Geophysical Journal International

Geophys. J. Int. (2016) **207,** 1596–1608
Advance Access publication 2016 September 16
GJI Geomagnetism, rock magnetism and palaeomagnetism

doi: 10.1093/gji/ggw349

The pseudo-Thellier palaeointensity method: new calibration and uncertainty estimates

Greig A. Paterson,[1] David Heslop[2] and Yongxin Pan[1]

[1]*Key Laboratory of the Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing* 100029, *China.*
E-mail: greig.paterson@mail.iggcas.ac.cn
[2]*Research School of Earth Sciences, Australian National University, Bldg.* 142, *Mills Rd., Acton, ACT* 2601, *Australia*

Accepted 2016 September 16. Received 2016 September 13; in original form 2016 June 12

SUMMARY

Non-heating palaeointensity methods are a vital tool to explore magnetic field strength variations recorded by thermally sensitive materials of both terrestrial and extraterrestrial origin. One such method is the calibrated pseudo-Thellier method in which a specimen's natural remanent magnetization is alternating field demagnetized and replaced with a laboratory induced anhysteretic remanent magnetization (as an analogue of a thermoremanent magnetization, TRM). Using a set of 56 volcanic specimens given laboratory TRMs in fields of 10–130 μT, we refine the calibration of the pseudo-Thellier method and better define the uncertainty associated with its palaeointensity estimates. Our new calibration, obtained from 32 selected specimens, resolves the issue of non-zero intercept, which is theoretically predicted, but not satisfied by any previous calibration. The range of individual specimen calibration factors, however, is relatively large, but consistent with the variability expected for SD magnetite. We explore a number of rock magnetic parameters in an attempt to identify selection thresholds for reducing the calibration scatter, but fail to find a suitable choice. We infer that our careful selection process, which incorporates more statistics then previous studies, may be largely screening out any strong rock magnetic dependence. Some subtle grain size or mineralogical dependencies, however, remain after selection, but cannot be discerned from the scatter expected for grain size variability of SD magnetite. As a consequence of the variability in the calibration factor, the uncertainty associated with pseudo-Thellier results is much larger than previously indicated. The scatter of the calibration is ∼25 per cent of the mean value, which implies that, when combined with the scatter of results typically obtained from a single site, the uncertainty of averaged pseudo-Thellier results will always be >25 per cent. As such, pseudo-Thellier results should be complementary to, and cross-validated with results from other methods. Nevertheless, the pseudo-Thellier method remains a valuable tool for obtaining palaeointensity estimates from thermally sensitive terrestrial and extraterrestrial materials and with careful data selection and analysis can yield results that are accurate to within a factor of 4 or better.

Key words: Archaeomagnetism; Palaeointensity.

1 INTRODUCTION

The challenges associated with obtaining estimates of absolute palaeointensity have led to a wide range of innovative methods to help improve data quality and quantity. In particular, specimens that are highly susceptible to thermal alteration during laboratory heating (e.g. meteorites and other extraterrestrial materials) have driven an upsurge in non-heating methods (e.g. Gattacceca & Rochette 2004; Muxworthy & Heslop 2011). One such approach is the calibrated pseudo-Thellier method, which uses laboratory induced anhysteretic remanent magnetization (ARM) to avoid heating (e.g. Garrick-Bethell *et al*. 2009; Yu 2010; de Groot *et al*. 2013; Lappe *et al*. 2013).

Although non-heating ARM based methods have been used for some time (Stephenson & Collinson 1974), they have recently been gaining more attention in terrestrial applications (e.g. de Groot *et al*. 2014). The basic premise of the most recent variant of the pseudo-Thellier method (de Groot *et al*. 2013) is that a specimen's natural remanent magnetization (NRM) is progressively alternating field (AF) demagnetized, before a laboratory induced ARM is progressively imparted using the same AF steps. This ARM is subsequently AF demagnetized at the same AF levels. By analogy

with the analysis of thermoremanent magnetization (TRM) based absolute palaeointensity methods, the NRM lost is plotted against ARM gained to form a pseudo-Arai plot (*cf*. Nagata *et al*. 1963). The pseudo-Arai slope (b_{PA}) represents a measure of the relative palaeointensity (when compared with other specimens with similar rock magnetic properties), such that

$$b_{PA} = \frac{\text{NRM lost}}{\text{ARM gained}}. \quad (1)$$

Numerous authors have explored absolute calibrations of the pseudo-Thellier method, particularly in respect to extraterrestrial materials (e.g. Yu *et al*. 2003; Yu 2010; Lappe *et al*. 2013), but many of these studies only explore the demagnetization of ARM and not its acquisition. Most recently, de Groot *et al*. (2013) explored the calibration of the pseudo-Thellier method and its application to lava flows and utilized ARM acquisition characteristics to provide a means of screening out non-ideal factors (discussed further in later sections). Hereafter, in this work we use the term pseudo-Thellier to refer the calibrated pseudo-Thellier method used to obtain absolute palaeointensity estimates.

An absolute palaeointensity (B_{Anc}) can be obtained by the use of an appropriate calibration factor (c) using:

$$B_{Anc} = |b_{PA}| \frac{B_{ARM}}{c} \quad (2)$$

where B_{ARM} is the DC bias field during ARM acquisition. de Groot *et al*. (2013) measured pseudo-Thellier data from historic lava flows from Hawaii and compared the pseudo-Arai slopes to the known field, but using a fixed ARM bias field of 40 μT. The above relation can then be expressed as

$$B_{Anc} = \frac{|b_{PA}|}{c_{40}}, \quad (3)$$

where c_{40} is the calibration in 40 μT DC field ($c = c_{40} \times B_{ARM}$). Over a narrow field range (34.7–37.7 μT), de Groot *et al*. (2013) introduced an intercept term and obtained the following calibration relationship: B_{Anc} (μT) = 7.371 × $|b_{PA}|$ + 14.661 (c_{40} = 0.1357). In this work, we also use an ARM bias field of 40 μT and primarily focus on determining the c_{40} calibration factor.

Subsequent work by de Groot *et al*. (2015) and de Groot *et al*. (2016) compared pseudo-Arai slopes with palaeointensities obtained from thermal based methods (e.g. Tauxe & Staudigel 2004; Dekkers & Böhnel 2006) in an attempt to improve the calibration. Although these improvements do not change the calibration notably, we view the original calibration as the most appropriate. This is because in all subsequent work, the fields that the pseudo-Arai slopes are compared with are unknown, thus this is not strictly calibration of the pseudo-Thellier method, but rather a measure of the consistency with results obtained from different methods.

Despite the attempts to improve the calibration, all previously proposed relationships have a non-zero intercept of ∼14 μT. Theoretically, however, it is expected that such a calibration should have zero intercept and should be a simple scaling relationship (i.e. a specimen with zero NRM should have a pseudo-Arai slope of zero). To explain this, de Groot *et al*. (2013) suggest that the calibration is highly nonlinear in the low field range. Although acknowledged as being far from ideal, to-date, there is no satisfactory explanation of the physical significance of a non-zero intercept.

In this study, we report new pseudo-Thellier data obtained from a suite of 56 well characterized specimens imparted with laboratory induced TRMs. Using these results, we propose a new data selection process and a new calibration of the pseudo-Thellier method that resolves the non-zero intercept problem and provides a better estimate of the uncertainty associated with the calibration and pseudo-Thellier results.

2 METHODS

2.1 Specimens and measurements

A total of 56 specimens were used in this study (1 cm cores). Forty of the specimens were taken from studies of Paterson *et al*. (2010a) and Paterson *et al*. (2010b) and are from Mt. St. Helens, USA and Vesuvius, Italy. In addition, we use 16 unpublished basaltic specimens from the Emeishan large igneous province. Specimen names are grouped by locality. 'A' and 'C' specimens are from Emeishan, 'D' specimens are from Mt. St. Helens, and 'L' specimens are from Vesuvius.

Prior to measurement for this study, all specimens had undergone at least two full thermal palaeointensity experiments and multiple heatings to 700 °C to ensure thermal stability. In Supporting Information Fig. S1, we compare the hysteresis parameters from our thermally stabilized specimens to fresh sister specimens (Paterson *et al*. 2010b; and unpublished measurements of the Emeishan basalts). We find that although there is slight hardening of coercivity (a median increase of ∼15 per cent) that results in lower B_{cr}/B_c, M_{rs}/M_s is not affected by a systematic change. When compared to hysteresis data from unheated volcanic material (Paterson *et al*. 2010b; Muxworthy *et al*. 2011), our heated specimens are indistinguishable from the broad range behaviour exhibited by natural specimens, which indicates that our specimens are suitable for comparison with unheated natural materials.

To produce NRMs, the specimens were given full TRMs (imparted at 700 °C) in fields of 10, 50, 90 and 130 μT using a Magnetic Measurements MMTD-SC furnace. For specimens L5 and L9, no 50 μT data are available. All remanence measurements were performed using a 2G Enterprises 760 cryogenic magnetometer. AF demagnetization (along three axes: *X*-, *Y*-, then *Z*-) was carried out in 14 steps (2.5, 7.5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 80, 100 and 120 mT) using the inline static AF demagnetizer of the magnetometer (model 2G600). This is long-core AF system where the AF amplitude spatially decays. The AF system operates at a frequency of ∼200 Hz (*X*- and *Y*-axes) or ∼150 Hz (*Z*-axis) and a translation speed of 25 cm s^{-1} was used. ARMs were progressively imparted using the magnetometer's inline coil (model 625 ARM) with a bias field of 40 μT using the peak AFs given above. The ARMs were subsequently AF demagnetized at the same AF steps. For each NRM field, the measurements of ARM acquisition and demagnetization were repeated. Both NRM and ARM were acquired along the specimens' *Z*-axis and during ARM acquisition the AF was also along the *Z*-axis. We note that in our system no notable currents are induced in the ARM bias field due to the AF signal.

Magnetic hysteresis and isothermal remanent magnetization (IRM) acquisition were measured on a Princeton Measurements Corporation vibrating sample magnetometer (model 3900) up to a peak field of 1.4 T. Hysteresis data were processed following the recommendations of Jackson & Solheid (2010) and where necessary the approach to saturation method of Fabian (2006) was applied. Saturation IRMs (SIRMs) were determined from the hysteresis data after correction for dia-/paramagnetic contributions and approach to saturation. All rock magnetic measurements were performed on exactly the same specimens as the pseudo-Thellier experiments.

Table 1. Summary of employed selection statistics.

Statistic	Threshold	Applied Plot	Reason
n	≥ 6	All	Sufficient data points should be used for fitting
f	≥ 0.45	Pseudo-Arai	Sufficient NRM should be used
β	≤ 0.1	Pseudo-Arai	Pseudo-Arai fit should not be scattered
R^2_{corr}	≥ 0.995	All	All plots should be linear
$\lvert\vec{k}'\rvert$	≤ 0.2	All	All plots should be linear
f_{Resid}	<0.15	Demag–Demag	To ensure NRM–ARM demag plot trends towards to origin
b_{AA}	$0.85 \leq \lvert b_{\text{AA}}\rvert \leq 1.15$	ARM–ARM	ARM–ARM plot should have an absolute slope of 1

2.2 Data analysis and selection

As will be discussed below, our specimens have variable amounts of NRM remaining after demagnetization to 120 mT, which is due to the presence of a magnetically hard mineral (hematite). For our analyses, we therefore remove the residual magnetizations from all remanence data via vector subtraction. We note that this makes negligible difference to the final pseudo-Thellier calibration, but is necessary to obtain consistent characterization of the ARM acquisition and demagnetization data, which is relevant when considering rock magnetic selection (discussed in Section 4.3). For example, specimen C2, which is a basaltic lava, has residual NRMs of ∼4–7 per cent after AF demagnetization at 120 mT. Before subtracting the residual NRM from the ARM demagnetization, the ARM median destruction fields (MDFs) for C2 are 28.0, 32.5, 37.6 and 45.4 mT, for the 10, 50, 90 and 130 μT fields, respectively. After subtracting the NRM residuals the ARM MDFs are 26.5, 27.2, 27.1 and 27.3 mT, respectively.

To ensure reliable estimates of pseudo-Arai plot slope, de Groot *et al*. (2013) recommend identifying a linear segment on the plot of ARM left versus NRM left (what we refer to as the demag–demag plot). This segment could then be used for analysing the pseudo-Arai plot. Here we introduce additional selection requirements, which are all based on desirable or theoretically expected behaviour. These are summarized in Table 1 and we follow the Standard Paleointensity Definition (SPD) terminology (Paterson *et al*. 2014). First, to ensure that sufficient data points and NRM fraction are used to make an estimate of the pseudo-Arai slope we require $n \geq 6$ and $f \geq 0.45$. By analogy with traditional double heating Arai plot analysis, we also require that $\beta \leq 0.1$.

Following de Groot *et al*. (2013), we note that in the ideal case, over the selected AF interval, the demag–demag plot, the pseudo-Arai plot, and the ARM left versus ARM gained plot (what we refer to as the ARM–ARM plot), should all be linear. The expectation of linearity comes from Thellier's law of reciprocity (Thellier & Thellier 1959). This fundamental assumption of all palaeointensity studies, states that the blocking spectrum activated by the NRM and by the laboratory remanence should be identical. This assumption is irrespective of the mineralogy carrying the remanence or whether it is blocking temperature or coercivity spectra that are activated. Violation of this requirement would result in nonlinear plots when different remanences are compared with each other and therefore linearity is a simple test that this basic assumption holds true. Therefore, when making fits to the pseudo-Arai plots we consider the linearity of all three plot types. We quantify linearity using the curvature of the fitted segment ($\lvert\vec{k}'\rvert$) (Paterson 2011; Paterson *et al*. 2015) as well as the square of the Pearson linear correlation coefficient (R^2_{corr}). A segment is chosen if $R^2_{\text{corr}} \geq 0.995$ and $\lvert\vec{k}'\rvert \leq 0.2$. These two linearity requirements are applied to all three analysis plots (i.e. the demag–demag, pseudo-Arai, and ARM–ARM plots).

de Groot *et al*. (2013) also note that the best-fit segment on the demag–demag plot should trend towards the origin if the two remanences are being properly demagnetized. To quantitatively assess this, here we introduce f_{Resid} as measure of the residual fraction. This is an analogy of the NRM fraction (f) of Coe *et al*. (1978) and is given by

$$f_{\text{Resid}} = \frac{|Y_{\text{Int.}}|}{\Delta y'}, \tag{4}$$

where $Y_{\text{Int.}}$ is the y-intercept on the demag–demag plot and $\Delta y'$ is the change in the ARM lost over the selected segment (see SPD; Paterson *et al*. 2014). This statistic biases against fits with large intercepts, or fits where only a small amount of ARM is lost. We only select fits with $f_{\text{Resid}} \leq 0.15$.

Finally, we note that, not only should the ARM–ARM plot be linear, but, in the absence of non-ideal factors, this plot should have a slope of -1. That is, for a selected AF range, ARM lost should be identical to the ARM gained. Violation of this is an indication of non-reciprocal ARM, which is a prerequisite of ARM based palaeointensity methods (Yu *et al*. 2002). We quantify this by requiring that the ARM–ARM slope (b_{AA}) should be in the range $0.85 \leq |b_{\text{AA}}| \leq 1.15$ (Table 1). The slope is calculated using stand major axis regression (see SPD; Paterson *et al*. 2014).

To account for rock magnetic variability and to improve the consistency of the calibration, de Groot *et al*. (2013) proposed $B_{1/2\text{ARM}}$, which is half the field required to impart the saturation ARM, as means of rock magnetic screening. They recommended that $B_{1/2\text{ARM}}$ should lie between 23–63 mT. We begin by first exploring the calibration without rock magnetic consideration (Section 3.2) and discuss rock magnetic factors in more detail in Section 4.3.

3 RESULTS

3.1 Remanence and demagnetization behaviour

For all 56 specimens, TRM acquisition is linear with applied field ($R \geq 0.992, p \leq 0.035$). Therefore, over the field range explored here (10–130 μT), nonlinear TRM acquisition need not be considered. Similarly, since all remanences are acquired along the same axis, anisotropy does not need to be considered.

de Groot *et al*. (2013) were able to demagnetize their specimens up to 300 mT, but we were only able to demagnetize our specimens to 120 mT. As a result, for our specimens variable amounts of NRM remain after demagnetization (5–40 per cent). Eleven specimens (∼20 per cent of all specimens) have mean residual NRMs >15 per cent, indicating magnetically hard minerals. This is most likely hematite, which is known to be present in some specimens before thermal stabilization (Paterson *et al*. 2010a,b), but may also form during the stabilization process. We note that the median

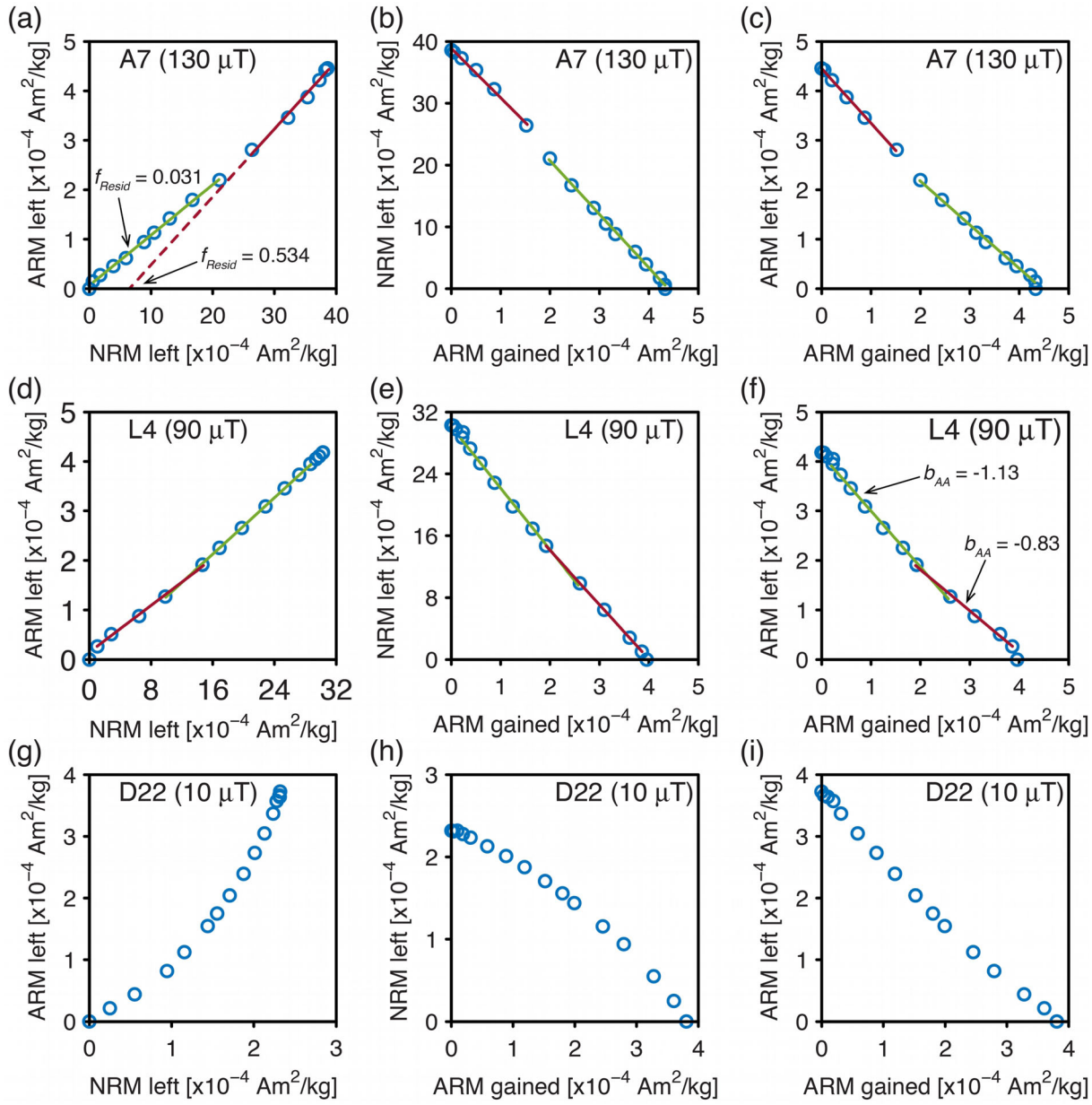

Figure 1. Example demag–demag (a, d, g), pseudo-Arai (b, e, h) and ARM–ARM pots (c, f, i) that illustrate a range of remanence behaviour. Green lines correspond to example fits that pass the applied selection criteria (Table 1) and red lines are example fits that fail.

residual NRM is ∼12 per cent, which means that half of our specimens have residual NRMs that would be broadly consistent with magnetically soft minerals such as magnetite and maghemite. After subtraction of residual NRM, reproducibility of the total ARM is within <10 per cent of the mean (generally <6 per cent), which confirms the thermal stability of the specimens.

Example demag–demag, pseudo-Arai, and ARM–ARM plots are shown in Fig. 1. These illustrate the range of behaviour that we observe and we highlight examples of fits that pass and fail selection. For the 130 μT experiment for specimen A7 (Figs 1a–c), we find that although the low coercivity range yields linear fits, they fail to trend to the origin on the demag–demag plot (Fig. 1a). The higher coercivity ranges, however, pass selection. Specimen L4 (90 μT; Figs 1d–f) illustrates a case where the high coercivity range is linear and trends close to the origin of the demag–demag plot (f_{Resid} ∼ 0.06), but where the slope of the ARM–ARM plot is −0.83 and not near the expected value of −1 (Fig. 1f). In this case, the lower coercivity ranges tend to pass our selection. Lastly, specimen D22 (10 μT; Figs 1g–h) is an example of where all plots are highly nonlinear and no fit can be found that passes our selection requirements. This indicates that the ARM acquisition and demagnetization and NRM demagnetization are not activating the same coercivity spectra, violating the assumption of reciprocity.

3.2 Pseudo-Thellier calibration

For most specimens, we identify multiple AF ranges that pass our selection criteria (Table 1). We therefore analyse all possible fits. Of the 222 experiments we performed, 136 yield at least one fit that passes selection. This represents 46 of 56 specimens. To determine the calibration of each specimen, we accept only specimens that yield acceptable results from at least three different TRM fields.

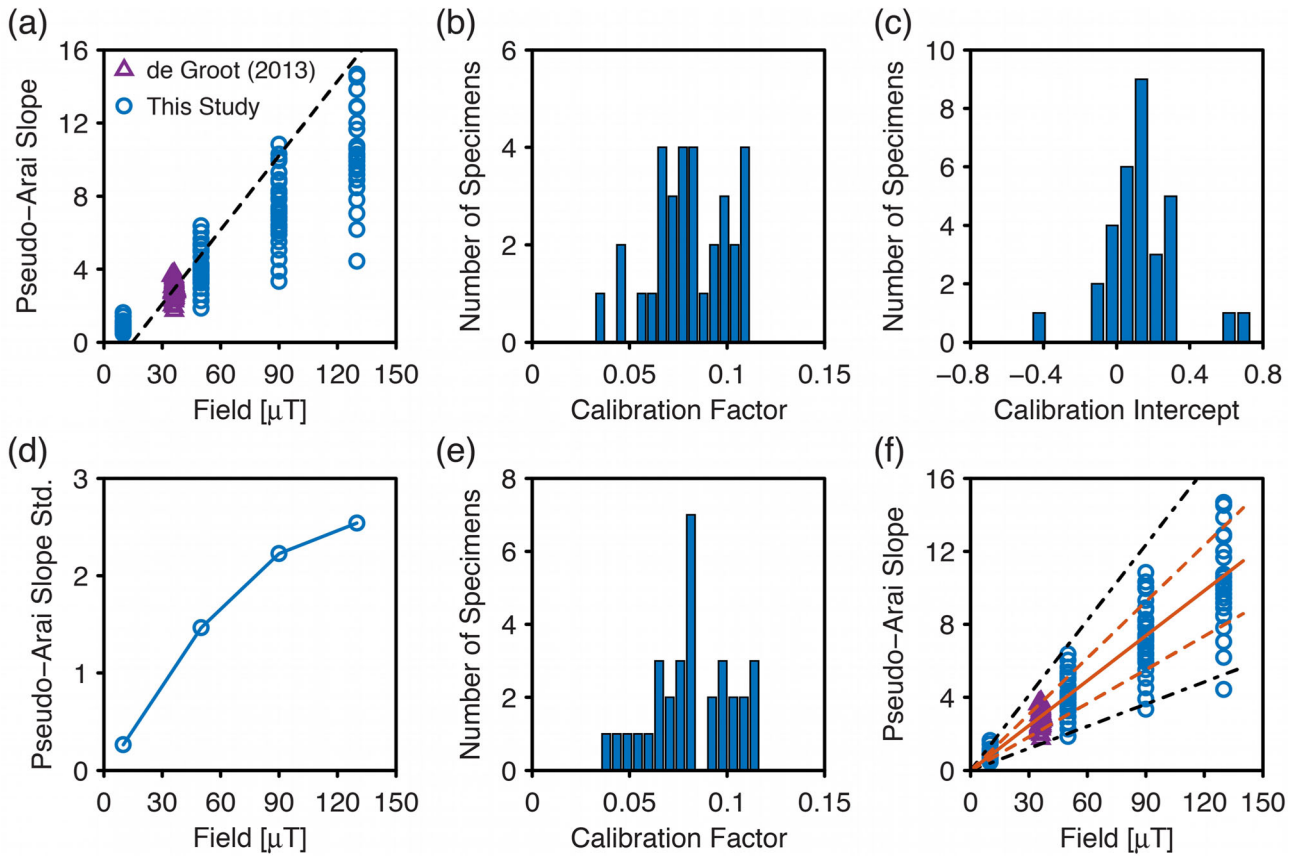

Figure 2. (a) Accepted mean pseudo-Arai slopes as a function of the NRM field. Blue circles are our new data, the purple triangles are from de Groot *et al*. (2013) and the dashed line is the calibration of de Groot *et al*. (2013). (b, c) Histograms of the calibration factor and intercept, respectively, from a non-anchored robust linear fit to data from each individual specimen. (d) The standard deviation of the pseudo-Arai plot slopes as a function of the NRM field. (e) Histogram of the calibration factors from an anchored robust linear fit to data from each individual specimen. (f) The accepted pseudo-Arai slopes as a function of the NRM field, alongside our new calibration. The solid orange line represents our new mean calibration and the orange dashed lines denote the one standard deviation envelope. The black dash-dot lines are the expected calibrations from 30 nm (lower) and 57 nm (upper) magnetite (Dunlop & West 1969). Other symbols are as in part (a).

This leaves a total of 32 out of 56 specimens that yield a calibration estimate, which represents 117 out of 222 experiments.

Given that multiple fits for each experiment are possible, we take the average of all pseudo-Arai slope estimates for each experiment. For example, for the 130 μT experiment for specimen L2 (Figs 1a–c), 38 possible fits pass our selection criteria with an average of 9.17 ± 0.23 (all uncertainties are given as ± one standard deviation, unless stated otherwise). This average value is taken to be our pseudo-Arai slope estimate for the 130 μT experiment for specimen L2. The effect in choosing different possible best-fit segments is found to be negligible and is discussed in Section 4.4. The accepted mean slopes and their standard deviations are given in Supporting Information Table S1.

All 117 accepted results are shown in Fig. 2(a) along with the accepted Hawaiian data and pseudo-Thellier calibration from de Groot *et al*. (2013). Although the Hawaiian data are consistent with our new data, the original calibration of de Groot *et al*. (2013), which differs only slightly from later modifications, systematically differs from our observations.

For each specimen, we perform robust linear fitting with an intercept term and histograms of the slope and intercepts are shown in Figs 2(b) and (c), respectively. The mean slope (the calibration factor c_{40}) and intercept are 0.0806 and 0.1349, respectively. The mean intercept corresponds to a zero pseudo-Arai plot slope field estimate of −1.7 μT. This mean field intercept is almost an order of magnitude less than previous calibrations, which yielded values of ∼14 μT (de Groot *et al*. 2013, 2015, 2016). The histogram of the intercepts has a clustering of values close to zero (Fig. 2c), but with a tendency to be slightly larger than zero. We also robustly fit each specimen with a linear model without an intercept term (i.e. anchored to the origin). An F-test comparison of variance between the two models reveals that for 31 out of the 32 specimens the simpler model without an intercept term yields a statistically better fit to the data at the 5 per cent significance level. The theoretical expectation of zero-intercept and the fact that it is statistically supported by the data strongly suggest that the calibration of the pseudo-Thellier method should be a simple scaling rule without an intercept term (as given in eqs 3 and 4). We therefore prefer to adopt the calibration factors obtained from a robust linear fit without an intercept term.

The histogram of calibration factors is shown in Fig. 2(e). The values range from 0.0353 to 0.1167, with a mean of 0.0820 ± 0.0207. The mean calibration value and the standard deviation interval are plotted in Fig. 2(f). This new calibration is consistent with the Hawaiian data used by de Groot *et al*. (2013) to develop their initial calibration (discussed further in Section 4.4). We note that using the same selected specimens and pseudo-Arai plot segments, but not removing the residual NRMs, yields a mean value of 0.0819 ± 0.0205,

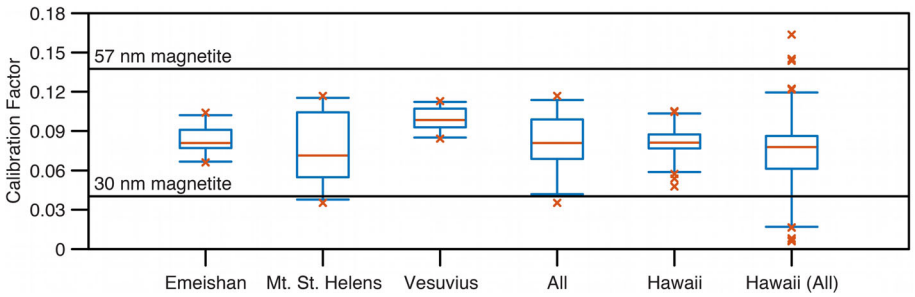

Figure 3. Box-whisker plots of our accepted specimen calibration factors for the Emeishan basalts (specimen groups A and C), Mt. St. Helens (group D), Vesuvius (group L), and all groups combined. The Hawaiian data are from de Groot *et al*. (2013) and are their selected ('Hawaii') and unselected ('Hawaii (All)') data. The 30 nm and 57 nm solid lines are the expected calibrations for SD magnetite (Dunlop & West 1969). In this plot, the boxes denote the interquartile ranges, the whiskers the 95 per cent ranges and the orange lines are the median values. The orange crosses represent values that lie outside the 95 per cent ranges.

Table 2. Descriptive statistics of the calibration factors from the different localities. *N* is the total number of specimens and *n* is the number of specimens that passed selection. The Hawaii data are from de Groot *et al*. (2013).

			Calibration factor				
Location	Group	*n*/*N*	Min.	Max.	Median	Mean	Std. Dev.
Emeishan LIP, China	A, C	16/16	0.0662	0.1040	0.0807	0.0827	0.0108
Mt. St. Helens, USA	D	13/32	0.0353	0.1167	0.0692	0.0771	0.0287
Vesuvius, Italy	L	3/8	0.0841	0.1127	0.1013	0.0994	0.0144
All	A, C, D, L	32/56	0.0353	0.1167	0.0807	0.0820	0.0207
Hawaii	Selected	93/197	0.0479	0.1052	0.0812	0.0811	0.0107
Hawaii	All	197/197	0.0059	0.1637	0.0776	0.0733	0.0242

which is consistent with the value after residual NRMs are subtracted.

4 DISCUSSION

4.1 Locality variability

Taken en masse, our new calibration data and that of de Groot *et al*. (2013) are consistent, but they come from four distinct localities. Fig. 3 is a box-whisker plot of the specimen calibration factors broken down by locality and some key values are summarized in Table 2. All three of our localities yield consistent results, but the scatter of the calibration factors from Mt. St. Helens is higher than the other locales. This is likely related to a wider range of lithologies and hence magnetic properties from this locality. Mt. St. Helens has the lowest mean/median values, followed by Emeishan, then Vesuvius (Table 2). The box-whisker plot for all of the selected specimens has a consistent overlap with all three localities, which suggests that no locality is distinctly different from any other.

Fig. 3 and Table 2 also summarize the Hawaiian data from de Groot *et al*. (2013). Here, we simply estimate the individual specimen calibration factors as the pseudo-Arai slope divided by the known ambient field during NRM acquisition. The selected data of de Groot *et al*. (2013) are consistent with our combined data as well as the Emeishan and Mt. St. Helens data, but have slightly lower calibration factors than those from our Vesuvius specimens. The unselected data, however, exhibit a much wider range of variability that is consistent with all three of our individual localities and consistent with our pooled data. Slight differences between our data and that of de Groot *et al*. (2013) are therefore likely due to different data selection processes.

We also note that the data from de Groot *et al*. (2013) were measured at the palaeomagnetic laboratory in Utrecht using the AF and ARM systems attached to their robotized 2G magnetometer (Mullender *et al*. 2016). Their AF system operates at a frequency of ∼200 Hz and AF demagnetization was performed using a translation speed of 2 cm s^{-1}, while ARM was acquired holding the specimen static as the fields were ramped up and down (de Groot, private communication, 2016). Despite notable differences in the experimental setup, our results are consistent (e.g. Figs 2, 3 and 7), which suggests that AF and ARM decay dependences may not be significant for these specimens. Alternatively, the general scatter of results may hide or be the result of decay rate dependences and future work should investigate this further.

4.2 Inherent variability of k_{TRM}/k_{ARM}: grain size effects

After selection our calibration factors remain scattered; the standard deviation is ∼25 per cent of the mean. It has long been recognized that the susceptibility of both TRM and ARM (k_{TARM} and k_{ARM}, respectively) depend on grain size, but k_{ARM} has a much stronger dependence (e.g. Dunlop & West 1969; Levi & Merrill 1976; Yu *et al*. 2003). As a result, the ratio of k_{TARM} to k_{ARM}, which is essentially the reciprocal of the calibration factor, is strongly grain size dependent. Available data from SD magnetite (≤65 nm) indicate that the ratio of TRM to ARM ranges from ∼1.6 to ∼5.5 (6 data from Dunlop & West 1969; Dunlop *et al*. 1975; Levi & Merrill 1976; Yu *et al*. 2003). Assuming a DC bias field of 40 μT during ARM acquisition, pseudo-Thellier c_{40} calibration factors can be estimated for these specimens. The estimated calibration factors range from 0.0404 to 0.1374 with a mean of 0.0956 ± 0.0325, which is slightly larger than our estimate (0.0820). The maximum and minimum calibrations factors, which come from the 30 and 57 nm specimens of Dunlop & West (1969), respectively, are plotted in Figs 2(f) and 3. The range defined by these specimens encompasses our data well, suggesting that our observed variability is consistent with, or at least

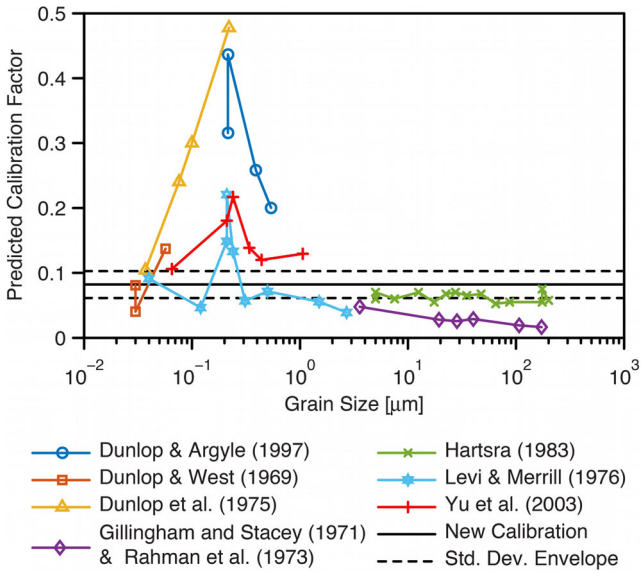

Figure 4. Predicted pseudo-Thellier calibration factors from sized magnetite specimens alongside our new calibration (0.0820 ± 0.0207) from volcanic materials. The predicted calibration factors are based on the reported values of the susceptibility of TRM and ARM and are determined for an ARM bias field of 40 μT.

indistinguishable from that expected for SD magnetite. Similarly, these bounds also encompass all of the Hawaiian data selected by de Groot *et al*. (2013) and much of their unselected results (Fig. 3).

In Fig. 4, we plot estimated calibration factors from available sized magnetite data (Dunlop & West 1969; Gillingham & Stacey 1971; Rahman *et al*. 1973; Dunlop *et al*. 1975; Levi & Merrill 1976; Hartstra 1983; Dunlop & Argyle 1997; Yu *et al*. 2003). This illustrates a sharp spike in the calibration factors in the pseudo-single domain (PSD) grain size range. It also illustrates, however, that the calibration factors that we observed are also consistent with some large PSD to small multidomain (MD) magnetite grain sizes (on the order of 0.3 to ∼3 μm), which would yield calibration factors of ∼0.0391 to ∼0.1385 (Levi & Merrill 1976; Yu *et al*. 2003). Larger grain sizes are likely to yield low values of ≤0.075 (Gillingham and Stacey; Rahman *et al*. 1973; Hartstra 1983) and may contribute to some of our lower estimates. Our selection process, which is discussed further below, is likely biasing against larger grain sizes, where the violation of reciprocity will be detected by our plot linearity criteria (e.g. Yu *et al*. 2003).

4.3 Rock magnetic correlations and screening

Given the abovementioned sensitivity of ARM to grain size it is desirable to attempt to restrict the range of specimens suitable for use with the pseudo-Thellier method. For this reason, de Groot *et al*. (2013) proposed the use of $B_{1/2\mathrm{ARM}}$ as a means of rock magnetic screening ($B_{1/2\mathrm{ARM}}$ is the field required to impart half of the saturation ARM). It should be noted that since our ARMs are not fully saturated some of our ARM quantifications might be underestimates. We therefore specifically note that these are the values for ARM acquired up to 120 mT and are denoted as ARM_{120} MDF and $B_{1/2\mathrm{ARM120}}$. We also use a high-field equivalent to $B_{1/2\mathrm{ARM}}$, $B_{1/2\mathrm{SIRM}}$, which is the field required to impart half the SIRM. We calculate $B_{1/2\mathrm{SIRM}}$ by normalizing the IRM acquisition curves by the SIRM obtained from the hysteresis measurements after correction for approach to saturation.

We first consider various rock magnetic parameters and the calibration factors without applying our new selection process. For each experiment, we take the mean of all possible pseudo-Arai plot fits with $n \geq 4$ and perform a robust linear regression on each specimen to determine the unselected calibration. We find that the low-field parameters (ARM_{120} MDF, $B_{1/2\mathrm{ARM120}}$, and residual NRM) have no significant correlation (at the 5 per cent significance level) with the unselected calibration factors (Figs 5a–c). However, high-field parameters (i.e. squareness, coercivity, and $B_{1/2\mathrm{SIRM}}$) are strongly and significantly correlated with the unselected specimen calibrations (Figs 5d–f). This suggests that the high-field measurements, which should be dominated by highly magnetic minerals (e.g. magnetite and maghemite), are picking up domain state (grain size and/or magnetic interaction) variations in these minerals. The high-field correlations are all positive (i.e. large values correspond to larger calibration factors). The more MD-like specimens (i.e. low squareness and coercivity) tend to yield lower calibration factors (Figs 5d–f), which is consistent with the values expected from large PSD to MD magnetite (Fig. 4).

In Fig. 6, we compare our accepted specimen calibration factors with the various rock magnetic parameters (the data for the selected specimens are reported in Supporting Information Table S2). We find that the low-field parameters (ARM_{120} MDF, $B_{1/2\mathrm{ARM120}}$, and residual NRM) have significant (5 per cent level) correlations with the calibrations factors, with higher values of the low-field parameters having lower calibration factors (Figs 6a–c). The residual NRM is also correlated with both ARM_{120} MDF ($R = 0.498$, $p = 0.004$) and $B_{1/2\mathrm{ARM120}}$ ($R = 0.445$, $p = 0.011$), which suggests that these parameters may be sensitive to the effectiveness of NRM demagnetization. For the high-field parameters (Figs 6d–f), however, we find no significant correlation with the calibration factors.

After applying our selection criteria described in Section 2.2, we observe a reduction in the correlation between the calibration factors and high-field rock magnetic properties, which we attribute to our selection process. The pseudo-Thellier data from Yu *et al*. (2003) exhibit an increase in plot curvature with increasing grain size for both the pseudo-Arai and ARM–ARM plots (the increase for the ARM–ARM plots is most distinct; Supporting Information Table S3). This increase in nonlinearity with increasing grain size is analogous to curvature on a traditional Arai plot (e.g. Levi 1977) and is a consequence of different coercivity spectra being activated by the NRM and laboratory ARM acquisition processes (i.e. a violation of reciprocity). Therefore, by applying a curvature specific selection criterion (Paterson 2011) we should be more effective in discriminating against larger grain sizes, which supports the suggestion that our calibration factors are consistent with the expected range of values from SD magnetite.

The increase in the correlation between the calibration factors and low-field remanence characteristic (e.g. Figs 5a and 6a) may be related to subtle variations in magnetic mineralogy of these particular specimens. Hematite is known to be present in some specimens and has likely formed during the thermal stabilization process. Our screening against nonlinear remanence behaviour most likely removed the dominant effect of domain state variations in highly magnetic minerals and may have revealed a more subtle influence of mineralogy. However, given that our range of calibration factors is consistent with grain size variability of magnetite, we cannot fully exclude the possibility that this is still related to a domain state effect (e.g. grain size).

de Groot *et al*. (2013) found $B_{1/2\mathrm{ARM}}$ to be a useful selector and Figs 6(a) and (b) suggest that both ARM_{120} MDF and $B_{1/2\mathrm{ARM120}}$ may help exclude some of the lower calibration factors we obtain.

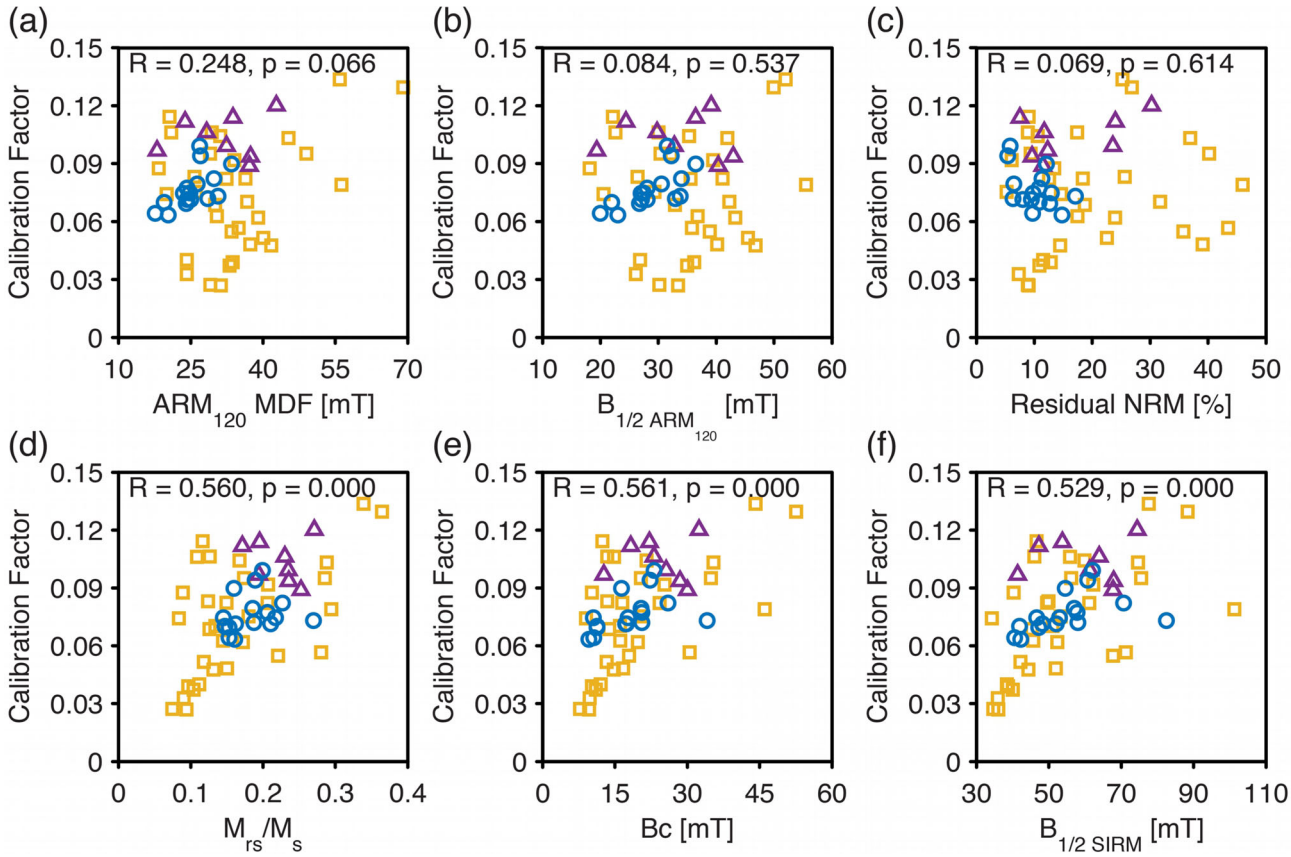

Figure 5. All specimen calibration factors as a function of (a) the ARM_{120} MDF, (b) $B_{1/2ARM120}$, (c) $B_{1/2SIRM}$, (d) M_{rs}/M_s (squareness), (e) coercivity (B_c), and (f) the mean residual NRM after demagnetization to 120 mT (as a percentage of the original NRM). In all plots, the blue circles are the Emeishan specimens (groups A and C), the yellow squares are the Mt. St. Helens specimens (group D) and the purple triangles are the Vesuvius specimens (group L). *R* and *p* are the Pearson linear correlations and their significance, respectively.

Specifically, requiring ARM_{120} MDF $\leq$ 37.2 mT or $B_{1/2ARM120}$ $\leq$ 36.5 mT yields calibrations of 0.0857 $\pm$ 0.0178 and 0.0891 $\pm$ 0.0158, respectively. This further rejects three or seven specimens, respectively. The validity of these calibrations is discussed below.

4.4 Validity of the new calibration

4.4.1 Sensitivity to the choice of pseudo-Arai segment

For each individual experiment (e.g. specimen D21 in 10 μT), multiple AF segments yield fits that pass our selection criteria (Table 1) and we simply estimate the calibration factor using the average pseudo-Arai slope from all acceptable fits for a particular experiment. To test the validity and robustness of this approach, we adopt a Monte Carlo resampling approach, whereby, for each specimen, we randomly select one acceptable fit for each field experiment and obtain a calibration factor. This is repeated for all 32 specimens and we can obtain an average calibration and its standard deviation. We repeat this process 10^4 times to obtain distributions for both the mean calibration factor and standard deviation, which are shown in Figs 7(a) and (b), respectively.

The mean calibration factors range from 0.0810 to 0.0832 (min. to max.; 0.0818–0.0822 inter-quartile range) with a mean of 0.0820 $\pm$ 0.0003. The calibration uncertainties range from 0.0197 to 0.0217 (min. to max.; 0.0206–0.0209 inter-quartile range) with a mean of 0.0208 $\pm$ 0.0003. These values compare well with our final calibration of 0.0820 $\pm$ 0.0207 and confirm that, after data selection, our calibration it is robust against the choice of pseudo-Arai plot segment.

4.4.2 An independent test

de Groot *et al.* (2013) reported 197 pseudo-Thellier results from Hawaiian lava flows where the true field intensity was known. Of these, 93 passed their selection process (23 mT $\leq B_{1/2ARM} \leq$ 63 mT) and were emplaced in fields of 34.7–37.7 μT. Using both the selected and unselected pseudo-Arai plot slopes we determine new palaeointensity estimates and test two of our proposed calibrations (before and after selection using ARM_{120} MDF). For fair comparison, we normalize the estimates by the known fields and take the logarithm (referred to as the deviation, *cf.* Paterson *et al.* 2014). Kernel density plots of the deviations are shown in Figs 7(c) and (d): zero deviation corresponds to the correct result, negative and positive values correspond to under- and overestimates, respectively.

For the calibration before rock magnetic selection (0.0820 $\pm$ 0.0207), the deviations of the selected results peak close to zero and the deviation of mean values is -0.012, which is an underestimate of $\sim$1.2 per cent (Fig. 7c). Similarly, the deviations of the unselected results peak close to zero and the deviation of mean values is -0.113 (Fig. 7d). For the ARM_{120} MDF

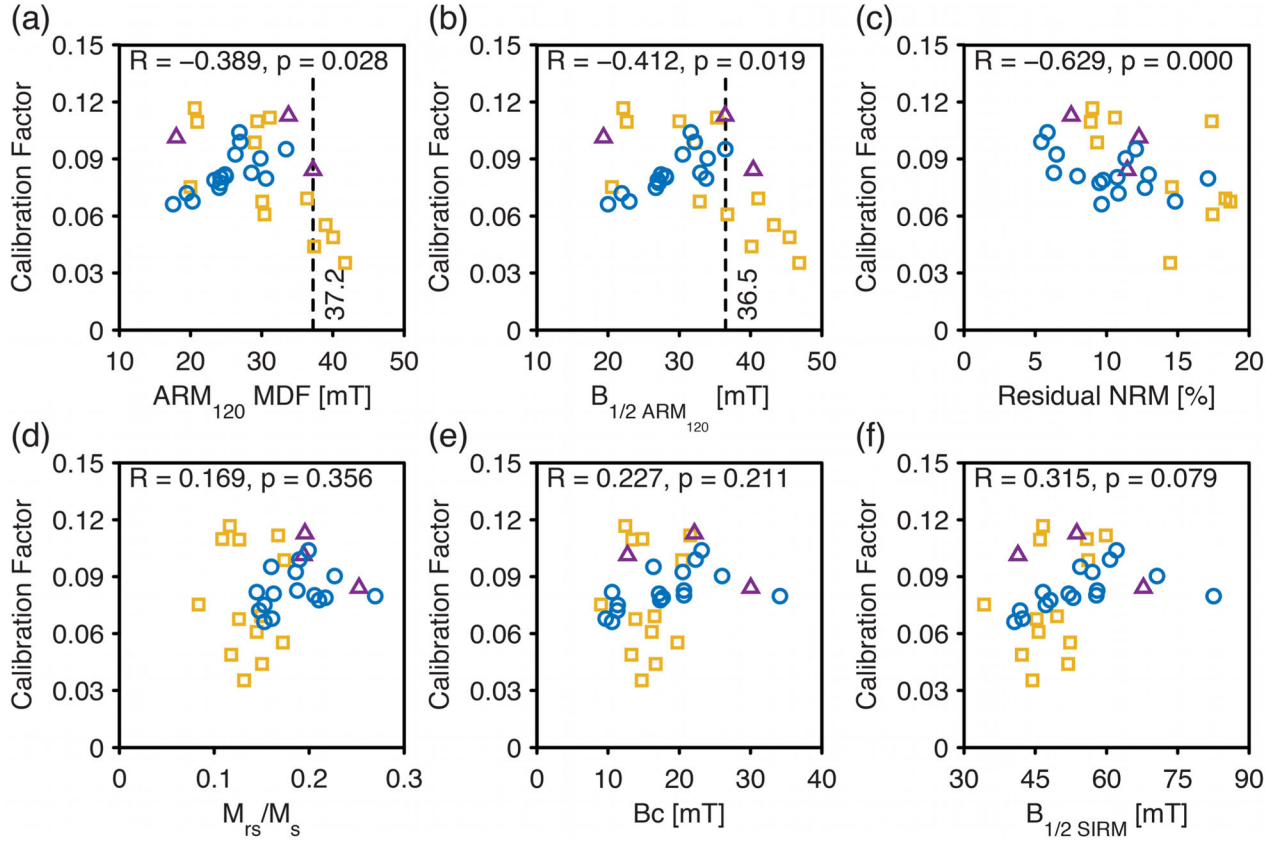

Figure 6. Accepted specimen calibration factors as a function of (a) the ARM_{120} MDF, (b) $B_{1/2ARM120}$, (c) $B_{1/2SIRM}$, (d) M_{rs}/M_s (squareness), (e) coercivity (B_c), and (f) the mean residual NRM after demagnetization to 120 mT (as a percentage of the original NRM). The dashed lines in parts (a) and (b) represent possible thresholds for rock magnetic selection. All other symbols are the same as in Fig. 5.

$\leq$ 37.2 mT screened calibration (0.0857 $\pm$ 0.0178), however, the palaeointensity estimates are systematically lower than the expected field (Figs 7c and d). For the selected data, this corresponds to a deviation of mean of -0.056 and for the unselected data -0.156: underestimates of $\sim$5 per cent and $\sim$14 per cent, respectively.

It should be noted that, if the $B_{1/2ARM120} \leq 36.5$ mT threshold is used, the mean calibration factor increases, which further increases the underestimation of the rock magnetic selected calibration. Although the systematic underestimation from our rock magnetically selected calibrations is small, for this particular data set, we find no compelling reason to support the use of a rock magnetic selection statistic. This is likely due to the preferential rejection of non-ideal data through the stricter selection process applied in deriving our calibration.

We cannot, however, exclude the possibility that further rock magnetic investigation or restriction of material types may lead a better constrained calibration. However, given the wide range of variability of TRM/ARM across the SD size range, the range that is traditional viewed as being ideal for palaeointensity determinations, it is difficult to assess what is an appropriate target value for the calibration of the pseudo-Thellier method, other than the range of values expected for SD grains, which our data are comparable to.

Due to the strong dependence on grain size, ARM based palaeointensity methods are often viewed as order-of-magnitude methods (*cf.* Fig. 4). Using our preferred calibration, 95 per cent of de Groot *et al.*'s unselected results are all within a factor of 4 of the known field. For the selected results, 95 per cent are within a factor of 1.3 (deviations of $\pm$0.26) and all are within a factor of 1.7 (deviations of $\pm$0.53). This suggests that the pseudo-Thellier method is accurate to within a factor 4, but may be much more accurate with careful data analysis and selection.

4.4.3 Comparison to thermal based palaeointensity methods

Both de Groot *et al.* (2015) and de Groot *et al.* (2016) attempted to refine the pseudo-Thellier calibration by cross-calibrating to thermal based methods. As we have noted previously, although useful as a cross-validation tool, this approach is not a true calibration of the pseudo-Thellier method. In Fig. 7(e), we plot our new calibration alongside the original and the data of de Groot *et al.* (2015) and de Groot *et al.* (2016). Although the data close to the original calibration, all are approximately within our one standard deviation envelope. When accounting for the uncertainty of both the calibration and the thermal palaeointensity estimates themselves, our newly calibrated pseudo-Thellier method yields results that are consistent with traditional thermal-based palaeointensity methods.

4.5 The uncertainty of pseudo-Thellier estimates

Our new data set provides a new measure of the uncertainty associated with the calibration of the pseudo-Thellier method (0.0820 $\pm$ 0.0207). This uncertainty, however, must be correctly propagated into the final a palaeointensity estimate. An average pseudo-Thellier palaeointensity estimate (m_B) is given by $m_B = \frac{m_s}{c}$,

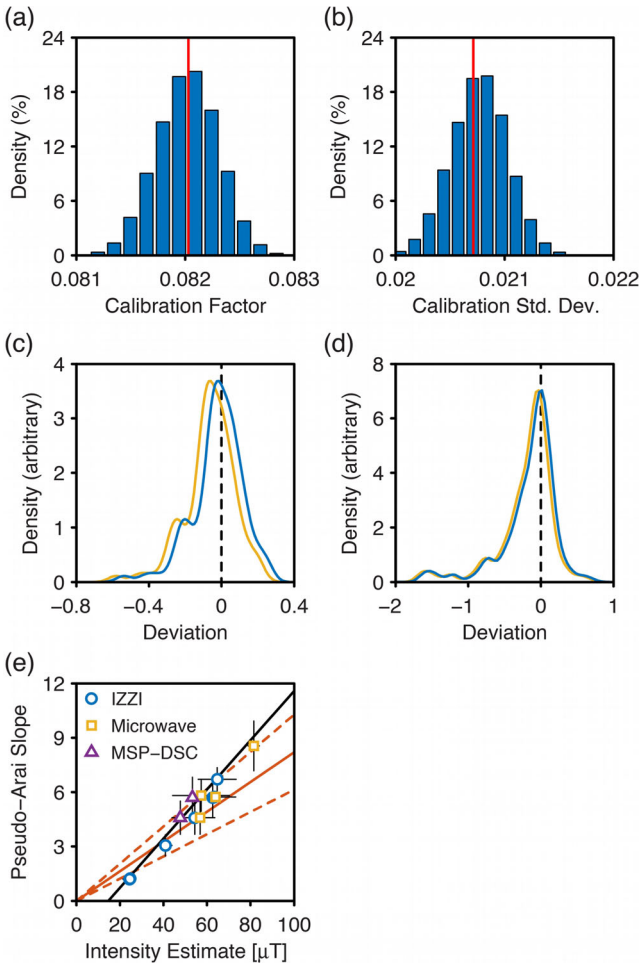

Figure 7. (a) Histogram of the Monte Carlo resampled calibration factors and (b) their associated standard deviations. In panels (a) and (b), the red lines are the values used for our preferred calibration. (c) Kernel density curves of the deviation from the known field for the selected Hawaiian data of de Groot *et al*. (2013) using the calibration before (blue) and after (yellow) rock magnetic selection based on the ARM_{120} MDF. (d) As for part (c), but applied to the unselected data of de Groot *et al*. (2013). (e) Comparison of the pseudo-Arai slopes with the palaeointensity results obtained from thermal based methods: thermal based IZZI protocol (Tauxe & Staudigel 2004), microwave method (Walton *et al*. 1996) and domain state corrected multispecimen method (MSP-DSC; Dekkers & Böhnel 2006; Fabian & Leonhardt 2010). The solid orange line represents our new mean calibration and the orange dashed lines denote the one standard deviation envelope and the solid black line is the calibration of de Groot *et al*. (2013). Data are site averages calculated from de Groot *et al*. (2015, 2016). Error bars are ± one standard deviation. For the MSP-DSC data, error bars are the average of the reported upper and lower errors.

where m_s is the mean slope of the pseudo-Arai plots and c is the calibration factor (0.0820). Because both m_s and c have associated uncertainties, both must be considered in the estimation of the uncertainty of the palaeointensity estimate. If s_s and s_c represent the standard deviations of the mean slope and calibration factor, respectively, then, via Gaussian error propagation, the standard deviation of the palaeointensity estimate (s_B) is given by

$$s_B = m_B \sqrt{\left(\frac{s_s}{m_s}\right)^2 + \left(\frac{s_c}{m_c}\right)^2}. \quad (5)$$

As an example, we consider the 1090 AD data from Hawaii, which record a relatively high palaeointensity (63.5 ± 1.8 μT; de Groot *et al*. 2013). Using the preferred data from (de Groot *et al*. 2013) the mean pseudo-Thellier slope is 6.631 ± 0.247 (±3.7 per cent). With our new calibration the mean intensity is 80.7 μT, which is higher than previously estimated. The uncertainty on this estimate, however, is much larger ±20.6 μT (±25.5 per cent) and is dominated by the inherent variability in the calibration factor.

4.6 Implications for previous results

Although a relatively new method, a considerable number of results have already been obtained using the pseudo-Thellier method (de Groot *et al*. 2013, 2014, 2015, 2016). In comparison to previous calibrations, our new calibration indicates that low palaeointensity estimates will be systematically overestimated and that high palaeointensity estimates will be underestimated (e.g. the 1090 AD data above), but in all cases, the uncertainty will be higher.

de Groot *et al*. (2015) produced a number of new data from the Canary Islands, which were contemporaneously studied by Kissel *et al*. (2015) who verified and contested some results. In particular, the Montaña Reventada lava flow yielded discordant results between de Groot *et al*. (2015) and Kissel *et al*. (2015), who obtained 81.7 ± 8.7 μT and 57.9 ± 6.4 μT, respectively. The result of de Groot *et al*. (2015) comes from both pseudo-Thellier and microwave Thellier data, but the pseudo-Thellier average from 6 results is 81.7 ± 10.9 μT. With our new calibration and uncertainty, this becomes 104.3 ± 31.2 μT. Although the intensity value has been shifted higher, due to the increased uncertainty, it is still consistent with de Groot *et al*.'s microwave results (81.6 ± 2.6 μT, $N = 3$) and within less than 1.5 standard deviations of the result of Kissel *et al*. (2015). Although the increase in the error of the pseudo-Thellier estimation does not fully reconcile the differences from these two studies, it illustrates the importance of correctly assessing the uncertainty associated with pseudo-Thellier results. It should be noted that physical sampling strategy has also been suggested as an important factor for why these two studies differ (Kissel *et al*. 2015).

de Groot *et al*. (2015) also note that one of their sites, TF-1909, records a pseudo-Thellier palaeointensity estimate that is low (22.0 μT) with respect to the expected IGRF value of ∼40 μT. Using their selected data and our calibration we estimate a value of 12.1 ± 4.0 μT, which is much lower than the expected value. We cannot reconcile this discrepancy and, following with de Groot *et al*. (2015), we call upon some, as yet unknown, rock magnetic behaviour to explain this unusually low value. We also note that employing our data selection procedures may yield a different result, but this remains untested. In the absence of such an analysis, however, this result implies that the pseudo-Thellier method may only be accurate to within a factor of ∼4.

The examples explored here illustrate the need for a more complete reassessment of the previously published pseudo-Thellier results and, due to the larger uncertainty associated with these results, effort should be made to cross-validate pseudo-Thellier palaeointensities with results from other methods. We also recommend that all future pseudo-Thellier analyses should be combined with comprehensive rock magnetic analyses, which might reveal more useful rock magnetic screening methods not identified here. Given the sensitivity of magnetometers, which allows measurement of small specimens, and the non-destructive nature of the pseudo-Thellier

method, we recommend that such rock magnetic analyses be performed on the same specimen, as was done here.

4.7 Application to extraterrestrial materials

One of the primary uses of non-heating palaeointensity methods is the study of extraterrestrial materials, which can be extremely sensitive to laboratory heating. Although palaeointensity methods based on SIRM (e.g. Gattacceca & Rochette 2004) are more common in the study of extraterrestrial materials, recently, ARM methods have become more popular (e.g. Garrick-Bethell *et al*. 2009; Lappe *et al*. 2013; Fu *et al*. 2014).

Our calibration is applicable to pseudo-Thellier experiments using a 40 μT DC bias field during ARM acquisition, but for comparison with calibrations derived from studies focusing on extraterrestrial materials it needs to be generalized for any bias field. By multiplying out the calibration factor ($c_{40} = 0.0820 \pm 0.0207$) by the ARM bias field (40 μT) we obtain a field generalized calibration of 3.280 ± 0.828, such that $B_{\mathrm{Anc}}\ (\mu\mathrm{T}) = B_{\mathrm{ARM}}\ (\mu\mathrm{T}) \times |b_{\mathrm{PA}}| / 3.280$. This value of 3.280 is slightly higher than that obtained by Yu (2010), but it is still consistent with his value of 2.60 ± 1.32 derived from thermally stabilized gabbro specimens. Our value, however, is much higher than that of Lappe *et al*. (2013), who obtained values of 0.7–1.85 for natural and synthetic dusty olivine, which may have a more complicated mineralogy than the specimens used in our study. In addition, Lappe *et al*.'s experiments used TRM bias fields of 200–1500 μT and ARM bias fields of 20–400 μT. Therefore, the effects of nonlinear remanence acquisition in large fields may be significant.

In both of these studies, and when using ARM based methods on extraterrestrial materials in general, comparison of ARM acquisition and demagnetization spectra (*cf*. Fig. 1f) is rarely performed (Garrick-Bethell *et al*. 2009; Yu 2010; Lappe *et al*. 2013). Our analysis further demonstrates that the coercivity spectra of TRM can be notably different from that of ARM (e.g. Fig. 1a), but also that ARM demagnetization spectra can be different from ARM acquisition spectra (e.g. Fig. 1f). The requirement of reciprocity, where remanences activate the same blocking spectra, whether it is blocking temperature or coercivity, is a fundamental premise of all palaeointensity studies that should be tested. The selection process that we introduce here allows for this and we recommend the same, or similar selection procedures be adopted for all ARM based palaeointensity methods.

By utilizing our selection criteria it is possible to use the pseudo-Thellier method to obtain results that are within a factor 4 (*cf*. the discussion of site TF-1909 in Section 4.6) or better (*cf*. the Hawaiian palaeointensity estimates in Figs 7c and d). This makes the accuracy of ARM methods comparable to the suggested accuracy of SIRM based methods (Gattacceca & Rochette 2004). However, to be able to better generalize the pseudo-Thellier method a more extensive data set that covers materials relevant to extraterrestrial specimens combined with a detailed exploration of ARM bias field dependence and nonlinear remanence acquisition are needed.

5 CONCLUSIONS

We have re-estimated the calibration of the pseudo-Thellier method using a collection 56 well-characterized specimens that represent a wide range of volcanic materials used for obtaining absolute palaeointensity estimates. Following careful data selection that tests fundamental assumptions about palaeointensity behaviour, 32 specimens are deemed as acceptable and yield calibration behaviour that is consistent with a zero intercept term (i.e. a specimen with zero NRM should have a pseudo-Arai slope of zero), which is a theoretical expectation that had not been met by some previous calibrations. Our final calibration factor (0.0820 ± 0.0207) exhibits variability that is consistent with other estimates and falls within the range of variability expected from SD magnetite. This calibration is applicable to pseudo-Thellier experiments using a 40 μT DC bias field during ARM acquisition, but can be generalized for any bias field (3.280 ± 0.828). Future studies, however, need to systematically explore the influence of ARM bias field to verify this calibration of the pseudo-Thellier method. Similarly, other important factors such as ARM anisotropy and the influence of AF decay rates need to be better quantified.

A systematic rock magnetic investigation reveals that, without data selection, there is a strong correlation between high-field rock magnetic behaviour and the calibration factors obtained from individual specimens. However, following data selection, we observe that the calibration factors are no longer correlated with high-field rock magnetic properties, but are correlated with weak-field parameters. This suggests that the pseudo-Thellier calibration factors are dominated by domain state variations in highly magnetic minerals (i.e. grain size and/or magnetic interactions in magnetite and/or maghemite). However, by using a new set of selection criteria (Table 1), we are able to effectively screen out much of this variation. The calibration factors remaining after applying these criteria are correlated with low-field remanence properties, which may be related to subtle variations in magnetic mineralogy of these particular specimens. Nevertheless, such subtle variations are difficult to distinguish from subtle grain size variations over a very narrow range (e.g. Fig. 3). Further study on a larger data set with a wider range of mineralogical variations may aid in identifying the effect of such subtleties.

Despite a differing approach to data selection and different ARM and AF equipment, our new calibration yields highly accurate results when applied to the independent data set of de Groot *et al*. (2013), which serves as demonstration of the veracity of our calibration. It is also consistent with results from more traditional thermal based palaeointensity methods and illustrates that the pseudo-Thellier method is a useful and valuable tool for obtaining absolute palaeointensity estimates. However, pseudo-Thellier results carry with them an inherent uncertainty associated with the calibration factor, which stems from the strong dependence of ARM on magnetic grain size. This uncertainty is at least 25 per cent and is larger than most thermal based palaeointensity methods.

An analysis of data from historical lavas suggests that the pseudo-Thellier method is accurate to within a factor of 4, but with careful data selection this may be reduced to a factor of 2. The pseudo-Thellier method, as applied here, may therefore have important applications in obtaining palaeointensities from extraterrestrial materials. Nevertheless, because of large uncertainties, the pseudo-Thellier method should be used to complement and cross-validate results from other methods and should only be used as a standalone method in cases where other methods fail or where non-destructive treatment is required.

ACKNOWLEDGEMENTS

GAP acknowledges funding from NSFC grants 41374072 and 41574063. DH acknowledges support from Australian Research Council grant DP160100805. YP and GAP acknowledges

support from CAS project XDB18010203. We thank Huafeng Qin for his assistance. We also thank reviewers Roger Fu and Yongjae Yu for their useful suggestions. Data generated in this work will be made publicly available through the MagIC database (earthref.org/MagIC/) following completion of ongoing work. Direct data requests are welcome.

REFERENCES

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.

de Groot, L.V., Biggin, A.J., Dekkers, M.J., Langereis, C.G. & Herrero-Bervera, E., 2013. Rapid regional perturbations to the recent global geomagnetic decay revealed by a new Hawaiian record, *Nat. Commun.,* **4,** doi:10.1038/ncomms3727.

de Groot, L.V., Dekkers, M.J., Visscher, M. & ter Maat, G.W., 2014. Magnetic properties and paleointensities as function of depth in a Hawaiian lava flow, *Geochem. Geophys. Geosyst.,* **15,** 1096–1112.

de Groot, L.V. *et al.*, 2015. High paleointensities for the Canary Islands constrain the Levant geomagnetic high, *Earth planet. Sci. Lett.,* **419,** 154–167.

de Groot, L.V., Pimentel, A. & Di Chiara, A., 2016. The multi-method paleointensity approach applied to volcanics from Terceira: full-vector geomagnetic data for the past 50 kyr, *Geophys. J. Int.,* **206**(1), 590–604.

Dekkers, M.J. & Böhnel, H.N., 2006. Reliable absolute palaeointensities independent of magnetic domain state, *Earth planet. Sci. Lett.,* **248,** 508–517.

Dunlop, D.J. & Argyle, K.S., 1997. Thermoremanence, anhysteretic remanence and susceptibility of submicron magnetites: nonlinear field dependence and variation with grain size, *J. geophys. Res.,* **102,** 20 199–20 210.

Dunlop, D.J. & West, G.F., 1969. An experimental evaluation of single domain theories, *Rev. Geophys.,* **7,** 709–757.

Dunlop, D.J., Bailey, M.E. & Westcott-Lewis, M.F., 1975. Lunar paleointensity determinations using anhysteretic remanence (ARM): a critique, in *Proc. Lunar Sci. Conf. 6th,* Pergamon Press, Inc., Houston, TX, pp. 3063–3069.

Fabian, K., 2006. Approach to saturation analysis of hysteresis measurements in rock magnetism and evidence for stress dominated magnetic anisotropy in young mid-ocean ridge basalt, *Phys. Earth planet. Inter.,* **154,** 299–307.

Fabian, K. & Leonhardt, R., 2010. Multiple-specimen absolute paleointensity determination: an optimal protocol including pTRM normalization, domain-state correction, and alteration test, *Earth planet. Sci. Lett.,* **297,** 84–94.

Fu, R.R. *et al.*, 2014. Solar nebula magnetic fields recorded in the Semarkona meteorite, *Science,* **346,** 1089–1092.

Garrick-Bethell, I., Weiss, B.P., Shuster, D.L. & Buz, J., 2009. Early Lunar Magnetism, *Science,* **323,** 356–359.

Gattacceca, J. & Rochette, P., 2004. Toward a robust normalized magnetic paleointensity method applied to meteorites, *Earth planet. Sci. Lett.,* **227,** 377–393.

Gillingham, D.E.W. & Stacey, F.D., 1971. Anhysteretic remanent magnetization (ARM) in magnetite grains, *Pure appl. Geophys.,* **91,** 160–165.

Hartstra, R.L., 1983. TRM, ARM and Isr of two natural magnetites of MD and PSD grain size, *Geophys. J. R. astr. Soc.,* **73,** 719–737.

Jackson, M. & Solheid, P., 2010. On the quantitative analysis and evaluation of magnetic hysteresis data, *Geochem. Geophys. Geosyst.,* **11,** Q04Z15, doi:10.1029/2009GC002932.

Kissel, C., Laj, C., Rodriguez-Gonzalez, A., Perez-Torrado, F., Carracedo, J.C. & Wandres, C., 2015. Holocene geomagnetic field intensity variations: contribution from the low latitude Canary Islands site, *Earth planet. Sci. Lett.,* **430,** 178–190.

Lappe, S.-C.L.L., Feinberg, J.M., Muxworthy, A. & Harrison, R.J., 2013. Comparison and calibration of nonheating paleointensity methods: a case study using dusty olivine, *Geochem. Geophys. Geosyst.,* **14,** 2143–2158.

Levi, S., 1977. Effect of magnetite particle size on paleointensity determinations of the geomagnetic field, *Phys. Earth planet. Inter.,* **13,** 245–259.

Levi, S. & Merrill, R.T., 1976. A comparison of ARM and TRM in magnetite, *Earth planet. Sci. Lett.,* **32,** 171–184.

Mullender, T.A.T., Frederichs, T., Hilgenfeldt, C., de Groot, L.V., Fabian, K. & Dekkers, M.J., 2016. Automated paleomagnetic and rock magnetic data acquisition with an in-line horizontal "2G" system, *Geochem. Geophys. Geosyst.,* **17,** doi:10.1002/2016GC006436.

Muxworthy, A.R. & Heslop, D., 2011. A Preisach method for estimating absolute paleofield intensity under the constraint of using only isothermal measurements: 1. Theoretical framework, *J. geophys. Res.,* **116,** B04102, doi:10.1029/2010JB007843.

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, *J. geophys. Res.,* **116,** B04103, doi: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,** 5277–5281.

Paterson, G.A., 2011. A simple test for the presence of multidomain behaviour during paleointensity experiments, *J. geophys. Res.,* **116,** B10104, doi:10.1029/2011JB008369.

Paterson, G.A., Muxworthy, A.R., Roberts, A.P. & Mac Niocaill, C., 2010a. Assessment of the usefulness of lithic clasts from pyroclastic deposits for paleointensity determination, *J. geophys. Res.,* **115,** B03104, doi:10.1029/2009JB006475.

Paterson, G.A., Roberts, A.P., Muxworthy, A.R., Mac Niocaill, C., Gurioli, L., Viramonté, J.G., Navarro, C. & Weider, S., 2010b. Paleomagnetic determination of emplacement temperatures of pyroclastic deposits: an under-utilized tool, *Bull. Volcanol.,* **72,** 309–330.

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,** 1180–1192.

Paterson, G.A., Biggin, A.J., Hodgson, E. & Hill, M.J., 2015. Thellier-type paleointensity data from multidomain specimens, *Phys. Earth planet. Inter.,* **245,** 117–133.

Rahman, A.A., Duncan, A.D. & Parry, L.G., 1973. Magnetization of multidomain magnetite particles, *Riv. Ital. Geofis.,* **22,** 259–266.

Stephenson, A. & Collinson, D.W., 1974. Lunar magnetic field palaeointensities determined by an anhysteretic remanent magnetization method, *Earth planet. Sci. Lett.,* **23,** 220–228.

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,** Q02H06, doi:10.1029/2003GC000635.

Thellier, E. & Thellier, O., 1959. Sur l'intensité du champ magnétique terrestre dans le passé historique et géologique, *Ann. Geophys.,* **15,** 285–376.

Walton, D., Snape, S., Rolph, T.C., Shaw, J. & Share, J., 1996. Application of ferrimagnetic resonance heating to palaeointensity determinations, *Phys. Earth planet. Inter.,* **94,** 183–186.

Yu, Y., 2010. Paleointensity determination using anhysteretic remanence and saturation isothermal remanence, *Geochem. Geophys. Geosyst.,* **11,** Q02Z12, doi:10.1029/2009GC002804.

Yu, Y., Dunlop, D.J. & Özdemir, Ö., 2002. Partial anhysteretic remanent magnetization in magnetite 2. Reciprocity, *J. geophys. Res.,* **107,** EPM 8-1–EPM 8-9, doi:10.1029/2001JB001269.

Yu, Y., Dunlop, D.J. & Özdemir, Ö., 2003. Are ARM and TRM analogs? Thellier analysis of ARM and pseudo-Thellier analysis of TRM, *Earth planet. Sci. Lett.,* **205,** 325–336.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this paper:

Figure S1. Comparison of hysteresis (a) coercivity ratio and (b) squareness before and after heating. Data before heating are from sister specimens from Paterson *et al*. (2010b) and unpublished Emeishan basalt data. We note that specimen heterogeneity may cause some of the differences. (c) Day plot comparison of our heated specimens (coloured symbols) and unheated natural volcanic materials (black dots). In all plots, the blue circles are the Emeishan specimens (groups A and C), the yellow squares are the Mt. St. Helens specimens (group D) and the purple triangles are the Vesuvius specimens (group L). In parts (a) and (b) the dashed lines and 1-to-1 ratio lines. The unheated natural volcanic hysteresis data in part (c) are sourced from Paterson *et al*. (2010b) and Muxworthy *et al*. (2011) and include lava flows and pyroclastic lithic clasts from seven localities.

Table S1. Mean selected pseudo-Arai slopes (± one standard deviation) from the used specimens in the four different NRM fields.

Table S2. The accepted specimens and their mean calibration factors and rock magnetic properties. ARM120 MDF, $B_{1/2\mathrm{ARM120}}$ and Residual NRM are calculated as the averages values from the selected field experiments.

Table S3. Pseudo-Arai and ARM–ARM plot curvature from the data of Yu *et al*. (2003). Curvature is calculated from all data points. Data were digitized from Yu *et al*. (2003) and the entire plot curvature was calculated following Paterson (2011).

(http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggw349/-/DC1)