UNIVERSITY OF CALIFORNIA SAN DIEGO

In search of the “Missing Dipole” - unravelling the paradoxes of Earth’s magnetic field strength.

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Earth Sciences

by

Brendan John Joseph Cych

Committee in charge:

Professor Lisa Tauxe, Chair
Professor Catherine Constable
Professor Eric Fullerton
Professor Jeffrey Gee
Professor Matthias Morzfeld

2022

The Dissertation of Brendan John Joseph Cych is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Lisa Tauxe, the chair of my committee for being an excellent advisor. Not only is she an incredibly experienced mentor to learn from, but she is also supportive and had my interests at heart throughout the course of my PhD. I would also like to thank the rest of my committee Cathy Constable, Eric Fullerton, Jeff Gee and Matti Morzfeld for their helpful guidance on all my projects throughout my time at Scripps, as well as their helpful reviews of the dissertation. All of my coauthors have my thanks for their contributions to each of the papers that form this dissertation. Finally, I would like to thank my family, and my friends in the Scripps community for supporting me, and keeping me sane during the last few tumultuous years.

Chapter 2, in full is a reprint of material from Cych, B., Morzfeld, M., & Tauxe, L. (2021). Bias Corrected Estimation of Paleointensity (BiCEP): An improved methodology for obtaining paleointensity estimates." Geochemistry, Geophysics, Geosystems, 22, e2021GC009755. https://doi.org/10.1029/2021GC009755. The dissertation author was the primary investigator and author of this paper.

Chapter 3, is in preparation for submission with Geochemistry, Geophysics Geosystems as Cych, B., Morzfeld, M., Heslop, D., Maher, S., Gee, J., & Tauxe, L. Thermal Resolution of Unblocking Temperatures (TROUT): A method for "unmixing" multi component magnetizations (In Preparation). The dissertation author was the primary investigator and author of this paper.

Chapter 4, is in preparation submission with Proceedings of the National Academy of Sciences as Cych, B., Tauxe, L., Cromwell, G., Sinton, H., & Koppers, A. Non-dipolar time averaged field for the past ‘million years from new paleointensity results from Hawai‘i, compared to global datasets (In Preparation). The dissertation author was the primary investigator and author of this paper.

VITA

2012-2016 MEarthSci, University of Oxford, United Kingdom

2022 PhD in Earth Sciences, University of California, San Diego

PUBLICATIONS

Cych, B., Tauxe, L., Cromwell, G., Sinton, H., & Koppers, A. Non-dipolar time-averaged field for the past 1.5 million years from new paleointensity results from Hawai'i, compared to global datasets (In preparation for submission)

Cych, B., Morzfeld, M., Heslop, D., Maher, S., & Tauxe, L. Thermal Resolution of Unblocking Temperatures (TROUT): A method for "unmixing" multi component magnetizations (In preparation for submission)

Cych, B., Morzfeld, M., & Tauxe, L. (2021). Bias Corrected Estimation of Paleointensity (BiCEP): An improved methodology for obtaining paleointensity estimates." Geochemistry, Geophysics, Geosystems, 22, e2021GC009755. https://doi.org/10.1029/2021GC009755

Tauxe, L., Santos, C. N., Cych, B., Zhao, X., Roberts, A. P., Nagy, L., & Williams, W. (2021). Understanding nonideal paleointensity recording in igneous rocks: Insights from aging experiments on lava samples and the causes and consequences of "fragile" curvature in Arai plots. Geochemistry, Geophysics, Geosystems, 22, e2020GC009423. https://doi.org/10.1029/2020GC009423

ABSTRACT OF THE DISSERTATION

In search of the “Missing Dipole” - unravelling the paradoxes of Earth’s magnetic field strength.

by

Brendan John Joseph Cych

Doctor of Philosophy in Earth Sciences

University of California San Diego, 2022

Professor Lisa Tauxe, Chair

The structure of the Earth’s magnetic field today is constantly changing, driven by convection currents in the Earth’s outer core. The dominant force in these convection currents is the rotation of the Earth, and so many paleomagnetists assume that when averaged over long enough timescales, the field behaves like a bar magnet centered on the Earth’s spin axis. This is known as the Geocentric Axial Dipole (GAD) hypothesis.

Igneous rocks and other materials can preserve a record of the direction and strength of the Earth’s magnetic field at the time that they formed. Global records of Earth’s magnetic field direction (paleodirection) from igneous rocks over the last 10 million years require only a small deviation from a GAD field, whereas global records of the field strength (paleointensity) display

a latitudinal distribution which is highly inconsistent with a dipole.

One problematic aspect of paleointensity records is that some rocks violate the assumptions of the paleointensity experiment, leading to inaccurate estimates of paleointensity. We describe two statistical methods which deal with different behaviors that are related in some way to these inaccurate results. The Bias Corrected Estimation of Paleointensity (BiCEP) method tries to find a relationship between non-ideal behavior in the experiment and bias in the results, allowing us to obtain accurate paleointensities without excluding specimens from our analyses. The Thermal Resolution of Unblocking Temperatures (TROUT) method allows us to isolate a range of measurement temperatures that are appropriate to use in a paleointensity analysis.

We present a set of high quality paleointensity data from Hawai‘i, which we analyze using the aforementioned BiCEP method. We also reanalyze data from paleointensity studies targeting similar lithologies from northern Israel and Antarctica using the same methodology. We account for the different temporal distributions of these studies by taking the time average of a paleointensity curve constructed for each location. Despite having large uncertainties, our time averages show that the Earth's magnetic field cannot be fit by a dipole when averaged over the last 1.5 Ma, but may have been dipolar before this time.

Chapter 1

Introduction

1.1 Paleomagnetic background

Earth's magnetic field is produced by convection of the liquid iron outer core. This dynamic system leads to a field at the surface which is constantly changing. Since the 16th century, global observations of Earth's magnetic field have been recorded, originally by navigators on ships (Jackson et al., 2000), and most recently by geomagnetic observatories and satellites around the globe. Not many deliberate records of the global magnetic field exist which are older than this (see Stern, 2003 for an historical view).

Paleomagnetism is the study of Earth's magnetic field in the geological past. Rocks can preserve an accidental record of the history of this field, as they contain magnetic minerals which act as a recording medium for the field they were magnetized in. Paleomagnetists use these accidental records to construct timescales which can be used for dating, to reconstruct the motion of tectonic plates, and to make inferences about the history of Earth's core and deep interior in the earliest parts of Earth's history.

Rocks can obtain a magnetization that records Earth's magnetic field in a number of ways. This dissertation deals chiefly with igneous rocks, which acquire a magnetization by cooling in the presence of Earth's magnetic field, a process known as Thermal Remanent Magnetization (TRM). To first order, a TRM obtained by a rock is aligned with the direction of, and proportional to the strength of, the magnetic field that it cooled in. Because of these properties, a TRM can be used to obtain an estimate of the Earth's magnetic field direction (known as the paleodirection), and an estimate of the Earth's magnetic field strength (known as the paleointensity) by normalizing the natural remanent magnetization (NRM) to a TRM acquired in a known lab field. In rapidly cooled igneous rocks like lava flows, TRMs provide an instantaneous snapshot of the Earth's magnetic field at the time of eruption.

1.2 The structure of Earth's magnetic field and the missing dipole

Many diagrams of the Earth's magnetic field have multiple field lines which are vertical at the Earth's poles, and horizontal at the equator, as is shown in Figure 1.1a. This is the type of field the Earth would have if it had a large bar magnet centered at its spin axis; this is known as a geocentric axial dipole or GAD. The present day field of the Earth is not produced by a GAD (some 10% of the structure demands more complicated field components), and the field has changed significantly over the last hundred years, with the magnetic North pole currently moving from Alaska towards Russia. Geomagnetists represent the Earth's magnetic field by a series of spherical harmonic terms (dipole, quadrupole, octupole, etc., see Figure 1.1b-d). The axial dipole term (represented by the symbol g_1^0) is the dominant term in the present field.

Reconstructions of the location of the Earth's tectonic plates in the past rely strongly on the assumption that Earth's magnetic field is a GAD when averaged over sufficiently long timescales. This assumption is known as the GAD hypothesis. Global compilations of paleomagnetic directions averaged over the past 10 Million years indicate that the Earth's average magnetic field is close to a GAD with a small amount of asymmetry in the axial dipole term (e.g. Cromwell et al., 2018). Conversely, global compilations of paleointensity data have a distribution with latitude which is incompatible with a predominantly dipolar field. In the Northern Hemisphere, time-averaged paleointensities in polar regions conform to a dipole field strength of 40-60 ZAm^2, whereas those at low latitudes conform to a dipole field strength of 60-80 ZAm^2 (Figure 1.2). Paleointensities in the Southern hemisphere are also weaker than those in the Northern hemisphere, although there are significantly fewer Southern hemisphere data. This hemispheric asymmetry in paleointensity results was previously reported by Cromwell et al., 2013 and is still found in recent data (e.g. Tauxe et al., 2022).

In Chapter 4, we discuss three potential hypotheses for the non dipolar distribution of paleointensity data with latitude. The hypotheses are as follows:

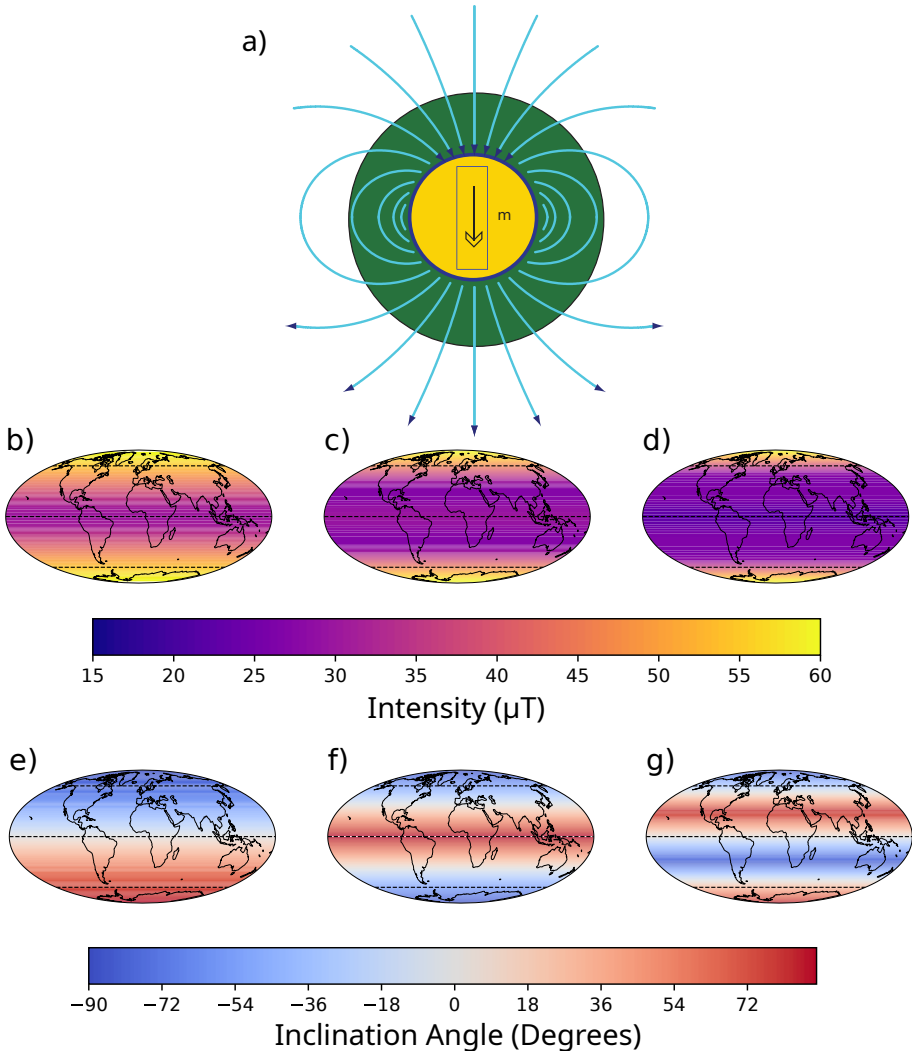

Figure 1.1. a) Lines of flux for a Geocentric Axial Dipole (GAD) field, which behaves like a bar magnet centered on Earth's spin axis (reproduced from Ben-Yosef et al., 2008). b), c), d) Latitudinal distribution of intensity (in μT) and e), f), g) inclination (the field angle to the horizontal) for b), e) a dipole field with a strength of 30 ZAm^2, c), f) a quadrupole field with a strength of 20 ZAm^2, d), g) an octupole field with a strength of 15 ZAm^2.

1. Global databases of paleointensity may contain inaccurate paleointensities which are analyzed using inconsistent methodologies.

2. The temporal sampling of paleointensities is not good enough to obtain a 5 Myr average.

3. The time averaged field is genuinely non-dipolar.

We test the validity of these hypotheses in Chapter 4.

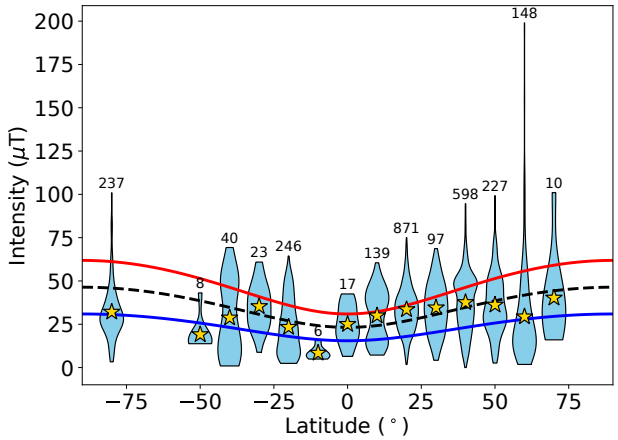

Figure 1.2. Distribution of paleointensities from the MagIC database (Tauxe et al., 2016 grouped into 10 degree latitude bins with means (yellow stars) over the past 5 Ma. Red, black and blue curves show the expected field for a GAD with a dipole moment of 80, 60 and 40 ZAm^2 respectively. Violins represent the density functions of the data with the number of data points in each violin shown above. Reproduced from Figure 4.1a.

1.3 Acquisition of a Thermal Remanent Magnetization

In order to understand why errors in paleointensity estimation may occur, it is important to understand how an igneous rock acquires a TRM. The theory of TRM acquisition was determined by French physicist Louis Néel in several papers in the 1940s and 50s. His 1949 paper (Néel, 1949) lays out the groundwork for the "uniaxial single domain theory" of TRM acquisition. In this theory, rocks, and other magnetic materials are composed of many small magnetic particles or "grains" which are each uniformly magnetized. Grains have an axis along which they are preferentially magnetized, known as the "easy" axis, which occurs because of anisotropies in the grain's shape, crystalline structure and internal stresses. For elongate grains, the easy axis is typically aligned with the elongation direction.

Under Néel's theory, a grain can be magnetized in one of two opposing directions along its easy axis. If supplied with enough energy, the magnetization in the grain can "flip" to the opposite direction. Néel defines the average amount of time for a grain to flip, also known as the relaxation time (τ), by the relation:

$$\tau = \frac{1}{C} \exp\left(\frac{Kv}{kT}\right), \tag{1.1}$$

where K is an anisotropy energy density and v is volume. Kv is an energy barrier which needs to be overcome to flip the grain's magnetization, k is the Boltzmann constant, T is the temperature and C is a constant frequency known as the "attempt frequency".

In an igneous rock, the magnetic minerals precipitate out of magma and lava and are not magnetized until they cool below a temperature known as the Curie temperature, at which they undergo a phase transition and become ferro/ferrimagnetic (see D. J. Dunlop and Özdemir, 1997 for details). On cooling below the Curie temperature, the magnetic minerals are initially exposed to large amounts of thermal energy (the kT term in Equation 1.1), which allows their magnetizations to flip rapidly. As they cool, individual particles will no longer havesufficient thermal energy to flip over the timescale of observation, becoming what is known as "blocked". Each grain therefore has it's own "Blocking Temperature" (T_B), below which τ becomes large enough for the magnetization to be preserved over geological timescales. When a grain becomes blocked, the minimum energy state is in one of two easy axis directions, with a greater proportion along the direction most closely aligned with the external magnetic field. A large number of randomly oriented grains in a rock will therefore have a net magnetization which is oriented parallel to the field in which they cooled. The strength of the magnetization is a hyperbolic tangent function of the strength of the field, but at Earth-like fields, the relationship between magnetization and field is normally linear, with the magnetization "saturating" at fields many orders of magnitude stronger than Earth-like fields.

Equation 1.1 explains the stability of the magnetic record in igneous rocks. Grains that are blocked at high temperature require large amounts of energy for their magnetizations to "flip" and so have relaxation times longer than the age of the universe at room temperature, leading to stable magnetizations being preserved over geological timescales. Pullaiah et al., 1975 constructed a series of curves which show that a grain of magnetite with a blocking temperature

of $\sim$250 °C has a relaxation time of a million years, and grains with blocking temperatures higher than 300 °C have blocking temperatures longer than 4.5 billion years.

1.4 The paleointensity experiment and sources of error in paleointensity estimation

The linear relationship between magnetization and ancient field is used to estimate a paleointensity. In theory, a paleointensity could be obtained by recording the NRM of an igneous rock, then remagnetizing the rock by giving it a TRM in a known lab field. The ratio of the TRM to the original NRM would be the ratio of the ancient field to the lab field. Multiple issues, which we will detail later in this section, mean that a simple ratio of these two magnetizations does not yield an accurate paleointensity. The Königsberger-Thellier-Thellier (KTT) family of experiments was devised (Königsberger, 1938, Thellier and Thellier, 1959) to obtain more accurate paleointensity estimates. KTT experiments involve a stepwise double heating, whereby a specimen is heated to a temperature and cooled in zero field to demagnetize it, and heated to the same temperature and cooled in a known lab field to acquire a partial TRM or pTRM. The order of the heating steps in a paleointensity experiment can matter, with the Thellier-Coe (Coe, 1967) protocol performing the zero-field heating before the in-field heating and the Thellier-Aitken (Aitken et al., 1988) protocol performing these treatments in the opposite order. The IZZI-Thellier protocol (Yu et al., 2004, Tauxe and Staudigel, 2004) alternates the order of the in-field and zero-field steps at each heating temperature, to check for effects which may be different between these two protocols.

In a KTT experiment, the magnetization remaining at each zero field step is plotted against the magnetization gained at each in-field step on what is known as an Arai plot (Nagata et al., 1963). In an idealized paleointensity experiment, the magnetization lost after heating to each temperature and cooling in zero field is proportional to the magnetization gained when heating and cooling in a known lab field at every temperature. This leads to the paleointensity

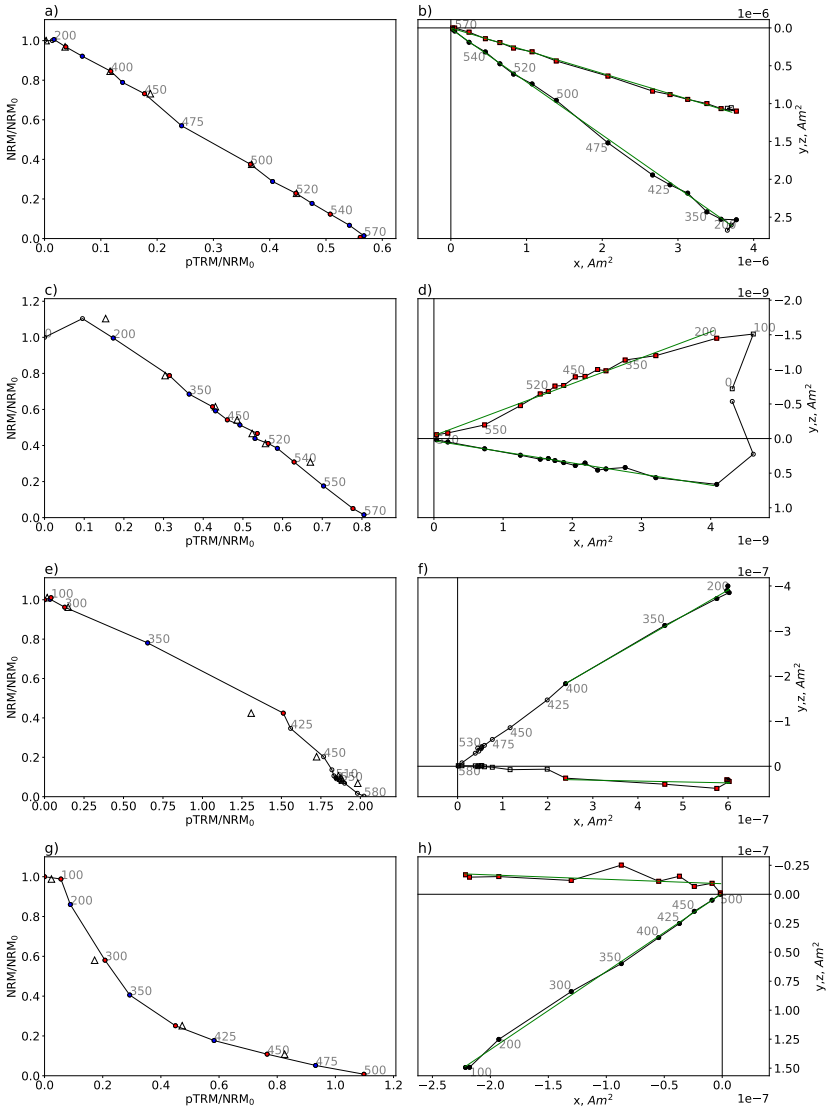

Figure 1.3. Left column, example Arai plots and right column, example Zijderveld plots for specimens from the study in Chapter 4. Zijderveld plots (Zijderveld, 1967a) show the direction of the magnetization vector after each zero-field heating with a horizontal projection (x-y plane) of the magnetization displayed as black dots, and a vertical projection of the magnetization (x-z plane) displayed as red squares. a-b) A specimen which conforms to our experimental assumptions, with a reasonably linear Arai (Nagata et al., 1963) and Zijderveld plot. c-d) A specimen which has acquired a VRM up to 200°C, leading to a change in direction on the Zijderveld plot, and a nonlinear Arai plot up to 200°C. e-f) A specimen which underwent thermal alteration at 400°C, leading to a change in the slope of the Arai plot above this temperature and an inconsistent repeat of the in-field step, (pTRM check: triangle at same NRM value as 400°C). g-h) A specimen containing predominantly non single-domain particles, leading to a curved Arai plot.

data plotting as a straight line on the Arai plot, the slope of which is proportional to the ratio of the ancient field to the lab field. An example of an idealized Arai plot is shown in Figure 1.3a. We identify three processes that can cause Arai plots to deviate from this idealized linear behavior, leading to errors in paleointensity estimation. We spend the remainder of this section outlining these processes.

1.4.1 Recording of multiple fields

Unfortunately, the NRM in a rock does not necessarily represent a magnetization acquired solely in an ancient field. A paleomagnetic specimen which formed millions of years ago may contain grains which have relaxation times of years to hundreds of years at room temperature. The magnetization of these grains will therefore 'relax' towards a more modern field direction, acquiring what is known as a Viscous Remanent Magnetization (VRM). Instead of recording a TRM from the time the rock cooled, the NRM instead records a sum of the original TRM and the newer VRM. Similarly, rocks can acquire a second, partial TRM (or pTRM) in a younger field by being partially reheated and cooled in that younger field. In igneous rocks, this commonly occurs in sequences of lava flows, where an older lava flow is reheated by a younger one flowing over the top of it. In both these cases, the slope of the Arai plot will only be constant at temperatures higher than the acquisition temperature of the pTRM (or the equivalent temperature for the VRM calculated using Equation 1.1). Paleomagnetists therefore look for a linear section of the Arai plot. An example of an Arai plot with a significant VRM component is shown in Figure 1.3c.

Another plot used frequently in paleointensity experiments is known as the Zijderveld plot (Zijderveld, 1967b). On this plot, the magnetization vector at each zero field step is plotted. The Zijderveld plot displays two projections of the magnetization vector on a 2d plot, one set of data shows the x-y plane of the data (e.g. black dots in Figure 1.3b), and the other shows the x-z plane (e.g. red squares in Figure 1.3b). For a specimen with a single TRM and no VRM, the Zijderveld plot will look like a straight line which trends towards the origin as the heating temperature increases and the specimen is demagnetized. The direction of this line will

be parallel to the field in which the specimen formed. An example of this behavior is shown in Figure 1.3b. A specimen containing a secondary TRM or VRM will have a Zijderveld plot which trends towards the origin only at high temperatures. At low temperatures, the direction on the Zijderveld plot will change, with the Zijderveld plot looking like two distinct lines. An example of this is given in Figure 1.3d.

1.4.2 Thermochemical alteration

The vast majority of particles carrying the magnetization in rocks are generally iron oxides. These minerals are particularly sensitive to oxidation-reduction reactions, and because the KTT family experiments involve heating rocks to high temperatures, thermochemical alteration of the magnetic minerals can occur during an experiment. If more magnetic minerals are created, or magnetic minerals undergo a reaction to form a non-magnetic mineral, then the magnetization of the specimen will be drastically changed. This is problematic for paleointensity experiments as the magnetization is no longer just a function of field and temperature, and so the original field can no longer be reproduced. To check for thermochemical alteration, paleomagnetists typically incorporate additional in-field heating steps which repeat previous heating steps made at lower temperature, known as pTRM checks. The magnetization gained during a pTRM check should be the same as the magnetization gained during the original in-field step. If thermochemical alteration occurred during any heating between the original in-field step and the pTRM check, then the pTRM check will differ from the original pTRM. An example of an Arai plot for a paleointensity experiment where the specimen underwent thermal alteration is shown in Figure 1.3e. Fortunately, any measurements made before thermochemical alteration has occurred can still record an accurate paleointensity. Paleomagnetists therefore look for a straight line on the Arai plot, and will not use temperature steps after a failed pTRM check.

1.4.3 Non single-domain particles

Up until this point, we have made the assumption that the particles that carry the magnetization in paleomagnetic samples obey Néel theory, i.e. that they behave like small bar magnets that can be uniformly magnetized in one of two directions (a grain of this type is described as being uniaxial single domain). Micromagnetic modeling (see e.g. Schabes and Bertram, 1988, Williams and Dunlop, 1989, Nagy et al., 2017) shows that uniaxial single domain theory only applies to a small size range of grains, which for magnetite is on the order of 10s of nanometers. The range of sizes, shapes and compositions of magnetic minerals in natural paleomagnetic samples is highly varied, and so Néel theory theory will not apply perfectly to many specimens. Grains which have multiple magnetization states may have inconsistent blocking and unblocking temperatures and so the magnetization of a specimen acquired in an ancient field may not be reproducible in a known lab field. D. Dunlop and Özdemir, 2001 performed paleointensity experiments on synthetically prepared powders of magnetite grains, and found that at larger grain sizes, the Arai plots exhibited a "sagging" behavior leading them to appear curved, see e.g. Figure 1.3g. They also found that a magnetization acquired by these specimens over a narrow range of temperatures was demagnetized over a much wider range of temperatures, indicating inconsistency of blocking and unblocking temperatures, an effect dubbed "pTRM tails". This curvature is particularly problematic, as paleomagnetists analyzing the plot may be tempted to select a range of temperatures over which the plot looks linear, assuming that the sagging effect is due to VRM or thermochemical alteration. However, neither of these slopes will yield an accurate paleointensity, leading to potential bias when a either the low temperature or high temperature slope is consistently used for paleointensity estimation (see e.g. Paterson, 2011).

1.4.4 Selection criteria and new approaches

Paleomagnetists have long used sets of statistics that describe non-ideal behavior in a paleointensity experiment, known as "Selection Criteria" to determine whether a specimen will

yield an accurate paleointensity estimate or not (Coe, 1967). A specimen is excluded from the analysis if it violates the threshold value. Since the original KTT experiments, progress has been made to identify problematic behaviors that cause erroneous paleointensity results, leading to the creation of new criteria. Levi, 1977 noted that large (so-called 'multi-domain' or MD) grains resulted in Arai plot curvature but suggested that the slope of the total TRM gave unbiased results. In the early 2000s, this behavior became more widely recognized, with studies such as D. Dunlop and Özdemir, 2001 and Krása et al., 2003. Krása et al., 2003 further showed that Arai plots exhibiting curvature yielded biased paleointensity estimates. To exclude specimens with curved Arai plots, researchers initially used criteria that constrain the fraction of the NRM on an Arai plot that can be used for a paleointensity experiment (e.g. Coe et al., 1978, Shaar and Tauxe, 2013). More recently, the curvature criterion $|\vec{k}|$ of Paterson, 2011 has been used to provide a mathematical description of curvature. A specimen with $|\vec{k}| = 0$ has a linear Arai plot, with larger $|\vec{k}|$ being indicative of curved Arai plots like the one shown in Figure 1.3g. Many older studies included in global paleomagnetic databases do not consider Arai plot curvature when conducting their analyses, and so may contain biased paleointensity results.

The threshold value for each selection criterion is often arbitrarily determined, and the sets of criteria used differ between laboratories and studies. Global compilations of paleointensity data over the last 5 million years are therefore analyzed using inconsistent methodologies. This is problematic as criteria which do not exclude many specimens may include erroneous results, and criteria which exclude too many specimens may yield too few results to get a good average. Paterson et al., 2014 compiled a list of over forty selection criteria currently in use. The authors then applied these criteria to a test data set of paleointensity results from historical lava flows and synthetic samples, for which the actual ancient field was approximately known. Using cross validation, they found a set of criteria which minimized both the number of accepted paleointensity results which were inaccurate, and the number of rejected paleointensity results which were accurate. However, the modified sets of criteria suggested by this study still include on average 19% of all inaccurate results while excluding 60% of accurate results. There is a clear

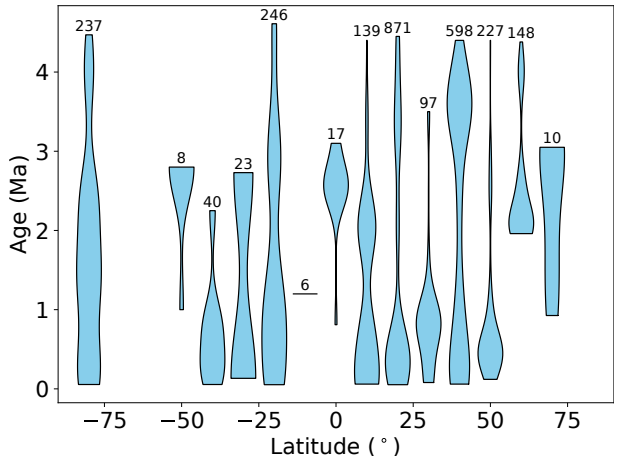

Figure 1.4. Age distributions (blue violins) of paleointensities in the MagIC database with ages between 0 and 5 Ma, binned for every 10° latitude. Reproduced from Figure 4.1b. Numbers represent number of intensity estimates in each latitude bin.

need for a more consistent approach for analyzing paleointensity behavior which does not rely on excluding data. Consistent analysis would allow for direct comparison between studies which would make it possible us to test our hypotheses about the distribution of the time-averaged field strength over the past 5 million years.

1.5 Temporal averaging of Earth's magnetic field

Paleointensity experiments generally represent snapshots of the Earth's magnetic field taken at the time a rock cooled. As the field's strength at a given location is constantly varying, many estimates are needed to build up an average at a particular time. Averages in Figure 1.2 were determined by simply taking the mean value of all paleointensities in that latitude bin for the last 5 million years. This would be a reasonable approach if the paleointensities in a particular bin were evenly distributed in time or if there was no variability in the average over time, but we are limited by the age distribution of rocks at a particular latitude, and the results that are published from different studies. The age distribution of paleointensities from MagIC in each latitude bin is shown in Figure 1.4. It is clear that some latitudes have age distributions that are significantly skewed toward more recent ages, whereas others have more uniform age

distributions. As with analysis of paleointensities, a more robust statistical approach is needed to average paleointensities from data sets which have differing age distributions.

1.6 Outline of this dissertation

The first part of this dissertation presents two methods which can be used to analyze paleointensity data in a more consistent way than standard selection criteria. In Chapter 2, we present the Bias Corrected Estimation of Paleointensity (BiCEP) method, which is a statistical method for dealing with paleointensity experiments where specimens contain non single-domain particles, leading to curvature on the Arai plot, as outlined in Section 1.4.3. BiCEP operates under the assumption that there is a relationship between curvature on the Arai plot (Paterson, 2011) and the estimated paleointensity of a specimen. By assuming that some specimens may have a biased paleointensity, BiCEP can obtain accurate paleointensity estimates without excluding specimens from the analysis owing to Arai plot curvature. Additionally, BiCEP propagates uncertainties regarding the curvature and intensity of the specimen up to the site level, leading to more statistically robust uncertainty estimates than traditional methods. Chapter 2 was published in Geochemistry, Geophysics, Geosystems in July 2021.

In Chapter 3 we present another method which we call Thermal Resolution of Unblocking Temperatures (TROUT). The TROUT method can be used to automatically find a range of temperatures suitable for use in a paleointensity experiment based on the Zijderveld plot, and so deals with the problems with paleointensity experiments with multiple recorded fields outlined in Section 1.4.1. TROUT can be used in paleointensity experiments where non single-domain grains cause the range of temperatures of two magnetizations to overlap, leading to a curved Zijderveld plot as well as potentially curved Arai plot.

The last part of this dissertation is an application of the BiCEP method described in Chapter 2. In Chapter 4 we describe how we collected samples from Hawai‘i, specifically targeting the type of lithologies we think are most likely to contain single-domain grains. We

performed a KTT-type experiment on specimens from these samples in the laboratory, and used BiCEP to analyze the results. We compared these results to studies targeting similar lithologies at other latitudes, reanalyzed these data using the same BiCEP method, and generated time averaged paleointensities with uncertainties by fitting curves to the paleointensity data, using smoothing methods to obtain the average values of these curves. In this way, we outline a methodology for obtaining comparable paleointensity estimates from different studies, involving field sampling strategies, laboratory protocols, and robust statistical techniques both for analyzing paleointensity data and obtaining a temporal average. Doing this allows us to test our hypotheses for the "missing dipole" from Section 1.2.

References

Aitken, M. J., Allsop, A. L., Bussell, G. D., & Winter, M. B. (1988). Determination of the intensity of the Earth's magnetic field during archaeological times: Reliability of the Thellier Technique. *Rev. Geophys.*, *26*(1), 3–12. https://doi.org/10.1029/RG026i001p00003

Ben-Yosef, E., Tauxe, L., Ron, H., Agnon, A., Avner, U., Najjar, M., & Levy, T. (2008). A new approach for geomagnetic archeointensity research: Insights on ancient matellurgy in the southern levant. *J. Archaelogical Science*, *35*, 2863–2879.

Coe, R. S. (1967). The determination of paleo-intensities of the earth's magnetic field with emphasis on mechanisms which could cause non-ideal behavior in thellier's method. *J. Geomag. Geoelectr.*, *19*, 157–178.

Coe, R. S., Grommé, S., & Mankinen, E. A. (1978). Geomagnetic paleointensities from radiocarbon-dated lava flows on hawaii and the question of the pacific nondipole low. *J. Geophys. Res.*, *83*, 1740–1756.

Cromwell, G., Johnson, C. L., Tauxe, L., Constable, C. G., & Jarboe, N. A. (2018). PSV10: A Global Data Set for 0–10 Ma Time-Averaged Field and Paleosecular Variation Studies. *Geochem. Geophys. Geosyst.*, *19*(5), 1533–1558. https://doi.org/10.1002/2017GC007318

Cromwell, G., Tauxe, L., Staudigel, H., Constable, C. G., Koppers, A. A. P., & Pedersen, R.-B. (2013). In search of long-term hemispheric asymmetry in the geomagnetic field: Results from high northern latitudes. *Geochem. Geophys. Geosyst.*, *14*(8), 3234–3249. https://doi.org/10.1002/ggge.20174

Dunlop, D. J., & Özdemir, Ö. (1997). *Rock magnetism: Fundamentals and frontiers*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612794

Dunlop, D., & Özdemir, Ö. (2001). Beyond Néel's theories: Thermal demagnetization of narrow-band partial thermoremanent magnetization. *Phys. Earth Planet. Int.*, *126*, 43–57.

Jackson, A., Jonkers, A. R. T., & Walker, M. R. (2000). Four centuries of geomagnetic secular variation from historical records. *Philos. Trans. R. Soc. London, Ser. A*, *358*(1768), 957–990. https://doi.org/10.1098/rsta.2000.0569

Königsberger, J. G. (1938). Natural residual magnetism of eruptive rocks. *Terrestrial Magnetism and Atmospheric Electricity*, *43*(3), 299–320. https://doi.org/10.1029/TE043i003p00299

Krása, D., Heunemann, C., Leonhardt, R., & Petersen, N. (2003). Experimental procedure to detect multidomain remanence during thellier–thellier experiments [Paleo, Rock and Environmental Magnetism 2002]. *Phys. Chem Earth (A/B/C)*, *28*(16), 681–687. https://doi.org/10.1016/S1474-7065(03)00122-0

Levi, S. (1977). The effect of magnetite particle size on paleointensity determinations of the geomagnetic field. *Phys, Earth. Planet. Int.*, *13*, 245–259.

Nagata, T., Arai, Y., & Momose, K. (1963). Secular variation of the geomagnetic total force during the last 5000 years. *J. Geophys. Res.*, *68*(18), 5277–5281. https://doi.org/0.1029/j.2156-2202.1963.tb00005.x

Nagy, L., Williams, W., Muxworthy, A. R., Fabian, K., Almeida, T. P., Conbhuı, P. Ó., & Shcherbakov, V. P. (2017). Stability of equidimensional pseudo–single-domain magnetite over billion-year timescales. *Proc. Natl. Acad. Sci. U.S.A.*, *114*(39), 10356–10360. https://doi.org/10.1073/pnas.1708344114

Néel, L. (1949). Théorie du traınage magnétique des ferromagnétiques en grains fins avec applications aux terres cuites. *Ann. géophys.*, *5*, 99–136.

Paterson, G. A. (2011). A simple test for the presence of multidomain behavior during paleointensity experiments. *J. Geophys. Res.*, *116*. https://doi.org/10.1029/2011JB008369

Paterson, G. A., Tauxe, L., Biggin, A. J., Shaar, R., & Jonestrask, L. C. (2014). On improving the selection of Thellier-type paleointensity data. *Geochem., Geophys., Geosyst.*, *15*(4), 1180–1192. https://doi.org/10.1002/2013GC005135

Pullaiah, G., Irving, E., Buchan, K. L., & Dunlop, D. J. (1975). Magnetization changes caused by burial and uplift. *Earth Planet. Sci. Lett.*, *28*(2), 133–143. https://doi.org/10.1016/0012-821X(75)90221-6

Schabes, M. E., & Bertram, H. N. (1988). Magnetization processes in ferromagnetic cubes. *J. Appl. Phys.*, *64*(3), 1347–1357. https://doi.org/10.1063/1.341858

Shaar, R., & Tauxe, L. (2013). Thellier_gui: An integrated tool for analyzing paleointensity data from thellier-type experiments. *Geochem. Geophys. Geosys.*, *14*, 677–692. https://doi.org/doi:10.1002/ggge.20062

Stern, D. P. (2003). A millennium of geomagnetism. *Reviews of Geophysics*, *41*(2), 1007.

Tauxe, L., Asefaw, H., Behar, N., Koppers, A. A. P., & Shaar, R. (2022). Paleointensity Estimates from the Pleistocene of Northern Israel: Implications for hemispheric asymmetry in the time-averaged field. *Geochem. Geophys. Geosyst.*, *n/a*(n/a), e2022GC010473. https://doi.org/10.1029/2022GC010473

Tauxe, L., Shaar, R., Jonestrask, L., Swanson-Hysell, N., Minnett, R., Koppers, A. A. P., Constable, C. G., Jarboe, N., Gaastra, K., & Fairchild, L. (2016). Pmagpy: Software package for paleomagnetic data analysis and a bridge to the magnetics information consortium (magic) database. *Geochem. Geophys. Geosys.*, *17*. https://doi.org/10.1002/2016GC006307

Tauxe, L., & Staudigel, H. (2004). Strength of the geomagnetic field in the Cretaceous Normal Superchron: New data from submarine basaltic glass of the Troodos Ophiolite. *Geochem. Geophys. Geosyst.*, *5*(2). https://doi.org/10.1029/2003GC000635

Thellier, E., & Thellier, O. (1959). Sur l'intensité du champ magnétique terrestre dans le passé historique et géologique. *Annales de Geophysique*, *15*, 285.

Williams, W., & Dunlop, D. J. (1989). Three-dimensional micromagnetic modelling of ferromagnetic domain structure. *Nature*, *337*, 634–637. https://doi.org/10.1038/337634a0

Yu, Y., Tauxe, L., & Genevey, A. (2004). Toward an optimal geomagnetic field intensity determination technique. *Geochem. Geophys. Geosyst.*, *5*(2). https://doi.org/10.1029/2003GC000630

Zijderveld, J. D. A. (1967a). A. C. demagnetization of rocks: Analysis of results. In D. W. Collinson, K. M. Creer, & S. K. Runcorn (Eds.), *Developments in Solid Earth Geophysics* (pp. 254–286). Elsevier. https://doi.org/10.1016/B978-1-4832-2894-5.50049-5

Zijderveld, J. D. A. (1967b). A. c. demagnetization of rocks: Analysis of results. In D. W. Collinson, K. M. Creer, & S. K. Runcorn (Eds.), *Developments in solid earth geophysics* (pp. 254–286). Elsevier. https://doi.org/10.1016/B978-1-4832-2894-5.50049-5

Chapter 2

Bias Corrected Estimation of Paleointensity (BiCEP): An improved methodology for obtaining paleointensity estimates

Brendan Cych[1], Matthias Morzfeld[1], Lisa Tauxe[1]

[1] Scripps Institution of Oceanography, University of California, San Diego

Chapter 2, in full is a reprint of material from Cych, B., Morzfeld, M., & Tauxe, L. (2021). Bias Corrected Estimation of Paleointensity (BiCEP): An improved methodology for obtaining paleointensity estimates." Geochemistry, Geophysics, Geosystems, 22, e2021GC009755. https://doi.org/10.1029/2021GC009755. The dissertation author was the primary investigator and author of this paper.

Abstract

The assumptions of paleointensity experiments are violated in many natural and archaeological materials, leading to Arai plots which do not appear linear and yield inaccurate paleointensity estimates, leading to bias in the result. Recently, paleomagnetists have adopted sets of "selection criteria" that exclude specimens with non linear Arai plots from the analysis, but there is little consensus in the paleomagnetic community on which set to use. In this paper, we present a statistical method we call Bias Corrected Estimation of Paleointensity (BiCEP), which assumes that the paleointensity recorded by each specimen is biased away from a true answer by an amount that is dependent a single metric of nonlinearity (the curvature parameter $\vec{k}$) on the Arai plot. We can use this empirical relationship to estimate the recorded paleointensity for a specimen where $\vec{k} = 0$, i.e., a perfectly straight line. We apply the BiCEP method to a collection of 30 sites for which the true value of the original field is well constrained. Our method returns accurate estimates of paleointensity, with similar levels of accuracy and precision to restrictive sets of paleointensity criteria, but accepting as many sites as permissive criteria. The BiCEP method has a significant advantage over using these selection criteria because it achieves these accurate results without excluding large numbers of specimens from the analysis. It yields accurate, albeit imprecise estimates from sites whose specimens all fail traditional criteria. BiCEP combines the accuracy of the strictest selection criteria with the low failure rates of the less reliable 'loose' criteria.

Plain Language Summary

Paleomagnetists perform experiments on rocks and pottery sherds (among other things) to estimate the strength of the ancient Earth's magnetic field (the paleointensity) through time. These make assumptions that are frequently violated, leading to bias. Quantitative metrics (selection criteria) attempt to screen out 'bad' data. If a particular experiment fails the criteria, the results are ignored. However, there is a lack of agreement as to which set of criteria are the most important and what is considered a failure. One of these criteria quantifies the deviation

from the fundamental assumption of linearity between the ancient and laboratory magnetizations. We present a new Bayesian method called Bias Corrected Estimation of Paleointensity (BiCEP), in which we assume that the estimated paleointensity depends on this deviation. We can then use this dependency to correct the paleointensity made on an ensemble of specimens with differing deviations from ideal behavior. BiCEP allows us to calculate accurate estimates of the ancient magnetic field, without ignoring results from non-ideal specimens. We test BiCEP on paleomagnetic data for which the original field strength is well constrained. BiCEP recovers the field strength with similar accuracy to stricter sets of criteria, but gets results for a greater number of sites.

2.1 Introduction

Estimates of the strength of the ancient Earth's magnetic field are currently made by performing experiments that compare the natural remanent magnetization (NRM) acquired by a specimen while cooling in the Earth's field, to a remanence known as thermal remanent magnetization (TRM) acquired by the specimen while cooling in a known laboratory field. Such experiments include the Königsberger-Thellier-Thellier (KTT) family of experiments (Königsberger, 1938; Thellier and Thellier, 1959), the Shaw family of experiments (Shaw, 1974), and the multi-specimen family of experiments (Hoffman et al., 1989), among others. All of these experimental families make assumptions about the relationship between the magnetic field and the remanent magnetization which may or may not be applicable (see the review by Tauxe and Yamazaki, 2015). In this paper, we will focus on the KTT family of experiments.

KTT type experiments involve a double heating protocol in which a specimen is heated two or more times to a series of temperatures up to the Curie Temperature. At each temperature, the specimen is cooled in two different fields. This has the effect of replacing the NRM with a TRM acquired in a known laboratory field. Data from KTT-type experiments are normally represented by the Arai diagram (Nagata et al., 1963), which plots the NRM magnetization

remaining at each temperature step against the magnetization imparted in the laboratory (often referred to as partial TRM or pTRM). The ratio of these two magnetizations, as represented by the slope of the best fitting line to the Arai plot data, is generally taken to be the ratio of the two magnetizing fields (ancient, B_{anc} and laboratory, B_{lab}).

KTT-type experiments rely on several assumptions which are frequently violated in paleointensity experiments. These include thermochemical alteration of specimens which may lead to the production of new magnetic minerals, and an assumption known as reciprocity, which requires that the blocking temperature (the temperature below which grains retain their magnetization after an external field is removed) is the same as the unblocking temperature (the temperature above which grains equilibrate with the external field).

The reciprocity assumption of Thellier and Thellier, 1959 is fundamental to Néel's theory for uniaxial single domain grains (Néel, 1949). Néel theory assumes that the electronic spins within magnetic grains are fully aligned, and that the alignment is in one of two directions along an energetically favorable 'easy' axis. In zero field, there is no preference for either direction, but in the presence of a field there is a slight preference for the direction along the easy axis with the smallest angle to the applied field. If the reciprocity assumption is met, then the energy required for the magnetization to change directions along the easy axis is always the same regardless of whether the specimen is cooled from higher temperature (blocking) or heated from room temperature (unblocking) and the two temperatures are identical.

By assuming that electronic spins within magnetic grains are fully aligned, Néel theory fails to take into account a term in the magnetic energy of grains which causes deviations from full alignment, resulting in structures such as the vortex state of, e.g., Williams and Dunlop, 1989. Although this effect is present in nearly all magnetic grains, it is insignificant over short length scales (10s of nm) and so uniaxial single domain theory may be a reasonable approximation for smaller, elongate grains. Specimens in paleointensity experiments contain mixtures of grains with different sizes and shapes and a specimen used for paleointensity is likely to include grains for which the applicability of single domain theory does not hold.

Failure of reciprocity and other fundamental assumptions embedded in the KTT family of experiments (laid out by e.g., Thellier and Thellier, 1959) provides a challenge for those analyzing paleointensity data. Paleomagnetists generally use a set of selection criteria which reject an intensity result if the NRM and pTRM data behave in a way which deviate from single domain theory (linear on the Arai plot, see Figure 2.1a) by more than some arbitrarily chosen threshold value. This is because data that contain a large proportion of non single domain-like grains or which otherwise violate the assumptions of the experiment are likely to give biased results (Tauxe et al., 2021). Selection criteria generally operate in a binary way, with specimens either being 'accepted' or 'rejected' from the estimation of the site mean, where 'site' is the collection of specimens assumed to have cooled in identical external magnetic fields (say, a lava flow or ceramic fragment).

Figure 2.1 gives a demonstration of biased results in specimens from prepared magnetite powders of increasing grain size that were magnetized in a 60 μT field (Krása et al., 2003). If all assumptions of Thellier and Thellier, 1959 were obeyed, we would expect the best fitting lines to data on Arai plots to give a range of values distributed closely about a mean of 60 μT. As the grain size of the powder increases, the Arai plot becomes more curved and the best fitting line to the Arai plot yields a progressively lower intensity estimate. As all the paleointensities estimated from the curved plots are below the expected value, the estimate for the ensemble can be biased, with the high temperature segment having an even lower mean value, and the low temperature segment having a high mean value. The data of Tauxe et al., 2021 also demonstrate downward curved Arai plots in natural samples are biased so this problem may effect many of the results compiled in paleointensity databases like the MagIC database (Tauxe et al., 2016) or PINT (Biggin, 2010).

The curvature of an Arai plot can be quantified using the curvature criterion ($\vec{k}$) of Paterson, 2011 (see also Paterson et al., 2014). Curvature is calculated using the reciprocal of the radius of a circle fit to scaled Arai plot data (see Section 2.2.2). While there is no theoretical basis for a circular fit (as opposed to the linear fit, which is firmly rooted in Néel theory), it is a

useful approximation that we will exploit in this paper.

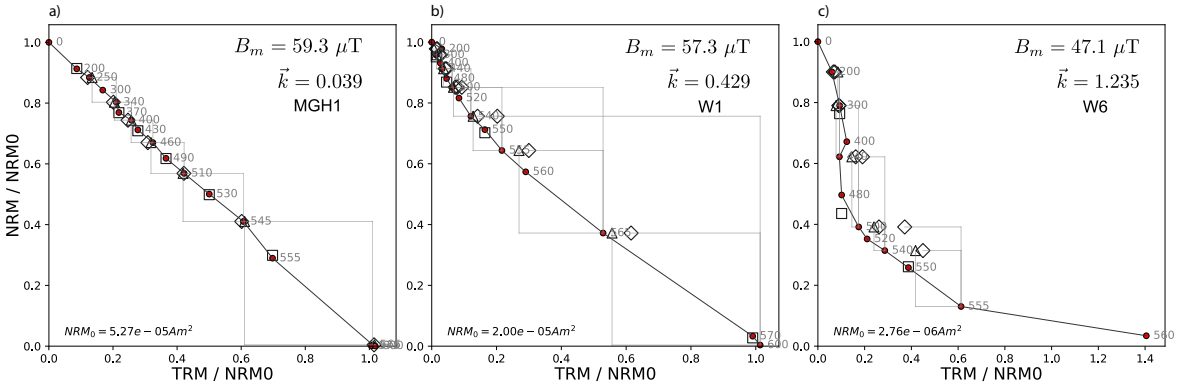

Figure 2.1. Arai plots from prepared magnetite powders given a TRM in a 60 μT field (Krása et al., 2003). The curvature criterion, $\vec{k}$ (Paterson, 2011) and specimen level paleointensity estimate B_m estimated from fitting a line to the entire Arai plot are plotted on the figure as text. The grain size of the magnetite powders increases from left to right. The coarser grains have non ideal domain state, leading to curved Arai plots and estimates of paleointensity which are biased to lower values than the expected 60 μT. a) Nominal grain size of 23 nm. b) Mean grain size of 70 nm. c) Mean grain size of 12.1 μm.

The practice of using binary (pass/fail) selection criteria is problematic for many reasons. Paleomagnetic specimens generally contain magnetic carriers which span a range of grain sizes and may or may not conform to the assumptions of the method. In addition, micromagnetic simulations (e.g., Nagy et al., 2017; Williams and Dunlop, 1989) demonstrate that the change in magnetic domain state with grain size is a continuum, and so one individual grain's behaviour may be more or less ideal than any other's. With binary pass/fail criteria, the distinction between 'good' and 'bad' data must be assessed with an arbitrary threshold value, which does not reflect the range of behaviors within both groups. Consequently there are a large number of selection criteria in common use (over 40 in Paterson et al., 2014), most of which have some empirical rationale, but there is little agreement on which set to use or their threshold values.

In this paper, we describe a new approach for paleointensity estimation that treats the quality of paleointensity data as a continuum as opposed to the binary 'in' or 'out' approach using selection criteria. We assume that paleointensities become more biased as specimens' magnetic behaviors become more non-ideal and their Arai plots become less linear. By allowing the data

interpretation for specimens to be based on the shape of their Arai plots, we are able to obtain unbiased estimates of paleointensity without the need for many specimen level (binary) selection criteria. We call this method the 'Bias Corrected Estimation of Paleointensity' or BiCEP. In the next section, we develop a Bayesian approach to obtain accurate paleointensity estimates with realistic uncertainties, using $\vec{k}$ as a metric of bias, and show how to combine data at the site level. In Section 2.3 we compare results from the BiCEP method to those of more traditional selection criteria based approaches. We discuss the results in Section 2.4 and summarize our conclusions in Section 2.5. Accompanying this paper, we release a Graphical User Interface (GUI) which can apply the BiCEP method to MagIC formatted data. Links and instructions on how to access the code can be found in Appendix 2.6.3.

2.2 Methods

2.2.1 Accounting for bias in paleointensity experiments

Paleomagnetists determine the paleointensity for a site by performing a Thellier-type double heating experiment on multiple specimens from that site. According to the theory for single domain grains (assuming no alteration of the specimen during heating), the ratio of NRM lost to pTRM gained is the ratio of the ancient field to the laboratory field. If the specimen conforms to theory, the Arai plot data will fall along a line the slope of which is equal to the ratio of ancient to the laboratory field (see Figure 2.1a).

We expect that the field strength predicted by the slope of the line on the Arai plot for each specimen (here called B_m) will be distributed about the true (expected) ancient field (B_{exp}) at the site with a Gaussian distribution. However, rarely do a set of specimens from a site all produce linear Arai plots that are easily interpretable. For example, interpretation of data from specimens with magnetic grains exhibiting non single domain magnetic domain states produce non-linear Arai plots which violate the assumptions of the method (e.g., Dunlop and Özdemir, 2001). Fitting lines to the data on such Arai plots often produces estimates of paleointensity which are biased

(see Figure 2.1c, Krása et al., 2003), which in turn would bias site level estimates.

Paleomagnetists generally approach non-ideal data by using certain quantitative criteria chosen to eliminate results suffering from one or more pathologies (Paterson et al., 2014). If a particular criterion calculated for a specimen fails to meet some threshold value, then the specimen is excluded from the analysis. In this paper, we present an alternative approach in which we allow for specimens to behave in a non-ideal (non-linear) fashion when considering how specimen intensity estimates are distributed about a site mean and weight the contribution of individual specimen estimates according to linearity. Under such a scheme, we start by predicting a bias for each specimen, and the specimens with the smallest predicted bias most strongly determine the paleointensity at that site. In this way, biased specimens do not strongly affect our site intensity estimate, as they are down-weighted, yet provide useful constraints on the uncertainty.

To predict the amount of bias a specimen is likely to have, we require a proxy for bias in paleointensity experiments. For this we use the curvature criterion $\vec{k}$ of Paterson, 2011 (see Section 2.2.2). There are several reasons that make this criterion a useful proxy for bias in paleointensity experiments:

- Specimens that are highly linear have, by definition, low values for $|\vec{k}|$ and will generally give unbiased paleointensity estimates (e.g., Cromwell et al., 2015).
- By contrast, specimens with higher $|\vec{k}|$ yield biased paleointensities, with the magnitude of the bias generally increasing with the magnitude of $|\vec{k}|$ (e.g., Tauxe et al., 2021).
- $|\vec{k}|$ has an empirical correlation with magnetic grain size (Paterson, 2011).

To predict bias, we can use a method by which we minimize the misfit to a model assuming that B_m is linearly related to $\vec{k}$ for all specimens. In other words, we say that:

$$B_m = B_{exp} + c\vec{k}_m + \varepsilon \tag{2.1}$$

Table 2.1. Table of sites used for analysis in this study, including original study locations, latitude, longitude and year of magnetization (where applicable), expected field at that location (B_{exp}), number of specimens used for analysis at that site M. Lat.: site latitude (°N). Long. site longitude (°E. N/A: Not Applicable (Synthetic). B_{exp} is either a known laboratory field, from the International Geomagnetic Reference Field (IGRF, Thébault et al., 2015 or in two cases (hw226,hw108) using the Arch3k.1 model of Korte et al., 2009

Site	Citation	Material	Lat.	Long.	Year	B_{exp}	M
1991-1992 Eruption Site	Bowles et al., 2006	lava flow	9.8	-104.3	1991	36.2	53
hw108	Cromwell et al., 2015	lava flow	19.9	-155.9	1859	39.3	23
hw123	Cromwell et al., 2015	lava flow	19.1	-155.7	1907	37.7	12
hw126	Cromwell et al., 2015	lava flow	19.7	-155.5	1935	36.4	13
hw128	Cromwell et al., 2015	lava flow	19.3	-155.9	1950	36.2	26
hw201	Cromwell et al., 2015	lava flow	19.4	-155.0	1990	35.2	12
hw226	Cromwell et al., 2015	lava flow	19.6	-155.5	1843	39.9	11
hw241	Cromwell et al., 2015	lava flow	19.5	-155.8	1960	36.0	18
BR06	Donadini et al., 2007	brick	60.1	24.9	1906	49.7	3
P	Muxworthy et al., 2011	lava flow	19.3	-102.1	1943	44.6	36
VM	Muxworthy et al., 2011	lava flow	40.8	14.5	1944	43.8	18
BBQ	Pick and Tauxe, 1993	submarine lava flow	9.8	-104.3	1990	36.2	12
rs25	Shaar et al., 2010	synthetic	N/A	N/A	N/A	30.0	5
rs26	Shaar et al., 2010	synthetic	N/A	N/A	N/A	60.0	5
rs27	Shaar et al., 2010	synthetic	N/A	N/A	N/A	90.0	10
remag-rs61	Shaar et al., 2011	synthetic	N/A	N/A	N/A	40.0	6
remag-rs62	Shaar et al., 2011	synthetic	N/A	N/A	N/A	60.0	6
remag-rs63	Shaar et al., 2011	synthetic	N/A	N/A	N/A	80.0	5
remag-rs78	Shaar et al., 2011	synthetic	N/A	N/A	N/A	20.0	4
kf	Tanaka et al., 2012	lava flow	65.7	-16.8	1984	52.0	3
Hawaii 1960 Flow	Yamamoto et al., 2003	lava flow	19.5	-155.8	1960	36.0	22
SW	Yamamoto and Hoshi, 2008	lava flow	31.6	-130.6	1946	46.4	19
TS	Yamamoto and Hoshi, 2008	lava flow	31.6	-130.6	1914	47.8	53
ET1	Biggin et al., 2007	basaltic lava	37.8	15.0	1950	43.3	3
ET2	Biggin et al., 2007	basaltic lava	37.8	15.0	1979	44.1	2
ET3	Biggin et al., 2007	basaltic lava	37.8	15.0	1983	44.2	4
Synthetic60	Krása et al., 2003	synthetic	N/A	N/A	N/A	60.0	7
LV	Paterson et al., 2010	Lithic Clasts	-23.4	67.7	1993	24.0	45
MSH	Paterson et al., 2010	Lithic Clasts	46.2	-122.2	1980	55.6	19
FreshTRM	Santos and Tauxe, 2019	remagnetized/synthetic	N/A	N/A	N/A	70.0	24

where m is an index reflecting the specimen number, ε is an error term and B_{exp} is the true value of B. Effectively, our model just becomes a linear fit between the specimen estimate B_m and $\vec{k}$, the y-intercept of which is the true value of the field B_{exp} and c is a slope constant. While there is no theoretical justification (yet) for why B_m would be related to $\vec{k}_m$, although it has been observed empirically (by Paterson, 2011 using the data in Figure 2.1, and more recently by Tauxe et al., 2021), a linear model is the simplest one to relate the two. We demonstrate in Section 2.3.3 that more complex models with a quadratic and cubic fit relating B_m to $\vec{k}_m$ perform worse than the linear model when predicting the paleointensity for sites for which the paleointensity is well constrained (historical lava flows or laboratory remanences).

Arai plot curvature is not the sole cause of bias in paleointensity experiments. In some cases, specimens with Arai plots which do not have high $|\vec{k}|$ but are still non linear (e.g., 'zig-zagged' as in, e.g., Yu et al., 2004), may still cause bias in paleointensity experiments. To counteract this, we use a Bayesian method of calculating $\vec{k}_m$ and B_m which provides an uncertainty for both of these parameters. The benefit of this approach is that specimens whose Arai plots are not well fit by a line or an elliptical arc have less influence on the linear fit. Therefore, the specimens with the lowest uncertainty in $\vec{k}$ are generally the most linear, and will have the most influence on the linear fit. Yet, for each specimen, there is a trade off between minimizing the circle fit at a specimen level and the linear fit between B_m and $\vec{k}$ for specimens from the same site, an issue we will deal with in Section 2.2.2.

Figure 2.2 shows results from our method (detailed in Section 2.2.2) applied to several sites for which the true value of B_{anc} (here, B_{exp}) is either calculated from the International Geomagnetic Reference Field (IGRF, Thébault et al., 2015) or Arch3k.1 (Korte et al., 2009) for historical flows, or known as the NRM is a laboratory TRM imparted to the specimens. Following Equation 2.1, the uncertainty in the intercept value of these linear fits gives us the uncertainty for our site value of B_{anc}. In this way, we can obtain an unbiased estimate of B_{anc} without relying on arbitrary binary (accept/reject) criteria to exclude specimen results.

In the following, we detail how the specimen level circle fit $\vec{k}$ and site level paleointensity

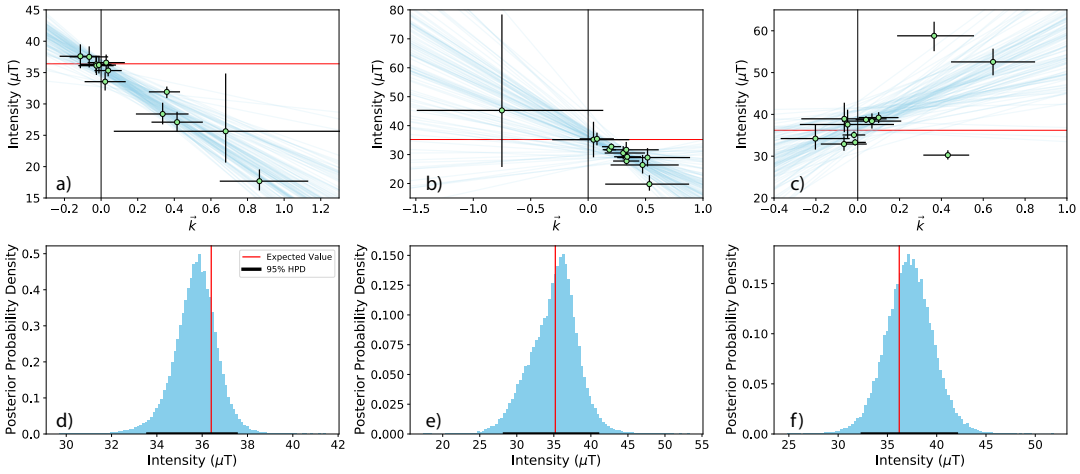

Figure 2.2. Example of results from the BiCEP method for several sites used as examples in this study. Lines (in blue) are fit to the values of B_m and $\vec{k}$ for each specimen (blue dots, with uncertainties as black lines). The values of linear fits at $\vec{k} = 0$ (blue histograms) provide an unbiased estimate of the expected paleointensity value at the site from the known field (red lines). a,d) hw126. b,e) hw201. c,f) BBQ. See Table 2.1 for sampling and citation details and Section 2.3 for comparison with the expected field values, B_{exp}.

for unknown values for B (here called B_{anc}) can be calculated. We then compare the efficacy of several different versions of our model to classical selection criteria. We do this using a data compilation from 30 sites updated from Paterson et al., 2014 and Tauxe et al., 2016 for which B_{exp} is well constrained (see Table 2.1 for details concerning the original publications of the data).

2.2.2 Statistical Methodology

Estimating curvature

Paterson, 2011 proposed a least squares fit of circles in Arai plot data. The parameter $\vec{k}$ of Paterson, 2011 is defined as the reciprocal of the radius of a best-fitting circle through the data. It is positive if the circle center is to the upper right of the Arai plot data (concave up, Figure 2.3a) and negative if the circle center is below and to the left of the Arai plot data (concave down, Figure 2.3b).

Before fitting to the Arai plot data, Paterson, 2011 scales the pTRMs by the maximum

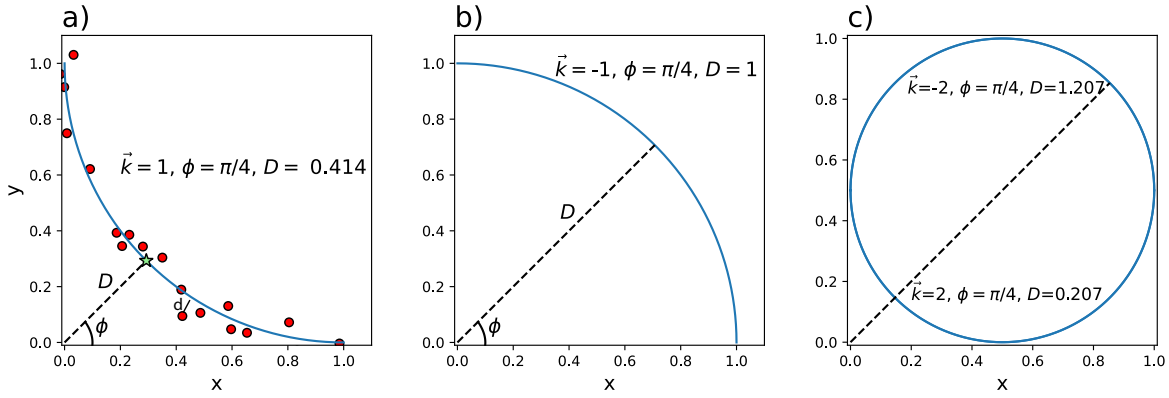

Figure 2.3. Example circles with different values for parameters $\vec{k}$ and D with the same ϕ, showing how these parameters define a circle. a) Positive $\vec{k}$. Red dots are example data, and the green star is the intersection of D, ϕ with the circle edge (see text for definitions). d is the distance of an individual data point from the best-fit curve (blue). b) Negative $\vec{k}$. Note that in this case, ϕ could take any value as the circle center is at the origin, making the definition of ϕ meaningless in this case. c) Example showing how two sets of the parameters $\vec{k}, \phi, D$ can describe the same circle.

pTRM to ensure that the paleointensity data are independent of the laboratory field. For estimating $\vec{k}$, we also subtract the minimum remaining NRM (NRM_{min}) for specimens for which full demagnetization has not been completed and we subtract the minimum pTRM (pTRM_{min}) for specimens for which the low temperature steps were excluded from the analysis (e.g., because of viscous remanent magnetization).

For the BiCEP method, we define two sets of data vectors x and y:

$$x_n = \frac{\text{pTRM}_n - \text{pTRM}_{min}}{\text{pTRM}_{max}}, \quad y_n = \frac{\text{NRM}_n - \text{NRM}_{min}}{\text{NRM}_0}, \tag{2.2}$$

where n is the index of the data point. Because scaling should be by the total (original) TRM (the NRM), we also exclude specimens whose NRM_{min} is more than 25% of the initial NRM. This is justified by the assumption that the experimenter did not carry out demagnetization to fully replace the NRM. Then, to fit a circle with center x_c, y_c and radius R to the data, we try to minimize the squared perpendicular distance d_n^2 (Figure 2.3a) of all the n data points to the circle edge:

$$\sum_{n=1}^{N} d_n^2 \quad \text{where} \quad d_n^2 = (\sqrt{(x_n - x_c)^2 + (y_n - y_c)^2} - R)^2. \tag{2.3}$$

In a total least squares fit, Equation 2.3 would be our objective function that we would minimize. To fit circles to the Arai plot using a Bayesian method, we use Bayes' formula (Equation 2.4). This formula allows us to assign a probability distribution to the values of different parameters (in this case, $\vec{k}_m$ and B_m), rather than just finding the 'best' value of the parameters. In a Bayesian context, we can simply assume that the data have some Gaussian noise distribution with some unknown standard deviation σ and apply Bayes' formula (e.g., Gelman et al., 2004):

$$P(\text{Parameters}|\text{Data}) = \frac{P(\text{Data}|\text{Parameters})P(\text{Parameters})}{P(\text{Data})}, \tag{2.4}$$

where the left hand side is the probability of the parameters given the data and the right hand side is the probability of the data given the parameters times the probability of the parameters, normalized by the probability of the data. In our case, the parameters are x_c, y_c, R and σ and our data are x and y so we rewrite Equation 2.4 as:

$$P(x_c, y_c, R, \sigma|x, y) = \frac{P(x, y|x_c, y_c, R, \sigma)P(x_c, y_c, R, \sigma)}{P(x, y)}. \tag{2.5}$$

The term $P(x, y|x_c, y_c, R, \sigma)$ is known as the "likelihood" and is based on the probability of generating the observed data from a given set of parameters using the assumed Gaussian distribution. The term $P(x_c, y_c, R, \sigma)$ is known as the "prior" and is a probability distribution for values of x_c, y_c, R and σ we consider to be reasonable before we see any data. We consider the priors on these parameters to be independent of one another, so we could rewrite this as $P(x_c)P(y_c)P(R)P(\sigma)$. The term $P(x, y)$ is known as the "evidence", and is simply a normalizing constant that makes the "posterior" probability distribution, $P(x_c, y_c, R, \sigma|x, y)$, integrate to 1. In our application, we can simplify the relationship by ignoring the normalization. Furthermore, we

can say from the definition of the Gaussian distribution that:

$$P(x,y|x_c,y_c,R,\sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left(\sum_{n=1}^{N} -\frac{d_n^2}{\sigma^2}\right). \tag{2.6}$$

Now we have an expression for our posterior probability distribution:

$$P(x_c,y_c,R,\sigma|x,y) \propto \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left(\sum_{n=1}^{N} -\frac{d_n^2}{\sigma^2}\right) P(x_c,y_c,R)P(\sigma). \tag{2.7}$$

Because the actual noise distribution of the Arai plot data is quite complicated (Paterson et al., 2012), we do not know the value of σ, so we use the uninformative prior $P(\sigma) \propto \frac{1}{\sigma}$; in other words, the smaller σ, the more likely the result. We can then substitute this prior into Equation 2.7 and integrate out σ to obtain:

$$P(x_c,y_c,R|x,y) \propto \left(\sum_{n=1}^{N} d_n^2\right)^{-N/2} P(R,x_c,y_c) \tag{2.8}$$

where N is the total number of measurements considered.

The set of parameters x_c, y_c and R is not easy to solve for, because Equation 2.3 has multiple local minima (see Chernov and Lesort, 2005 for a more detailed discussion). Consider the simple case of a specimen with a linear Arai plot; in even this simplest case, there are four minima, as both R and x_c, y_c will be either positive or negative and very large. To avoid this complexity, we can use instead a change of parameters similar to that of Chernov and Lesort, 2005 which Paterson, 2011 used as a basis for the circle fitting protocol. Based on this, we define a set of three new parameters which avoid the problem of multiple minima.

Firstly, we require a point on the Arai plot which can be related to a unimodal distribution. We know that linear data will plot along the edge of a circle (the tangent), so if we draw a line from the origin toward the center (x_c, y_c) (not shown), this will touch the edge of the circle at some distance D (green star in Figure 2.3a). The angle to the horizontal of this line we call ϕ and we can directly estimate the $\vec{k}$ parameter of Paterson, 2011 using Equations 2.9,2.10,2.11.

We can then establish equations for transforming between these two sets of parameters (see Appendix 2.6.1 for a more detailed derivation):

$$x_c = \left(D + \frac{1}{\vec{k}}\right)\cos(\phi), \tag{2.9}$$

$$y_c = \left(D + \frac{1}{\vec{k}}\right)\sin(\phi), \tag{2.10}$$

$$R = \frac{1}{|\vec{k}|}. \tag{2.11}$$

Despite this transformation, the circle fitting equation can still have multiple minima, even with $\vec{k}, D, \phi$ as our parameters, as the line connecting the origin to the horizontal touches the circle edge in two locations (see Figure 2.3c). However, we can use prior distributions to avoid this.

Chernov and Lesort, 2005 define a function of the data d_{max} to define the region of possible values for $\vec{k}$:

$$d_{max} = \max_{i,j}\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{2.12}$$

Additionally, we define distance from the origin to the centroid of the data, d_{cent}:

$$d_{cent} = \sqrt{\bar{x}^2 + \bar{y}^2} \tag{2.13}$$

Using this function, we can assume that $D < 2d_{cent}$ and $|\vec{k}| < N/d_{max}$ and can define priors for our parameters:

$$P(D) \sim \text{Uniform}(0, 2d_{cent}), \tag{2.14}$$

$$P(\phi) \sim \text{Uniform}(0, \pi), \tag{2.15}$$

and

$$P(\vec{k}) \sim \text{Uniform}(-N/d_{max}, N/d_{max}). \tag{2.16}$$

Using these priors gives us a posterior with a single maximum in most cases, which makes the problem much easier to solve computationally.

We can now apply a Bayesian approach to estimate $\vec{k}$ for all temperature steps for a given specimen m. It is frequently useful to choose a subset of the temperature steps (e.g., if there is evidence for multiple components of the NRM or heating related alteration, as detected by repeated lower temperature pTRM steps). When using a subset of steps, we scale by the maximum pTRM for all temperature steps and the NRM at room temperature; in this way we can predict the curvature for the part of the Arai plot that is missing. This means that interpretations based on a small fraction of the Arai plot will have large uncertainties in the value of $\vec{k}$. Therefore, our circle fit can prioritize interpretations using the largest fraction of the NRM.

Figures 2.4a,c,e show circle fits sampled from the posterior distributions for specimens from site hw126 (site level results shown in Figure 2.2a). The probability densities of all the $\vec{k}$ values for each specimen are plotted in Figures 2.4b,d,f. The plot demonstrates how a straight Arai plot (Figure 2.4a) produces a narrow posterior about $\vec{k} = 0$ (Figure 2.4b), while a curved one (Figure 2.4c) produces a posterior which does not contain $\vec{k} = 0$ (Figure 2.4d). In the example with failed pTRM checks at higher temperatures (offset triangles in Figure 2.4e), we exclude the data points represented by open circles and use a linear segment with only a portion of the results, the posterior distribution of $\vec{k}$ has a larger uncertainty on the value, translating to a larger uncertainty in the bias for that specimen. We do not advocate for any particular method of checking for alteration, and do not exclude any measurement steps in our results section. However, our circle fitting routine allows for measurement steps to be excluded and accounts for the increased uncertainty in doing so.

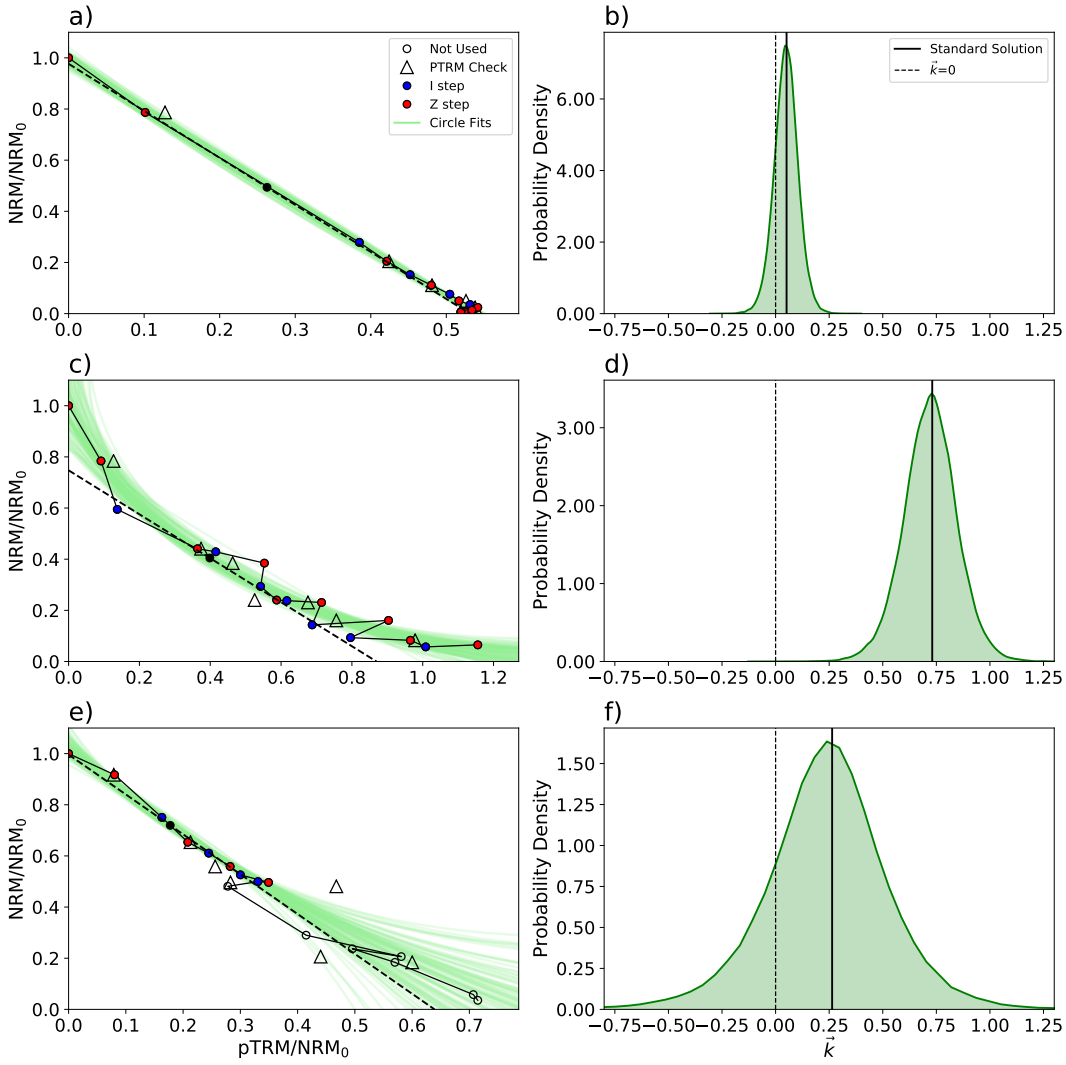

Figure 2.4. Examples of circle fits to Arai plots (left column) and approximate probability densities of $\vec{k}$ (right column). Dashed lines in the left hand plots are the tangents to circles with the median values for ϕ and D. We use tangents to the circle to get an estimate for B_m as outlined in Section 2.2.2. Triangles in a), c), e) are repeated lower temperature steps (pTRM checks) that indicate alteration of magnetic minerals during the experiment when offset from the original measurements (red dots). a) Specimen hw126a1. A fit to a straight line yields a precise $\vec{k}$ distribution with a maximum close to zero (b). c) Specimen hw126a7. A curved Arai plot with a high amount of scatter/zigzagging (left) results in a higher uncertainty in the value of $\vec{k}$ (d). e) Specimen hw126a6. Arai plot for a specimen that underwent thermochemical alteration at high temperature. A circle fit to just the low temperature steps results in a high uncertainty in the value of $\vec{k}$ (f). Note that we do not exclude any measurements due to thermal alteration in our results section, and that this is only done here for illustrative purposes.

Obtaining a specimen level paleointensity estimate

Analogous to the case in which paleointensity estimates are made using the slope of a fitted line to the Arai plot data, we can obtain a similar "slope" value for a circular arc fit to the data. Consider the case in which the edge of the circle forms an exact line ($\vec{k}$=0, see Figure 2.4a). In this case, the slope of the line can be given by the tangent to the circle at the point where it intersects a line drawn from the origin (0,0) to the circle center (Figure 2.3a). In other words, the "slope" of the Arai plot can be estimated as $\cot\phi$, which gives the tangent to the circle. We can then turn this into an intensity estimate B_m using the formula:

$$B_m = \frac{B_{lab}\cot(\phi)}{\mathrm{pTRM}_{max}}, \tag{2.17}$$

where B_{lab} is the laboratory field used to impart a pTRM to the specimen. If a specimen is corrected for anisotropy, cooling rate, or non-linear acquisition of TRM, we apply this correction to Equation 2.17.

We now have a way of obtaining estimates for B_m and $\vec{k}_m$ for each specimen. We use the methodology laid out in Sections 2.2.2 and 2.2.2 to plot the median value of the posterior for these parameters (with error bars) in Figure 2.5a, and examples of circle fits in Figures 2.5c, e, g. For specimens with values of $\vec{k}$ that are approximately 0 (Figure 2.5g), the B_m values are quite accurate. There appears to be a bias for specimens with large $\vec{k}$, with the amount of bias increasing as $\vec{k}$ increases. In this example, large positive values of $\vec{k}$ lead to a large underestimates of B_m while negative values of $\vec{k}$ lead to overestimates of B_m (although small in this example).

Obtaining a site level paleointensity estimate

The main problem with the method presented thus far is that we still do not have a way of obtaining an estimate for B_{anc}, the unknown value at the site level. However, in Figure 2.5a there appears to be a dependence between $\vec{k}_m$ and B_m as suggested earlier, with most of the specimens showing a quasi-linear relationship (the only exception being the point labeled e) whose Arai

plot is shown in Figure 2.5e) and suggests there is a great deal of uncertainty in the value of $\vec{k}$ itself. Because of this, we can modify our model slightly by imposing the extra restriction that B_m must be linearly dependent on $\vec{k}_m$ (with noise) using Equation 2.1 (substituting B_{anc} for the unknown value of B_{exp}).

Previous papers have assumed that B_{anc} for selected specimens follows a Gaussian distribution and we can also make this assumption here. In the following, we will show how this modification can shift results from specimens that are offset from the linear relationship toward the line (as in the point labeled 'f' in Figure 2.5b) and produce models (shown as blue lines) that estimate all of our B_m. We can then use the resulting models to estimate the probability distribution for B_m as:

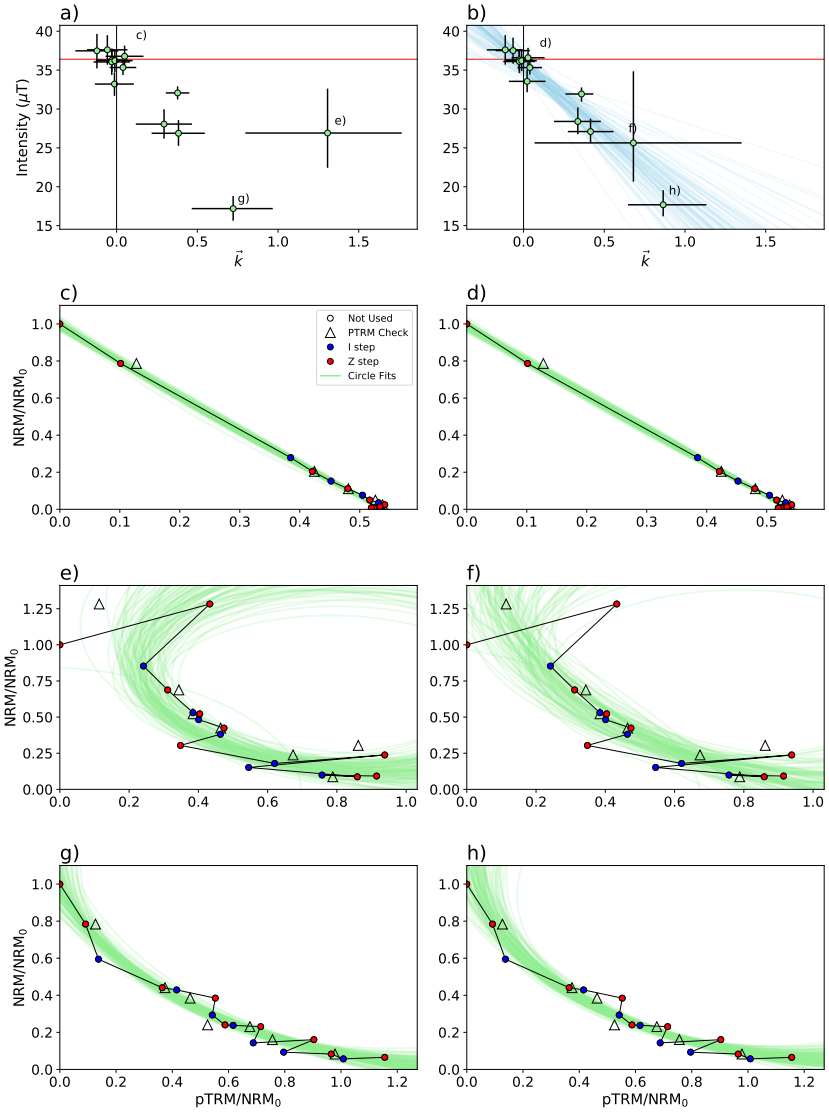

Figure 2.5. Examples demonstrating how the predicted $\vec{k}$ and B_m for each specimen are modified for a site by using a hierarchical model (Equation 2.14). The left column shows draws from the posterior for an "unpooled" model where we estimate B_m and $\vec{k}_m$ independently. The right column shows draws from the posterior for the BiCEP method where we assume a linear relationship between B_m and $\vec{k}_m$. a) Red horizontal line is B_{exp} (hw126, see Table 2.1). 95% credible intervals for $\vec{k}_m$ and B_m are plotted using black error bars, with the medians as green points. b) Representative draws from the posterior distribution are plotted as blue lines assuming that the individual specimen values B_m follow the relationship stated in Equation 2.14. Note that the higher curvature specimens with large uncertainty in $\vec{k}$ follow a linear trend away from B_{exp}. c),e),g): [Symbols same as in Figure 2.4.] Arai plots of particular specimens are shown with circle fits sampled from the posterior of the unpooled model shown in a) and plotted in green. In d), f), h), same specimens as in c), e), g) but using the posterior of the BiCEP model in b). Note that there is little change in the specimen in d) for which a close fit to the data is possible, but in f) and h) the curvature (and intensity) of the specimen are modified to fit the line better.

$$P(B_m|k_m, B_{anc}, \sigma_{site}, c) = \frac{1}{\sqrt{2\pi\sigma_{site}^2}} \exp\left(- \frac{(B_{anc} + c\vec{k}_m - B_m)^2}{2\sigma_{site}^2} \right). \quad (2.18)$$

Now we can combine our expressions for B_m and $\vec{k}_m$ (Equations 2.17, Sections 2.2.2 and 2.2.2) with the new constraint of a linear relationship between B_m and $\vec{k}_m$ (Equation 2.18). This allows us to obtain an expression for the site level intensity estimate B_{anc}:

$$P(B_{anc}, \sigma_{site}, c, B_m, k_m, D_m|x_m, y_m) \propto$$

$$P(x_m, y_m|k_m, D_m, B_m)P(B_m|k_m, B_{anc}, \sigma_{site}, c)P(B_{anc}, \sigma_{site}, c)P(D_m, k_m). \quad (2.19)$$

Equation 2.19 may look complicated, but we defined each of the terms already. The benefit of this treatment is that we can obtain $P(x_m, y_m|k_m, D_m, B_m)$ from our circle fitting in Equation 2.8 (see also Appendix 2.6.1). We defined $P(B_m|k_m, B_{anc}, \sigma_{site}, c)$ in Equation 2.18. The values of $\vec{k}$ and B_m for each specimen are needed to fit both of these terms. This means that specimens with large scatter in their Arai plots (those which have Arai plots that are not fit well by a line or a circle) are more strongly affected by the site level fit B_{anc}, and therefore by the specimens with more linear (or circular) Arai plots. Conversely, those specimens with a small uncertainty in $\vec{k}$ or B_m are tightly constrained by the Arai plot fit and so have more control over the fit at the site level.

The other two terms on the right side of Equation 2.19 ($P(B_{anc}, \sigma_{site}, c)P(D_m, k_m)$), are priors. $P(D_m, k_m)$ were defined in Equations 2.14 and 2.16 respectively. Now, we need to define priors for $P(B_{anc}, \sigma_{site}, c)$. For this purpose, we use a poorly constrained prior for the slope, c, where $P(c) \propto 1$. Although this is not a probability distribution, the resulting posterior distribution for B_{anc} is always a real probability distribution if the number of specimens is greater than one. We use a uniform prior between 0 and 250 μT for $P(B_{anc})$ as intensity values can never be negative and in databases such as the MagIC database (Tauxe et al., 2016) or the PINT database of Biggin, 2010 rarely (if ever) exceed 250 μT. For $P(\sigma_{site})$ we use a normal distribution with

zero mean and standard deviation of 5 μT, truncated to always be positive.

Figure 2.5b shows our median estimates for B_m and $\vec{k}_m$ after applying the linear restriction. Here, there is a tradeoff between fitting the Arai plot data with the circle, and fitting the linear trend at a site level. The effect of the linear fitting is apparent when compared to estimating $\vec{k}_m$ and B_m for each specimen in isolation, which is shown in Figure 2.5a. With the linear restriction, the $\vec{k}$ and B_m of specimens are "pulled" closer to a linear trend by modifying the Arai plot fits; specimens with more uncertain $\vec{k}_m$ are more strongly affected (e.g., specimen labeled e) and f) in Figure 2.5a and b). The specimens with highly linear Arai plots (for which we have small uncertainty in $\vec{k}_m$), the circle fits (see g and h) are mostly unchanged. Despite this modification of the circle fits to the Arai plots by the linear model, the circle fits to those specimens do not look unreasonable.

2.2.3 Metrics of success

In order to 'ground-truth' the method, we rely on a compilation of paleointensity data updated from that of Paterson et al., 2014 and Tauxe et al., 2016. This compilation has data from 30 sites for which B_{anc} is well constrained (hence we use B_{exp}), either through the IGRF, or because the specimens were given TRMs in a known lab field before the Thellier experiment. One exception to this is for hw226 and hw108, lava flows erupted in Hawaii in 1843 and 1859, prior to the range included in the IGRF. For these sites, we used the Arch3k.1 model of Korte et al., 2009. A list of sites used here is given in Table 2.1. Instead of choosing a range of temperatures for each site, we simply use every temperature on the Arai plot for all specimens.

Because we have to estimate multiple parameters for each specimen, our method involves a high dimensional optimization problem. Therefore, we generate the estimates for B_{anc} for a given site using a Markov chain Monte Carlo (MCMC) method which approximates the posterior distribution by generating pseudosamples from it (see Appendix 2.6.2 for details). MCMC techniques are frequently used to solve high dimensional problems of this kind.

For each site, we quantify the effectiveness of the BiCEP method using several metrics,

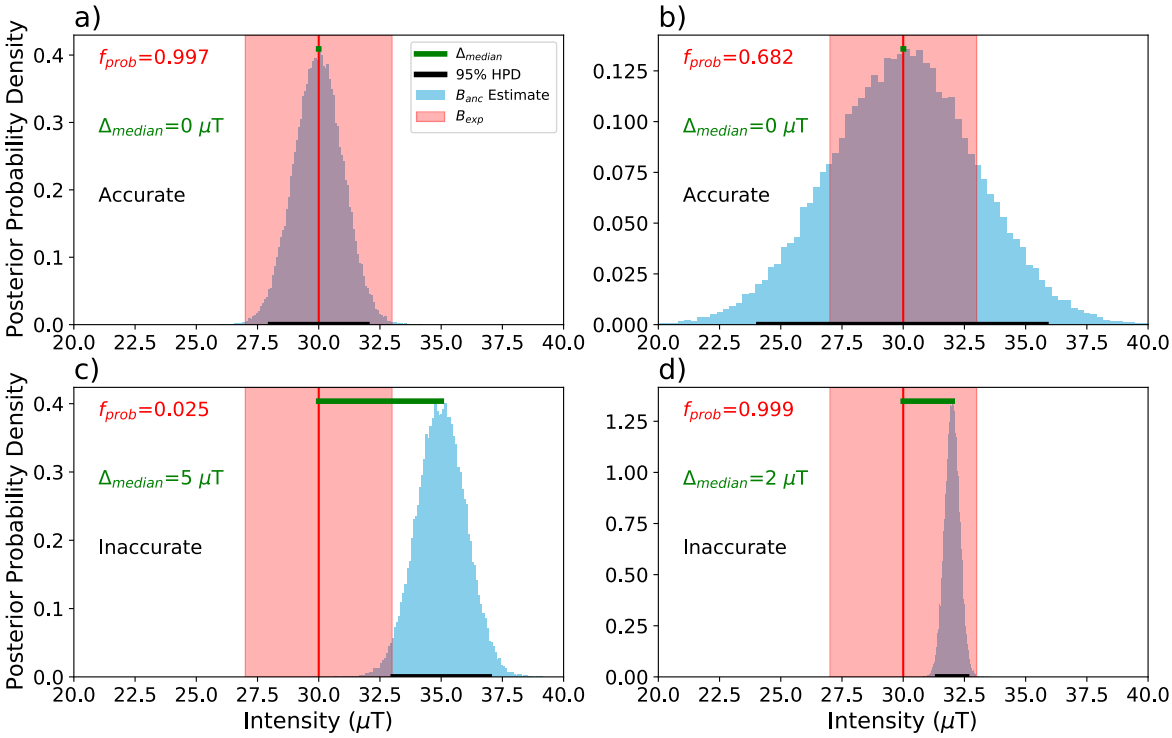

Figure 2.6. Examples of accuracy and precision metrics used in this study with simulated Gaussian distributions of B_{anc} for illustration. a) An accurate and precise estimate, b) An accurate but imprecise estimate, c) An inaccurate and imprecise estimate. d) A slightly inaccurate and highly precise estimate. Accuracy check used for n_{acc} checks whether the black line intersects the expected value (B_{exp}). f_{prob} is the area of the blue histogram that lies within the red shaded area. Δ_{median} is the length of the green line.

f_{prob}, Δ_{median} (see Figure 2.6 for graphical representation), $\bar{f}_{prob}$, and n_{acc}:

1. f_{prob}: We report the median value of our posterior distribution and the 2.5^{th} and 97.5^{th} percentile of the Monte Carlo sample (95% credible interval) as error bars. To quantify the effectiveness of our method, we look at the proportion of the posterior distribution that lies within 3 μT of the expected value of B (B_{exp}) and call this proportion f_{prob}.

2. $\bar{f}_{prob}$: the mean value of f_{prob} over all sites included in the study. A value of 1 is the best possible value and means all our results are accurate and precise to better than 3 μT.

3. Δ_{median}: the difference (in μT) between the median value of the MCMC sample (see Section 2.6.2 for explanation) and B_{exp}. The median value of Δ_{median} is $\tilde{\Delta}_{median}$. Values of

$\tilde{\Delta}_{median}$ close to zero are best.

4. n_{acc}: the number of sites for which B_{exp} lies within our 95% credible interval. A related parameter, f_{acc} is the fraction of results that are accurate (n_{acc}/n_{sites}), where n_{sites} is the total number of sites analyzed. We expect this number to be 0.95 in ideal circumstances.

We use these metrics to compare the BiCEP results to those obtained by several different sets of selection criteria: CCRIT (Cromwell et al., 2015), Paterson's modified PICRIT03 (here called PICRITMOD) and SELCRIT Criteria (here called SELCRITMOD, Paterson et al., 2014). For this exercise, we also calculated these two criteria with the addition of the curvature criterion of $|\vec{k}| < 0.270$, which we refer to as PICRITMODk and SELCRITMODk. We apply these criteria using the standard deviation optimization method in Thellier GUI. Most sets of commonly used selection criteria rely on an assumption of a Gaussian probability distribution for the site level estimate B_{anc}, which allows us to calculate these same metrics.

For our analyses of our success metrics, we exclude sites that contain fewer than three specimens. For fair comparison, we do not exclude sites from our analyses with traditional selection criteria which have high standard deviation, as we do not do this for BiCEP. If a site fails to produce an estimate of B_{anc} for any reason (for example, selection criteria passed less than two specimens), we assume the prior distribution of a uniform distribution between 0 and 250 μT. This allows us to compare methods directly, with a penalty applied for excluding sites. An excluded site will have f_{prob}=0.012, whereas a site with a highly inaccurate result can have f_{prob} of 0, so exclusion is considered only slightly better than an inaccurate result in this scheme. We discuss the results of this comparison in Section 2.3.1.

2.2.4 Width of prior and order of fit

Here we consider several alternative contingent models in order to explore our choices for $P(\sigma_{site})$ and assumptions about the relationship of B_m and $\vec{k}$. In addition to using a standard deviation of 5 μT for $P(\sigma_{site})$, we use standard deviations of 10 μT and 20 μT. The effect of this

is hard to conceptualize, but wider priors will prioritize fitting circles to the individual specimens over fitting the linear relationship between B_m and $\vec{k}_m$ at a site level. The practical effect of this is wider posteriors for sites where the number of specimens is small.

So far, we have assumed *a priori* that B_m is linearly dependent on $\vec{k}_m$. Because there is no theoretical reason why this should be the case, we test models for which the relationship between B_m and $\vec{k}_m$ is described by a quadratic polynomial and a cubic polynomial. We would expect a higher order model to more closely fit the individual $\vec{k}_m$ and B_m values, but with a loss of precision due to the more complicated model.

Results for our method, as well as for two sets of selection criteria, are given in Table 2.2. For each model, we calculate $\bar{f}_{prob}$, $\tilde{\Delta}_{median}$ and f_{acc} for comparison. In this table, our models are named for the value of the standard deviation of $P(\sigma_{site})$ as well as the order of the fit. Our preferred model is referred to as "Linear 5 μT", and this is the model used in this paper where otherwise unspecified.

2.2.5 MCMC sampler diagnostics

MCMC samplers are only ever an approximation of the posterior distribution, and the number of Monte Carlo samples needed to make an accurate approximation is not the same for every site, or every run of the sampler. To determine whether we are accurately sampling the posterior distribution, we look at three diagnostics which are also described in Appendix 2.6.2:

1. $\hat{R}$: (Gelman and Rubin, 1992) quantifies convergence between chains in the MCMC method. This parameter is required to be between 1.1 and 0.9 for the sampler to converge.

2. n_{eff}: the effective MCMC sample size. We are using 30,000 Monte Carlo samples and n_{eff} should be large (> 1000) to have a good representation of our parameters.

3. f_{div}: the proportion of divergent transitions f_{div} in the MCMC sample. This should ideally be zero, but it does not appear to cause large problems for the estimate of B_{anc} if it is non zero (see Section 2.6.2).

The diagnostics n_{eff} and $\hat{R}$ are produced for each of our parameters (each of our $B_m, \vec{k}_m, D_m$ and B_{anc}, σ_{site}). When reporting these values, we look at the worst value of $\hat{R}$ (furthest from unity) and the value of n_{eff} for B_{anc}. If $\hat{R}$ >1.1, we replace the distribution on B_{anc} with a uniform distribution between 0 and 250 μT (the prior). The results of the MCMC sampler are presented in Section 2.3.4.

2.3 Results

2.3.1 Comparison of BiCEP to Selection Criteria

In this section, we compare the BiCEP to several sets of selection criteria (see Section 2.2.3). The full set of results for all sites can be seen in Figure 2.7, and are summarized in Supplementary Data Set S1.

Table 2.2. Results comparing the models used in this study to results using CCRIT (Cromwell et al., 2015) as well as PICRITMOD and SELCRITMOD (Paterson et al., 2014), both with and without the $\vec{k}$ criterion. See details in text and Figure 2.6 for explanations of the different parameters presented here. Results are sorted by the number of specimens in the site used to make the estimate using our method.

Model Name	$\bar{f}_{prob}$	$\tilde{\Delta}_{median}$ (μT)	f_{acc}	Number of Sites
Linear, 5 μT (BiCEP)	0.63	1.7	0.85	25
Linear, 10 μT	0.62	1.7	0.85	25
Linear, 20 μT	0.61	1.7	0.85	25
Quadratic, 5 μT	0.56	1.7	0.81	25
Quadratic, 10 μT	0.55	1.6	0.85	25
Quadratic, 20 μT	0.55	1.8	0.85	25
Cubic, 5 μT	0.45	2.6	0.85	22
Cubic, 10 μT	0.45	2.3	0.85	24
Cubic 20 μT	0.44	2.5	0.85	24
CCRIT	0.47	1.9	0.88	22
PICRIT (Modified)	0.56	2.2	0.77	23
PICRIT (Modified with $\vec{k}$)	0.61	1.9	0.69	21
SELCRIT (Modified)	0.53	2.9	0.58	25
SELCRIT (Modified with $\vec{k}$)	0.58	2.3	0.58	23

Figure 2.7 shows the 95% credible intervals for each method, normalized by the expected value at the site. The median values of our results are generally similar to those found by our

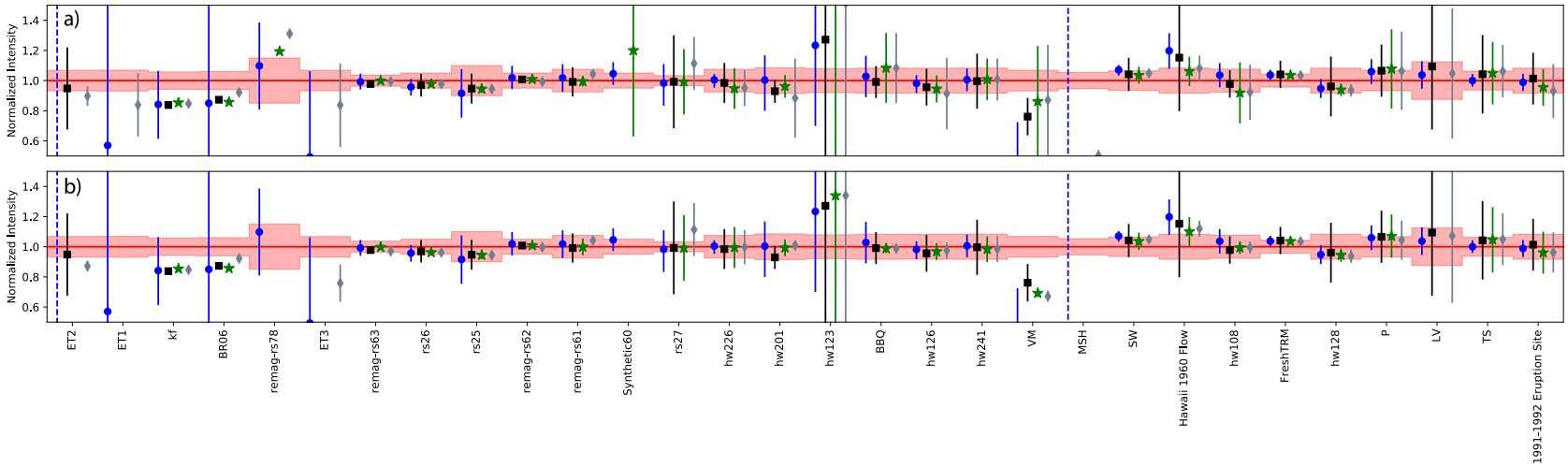

Figure 2.7. In a) for our collection of sites, we plot (Table 2.1) paleointensity estimates using BiCEP (blue circles) with 95% confidence interval compared to results using CCRIT (black squares), Paterson's modified PICRITMOD (green stars) and SELCRITMOD (grey diamonds). In b), the same information is plotted, but green stars represent results from PICRITMODk and grey diamonds represent results from SELCRITMODk; the $\vec{k} < 0.270$ criterion has been applied to PICRITMOD and SELCRITMOD. A dashed blue line indicates a site where the sampler failed with $\hat{R} > 1.1$, so the prior distribution (a uniform distribution between 0 and 250 μT was used. The results are normalized to the expected field value for each site (black line, by definition 1) and the pink shaded region represents $\pm$ 3 μT. Sites are ordered by the number of specimens used by BiCEP for paleointensity analysis in that site with the number of specimens increasing to the right. Results are summarized in Supplementary Data Set S1.

selection criteria. BiCEP yields the largest number of accurate and precise results, with CCRIT being generally less precise and slightly less accurate. PICRITMOD and SELCRITMOD are generally less accurate and precise than BiCEP, however introducing the curvature criterion for PICRITMODk and SELCRITMODk improve the accuracy and precision significantly. Both PICRITMODk and SELCRITMODk boast highly precise estimates for passing sites, with similar levels of accuracy to BiCEP. However, this improved accuracy and precision is achieved by excluding more sites, which penalizes these methods using our success metrics.

Sites in Figure 2.7 are sorted by the number of specimens used by BiCEP for the analysis. Unique to our method, sites with low numbers of specimens (M) have wide credible intervals and sites with high M have narrow credible intervals, so the estimate of B_{anc} becomes more precise as more specimens are measured. This is because calculating the credible interval for a B_{anc} is more similar to calculating the standard error of the mean than the site level standard deviation, which is done for our traditional selection criteria. The increasing precision on B_{anc} leads to some sites with high M having estimates of B_{anc} which are seemingly too precise. These estimates are still generally only a few μT away from the expected value, however, and we discuss potential reasons for this in Section 2.4.4.

Our results in Table 2.2 indicate that BiCEP is the method that yields the largest number of accurate and precise results, having a higher $\bar{f}_{prob}$ and lower $\tilde{\Delta}_{median}$ than all of our sets of selection criteria. For selection criteria which include a curvature criterion, much of this improvement comes from BiCEP's inclusion of accurate results for two sites, remag-rs78 and Synthetic60. If we look exclusively at the sites which passed each criterion, PICRITMODk and SELCRITMODk achieve higher levels of precision for those sites (higher $\bar{f}_{prob}$ than BiCEP if only passing sites considered), with PICRITMODk achieving similar levels of accuracy to BiCEP (similar $\tilde{\Delta}_{median}$ for passing sites). This higher level of precision is likely an outcome of using the standard deviation optimization procedure, and is probably not reflective of the true uncertainty judging by the low f_{acc} for both PICRITMODk and SELCRITMODk. CCRIT still achieves lower $\bar{f}_{prob}$ and higher $\tilde{\Delta}_{median}$ than BiCEP even if only passing sites are considered,

indicating a slightly lower accuracy and precision overall. Our two selection criteria which do not include a curvature criterion (PICRITMOD and SELCRITMOD) have a larger number of passing sites, including remag-rs78 and Synthetic60, but still have reduced $\bar{f}_{prob}$ and $\tilde{\Delta}_{median}$. Ultimately it seems that BiCEP offers the best of both worlds, passing at least as many sites as the more permissive criteria, and achieving higher accuracy and more realistic precision than the more restrictive criteria.

2.3.2 Width of the prior

To investigate the role of the prior distribution ($P(\sigma_{site})$), we apply the BiCEP method on the data compilation using a variety of values for its standard deviation (see Table 2.2). The main effect of varying σ_{site} is that for smaller values, the estimates of B_m and $\vec{k}_m$ for specimens are "pulled" closer to the line being fitted at a site level (see Figure 2.5a,b). For our estimate of B_{anc}, this means that sites with fewer specimens will be more precise, as it is unlikely that specimen B_m will deviate strongly from the mean. For sites with many specimens, there is little effect as σ_{site} is well constrained by the data.

From Table 2.2, we see that changes to $P(\sigma_{site})$ seem to have little influence on the effectiveness of the model, as all our f_{acc} values are the same for our linear model regardless of the prior distribution used. We can also see graphically in Figure 2.7 that our precision is low for sites with small number of specimens (M). Because of this, we favor the version of the model with a 5 μT standard deviation on $P(\sigma_{site})$, as models with higher standard deviations reduce precision without capturing any more sites within their 95% credible intervals.

2.3.3 Order of polynomial fit

The results for our test sites (Table 2.2) demonstrate that increasing the order of the polynomial fit decreases the precision of the estimate as demonstrated by reduced values of $\bar{f}_{prob}$. This is expected as there are more parameters to be estimated with the same number of data. The level of accuracy is not significantly improved by increasing the model order. The best quadratic

model produced a $\tilde{\Delta}_{median}$ of 1.6 μT, which is not a significant improvement over the value of 1.7 μT for the best linear model to account for the reduction in precision. The number of passing sites is reduced for the cubic model, indicating that the sampler is struggling to fit this model. Consequently, the cubic model produces more inaccurate and less precise results. For this reason, we assume a linear relationship between B_m and $\vec{k}_m$.

2.3.4 Sampler Diagnostics

The sampler diagnostics for each site are given in Table 2.3. Indicators of poor MCMC sampler performance (worst $\hat{R}$ >1.1, low n_{eff}, high f_{div}) tend to occur at sites with four or fewer specimens, or for specimens where the Arai plots are extremely scattered and the sampler struggles to fit them. In the latter case, it may be possible to exclude these specimens by looking at which specimen level parameters have high $\hat{R}$, as this indicates that fitting a circle to these specimens is inappropriate. We did not exclude specimens on this basis in our analysis, however, we include an option to do this in the BiCEP GUI software (see Appendix 2.6.3).

The prevalence of high $\hat{R}$ for sites with low numbers of specimens indicates that to get a strongly reproducible answer from this method, paleomagnetists ought to measure five or more specimens per site. In practice, most studies already do this in order to have enough specimens that pass the chosen selection criteria, yet many specimens may be excluded from analysis. Here, we can use all of the specimens measured so there may be no additional burden.

2.3.5 Summary of Results

After testing all of our contingent models, we prefer the model which assumes the relationship between B_m and $\vec{k}_m$ is linear, and which uses a 5 μT standard deviation on $P(\sigma_{site})$. This model performs better than classical sets of selection criteria, either passing a greater number of sites (than CCRIT, PICRITMODk, SELCRITMODk) or having significantly higher accuracy and precision (than PICRIT, SELCRIT). Our precision increases for sites for which the number of specimens is large, similar to calculating the standard error of the mean when

Table 2.3. Sampler diagnostics (see Section 2.2.5 for an explanation of each diagnostic) for each site using the BiCEP method.

Site Name	Worst $\hat{R}$	n_{eff}	f_{div}
1991-1992 Eruption Site	1.00	59741	0.00
hw108	1.00	77959	0.00
hw123	1.01	11687	0.00
hw126	1.00	36130	0.01
hw128	1.00	78978	0.00
hw201	1.00	10641	0.01
hw226	1.00	7139	0.05
hw241	1.00	66565	0.00
BR06	1.01	451	0.00
P	1.00	62252	0.00
VM	1.05	1447	0.00
BBQ	1.00	63082	0.00
rs25	1.00	5614	0.00
rs26	1.00	11866	0.00
rs27	1.00	22211	0.00
remag-rs61	1.00	26746	0.00
remag-rs62	1.00	16916	0.00
remag-rs63	1.00	3788	0.00
remag-rs78	1.00	12388	0.00
kf	1.02	2712	0.00
Hawaii 1960 Flow	1.00	60184	0.00
SW	1.00	36390	0.00
TS	1.00	56518	0.00
ET1	1.01	995	0.00
ET2	6.93	6	0.03
ET3	1.01	424	0.00
Synthetic60	1.00	36572	0.01
LV	1.02	5931	0.08
MSH	2.78	24	0.45
FreshTRM	1.00	81007	0.00

using selection criteria. Unlike selection criteria, the BiCEP method does not require exclusion of large numbers of specimens to obtain an accurate result, which leads us to prefer it over those methods.

2.4 Discussion

2.4.1 Advantages of BiCEP compared to selection criteria

BiCEP has significant advantages over the classical selection criteria approach. Firstly, we obtain paleointensity estimates for all sites with at least three specimens, including some which do not contain any specimens that would pass classical selection criteria (see Figure 2.7). In most cases, our estimates have similar or higher accuracy than the selection criteria approach (evidenced by $\tilde{\Delta}_{median}$ and Figure 2.7), and this is accomplished while only excluding specimens from the analysis which were not fully demagnetized. In some cases, our method yields results even if none of the selection criteria accept any specimens or are inaccurate. For example, for sites remag-rs78 and Synthetic60, our strict criteria (CCRIT, PICRITMODk, SELCRITMODk) produce no results, and our more permissive criteria (PICRITMOD, SELCRIT) produce less accurate (and in the case of Synthetic60, much less precise) results than BiCEP.

Secondly, the increasing precision of our paleointensity estimate as the number of specimens increases allows for an improved workflow when compared to classical paleointensity criteria. Instead of needing a minimum number of specimens to pass our selection criteria, we can keep measuring specimens until we reach a desired level of precision. We discuss this workflow in more detail in Section 2.4.3. The property of increasing precision with number of specimens is inherent to Bayesian models and can also be found in the method of Kosareva et al., 2020, although their method does not include the bias correction found in our method.

Thirdly, the BiCEP method propagates the uncertainties from a specimen to the site level. Specimens with more scattered (or non linear, or non circular) Arai plots will have less influence over the specimen mean than those with highly linear Arai plots. In addition to this, the BiCEP method foregoes the need for criteria which are concerned with the length of the line on the Arai plot used to make an interpretation, like the NRM Fraction (e.g., FRAC of Shaar and Tauxe, 2013). Using a set of temperatures with small FRAC will cause an increase in the uncertainty in $\vec{k}$ (see Figure 2.4e, f), which will cause this specimen to have less effect on the estimate of B_{anc},

without excluding it from the analysis entirely. We discuss this further in Section 2.4.5.

2.4.2 Predictive ability of the method

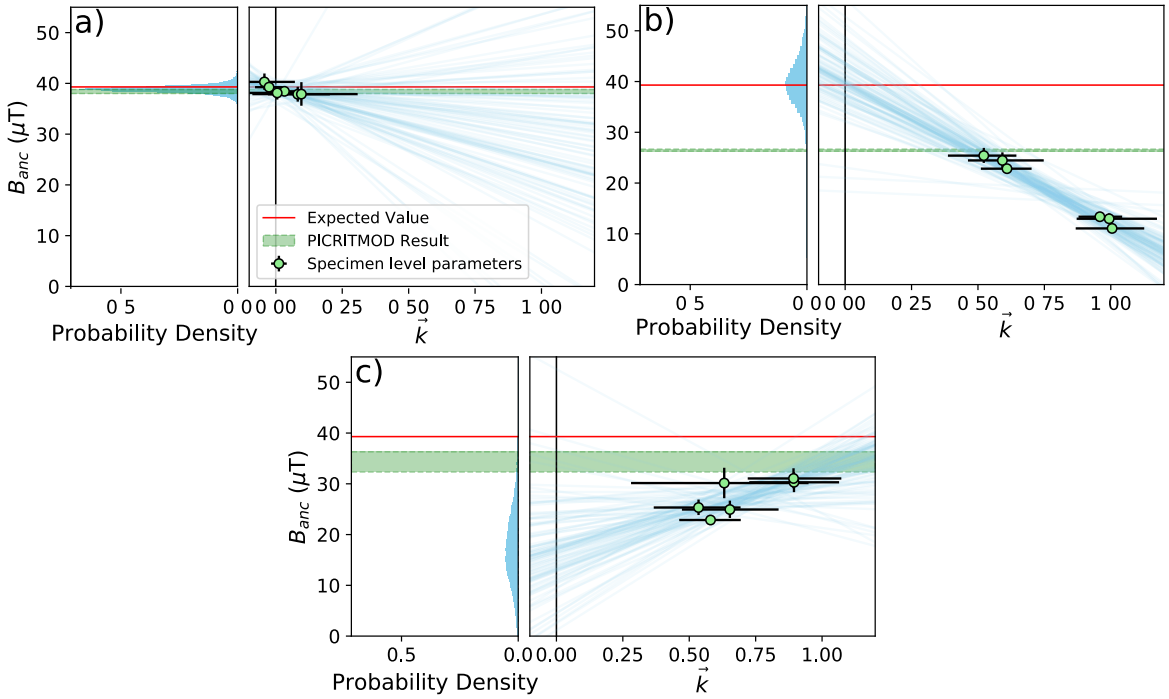

Figure 2.8. Example of the BiCEP method applied to three subsets of 6 specimens from site hw108 (B_{exp} = 39.3 μT). The left column in each subplot shows histograms of the BiCEP results, and the right column shows plots of specimen $\vec{k}$ vs B_{anc} with the BiCEP line fits. Light green shaded regions with dashed edges represent the 2σ interval of the PICRITMOD estimate for these subsamples. In a) there is a small range of $\vec{k}$ and B_{anc} values which leads to an imprecise estimate of c, but an accurate and precise estimate of B_{anc}. In b) there is a large range of values on $\vec{k}$, but all specimens have high $\vec{k}$. This leads to an estimate with a relatively precise estimate of c, and an accurate but imprecise estimate of B_{anc}. In c) there is a reasonably small range of values on B_{anc}, and the relationship between B_{anc} and $\vec{k}$ is not linear, but BiCEP attempts to find a linear model. This leads to an imprecise and inaccurate estimate of both c and B_{anc}.

Although our results are promising, it is worth noting that traditional selection criteria also perform well for the majority of our sites. To see if the BiCEP method offers accurate results with poorer quality data, we subsampled results from site hw108, which had a range of good and poor quality specimens. Figure 2.8 shows the results of BiCEP applied to three different subsets of six specimens taken from this site, along with the results of the PICRITMOD criteria applied

to this site (in green). It is worth noting that only the specimens in Figure 2.8a would pass the CCRIT criteria which gave a highly accurate result (within 1 μT), or any of our more restrictive criteria.

We identify three behaviours for which BiCEP results deviate from a linear model with high precision on the slope and intercept. Figure 2.8a shows a subset of specimens for which the range of $\vec{k}$ values of the specimens is very small, and so the uncertainty of the slope of the linear relationship between $\vec{k}_m$ and B_m (c) is high. In this case, however, because these specimens all have $\vec{k}$ close to zero, the estimate of B_{anc} is accurate and precise. Figure 2.8b shows a different subset of specimens for which the range of $\vec{k}$ values is large, but there are no $\vec{k}$ values close to zero. This results in an estimate of B_{anc} which is still accurate, but imprecise due to the uncertainty in extrapolating the linear relationship between $\vec{k}_m$ and B_m back to zero. The PICRITMOD result for this subset returns an average value which underestimates B_{exp} by around $\sim$13 μT or $\sim$30%, and criteria using the curvature criterion return no values, as all specimens have curvature values higher than the threshold. The high uncertainty in B_{anc} might still be considered a problem, but this result indicates that measuring more specimens would likely yield a more precise result.

Figure 2.8c shows a set of specimens where the range of $\vec{k}$ is low, so the $\vec{k}_m$ versus B_m relationship is not particularly linear. BiCEP attempts to find a linear trend with these data, and extrapolates back to a B_{anc} which is both highly inaccurate and imprecise. This might be considered a problem for BiCEP, but it is possible to detect such behavior as the uncertainty on both B_{anc} and the slope relating the B_m versus $\vec{k}_m$ (c) are large. This indicates to us that we can use a metric of the uncertainty in both the slope and intercept of the linear fit in BiCEP to decide whether a site level result is accurate or not. This leads us to a laboratory workflow which uses BiCEP results to decide if a site is acceptable, might be acceptable with further work or is unlikely to give a reasonable result.

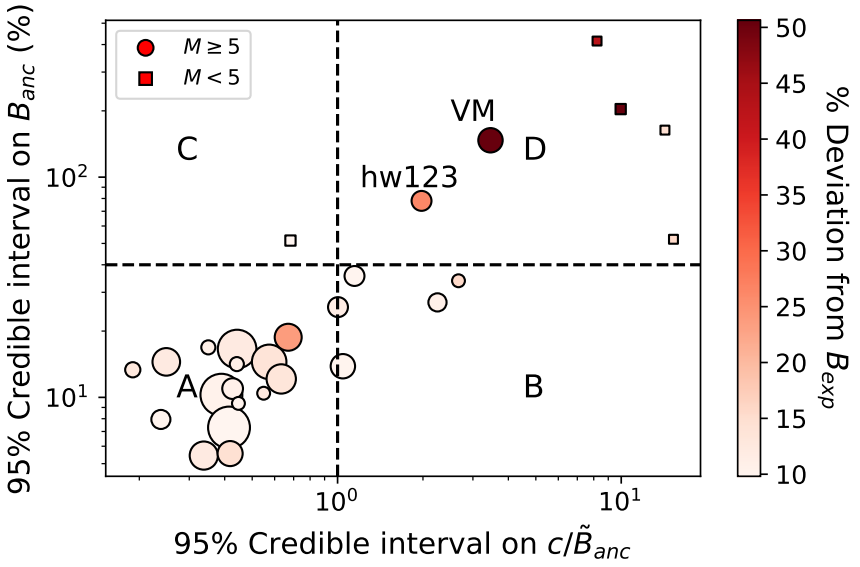

Figure 2.9. Plot of the 95% credible interval on B_{anc} against the 95% confidence interval on c (slope between intensity estimate and $\vec{k}$), normalized by the median B_{anc} for all sites with $\hat{R} < 1.1$. Circles indicate sites for which the number of specimens $M \geq 5$, and squares indicate sites where $M < 5$. Colors indicate the deviation of the median value of the estimate from the expected site value (B_{exp}) as a percentage. The size of markers is used to represent M. The horizontal dashed line indicates a value of 40% for the full width of the 95% credible interval on B_{anc}, which for a Gaussian distribution would correspond to a standard deviation of $\pm$10%. The vertical dashed line represents a value of 1 for the 95% confidence interval of $c/\tilde{B}_{anc}$. Suggested workflow for sites in regions: A) or B), accept the site or continue measuring if improve precision is desired. C) Continue measuring specimens, as improved precision is likely. D) If $M \geq 5$ stop measuring the site as further effort is likely to be futile. Otherwise continue measuring specimens until $M = 5$.

2.4.3 Workflow with BiCEP

Figure 2.9 plots the 95% credible interval on B_{anc} as a percentage against the 95% credible interval on c as a proportion of the median $\tilde{B}_{anc}$ for all sites where $\hat{R} > 1.1$. The sizes of the points on the plot represents the number of specimens per site (M), with squares representing sites with $M < 5$. The colors show the percentage deviation from B_{exp} using BiCEP, with redder colors for more inaccurate results. With the exception of two sites (VM and hw123), as the number of specimens increases, sites trend towards the bottom left region of this plot, indicating an increase in precision. This has diminishing returns as the number of specimens increases

above five. The increase in precision is also accompanied with an increase in accuracy. Almost all sites with an estimated precision on B_{anc} better than 40% have median values within 20% of B_{exp}. Our outlier sites VM and hw123, which are imprecise despite having large numbers of specimens (M=12 and 18 respectively), are also inaccurate. This indicates that the width of the 95% credible intervals is a useful statistic for diagnosing inaccuracy in the BiCEP method.

We have divided Figure 2.9 into four regions (labeled A-D). Sites in region A have high precision on both B_{anc} and c and are representative of the results for the majority of sites in this study; sites in this region are highly accurate. Sites in region B have high precision on B_{anc} (better than 40%, which for a Gaussian distribution would be equivalent to a standard deviation of $\pm$10%) but low precision on c (95% credible interval on $c/\tilde{B}_{anc} > 1$). These sites are usually analogous to the example shown in Figure 2.8a, with low Arai plot curvature and similar intensities for all specimens. Sites in region C have high precision on B_{anc} but a low precision on c. These sites may have a large number of curved specimens which follow a linear trend that can be extrapolated back to the correct B_{exp}, and are analogous to our example in Figure 2.8b. Region D is representative of the worst constrained estimates, with low precision on B_{anc} and c. Sites in this region may have highly inaccurate estimates of B_{anc}, often with low M. If these sites have high M, they may be similar to our example in Figure 2.8c in which a linear relationship between B_{anc} and c is not well determined, and the average $|\vec{k}|$ is large, leading to an inaccurate estimate of B_{anc}.

Considering the region in which a particular site plots leads to a workflow based on the likelihood of success. In general, sites with very low numbers of specimens, ($M = 2$ or 3), will begin in region D, and migrate to regions C, B or A as M increases to around five. If a site has migrated to region A or B after five specimens have been measured, then we likely have an accurate and precise estimate of B_{anc}, and we can finish measuring specimens (or continue to measure if a higher level of precision is desired). If a site has migrated to region C, it is likely that our estimate is accurate and that our uncertainty in B_{anc} can be reduced by increasing the number of specimens. If our site remains in region D after five specimens have been measured,

the site level estimate may be inaccurate, and measuring more specimens would be unlikely to reduce the site level uncertainty.

Because the regions in Figure 2.9 define a workflow based on measuring five specimens, we wanted to test whether our methodology could identify sites which will remain in region D after a large number of specimens were measured. We randomly subsampled 100 sets of 5 specimens from sites hw108, VM and hw123 and calculated B_{anc} and c using BiCEP. Site hw108 was chosen because contains specimens which exhibit a wide range of behaviours, with a large number of specimens having high $\vec{k}$, but yields an accurate result. Sites VM and hw123 were chosen because these are our sites which remain in region D after measuring a large number of specimens. Our results for these three sites are given in Supplemental Figure S2. Our subsampled hw108 obtained results in region D 5 times out of 100, whereas our subsampled VM and hw123 obtained results in region D 94 and 90 times respectively. This indicates that sites which remain in region D after measuring 5 specimens are likely to remain there after measuring many more, and so measuring more specimens is usually a futile effort.

2.4.4 Overly precise estimates of B_{anc}

The BiCEP method has a lower f_{acc} than CCRIT, despite having a similar degree of accuracy when using a metric like $\tilde{\Delta}_{median}$. The reason for this is that the increasing precision on the BiCEP estimate leads to estimates which are highly precise when M is large. This is the case shown in Figure 2.6d.

Labeling sites with extremely high precision in the estimate as inaccurate may be misleading, as we have not taken into account uncertainties in the value of the expected fields at the sites in this study. For example, using differences between the observed directions and the IGRF, Yamamoto and Hoshi, 2008 quoted the expected value at the site "SW" as 46.0$\pm$2.6 μT, which is just consistent with the 95% credible interval for our specimen (48.2-49.7 μT). Because of this, we prefer to use $\bar{f}_{prob}$ as a metric of how well a model performs as it allows for a few μT of uncertainty in the expected field value. Additionally, Yamamoto and Yamaoka, 2018

suggested that the IZZI-Thellier results for sites SW and TS may be biased slightly high due to acquisition of a thermo-chemical remanent magnetization (TCRM), which is not detectable by our method. Yamamoto et al., 2003 also invoke a TCRM mechanism to explain the paleointensity overestimate for the Hawaii 1960 Flow, which is another of their sites for which we overestimate the expected intensity (see Figure 2.7 and Supplementary Data Set S1). We note that Cromwell et al., 2015 also sampled the 1960 flow (hw241 which targeted the fine grained flow top) and all selection criteria resulted in accurate results, with BiCEP producing the tightest confidence interval.

2.4.5 Exclusion of measurement level data

It is frequently possible to improve the accuracy and precision of results by finding the 'best' set of temperature steps to use in the intensity interpretation. Two situations frequently occur for which this might be justified. The first is the case in which thermochemical alteration occurs at high temperature (e.g., Figure 2.4e). For such specimens, the low temperature measurements can be used to make a paleointensity estimate (colored dots in the figure). Figures 2.4e and f show how our method can be used on a reduced range of temperature steps on the Arai plot at the cost of precision. The plot of circle fits (green lines in Figure 2.4e) demonstrates that the Arai plot interpretations are poorly constrained and can continue in any direction after the last temperature step chosen. This results in a higher uncertainty in the curvature associated with this (Figure 2.4f). The second case in which a portion of the data could be excluded from the calculation, would be when the magnetization has multiple components (Figure 2.10a). In such a case, a paleointensity estimate can only be made using the small range of temperature steps that correspond to the characteristic component. We currently do not have an objective method to choose which set of temperature steps on the Arai plot to use. We suggest that decisions about which data points to include should not be made based on the original in-field or zero field Arai plot measurements (dots in the Arai plots), but rather exclusively on deviating pTRM checks (triangles in, e.g., Figure 2.4e) or other indicators of alteration for the first case and on

the directions of the magnetization vector (it must trend to the origin and be well defined) in the second case, e.g., Figure 2.10a.

Caution should be used when excluding a particular temperature steps for reasons other than this. If the set of temperature steps chosen does not represent the characteristic component of magnetization, this can alter the outcome of the BiCEP method, especially if a large part of the Arai plot is excluded. Additionally, excluding more points on the Arai plot tends to increase the chance that a specimen will cause $\hat{R}$ failure. As such, we recommend using as many points on the Arai plot as possible unless done for one of the reasons stated above.

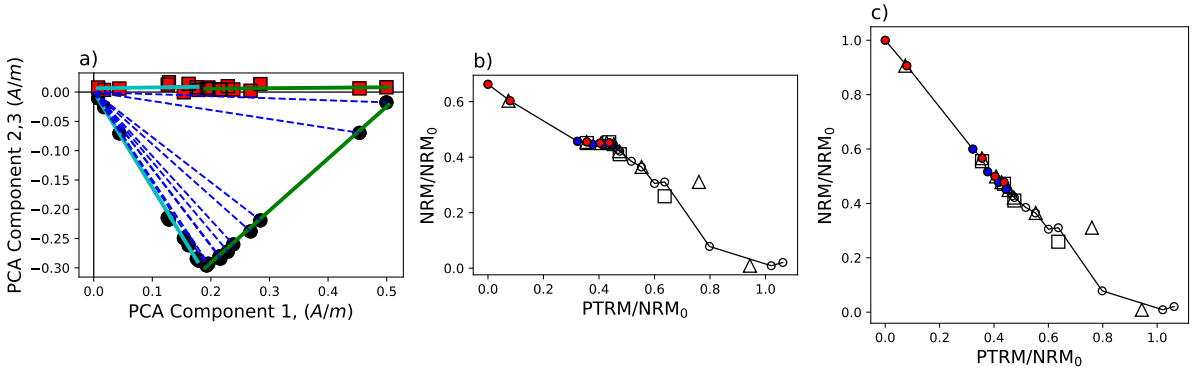

Figure 2.10. a) Example of vector endpoint diagram for specimen FB2-B1 from Lisé-Pronovost et al., 2020. The magnetization is rotated so that the principal component of the TRM direction for all steps lies along the x axis. Green line fit to the low temperature component and cyan line fit to the high temperature component. b) Arai plot and c) "corrected" Arai plot for a specimen from the data shown in b). NRM values for the low temperature component (filled circles) are usually calculated by taking the magnitude of the vector endpoint (blue dashed lines in the vector endpoint diagram in a). In c), these NRM values are calculated by vector subtracting the high temperature component (cyan line), taking the magnitude of our new NRM vectors (distance along green line), and adding the magnitude of the low temperature component (length of cyan line). Both b) and c) are scaled by the total NRM distance along both components (total distance along both green and cyan lines).

2.4.6 Application to multi-component magnetizations

We test an application of the BiCEP method on data with multi-component directions as shown in Figure 2.10a using the data of Lisé-Pronovost et al., 2020. The data are from Scottish firebricks which were used in a foundry in Australia. The date and location of firing are both well

constrained, hence we have a reasonably well constrained value for B_{exp}. The bricks all contained a low temperature component associated with the Australian field. Some also displayed a high temperature component associated with the original firing in Scotland as shown in Figure 2.10a. Lisé-Pronovost et al., 2020 already have interpretations which separate these components in the original study. To account for the change in direction of the NRM, we subtract the high temperature component from the low temperature component, and then add the magnitude of these values to the magnitude of the low temperature component (see Figure 2.10 for a graphical explanation). The vector subtraction is necessary for the low-temperature component as we need a total TRM ($pTRM_{max}$) to scale by in order to penalize the result for shorter components. We then proceed to use the BiCEP method as previously described, using the original interpretations for the different components. For the sake of simplicity, we do not perform the magnetomineralogical change (MMC) correction (Valet et al., 1996). We also do not apply the corrections for anisotropy of TRM or cooling rate with these data, as they appeared to be negligible. Of course these could be applied in the usual fashion if necessary.

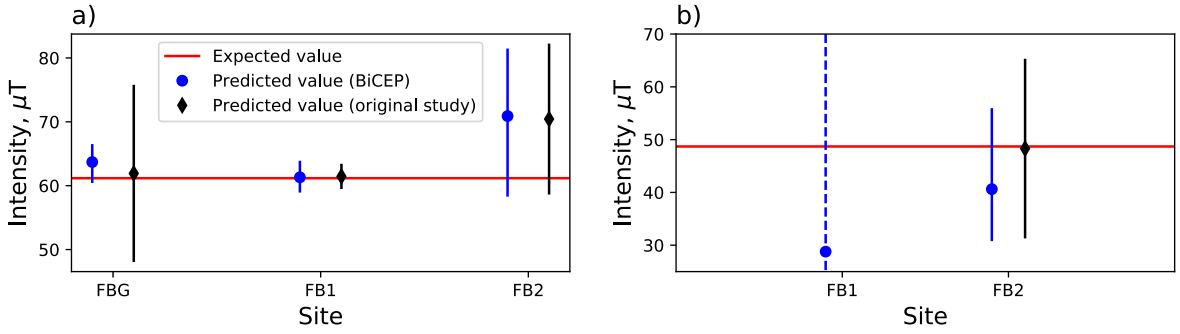

Figure 2.11. Expected and predicted intensities on the data of Lisé-Pronovost et al., 2020 using BiCEP (blue circles) and the method used in the original study (black diamonds). a) Results for the low temperature component (Australia, expected field value 61.17 μT) for each firebrick. b) Results for the high temperature component (Scotland, expected field value 48.3 μT), where this component was present. The dashed blue line indicates that the MCMC sampler failed to converge for site FB1.

We display the results from multi-component remanences in Figure 2.11. We find that for the low temperature, Australian field, component (Figure 2.11a), our estimates for all firebricks

contain the expected answer (61.17 μT) within the 95% credible interval. Our interpretation for site FBG is slightly less accurate than the original analysis but with much higher precision. This difference is likely caused by not applying the MMC correction, as the specimens at this site were mostly of good quality, with none being excluded from the original analysis.

For the high temperature component (Figure 2.11b) our results behave differently. The sampler does not converge for site FB1, indicating too few specimens in the analysis. For site FB2, we have a result that is less accurate, but more precise than in the original study. The lack of MMC correction may contribute to the decreased accuracy in this example, whereas the reduced precision is likely caused by the smaller length of the interpretation on the Arai plot, leading to a higher uncertainty in the curvature for that specimen. Our results for this study demonstrate that BiCEP will obtain precise estimates for components which represent most of the magnetization, and be imprecise for components which have small NRM fraction.

2.4.7 Implications for bias in curved Arai plots

The success of our method demonstrates that Arai plot "curvature" or sagging does lead to a progressive bias in paleointensity estimation which increases as the amount of curvature increases as described by Tauxe et al., 2021 and strongly suggested by the data of Krása et al., 2003 (see Figure 2.1). Our estimates are made by using the tangent to a circle fit rather than fitting a line to part of the data, so one might expect them to be biased. However, it has been demonstrated by e.g. the data of (Krása et al., 2003) that fitting lines to the high temperature or low temperature slope of Arai plots yields even more biased results than using the total TRM, which is more similar to the tangent. The scaling used by our method incorporates the added uncertainty in the line slope and $\vec{k}$ associated with choosing one of these slopes, which allows for more consistent analysis between specimens with interpretations of varying quality. The bias seen generally underestimates paleointensity with higher (positive) curvature, but this is not the case for all sites, some of which exhibit the opposite trend.

The assumption of a quasi-linear dependence between the specimen level paleointensities

and the curvature of the Arai plot does not have any theoretical basis. This does imply that the curvature is linearly related to the change in TRM susceptibility (or decay of the original magnetization) between the original and lab coolings, a relationship which should be further investigated. We stress that this relationship only needs to be loosely followed for our method to work. In cases where there does not appear to be a strong linear relationship between B_m and $\vec{k}_m$ (e.g. in Figure 2.8a), an accurate paleointensity estimate is still possible if there are enough specimens with low $|\vec{k}|$, as the intercept of the linear fit is still well constrained even if the slope is not. Conversely, if there are few specimens with low $|\vec{k}|$ and there is a poor linear relationship, then both the slope and intercept are poorly constrained, resulting in a huge uncertainty in B_{anc}, as is seen in Figure 2.8c.

2.5 Conclusions

- We present a new Bayesian method (BiCEP) which accounts for bias in paleointensity estimates in specimens.

- Instead of excluding specimens from the paleointensity analysis in the traditional (binary) selection criteria based approach, our method predicts an amount of bias for each specimen, using the curvature of the Arai plot as a metric of non-linearity and a predictor of bias. In this way, the BiCEP method is quite different from the recently published Bayesian approach of Kosareva et al., 2020.

- When tested on a compilation of sites for which an approximate paleointensity is known *a priori*, our method yields levels of accuracy and precision similar to, or better than restrictive paleointensity criteria, whilst accepting as many results as permissive criteria.

- Our method generates some slightly inaccurate paleointensity estimates with high levels of precision, but these can generally be explained with inaccuracies in the expected field (see Section 2.4.4).

- The BiCEP method handles uncertainties in a different way than using classical selection criteria, as the uncertainty in site level estimates decreases as the number of specimens increases, but this uncertainty remains high when the number of specimens is low due to inclusion of prior information. The Bayesian uncertainties are in this way more similar to the 'extended error bars' in the Thellier_GUI auto-interpreter of Shaar and Tauxe, 2013.

- We propose a workflow in which sites are accepted and measurement of specimens can cease once a desired level of confidence in the site level estimate has been reached. Sites which do not reach this level of confidence after measuring several (> 5) specimens likely do not contain useful information and can be discarded.

Data Availability Statement

Data used in this paper may be found in the MagIC database at: https://earthref.org/MagIC/17104/0326fdaa-4bcf-44f3-989d-0116b9a2fb75 for review and will be available to the public at https://earthref.org/MagIC/17104 on publication.

2.6 Appendix

2.6.1 Change of variables

In Section 2.2.2 we mention that we need to use a change of variables to get from our original circle fitting parameters R, x_c, y_c to our new set of parameters $\vec{k}, D, \phi$. We can use the Jacobian of the parameter change to get the new formula for the posterior probability under our new parameters:

$$P(D,\phi,\vec{k}|x,y) = P(x_c,y_c,R|x,y)\left|\frac{\partial(x_c,y_c,R)}{\partial(D,\phi,\vec{k})}\right|. \tag{2.20}$$

We can evaluate this Jacobian as:

$$\left|\frac{\partial(x_c, y_c, R)}{\partial(D, \phi, \vec{k})}\right| = \left|\frac{\vec{k}}{|\vec{k}^3|}\left(D + \frac{1}{\vec{k}}\right)(\cos\phi + \sin\phi)\right|. \tag{2.21}$$

So our posterior looks like:

$$P(D, \phi, \vec{k}|x, y) \propto \left(\sum_{n=1}^{N} \sqrt{((D + \frac{1}{\vec{k}}\cos\theta) - x_n)^2 + ((D + \frac{1}{\vec{k}}\sin\theta) - y_n)^2} - \frac{1}{|\vec{k}|})^2\right)^{-N/2}$$

$$\left|\frac{\vec{k}}{|\vec{k}^3|}\left(D + \frac{1}{\vec{k}}\right)(\cos\phi + \sin\phi)\right| P(\vec{k}, \phi, D). \tag{2.22}$$

2.6.2 Markov chain Monte Carlo sampling

The Markov chain Monte Carlo (MCMC) sampling method generates a set of samples from the posterior probability distribution of B_{anc} which allows us to approximate it. We use the python bindings for the Stan software package (http://mc-stan.org) to generate these samples which provides diagnostic information and runs relatively quickly. For each site we run four Markov chains and generate 30,000 samples of B_{anc} in each chain. We discard the first half of the chain as 'burn in' for a total of 60,000 samples.

Stan provides several diagnostics that tell us whether we have successfully sampled the posterior distribution. These include the $\hat{R}$ score (Gelman and Rubin, 1992) which tells us about the convergence between chains, and is required to be between 1.1 and 0.9 which is necessary for convergence, the effective sample size, n_{eff} which should be large (> 1000) for a good sample and the number of divergent transitions (f_{div}) which should be zero in ideal cases. In most cases our results display high degrees of convergence with $\hat{R}$ close to 1 and high effective sample sizes. Some sites included divergent transitions in small numbers. These seem to occur at a specimen level for specimens where the posterior distribution of one of the circle parameters is long-tailed. In theory this can mean the posterior was inefficiently sampled, but because these specimens generally have large uncertainties on their $\vec{k}$ parameter, the final results do not change, even under

a change of parameters. The sampler struggled to converge, with $\hat{R} > 1.1$ for several sites with very few specimens, where once again the distributions are extremely long tailed. The sampler also did not converge for site MSH, where the Arai plots were so non linear, with few points, that BiCEP struggled to fit circles to them. We consider these sites to have "failed" using our method (grade of 'D' in Figure 2.9) and use the prior distribution on B_{anc} (uniform between 0 and 250 μT) as an estimate of their intensity. We calculate the $\hat{R}$ furthest from unity, the n_{eff} for B_{anc} and the proportion of divergent samples f_{div} for our model.

2.6.3 Code and GUI

We present a simple GUI that can perform the BiCEP method on data in the MagIC format. The code uses Jupyter notebooks and can be found at (http://github.com/bcych/BiCEP_GUI) and contains a readme file detailing how to use the notebook. The GUI can also be accessed at the Earthref JupyterHub site (http://jupyterhub.earthref.org). To access the GUI this way:

- Sign up to Earthref at (http://earthref.org)
- Navigate to the Earthref JupyterHub site at (http://jupyterhub.earthref.org)
- Open and run all the cells in the "BiCEP GUI - Setup.ipynb" notebook.
- Upload MagIC formatted "sites", "samples", "specimens" and "measurements" files to the BiCEP_GUI directory in JupyterHub. These can be formatted using pmag_gui (Tauxe et al., 2016).
- Open the BiCEP GUI notebook and press the "App Mode" button.

For more detailed instructions, read the included readme file at the github site.

2.7 Acknowledgements

We are deeply grateful to Lennart de Groot and Greig Paterson for their very helpful reviews and for the advice and guidance given by Andrew Roberts, David Heslop and Joseph

Wilson. This research was supported in part by NSF Grant EAR1827263 to LT. We are also grateful to Agnes Lisé-Pronovost for sharing her measurement level data for use in section 4.6.

Chapter 2, in full is a reprint of material from Cych, B., Morzfeld, M., & Tauxe, L. (2021). Bias Corrected Estimation of Paleointensity (BiCEP): An improved methodology for obtaining paleointensity estimates." Geochemistry, Geophysics, Geosystems, 22, e2021GC009755. https://doi.org/10.1029/2021GC009755. The dissertation author was the primary investigator and author of this paper.

References

Biggin, A. J. (2010). Paleointensity database updated and upgraded. *EOS*, *91*, 15.

Biggin, A. J., Perrin, M., & Dekkers, M. J. (2007). A reliable absolute palaeointensity determination obtained from a non-ideal recorder. *Earth and Planetary Science Letters*, *257*(3), 545–563. https://doi.org/10.1016/j.epsl.2007.03.017

Bowles, J., Gee, J. S., Kent, D. V., Perfit, M. R., Soule, S. A., & Fornari, D. J. (2006). Paleointensity applications to timing and extent of eruptive activity, 9°–10°n east pacific rise. *Geochemistry, Geophysics, Geosystems*, *7*(6). https://doi.org/10.1029/2005GC001141

Chernov, N., & Lesort, C. (2005). Least squares fitting of circles. *Journal of Mathematical Imaging and Vision*, *23*(3), 239–252. https://doi.org/10.1007/s10851-005-0482-8

Cromwell, G., Tauxe, L., Staudigel, H., & Ron, H. (2015). Paleointensity estimates from historic and modern hawaiian lava flows using glassy basalt as a primary source material. *Phys. Earth Planet. Int.*, *241*, 44–56. https://doi.org/10.1016/j.pepi.2014.12.007

Donadini, F., Kovacheva, M., Kostadinova, M., Casas, L., & Pesonen, L. (2007). New archaeointensity results from scandinavia and bulgaria: Rock-magnetic studies inference and geophysical application. *Physics of the Earth and Planetary Interiors*, *165*(3), 229–247. https://doi.org/10.1016/j.pepi.2007.10.002

Dunlop, D., & Özdemir, Ö. (2001). Beyond Néel's theories: Thermal demagnetization of narrow-band partial thermoremanent magnetization. *Phys. Earth Planet. Int.*, *126*, 43–57.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (Second). Chapman & Hall/CRC, Boca Raton, FL.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Hoffman, K. A., Constantine, V. L., & Morse, D. L. (1989). Determinaton of absolute palaeointensity using a multi-specimen procedure. *Nature*, *339*, 295–297.

Königsberger, J. G. (1938). Natural residual magnetism of eruptive rocks. *Terrestrial Magnetism and Atmospheric Electricity*, *43*(3), 299–320. https://doi.org/10.1029/TE043i003p00299

Korte, M., Donadini, F., & Constable, C. G. (2009). Geomagnetic field for 0–3 ka: 2. a new series of time-varying global models. *Geochemistry, Geophysics, Geosystems*, *10*(6). https://doi.org/10.1029/2008GC002297

Kosareva, L. R., Kuzina, D. M., Nurgaliev, D. K., Sitdikov, A. G., Luneva, O. V., Khasanov, D. I., Suttie, N., & Spassov, S. (2020). Archaeomagnetic investigations in Bolgar (Tatarstan). *Stud. Geophys. Geod.*, *64*(2), 255–292. https://doi.org/10.1007/s11200-019-0493-3

Krása, D., Heunemann, C., Leonhardt, R., & Petersen, N. (2003). Experimental procedure to detect multidomain remanence during thellier–thellier experiments [Paleo, Rock and Environmental Magnetism 2002]. *Phys. Chem Earth (A/B/C)*, *28*(16), 681–687. https://doi.org/10.1016/S1474-7065(03)00122-0

Lisé-Pronovost, A., Mallett, T., & Herries, A. I. R. (2020). Archaeointensity of nineteenth-century scottish firebricks from a foundry in melbourne, australia: Comparisons with field models and magnetic observatory data. *Geological Society, London, Special Publications*, *497*(1), 27–45. https://doi.org/10.1144/SP497-2019-72

Muxworthy, A. R., Heslop, D., Paterson, G. A., & Michalk, D. (2011). A preisach method for estimating absolute paleofield intensity under the constraint of using only isothermal measurements: 2. experimental testing. *Journal of Geophysical Research: Solid Earth*, *116*(B4). https://doi.org/10.1029/2010JB007844

Nagata, T., Arai, Y., & Momose, K. (1963). Secular variation of the geomagnetic total force during the last 5000 years. *J. Geophys. Res.*, *68*(18), 5277–5281. https://doi.org/0.1029/j.2156-2202.1963.tb00005.x

Nagy, L., Williams, W., Muxworthy, A. R., Fabian, K., Almeida, T. P., Conbhuı, P. Ó., & Shcherbakov, V. P. (2017). Stability of equidimensional pseudo–single-domain magnetite over billion-year timescales. *Proc. Natl. Acad. Sci. U.S.A.*, *114*(39), 10356–10360. https://doi.org/10.1073/pnas.1708344114

Néel, L. (1949). Théorie du traınage magnétique des ferromagnétiques en grains fins avec applications aux terres cuites. *Ann. géophys.*, *5*, 99–136.

Paterson, G. A. (2011). A simple test for the presence of multidomain behavior during paleointensity experiments. *J. Geophys. Res.*, *116*. https://doi.org/10.1029/2011JB008369

Paterson, G. A., Biggin, A. J., Yamamoto, Y., & Pan, Y. (2012). Towards the robust selection of Thellier-type paleointensity data: The influence of experimental noise. *Geochem. Geophys. Geosyst.*, *13*(5). https://doi.org/10.1029/2012GC004046

Paterson, G. A., Muxworthy, A. R., Roberts, A. P., & Mac Niocaill, C. (2010). Assessment of the usefulness of lithic clasts from pyroclastic deposits for paleointensity determination. *Journal of Geophysical Research: Solid Earth*, *115*(B3). https://doi.org/10.1029/2009JB006475

Paterson, G. A., Tauxe, L., Biggin, A. J., Shaar, R., & Jonestrask, L. C. (2014). On improving the selection of thellier-type paleointensity data. *Geochem. Geophys. Geosyst.*, *15*(4), 1180–1192. https://doi.org/10.1002/2013GC005135

Pick, T., & Tauxe, L. (1993). Holocene paleointensities: Thellier experiments on submarine basaltic glass from the east pacific rise. *Journal of Geophysical Research: Solid Earth*, *98*(B10), 17949–17964. https://doi.org/10.1029/93JB01160

Santos, C. N., & Tauxe, L. (2019). Investigating the accuracy, precision, and cooling rate dependence of laboratory-acquired thermal remanences during paleointensity experiments. *Geochem., Geophys., Geosyst.*, *20*(1), 383–397. https://doi.org/10.1029/2018GC007946

Shaar, R., & Tauxe, L. (2013). Thellier_gui: An integrated tool for analyzing paleointensity data from thellier-type experiments. *Geochem. Geophys. Geosys.*, *14*, 677–692. https://doi.org/doi:10.1002/ggge.20062

Shaar, R., Ron, H., Tauxe, L., Kessel, R., & Agnon, A. (2011). Paleomagnetic field intensity derived from non-sd: Testing the thellier izzi technique on md slag and a new bootstrap procedure. *Earth and Planetary Science Letters*, *310*(3), 213–224. https://doi.org/10.1016/j.epsl.2011.08.024

Shaar, R., Ron, H., Tauxe, L., Kessel, R., Agnon, A., Ben-Yosef, E., & Feinberg, J. M. (2010). Testing the accuracy of absolute intensity estimates of the ancient geomagnetic field using copper slag material. *Earth and Planetary Science Letters*, *290*(1), 201–213. https://doi.org/10.1016/j.epsl.2009.12.022

Shaw, J. (1974). A new method of determining the magnitude of the paleomanetic field application to 5 historic lavas and five archeological samples. *Geophys. J. R. astr. Soc.*, *39*, 133–141.

Tanaka, H., Hashimoto, Y., & Morita, N. (2012). Palaeointensity determinations from historical and Holocene basalt lavas in Iceland. *Geophysical Journal International*, *189*(2), 833–845. https://doi.org/10.1111/j.1365-246X.2012.05412.x

Tauxe, L., Santos, C., Cych, B., Zhao, X., Roberts, A., Nagy, L., & Williams, W. (2021). Understanding non-ideal paleointensity recording in igneous rocks: Insights from aging experiments on lava samples and the causes and consequenes of 'fragile' curvature in arai plots. *Geochem. Geophys. Geosyst.*, *22*, e2020GC009423. https://doi.org/10.1029/2020GC009423

Tauxe, L., Shaar, R., Jonestrask, L., Swanson-Hysell, N. L., Minnett, R., Koppers, A. a. P., Constable, C. G., Jarboe, N., Gaastra, K., & Fairchild, L. (2016). PmagPy: Software package for paleomagnetic data analysis and a bridge to the magnetics information consortium (MagIC) database. *Geochem., Geophys., Geosyst.*, *17*(6), 2450–2463. https://doi.org/10.1002/2016GC006307

Tauxe, L., & Yamazaki, T. (2015). Paleointensities. In M. Kono (Ed.), *Geomagnetism* (2nd Edition, pp. 461–509). Elsevier.

Thébault, E., Finlay, C. C., Beggan, C. D., Alken, P., Aubert, J., Barrois, O., Bertrand, F., Bondar, T., Boness, A., Brocco, L., Canet, E., Chambodut, A., Chulliat, A., Coısson, P., Civet, F., Du, A., Fournier, A., Fratter, I., Gillet, N., … Zvereva, T. (2015). International Geomagnetic Reference Field: the 12th generation. *Earth Planets Space*, *67*(1), 79. https://doi.org/10.1186/s40623-015-0228-9

Thellier, E., & Thellier, O. (1959). Sur l'intensité du champ magnétique terrestre dans le passé historique et géologique. *Ann. Geophys.*, *15*, 285.

Valet, J.-P., Brassart, J., Le Meur, I., Soler, V., Quidelleur, X., Tric, E., & Gillot, P.-Y. (1996). Absolute paleointensity and magnetomineralogical changes. *Journal of Geophysical Research: Solid Earth*, *101*(B11), 25029–25044. https://doi.org/10.1029/96JB02115

Williams, W., & Dunlop, D. J. (1989). Three-dimensional micromagnetic modelling of ferromagnetic domain structure. *Nature*, *337*, 634–637.

Yamamoto, Y., & Hoshi, H. (2008). Paleomagnetic and rock magnetic studies of the sakurajima 1914 and 1946 andesitic lavas from japan: A comparison of the ltd-dht shaw and thellier paleointensity methods. *Phys. Earth and Planet. Inter.*, *167*, 118–143.

Yamamoto, Y., Tsunakawa, H., & Shibuya, H. (2003). Palaeointensity study of the hawaiian 1960 lava: Implications for possible causes of erroneously high intensities. *Geophys J Int*, *153*(1), 263–276.

Yamamoto, Y., & Yamaoka, R. (2018). Paleointensity study on the Holocene surface lavas on the Island of Hawaii using the Tsunakawa-Shaw method. *Front. Earth Sci.*, *6*. https://doi.org/10.3389/feart.2018.00048

Yu, Y., Tauxe, L., & Genevey, A. (2004). Toward an optimal geomagnetic field intensity determination technique. *Geochemistry, Geophysics, Geosystems*, *5*(2). https://doi.org/10.1029/2003GC000630

Chapter 3

Thermal Resolution of Unblocking Temperatures (TROUT): A method for "unmixing" multi component magnetizations

Brendan Cych[1], Matthias Morzfeld[1], David Heslop[2], Sarah Maher[1], Jeffrey Gee[1], Lisa Tauxe[1]

[1] Scripps Institution of Oceanography, University of California, San Diego

[2] Research School of Earth Sciences, Australian National University, Canberra, Australia

Chapter 3, is in preparation for submission to Geochemistry, Geophysics Geosystems as Cych, B., Morzfeld, M., Heslop, D., Maher, S., Gee, J., & Tauxe, L. Thermal Resolution of Unblocking Temperatures (TROUT): A method for "unmixing" multi component magnetizations (In Preparation for Submission). The dissertation author was the primary investigator and author of this paper.

Abstract

We present a model called Thermal Resolution Of Unblocking Temperatures (TROUT) that describes the process of thermal demagnetization as a function of temperature (or alternating field demagnetization as a function of coercivity). Some rocks contain multiple remanence "components", each of which preserve a record of a different magnetic field. The temperature ranges over which these remanence components demagnetize can overlap, making it difficult to determine their directions. TROUT models the unblocking temperature distributions of components in a thermal demagnetization, allowing for these distributions to overlap. It can be used to find the temperature at which the paleomagnetic direction changes, as well as ranges of temperatures where two directional components overlap, leading to curved demagnetization diagrams. When applied to specimens given multi-component TRMs in the laboratory, the TROUT method is able to estimate the temperature at which the pTRMs were acquired to within one temperature step, even for specimens with significant overlap. The TROUT model has numerous applications: knowing the temperature at which the direction changes is useful for experiments in which the thermal history of a specimen is of interest (e.g. emplacement temperature of pyroclastic deposits, re-heating of archaeological artifacts, reconstruction of cooling rates of igneous bodies). The ability to determine whether a single component or multiple components are demagnetizing at a given temperature is useful for choosing appropriate ranges of temperatures to use in paleointensity experiments. Finally, the width of the range of temperatures where two components overlap may be useful for inferring the domain state of specimens.

Plain Language Summary

Rocks can contain records of multiple magnetization events or "components". Paleomagnetists often attempt to separate these components by heating their specimens to progressively higher temperatures, and cooling in the absence of a magnetic field to remove part of the magne-

tization. Paleomagnetists are often interested in the temperature range over which one of the magnetization components is demagnetized, as it can tell them about the thermal history of the specimen (e.g. the temperature it was reheated to in the past). Unfortunately, two components can demagnetize with overlapping temperature ranges, leading to ambiguity in the temperatures of interest. In this paper we create a model called Thermal Resolution of Unblocking Temperatures (TROUT) which can account for this overlap and find the temperatures that paleomagnetists are interested in. We tested TROUT on data sets where we knew the temperatures the specimens were reheated to, and it accurately estimated those temperatures. These test cases indicate that TROUT will be useful for many different applications where paleomagnetists want to know about the thermal history of a specimen.

3.1 Introduction

Paleomagnetists use the magnetizations of rocks to make inferences about the Earth's ancient magnetic field. A rock that cools in a magnetic field obtains a magnetization in the same direction as that field. In theory, a paleomagnetist could take an oriented paleomagnetic specimen and measure its magnetization to obtain the direction of the ancient magnetic field. Unfortunately, many rocks have experienced multiple magnetization events in different field directions, causing them to have a total magnetization which is the sum of multiple magnetic "components" acquired in these different fields. To separate out the individual components, paleomagnetists demagnetize specimens in a stepwise fashion by heating them at progressively higher temperatures and cooling them in the absence of a magnetic field to demagnetize them. Strong alternating magnetic fields may also be used to demagnetize a specimen. Experiments like these that involve some form of stepwise demagnetization are ubiquitous in paleomagnetic studies.

Data from a thermal or alternating field (AF) demagnetization experiments are typically plotted on the Zijderveld diagram (Zijderveld, 1967), whereby the X coordinate of the direction is

plotted against the Y and Z coordinates (see Figure 3.1) allowing visualization of inherently three dimensional behavior on a two-dimensional plot. Data with a single magnetic component will plot as a pair of straight lines trending toward the origin. For a two-component magnetization, as the specimen is demagnetized the data will plot as straight lines which do not trend toward the origin at low temperatures (fields). At higher temperatures (fields), the direction of the lines will change, with a trend toward the origin. The directions of these lines are parallel to the direction of the two fields in which the specimen was magnetized. An example of this two component behavior for a simulated thermal demagnetization expriment is shown in Figure 3.1a. In some cases, instead of distinct lines with sharp boundaries (as in Figure 3.1a), the change in direction appears to happen over several temperature steps, leading to a "curved" appearance in the Zijderveld plot, as is shown in Figure 3.1b. This behavior makes multi-component Zijderveld plots more difficult to interpret, as the demagnetization experiment has failed to completely separate the two components.

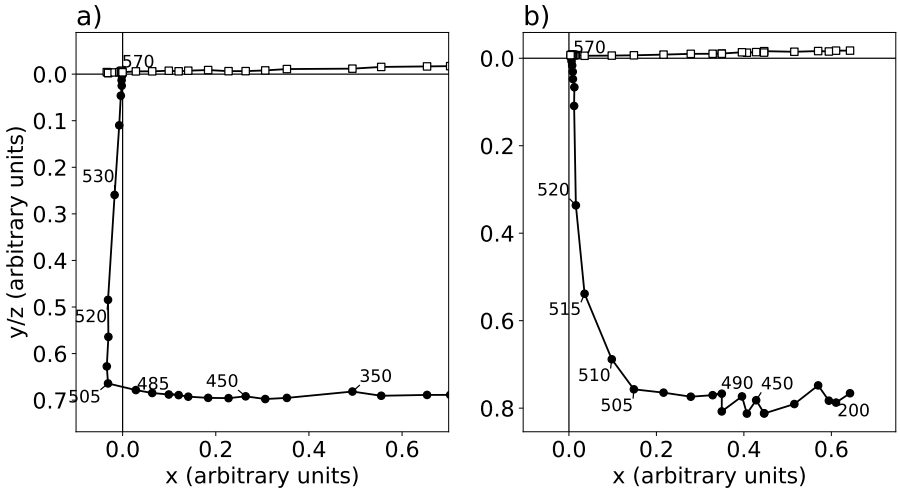

Figure 3.1. Examples of Zijderveld plots. Solid (open) symbols: x versus y (z). a) A specimen with two relatively straight, resolvable magnetization components in perpendicular directions. b) Another specimen, magnetized in the same way as in a), but displaying two components with overlapping unblocking temperatures. The resulting Zijderveld plot has a "curved" appearance.

Demagnetization experiments are able to separate multiple magnetic components because rocks are composed of many nano- to micrometer scale magnetic particles. Louis Néel proposed

a theory (Néel, 1949) that these magnetic particles can be magnetized in one of two directions, and some amount of energy is required to "flip" their magnetic moment from one 'easy' direction to the other. Partial demagnetization of a specimen is achieved by randomizing the moments of particles with energy barriers below the thermal or AF treatment step. In a thermal demagnetization experiment, this occurs at a particular temperature, known as the "unblocking temperature" (T_{ub}). In AF demagnetization, the field provides the energy to flip the moment of the particle. Néel theory suggests that a magnetic component acquired by heating to a particular temperature, the blocking temperature (T_b), and cooling in a field would be removed by thermally demagnetizing that specimen to the same temperature $T_b = T_{ub}$. However, Néel theory applies only to uniformly magnetized or "single domain" magnetic particles. Micromagnetic modeling has shown that only small particles on the scale of 10s of nanometers (Williams and Dunlop, 1989) are magnetized in this way. For larger particles, which occupy vortex and multi-domain states, differences between the temperature at which particles block and unblock have been observed (Dunlop and Özdemir, 2001). This "overlap" of the unblocking temperatures between components can lead to the curvature observed on the Zijderveld plot in Figure 3.1b. Overlapping of the unblocking temperatures may also occur if two components are not magnetized in the same way, e.g. a specimen containing both a thermally acquired and a chemically acquired remanence. Additionally, specimens which moved during cooling, or which cooled slowly as the direction of the Earth's magnetic field changed, may have curved Zijderveld plots due to the rotation of the specimen relative to the field.

Previous studies have dealt with isolating individual directions from multi-component magnetizations with overlapping unblocking temperatures. Hoffman and Day, 1978 created a generalized method for separating overlapping multi-component directions by fitting great circle paths to the vector differences of Zijderveld data. Kirschvink, 1980 created a method for obtaining best fitting lines or planes to paleomagnetic data using principal component analysis, and McFadden and McElhinny, 1988 extended this methodology to find a common paleomagnetic direction using the intersection of lines and planes from different paleomagnetic specimens.

None of these methods quantitatively provide information about the unblocking temperature ranges over which different components demagnetize when the unblocking temperatures overlap. However, this information is desirable in several use cases, for example, when performing paleodirectional/intensity analysis, paleomagnetists attempt to choose a range of temperatures on the Zijderveld plot where a single remanence component is being demagnetized. The blocking temperature distribution of each component is also useful for determining the temperature at which a pyroclastic flow came to rest (also known as the emplacement temperature), or to find the temperature a host rock was reheated to by an intruding igneous dike (a form of "baked contact test", as per Everitt and Clegg, 1962).

In this paper, we present a method for treating demagnetization data which can obtain the direction, relative magnitude, and unblocking temperature (or coercivity) distribution for each component in a demagnetization experiment. If the components are easily resolvable, then the unblocking temperature (coercivity) distributions will be completely separate, which is shown in Figure 3.2c, leading to a Zijderveld plot like the one in Figure 3.2a. If the two components are hard to resolve then the distributions will overlap with one another, as is shown in Figure 3.2d, causing the curved Zijderveld plot in Figure 3.2b. We call our approach "Thermal Resolution of Unblocking Temperatures" (TROUT). The construction of this model is explained in Section 3.2 and we explain how to fit the model to real demagnetization data in Section 3.2.2.

One advantage that TROUT has over traditional methods of analyzing demagnetization data is the ability to obtain information about the range of unblocking temperatures for a particular component. For example, TROUT can obtain an estimate of the temperature to which a specimen was remagnetized, as well as the range of temperatures over which two components overlap (see Section 3.2.4). In Section 3.3 we apply the TROUT method to several data sets that make use of this thermal information. In Section 3.3.1 we extend the work of Maher et al., 2021 to test whether their specimens have reproducible unblocking temperatures that could be used to estimate a quantitative cooling rate for fast spreading lower oceanic crust. In Section 3.3.2, we remagnetized specimens from Tauxe et al., 2021 to test TROUT's ability to

obtain the remagnetization temperatures and directions for a set of specimens which may have different blocking and unblocking temperature distributions. And, in Section 3.3.3, we replicate the experimental results of Kent and Gee, 1994 in which specimens were given a secondary magnetization at low temperatures to test whether high blocking temperatures observed in previous thermal demagnetization experiments were a result of specimens undergoing chemical alteration during heating. We discuss our findings in Section 2.4.

3.2 Methodology

3.2.1 Modeling magnetization as a function of demagnetizing energy

Our goal is to model thermal demagnetization in a specimen with multiple components, but it is simpler to first consider the uni-vectorial case. A specimen with a single magnetic component will have a magnetization which reduces in magnitude but maintains a constant direction as it is demagnetized. If we call the energy used to demagnetize the specimen x, then we can come up with an expression for the magnetization remaining as a function of x, $\vec{M}(x)$. We can call the field direction $\hat{B}$ (a unit vector) and the original magnitude of the magnetization before demagnetizing c. We then need an expression for the amount of the magnetization that survives after demagnetizing to x. It should vary between 1 at $x = 0$ and 0 at some maximum value for x. We can call this expression the "demagnetization function" $F(x)$. The full equation of the magnetization is the product of these three terms:

$$\vec{M}(x) = c\hat{B}F(x). \tag{3.1}$$

Alternatively, a specimen may record multiple (re)magnetization events, and so may have multiple (K) components. In this case, the magnetization is the sum of the magnetizations of the components, i.e.

$$\vec{M}(x) = \sum_{k=1}^{K} c_k \hat{B}_k F_k(x). \tag{3.2}$$

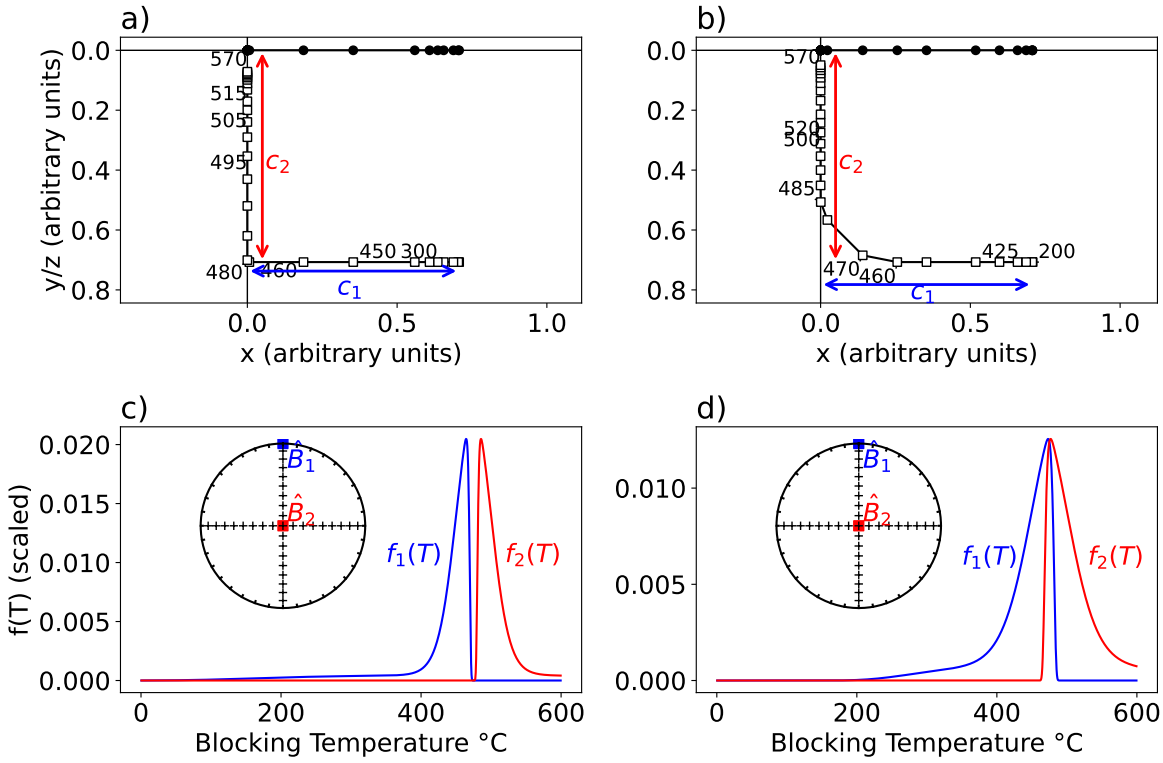

Figure 3.2. Simulated two-component thermal demagnetization data subjected to TROUT. For more information on the different parameters referenced in this figure, see Section 3.2. Panels a) and b): Simulated Zijderveld plots with the magnitude of each component (c_k) indicated by blue and red arrows. Solid (open) symbols are X,Y (X,Z) components. Panels c) and d): unblocking temperature density functions $f(T, \mu_k, s_k, p_k, q_k)$ for each component obtained from the TROUT method. Left column: Example without overlapping unblocking temperatures. Right column: Example with overlapping unblocking temperatures. High (low) temperature components indicated in red (blue). In d) the unblocking temperature density clearly displays overlap between the unblocking temperature functions. Insets: Equal area plots of the directions $\hat{B}_k$ of each component. Center of diagram is the Z (vertical) direction, top (right) edge is the X (Y) direction. B_1 (B_2) was acquired along the X (Z) specimen direction.

There is no requirement that any two demagnetization functions F_i and F_j operate over independent temperature ranges. Because of this, we can use this model to simulate specimens with overlapping unblocking temperatures, provided we have some expression for F.

The way to construct the problem of mixed TRM components is very similar to the problem of unmixing Isothermal Remanent Magnetization (IRM) components (e.g., Egli, 2003). These experiments use an alternating field to demagnetize a magnetization acquired in a strong

magnetic field at room temperature. The magnitude of the magnetization at each demagnetization step is used to infer several "components" which have different coercivity distributions. In the IRM unmixing literature, a "component" usually refers to a subpopulation of magnetic particles interpreted as representing a particular magnetic mineral, whereas in TRM unmixing, a component refers to a set of particles magnetized in a particular geomagnetic field. Both types of components require a flexible function to model the wide range of demagnetization curves observed in real data.

For the purposes of IRM unmixing, Egli, 2003 created a four parameter Skew Generalized Gaussian (SGG) Distribution given by:

$$f(x,\mu,s,p,q) = \frac{1}{2^{1+1/p}s\Gamma(1+1/p)} \frac{|qe^{qz}+q^{-1}e^{z/q}|}{e^{qz}+e^{z/q}} \exp\left[-\frac{1}{2}\left|\ln\left(\frac{e^{qz}+e^{z/q}}{2}\right)\right|^{p}\right], \quad (3.3)$$

where $z = \frac{x-\mu}{s}$ and Γ is the gamma function; the scalar parameters μ, s, p, q determine the "shape" of the distribution. To first order, μ controls the location of the distribution, s controls the scale, p controls the kurtosis and q controls the skewness, although interactions between these parameters mean that they do not solely affect these properties of the distribution. SGG distributions are able to approximate a wide range of other distributions by selecting the parameters μ, s, p, q appropriately. For this reason, SGGs are useful for our purposes. In the work of Egli, 2003, the SGG distribution is scaled and fit to the negative derivative of the magnitude of the demagnetization data with respect to x. In the case of IRMs, this quantity represents the coercivity distribution, and in the TRM case, the unblocking temperature distribution of the magnetization. The demagnetization function is therefore given by:

$$F(x,\mu,s,p,q) = 1 - \int f(x,\mu,s,p,q)\, dx. \quad (3.4)$$

In Figure 3.2 we show simulated thermal demagnetization experiments for the case of non-overlapping unblocking temperatures (Figure 3.2a, c) and the case of overlapping unblocking

temperatures (Figure 3.2b, d). In the following sections, we explain how we can use the TROUT approach to find the unblocking temperature distributions and directions that provide the best fit to thermal demagnetization data, allowing us to quantify the temperature that separates the two components and the amount of overlap in unblocking temperatures between two components.

3.2.2 Fitting TROUT to data

The TROUT model in Equation 3.2 describes magnetization as a function of temperature. The model includes several "model parameters" that define the unblocking temperature distributions, directions and relative magnitudes of each component. We need a framework for fitting the TROUT model to the data by selecting model parameters that lead to model outputs that are comparable to the data. We do the model fitting via optimization of a Bayesian posterior distribution, which synthesizes prior information about the parameters with model-data misfit (see, e.g., T05). Specifically, the posterior distribution is proportional to the product of a likelihood that describes the model-data misfit, and a prior distribution that specifies additional information about the model parameters. By finding the maximum value of the posterior distribution, TROUT thus finds the model that best fits the data, while simultaneously satisfying prior constraints and, therefore, avoids overfitting the data. We now describe how we set up the likelihood and prior distributions and how to solve the resulting optimization problem.

Likelihood

We start with the likelihood that specifies how we interpret model-data misfit. The unknowns of the TROUT model in Equation 3.2 are $c, \hat{B}$ and the function $F(x)$, which is parameterized via Equations 3.3 and 3.4, adding the unknowns μ, s, p and q. To simplify notation, we collect all unknowns in a vector $\theta = (c, \hat{B}, \mu, s, p, q)$. Then the model data misfit is usually defined by the quadratic:

$$\text{Misfit} = \frac{1}{2}(\vec{M}(x,\theta)_{pred} - \vec{M}_{obs})^T C^{-1} (\vec{M}(x,\theta)_{pred} - \vec{M}_{obs}), \tag{3.5}$$

where $\vec{M}_{pred}(x)$ is given by Equation 3.2, $\vec{M}_{obs}(x)$ is our observed magnetization, and C is a matrix that describes errors in the data. The misfit and likelihood are connected via an exponential

$$p(M_{obs}|\theta) = (\det 2\pi C)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\vec{M}(x,\theta)_{pred} - \vec{M}_{obs})^T C^{-1}(\vec{M}(x,\theta)_{pred} - \vec{M}_{obs})\right). \quad (3.6)$$

To specify the error covariance matrix C, we recall that when we make paleomagnetic measurements, a significant proportion of measurement error comes from misorientation of specimens in the sample holder (see for example Paterson et al., 2012 for a discussion) which is an angular misfit. This should be correctly reflected by the choice of C in the likelihood and for that purpose we adapt the work of Holme and Bloxham, 1996, who devised an error matrix for satellite data, where there are similarly two kinds of noise, measurement errors as well as "attitude errors," caused by misorientation. Specifically, we use the error covariance matrix:

$$C = I(\sigma^2 + |\vec{M}|^2\psi^2) - \vec{M}\vec{M}^T\psi^2, \quad (3.7)$$

where ψ (an angle in radians) and σ (a constant measurement uncertainty for all data) are unknown parameters that define the noise distribution. Note that the fact that the error covariance matrix depends on unknown parameters can be addressed by simply appending these unknowns to the model parameter vector $\theta = (c, \hat{B}, \mu, s, p, q, \sigma, \psi)$ and subsequently defining the posterior jointly over all unknowns (see below). Finally, if we have a set of N measurements, the likelihood is the product of the likelihoods for each measurement:

$$P(\vec{M}_{obs}|\theta) = \prod_{i=1}^{N} P(\vec{M}_{obs}(x_i)|\theta), \quad (3.8)$$

where we emphasize again that the unknowns include the unknown parameters defining the measurement noise. Two examples of this noise model and the resulting likelihoods are shown in Figure 3.3.

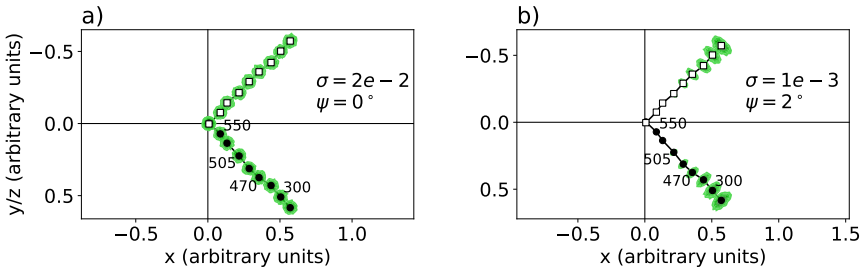

Figure 3.3. Examples of the likelihood distribution (green shaded areas) for an idealized single component Arai plot (Zijderveld data). The TROUT method prioritizes solutions which have maximum likelihood, i.e., are closest to the center of each green area for each data point. a) shows the likelihood distribution with constant noise $\sigma = 2e-2$, $\psi = 0^\circ$, whereas b) shows the likelihood distribution with predominantly angular noise $\sigma = 1e-3$, $\psi = 2^\circ$. In b) the angular noise causes the likelihood distribution to become spread out in a plane perpendicular to the direction of the Zijderveld data.

Table 3.1. Lower and upper bounds used in calculating the prior distributions of TROUT model parameters.

Parameter	c	μ	s	p^*	q^*	ψ	σ
Lower bound	0	$x_{\min}$	0	0	-5	0	0
Upper bound	$2\max(\lvert\vec{(M)}_{obs}\rvert)$	$x_{\max}$	$\frac{10}{6}(x_{\max}-x_{\min})$	5	5	$\frac{\pi}{18}$	∞

Prior Distribution

A prior distribution incorporates all the information we may have about parameters before any data are collected. Priors impose constraints on the model parameters and ameliorate issues of non-uniqueness by focusing on solutions that are in line with any of this additional information. We now describe how we construct the prior in TROUT.

First, we impose upper and lower bounds on all parameters, which are listed in Table 3.1. Here, we work with re-scaled variables p^* and q^* defined as

$$q^* = \cot(\frac{\pi}{2}q), \quad p^* = \ln(p). \tag{3.9}$$

We do the re-scaling of q because SGG distributions with q values of both 1 and -1 have zero

skewness, and the skewness of the distribution tends towards infinity or negative infinity as q tends towards zero from either side. For the rescaled variable, $q^* = 0$ corresponds to an SGG with zero skewness, and $|q^*| \to \infty$ implies that the skewness tends to infinity, with the direction being determined by the sign of q^*. The log-transformation on p is for convenience.

For the noise parameter σ, we use an "improper prior" $P(\sigma) \propto 1/\sigma$ which is a popular choice and equivalent to a uniform prior over $\log \sigma$, hence easily enforcing the constraint that $\sigma > 0$. For the parameter ψ we impose a uniform prior $P(\psi)$, which means that all values between the lower and upper bounds have equal probability. The parameter $\hat{B}$ is a unit vector and its prior distribution is uniform on the unit sphere. Moreover, σ, $\hat{B}$ and ψ are independent of each other and of all other parameters.

For the model parameters c, μ, s, p, q we define a more complex prior that promotes non-overlapping distributions. The reason is that if two models fit the data equally well, we expect the one with the smallest overlap between unblocking temperatures to be the correct one (in the majority of cases, see also below). We want to find a prior that gives us the least overlapping solution that fits the data well.

To quantify "overlap" between two distributions, we use a modified version of the non-parametric overlap coefficient of Inman and Bradley, 1989. For any two distributions, this coefficient is given by:

$$\eta(i,j) = \int_{-\infty}^{\infty} \min\left[f_i(x), f_j(x)\right] dx. \tag{3.10}$$

The TROUT modification accounts for the scaling of the distributions:

$$\eta(i,j) = \frac{1}{\min(c_i, c_j)} \int_{-\infty}^{\infty} \min\left[c_i f_i(x), c_j f_j(x)\right] dx. \tag{3.11}$$

The coefficient $\eta(i,j)$ gives a value between 0 (no overlap between f_i and f_j) and 1 (total overlap, one distribution entirely contained within the other).

Examples of the overlap coefficient for pairs of distributions are given in Figure 3.4. To

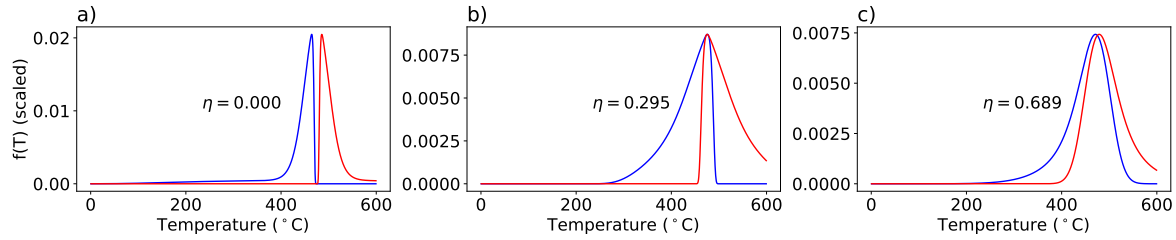

Figure 3.4. Examples of the overlap coefficient (η) for three different pairs of SGG distributions. In a), there is no overlap between the two distributions, so $\eta = 0$. In b) there is a significant overlap between the two distributions, so $\eta = 0.295$. In c) the two distributions are highly overlapping, so $\eta = 0.689$

penalize pairs of distributions with strongly overlapping unblocking temperatures in our model, we use the informative prior distribution $P(\eta) \sim \text{Beta}(1,10)$. For specimens with more than two components, we compute η for every pair of components. Since η is a function of the parameters c, μ, s, p, q (via f in Equation 3.3), $P(\eta)$ defines the prior distribution on c, μ, s, p, q, and we note that c, μ, s, p, q are not independent of each other, but they are independent of σ, $\hat{B}$ and ψ. The prior over all parameters is thus

$$P(c, \hat{B}, \mu, s, p^*, q^*, \psi, \sigma) = P(\sigma)P(\hat{B})P(\psi)P(c, \mu, s, p^*, q^*) \tag{3.12}$$

Finally, we note that, in some cases, (e.g., when two components that are not TRMs are thermally demagnetized), a model with more overlap may be correct, which is in contradiction to our prior. In this case, however, information about the unblocking temperature distributions may not be particularly helpful, and other methods such as the plane fitting method of McFadden and McElhinny, 1988 exist for obtaining directions. For this reason, promoting little overlap by construction of the prior is justified in the majority of cases we consider.

Estimating the maximum of the posterior distribution

The posterior distribution is the product of likelihood and prior, specifically,

$$P(c, \hat{B}, \mu, s, p^*, q^*, \sigma, \psi | \vec{M}_{obs}) = P(\vec{M}_{obs} | \theta)P(\theta), \tag{3.13}$$

where the likelihood $P(\vec{M}_{obs}|\theta)$ is as in Equation 3.6 and $P(\theta)$ is shorthand notation for the prior distribution (recall that $\theta = (c, \mu, s, p^*, q^*, \hat{B}, \sigma, \psi)$). Specifically, we can write $P(\theta) = P(c, \mu, s, p^*, q^*)P(\hat{B}, \psi)P(\sigma)$, where $P(\hat{B}, \psi)$ is a uniform prior, $P(\sigma)$ is the improper prior, and where $P(c, \mu, s, p^*, q^*)$ enforces the non-overlap condition via a prior in η in Equation 3.11.

To obtain unblocking temperature distributions, directions and magnitudes of each of the components, we find the set of model parameters with maximum posterior probability. To do this, we first compute an informed guess for the maximum posterior solution, assuming that there is no overlap between the components. Details of how we compute this best guess are given in 3.6. To obtain the set of model parameters that maximize our posterior, we use the Broyden-Fletcher-Goldfarb-Shanno (BFGS, Nocedal, pp. 136) method of optimization, which uses the gradient of the posterior distribution to find a local maximum. We compute the gradient via finite differencing. Because the BFGS algorithm is designed to find a local maximum for a posterior distribution, and our posterior distribution may have multiple maxima, we run the BFGS multiple times on randomly generated sets of model parameters close to our best guess. A similar scheme is used by Zhao et al., 2018 to initialize optimizations for IRM unmixing. The BFGS method considers itself to have found a "maximum" when the gradient of the function reaches some minimum threshold, which may not necessarily be zero. To improve on the results of the BFGS optimization, we pick the optimization that leads to the largest posterior and the run a second optimization, using the gradient free algorithm of Nelder and Mead, 1965 to get even closer to the (global) maximum of the posterior distribution than the gradient-based BFGS method.

3.2.3 Rescaling of data

To be able to more easily fit SGG distributions to IRM data, Egli, 2003 scaled coercivities x by a power β; $x' = x^{\beta}$. This is done because coercivity distributions may often be highly skewed toward values near $x = 0$, but may not be well approximated by a log distribution. Egli performed this scaling such that the IRM as a function of x' "has symmetry closest to a hyperbolic

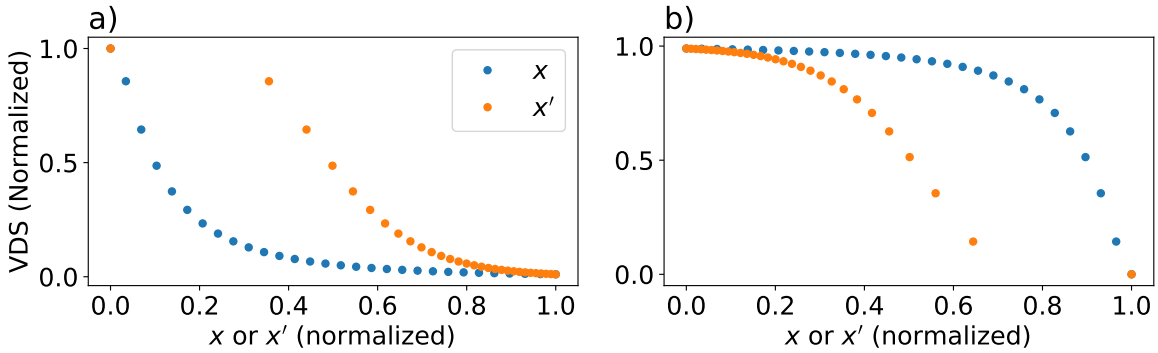

Figure 3.5. Examples of rescaling data from x (blue) to x' (orange) when a) $\gamma < 0$ and b) $\gamma > 0$. The rescaling improves the fitting of distributions to the data by increasing the spacing between points where the magnetization varies strongly and reducing spacing between points where the magnetization does not vary.

tangent distribution".

In the TROUT method, we employ a similar scaling of data to fit SGG distributions better. In contrast to coercivity distributions, unblocking temperatures may be highly skewed such that most of the unblocking occurs close to the Curie temperature. To account for the family of possible cases, we scale the thermally unblocked data by a constant γ to produce a scaled version of our demagnetization energy $x'(x,\gamma)$:

$$x' = \begin{cases} x^{\exp(-|\gamma|)} & \gamma < 0 \\ x & \gamma = 0 \\ x'_{max} - (x_{max} - x)^{\exp(-|\gamma|)} & \gamma > 0. \end{cases} \tag{3.14}$$

To know which value of γ to use, we minimize the misfit to a hyperbolic tangent function, by minimizing:

$$[\mathrm{VDS}(x') - (1 - \tanh(\alpha x'/x'_{max} - \alpha/2))/2]^2, \tag{3.15}$$

where $\mathrm{VDS}(x')$ is the vector difference sum ((Tauxe and Staudigel, 2004)) of all $M(x \geq x')$. Here, α is a free parameter that controls the width of the hyperbolic tangent function. (see Figure 3.5

for an example). In this case, we fit the SGG distributions to the scaled data, to obtain a function $F'(x')$. This can then be converted back to $F(x)$ using a change of parameters.

3.2.4 Understanding results from the TROUT model

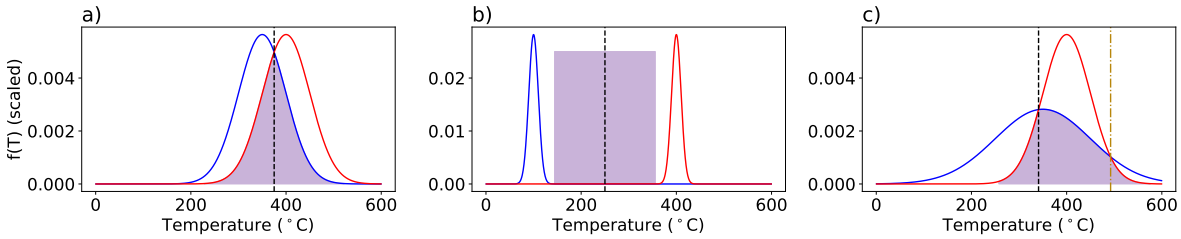

Figure 3.6. Figure illustrating the "crossover temperature" (vertical line) and "mixed region" (MR, purple shaded region) for three different pairs of unblocking temperature distributions. a) Shows the simplest case, where the two distributions overlap. b) Shows another common case, where the two distributions do not overlap, and the MR is interpreted as being in between the two distributions, and the crossover temperature is in the center of the MR. c) Shows a rare case, where both distributions overlap in such a way that there are two crossover temperatures. The crossover temperature where the unblocking temperature distributions are largest is preferred in this case.

From the TROUT model we have defined, we want to obtain two pieces of information. Firstly, we want to have an estimate of the temperature a specimen was reheated to or the temperature the specimen reached as the field changed. Secondly, we want to know the range of temperatures where only one component is unblocking, e.g. for use in a paleointensity experiment where we want to retrieve a single field strength. To assess which components are unblocking at a given temperature, we define a parameter π for each component. For the k_{th} component, π_k is the ratio of the unblocking temperature distribution to the sum of all the unblocking temperature distributions, or: $\pi_k(x) = \frac{c_k f_k(x)}{\sum_{k=1}^{K} c_k f_k(x)}$. Thus for the k_{th} component, $\pi_k \approx 1$ when that component is unblocking. If $\pi_k \approx 0$ then a different component is unblocking. If more than one component is unblocking at the same time, then $0 < \pi_k < 1$.

For any two components, we define a "Crossover Temperature" (CT) at which $\pi_i = \pi_j$, which represents the temperature to which a specimen was reheated (if the blocking and

unblocking temperatures are equal). In addition to CT, we can also define a "mixed region" (MR), defined as being the region where $\max[\pi_i, \pi_j] < 0.95$ and $f_i, f_j > 0$.

Outside the mixed regions, a single component is unblocking. If there is no overlap between the two components, then we define the crossover temperature slightly differently. We first find temperature ranges where $f_i, f_j > 0$ and $\max[\pi_i, \pi_j] > 0.5$. We define the MR as being the range of temperatures between those two points and the CT as being the center point of this range. Conversely, some distributions may have two crossover temperatures as they cross one another twice. If a single crossover temperature is required by our method, we pick the one for which $f_i = f_j$ is highest. Examples of the crossover temperature and mixed region in different cases are shown in Figure 3.6.

3.3 Applications

We now illustrate how to use TROUT by applying it to three different data sets and discussing the TROUT procedure and results in detail. The first data set is from a set of specimens from the Pito Deep in the Pacific Ocean remagnetized in the laboratory by Maher et al., 2021. The second data set is from specimens exhibiting 'fragile curvature' described by Tauxe et al., 2021. And the third set is a from Mid Ocean Ridge Basalt (MORB) specimens initially described by Kent and Gee, 1994.

3.3.1 Pito Deep Specimens

Maher et al., 2021 documented numerous examples of samples with multicomponent remanences from tectonic exposures of the lower oceanic crust at Pito Deep (∼23S, 112W). These multicomponent remanences provide information on the thermal history of the lower crust, as they imply that the rocks cooled over a long time during which the Earth's magnetic field reversed. Because the ages of these reversals are known, the authors were able to qualitatively determine that the crust cooled more slowly than previously predicted from solely conductive cooling models. The temperatures of the remanence "components" are representative of a

particular time interval, and so were used as an estimate of cooling rate for this lower crustal section in the doctoral dissertation of S. Maher (Maher, 2021). Our analysis above using the TROUT method corroborates the significance of these temperature estimates and additionally provides uncertainty estimates."

To explore the meaning of remagnetization temperatures, specimens from the study of Maher et al., 2021 were given two or three approximately orthogonal TRMs at temperatures close to the apparent reversal temperatures in the original thermal demagnetization experiments. If these temperatures are reproducible when thermally demagnetizing the newly acquired pTRMs, then this is evidence that the original NRM components unblock over independent temperature ranges and so can be used to obtain a quantitative cooling rate. Here, we apply the TROUT method to obtain the unblocking temperature distributions, and the different CTs and MRs. Note that orthogonal magnetizations of the TRMs were not always possible, due to the specimens not being perfect cubes.

We fit the TROUT model to the 72 specimens given pTRMs by Maher et al., 2021. Some examples of fitted models are given in Figure 3.7 and the full set of crossover temperatures and mixed regions is shown in Figure 3.8. To first order, the crossover temperatures are good estimates of the original temperatures at with the pTRMs were imparted, with almost all CTs being within ±1 temperature step of the expected value and no CT being more than two steps from the expected value. The vast majority of our MR ranges have a full width that spans two or fewer temperature steps, but a significant number are wider than this. In some cases, this is due to curvature observed in the directions, where there may be overlap in the unblocking temperature distributions of the two components (see Figure 3.7c-f). In a few cases, there appears to be overlap in the unblocking temperature distribution in regions where the directions are a straight line. This seems to occur in situations where the ratio of two unblocking temperature distributions is constant over a wide range, leading to an overestimate of the overlap (see e.g. Figure 3.7g,h).

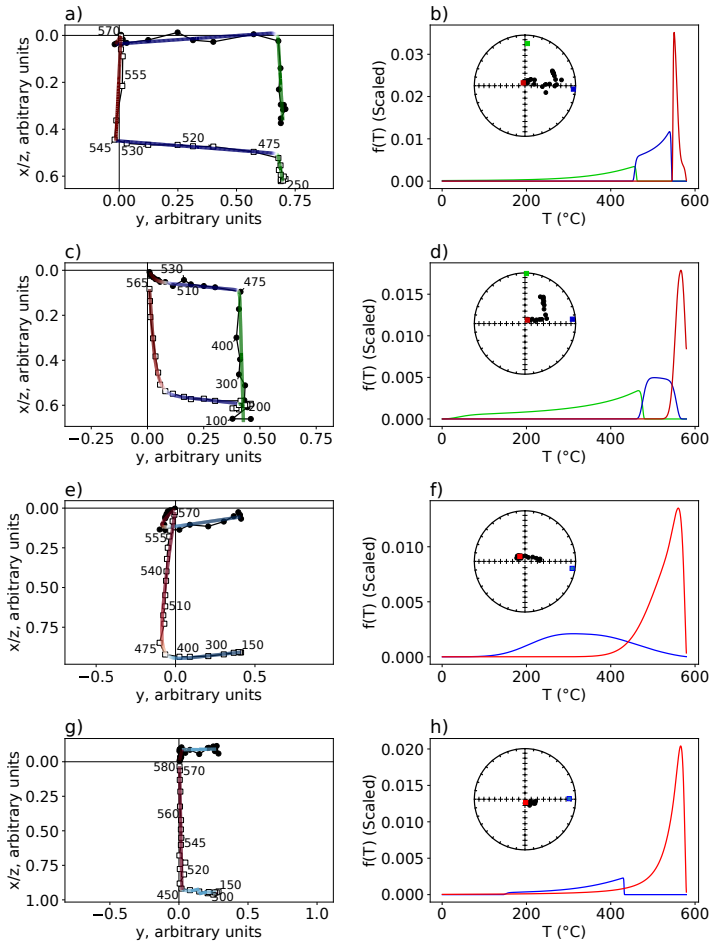

Figure 3.7. Results from the Pito Deep data set from specimens (top row) PD036a1, (second row) PD135a2, (third row) PD014d2 and (bottom row) PD142a2. Left column: Zijderveld plots of the data, with the model fit to the data superimposed. Colors represent the π ratio (see Section 3.2.4) indicating the dominant component, with red indicating the highest temperature TRM, blue indicating the moderate temperature TRM and green indicating the low temperature TRM. Center column: unblocking temperature distributions of each component. Right hand column: Equal area plots of the Zijderveld data and the directions of each component. Specimen PD036a1 contains three easily resolvable components, which appear as straight lines with sharp corners on the Zijderveld plot (a) which correspond to three unblocking temperature distributions with little overlap in b). Specimens PD135a2 and PD014d2 have two or three component magnetizations with curved directions between the high and moderate (red and blue respectively) component on the Zijderveld plots (d,g), these cause overlapping unblocking temperature distributions (e,h). Specimen PD142a2 appears to have overlapping unblocking temperatures (k), despite having no curved directions on the Zijderveld plot (j). This is likely because the ratio between the two components remains fairly stable over most of the range of the low temperature component.

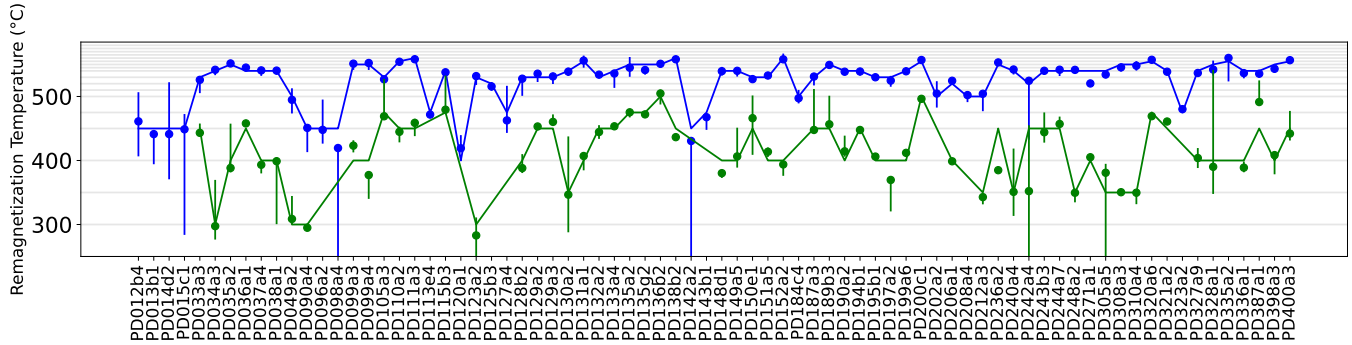

Figure 3.8. Estimates of the remagnetization temperatures from the Pito Deep data set using TROUT. Blue and green lines represent the temperatures at which the second and third pTRMs were applied. Blue and green dots represent our estimated crossover temperatures for the same specimens, and blue and green error bars represent our MRs for each specimen. The grey horizontal lines represent the set of temperature steps used to thermally demagnetize these specimens. In almost all cases, our crossover temperatures reproduce the original temperatures to which the TRM was heated within ±1 step. The majority of the MRs are also confined to ±1 step although there are several which are indicative of overlapping unblocking temperatures in these specimens.

Although uncommon, there are a few of these cases in the Pito Deep dataset, and so the MR should be taken as an upper bound on the range of overlapping unblocking temperatures. The highly reproducible crossover temperatures and generally narrow MRs estimated for the Pito Deep data indicate that the majority of these specimens would be appropriate for use in estimating a cooling rate. Histograms of the CTs relative to the actual remagnetization temperatures, and the MR ranges are shown in Figure 3.9.

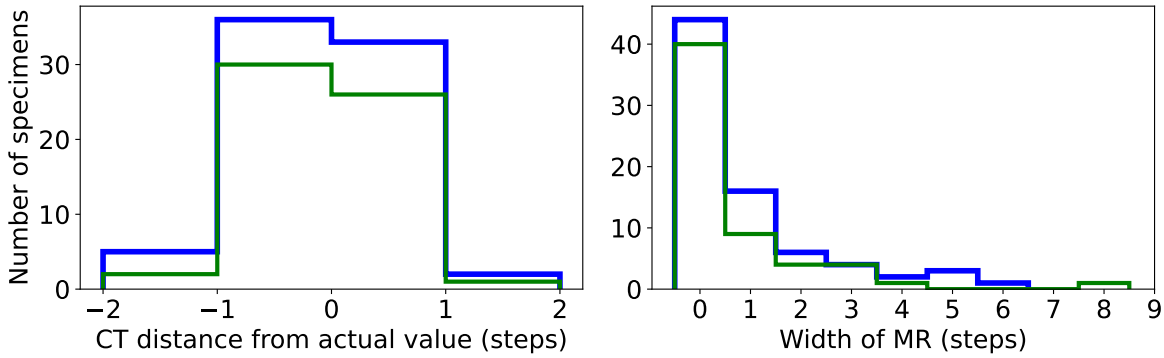

Figure 3.9. Plot showing histograms of a) distance of CTs from the known pTRM acquisition temperature for each specimen, measured in temperature steps, b) number of temperature steps that the MR spans for each specimen. Blue and green represent the peak temperatures of the second and third pTRM respectively. The CT is generally less than ± 1 step from the expected temperature for both components. The MR is also generally narrow ($\leq$ 2 steps) with some exceptions. The higher temperature pTRM has a slightly wider average MR, likely due to the 5 $^{\circ}$C temperature steps in this range.

3.3.2 Fragile Curvature Specimens

Santos and Tauxe, 2019 characterized a set of paleomagnetic specimens which exhibited a range of behaviors in paleointensity experiments. In some specimens, the Arai plots (Nagata et al., 1963) in the original experiments were quite curved, but became straight when given a fresh TRM and subjected again to the paleointensity experiment. Then, Tauxe et al., 2021 showed that these specimens became more curved when a laboratory TRM was allowed to 'age' over a period of several years. This behavior was called 'fragile' curvature. We applied two component TRMs to specimens from the same samples, expecting that some of the specimens

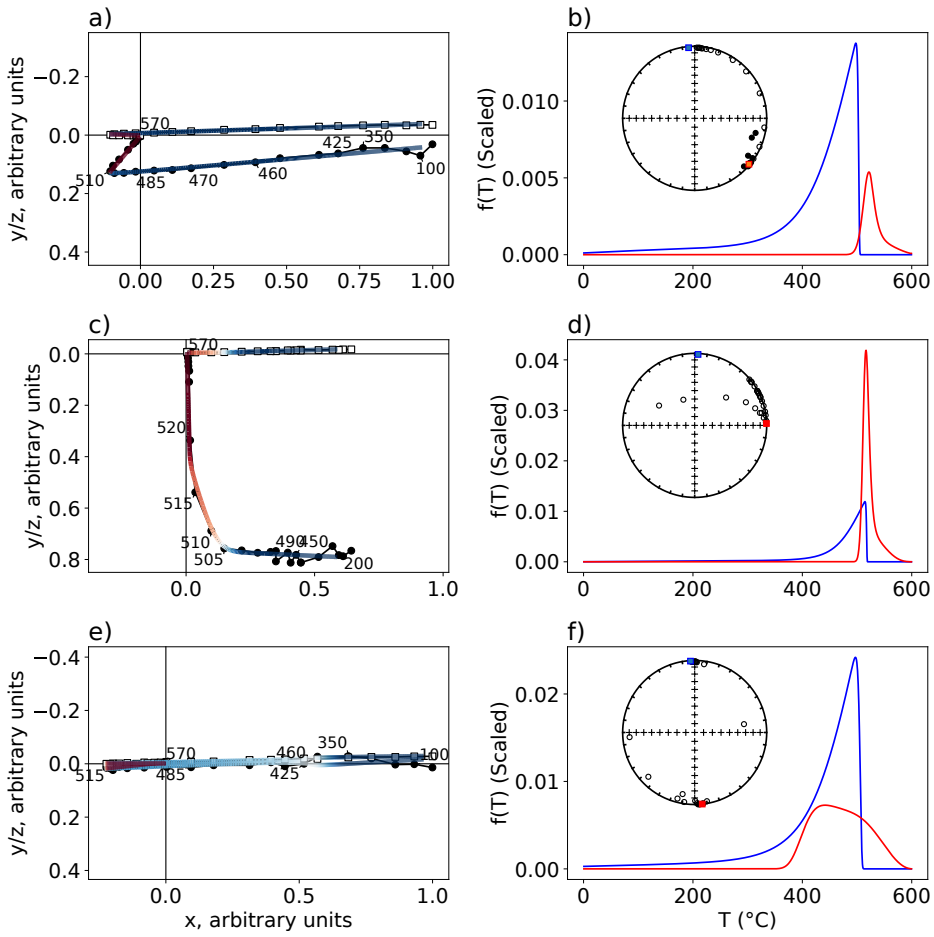

Figure 3.10. Examples of results from the fragile curvature dataset from specimens (top row) mc117d, (middle row) jm009f and (bottom row) mc117e. A description of this layout is given in Figure 3.7. Note that specimen mc117e shows a strong overlap in the unblocking temperatures, which is due to the anti-parallel directions of the two components, which makes the degree of overlap impossible to determine. However, the crossover temperature of this specimen is still close to 500°C.

which produced non-ideal behavior in the paleointensity experiments would also have curved Zijderveld plots. A single TRM was applied at 600°C in one of seven directions in the laboratory x, y plane, spaced every 45°, excluding the $+x$ direction. A second pTRM was then applied at 500°C in the remaining $+x$ direction, allowing us to test effects of the angle between our two fields on our results.

After giving specimens from the dataset of Tauxe et al., 2021 two component TRMs, we

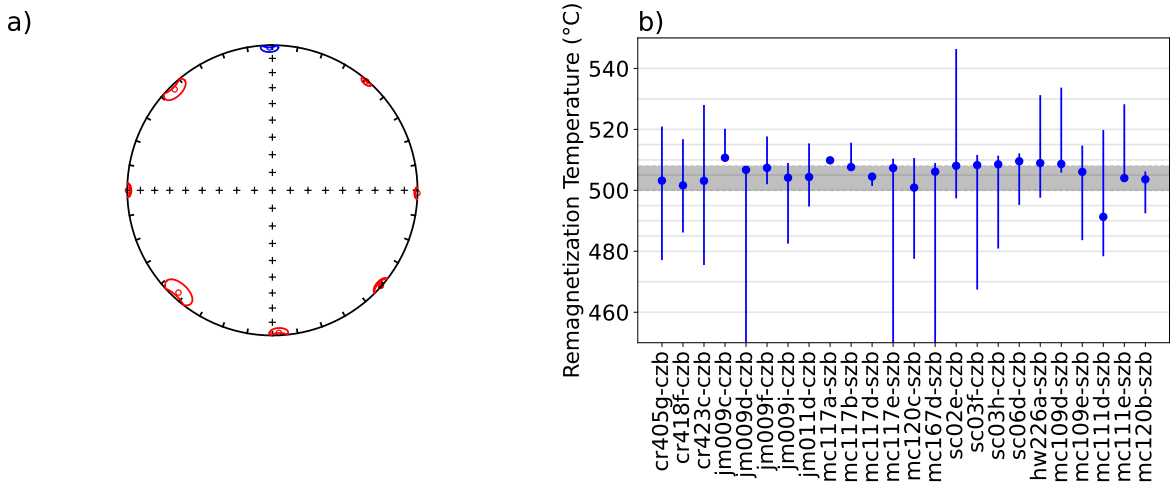

Figure 3.11. a) Equal area plot of mean directions obtained by TROUT for the remagnetized specimens from Tauxe et al. (2021) . Blue: Low Temperature component (n=24) and Red: High temperature directions (n=3 or 4). Means overlap with the expected mean for each direction, except for the direction with a declination of 45°. b) Estimates of the remagnetization temperatures for the fragile curvature specimens. Grey shaded area is the expected range of remagnetization temperatures of 500-508°C. Blue dots are the crossover temperatures from our method and blue error bars are the MRs. Faint horizontal lines are the temperature steps for the thermal demagnetization experiment. specimens mc117e, mc167d, mc109e, jm009d and sc02e were magnetized in anti-parallel fields to the magnetization direction, leading to wide MRs. In general the MRs span several temperature steps, wider than the results from the Pito Deep data set (see Figure 3.8).

used TROUT to find directions, CTs and MRs. Example results from the TROUT model are shown in Figure 3.10. The mean direction obtained for each pTRM orientation is plotted on an equal area projection in Figure 3.11a, with the high temperature directions shown in red and the low temperature direction shown in blue. The mean directions have α_{95}s within uncertainty of the expected direction, except for the direction with a declination of 45°. After the initial heating, the direction of the high temperature TRM was measured for each specimen. The mean measured direction of the high temperature TRMs is within uncertainty of the mean direction from TROUT for the group of 45° declination specimens, and so it is likely that the few degrees deviation of the TRMs from the expected direction is due to small misorientations of the specimens in the oven.

Estimated CTs and MRs are shown in Figure 3.11b. It is noticeable that the average crossover temperature is significantly higher than the expected value of 500°C by about 5-10°C. The likely reason for this is that the specimens were given pTRMs in a sample holder constructed from a thick titanium bar which allowed them to remain upright in the paleomagnetic oven, but were demagnetized in a different sample holder in which the specimens are kept horizontally. Although the thermal gradient in the Scripps paleomagnetic oven is small, the different thermal masses of these two sample holders could explain the 5-10°C overestimate of the remagnetization temperatures observed in these specimens. No comparable temperature offset was observed for the Pito Deep samples, which were given pTRMs using a different (lower mass) sample holder.

The MRs for the specimens of Tauxe et al., 2021 are often considerably wider than those for the Pito Deep specimens, even for some of the specimens which do not exhibit curvature in their original Arai plots ("hw" and "mc" specimens). Where the two TRM directions are anti-parallel, the method generally struggles to find reasonable unblocking temperature distributions, with a wide MR often found to be favorable. However, the crossover temperature for these specimens is similar to other specimens. This makes sense, as the magnetization is effectively confined to a single dimension, overlapping unblocking temperatures will affect the magnitude but not the direction of measurements, making the unblocking temperature distributions harder to invert for. An example of this is shown with specimen mc117e in the bottom row of Figure 3.10. Similarly to some of the results for the Pito Deep specimens, this demonstrates that the MR may be taken as an upper bound on the range over which unblocking temperatures overlap, as it tends to be overestimated for two anti-parallel magnetizations.

3.3.3 MORB Reheating Experiment

Kent and Gee, 1994 noted unusual behavior in a set of samples from the East Pacific Rise whereby the unblocking temperatures extended well beyond what was expected for the observed titanomagnetite Curie temperature. At the time of the study, the predominant hypothesis was that these high unblocking temperatures were carried by a new phase formed by thermochemical

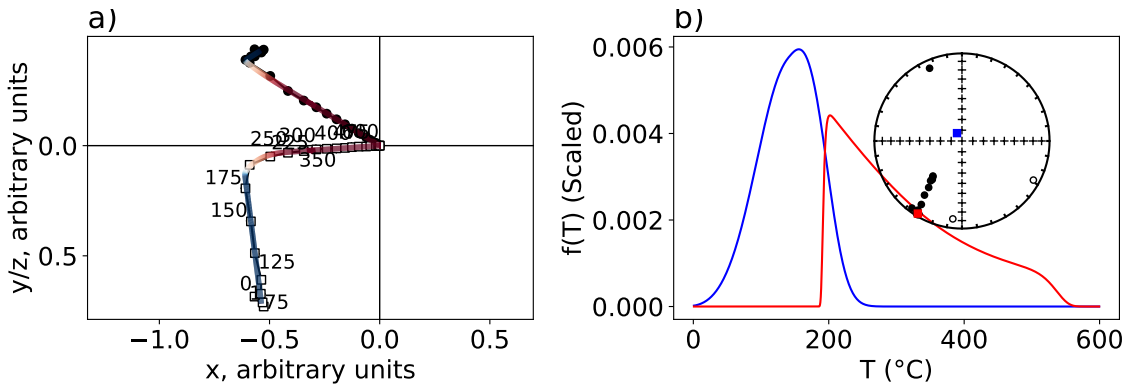

Figure 3.12. a) Zijderveld diagram of thermal demagnetization data of NRM plus pTRM imparted at 200°C from Figure 3 of Kent and Gee (1994). Superimposed on the original data (shown as black circles (X,Y) and open squares (X,Z) are TROUT model fits. b) Unblocking temperature distributions from the components shown in a). Inset) Equal Area plot of the directions from a) and the TROUT fit directions.

alteration during heating. In order to test this, Kent and Gee, 1994 gave a Mid Ocean Ridge Basalt (MORB) specimen a pTRM at 200°C, the Curie temperature of titanomagnetite. Thermal demagnetization revealed a two component magnetization, with the high temperature component in the direction of the original NRM, indicating that the high unblocking temperature component could not have been newly formed during thermal demagnetization. The Kent and Gee (1994) result is shown in Figure 3.12. We tested TROUT on the data shown in Figure 3.12a as this dataset presents an excellent opportunity to explore the value of estimating CTs and MRs in a quantitative and reproducible manner. The TROUT results are plotted on top of the original data in Figure 3.12a. And the unblocking temperature distributions are shown in Figure 3.12b. The TROUT estimates are consistent with the original study.

A set of hitherto unpublished data was produced as part of the Kent and Gee (1994) investigation in which the authors imparted pTRMs to several specimens at temperatures at 50°C steps ranging from 100 to 300°C and thermally demagnetized them as in Figure 3.12a. We applied TROUT to that data set, estimating the CTs and MRs. The TROUT results are shown in Figure 3.13. The crossover temperatures are all within one step of the temperatures at which the pTRMs were imparted, and the MRs are all within $\pm$ 1-2 temperature steps. There is no evidence

for any part of the pTRM component persisting to high unblocking temperatures, but the width of the MRs increases with the temperature of the pTRM. This could be caused by small amounts of thermochemical alteration, as the amount of material produced could increase with increased temperatures. This effect could also be caused by an inequality of blocking temperatures (the temperature at which the pTRM was acquired) and unblocking temperatures (the temperature at which the pTRM is demagnetized). Under this interpretation, the width of the MRs would be related to grain size. Blocking temperatures of 150°C and under have 'tight' MRs, expected for single domain grains with blocking temperatures equal to the unblocking temperature. Blocking temperatures from 200°C and above have wider MRs, which are skewed to higher temperatures, consistent with unblocking temperatures in excess of the blocking temperature, a hallmark of so-called 'high temperature' pTRM tails (Dunlop and Özdemir, 2000).

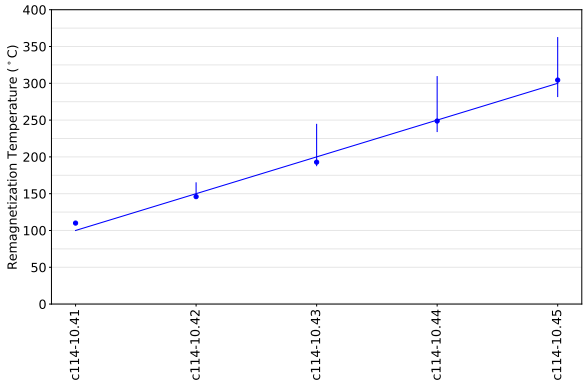

Figure 3.13. Estimates of the remagnetization temperatures from the MORB reheating experiment, unpublished data of Kent and Gee (1994). Blue line: pTRM remagnetization temperatures used in the MORB reheating experiment. Blue dots: crossover temperatures obtained from our method. Blue errorbars: mixed regions obtained from the TROUT method. Thin horizontal lines: Temperature steps used in the thermal demagnetization experiment. The crossover temperatures are all within one step, and the MRs all span 1 or 2 steps. Interestingly, the MRs widen as the pTRM temperature increases, which may be a results of high temperature pTRM tails.

3.4 Discussion

In this paper, we have presented a method which provides a full mathematical description for a thermal or alternating field demagnetization experiment, with the ability for the unblocking

temperature or coercivity spectra of components to overlap. This method can be used to obtain the direction and magnitude of each directional "component" of the magnetization, as well as information about the temperatures over which each component demagnetizes. In laboratory experiments, the TROUT method can reproduce the temperature at which a pTRM was acquired to high accuracy (usually ±1 temperature step), even when that pTRM demagnetizes over a range of temperatures leading to a curved Zijderveld plot.

In addition to accounting for overlapping unblocking temperature ranges, the TROUT method considers both instrument noise and misorientation errors when fit to specimens, which offers a more robust error model than principal component analysis. It can also be used to estimate a range of temperatures over which more than one magnetic component demagnetizes simultaneously, which we call the "mixed region" or MR, obtained from the overlap between the unblocking temperature distributions of each component. In general, the MR obtained from the TROUT method is controlled by the curvature of measurements on the Zijderveld plot.

There are some cases in which TROUT should not be applied. Because the model is non-unique, it sometimes returns MRs that operate over a broader range of temperatures than expected. To reduce the non-uniqueness, we utilize a prior that penalizes unblocking temperature distributions which overlap strongly. In some cases, TROUT still produces counter-intuitive MRs, particularly when two components are parallel to or antiparallel to one another (see e.g., Figure 3.10, bottom row) or where the two unblocking temperature distributions may completely overlap and of very different magnitudes (e.g. with an IRM from a lightning strike and a TRM), where the method of intersection of planes (McFadden and McElhinny, 1988) may be more applicable. In scenarios where a component is not acquired thermally, there is generally little information to be gained from the unblocking temperature distributions of the components. In these cases TROUT still picks accurate crossover temperatures, but will have large uncertainties in the blocking temperature distributions and magnitudes for each component.

In this paper, we have estimated a "best" set of TROUT model parameters where the posterior distribution is at a maximum. Uncertainties in these parameters can be determined by

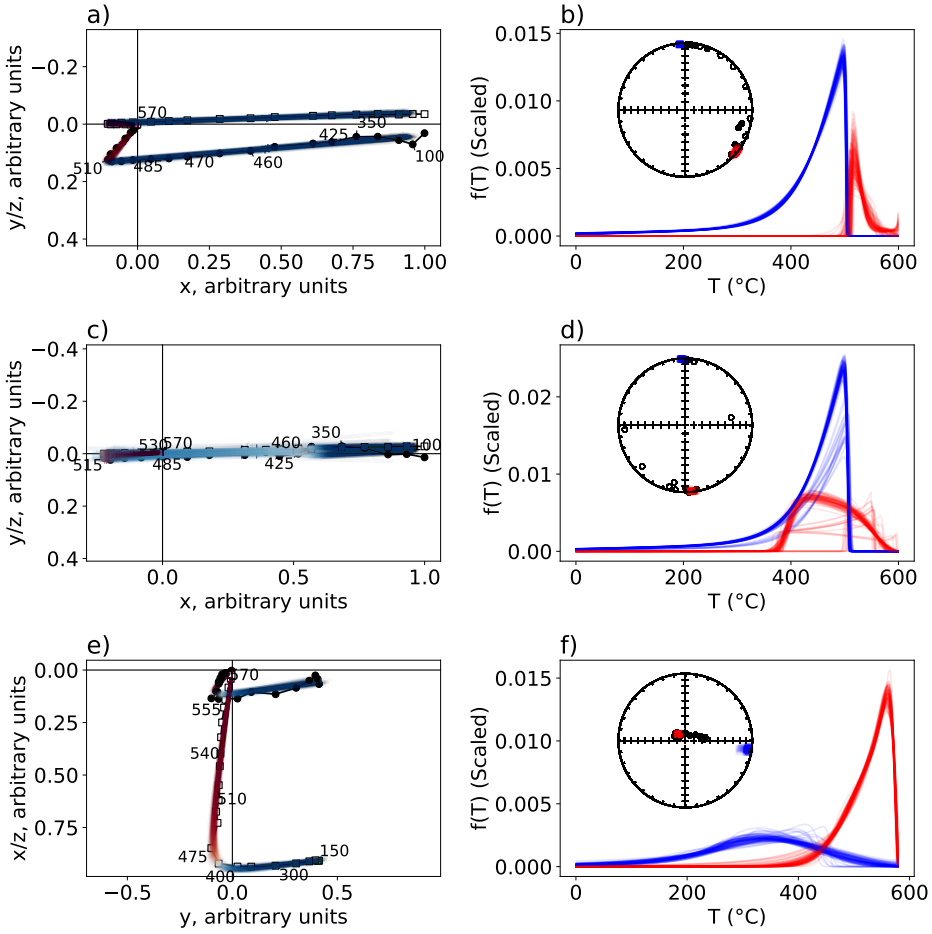

Figure 3.14. Samples from the TROUT posterior distribution using the Markov Chain Monte Carlo method of Goodman and Weare (2010), Zijderveld plots (left column), unblocking temperature distributions (right column) and Equal Area projections (insets) are displayed. a)-b) Samples from the posterior distribution for specimen mc117d-szb. There is little uncertainty in the unblocking temperature distribution, directions or magnitudes for this specimen. c)-d) Results for specimen mc117e-szb. The two antiparallel components in this specimen lead to uncertainties in the blocking temperature distributions and relative magnitudes (c) of the two components. However, the crossover temperature and directions of the two components have relatively low uncertainty. e)-f) Results for specimen PD014d2- there is some uncertainty in the unblocking temperature distribution of the weaker, low temperature component for this specimen, with the unblocking temperature distribution in f) showing a possible solution with less unblocking temperature overlap. Note that these samples are unlikely to sample the relative sizes of the different posterior modes equally, and so should only be used as a qualitative estimate of uncertainty.

directly sampling from the posterior distribution, using a Markov Chain Monte Carlo (MCMC) method. However, results from the TROUT model have multiple posterior modes, a behavior

which requires a prohibitively large number of samples for an MCMC method to accurately represent the posterior distribution. Despite this setback, MCMC sampling can be performed to get a qualitative idea of the number of posterior modes (and therefore reasonable solutions) to the TROUT model. Examples of samples drawn from the posterior distribution using the ensemble MCMC method of Goodman and Weare, 2010 for specimens presented earlier in this study are shown in Figure 3.14. For each of these plots, an ensemble of 64 markov chains each with 10000 samples was used to sample from the posterior, with the first 5000 samples from each chain discarded as "burn in". It is apparent that for specimens where the TROUT method yields counterintuitive results (e.g. for specimen mc117e-szb with two antiparallel components in Figure 3.14c-d), that there are multiple possible solutions. TROUT simply picked the solution which best fit the data. For specimen PD014d2 in Figure 3.14e-f), there is also a weaker alternative mode with less strongly overlapping blocking temperatures. Alternative posterior modes can also be assessed more rapidly by looking at the range of solutions obtained by initializing the BFGS optimizer from different starting points, instead of just the best solution obtained by TROUT.

The source of overlapping unblocking temperatures in paleomagnetic specimens is likely related to the presence of magnetic grains which have different blocking and unblocking temperatures, described in Dunlop and Özdemir, 2001, and which cause the "fragile curvature" behavior in paleointensity experiments, described by Tauxe et al., 2021. Micromagnetic modelling has demonstrated that certain sizes of magnetite have multiple minimum energy states which lead to unstable behavior (Nagy et al., 2017). This mechanism is consistent with the results described in Section 3.3.3. It could therefore be possible to use this framework to estimate the similarity between the blocking and unblocking temperatures in a paleointensity experiment. However, this is beyond the scope of this paper.

3.5 Conclusions

We present a mathematical model called TROUT which fully describes a thermal demagnetization experiment for specimens with multi component magnetization. This model can be inverted to give us directional and thermal information about a paleomagnetic experiment. A particular advantage of the TROUT method is that it can obtain an estimate of the range of temperatures over which a remanence component demagnetizes. TROUT allows for the unblocking temperature ranges of multiple components to overlap, a behavior that has long been observed in real Zijderveld data (e.g. Hoffman and Day, 1978). The ability for TROUT to obtain blocking temperature ranges is useful for a number of applications. Paleomagnetists wish to isolate a range of temperatures over which a single component is demagnetizing for analyzing paleodirection and paleointensity experiments, which TROUT enables. Additionally, TROUT can be used to find the emplacement temperature of tuffs or pyroclastic flows, as well as the temperature to which a specimen was reheated, e.g. in a host rock remagnetized by an intruding dike, or an archaeological specimen where the temperature of cooking is of interest. One of the advantages of TROUT is that it can be used to perform these analyses when Zijderveld plots are curved and without easily resolvable components.

We analysed three data sets containing specimens that were partially remagnetized in laboratory fields at known temperatures using TROUT. TROUT can estimate the temperature at which a laboratory pTRM was acquired to within ± 1 temperature step, even for specimens where there is overlap in the unblocking temperature distributions of two components. TROUT's ability to model the unblocking temperature distributions of different components may be a useful proxy of domain state in some specimens, although care should be taken when doing this, as overlapping unblocking temperatures can be caused by effects not related to domain state, and the unblocking temperature distributions obtained by TROUT may be non-unique.

3.6 Initialization of the optimization algorithm

In Section 3.2.2 we discuss our method for maximizing the posterior distribution. Any optimization requires initialization of the search algorithms and we mentioned in Section 3.2.2 that we computed an "informed guess" for the set of parameters that best fit the model. Here, we describe the procedure of coming up with an informed guess in detail. First, we use the more traditional technique of obtaining directions from a thermal demagnetization experiment, using principal component analysis (PCA). For a specimen with K components, we split the data into K pieces, where the minimum temperature of piece $k+1$ and the maximum temperature of piece k are shared. For each set of K pieces we compute the line defined by the principal component of each piece. We find the closest point of intersection between these lines, and for each intersection, we translate the lower temperature line so that the two lines intersect. We set the end point of the lowest temperature line to the NRM of the specimen, and we set the end point of the highest temperature line to the magnetization of the final demagnetization step. This allows us to describe the demagnetization experiment by a set of lines with connected endpoints, which is what we would expect to see in a Zijderveld plot with no unblocking temperature overlap.

For each of the K lines we obtain from our data-splitting approach, we take the sum of squared distances of the measurements in each piece from each line. We repeat our procedure for all possible sets of K temperature ranges, and choose the set of lines that minimizes the sum of squared distances. We take the direction of these lines as an estimate of $\vec{B}$ and their lengths as an estimate of c. The lines we have chosen should fit the data reasonably well and will approximate the maximum posterior $\vec{B}$ and c for a specimen with no overlap between the components.

To obtain estimates of μ, s, p and q, we take the vector difference sum of the data in each of our partitions, divide by c and find the set of μ, s, p^* and q^* that minimizes the squared difference between the vector difference sum and $F(x, \mu, s, p, q)$ using a single run of the BFGS method. By treating each component as its own demagnetization experiment, we can obtain a

good estimate of the μ, s, p and q that would maximize the posterior distribution if there were no overlap between components.

For σ, and ψ, we use values of $0.1|\vec{M}|_{\max}$ and $\pi/36$ as our "best guess". These parameters are intentionally set to larger values than we expect, as initializing with a small noise tolerance may cause BFGS optimizations initialized around the guess to get trapped in local minima more frequently. Although our method of generating a best guess is complex, initializing the BFGS optimizer at a guess close to the one which maximizes the posterior greatly increases the success rate of the optimization. This is because over a lot of the prior space, the posterior probability is effectively zero.

3.7 Acknowledgements

This work was supported in part by NSF grant EAR1827263 to LT. I would like to thank my doctoral dissertation committee, particularly Cathy Constable for her helpful feedback on this paper. We also thank Andrew Roberts for many thoughtful discussions on the uses of Bayesian techniques in paleomagnetism. All data used in this paper will be available in the MagIC database upon acceptance of the article.

Chapter 3, is in preparation for submission with Geochemistry, Geophysics Geosystems as Cych, B., Morzfeld, M., Heslop, D., Maher, S., Gee, J., & Tauxe, L. Thermal Resolution of Unblocking Temperatures (TROUT): A method for "unmixing" multi component magnetizations (In Preparation). The dissertation author was the primary investigator and author of this paper.

References

Dunlop, D. J., & Özdemir, Ö. (2000). Effect of grain size and domain state on thermal demagnetization tails. *Geophys. Res. Lett.*, *27*, 1311–1314.

Dunlop, D. J., & Özdemir, Ö. (2001). Beyond Néel's theories: thermal demagnetization of narrow-band partial thermoremanent magnetizations. *Phys. Earth Planet. Inter.*, *126*(1), 43–57. https://doi.org/10.1016/S0031-9201(01)00243-6

Egli, R. (2003). Analysis of the field dependence of remanent magnetization curves. *J. Geophys. Res. Solid Earth*, *108*(B2). https://doi.org/10.1029/2002JB002023

Everitt, C. W. F., & Clegg, J. A. (1962). A field test of palaeomagnetic stability. *Geophys. J. Int.*, *6*(3), 312–319. https://doi.org/10.1111/j.1365-246X.1962.tb00354.x

Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, *5*(1), 65–80. https://doi.org/10.2140/camcos.2010.5.65

Hoffman, K. A., & Day, R. (1978). Separation of multi-component NRM: A general method. *Earth Planet. Sci. Lett.*, *40*(3), 433–438. https://doi.org/10.1016/0012-821X(78)90166-8

Holme, R., & Bloxham, J. (1996). The treatment of attitude errors in satellite geomagnetic data. *Phys. Earth Planet. Inter.*, *98*(3), 221–233. https://doi.org/10.1016/S0031-9201(96)03189-5

Inman, H. F., & Bradley, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Comm. Statist. Theory Methods*, *18*(10), 3851–3874. https://doi.org/10.1080/03610928908830127

Kent, D. V., & Gee, J. (1994). Grain size-dependent alteration and the magnetization of oceanic basalts. *Science*, *265*(5178), 1561–1563. https://doi.org/10.1126/science.265.5178.1561

Kirschvink, J. L. (1980). The least-squares line and plane and the analysis of palaeomagnetic data. *Geophys. J. Int.*, *62*(3), 699–718. https://doi.org/10.1111/j.1365-246X.1980.tb02601.x

Maher, S. M., Gee, J. S., Cheadle, M. J., & John, B. E. (2021). Three-dimensional magnetic stripes require slow cooling in fast-spread lower ocean crust. *Nature*, *597*, 511–515. https://doi.org/10.1038/s41586-021-03831-6

Maher, S. M. (2021). Using magnetic methods to better understand the thermal structure of fast-spread, lower ocean crust at Pito Deep [[Online; accessed 17. Oct. 2022]]. https://escholarship.org/uc/item/01x3z748

McFadden, P. L., & McElhinny, M. W. (1988). The combined analysis of remagnetization circles and direct observations in palaeomagnetism. *Earth Planet. Sci. Lett.*, *87*(1), 161–172. https://doi.org/10.1016/0012-821X(88)90072-6

Nagata, T., Arai, Y., & Momose, K. (1963). Secular variation of the geomagnetic total force during the last 5000 years. *J. Geophys. Res.*, *68*, 5277–5282.

Nagy, L., Williams, W., Muxworthy, A., Fabian, K., Almeida, T., Conbhuí, P., & Shcherbakov, V. (2017). Stability of equidimensional pseudo–single-domain magnetite over billion-year timescales [[Online; accessed 25. Aug. 2022]]. *Proc. Natl. Acad. Sci. U.S.A.* https://doi.org/10.1073/pnas.1708344114

Néel, L. (1949). Théorie du traınage magnétique des ferromagnétiques en grains fins avec applications aux terres cuites. *Ann. géophys.*, *5*, 99–136.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Comput. J.*, *7*(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

Paterson, G. A., Biggin, A. J., Yamamoto, Y., & Pan, Y. (2012). Towards the robust selection of Thellier-type paleointensity data: The influence of experimental noise. *Geochem. Geophys. Geosyst.*, *13*(5). https://doi.org/10.1029/2012GC004046

Santos, C., & Tauxe, L. (2019). Investigating the accuracy, precision, and cooling rate dependence of laboratory acquired thermal remanences during paleointensity experiments. *Geochem. Geophys. Geosys.*, *20*, 383–397. https://doi.org/10.1029/2018GC007946

Tauxe, L., Santos, C. N., Cych, B., Zhao, X., Roberts, A. P., Nagy, L., & Williams, W. (2021). Understanding nonideal paleointensity recording in igneous rocks: Insights from aging

experiments on lava samples and the causes and consequences of 'fragile' curvature in arai plots. *Geochem. Geophys. Geosyst.*, *22*(1), e2020GC009423. https://doi.org/10.1029/2020GC009423

Tauxe, L., & Staudigel, H. (2004). Strength of the geomagnetic field in the Cretaceous Normal Superchron: New data from submarine basaltic glass of the Troodos Ophiolite. *Geochem. Geophys. Geosyst.*, *5*(2), Q02H06, doi:10.1029/2003GC000635.

Williams, W., & Dunlop, D. J. (1989). Three-dimensional micromagnetic modelling of ferromagnetic domain structure. *Nature*, *337*, 634–637. https://doi.org/10.1038/337634a0

Zhao, X., Fujii, M., Suganuma, Y., Zhao, X., & Jiang, Z. (2018). Applying the Burr Type XII distribution to decompose remanent magnetization curves. *J. Geophys. Res. Solid Earth*, *123*(10), 8298–8311. https://doi.org/10.1029/2018JB016082

Zijderveld, J. D. A. (1967). A. C. demagnetization of rocks: Analysis of results. In D. W. Collinson, K. M. Creer, & S. K. Runcorn (Eds.), *Developments in Solid Earth Geophysics* (pp. 254–286). Elsevier. https://doi.org/10.1016/B978-1-4832-2894-5.50049-5

Chapter 4

Non-dipolar time-averaged field for the past 1.5 million years from new paleointensity results from Hawai‘i, compared to global datasets

Brendan Cych[1], Lisa Tauxe[1], Geoff Cromwell[2], John Sinton[2], Anthony Koppers[4]

[1]Scripps Institution of Oceanography, University of California, San Diego

[2]United States Geological Survey

[3]Department of Earth Sciences, University of Hawai‘i at Manōa, Honolulu, HI

[4]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvalis OR

Chapter 4 is in preparation for submission to Proceedings of the National Academy of Sciences as Cych, B., Tauxe, L., Cromwell, G., Sinton, H., & Koppers, A. Non-dipolar time-averaged field for the past 1.5 million years from new paleointensity results from Hawai‘i, compared to global datasets (In Preparation). The dissertation author was the primary investigator and author of this paper.

Significance Statement

Reconstructions of tectonic plates rely on the assumption that the time-averaged geomagnetic field behaves like a axial geocentric dipole. Global compilations of field directions are close to a GAD, but field strengths are not. Obtaining estimates of field strength is difficult because of high experimental failure rates and inconsistent analysis methods between studies. Here we present new data collected from igneous rocks in Hawai‘i and use consistent analytical methods to compare them to published data from Antarctica and Israel. Our results indicate a persistent non-dipolar component in the Earth's magnetic field over the past 1.5 Ma, but more dipolar behavior prior to that. This is surprising, given our current understanding of the processes that give rise to the field.

Abstract

A foundational assumption in paleomagnetism is that the Earth's magnetic field behaves as a geocentric axial dipole (GAD) when averaged over sufficient timescales. Compilations of directional data averaged over the past 5 Ma yield a distribution close to this assumption, but the distribution of paleointensity data over this timescale is not. Reasons for the failure of GAD include: 1) Arbitrary "selection criteria" to eliminate "unreliable" data vary between studies, so the paleointensity database may include biased results. 2) The age distribution of existing paleointensity data varies from latitude to latitude so different latitudinal averages likely represent different time periods. 3) The time-averaged field could be truly non-dipolar. Here, we present a consistent methodology for analyzing paleointensity results and comparing time-averaged paleointensities from different studies. We apply it to data from Plio/Pleistocene Hawai‘ian igneous rocks, sampled from fine-grained, quickly cooled material (lava flow tops, dike margins and scoria cones) and subjected to the IZZI-Thellier technique; the data were analyzed using the BiCEP method of Cych et al (2021, doi:10.1029/2021GC009755), which produces accurate paleointensity estimates without arbitrarily excluding specimens from the

analysis. We constructed a paleointensity curve for Hawai‘i over the Plio/Pleistocene using the method of Livermore et al (2018, doi:10.1093/gji/ggy383), which accounts for age distribution and has robust uncertainties. We demonstrate that even with the large uncertainties associated with obtaining a mean field from temporally sparse data, our average paleointensities obtained from Hawai‘i and Antarctica (from Asefaw et al., 2021, doi:10.1029/2020JB020834, reanalyzed here) are not GAD-like after about 1.5 Ma.

Paleomagnetists use the direction of the magnetization acquired in the Earth's ancient magnetic field to obtain estimates of the ancient latitude at which the rock formed. Calculation of a latitude relies on an assumption that the Earth's magnetic field is structured like a bar magnet when averaged over sufficiently long timescales, so that the magnetic field is vertical at the poles, and horizontal at the equator, also termed a Geocentric Axial Dipole (GAD). Estimates of the Earth's magnetic field direction, taken from different latitudes over the past 10 Myr conform relatively well to a GAD field, with a small hemispheric asymmetry (Cromwell et al., 2018). On the other hand, estimates of the Earth's magnetic field strength (the paleointensity) averaged over the last 5 Myr consistently show a behaviour incompatible with a strongly dipolar field. A seemingly persistent feature in paleointensity data is the presence of weak paleofields at high southern latitudes (Asefaw et al., 2021; Lawrence et al., 2009; Tauxe et al., 2022), which causes a hemispheric asymmetry in the paleointensity data. This is seen in paleointensities from the MagIC database over the last 5 Ma (plotted in Figure 4.1a) where the mean paleointensity at 80° S would be produced by a centered magnetic dipole with a moment of around 40 ZAm^2, whereas the mean paleointensity at 20° N would require a dipole moment with a magnitude closer to 80 ZAm^2. Attempts to fit Giant Gaussian Process (GGP) models to paleointensity data to determine the structure of the time-averaged field have found that the field consistently requires a strong quadrupole term 15-30% the strength of the dipole field (Muxworthy, 2017; Shcherbakov et al., 2019), producing this asymmetry. However, such a large quadrupole is completely incompatible with the directional data.

We have identified three hypotheses that could explain the non-dipole like behaviour of

global time-averaged paleointensity records: bias in paleointensity estimation, comparison of temporally distinct data in a time varying field, and genuine non-dipole field behavior. Regarding the issue of bias, paleointensity estimation involves normalizing the observed natural remanent magnetization (NRM) to a magnetization acquired in a known laboratory field. The accurate determination of a paleointensity therefore requires that the acquisition of a magnetization be reproducible. However, it has been shown (e.g., Dunlop and Özdemir, 2001; Krása et al., 2003; Levi, 1977; Tauxe et al., 2021) that some rocks have non reproducible magnetizations, which can lead to biased paleointensity estimates. Global paleointensity records may be confounded by these biased estimates, leading to an apparent non dipole signature. Alternatively, geomagnetic intensity variations through time may not be well averaged. The majority of paleointensity determinations are made with volcanic rocks, which record an instantaneous snapshot of the magnetic field at the time they cool. Archeomagnetic data indicate that the Earth's magnetic field strength can vary strongly over decades to centuries, so numerous paleointensity estimates are necessary for a good average. If the field strength varies over long timescales (e.g. millions of years), then comparing the "average" of two studies may not be meaningful if the units sampled are of different ages. And finally, it is also possible that the geomagnetic field is not in fact GAD-like but has long-term non-axial dipole contributions (as suggested by Cromwell et al., 2013; Tauxe et al., 2022; Wilson, 1970).

Recent paleointensity studies have made a greater effort to eliminate bias from the results. Paleomagnetists have identified behaviors in a paleointensity experiment that deviate from theoretical expectations and may lead to bias. In most paleointensity studies, results from paleomagnetic specimens are excluded from the analysis if they fail a set of "selection criteria" which are phenomenological descriptions of these behaviors. Alternatively, the BiCEP method (Cych et al., 2021) attempts to find a relationship between the apparent paleointensity and one of these commonly used selection criteria (curvature G. A. Paterson, 2011), and attempts to correct for the bias induced by the non-ideal behavior, obtaining accurate results without excluding data from the analysis based on arbitrary criteria. Recently, a study (Tauxe et al., 2022) which

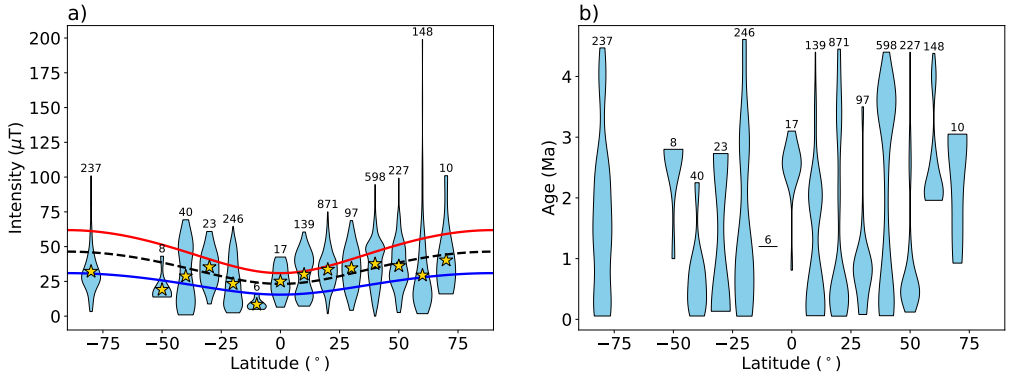

Figure 4.1. Violin plots showing latitudinal binned distributions of a) paleointensity and b) age for reported paleointensity results from the MagIC database aged between 50 ka and 5 Ma. In a violin plot, the width of the violin represents the frequency of intensities in that latitude bin, with the widest point in the violin representing the modal value. The number of data points in each bin are noted above the violins. The yellow stars in a) are the mean paleointensity value at each latitude bin and the solid blue, dashed black and solid red lines represent the expected mean values for a dipole field with a strength of 40, 60 and 80 ZAm^2 respectively.

used the strict CCRIT criteria (Cromwell et al., 2015) and the BiCEP method on paleointensity studies from several latitudes found that there is still a discrepancy between these time-averaged paleointensities and those expected for a GAD field, making our first hypothesis (apparent non-dipole behavior is caused by bias in paleointensity estimation) unlikely to be the cause of inaccurate paleointensities.

Figure 4.1b shows the age distribution of latitudinally binned absolute paleointensity data in the MagIC database (without selection). It is apparent that different latitude bins have different age distributions. Because of this, the average paleointensity from each bin is representative of a different time period, and is not an average paleointensity for the whole of the last 5 Ma. High quality paleointensity data, analyzed in a consistent manner, are needed to determine whether temporal sampling is the cause of apparent non-dipolar behavior, or if the time-averaged field is truly non-dipolar, as outlined in our third hypothesis.

In this paper, we present high quality paleointensity estimates from rapidly cooled volcanic material from lava flows, dikes and vent deposits (scoria and spatter cones) aged 0-4 Myr from the Hawai‘ian islands. In Section 4.1, we describe how we collect samples in the field

(4.1.1), how we conduct paleointensity experiments (4.1.2) on specimens therefrom, how we analyze our results using the BiCEP method which produces accurate estimates for specimens magnetized in known fields (4.1.3), and how we obtain ages for our samples using $^{40}Ar/^{39}Ar$ dating (4.1.4). In Section 4.2, we show the results of our paleointensity study in Hawai‘i. Section 4.3.2 discusses how our results show that scoria may be a useful lithology for obtaining high quality paleointensity estimates, and are in agreement with estimates from other lithologies. In Section 4.3.3 we fit a model to our paleointensity data in an attempt to derive a time average that accounts for uneven temporal sampling. We then apply the same methodology to studies from Northern Israel and in Antarctica. This allows us to test whether poor temporal sampling or non dipole behavior is responsible for the weaker paleointensity at high latitudes. Our results indicate that there is a persistent non-dipole component in the Earth's magnetic field over at least the past 1.5 Myr with older data being much more consistent with a GAD field.

4.1 Methods

4.1.1 Field Methods

Our results come from samples collected over three field seasons from outcrops on the Hawai‘ian islands. Samples were collected from the islands of Hawai‘i, Maui, Moloka‘i, and O‘ahu in an attempt to get a representative average paleointensity over the past 4 Myr. This study targeted predominantly glassy and fine grained igneous material from lava flow tops and bottoms, scoria cones and dike margins. Néel theory (Néel, 1949) predicts the physics of "uniaxial single domain" grains which should behave ideally in a paleointensity experiment. Only very small magnetic particle sizes exhibit single domain behavior, and so we sampled rapidly quenched materials most likely to contain these fine grains.

In the field, we collected small unoriented hand samples using a hammer and chisel; this allowed us to obtain smaller pieces of material and was less destructive than obtaining oriented specimens with a drill. Maps of our sampling localities are shown in Figure 4.2 and details

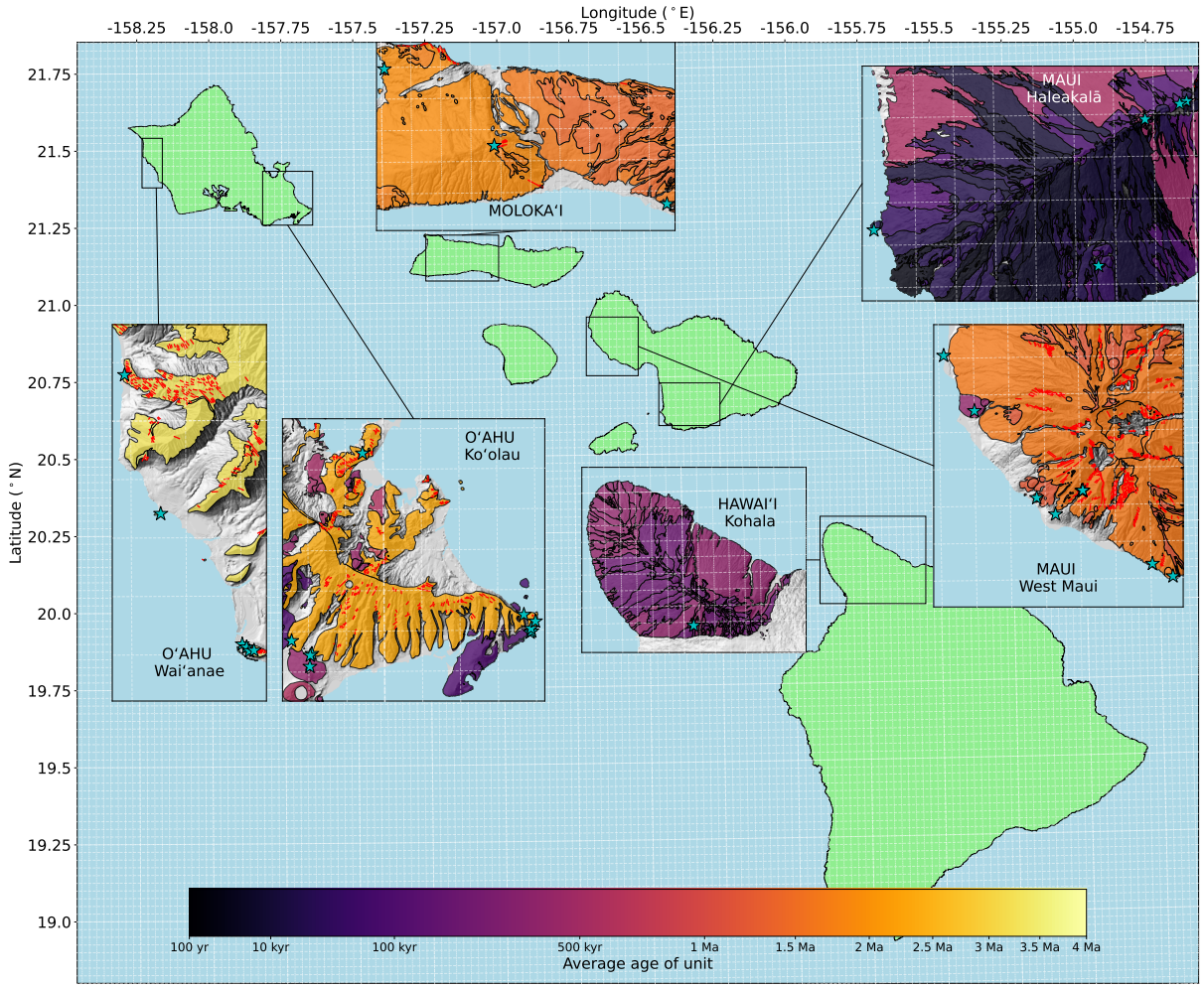

Figure 4.2. Maps showing sampling localities for successful sites used in this study (blue stars). Insets are labeled with the name of each island in capital letters and the name of the volcano (if applicable) in lowercase. Each map shows samples from a different Volcano/Island. Colors represent average ages of units (Sherrod et al., 2007), with darker colors indicating younger flows (see colorbar), and dike locations indicated by red lines. Topographic data: U.S. Geological Survey (USGS). 2015. USGS 10-m Digital Elevation Model (DEM): Hawai'i. Coastline data: Hawai'i Statewide GIS Program.

regarding location, age and material are given in Table 4.1.

4.1.2 Laboratory Work

Each sample was crushed with a mortar and pestle to produce multiple paleomagnetic specimens with masses on the order of 0.1 g. Specimens were weighed and glued into 1 cm wide borosilicate glass tubes using a high temperature, low magnetic moment glue (KaSil). We

Table 4.1. Ages and locations for sites from this study that passed CCRIT or BiCEP. Locations for all sites, including those that did not pass CCRIT or BiCEP are listed in the supporting information. Latitudes and Longitudes are referenced to the WGS84 standard.

Site	Island	Lithology	Lat. (°N)	Lon. (°E)	Age (Ma)	$\pm 2\sigma$
HW306	Hawai‘i	Vent Deposit	20.04470	-155.73437	0.1900	0.0700
ML001	Moloka‘i	Dike	21.13719	-157.15547	2.0700	0.0200
ML012	Moloka‘i	Vent Deposit	21.08955	-157.01053	1.6100	0.0300
ML015	Moloka‘i	Vent Deposit	21.19876	-157.24734	1.7700	0.0200
MU004	Maui	Vent Deposit	20.77605	-156.53433	1.4300	0.0200
MU009	Maui	Vent Deposit	20.81885	-156.61782	0.6100	0.0120
MU011	Maui	Vent Deposit	20.83016	-156.63110	1.2300	0.0690
MU012	Maui	Vent Deposit	20.88931	-156.67484	0.3000	0.0216
MU013	Maui	Vent Deposit	20.92685	-156.69633	0.5840	0.0100
MU023	Maui	Vent Deposit	20.61085	-156.31100	0.0765	0.0635
MU025	Maui	Vent Deposit	20.70692	-156.25424	0.0950	0.0450
MU027	Maui	Vent Deposit	20.70551	-156.25857	0.0950	0.0450
MU031	Maui	Vent Deposit	20.69669	-156.28040	0.0670	0.0404
MU036	Maui	Vent Deposit	20.63397	-156.45102	0.0106	0.0085
MU106	Maui	Dike	20.83446	-156.59879	1.4900	0.0500
MU109	Maui	Dike	20.83440	-156.59798	1.5500	0.0500
MU111	Maui	Dike	20.83471	-156.59808	1.4500	0.0600
MU113	Maui	Lava Flow	20.78467	-156.54893	1.1000	0.0600
OA003	O‘ahu	Flow	21.29434	-157.81123	2.5500	0.0800
OA008	O‘ahu	Flow	21.40440	-158.17461	3.7100	0.0600
OA014	O‘ahu	Dike	21.51972	-158.22772	3.4900	0.1700
OA015	O‘ahu	Flow	21.46033	-158.21154	3.1000	0.0300
OA019	O‘ahu	Flow	21.30938	-157.65783	2.8400	0.0600
OA026	O‘ahu	Flow	21.29836	-157.65380	2.7700	0.1300
OA028	O‘ahu	Flow	21.29907	-157.65273	2.7200	0.0800
OA030	O‘ahu	Vent Deposit	21.27831	-157.79929	0.3800	0.1100
OA100	O‘ahu	Vent Deposit	21.28628	-157.79791	0.4800	0.0400
OA101	O‘ahu	Vent Deposit	21.28521	-157.79900	0.4800	0.0400
OA104	O‘ahu	Flow	21.30080	-157.65320	2.1800	0.3500
OA108	O‘ahu	Dike	21.30527	-157.65027	2.2500	0.1700
OA114	O‘ahu	Dike	21.41002	-157.76354	2.8700	0.0600
OA116	O‘ahu	Dike	21.40308	-158.17264	3.7200	0.0500
OA117	O‘ahu	Dike	21.40308	-158.17264	3.7200	0.0500
OA123	O‘ahu	Sill Margin?	21.40149	-158.17141	2.5900	0.0900
OA124	O‘ahu	Dike	21.40168	-158.16927	3.2500	0.0100

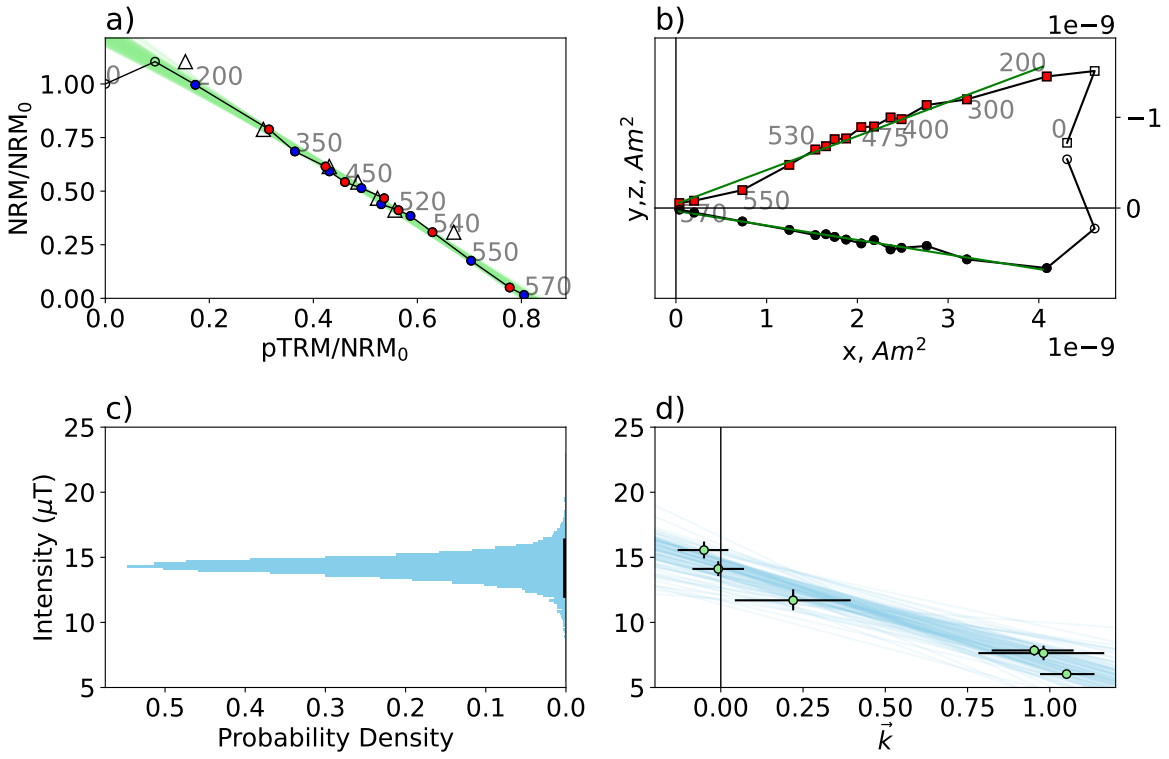

Figure 4.3. Example of BiCEP being used to obtain a paleointensity for site MU111. a) Arai plot (Nagata et al., 1963) for specimen MU111A05, red dots represent steps where the zero-field measurement was made first, and blue dots represent in-field first steps. Open circles represent temperature steps not used for this analysis. Triangles are pTRM checks and green curves are BiCEP's circular fits to the data. b) Zijderveld plot (Zijderveld, 1967) showing magnetic direction data. Open symbols are steps where the temperature steps were not used. Green line is a principal component analysis fit to the directional data. c) Histogram of possible site mean intensities from BiCEP. d) BiCEP fit showing the predicted relationship (blue lines) between intensity (y axis) and the curvature criterion ($\vec{k}$, x axis).

subjected each specimen to the IZZI-Thellier method (Tauxe and Staudigel, 2004; Yu et al., 2004). This is a step-wise double heating experiment in which the NRM is replaced by a Thermal Remanent Magnetization (TRM) acquired in a known lab field. Under the IZZI protocol, the order of the in-field and zero-field steps alternates at each temperature step. Under ideal conditions, the ratio of the magnetization lost in a zero-field step to the magnetization gained in an in-field step is the ratio of the ancient field (B_{anc}) to the laboratory field (B_{lab}). For this study, multiple lab fields were used for different specimens, as we observed that the choice of

B_{lab} affected whether our specimens passed or failed some of our criteria (see Section 1.4.1.3).

4.1.3 Analysis of Data

To make sure that we have unbiased results, we used two different analysis methods on our data to obtain an estimate of the ancient field. Primarily, we used the BiCEP method (Cych et al., 2021) of estimating paleointensities, but we also looked at results using the CCRIT criteria of Cromwell et al., 2015. BiCEP assumes that the magnetization records a single field, and thermochemical alteration of the specimen has not occurred. To make certain of this, we used the minimal selection criteria (see G. Paterson et al., 2014 for definitions and references) MAD_{Coe} <5 (see Section 4.3.1), DANG<10, DRAT<10 and selected the set of temperature steps on the Arai plot which maximize the FRAC criterion while passing the MAD_{Coe}, DANG and DRAT criteria. The vast majority of our specimens pass these criteria with ease, and the ones that do not would unambiguously be rejected by almost any other set of criteria. Site results from BiCEP have a 95% credible interval which is equivalent to the full width of the 2σ interval from traditional selection criteria methods (e.g., CCRIT). We considered a site level result from BiCEP acceptable if it has a credible interval with a full width less than 40% of the median value, or 16 μT, whichever is greater (the original BiCEP A or B criteria of Cych et al., 2021 only include the former criterion). This is equivalent to criteria of $\pm$10% or 4 μT used for the CCRIT at a site level. An example of BiCEP being used to estimate B_{anc} and its uncertainty for a site is shown in Fig. 4.3.

4.1.4 Age Constraints

We obtained a range of radiometric ages for our samples that span the past 4 Ma. Rocks from 23 of our successful sites were analyzed at the Argon Geochronology lab at Oregon State University (OSU) for age determination. 200-300 μm pieces from each sample were prepared by acid leaching in an ultrasonic bath according to the procedure of Koppers et al., 2000. This was followed by irradiation of the samples in the OSU TRIGA CLICIT nuclear reactor. Samples

were then incrementally heated using a defocused CO_2 laser, and the isotopic composition of the released argon was measured using an ARGUS-VI multi-collector mass spectrometer. Seventeen of our ages were calculated using argon-argon (Ar-Ar) plateaus. Three ages from sites OA019, OA116 and OA124 were calculated using Mini-Plateau ages. Sites MU011 and MU036 were calculated using inverse isochron ages and site ML001 was calculated using a total fusion age. For sites OA030, OA100 and OA101, we used existing potassium-argon (K-Ar) ages, (Ozawa et al., 2005) and on West Maui, existing K-Ar ages (Tagami et al., 2003) were similarly used for sites MU009 and MU013. Mapped scoria cones at sites MU023, MU025 and MU027 have good age constraints over the timescale we are interested in from K-Ar dating and stratigraphic relationships outlined in (Sherrod et al., 2003). Finally, site OA026 has its ages constrained by stratigraphic relationship with our other Ar-Ar dated flows. A full table of ages is given in Table 4.1.

4.2 Results

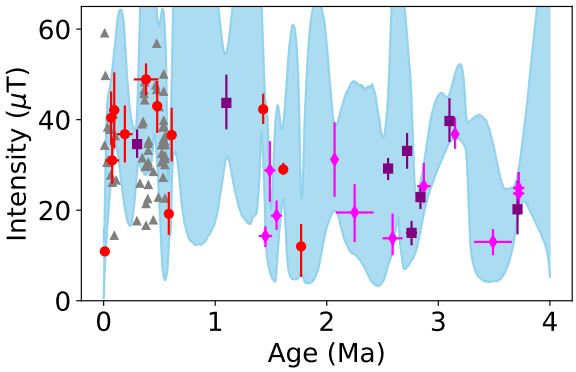

Figure 4.4. Paleointensity and age estimates from this study using the BiCEP method from lava flows (purple squares), scoria cones (red circles) and dike margins (pink diamonds) . Error bars represent the 95% credible interval for intensity estimates, and the 2σ interval for age estimates. Grey triangles are other Hawai'ian results from the HSDP2 core (Cai et al., 2017), which have a similar distribution over this time period to our results. Blue envelope represents the 95% credible interval for the AH-RJMCMC model (Livermore et al., 2018) fit to the data (see section 4.3.3).

Results are listed in Table 4.2. We obtain passing results from 35 sites (Table 4.2): 31 passed BiCEP and 21 passed CCRIT. Some of the results that pass CCRIT do not pass BiCEP, but those sites that pass both methods exhibit good agreement between one another. Because BiCEP gives a more objective analysis, and because we obtain more passing results with this method, we use only the results that pass BiCEP for the rest of our analyses.

We plot our results versus age in Fig. 4.4. It is apparent that our results support the hypothesis that the more recent field (over the past 1.5 Ma) is considerably higher than that from 1.5-4 Ma (e.g., Tauxe, 2006, supporting the observation of a potential long period variation in the field strength (Selkin and Tauxe, 2000; Tauxe, 2006; Ziegler et al., 2011). It is also worth noting that in Fig. 4.1, latitudes which have age distributions which skew towards ages older than 1 Ma (e.g. 80° S, 60° N, 0°) tend to have averages that agree with a $\sim$40 ZAm^2 dipole, whereas the majority of latitudes with mostly younger results tend to agree with a 60-70 ZAm^2 dipole moment, so qualitatively our hypothesis that the missing dipole may be caused by temporal sampling seems plausible. However, the data from Antarctica (Asefaw et al., 2021) span the entire last 4 Ma but also have an average field consistent with a 40 ZAm^2 axial dipole strength, so temporal sampling alone does not explain all of the deviation from a GAD field.

Our high paleointensity results over the past 1.5 Ma come predominantly from vent deposits (scoria and spatter cones), whereas our older results come predominantly from dikes and lava flows. The dikes and lava flows come predominantly from the early shield building stages of Hawai‘ian volcanoes, whereas the vent deposits are predominantly from the later stages of volcanic construction. The difference in lithology being coupled with a difference in field strength may be concerning, however our young, high field strength results agree well with results from lava flows in the HSPD2 core (Cai et al., 2017; Tauxe and Love, 2003, reanalyzed in Tauxe et al., 2022), shown as grey triangles in Figure 4.4. Additionally, results from several scoria cones yielded much weaker fields, including for two cones on Moloka‘i older than 1.5 Ma. This leads us to believe that our results from scoria are accurate.

Table 4.2. Paleointensity results from specimens in this study which passed BiCEP and CCRIT. n_{pass}: Number of passing specimens. n_{tot}: Total number of specimens. For CCRIT results B_{min} and B_{max} represent the bounds of the 2σ interval, and so a full width of 40% or 16 μT is considered to have passed. The method column represents the preferred paleointensity result (BiCEP) when a site passed both BiCEP and CCRIT

Site	n_{pass}/n_{tot}	B_{min}	B_{anc}	B_{max}	Method
HW306	8/8	31.00	36.90	42.90	BiCEP
ML001	7/7	23.50	31.20	39.10	BiCEP
ML012	6/6	28.10	29.00	30.20	BiCEP
ML015	5/5	5.00	11.80	17.10	BiCEP
MU004	11/11	39.30	42.20	45.40	BiCEP
MU009	6/6	29.80	35.10	40.90	BiCEP
MU011	5/9	19.21	26.51	33.81	CCRIT
MU012	6/6	31.60	34.70	37.60	BiCEP
MU013	8/8	14.90	19.20	23.80	BiCEP
MU023	5/8	31.50	34.00	37.20	BiCEP
MU025	7/7	33.70	42.10	50.30	BiCEP
MU027	6/6	18.70	24.70	30.70	CCRIT
MU031	10/10	34.60	40.50	46.10	BiCEP
MU036	9/9	10.40	10.90	11.30	BiCEP
MU106	7/12	21.90	30.10	36.30	BiCEP
MU109	5/7	11.30	16.90	22.20	BiCEP
MU111	6/6	10.60	13.70	16.10	BiCEP
MU113	7/8	37.40	43.40	49.60	BiCEP
OA003	10/11	27.00	30.10	33.20	BiCEP
OA008	4/4	14.80	20.20	26.60	BiCEP
OA014	5/10	7.40	13.00	18.60	BiCEP
OA015	8/8	35.20	39.70	44.50	BiCEP
OA019	15/15	20.20	22.80	25.30	BiCEP
OA026	8/8	12.50	15.00	17.50	BiCEP
OA028	8/8	28.50	32.30	36.00	BiCEP
OA030	16/16	45.60	48.90	52.10	BiCEP
OA100	6/12	50.02	51.02	52.02	CCRIT
OA101	9/9	37.20	43.00	48.30	BiCEP
OA104	3/8	15.79	17.55	19.31	CCRIT
OA108	8/8	14.60	20.00	25.50	BiCEP
OA114	6/6	21.90	25.30	29.60	BiCEP
OA116	7/8	22.40	25.50	29.10	BiCEP
OA117	5/5	20.80	24.30	27.90	BiCEP
OA123	5/8	10.40	13.30	17.20	BiCEP
OA124	7/7	33.70	36.80	40.10	BiCEP

4.3 Discussion

4.3.1 Pitfalls of selection criteria

We used the BiCEP method to obtain site level paleointensity estimates, and prefer this over the CCRIT method (and all other sets of selection criteria in use by various authors) as BiCEP produces many more site level results than CCRIT. Often, BiCEP passed sites where specimens failed the FRAC criterion of CCRIT, which specifies that a large proportion of the total magnetization of the specimen is needed to make a paleointensity estimate. BiCEP accounts for the uncertainty in curvature (and therefore bias), introduced by using only part of a specimen's Arai plot for a paleointensity estimate. This can be seen in Fig. 4.3a, where specimen MU111A05 fails CCRIT due to low FRAC, but using a smaller part of the Arai plot translates to only a small increase in the uncertainty in curvature, shown by the green curves fit to the data.

In addition to the FRAC criterion in CCRIT, we identify cases in which criteria may reject a specimen if it has an ancient field much lower than the lab field. The MAD criterion may be exceeded if the laboratory magnetization acquired in an in-field step is not fully removed during a zero-field step, a consequence of a "high temperature pTRM tail" (Dunlop and Özdemir, 2000). This behavior is very noticeable in IZZI experiments (Fig. 4.5), as the in-field first steps are more strongly affected by this effect. This leads to a zig-zag appearance in the Zijderveld plot. The sizes of these tails are dependent on both the magnitude of the lab field, and the effect the tails have on MAD is dependent on the angle between lab and ancient field. If we call this angle θ, then the perpendicular part of the tails will be controlled by $B_{lab} \sin\theta$. If we assume no other sources of deflection to the MAD angle, the equation for the effect is:

$$\tan(\mathrm{MAD}) \propto \frac{B_{lab}}{B_{anc}} \sin\theta. \tag{4.1}$$

This equation demonstrates that in the same lab field, sites with low ancient fields will be preferentially rejected with higher MAD, and sites with high ancient fields will be preferentially

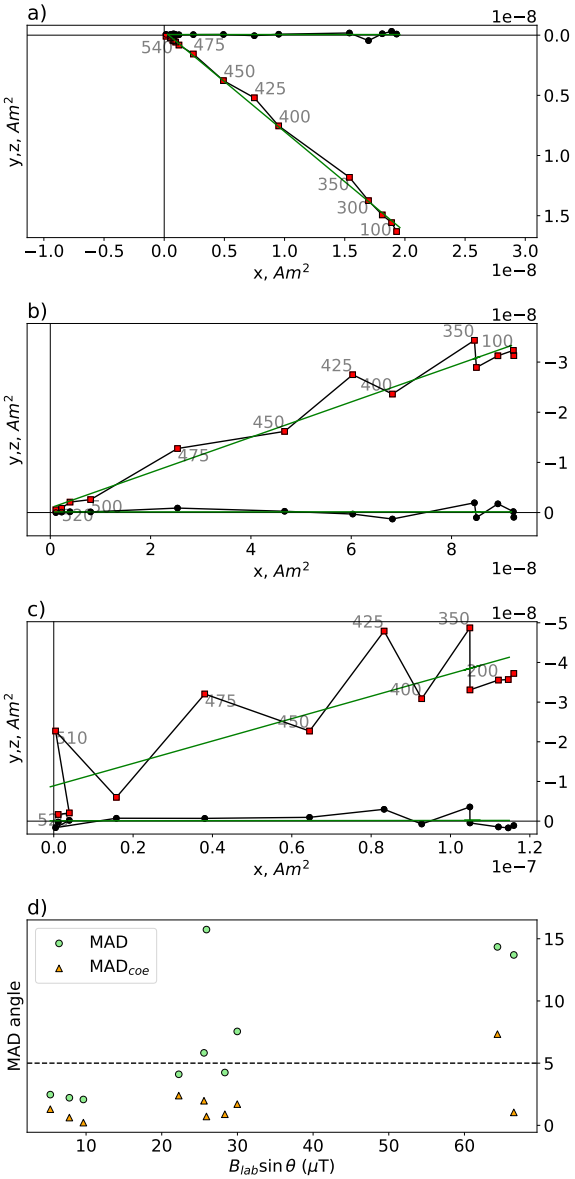

Figure 4.5. a)-c) Zijderveld plots of specimens from site OA014, showing zig-zagging behavior that progressively increases with lab field and d) Scatter plot showing the relationship between the MAD criterion, and the magnitude and angle of the lab field for all ten fully demagnetized specimens from this site. Paleointensity experiments were performed laboratory fields of a) a 10 μT, b) 30 μT and c) 70 μT. d) MAD (green circles) angle against the strength of the component of the lab field perpendicular to the ancient field direction (calculated by the PCA of the zero-field first steps). Orange triangles are the MAD of the zero-field first steps only (MAD_{Coe}). Horizontal dashed line represents the selection criterion (5) used in this study. Using MAD_{Coe} improves, though does not completely eliminate, the lab-field dependence of MAD. All MADs were calculated using temperature steps above 400-600°C to avoid any potential viscous remanent magnetization (VRM).

accepted.

To counteract the lab field-dependent effects, we used 10, 30 and 70 μT fields in our studies, which captures the range of the ancient field. At some sites with low estimated B_{anc}, there was an observably higher pass rate in lower fields. An example of this for site OA014 is illustrated in Fig. 4.5. To treat specimens magnetized in different fields fairly, it is tempting to come up with a criterion for MAD which is dependent on Equation 4.1. However, effects that we may be using MAD to look for (e.g. two component magnetizations) will not be dependent on the lab field, and so we suggest calculating MAD for exclusively the zero-field first or "Coe" type steps (Coe, 1967). Although pTRM tails may still be present in these steps, they will be significantly reduced in in-field first steps. We call a MAD calculated using these steps MAD_{Coe} and how it compares to MAD for site OA014 is shown in Fig. 4.5d. This significantly reduces the lab field-dependent effects, but does not eliminate them entirely. Because pTRMs scale with the lab field used, there may be other unrecognized pTRM dependent effects. We recommend using a range of lab fields in paleointensity studies as the most robust way of compensating for these effects.

4.3.2 Sample Characterization

We have demonstrated our ability to obtain high quality paleointensity results from our samples using the BiCEP method. However, it is not clear what the primary carriers of the magnetization are for these samples, particularly for samples from vent deposits, which are relatively unstudied in the paleointensity literature. To attempt to characterize the domain state of our samples, we obtained First Order Reversal Curves (FORCs, Pike et al., 1999) using the xFORC protocol of Zhao et al., 2017 on selected material from sites which passed BiCEP (and from some which failed). For this analysis we used sister specimens from the same samples for which the paleointensity results were acquired. FORCs are a qualitative way of assessing the domain state of a specimen using its hysteresis properties. Specimens which contain "Single-Domain" (SD) grains which are ideal for the paleointensity experiment will have FORCs with

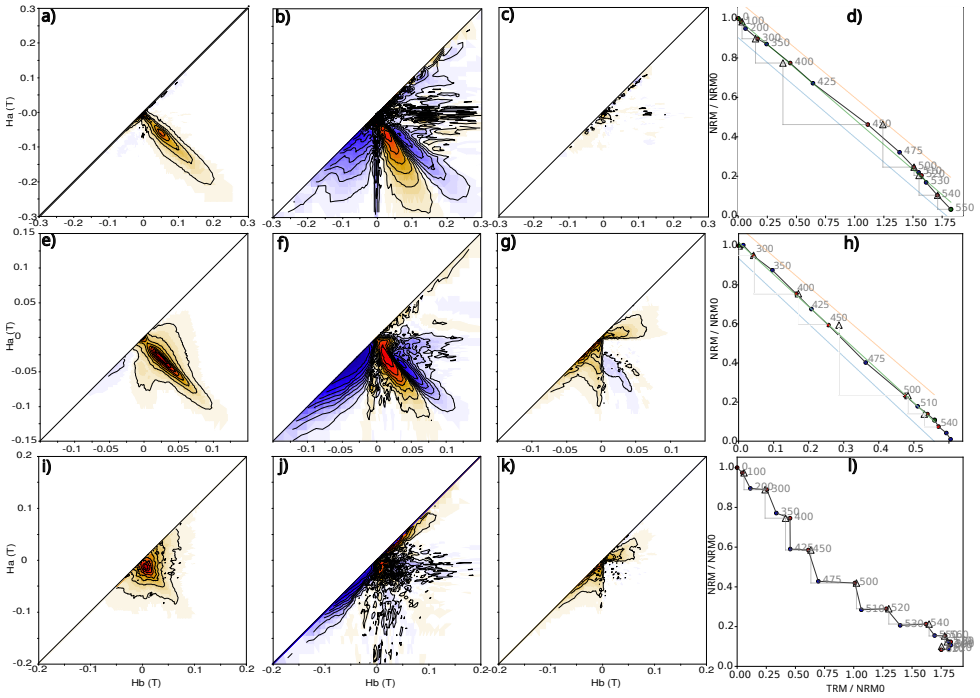

Figure 4.6. First Order Reversal Curves (FORCs) a),e),i), iFORCs b),f),j), and tForcs c),g),j) calculated using the xFORC protocol (Zhao et al., 2017). All FORCs calculated using a smoothing factor of 2 and a non-linear color scale of 1, except for iFORCs which were calculated using a smoothing factor of 3 and a non-linear color scale of 10. Arai plots are plotted in d),h),k). FORCs use sister specimens from two sites that yielded passing results: OA030 (top row), OA014 (center row) and a site which did not pass CCRIT or BiCEP, HW305 (bottom row). Sites which yielded specimens with linear Arai plots tend to have an elongated central ridge and have 3 lobes in the iFORC (top and center rows), whereas sites with curved Arai plots tend to have more spread along the $H_a = H_b$ direction and have extremely noisy iFORCs with little information.

a central ridge of positive values along the H_a=-H_b axis (see e.g. Fig. 4.6a). Specimens with higher numbers of non SD grains will have FORCs which have a spread along the H_a=H_b axis. The iFORC which represents the induced part of the magnetization displays a pattern of three distinct "lobes" (e.g. Fig 4.6b,e) for a sample containing SD grains, whereas it may display four "lobes" or be extremely noisy for samples containing non-SD grains. The tFORC represents "transient hysteresis" which occurs in non-SD grains, specimens with just noise on the tFORC (e.g. Fig. 4.6c) are most likely to be single domain.

Examples of FORCs and Arai plots for different samples are displayed in Fig. 4.6. Our FORC interpretations predominantly agree with our paleointensity results. In general, FORCs

we obtained from dike samples have pronounced central ridges and three lobes in the iFORC if visible, and effectively no tFORC (Fig. 4.6a-d). These samples generally had Arai plots which were straight lines, but sometimes underwent thermochemical alteration at high temperatures. Samples from our lava flows and vent deposits had central ridges, with small amounts of transient hysteresis and spreading along the $H_a = H_b$ axis. These samples still have linear Arai plots, and often have three lobes present in the iFORC, which suggests that the majority of carriers in these specimens are single domain. An example is shown in Fig. 4.6e-h. An example from a relatively coarse grained lava flow is given in Fig. 4.6i-l. Samples like these had highly curved or zig-zagging Arai plots (Fig. 4.6l) generally had no central ridge and lots of spreading along the $H_a = H_b$ axis (Fig. 4.6i). These samples had pronounced tFORCs(Fig. 4.6k) , and only noise in the iFORCs away from the H_a axis (Fig. 4.6j), observations which are consistent with the curved and zig-zagging Arai plots.

We also obtained Back Scattered Electron (BSE) images using an Scanning Electron Microscope (SEM), and Electron Dispersive X-Ray Spectroscopy (EDS) element maps to identify iron oxides in several thin sections taken from our samples. Several pictures from these analyses are displayed Fig. 4.7. Dike samples we analyzed contained no visible iron oxides in the glass, and almost no iron oxides in the groundmass. This is consistent with our FORCs and Arai plots (Fig. 4.6a-d), which are indicative of this specimen containing a predominance of single domain grains, which are 10s of nm in scale and not resolvable by the SEM used in this analysis. By contrast, samples from vent deposits contained numerous micron-scale iron bearing oxides in the groundmass, and in some cases, larger iron oxides on the scales of 100s of microns (Fig. 4.7a-d), size ranges where we would expect the grains to yield curved Arai plots. Many of these grains have elongated “cruciform” textures (Fig. 4.7c) or have heterogeneous compositions (Fig. 4.7a,d). One possibility is that these textures may persist to smaller scales, causing our larger grains to behave like assemblages of smaller, single domain, grains, due to their elongation or having smaller magnetic subregions separated by nonmagnetic lamellae. Another possibility is that these large grains do not contribute to the remanence. However, our lava flows and vent deposits

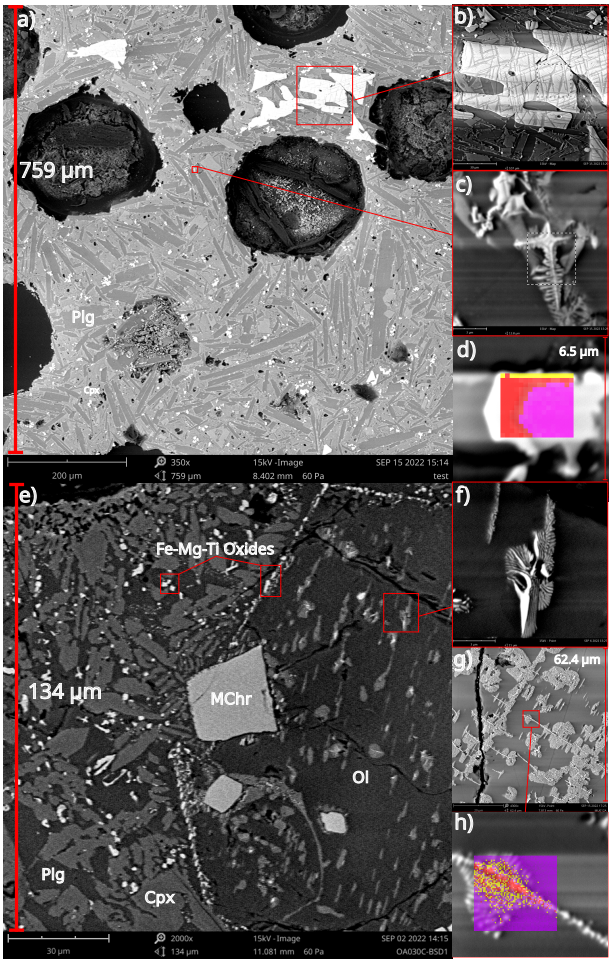

Figure 4.7. Back Scattered Electron (BSE) images and Electron Dispersive X-Ray Spectroscopy (EDS) maps of sister specimens from selected samples used in this study. Red text gives vertical field of view (FOV) for each image. a) BSE image of sample ML015A, a scoria vent deposit. b) Zoomed in image of large oxide in a), showing Fe-Ti exsolution textures. c) Zoomed in image of small oxide in a), showing elongate skeletal/cruciform structure. d) EDS element map of a typical oxide from another scoria vent deposit, ML012A, showing heterogeneous composition in the Iron-Titanium oxides. The atomic content of Si is shown in yellow, Fe in red, and Ti in pink. e) BSE image of sample OA030A, an agglutinated basanite vent deposit. f) Close up of high temperature alteration texture in olivine phenocryst. g) The same texture present in sample MU012A, a breccia from the bottom of a basanite lava flow. h) Close up of this texture with EDS element map. Colors are the same as d), with purple representing Mg. Note that the light colors in the BSE image represent an iron rich phase (interpreted as magnetite), which is surrounded by a phase richer in silicon than the surrounding olivine, interpreted as enstatite. Dominant mineral phases written on a) and e): Plg: plagioclase feldspar, Cpx: clinopyroxene, Ol: olivine, MChr: chrome spinel. Horizontal banding present in b),c),d),f),h) is an artifact of charging the sample that occurs in the SEM's EDS element mapping mode.

have much higher NRM moments than our dikes, with mass normalized NRMs on the order of 10^{-2} to 10^{-3} Am^2/kg, as opposed to the dikes which have moments on the order of 10^{-4} to 10^{-5} Am^2/kg.

Two thin sections from sites MU012 and OA030 have numerous olivine grains which exhibit an unusual texture, as displayed in Fig. 4.7e-h. This texture has been observed previously (e.g. Blondes et al., 2012; Ejima et al., 2017) and is interpreted as being caused by oxidation of olivine at temperatures above 800°C, which causes breakdown into an iron oxide (magnetite or hematite depending on formation conditions) and enstatite (see Fig. 4.7h and figure caption). The temperature of the oxidation means that the samples were oxidized prior to gaining a magnetization, which means the NRM is a primary TRM acquired during cooling. Oxidation of this kind seems to typically occur in fire fountaining strombolian type eruptions (e.g. Del Moro et al., 2013) where the lavas remain at high temperatures in an oxidizing environment for a while (e.g. 950 °C for 24-48 hours as per Haggerty and Baker, 1967). OA030 is an agglutinated basanitic vent deposit, agreeing with this oxidative environment, whereas the MU012 sample was taken from breccia/clinkers in an a‘ā lava flow, which may also undergo high temperature oxidation although the source is less clear.

Both sites with evidence for high temperature oxidation of olivines had highly linear Arai plots (see Figure 4.6h), with 16/16 specimens passing the strict CCRIT criteria for OA030, and 6/6 passing for MU012. Additionally a sample from OA030 has a FORC indicative of single-domain to single-vortex domain state, with a central ridge and three lobes in the iFORC (see Fig. 4.6, middle row). This indicates that the oxides formed by this breakdown may have extremely desirable properties for paleointensity experiments. Similar to the smaller oxides found in our other vent deposits (Fig. 4.7c), the elongation and finger-like structures present in these oxides could also explain their ideal behavior in the paleointensity experiment. These thin sections also contained numerous micron scale iron-titanium-magnesium oxides (interpreted as magnesioferrite) in the groundmass and around the outside of the olivine grains (Fig. 4.7e), but because the majority of the remanence unblocks between 400 and 600°C (see Fig. 4.6d), we

believe that magnetite is the dominant remanence carrier in these specimens.

Despite the large iron oxides observed in vent deposits and lava flows from this study, we conclude that these lithologies provide a good source for paleointensity estimates, as they have a high success rate relative to our other lithologies owing to their strikingly linear Arai plots (see Fig. 4.6, top row). Site MU113 provides further evidence for this, as material sampled from the inside of a lava tube gave an identical result to material sampled from a scoriaceous bomb entrained in the same flow. There are other reasons to favour these types of lithologies: The formation of these samples in an oxic environment at high temperature may help prevent thermochemical alteration during the paleointensity experiment, and fresh scoria is also easy to come by in Hawai‘i, as many scoria cones are quarried. However, most preserved vent deposits are typically formed during the later stages of Hawai‘ian volcanism, and consequently we have no results from scoria older than 2 Ma.

4.3.3 Temporal Distributions of Intensity

Mismatch between the observed distribution of paleointensities with latitude and the expected distribution for a GAD field Fig 4.1a) could potentially be caused by inconsistencies in treatment of data among different paleointensity studies. To compare the time-averaged field from our model to data from different latitudes, we reanalyzed results from recent paleomagnetic studies in Northern Israel (Tauxe et al., 2022) and Antarctica (Asefaw et al., 2021) using the BiCEP method and the same criteria used for the Hawai‘i samples. Tables of results from these re-analyses can be found in the Supporting Information. Each of these studies yielded passing sites with results spanning the past 2.5 Myr. For direct comparisons between locations, we convert each paleointensity result to a Virtual Axial Dipole Moment (VADM) which is the moment of the geocentric dipole (measured in ZAm^2) that would yield the observed paleointensity at a given latitude. Our average VADM for Hawai‘i is 62.4 ZAm^2, which is similar to the 64.2 ZAm^2 value from Israel, but is significantly higher than the average in Antarctica (39.6 ZAm^2). Plots of VADMs with age for each location are shown in Figures 4.8a)-c), with average VADMs plotted

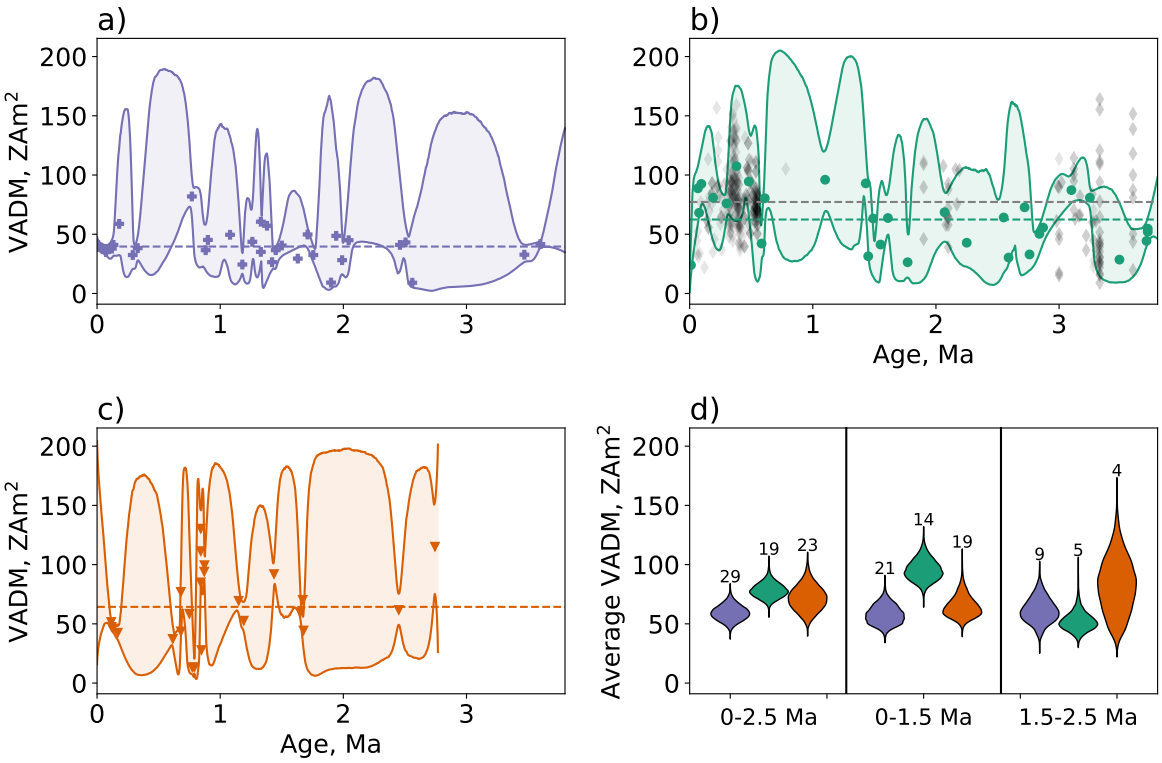

Figure 4.8. a) - c) Plots of VADM against age (symbols), and 95% credible envelopes for AH-RJMCMC models (Livermore et al., 2018) (shaded areas) for studies from a) Antarctica (purple plus symbols), b) Hawai'i (green dots), and c) Israel (orange triangles). Horizontal dashed lines are the average VADM of all paleointensity estimates (symbols) for each plot. In b), all unfiltered data in the MagIC database from Hawaii aged between 50 ka and 3.8 Ma are plotted as grey diamonds, and the average VADM from these data are plotted as a grey horizontal line. d) Violin plots showing the distribution of averaged VADMs over different time periods, numbers refer to the number of paleointensity within these temporal ranges, although data outside these ranges may also contribute to these averages. Data from Hawai'i have a significantly higher average VADM than in Israel and Antarctica over the past 1.5 Ma, which is reflected in the averages from 0-2.5 Ma. Average VADMs for data older than 1.5 Ma appears to agree for all three locations.

as horizontal dashed lines. In Figure 4.8b we also plot all the data from Hawai‘i in the MagIC database from this time interval in grey. The unfiltered data have a significantly higher variance than our data, and the weaker field seen prior to 1.5 Ma in our data is not apparent in the unfiltered Hawai‘ian data, which have an average VADM of 77.2 ZAm^2. These differences could occur because more field variation is being captured by the larger dataset, or because the unfiltered data have more variance due to inconsistency in their analysis (for example, preferentially taking the low temperature steps in a potentially sagging Arai plot). Despite the consistency in analysis of our data, the average VADM in Hawai‘i and Israel is still very different to that found in Antarctica, indicating that inconsistency in analyses and biased paleointensities caused by Arai plot curvature are not the source of this mismatch.

Taking an average VADM of the entire age range of our data may not be representative of the time-averaged field, because our data have different temporal distributions, with no data in Israel older than 2.75 Ma. In Hawai‘i, this average does not capture the change in average field strength seen at 1.5 Ma, and in Israel, we have many paleointensity data which record a strong field and come from a small range of time around 850 ka B.P. Because this time interval is oversampled, it will bias our average VADM towards these higher values. To account for these problems, it would make sense to fit a curve to our VADMs and take an average of the curve over an interval of interest. We do this using the “Age Hyperparameter Reversible Jump Markov Chain Monte Carlo” (AH-RJMCMC) method (Livermore et al., 2018). This model fits piecewise linear curves to paleointensity data in a probabilistic fashion, with curves with less linear pieces being preferable for the model. At times when there are few data, the model uncertainties become very large and revert to a uniform prior distribution, which we set as 0-220 ZAm^2. At times where we have no data, the uncertainty in the average VADM will increase, and so any differences in the average VADM using this method are driven by the data.

We computed the AH-RJMCMC models, which output a series of possible piecewise linear curves at each locality. We took the average value of each curve over the past 2.5 Ma, and converted these averages to VADMs. The models produced by this analysis are shown in

Fig. 4.8a-c, and the distributions of the time-averaged VADMs for each locality are plotted on the violin plots in Figure 4.8d. Using this methodology, it is apparent that the time-averaged VADMs over the last 1.5 Ma from Hawai‘i and Antarctica are indeed not consistent with each other, but the time-averaged VADM in Israel could be compatible with either of the other latitudes. However, there is not enough evidence to confirm a difference in the temporal average between Hawai‘i and Antarctica from 1.5-2.5 Ma, with the average VADMs appearing consistent. This implies that poor temporal sampling is not the reason for inconsistent paleointensities at different latitudes, but that some form of genuine non-dipolar field behavior that causes higher fields in Hawai‘i than Antarctica at least until 1.5 Ma. More paleointensity studies with high quality paleointensity data at different latitudes (especially from the southern hemisphere) are needed to better understand the sources of this non-dipolar behavior.

4.4 Conclusions

In this paper, we obtained 31 high quality paleointensity results from dikes, lava flow tops and vent deposits collected in the Hawai‘ian islands, with ages ranging from 0-4 Ma. We demonstrate a methodology for obtaining accurate time-averaged paleointensities, with uncertainties which allow direct comparison between paleointensity studies at different latitudes. The use of BiCEP allows for consistent comparison of results between different studies, and using the methodology of Livermore et al., 2018 allows us to obtain a time-averaged intensity, with uncertainty, which accounts for the temporal distribution of our paleointensity. Because these robust statistical approaches are used for calculating time-averaged paleointensities, we are able to exclude the hypotheses that inconsistency of our time-averaged VADMs is due to either biased paleointensity data, or inconsistent temporal sampling of paleointensities.

Applying the new methodology to data from the Hawai‘ian islands, we find that the time-averaged paleointensity in Hawai‘i over the past 1.5 Ma was higher than during the period from 1.5-4 Ma. Comparing results from paleointensity studies at three latitudes, we find that

this period of high paleointensity is not recorded in rocks from Antarctica or Israel. We reiterate the conclusion of other recent papers (e.g. Tauxe et al., 2022) that the Earth's magnetic field averaged over the past 1.5 Ma does not conform to a Geocentric Axial Dipole. Further time averages at a greater range of latitudes and times will be needed to obtain better estimates of the structure of this time-averaged field.

Our results also indicate that vent deposits containing scoria and olivine bearing rocks which are oxidized at high temperatures are potentially good lithologies for obtaining high quality paleointensity estimates, with higher success rates in the paleointensity experiment. Specimens from these lithologies have strong magnetizations and tend to alter less in paleointensity experiments. Additionally, these deposits are frequently quarried, allowing for easy access to fresh material in the field. Despite their useful properties in paleointensity experiments, and their single-domain like FORCs, the size of iron oxides in these samples when viewed under a microscope is orders of magnitude larger than would be expected for single domain grains. Further study of the magnetic carriers in these samples should be undertaken to understand why they have such ideal rock magnetic properties.

Data Archival

All data and interpretations are available at https://earthref.org/MagIC/19614/9208acad-0f62-4d9e-b265-4c8907d40eb7 and will be made available in the MagIC database at earthref.org/MagIC/19614 on acceptance of this paper.

4.5 Acknowledgements

This work was partially supported by EAR1827263 to LT. We would like to thank the Hawai‘i Department of Land and Natural Resources' Forestry and Wildlife Program for issuing sampling permits in West Maui, and Moloka‘i Ranch Ltd. and Moloka‘i Land Trust for allowing us to sample on their land. Finally we would like to thank the late Jasper Konter for his help in the field and hospitality during our field work on Oahu. He will be missed.

Chapter 4, is in preparation submission with Proceedings of the National Academy of Sciences as Cych, B., Tauxe, L., Cromwell, G., Sinton, H., & Koppers, A. Non-dipolar time averaged field for the past ‘million years from new paleointensity results from Hawai‘i, compared to global datasets (In Preparation). The dissertation author was the primary investigator and author of this paper.

References

Asefaw, H., Tauxe, L., Koppers, A. A. P., & Staudigel, H. (2021). Four-Dimensional Paleomagnetic Dataset: Plio-Pleistocene Paleodirection and Paleointensity Results From the Erebus Volcanic Province, Antarctica. *J. Geophys. Res. Solid Earth*, *126*(2), e2020JB020834. https://doi.org/10.1029/2020JB020834

Blondes, M. S., Brandon, M. T., Reiners, P. W., Page, F. Z., & Kita, N. T. (2012). Generation of Forsteritic Olivine (Fo99·8) by Subsolidus Oxidation in Basaltic Flows. *J. Petrol.*, *53*(5), 971–984. https://doi.org/10.1093/petrology/egs006

Cai, S., Tauxe, L., & Cromwell, G. (2017). Paleointensity From Subaerial Basaltic Glasses From the Second Hawaii Scientific Drilling Project (HSDP2) Core and Implications for Possible Bias in Data From Lava Flow Interiors. *J. Geophys. Res. Solid Earth*, *122*(11), 8664–8674. https://doi.org/10.1002/2017JB014683

Coe, R. S. (1967). The determination of paleo-intensities of the earth's magnetic field with emphasis on mechanisms which could cause non-ideal behavior in thellier's method. *J. Geomag. Geoelectr.*, *19*, 157–178.

Cromwell, G., Johnson, C. L., Tauxe, L., Constable, C. G., & Jarboe, N. A. (2018). PSV10: A Global Data Set for 0–10 Ma Time-Averaged Field and Paleosecular Variation Studies. *Geochem. Geophys. Geosyst.*, *19*(5), 1533–1558. https://doi.org/10.1002/2017GC007318

Cromwell, G., Tauxe, L., Staudigel, H., Constable, C. G., Koppers, A. A. P., & Pedersen, R.-B. (2013). In search of long-term hemispheric asymmetry in the geomagnetic field: Results from high northern latitudes. *Geochem. Geophys. Geosyst.*, *14*(8), 3234–3249. https://doi.org/10.1002/ggge.20174

Cromwell, G., Tauxe, L., Staudigel, H., & Ron, H. (2015). Paleointensity estimates from historic and modern Hawaiian lava flows using glassy basalt as a primary source material. *Phys. Earth Planet. Inter.*, *241*, 44–56. https://doi.org/10.1016/j.pepi.2014.12.007

Cych, B., Morzfeld, M., & Tauxe, L. (2021). Bias Corrected Estimation of Paleointensity (BiCEP): An Improved Methodology for Obtaining Paleointensity Estimates. *Geochem. Geophys. Geosyst.*, *22*(8), e2021GC009755. https://doi.org/10.1029/2021GC009755

Del Moro, S., Renzulli, A., Landi, P., La Felice, S., & Rosi, M. (2013). Unusual lapilli tuff ejecta erupted at Stromboli during the 15 March 2007 explosion shed light on the nature and thermal state of rocks forming the crater system of the volcano. *J. Volcanol. Geotherm. Res.*, *254*, 37–52. https://doi.org/10.1016/j.jvolgeores.2012.12.017

Dunlop, D. J., & Özdemir, Ö. (2000). Effect of grain size and domain state on thermal demagnetization tails. *Geophys. Res. Lett.*, *27*, 1311–1314.

Dunlop, D., & Özdemir, Ö. (2001). Beyond Néel's theories: Thermal demagnetization of narrow-band partial thermoremanent magnetization. *Phys. Earth Planet. Int.*, *126*, 43–57.

Ejima, T., Yoneda, M., Akasaka, M., Ohfuji, H., Kon, Y., Nagashima, M., & Nakamuta, Y. (2017). Precipitates within olivine phenocrysts in oxidized andesitic scoria from Kasayama volcano, Hagi, Japan. *J. Mineral. Petrol. Sci.*, *112*(3), 116–126. https://doi.org/10.2465/jmps.161219

Haggerty, S. E., & Baker, I. (1967). The alteration of olivine in basaltic and associated lavas. *Contrib. Mineral. Petrol.*, *16*(3), 233–257. https://doi.org/10.1007/BF00371094

Koppers, A. A. P., Staudigel, H., & Wijbrans, J. R. (2000). Dating crystalline groundmass separates of altered Cretaceous seamount basalts by the 40Ar/39Ar incremental heating technique. *Chem. Geol.*, *166*(1), 139–158. https://doi.org/10.1016/S0009-2541(99)00188-6

Krása, D., Heunemann, C., Leonhardt, R., & Petersen, N. (2003). Experimental procedure to detect multidomain remanence during thellier–thellier experiments [Paleo, Rock and Environmental Magnetism 2002]. *Phys. Chem Earth (A/B/C)*, *28*(16), 681–687. https://doi.org/10.1016/S1474-7065(03)00122-0

Lawrence, K. P., Tauxe, L., Staudigel, H., Constable, C., Koppers, A., McIntosh, W. C., & Johnson, C. L. (2009). Paleomagnetic field properties near the southern hemisphere

tangent cylinder. *Geochem. Geophys. Geosyst.*, *10*, Q01005. https://doi.org/doi:10.1029/2008GC002072

Levi, S. (1977). The effect of magnetite particle size on paleointensity determinations of the geomagnetic field. *Phys. Earth Planet. Inter.*, *13*(4), 245–259. https://doi.org/10.1016/0031-9201(77)90107-8

Livermore, P. W., Fournier, A., Gallet, Y., & Bodin, T. (2018). Transdimensional inference of archeomagnetic intensity change. *Geophys. J. Int.*, *215*(3), 2008–2034. https://doi.org/10.1093/gji/ggy383

Muxworthy, A. R. (2017). Considerations for Latitudinal Time-Averaged-Field Palaeointensity Analysis of the Last Five Million Years. *Front. Earth Sci.*, *0*. https://doi.org/10.3389/feart.2017.00079

Nagata, T., Arai, Y., & Momose, K. (1963). Secular variation of the geomagnetic total force during the last 5000 years. *J. Geophys. Res.*, *68*, 5277–5282.

Néel, L. (1949). Théorie du traınage magnétique des ferromagnétiques en grains fins avec applications aux terres cuites. *Ann. géophys.*, *5*, 99–136.

Ozawa, A., Tagami, T., & Garcia, M. O. (2005). Unspiked K–Ar dating of the Honolulu rejuvenated and Ko‘olau shield volcanism on O‘ahu, Hawai‘i. *Earth Planet. Sci. Lett.*, *232*(1), 1–11. https://doi.org/10.1016/j.epsl.2005.01.021

Paterson, G., Tauxe, L., Biggin, A., Shaar, R., & Jonestrask, L. (2014). On improving the selection of thellier-type paleointensity data. *Geochemistry Geophysics Geosystems*, *15*(4). https://doi.org/10.1002/2013GC005135

Paterson, G. A. (2011). A simple test for the presence of multidomain behavior during paleointensity experiments. *J. Geophys. Res. Solid Earth*, *116*(B10). https://doi.org/10.1029/2011JB008369

Pike, C., Roberts, A., & Verosub, K. (1999). Characterizing interactions in fine magnetic particle systems using first order reversal curves. *J. Appl. Phys.*, *85*, 6660–6667.

Selkin, P., & Tauxe, L. (2000). Long-term variations in paleointensity. *Phil. Trans. Roy. Soc. Lond.*, *358*, 1065–1088.

Shcherbakov, V. P., Khokhlov, A. V., & Sycheva, N. K. (2019). Analysis of the Hypothesis of a Giant Gaussian Process as a Means for Describing Secular Variations of the Geomagnetic Field Vector. *Izv. Phys. Solid Earth*, *55*(1), 182–194. https://doi.org/10.1134/S1069351319010099

Sherrod, D. R., Nishimitsu, Y., & Tagami, T. (2003). New K-Ar ages and the geologic evidence against rejuvenated-stage volcanism at Haleakala, East Maui, a postshield-stage volcano of the Hawaiian island chain. *GSA Bulletin*, *115*(6), 683–694. https://doi.org/10.1130/0016-7606(2003)115⟨0683:NKAATG⟩2.0.CO;2

Sherrod, D. R., Sinton, J. M., Watkins, S. E., & Brunt, K. M. (2007). Geologic map of the state of hawaii. *US geological survey open-file report*, *1089*, 83.

Tagami, T., Nishimitsu, Y., & Sherrod, D. R. (2003). Rejuvenated-stage volcanism after 0.6-m.y. quiescence at West Maui volcano, Hawaii: new evidence from K–Ar ages and chemistry of Lahaina Volcanics. *J. Volcanol. Geotherm. Res.*, *120*(3), 207–214. https://doi.org/10.1016/S0377-0273(02)00385-2

Tauxe, L. (2006). Long-term trends in paleointensity: The contribution of dsdp/odp submarine basaltic glass collections. *Physics of the Earth and Planetary Interiors*, *156*(3-4), 223–241.

Tauxe, L., Asefaw, H., Behar, N., Koppers, A. A. P., & Shaar, R. (2022). Paleointensity Estimates from the Pleistocene of Northern Israel: Implications for hemispheric asymmetry in the time-averaged field. *Geochem. Geophys. Geosyst.*, *n/a*(n/a), e2022GC010473. https://doi.org/10.1029/2022GC010473

Tauxe, L., Santos, C. N., Cych, B., Zhao, X., Roberts, A. P., Nagy, L., & Williams, W. (2021). Understanding nonideal paleointensity recording in igneous rocks: Insights from aging experiments on lava samples and the causes and consequences of 'fragile' curvature in

arai plots. *Geochem. Geophys. Geosyst.*, *22*(1), e2020GC009423. https://doi.org/10.1029/2020GC009423

Tauxe, L., & Love, J. J. (2003). Paleointensity in Hawaiian Scientific Drilling Project Hole (HSDP2): Results from submarine basaltic glass. *Geochem. Geophys. Geosyst.*, *4*(2). https://doi.org/10.1029/2001GC000276

Tauxe, L., & Staudigel, H. (2004). Strength of the geomagnetic field in the Cretaceous Normal Superchron: New data from submarine basaltic glass of the Troodos Ophiolite. *Geochem. Geophys. Geosyst.*, *5*(2). https://doi.org/10.1029/2003GC000635

Wilson, R. L. (1970). Permanent Aspects of the Earth's Non-dipole Magnetic Field over Upper Tertiary Times. *Geophys. J. Int.*, *19*(4), 417–437. https://doi.org/10.1111/j.1365-246X.1970.tb06056.x

Yu, Y., Tauxe, L., & Genevey, A. (2004). Toward an optimal geomagnetic field intensity determination technique. *Geochem. Geophys. Geosyst.*, *5*(2). https://doi.org/10.1029/2003GC000630

Zhao, X., Roberts, A. P., Heslop, D., Paterson, G. A., Li, Y., & Li, J. (2017). Magnetic domain state diagnosis using hysteresis reversal curves. *J. Geophys. Res. Solid Earth*, *122*(7), 4767–4789. https://doi.org/10.1002/2016JB013683

Ziegler, L., Constable, C., Johnson, C. L., & Tauxe, L. (2011). Padm2m: A penalized maximum likelihood model of the 0-2 ma paeomagneitc axial dipole moment. *Geophys. J. Int.*, *184*, 1069–1089.

Zijderveld, J. D. A. (1967). *A.c. demagnetization of rocks: Analysis of results*. Chapman; Hall.

Conclusions

In this dissertation I outline a clear methodology for producing a robust time averaged paleointensity, with uncertainties, from accurate and precise paleointensity data. In Chapter 4, I outlined how targeting the right kind of materials in the field and using the correct lab protocols can be used to obtain good quality results from paleointensity experiments. In Chapter 2 I explained how to analyze paleointensity data to obtain accurate and precise estimates of paleointensity without excluding specimens from the analysis based on arbitrary selection criteria. In Chapter 3, I demonstrated how to select the correct range of temperature steps to isolate a paleointensity acquired in a single field. Finally, I demonstrate how to obtain a time averaged paleointensity in Chapter 4.

Put together, this dissertation gives us a protocol that can be followed and reproduced by undertaking paleointensity studies at other latitudes. I present data from Hawai‘i, and compare it to two other such studies where we reanalyzed the data consistently in Chapter 4. For the three hypotheses that I outlined in Section 1.4.1, I can rule out the hypothesis that the temporal distribution of paleointensity estimates is not the reason for apparent non-dipole field behavior over the past 1.5 Ma, as Antarctica has a consistently different Virtual Axial Dipole Moment (VADM) to Hawai‘i over this time period. Because the robust BiCEP method I used to estimate our paleointensities yields accurate paleointensities when applied to historical lava flows, it is very unlikely that this difference between Hawai‘i and Antarctica is caused by bias in the paleointensity estimates. Consequently, I must conclude that there is a non-dipole component to the time averaged field over this timescale. This is an unexpected result, and it is hard to imagine what kind of field behavior or mechanism could cause this long term non-dipole behavior.

Determining the exact structure of this time averaged field is not possible from just three study locations, and so more paleointensity studies (particularly at poorly sampled latitudes in the database) are needed to constrain this structure.

Appendix A

Supplementary Information for Chapter 2

Contents of this file

1. Figure A1- plots of B_{exp} vs B_{anc} results using the BiCEP method and other sets of selection criteria used in this study.

2. Figure A2- plots of the credible intervals of BiCEP results for subsamples of 5 specimens from sites hw108, VM and hw123.

3. Figures A3 to A32- plots of unpooled B_{anc} and $\vec{k}$ fits and the BiCEP method applied to all sites used in this study.

Additional Supporting Information (Files uploaded separately)

1. Caption for Data Set S1 - This dataset contains the full set of results shown in Figure 2.7 and described in Section 2.3. It is available for download at the original version of this paper at https://doi.org/10.1029/2021GC009755

Data Set S1 This data set is a csv file containing the full set of results for each version of the BiCEP method (Linear, Quadratic and Cubic models with 5, 10 and 20 μT standard deviations on the prior for σ_{site}, vs classic selection criteria. Columns are named by the model class (Linear, Quadratic, Cubic, or Selection Criteria, followed by the criteria name/prior standard deviation in μT) and then the parameter in question, either the percentile (2.5%, 50%, 97.5%) or f_prob for the f_{prob} value. The first three columns of the csv file contain the site name, B_{exp} and number of specimens (M).

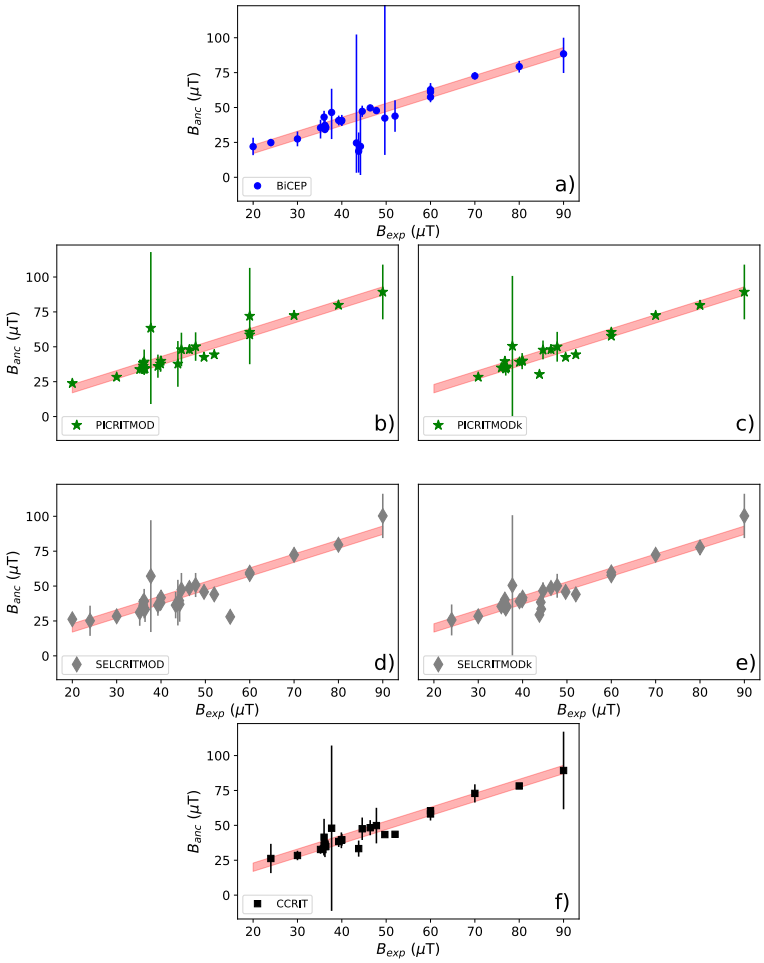

Figure A.1. Plot of B_{anc} vs B_{exp} for the passing sites for BiCEP and each of our criteria. a) BiCEP results, b) PICCRITMOD results, c) PICRITMODk results, d) SELCRIT results, e) SELCRITMODk results, f) CCRIT results. Red shaded area represents B_{exp} ± 3 μT. Note that for BiCEP, most of the severely underestimated and highly inaccurate results are for sites with low numbers of specimens ($M < 5$) which did not pass the other criteria.

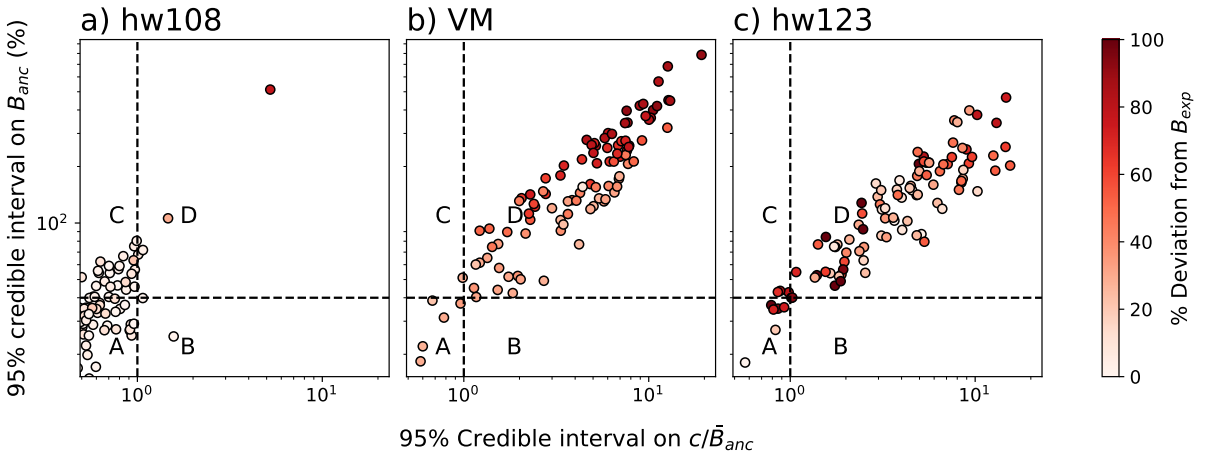

Figure A.2. Figure in the style of Figure 9 from the main manuscript. A set of 100 random subsamples of 5 specimens were taken from a) site hw108, b) site VM and c) site hw123. We applied the BiCEP method to these subsamples and plot the 95% credible interval on B_{anc} as a percentage against the 95% credible interval on $c/\tilde{B}_{anc}$. Dashed lines represent the boundaries between regions A, B, C and D, defined in Section 4.3, and point colors represent the percentage deviation of the median from B_{exp}. It is apparent that for our sites where we obtain an inaccurate answer for a large number of specimens (VM and $hw123$), our subsamples fall in region D the majority of the time (94/100 and 90/100 respectively). Conversely, our site hw108, which returns an accurate result with higher numbers of specimens only has 5 results in region D. This indicates that we should either continue measuring specimens (region C, 24/100 results) or that we have a relatively accurate and precise result already (regions A and B, 66/100 results). Although few sites/subsamples were used for this analysis, this test indicates that our labelling scheme has a predictive accuracy of 90-95%

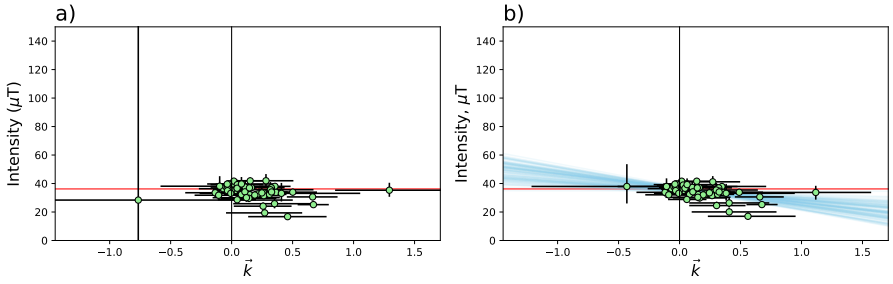

Figure A.3. Unpooled and BiCEP models applied to site 1991-1992 Eruption Site a). Independent estimates of B_{anc} vs $\vec{k}$ are plotted as green circles. These estimates are made without assuming a linear relationship between B_{anc} and $\vec{k}$, similar to the analysis in Figure 9a) Black error bars represent 95% credible intervals. b) The BiCEP method applied to this site. Green circles and error bars represent the 95% credible interval, blue lines represent draws from the posterior distribution. The $\vec{k}$=0 axis is plotted as a black vertical line, and the value of B_{exp} is plotted as a red horizontal line. For an accurate estimate, the blue lines should cross the point where these two lines intersect.

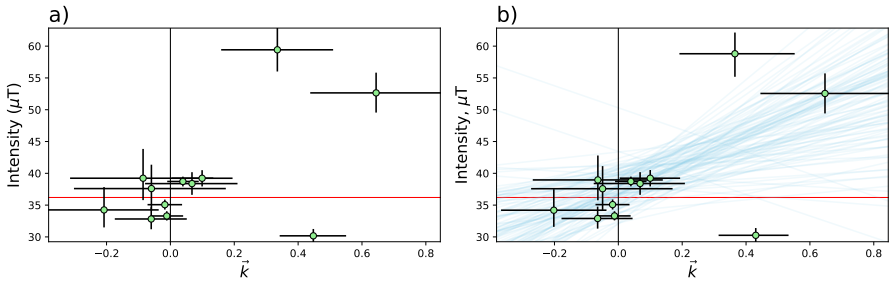

Figure A.4. The same methodology described in the caption for Figure A3 applied to site BBQ

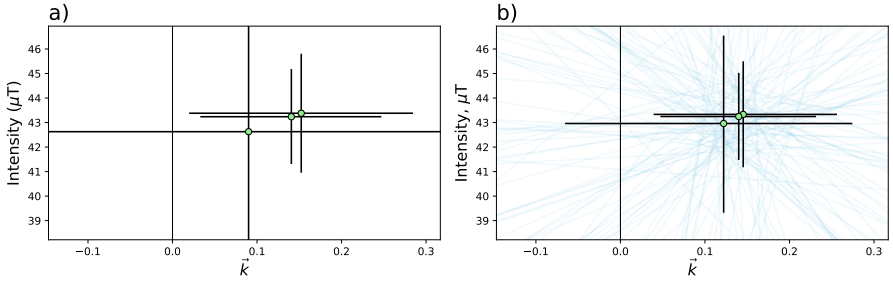

Figure A.5. The same methodology described in the caption for Figure A3 applied to site BR06

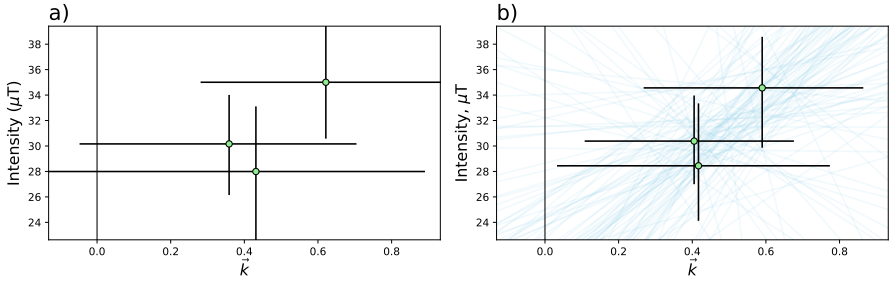

Figure A.6. The same methodology described in the caption for Figure A3 applied to site ET1

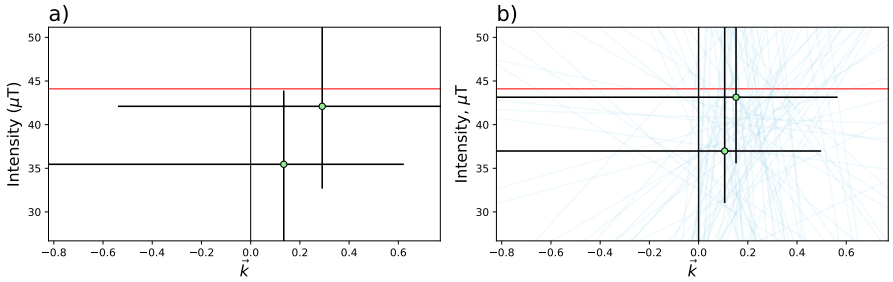

Figure A.7. The same methodology described in the caption for Figure A3 applied to site ET2

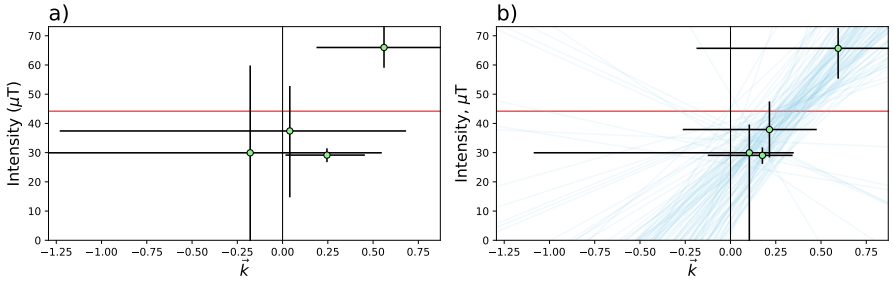

Figure A.8. The same methodology described in the caption for Figure A3 applied to site ET3

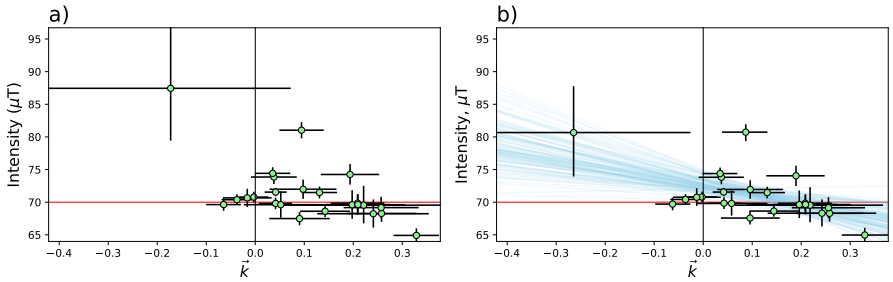

Figure A.9. The same methodology described in the caption for Figure A3 applied to site FreshTRM

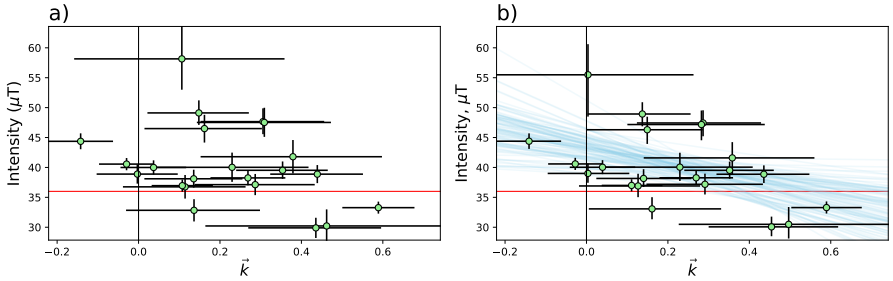

Figure A.10. The same methodology described in the caption for Figure A3 applied to site Hawaii 1960 Flow

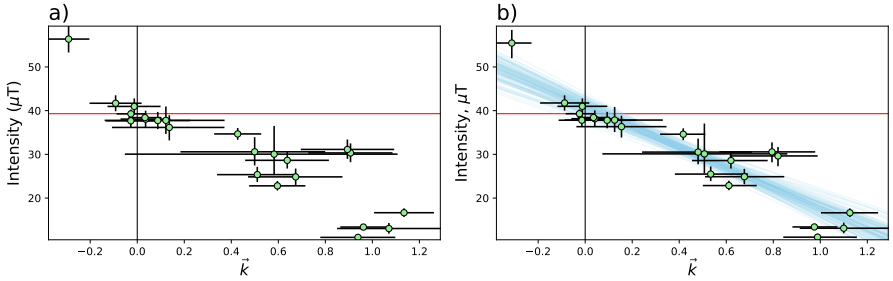

Figure A.11. The same methodology described in the caption for Figure A3 applied to site hw108

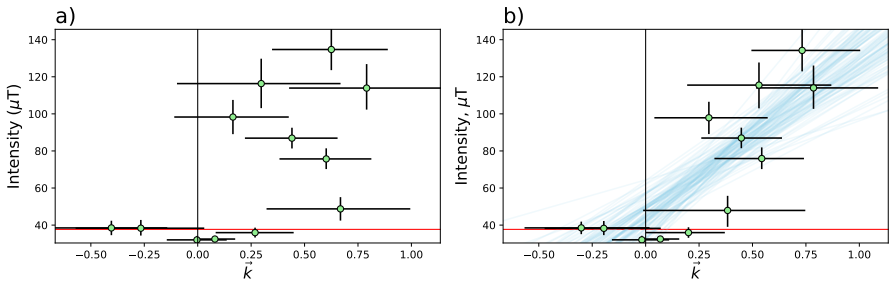

Figure A.12. The same methodology described in the caption for Figure A3 applied to site hw123

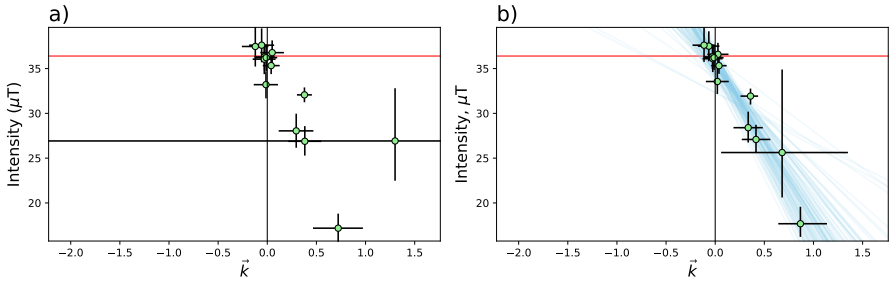

Figure A.13. The same methodology described in the caption for Figure A3 applied to site hw126

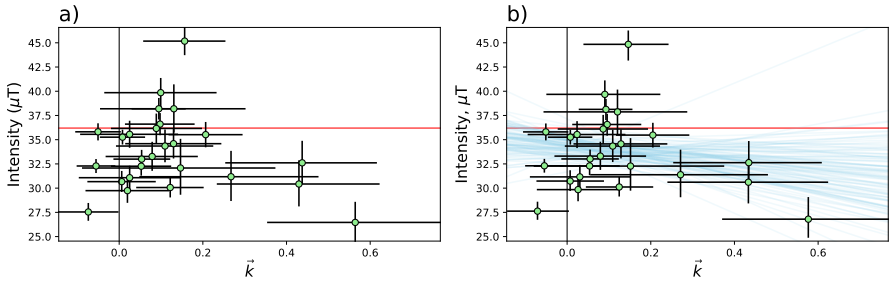

Figure A.14. The same methodology described in the caption for Figure A3 applied to site hw128

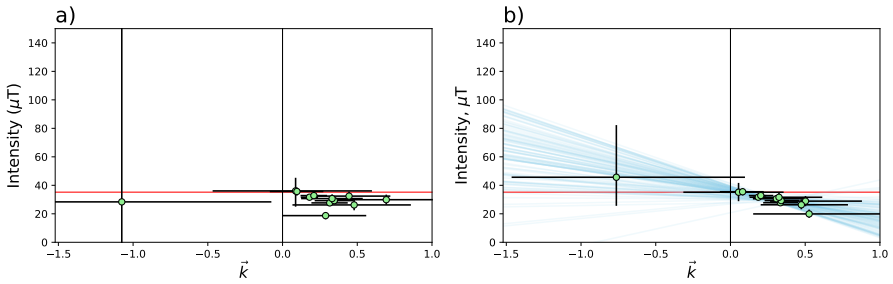

Figure A.15. The same methodology described in the caption for Figure A3 applied to site hw201

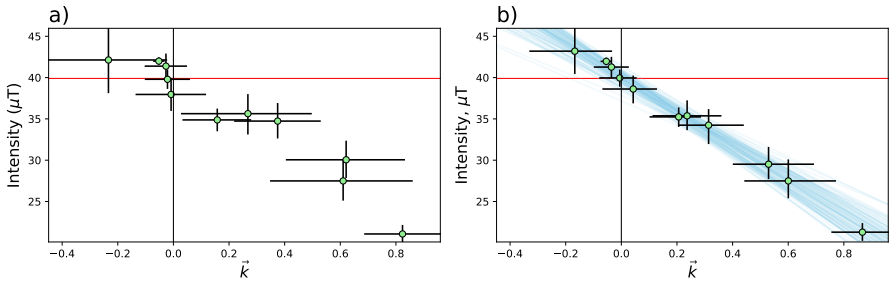

Figure A.16. The same methodology described in the caption for Figure A3 applied to site hw226

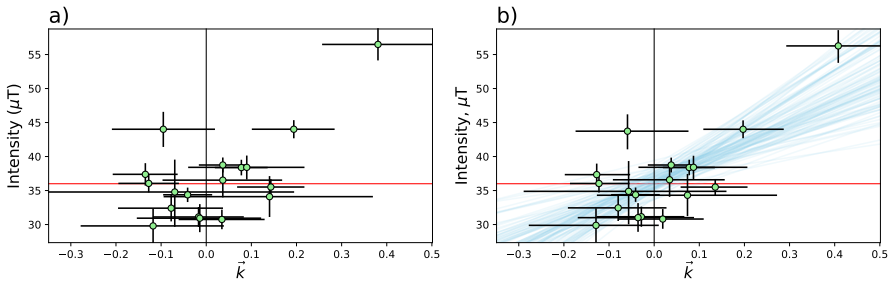

Figure A.17. The same methodology described in the caption for Figure A3 applied to site hw241

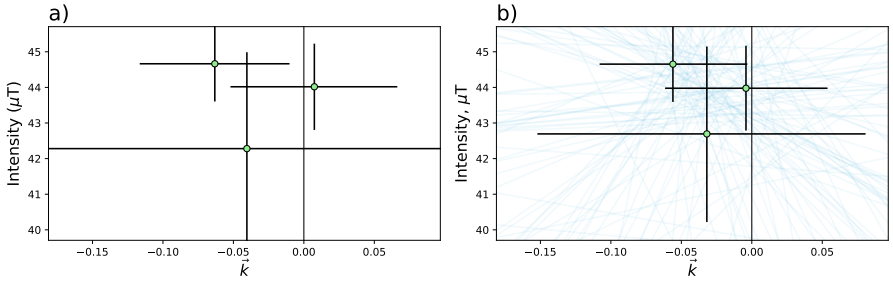

Figure A.18. The same methodology described in the caption for Figure A3 applied to site kf

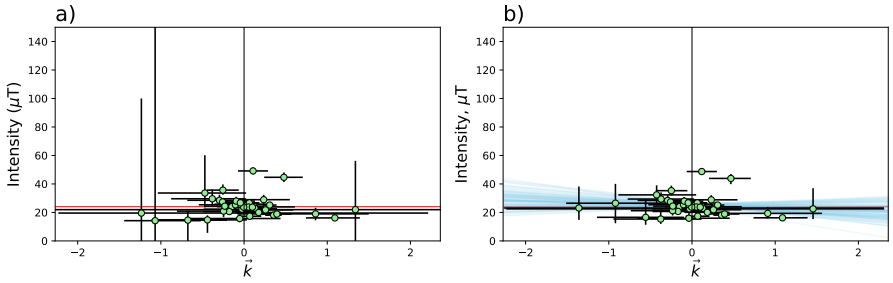

Figure A.19. The same methodology described in the caption for Figure A3 applied to site LV

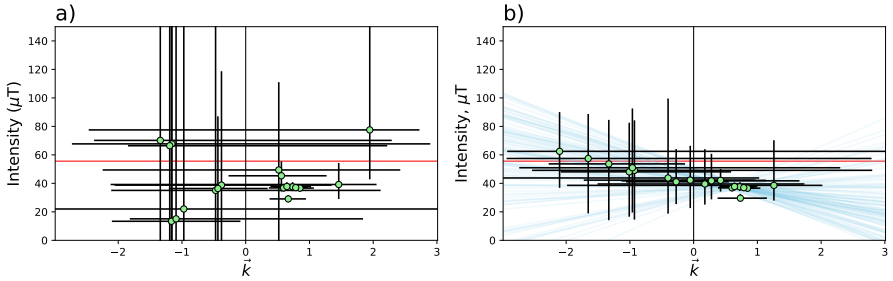

Figure A.20. The same methodology described in the caption for Figure A3 applied to site MSH

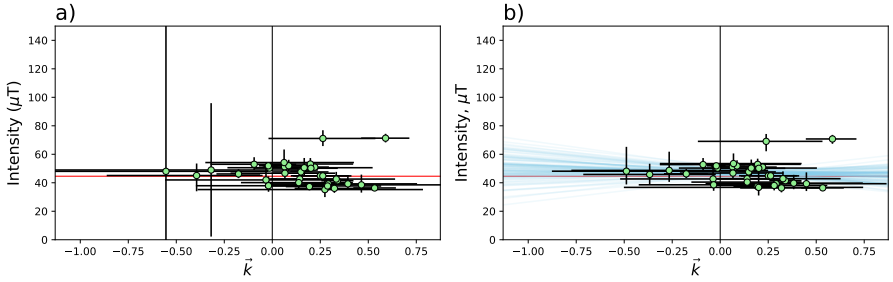

Figure A.21. The same methodology described in the caption for Figure A3 applied to site P

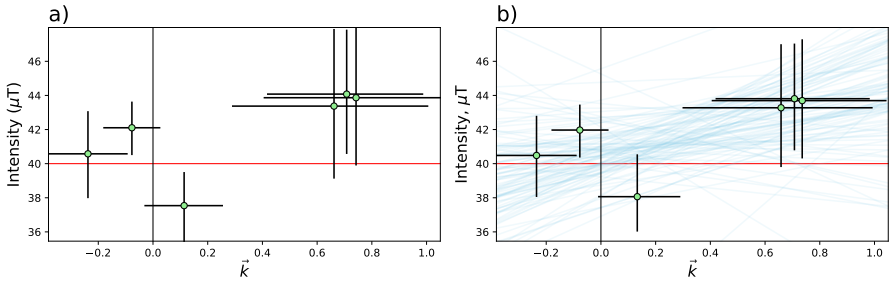

Figure A.22. The same methodology described in the caption for Figure A3 applied to site remag-rs61

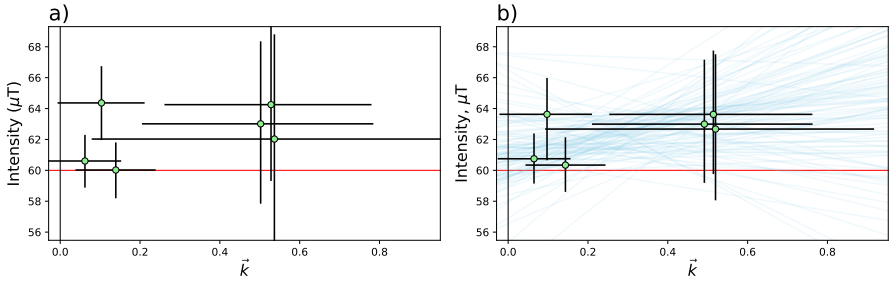

Figure A.23. The same methodology described in the caption for Figure A3 applied to site remag-rs62

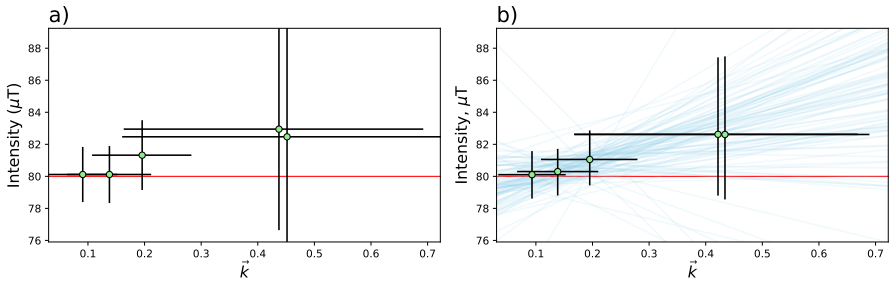

Figure A.24. The same methodology described in the caption for Figure A3 applied to site remag-rs63

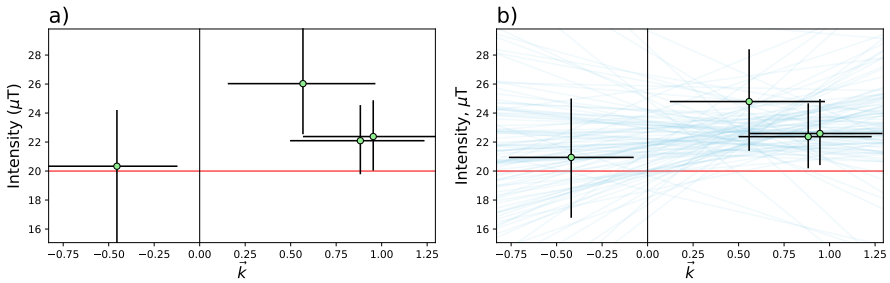

Figure A.25. The same methodology described in the caption for Figure A3 applied to site remag-rs78

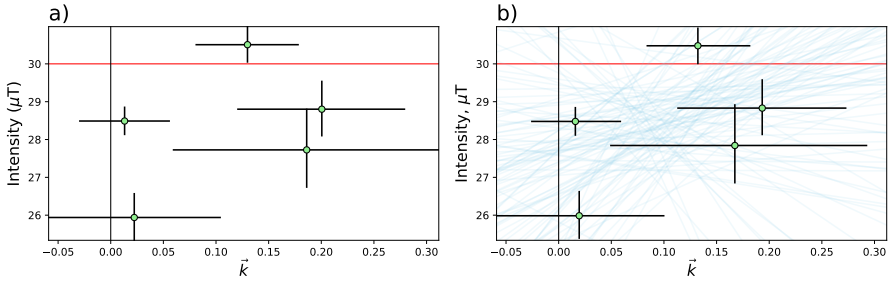

Figure A.26. The same methodology described in the caption for Figure A3 applied to site rs25

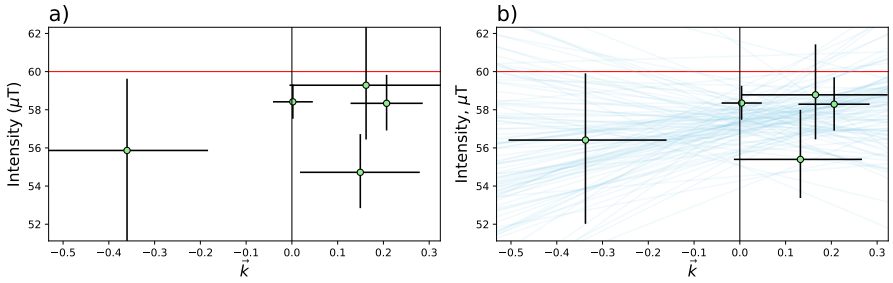

Figure A.27. The same methodology described in the caption for Figure A3 applied to site rs26

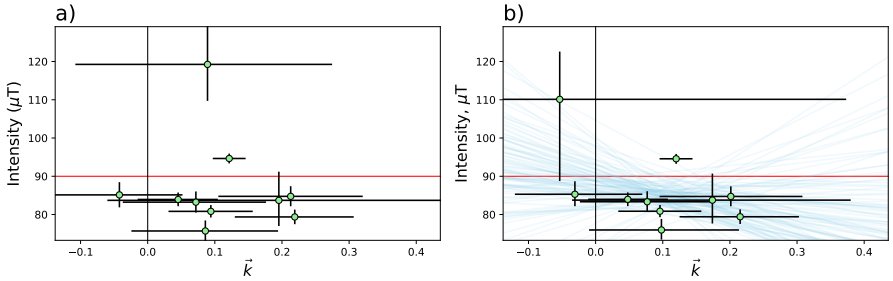

Figure A.28. The same methodology described in the caption for Figure A3 applied to site rs27

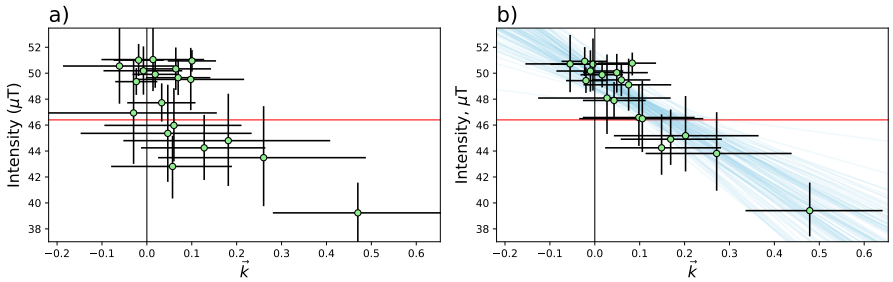

Figure A.29. The same methodology described in the caption for Figure A3 applied to site SW

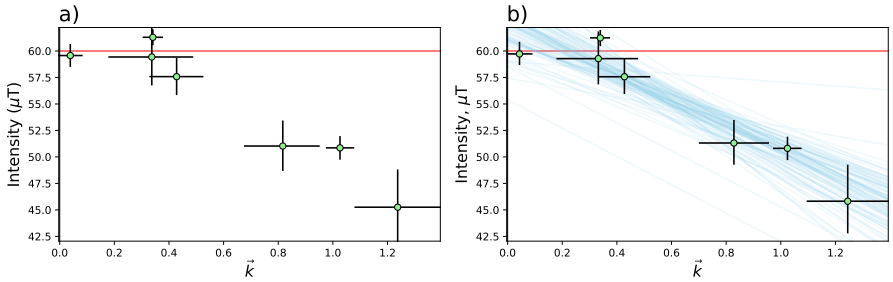

Figure A.30. The same methodology described in the caption for Figure A3 applied to site Synthetic60

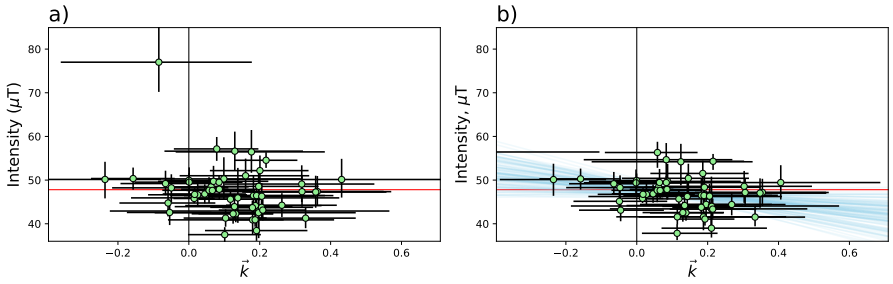

Figure A.31. The same methodology described in the caption for Figure A3 applied to site TS

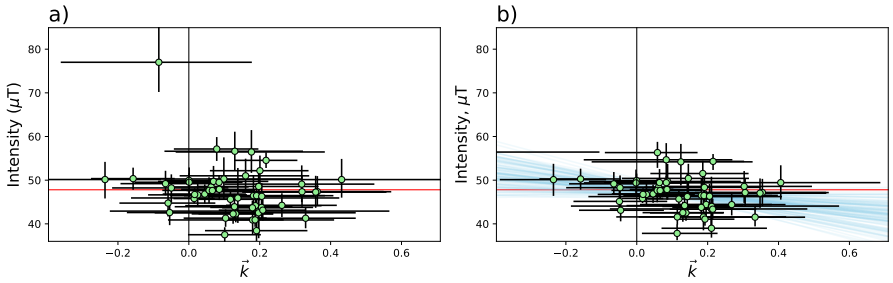

Figure A.32. The same methodology described in the caption for Figure A3 applied to site VM

Appendix B

Supplementary Information for Chapter 4

Table B.1. Results analyzed using BiCEP from Antarctica (Asefaw et al., 2021, sites with the prefix 'mc') and Northern Israel (Tauxe et al., 2021, sites with the prefix 'GHI')

Site	Latitude	Longitude	Age (Ma)	Age 2σ	B_{min}	B_{median}	B_{max}
mc1001	-77.850000	166.640000	1.1800	0.0100	12.7	18.6	24.9
mc1002	-77.850000	166.690000	0.3300	0.0200	22.3	29.1	36.3
mc1009	-77.550000	166.200000	0.0740	0.0150	23.9	28.0	35.1
mc1015	-77.470000	169.230000	1.3300	0.0200	22.7	26.7	30.6
mc1020	-77.880000	165.020000	0.7700	0.0320	55.6	62.3	69.3
mc1029	-78.310000	164.800000	0.1800	0.0800	41.7	44.7	48.0
mc1031	-78.350000	164.300000	0.1330	0.0117	23.5	31.2	39.0
mc1032	-78.360000	164.300000	0.0078	0.0120	27.9	31.2	34.9
mc1036	-78.390000	164.270000	0.1200	0.0400	22.2	28.8	35.3
mc1103	-78.240000	163.360000	1.4210	0.0300	14.5	20.1	27.7
mc1109	-78.280000	163.540000	1.2610	0.0400	29.5	33.2	37.0
mc1111	-78.220000	162.790000	1.9900	0.0400	18.5	21.4	24.6
mc1115	-78.240000	162.960000	2.4600	0.3100	26.7	31.2	35.8
mc1119	-78.240000	162.960000	1.0800	0.2200	34.6	37.7	40.9
mc1120	-78.240000	163.090000	1.7560	0.0500	22.4	24.7	27.0
mc1121	-78.240000	162.950000	2.5050	0.0600	28.8	32.8	35.7
mc1127	-78.250000	163.730000	1.9420	0.0680	33.7	37.0	40.4
mc1135	-78.230000	166.560000	3.6000	0.0100	29.4	31.7	33.9
mc1139	-78.260000	163.080000	0.8820	0.0800	24.6	27.8	31.0
mc1140	-78.280000	163.000000	2.0430	0.0900	28.4	34.2	39.1
mc1145	-78.240000	162.893000	1.9000	0.1200	3.3	7.0	10.2
mc1147	-78.200000	162.960000	1.6300	0.3200	16.5	22.4	27.2
mc1155	-77.700000	162.250000	1.5000	0.0500	23.3	30.8	38.3

Table B.1. Results analyzed using BiCEP from Antarctica (Asefaw et al., 2021, sites with the prefix 'mc') and Northern Israel (Tauxe et al., 2021, sites with the prefix 'GHI') (Continued)

mc1157	-77.700000	162.260000	1.7100	0.0100	31.4	37.9	44.9
mc1160	-77.690000	162.350000	3.4700	0.0500	18.1	24.9	31.4
mc1165	-77.510000	169.330000	1.4510	0.0600	20.6	27.8	35.1
mc1167	-77.490000	169.290000	1.3800	0.1000	38.9	43.5	48.4
mc1200	-77.550000	166.160000	0.0730	0.0100	21.2	26.6	31.8
mc1302	-78.190000	165.320000	0.0400	0.0200	23.7	28.9	34.3
mc1304	-78.240000	163.360000	0.2900	0.0400	18.8	24.7	29.5
mc1305	-78.240000	163.230000	0.9000	0.2000	30.6	34.4	38.2
mc1306	-77.700000	162.690000	2.5600	0.2600	4.5	6.9	9.5
mc1307	-77.850000	166.670000	1.3300	0.2400	39.7	46.1	53.5
GHI01	33.126350	35.782270	0.1177	0.0358	20.1	25.2	30.2
GHI02	33.158050	35.776730	0.1296	0.0012	20.5	24.5	27.9
GHI03B	33.122790	35.724160	0.8420	0.0233	66.7	69.3	72.2
GHI03C	33.122790	35.724160	0.8420	0.0233	36.8	45.2	52.5
GHI03D	33.122790	35.724160	0.8420	0.0233	47.0	59.2	70.1
GHI05	32.960510	35.862240	0.1679	0.0255	19.7	22.6	25.0
GHI06	33.069580	35.771430	0.1145	0.0085	26.4	27.4	28.5
GHI07	33.085810	35.755890	0.6805	0.0183	33.6	40.9	47.7
GHI07C	33.085810	35.755890	0.6805	0.0183	21.0	23.2	25.2
GHI10	33.051680	35.849680	0.6149	0.0349	18.1	19.8	21.5
GHI18	33.025833	35.494912	1.6700	0.0400	30.9	37.3	43.6
GHI19	32.995278	35.525986	2.4500	0.0226	27.1	32.8	39.4
GHI20	32.926290	35.849940	1.6500	0.0200	29.9	31.8	33.5
GHI21	32.926290	35.849940	1.6765	0.0302	21.7	23.6	25.6

Table B.1. Results analyzed using BiCEP from Antarctica (Asefaw et al., 2021, sites with the prefix 'mc') and Northern Israel (Tauxe et al., 2021, sites with the prefix 'GHI') (Continued)

GHI25	33.218726	35.777062	0.8723	0.0053	44.7	52.9	60.8
GHI26	33.220000	35.776833	0.8704	0.0169	46.2	50.0	53.7
GHI27	33.212500	35.786157	1.1498	0.0348	33.5	36.9	40.6
GHI28	33.212500	35.786157	1.1912	0.0152	21.5	28.0	33.7
GHI29	33.179444	35.793218	0.7496	0.0945	28.7	31.1	33.2
GHI39	33.141000	35.682000	0.8476	0.1165	5.9	14.8	21.7
GHI40	33.141000	35.682000	0.7736	0.1949	4.7	7.1	9.8
GHI41	33.141000	35.683000	0.7902	0.0058	4.7	7.1	10.0
GHI44	33.042000	35.836000	1.4369	0.0195	45.6	48.9	52.6
GHI46	32.868290	35.829050	2.7442	0.0475	51.2	61.2	75.4

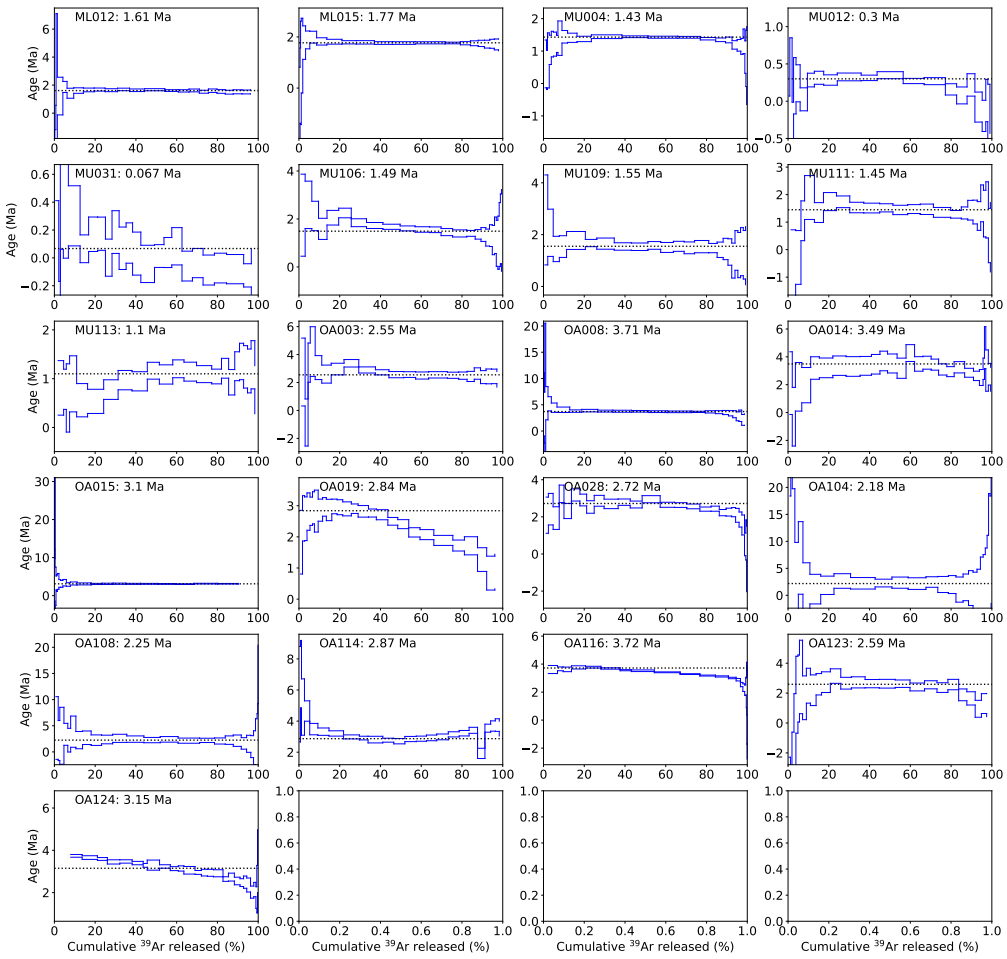

Figure B.1. Plots of Age against cumulative Argon released for age plateau and mini-plateau (OA019, OA116, OA124) age experiments. The mini-plateau ages for OA124 are concordant with total fusion ages from other samples from same site, and the plateau age for OA019 is close to the age for OA028, a nearby lava flow which it is stratgraphically below.