Geophys. J. Int. (2013) doi: 10.1093/gji/ggt033
Geophysical Journal International Advance Access published February 19, 2013

The effects of anisotropic and non-linear thermoremanent magnetizations on Thellier-type paleointensity data

Greig A. Paterson

Paleomagnetism and Geochronology Laboratory, Key Laboratory of Earth's Deep Interior, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 10029, *China. E-mail: greig.paterson@mail.iggcas.ac.cn*

Accepted 2013 January 25. Received 2013 January 11; in original form 2012 November 15

SUMMARY

Numerous non-ideal factors can influence paleointensity data, but the detection of these factors remains problematic and new approaches to understanding how paleointensity data behave are needed. In this study, a recently developed stochastic model of single domain (SD) paleointensity behaviour is expanded to investigate the effects that anisotropic and non-linear thermoremanent magnetizations (TRMs) have on the paleointensity results and the parameters used to select data. The model results indicate that before applying any form of correction these non-ideal factors can produce results that are self-consistent, but highly inaccurate. The methods that are currently used to correct for anisotropic and non-linear TRMs are effective and greatly increase the likelihood of obtaining accurate results. The corrections, however, do not restore the results to those of ideal SD samples measured with the same laboratory-to-ancient field ratio, but the data are restored to those of ideal SD samples with the equivalent laboratory-to-ancient magnetization ratios (M_{Lab}/M_{Anc}). The simulations indicate that non-linear and anisotropic TRM have no or only a weak influence on the parameters commonly used to select paleointensity data, which means that these non-ideal factors are effectively undetectable. These new models suggest that the paleointensity behaviour of thermally/chemically stable SD samples, whether they are ideally behaved, anisotropy or non-linear TRM corrected, is near universal and depends only on M_{Lab}/M_{Anc} and the choice of paleointensity protocol (i.e. Coe-type versus Thellier). Given the high self-consistency and highly inaccurate results that anisotropic and non-linear TRM can yield, it is essential to test for such effects and all Thellier-type paleointensity studies *must* include tests for anisotropic and non-linear TRM to assert the reliability of the data obtained.

Key words: Archaeomagnetism; Palaeointensity.

1 INTRODUCTION

The study of paleointensity (the strength of the ancient geomagnetic field) is a challenging endeavour, but is important for understanding the temporal evolution of the geodynamo. Obtaining and demonstrating a reliable paleointensity estimate, however, is hindered by several complicating factors that often lead to the failure of experiments or uncertainty that inaccurate results have been successfully rejected. These factors are collectively termed non-ideal behaviour and include such influences as large magnetic grains, alteration in the laboratory or in nature, non-linear acquisition of thermoremanent magnetization (TRM) with applied field, anisotropic TRM, or magnetic interactions between magnetic grains, among others.

Our primary means of identifying data that suffer non-ideal effects lies in the choice of experimental procedure used to obtain paleointensity estimates and a suite of selection parameters derived from various checks and intuitive notions of how ideal data should behave. The range of selection parameters used in paleointensity studies is diverse and expanding (e.g. Coe *et al.* 1978; Prévot *et al.* 1985; Leonhardt *et al.* 2004; Tauxe & Staudigel 2004; Paterson 2011). There is, however, no consensus as to which parameters and which cut-off values are most appropriate to select data. Traditionally, paleointensity analysts choose their preferred set of selection parameters and define cut-off values in a somewhat arbitrary fashion. Although the choice of selection cut-off values is influenced by our physical understanding of non-ideal behaviour (e.g. higher repeatability of laboratory TRM should indicate less alteration), the ability to objectively define selection criteria is essential to asserting the reliability of paleointensity studies. Our experimental knowledge of the various physical mechanisms, however, is insufficient to fully characterize the effects that non-ideal behaviour has on paleointensity data. Isolating individual non-ideal factors in laboratory control experiments can be difficult and the time-consuming nature of paleointensity experiments often limits the number of data

available, which hinders statistical analysis. Numerical simulations provide an alternative means to investigate non-ideal behaviour, but the application of numerical results to real world situations is limited by the lack of variability of numerical results when compared with real data.

Recently, Paterson *et al.* (2012) developed a stochastic model to investigate the statistical nature of paleointensity data. The model captures the variability of real paleointensity data by incorporating the effects of experimental noise into simulations of paleointensity experiments. To construct the most realistic model possible, the sources and magnitudes of noise variations were constrained using over 75,000 real data measurements and sample (un)blocking behaviour was based on the demagnetization data from ∼2000 real samples. The model assumes hypothetical ideal samples, that is, non-interacting single domain (SD) grains that obey Thellier's laws of additivity, independence and reciprocity (Thellier & Thellier 1959), and do not chemically alter during the experiment. Under this assumption, Paterson *et al.* (2012) demonstrated how experimental noise could influence paleointensity results and the parameters that are commonly used to select data. They found that under certain circumstances (often, but not limited to, cases where the laboratory field strength was weaker than the ancient field), some widely used selection criteria are likely to be overly strict and result in the rejection of well-behaved data that yield accurate results, but are only subject to expected levels of experimental noise. Paterson *et al.* (2012) proposed the use of a natural remanent magnetization (NRM) fraction (f) of $\geq$0.35 for data selection to minimize the manifestation of noise. They also defined lower threshold values for selection parameters, which represent values below which non-ideal behaviour cannot be distinguished from expected levels of experimental noise. These thresholds are one step towards objectively constraining paleointensity data selection.

In this study, the model of Paterson *et al.* (2012) is extended to investigate the effects of two types of non-ideal behaviour on SD samples: anisotropic TRM and non-linear TRM acquisition with applied field. In Section 2, the basic principles of the original model, the theory behind these two non-ideal factors, and how they have been incorporated into the model are outlined. The influence of anisotropic TRM and non-linear TRM acquisition are investigated for a number of frequently used experimental protocols in Section 3 and the implications that these results have for paleointensity studies are discussed in Section 4.

2 MODEL DESCRIPTION AND BACKGROUND THEORY

2.1 The basic stochastic model

Full details of the stochastic model are given by Paterson *et al.* (2012) and only the key points are summarized here. During a typical paleointensity experiment, the NRM of a sample is progressively replaced by a TRM acquired in the laboratory using a known field, $B_{\rm Lab}$. For an ideal sample the ratio of the NRM lost to the TRM gained at a given temperature is equivalent to the ratio of ancient magnetic field strength ($B_{\rm Anc}$) to $B_{\rm Lab}$:

$$\frac{\rm NRM_{Anc}}{\rm TRM_{Lab}} = \frac{B_{\rm Anc}}{B_{\rm Lab}}. \quad (1)$$

By progressively replacing the NRM, multiple estimates of this ratio can be made and when plotted on an Arai diagram (Nagata *et al.* 1963) an ideal sample will plot as a straight line with slope $-\frac{B_{\rm Anc}}{B_{\rm Lab}}$.

Although SD theory can be used to simulate a paleointensity experiment (Néel 1949), in the stochastic model a phenomenological approach is used. Under the assumption that Thellier's laws hold true, the unblocking and blocking of an ideal sample can be described by identical distributions of (un)blocking temperatures. After heating to temperature T_i the NRM remaining (after demagnetization) and the TRM gained (after remagnetization) can be described by:

$$\mathrm{NRM}(T_i) = B_{\rm Anc}\int_{T_i}^{T_c} f(T)\,\mathrm{d}T, \quad (2)$$

$$\mathrm{TRM}(T_i) = B_{\rm Lab}\int_{T_0}^{T_i} f(T)\,\mathrm{d}T, \quad (3)$$

where T_c is the Curie temperature, T_0 is room temperature and $f(T)$ is the distribution of (un)blocking temperatures, which is taken to follow a beta distribution with parameters constrained by real demagnetization data (Paterson *et al.* 2012). In practice the TRM gained cannot be measured directly, but must be calculated from the total magnetization (J), which is the vector summation of NRM(T_i) and TRM(T_i):

$$\begin{aligned} J(T_i) &= \mathrm{NRM}(T_i) + \mathrm{TRM}(T_i) \\ &= B_{\rm Anc}\int_{T_i}^{T_c} f(T)\,\mathrm{d}T + B_{\rm Lab}\int_{T_0}^{T_i} f(T)\,\mathrm{d}T, \end{aligned} \quad (4)$$

$$\mathrm{TRM}(T_i) = J(T_i) - \mathrm{NRM}(T_i). \quad (5)$$

A 3-D remanence vector can be obtained by assuming that $f(T)$ is isotropic and describing the field directions and magnetizations as Cartesian vectors. The assumption that $f(T)$ is isotropic is equivalent to assuming a random assemblage of uniaxial SD grains in Néel theory. Details of how the model has be modified to simulate anisotropic and non-linear TRM are given in Sections 2.2 and 2.3, respectively.

The stochastic model incorporates 10 sources of experimental noise and includes the repeatability of furnace peak temperature, furnace thermal gradients, variations in furnace hold time and cooling rate, variations in the applied laboratory field (magnitude and direction with respect to a sample), residual fields, magnetometer measurement and background noise levels and variations of sample position between steps with respect to the magnetometer measurements axes. The data and methods used to constrain the magnitudes of these sources of experimental noise are outlined in detail by Paterson (*et al.* 2012). Each noise source is added to the paleointensity simulations at the level at which it was measured (i.e. temperature noise is added to the temperature steps, etc.) and in a sequence that represents the physical sequence of a real experiment. The noise is then numerically propagated into variations of magnetization and selection parameters.

Given that noise is a statistical phenomenon (i.e. it follows a distribution) it is necessary to undertake Monte Carlo analyses to characterize the resultant distributions of the paleointensity data and parameters. The analysis undertaken in this study uses the same basic procedure as Paterson *et al.* (2012), which is as follows:

(1) Randomly select a (un)blocking spectrum from the real data fits.
(2) Create a randomly oriented NRM.
(3) Simulate the paleointensity experiment with $B_{\rm Lab}$ applied along the z-axis.

(iv) Randomly select a segment comprising at least 4 points, with a negative slope, with a fraction (f) ≥ 0.35, and a gap factor (g) > 0.

(4) Calculate the paleointensity results for the best-fit segment.

(5) Repeat steps 1–5 for 10^4 simulations.

The criteria in step 4 are necessary to avoid unrealistic fits and the minimum fraction is that recommended by Paterson *et al.* (2012).

For simplicity, in the original models of Paterson *et al.* $B_{\rm Anc}$ and $B_{\rm Lab}$ were scaled such that $B_{\rm Lab} = 1$ was equivalent to a 40 μT field. $B_{\rm Lab}$ was used to scale field uncertainties to make the model results dimensionless. In these models, this scaling is removed and all fields are rescaled to values equivalent to μT, but given the phenomenological nature of the models and for simplicity the results are still assumed to be dimensionless. This allows the non-linear TRM models to be compared more easily to real data. For the anisotropy models, $B_{\rm Lab}$ is taken to be constant ($B_{\rm Lab} = 40$ μT) and $B_{\rm Anc}$ is varied (20, 40 and 80 μT). For the non-linear TRM models both $B_{\rm Anc}$ and $B_{\rm Lab}$ are varied. Other modifications that are specific to the modelled non-ideal factors are detailed below.

Full definitions of the various selection parameters investigated are given in the Supporting Information. To quantify accuracy, paleointensity estimates are described as a deviation from the expected value. The deviation is the logarithm of the estimate normalized by the expected value. Zero deviation is exactly the expected value; positive and negative values are over- and underestimates, respectively. Results are defined as accurate if they are within a factor of 1.1 of the expected paleointensity value (i.e. within ~10 per cent). The scatter of results is the standard deviation defined as a percentage of the mean result. The resultant distributions of paleointensity selection parameters are analyzed by taking the 95th percentiles of the distributions, as was done by Paterson *et al.* (2012). These 95 per cent thresholds represent the upper limit of variability of the selection parameters that results from biasing factors (i.e. anisotropic or non-linear TRM) and/or experimental noise.

2.2 Anisotropic TRM

A frequent, and often valid, assumption of paleo- and archeomagnetic studies is that the samples are magnetically isotropic (i.e. the magnetic recording is independent of the direction of the applied field). Remanence anisotropy, however, can be significant in some paleo- and archeomagnetic samples and has long been recognized as a potential biasing factor for paleomagnetic directions (e.g. Uyeda *et al.* 1963) and for paleointensity determinations (e.g. Rogers *et al.* 1979; Aitken *et al.* 1981). Anisotropy of magnetic remanence affecting paleointensity data has been widely recognized in archeological artifacts (e.g. Rogers *et al.* 1979; Aitken *et al.* 1981), geological materials (e.g. Selkin *et al.* 2000; Yu & Dunlop 2001) and recently in copper slag deposits, which are the tailings of copper smelting (e.g. Ben-Yosef *et al.* 2008; Shaar *et al.* 2010).

For weak magnetic fields, anisotropy of susceptibility and remanent magnetization is the result of an anisotropic distribution of magnetic easy axes within a sample (e.g. Fanjat *et al.* 2012). TRM anisotropy can form through several processes all of which produce a preferential alignment of magnetic grains that can deflect the remanent magnetization away from the geomagnetic field vector. Such processes include manufacturing induced fabrics in bricks and pottery (e.g. Aitken *et al.* 1981), deformation fabrics (common in metamorphosed rocks, e.g. Selkin *et al.* 2000), elongated particles (e.g. Selkin *et al.* 2000), exsolution textures (e.g. Feinberg *et al.* 2006) or through dendritic crystal growth (e.g. Shaar *et al.* 2010).

The first correction for anisotropic paleointensity estimates was proposed by Veitch *et al.* (1984), but later expanded by Selkin *et al.* (2000). Both methods are theoretically identical, but differ in the details of their application. The methods are based on the premise that in a weak magnetic field a TRM vector (**TRM**) is related to the applied field vector (**B**) by:

$$\mathbf{TRM} = \chi_{TRM}\mathbf{B}, \tag{6}$$

where χ_{TRM} is the TRM anisotropy tensor, which is assumed to be temperature invariant. The anisotropy tensor can be experimentally determined by giving a sample a full TRM in six different directions; the reader is referred to Tauxe (2010) for full details on measuring anisotropy tensors.

The first step to correct for paleointensity anisotropy is to determine a unit vector in the direction of the ancient field ($\hat{\mathbf{B}}_{\rm Anc}$, where the hat denotes a unit vector). Given a unit vector in the direction of the characteristic NRM direction ($\hat{\mathbf{M}}_{\mathbf{ChRM}}$) isolated from an NRM thermal demagnetization experiment (or the NRM steps of a paleointensity experiment), $\hat{\mathbf{B}}_{\rm Anc}$ can be calculated from:

$$\hat{\mathbf{B}}_{\rm Anc} = \frac{\chi_{TRM}^{-1}\hat{\mathbf{M}}_{\mathbf{ChRM}}}{\left|\chi_{TRM}^{-1}\hat{\mathbf{M}}_{\mathbf{ChRM}}\right|}. \tag{7}$$

The anisotropy tensor can be used to calculate the paleointensity correction factor, c, which is the ratio of a magnetization gained in the direction of $\hat{\mathbf{B}}_{\rm Lab}$ to a magnetization gained in the direction of $\hat{\mathbf{B}}_{\rm Anc}$:

$$c = \frac{\left|\chi_{TRM}\hat{\mathbf{B}}_{\rm Lab}\right|}{\left|\chi_{TRM}\hat{\mathbf{B}}_{\rm Anc}\right|}. \tag{8}$$

When $\hat{\mathbf{B}}_{\rm Anc}$ and $\hat{\mathbf{B}}_{\rm Lab}$ are parallel $c = 1$. c values >1 result when $\hat{\mathbf{B}}_{\rm Anc}$ is closer to the minor anisotropy axis and $\hat{\mathbf{B}}_{\rm Lab}$ is closer to the major axis, and c values <1 occur when $\hat{\mathbf{B}}_{\rm Lab}$ is closer to the minor anisotropy axis and $\hat{\mathbf{B}}_{\rm Anc}$ is closer to the major axis.

The method of Veitch *et al.* (1984) corrects the final paleointensity estimate directly, such that the corrected estimate ($B^*_{\rm Anc}$) is given by:

$$B^*_{\rm Anc} = cB_{\rm Lab}\frac{\rm NRM_{Anc}}{\rm TRM_{Lab}}. \tag{9}$$

The method of Selkin *et al.* (2000), however, corrects each TRM acquisition before analysis on the Arai plot. The laboratory TRM vector ($\mathbf{TRM}_{\rm Lab}$) is scaled by R ($= 1/c$) to yield the anisotropy corrected magnetization vector ($\mathbf{TRM}^*_{\rm Lab}$):

$$\mathbf{TRM}^*_{\rm Lab} = R \times \mathbf{TRM}_{\rm Lab}. \tag{10}$$

The analysis of the paleointensity data can proceed as normal (i.e. fitting a best-fit line segment on an Arai plot), but using $\mathbf{TRM}^*_{\rm Lab}$ in place of $\mathbf{TRM}_{\rm Lab}$. Using this approach a corrected paleointensity estimate can be determined from the slope of the corrected Arai plot:

$$B^*_{\rm Anc} = B_{\rm Lab}\frac{\rm NRM_{Anc}}{\rm TRM^*_{Lab}} = B_{\rm Lab}\frac{\rm NRM_{Anc}}{R \times \rm TRM_{Lab}} = cB_{\rm Lab}\frac{\rm NRM_{Anc}}{\rm TRM_{Lab}}. \tag{11}$$

Therefore, both corrections are theoretically equivalent. The main difference between the two methods is the calculation of the paleointensity selection parameters. The correction of Veitch *et al.* (1984) is applied only to the paleointensity estimate after analysis, whereas the method of Selkin *et al.* (2000) is applied to all TRM data before any analysis, which includes any TRM checks used to test for non-ideal behaviour. Therefore the method of Veitch *et al.* (1984) will not influence paleointensity selection parameters,

however, the method of Selkin *et al.* (2000) influences some parameters and these effects are investigated.

Anisotropic TRM can be phenomenologically modelled by:

$$\mathbf{TRM}(T_i) = \chi_{TRM}\mathbf{B}_{\rm Lab}\int_{T_0}^{T_i} f(T)\,{\rm d}T. \quad (12)$$

The anisotropy tensor, χ_{TRM}, describes a shape ellipsoid with three orthogonal eigenvectors $\mathbf{V}_i$ and corresponding (unnormalized) eigenvalues s_i. Although Fanjat *et al.* (2012) describe a method to theoretically derive χ_{TRM} from a specified distribution of magnetic easy exes, for simplicity χ_{TRM} is determined based on the eigenvalues necessary to model a desired degree of anisotropy ($P = \frac{s_1}{s_3}$; Nagata 1961). For an isotropic sample $s_1 = s_2 = s_3$ and $P = 1$. For anisotropic samples the shape ellipsoid can take on three general forms: prolate (i.e. $s_1 > s_2 = s_3$), oblate (i.e. $s_1 = s_2 > s_3$) and triaxial (i.e. $s_1 > s_2 > s_3$). However, since the effects of remanence anisotropy on paleointensity estimates depend only on the relative effects along the $B_{\rm Anc}$ and $B_{\rm Lab}$ directions they are independent of the shape of the ellipsoid and the absolute scale of the eigenvalues. Therefore, for simplicity, it is assumed that $s_1 = Ps_3$ and $s_3 = 1$ and a prolate shape ellipsoid ($s_{1,2,3} = [P, 1, 1]$) is modelled. For a given sample the degree of anisotropy controls the range of possible c values experienced during a paleointensity experiment, such that $c_{\rm min} = 1/P$ and $c_{\rm max} = P$. The degrees of anisotropy measured from metamorphic rocks Selkin *et al.* (2000) and from copper slag Shaar *et al.* (2010) range from ∼1.04 to ∼2.42. In these simulations, 11 degrees of anisotropy up to $P = 3.0$ are modelled. For each simulated experiment the orientation of the anisotropy tensor is randomly generated.

The simulated measurement of the TRM anisotropy tensor is separate from the paleointensity experiment. The samples are 'heated' above the Curie temperature to simulate a full TRM. The magnitude of applied field is taken to be the same as for the paleointensity experiment ($B_{\rm Lab} = 40\ \mu$T), but is applied in six different orientations ($\pm x$, $\pm y$, $\pm z$). The anisotropy tensor is calculated following Tauxe (2010). For real samples these extra heating steps increase the potential for a sample to chemically alter and an additional step to test the reproducibility of TRM acquisition is typically performed. A repeat heating in the $+x$ orientation is simulated to test TRM acquisition repeatability. Given that the samples are 'heated' above T_c, temperature variations are taken to be negligible and not included (see Paterson *et al.* 2012), but all other noise sources (e.g. measurement noise, orientation variations, etc.) are included.

2.3 Non-linear TRM acquisition

The implicit assumption in eqs (1) and (6) is that magnetization is linearly proportional to the field in which it was acquired. Single domain theory (Néel 1949), however, predicts that TRM is proportional to the hyperbolic tangent of the applied field (B) as described by:

$${\rm TRM} = M_{rs}\tanh\left(\frac{VM_s(T_b)B}{kT_b}\right), \quad (13)$$

where M_{rs} is the saturation remanent magnetization, V is the grain volume, $M_s(T_b)$ is the saturation magnetization at the blocking temperature (T_b), and k is the Boltzmann constant.

The linear approximation generally holds true for most SD grains in weak fields that are on the order of geomagnetic field strengths (i.e. $\lesssim 100\ \mu$T). However, Selkin *et al.* (2007) theoretically and experimental demonstrated that large elongate particles exhibited high degrees of non-linearity in fields as low as 40–60 μT. They proposed that paleointensity measurements be accompanied by additional measurements of TRM as a function of applied field and that the data could be approximated by the simple relation:

$${\rm TRM} = A_1\tanh(A_2B), \quad (14)$$

where A_1 and A_2 are scaling coefficients. Selkin *et al.* (2007) demonstrated that this simple approximation fits real data well, a conclusion corroborated by Shaar *et al.* (2010).

For an SD sample, in the absence of chemical alteration (i.e. the coefficients A_1 and A_2 do not change), the slope of the line (b) on an Arai plot can be described by:

$$|b| = \frac{{\rm NRM_{Anc}}}{{\rm TRM_{Lab}}} = \frac{\tanh(A_2B_{\rm Anc})}{\tanh(A_2B_{\rm Lab})}, \quad (15)$$

and $B_{\rm Anc}$ is:

$$B_{\rm Anc} = \frac{\tanh^{-1}(|b|\tanh(A_2B_{\rm Lab}))}{A_2}. \quad (16)$$

It can be noted that the linear approximation is valid in the limit as A_2 tends to zero and that linearity of magnetization with applied field is a special case of the more general non-linear form.

From eqs (13) and (14) it can be seen the A_1 represents the saturation remanence and that $\tanh(A_2B)$ can be viewed as the percentage saturation in a field of B. Seven degrees of non-linearity are modelled based on the percentage saturation in a field of 80 μT: 0 per cent (i.e. linear TRM), 20, 40, 60, 80, 90 and 99 per cent. Using these saturations and eq. (14), A_2 can be calculated. A_1 is a scaling constant, which, for simplicity, is set to $1/A_2$ (this relation is seen in the data from Shaar *et al.* 2010). The copper slag samples studied by Shaar *et al.* (2010) have 80 μT saturations of 6–58 per cent with an average of ∼41 per cent. The samples investigated by Selkin *et al.* (2008) have 80 μT saturations of 2–99 per cent with an average of ∼77 per cent.

The modelling of non-linear TRM is straightforward and can be achieved by modifying eqs (2) and (3) to:

$${\rm NRM}(T_i) = A_1\tanh(A_2B_{\rm Anc})\int_{T_i}^{T_c} f(T){\rm d}T, \quad (17)$$

$${\rm TRM}(T_i) = A_1\tanh(A_2B_{\rm Lab})\int_{T_0}^{T_i} f(T){\rm d}T. \quad (18)$$

It should be noted that an implicit assumption of eqs (14)–(18) is that the degree of non-linearity is independent of the blocking temperature, which, given that $A_2 \propto \frac{M_s(T_b)}{T_b}$ (eq. 13), is strictly not the case. For the purposes of this phenomenological model, however, it is assumed that this temperature dependence is negligible. This is equivalent to assuming that the degree of non-linearity of all partial TRMs (pTRMs) is identical to that of a total TRM.

Following the paleointensity experiment the linearity of TRM is tested at six fields (B_{NLT}): 10, 20, 40, 60, 80 and 100 μT. At each field a full TRM is imparted ($\rm TRM_{NLT}$). If the linear correlation (r^2) between B_{NLT} and $\rm TRM_{NLT}$ is >0.999 it is assumed that TRM is linear with applied field, otherwise eq. (14) is fitted to the data to obtain estimates for A_1 and A_2. These best-fit coefficients are then used to determine the corrected paleointensity results using eq. (16). As is the case for anisotropic TRM, the repeatability of the first non-linear TRM measurement is checked to test for alteration. This step typically tests the TRM acquired in the first applied field, but since true alteration is absent from this model the sequence of steps is not a factor and TRM repeatability tested at all fields strengths.

In the absence of experimental noise eq. (16) always yields the correct paleointensity, irrespective of the degree of non-linearity. In the presence of noise, however, $|b|\tanh(A_2 B_{\rm Lab})$ can be greater than unity. The inverse hyperbolic tangent of this, however, is a complex number. In such cases only the real component is taken. In their study of non-linear and anisotropic TRM from single crystals of plagioclase, Usui & Nakamura (2009) noted that for some samples the errors were 'too large to give meaningful values'. This was due to the inverse hyperbolic tangent yielding complex numbers (Y. Usui, personal communication, 2012). Although Usui & Nakamura (2009) used a combined anisotropy and non-linear TRM correction, this demonstrates that this type of behaviour can be seen in real data. In general, these situations are uncommon and occur only when $B_{\rm Anc}$ is close to a saturation field. The overall effect of these results, however, is negligible; excluding such cases does not greatly change the final results. Examples are given in the Supporting Information.

2.4 Experimental protocols

The simulations presented here use the same experimental procedure as Paterson *et al.* (2012). The paleointensity experiments consist of 14 temperature steps up to the Curie temperature (T_c). The model assumes that SD magnetite ($T_c = 580\,^\circ$C) is the only magnetic carrier and the procedure uses temperature steps of 0, 75, 150, 225, 300, 375, 450, 500, 530, 560, 565, 570, 575 and 580 °C. Where possible, the experiments include both pTRM and pTRM tail checks. The checks are performed at alternating temperature steps (i.e. pTRM checks at 75, 150, 300, etc. and pTRM tail checks at 75, 225, 375, etc.). The pTRM check temperatures are chosen to cover a continuous range. For example, the check at 75 °C is performed after a peak temperature of 150 °C, the check at 150 °C is performed after a peak temperature of 300 °C, and so on.

Four of the mostly commonly used paleointensity protocols are modelled: the Thellier protocol (Thellier & Thellier 1959), the Coe protocol (Coe 1967), the Aitken/Walton protocol (Walton 1979; Aitken *et al.* 1988) and the IZZI protocol (Yu *et al.* 2004). The order in which the sample is de-/remagnetized differs for each protocol. In the Thellier-Thellier protocol a sample is first heated and cooled in an applied field of $B_{\rm Lab}$, before a second heating and cooling to the same temperature, but with an applied field of -$B_{\rm Lab}$. The NRM remaining can then be calculated as half vector sum of these two resultant magnetizations and half the vector difference is the TRM gained. In the Coe protocol, the NRM remaining is directly measured by heating the sample first in zero-field. The second heating is in an applied field, which is used to calculate the TRM gained. The Aitken protocol reverses this sequence with the first step being the in-field step. The IZZI protocol alternates between the Aitken sequence (in-field, zero-field; IZ) and the Coe sequence (zero-field, in-field; ZI).

Paterson *et al.* (2012) demonstrated that the propagation of noise for the Coe, Aitken and IZZI protocols was near identical and these three protocols yielded the same results. For the Thellier protocol, however, due to the vector calculations used to determine both the NRM and TRM the propagation of noise is different. In general NRM variance from the Thellier protocols is higher than the other protocols, but the TRM variance is lower. In the Coe, Aitken and IZZI protocols the NRM remaining is directly measured and has noise that is associated with a single experimental step. For the Thellier protocol two TRM acquisition steps are used to determine the NRM remaining, which means that noise from two experimental steps are propagated into the NRM determination. For all protocols the TRM gained is not measure directly, but is calculated by the subtraction of two vectors. For the Coe, Aitken and IZZI protocols the vector calculation is a straight vector subtraction (eq. 5). For the Thellier protocol, however, the TRM acquisition is measured twice and the vector calculation must be averaged to yield the TRM gained. This also averages the effects of noise and tends to reduce the TRM variance of the Thellier protocol when compared to other protocols.

For the models investigated here, the results for the Coe, Aitken and IZZI protocols are near identical. For brevity, the results for only the Coe protocol are presented, but these results are representative of those from the Aitken and IZZI protocols. The results for the Thellier protocol are discussed, but only the selected figures are presented; the full figures are presented in the Supporting Information.

3 RESULTS

3.1 Anisotropic TRM

The descriptive statistics from the Coe protocol simulations with $B_{\rm Lab} = B_{\rm Anc}$ are shown in Fig. 1. In this figure the results are plotted as functions of c and P. For the c plots the data are binned by c with each bin 0.2 units wide. All degrees of anisotropy are analyzed. For the P plots, only simulated results where $c = \frac{1}{1.05}$–1.05 are used. This range of c values represents results when $B_{\rm Lab}$ is close to parallel to $B_{\rm Anc}$ and the effects of anisotropy should be close to a minimum and therefore highlight s the influence of P. In both cases, bins with fewer than 10^3 results are excluded.

Considering the proportion of inaccurate results (Figs 1a and b), before correction there is a strong dependence on c, with $c \lesssim 0.83$ ($= \frac{1}{1.2}$) and $c \gtrsim 1.2$ yielding 100 per cent inaccurate results. After correction, the proportion of inaccurate results is greatly reduced, with $c \geq 1$ yielding the lowest proportions. The maximum percentage of inaccurate results is ∼14 per cent and occurs when $c \approx 0.45$. Although the relative effects of increasing degree of anisotropy are substantial (i.e. a fourfold increase), the absolute increase from ∼1 to ∼4 per cent is small (Fig. 1b). The deviation of the mean result before correction is strongly dependent on c with values < 1 overestimating the true paleointensity and $c > 1$ yielding underestimates (Fig. 1c). After correction, however, all values of c yield accurate mean results. The deviation of the mean result, however, is not greatly influenced by the degree of anisotropy. The scatter of the results has only a weak dependence on c and no dependence on P (Figs 1 e and f). The scatter decreases with increasing c and is reduced from ∼7 to ∼4 per cent as c increases from 0.45 to 2 (Fig. 1 e). Anisotropy correction has only a small effect, but consistently reduces the scatter.

The results for the Thellier protocol are shown in Fig. 2. The behaviour of the results with c is near identical to Coe results, but the percentage of inaccurate results after correction increases with c and reaches a maximum of ∼13 per cent at $c = 2$. The results of the Thellier protocol, however, have a stronger dependence on the degree of anisotropy. The proportion of inaccurate results increases from ∼1 to ∼9–12 per cent as P increases from 1.0 to 3.0 (Fig. 2b). The deviation of the mean results increasingly underestimates the true paleointensity as P increases (Fig. 2d). The effect, however, is small and the mean results are consistently accurate to within $\lesssim$2–3 per cent of the expected value. The scatter of the results noticeably increases with the degree of anisotropy and increases from ∼3 per cent at $P = 1.0$ to ∼5 per cent at $P = 3.0$.

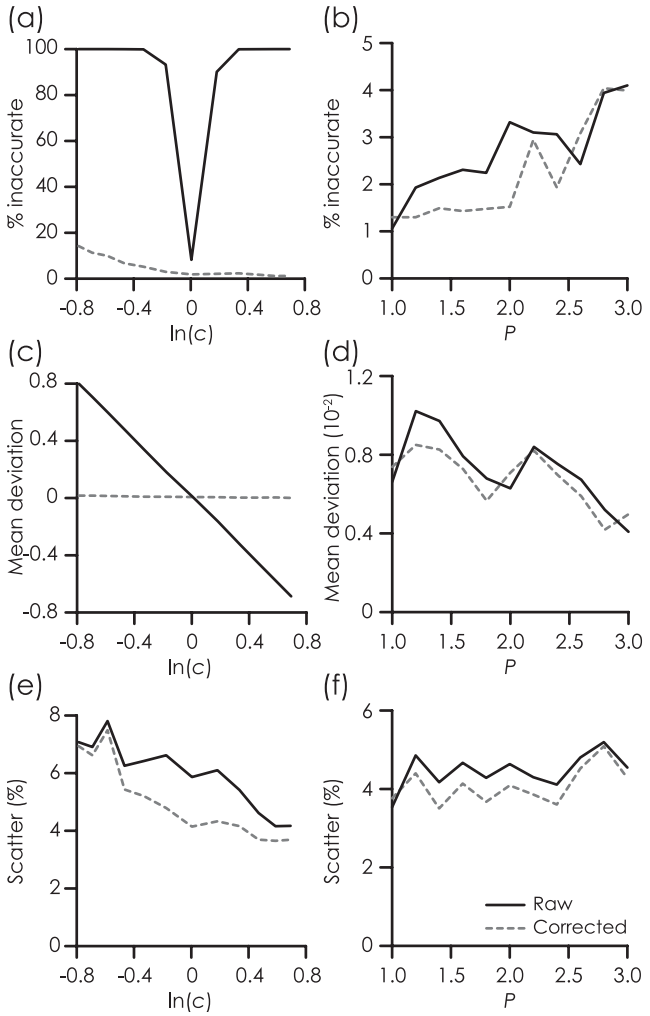

Figure 1. Variation of the descriptive statistics from the Coe protocol simulations as a function of c and P, before and after correction ($B_{\mathrm{Lab}} = B_{\mathrm{Anc}}$). (a)–(b) The percentage of inaccurate results. (c)–(d) The deviation of the mean results from the expected value. (e)–(f) The scatter of the results as a percentage of the mean. In plots b, d and f the results are from simulations with $c = \frac{1}{1.05}$–1.05.

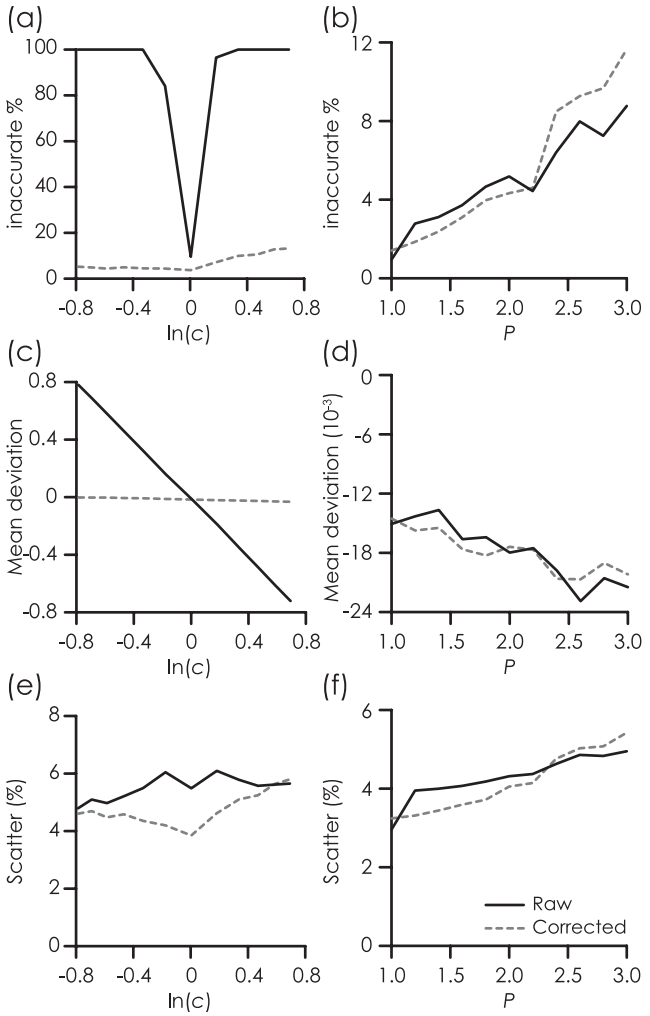

Figure 2. Variation of the descriptive statistics from the Thellier protocol simulations as a function of c and P, before and after correction ($B_{\mathrm{Lab}} = B_{\mathrm{Anc}}$). (a)–(b) The percentage of inaccurate results. (c)–(d) The deviation of the mean results from the expected value. (e)–(f) The scatter of the results as a percentage of the mean. In plots b, d and f the results are from simulations with $c = \frac{1}{1.05}$–1.05.

The variation of the descriptive statistics with P is due, in part, to the dependence of the 'measured' c value ($\tilde{c}$) on the degree of anisotropy. The empirical cumulative distribution functions (ECDFs) for $\tilde{c}/c$ for selected degrees of anisotropy are shown in Fig. 3 for both the Coe and Thellier protocols. For low degrees of anisotropy $\tilde{c}$ is consistently close to the expected value, but as P increases $\tilde{c}$ increasingly deviates from c. This increasing deviation with P produces a greater proportion of inaccurate results and increases the scatter of the results. This effect is more pronounced for the Thellier method than for the Coe due to the greater effect of noise on the estimation of the NRM and hence $\hat{\mathbf{B}}_{\mathrm{Anc}}$. For most practical purposes, however, the effect should be small: most archeological materials have $P \lesssim 2.0$ (e.g. Genevey & Gallet 2002; Ben-Yosef *et al.* 2009; Shaar *et al.* 2010).

In all of the above cases, after correction the descriptive statistics have a dependence on c, which means that the effectiveness of the correction is variable with c. The expectation, however, is that the anisotropy correction can fully restore the results to the equivalent results for an isotropic sample measured with the same $B_{\mathrm{Lab}}/B_{\mathrm{Anc}}$ combination. That is to say, the corrected results should be independent of c and correspond to the values at $c = 1$. The behaviour of the corrected descriptive statistics, however, can be

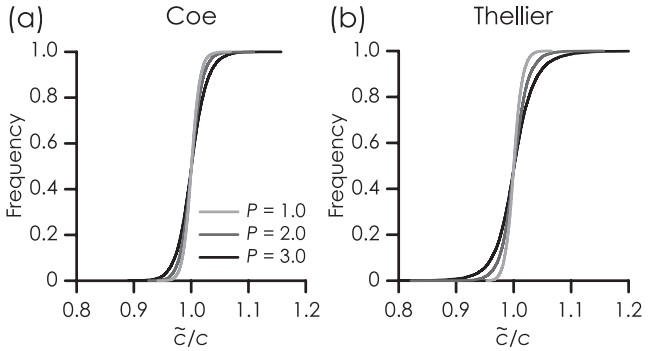

Figure 3. The empirical cumulative distributions of $\tilde{c}/c$ for various degrees of anisotropy. For (a) the Coe protocol and (b) the Thellier protocol.

understood in the context of isotropic behaviour by considering the ratio of the total magnetization gained in $B_{\rm Lab}$ ($M_{\rm Lab}$) to the total magnetization gained in $B_{\rm Anc}$ ($M_{\rm Anc}$):

$$\frac{M_{\rm Lab}}{M_{\rm Anc}} = \frac{|\chi_{TRM}\mathbf{B}_{\rm Lab}|}{|\chi_{TRM}\mathbf{B}_{\rm Anc}|} = c\frac{B_{\rm Lab}}{B_{\rm Anc}}. \quad (19)$$

The descriptive statistics from the Coe and Thellier protocols are plotted against $M_{\rm Lab}/M_{\rm Anc}$ in Fig. 4. In this figure all three $B_{\rm Lab}/B_{\rm Anc}$ combinations are used and the data are split into 18 bins each with $\geq 10^3$ results. With the exception of the isotropic results, all degree of anisotropy are used (i.e. $P = 1.2$–3.0). The circles on each plot are the 95 per cent thresholds from isotropic simulation at five $B_{\rm Lab}/B_{\rm Anc}$ strength combinations ($B_{\rm Lab} = \frac{1}{4}B_{\rm Anc}$ to $B_{\rm Lab} = 4B_{\rm Anc}$). The three sharp drops in the percentage of inaccurate results before correction (Fig. 4a and b) are a result of the simulations where $c \approx 1$ (i.e. $B_{\rm Lab}$ is near parallel to $B_{\rm Anc}$) and the effects of anisotropy are at a minimum. These results plot as three distinct drops due the three $B_{\rm Lab}/B_{\rm Anc}$ combinations used in the anisotropy simulations.

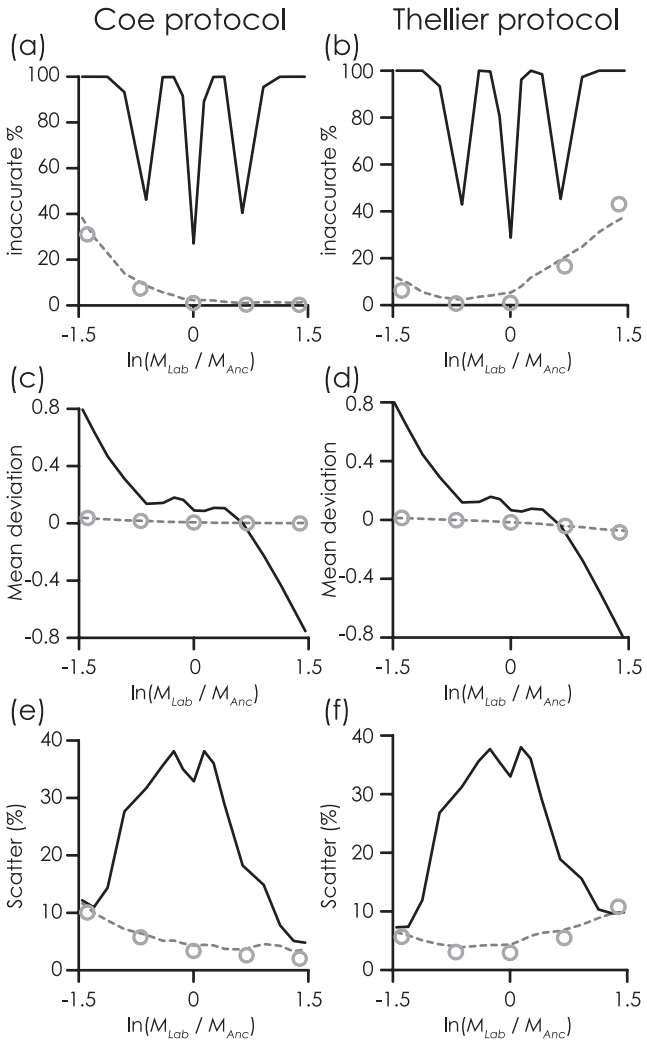

Figure 4. Variation of the descriptive statistics from the Coe and Thellier protocols as a function of the ratio of the total TRM acquired in $B_{\rm Lab}$ ($M_{\rm Lab}$) to that acquired in $B_{\rm Anc}$ ($M_{\rm Anc}$). (a)–(b) The percentage of inaccurate results. (c)–(d) The deviation of the mean results from the expected value. (e)–(f) The scatter of the results as a percentage of the mean. The solid (dashed) lines represent the results before (after) correction. The plots include all three $B_{\rm Lab}/B_{\rm Anc}$ strength combinations, but the results from the isotropic simulations ($P = 1$) are not included. The circles represent the results from simulations of isotropic samples using five $B_{\rm Lab}/B_{\rm Anc}$ combinations. In parts (c)–(f) the paleointensity estimates from different $B_{\rm Lab}/B_{\rm Anc}$ strength combinations have been normalized by the expected before combining the results.

For the anisotropy corrected results in Fig. 4 it can be seen that instead of restoring the results to the expected isotropic results from the same $B_{\rm Lab}/B_{\rm Anc}$ ratio, the corrected results are restored to expected isotropic results from the equivalent $M_{\rm Lab}/M_{\rm Anc}$. That is, anisotropy corrected results behave identically to ideal SD results measured with a different $B_{\rm Lab}/B_{\rm Anc}$ combination. The small differences between the corrected anisotropic results and the isotropic results are due to influence of the degree of anisotropy discussed above. Therefore the behaviour of anisotropy corrected paleointensity estimates can be understood in terms of the behaviour of isotropic data.

The paleointensity selection parameters are plotted as a function of $M_{\rm Lab}/M_{\rm Anc}$ for the Coe protocol in Fig. 5. The equivalent figure for the Thellier protocol is given in the Supporting Information. The correction for anisotropic TRM only affects as small number of parameters. The reasons for this are discussed below, but in the following description only the uncorrected results are considered. As is the case for the descriptive statistics, most parameters from the anisotropic simulations follow the trend of those from isotropic samples (Fig. 5). The parameters from the Thellier protocol also follow the same trends as those from the isotropic Thellier simulations. Many parameters, however, do not follow the isotropic trend exactly and some (e.g. the quality factor, q, Fig. 5b) deviate from isotropic behaviour. Deviations from the isotropic values are due in part to binning of the data, but are mostly due to the degree of anisotropy. The effects of P on some paleointensity selection parameters are illustrated in Fig. 6. For the β, q and *DRAT* parameters (Figs 6a)–c), as the degree of anisotropy increases the parameter thresholds increase (decrease for q) with the effect being most pronounced at higher $B_{\rm Lab}$ strengths. For other parameters (e.g. $DRAT_{Tail}$, Fig. 6d) the degree of anisotropy has no effect and the parameter thresholds for anisotropic samples follow the isotropic trend (e.g. Fig. 5k). In the case of β and *DRAT*, increasing anisotropy increases the parameter thresholds. At high $B_{\rm Lab}/B_{\rm Anc}$ ratios, and hence high $M_{\rm Lab}/M_{\rm Anc}$ ratios, high degrees of anisotropy will produce parameter thresholds that are higher than for isotropic samples, which can been seen in Figs 5(a) and (g). For q the opposite is true, with high P producing lower threshold values (Fig. 6b), hence anisotropic results with high $M_{\rm Lab}/M_{\rm Anc}$ ratios will underestimate the isotropic values (Fig. 5b). The influence of P on the parameters thresholds, however, is too small to be detected by most typical selection criteria thresholds. For example, $\beta \leq 0.1$ or $q \geq 1$ are commonly use selection criteria.

The anisotropy correction of Selkin *et al.* (2000) only affects the *DRAT*-style parameters and the pTRM tail check δt^* (Figs 5 g–i, k and m); all other parameters are unaffected. This correction method rescales the TRM measurements only (eq. 10), therefore parameters that are based solely on NRM measurements are not affected (e.g. *MAD* or α). Similarly, parameters where the raw check value and the normalizer are both scaled by R (e.g. β or δCK), or where the data are re-scaled to calculate the parameter (e.g. k or *SSE*), are unaffected by the correction. For the *DRAT*-style parameters the raw check values are either scaled by R or not at all (i.e. $DRAT_{Tail}$). The length of the best-fit line on the Arai plot, which is the normalizer for *DRAT*-style parameters, is only partially scaled by R and is transformed from $\sqrt{\Delta x^2 + \Delta y^2}$ before correction to $\sqrt{R\Delta x^2 + \Delta y^2}$ after correction (Δx and Δy are the TRM and NRM lengths of the best-fit line on the

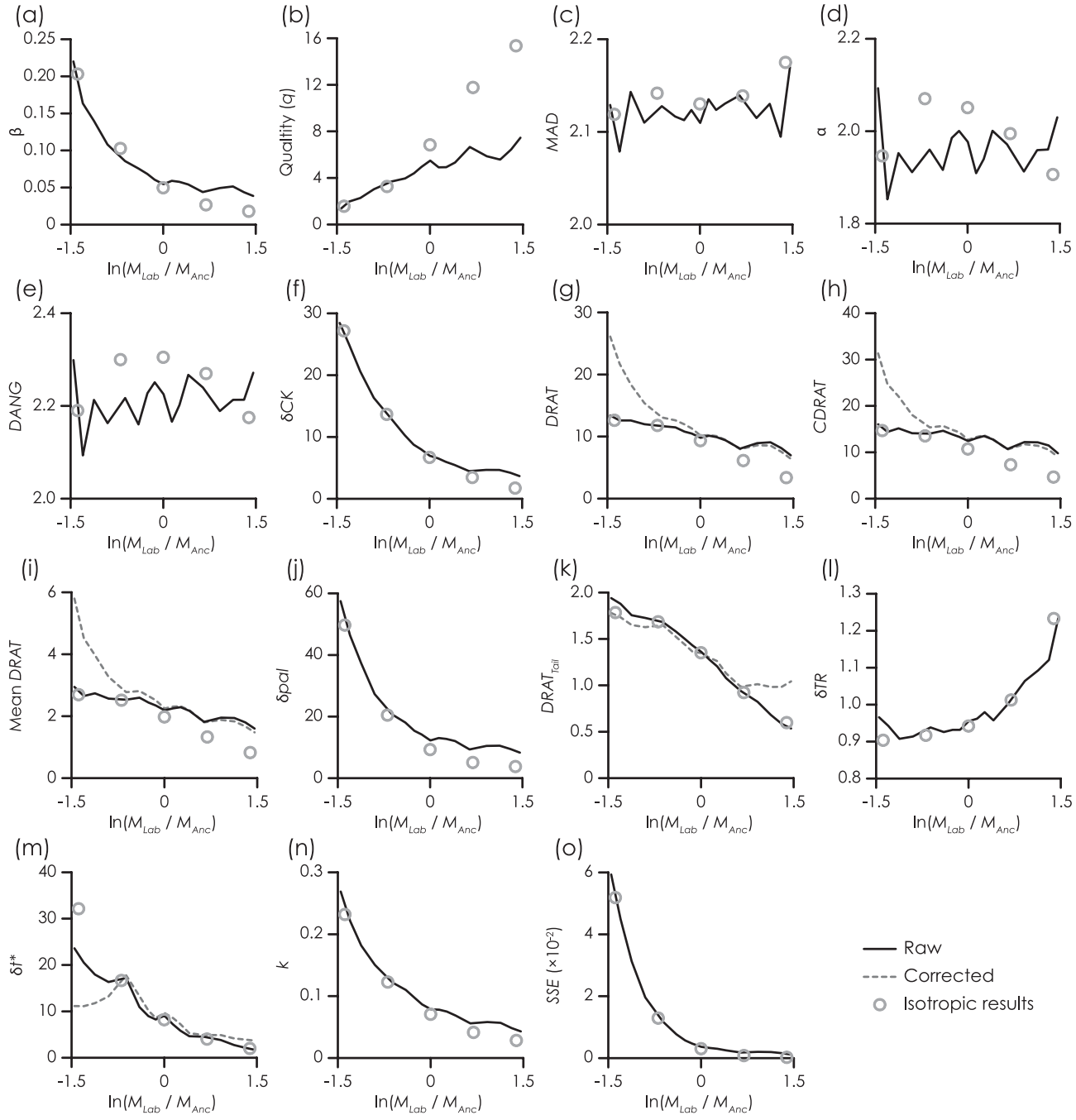

Figure 5. Behaviour of paleointensity selection parameter 95 per cent thresholds from the anisotropic TRM simulations of the Coe protocol. Parameters are plotted as a function of $M_{\rm Lab}/M_{\rm Anc}$. (a)–(e) Arai plot and directional parameters. (f)–(j) pTRM check parameters. (k)–(m) pTRM tail check parameters. (n)–(o) Arai plot curvature parameters. Only the *DRAT*-style parameters and $\delta t*$ are affected by the correction. Symbols are the same as in Fig. 4.

Arai plot, respectively). Therefore, when calculating *DRAT*-style parameters, the *R*s do not cancel. The effect of the correction, however, is generally detrimental and increases the parameter values (Figs 5 g–i, k and m). The slope on the Arai plot is used in the calculation of pTRM tail check δt^*, which is why it is affected by the correction. With the exception of $DRAT_{Tail}$, the anisotropy correction only has an effect at low $M_{\rm Lab}/M_{\rm Anc}$ values. Given the deterioration of the *DRAT*-style thresholds and in all cases the increased deviation from the expected isotropic values, applying the anisotropy correction to paleointensity selection parameters should be avoided.

3.2 Non-linear TRM acquisition

The paleointensity descriptive statistics as a function of non-linearity from the Coe protocol with various $B_{\rm Lab}/B_{\rm Anc}$ combinations are shown in Fig. 7. When the degree of non-linearity (measured as the percentage of saturation in a 80 μT field) is low, $\lesssim$ 40 per cent, corrected and uncorrected results are near identical and well behaved (low proportion of inaccurate results, accurate mean and low scatter). This indicates that these degrees of non-linearity have a negligible effect on the final results. With the exception of the case where $B_{\rm Lab} = B_{\rm Anc} = 40\ \mu$T (Figs 7b, f and j), all uncorrected results

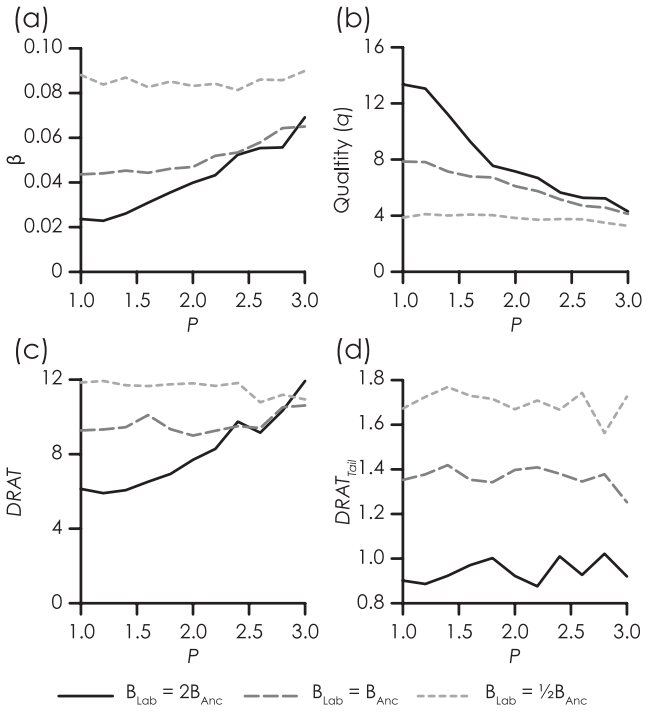

Figure 6. Behaviour of selected paleointensity parameter thresholds from the Coe protocol simulations as a function of *P*. (a)–(b) Arai plot parameters, (c) pTRM check parameter *DRAT*, (d) pTRM tail check parameter $DRAT_{Tail}$.

have a high proportion of inaccurate results and yield mean results that deviate greatly for the expected results when the degree of non-linearity is high. The scatter of the results, however, is consistently low (< 6 per cent) irrespective of the degree of non-linearity or B_{Lab}/B_{Anc}. The exception is when $B_{Lab} = B_{Anc} = 40$ where, before correction, the results are accurate and the scatter is low at all degrees of non-linearity. This is a general feature of the case where $B_{Lab} = B_{Anc}$ and is not limited to presented the case.

After correction the percentage of inaccurate results is reduced and the mean results are accurate (within ∼10 per cent of the excepted values) irrespective of the degree of non-linearity. The exception to this improvement is when $B_{Lab} = B_{Anc}$, where the percentage of inaccurate results and the deviation of the mean are higher than the uncorrected values at high degrees of non-linearity (Figs 7 f and i), but the values are still relatively low. For all B_{Lab}/B_{Anc} combinations the scatter of results after correction increases with increasing non-linearity. The highest scatter occurs when B_{Lab} is half of B_{Anc} (Fig. 7k).

Considering the $B_{Anc} = 80$, $B_{Lab} = 40$ and the $B_{Anc} = 40$, $B_{Lab} = 80$ cases (the last two columns in Fig. 7, respectively). When $B_{Lab} < B_{Anc}$ the uncorrected results behave poorly at high degrees of non-linearity (nearly all results are inaccurate and the mean results greatly underestimate the true paleointensity; Figs 7c and g). Even after correction, however, a high degree of non-linearity yields a large proportion of inaccurate results and a high scatter (Figs 7c and k). When $B_{Lab} > B_{Anc}$ the uncorrected results also yield high proportions of inaccurate results, but the mean result overestimates the true paleointensity (Figs 7 d and h). The corrected results, however, represent a large improvement at all degrees of non-linearity, with a low proportion of inaccurate resusts (< 12 per cent) and a consistently low scatter ($\lesssim$ 8 per cent). Combined with the results when $B_{Lab} = B_{Anc}$, this suggests that $B_{Lab} \geq B_{Anc}$ is the optimum field configuration for SD paleointensity experiments using the Coe, Aitken or IZZI protocols, as was suggested by Paterson *et al.* (2012).

The equivalent of Fig. 7 for the Thellier protocol is given in the Supporting Information. The same general trends that are seen from the Coe protocol simulations are present in the Thellier results. $B_{Lab} = B_{Anc}$ yields well-behaved results before correction at all degrees of non-linearity, but at high degrees, corrected results have higher proportions of inaccurate results and a higher scatter. When $B_{Lab} < B_{Anc}$, at high degrees of non-linearity (99 per cent saturation at $B = 80\ \mu$T), corrected results have high proportions of inaccurate results (> 70 per cent) and high scatters (>20 per cent). As is the case for the Coe protocol, when $B_{Lab} > B_{Anc}$ the corrected results are consistently well-behaved irrespective of the degree of non-linearity. The mean results are all accurate and the scatter is low, however, the proportion of inaccurate results is relatively high compared with the Coe protocol, typically ∼15–17 per cent for all degrees of non-linearity. This is a result of the sensitivity of the Thellier protocol to noise, which tends to produce more inaccurate results when $B_{Lab} > B_{Anc}$ (Paterson *et al.* 2012).

The descriptive statistics from all B_{Lab}/B_{Anc} and non-linearity combinations (63 models of the Coe protocol) are shown in Fig. 8. The results are plotted as scatter plots of the degree of saturation in B_{Lab} and B_{Anc}. For the uncorrected simulations (Figs 8a–c), when the saturations in B_{Lab} and B_{Anc} are both $\lesssim$ 50 per cent the results tend to have low proportions of inaccurate results (< 20 per cent), absolute mean deviations < 0.2 (with most results being < 0.1, i.e. accurate), and with low scatters (< 3–6 per cent). Outside of this region and when $B_{Lab} \neq B_{Anc}$ uncorrected results are much more inaccurate and scattered. After correction (Figs 8d–f) the proportion of inaccurate results decreases, the accuracy of the mean values increases, but scatter increases. A number of conditions, however, still yield poor data, but all of the poorer results (i.e. higher percentage of inaccurate results, more inaccurate mean and greater scatter) tend to occur when B_{Anc} is close to a saturation field ($\gtrsim$ 80–85 per cent saturation). Along the $B_{Lab} = B_{Anc}$ lines (dashed lines in Figs 8d–f) the results tend to behave better, but at B_{Anc} saturations of > 60 per cent the scatter of results increases and at saturations of ≥ 80 per cent more than 20 per cent of the results are inaccurate. Above the $B_{Lab} = B_{Anc}$ lines (i.e. when $B_{Lab} > B_{Anc}$) results are consistently well behaved with low proportions of inaccurate results, accurate mean results and low scatters, even at B_{Anc} saturations of ∼85 per cent. The corrected results are also plotted as a function of M_{Lab}/M_{Anc} (Figs 8g–i). In general, when $\ln(M_{Lab}/M_{Anc}) > 0$ (i.e. $B_{Lab} > B_{Anc}$) the corrected data from non-linear TRMs (the lines in Figs 8g–i) follow the same trends as the results from the linear TRMs (the symbols in Figs 8g–i). When $\ln(M_{Lab}/M_{Anc}) \leq 0$ (i.e. $B_{Lab} \leq B_{Anc}$) the corrected statistics become more erratic and the data from non-linear TRMs do not perfectly follow the same trends as the results from the linear TRMs. This erratic behaviour is a consequence of the poor results obtained when B_{Anc} is close to a saturation field.

The correction for non-linear TRM is applied only to the final paleointensity estimate and does not influence the selection parameters (i.e. corrected and uncorrected selection parameters are identical), therefore only selection parameters from the uncorrected results are considered. The variation of the 95 per cent thresholds for commonly used selection parameters from the Thellier, Coe and Aitken protocols are shown in Fig. 9. The parameters are plotted as a function of the ratio of M_{Lab}-to-M_{Anc}; no clear trend exists if we consider the ratio of B_{Lab}-to-B_{Anc}. The fact that the parameters from non-linear TRM paleointensity data vary with the magnetization ratio and not of the field ratio is the general trend for non-linear TRM paleointensity selection parameters and statistics; observable trends

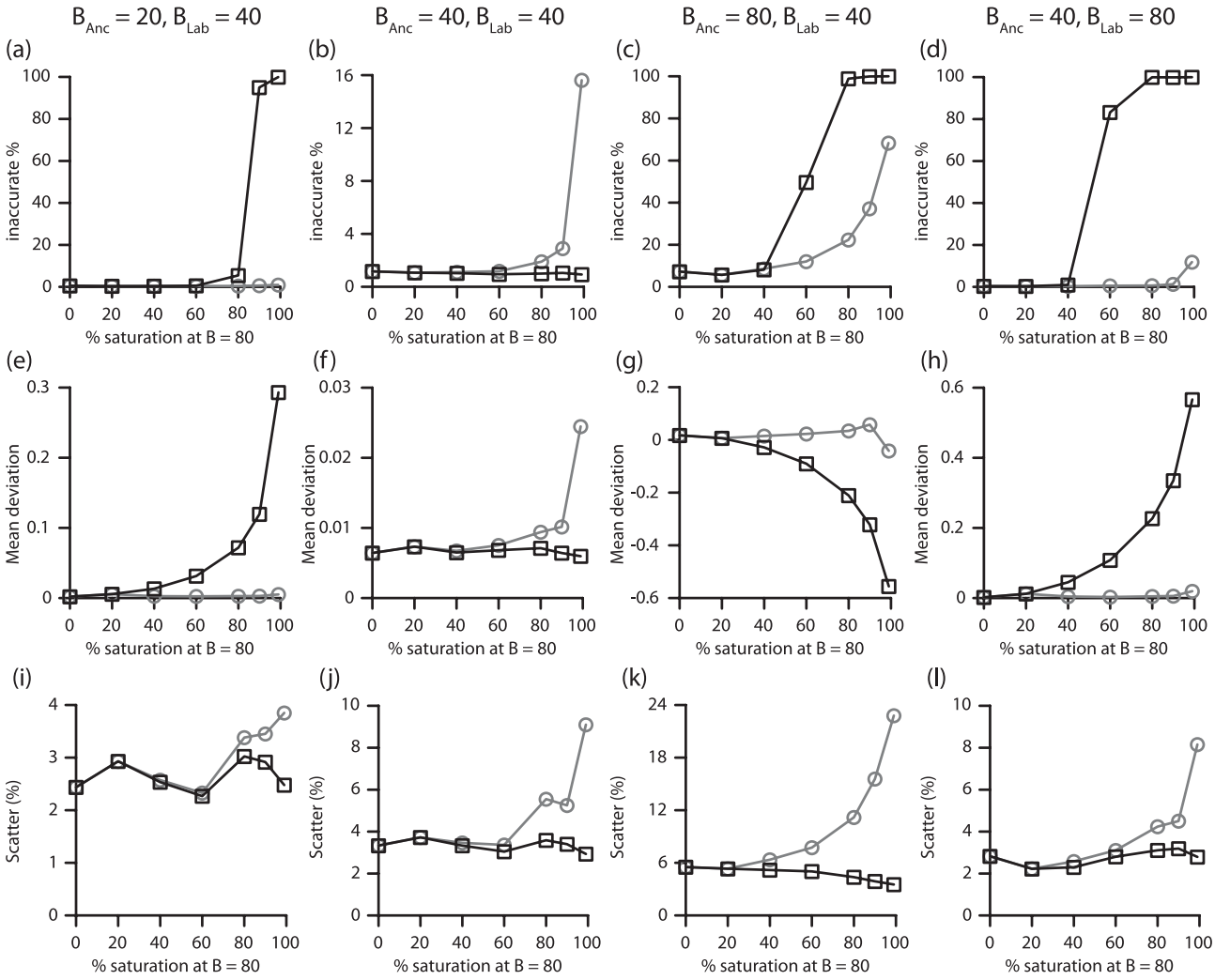

Figure 7. The paleointensity descriptive statistics for the effects of non-linear TRM at various combinations of B_{Anc} and B_{Lab}. (a)–(d) The percentage of inaccurate results. (e)–(h) The deviation of the mean results. (i)–(l) The scatter of the results as a percentage of the mean results. Squares (circles) represent uncorrected (corrected) results.

as a function of field ratio are the special case where TRM gained is linearly proportional to the applied field. In these plots the solid lines represent the results from the non-linear TRM models and the symbols represent the results from the linear TRM models (i.e. 0 per cent saturation). The IZZI protocol results are omitted for clarity as they coincide with the results from the Coe protocol. For all parameters and for all protocols the results from non-linear TRMs follow the trends of the linear TRMs. This means that non-linear TRMs do not affect the parameters used to select paleointensity data and that all variations in the parameter 95 per cent thresholds are controlled by the ratio of M_{Lab}-to-M_{Anc} and experimental noise. In short, non-linear TRM cannot be detected by commonly used selection parameters and therefore, should be tested for separately.

The results from these simulations have a magnetization/field range ($M_{\mathrm{Lab}} = \frac{1}{4}M_{\mathrm{Anc}}$ to $M_{\mathrm{Lab}} = 4M_{\mathrm{Anc}}$) that is larger than that explored by Paterson *et al.* (2012) ($M_{\mathrm{Lab}} = \frac{1}{2}M_{\mathrm{Anc}}$ to $M_{\mathrm{Lab}} = 2M_{\mathrm{Anc}}$). With this extended range, parameter behaviour not previous seen by Paterson *et al.* (2012) is evident, particularly for the Aitken and Thellier protocols. For the Thellier protocol when $M_{\mathrm{Lab}}/M_{\mathrm{Anc}} < 0.5$ (logarithm < -0.69) the β, k and *SSE* thresholds start to rise as opposed to continuously decrease as was suggested before. The quality factor (q) 95 per cent threshold reaches a peak at $M_{\mathrm{Lab}}/M_{\mathrm{Anc}} \approx 0.5$, but then falls at higher ratios. For the Aitken protocol, Paterson *et al.* (2012) noted that the directional parameter thresholds (*MAD*, α and *DANG*) were higher than those for the Coe protocol when $B_{\mathrm{Lab}} = 2B_{\mathrm{Anc}}$ ($M_{\mathrm{Lab}}/M_{\mathrm{Anc}} = 2$; logarithm $= 0.69$). This trend continues and the 95 per cent thresholds for these parameters from the Aitken protocol increase as $M_{\mathrm{Lab}}/M_{\mathrm{Anc}}$ increases. When $M_{\mathrm{Lab}}/M_{\mathrm{Anc}} > 2$ the β and k thresholds begin to rise and the q threshold decreases. Although the behaviour of the Coe and IZZI protocols is near identical, the behaviour of the Aitken protocol becomes discernibly different when $B_{\mathrm{Lab}} > B_{\mathrm{Anc}}$.

3.3 TRM reproducibility checks

The repeatability of TRM acquisition (δTRM) during measurement of an anisotropy tensor and the degree of non-linearity is shown in Fig. 10. The values plotted are the 95th percentiles of the difference between two repeat TRM acquisition steps as a percentage of the first acquisition. That is to say, for 95 per cent of samples subject to expected levels of experimental noise, TRM should be reproducibly to within $\leq \delta TRM$. For anisotropic samples δTRM varies as a

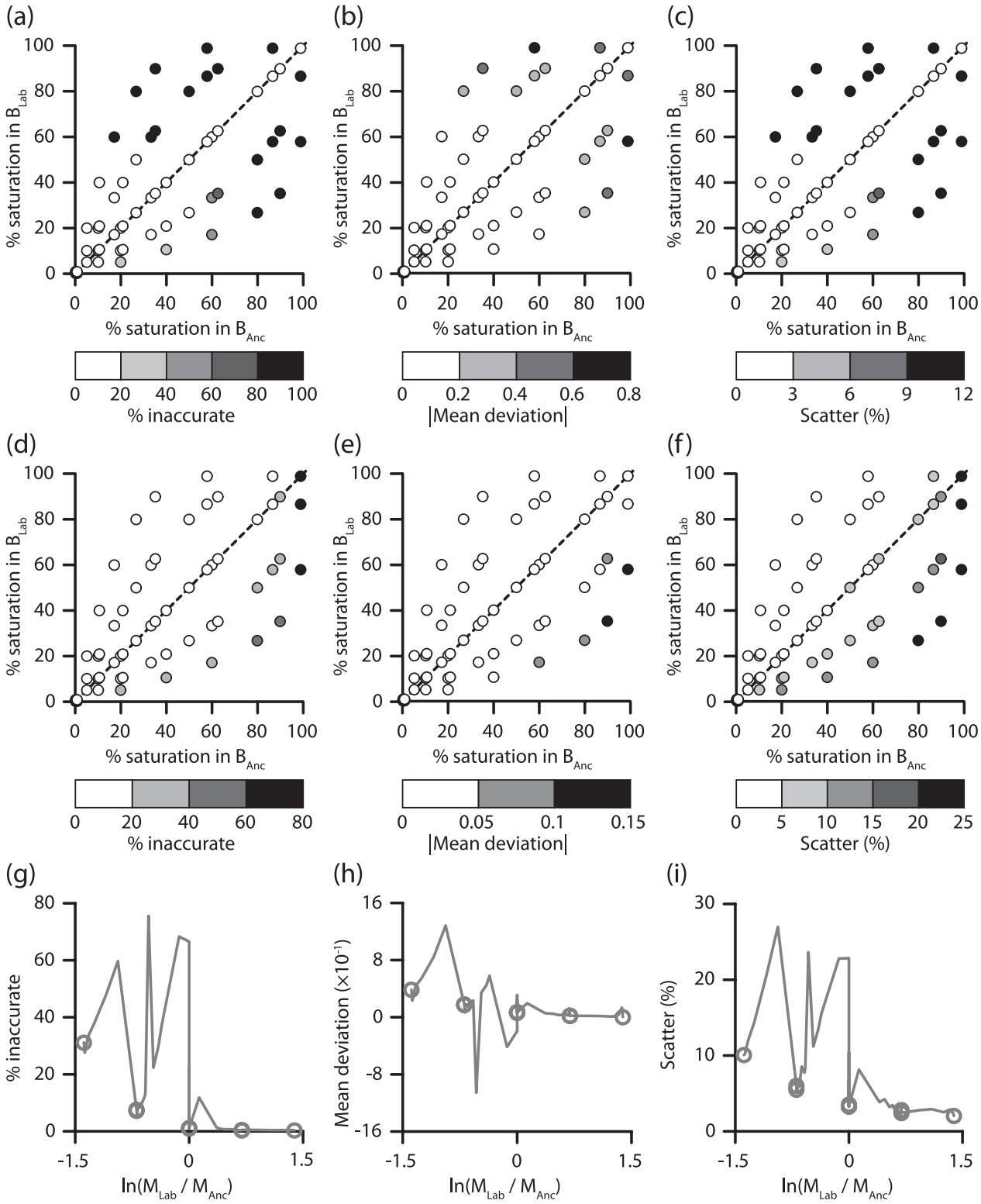

Figure 8. The paleointensity descriptive statistics from the Coe protocol simulations as colour-coded scatter plots of percentage saturation in B_{Lab} against percentage saturation in B_{Anc}. (a)–(c) Before and (d)–(f) after correction for the effects of non-linear TRM. (h)–(i) The descriptive statistics after correction as a function of $M_{\mathrm{Lab}}/M_{\mathrm{Anc}}$. The dashed lines in parts (a)–(f) mark where $B_{\mathrm{Lab}} = B_{\mathrm{Anc}}$. In parts (h)–(i) the circles denote results from simulations with linear TRM.

function of P, but is independent of applied field strength. For isotropic samples 95 per cent of samples will have TRM acquisitions that are reproducibly to within better than ∼1 per cent, but this increases to a maximum of ∼6 per cent for $P = 3$ (Fig. 10a). For non-linear TRM, the repeatability of TRM to check for possible alteration is approximately independent of the applied field, but has a weak dependence of the degree of non-linearity (Fig. 10b). For samples that that reach 0–80 per cent TRM saturation in an 80 μT

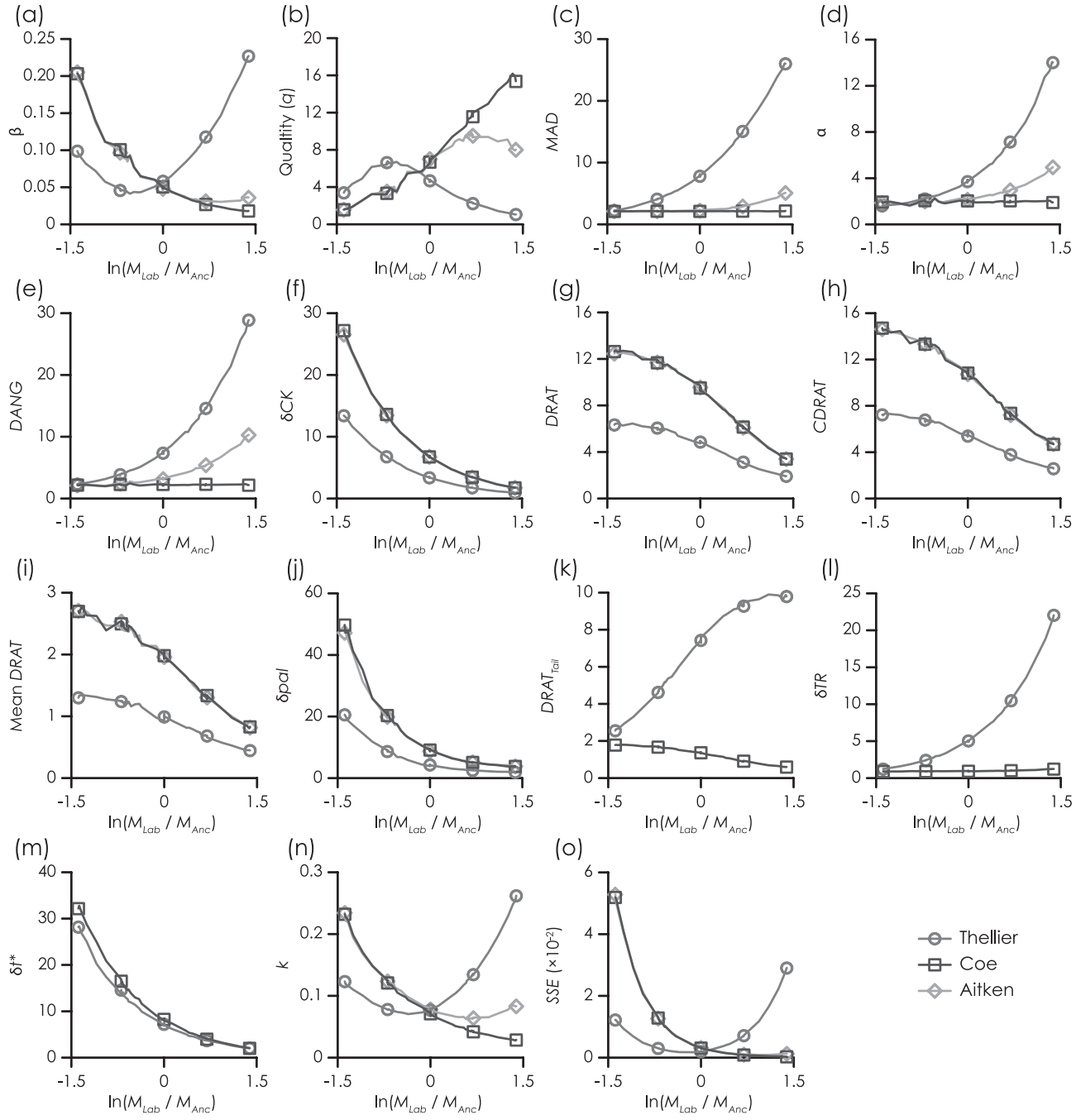

Figure 9. The paleointensity selection parameter 95 per cent thresholds from the non-linear TRM simulations as a function of $M_{\rm Lab}/M_{\rm Anc}$. (a)–(e) Arai plot and directional parameters. (f)–(j) pTRM check parameters. (k)–(m) pTRM tail check parameters. (n)–(o) Arai plot curvature parameters. Squares represent results from the Coe protocol, circles from the Thellier protocol and diamonds from the Aitken protocol. In all plots the symbols denote results from simulation with linear TRM. For clarity the IZZI protocol results, which coincide with the Coe results, have been omitted. No pTRM tail checks (k)–(m) were included in the Aitken simulations. For many parameters the Aitken and Coe results coincide. When $\ln(M_{\rm Lab}/M_{\rm Anc}) > 0$, $B_{\rm Lab} > B_{\rm Anc}$ and when $\ln(M_{\rm Lab}/M_{\rm Anc}) < 0$, $B_{\rm Lab} < B_{\rm Anc}$.

field, 95 per cent of TRM measurements are reproducibility to with ∼1 per cent of the first measurement. At 90 and 99 per cent saturation, the error of reproducibility increases, but reaches a maximum of only 1.2 and 99 per cent saturation for all applied fields. When rejecting anisotropy tensor or non-linearity measurements on the basis of potential thermal alteration the rejection thresholds should not be stricter than these values.

4 DISCUSSION

4.1 Comparison to real data

The number of studies that investigate the effects of anisotropic and non-linear TRM is limited. Selkin *et al.* (2000) investigated the effects of anisotropy on eight samples from the Stillwater Complex,

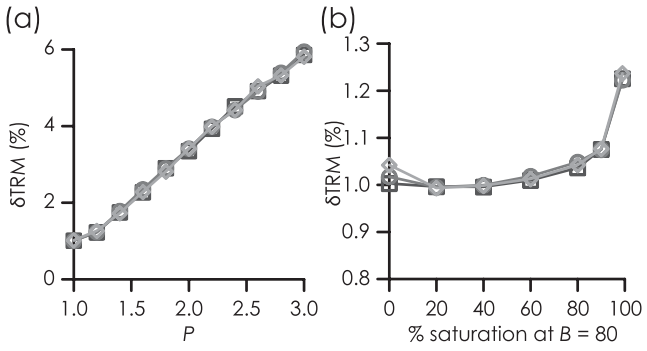

Figure 10. The repeatability of TRM acquisition for (a) the determination of the anisotropy tensor and (b) the determination of the degree of non-linearity. These values represent the 95th percentiles of the difference between two repeat TRM acquisition steps as a percentage of the first acquisition. Diamonds represent $B_{Lab} = 20\ \mu$T, circles represent $B_{Lab} = 40\ \mu$T and squares represent $B_{Lab} = 80\ \mu$T.

MT, USA. The average degree of anisotropy from these samples was ∼2.0 (min./max. = 1.79/2.16), which should yield a range of c values from 0.5 to 2. The measurements of Selkin *et al.* (2000), however, were restricted to a single plane and the measured c values range from ∼0.5 to ∼1.0. Each sample was imparted with a laboratory TRM to simulate an ancient TRM and paleointensity estimates were made using the Coe protocol. The experiments were performed using $B_{Anc} = B_{Lab} = 25\ \mu$T, therefore any non-linear TRM effects should be negligible. Considering the uncorrected results, none are accurate (within ∼10 per cent of the expected value), the deviation of the mean result is inaccurate (0.40), and scatter is 39.5 per cent of the mean. For the corrected results, five results (62.5 per cent) are inaccurate, the deviation of the mean is −0.04, and the scatter is 16.2 per cent.

To compare with modelled results, the Coe protocol experiment with $B_{Lab} = B_{Anc}$ and $P = 2.0$ is used. Only simulations where the 'measured' c values are between 0.5 and 1.0 are retained for the comparison (5736 simulated experiments). For the uncorrected results, 78.5 per cent are inaccurate, the mean deviation is inaccurate (0.30), and the scatter is 20.6 per cent. For the corrected results, 3.9 per cent are inaccurate, the mean deviation is 0.01, and the scatter is 5.7 per cent.

If the data and the simulations are normalized by the respective expected results, the distributions of paleointensities can be compared. For the uncorrected results the two-sample Kolmogorov-Smirnov (KS) test cannot reject the null hypothesis (at the 5 per cent significance level) that the real data from Selkin *et al.* (2000) are from the same population distribution as the model results ($p = 0.166$). The KS test, however, rejects the null hypothesis for the corrected results ($p = 0.048$). The Wilcoxon rank sum test (a non-parametric test for common medians) cannot reject the null hypothesis (5 per cent significance level) that the real data and model results have equal medians for both uncorrected and corrected results ($p = 0.534$ and $p = 0.103$, respectively). These analyses suggest that the simulated paleointensity results are comparable to real data. The averaging of anisotropy degree, small effects from other non-ideal factors (i.e. alteration and/or large grain sizes), or the disparity between the number of data available (eight samples compared with 5736 simulations) could explain the differences seen in these comparisons.

Shaar *et al.* (2010) investigated the effects of non-linear TRM on copper slag samples where the true paleointensity was known (the NRM was imparted in the laboratory). These samples, however, also suffered from anisotropic TRM. For the 18 samples investigated the maximum degree of anisotropy was $P = 1.48$ ($c \approx 0.68$–1.48), which can be adequately corrected for and therefore the results after anisotropy correction are considered for comparison with the non-linear TRM models. Shaar *et al.* performed paleointensity experiments using the IZZI protocol, but using four different B_{Lab}/B_{Anc} combinations with a maximum of five samples per field combination. For four samples the experiments used $B_{Anc} = B_{Lab} = 90\ \mu$T and no non-linear TRM correction was applied. Three of these four samples yield accurate results. TRM non-linearity was determined for the remaining 14 samples and the average percentage of saturation in an 80 μT field was ∼41.0 per cent (min./max. = 6.5/58.6 per cent). The maximum saturations in B_{Lab} or B_{Anc} were ∼32 and ∼64 per cent, respectively (maximum applied fields of 40 and 90 μT, respectively). For the paleointensity experiments performed on these 14 samples $B_{Anc} \neq B_{Lab}$.

If we consider the saturation plots in Figs 8(a)–(f) (the equivalent figure for the IZZI protocol is given in the Supporting Information and is near identical to Fig. 8). The saturation values from Shaar *et al.* (2010) are in a region that, before correction is expected to yield $\lesssim 60$ per cent inaccurate results (with most results < 40 per cent), absolute mean deviations < 0.2 (mostly < 0.1), and scatters of 3–6 per cent. The uncorrected results from Shaar *et al.* have 57 per cent inaccurate estimates, an absolute mean deviation of 0.07, and a scatter of 5.8 per cent. After non-linear TRM correction, 7 per cent of estimates are inaccurate, the absolute mean deviation is 0.05 and the scatter is 3.9 per cent. The corrected results from the model yield < 20 per cent inaccurate results (with most results < 10 per cent), absolute mean deviations of < 0.05, and scatters of 5–10 per cent. Although the number of real data available is limited, the model predictions compare well with the results from Shaar *et al.* (2010).

4.2 Implications for experiments and data selection

4.2.1 Anisotropic TRM

Paleointensity samples that suffer from remanence anisotropy can yield results that are highly inaccurate, but with a low scatter, irrespective of the paleointensity protocol used. Despite these poorly behaved results, the 95 per cent thresholds for commonly used selection parameters are less than typically used values. That is to say, ≥ 95 per cent of uncorrected results have parameter values that would not be detected by typical selection criteria.

The currently used methods to correct for anisotropic TRM (Veitch *et al.* 1984; Selkin *et al.* 2000) are effective. These methods can successfully reduce the proportions of inaccurate results and yield an accurate mean result. The effectiveness of the correction at minimizing inaccurate results depends on the angles that both B_{Anc} and B_{Lab} make with respect to the principal anisotropy axes, the degree of anisotropy, and the relative strengths of the two fields (discussed in Section 4.2.4). For the Coe protocol, $c \geq 1$ yields the lowest proportion of inaccurate results (Fig. 1a), which corresponds to B_{Lab} lying closer to the major anisotropy axis than B_{Anc}. For the Thellier protocol the opposite is true, the lowest proportion of inaccurate results occurs when B_{Anc} lies closer to the major anisotropy axis than B_{Lab} (i.e. $c \leq 1$). If the anisotropy tensor is approximately known before the paleointensity experiment (e.g. measured from a sister sample or from anisotropy of magnetic susceptibility), by simply inverting the sample z-axis (or whatever sample axis B_{Lab} is applied along) it is possible to orient B_{Lab} to within $\leq 90^\circ$ of any

specific anisotropy axis and hence approximately control the range of c. Such a procedure may help to increase the success rate of experiments particularly for samples with high degrees of anisotropy.

The degree of anisotropy influences the behaviour of the paleointensity results and the effectiveness of the correction diminishes at high degrees of anisotropy due to increasing uncertainties in the estimation of c (e.g. Fig. 3). For most practical purposes the increased inaccuracy is small and only occurs at extremely high degrees of anisotropy, which are uncommon in most materials. Nevertheless, extremely anisotropic materials are being considered and used for paleointensity determinations. Studying magnetite exsolution in single crystals of pyroxene, Feinberg *et al.* (2006) reported magnetization variations in a single plane that are equivalent to anisotropy degrees of ∼3.6. Similarly, Usui & Nakamura (2009) reported anisotropy degrees as high as 4.67 from magnetite exsolved within crystals of plagioclase. Studies that use such highly anisotropic samples may be affected by the reduced effectiveness of the anisotropy correction. In these cases the results should be complimented by samples with lower degrees of anisotropy in order to test the consistency of the results.

Paterson *et al.* (2012) noted that the main source of experimental noise in all Thellier-type paleointensity protocols came from re-orientation noise between measurement steps. The reduced effectiveness of the correction with increasing degree of anisotropy can, in part, be attributed to the same effect. The variation of $\chi_{TRM}\mathbf{B}$ for different degrees of anisotropy as a function of the angle between $\mathbf{B}$ and the major anisotropy axis, $\mathbf{V}_1$, is shown in Fig. 11a. The plots in Fig. 11 do not include any experimental noise. As $\mathbf{B}$ is rotated from $\mathbf{V}_1$ to $\mathbf{V}_3$ (i.e. 90°), $\chi_{TRM}\mathbf{B}$ decreases, but in a non-linear fashion. The gradient of this variation, shown in Fig. 11(b), gives an approximate indication of the effects of re-orientation noise on $\chi_{TRM}\mathbf{B}$. For example, for $P = 3.0$ with $\mathbf{B}$ at 60° to $\mathbf{V}_1$, a one degree variation in $\mathbf{B}$ (or the estimation of $\mathbf{B}$) will produce a ∼2 per cent error in $\chi_{TRM}\mathbf{B}$. The relative effect of re-orientation noise increases as $\mathbf{B}$ moves away from $\mathbf{V}_1$, but rapidly drops as $\mathbf{B}$ becomes parallel to $\mathbf{V}_3$. At all angles, however, higher degrees of anisotropy amplify the effect of re-orientation noise. This increased uncertainty in $\chi_{TRM}\mathbf{B}$ with increasing P is one of the factors that leads to increased variation of $\tilde{c}$ with P and hence an increased likelihood of inaccurate results. This re-orientation noise effect is also a factor in producing poorer parameter thresholds from high degrees of anisotropy. As noted by Paterson *et al.* (2012), however, methods that fix a sample throughout the entire paleointensity experiment or that utilize specialized sample holders should reduce this effect.

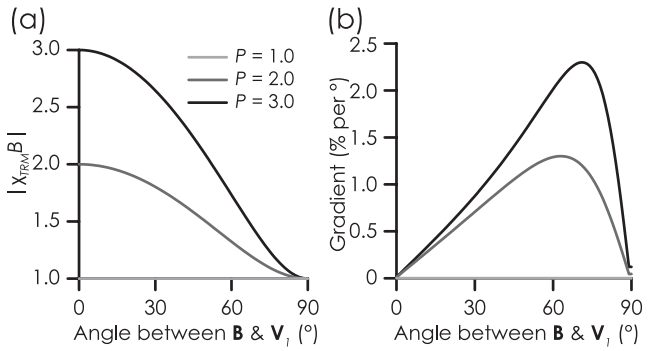

Figure 11. (a) The variation of $\chi_{TRM}\mathbf{B}$ as a function of the angle between $\mathbf{B}$ and the principal anisotropy axis, $\mathbf{V}_1$. (b) The gradient of $\chi_{TRM}\mathbf{B}$ as a function of angle. These plots are theoretical variations and contain no experimental noise.

4.2.2 *Non-linear TRM*

The effect of non-linear TRM on paleointensity descriptive statistics depends on the degree of saturation in B_{Anc} and B_{Lab}. If B_{Anc} and B_{Lab} are at less than ∼40–50 per cent saturation, or if $B_{\mathrm{Lab}} \approx B_{\mathrm{Anc}}$, uncorrected results are well behaved and yield accurate means, with a low scatter, and a low proportion of inaccurate results. This is the case for all Thellier-type protocols. As the degree of non-linearity increases, uncorrected results for all protocols have increasing proportions of inaccurate results ($\geq$ 80 per cent) and the mean results become highly inaccurate, but the scatters are low (typically $<$ 10 per cent). Although a high degree of self-consistency is necessary to assert the reliability of paleointensity results, alone it is an insufficient criterion.

After correction the results are generally improved. When B_{Anc} is close to a saturation field ($\gtrsim$ 85 per cent saturation), however, corrected results tend to behave poorly, with high proportions of inaccurate results, inaccurate means and higher scatters. For the Coe, Aitken and IZZI protocols the greatest improvement occurs when $B_{\mathrm{Lab}} > B_{\mathrm{Anc}}$. For the Thellier protocol, which in the absence of non-ideal behaviour yields optimal results when $B_{\mathrm{Lab}} \leq B_{\mathrm{Anc}}$, also has the largest improvement when $B_{\mathrm{Lab}} > B_{\mathrm{Anc}}$. As long as B_{Anc} is not close to saturation, however, $B_{\mathrm{Lab}} \leq B_{\mathrm{Anc}}$ still yields the best results for the Thellier protocol.

4.2.3 *Data selection*

Fundamentally, the behaviour of thermally/chemical stable SD paleointensity data, whether it is ideally behaved or influenced by anisotropic or non-linear TRM, is controlled by the ratio of M_{Lab}-to-M_{Anc}. As a consequence ideal, anisotropic or non-linear SD behaviour are essentially indistinguishable using typical paleointensity selection parameters. Only in the extreme, but generally rare, cases of highly anisotropic behaviour (i.e, $P \gtrsim 2$) will there be a noticeable effect. This effect, however, will not be detected by common selection criteria. The failure of anisotropic and non-linear TRM to manifest in commonly used selection parameters renders these non-ideal factors undetectable. In both cases the results can be highly self-consistent, but highly inaccurate. It is therefore a necessity of all paleointensity studies to test for anisotropic and non-linear TRM.

4.2.4 *The choice of laboratory field*

The ratio of B_{Lab}-to-B_{Anc}, or more generally the ratio of M_{Lab}-to-M_{Anc}, plays an important role in determining the overall behaviour of paleointensity data, even in the presence of expected levels of experimental noise alone (Paterson *et al.* 2012). The effect of B_{Lab} on the accuracy and scatter of the anisotropy corrected results can be understood in terms of $M_{\mathrm{Lab}}/M_{\mathrm{Anc}}$ and is identical to that of isotropic SD behaviour (Fig. 4). For non-linear TRM, the same underlying trend exists, but is partially obscured by results from high degrees of non-linearity (e.g. Figs 8g–i). For the Coe, Aitken and IZZI protocol $M_{\mathrm{Lab}} \gtrsim M_{\mathrm{Anc}}$, and $M_{\mathrm{Lab}} \lesssim M_{\mathrm{Anc}}$ for the Thellier protocol, yield the best results in terms of descriptive statistics. For the behaviour of the selection parameters, non-linear or anisotropic TRM has no or little effect and the 95 per cent thresholds vary as would be expected for linear/isotropic TRMs subject only to expected levels of noise. The choice of laboratory field should not only be such as to maximize the likelihood of obtaining accurate results, but also maximize the sensitivity of selection parameters to other non-ideal factors.

For the Coe and IZZI protocols the suggestion of Paterson *et al.* (2012), that $M_{Lab} \gtrsim M_{Anc}$, remains valid. For the Thellier protocol when $M_{Lab} < \frac{1}{2} M_{Anc}$ the Arai plot data become more scattered and more curved, and the quality factor decreases when compared with the results from $M_{Lab} \approx M_{Anc}$ (Figs 9a, b, n and o). For the Aitken protocol, when $M_{Lab} > 2M_{Anc}$ the quality factor decreases and directional parameter 95 per cent thresholds increase notably compared to the Coe and IZZI protocols. In practice M_{Lab}/M_{Anc} is only known after measurement. It can be noted, however, that for SD behaviour $\ln(M_{Lab}/M_{Anc}) \approx \ln(|\tilde{b}|)$, where $\tilde{b}$ is the measured slope on the Arai plot before applying any correction. Converting M_{Lab}/M_{Anc} to B_{Lab}/B_{Anc} depends on the intrinsic properties the sample. For anisotropic behaviour, M_{Lab}/M_{Anc} depends on the orientation of both B_{Anc} and B_{Lab} with respect to the anisotropy axes (characterized by c) as well as the relative strengths of both fields. For non-linear TRMs, M_{Lab}/M_{Anc} depends on the degree of saturation in both B_{Anc} and B_{Lab}. For simplicity, assuming isotropic and linear TRM acquisition, the following field laboratory ranges are recommended to minimize the effects of noise and increase sensitivity to other non-ideal factors.

(i) Thellier: $\frac{1}{2} B_{Anc} \lesssim B_{Lab} \lesssim B_{Anc}$
(ii) Coe/IZZI: $B_{Anc} \lesssim B_{Lab}$
(iii) Aitken: $B_{Anc} \lesssim B_{Lab} \lesssim 2B_{Anc}$

4.3 A combined anisotropic and non-linear TRM model

In natural samples anisotropic and non-linear TRM can result from the same mechanism (e.g. highly elongate SD grains) and often paleointensities suffer from both effects. There is a vast range of possible combinations that can be modelled, instead a situation analogous to the conditions reported by Ben-Yosef *et al.* (2009) is simulated. Ben-Yosef *et al.* (2009) used archeological materials (pottery, furnace fragments, tuyères and copper slag) to obtain Iron Age paleointensity estimates from the Southern Levant. They reported paleointensity estimates from ∼60 μT to as high as 110–130 μT. This suggests geomagnetic intensity variations of a factor two within less than a couple of hundred years, which are the most rapid temporal variations of the magnetic field known in the archeomagnetic record.

Ben-Yosef *et al.* (2009) measured the anisotropy tensors for 89 samples and the degree of anisotropy varied from 1.03 to 1.48 ($c \approx 0.68$–1.48). Although the range of c values for the 40 anisotropy corrected samples that were accepted by Ben-Yosef *et al.* (2009) range from 0.83 to 1.09, these are taken to represent a subsampling of the full range, which is controlled by P. Forty samples were also measured for non-linear TRM with the degree of saturation in a 80 μT field ranging from 0.7 to 49.7 per cent. The distribution of P values and degree of non-linearity can both be approximated by Weibull distributions, which are used as inputs into the model (further details of the distribution fitting are given in the Supporting Information). Ben-Yosef *et al.* used the IZZI protocol and used various laboratory fields (30, 50 and 70 μT). The simulations are restricted to only the IZZI protocol and model the extreme $B_{Anc} =$ 130 μT paleointensity results using two laboratory fields (30 and 70 μT).

Table 1 summarizes the descriptive statistics of the simulations before and after correction, and after applying the criteria used by Ben-Yosef *et al.* to select the data ($DRATS \leq 20$, $\beta \leq 0.1$, $f_{vds} \geq 0.7$, $MAD \leq 10$). In all cases, before and after correction and selection, the average results are accurate (within $\lesssim 4$ per cent) and the scatters are low (≤ 14.0 per cent). Before selection, the proportion of inaccurate results when $B_{Lab} = 30$ μT is high (≥ 38.7 per cent) even after correction for anisotropy and non-linear effects. After correction, the proportion of inaccurate results when $B_{Lab} = 70$ μT is much lower (∼13 per cent). For the $B_{Lab} = 30$ μT simulation the percentage of inaccurate results is greatly reduced after selection. This reduction also occurs for uncorrected data and is due to the choice of a large fraction (f_{vds}). Paterson *et al.* (2012) investigated the effects of fraction and showed that, for all protocols, larger fractions produced fewer inaccurate results. They also noted that the reduction trend was dependent on the ratio of B_{Lab}-to-B_{Anc}, with low ratios (i.e. $B_{Lab} < B_{Anc}$) yielding more inaccurate results than higher ratios for the same choice of fraction. For example, IZZI protocol results with a minimum fraction (f) of 0.35, the $B_{Lab} = B_{Anc}$ model yielded ∼1.5 per cent inaccurate results, whereas the $B_{Lab} = \frac{1}{2} B_{Anc}$ model yielded ∼7.4 per cent. At a minimum fraction of 0.70 this is reduce to ∼0.5 and ∼3.3 per cent, respectively. For the $B_{Lab} = 30$ μT simulation, B_{Lab} is more than four times smaller than B_{Anc} so it should be expected the proportion of inaccurate results should be high with a minimum fraction of 0.35 (> 30 per cent as a result of noise alone) and will be reduced after applying a stricter fraction criterion, which is the case here. Even in the extreme case of $B_{Lab} = 30$ μT and $B_{Anc} =$ 130 μT, the use of a large fraction greatly increases the likelihood of obtaining an accurate paleointensity from the samples studied by Ben-Yosef *et al.* (2009). The choice of a higher B_{Lab} in future experiments will allow the use of a smaller fraction, which will help to minimize unnecessary data rejection.

Table 1. Descriptive statistics from the simulation of the data from Ben-Yosef *et al.* (2009), which are affected by both anisotropic and non-linear TRM. In all cases $B_{Anc} = 130$ μT and the minimum fraction is 0.35.

	$B_{Lab} = 30$ μT		$B_{Lab} = 70$ μT	
	Raw	Corrected	Raw	Corrected
Unselected				
Mean deviation	−0.04	0.00	−0.04	−0.02
Scatter	13.1	14.0	8.8	7.0
Per cent inaccurate	45.8	38.7	29.0	13.1
Selected				
Per cent selected	21.3		28.5	
Mean deviation	−0.03	0.02	−0.03	−0.01
Scatter	9.6	8.8	7.8	5.4
Per cent inaccurate	6.8	5.2	6.2	1.6

From the perspective of SD anisotropic and non-linear TRM, the high paleointensities reported by Ben-Yosef *et al.* (2009) can be viewed as a reliable observation. However, the effects of other non-ideal factors and unwanted bias that may be introduced by the selection process need to be further investigated.

5 CONCLUSIONS

Paleointensity studies are infamous for having high-failure rates that result in the rejection of large amounts of data. The factors that can contribute the failure of paleointensity studies are diverse and complex. The development of new stochastic models, however, allow the statistical behaviour of non-ideal factors to be explored in great detail. Using such a model the effects of anisotropic and non-linear TRM have been investigated and the following conclusions have been reached:

(1) The effects of uncorrected anisotropic and non-linear TRM on the accuracy of paleointensity estimates can be highly

detrimental and can lead to large over- or underestimates depending on the ancient field, the choice of laboratory field and the degree of non-ideal behaviour. The scatter of the results, however, can be highly self-consistent.

(2) The correction methods for both anisotropic and non-linear paleointensities are effective. Their effects, however, are best understood as a function of magnetizations (i.e. M_{Lab}/M_{Anc}). When considered this way, the descriptive statistics of ideal, anisotropy corrected or non-linear TRM corrected SD paleointensities are identical.

(3) For non-linear TRMs the M_{Lab}/M_{Anc} relation breaks down for high degrees of non-linearity, but provided that B_{Anc} is not close to a saturation field, non-linear TRM can be successfully corrected for and yield precise and accurate results that are consistent with those from ideal SD samples.

(4) The effectiveness of the anisotropy correction diminishes with increasing anisotropy degree. Samples exhibiting extreme degrees of anisotropy ($P \gtrsim 3.0$) should only be accepted if their results are consistent with those from samples with lower degrees of anisotropy.

(5) Anisotropy correction methods that correct the paleointensity selection parameters should be avoided as they can have a detrimental effect on our assessment of the reliability of the data used. The original method of Veitch *et al.* (1984) should be the preferred approach.

(6) The behaviour of paleointensity selection parameters from anisotropic and non-linear TRM is near identical to that from ideal SD samples subject to expected levels of experimental noise. In effect, these non-ideal factors are undetectable by any commonly used selection parameter and as a result may be much more prevalent in paleointensity studies than has been previously recognized.

(7) The lack of distinct influence on common paleointensity parameters means that all paleointensity studies *must* test for anisotropic and non-linear TRM.

ACKNOWLEDGMENTS

GAP acknowledges funding from a Young International Scientist Fellowship from the Chinese Academy of Sciences (grant number 2009Y2BZ5) and NSFC Grant 41050110132. Some data were obtained from the MagIC Paleomagnetic database (http://earthref.org/MAGIC/) and the author is grateful for the efforts of the authors and the MagIC team for making the data available. Pierre Camps, Valera Shcherbakov and an anonymous reviewer are thanked for their careful reviews that improved this work. Eduard Petrovsky is thanked for his editorial handling.

REFERENCES

Aitken, M.J., Alcock, P.A., Bussell, G.D. & Shaw, C.J., 1981. Archaeomagnetic determination of the past geomagnetic intensity using ancient ceramics: allowance for anisotropy, *Archaeometry,* **23,** 53–64.

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,** 3–12.

Ben-Yosef, E., Ron, H., Tauxe, L., Agnon, A., Genevey, A., Levy, T.E., Avner, U. & Najjar, M., 2008. Application of copper slag in geomagnetic archaeointensity research, *J. geophys. Res.,* **113,** B08101, doi:10.1029/2007JB005235.

Ben-Yosef, E., Tauxe, L., Levy, T.E., Shaar, R., Ron, H. & Najjar, M., 2009. Geomagnetic intensity spike recorded in high resolution slag deposit in Southern Jordan, *Earth planet. Sci. Lett.,* **287,** 529–539.

Coe, R.S., 1967. Paleo-intensities of the Earth's magnetic field determined from Tertiary and Quaternary rocks, *J. geophys. Res.,* **72,** 3247–3262.

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.

Fanjat, G., Camps, P., Shcherbakov, V., Barou, F., Sougrati, M.T. & Perrin, M., 2012. Magnetic interactions at the origin of abnormal magnetic fabrics in lava flows: a case study from Kerguelen flood basalts, *Geophys. J. Int.,* **189,** 815–832.

Feinberg, J.M., Harrison, R.J., Kasama, T., Dunin-Borkowski, R.E., Scott, G.R. & Renne, P.R., 2006. Effects of internal mineral structures on magnetic remanence of silicate-hosted titanomagnetite inclusions: an electron holography study, *J. geophys. Res.,* **111,** B12S15, doi:10.1029/2006JB004498.

Genevey, A. & Gallet, Y., 2002. Intensity of the geomagnetic field in western Europe over the past 2000 years: new data from ancient French pottery, *J. geophys. Res.,* **107,** 2285, doi:10.1029/2001jb000701.

Leonhardt, R., Heunemann, C. & Krása, D., 2004. Analyzing absolute paleointensity determinations: acceptance criteria and the software ThellierTool4.0, *Geochem. Geophys. Geosyst.,* **5,** Q12016, doi:10.1029/2004GC000807.

Nagata, T., 1961. *Rock Magnetism,* Maruzen, Tokyo.

Nagata, T., Arai, Y. & Momose, K., 1963. Secular variation of the geomagnetic total force during the last 5,000 years, *J. geophys. Res.,* **68,** 5277–5281.

Néel, L., 1949. Théorie du traînage 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 behaviour during paleointensity experiments, *J. geophys. Res.,* **116,** B10104, doi: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,** Q05Z43, doi:10.1029/2012GC004046.

Prévot, M., Mankinen, E.A., Coe, R.S. & Grommé, C.S., 1985. The Steens Mountain (Oregon) geomagnetic polarity transition: 2. Field intensity variations and discussion of reversal models, *J. geophys. Res.,* **90,** 10 417–10 448.

Rogers, J., Fox, J.M.W. & Aitken, M.J., 1979. Magnetic anisotropy in ancient pottery, *Nature,* **277,** 644–646.

Selkin, P.A., Meurer, W.P., Newell, A.J., Gee, J.S. & Tauxe, L., 2000. The effect of remanence anisotropy on paleointensity estimates: a case study from the Archean Stillwater Complex, *Earth planet. Sci. Lett.,* **183,** 403–416.

Selkin, P.A., Gee, J.S. & Tauxe, L., 2007. Nonlinear thermoremanence acquisition and implications for paleointensity data, *Earth planet. Sci. Lett.,* **256,** 81–89.

Selkin, P.A., Gee, J.S., Meurer, W.P. & Hemming, S.R., 2008. Paleointensity record from the 2.7 Ga Stillwater Complex, Montana, *Geochem. Geophys. Geosyst.,* **9,** Q12023, doi:10.1029/2008gc001950.

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 planet. Sci. Lett.,* **290,** 201–213.

Tauxe, L., 2010. *Essentials of Paleomagnetism,* University of California Press, Berkeley.

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. Géophys.,* **15,** 285–376.

Usui, Y. & Nakamura, N., 2009. Nonlinear thermoremanence corrections for Thellier paleointensity experiments on single plagioclase crystals with exsolved magnetites: a case study for the cretaceous normal superchron, *Earth Planets Space,* **61,** 1327–1337.

Uyeda, S., Belshé, J.C., Fuller, M.D. & Girdler, R.W., 1963. Anisotropy of magnetic susceptibility of rocks and minerals, *J. geophys. Res.,* **68,** 279–291.

Veitch, R.J., Hedley, I.G. & Wagner, J.J., 1984. An investigation of the intensity of the geomagnetic-field during Roman times using magnetically anisotropic bricks and tiles, *Arch. Sci.,* **37,** 359–373.

Walton, D., 1979. Geomagnetic intensity in Athens between 2000 BC and AD 400, *Nature,* **277,** 643–644.

Yu, Y. & Dunlop, D.J., 2001. Paleointensity determination on the Late Precambrian Tudor Gabbro, Ontario, *J. geophys. Res.,* **106,** 26 331–26 343.

Yu, Y.J., Tauxe, L. & Genevey, A., 2004. Toward an optimal geomagnetic field intensity determination technique, *Geochem. Geophys. Geosyst.,* **5,** Q02H07, doi:10.1029/2003GC000630.

6 SUPPORTING INFORMATION

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

Figure S1. Behaviour of paleointensity selection parameter 95 per cent thresholds from the anisotropic TRM simulations of the Thellier protocol.

Figure S2. Behaviour of selected paleointensity parameters from the Thellier protocol simulations as a function of P.

Figure S3. The percentage of results that yield a paleointensity estimate that contains an imaginary component after correction for non-linear TRM.

Figure S4. Comparison of results, after non-linear TRM correction, from the Coe protocol (a)–(c) using all results, which include using the real component of complex results, and (d)–(e) using only real results (i.e. all complex results are excluded).

Figure S5. The paleointensity descriptive statistics for the effects of non-linear TRM at various combinations of B_{Anc} and B_{Lab} from the Thellier protocol.

Figure S6. The paleointensity descriptive statistics from the Thellier protocol simulations as colour-coded scatter plots of percentage saturation in B_{Lab} against percentage saturation in B_{Anc}. (a)–(c) Before and (d)–(f) after correction for the effects of non-linear TRM.

Figure S7. The paleointensity descriptive statistics from the Aitken protocol simulations as colour-coded scatter plots of percentage saturation in B_{Lab} against percentage saturation in B_{Anc}

Figure S8. The paleointensity descriptive statistics from the IZZI protocol simulations as colour-coded scatter plots of percentage saturation in B_{Lab} against percentage saturation in B_{Anc}.

Table S1. Definitions of the experimental paleointensity selection parameters used in this study (http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggt033/-/DC1)

Please note: Oxford university press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.