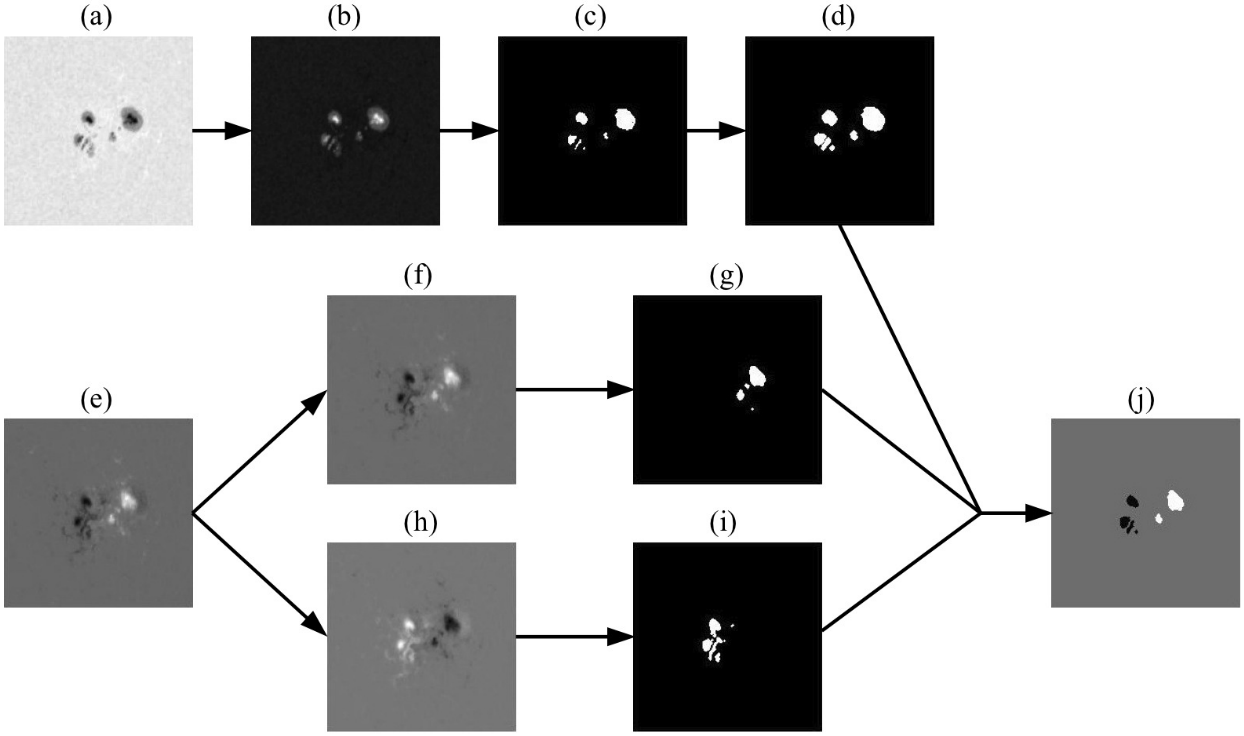


Figure 2.3: *Identifying Active Region Pixels.* *(a)* The raw β white-light image. *(b)* The inverted white-light image after applying a morphological opening operation using a spherical SE with radius 5. *(c)* The pixels belonging to the sunspot group identified by thresholding. *(d)* The sunspot area found by twice dilating the previous image using a disk shaped SE with radius 1. *(e)* The raw β magnetogram. *(f)* The magnetogram after applying a morphological opening operation using a spherical SE with radius 1. *(g)* The positive polarity active region pixels identified by thresholding. *(h)* The inverted magnetogram after applying a morphological opening operation using a spherical SE with radius 1. *(i)* The negative polarity active region pixels identified by thresholding. *(j)* The simple active region representation found by combining the positive and negative polarity active region pixels and excluding any pixels that are not also identified as part of the sunspot area in image (d).

SE because of the circular appearance of sunspots and active regions. The radius is chosen so that small structures are filtered out and larger round structures (i.e. the sunspots) are smoothed. Since at this point in our procedure we are only concerned with identifying the general location of sunspots in the white-light image, we are not concerned with possible destruction of features. Next, a thresholding operation is applied by setting pixels with values above $\bar{x} + 4s$ to one and the rest to zero, where $\bar{x}$ and s are, respectively, the mean and sample standard deviation of all the pixel values in the image. The resulting binary image is dilated twice using a disk-shaped SE of radius one, which slightly increases the total area of pixels with value one. The location of sunspots in the white light image are now identified by the pixels in the binary image with value one, which we call the *sunspot area*. This process for identifying the location of sunspots is demonstrated on a β sunspot group in Figure 2.3(a)-(d).

The second part of our procedure involves obtaining simple representations of active region morphology by identifying general regions of positive and negative polarity in the magnetogram. We first clean the magnetogram with a morphological opening operation using a spherical SE of radius one. This is the smallest radius we can choose, and it is preferable since we want to remove noise while preserving and only slightly smoothing the areas of identifiable magnetic flux. We want the morphological opening to smooth the areas of positive magnetic polarity that appear white in the magnetogram so that cleaner boundaries can be obtained via thresholding (now setting pixels with values greater than $\bar{x} + 3s$ to one and the rest to zero), which identifies the *positive polarity pixels*. The same operations are applied to the inverted magnetogram (obtained by multiplying each pixel value by -1) to obtain the *negative polarity pixels*, and since we want to distinguish the two polarities these pixels are given a value of two instead of one. Any positive or negative polarity pixels (hereafter *white pixels* and *black pixels*, respectively) that fall outside the sunspot area are deemed to be part of the plage network and are set to the background value of zero. The result is a trinary sketch detailing the morphology of magnetically active regions as seen in the magnetogram. Figure 2.3 demonstrates the entire process on the β active region, with arrows to illustrate how the images created at different steps of our procedure are utilized. Although the intensity information present in magnetograms is lost, this is not a detriment to our method. The type of sunspot is mainly determined by its morphological characteristics and not by the magnitude of the magnetic field. In particular, we focus on how well the magnetic flux is concentrated or scattered and how well the positive and negative polarities are mixed.

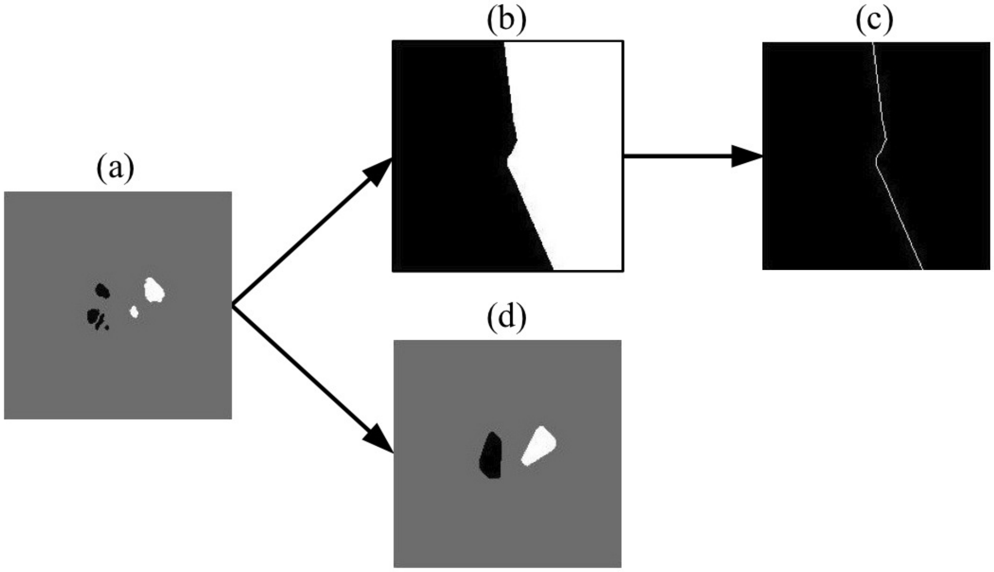


Figure 2.4: *Extracting Numerical Summaries of Active Regions. (a)* The simple active region representation obtained through the process demonstrated in Figure 2.3. *(b)* The separating boundary between regions of opposite magnetic polarity obtained via seeded region growing. *(c)* The polarity separating line obtained by removing interior and border pixels, followed by applying both a morphological opening (using a disk-shaped SE of radius one) and a morphological pruning to reduce jaggedness. *(d)* The simple active region representation in (a) after putting separate convex hulls around opposite polarity active region pixels.

2.3.2.2 Numerical Features Extraction

The trinary representation of the active region, Figure 2.3(j) and Figure 2.4(a), is our starting point for extracting numerical features that characterize the distribution of magnetic polarity. Since it is obviously important to distinguish between unipolar sunspots and bipolar sunspot groups, our first numerical feature is an extreme ratio of the number of extracted white and black pixels (N_{W} and N_{B}, respectively), denoted $|N_{\mathrm{W}}/N_{\mathrm{B}}|$. The ratio is expected to be close to one for bipolar groups (β, $\beta\gamma$, and $\beta\gamma\delta$) and close to zero for α spots.

Next, we can use the amount of scatter of the white and black pixels as an indicator of active region complexity. We do this by introducing a spatial complexity measure. In particular, let W be the set of white pixels. We then compute the center of mass, c, of W. For each pixel $w \in W$, the number of pixels that a line segment from w to c passes through is denoted $L(w)$

and of these, the number of white pixels is denoted $l(w)$. The spatial complexity measure, $A(W)$, is computed as

$$A(W) = 1 - \frac{1}{|W|} \sum_{w \in W} \frac{l(w)}{L(w)},$$

where $|W|$ is the number of pixels in W. Similarly, we compute $A(B)$ for the set of black pixels. In general, we expect higher spatial complexity for $\beta\gamma$ and $\beta\gamma\delta$ sunspot groups when compared to α sunspots and β sunspot groups.

The distinction between β sunspot groups and the other bipolar sunspot groups is the presence of a single distinct North-South polarity inversion line. By treating the white and black pixels as seeds in a standard seeded region growing operation (Adams and Bischof, 1994), we can produce a binary image that is segmented into regions of opposite polarity dominance, as in Figure 2.4(b). A separating boundary for regions of opposite polarity is obtained by setting all interior and border pixels to zero, as in Figure 2.4(c). A morphological opening operation using a disk-shaped SE of radius one is applied to reduce the jaggedness of the separating line, and a morphological pruning operation is applied to remove any offshoots. As sunspot groups evolve from β to $\beta\gamma$, the separating line between the two polarity regions is expected to become more complex as quantified by its curvature. We determine the amount of curvature by first tracing the separating line. The top-most pixel of the separating line is labeled as the starting point, then we look at all the neighboring pixels to find the next pixel that is part of the separating line. This process is repeated until the entire separating line has been traced, and the separating line pixel coordinates are recorded in order as (x_i, y_i), which denotes that the i^{th} pixel along the separating line is in row x and column y. Then, a quantity C_1 is computed as

$$C_1 = 1 - \frac{1}{N-1} \sum_{i=2}^{N} \frac{|x_i - x_1|}{\sqrt{(x_i - x_1)^2 + (y_i - y_1)^2}}.$$

This is a measure of curvature because the ratio inside the summation is one for a pixel lying on the 90° vertical line that intersects the starting point, so that taking the average value of this ratio for all the pixels along the separating line will yield a value of one for perfectly straight lines and smaller values as the amount of curvature increases. The average is subtracted from one so that higher values for C_1 are associated with higher curvature. We also compute C_2 in the same way as C_1, but the tracing of the separating line is initialized at its bottom-most pixel. Taking the average of C_1 and C_2 yields C, the amount of curvature of the separating line. If it happens that more than one separating line is found (as might happen with more complex sunspot groups), all the separating lines are traced and pixel coordinates are appended into a single list in order of separating line length.

The degree to which the areas of opposite magnetic polarity are "mixing" is another useful measure of active region complexity. To determine polarity mixture, a convex hull is placed separately around the sets of white and black pixels, as in Figure 2.4(d). The ratio of the number of pixels contained inside the intersection of the two hulls, N_{in}, to the total number of pixels in the area constrained by the two hulls (where the area of overlap is only counted once), N_{out}, is a quantification of the degree of mixture, denoted by $N_{\text{in}}/N_{\text{out}}$.

The last step in our numerical feature extraction routine is to identify the presence of delta spots that distinguish $\beta\gamma$ sunspot groups from $\beta\gamma\delta$ sunspot groups. To do this, we clean the inverted white-light image with a morphological operation using a spherical SE of radius one. We use a less destructive SE than in the first part of our procedure since we are concerned with distinguishing umbrae and penumbrae boundaries and do not wish to over-smooth these areas. Once we have a slightly smoothed image, we threshold to identify the *sunspot pixels*. We apply a second thresholding to the sunspot pixels to obtain *umbrae* pixels, and pixels that belong to the sunspot pixels set but not the set of umbrae pixels are designated *penumbrae pixels*. The procedure for reducing the raw white-light image to a simple representation of umbra and penumbra structure is demonstrated in Figure 2.5. Since we have already identified pixels corresponding to regions of opposite magnetic polarity in the magnetogram, as in Figure 2.3(j), we can examine the polarity of umbrae pixels within enclosed sets of penumbrae pixels and identify delta spots. We denote the number of delta spots detected for a particular sunspot group by N_{delta}. The size of the delta spots is also useful for distinguishing sunspots that are borderline between $\beta\gamma$ and $\beta\gamma\delta$. If delta spots are identified, then the total number of umbrae pixels associated with the delta spots is denoted S_{delta}. In cases where no delta spots are identified, both N_{delta} and S_{delta} are given values of zero.

A final caveat is that for certain unipolar sunspot groups in which only a single polarity is identified, C, $N_{\text{in}}/N_{\text{out}}$, and either $A(W)$ or $A(B)$ cannot be calculated. In these cases we assign the physically unrealistic value of negative one. We made a decision to form ratios etc., e.g. $|N_{\text{W}}/N_{\text{B}}|$ and $N_{\text{in}}/N_{\text{out}}$, prior to feeding them into a learning algorithm for classification because we want interpretable numerical features. The learning algorithm will essentially be treated as a black box that takes scientifically meaningful inputs and returns the Mt. Wilson classification.

2.3.3 Automatic Classification

Our numerical features, $|N_{\text{W}}/N_{\text{B}}|$, $A(W)$, $A(B)$, C, $N_{\text{in}}/N_{\text{out}}$, N_{delta}, and S_{delta}, can be used as covariates in supervised learning algorithms, using the manually determined labels. For the purposes of this chapter, we do not attempt to provide an optimal classifier but merely demonstrate the efficacy of our general techniques for reducing complex images into

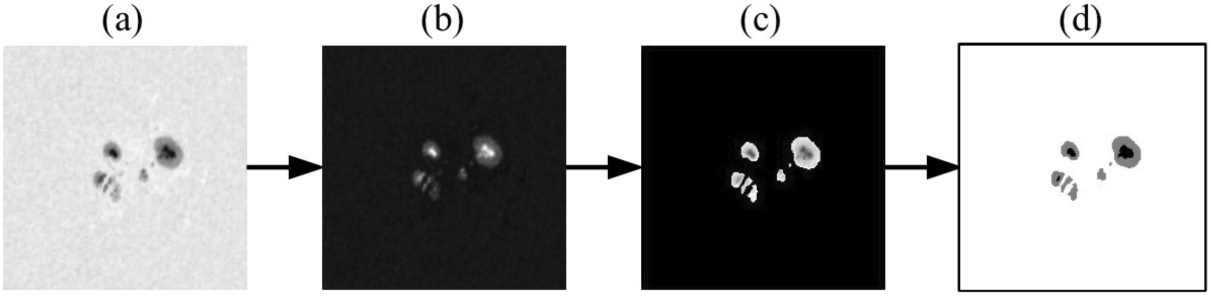


Figure 2.5: *Identifying Delta Spots (a)* The original β white-light image. *(b)* The inverted white-light image after applying a morphological opening operation using a spherical SE with radius one. *(c)* The pixels belonging to the sunspot group, with non-zero pixel values assigned by point-wise multiplication of the binary image obtained by thresholding image (b) and the smoothed white-light image. *(d)* A simple representation of the umbra and penumbra regions obtained by thresholding on only the non-zero pixels from image (c). This is used in conjunction with image (j) from Figure 2.3 to determine if there are umbrae with opposite polarity within a single enclosed penumbra, which is then identified as a delta spot.

scientifically meaningful and statistically useful numerical summaries. We use Breiman's standard *random forest* (Breiman, 2001)—a state-of-the-art nonparametric classifier that is an ensemble of individual decision trees—to exhibit how we can provide an automatic classification that is comparable to the manual classification. With N cases in the training set and p features, each tree of the random forest proceeds by sampling $n = N$ cases from the training set with replacement and randomly selecting $\sqrt{p}$ features to make a decision at each node. The desired number of trees are grown to completion and the classification of a new case is the majority vote of all the trees. Random forest is a sensible choice for a classifier since our features were tailored to make "if-else" type decisions (e.g. if $|N_{\mathrm{W}}/N_{\mathrm{B}}| < \epsilon$ then classify as α, else continue, for a value ϵ determined by the classifier). We implement the random forest classifier using the "randomForest" package in R. Two input variables are randomly selected as candidates for splitting at each node, and each individual tree is grown to completion. We use 1000 decision trees in total and the classifications are assigned based on majority vote.

We evaluate the effectiveness of our numerical features on a dataset consisting of 119 magnetogram and white-light image pairs taken by SoHO between May 1996 and December 1999.

		Manual Classification			
		α	β	$\beta\gamma$	$\beta\gamma\delta$
Automatic Classification	α	25	1	0	0
	β	2	63	5	0
	$\beta\gamma$	0	1	11	1
	$\beta\gamma\delta$	0	0	2	8

Table 2.1: Confusion matrix of the random forest predictions on out-of-bag data.

The dataset was constructed by choosing magnetograms displaying individual distinct active regions that have been manually classified according to the Mt. Wilson scheme by experts at the Space Weather Prediction Center (`http://www.swpc.noaa.gov/`). Since the training set for a particular tree in the random forest is a bootstrap sample drawn with replacement, the samples not used for that tree can be used as an out-of-bag test set for that tree. In this way, we can evaluate the random forest classifier based on the predictions on out-of-bag data as presented in Table 2.1. The manual and automatic classifications agree on most of the sunspot groups (the overall agreement rate is 90%), with disagreement most prominent for groups labelled manually as $\beta\gamma$. All disagreements between the automatic and manual classifiers are over adjacent classes, e.g., sunspot groups labeled manually as belonging to the $\beta\gamma$ class were placed by the automatic classifier into the β class or the $\beta\gamma\delta$ class, etc. There is a continuum of the complexity of active region morphology as we proceed from one class to the next as follows: $\alpha \to \beta \to \beta\gamma \to \beta\gamma\delta$. Since we are more concerned with capturing active region complexity in scientifically meaningful numerical summaries than exactly mimicking the manual classification, disagreements of this type are not particularly worrisome.

The true performance of any automatic classifier will be difficult to determine since the manual classification is prone to known biases and inconsistencies. We stress that exact agreement is not necessarily the "gold standard" when automating a manual classification that is both artificial and subjective. The true morphologies of active regions are continuous

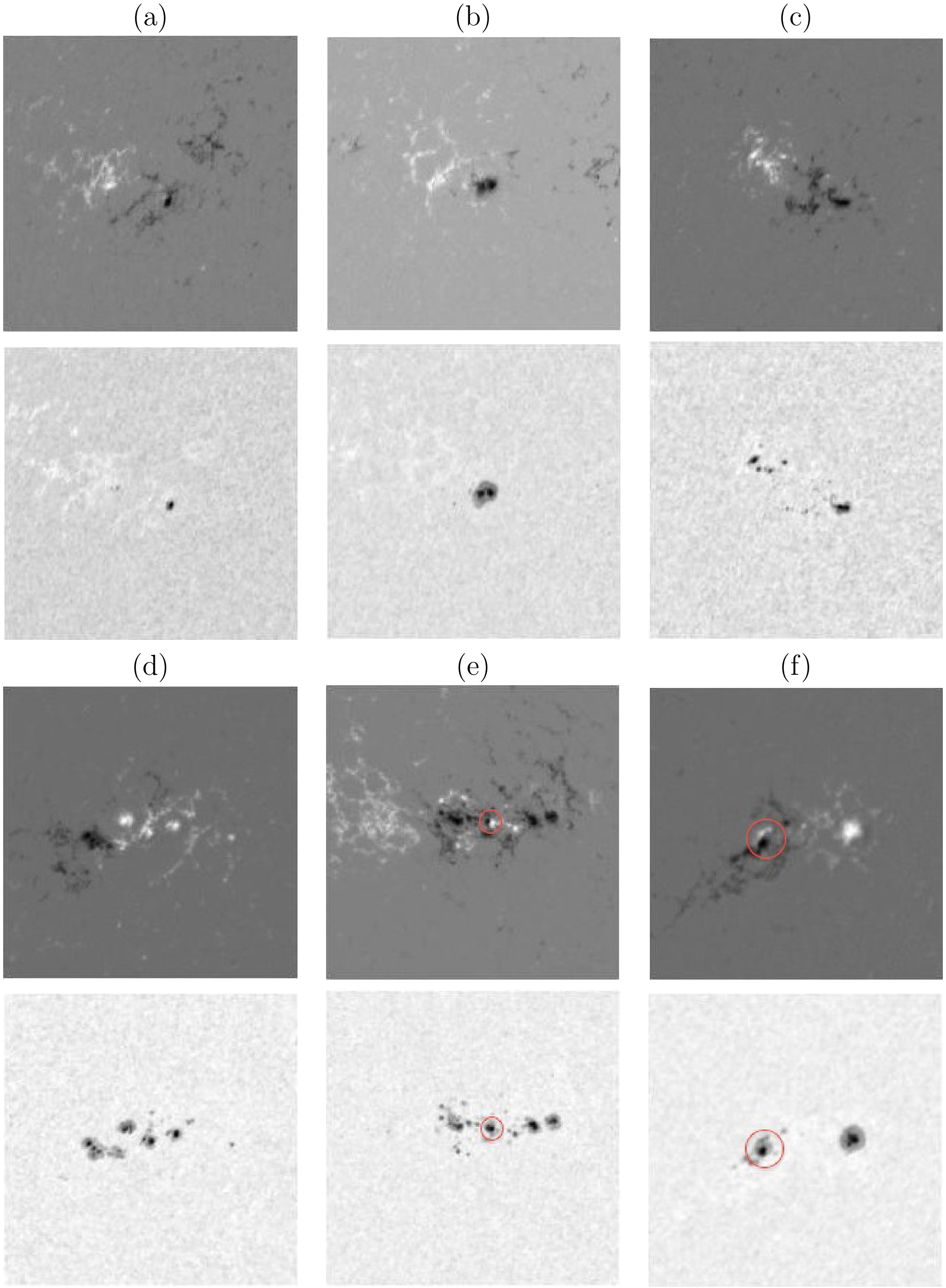


Figure 2.6: Examples of the disagreements between manual and automatic classification, with magnetograms in the top rows and white-light images in the bottom rows. Classifications are given as *manual/automatic*. *(a)* α/β. *(b)* β/α. *(c)* $\beta/\beta\gamma$. *(d)* $\beta\gamma/\beta$. *(e)* $\beta\gamma/\beta\gamma\delta$. *(f)* $\beta\gamma\delta/\beta\gamma$.

and sunspot groups can smoothly evolve from one class to another in short periods of time. As a result, there is often ambiguity as to the "true" classification of a particular sunspot group. In Figure 2.6, we present six of the sunspot groups for which there was a disagreement. The sunspot group in Figure 2.6(a) was classified manually as α and automatically as β, and appears intermediate between α and β since the negative polarity is not necessarily dominant. The sunspot group in Figure 2.6(b) was classified manually as β and automatically as α, but there appears to be a dominant (negative) polarity. Both Figure 2.6(a) and Figure 2.6(b) are intermediate between α and β, although Figure 2.6(b) is closer to unipolar despite Figure 2.6(a) having the α designation. A similar level of ambiguity exists between the sunspot groups in Figures 2.6(c) and 2.6(d), with Figure 2.6(c) having a manual classification of β and an automatic classification of $\beta\gamma$, and Figure 2.6(d) having a manual classification of $\beta\gamma$ and an automatic classification of β. Both examples are intermediate between β and $\beta\gamma$. The sunspot group in Figure 2.6(e) was classified manually as $\beta\gamma$ and automatically as $\beta\gamma\delta$. However, if we examine the circled areas in the magnetogram and white-light images, the umbrae appear to consist of both positive and negative polarities, which would indicate the presence of a δ spot. At the very least, the presence of the δ spot is ambiguous. We can contrast this with Figure 2.6(f), where the circled areas do not appear to reveal a δ spot. Since Figure 2.6(f) was classified manually as $\beta\gamma\delta$ but automatically as $\beta\gamma$, this demonstrates that tuning the automatic classifier to exactly mimic the manual classification is not desirable.

Manual classification schemes must rely on an artificial discretization of active region morphologies, but an automated procedure need not be likewise hindered. With the science-driven numerical features we have extracted from magnetogram and white-light images, it is possible to abandon discrete classification schemes in favor of a "classification" based on a continuum of classes. We expect this approach to better capture the complex evolutionary patterns of sunspots and allow for better prediction of solar flares and CMEs.

2.4 Discussion

The general procedures we have outlined in this chapter are designed to perform science-based feature extraction. While a variety of fields are producing high-quality digital image data, most image analysis techniques rely on algorithmic operations that are not tuned to the specific scientific information encoded in the images. We aim to purposefully extract the scientific content and use it to guide our choice of numerical features. For example, we believe there is useful information in the imperfectly performed manual sunspot classification and the Mt. Wilson scheme itself. The classes of the Mt. Wilson scheme were constructed to represent increasing complexity of magnetic field structure, but that complexity can also be captured with a set of numerical features as we demonstrated in this chapter. Our automatically produced Mt. Wilson classification agrees with manually assigned classifications on 107 out of 119 images in our dataset, with disagreements occurring over adjacent classes. We also highlight that while disagreements between manual and automatic classifications will inevitably occur, it is not advantageous to tune the automatic classification routine to exactly mimic the manual classifications due to the presence of ambiguous classes and human observer bias. Insofar as the manually obtained Mt. Wilson classifications encode information about the complexities of the Sun's magnetic field, we can obtain similar information using a self-consistent and reproducible automated method.

The fact that the extracted information is now contained in scientifically meaningful numerical feature vectors is also important because it allows for further downstream analysis. An example is to use the vectors we obtain for each sunspot group in conjunction with space-weather data to determine correlations between our numerical features and solar flares. We know that the Mt. Wilson scheme has some power to predict flares when combined with other space-weather data (Ireland *et al.*, 2008), and it would be interesting if any of the numerical features we have extracted are shown to be particularly effective for predicting solar flares. One approach would be to predict flares by again feeding the numerical features into a random forest, but this time predict whether or not a solar flare occurs in a given period of time. Random forests provide an easy method of determining which variables are the most important (see `http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#varimp` for a description), so the random forest could be run again using only the variables that are deemed important. If the accuracies are similar then we have determined which numerical features (assuming there is an acceptable level of accuracy to begin with) are useful for predicting solar flares. A researcher could also apply our methodology and compute additional numerical summaries that might be useful for different scientific investigations. An example of this might be to calculate the average gradient along the separating line in Figure 2.4(c) or the maximum intensity for the umbrae pixels using Figure 2.4(d) and the raw white-light image. We do not extract these numerical features since they are not directly tied to the Mt. Wilson scheme, but these or similar quantities may be useful when pursuing other goals and our methodology makes them easy to calculate.

Another possibility for further downstream analysis is to take advantage of the high cadence of SDO and apply our feature extraction routine to time series of images of evolving sunspots and active regions to produce vectors of features. Dynamic linear models (DLMs) can then be developed to model the evolving morphology of active regions. There exist DLMs for the solar cycle (e.g., Yu *et al.*, 2012; Stenning *et al.*, 2015), as well as new efficient algorithms for fitting DLMs based on interweaving data augmentations schemes (Simpson *et al.*, 2015). We can then explore clusters in the patterns of evolution of sunspots and active regions—existing classification schemes only consider static images. Prior to clustering, we could apply nonlinear dimension reduction (e.g., *diffusion map*) to reveal structure in both vectors of image features and in fitted DLM parameters, a technique used successfully to analyze star formation history in galaxies (Richards *et al.*, 2009). This would be a first step towards a new classification scheme that treats active regions as dynamic features and will aid in the prediction of space weather.

Chapter 3

A Bayesian Analysis of the Solar Cycle Using Multiple Proxy Variables

3.1 Introduction

The solar cycle was first noticed in the 11-year cyclic pattern in the average number of sunspots visible on the solar disk as viewed from Earth (Wolf, 1852). Since sunspots form when intense magnetic fields inhibit convection, they are linked to the overall magnetic activity of the Sun and have long been a valuable proxy for solar activity. This value is partly derived from the fact that sunspot numbers (SSNs) comprise the longest uninterrupted set of observations in astronomy, with records starting in the early seventeenth century and available as monthly estimates since 1749. Correlations with solar activity have been established in other proxies such as the 10.7cm flux (i.e. the solar radio flux per unit frequency with a wavelength of 10.7cm), solar flare numbers, sunspot areas, *etc.* (Hudson, 2007). Still older activity can be reconstructed using radiocarbon measurements that are dated using tree-ring data (Bonev *et al.*, 2004; Solanki *et al.*, 2004). Yet the SSNs are the baseline for establishing properties of the solar cycle, and predictions for future cycle maxima are generally based off the SSNs.

Although the oldest SSNs were collected 265 years ago, the data are surprisingly reliable. The sunspot number is $R = k(10g + s)$, where s is the number of individual spots, g is the number of sunspot groups, and k is a factor that corrects for systematic differences between instruments and observatories. Despite advances in technology and the advent of higher resolution images of the Sun, the historical values of R are not expected to have significantly higher uncertainty for two reasons: (1) sunspot size visibility, which affects our ability to see the faintest individual spots, is limited by atmospheric conditions and that limit was reached a long time ago and (2) sunspot groups are always counted as ten individual spots, regardless of the actual number of spots in the group.

Using this data, Waldmeier determined that, within a cycle, the time for SSNs to rise to maximum is less than the time to fall to minimum (Waldmeier, 1935). Other relations, such as the amplitide-period effect—the correlation between the duration of a cycle and the amplitude of the following cycle—were established using SSNs, and can be utilized to predict characteristics of future cycles (e.g., Hathaway *et al.*, 1994, 2002; Watari, 2008). Yu *et al.* (2012) analyzes the SSNs to empirically derive statistically meaningful correlations between several parameters that they use to describe the solar cycle as part of building a Bayesian multilevel model that accounts for uncertainties in both the average monthly sunspot numbers, and in predicting the characteristics of future cycles. Such correlations help constrain physical models of the solar dynamo that attempt to explain the general solar cycle (Schüssler, 2007).

The statistical model of Yu *et al.* (2012) uses a multilevel structure to capture complex patterns in the solar cycle. The first level of the model parameterizes the solar cycle using using cycle-specific parameters, and describes the distribution of the observed monthly average SSNs around the parameterized cycle. The second level of the model incorporates relationships between the parameters of consecutive cycles, resulting in a hidden Markov model that generates characteristics of cycle $i + 1$ given the characteristics of cycle i. While this model was initially fit using observed SSNs, the model can in principle be fit using other proxies that follow the same underlying solar cycle.

A plot of the estimates of monthly average SSNs going back to the mid-nineteenth century is presented in the top panel of Figure 3.1. (This is a subset of the data plotted in Figure 1.3.) The cyclic pattern of SSNs that led to the discovery of the solar cycle is clearly visible. The available data extends back to January, 1749, and is maintained and made available by the Solar Influences Data Analysis Center in Belgium (`http://sidc.oma.be`). The bottom two panels of Figure 3.1 present the available data for two additional proxies of solar activity: (1) monthly average total sunspot areas, extending back to May, 1874, also available from the Solar Influences Data Analysis Center and (2) the monthly average 10.7cm flux, which can be obtained from the National Oceanic and Atmospheric Administration's National Geophysical Data Center (`http://www.ngdc.noaa.gov/stp/solar/flux.html`). Although the SSNs, sunspot areas, and the 10.7cm flux have differing temporal coverage, from the period of overlap it is clear that all three of these proxies follow a similar underlying pattern, which is the solar cycle. There are differences, however, in the cycle properties implied by the proxies. For example, the sunspot areas appear to have shorter cycle lengths and less pronounced peaks, especially in the first few observed cycles. The proxies are nonetheless highly correlated and believed to be associated with a common underlying solar cycle. Our goal is to combine information from the multiple proxies into a single omnibus estimate of the underlying cycle.

There are physical explanations as to why these proxies are correlated; the total areas of sunspots will obviously depend on the number of sunspots, and there is evidence that the 10.7cm flux values are influenced by the magnetic fields associated with sunspots (Greenkorn, 2012). Because such proxies have varying temporal coverages and may have varying cadences,

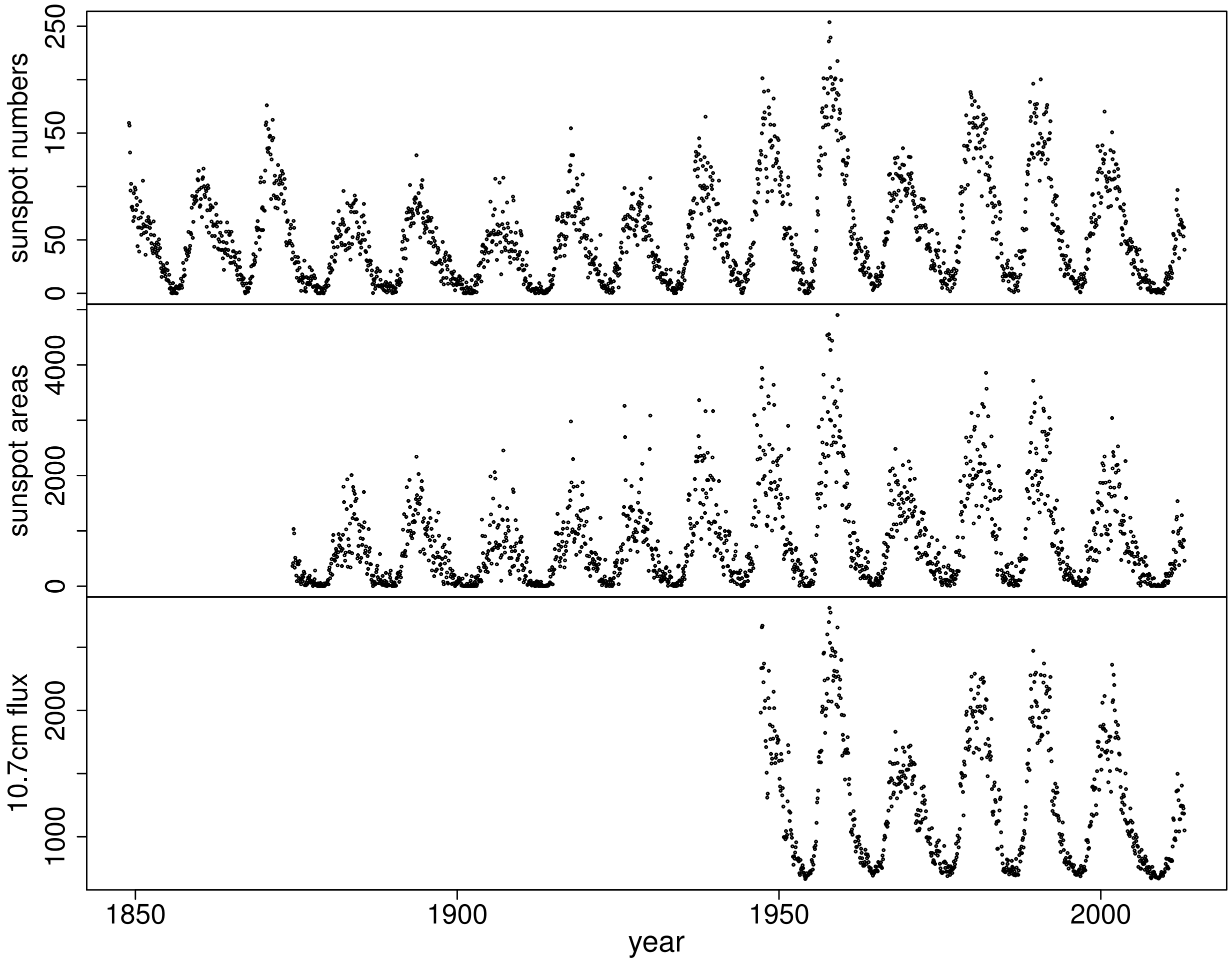


Figure 3.1: The observed proxies. *Top row*: monthly average sunspot numbers. *Middle row*: monthly average sunspot areas. *Bottom row*: monthly average 10.7cm flux. The roughly 11-year cycle of sunspot numbers follows the overall solar cycle, and we observe similar patterns in the sunspot areas and the 10.7cm flux.

it is challenging to combine them to model the underlying solar cycle in a sophisticated statistical analysis. Ideally, we would like to take advantage of the depth of high quality data that has become available in recent years, while incorporating the long record of SSNs.

Given the interest in predicting future solar maxima, diverse methodologies have been utilized including those based on i) models of the solar dynamo (e.g., Choudhuri, 1992; Charbonneau and Dikpati, 2000; Dikpati and Gilman, 2006; Choudhuri *et al.*, 2007; Charbonneau, 2007), ii) measurements of geomagnetic activity that act as precursors (e.g. Hathaway and Wilson, 2006), and iii) statistical analysis of historical and current data (e.g., Hathaway *et al.*, 1994; Benestad, 2005; Gil-Alana, 2009; Yu *et al.*, 2012), among others. An overview of the various predictions made for the current solar cycle, cycle 24, is given in Pesnell (2012). Within the solar physics community there has been debate over the amplitude of cycle 24, with different physical models producing a large range of predictions (e.g., Dikpati *et al.*, 2006; Choudhuri *et al.*, 2007). For the previous cycle, cycle 23, Kane (2001) notes that among twenty predictions of the smoothed maximum sunspot number made by different researchers, only eight were within an acceptable range of the later observed value. The Bayesian method of Yu *et al.* (2012) showed that there is considerable uncertainty in predicting an upcoming solar maximum using data up to the current solar minimum.

This chapter is divided into five sections. In Section 3.2 we review the Yu *et al.* (2012) Bayesian multilevel model for the solar cycle. In Section 3.3 we develop a systematic strategy for combining multiple proxies, and in particular describe how the pattern of missing data can be exploited to obtain coherent statistical inference. We present the results of fitting the solar cycle model to data that combines multiple proxies in Section 3.4. Finally, in Section 3.5 we summarize our results and discuss directions of future research.

3.2 Modeling the Solar Cycle with Sunspot Numbers

Yu *et al.* (2012) proposes a Bayesian multilevel model for fitting the solar cycle using monthly average SSNs as a proxy for the solar activity level. The solar cycle is parameterized with a set of cycle-specific parameters that together describe the total length, rising time, and amplitude for a given cycle. In the first level of the multilevel model, the observed SSNs are related to the parameterized solar cycle. The second level of the model incorporates a Markov structure that links parameters of consecutive cycles and encapsulates known features of the sunspot cycle in a series of sequential relations. These two stages are combined into a coherent statistical model, which is fit using Markov chain Monte Carlo methods. This structure allows for straightforward prediction of the characteristics of current cycles, even with data only extending to the beginning of the cycle. This is an important facet of the predictive capability of the Bayesian multilevel model.

3.2.1 Level One: Modeling the Cycles

Figure 3.2 illustrates our parameterized model for a single solar cycle. In it, for cycle i, suppose $t_0^{(i)}$ is the starting time, $t_{\max}^{(i)}$ is the time of the cycle maximum, $t_1^{(i)}$ is the end time, $c^{(i)}$ is the amplitude, and $U^{[t]}$ is the "average solar activity level" at time t. Here t is recorded in units of months; although the exact number varies, there are roughly $11 \times 12 = 132$ months per cycle. Under the Yu *et al.* (2012) model, the rising phase of the cycle is described as

$$U^{[t]} = c^{(i)} \left(1 - \left(\frac{t_{\max}^{(i)} - t}{t_{\max}^{(i)} - t_0^{(i)}} \right)^{\alpha_1} \right) \qquad \text{for } t < t_{\max}^{(i)}, \tag{3.1}$$

and the declining phase as

$$U^{[t]} = c^{(i)} \left(1 - \left(\frac{t - t_{\max}^{(i)}}{t_1^{(i)} - t_{\max}^{(i)}} \right)^{\alpha_2} \right) \qquad \text{for } t > t_{\max}^{(i)}, \tag{3.2}$$

where $\alpha_1, \alpha_2 > 1$ are shape parameters assumed to be constant for all cycles. Together, (3.1) and (3.2) parameterize the solar cycle, and the curve described by (3.1) and (3.2) is the curve in Figure 3.2. An important feature of this parameterization is that the starting point of the the next cycle, $t_0^{(i+1)}$, is not necessarily identical to the end point of the current cycle, $t_1^{(i)}$. When two cycles overlap (i.e. when $t_1^{(i)} > t_0^{(i+1)}$), the activity level, $U^{[t]}$, is given by the sum of the contributions of the form (3.2) and (3.1) from the respective cycles; when $t_1^{(i)} < t_0^{(i+1)}$, $U^{[t]} = 0$ for $t_1^{(i)} < t < t_0^{(i+1)}$.

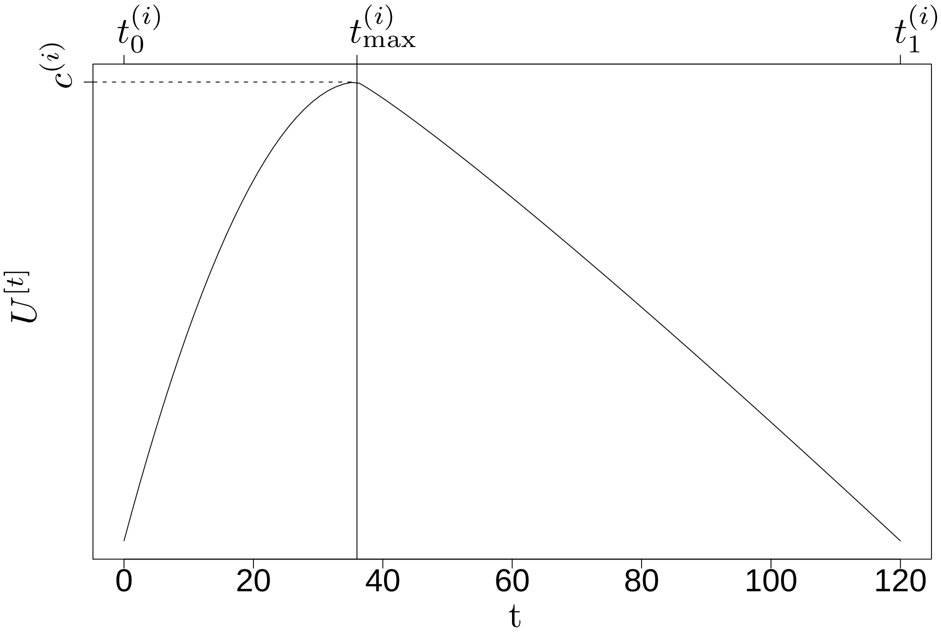


Figure 3.2: Parameterized form of a solar cycle. We illustrate $U^{[t]}$ with $c^{(i)} = 10, t_0^{(i)} = 0, t_{\max}^{(i)} = 36, t_1^{(i)} = 120,\ \alpha_1 = 1.9$ and $\alpha_2 = 1.1$, where $U^{[t]}$ is specified by (3.1) and (3.2).

The observed SSNs span a total of 25 cycles that are designated cycle 0 through cycle 24. The parameters that are specific to cycle i are the set $\theta^{(i)} = (t_0^{(i)}, t_{\max}^{(i)}, t_1^{(i)}, c^{(i)})$. The collection of cycle-specific parameters is contained in the set $\Theta = (\theta^{(0)}, \ldots, \theta^{(24)})$. Then, the full set of parameters that characterize the solar cycle is given by (Θ, α), where $\alpha = (\alpha_1, \alpha_2)$ does not vary from cycle to cycle and is therefore not included in Θ. With this, the distribution of the observed SSNs given Θ and the cycle-invariant parameters is

$$\sqrt{y}^{[t]} | (\Theta, \alpha, \beta, \sigma^2) \overset{\text{ind}}{\sim} N(\beta + U^{[t]},\ \sigma^2), \tag{3.3}$$

where $y^{[t]}$ is the monthly average SSN at time t, and the parameter β may be regarded as a baseline. The SSNs are modeled after a square-root transformation in order to stabilize the variance. The independence assumption in (3.3) is valid since sunspots disappear or rotate over the edge of the solar disk over timescales shorter than the observed monthly average SSNs. The independence assumption is not valid when analyzing daily fluctuations since the same sunspot or group of sunspots is counted every day until it vanishes. The decision to use monthly averages was partly motivated to avoid complex modeling of daily correlations.

3.2.2 Level Two: Relationships Between Consecutive Cycles

The evolution of the solar cycle is modeled via a Markov structure on the cycle-specific parameters $\{\theta^{(i)}, i = 0, \ldots, 24\}$. In particular, we model

$$p(\theta^{(0)}, \ldots, \theta^{(24)}|\eta) = p(\theta^{(0)}|\eta) \prod_{i=1}^{24} p(\theta^{(i)}|\theta^{(i-1)}, \eta), \tag{3.4}$$

where η is a set of hyper-parameters that we describe below. The distribution $p(\theta^{(i)}|\theta^{(i-1)}, \eta)$ is further factored in that we model each of the components in cycle i in their temporal order within the cycle. That is, we first model the cycle's start time, $t_0^{(i)}$, given the parameters of the previous cycle, $\theta^{(i-1)}$, then model its amplitude, $c^{(i)}$, given $t_0^{(i)}$ and $\theta^{(i-1)}$, then model the time at which it reaches maximum, $t_{\max}^{(i)}$, given $t_0^{(i)}$, $c^{(i)}$ and $\theta^{(i-1)}$, and finally model its end time, $t_1^{(i)}$, given $t_0^{(i)}$, $c^{(i)}$, $t_{\max}^{(i)}$ and $\theta^{(i-1)}$.

Beginning with the start time of cycle i, Yu *et al.* (2012) allowed the start time, $t_0^{(i)}$, to be different from, but dependent on, the end time of the previous cycle, $t_1^{(i-1)}$. Given $\theta^{(i-1)}$, $t_0^{(i)}$ is modeled as

$$t_0^{(i)} \mid t_1^{(i-1)} \sim t_1^{(i-1)} + N(0, \tau_0^2), \tag{3.5}$$

where the hyper-parameter τ_0^2 regulates the time difference between $t_0^{(i)}$ and $t_1^{(i-1)}$. The conditional distribution of $t_0^{(i)}$ depends on $\theta^{(i-1)}$ only through $t_1^{(i-1)}$.

To formulate $p(c^{(i)}, t_{\max}^{(i)}, t_1^{(i)}|t_0^{(i)}, \theta^{(i-1)}, \eta)$, Yu *et al.* (2012) conducted an exploratory analysis of the observed relationships among the parameters of consecutive cycles. In particular, they fit the model described by (3.3) to each of the 25 cycles individually and used the observed correlations among the cycle-specific fitted parameters to specify the parametric form of the distribution of $\theta^{(i)}$ given $\theta^{(i-1)}$. For example, there is a positive correlation between consecutive amplitudes $c^{(i)}$ and $c^{(i-1)}$, and a negative correlation between $c^{(i)}$ and $t_0^{(i)} - t_{\max}^{(i-1)}$. The predictive power of the positive correlation between consecutive amplitudes can therefore be enhanced by combining it with the negative correlation with $t_0^{(i)} - t_{\max}^{(i-1)}$. This means that given $\theta^{(i-1)}$, the amplitude of cycle i, $c^{(i)}$, depends on both $c^{(i-1)}$ and $t_0^{(i)} - t_{\max}^{(i-1)}$. Therefore, the distribution of $c^{(i)}$ given $\theta^{(i-1)}$ and $t_0^{(i)}$ is modeled as

$$c^{(i)} \mid (c^{(i-1)}, t_0^{(i)}, t_{\max}^{(i-1)}) \sim \delta_1 + \gamma_1 \frac{c^{(i-1)}}{t_0^{(i)} - t_{\max}^{(i-1)}} + N(0, \tau_1^2). \tag{3.6}$$

Additional correlations are observed among the components of $\theta^{(i)}$. For example, Yu *et al.* (2012) observe a negative correlation between the rising time of a cycle, $t_{\max}^{(i)} - t_0^{(i)}$, and the amplitude reached during the same cycle. This negative correlation was first discovered by Waldmeier (1935) and is hence known as the "Waldmeier effect." This effect means that the time at which cycle i reaches a peak, $t_{\max}^{(i)}$, is dependent on the starting time and the amplitude of that cycle. The distribution of $t_{\max}^{(i)}$ given $\theta^{(i-1)}$, $t_0^{(i)}$ and $c^{(i)}$ is thus modeled as

$$t_{\max}^{(i)} \mid (t_0^{(i)}, c^{(i)}) \sim t_0^{(i)} + \delta_2 + \gamma_2 c^{(i)} + N(0, \tau_2^2). \tag{3.7}$$

Notice that $t_{\max}^{(i)}$ is conditionally independent of $\theta^{(i-1)}$.

Finally, Yu *et al.* (2012) observe and incorporate a correlation between the amplitude, $c^{(i)}$, and the time-to-decline, $t_1^{(i)} - t_{\max}^{(i)}$, of that cycle. This means that $t_1^{(i)}$ depends on $c^{(i)}$ and $t_{\max}^{(i)}$. With this, the distribution of $t_1^{(i)}$ given $t_{\max}^{(i)}$ and $c^{(i)}$ is modeled as

$$t_1^{(i)} \mid (t_{\max}^{(i)}, c^{(i)}) \sim t_{\max}^{(i)} + \delta_3 + \gamma_3 c^{(i)} + N(0, \tau_3^2), \tag{3.8}$$

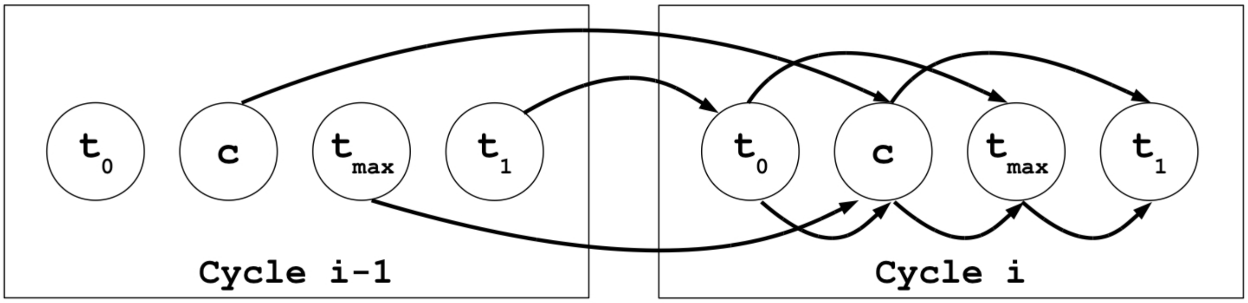


Figure 3.3: Markov structure relating the cycle-specific parameters $\theta^{(i-1)}$ and $\theta^{(i)}$.

and $t_1^{(i)}$ is conditionally independent of $\theta^{(i-1)}$.

The relations described by (3.5) to (3.8) can be encapsulated with the Markov structure illustrated in Figure 3.3. This structure allows for straightforward prediction of the characteristics of current cycles, even with data only extending to the beginning of the cycle. This is an important facet of the predictive capability of the Bayesian multilevel model. Together, (3.5) to (3.8) define the joint distribution of $p(\theta^{(i)}|\theta^{(i-1)}, \eta)$, where $\eta = (\tau_0^2, \gamma_j, \delta_j, \tau_j^2, j = 1, 2, 3)$ are the hyper-parameters. Yu *et al.* (2012) also examined correlations between non-adjacent cycles (*i.e.*, between cycle i and cycle $i \pm 2$). However, no evidence was found to suggest more than a lag-one dependence. This has important scientific implications since it suggests that the solar dynamo does not retain memory beyond one cycle. Further evidence for this property has been discovered by examining magnetic proxies of solar activity (Muñoz-Jaramillo *et al.*, 2013), and by computer simulation of the solar dynamo (Karak and Nandy, 2012).

3.2.3 Prior Distiributions

In (3.3), both β and $\log \sigma$ are given independent uniform prior distributions. To allow a wide range of cycle shapes, a uniform prior distribution on the interval $(1, 3)$ is used for both α_1 and α_2. The cycle-specific parameters for cycle 0, *i.e.* $t_0^{(0)}$, $t_1^{(0)}$, $t_{\max}^{(0)}$, and $c^{(0)}$ are assigned non-informative uniform prior distributions, subject to physical constraints on their ranges.

We consider two prior distributions on the hyper-parameters, η, namely $p(\gamma_j, \delta_j, \tau_j^2) \propto \frac{1}{\tau_j}$, $j = 1, 2, 3$, and $p(\gamma_j, \delta_j, \tau_j^2) \propto 1$, $j = 1, 2, 3$. In our numerical analyses, results are not sensitive to this choice and we therefore only report results obtained using $p(\gamma_j, \delta_j, \tau_j^2) \propto \frac{1}{\tau_j}$, $j = 1, 2, 3$.

3.3 Incorporating Multiple Proxies of Solar Activity

An inherent difficulty with combining multiple proxies to model the solar cycle is the varying temporal coverages of the proxies. SSNs, for example, are available as monthly estimates extending back to January, 1749, with no missing data. As technology improved, more proxies began to be observed and, like the SSNs, have been recorded up to the present. For clarity here and in our numerical illustrations, we consider three proxies: monthly average SSNs, monthly average total sunspot areas, and monthly average 10.7cm flux. Estimates of monthly average total sunspot areas extend back to May, 1874. Recordings of the 10.7cm flux began more recently, and estimates of the monthly average are available since February, 1947. It is important to note that, generally, once a proxy comes online (i.e., once a proxy begins to be recorded) it stays online and so there are no gaps in the data for individual proxies, resulting in a *monotone missing data pattern* (e.g., Little and Rubin, 2002). This pattern is readily apparent by examining the time series of the three proxies presented in Figure 3.1, and is described in detail in Section 3.3.2. Before we discuss our strategy for dealing with missing data, we describe how we would handle multiple proxies that were all observed over the same time period.

3.3.1 Complete-Data Analysis

With no missing data, we observe $Y^{[t]} = (y_1^{[t]}, y_2^{[t]}, y_3^{[t]})$ at each t, where y_1 is the monthly average SSN, y_2 is the monthly average sunspot area, and y_3 is the monthly average 10.7cm flux. Since the observed data represent monthly averages, $t = 1, \ldots, 3168$ indexes month. With this scheme, $t = 0$ corresponds to January, 1749, and $t = 3168$ corresponds to December, 2012.

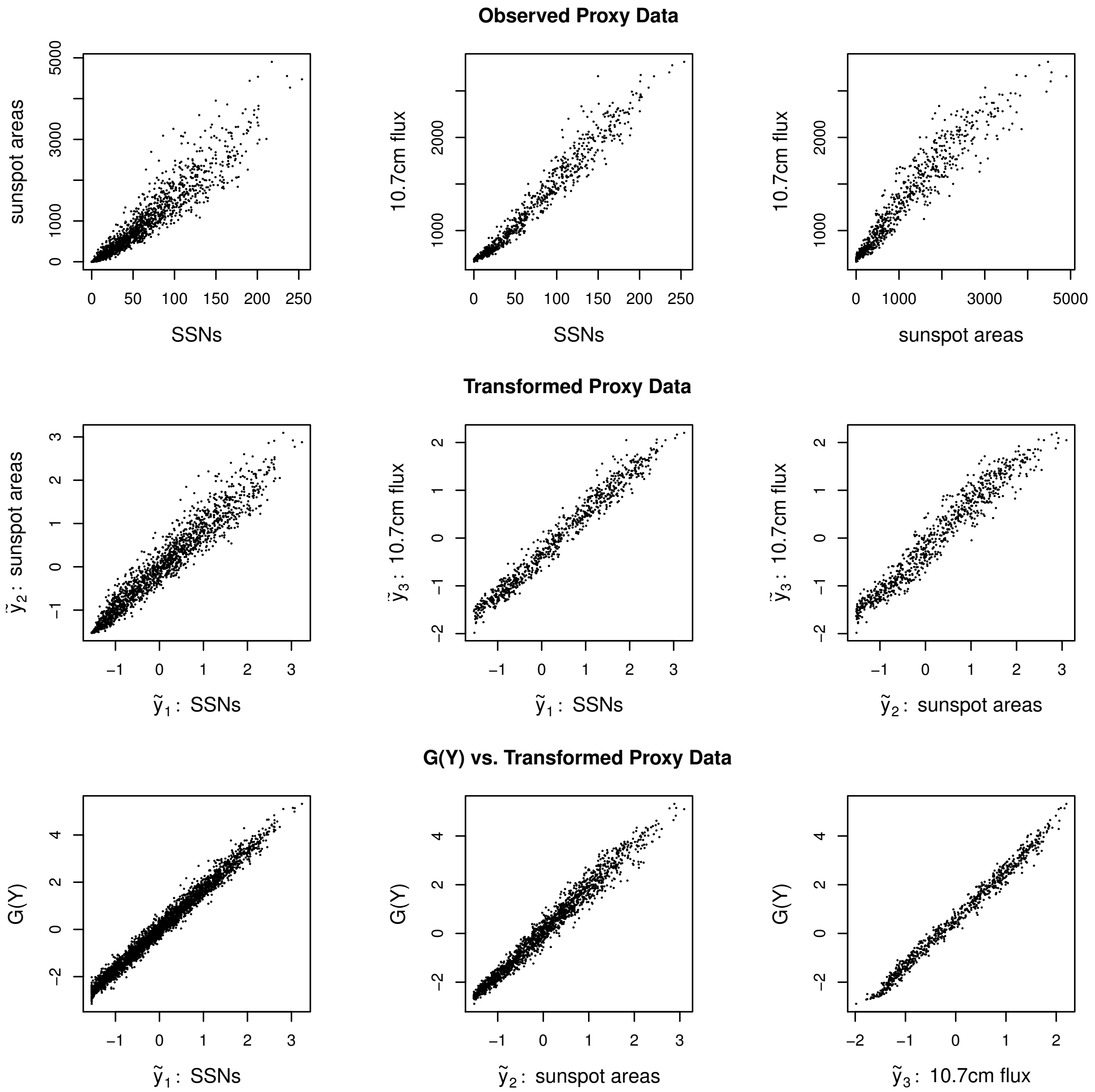
The distribution of $\sqrt{y}^{[t]}$ in (3.3) can be used to model any proxy, perhaps transformed, that follows the same underlying solar cycle. In this way, (3.3) can be generalized to

$$G(Y^{[t]}) \mid (\Theta, \alpha, \beta, \sigma^2) \overset{\text{ind}}{\sim} N(\beta + U^{[t]},\ \sigma^2), \tag{3.9}$$

where $G(Y^{[t]})$ is a mapping from the multivariate $Y^{[t]}$ to a scalar value, and the underlying parameters are modeled in the same way as in Sections 3.2.2 and 3.2.3. Yu *et al.* (2012) used $G(Y^{[t]}) = \sqrt{y_1}^{[t]}$ to obtain (3.3), but we are interested in finding a $G(Y^{[t]})$ that incorporates information from all available proxies.

The top row of Figure 3.4 displays scatterplots of the observed proxies and illustrates their strong linear correlations. Most of the variability in the data is in one linear dimension. Thus, it is appropriate to employ principal component analysis (PCA) to project the multivariate data onto the one-dimensional manifold defined by the direction of maximum variance.

Before deploying PCA, we transform the proxies to stabilize their variances, which all increase with their mean, see Figure 3.4. For the SSNs and sunspot areas we use the transformation $\sqrt{y_j^{[t]} + 10}$ for $j = 1, 2$. For the 10.7cm flux, we apply the transformation $\sqrt{y_3^{[t]} - \min_t(y_3^{[t]})}$. The constant offsets in the transformations were empirically chosen to improve linearity. We also normalize each transformed proxy by subtracting off its mean and dividing by its standard deviation. Subtracting off the mean is a necessary step in performing PCA,


Observed Proxy Data

Transformed Proxy Data

G(Y) vs. Transformed Proxy Data

Figure 3.4: Transforming the proxies. The top row displays the two-dimensional scatterplots of the observed proxy data, Y. The middle row displays the two-dimensional scatterplots of the transformed proxy data, $\tilde{Y}$. The bottom row displays the computed value of $G(Y) = \omega^T\tilde{Y}$ versus each of the transformed proxies. Notice that $G(Y)$ is highly correlated with each of the transformed proxies.

and dividing by the standard deviation controls for the differences in scale between the proxies. We denote the values of the transformed and normalized proxies at time t by

$\tilde{Y}^{[t]} = (\tilde{y}_1^{[t]}, \tilde{y}_2^{[t]}, \tilde{y}_3^{[t]})$. Scatterplots of the transformed proxy data, $\tilde{Y}$, are displayed in the second row of Figure 3.4. From these plots we note that the relationships remain linear, the correlations remain strong, but the variances are more stable.

With ω denoting the weights associated with the first principal component of the $\tilde{Y}^{[t]}$, we let $G(Y^{[t]}) = \omega^T \tilde{Y}^{[t]}$, where $G(Y^{[t]})$ is a scalar value representing the "solar activity level" at time t. The bottom row of Figure 3.4 shows that $G(Y)$ is highly correlated with each of the transformed proxies, which demonstrates its efficacy as a representation of the overall solar activity level. The first principal component accounts for 98% of the total variability in the proxies, so little information is lost in the PCA-based dimension reduction.

Once the $G(Y^{[t]})$ are obtained they are treated as observed data with distribution given in (3.9). We use a Gibbs sampler to explore the posterior distribution, incorporating Metropolis-Hastings updates as necessary (see Yu *et al.*, 2012). First, however, we must devise a scheme for handling missing observations of sunspot areas and the 10.7cm flux.

3.3.2 Multiple Imputation Strategy for Missing Data

The monthly average SSNs contain no missing data and are observed for all $\tilde{y}_1^{[t]}$, $t = 1, \ldots, 3168$. Records for the monthly average sunspot areas begin at month $t = 1505$ and contain 1664 observations, and records for the monthly average 10.7cm flux begin at month $t = 2378$ and contain 791 observations. Therefore, the SSNs are observed whenever the sunspot areas are observed, and both the SSNs and sunspot areas are observed whenever the 10.7cm flux is observed. This monotone missing data pattern (see Figure 3.5) allows the development of a straightforward strategy to account for missing data.

Let $\tilde{Y}_{\text{mis}}$ be the missing data, $\tilde{Y}_{\text{mis}} = \{\tilde{y}_2^{[t]}, t = 1, \ldots, 1504;\ \tilde{y}_3^{[t]}, t = 1, \ldots, 2377\}$, and $\tilde{Y}_{\text{obs}}$ be the observed data. A fully Bayesian strategy for handling the missing data would base

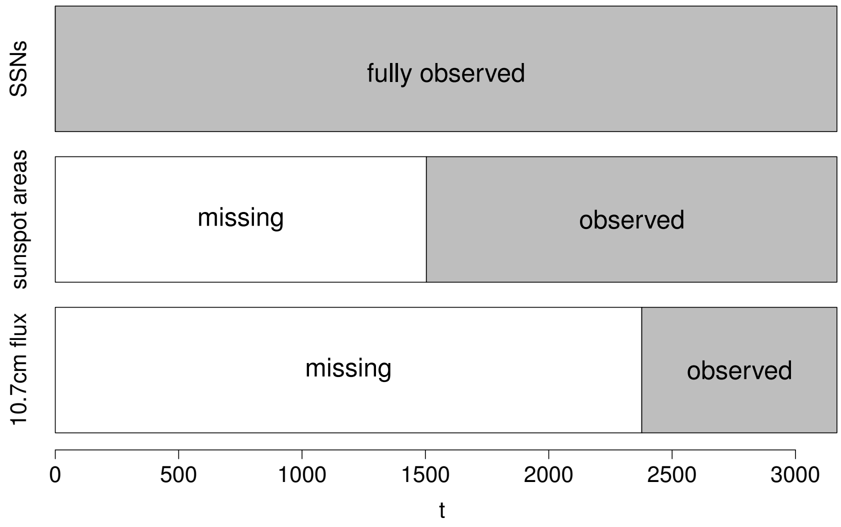


Figure 3.5: Illustration of the monotone missing data pattern. The solid gray bars indicate the range of observation times t for each of the three solar activity proxies: SSNs, sunspot areas, and 10.7cm flux. White bars indicate the range for which observations are missing. The SSNs are fully observed for months $t = 1, \ldots, 3168$. The sunspot areas are missing for months $t = 1, \ldots, 1504$. The 10.7cm flux is missing for months $t = 1, \ldots, 2377$. SSNs are observed whenever sunspot areas are observed, and both SSNs and sunspot areas are observed whenever the 10.7cm flux is observed. This is a monotone pattern of missing data.

inference on $p(\tilde{Y}_{\text{mis}}, \Theta, \alpha, \beta, \sigma, \eta \mid \tilde{Y}_{\text{obs}})$. This could be done by constructing a Gibbs sampler that at each iteration first updates $\tilde{Y}_{\text{mis}} \sim p(\tilde{Y}_{\text{mis}} \mid \tilde{Y}_{\text{obs}}, \Theta, \alpha, \beta, \sigma, \eta)$ and the updates $(\Theta, \alpha, \beta, \sigma, \eta) \sim p(\Theta, \alpha, \beta, \sigma, \eta \mid \tilde{Y}_{\text{mis}}, \tilde{Y}_{\text{obs}})$ using existing computer code. This would, however, require us to specify a model for the multivariate $\tilde{Y}$ rather than simply for the univariate $G(Y)$.

Luckily, multiple imputation (e.g., Little and Rubin, 2002) provides a principled way to use the univariate model and existing computer code to infer the solar cycle using multiple proxies. We first specify a separate simple local missing data model $p(\tilde{Y}_{\text{mis}}|\tilde{Y}_{\text{obs}}, \phi, \zeta)$ that incorporates the Markovian structure inherent in the data . In particular, we model

$$\tilde{y}_2^{[t]} \mid (\tilde{y}_1^{[t]}, \tilde{y}_2^{[t+1]}) \sim N(\phi_{01} + \phi_{11}\tilde{y}_1^{[t]} + \phi_{21}\tilde{y}_2^{[t+1]},\ \zeta_1) \tag{3.10}$$

for $t = 1, \ldots, 1504$, and

$$\tilde{y}_3^{[t]} \mid (\tilde{y}_1^{[t]}, \tilde{y}_2^{[t]}, \tilde{y}_3^{[t+1]}) \sim N(\phi_{02} + \phi_{12}\tilde{y}_1^{[t]} + \phi_{22}\tilde{y}_2^{[t]} + \phi_{32}\tilde{y}_3^{[t+1]},\ \zeta_2) \tag{3.11}$$

for $t = 1, \ldots, 2377$. Together, (3.10) and (3.11) define the distribution of $p(\tilde{Y}_{\text{mis}} \mid \tilde{Y}_{\text{obs}}, \phi, \zeta)$. We fit (3.10) using only the observations for which both $\tilde{y}_1$ and $\tilde{y}_2$ are observed, and fit (3.11) using only the observations for which all three quantities are observed. With the fitted models in place, $\tilde{Y}_{\text{mis}}$ can be imputed by drawing values from $p(\tilde{Y}_{\text{mis}} \mid \tilde{Y}_{\text{obs}}, \hat{\phi}, \hat{\zeta})$.

Multiple imputation coherently accounts for two sources of uncertainty: the uncertainty that would be present even if all proxies were observed for the same time period (*i.e.*, if there was no missing data), and the uncertainty that arises from imputing missing data. With multiple imputation, we obtain M imputations of the missing data, $\tilde{Y}_{\text{mis}}^{(m)} \sim p(\tilde{Y}_{\text{mis}} \mid \tilde{Y}_{\text{obs}}, \hat{\phi}, \hat{\zeta})$ and apply the complete-data analysis described in Section 3.3.1 to each of the imputed data sets. Parameter estimates and uncertainties are based on the multiple imputation combining rules (e.g., Harel and Zhou, 2007; Little and Rubin, 2002). Under this scheme, we obtain M estimates, $\hat{\psi}_m$, of any particular model parameter, ψ, along with their associated variances, V_m. Since we sample from the posterior distribution to fit the Bayesian multilevel model described in Section 3.2, natural candidates for $\hat{\psi}_m$ and V_m are the posterior means and posterior variances of ψ, under each of the M imputed data sets. The multiple-imputation estimate of ψ is $\hat{\psi} = \frac{1}{M}\sum_{m=1}^{M}\hat{\psi}_m$. The estimate of the variance is a combination of the average within-imputation variance, $W = \frac{1}{M}\sum_{m=1}^{M} V_m$, and the between imputation variance, $B = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\psi}_m - \hat{\psi})^2$, and is given by $T = W + \frac{M+1}{M}B$. Interval estimates are computed from a reference t-distribution, $(\psi - \hat{\psi})T^{-1/2} \sim t_\nu$, where the degrees of freedom is given by $\nu = (M-1)\left(1 + \frac{M}{M+1}\frac{W}{B}\right)^2$.

Our multiple imputation procedure relies upon the assumption that ζ_1 and ζ_2 do not vary over time. To test this assumption, we fit the local missing data model for different time

	Newest Third $\hat{\zeta}_l$ (S.E.)	Oldest Third $\hat{\zeta}_l$ (S.E.)	F-test p-value
Imputing Sunspot Areas: $\tilde{y}_2$	3.88 (0.14)	3.17 (0.11)	< 0.01
Imputing 10.7cm Flux: $\tilde{y}_3$	1.37 (0.07)	1.34 (0.07)	0.74

Table 3.1: Examining the multiple imputation assumptions. The first column displays the estimates of ζ_1 and ζ_2, along with their estimated standard errors (S.E.), using the *newest third* of the observed data for the sunspot areas and 10.7cm flux, respectively. The second column displays the same, but using the *oldest third* of the observed data for the sunspot areas and 10.7cm flux, respectively. We use an F-test of equality of variances with the null hypothesis that ζ_l for the oldest time epoch is equal to ζ_l for the newest time epoch, $l = 1, 2$. The resulting p-values are given in the third column.

periods of data separately and compared the results. We first fit (3.10) using the most recent third of the available $\tilde{y}_2$ data and obtain an estimate and standard error of ζ_1. We then fit (3.10) again using the oldest third of the available $\tilde{y}_2$ data and obtain a second estimate and standard error of ζ_1. We then perform an F-test for the hypothesis that ζ_1^2 for the older time epoch is equal to ζ_1^2 for the newer time epoch. We repeat this procedure for (3.11) using the oldest and most recent third of the $\tilde{y}_3$ data. The results are summarized in Table 3.1. We find that ζ_1 appears to change over time, but that it is larger for the newer data. Since most of $\tilde{y}_2$ is observed, not imputed, we are not particularly worried about overestimating ζ_1 for the older data. In addition, ζ_2 does not appear to vary over time, and this is more important since, unlike $\tilde{y}_2$, the majority of the values of $\tilde{y}_3$ that go into our final analysis are imputed.

3.4 Results

We now discuss the fit of the Bayesian multilevel model for the solar cycle. To allow for comparison we obtain model fits using both $G(Y^{[t]}) = \sqrt{y_1}^{[t]}$ (*i.e.*, the SSN model) and $G(Y^{[t]}) = \omega^T \tilde{Y}^{[t]}$ (*i.e.*, the multiple-proxy model). When $G(Y^{[t]}) = \sqrt{y_1}^{[t]}$, we do not need to perform multiple imputation since there are no missing SSNs. In this case, the fitted values of all quantities are given by their posterior means and credible intervals are given by their 2.5% and 97.5% posterior quantiles. When $G(Y^{[t]}) = \omega^T \tilde{Y}^{[t]}$, we use the multiple imputation strategy described in Section 3.3.2, with $M = 5$ imputations. The estimates for all quantities are computed by setting $\hat{\psi}_m$ and V_m equal to their posterior mean and posterior variances, respectively, and following the multiple imputation combining rules. In particular, the estimate for the solar activity level at time t is computed by setting $\hat{\psi}_m$ and V_m equal to the posterior mean and posterior variance of $U^{[t]} + \beta$. The fitted values of $U^{[t]} + \beta$ are given by the average value of the M within-imputation posterior means and a 95% interval at time t is computed from the reference t-distribution as described in Section 3.3.2.

The fitted values of $U^{[t]} + \beta$ and associated 95% intervals for the multiple-proxy model are plotted in the top panel of Figure 3.6, for the time interval with all three proxies observed (*i.e.*, $t = 2378, \ldots, 3168$). The data presented in this panel are the $\omega^T \tilde{Y}$ values for the given time interval. The bottom panel of Figure 3.6 shows the fitted values of $U^{[t]} + \beta$ and associated 95% intervals under the SSN model for the same time interval; the data are the observed $\sqrt{y_1}^{[t]}$ values. The solid vertical lines in both panels are the fitted values for $t_{\text{max}}^{(i)}$ for cycles $i = 18, \ldots, 23$, and the dashed vertical lines are their 95% intervals. The estimates of $t_{\text{max}}^{(i)}$ under the multiple-proxy model are later than the estimates under the SSN model, although in some cases their 95% intervals overlap. We discuss further comparisons of the timing of the fitted cycles below.

To evaluate the quality of the model fits we plotted the residuals versus time and versus the fitted values, but did not observe any patterns that might call the models into question. The plot of the residuals versus time for the SSN model did not show any significant patterns or evidence of heteroscedasticity that would lead us to question the reliability of the historical sunspot numbers. We also simulate the full time series from the posterior predictive distribution 5000 times. Figure 3.7 displays the 95% posterior predictive intervals from the simulated series along with the $\omega^T \tilde{Y}$ values for the time period when all three proxies are

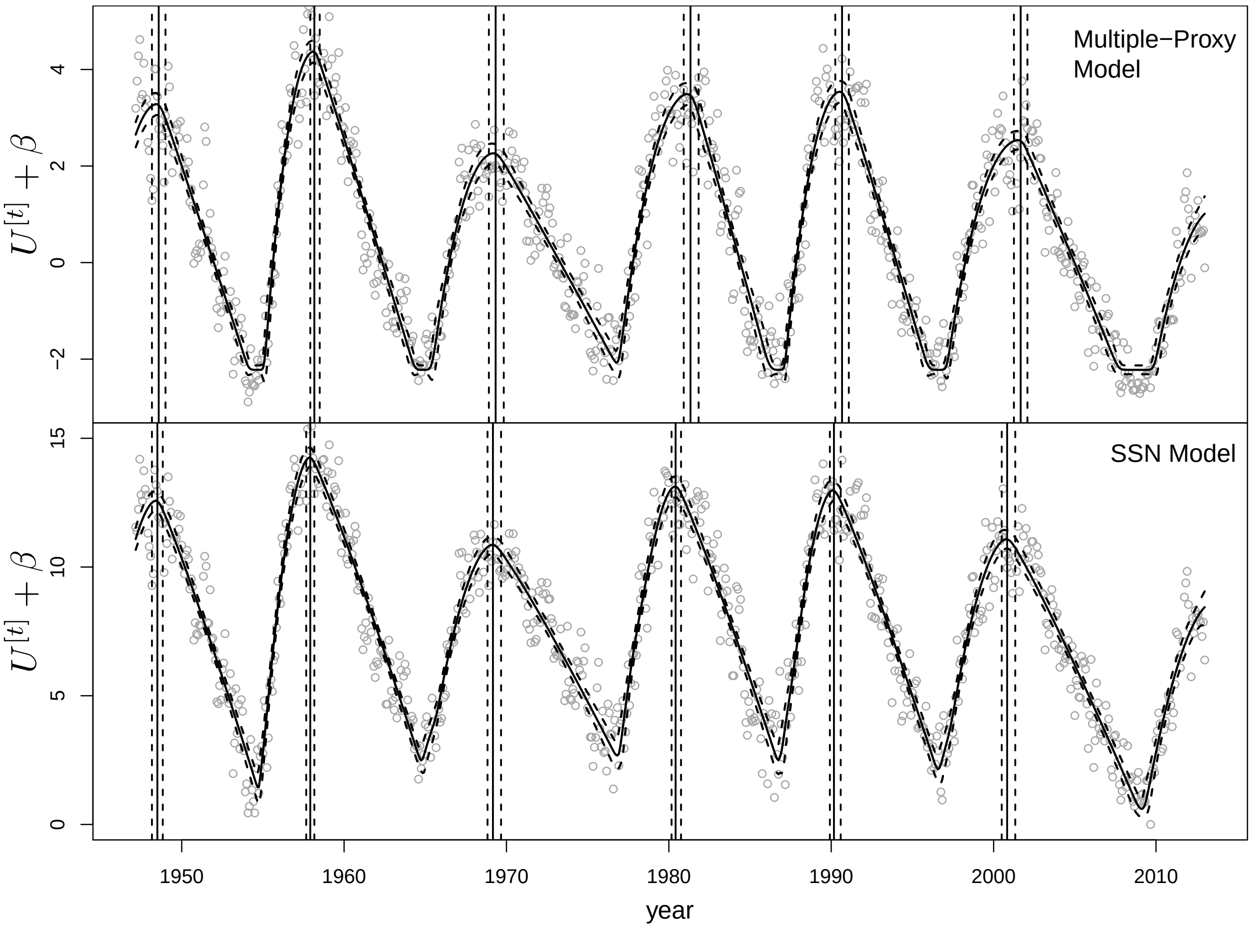


Figure 3.6: The fitted solar cycle. The two panels compare the fitted models and data for the multi-proxy (top) and SSN (bottom) models. They include $U^{[t]} + \beta$ (solid curves) and their 95% intervals (dashed curves), along with fitted values for $t_{\max}^{(i)}$ (solid vertical lines) and their 95% intervals (dashed vertical lines). Gray circles represent $\omega^T \tilde{Y}$ (top panel) and $\sqrt{y_1}^{[t]}$ (bottom panel). The time interval displayed covers the period when all three proxies are observed. Using multiple proxies consistently results in later fitted times for the solar maxima.

observed. The simulated series are consistent with the observed data.

Since solar physicists are concerned with predicting the timing of solar cycles, we can obtain fitted values and 95% intervals for the rising time of each cycle, $t_{\max}^{(i)} - t_0^{(i)}$, $i = 0, \ldots, 24$, under both the SSN and multiple-proxy models. These are presented in Figure 3.8. The left panel displays the fitted values and 95% intervals for both model fits over time, and

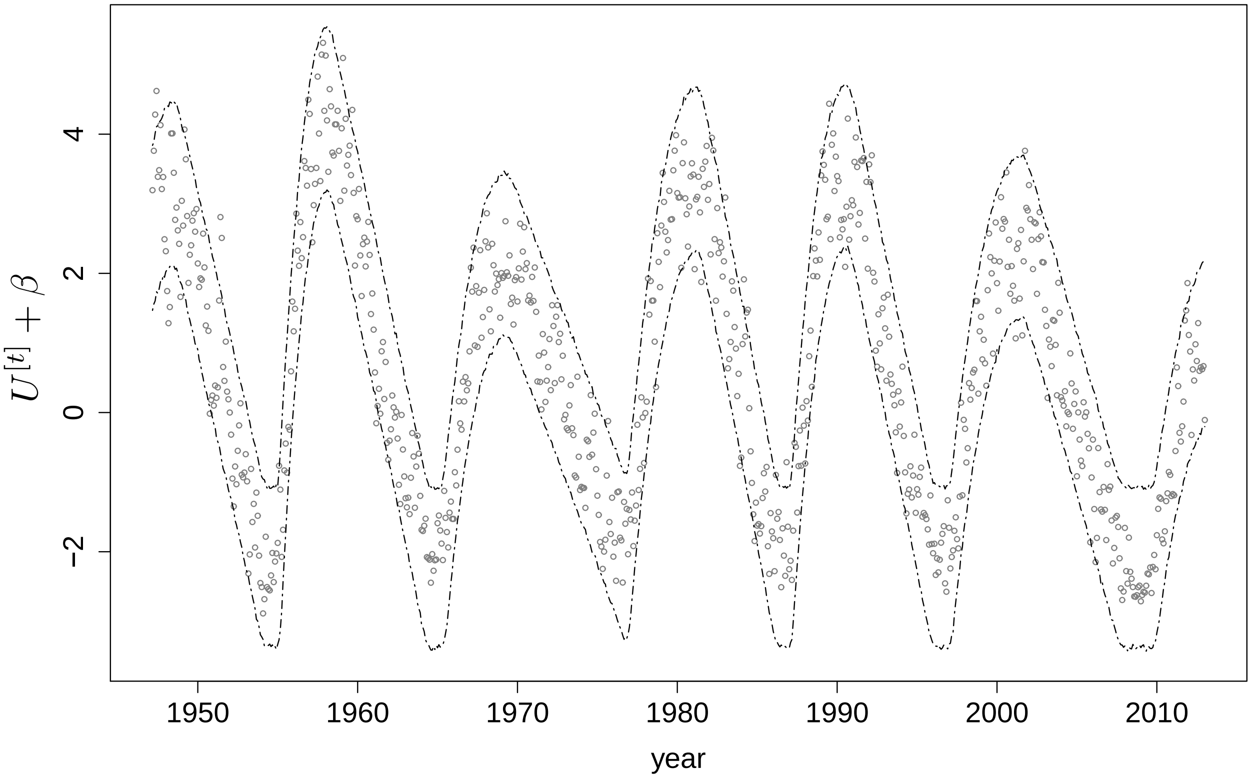


Figure 3.7: Posterior predictive check. The full time series is simulated from the posterior predictive distribution 5000 times, and 95% pointwise posterior predictive intervals are given by the dashed lines. Gray circles represent the observed $G(Y) = \omega^T \tilde{Y}$. The observed data are consistent with the posterior predictive distribution.

the right panel displays a scatterplot of the fitted values under the two models along with their associated 95% intervals. We do not include results for cycle 0 since this initial cycle is incomplete and has relatively large 95% intervals. The 95% intervals for cycle 24 are also larger than those of other cycles since the cycle is ongoing. That cycle 0 and 24 have fewer observed neighboring cycles than the other cycles also contributes to their larger 95% intervals.

Overall, the rising times do not appear to differ significantly between the two model fits. This is not the case when examining the falling times of each cycle, $t_1^{(i)} - t_{\max}^{(i)}$, which are displayed in Figure 3.9. They are significantly shorter for the model fit with multiple proxies since the 95% intervals rarely intersect the 45° line plotted in the right panel. Taken together, these

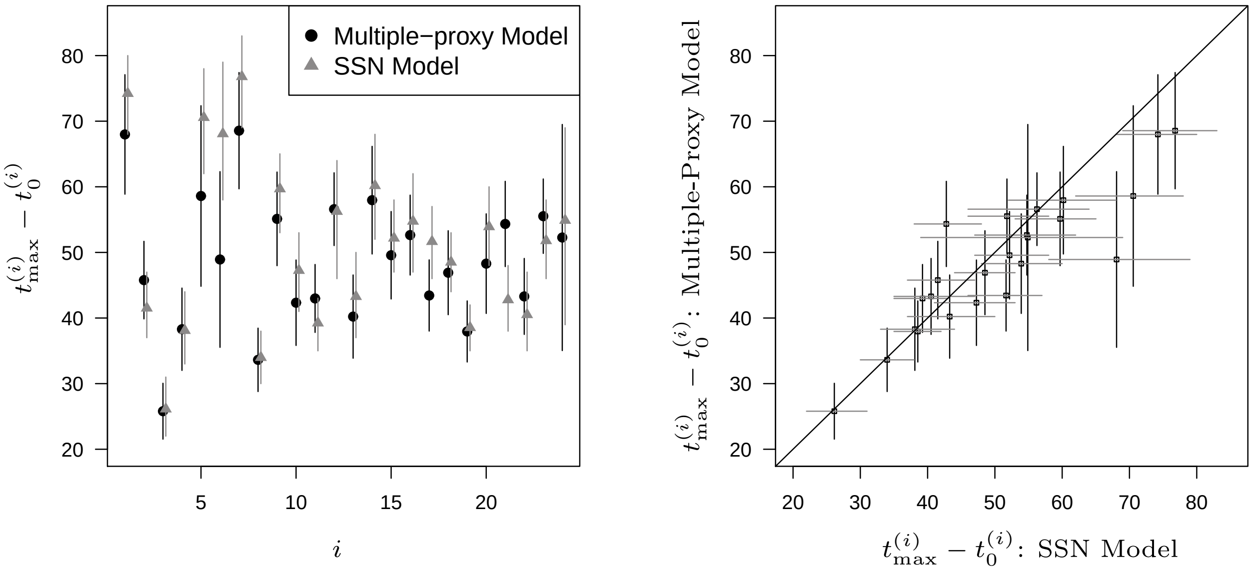


Figure 3.8: Fitted values and 95% intervals for the cycle rising times under the multiple-proxy and SSN models. There does not appear to be a significant difference in the rising times between the two model fits.

results suggest that fitting multiple proxies instead of only using the SSNs yields shorter overall cycle lengths, $t_1^{(i)} - t_0^{(i)}$. We display fitted values and 95% intervals for the total cycle lengths in Figure 3.10, and confirm that the model fit with multiple proxies generally has significantly shorter total cycle lengths.

There has been speculation that recent solar cycles represent a period of relatively high activity. Temmer (2010) and Shapoval *et al.* (2013), for example, suggested that the Sun was in a "low-activity" phase from around 1850 to 1915, while the Sun was in a "high-activity" phase from 1915 to the most recent solar minimum. We can evaluate this claim by obtaining an estimate and 95% interval for the mean amplitude during the high-activity phase, cycles 15 though 23, and the low-activity phase, cycles 10 through 14. Let $\bar{c}^{(j:k)}$ represent the mean amplitude from cycles j to k. Then, under the multiple-proxy model, the estimate of $\bar{c}^{(15:23)}$ is 5.01, with 95% interval $(4.86, 5.17)$, and the estimate of $\bar{c}^{(10:14)}$ is 3.86, with 95% interval $(3.70, 4.02)$. Thus, there is evidence that cycles 15 through 23 exhibit higher average solar activity than do cycles 10 through 14. Temmer (2010) and Shapoval *et al.* (2013) also

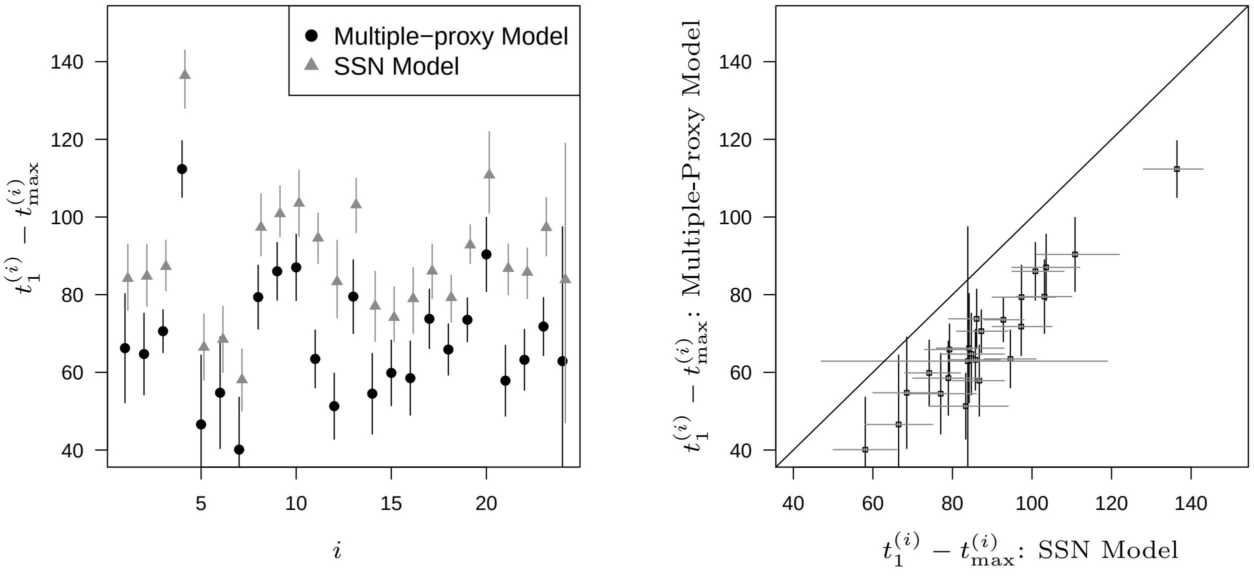


Figure 3.9: Fitted values and 95% intervals for the cycle falling times under the multiple-proxy and SSN models. Unlike the cycle rising times, the falling times do appear to differ between the two model fits. Specifically, the falling times are significantly shorter under the model fit with multiple proxies.

hypothesized that cycle 24 represents a shift back to a low-activity phase. Our Bayesian approach allows for straightforward calculation of the probability that the amplitude for cycle 24 will be below $\bar{c}^{(15:23)}$: $Pr(c^{(24)} < \bar{c}^{(15:23)}) = 1$. Futhermore $Pr(c^{(24)} < \bar{c}^{(10:14)}) = 0.88$, which suggests that the amplitude for cycle 24 may be unusually low, even when compared to the low-activity regime.

The Markov structure of our model allows for straightforward prediction of the current cycle even when little data for this cycle is observed. In Figure 3.11, we show the prediction for cycle 24, using data extending up to May 2010 (left panel) and up to December 2012 (right panel). When less data is available from the current cycle, predictions rely more on the Markov structure of the model and are thus more uncertain. As data become available, the predictions are increasingly driven by the current cycle and the uncertainties diminish. The fitted solar cycle is similar in both cases, but the 95% intervals are noticeably narrower with more data. This shows that the cycle-to-cycle relationships we learn in the second stage of the model are consistent with the most recent data and we can make reasonable predictions

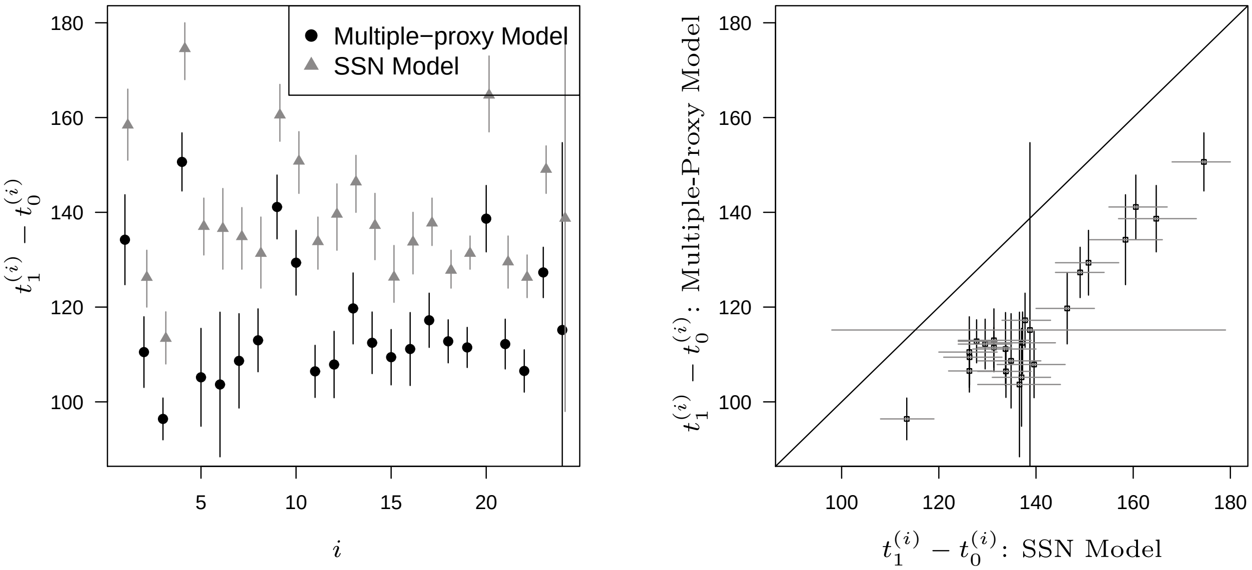


Figure 3.10: Fitted values and 95% intervals for the total cycle lengths under the multiple-proxy and SSN models. Following from the results displayed in Figures 3.8 and 3.9, the multiple-proxy model generally has significantly shorter cycle lengths than the SSN model has.

of cycle characteristics at the start of the cycle, albeit with considerable uncertainty.

3.5 Summary and Discussion

We have carried out a fully Bayesian analysis of the solar cycle using multiple proxy variables by generalizing the model of Yu *et al.* (2012). After suitably transforming the data to stabilize the variance and increase linearity, we multiply-imputed missing data by specifying a simple local missing data model that incorporates the Markovian structure of the data. The dimensionality of each imputed data set is reduced using PCA to project the multivariate proxy observations onto a one-dimensional subspace along the direction of highest variance. In this way, we obtain a univariate summary of solar activity at each time point, allowing us to utilize the existing univariate model to infer properties of the solar cycle using multiple proxies. This approach is based on the current understanding that there is a single underlying solar cycle, and the several proxies all provide information about the cycle.

It is necessary to use the long history of the SSNs and sunspot area observations in order to learn the patterns among consecutive solar cycles. Multiple imputation is used in order to easily derive estimates based on the posterior distribution of the model parameters given all of the data. If only complete observations are used, meaning only the data with all three proxies observed, the model is overfit. This problem would be compounded in future analyses that may also include other proxies that do not extend back for more than a couple of cycles. One of our primary aims is to allow additional proxies, even if they are largely missing, so that all available data can be used in a coherent statistical framework.

We compare fits of the Bayesian multilevel model of the solar cycle based on (i) multiple proxies and (ii) the SSNs alone. We observe significant differences in the inferred cycle properties. In particular, we find that the model fit with multiple proxies has shorter falling times than the model fit with the SSNs. Since we do not find significant differences in the rising times, the shorter falling times from the multiple-proxy model also imply shorter total cycle lengths, which we also observe. Shorter cycle lengths is turn imply longer solar minima. It has been observed elsewhere that the Sun can remain in a prolonged state of minimum activity, and there is evidence that the most recent solar minimum was unusual in its depth and duration (Basu, 2013). During the most recent solar minimum the 10.7cm flux was the lowest ever recorded, and physical characteristics of the solar surface and interior were unusual when compared to previous solar minima (Basu, 2013). It is clear from Figure 3.6 that the most recent minimum of the fitted solar cycle has a longer duration under the multiple-proxy model. In this regard the multiple-proxy fit captures an important feature in the solar cycle that is missed by the SSN model. The use of additional proxies may further illuminate this effect.

Future work will also consider additional functional forms for the solar cycle. Recent studies have presented evidence that many solar cycles have double maxima (e.g., Georgieva, 2011;

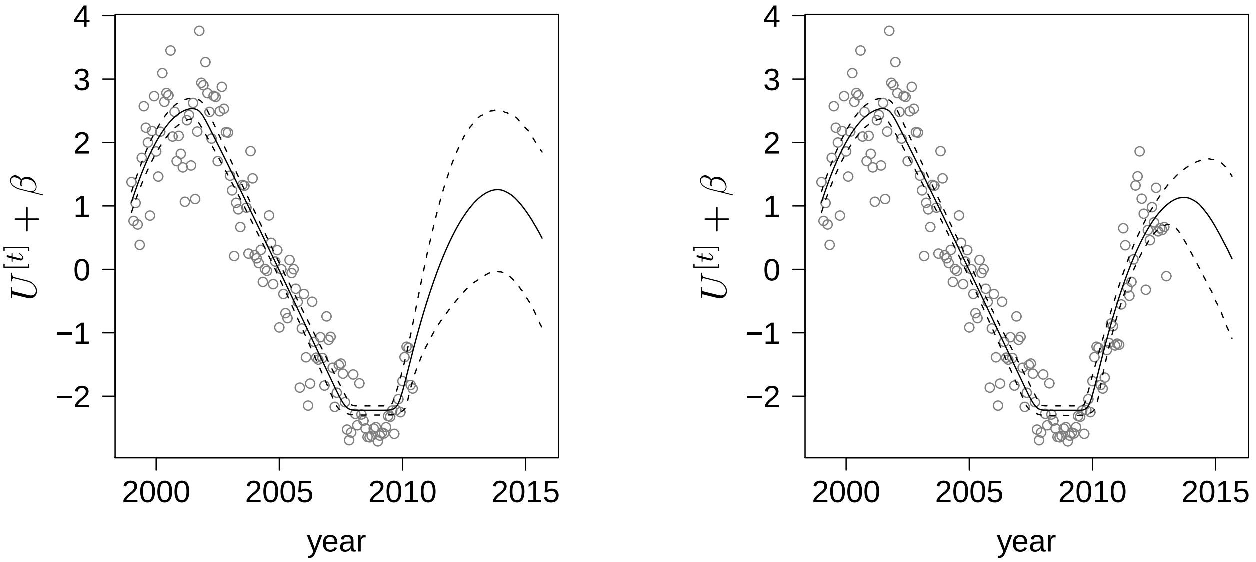


Figure 3.11: Prediction for cycle 24, using data up to May 2010 (left) and up to December 2012 (right). The solid curve is the fitted $U^{[t]} + \beta$, and 95% posterior (predictive) intervals are given by the dotted lines. Gray circles represent the observed $G(Y) = \omega^T \tilde{Y}$. As more data for the cycle is obtained, the uncertainty in the predictions is reduced.

Kilcik and Ozguc, 2014), which our current parameterization does not capture. Gnevyshev (1967) suggests that complex physical processes produce double peaks in all cycles, but often the gap between them is too short for the two within-cycle maxima to be distinguished. One possible explanation for double maxima that we can explore is the existence of separate cycles acting on each hemisphere of the Sun. Under this scenario the separate northern-hemisphere and southern-hemisphere cycles are parameterized by (3.1) and (3.2), with total activity being the sum over the two hemispheres and double maxima appearing when the two hemispheres reach peak activity at different times.

Our multiple-proxy Bayesian multilevel model of the solar cycle provides the flexibility needed to dynamically describe the complex structure of cycles and their varying shapes, duration, and amplitudes, while capturing the predictable way in which these features evolve over time. The effective combination of multiple imputation and PCA-based dimension reduction makes it straightforward to incorporate additional proxies, all the while taking advantage of the long history of SSN observations.

Part II

Bayesian Analysis of Stellar Evolution

Chapter 4

Background

4.1 Introduction[1]

4.1.1 The Lifecycle of a Star

For most of their lives, stars are powered by the thermonuclear fusion of hydrogen into helium in their cores, the byproduct of which is the release of energy into space. The evolution and eventual fate of a star is almost entirely determined by its initial mass. According to current astrophysical understanding, a star forms when a dense region of a molecular cloud collapses under its own gravity, forming a *protostar*. If the mass of the resulting protostar is greater than approximately 10% of the mass of the Sun, M_{Sun}, thermonuclear fusion is ignited and lasts for millions to billions of years, depending on the initial mass of the star; more massive stars burn hotter and use up their hydrogen fuel faster than less massive stars. When the hydrogen in the stellar core is mostly depleted, the core collapses and its temperature increases. At the same time, the radius of the star increases and its surface cools, resulting

---

[1] This introductory material is adapted from a more detailed introduction to stellar evolution aimed at statisticians that is given in Sections 1 and 2 of van Dyk *et al.* (2009).

in a *red giant* star. For stars with sufficient mass, the core becomes hot enough to fuse helium into even heavier elements. The helium-burning phase is relatively short, lasting about one-tenth as long as the hydrogen-burning phase. During the helium-burning phase the star will undergo mass loss due to low surface gravity (a result of the expanded stellar radius), resulting in a short-lived *planetary nebula*.

The end of a star's lifecycle depends on its initial mass. For stars with an initial mass less than about 8 M_{Sun}, the collapse of the core following the depletion of fuel for further thermonuclear reactions is eventually balanced by the outward force of electron degeneracy pressure. After the outer layers of the star blow away, all that is left is the now stable core, (typically) composed of carbon and oxygen, which is known as a *white dwarf* star. For stars with an initial mass greater than about 8 M_{Sun}, electron degeneracy pressure is not enough to halt the collapse of the core; stability is not achieved until after electrons are forced into atomic nuclei, combining with protons to form neutrons. The resulting neutron degeneracy pressure halts further collapse, producing a core that is composed almost entirely of neutrons—a *neutron star*. Neutron stars are the smallest and densest stars known to exist; for stars with a still higher initial mass, even neutron degeneracy pressure cannot prevent further (infinite) collapse, leading to the formation of a *black hole*.

4.1.2 Star Clusters

Stars tend to form in groups known as *star clusters*, of which there are two main types: *open clusters* and *globular clusters*. Open clusters are (loosely) gravitationally bound conglomerations of up to a few thousand stars that all formed at roughly the same time from the same molecular cloud. Open clusters are held together by mutual gravitational attraction and last for a few hundred million to a few billion years, depending on the combined mass of the stars. Globular clusters are similar to open clusters, but can contain millions of stars.

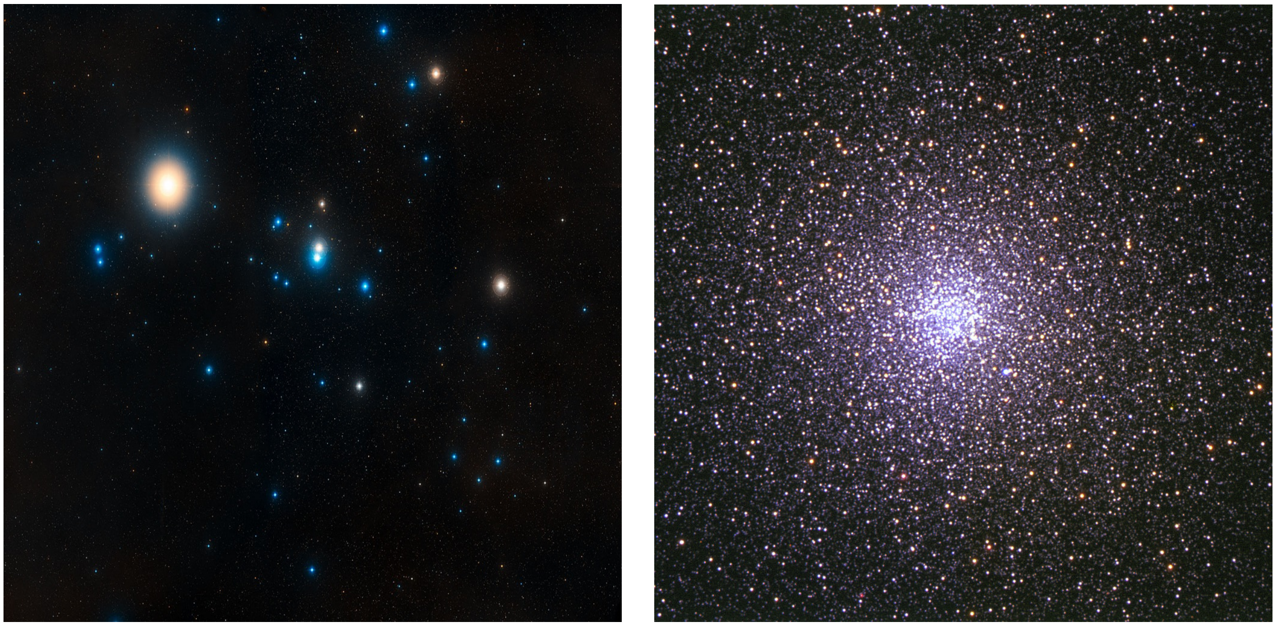


Figure 4.1: Star Clusters. *Left:* The Hyades open cluster in the constellation Taurus. (Image courtesy of NASA and ESA.) *Right:* The globular cluster 47 Tuc in the constellation Tucana. (Image courtesy of ESA.) Compared to the Hyades, 47 Tuc contains many more stars that are bound in a tighter group.

The higher combined stellar mass results in more tightly bound clusters that can last for longer timespans. Open and globular clusters differ in the types of stars that comprise the cluster, as well as in their locations within the Galaxy and their orbits, but such differences are unimportant for our statistical analyses.

Examples of star clusters are presented in Figure 4.1. The left panel of Figure 4.1 is an image of the *Hyades* open cluster. At a distance of about 153 lightyears, the Hyades is the nearest open cluster to the Earth (Perryman *et al.*, 1998). Its age has been estimated at around 625 million years (Perryman *et al.*, 1998; van Dyk *et al.*, 2009), and it has a radius of about 8.8 lightyears. The right panel of Figure 4.1 is an image of the globular cluster *47 Tucanae* (47 Tuc), which we analyze in Chapter 6. Like all globular clusters, 47 Tuc is very old, with an estimated age of around 13 billion years (Zoccali *et al.*, 2001). It is located at a distance of about 16,700 lightyears from the Earth (Carretta *et al.*, 2000), and has a radius of about 60 lightyears. Both the Hyades and 47 Tuc can be seen with the naked eye, in the constellations of Taurus and Tucana, respectively. Comparing the image of the Hyades with the image of 47 Tuc, we find that the latter contains many more stars in a denser grouping.

Star clusters are important for the study of stellar evolution. Since stars in a cluster are born from the same parent molecular cloud at roughly the same time, we can consider them to have the same age and chemical composition. Because star clusters are located far away, and the separation between the stars is relatively small, we can consider them to be equally distant from the Earth. (The Hydades is an exception, as its radius is about 6% of its mean distance from the Earth (De Gennaro, 2009). For comparison, the radius of 47 Tuc is about 0.4% of its mean distance from the Earth.) Lastly, because the stars are physically co-located, we can assume that their light travels through the same interstellar material before reaching our telescopes. Under these conditions, stars in a cluster differ only in their initial masses. That we can consider stars in a cluster to be similar in all these ways simplifies statistical modeling and inference (van Dyk *et al.*, 2009; Stein *et al.*, 2013).

4.1.3 Color-Magnitude Diagrams

A *color-magnitude diagram* (CMD) is a plot of temperature versus brightness for stars belonging to a cluster. A star's brightness is quantified by its *luminosity*, which is the amount of energy radiated per unit time. A CMD is constructed using a star's *magnitude*—a negative logarithmic transformation of luminosity, so that smaller magnitudes correspond to brighter stars—in several color bands, called *photometric magnitudes.* A difference in photometric magnitudes, the horizontal axis of a CMD, is a star's spectral type (i.e., color), which is a proxy for temperature: bluer stars are hotter and redder stars are cooler. The original type of CMD, referred to as a Hertzsprung-Russell diagram (H-R diagram) after its inventors, plots a difference in absolute magnitude on the horizontal axis versus an absolute magnitude on the vertical axis. H-R diagrams are the original CMDs, but require precise

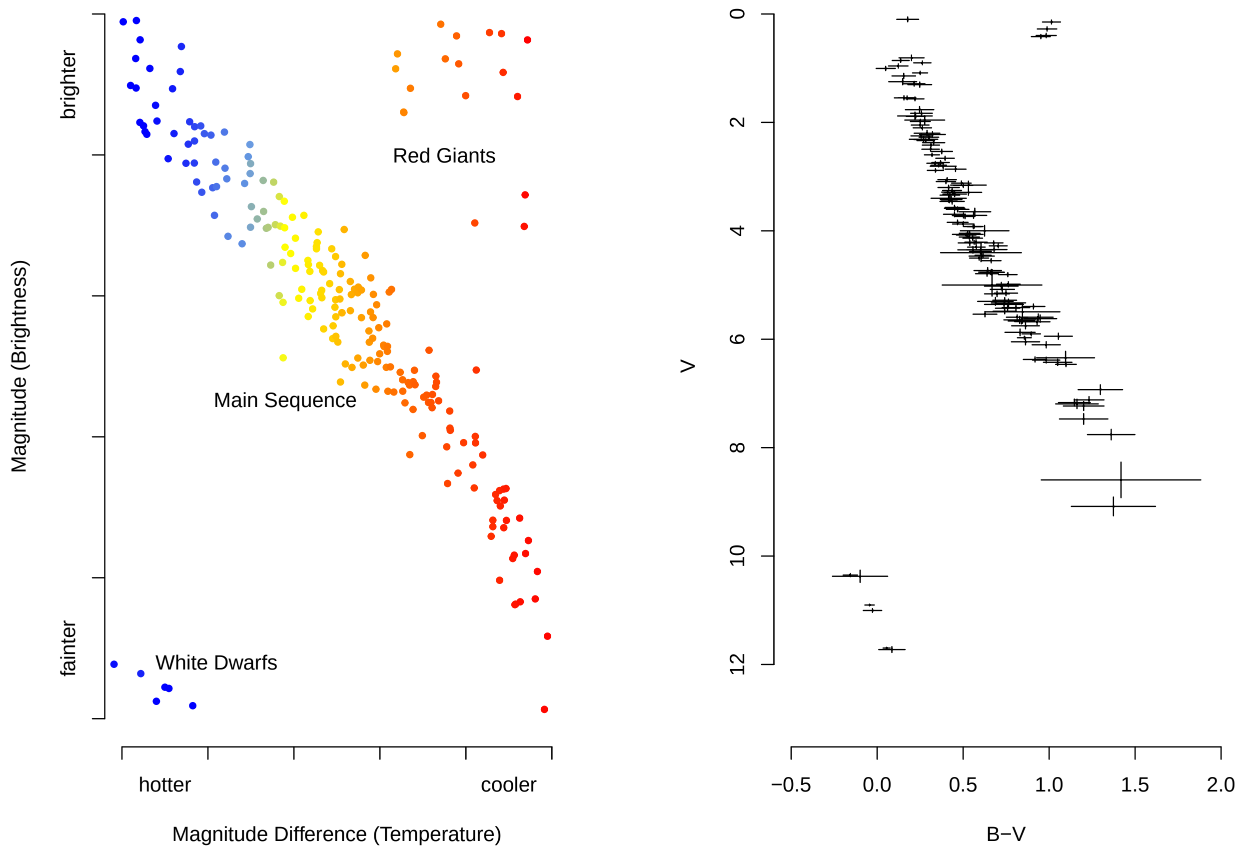


Figure 4.2: Color-Magnitude Diagrams. *Left:* A schematic color-magnitude diagram (CMD). The main-sequence stars, red giants, and white dwarfs assemble into distinct and easily recognizable groups. The artificial coloring corresponds to the temperature of stars in the CMD: cooler stars are colored towards the red end of the spectrum and hotter stars are colored towards the blue end of the spectrum. *Right:* The CMD for the Hyades open cluster, with the main-sequence stars, red giants, and white dwarfs easily discernible. Each star is plotted with 68% (i.e. 1σ) Gaussian measurement errors. The horizontal axis is the difference in magnitude between the B ("blue") and V ("visual") filters, which corresponds to stellar spectral type (i.e. color) or temperature. The vertical axis is the magnitude in the V filter, which indicates brightness. This figure is adapted from Figure 1 in van Dyk *et al.* (2009).

distance measurements to calculate absolute magnitudes that are generally difficult, if not impossible, to obtain. Since stars in a cluster are located at roughly the same distance from Earth, a CMD uses the *apparent magnitudes* that are actually measured, as opposed to the *absolute magnitudes* that would be measured if the cluster were at a distance of 10 parsecs (approximately 33 light years). Therefore, H-R diagrams are generalized by CMDs.

Since a star's luminosity—and hence its magnitude—changes throughout its lifecycle, CMDs can be used to identify stars in different evolutionary states. A schematic CMD is presented in the left panel of Figure 4.2, and a CMD constructed using photometric data from the Hyades is presented in the right panel. The *main-sequence* stars are those in their initial hydrogen-burning phase of life. There is a continuum of initial mass moving from the lower-right to the upper-left of the main sequence. Stars to the upper-left are massive, hotter, and brighter; they burn their hydrogen fuel more quickly and transition to the "Red Giant" group located at the upper-right of the CMD before the less massive stars found lower on the main sequence. Red giant stars are both cooler—which makes them appear "redder"—and more luminous due to their expanded radius. After a star depletes its thermonuclear fuel and loses its outer layers, the remaining core migrates to the dim and cool "White Dwarf" group at the lower-left of the CMD. (Neutron stars are not nearly as numerous as white dwarfs and do not typically have detectable luminosities, while black holes have no luminosity. As a result, neutron stars and black holes do not appear in CMDs.) The main-sequence stars, red giants, and white dwarfs are labeled on the schematic CMD; the same groups are easily identified in the CMD for the Hyades.

In addition to age and initial mass, a number of other parameters affect a star's location on the CMD. *Metallicity* is $\log_{10}$ of the ratio of iron and hydrogen atoms in a star, and represents the chemical composition of the star in terms of the abundance of elements heavier than helium. Iron is used as a proxy for all elements heavier than helium due to the relative ease at which it can be identified in spectral analysis (van Dyk *et al.*, 2009). Since heavier elements absorb "bluer" light and inhibit thermal radiation, stars that differ in metallicity will have different sets of photometric magnitudes, even when their other parameters are equal. Similarly, the *helium abundance*—$\log_{10}$ of the ratio of helium and hydrogen atoms—in stellar cores affects the efficiency of hydrogen-burning, since higher amounts of helium make the thermonuclear fuel less "pure." Higher helium abundance corresponds to stellar cores that are somewhat smaller, denser, and hotter, which again affects their photometric magnitudes (van Dyk *et al.*, 2009).

The photometric magnitudes of white dwarfs are affected by the *carbon fraction*—the proportion of the star composed of carbon atoms. Carbon is synthesized via thermonuclear fusion late in a star's lifecycle. If it has sufficient mass, a star fuses helium into carbon and oxygen during its red-giant phase. Only 3% of stars have sufficient mass to fuse carbon into even heavier elements (Fontaine *et al.*, 2001). Thus, for a typical red giant the carbon and oxygen produced from helium fusion builds up in the stellar core, which forms a white dwarf after the star has shed its outer layers.

The parameters described so far—age, initial mass, metallicity, helium abundance, and carbon fraction—affect a star's absolute magnitude and, consequently, its apparent magnitude. Two additional parameters affect only apparent magnitudes. Since space is not a complete void, some of the light emitted by a star is blocked by interstellar matter, which is quantified by *absorption*. Lastly, the *distance* of a star affects its apparent magnitude, since stars that are father away appear less luminous. Thus, we consider up to seven parameters that affect a star's photometric magnitudes and therefore determine its placement on the CMD: age, initial mass, metallicity, helium abundance, distance, absorption, and, for white dwarfs, carbon fraction. We refer to the parameters that are common to all stars in a cluster—age, metallicity, distance, absorption, and carbon fraction—as *cluster parameters*. For reasons that become clear in Chapter 5, we do not consider helium abundance to be a cluster parameter. The stellar evolution parameters that affect a star's photometric magnitudes and placement on the CMD are summarized in Table 4.1. The notation in Table 4.1 is introduced in Section 4.2.1.

A closed-form function for determining a star's photometric magnitudes given the stellar evolution parameters does not exist. Instead, the intricate physics calculations required to

Table 4.1: Stellar Evolution Parameters

Parameter	Description	Notation
Age	$\log_{10}$ of the cluster age	θ_{age}
Distance	distance modulus (difference between apparent and absolute magnitudes)	θ_{m-M_V}
Absorption	loss of light due to interstellar matter	θ_{A_V}
Metallicity	$\log_{10}$ of the ratio of iron and hydrogen atoms	$\theta_{[\text{Fe/H}]}$
Carbon Fraction	proportion of a white dwarf composed of carbon atoms	θ_C
Helium Abundance	$\log_{10}$ of the ratio of helium and hydrogen atoms	ϕ_Y
Initial Mass	initial mass in units of solar mass, M_{Sun}	M_i

produce photometric magnitudes are executed in complex computer-based models of stellar evolution. These models are discussed in Section 4.1.4

4.1.4 Bayesian Analysis of Stellar Evolution (BASE-9)

BASE-9 is a software suite developed for Bayesian analysis of stellar evolution. A combination of several computer-based stellar evolution models is used to predict a star's photometric magnitudes, given the set of input parameters described in Section 4.1.3. BASE-9 has the capability to simulate a star cluster with specific physical properties and realistic Gaussian measurement errors. However, its primary purpose is to recover star cluster parameters from photometric data. For this aim, BASE-9 includes sophisticated Markov chain Monte Carlo (MCMC) routines to explore the marginal posterior distribution of the parameters of scientific interest. BASE-9 is available as open source code from GitHub (`https://github.com/argiopetech/base/releases`), and is also available through Amazon Web Services. Additional technical details can be found in the BASE-9 Manual (`http://arxiv.org/pdf/1411.3786.pdf`).

For main-sequence and red giant stars, BASE-9 includes the state-of-the-art models by Dotter *et al.* (2008, and updated at `http://stellar.dartmouth.edu/~models/`), Girardi *et al.* (2000), and Yi *et al.* (2001). Other models are included for white dwarfs; the models for main-sequence and red giant stars do not capture the chaotic transformation that stars undergo during the transition from red giants to white dwarfs. BASE-9 therefore includes the white dwarf atmosphere models of Bergeron *et al.* (1995), and the white dwarf interior (i.e. white dwarf cooling) models of Wood (1992), Althaus and Benvenuto (1998), Renedo *et al.* (2010), and Montgomery *et al.* (1999). In addition, M. Montgomery created a new white dwarf interior model specifically for use with BASE-9, using the codes described in Montgomery *et al.* (1999). This is the only white dwarf interior model that includes carbon fraction as an adjustable parameter. We make use of this model in Chapter 6, where we refer to it as the "updated" Montgomery *et al.* (1999) model.

Finally, an empirical mapping known as the *initial-final mass relation* is needed to link the initial mass of a main-sequence star with the mass of its resulting white dwarf. BASE-9 includes the initial-final mass relations of Salaris *et al.* (2009), Weidemann (2000), and Williams *et al.* (2009). An option also exists to fit the initial-final mass relation, which is parameterized as a line, broken line, or low-order polynomial; see Stein *et al.* (2013) for more details. An atmosphere model, interior model, and initial-final mass relation together take the seven input parameters from Section 4.1.3 and output the predicted magnitudes for a white dwarf star.

Combining several computer models for stars at different evolutionary stages into one comprehensive computer-based model of stellar evolution was first proposed by von Hippel *et al.* (2006), and was further developed and refined by DeGennaro *et al.* (2009), van Dyk *et al.* (2009) and Stein *et al.* (2013). Following van Dyk *et al.* (2009), we hereafter refer to the combination of (i) the models for main-sequence and red giant stars with (ii) the white dwarf interior and atmosphere models, which are bridged by (iii) an initial-final mass relation as

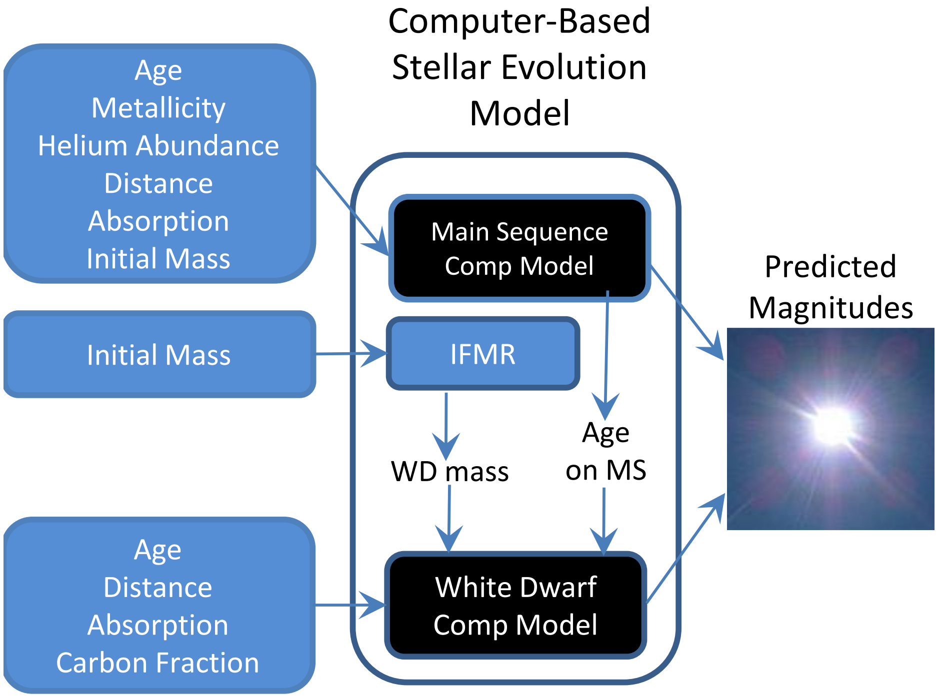


Figure 4.3: Computer-based stellar evolution model. The seven unique input parameters are listed on the left. These are fed into the computer-based stellar evolution model that combines models for main-sequence and red giant stars with models for white dwarfs, bridged by the initial-final mass relation. The computer model outputs the predicted magnitudes for a star with the physical properties specified by the inputs. This figure is adapted from Figure 2 in (Stein *et al.*, 2013).

the *computer-based stellar evolution model*. This combined model accounts for all stages in the lifecycle of a star, as illustrated in Figure 4.3. Readers interested in the physics underlying the computer models are referred to van Dyk et al. (2009) and in the online supplement and references therein.

BASE-9 has been successful in a variety of applications involving open clusters. For example, BASE-9 has been used to (i) provide the most accurate white dwarf age for the Hyades to date (DeGennaro *et al.*, 2009; van Dyk *et al.*, 2009), (ii) derive the best age for the open cluster NGC 2477 (Jeffery *et al.*, 2011), and (iii) derive the initial-final mass relation (Stein *et al.*, 2013). A full list of references for successful applications of BASE-9 can be found in the manual cited above.

4.2 Review of Models and Methods Used in BASE-9

4.2.1 Notation

For each star in a dataset we observe photometric magnitudes in different color bands (i.e., with different filters). More precisely, we refer to the observed photometric magnitude in color band j for star i as x_{ij} for $j = 1, \ldots, n$ and $i = 1, \ldots, N$, where N is the number of stars in the dataset and n is the number of color bands. The observed photometric magnitudes for star i are tabulated in the vector $\boldsymbol{X}_i = (x_{i1}, \ldots, x_{in})^\top$, and the known Gaussian measurement errors in the variance-covariance matrix, $\boldsymbol{\Sigma}_i$. The cluster parameters—age, metallicity, distance, absorption, and carbon fraction—are contained in the vector $\boldsymbol{\Theta} = (\theta_{\text{age}}, \theta_{[\text{Fe/H}]}, \theta_{m-M_V}, \theta_{A_V}, \theta_C)$. The inclusion of carbon fraction, θ_C, is an extension of BASE-9 that has not appeared before. Helium abundance, not contained in $\boldsymbol{\Theta}$ for reasons that will become clear in Chapter 5, is denoted by ϕ_Y. So far, the only *stellar parameter* specific to star i is its initial mass, M_i. The computer-based stellar evolution model, $\boldsymbol{G}$, takes $(M_i, \boldsymbol{\Theta}, \phi_Y)$ and outputs a $1 \times n$ vector containing the predicted magnitudes for a star with those parameters. We express the vector of predicted magnitudes as $\boldsymbol{G}(M_i, \boldsymbol{\Theta}, \phi_Y)$.

Current astrophysical understanding suggest that a significant portion of all stars are actually *binary stars*—two stars orbiting the center of mass of the system (van Dyk *et al.*, 2009). Binary stars often appear as a single point of light, in which case the luminosity we observe is the sum of the luminosity from both stars. The added luminosity of a binary companion shifts a star's position off the main sequence on the CMD, which can result in systematic errors if not properly handled. Following van Dyk *et al.* (2009), we account for unresolved binary stars by treating every observed star as a possible binary system and fitting its component masses; a low mass for the secondary companion star is treated as evidence for a unitary system. We denote the mass of the primary star (i.e., the more massive star) in the system as M_i, and we denote the ratio of the secondary mass (i.e., the mass of the less-massive binary companion star) to the primary mass as R_i. Since luminosity is related to magnitude by magnitude $= -2.5\log_{10}$(luminosity), the predicted magnitudes for (binary) star i can be expressed as

$$\boldsymbol{\mu}_i = -2.5\log_{10}\left(10^{-\boldsymbol{G}(M_i,\boldsymbol{\Theta},\phi_Y)/2.5} + 10^{-\boldsymbol{G}(M_iR_i,\boldsymbol{\Theta},\phi_Y)/2.5}\right), \tag{4.1}$$

where $\boldsymbol{\mu}_i$ is the mean vector of the photometric magnitudes, and $\boldsymbol{G}(M_iR_i,\boldsymbol{\Theta},\phi_Y)$ is the vector of predicted magnitudes for the secondary star since M_iR_i is the secondary mass. (Existing computer-based models of stellar evolution do not account for the complicated physics of binary systems involving white dwarfs. Since such systems are rare, however, we do not include them in our model.)

4.2.2 The Likelihood Function

Observed data from star clusters are contaminated by *field stars*—stars which are in the line-of-sight of the cluster, but are actually behind or in front of the cluster (van Dyk *et al.*, 2009). By observing the motion and velocity of individual stars, astrophysicists can attempt to separate field stars from cluster stars. However, such measurements are not typically available for all stars, and there is some uncertainty even for well-studied clusters. Following van Dyk *et al.* (2009), we account for field star contamination by modeling the observed data using a two-component finite mixture distribution. In particular, we define $\boldsymbol{Z} = (Z_1,\ldots,Z_N)$, where $Z_i = 1$ if star i is a cluster star and $Z_i = 0$ if star i is a field star. This allows us to specify separate conditional distributions of the observed photometric magnitudes for cluster stars and field stars. For cluster stars, we model the observed photometric magnitudes as n independent Gaussian random variables with means given in $\boldsymbol{\mu}_i$ and variances in the variance-covariance matrix $\boldsymbol{\Sigma}_i$ of known measurement errors. While this model appears simple at first glance, it is actually quite complex due to the dependence of $\boldsymbol{\mu}_i$ on the stellar evolution parameters, and the complex interdependencies therein. That $\boldsymbol{G}$ cannot be expressed in closed form yields challenges for inference and computation.

Following van Dyk *et al.* (2009), we specify a simple model for field stars that does not depend on any of the parameters of interest. Each field star may have its own values for θ_{age}, $\theta_{\text{[Fe/H]}}$, θ_{m-M_V}, θ_{A_V}, θ_C, and ϕ_Y, and we cannot identify these parameters. We therefore simply assume that each field star magnitude is uniformly distributed over the range of the data, such that

$$P(\boldsymbol{X}_i|Z_i = 0) = c \quad \text{if } \min_j \leq x_{ij} \leq \max_j, j = 1, \ldots, n,$$

and zero elsewhere, where $(\min_j, \max_j)$ is the range of magnitude values for color band j, and $c = \left[\prod_{j=1}^{n}(\max_j - \min_j)\right]^{-1}$. We could of course incorporate a more complex and realistic model for field stars; properties of field stars for specific Galactic fields exist and may assist in tuning the model (Stein *et al.*, 2013). However, our work to date has not necessitated the additional effort since the simple model adequately identifies field stars.

A preliminary likelihood function can now be written,

$$\begin{aligned}
&L_{\text{p}}(\boldsymbol{M}, \boldsymbol{R}, \boldsymbol{Z}, \boldsymbol{\Theta}, \phi_Y|\boldsymbol{X}, \boldsymbol{\Sigma}) \\
&= \prod_{i=1}^{N} \Bigg[Z_i \times \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{1}{2}\Big(\boldsymbol{X}_i - \boldsymbol{\mu}_i\Big)^{\top} \boldsymbol{\Sigma}_i^{-1} \Big(\boldsymbol{X}_i - \boldsymbol{\mu}_i\Big)\right) \\
&\qquad + (1 - Z_i) \times P(\boldsymbol{X}_i|Z_i = 0) \Bigg] \\
&= \prod_{i=1}^{N} \Bigg[Z_i \times P(\boldsymbol{X}_i|\boldsymbol{\Sigma}_i, M_i, R_i, \boldsymbol{\Theta}, \phi_Y, Z_i = 1) \\
&\qquad + (1 - Z_i) \times P(\boldsymbol{X}_i|Z_i = 0) \Bigg],
\end{aligned} \tag{4.2}$$

where $\boldsymbol{M} = (M_1, \ldots, M_N)$, $\boldsymbol{R} = (R_1, \ldots, R_N)$, $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)$, $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_N)$, and $P(\boldsymbol{X}_i | M_i, R_i, \boldsymbol{\Theta}, \phi_Y, Z_i = 1)$ is the conditional distribution of observed photometric magnitudes for a cluster star. We refer to this as a preliminary likelihood, using a 'p' in the subscript, since $P(\boldsymbol{X}_i | M_i, R_i, \boldsymbol{\Theta}, \phi_Y, Z_i = 1)$ is modified in Chapters 5 and 6 to suit specific scientific applications.

4.2.3 Bayesian Model Fitting and the Prior Distributions

We take a Bayesian perspective for model fitting. A key advantage to adopting a Bayesian approach is the ability to directly incorporate substantive astrophysical knowledge through the prior distributions for the stellar evolution parameters. For example, Miller and Scalo (1979) derived the distribution of initial stellar masses in the Milky Way. Since the globular clusters we study orbit the Milky Way and are thought to have the same distribution of initial stellar masses, we can use this prior analysis to construct an informative prior distribution on initial mass for stars in our dataset. In particular, we specify a Gaussian prior distribution on the base 10 logarithm of primary initial masses:

$$P\left(\log_{10}(M_i)\right) \propto \exp\left(-\frac{1}{2}\left(\frac{\log_{10}(M_i) + 1.02}{0.677}\right)^2\right), \tag{4.3}$$

truncated to 0.1 M_{Sun} to 8 M_{Sun}, where the numerical constants are taken from Miller and Scalo (1979). The lower truncation is due to the fact that an initial mass of less than about 0.1 M_{Sun} is not sufficient to initiate the fusing of hydrogen into helium necessary to form a star. The upper truncation is because the star clusters we study are sufficiently old that any stars with an initial mass above 8 M_{Sun} would have used up their nuclear fuel long ago and become a neutron star or black hole, and thus would not be included in our observed data (van Dyk *et al.*, 2009). In general, this prior distribution for M_i reflects the knowledge that there are many more low mass stars than high mass stars. For the mass ratio of the secondary and primary masses we use a uniform prior distribution on $[0, 1]$. We need not truncate the lower end of the secondary mass because low secondary masses provide evidence that the star is a unitary system (van Dyk *et al.*, 2009).

For the cluster parameters, $\mathbf{\Theta}$, we incorporate ancillary information to specify informative prior distributions when available, and use relatively diffuse prior distributions when such information is lacking. In particular, for $\theta_{[\mathrm{Fe/H}]}$, θ_{m-M_V}, and θ_{A_V}, we use Gaussian prior distributions (truncated to be positive in the case of θ_{A_V} since absorption cannot be negative), with means set according to prior astrophysical knowledge and standard deviations chosen to be reasonably diffuse. We use a uniform prior distribution on $[0, 1]$ on the carbon fraction in white dwarfs, θ_C, since ancillary information necessary to specify a more informative prior distribution does not currently exist.

For age, θ_{age}, we use a uniform prior distribution on the $\log_{10}$ of cluster age in years, truncated to the range 1 billion years to 15 billion years. Based on current astrophysical knowledge, the globular clusters we study are all at least 10 billion years old (Forbes and Bridges, 2010), so our lower bound on cluster age is quite conservative. We are similarly conservative with the upper bound on age; the current best estimate of the age of the universe, and therefore the oldest possible age of any globular cluster, is 13.798 ± 0.037 billion years (Planck Collaboration *et al.*, 2014). Specifying the uniform prior distribution on the $\log_{10}$ of cluster age in years reflects the fact that older clusters are less common than younger ones (van Dyk *et al.*, 2009).

When helium abundance, ϕ_Y, is not the primary parameter of scientific interest, we use a Gaussian prior distribution with a mean and standard deviation set according to whatever prior astrophysical information is available. When ϕ_Y is the primary parameter of scientific interest, we use a uniform prior distribution on a physically realistic range. Additional discussion is provided in Chapter 5.

Ancillary measurements can be used to probabilistically separate field stars from cluster stars. Because stars in a cluster are gravitationally bound, they move as a pack through the galaxy. Astrophysicists can then analyze their movement and velocity to determine which stars belong to the cluster. When such ancillary measurements are unavailable, we use $P(Z_i = 1) = \alpha$ for $i = 1, \ldots, N$, where α is based on the expected fraction of cluster stars in the dataset.

4.2.4 Statistical Computation

Observed data from globular clusters can contain hundreds of thousands of stars. Since each star has a primary mass, M_i, mass ratio, R_i, and cluster membership status, Z_i, there are at least $3N + S_\Theta$ parameters, where S_Θ is the number of cluster parameters. Thus, we are dealing with a high-dimensional posterior distribution. Furthermore, the computer-based models for stellar evolution are not available in closed form, and there are complex nonlinear relationships in the parameters. To analyze open clusters, which contain fewer stars than globular clusters, DeGennaro et al. (2009) and van Dyk et al. (2009) construct a Metropolis-within-Gibbs sampler that uses dynamic reparameterizations and is implemented entirely sequentially: the cluster membership indicators are updated first, followed by the primary masses, mass ratios, and finally the cluster parameters. A sequence of initial runs is used to explore the posterior correlations among the parameters, which are then used to automatically derive transformations to reduce the correlations, thereby improving convergence.

The dynamic reparameterization approach has been successful with both simulated and observed open clusters, including the Hyades (DeGennaro *et al.*, 2009) and NGC 2477 (Jeffery *et al.*, 2011). However, some clusters yield posterior distributions containing multiple modes and extreme nonlinear correlations. In particular, multiple modes may result from varying the set of cluster stars. In particular, the MCMC sampler may stick in a local mode that results from a particular configuration of stars determined to be cluster stars, unable to escape to another mode even if the latter has a larger joint posterior probability. To tackle these challenges, Stein *et al.* (2013) marginalize over $\boldsymbol{M}$, $\boldsymbol{R}$, and $\boldsymbol{Z}$, thereby collapsing the parameter space from $3N+S_\Theta$ dimensions down to S_Θ dimensions. Stein *et al.* (2013) exploit conditional independence to reduce the $2N$-dimensional integration required to marginalize over $\boldsymbol{M}$ and $\boldsymbol{R}$ to the product of N 2-dimensional integrations. This provides a significant speedup in computation since the N 2-dimensional integrations can be performed in parallel. A Metropolis algorithm is then used to update the cluster parameters in a single jump. This overall approach has been successful for analyzing the initial-final mass relation using data from a variety of open clusters (Stein *et al.*, 2013).

4.3 Statistical Properties of the Parameter Estimates

In this section we explore non-standard asymptotic behavior of the parameter estimates—the influence of the prior distribution does not diminish as the sample size grows. To do so, we simulate a globular cluster with "average" parameters that are based on previously published studies compiled in Harris (1996, and updated in 2010 at `http://www.physics.mcmaster.ca/~harris/mwgc.dat`). The true parameter values under the simulation for the cluster parameters, as well as the prior distributions we assign for each parameter, are summarized in Table 4.2. In the table, $N(\mu, \sigma^2)$ is a Gaussian distribution with mean μ and standard deviation σ, and $TN(\mu, \sigma^2; 0)$ is a Gaussian distribution with mean μ and standard deviation σ, truncated to be positive. The standard deviations for the Gaussian prior distributions were chosen to reflect reasonable uncertainty on the parameter estimates from the Harris (1996) globular cluster catalogue. The prior distributions on age and initial mass are as described in Section 4.2.3. Because we do not include white dwarfs in this simulation study,

Table 4.2: Simulation Cluster Parameters and Prior Distributions

Parameter	Simulation Value	Prior Distribution
θ_{age}	10.08	Uniform(9.0, 10.176)
θ_{m-M_V}	15.375	$N(15.375, 0.02^2)$
θ_{A_V}	0.372	$TN(0.372, 0.006^2; 0)$
$\theta_{\text{[Fe/H]}}$	-1.5	$N(-1.5, 0.02^2)$
ϕ_Y	0.27	$N(0.27, 0.2^2)$

we do not specify a prior distribution for the carbon fraction. We also do not include binary star systems or field stars in this investigation, and therefore specify $P(Z_i = 1) = 1.0$ for $i = 1, \ldots, N$.

To simulate the cluster, we first draw 5000 initial masses from the prior distribution on initial mass, which is specified in Section 4.2.3. We therefore know the true initial mass of each star under the simulation, although we do not specify them directly. Each of these initial masses is then fed into the computer-based stellar evolution model, $\boldsymbol{G}$, along with the values for the cluster parameters given in Table 4.2. The specific computer model we use is the updated Dotter *et al.* (2008) model. The computer-based stellar evolution model outputs the predicted magnitudes for each star in several wavelength bands (i.e., color bands), of which we select a subset of five that correspond to the wavelength bands for the observed data we analyze in Chapter 5 (i.e., $F275W$, $F336W$, $F438W$, $F606W$, and $F814W$). We can then plot a CMD for the simulated cluster, which is presented in Figure 4.4(a); black dots represent the position of the simulated stars. The dark-red curve is an *isochrone*—a theoretical curve that represents the CMD positions of stars having the same cluster parameters but varying in initial mass. Initial mass increases monotonically from the bottom of the isochrone at the lower main sequence (bottom right) to the top of the isochrone at the end of the red giant branch (top right).

Using the simulated photometry presented in Figure 4.4(a) as a baseline, we consider a partial 2 × 2 simulation study design. In this design, we consider two magnitude depths at which we truncate the data prior to model fitting: F275W = 23 and F275W = 27.5. For model fitting, we assume Gaussian measurement errors and consider two values for the standard deviation, σ: 0.01 and 0.03. In total, we consider three scenarios:

Scenario 1 truncation at F275W = 23; assumed σ for model fitting = 0.01.

Scenario 2 truncation at F275W = 23; assumed σ for model fitting = 0.03.

Scenario 3 truncation at F275W = 27.5; assumed σ for model fitting = 0.03.

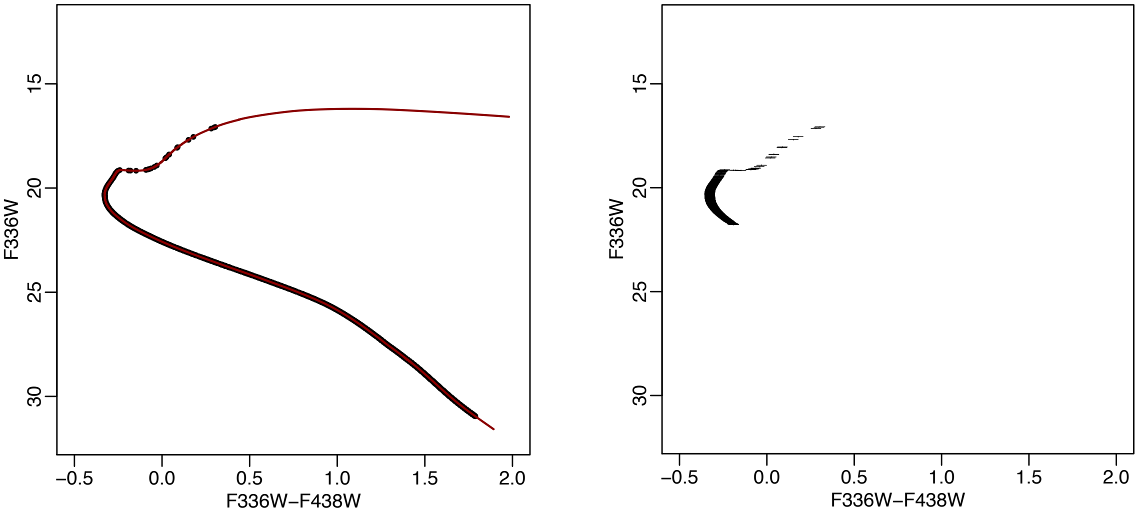
These three scenarios are illustrated in Figures 4.4(b)-(d). Our goal in this study is to examine the posterior means of initial masses, which we refer to as simply "masses" for the remainder of this section, and their effects on the other fitted parameters. We choose the truncation point at $F275W = 23$ to be representative of the truncation we apply to observed data from multiple-population globular clusters. As we discuss in Section 5.4, we truncate observed data since we believe the stellar evolution models to be inaccurate for low-mass (i.e., high-magnitude) stars. We choose the second truncation point at $F275W = 27.5$ so that Scenario 3 includes stars with masses lower than their prior mean. This allows us to examine whether the fitted masses are being shrunk towards the prior mean. We consider two different values for the assumed Gaussian measurement error because we want to isolate the influence of the prior distribution on (initial) mass. When the assumed Gaussian measurement error is larger, we expect that the prior distribution on mass will have more influence. Therefore, we construct Scenario 1 and Scenario 2 so that they only differ in the Gaussian measurement error assumed in model fitting. This difference is illustrated by comparing Figure 4.4(b) and Figure 4.4(c), where the horizontal and vertical bars represent the assumed Gaussian measurement error. Notice that for both scenarios the positions of the stars on the CMD, indicated by the intersection of the horizontal and vertical bars, have not shifted from their true positions in Figure 4.4(a). Finally, the only difference between Scenario 2 and Scenario 3 is the data depth; lower-mass (i.e. fainter) stars are included in Scenario 3 that were removed in Scenario 2. This is made clear by comparing Figure 4.4(c) and Figure 4.4(d).

To summarize, all three scenarios come from the simulated cluster presented in Figure 4.4(a), which is simulated without errors on the photometric magnitudes. For Scenarios 1 and 2 we assume different Gaussian measurement errors on the simulated photometry, which is truncated at $F275W = 23$ in both scenarios. We then fit the intentionally misspecified models, which are misspecified since there are no actual measurement errors, in order to examine how the prior distribution on mass influences the posterior distribution. For Scenarios 2 and 3 we use the same assumed Gaussian measurement errors in model fitting, but truncate the simulated photometry at different depths in order to detect any shrinkage in the fitted masses towards the prior mean. For each scenario, we explore the posterior distribution using an adaptive Metropolis algorithm that is described in detail in Section 5.3.2.2.

Since for this investigation we are interested in the posterior means of the masses and their effects on the other fitted parameters, we begin by comparing the marginal posterior distributions for each parameter under each scenario. In Figure 4.5 we present overlapping histograms representing the marginal posterior distributions for each parameter based on its MCMC draws. Yellow bars correspond to the marginal posterior distributions under Scenario 1, and blue bars correspond to the same under Scenario 2. Areas where the two distributions overlap appear green. These two scenarios have the same data depth, but differ in the Gaussian measurement errors assumed in model fitting. The solid dark-red vertical lines indicate the true value of each parameter under the simulation. As expected, the larger Gaussian measurement errors assumed when fitting the model under Scenario 2 results in wider marginal posterior distributions for each parameter. However, we also notice that while the posterior means, which we consider to be the fitted parameter values, under Scenario 1 are close to the true values of the parameters under the simulation, for Scenario 2

(a) Simulated Cluster CMD

(b) CMD for Scenario 1



(c) CMD for Scenario 2

(d) CMD for Scenario 3

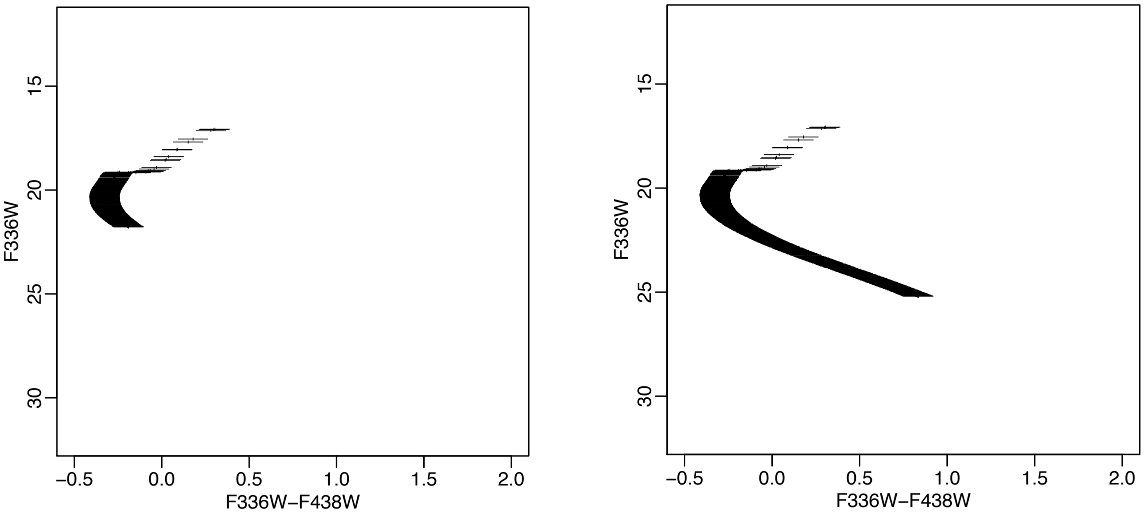


Figure 4.4: A simulated globular cluster. *(a)* The CMD from a simulated globular cluster containing 5000 stars, with parameters $\theta_{\text{age}} = 10.08$, $\theta_{m-M_V} = 15.375$, $\theta_{A_V} = 0.372$, $\theta_{\text{[Fe/H]}} = -1.5$, and $\phi_Y = 0.27$. Black dots represent the position of the simulated stars, and the dark-red curve is the isochrone given the true parameter values under the simulation. *(b)* The CMD corresponding to Scenario 1. The simulated photometry presented in (a) is truncated at a depth of $F275W = 23$. The horizontal and vertical bars represent the Gaussian measurement error assumed in model fitting, which has a standard deviation equal to 0.01. Note that the vertical errors are difficult to distinguish due to the scale of the vertical axis. *(c)* The CMD corresponding to Scenario 2. The only difference from (b) is in the Gaussian measurement error assumed in model fitting, which has a standard deviation of 0.03 for this scenario. *(d)* The CMD corresponding to Scenario 3. The only difference from (c) is that the simulated photometry is truncated at $F275W = 27.5$ instead of at $F275W = 23$.

there is a perceptible deviation for every parameter. This deviation is particularly noticeable for age, metallicity, and helium abundance.

To explore this phenomena further, we compare the marginal posterior distributions under Scenario 2 to those under Scenario 3, where now the only difference between the two scenarios is the data depth. The overlapping histograms for this comparison are presented in Figure 4.6. There, the blue bars again correspond to the marginal posterior distributions under Scenario 2. The magenta bars correspond to the marginal posterior distributions under Scenario 3, and overlap between the marginal posterior distributions under the different scenarios appears lavender. The true parameter values under the simulation are again designated by the dark-red vertical lines. Here, we observe a slight decrease in the width of the marginal posterior distribution under Scenario 3, which we attribute to the greater number of stars (i.e., larger sample size). However, the deviation in the fitted parameters remains.

Finally, in Figure 4.7 we examine the fitted mass values. The blue 'x's denote the difference between the fitted mass and the true mass for each star, and the grey vertical bars are 95% intervals. The dark-red horizontal line indicates where the difference between the fitted mass and the true mass is equal to zero. The results we obtain indicate non-standard asymptotics. While we typically expect the deviation between the fitted parameters and the true parameter values under the simulation to decrease as the amount of data increases, we instead notice that the deviation increases with data depth. Furthermore, when the Gaussian measurement error assumed for model fitting is small, the fitted masses are close to the true masses; see the left panel of Figure 4.7. When the only change to the scenario is an increase in the Gaussian measurement error assumed for model fitting, there is more deviation between the fitted masses and the true masses under the simulation; see the middle panel of Figure 4.7. This suggests that the prior distribution on mass is influencing the posterior means. However, our initial suspicion that the fitted masses are shrinking towards the prior mean is incorrect; see the right panel of Figure 4.7. Although the prior mean is 0.55 M_{Sun}, the fitted masses for all

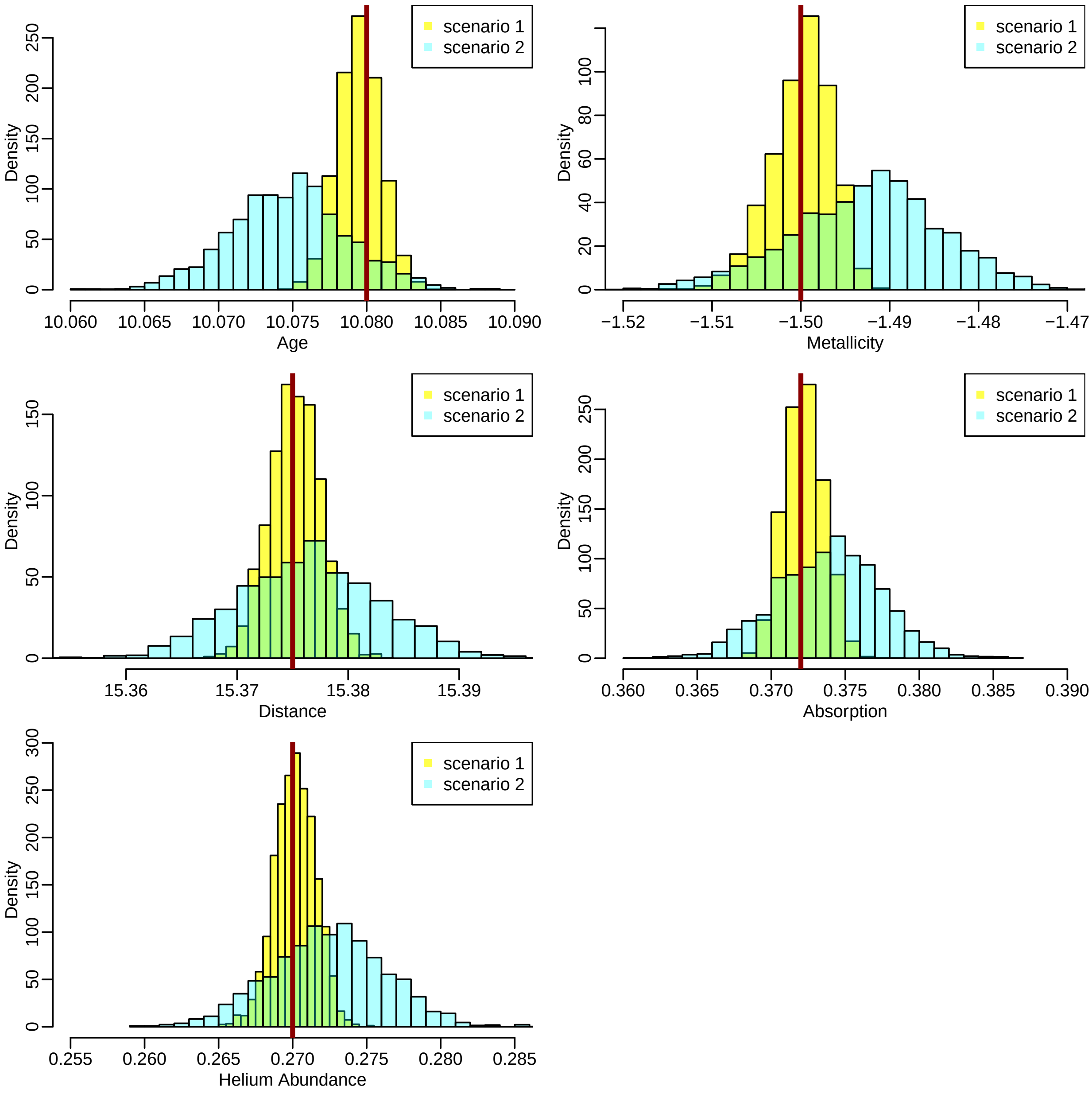


Figure 4.5: Comparison of Scenario 1 and Scenario 2. The dark-red vertical lines indicate the true parameter values under the simulation. The yellow histogram is the marginal posterior distribution under Scenario 1, in which we assume small Gaussian measurement errors in model fitting. The blue histogram is the marginal posterior distribution under Scenario 2, in which we assume larger Gaussian measurement errors in model fitting. When the assumed error is smaller, the posterior means (a.k.a. the fitted parameters) are close to the true parameter values under the simulation. When the assumed error is larger, the marginal posterior distributions are wider and the posterior means exhibit noticeable deviation from the true parameter values under the simulation.

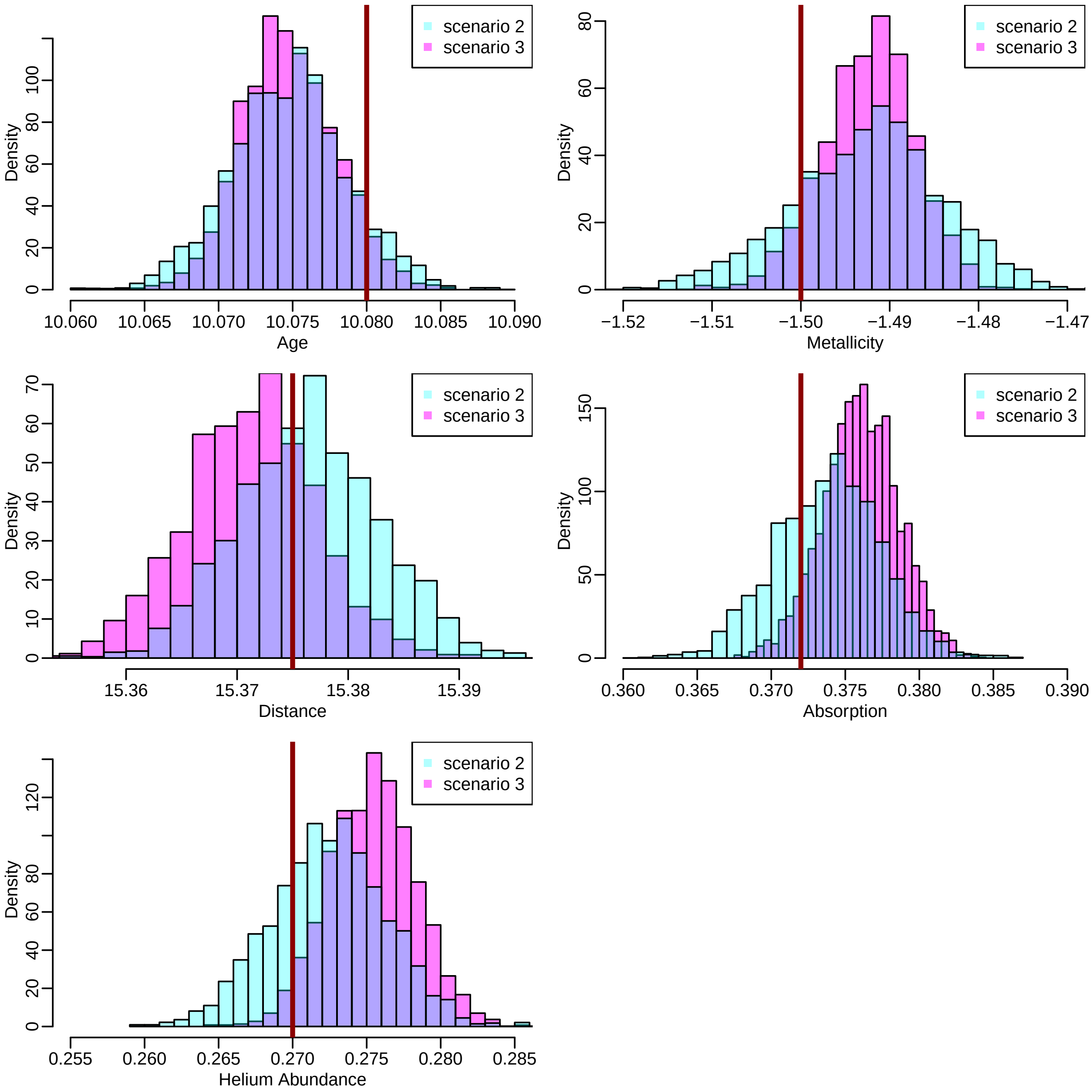


Figure 4.6: Comparison of Scenario 2 and Scenario 3. As in Figure 4.5, the dark-red vertical lines indicate the true parameter values under the simulation, and the blue histogram represents the marginal posterior distribution under Scenario 2. Here, the magenta histogram represents the marginal posterior distribution under Scenario 3. The overlap of the two distributions appears lavender. We notice that although there is a slight reduction in the width of the marginal posterior distributions from Scenario 2 to Scenario 3, attributed to the greater number of stars under Scenario 3, both scenarios exhibit deviations in the fitted parameters from the true parameter values under the simulation.

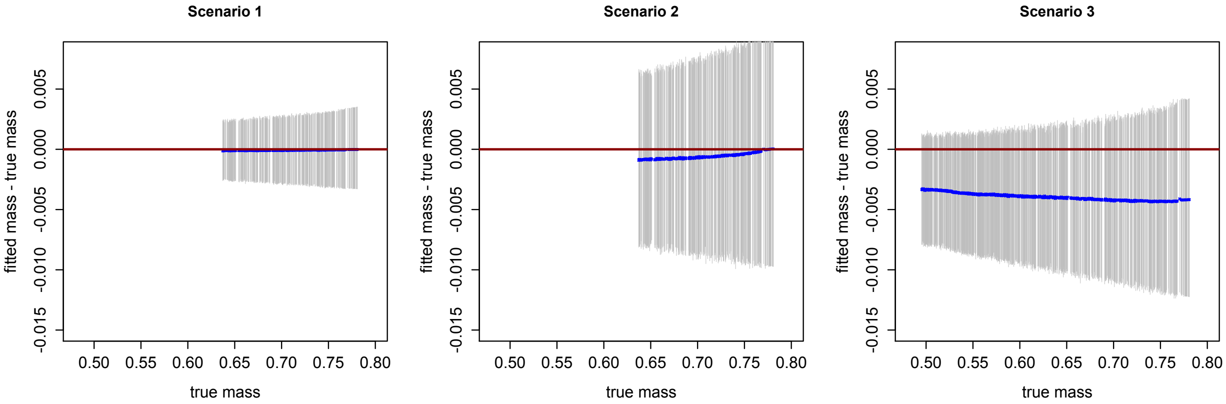


Figure 4.7: Difference between the fitted masses and the true masses. The blue ‘x’s denote the difference between the fitted mass and the true mass for each star, and the vertical grey bars are 95% intervals. We notice deviations for the fitted masses that are likely filtering through to other parameter estimates.

the stars are smaller than their true masses, including stars for which the true mass is less than the prior mean. As a result, we conclude that the fitted masses are instead shrinking towards the prior mode, which is at 0.1 M_{Sun}. The cause is that as the sample size increases, the influence of the prior distribution on mass does not diminish because there is only one observation (i.e. one star) per mass parameter. There is no ability to share strength for these parameters because we do not fit the distribution of the masses. By considering Figures 4.5, 4.6, and 4.7 together, we believe that deviations in the fitted masses from the true masses are filtering through to the other parameter estimates.

To gain intuition into the mechanism producing the deviations between the fitted masses and the true masses under the simulation, we consider the following toy-example. First, we model mass, m, as a Gaussian random variable with mean $\tilde{m}$ and known standard deviation ξ; i.e., $m \sim N(\tilde{m}, \xi)$. We then specify a Gaussian prior distribution on the base 10 logarithm of $\tilde{m}$:

$$P\left(\log_{10}(\tilde{m})\right) \propto \exp\left(-\frac{1}{2}\left(\frac{\log_{10}(\tilde{m}) + 1.02}{0.677}\right)^2\right), \tag{4.4}$$

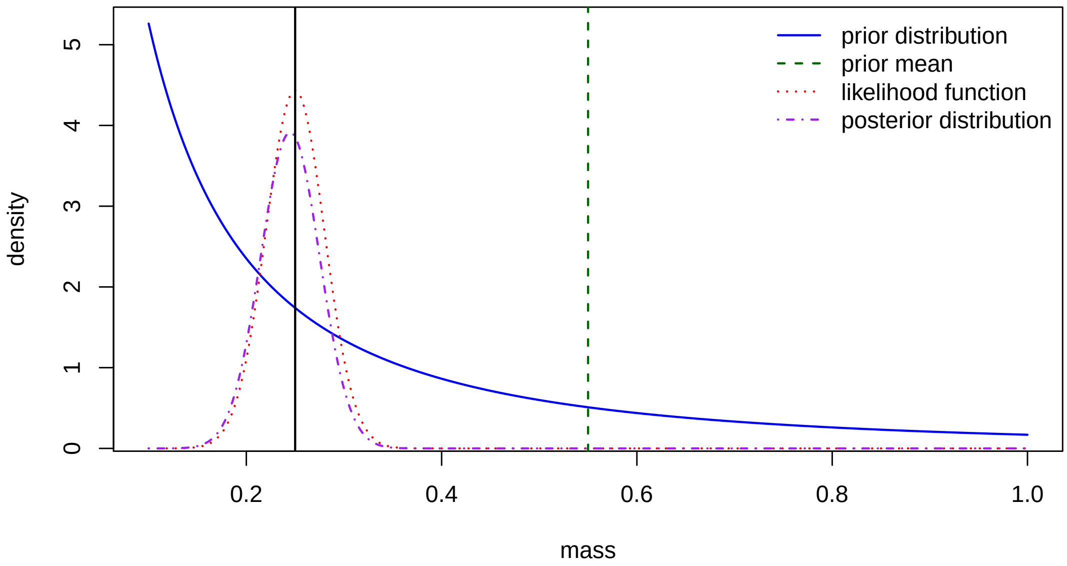


Figure 4.8: Toy-example. The solid blue curve is the prior distribution on $\tilde{m}$, with the prior mean indicated by the dashed green vertical line. The dotted red curve is the likelihood function. We observe an initial mass $m = \tilde{m} = 0.25\ M_{\text{Sun}}$, indicated by the solid black vertical line. The posterior distribution of $\tilde{m}$ is the purple dashed-dotted curve, which we notice is skewed towards the prior mode.

truncated to the range $0.1\ M_{\text{Sun}}$ to $8\ M_{\text{Sun}}$. Now, suppose we observe $m = \tilde{m}$. This scenario is presented in Figure 4.8, with $\tilde{m} = 0.25\ M_{\text{Sun}}$ and $\xi = 0.03$. There, the solid blue curve is the prior distribution on $\tilde{m}$, and the dashed green vertical line is the prior mean. The dotted red curve is the likelihood function, and the purple dashed-dotted curve is the posterior distribution. We observe that the posterior distribution of $\tilde{m}$ is slightly skewed towards the prior mode, which is where the prior distribution is truncated at $0.1\ M_{\text{Sun}}$.

The situation we are investigating is more complicated than the simple story we tell with the above toy-example. We do not measure masses directly, and there are complex nonlinear correlations between the masses and the other parameters. Nevertheless, we believe that a similar effect is leading to the non-standard asymptotics we observe in the parameter estimates.

When conducting numerical studies such as those above, we do not need to worry about systematic errors stemming from the computer-based stellar evolution model; the same computer model is used for both simulation and model fitting. However, for a real data analysis we must consider that the computer-based stellar evolution model contains inaccuracies and/or approximations. As a result, posterior intervals computed under a particular model underestimate the actual uncertainty (van Dyk *et al.*, 2009). We expect the deviations we observe in our above investigation to be small compared to the systematic errors in the underlying physics models. In that sense, the deviations we observe are not astrophysically meaningful. Nonetheless, the deviations caused by the influence of the prior distribution on mass will affect our numerical results in subsequent chapters.

Chapter 5

Multiple Stellar Populations in Galactic Globular Clusters

5.1 Introduction

Globular clusters have long been used by astrophysicists as probes into the formation and evolution of galaxies (e.g., Sandage, 1962; Searle and Zinn, 1978; Janes and Demarque, 1983; Lee *et al.*, 2001; Marín-Franch *et al.*, 2009; Forbes and Bridges, 2010). Much of the past work on globular clusters has assumed the "simple model" of star cluster formation described in Section 4.1.2—stars in a cluster form at (roughly) the same time from a common molecular cloud. This classical formation history produces globular clusters that are comprised of *simple stellar populations*—groups of stars with the same age that share certain physical properties (e.g., helium abundance) due to their being born from the same molecular cloud. However, within the past decade, this assumption has been called into question as numerous studies have produced evidence that globular clusters in fact host multiple distinct stellar populations (e.g., Bedin *et al.*, 2004; Gratton *et al.*, 2004; Carretta *et al.*, 2006; Villanova *et al.*, 2007; Piotto *et al.*, 2007; Piotto, 2009; Milone *et al.*, 2012). The implication is that (at least some) globular clusters have undergone multiple epochs of star formation. As a result, the "simple model" now describes the formation of a single stellar population, with globular clusters being a mix of multiple populations.

The first hints of multiple populations in globular clusters were provided by Cannon and Stobie (1973) in their investigation of the globular cluster Omega Centauri (ω Cen). Their evidence was that the red giants in the CMD were more spread out than known sources of variation (e.g., photometric errors, field stars) could explain. Similarly, Hesser and Bell (1980) found discrepancies in the CMD of the globular cluster 47 Tuc that could not be explained given the simple stellar population assumption.

While the aforementioned studies, and several others, suggested that the simple model of globular cluster formation may be inaccurate, unequivocal evidence was not discovered until 2004. Using the Hubble Space Telescope, Bedin *et al.* (2004) obtained high-precision photometric magnitudes of the stars in ω Cen. The CMD produced from these photometric magnitudes show several distinct "strands" in the main-sequence and red-giant groups, which is incontrovertible evidence of multiple stellar populations (see panel (b) of Figure 1 of Bedin *et al.*, 2004). Two CMDs for the globular cluster NGC 6352, which we analyze in Section 5.5.3, are presented in Figure 5.1. The left panel of Figure 5.1 is a typical CMD, constructed with a difference of two apparent magnitudes on the horizontal axis. The red-giant stars are more spread out than we would expect for a simple stellar population. However, visually separating the two populations using only this CMD is challenging. To reveal the multiple populations, we construct a non-typical CMD using a difference of differences in apparent magnitudes on the horizontal axis; see the right panel of Figure 5.1. With this construction, the distinct strands in the main-sequence and red-giant groups that are indicative of multiple populations are now apparent. (As there are no white dwarfs in the observed data, both CMDs lack a white-dwarf group.)

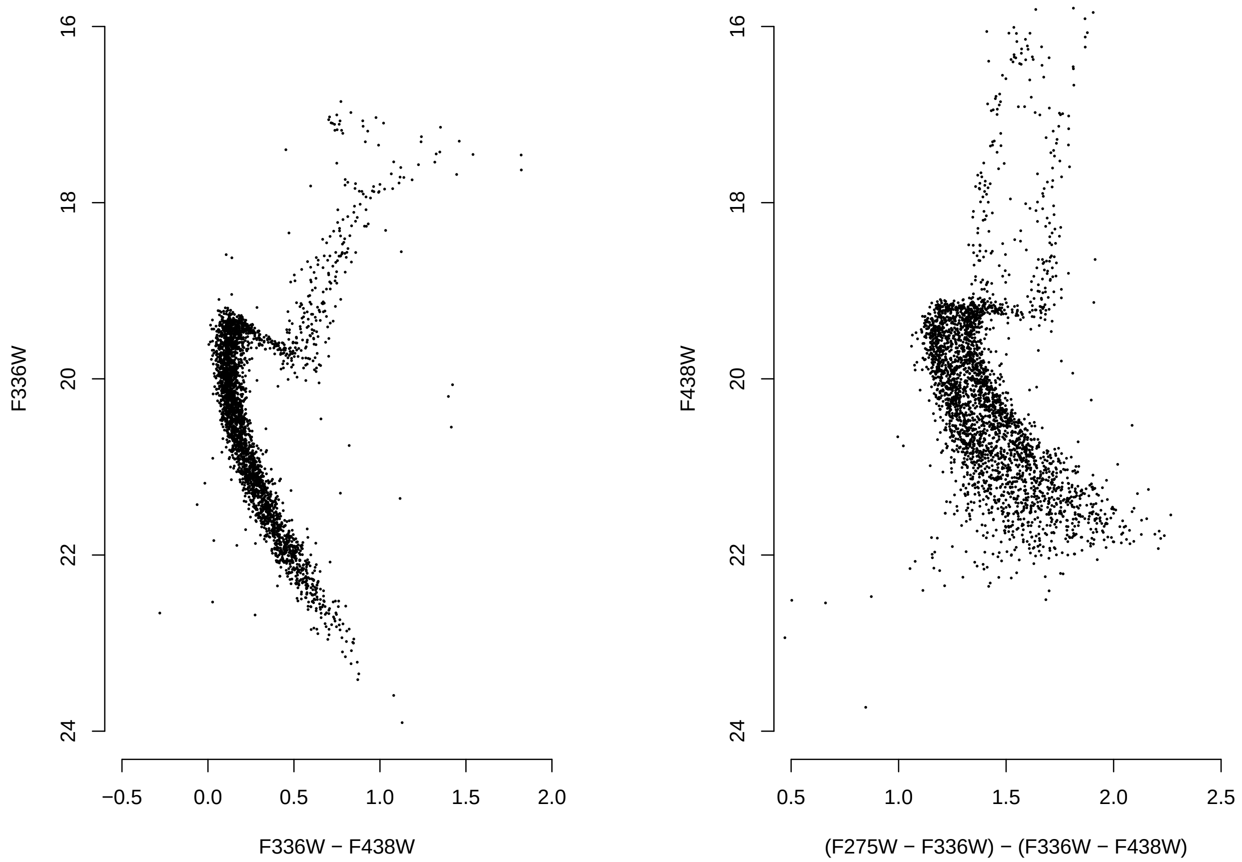


Figure 5.1: Two CMDs for the globular cluster NGC 6352. *Left*: A typical CMD, constructed using a difference of two apparent magnitudes on the horizontal axis. While the spread of the red giant stars is greater than expected for a simple stellar population, multiple populations are not easily distinguishable. *Right*: A CMD constructed using a difference of differences in apparent magnitudes on the horizontal axis. With this construction, the multiple strands in the main-sequence and red-giant groups that indicate multiple stellar populations are revealed.

Multiple populations in globular clusters imply multiple epochs of star formation, which produce at least two distinct populations of stars. When multiple epochs of star formation occur, it is expected that later generations of stars "recycle" material from earlier generations. As described in Section 4.1.1, helium and even heavier elements are produced in the cores of stars via nucleosynthesis. When very massive stars explode as supernovas, these elements are released and enrich the molecular clouds from which the next generation of stars will form.

That is, "second generation" stars form from the ejecta of "first generation" stars (e.g., Bekki *et al.*, 2007; Prantzos *et al.*, 2007; D'Antona and Caloi, 2008; Carretta *et al.*, 2010; Gratton and Carretta, 2010; Bekki, 2011). This reuse of stellar material leads to different elemental abundances, which can then be observed via spectroscopy or photometry. As a result of complex astrophysical processes, the major observed differences of multiple populations are driven by variation in helium abundance, ϕ_Y in Table 4.1.

When working with photometric magnitudes, the multiple populations are most prominent in the ultraviolet wavelengths (UV). While many previous studies of globular clusters were in the visual wavelengths, recent high-quality UV photometric data from the Hubble Space Telescope allow us to better investigate the presence of multiple stellar populations. In fact, the vast majority of globular clusters that have been studied in the UV to sufficient accuracy display characteristics that can be attributed to multiple populations. UV filters are particularly sensitive to variations in elemental abundances, such as the difference in helium abundance. For this reason, we incorporate the improved main-sequence and red-giant models of Dotter *et al.* (2008, as updated at `http://stellar.dartmouth.edu/~models/`) into BASE-9 that predict UV photometric magnitudes. While spectral data may provide better abundance accuracy for individual elements, photometric data is actually preferable due to the significantly greater number of stars per cluster for which photometric magnitudes are available. It is partly for this reason that we use photometric data in our analyses.

Hundreds of orbits of the Hubble Space Telescope and many nights on large ground-based telescopes have been devoted to observing globular clusters in particular. Careers in astrophysics have been devoted to developing models of stellar evolution. The statistical methods employed to fit the costly models to the expensive data, however, are rudimentary. Astrophysicists typically plot the stellar evolution models on top of the observed data and adjust the parameters until an acceptable fit is achieved, where the goodness of the fit is determined by visual inspection. This technique, known as a "chi-by-eye" approach, yields inaccurate results and cannot capture uncertainties in the model fits even when studying open clusters with single stellar populations (van Dyk *et al.*, 2009). When studying globular clusters that host multiple stellar populations, the “chi-by-eye” method completely fails since each population may have subtle differences in (at least) one parameter, and the stellar populations overlap in non-obvious ways in CMDs.

In this chapter we present a Bayesian hierarchical model for multiple-population globular clusters with two stellar populations, extending the model for single-population open clusters developed by DeGennaro *et al.* (2009), van Dyk *et al.* (2009), and Stein *et al.* (2013). This multiple-population model assumes that a globular cluster hosts two stellar populations that differ in helium abundance. Estimating the difference in helium abundance will provide astrophysicists insight into the possible mechanisms that produce multiple-population clusters. To fit the more complex multiple-population model, we implement an adaptive Metropolis algorithm (e.g., Haario *et al.*, 2001; Roberts and Rosenthal, 2009; Rosenthal *et al.*, 2011). This algorithm has the benefit of improving convergence without requiring much tuning by the user. The latter is important since the multiple-population model and improved MCMC techniques are incorporated into the current version of BASE-9, which is currently being used by the astrophysics community to explore the astrophysical implications of multiple-population globular clusters. As far as we are aware, ours is the only statistically principled approach for estimating the parameters of multiple-population globular clusters.

The rest of this chapter is divided into five sections. In Section 5.2 we present our statistical model for a multiple-population globular cluster with two stellar populations that differ in helium abundance. In Section 5.3 we discuss the computational challenges involved with fitting our statistical model, and in particular describe how adaptive MCMC techniques improve convergence. In Section 5.4 we present the capabilities of our model and methods with a series of numerical studies. In Section 5.5 we present the results of fitting our multiple-population model to UV photometric magnitudes from three globular clusters: NGC 6352, NGC 5272, and NGC 0288. Finally, in Section 5.6 we summarize our results and discuss directions of future research.

5.2 Statistical Model for Multiple Population Globular Clusters

5.2.1 Extending the Hierarchical Model

The model in Chapter 4 is based on a hierarchy of n photometric magnitudes within stars and of stars within a cluster. Each star is described by is primary mass, relative mass, and cluster-membership indicator. The cluster is defined by its age, metallicity, distance, and absorption. (We can ignore carbon fraction since the multiple-population clusters we study do not contain white dwarfs in the observed data.) Here, we add a layer to that hierarchy. Specifically, we add *population parameters*—parameters which are common to all stars belonging to a particular stellar population, but which vary across populations in a cluster. The population parameters that we add are the helium abundance and proportion of stars for population k, which we denote with ϕ_{Yk} and ϕ_{pk}, respectively. We refer to the population with the smaller helium abundance as "Population 1," and the population with the larger helium abundance as "Population 2." The hierarchy of cluster parameters, population parameters, and stellar parameters is illustrated in Figure 5.2.

5.2.2 The Likelihood Function

The likelihood for a "single" population cluster is given in (4.2), which itself can be viewed as a mixture of two populations: cluster stars and field stars. We now focus on clusters that host two distinct stellar populations and model each cluster star with a two-component finite

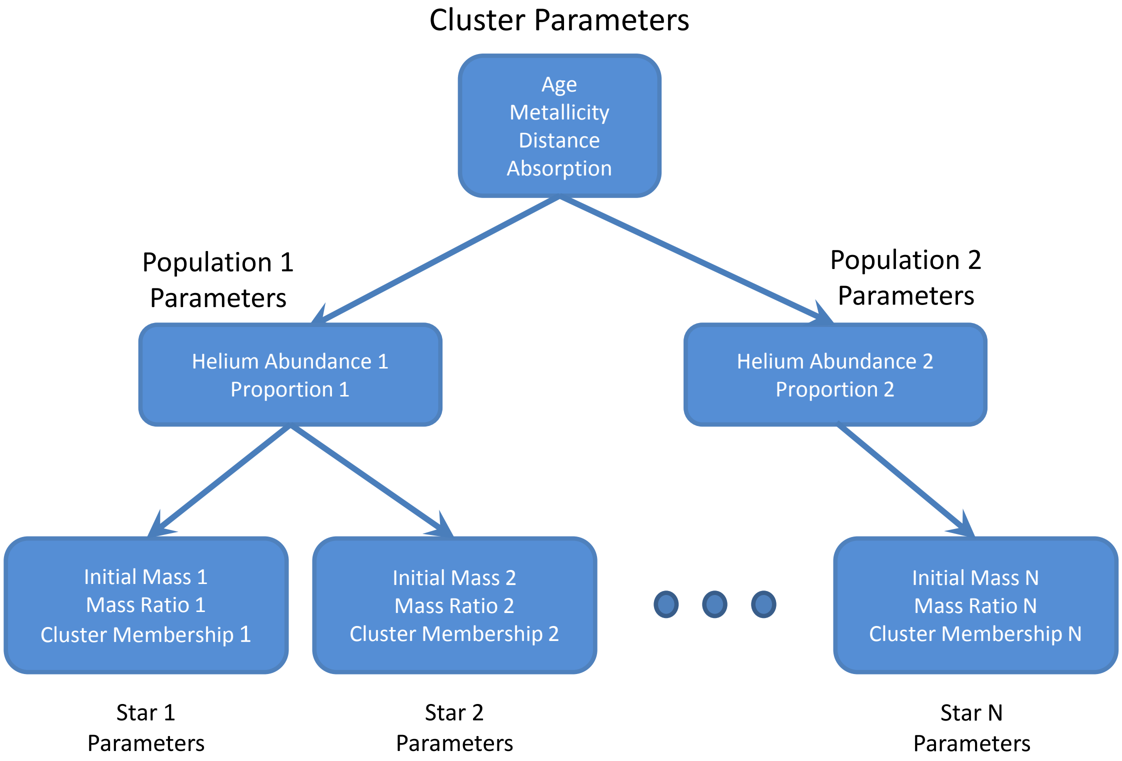


Figure 5.2: Hierarchy of cluster, population, and stellar parameters for a two-population globular cluster. The cluster parameters—age, metallicity, distance, and absorption—are common to all stars in the cluster. The population parameters—helium abundance and the proportion of stars in a particular population—are common to all stars in a population but may be different between populations. The stellar parameters—initial mass, mass ratio, and cluster membership indicator—are allowed to vary on a star-by-star basis.

mixture of n-dimensional multivariate Gaussian distributions. This results in three stellar populations: field stars and two cluster populations. The likelihood function is then

$$\begin{aligned} & L(\boldsymbol{M},\boldsymbol{R},\boldsymbol{\Theta},\boldsymbol{\Phi},\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\Sigma}) \\ = \quad & \prod_{i=1}^{N}\Bigg[Z_i \times \Bigg\{ \frac{\phi_{p1}}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{1}{2}\Big(\boldsymbol{X}_i-\boldsymbol{\mu}_{i1}\Big)^\top \boldsymbol{\Sigma}_i^{-1}\Big(\boldsymbol{X}_i-\boldsymbol{\mu}_{i1}\Big)\right) \\ & + \frac{\phi_{p2}}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{1}{2}\Big(\boldsymbol{X}_i-\boldsymbol{\mu}_{i2}\Big)^\top \boldsymbol{\Sigma}_i^{-1}\Big(\boldsymbol{X}_i-\boldsymbol{\mu}_{i2}\Big)\right) \Bigg\} \\ & + (1-Z_i)\times P(\boldsymbol{X}_i|Z_i=0)\Bigg], \end{aligned} \tag{5.1}$$

where $\mathbf{\Theta} = (\theta_{\text{age}}, \theta_{[\text{Fe/H}]}, \theta_{m-M_V}, \theta_{A_V})$, and $\mathbf{\Phi} = (\phi_{Y1}, \phi_{Y2}, \phi_{p1}, \phi_{p2})$. Here, $\boldsymbol{\mu}_{ik}$ is given by

$$\boldsymbol{\mu}_{ik} = -2.5\log_{10}\left(10^{-\boldsymbol{G}(M_i,\mathbf{\Theta},\phi_{Yk})/2.5} + 10^{-\boldsymbol{G}(M_iR_i,\mathbf{\Theta},\phi_{Yk})/2.5}\right), \tag{5.2}$$

where $\boldsymbol{G}$ is the computer-model for stellar evolution with the updated Dotter *et al.* (2008) models for main-sequence and red giant stars that output UV predicted magnitudes. Since we are modeling clusters that are comprised of two stellar populations, $\phi_{p1} + \phi_{p2} = 1$.

5.2.3 Prior Distributions

For $\mathbf{\Theta}$, $\boldsymbol{M}$, and $\boldsymbol{R}$, we use the same prior distributions as described in Section 4.2.3. Since the populations parameters, $\mathbf{\Phi} = (\phi_{Y1}, \phi_{Y2}, \phi_{p1}, \phi_{p2})$, are the primary parameters of scientific interest, we use uniform prior distributions subject to physical constraints on their ranges. A uniform prior distribution on the interval $[0.15, 0.3]$ is used for ϕ_{Y1}. Similarly, a uniform prior distribution on the interval $[0.15, 0.4]$ is used for ϕ_{Y2}, and we impose the constraint $\phi_{Y2} > \phi_{Y1}$. In both cases the intervals for the uniform prior distributions were determined based on current astrophysical understanding of the formation and evolution of the Universe. Since $\phi_{p2} = 1 - \phi_{p1}$, we only need to specify a prior distribution for ϕ_{p1}. Because we do not typically have any a priori knowledge about the proportion of stars in each population, ϕ_{p1} is given a uniform prior distribution on the interval $[0, 1]$.

Many of the multiple-population clusters we study will have ancillary information that we can use to help separate field stars from cluster stars in our observed data. Until such information is made available, we assume $P(Z_i = 1) = \alpha$ for $i = 1, \ldots, N$ and investigate sensitivity to the choice of α.

5.3 Statistical Computation

5.3.1 Marginalization via Numerical Integration

For analyzing two-population globular clusters, we adopt an approach that is initially similar to Stein *et al.* (2013), in that we marginalize over $\boldsymbol{M}$, $\boldsymbol{R}$, and $\boldsymbol{Z}$. With the full joint posterior distribution denoted by $P(\boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{M}, \boldsymbol{R}, \boldsymbol{Z} \mid \boldsymbol{X})$, the marginal posterior distribution of $(\boldsymbol{\Theta}, \boldsymbol{\Phi})$ is given by

$$P(\boldsymbol{\Theta}, \boldsymbol{\Phi} \mid \boldsymbol{X}) = \int \cdots \int \left(\sum_{Z_1} \cdots \sum_{Z_N} P(\boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{M}, \boldsymbol{R}, \boldsymbol{Z} \mid \boldsymbol{X}) \right) d\boldsymbol{M} d\boldsymbol{R} \tag{5.3}$$

$$\propto P(\boldsymbol{\Theta}, \boldsymbol{\Phi}) \prod_{i=1}^{N} \left[c(\boldsymbol{\Theta}, \boldsymbol{\Phi}) P(Z_i = 1) + P(\boldsymbol{X}_i | Z_i = 0) P(Z_i = 0) \right], \tag{5.4}$$

where

$$c(\boldsymbol{\Theta}, \boldsymbol{\Phi}) = \int \int P(\boldsymbol{X}_i \mid M_i, R_i, \boldsymbol{\Theta}, \boldsymbol{\Phi}, Z_i = 1) P(M_i, R_i \mid Z_i = 1) dM_i dR_i. \tag{5.5}$$

When $P(Z_i = 1) = \alpha$ for $i = 1, \ldots, N$, the marginalization over $\boldsymbol{Z}$ simplifies and (5.3) and (5.4) reduce to $P(\boldsymbol{\Theta}, \boldsymbol{\Phi} \mid \boldsymbol{X}) \propto$

$$\prod_{i=1}^{N} \int \int P(\boldsymbol{\Theta}, \boldsymbol{\Phi}) \Bigg[\alpha \times \Bigg\{ \frac{\phi_{p1}}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{1}{2} \Big(\boldsymbol{X}_i - \boldsymbol{\mu}_{i1} \Big)^{\top} \boldsymbol{\Sigma}_i^{-1} \Big(\boldsymbol{X}_i - \boldsymbol{\mu}_{i1} \Big) \right)$$
$$+ \frac{\phi_{p2}}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{1}{2} \Big(\boldsymbol{X}_i - \boldsymbol{\mu}_{i2} \Big)^{\top} \boldsymbol{\Sigma}_i^{-1} \Big(\boldsymbol{X}_i - \boldsymbol{\mu}_{i2} \Big) \right) \Bigg\} P(M_i, R_i \mid Z_i = 1) dM_i dR_i$$
$$+ (1 - \alpha) \times P(\boldsymbol{X}_i | Z_i = 0) \Bigg], \tag{5.6}$$

This integral cannot be evaluated analytically since $P(\boldsymbol{X}_i \mid M_i, R_i, \boldsymbol{\Theta}, \boldsymbol{\Phi}, Z_i = 1)$ depends on M_i and R_i through $\boldsymbol{G}$, the computer-based stellar evolution model (Stein *et al.*, 2013). Instead, we employ brute-force numerical integration via Riemann sums. By marginalizing out the $3N$ stellar parameters (or $2N+1$ when $P(Z_i = 1) = \alpha$ for $i = 1, \ldots, N$), we reduce the dimension of the posterior distribution from thousands to just seven.

5.3.2 Adaptive MCMC

Since the remaining parameter vector $(\boldsymbol{\Theta}, \boldsymbol{\Phi})$ after marginalizing out $\boldsymbol{M}$, $\boldsymbol{R}$, and $\boldsymbol{Z}$ is just seven-dimensional, we initially implemented a Metropolis algorithm (Metropolis *et al.*, 1953) to update all the parameters together in a single jump within each iteration. However, we found the trial-and-error approach to tuning the proposal distribution to be difficult, which is not surprising given the correlations among the the components of $(\boldsymbol{\Theta}, \boldsymbol{\Phi})$ in the computer-based stellar evolution model. Furthermore, although finding an optimal proposal distribution by fine-tuning may have eventually been successful, we intend BASE-9 to be an easy-to-use software suite for the astrophysics community. As such, requiring users to arduously hand-tune proposal distributions is suboptimal. To avoid this, we implement an *Adaptive Metropolis* (AM) algorithm (e.g., Haario *et al.*, 2001; Roberts and Rosenthal, 2009; Rosenthal *et al.*, 2011) to sample from the marginal posterior distribution, $P(\boldsymbol{\Theta}, \boldsymbol{\Phi} | \boldsymbol{X})$.

5.3.2.1 Ergodicity of Adaptive MCMC

We follow the example of Rosenthal *et al.* (2011) to devise an algorithm that improves our sampler. To start, suppose we have an MCMC algorithm that at iteration n updates $\mathbf{Y}_n$ to $\mathbf{Y}_{n+1}$ with a Markov chain kernel P_{Γ_n}, where $\mathbf{Y}_n$ is the state of the chain at iteration n. Here, each fixed kernel P_γ has the same stationary distribution, π, and Γ_n is the specific kernel at iteration n, such that, for $n = 0, 1, 2, \ldots$

$$P\left(\mathbf{Y}_{n+1} \in \mathbf{A} | \mathbf{Y}_n = \mathbf{y}, \Gamma_n = \gamma, \mathbf{Y}_{n-1}, \ldots, \mathbf{Y}_0, \Gamma_{n-1}, \ldots, \Gamma_0\right) = P_\gamma\left(\mathbf{y}, \mathbf{A}\right). \tag{5.7}$$

The goal is to update $\{\Gamma_n\}$ with an adaptive updating scheme in such a way that the adaptation does not destroy the stationary distribution of the chain.

As recounted in Rosenthal *et al.* (2011), Roberts and Rosenthal (2007) proved that two sufficient conditions guarantee proper convergence: (i) the *Diminishing* (a.k.a. *Vanishing*) *Adaptation* condition and (ii) the *Containment* (a.k.a. *Bounded Convergence*) condition. The latter condition is that

$$\{M_\varepsilon\left(\mathbf{Y}_n, \Gamma_n\right)\}_{n=0}^{\infty} \qquad \text{is bounded in probability,} \qquad \varepsilon > 0, \tag{5.8}$$

where $\{M_\varepsilon\left(\mathbf{y}, \gamma\right) = \inf\{n \geq 1 : ||P_\gamma^n\left(\mathbf{y}, \cdot\right) - \pi\left(\cdot\right)|| \leq \varepsilon\}$ is the convergence time of P_γ when starting in state $\mathbf{y} \in \mathcal{Y}$, where $\mathcal{Y}$ is the state space. This condition is almost always satisfied except is specially constructed examples, and so can usually be ignored in practice (Rosenthal *et al.*, 2011).

The Diminishing Adaptation condition, however, is key. Formally, the condition is

$$\lim_{n \to \infty} \sup_{\mathbf{y} \in \mathcal{Y}} ||P_{\Gamma_{n+1}}\left(\mathbf{y}, \cdot\right) - P_{\Gamma_n}\left(\mathbf{y}, \cdot\right)|| = 0 \qquad \text{in probability.} \tag{5.9}$$

As its name implies, the Diminishing Adaptation condition states that the amount of adaptation at iteration n must go to 0 as $n \to \infty$. Care must be taken to ensure that this condition holds for all adaptive MCMC schemes (Rosenthal *et al.*, 2011).

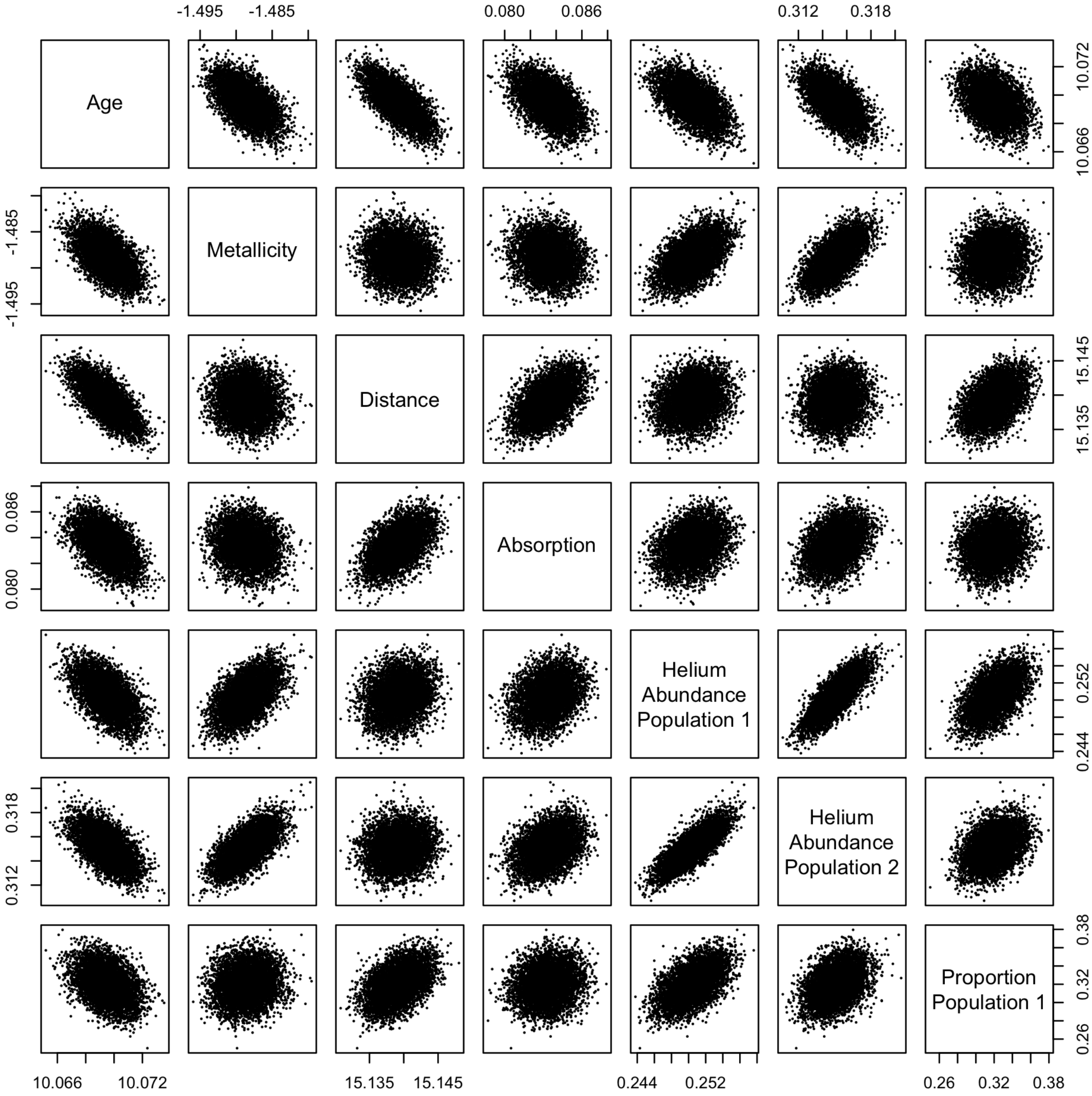


Figure 5.3: Posterior draws from the marginal posterior distribution $P(\mathbf{\Theta}, \mathbf{\Phi}|\boldsymbol{X})$. The data used are photometric magnitudes from the globular cluster NGC 5272. From these draws, $P(\mathbf{\Theta}, \mathbf{\Phi}|\boldsymbol{X})$ appears roughly Gaussian.

5.3.2.2 Adaptive Metropolis Implementation

After marginalizing over $\boldsymbol{M}$, $\boldsymbol{R}$, and $\boldsymbol{Z}$, the resulting marginal posterior distribution $P(\mathbf{\Theta}, \mathbf{\Phi}|\boldsymbol{X})$ appears roughly Gaussian. This is illustrated in Figure 5.3, in which we display the matrix of two-dimensional scatterplots of 50,000 posterior draws from the marginal posterior distribution $P(\mathbf{\Theta}, \mathbf{\Phi}|\boldsymbol{X})$. The observed data are photometric magnitudes from the globular cluster NGC 5272. We specify Gaussian prior distributions on the cluster parameters (truncated to be positive for θ_{A_V}), with means set according to values in the updated Harris (1996) catalogue and standard deviations chosen to be reasonably informative. Uniform prior distributions subject to physical constraints on their ranges are used for the population parameters. More details about the prior distributions and the run that produced these posterior draws are provided in Section 5.5.2.

Based on results in Gelman *et al.* (1996), the optimal proposal distribution for a $N(0, \Sigma)$ target distribution is $N\left(0, \left[(2.38)^2/d\right]\Sigma\right)$, where Σ is a d-dimensional covariance matrix. To allow for more efficient jumping, we use a multivariate t proposal distribution instead of the multivariate normal. Specifically, for iteration $n+1$ we use a multivariate t proposal distribution with 6 degrees of freedom, centered at the current value of $(\mathbf{\Theta}, \mathbf{\Phi})$ and with scale equal to $\boldsymbol{\xi}_n$. Here, $\boldsymbol{\xi}_n$ is the empirical covariance matrix of the first n iterations. Since we recalculate $\boldsymbol{\xi}_n$ at every iteration, the proposal distribution adapts at every iteration based on the past history of the chain. In this way, we have constructed an AM sampling algorithm. Since $\boldsymbol{\xi}_n$ is an empirical average over the first n iterations, the Diminishing Adaptation condition is satisfied (Rosenthal *et al.*, 2011).

When implementing the AM algorithm, we first run a "tuning" period. The goal of this tuning period is not to obtain an optimal proposal distribution, but rather to sufficiently explore the posterior distribution and generate a reasonable $\boldsymbol{\xi}_0$ for the AM algorithm. The tuning period proceeds as follows:

1. Set $j = 0$. Draw $\mathbf{Y}_{d+1} \sim N(\mathbf{Y}_d, 25\mathbf{\Delta}_j)$ for $d = 1, \ldots, 99$, where $\mathbf{Y}_1$ is the starting value of the chain and $\mathbf{\Delta}_0$ is a diagonal covariance matrix with fixed variances, both of which are specified by the user. The constant factor of 25 is chosen so that the chain

takes “big steps” to explore the parameter space.

2. Set $j = j + 1$. If $j = 20$, jump to Step 5. Else, draw $\mathbf{Y}_{100j+k} \sim N(\mathbf{Y}_{100j+k-1}, 5\boldsymbol{\Delta}_{j-1})$ for $k = 1, \ldots 50$. During these iterations the chain takes “medium steps” to explore the parameter space, which may assist in jumping between modes. Next, draw $\mathbf{Y}_{100j+k} \sim N(\mathbf{Y}_{100j+k-1}, \boldsymbol{\Delta}_{j-1})$ for $k = 51, \ldots, 100$.
3. Calculate the acceptance rate, a, of iterations $100j + 51$ to $200j$. If $0.2 < a < 0.4$, proceed to Step 4. Else, set $\boldsymbol{\Delta}_j = \varsigma(a)\boldsymbol{\Delta}_{j-1}$, where $\varsigma(a)$ is given in Table 5.1, and return to Step 2.
4. Set $\boldsymbol{\Delta}_j = \boldsymbol{\Delta}_{j-1}$, then set $j = j + 1$. Draw $\mathbf{Y}_{100j+k} \sim N(\mathbf{Y}_{100j+k-1}, \boldsymbol{\Delta}_{j-1})$ for $k = 1, \ldots, 100$ and calculate a for iterations $100j + 1$ to $200j$. If $0.2 < a < 0.4$, proceed to Step 5. Else, set $\boldsymbol{\Delta}_j = \varsigma(a)\boldsymbol{\Delta}_{j-1}$ and return to Step 2.
5. Discard the first 100 draws produced during Step 1 and calculate the empirical covariance matrix of all remaining draws, which is then denoted by $\boldsymbol{\xi}_0$. Then terminate the tuning period.

Once we have calculated $\boldsymbol{\xi}_0$ from the tuning period, the first 1000 iterations of the AM algorithm use the multivariate t proposal distribution described above, with $\boldsymbol{\xi}_n = \boldsymbol{\xi}_0$ for $n = 1, \ldots, 1000$. This non-adapting period is necessary to generate a sufficient number of draws to estimate posterior covariances prior to adapting the proposal distribution. At iteration 1001 and every subsequent iteration, $\boldsymbol{\xi}_n$ is the empirical covariance matrix of the previous n iterations.

5.3.2.3 Improving Convergence

The effectiveness of our AM algorithm in improving convergence is demonstrated in Figure 5.4. There, we compare the performance of our AM algorithm to that of a non-adaptive

Table 5.1: Tuning Period Scaling Factors

range of a	$\varsigma(a)$
$a > 0.9$	2
$0.7 < a \leq 0.9$	1.8
$0.5 < a \leq 0.7$	1.5
$0.4 < a \leq 0.5$	1.2
$0.15 \geq a < 0.2$	1/1.5
$0.05 \geq a < 0.15$	1/1.8
$a < 0.05$	1/2

Metropolis algorithm for fitting our multiple-population model using photometric data from the globular cluster NGC 5272, which we describe in more detail in Section 5.5.2. The tuning period for the Metropolis algorithm follows the same procedure as the AM algorithm. For both algorithms, the trace plots we present are constructed from $25,000$ MCMC iterations. The proposal distribution in the AM algorithm begins adapting at iteration 1001, after which the algorithm quickly becomes more efficient. This leads to more efficient sampling and improved convergence, and requires less tuning by the user.

5.4 Numerical Results

As an initial test of our method, we simulate two-population globular clusters under three scenarios, with ten replicate clusters per scenario. The three scenarios differ in the percentage of stars belonging to population 1: 50%, 80%, and 100% for scenario 1, 2, and 3, respectively. Scenario 3 therefore contains ten replicates of a single-population cluster, which we intentionally (mis-)fit with our multiple-population model to demonstrate how model misspecification can be diagnosed. Each cluster is simulated with $\theta_{\text{age}} = 10.08$, $\theta_{m-M_V} = 15.375$, $\theta_{A_V} = 0.372$, and $\theta_{[\text{Fe/H}]} = -1.5$. These are the same "average" cluster parameters for globular clusters that we used for our numerical study in Section 4.3. Each cluster is also simulated with

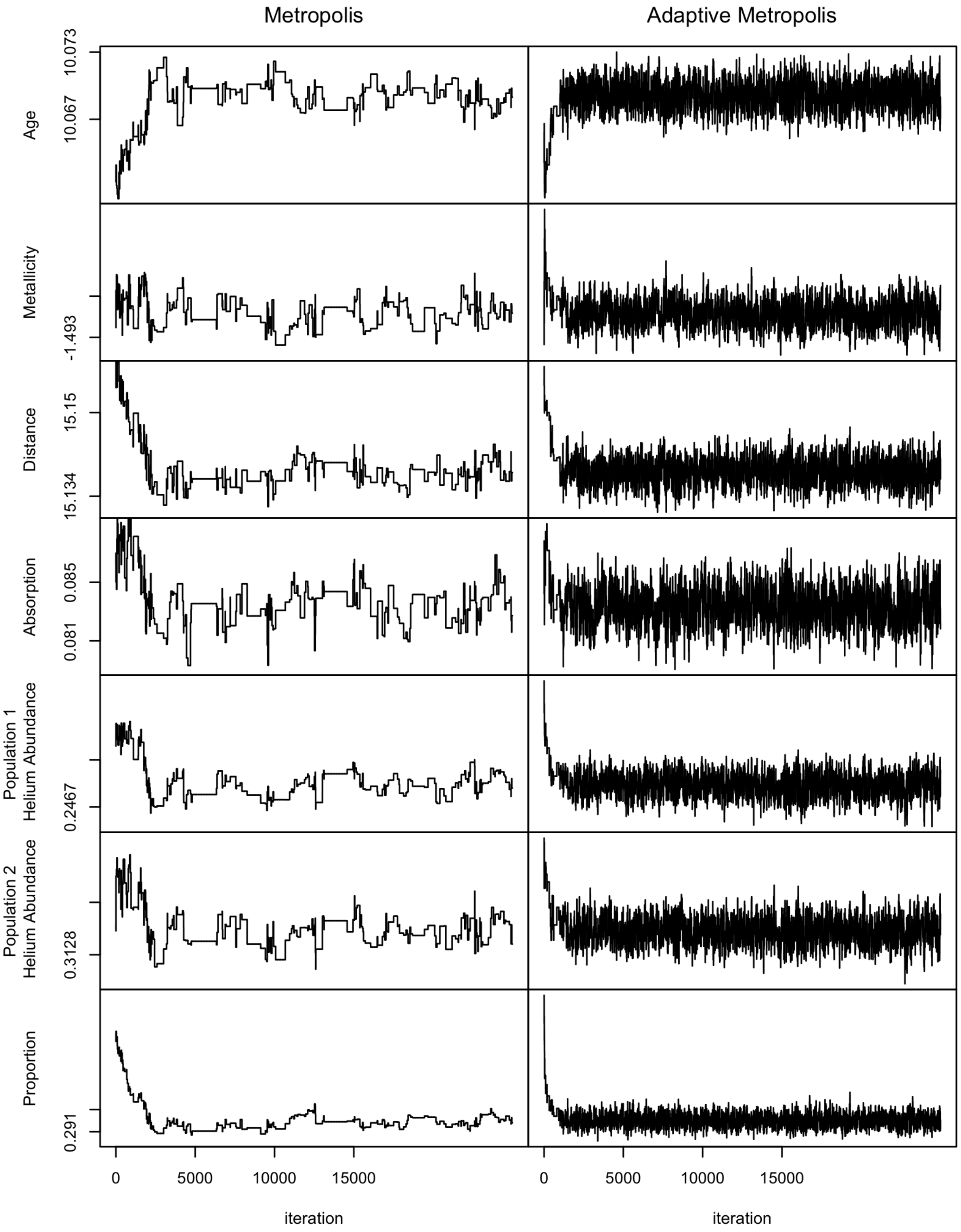


Figure 5.4: Improving convergence with an Adaptive Metropolis algorithm. The left column presents trace plots for a non-adaptive Metropolis algorithm, and the right column presents the trace plots for the Adaptive Metropolis algorithm we devised. Both algorithms used the same data and starting values, and followed the same tuning procedure.

$\phi_{Y1} = 0.24$ and $\phi_{Y2} = 0.29$, such that the true difference in helium abundance under the simulation is 0.05 for each cluster. We simulate 30,000 cluster stars and 1000 field stars per cluster, and every star is generated as a single-star system. Since the multiple-population clusters we analyze do not contain white dwarfs in the observed data, we do not include white dwarfs in our simulated clusters. For each cluster we generate photometric magnitudes in five wavelength bands, corresponding to the wavelength bands in the observed multiple-population data: $F275W$, $F336W$, $F438W$, $F606W$, and $F814W$. Details about these wavelength bands are provided in Section 5.5. The photometric magnitudes for each star are simulated with Gaussian measurement error that is representative of the measurement error we expect for observed data. The Gaussian measurement error is a function of both wavelength band and magnitude, as depicted in Figure 5.5.

Of the 31,000 stars per cluster that we generate, about 90% are dropped from the simulated cluster for two reasons. First, there is a threshold signal-to-noise ratio that identifies which stars are too faint to be detectable under realistic observational conditions. Because initial stellar masses are drawn from the prior distribution on initial mass, which is skewed towards low-mass/low-luminosity stars, most of the stars are not sufficiently bright to be detectable. Second, we believe the stellar evolution models to be inaccurate for higher-magnitude main-sequence stars (van Dyk *et al.*, 2009; DeGennaro *et al.*, 2009). Although there should not be any systematic effects from inaccuracies in the stellar evolution models when simulating clusters under those same models, for the data analyses in Section 5.5 we discard stars below a certain magnitude in the $F275W$ UV wavelength band. We therefore apply a similar data truncation to our simulated clusters so that our numerical results may be as illuminating as possible. The exact magnitude we use as a cutoff for a particular cluster depends on the assumed distance to the cluster, see Section 5.5. For our numerical results, we drop any stars with a photometric magnitude in the $F275W$ UV wavelength band less than 23. After losing stars due to low signal-to-noise and the $F275W$ magnitude cutoff, about 3000 simulated stars remain per cluster. This is on the lower-end for the typical number of stars

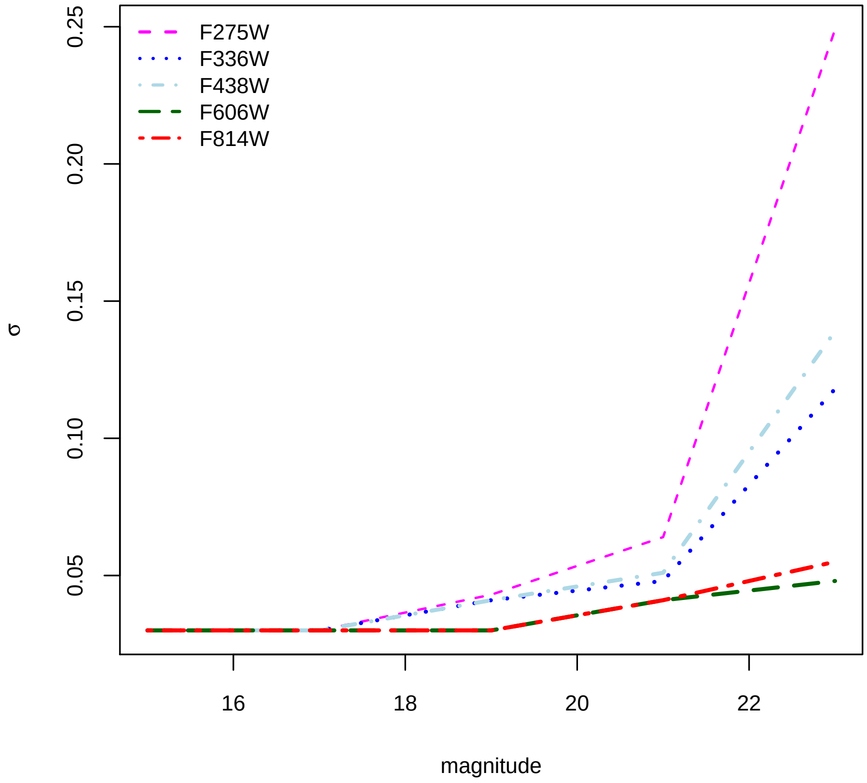


Figure 5.5: Gaussian measurement error for simulated two-population clusters. In our numerical studies, the photometric magnitudes for each star are simulated with Gaussian measurement error that is representative of the measurement error we expect for observed data, which is depicted above. Here, σ is the standard deviation of the Gaussian measurement error.

included in (cleaned) observed data for a multiple-population cluster. (The data-cleaning procedure is described in Section 5.5.)

The prior distributions for θ_{m-M_V}, θ_{A_V}, and $\theta_{\text{[Fe/H]}}$ are Gaussian distributions centered at the true values under the simulation, truncated to be positive for θ_{A_V}, with standard deviations of 0.02, 0.006, and 0.02, respectively. These standard deviations are typical values of the prior standard deviations for the multiple-population clusters we analyze in Section 5.5. The prior distributions on the population parameters are as described in Section 5.2.3. We assign

$P(Z_i = 1) = \alpha = 0.95$ for $i = 1, \ldots, N$. This is the value for α we use when analyzing the multiple-population clusters in Section 5.5 after testing the sensitivity to the choice of α for each cluster.

For each of the thirty simulated clusters, we use our AM algorithm to explore the marginal posterior distribution, $P(\boldsymbol{\Theta}, \boldsymbol{\Phi}|\boldsymbol{X})$. We run one chain per cluster for 26,000 iterations after the tuning period. Inspection of the trace plot for each chain shows that all the chains reach their apparent stationary distributions within the first 6000 iterations. We discard the first 6000 iterations of each chain as burn-in and base inference on the remaining 20,000 iterations. Results of the simulations for scenarios 1, 2, and 3 appear in Figures 5.6, 5.7, and 5.8, respectively.

In Figures 5.6 and 5.7, we observe that our method is performing reasonably well with respect to recovering the difference in helium abundance and the proportion of stars in each population. This is encouraging, as our main inferential goal is to recover the difference in helium abundance since it provides insight about the methods of formation of the populations within a cluster. While it is disappointing that there is bias between the fitted parameters and the true values of all the parameters under the simulation, the biases are relatively small and not unexpected based on our results in Section 4.3. Furthermore, as we discussed in Section 4.3, for a real data analysis such biases are small compared to systematic errors stemming from the computer models of stellar evolution. Nonetheless, we note that the bias in helium abundance affects the fitted values in the same way for both populations; see the third row of Figures 5.6 and 5.7. As a result, we can adequately recover the relative difference in helium abundance despite biases in the fitted parameters.

To check that the recovered helium abundance difference is due to the presence of multiple populations and not to our model specification, we intentionally fit simulated single-population clusters with our (now misspecified) two-population model. In Figure 5.8, we observe that we obtain similar results for the cluster parameters as we did when the repli-

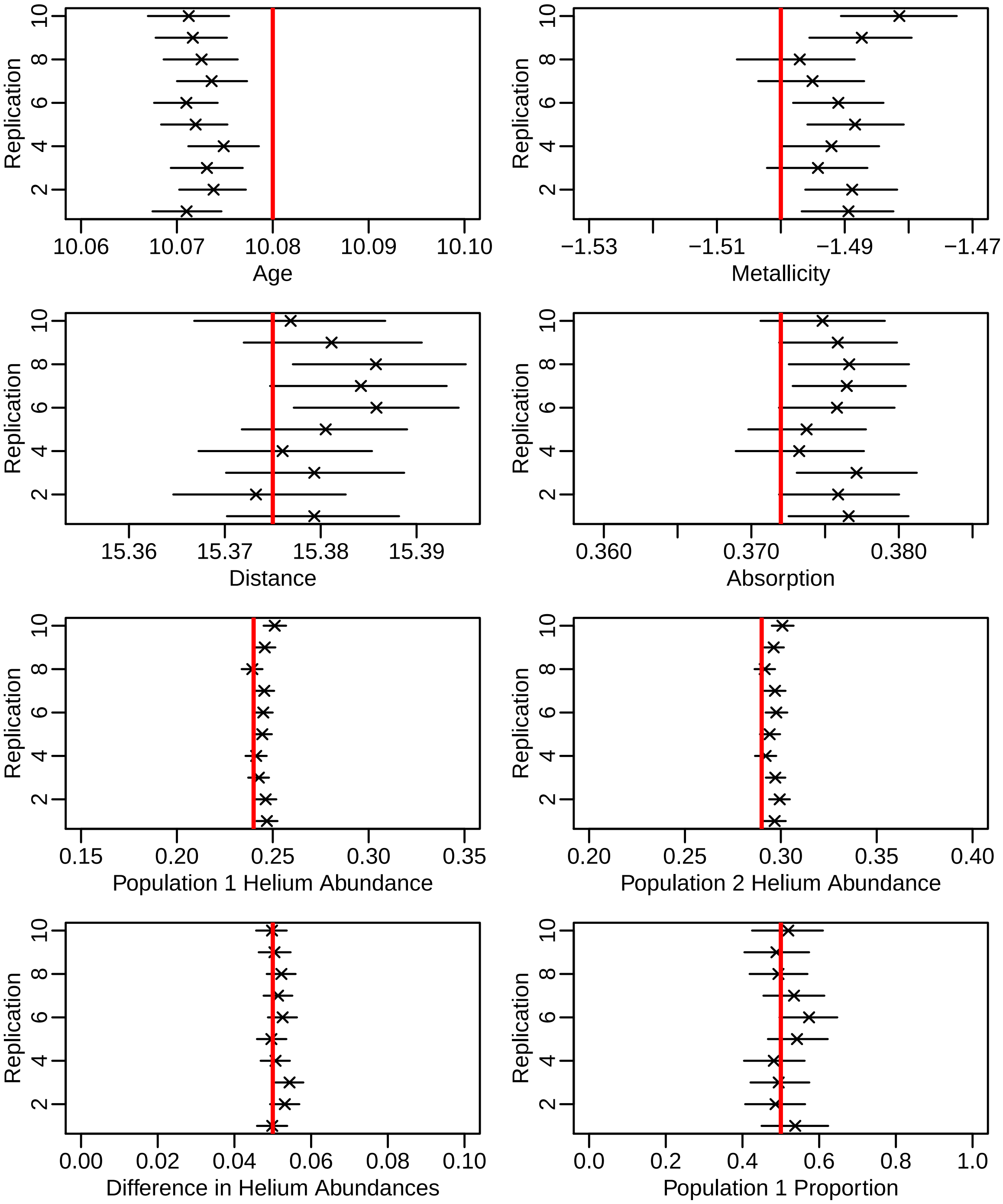


Figure 5.6: Scenario 1: 50% of stars in Population 1. The horizontal bars are 95% posterior intervals, with posterior means marked by an 'x.' The true parameter values under the simulation are indicated by the red vertical lines.

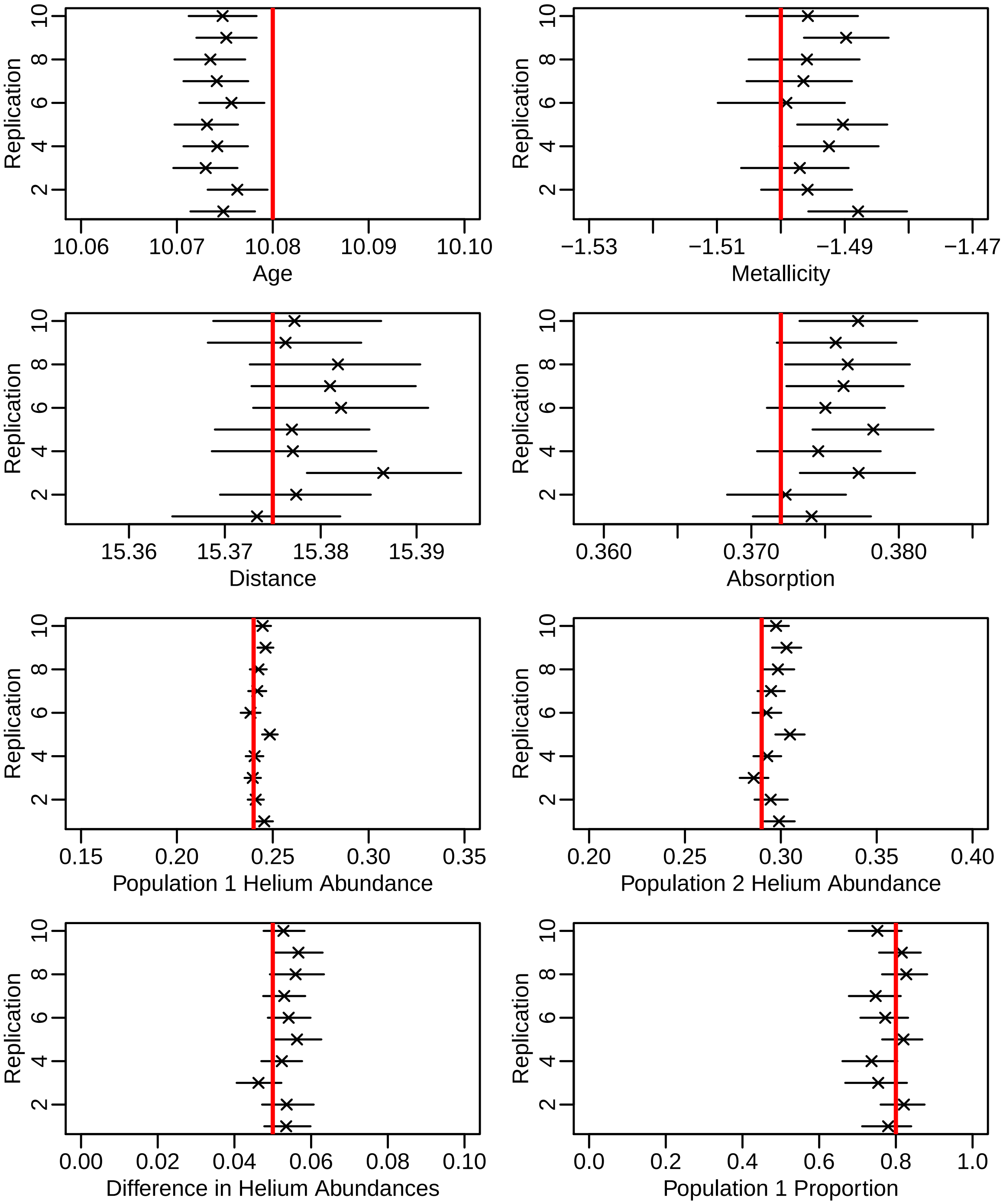


Figure 5.7: Scenario 2: 80% of stars in Population 1. The horizontal bars are 95% posterior intervals, with posterior means marked by an ‘x.’ The true parameter values under the simulation are indicated by the red vertical lines.

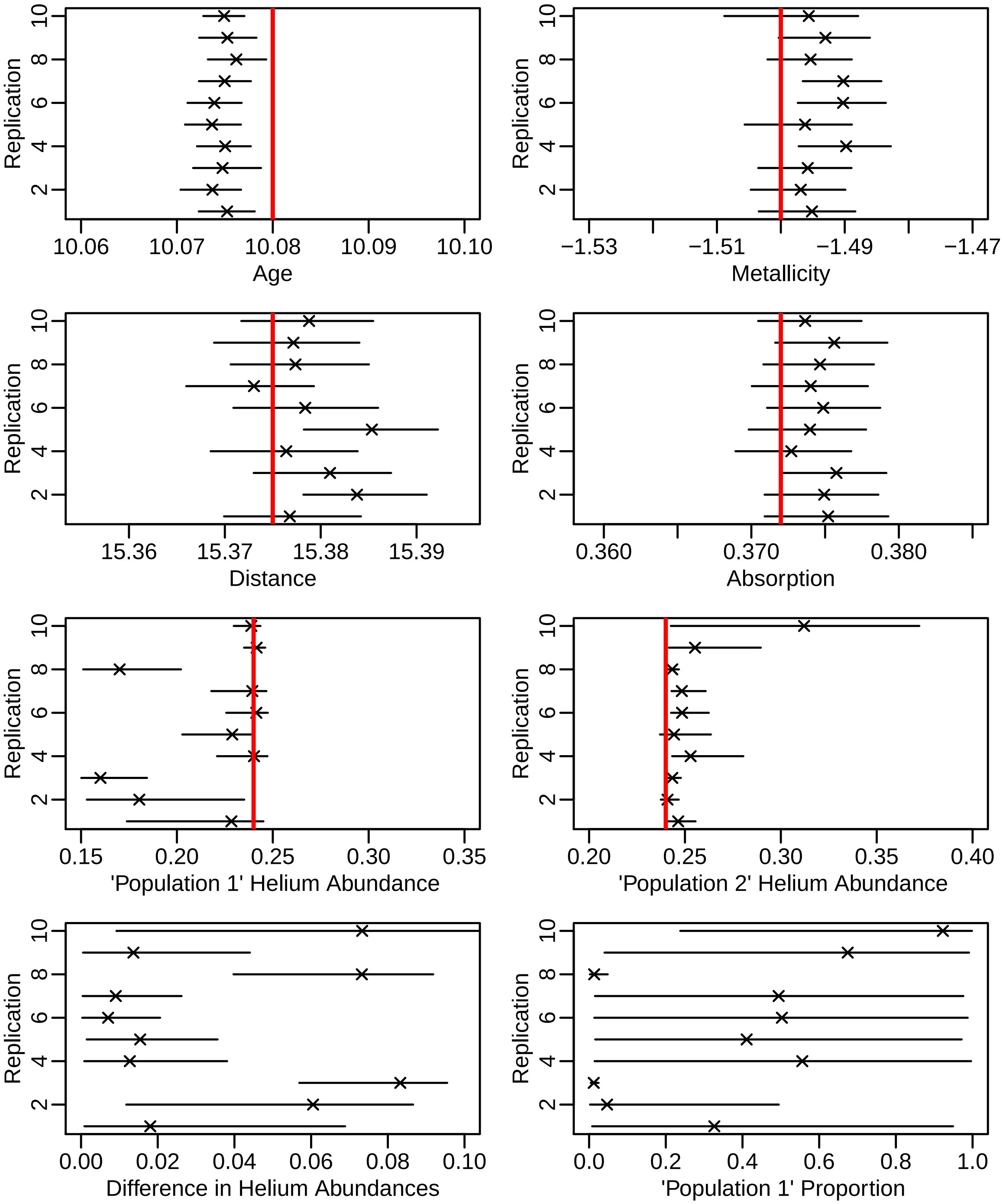


Figure 5.8: Scenario 3: 100% of stars in Population 1. The horizontal bars are 95% posterior intervals, with posterior means marked by an 'x.' The true parameter values under the simulation are indicated by the red vertical lines. For this scenario, there is only one stellar population, and so the "Population 1" and "Population 2" distinctions are meaningless.

cated datasets did indeed contain two stellar populations. This is expected, since the cluster parameters are common to both populations. However, the results for the proportion of stars in Population 1 indicate model misspecification. Specifically, we observe that the fitted value is close to either zero or one (replicates 2, 4, 8, and 10) and/or the 95% interval is very wide, spanning most of the range from $[0, 1]$ (replicates 1, 2, 4, 5, 6, 7, 9, and 10). Both of these outcomes suggests that a second population may not be present in the observed data. In practice, if we obtain a fitted value for the proportion of less than 0.1 or greater than 0.9, and/or a 95% interval with a width greater than a third of the range from $[0, 1]$, we would question our assumption that the cluster we are analyzing hosts two stellar populations.

5.5 Analyses of Two-Population Globular Clusters

In this section we apply our method to three globular clusters believed to host two stellar populations: NGC 288, NGC 5272, and NGC 6352. Our main objective is to estimate the difference in helium abundance between the two stellar populations, as well as the proportion of stars in each population. A secondary objective is to evaluate the underlying computer models of stellar evolution. This can be done in a preliminary way by examining how well the fitted models agree with the observed data. We are of course also interested in estimating the other cluster parameters, such as age. The observed data for each cluster consists of photometric magnitudes in five different wavelength bands per star. Two of the wavelength bands, $F606W$ and $F814W$, are in the visual part of the electromagnetic spectrum and were obtained with the Hubble Space Telescope's Advanced Camera for Surveys (ACS). The other three wavelength bands, $F275W$, $F336W$, and $F438W$, are in the UV. The UV photometric magnitudes were obtained with the Hubble Space Telescope's Wide Field Camera 3, and have been called a "magic trio" (e.g., Piotto *et al.*, 2015) for their ability to reveal multiple stellar populations. A detailed description of the data collection and pre-processing is given in Piotto *et al.* (2015).

We present CMDs for NGC 288, NGC 5272, and NGC 6352 in Figure 5.9; each star in the observed data is represented by a dot. All the clusters we analyze contain a large number of stars in the observed data: 30,170 for NGC 288, 179,330 for NGC 5262, and 29,025 for NGC 6352. Since fitting our multiple-population model with this amount of data is currently computationally infeasible, we prune the size of the observed data for each cluster. The stars that remain after pruning are indicated with black dots in the CMDs in Figure 5.9; stars that are cut from the observed data are indicated with grey dots. The general data-reduction routine proceeds as follows:

1. First, stars that are pretty clearly field stars are removed. The observed photometric magnitudes for each star in a cluster are compiled from a series of images. Stars exhibiting a large change in position from image to image are unlikely to belong to the globular cluster. The change in position from image to image is a proxy for *proper motion*—changes in the relative position of stars in the sky (van Dyk *et al.*, 2009). Stars belonging to a cluster will appear to move as a pack, so proper motion is one type of ancillary information we typically use to probabilistically separate cluster stars from field stars.

2. Next, by examining the CMD for each cluster, we can make general cuts to remove types of stars we cannot model. For example, the stars that appear to lie along a horizontal line at magnitude $F336W = 16$ in the CMDs for NGC 288 and NGC 5272 are *horizontal branch* stars. These stars are undergoing dynamic physical transformations as they transition from red giants to white dwarfs. We do not have computer models that can predict photometric magnitudes for stars in this transitionary phase of stellar evolution.

3. "Quality cuts" are also made to remove stars with very low signal-to-noise in one or more wavelength band, based on technical calibration details of the cameras onboard the Hubble Space Telescope.

4. Because we believe that the computer-based stellar evolution model is particularly inaccurate for fainter (i.e. higher-magnitude) stars, we impose a magnitude cutoff for each cluster. The cutoff depends on the distance to the cluster reported in the updated Harris (1996) globular cluster catalogue. We use a magnitude cutoff of $F275W = 21.815$ for NGC 288, $F275W = 22.074$ for NGC 5272, and $F275W = 22.930$ for NGC 6352.

5. Finally, we sample from the remaining stars such that the final photometry set contains approximately 3000 stars. This is done by first determining a point close to where stars "turn off" the main sequence. Stars at this so-called *main-sequence turn off* (MSTO) point are in the beginning stages of transforming into red giants. We determine this point by visual inspection to be $F336W = 19.0$ for NGC 288, $F336W = 18.8$ for NGC 5272, and $F336W = 20.0$ for NGC 6352. These points are indicated by the horizontal dotted lines in Figure 5.9. We are particularly conservative in that stars turning off the main sequence lie above the cutoff. We then sample approximately equal numbers of stars from above and below the horizontal dotted lines so that our final photometry sets will contain a good mix of main-sequence and post-main-sequence stars.

We describe an analysis of each of the three multiple-population globular clusters in turn. We begin with NGC 288, followed by NGC 5272, and conclude with NGC 6352. Unfortunately photometric errors (i.e. the Gaussian measurement errors on the photometric magnitudes) have not yet been made available for any of the multiple-population clusters we wish to study. As such, we assign tentative photometric errors. Specifically, we assume Gaussian measurement error with a standard deviation of 0.03 for the UV photometric magnitudes, and Gaussian measurement error with a standard deviation of 0.02 for the visual photometric magnitudes. We will repeat our analyses when the actual photometric errors are made available.

For all three clusters, the prior distributions for age and initial mass are described in Sec-

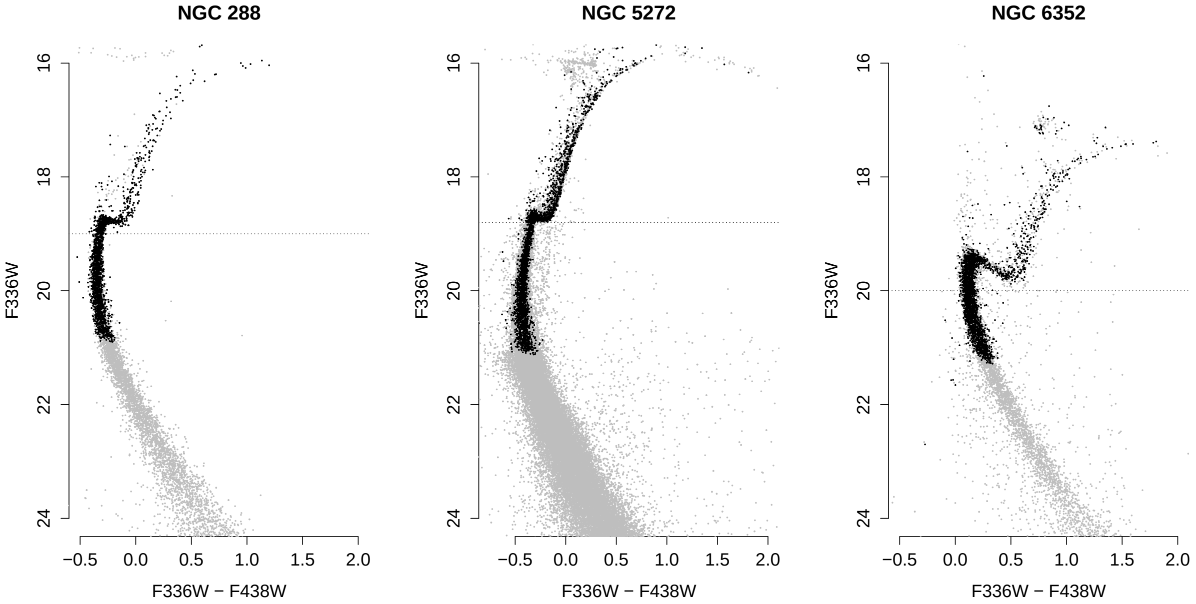


Figure 5.9: CMDs for NGC 288, NGC 5272, and NGC 6352. Grey dots represent the position on the CMD of stars from the full observed data. Black dots represent the sample of stars that remain after data pruning; we treat these stars as observed data in our analyses. The horizontal lines indicate the cutoff we use to separately sample main-sequence and post-main-sequence stars.

tion 4.2.3, and the prior distributions for the population parameters are given in Section 5.2.3. In the following subsections we report the prior distributions on metallicity, distance, and absorption.

5.5.1 NGC 288

NGC 288 can be found in the night sky in the constellation Sculptor. The prior distributions on metallicity, distance, and absorption are

$$\begin{aligned}
\theta_{[\mathrm{Fe/H}]} &\sim N(-1.32, 0.02^2),\\
\theta_{m-M_V} &\sim N(14.84, 0.02^2), \text{ and}\\
\theta_{A_V} &\sim TN(0.093, 0.003^2; 0).
\end{aligned}$$

The prior means for $\theta_{[\mathrm{Fe/H}]}$, θ_{m-M_V}, and θ_{A_V} for all the multiple-population clusters we study come from the updated Harris (1996) globular cluster catalogue, with standard deviations chosen to be relatively conservative. Before specifying $P(Z_i = 1) = \alpha$ for all $i = 1, \dots, N$, we investigate the sensitivity to the choice of α. Ancillary information such as proper motion is currently being processed and will soon be made available to researchers (Piotto *et al.*, 2015). When it is released, we can use this information to specify $P(Z_i = 1) = \alpha_i$ on a star-by-star basis. Until then, we specify $P(Z_i = 1) = \alpha$ for all $i = 1, \dots, N$ and investigate the sensitivity to that choice. Specifically, we test five values for α: 0.8, 0.9, 0.95, 0.99, and 0.999. Astrophysicists believe that a reasonable range for α is 0.9 to 0.99, but we test a wider range to be conservative. To fit each of the five resulting models, we run our AM algorithm for 51,000 iterations after the tuning period. Inspection of the trace plots shows that every chain converges to its apparent stationary distribution by iteration 11,000. We discard the first 11,000 iterations as burn-in, and base inference on the remaining 40,000 MCMC draws. (We follow this same procedure for NGC 5272 and NGC 6352 as well.) The results of the sensitivity analysis are presented in Figure 5.10; posterior means are indicated by an 'x,' and the horizontal bars are 95% posterior intervals. The choice of α has a noticeable effect on our inferences. However, the deviations in the fitted values are small and not scientifically meaningful. Thus, we use $\alpha = 0.95$ for the remainder of our analysis of NGC 288.

After specifying α, we explore $P(\boldsymbol{\Theta}, \boldsymbol{\Phi} | \boldsymbol{X})$ using five separate chains with different starting values. Each chain is run for 51,000 iterations after the tuning period. Inspection of the trace plots shows that every chain converges to the same apparent stationary distribution by iteration 11,000; see Figure 5.11. For each chain we discard the first 11,000 iterations as burn-in, and keep the remaining 40,000 iterations. (This same procedure is used for our analyses of NGC 5272 and NGC 6352). Fitted values and 95% intervals for $(\boldsymbol{\Theta}, \boldsymbol{\Phi})$, as well as the difference in helium abundance, $\phi_{Y2} - \phi_{Y1}$, are given in Table 5.2. The fitted values are posterior means based on the pooled MCMC draws, and 95% credible intervals are given by the 2.5% and 97.5% posterior quantiles of those draws.

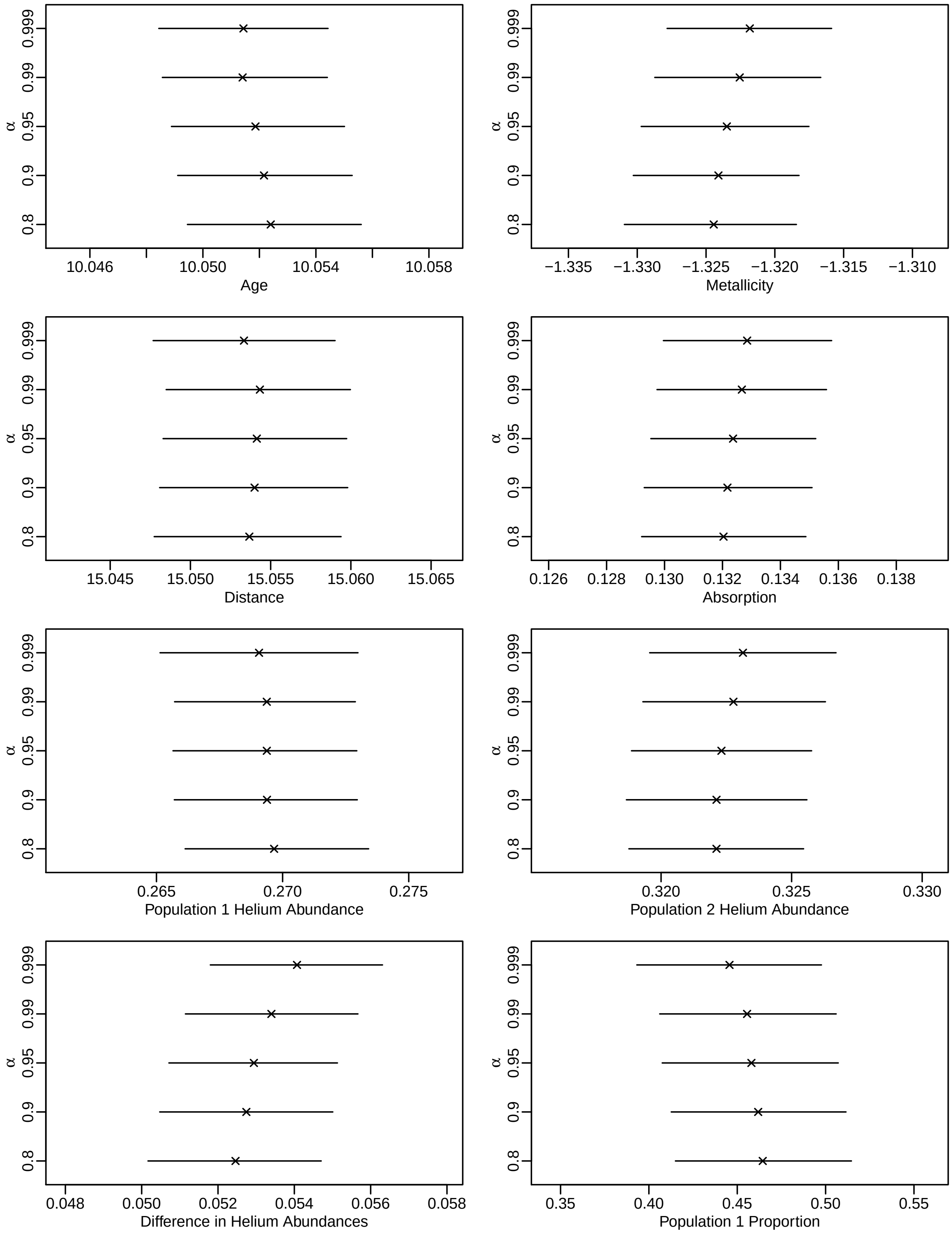


Figure 5.10: Sensitivity analysis of α for NGC 288. Posterior means are denoted by an 'x.' The horizontal bars are 95% posterior intervals.

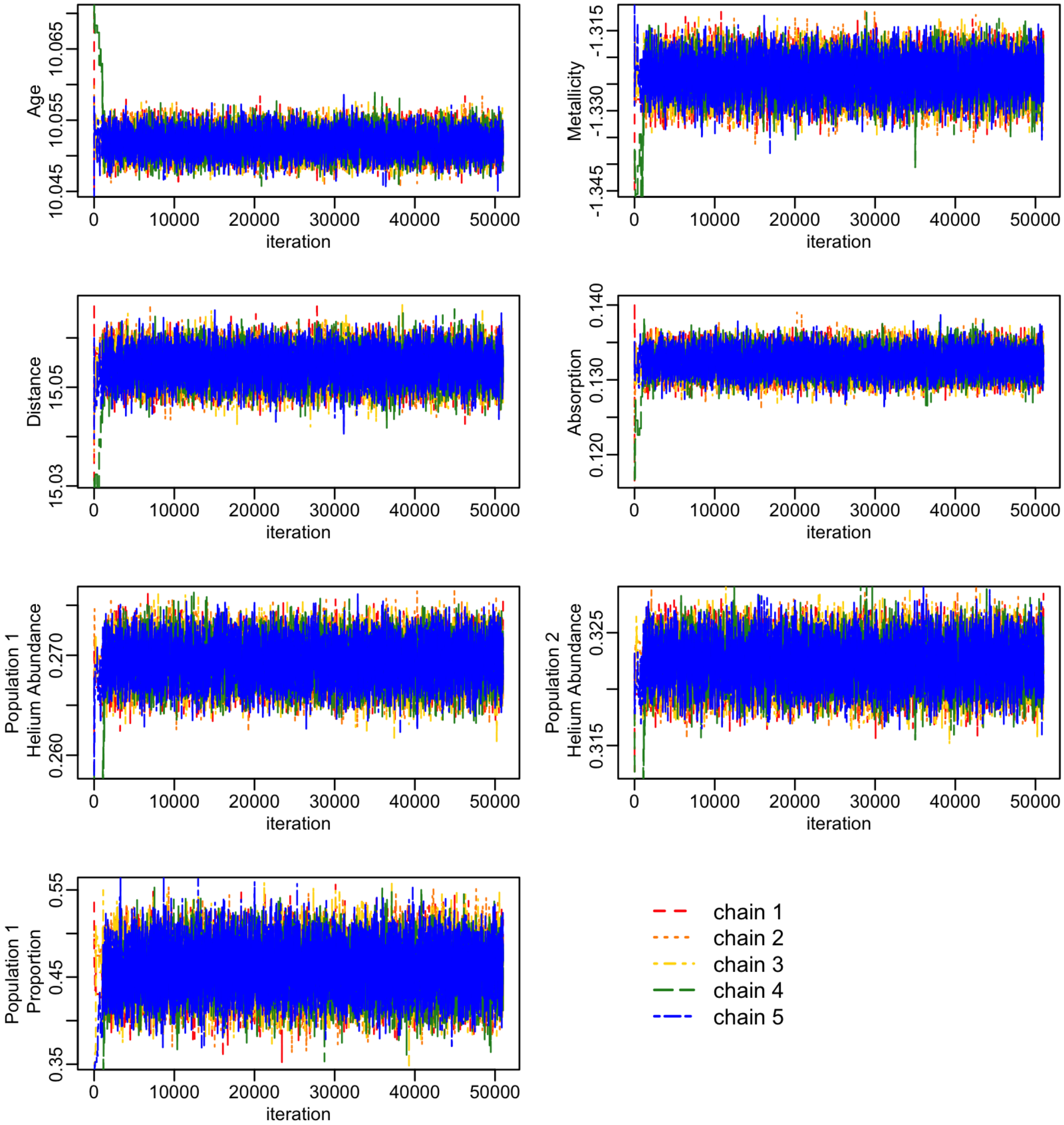


Figure 5.11: Five chains for NGC 288. It is apparent that the chains reach their apparent stationary distribution before iteration 11,000.

In Figure 5.12 we present a matrix with all possible CMDs for NGC 288, along with the fitted isochrones. The fitted isochrone for Population 1 is represented by cyan curves, and the fitted isochrone for Population 2 is represented by purple curves. The fitted isochrone for each population is the curve produced from the output of $\boldsymbol{G}$ over the entire range of initial mass and with the posterior means of both the cluster parameters and helium abundance

Table 5.2: Results for NGC 288

Quantity	Fitted Value	95% CI
θ_{age}	10.052	(10.049, 10.055)
$\theta_{\text{[Fe/H]}}$	-1.323	(-1.330, -1.318)
θ_{m-M_V}	15.054	(15.048, 15.060)
θ_{A_V}	0.132	(0.129, 0.135)
ϕ_{Y1}	0.269	(0.266, 0.273)
ϕ_{Y2}	0.322	(0.319, 0.326)
$\phi_{Y2} - \phi_{Y1}$	0.0529	(0.0506, 0.0552)
ϕ_{p1}	0.459	(0.410, 0.509)

as input. The specific computer model used to produce the fitted isochrones is the updated Dotter *et al.* (2008) model that was used for model fitting.

The fitted isochrones match the observed data well in some CMDs, and poorly in others. For example, the fitted isochrones appear to predict the two populations we observe in the CMD $F275W - F438W$ vs. $F275W$, which is marked with an orange circle. However, the fitted isochrones do not agree with the observed data in the CMD $F438W - F606W$ vs. $F438W$, which is marked with a green square. In general, we find that CMDs using only the UV wavelength bands—$F275W$, $F336W$, and $F438W$—tend to display better model fits. Examining these discrepancies is a useful first step towards designing computer models that can better predict the observed data.

5.5.2 NGC 5272

NGC 5272 can be found in the night sky in the constellation of Canes Venatici. Having been discovered in the mid-eighteenth century, it is one of the most well-studied globular clusters

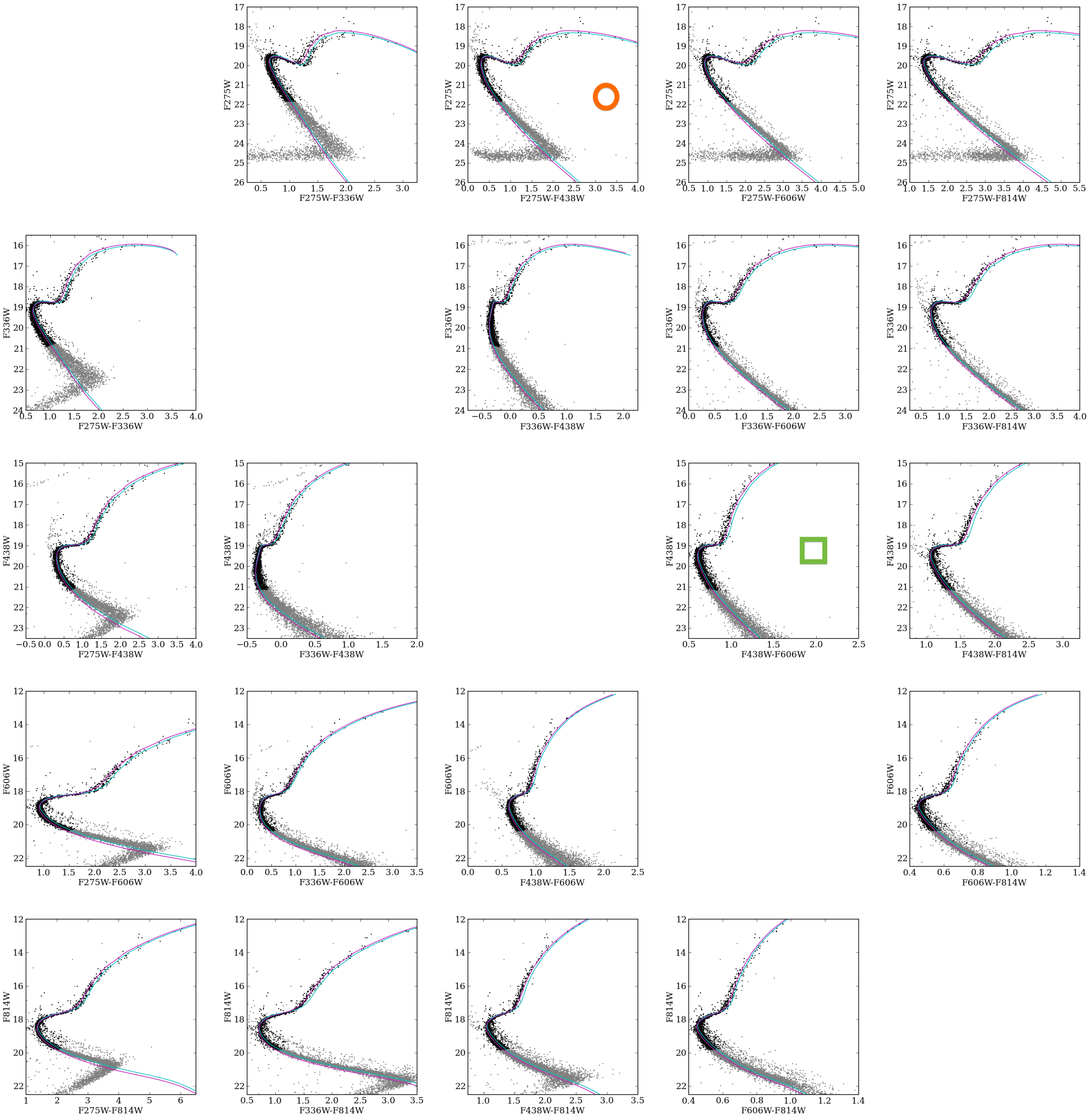


Figure 5.12: Fitted model with all possible CMDs for NGC 288. The black dots indicate stars which were used to fit the model, while grey dots indicate stars that did not pass the data pruning but were in the full observed photometry set. The fitted isochrones produced with the updated Dotter *et al.* (2008) model are represented by the cyan curve for Population 1, and by the purple curve for Population 2.

Table 5.3: Results for NGC 5272

Quantity	Fitted Value	95% CI
θ_{age}	10.069	(10.067, 10.072)
$\theta_{\text{[Fe/H]}}$	-1.489	(-1.493, -1.484)
θ_{m-M_V}	15.139	(15.135, 15.144)
θ_{A_V}	0.083	(0.081, 0.086)
ϕ_{Y1}	0.250	(0.247, 0.254)
ϕ_{Y2}	0.315	(0.312, 0.317)
$\phi_{Y2} - \phi_{Y1}$	0.0647	(0.0624, 0.0670)
ϕ_{p1}	0.320	(0.290, 0.351)

(Machholz, 2002). The prior distributions on metallicity, distance, and absorption are

$$\begin{aligned} \theta_{\text{[Fe/H]}} &\sim N(-1.5, 0.2^2), \\ \theta_{m-M_V} &\sim N(15.07, 0.02^2), \text{ and} \\ \theta_{A_V} &\sim TN(0.031, 0.003^2; 0). \end{aligned}$$

As with NGC 288, we test the sensitivity to the choice of α; the results are presented in Figure 5.13. We find that our inferences are a bit more sensitive to the choice of α for NGC 5272 than they were for NGC 288, particularly for the difference in helium abundance and the proportion of stars in Population 1. However, the deviations in the posterior means are still relatively small and not scientifically meaningful. This is especially true if we only consider the range of α that astrophysicists consider to be realistic for the clusters we study, which is $\alpha = 0.9$ to $\alpha = 0.99$. As a result, we are not particularly concerned about the influence of α. As with NGC 288, we use $\alpha = 0.95$ for the remainder of our analysis of NGC 5272.

Fitted values and 95% intervals based on MCMC draws from the five separate chains we use to explore $P(\boldsymbol{\Theta}, \boldsymbol{\Phi} | \boldsymbol{X})$ are given in Table 5.3. The matrix of all possible CMDs for NGC 5272, along with fitted isochrones, is presented in Figure 5.14. We find similar patterns in

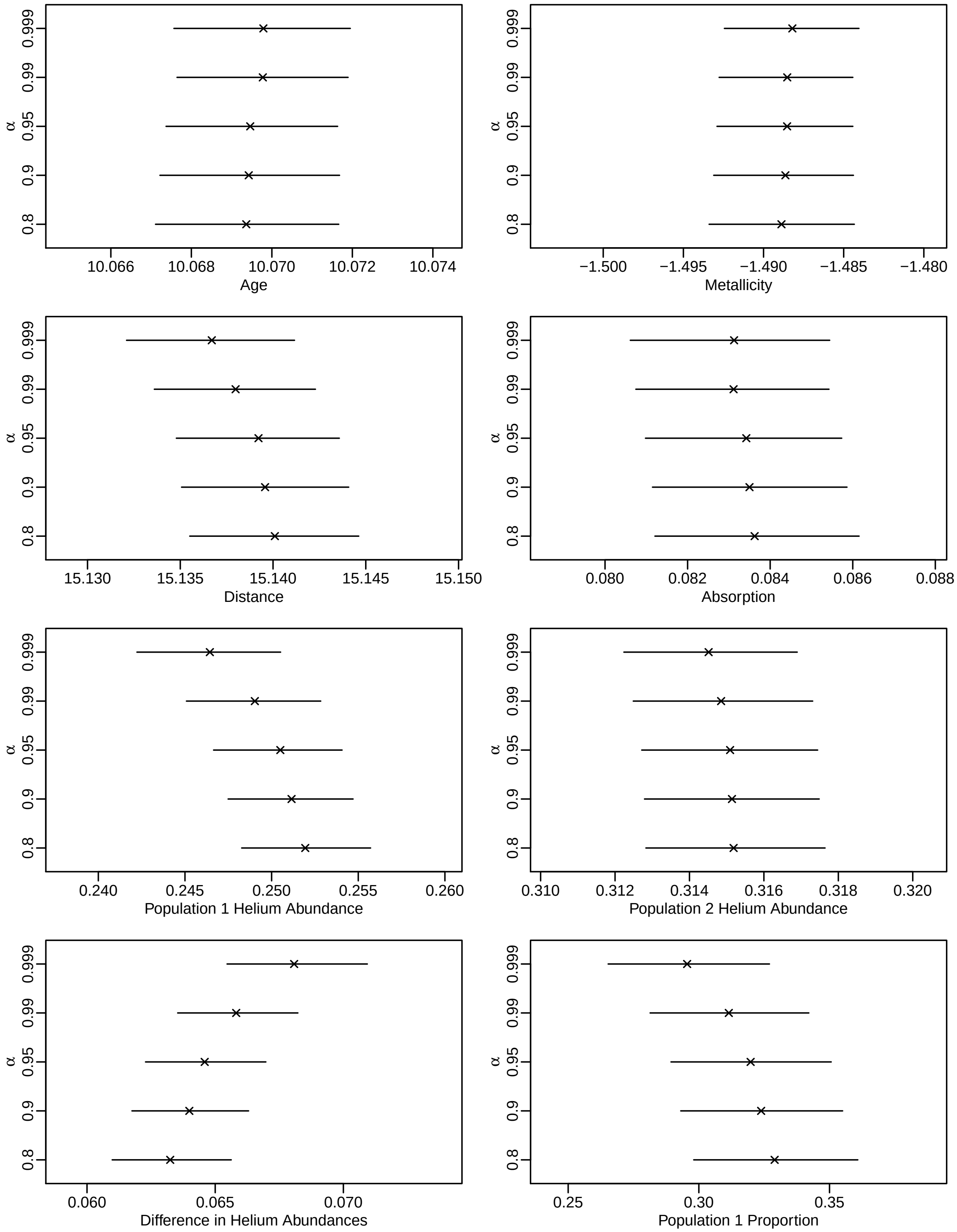


Figure 5.13: Sensitivity analysis of α for NGC 5272. Posterior means are denoted by an ‘x.’ The horizontal bars are 95% posterior intervals.

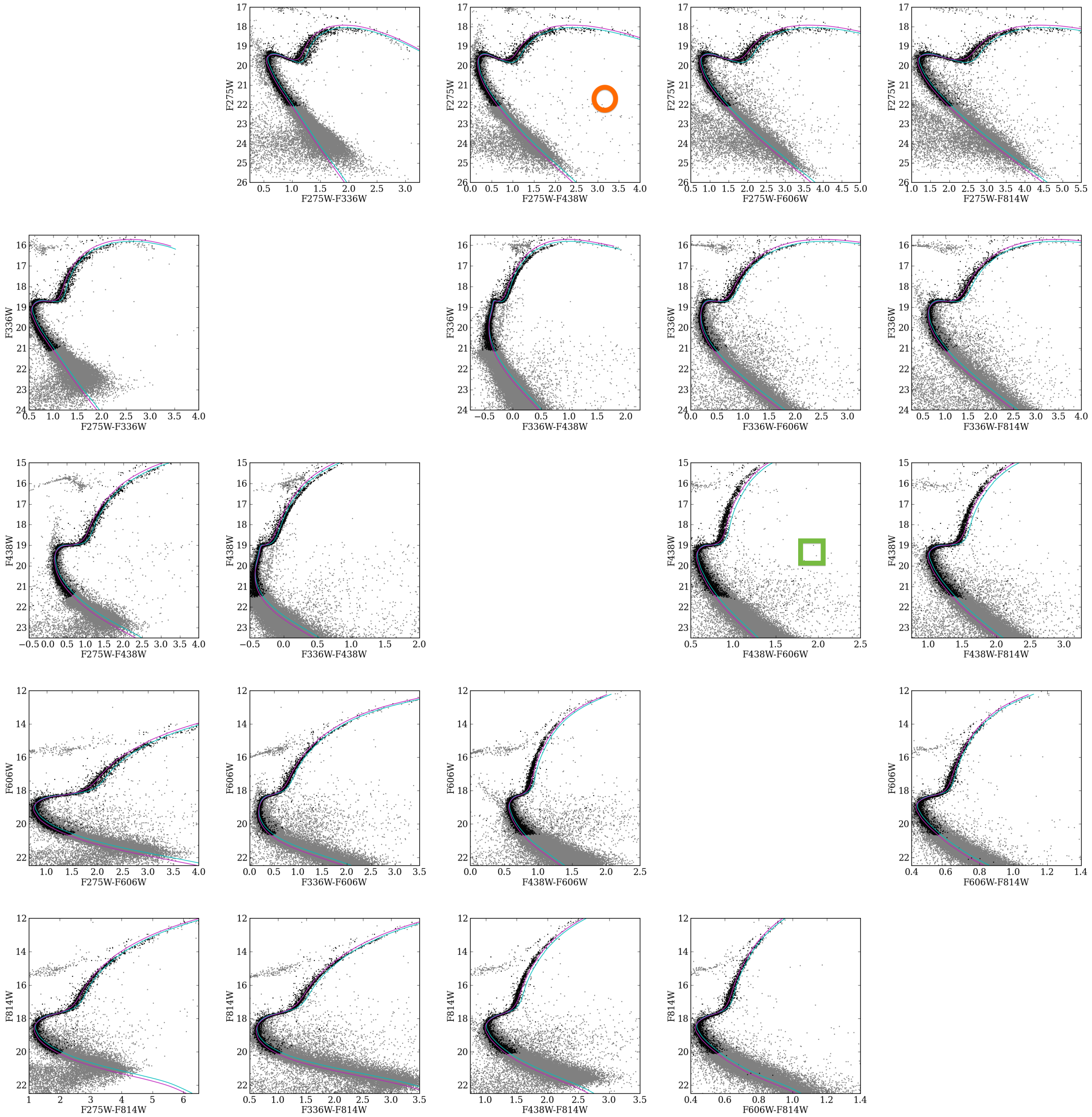


Figure 5.14: Fitted model with all possible CMDs for NGC 5272. The black dots indicate stars which were used to fit the model, while grey dots indicate stars that did not pass the data pruning but were in the full observed photometry set. The fitted isochrones produced with the updated Dotter *et al.* (2008) model are represented by the cyan curve for Population 1, and by the purple curve for Population 2.

the model fits that we noticed during our analysis of NGC 288. Specifically, the CMDs that only involve the UV wavelength bands tend to display better model fits than CMDs that involve F606W or F814W (or both). If we examine the same two CMDs that we highlighted for NGC 288, which like their counterparts are identified with an orange circle and a green square, we notice similar behavior in the fitted isochrones. Specifically, the fitted isochrones in the CMD with the orange circle appear to capture the two distinct stellar populations. On the other hand, the fitted isochrones in the CMD with the green square appear to completely miss the red giants altogether, and therefore fail to identify the two stellar populations.

5.5.3 NGC 6352

NGC 6352 can be found in the night sky in the constellation Ara. The prior distributions on metallicity, distance, and absorption are

$$\begin{aligned} \theta_{[\mathrm{Fe/H}]} &\sim N(-0.64, 0.02^2), \\ \theta_{m-M_V} &\sim N(14.43, 0.02^2), \text{ and} \\ \theta_{A_V} &\sim TN(0.68, 0.006^2; 0). \end{aligned}$$

As with our previous analyses, we test the sensitivity to the choice of α. The results of the sensitivity analysis are presented in Figure 5.15. Here, we find similar results to the sensitivity analysis for NGC 5272. Specifically, the difference in helium abundance grows, and the proportion of stars in population 1 shrinks, as α increases. However, as with NGC 5272, the deviations in the fitted values due to the choice of α are not scientifically meaningful. We use $\alpha = 0.95$ for the remainder of our analysis of NGC 6352.

Fitted values and 95% intervals based on MCMC draws from the five separate chains we use to explore $P(\boldsymbol{\Theta}, \boldsymbol{\Phi}|\boldsymbol{X})$ are given in Table 5.4. The matrix of all possible CMDs for

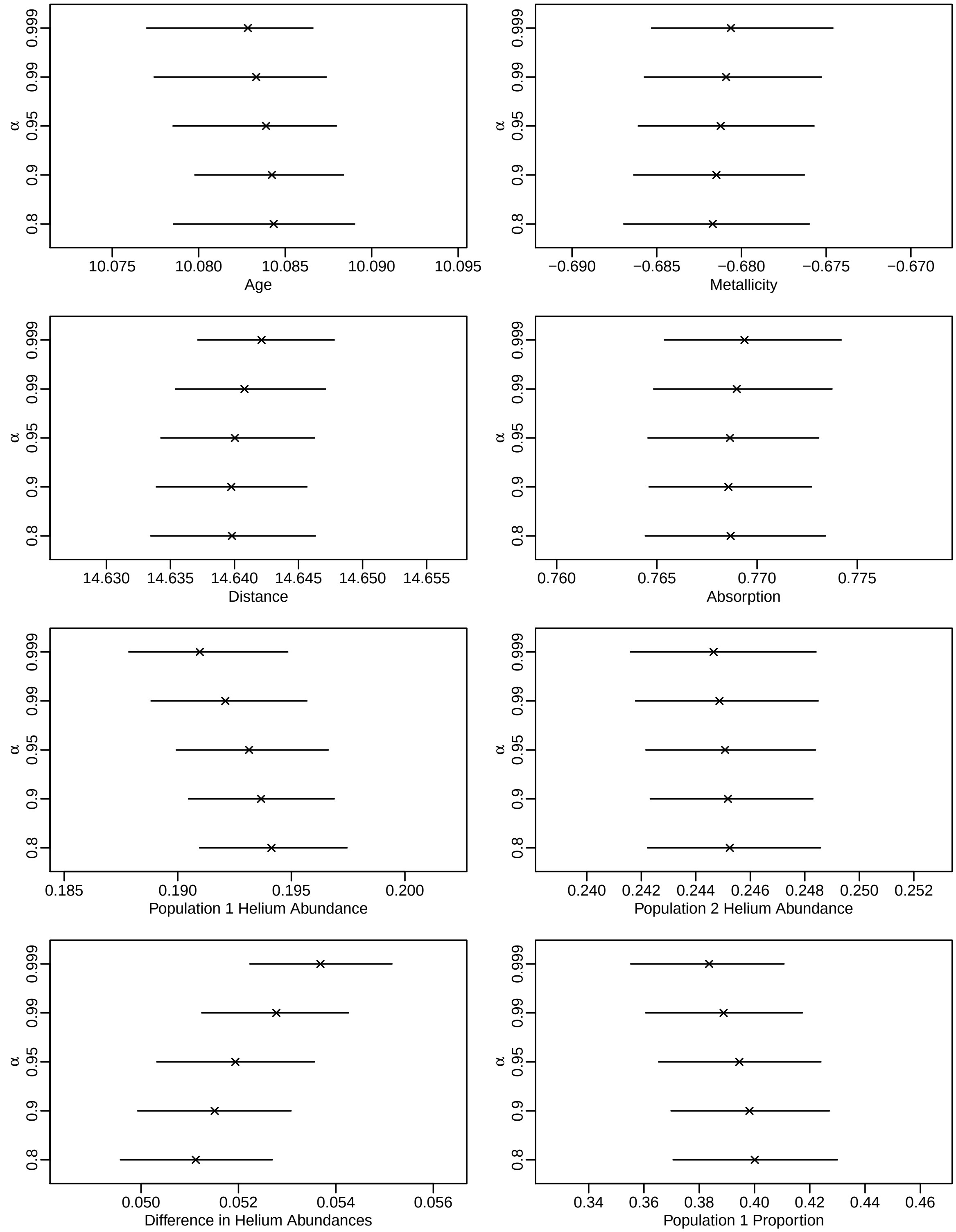


Figure 5.15: Sensitivity analysis of α for NGC 6352. Posterior means are denoted by an 'x.' The horizontal bars are 95% posterior intervals.

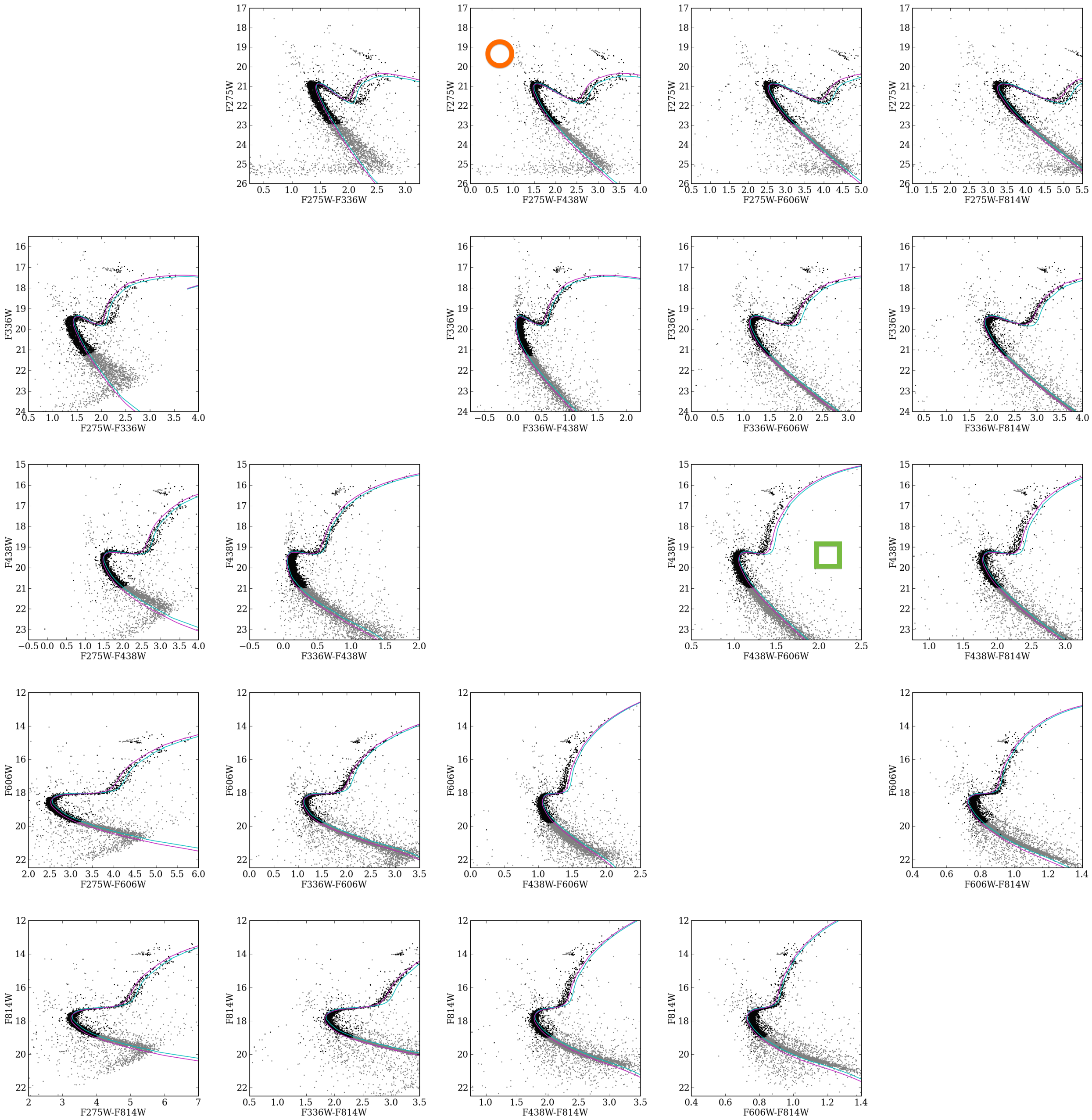


Figure 5.16: Fitted model with all possible CMDs for NGC 6352. The black dots indicate stars which were used to fit the model, while grey dots indicate stars that did not pass the data pruning but were in the full observed photometry set. The fitted isochrones produced with the updated Dotter *et al.* (2008) model are represented by the cyan curve for Population 1, and by the purple curve for Population 2.

Table 5.4: Results for NGC 6352

Quantity	Fitted Value	95% CI
θ_{age}	10.084	(10.078, 10.088)
$\theta_{[\text{Fe/H}]}$	-0.681	(-0.686, -0.676)
θ_{m-M_V}	14.640	(14.634, 14.646)
θ_{A_V}	0.769	(0.765, 0.773)
ϕ_{Y1}	0.193	(0.190, 0.197)
ϕ_{Y2}	0.245	(0.242, 0.248)
$\phi_{Y2} - \phi_{Y1}$	0.0519	(0.0503, 0.0535)
ϕ_{p1}	0.395	(0.366, 0.424)

NGC 6352, along with fitted isochrones, is presented in Figure 5.16. Of the three clusters we analyze, the model fits are perhaps the poorest for NGC 6352. This is particularly apparent if we compare the first CMD in both the first and second rows to their counterparts in NGC 288 and NGC 5272. Nonetheless, we observe the same general patterns in the model fits that we noticed with the previous two clusters. Examination of the CMDs marked with the orange circle and the green square in particular reveals similar fits to those for NGC 288 and NGC 5272. Additional discussion is reserved for Section 5.6.

5.6 Summary and Discussion

In this chapter we presented a Bayesian hierarchical model for multiple-population globular clusters. This is a substantial improvement over the common approach of plotting computer models on top of observed data and tuning the parameters until the two appear to agree. The latter “chi-by-eye” approach necessarily relies on two-dimensional CMDs for model fitting. By formulating a Bayesian hierarchical model, we do not need to rely on any or all two-dimensional projections of five-dimensional data during model fitting. This is important since the stellar populations overlap in complex and non-obvious ways in CMDs, even for clusters that only host two stellar populations. We demonstrated with numerical studies that our method can adequately recover the population parameters of two-population clusters. Specifically, we are successful at recovering the difference in helium abundance between the two stellar populations and the proportion of stars in each population. We also demonstrated how to diagnose model misspecification in the event that our two-population model is applied to a single-population cluster. In particular, we showed that (i) the fitted value for the proportion of stars is population 1 is close to zero or one, and/or (ii) the associated posterior interval extends over most of the range from zero to one.

After demonstrating the capabilities of our Bayesian hierarchical model for multiple-population globular clusters, we fit the model using observed photometric magnitudes from three clusters: NGC 288, NGC 5272, and NGC 6352. After verifying that the value we specify for α is not very influential, we explore the marginal posterior distribution of the cluster and population parameters using an AM algorithm that we devised for this purpose; the AM algorithm greatly improved convergence over earlier, non-adaptive approaches. To diagnose convergence we ran five separate chains per cluster, all with different starting values. For each cluster, we found that the five separate chains quickly converged to the same apparent stationary distribution.

Our analysis of each of the three two-population clusters in turn reveals several interesting features and patterns. First, if we examine the fitted values for the helium abundances of the three clusters, we notice that there is a range of values for both $\hat{\phi}_{Y1}$ and $\hat{\phi}_{Y2}$. For $\hat{\phi}_{Y1}$ the values are 0.269, 0.250, and 0.193 for NGC 288, NGC 5272, and NGC 6352, respectively; similarly, the values for $\hat{\phi}_{Y2}$ are 0.322, 0.315, and 0.245. Notice in particular that the larger helium abundance for NGC 6352 is less than the smaller helium abundances for NGC 288 and NGC 5272. However, the values for $\widehat{\phi_{Y2} - \phi_{Y1}}$ are 0.0529, 0.0647, and 0.0519 for NGC 288, NGC 5272 and NGC 6352, respectively. Thus, despite the range of fitted helium abundance values, and in particular the deviation in the helium abundance values for NGC 6352 compared to the other clusters, the fitted values for the difference in helium abundance between two populations in a particular cluster fall within a narrow range. This is not a result we expected a priori, and fitting our two-population model with observed data from additional clusters believed to host two stellar populations will either confirm of refute this finding. If we continue to find that the difference in helium abundance falls within a narrow range, this finding will shed light on the possible mechanisms that produce multiple-population globular clusters. Alternatively, we may find that the difference in helium abundance is correlated with more global properties of the cluster (e.g., total mass, metallicity, etc.). Either result will inform astrophysicists about the methods of formation of the populations within globular clusters.

In addition to estimating the difference in helium abundance in two-population globular clusters, a secondary objective was to examine the model fits to investigate properties of the underlying computer model of stellar evolution. In general, we find that the fitted models agree with the observed data in CMDs using only UV photometric magnitudes. This is not true for CMDs that use either or both of the $F606W$ and $F814W$ wavelength bands. While we do not automatically conclude that the mismatch between fitted models and observed data is due to systematic errors in the computer model, the pattern in model fits from the three clusters we analyzed suggests that further examination of this effect is warranted. If we notice the same pattern when many two-population clusters have been studied, and after the actual photometric errors are made available, this will provide evidence that the morphologies inherent to the computer model differ in some systematic way from the morphologies of observed data; we already know this to be true for fainter main-sequence stars (van Dyk *et al.*, 2009; DeGennaro *et al.*, 2009). Nevertheless, these results can provide key feedback on how the models are deficient—in what wavelength bands at at what evolutionary stages.

Now that we have demonstrated the capabilities of our Bayesian hierarchical model for two-population globular clusters, near-future work will focus on fitting the model with observed data from many clusters so that astrophysicists can extract scientific implications. The time is particularly ripe for this analysis, as observed data from multiple-population globular clusters with the necessary UV photometric magnitudes are just now being processed (see Piotto *et al.*, 2015). When it is available, we will incorporate ancillary information to specify $P(Z_i = 1) = \alpha_i$ on a star-by-star basis. Alternatively, a better strategy may be to specify α_i as a parameterized function of the ancillary information (e.g., proper motion data) and fit the parameters of that function. In addition, we intend to fit the population distribution of the initial masses so that there is ability to share strength for these parameters. In the long-term, we will extend the Bayesian hierarchical model to include more than two stellar populations. We will also incorporate additional population-level parameters (e.g., the carbon-nitrogen-oxygen abundance). While these extensions to the Bayesian hierarchical model are mathematically trivial, the additional complexity of the model will lead to interesting statistical challenges, particularly in computation and inference.

Chapter 6

The Carbon Fraction of White Dwarfs

6.1 Introduction

Carbon and oxygen, essential ingredients for life on Earth, are key elements in the later stages of a star's lifecycle. After a main-sequence star exhausts its hydrogen fuel, it (assuming an appropriate initial mass) expands into a red giant and fuses helium into carbon and oxygen. A typical red giant does not have sufficient mass to fuse carbon into even heavier elements, so the carbon and oxygen produced from helium fusion will build up in the core. After the outer layers of the red giant blow away, the core forms a white dwarf. As a result, most white dwarfs are composed almost entirely of carbon and oxygen (Fontaine *et al.*, 2001).

The carbon fraction, which is the fraction of a white dwarf composed of carbon atoms, is scientifically interesting because it provides insight into fundamental nuclear physics. Specifically, while the formation rate of carbon in nuclear fusion reactions is well understood (via the triple-α nuclear fusion process), the formation rate of oxygen has large experimental uncertainty (Rehm, 2012). It is the relative difference in these rates that determines the carbon fraction in white dwarfs. Therefore, the carbon fraction contains important information about experimentally uncertain nuclear reaction rates that produce carbon and oxygen as a byproduct.

A timely reason why gaining insight into the relative formation rates of carbon and oxygen is important is the connection to planetary chemistry. To date, over 1,800 *exoplanets*—planets outside our solar system that orbit stars other than the Sun—have been confirmed, and there are over 4,000 candidates awaiting verification (see the NASA Exoplanet Archive: `http://exoplanetarchive.ipac.caltech.edu/`). Better knowledge of the relative difference in the formation rates of carbon and oxygen, which we gain by empirically determining the carbon fraction in white dwarfs, informs astrophysicists about the composition of these exoplanets. Although most red giant stars end their lives as white dwarfs, the most massive ones explode in energetic supernovas. These explosions blow the star apart, and the carbon and oxygen that has been synthesized is ejected into space. These elements then enrich the molecular clouds that eventually form new stars and planets. In this way, the carbon fraction in white dwarfs provides indirect information about the composition of exoplanets. More specifically, if the abundance of gaseous carbon is more than fifty percent the abundance of gaseous oxygen, then all the oxygen is locked up in carbon dioxide. Otherwise, there is excess oxygen, which forms silicon dioxide. Carbon dioxide is gaseous unless at very low temperature, whereas silicon dioxide is solid unless at very high temperatures. As a result, high levels of oxygen result in planetary systems similar to our own solar system, with inner planets largely composed of silicon dioxide crusts and mantles. Conversely, planets such as the Earth cannot exist with high levels of carbon, and "diamond planets" may form instead (Kuchner and Seager, 2005).

In addition to providing feedback about fundamental physics, which has many interesting scientific implications beyond those already discussed, the carbon fraction of white dwarfs is important for the study of stellar evolution. As described in Section 4.1.4, white dwarf cooling models are among the components of our computer-based stellar evolution model.

The carbon fraction of white dwarfs affects the white dwarf cooling rate, and therefore the implied ages of star clusters determined via their white dwarfs. In fact, the carbon fraction is the largest source of uncertainty in white dwarf ages derived by *cosmochronometry*—a method that uses white dwarfs to directly determine the age of various stellar populations (Metcalfe, 2003). So determining the carbon fraction in white dwarfs will provide better estimates of star cluster ages, and also fill in gaps in our understanding of late-stage stellar evolution.

This chapter explores the use of our Bayesian hierarchical model for single-population globular clusters to estimate the carbon fraction of white dwarfs. Using observed photometry from globular clusters, as opposed to open clusters, is necessary since the former have many more white dwarfs. In particular, the observed data consists of photometric magnitudes from 47 Tuc in two visual wavelength bands: $F606W$ and $F814W$. The observed data is described in detail in Section 6.5. The updated Montgomery *et al.* (1999) white dwarf cooling model that we incorporated into BASE-9 was created for the specific purpose of allowing us to include the carbon fraction as a parameter in our Bayesian hierarchical model. As far as we are aware, this chapter contains the first-ever attempt to estimate the carbon fraction of white dwarfs.

The rest of this chapter is divided into five sections. In Section 6.2 we present a statistical model that incorporates separate computer models for main-sequence/red giant stars and white dwarfs. In Section 6.3 we review the computational methods we use to fit the model. In Section 6.4 we present a numerical study to demonstrate how the carbon fraction affects the implied white-dwarf ages of star clusters. Analysis of the globular cluster 47 Tuc is described in Section 6.5, followed by a discussion in Section 6.6.

6.2 Statistical Model

6.2.1 The Likelihood Function

Unlike the other cluster parameters, the carbon fraction is only a parameter for white dwarf stars. Because of this, it is useful to consider the separate computer models that provide predicted magnitudes for main-sequence/red giant stars (MS/RG stars) and white dwarfs (WDs). We can express the computer model for MS/RG stars as $\boldsymbol{G}_{\text{MS/RG}}(M_i, \boldsymbol{\Theta}, \phi_Y)$, where $\boldsymbol{\Theta}$ are the cluster parameters for MS/RG stars, $\boldsymbol{\Theta} = (\theta_{\text{age}}, \theta_{\text{[Fe/H]}}, \theta_{m-M_V}, \theta_{A_V})$. Since all MS/RG stars are modeled as binary stars,

$$\boldsymbol{\mu}_i = -2.5 \log_{10} \left(10^{-\boldsymbol{G}_{\text{MS/RG}}(M_i, \boldsymbol{\Theta}, \phi_Y)/2.5} + 10^{-\boldsymbol{G}_{\text{MS/RG}}(M_i R_i, \boldsymbol{\Theta}, \phi_Y)/2.5}\right) \tag{6.1}$$

is the vector of predicted magnitudes for star system $i \in \mathcal{I}_{\text{MS/RG}}$, where $\mathcal{I}_{\text{MS/RG}}$ is the set of indices for the MS/RG system. Our observed data indicates which stars are WDs. Stars can easily be pre-specified as MS/RG stars or WDs using their location on a CMD.

For star $i \in \mathcal{I}_{\text{WD}}$, where $\mathcal{I}_{\text{WD}}$ is the set of indices for WDs, the vector of predicted magnitudes is given by $\boldsymbol{G}_{\text{WD}}(M_i, \boldsymbol{\Theta}, \theta_C, \phi_Y)$, where $\boldsymbol{\Theta} = (\theta_{\text{age}}, \theta_{\text{[Fe/H]}}, \theta_{m-M_V}, \theta_{A_V})$ as above, and θ_C is the carbon fraction. Since we do not consider WD binaries, $\boldsymbol{\mu}_i = \boldsymbol{G}_{\text{WD}}(M_i, \boldsymbol{\Theta}, \theta_C, \phi_Y)$ for

$i \in \mathcal{I}_{\text{WD}}$; see Section 4.2.1. Thus, the likelihood function can be written as

$$
\begin{aligned}
&L(\boldsymbol{M}, \boldsymbol{R}, \boldsymbol{Z}, \boldsymbol{\Theta}, \theta_C, \phi_Y | \boldsymbol{X}, \boldsymbol{\Sigma}) \\
= & \prod_{i \in \mathcal{I}_{\text{MS/RG}}} \left[\frac{Z_i}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{1}{2} \Big(\boldsymbol{X}_i - \boldsymbol{\mu}_i \Big)^{\top} \boldsymbol{\Sigma}_i^{-1} \Big(\boldsymbol{X}_i - \boldsymbol{\mu}_i \Big) \right) + (1 - Z_i) P(\boldsymbol{X}_i | Z_i = 0) \right] \\
\times & \prod_{i \in \mathcal{I}_{\text{WD}}} \left[\frac{Z_i}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_i|}} \exp\Bigg(-\frac{1}{2} \Big(\boldsymbol{X}_i - \boldsymbol{G}_{\text{WD}}(M_i, \boldsymbol{\Theta}, \theta_C, \phi_Y) \Big)^{\top} \right. \\
& \left. \times \ \boldsymbol{\Sigma}_i^{-1} \Big(\boldsymbol{X}_i - \boldsymbol{G}_{\text{WD}}(M_i, \boldsymbol{\Theta}, \theta_C, \phi_Y) \Big) \Bigg) + \ (1 - Z_i) P(\boldsymbol{X}_i | Z_i = 0) \right], \qquad (6.2)
\end{aligned}
$$

where $\boldsymbol{Z} = (Z_1, \ldots, Z_N)$, $\boldsymbol{M} = (M_1, \ldots, M_N)$, $\boldsymbol{R} = \{R_i, i \in \mathcal{I}_{\text{MS/RG}}\}$, $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)$, $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_N)$, and $\boldsymbol{\mu}_i$ is given by (6.1) for $i \in \mathcal{I}_{\text{MS/RG}}$.

6.2.2 Prior Distributions

Here, the primary parameter of scientific interest, in addition to cluster age, is θ_C. Since we do not have any prior knowledge about θ_C, we specify a uniform prior distribution on the interval $[0, 1]$. For $\boldsymbol{\Theta}$, ϕ_Y, $\boldsymbol{M}$, and $\boldsymbol{R}$, we use the same prior distributions as described in Section 4.2.3. We specify $P(Z_i = 1) = \alpha = 0.9$ for $i = 1, \ldots, N$. Although there exists proper motion data that can help specify $P(Z_i = 1) = \alpha_i$ on a star-by-star basis (Kalirai *et al.*, 2013), it is not publicly available at this time.

6.3 Statistical Computation

As in Section 5.3, we marginalize over $\boldsymbol{M}$, $\boldsymbol{R}$, and $\boldsymbol{Z}$ to improve computation. The full joint posterior distribution is denoted by $P(\boldsymbol{\Theta}, \theta_C, \phi_Y, \boldsymbol{M}, \boldsymbol{R}, \boldsymbol{Z} \mid \boldsymbol{X})$, and the marginal posterior distribution of $(\boldsymbol{\Theta}, \theta_C, \phi_Y)$ is given by

$$P(\boldsymbol{\Theta}, \theta_C, \phi_Y \mid \boldsymbol{X}) = \int \cdots \int \left(\sum_{Z_1} \cdots \sum_{Z_N} P(\boldsymbol{\Theta}, \theta_C, \phi_Y, \boldsymbol{M}, \boldsymbol{R}, \boldsymbol{Z} \mid \boldsymbol{X}) \right) d\boldsymbol{M} d\boldsymbol{R} \tag{6.3}$$

$$\propto P(\boldsymbol{\Theta}, \theta_C, \phi_Y) \prod_{i=1}^{N} \left[c(\boldsymbol{\Theta}, \theta_C, \phi_Y) P(Z_i = 1) + P(\boldsymbol{X}_i | Z_i = 0) P(Z_i = 0) \right], \tag{6.4}$$

where

$$c(\boldsymbol{\Theta}, \theta_C, \phi_Y) = \int \int P(\boldsymbol{X}_i \mid M_i, R_i, \boldsymbol{\Theta}, \theta_C, \phi_Y, Z_i = 1) P(M_i, R_i \mid Z_i = 1) dM_i dR_i. \tag{6.5}$$

The resulting marginal posterior distribution $P(\boldsymbol{\Theta}, \theta_C, \phi_Y | \boldsymbol{X})$ is six-dimensional, and we explore this distribution using the AM algorithm described in Section 5.3.2.2.

6.4 Numerical Results

In this section we use a numerical simulation to demonstrate the potential folly of fixing the carbon fraction to a commonly used value when estimating the age of a star cluster based on its WDs. We simulate and fit a total of fifteen clusters under our model. Every cluster is simulated with $\theta_{\text{age}} = 10.079$, $\theta_{[\text{Fe/H}]} = -0.72$, $\theta_{m-M_V} = 13.37$, $\theta_{A_V} = 0.10$, and $\phi_Y = 0.27$. These values were chosen to be similar to the adopted values for 47 Tuc (e.g., Ferraro *et al.*, 2004; Grundahl *et al.*, 2002; di Criscienzo *et al.*, 2010). For θ_C we simulate five replicate clusters with $\theta_C = 0.2$, five replicate clusters with $\theta_C = 0.4$ and five replicate clusters with $\theta_C = 0.6$. For each replicate we generate photometric magnitudes in the $F606W$ and $F814W$ wavelength bands. The photometric magnitudes are simulated with $N(0, 0.01^2)$ measurement errors in each wavelength band. We generate 3000 stars per cluster, about 600 of which are WDs. We do not directly set the number of WDs; initial masses are drawn from the prior distribution on initial mass, and stars with an initial mass above a certain threshold are evolved into WDs. The exact number of stars per cluster varies slightly since typically 15-40 of the generated stars have a high enough initial mass that they evolve beyond the white-dwarf stage and are therefore discarded. We do not include field stars, and binary star systems are not allowed.

Since for this numerical study we are interested in estimating the age of a cluster based on its WDs, we discard main-sequence stars with magnitudes in the $F814W$ wavelength band outside the narrow range of 18.5 to 20. We do this so that the cluster age will be constrained by the WDs alone. We include some main-sequence stars only to help constrain the other cluster parameters. We also discard WDs with magnitude in the $F814W$ wavelength band greater than 30, as such WDs are too faint to be observed. For model fitting, we fix the value of θ_C to a commonly used value, specifically $\theta_C = 0.4$. Thus, the model is misspecified when fit using the replicates with $\theta_C = 0.2$ or $\theta_C = 0.6$. This is intentional, as our main goal with this study is to examine how this misspecification influences the fitted values of cluster ages. We also fix the values for θ_{m-M_V}, θ_{A_V}, and ϕ_Y to their true values under the simulation since they are nuisance parameters for our current investigation, and doing so simplifies computation and inference. For $\theta_{[\mathrm{Fe/H}]}$, we use a Gaussian prior distribution with a mean equal to the true value under the simulation and a standard deviation of 0.02. The prior distribution on θ_{age} is described in Section 4.2.3. Since field stars are not included, we set $P(Z_i = 1) = 1.0$ for each star in every cluster.

To explore the marginal posterior distribution $P(\theta_{\mathrm{age}}, \theta_{[\mathrm{Fe/H}]}|\boldsymbol{X})$, we use the AM algorithm described in Section 5.3.2.2. We run one chain per replicate for 26,000 iterations after the

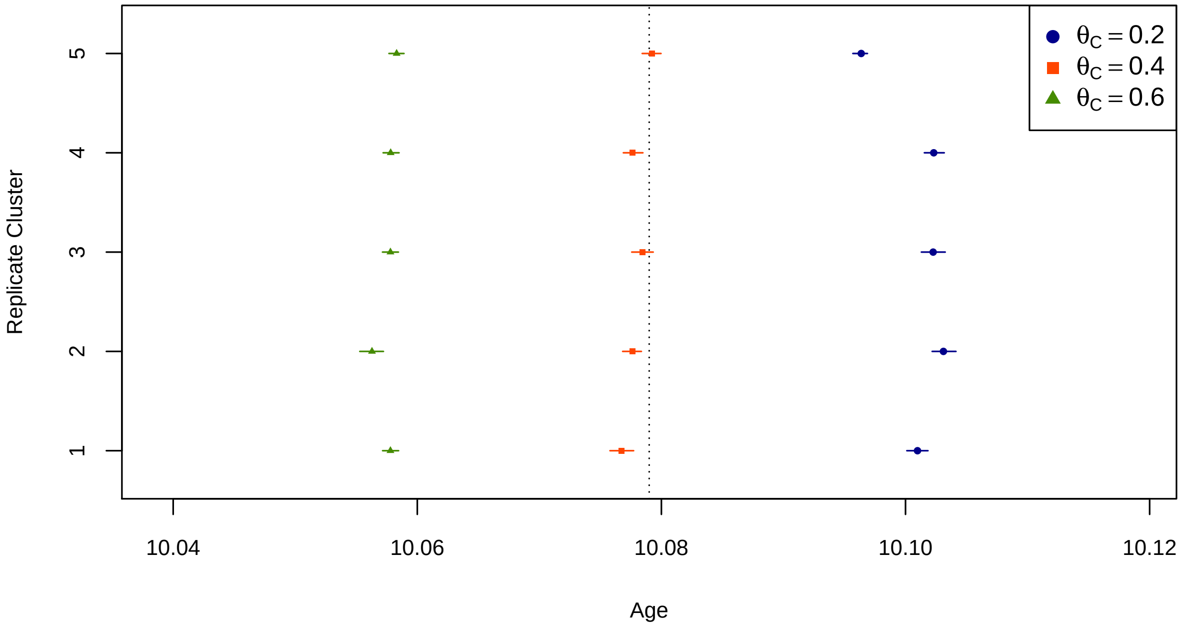


Figure 6.1: Numerical study results. The true value of age under the simulation is indicated by the black dotted vertical line. Blue circles, orange squares, and green triangles are posterior means corresponding to clusters simulated with $\theta_C = 0.2$, $\theta_C = 0.4$, and $\theta_C = 0.6$, respectively. Horizontal bars are 95% posterior intervals with the same color coding.

tuning period. Inspection of the trace plots shows that all the chains converged to their apparent stationary distribution by iteration 6000. We discard the first 6000 iterations of each chain as burn-in, and our inferences for each replicate are based on the MCMC draws of the remaining 20,000 iterations. The results of this numerical study are presented in Figure 6.1; blue circles, orange squares, and green triangles are posterior means corresponding to clusters simulated with $\theta_C = 0.2$, $\theta_C = 0.4$, and $\theta_C = 0.6$, respectively. Horizontal bars are 95% posterior intervals with the same color coding as the posterior means. We find that when the true value of θ_C under the simulation is 0.2, the fitted values for θ_{age} are all above the true value under the simulation. Likewise, when the true value of θ_C under the simulation is 0.6, the fitted values for θ_{age} under the misspecified model fall below the true value under the simulation. When the model is correctly specified, the fitted values for θ_{age}

are close to the true value under the simulation. While there is bias even when the model is correctly specified, which we attribute to the influence of the prior distribution on initial mass as discussed in Section 4.3, the bias is small compared to the systematic errors we observe under the misspecified models. In general, this numerical study demonstrates that determining the age of a star cluster based on its WDs requires an accurate estimate of the carbon fraction.

6.5 Analysis of 47 Tucanae

47 Tuc is the second brightest globular cluster in the night sky after ω Cen, and it contains millions of stars. As one of the most well-studied globular clusters, 47 Tuc is one of only a few globular clusters for which enough WDs have been observed to provide a meaningful estimate of their carbon fraction. The photometric magnitudes for 47 Tuc were obtained from 120 orbits of the Hubble Space Telescope using only two filters (Kalirai *et al.*, 2013). At roughly 40 minutes per orbit, this equates to exposure times of approximately 40 hours per filter (i.e. wavelength band). For comparison, the photometric magnitudes for the multiple-population globular clusters that we analyzed in Chapter 5 were obtained with exposure times of about 0.02 to 0.1 hours per filter. The extra exposure time per filter for 47 Tuc is necessary to reveal the WDs that are missing in the data for NGC 288, NGC 5272, and NGC 6352. The drawback is that the observed data contains only two photometric magnitudes per star.

A full CMD for 47 Tuc is presented in the left panel of Figure 6.2. It is immediately apparent that there are three distinct strands in the CMD. The longest strand, in the upper right of the CMD, is the main-sequence and red-giant branch for 47 Tuc. The shortest strand, in the lower left of the CMD, is the white dwarf cooling sequence for 47 Tuc. The medium-length strand, in between the other two, is the “main sequence” for a dwarf galaxy known as the

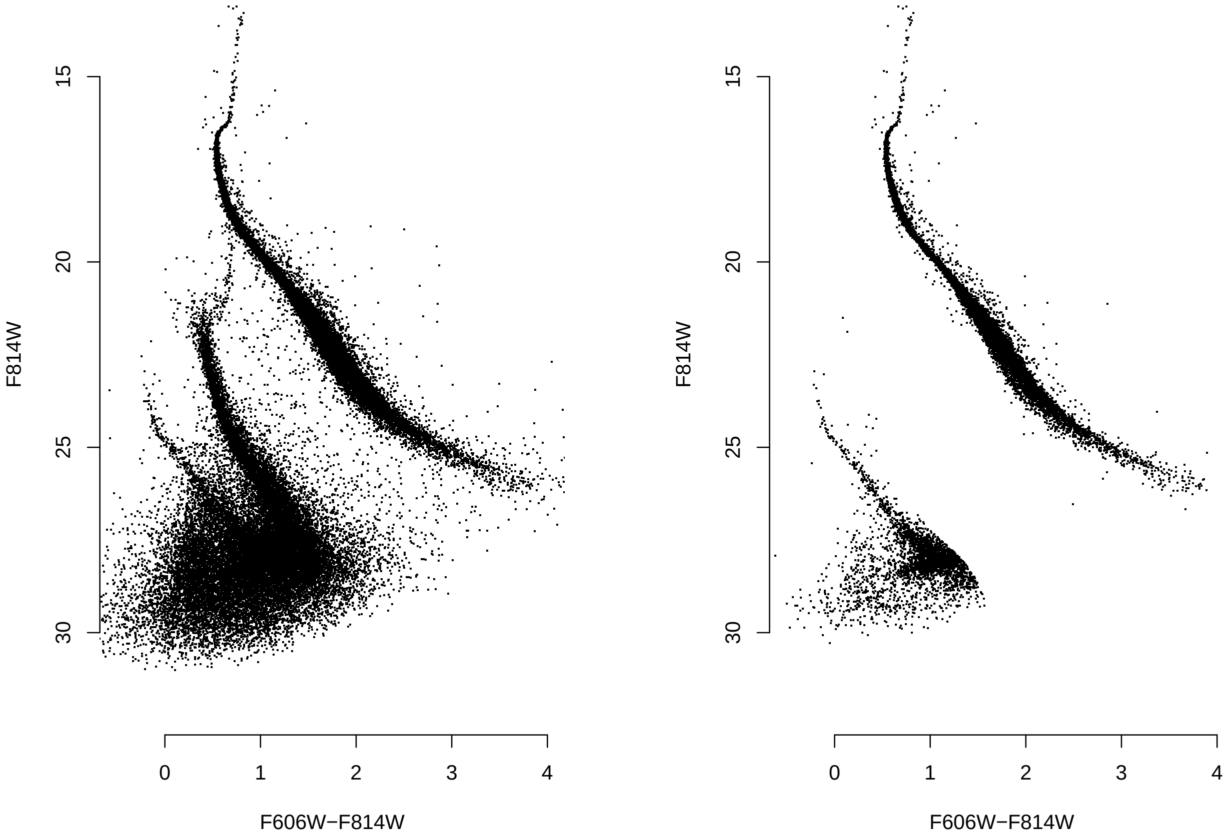


Figure 6.2: The CMD for 47 Tucanae. *Left*: The CMD constructed using the full observed dataset. The observed data are contaminated by stars from the Small Magellanic Cloud, and we observe what appears to be a main sequence for the SMC between the main sequence and white-dwarf group of 47 Tuc. *Right*: The CMD for 47 Tuc after removing stars that we are certain belong to the SMC. Although contamination remains, the white-dwarf group for 47 Tuc becomes more apparent.

Small Magellanic Cloud (SMC). The reason for this contamination is that the line of sight to 47 Tuc intersects the outer boundary of the SMC (Kalirai *et al.*, 2013). Luckily, because 47 Tuc is so well studied there is ancillary information that can be used to separate stars belonging to 47 Tuc from those belonging to the SMC (Kalirai *et al.*, 2013). We can use this information to probabilistically separate cluster stars from field stars. However, due to the large number of stars in the observed data, for computational reasons we instead choose to simply discard the stars that we are certain belong to the SMC. Stars that are ambiguous are not discarded. This data cleaning reduces the size of the observed data from 51,480 stars to a more manageable 14,439 stars. (Although 47 Tuc contains millions of stars, the Hubble Space Telescope only focuses on a portion of the cluster. This results in only 51,480 stars in the observed data instead of the millions that 47 Tuc is known to contain.) The resulting CMD after removing some of the stars belonging to the SMC is presented in the right panel of Figure 6.2. The white-dwarf sequence for 47 Tuc is more easily distinguishable in the cleaned CMD. We treat this cleaned data set as the observed data for the rest of this chapter.

When fitting our Bayesian hierarchical model, we fix the helium abundance to 0.27. This is a commonly adopted value for 47 Tuc (e.g., di Criscienzo *et al.*, 2010). This simplifies inference, as there is a relationship within the computer-based stellar evolution model that ties metallicity to helium abundance (Dotter *et al.*, 2008). The prior distributions on metallicity, distance, and absorption are

$$\begin{aligned}\theta_{[\mathrm{Fe/H}]} &\sim N(-0.72, 0.3^2),\\ \theta_{m-M_V} &\sim N(13.37, 0.2^2), \text{ and}\\ \theta_{A_V} &\sim TN(0.10, 0.3^2; 0),\end{aligned}$$

where the posterior means come from Grundahl *et al.* (2002) and Ferraro *et al.* (2004), and the prior standard deviations are chosen to be conservative.

For our analysis of 47 Tuc we proceed through a series of model fits under different conditions. Before describing these fits, we first describe our data-reduction approach. Since we do not believe the stellar evolution models to be accurate for faint (higher-magnitude) MS/RG stars, we discard MS/RG stars with photometric magnitudes in the $F814W$ wavelength band greater than 20. We also remove stars that we cannot model with the computer-based stellar evolution model. Specifically, we remove *blue stragglers*—main-sequence stars which appear younger than other stars on the main-sequence (Leonard, 1989). Several theories for blue stragglers have been put forth, which include (i) blue stragglers formed later than the other cluster stars and are therefore younger, and (ii) blue stragglers result from stellar collisions (e.g., Leonard, 1989). Regardless of which explanation for blue stragglers is correct, if any, these stars exhibit more complicated physics than the current computer models of stellar evolution can capture. Since 47 Tuc is known to contain at least twenty-one blue stragglers (Paresce *et al.*, 1991), we remove four that are obvious. If we do not remove these blue stragglers, we obtain fitted values for θ_{age} that are much too young. After pruning the observed data as described above, the final set of MS/RG stars that we use during model fitting is indicated by black points in the main-sequence portions of the CMDs presented in Figure 6.3. The four grey points that appear above the main-sequence "turn-off" point (at about magnitude $F814W = 17$) are the blue stragglers we removed. The panels $(a) - (d)$ in Figure 6.3 correspond to four model fits that we describe in turn below. Stars that are not included during a particular model fit are indicated with grey points, and the magenta curves are fitted isochrones.

Before considering the carbon fraction, we first fit our Bayesian hierarchical model using only the MS/RG stars and thus without including the carbon fraction as a model parameter. This is to establish a baseline model fit to allow for comparison. We run our AM algorithm for 51,000 iterations after the tuning period. Inspection of the trace plots indicates that the chain quickly converged to its apparent stationary distribution. We discard the first 1000 iterations as burn-in, and fitted parameters and 95% intervals are based on the remaining 50,000 MCMC draws. The CMD for 47 Tuc with the fitted isochrone from this baseline model fit is presented in Figure 6.3(a), with posterior means (i.e. fitted values) and 95% intervals given in the upper-right corner. We find that, in general, the fitted isochrone agrees with the observed data. We can therefore use the fitted values of the cluster parameters as baseline values when examining subsequent model fits. For every model fit that follows, we are always fitting the carbon fraction.

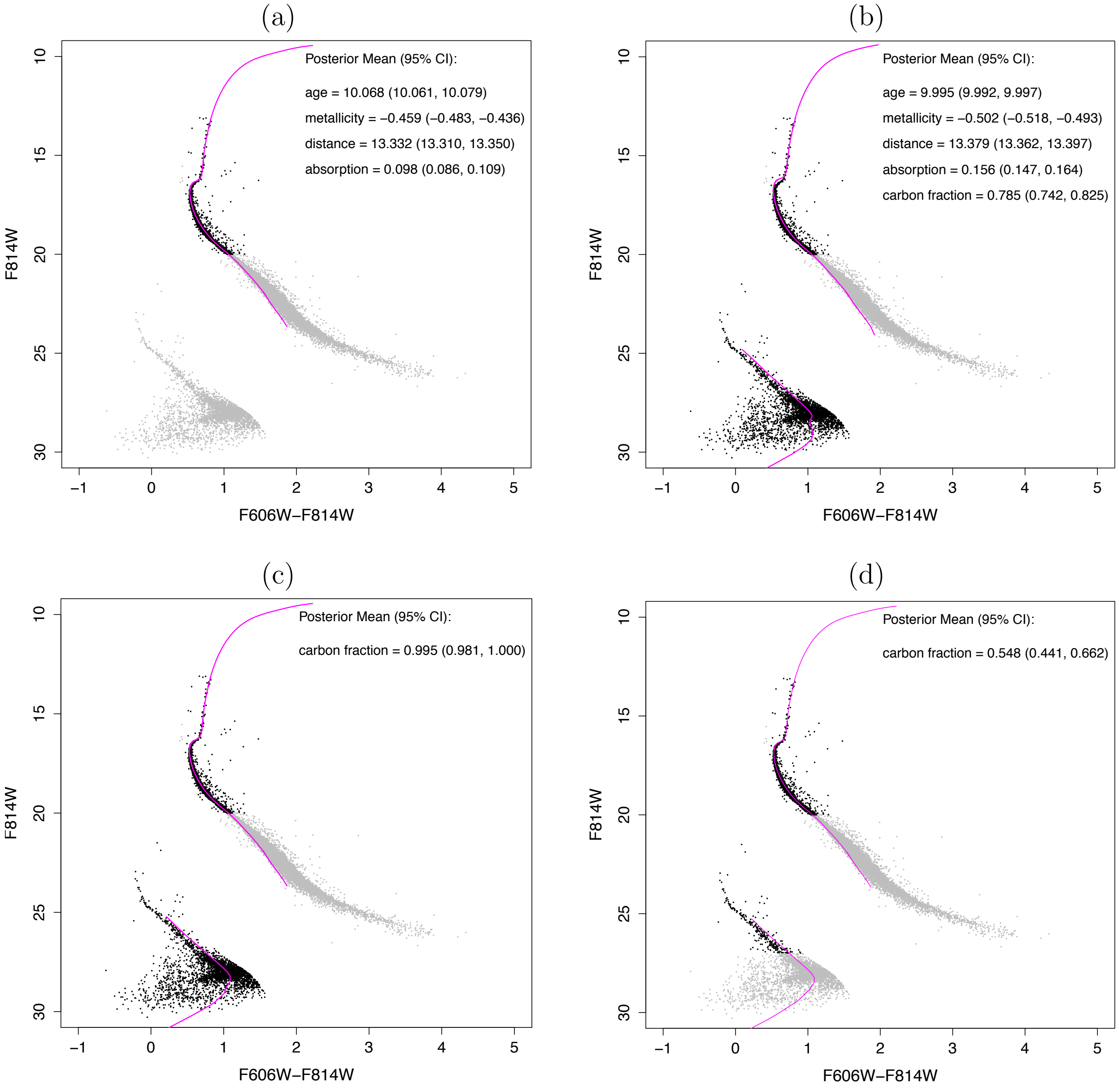


Figure 6.3: Four model fits for 47 Tuc. Black points indicate stars that were included in observed data for a particular model fit, while grey points indicate stars that were not included. The magenta curves are fitted isochrones for the particular model fits. *(a)* A baseline model fit, using only MS/RG stars, truncated to $F814W < 20$. *(b)* A model fit using the MS/RG stars from (a) in addition to all the WDs in the observed data. *(c)* A model fit with the same data as (b), but fixing the values of the cluster parameters except for θ_C to their posterior means under the baseline model. *(d)* The model fit after truncating the observed WDs to $F814W < 27$, and fixing the cluster parameters as in (c).

Next, we fit our Bayesian hierarchical model using all the WDs in addition to the MS/RG stars. We again run our AM algorithm for 51,000 iterations after the tuning period. Inspection of the trace plots indicates that this chain converged to its apparent stationary distribution by iteration 11000; we discard the first 11000 iterations as burn-in. The CMD with the fitted isochrone from this model fit is presented in Figure 6.3(b), with posterior means and 95% intervals given in the upper-right corner. In particular, the fitted value for θ_C is 0.785. For this model fit, we find that the fitted isochrone does not predict the observed data as well as the baseline model fit. The discrepancy between the observed data and the fitted isochrone is particularly apparent at the point where stars are turning off the main sequence, at around $F814W = 17$. There, we can distinguish a gap between the observed data and the fitted isochrone. Because the fitted isochrone "turns off" at a lower magnitude than in the baseline fit, we obtain a lower fitted value for θ_{age}. The fitted value we obtain of 9.995 $\log_{10}$(years), or approximately 9.89 billion years, is lower than the 10 billion years astrophysicists typically adopt as a lower limit on the age of globular clusters. We also observe that the white dwarf cooling sequence is "bluer" than the fitted isochrone (i.e., the stars in the white dwarf cooling sequence fall to the left of the fitted isochrone in the CMD).

For our next model fit, we fix the values of the cluster parameters to their posterior means under the baseline model fit, excluding the carbon fraction. With this, the fitted values for θ_{age}, $\theta_{[\text{Fe/H}]}$, θ_{m-M_V}, and θ_{A_V} are determined by the MS/RG stars alone. We run our AM algorithm for 51,000 iterations after the tuning period, and inspection of the trace plots shows that the chain quickly converged to its apparent stationary distribution. We discard the first 1000 iterations as burn-in, and the fitted value and 95% interval for θ_C are based on the remaining 50,000 MCMC draws. The CMD with the fitted isochrone from this model fit is presented in Figure 6.3(c), with the posterior mean and 95% interval for θ_C given in the upper-right corner. Notice that the fitted isochrone through the MS/RG stars is identical to the fitted isochrone in Figure 6.3(a), which is a result of fixing θ_{age}, $\theta_{[\text{Fe/H}]}$, θ_{m-M_V}, and θ_{A_V} to their posterior means under the baseline model fit. Unfortunately, the WDs are still bluer than the fitted isochrone in the white dwarf cooling sequence. In addition, the posterior mode of θ_C is 1.0, which is unrealistic.

After consideration of the model fits thus far, we believe that contamination in the WDs from stars belonging to the SMC is responsible for the unrealistic fitted values for θ_C. When ancillary information related to the proper motions of these stars is released, we can probabilistically separate 47 Tuc WDs from stars belonging to the SMC. For now, we discard WDs with a magnitude in the $F814W$ wavelength band greater than 27. We are now fairly confident that most of the WDs indicated by black points in Figure 6.3(d), which are the WDs we include for our final model fit, belong to 47 Tuc and not the SMC. We again fix θ_{age}, $\theta_{[\text{Fe/H}]}$, θ_{m-M_V}, and θ_{A_V} to their posterior means under the baseline model fit since we believe those to be reasonable values for 47 Tuc. We run our AM algorithm for 51,000 iterations after the tuning period, and inspection of the trace plot shows that the chain quickly converged to its apparent stationary distribution. After discarding the first 1000 iterations as burn-in, the posterior mean of θ_C from the remaining 50,000 MCMC draws is 0.548, with a 95% interval of (0.441, 0.662). This estimate of θ_C is much more credible. Nevertheless, the white dwarf cooling sequence remains bluer than the fitted isochrone. We attribute this discrepancy to inadequacies in $\boldsymbol{G}_{\text{WD}}$, which we discuss further in Section 6.6.

6.6 Summary and Discussion

Empirically determining the carbon fraction in white dwarfs is interesting because (i) it provides information about astrophysically important nuclear reaction rates that have large experimental uncertainties, (ii) it affects the white dwarf cooling rate and therefore the implied age of any white dwarf, and (iii) it is a test of the stellar evolution models, and fills in a gap in our understanding of late-stage stellar evolution. To estimate the carbon fraction of WDs, we extended our Bayesian hierarchical model for single-population globular clusters to include the carbon fraction as a model parameter, incorporating the updated Montgomery *et al.* (1999) white dwarf cooling model that was developed for this project. As far as we are aware, ours is the first-ever attempt to estimate the carbon fraction in WDs, as opposed to simply adopting a common value for model fitting. We demonstrated with a numerical study that adopting an incorrect value for the carbon fraction can lead to incorrect estimates of white-dwarf-based cluster ages. Finally, we fit our Bayesian hierarchical model with observed data from the globular cluster 47 Tuc under a variety of conditions to estimate the carbon fraction.

In all of our model fits involving WDs, the observed white dwarf cooling sequence is bluer than the fitted isochrone, which we attribute to inadequacies in $\boldsymbol{G}_{\text{WD}}$. In Figure 6.4, we present the plot that appears in Figure 6.3(a), corresponding to the baseline model fit. We add fitted isochrones for the white dwarf cooling sequence that were constructed using the values for θ_{age}, $\theta_{[\text{Fe/H}]}$, θ_{m-M_V}, and θ_{A_V} that produced the fitted isochrone for the MS/RG stars (i.e., the posterior means of the baseline model fit) and different values for θ_C. That is, we are presenting the range of possible model fits in the white-dwarf group, given the values for θ_{age}, $\theta_{[\text{Fe/H}]}$, θ_{m-M_V}, and θ_{A_V} that we obtained from fitting the model with the MS/RG stars alone. We notice that the observed white dwarf cooling sequence is bluer than the fitted isochrones, regardless of the chosen value for θ_C. Since the baseline model is fit with $\boldsymbol{G}_{\text{MS/RG}}$ alone, and θ_C is only an input to $\boldsymbol{G}_{\text{WD}}$, this points to a discrepancy between the separate computer models for MS/RG stars and WDs. While it is possible that misspecifying α or the population distribution of the initial masses is contributing to the discrepancy, we are more concerned about inaccuracies in the computer models themselves. Nevertheless, the discrepancy we observe provides feedback that is useful for designing computer models that better represent the observed data.

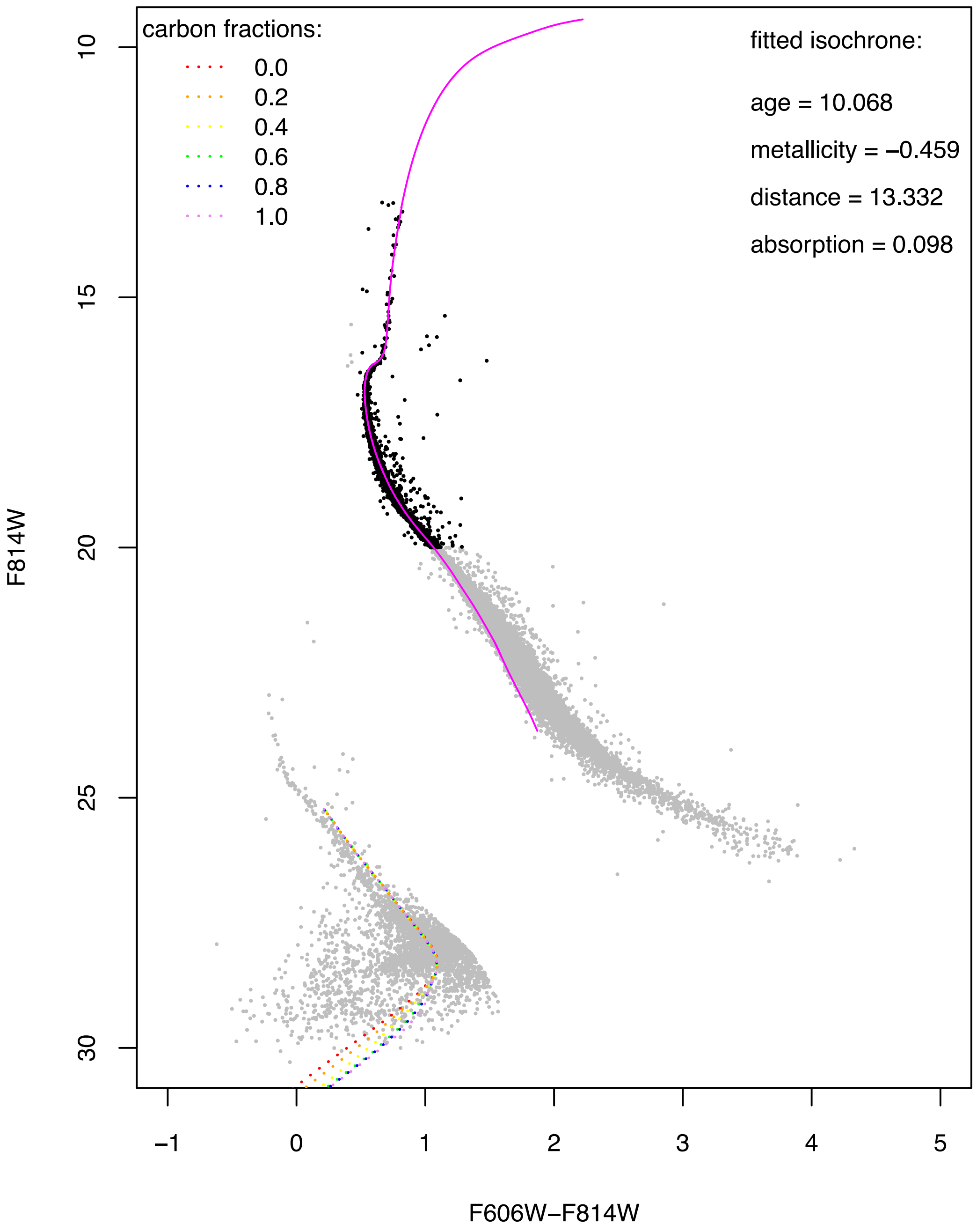


Figure 6.4: Examining $\boldsymbol{G}_{\mathrm{WD}}$. The solid magenta curve along the MS/RG stars is the fitted isochrone from the baseline model fit. The dotted curves in the white-dwarf group are the fitted isochrones produced using the posterior means for θ_{age}, $\theta_{[\mathrm{Fe/H}]}$, θ_{m-M_V}, and θ_{A_V}, along with different values for θ_C as indicated in the upper left. We find that regardless of the chosen value for θ_C, the white dwarf cooling sequence is bluer than the fitted isochrones, which points to inadequacies in $\boldsymbol{G}_{\mathrm{WD}}$.

Bibliography

Adams, R. and Bischof, L. (1994). Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.*, **16**(6), 641–647.

Althaus, L. G. and Benvenuto, O. G. (1998). Evolution of DA white dwarfs in the context of a new theory of convection. *Monthly Notices of the Royal Astronomical Society*, **296**, 206–216.

Aschwanden, M. (2010). Image processing techniques and feature recognition in solar physics. *Solar Physics*, **262**, 235–275. 10.1007/s11207-009-9474-y.

Basu, S. (2013). The peculiar solar cycle 24 - where do we stand? *Journal of Physics Conference Series*, **440**(1), 012001.

Bedin, L. R., Piotto, G., Anderson, J., Cassisi, S., King, I. R., Momany, Y., and Carraro, G. (2004). ω Centauri: The Population Puzzle Goes Deeper. *The Astrophysical Journal Letters*, **605**, L125–L128.

Bekki, K. (2011). Secondary star formation within massive star clusters: origin of multiple stellar populations in globular clusters. *Monthly Notices of the Royal Astronomical Society*, **412**, 2241–2259.

Bekki, K., Campbell, S. W., Lattanzio, J. C., and Norris, J. E. (2007). Origin of abundance inhomogeneity in globular clusters. *Monthly Notices of the Royal Astronomical Society*, **377**, 335–351.

Benestad, R. E. (2005). A review of the solar cycle length estimates. *Geophysical Research Letters*, **32**, 15714.

Bergeron, P., Wesemael, F., and Beauchamp, A. (1995). Photometric Calibration of Hydrogen- and Helium-Rich White Dwarf Models. *Publications of the Astronomical Society of the Pacific*, **107**, 1047.

Bolduc, L. (2002). GIC observations and studies in the Hydro-Québec power system. *Journal of Atmospheric and Solar-Terrestrial Physics*, **64**(16), 1793–1802.

Bonev, B. P., Penev, K. M., and Sello, S. (2004). Long-term solar variability and the solar cycle in the 21st century. *The Astrophysical Journal Letters*, **605**, L81–L84.

Breiman, L. (2001). Random forests. *Mach. Learn.*, **45**(1), 5–32.

Cannon, R. D. and Stobie, R. S. (1973). Photometry of southern globular clusters I. Bright stars in omega Centauri. *Monthly Notices of the Royal Astronomical Society*, **162**, 207.

Carretta, E., Gratton, R. G., Clementini, G., and Fusi Pecci, F. (2000). Distances, Ages, and Epoch of Formation of Globular Clusters. *The Astrophysical Journal*, **533**, 215–235.

Carretta, E., Bragaglia, A., Gratton, R. G., Leone, F., Recio-Blanco, A., and Lucatello, S. (2006). Na-O anticorrelation and HB. I. The Na-O anticorrelation in NGC 2808. *Astronomy and Astrophysics*, **450**, 523–533.

Carretta, E., Bragaglia, A., Gratton, R. G., Recio-Blanco, A., Lucatello, S., D'Orazi, V., and Cassisi, S. (2010). Properties of stellar generations in globular clusters and relations with global parameters. *Astronomy & Astrophysics*, **516**, A55.

Charbonneau, P. (2007). Babcock-Leighton models of the solar cycle: questions and issues. *Advances in Space Research*, **39**, 1661–1669.

Charbonneau, P. and Dikpati, M. (2000). Stochastic fluctuations in a Babcock-Leighton model of the solar cycle. *The Astrophysical Journal*, **543**, 1027–1043.

Choudhuri, A. R. (1992). Stochastic fluctuations of the solar dynamo. *Astronomy and Astrophysics*, **253**, 277–285.

Choudhuri, A. R., Chatterjee, P., and Jiang, J. (2007). Predicting solar cycle 24 with a solar dynamo model. *Physical Review Letters*, **98**(13), 131103.

Colak, T. and Qahwaji, R. S. R. (2009). Automated mcintosh-based classification of sunspot groups using MDI images. *Solar Physics*, **248**(2), 277–296.

Curto, J., Blanca, M., and Martnez, E. (2008). Automatic sunspots detection on full-disk solar images using mathematical morphology. *Solar Physics*, **250**, 411–429. 10.1007/s11207-008-9224-6.

D'Antona, F. and Caloi, V. (2008). The fraction of second generation stars in globular clusters from the analysis of the horizontal branch. *Monthly Notices of the Royal Astronomical Society*, **390**, 693–705.

De Gennaro, S. A. (2009). *White dwarfs and the ages of stellar populations*. Ph.D. thesis, The University of Texas at Austin.

DeGennaro, S., von Hippel, T., Jefferys, W. H., Stein, N., van Dyk, D., and Jeffery, E. (2009). Inverting Color-Magnitude Diagrams to Access Precise Star Cluster Parameters: A New White Dwarf Age for the Hyades. *The Astrophysical Journal*, **696**, 12–23.

di Criscienzo, M., Ventura, P., D'Antona, F., Milone, A., and Piotto, G. (2010). The helium spread in the globular cluster 47 Tuc. *Monthly Notices of the Royal Astronomical Society*, **408**, 999–1005.

Dikpati, M. and Gilman, P. A. (2006). Simulating and predicting solar cycles using a flux-transport dynamo. *The Astrophysical Journal*, **649**, 498–514.

Dikpati, M., de Toma, G., and Gilman, P. A. (2006). Predicting the strength of solar cycle 24 using a flux-transport dynamo-based tool. *Geophysical Research Letters*, **33**, 5102.

Dotter, A., Chaboyer, B., Jevremović, D., Kostov, V., Baron, E., and Ferguson, J. W. (2008). The Dartmouth Stellar Evolution Database. *The Astrophysical Journal Supplement*, **178**, 89–101.

Ferraro, F. R., Beccari, G., Rood, R. T., Bellazzini, M., Sills, A., and Sabbi, E. (2004). Discovery of Another Peculiar Radial Distribution of Blue Stragglers in Globular Clusters: The Case of 47 Tucanae. *The Astrophysical Journal*, **603**, 127–134.

Fontaine, G., Brassard, P., and Bergeron, P. (2001). The Potential of White Dwarf Cosmochronology. *Publications of the Astronomical Society of the Pacific*, **113**, 409–435.

Forbes, D. A. and Bridges, T. (2010). Accreted versus in situ Milky Way globular clusters. *Monthly Notices of the Royal Astronomical Society*, **404**, 1203–1214.

Gelman, A., Roberts, G., and Gilks, W. (1996). Efficient metropolis jumping rules. *Bayesian Statistics*, **5**(599-608), 42.

Georgieva, K. (2011). Why the sunspot cycle is double peaked. *ISRN Astronomy and Astrophysics*, **2011**.

Gil-Alana, L. A. (2009). Time series modeling of sunspot numbers using long-range cyclical dependence. *Solar Physics*, **257**, 371–381.

Girardi, L., Bressan, A., Bertelli, G., and Chiosi, C. (2000). VizieR Online Data Catalog: Low-mass stars evolutionary tracks and isochrones (Girardi+, 2000). *VizieR Online Data Catalog*, **414**, 10371.

Gnevyshev, M. (1967). On the 11-years cycle of solar activity. *Solar Physics*, **1**(1), 107–120.

Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing, 2nd ed.* Prentice Hall, Upper Saddle River, N.J. USA.

Gratton, R., Sneden, C., and Carretta, E. (2004). Abundance Variations Within Globular Clusters. *Annual Review of Astronomy & Astrophysics*, **42**, 385–440.

Gratton, R. G. and Carretta, E. (2010). Diluting the material forming the second generation stars in globular clusters: the contribution by unevolved stars. *Astronomy & Astrophysics*, **521**, A54.

Greenkorn, R. A. (2012). A comparison of the 10.7-cm radio flux values and the International Sunspot Numbers for solar activity cycles 19, 20, and 21. *Solar Physics*, **280**, 205–221.

Grundahl, F., Stetson, P. B., and Andersen, M. I. (2002). The ages of the globular clusters M 71 and 47 Tuc from Strömgren uvby photometry. Evidence for high ages. *Astronomy & Astrophysics*, **395**, 481–497.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, **7**(2), 223–242.

Harel, O. and Zhou, X.-H. (2007). Multiple imputation: review of theory, implementation and software. *Statistics in Medicine*, **26**(16), 3057–3077.

Harris, W. E. (1996). A Catalog of Parameters for Globular Clusters in the Milky Way. *The Astronomical Journal*, **112**, 1487.

Hathaway, D. H. and Wilson, R. M. (2006). Geomagnetic activity indicates large amplitude for sunspot cycle 24. *Geophysical Research Letters*, **33**, 18101.

Hathaway, D. H., Wilson, R. M., and Reichmann, E. J. (1994). The shape of the sunspot cycle. *Solar Physics*, **151**, 177–190.

Hathaway, D. H., Wilson, R. M., and Reichmann, E. J. (2002). Group sunspot numbers: sunspot cycle characteristics. *Solar Physics*, **211**, 357–370.

Hesser, J. E. and Bell, R. A. (1980). CN variations among main-sequence 47 Tucanae stars. *The Astrophysical Journal Letters*, **238**, L149–L153.

Hudson, H. S. (2007). The unpredictability of the most energetic solar events. *The Astrophysical Journal Letters*, **663**, L45–L48.

Ireland, J., Young, C. A., McAteer, R. T. J., Whelan, C., Hewett, R. J., and Gallagher, P. T. (2008). Multiresolution analysis of active region magnetic structure and its correlation with the Mt. Wilson classification and flaring activity. *Solar Physics*.

Janes, K. and Demarque, P. (1983). The ages and compositions of old clusters. *The Astrophysical Journal*, **264**, 206–214.

Jeffery, E. J., von Hippel, T., DeGennaro, S., van Dyk, D. A., Stein, N., and Jefferys, W. H. (2011). The White Dwarf Age of NGC 2477. *The Astrophysical Journal*, **730**, 35.

Kalirai, J. S., Anderson, J., Dotter, A., Richer, H. B., Fahlman, G. G., Hansen, B. M. S., Hurley, J., Reid, I. N., Rich, R. M., and Shara, M. M. (2013). Ultra-Deep Hubble Space Telescope Imaging of the Small Magellanic Cloud: The Initial Mass Function of Stars with $M < \sim 1M_{\text{sun}}$. *The Astrophysical Journal*, **763**, 110.

Kane, R. P. (2001). Did predictions of the maximum sunspot number for solar cycle 23 come true? *Solar Physics*, **202**, 395–406.

Karak, B. B. and Nandy, D. (2012). Turbulent pumping of magnetic flux reduces solar cycle memory and thus impacts predictability of the Sun's activity. *The Astrophysical Journal Letters*, **761**, L13.

Kilcik, A. and Ozguc, A. (2014). One possible reason for double-peaked maxima in solar cycles: is a second maximum of solar cycle 24 expected? *Solar Physics*, **289**(4), 1379–1386.

Kuchner, M. J. and Seager, S. (2005). Extrasolar Carbon Planets. *ArXiv Astrophysics e-prints*.

Lee, J.-W., Carney, B. W., Fullton, L. K., and Stetson, P. B. (2001). Near-Infrared Photometry of Globular Clusters NGC 6287 and NGC 6341 (M92): The Formation of the Galactic Halo. *The Astronomical Journal*, **122**, 3136–3154.

Leonard, P. J. T. (1989). Stellar collisions in globular clusters and the blue straggler problem. *The Astronomical Journal*, **98**, 217–226.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics. Wiley, New York, 2nd edition.

Machholz, D. (2002). *The Observing Guide to the Messier Marathon: A Handbook and Atlas*. Cambridge University Press.

Marín-Franch, A., Aparicio, A., Piotto, G., Rosenberg, A., Chaboyer, B., Sarajedini, A., Siegel, M., Anderson, J., Bedin, L. R., Dotter, A., Hempel, M., King, I., Majewski, S., Milone, A. P., Paust, N., and Reid, I. N. (2009). The ACS Survey of Galactic Globular Clusters. VII. Relative Ages. *The Astrophysical Journal*, **694**, 1498–1516.

McIntosh, P. S. (1990). The classification of sunspot groups. *Solar Physics*, **125**, 251–267.

Metcalfe, T. S. (2003). White Dwarf Asteroseismology and the ^{12}C(α,γ)^{16}O Rate. *The Astrophysical Journal Letters*, **587**, L43–L46.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, **21**, 1087–1092.

Milone, A. P., Piotto, G., Bedin, L. R., King, I. R., Anderson, J., Marino, A. F., Bellini, A., Gratton, R., Renzini, A., Stetson, P. B., Cassisi, S., Aparicio, A., Bragaglia, A., Carretta, E., D'Antona, F., Di Criscienzo, M., Lucatello, S., Monelli, M., and Pietrinferni, A. (2012). Multiple Stellar Populations in 47 Tucanae. *The Astrophysical Journal*, **744**, 58.

Montgomery, M. H., Klumpe, E. W., Winget, D. E., and Wood, M. A. (1999). Evolutionary Calculations of Phase Separation in Crystallizing White Dwarf Stars. *The Astrophysical Journal*, **525**, 482–491.

Muñoz-Jaramillo, A., Dasi-Espuig, M., Balmaceda, L. A., and DeLuca, E. E. (2013). Solar cycle propagation, memory, and prediction: insights from a century of magnetic proxies. *The Astrophysical Journal Letters*, **767**, L25.

NASA (2009). *SDO: Our Eye on the Sun. A Guide to the Mission and Purpose of NASA's Solar Dynamics Observatory*. National Aeronautics and Space Administration.

NRC (2008). *Severe Space Weather Events—Understanding Societal and Economic Impacts*. National Research Council.

Paresce, F., Meylan, G., Shara, M., Baxter, D., and Greenfield, P. (1991). Blue stragglers in the core of the globular cluster 47 Tucanae. *Nature*, **352**, 297–301.

Perryman, M. A. C., Brown, A. G. A., Lebreton, Y., Gomez, A., Turon, C., Cayrel de Strobel, G., Mermilliod, J. C., Robichon, N., Kovalevsky, J., and Crifo, F. (1998). The Hyades: distance, structure, dynamics, and age. *Astronomy & Astrophysics*, **331**, 81–120.

Pesnell, W. D. (2012). Solar cycle predictions (invited review). *Solar Physics*, **281**(1), 507–532.

Piotto, G. (2009). Observations of multiple populations in star clusters. In E. E. Mamajek, D. R. Soderblom, and R. F. G. Wyse, editors, *IAU Symposium*, volume 258 of *IAU Symposium*, pages 233–244.

Piotto, G., Bedin, L. R., Anderson, J., King, I. R., Cassisi, S., Milone, A. P., Villanova, S., Pietrinferni, A., and Renzini, A. (2007). A Triple Main Sequence in the Globular Cluster NGC 2808. *The Astrophysical Journal Letters*, **661**, L53–L56.

Piotto, G., Milone, A. P., Bedin, L. R., Anderson, J., King, I. R., Marino, A. F., Nardiello, D., Aparicio, A., Barbuy, B., Bellini, A., Brown, T. M., Cassisi, S., Cool, A. M., Cunial, A., Dalessandro, E., D'Antona, F., Ferraro, F. R., Hidalgo, S., Lanzoni, B., Monelli, M., Ortolani, S., Renzini, A., Salaris, M., Sarajedini, A., van der Marel, R. P., Vesperini, E., and Zoccali, M. (2015). The Hubble Space Telescope UV Legacy Survey of Galactic Globular Clusters. I. Overview of the Project and Detection of Multiple Stellar Populations. *The Astronomical Journal*, **149**, 91.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Alves, M. I. R., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Aussel, H., and et al. (2014). Planck 2013 results. I. Overview of products and scientific results. *Astronomy & Astrophysics*, **571**, A1.

Prantzos, N., Charbonnel, C., and Iliadis, C. (2007). Light nuclei in galactic globular clusters: constraints on the self-enrichment scenario from nucleosynthesis. *Astronomy & Astrophysics*, **470**, 179–190.

Rehm, K. E. (2012). The Origin of Oxygen in the Universe - A new approach to an Old Question. *Journal of Physics Conference Series*, **337**(1), 012006.

Renedo, I., Althaus, L. G., Miller Bertolami, M. M., Romero, A. D., Córsico, A. H., Rohrmann, R. D., and García-Berro, E. (2010). New Cooling Sequences for Old White Dwarfs. *The Astrophysical Journal*, **717**, 183–195.

Richards, J. W., Freeman, P. E., Lee, A. B., and Schafer, C. M. (2009). Accurate parameter estimation for star formation history in galaxies using sdss spectra. *Monthly Notices of the Royal Astronomical Society*, **399**(2), 1044–1057.

Roberts, G. O. and Rosenthal, J. S. (2007). Coupling and ergodicity of adaptive markov chain monte carlo algorithms. *Journal of Applied Probability*, pages 458–475.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, **18**(2), 349–367.

Rosenthal, J. S. *et al.* (2011). Optimal proposal distributions and adaptive mcmc. *Handbook of Markov Chain Monte Carlo*, pages 93–112.

Salaris, M., Serenelli, A., Weiss, A., and Miller Bertolami, M. (2009). Semi-empirical White Dwarf Initial-Final Mass Relationships: A Thorough Analysis of Systematic Uncertainties Due to Stellar Evolution Models. *The Astrophysical Journal*, **692**, 1013–1032.

Sandage, A. (1962). Photometric Data for the Old Galactic Cluster NGC 188. *The Astrophysical Journal*, **135**, 333.

Schüssler, M. (2007). Are solar cycles predictable? *Astronomische Nachrichten*, **328**, 1087.

Searle, L. and Zinn, R. (1978). Compositions of halo clusters and the formation of the galactic halo. *The Astrophysical Journal*, **225**, 357–379.

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press.

Shapoval, A., Le Mouël, J. L., Courtillot, V., and Shnirman, M. (2013). Two regimes in the regularity of sunspot number. *The Astrophysical Journal*, **779**, 108.

Simpson, M., Niemi, J., and Roy, V. (2015). Interweaving markov chain monte carlo strategies for efficient estimation of dynamic linear models. Submitted.

Soille, P. (2003). *Morphological Image Analysis: Principles and Applications*. Springer, Berlin, second edition.

Solanki, S. K., Usoskin, I. G., Kromer, B., Schüssler, M., and Beer, J. (2004). Unusual activity of the Sun during recent decades compared to the previous 11,000 years. *Nature*, **431**, 1084–1087.

Stein, N. M., van Dyk, D. A., von Hippel, T., DeGennaro, S., Jeffery, E. J., and Jefferys, W. H. (2013). Combining computer models to account for mass loss in stellar evolution. *Statistical Analysis and Data Mining*, **6**(1), 34–52.

Stenning, D., Kashyap, V., Lee, T., van Dyk, D., and Young, C. (2012). Morphological image analysis and its applications to sunspot classification. In E. D. Feigelson and J. Babu, editors, *Statistical Challenges in Modern Astronomy V*.

Stenning, D. C., Lee, T. C. M., van Dyk, D. A., Kashyap, V., Sandell, J., and Young, C. A. (2013). Morphological feature extraction for statistical learning with applications to solar image data. *Statistical Analysis and Data Mining*, **6**(4), 329–345.

Stenning, D. C., van Dyk, D. A., Yu, Y., and Kashyap, V. (2015). A bayesian analysis of the solar cycle using multiple proxy variables. In S. Upadhyay, D. Dey, U. Siggh, and A. Loganathan, editors, *Current Trends in Bayesian Methodology with Applications*. Chapman & Hall/CRC Press.

Temmer, M. (2010). Statistical properties of flares and sunspots over the solar cycle. In S. R. Cranmer, J. T. Hoeksema, and J. L. Kohl, editors, *SOHO-23: Understanding a Peculiar Solar Minimum*, volume 428 of *Astronomical Society of the Pacific Conference Series*, page 161.

van Dyk, D. A., Degennaro, S., Stein, N., Jefferys, W. H., and von Hippel, T. (2009). Statistical analysis of stellar evolution. *Annals of Applied Statistics*, **3**, 117–143.

Villanova, S., Piotto, G., King, I. R., Anderson, J., Bedin, L. R., Gratton, R. G., Cassisi, S., Momany, Y., Bellini, A., Cool, A. M., Recio-Blanco, A., and Renzini, A. (2007). The Multiplicity of the Subgiant Branch of ω Centauri: Evidence for Prolonged Star Formation. *The Astrophysical Journal*, **663**, 296–314.

von Hippel, T., Jefferys, W. H., Scott, J., Stein, N., Winget, D. E., De Gennaro, S., Dam, A., and Jeffery, E. (2006). Inverting Color-Magnitude Diagrams to Access Precise Star Cluster Parameters: A Bayesian Approach. *The Astrophysical Journal*, **645**, 1436–1447.

Waldmeier, M. (1935). Neue eigenschaften der sonnenfleckenkurve. *Astronomische Mitteilungen der Eidgenössischen Sternwarte Zurich*, **133**(105).

Watari, S. (2008). Forecasting solar cycle 24 using the relationship between cycle length and maximum sunspot number. *Space Weather*, **6**(12).

Weidemann, V. (2000). Revision of the initial-to-final mass relation. *Astronomy & Astrophysics*, **363**, 647–656.

Williams, K. A., Bolte, M., and Koester, D. (2009). Probing the Lower Mass Limit for Supernova Progenitors and the High-Mass End of the Initial-Final Mass Relation from White Dwarfs in the Open Cluster M35 (NGC 2168). *The Astrophysical Journal*, **693**, 355–369.

Wolf, R. (1852). Neue untersuchungen über die periode der sonnen-flecken und ihre bedeutung. *Viertel. Natur. Ges. Bern*, **254**(179).

Wood, M. A. (1992). Constraints on the age and evolution of the Galaxy from the white dwarf luminosity function. *The Astrophysical Journal*, **386**, 539–561.

Woodard, D. B., Crainiceanu, C., and Ruppert, D. (2012). Hierarchical adaptive regression kernels for regression with functional predictors. *Journal of Computational and Graphical Statistics*.

Yi, S., Demarque, P., Kim, Y.-C., Lee, Y.-W., Ree, C. H., Lejeune, T., and Barnes, S. (2001). Toward Better Age Estimates for Stellar Populations: The Y^2 Isochrones for Solar Mixture. *The Astrophysical Journal Supplement*, **136**, 417–437.

Yu, Y., van Dyk, D. A., Kashyap, V. L., and Young, C. A. (2012). A Bayesian analysis of the correlations among sunspot cycles. *Solar Physics*, **281**, 847–862.

Zoccali, M., Renzini, A., Ortolani, S., Bragaglia, A., Bohlin, R., Carretta, E., Ferraro, F. R., Gilmozzi, R., Holberg, J. B., Marconi, G., Rich, R. M., and Wesemael, F. (2001). The White Dwarf Distance to the Globular Cluster 47 Tucanae and its Age. *The Astrophysical Journal*, **553**, 733–743.

Appendix A

Morphological Operations

Morphological operations (see, e.g., Serra (1982); Soille (2003)) are set theoretic manipulations of images designed to enhance, recognize, and extract features of interest. MM provides tools that allow us to manipulate regions by filtering, thinning, pruning, etc., and describe the characteristics of regions, such as their shapes, boundaries, and skeletons. Here, we briefly describe the basics of morphological operations.

Morphological operators operate on either grayscale or binary images. Here, for simplicity and ease of explanation, we concentrate on binary images.

All morphological operations involve a structuring element (SE), which is usually applied as a filter or a convolution kernel to the image. The SE is typically a simple shape, such as a 3×3 square array, or subsets thereof.

The two fundamental morphological operations are erosion and dilation:

The Erosion of an image is a set of all pixels in an image that fully contain all possible translations of the SE. In other words, for a SE Y and an image I,

$$\text{Erode}(I|Y) = \{z|Y_z \subseteq I\}\,,$$

where z are all the points such that Y, translated by z, is contained in I. Erosion shrinks an image, removing points not within in the SE and enlarging the background of an image.

The Dilation of an image is the reverse of an erosion, when the image is expanded to include all points of the SE that may overlap any point in the image. Note that in this case the SE is flipped around the origin prior to the translations, in a manner very similar to that of a convolution. Thus,

$$\text{Dilate}(I|Y) = \{z|[\hat{Y}_z \cap I] \subseteq I\}\,,$$

where $\hat{Y}$ represents Y flipped about the origin. The dilation of an image is an exact dual to an erosion of the background,

$$\text{Erode}(I|Y)^c = \text{Dilate}(I^c|\hat{Y})\,,$$

where the superscript ^{c} refers to set complement.

Sequential applications of dilation and erosion are widely used morphological operators:

Opening: An erosion, followed by a dilation, is called an Opening, i.e.,

$$\text{Open}(I|S) = \text{Dilate}(\text{Erode}(I|Y))\,.$$

Thus, an opening operation ends up removing pixels from the image overall. Applying an opening operation has a smoothing effect on the image, enhancing contours and rounded structures while removing disconnected shapes.

Closing: The two operations carried out in reverse order is called a Closing, i.e.,

$$\text{Close}(I|Y) = \text{Erode}(\text{Dilate}(I|Y))\,.$$

The morphological closing operation adds in pixels to smooth the image.

Skeleton: A morphological skeleton is an irreducible set of points that represent an extended region, such that each pixel on the skeleton is in some sense at the maximal distance from the edge of the region. The skeleton of an image I, given a SE Y (usually a 3×3 cell suffices),

$$\begin{aligned} \text{Skeleton}(I|Y) &= \cup_{k=0}^{K} \text{Skeleton}(I|kY)\,, \\ \text{Skeleton}(I|kY) &= \text{Erode}^k(I|Y) - \text{Open}(\text{Erode}^k(I|Y))\,, \end{aligned}$$

where $\text{Erode}^k(I|Y)$ indicates k successive erosions of I with Y, and K is the largest value of k such that $\text{Erode}^{K+1}(I|Y) = \emptyset$, a null set. Thus, the Skeleton operator acts to "thin" an image, leaving only the spine of a region of interest.

Prune: Skeleton and other thinning operations tend to leave small offshoots that result from irregularities in the boundaries of the regions of interest. This is usually accomplished by, first, finding all points along a skeleton that has more then two neighbors, excluding them from the image temporarily, removing all contiguous sets of pixels with fewer than some required number of pixels (typically, 4), and then reinserting the points removed previously, and dilating the image and taking its intersection with the original skeleton until no new pixels are found:

$$\begin{aligned}
I_1 &= \text{convolve}(\text{Skeleton}(I|Y)|\text{3x3}) \\
I_2 &= \{x|I_1 \geq 3\} \\
I_3 &= \{x|\text{Area}(x) > 4\} \\
I_4 &= I_3 \cup I_2 \\
\text{Prune}(\text{Skeleton}(I|Y)) &= \{\text{Dilate}(I_4|\text{3x3}) \cap \text{Skeleton}(I|Y)\}^K ,
\end{aligned}$$

where the last operation is carried out K times such that

$$\{\text{Dilate}(I_4|\text{3x3}) \cap \text{Skeleton}(I|Y)\}^K \equiv \{\text{Dilate}(I_4|\text{3x3}) \cap \text{Skeleton}(I|Y)\}^{K+1} .$$