Are systematic differences between thermal and microwave Thellier-type palaeointensity estimates a consequence of multidomain bias in the thermal results?

Andrew John Biggin\textsuperscript{a,b,*}

\textsuperscript{a} Palaeomagnetic Laboratory Fort Hoofddijk, Universiteit Utrecht, Budapestlaan 17, 3584 CD Utrecht, The Netherlands
\textsuperscript{b} Geomagnetism Laboratory, School of Environmental Sciences, University of Liverpool, Oliver Lodge Labs, Oxford Street, Liverpool L69 7ZE, UK

\textbf{Article info}

Article history:
Received 24 July 2008
Received in revised form 25 February 2010
Accepted 10 March 2010

Edited by: Keke Zhang.

\textbf{Keywords:}
Palaeointensity
Microwave
Palaeomagnetism
Multidomain
Thermoremanence
Lavas

\textbf{Abstract}

A meta-analysis which directly compares 13 datasets of microwave and thermal Thellier palaeointensity measurements obtained from igneous rocks is reported. The results clearly indicate systematic differences in the measurements made using the two approaches: the thermal results tend to be higher (often by several tens of percent) and associated with slightly noisier Arai plots. Additionally, in nearly half of the cases considered, higher thermal palaeointensity estimates were reported to be calculated from the low-temperature portions of concave-up Arai plots while microwave measurements made on specimens from the same core sample or rock unit were derived from more linear Arai plots. The most plausible explanation for the bulk of these differences appears to be that non-ideal behaviour caused by multidomain, vortex state, and/or interacting ferrimagnetic grains is exaggerated in the results of the thermal experiments over their microwave counterparts. A detailed investigation of the individual studies concludes that such effects may be adversely affecting some of the measurements (thermal and/or microwave) in 10 of the 13 datasets considered. It is probable that these non-ideal effects are enhanced in the thermal results over their microwave counterparts, at least in part, because of the different Thellier-type protocols that are used in the two types of study. The findings of this study suggest that biasing of absolute palaeointensity estimates by multidomain effects may be much more prevalent than previously thought and that this effect might be responsible for certain discrepancies that have been observed in the GEOMAGIA50 database. They also clearly indicate that, in addition to taking the steps they already do to avoid biasing from secondary overprints and thermally-induced alteration, future absolute palaeointensity studies of any type must also take great care to avoid bias from these multidomain effects.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Measurements of the geomagnetic palaeointensity (PI) recorded in rocks can provide information about the behaviour of the Earth's magnetic field that is useful for understanding geodynamo theory, the core's thermochemical evolution, and the thermal interaction of the core with the mantle. PI measurements are also important, directly and indirectly, for the dating of young materials (e.g. Carlut et al., 2004). Being applicable to such a wide range of problems in the geosciences, it is crucially important that such measurements are as accurate as possible.

For this reason, the field of absolute palaeointensity determination is presently undergoing something of a revolution as large numbers of new techniques are proposed, tested, and adopted (Cottrell and Tarduno, 1999; Yamamoto et al., 2003; Le Goff and Gallet, 2004; Hoffman and Biggin, 2005; Dekkers and Böhnel, 2006; Biggin et al., 2007b; Dunlop, 2008). Prominent among these innovations is the approach where high-frequency (8–14 GHz) microwaves are used in place of conventional thermal energy to (de)magnetise the samples (Walton et al., 1996). The motivation behind the adoption of microwaves in palaeointensity experiments was to reduce the occurrence of magnetomineralogical alteration by reducing the temperature that the bulk sample is heated to as well as the duration of this heating. This thermally-induced alteration frequently causes palaeointensity experiments to fail (Kosterov and Prevot, 1998; Hill et al., 2002b) so the adoption of microwaves is expected to improve success rates. Comprehensive descriptions of the theory and practice of microwave palaeointensity determination are given elsewhere (e.g. Hill and Shaw, 1999) and will not be repeated here.

A considerable number of palaeointensity studies have been undertaken that allow the direct comparison of results produced...
using the thermal and microwave approaches on the same rock units or even the same core samples. In this study, the results of a meta-analysis (i.e. one that collates similar data from independent studies and analyses them en masse) will be used to show that there is a strong tendency for microwave measurements to be systematically lower than thermal measurements made on the same materials. This difference, observed in many studies and frequently reaching several tens of percent, produces the uncomfortable implication that at least one of these approaches is prone to frequent and significant bias. It is vital and urgent that the cause of the differences is identified so that the biasing factor(s) can be avoided in future studies. Thermal and microwave measurements made in future studies will then give the same, hopefully more accurate, results.

Section 2 provides essential background information and Section 3 describes the comparison between the thermal and microwave measurements. Section 4 presents an interpretation of the information obtained in Section 3 and identifies the potential candidates for causing microwave–thermal discrepancies. It concludes that ferrimagnetic grains within the igneous samples that do not fully obey Thellier’s laws of thermoremanence (loosely termed ‘multidomain grains’) may be affecting many of the studies considered here to a significant degree. Section 5 discusses some of the important implications that this conclusion has for published and future absolute palaeointensity studies.

2. Thellier-type palaeointensity experiments

2.1. Methodology and analysis

For weak field intensities such as that at the Earth’s surface, the intensity of a thermoremanent magnetisation (TRM) is linearly proportional to the imparting field (Thellier, 1938). The constant of proportionality relating these properties can, in practise, only be measured by observing the intensity of a fresh TRM imparted in the sample using a known field. Therefore all absolute palaeointensity experiments involve the measurement of a laboratory TRM alongside the natural remanent magnetisation (NRM) of the sample (itself presumed to be an ancient TRM).

All of the studies that will be considered here (both microwave and thermal) used a Thellier-type experimental protocol to measure the palaeointensity. Such protocols require that the natural remanent magnetisation (NRM) of each sample is removed in steps while a laboratory thermoremanence (TRM) or microwave equivalent (TRM) is simultaneously stepwise imparted.

In thermal experiments, demagnetisation and remagnetisation is achieved by heating and cooling the samples in a controlled magnetic field (zero field for demagnetisation) using a progressively higher peak temperature. In microwave experiments, the sample is instead subject to applications of high-frequency (8 or 14 GHz) microwaves, which increase progressively in power and/or duration.

The results of Thellier-type experiments are generally analysed using Arai plots (Nagata et al., 1963) where the NRM remaining in a sample is plotted (as the ordinate) against the laboratory TRM it has acquired at different stages of the experiment (on the abscissa). The slope of the line connecting the points on an Arai plot is then taken together with the intensity of the magnetic field applied to the samples in the experiment and used to obtain the palaeointensity measurement.

Thellier-type experiments are presently the most popular method of palaeointensity measurement because each point produced on an Arai plot is essentially an individual palaeointensity estimates (each from a different blocking temperature range in the sample). This is beneficial on two counts: firstly, the linearity of the line connecting the points acts as a built-in consistency check for the overall sample estimate and secondly, individual points that are demonstrably biased (generally by secondary magnetisations or magnetomineralogical alteration of the sample) can be excluded easily so that the line from which the estimate is produced is based on only those points that appear reliable.

Irrespective of whether microwave or thermal energy is used, there exist a number of different protocols by which a Thellier-type experiment can be performed. These can be separated into two broad groups: double-treatment and single-treatment methods. Double-treatment protocols require that the sample is given the same thermal or microwave treatment twice, one immediately after the other. The amount of NRM lost and laboratory TRM gained during these treatments is found from vector calculations involving the measurements made after each of the treatments. The individual protocols differ in how the magnetic field is set during the two treatments:

1. In Thellier’s original protocol (Thellier and Thellier, 1959), a field of the same non-zero intensity is applied in both treatments but its direction with respect to the sample is reversed.
2. In Coe’s protocol (Coe, 1967), the field is switched off in the first treatment (so that the sample is in a magnetic vacuum) and switched on in the second.
3. In Aitken’s protocol (Aitken et al., 1988), the field is switched on in the first treatment and off in the second.
4. In the IZZI protocol (Tauxe and Staudigel, 2004), the order of treatments alternates at each different temperature/powers step so that it is the same as that in Coe’s protocol at one step and the same as that in Aitken’s protocol at the next.

Single-treatment protocols are, in principle, able to produce the same information as double-treatment protocols in half as much time.

1. In the perpendicular protocol (Kono and Ueno, 1977) the sample is first stepwise demagnetised using thermal or microwave treatments and a zero field so that any magnetic overprint is removed. Subsequent treatments to higher temperature/powers steps are made with the field applied exactly perpendicular to the remaining NRM.
2. In the quasi-perpendicular (QP) protocol (Biggin et al., 2007b), magnetic overprints are first removed using a small thermal, microwave, or alternating-field demagnetisation treatment. The rest of the experiment then consists of microwave or thermal treatments made to progressively higher peak temperatures/powers with the field applied at an angle greater than 45° to the remaining NRM. After each in-field step, but before the measurement, a further treatment is applied, which is identical to the step that removed the overprint.

Each of these protocols has its own advantages and disadvantages and while some are much more widely used than others, no single protocol is overwhelmingly more popular than the rest. The perpendicular protocol has not been widely used in recent thermal studies but has been used in practically all microwave studies. This is partly for a historic reason—the original microwave system could not reproduce the power absorption of the cavity well enough to allow double-treatment protocols to be reliable. However, it is still routinely used in microwave experiments because it is faster than double-treatment protocols and because the field can very easily be aligned perpendicular to the NRM of a sample using the microwave apparatus (Hill and Shaw, 2007).

The single- and double-treatments outlined for each of the protocols described above are frequently supplemented by extra treatments designed to provide additional consistency checks. The
Table 1
Summary of acceptance criteria used in datasets analysed here (see Table 2 for details and note that the H60(M), E20, BAR, and SIM datasets were subject to no formal acceptance criteria).

<table>
<thead>
<tr>
<th>Code</th>
<th>Type</th>
<th>Study</th>
<th>a (%)</th>
<th>MAD (%)</th>
<th>n</th>
<th>f</th>
<th>g</th>
<th>β</th>
<th>q</th>
<th>pTRM check</th>
<th>pTRM tail check</th>
</tr>
</thead>
<tbody>
<tr>
<td>XIT</td>
<td>T, M</td>
<td>Böhnel et al. (2003)</td>
<td>&lt;15</td>
<td>&lt;15</td>
<td>≥4</td>
<td>≥0.15</td>
<td>&lt;0.10</td>
<td>≥1</td>
<td>DRAT &lt; 10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICE</td>
<td>T</td>
<td>Goguitchaichvili et al. (1999) Class 1</td>
<td>≥4</td>
<td>&lt;0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICE</td>
<td>M</td>
<td>Brown et al. (2006) Class 1*</td>
<td>≥4</td>
<td>≥0.30</td>
<td>≥0.5</td>
<td>&gt;0.97</td>
<td>n ≥ 3; ΔpTRM/NRM0 &lt; 5</td>
<td>n ≥ 3; ΔpTRM/NRM0 &lt; 5</td>
<td>≥&lt;0.10</td>
<td>DRAT &lt; 10%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
</tr>
<tr>
<td>SOH</td>
<td>T</td>
<td>Teaney et al. (2002)</td>
<td>‘Stable &amp; Primary’</td>
<td>≥4</td>
<td>≥0.30</td>
<td>≥0.5</td>
<td>&gt;0.99</td>
<td>≥5</td>
<td>DRAT &gt; 20%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
<td></td>
</tr>
<tr>
<td>SOH</td>
<td>M</td>
<td>Gratton et al. (2005a)</td>
<td>≥4</td>
<td>≥0.30</td>
<td>≥0.5</td>
<td>&gt;0.99</td>
<td>&gt;0.99</td>
<td>&gt;5</td>
<td>DRAT &gt; 20%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
<td></td>
</tr>
<tr>
<td>WAI</td>
<td>M</td>
<td>Hill et al. (2005, 2006)</td>
<td>≥4</td>
<td>≥0.15</td>
<td>&lt;0.10</td>
<td>&gt;1 &amp; &lt;0.40</td>
<td>≥0.15</td>
<td>&gt;1</td>
<td>DRAT &gt; 10%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
<td></td>
</tr>
<tr>
<td>WAI</td>
<td>T</td>
<td>Herrero-Bervera and Valet (2005)</td>
<td>‘Anchored to origin’</td>
<td>&gt;0.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWS</td>
<td>T, M</td>
<td>Garcia et al. (2006)</td>
<td>‘ChRM’</td>
<td>&lt;10</td>
<td>≥4</td>
<td>≥0.30</td>
<td>&lt;0.10</td>
<td>&gt;1</td>
<td>DRAT &gt; 10%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
<td></td>
</tr>
<tr>
<td>NIU</td>
<td>T</td>
<td>Shi et al. (2005)</td>
<td>‘Primary’</td>
<td>≥4</td>
<td>≥0.15</td>
<td>≥5</td>
<td>&gt;0.15</td>
<td>&gt;5</td>
<td>ΔpTRM bath &lt; 20%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
<td></td>
</tr>
<tr>
<td>H60</td>
<td>T</td>
<td>Yamamoto et al. (2003)</td>
<td>‘Primary’</td>
<td>≥4</td>
<td>≥0.15</td>
<td>≥5</td>
<td>&gt;0.15</td>
<td>&gt;5</td>
<td>ΔpTRM bath &lt; 20%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
<td></td>
</tr>
<tr>
<td>BGT</td>
<td>T</td>
<td>Presling et al. (2006)</td>
<td>‘Primary’</td>
<td>≥7</td>
<td>&gt;0.35</td>
<td>≥5</td>
<td>&gt;0.40</td>
<td>≥0.10</td>
<td>&gt;0.99</td>
<td>DRAT &gt; 20%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
</tr>
<tr>
<td>BGT</td>
<td>T</td>
<td>Presling et al. (2007) Class 1*</td>
<td>≥7</td>
<td>&gt;0.35</td>
<td>≥5</td>
<td>&gt;0.40</td>
<td>≥0.10</td>
<td>&gt;0.99</td>
<td>&gt;5</td>
<td>DRAT &gt; 20%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
</tr>
<tr>
<td>PRB</td>
<td>T, M</td>
<td>Thomas et al. (2004)</td>
<td>‘Primary’</td>
<td>&lt;10</td>
<td>≥4</td>
<td>≥0.30</td>
<td>&lt;0.10</td>
<td>&gt;1</td>
<td>DRAT &gt; 10%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
<td></td>
</tr>
<tr>
<td>JOR</td>
<td>T, M</td>
<td>Gratton et al. (2005b)</td>
<td>‘Primary’</td>
<td>&lt;10</td>
<td>≥4</td>
<td>≥0.30</td>
<td>&lt;0.10</td>
<td>&gt;1</td>
<td>DRAT &gt; 10%</td>
<td>ΔNRM/NRM0 &lt; 20%</td>
<td></td>
</tr>
</tbody>
</table>

pTRM check (Coe, 1967) is a repeat treatment (which may be carried out in or out of an applied field) designed to test for sample alteration. Specifically, it is designed to check whether the ability of the sample to acquire laboratory TRM during a particular treatment has changed subsequent to a higher temperature/power treatment being applied.

The pTRM tail check (Riisager and Riisager, 2001) is a treatment undertaken in zero field to check whether the NRM remaining in the sample has changed following an earlier treatment to the same peak temperature/power carried out in an applied field. This check is redundant in Aitken’s protocol which in any case, involves a demagnetisation treatment performed immediately after a remagnetisation treatment. While the pTRM tail check is generally used to detect high-temperature ‘tails’ of pTRM resulting from multidomain (MD) grains in the sample, it should be noted that it also serves to indicate certain types of magnetomineralogical alteration (McClelland and Briden, 1996).

PI measurements made using Thellier-type protocols are generally accompanied by a large number of parameters pertaining to the technical quality of the estimate (defined in numerous ways). The authors of a PI study generally define an arbitrary set of minimum criteria based on these parameters that all accepted data must meet. The criteria used by the studies considered in Sections 3 and 4 are listed in Table 1 and the parameters they refer to are briefly defined in the caption and elsewhere (e.g. Selkin and Tauxe, 2000).

Of particular relevance to the analysis that will be undertaken in Sections 3 and 4 are the parameters f and β. The f parameter is the fraction of the total NRM of the sample that is used to recover the palaeointensity estimate (Coe et al., 1978). The β parameter is a measure of the scatter of the points about the straight line on the Arai plot which is used to obtain the palaeointensity estimate (Selkin and Tauxe, 2000). Unlike the f parameter, β is not universally given with any set of published Thellier-type measurements.

General speaking, more reliable palaeointensity estimates have larger f and smaller β parameters associated with them. Another parameter that will be referred to throughout this study is the intensity error fraction (IEF; Biggin et al., 2007a) which is simply the deviation (positive high, negative low) of a palaeointensity estimate from the true value (if known) expressed as a percentage of the true value.

The checks and reliability criteria outlined above are deemed necessary because of the large potential for absolute PI measurements to be adversely affected by various processes. One of the most serious threats to the reliable and efficient measurement of palaeointensities is the possibility of thermochemical alteration of the sample occurring during the experiment itself. This will tend to alter the magnetic properties of the sample and therefore the slope of the Arai plot (which may increase or decrease depending on the alteration process). PI estimates are screened against bias due to sample alteration by checks on the linearity of the selected portion on the Arai plot as well as through the use of pTRM checks and pTRM tail checks. Other potential threats to PI estimation include the alteration of ferromagnetic grains below their Curie temperature in nature (e.g. Fabian, 2009), sample anisotropy (e.g. Tema, 2009), cooling rate differences between nature and the lab (e.g. Perrin, 1998), non-linear field dependence of TRM (Selkin et al., 2007), and multidomain grains. The last of these factors will be dealt with in detail in the next section because of the high degree of relevance it appears to have for the present study.

2.2. Non-ideal behaviour in Thellier-type palaeointensity experiments caused by multidomain grains: predictions from a phenomenological model

Samples that do not contain uniquely non-interacting single domain grains can violate Thellier’s laws of thermoremanence (Thellier, 1938) and behave non-ideally in both thermal (Levi, 1977) and microwave (Biggin et al., 2007b) Thellier-type paleointensity experiments. In this study, the abbreviation MD refers to all types of grains that exhibit this type of non-ideal behaviour. This probably includes vortex state grains (Shcherbakov and Shcherbakova, 2001) and interacting single domain and as well as genuinely MD grains.

This undesirable behaviour is unfortunately not rare. Numerous palaeointensity studies performed on young lavas have concluded...
that “multidomain grains” are responsible for biasing of the measurements (e.g. Calvo et al., 2002; Carlot and Kent, 2002; Biggin and Thomas, 2003; Böhnel et al., 2003; Coe et al., 2004; Chauvin et al., 2005; Biggin et al., 2007a; Michalk et al., 2008, 2010). Considerable effort has therefore been expended on further improving our understanding of the process of MD acquisition and demagnetisation so as to better and overcome the resulting problems in PI experiments (e.g. Fabian, 2001; Shcherbakov and Shcherbakova, 2001; Leonhardt et al., 2004; Biggin, 2006; Yu and Tauxe, 2006).

The non-ideal behaviour characterising MD grains in Thellier-type experiments have long been found to present itself most commonly in the form of ‘concave-up’ Arai plots. In such Arai plots, the points sag beneath the ideal line connecting the theoretical end-points representing a full natural TRM and a full laboratory-induced TRM, respectively. This effect is well known to become more pronounced as the average grain size of the sample in question increases (Levi, 1977). However, more recently, studies such as those given above have revealed that the degree to which MD grains affect an Arai plot and the exact nature of these effects also depends a great deal on the precise way in which the experiment was carried out.

This study will make use of a phenomenological model of MD TRM which was developed in an earlier study (Biggin, 2006) by the incorporation of first-order symmetry laws of pTRM (Biggin and Poidras, 2006) into a pre-existing phenomenological model (Fabian, 2001). This model has already been demonstrated to be a useful tool capable of accurately predicting qualitative behaviour in actual experiments performed on samples exhibiting MD behaviour (Biggin, 2006; Biggin et al., 2007b). It uses a characteristic function, $\chi$, to represent the (un)blocking behaviour of a sample in a thermal or microwave experiment. This function comprises a series of parameters, the most important of which to this study is $\lambda$, which defines the extent of the non-reciprocal components of $\text{pTRM}$ acquired by the sample (essentially its propensity to exhibit MD-like behaviour). More details about the model can be found elsewhere (Biggin, 2006) and will not be repeated here.

The phenomenological model makes no distinction between thermal and microwave energy but can help establish the effectiveness of different protocols used in either type of experiment. It will be used here to simulate different experimental routines on samples with identical magnetic properties so that their results can be compared. Given a series of simulations made using an identical $\chi$ function, any observed differences in the Arai plots is then uniquely a result of the following:

1. The experimental protocol used (see Section 2.1).
2. The angle ($\theta$) made between the laboratory field ($H_{lab}$) and the primary NRM of the sample (imported by field $H_{nrm}$).
3. The ratio ($H_{lab}/H_{nrm}$) of the intensities of these fields.
4. The number and spacing of points on the Arai plot (a function of both the temperature/power steps chosen for the simulation and the sample’s unblocking behaviour).
5. Whether and how pTRM and/or pTRM tail checks are used in the experiment.

Fig. 1 shows numerous predictions of the phenomenological model based on the same simulated sample undergoing Thellier-type PI experiments with different protocols and values of $\theta$ used. Only factors numbered 1 and 2 in the list above are varied in these simulations but the results are instructive nonetheless. The characteristic function (given in Fig. 1) was chosen to represent a sample with a ‘moderate’ degree of MD behaviour. For example, the shape of the Arai plot simulated using the Coe protocol with $\theta = \theta'$ is approximately in the middle of the range of shapes produced experimentally by Shcherbakov and Shcherbakova (2001). This same characteristic function will be used throughout this study except in a few (clearly identified) cases where the value of $\lambda$ will be increased or decreased, making the sample more or less MD-like, in order to highlight particular issues.

Some general observations made from Fig. 1 are given below. These are largely independent of factors 3–5 above and the simulated sample properties (providing these produce at least some MD behaviour) and have already been experimentally verified either directly or indirectly (Biggin, 2006; Biggin et al., 2007b).

(a) Single-treatment protocols produce results less affected by MD grains than do double-treatment protocols.
(b) For double-treatment protocols, MD effects become more pronounced as $\theta$ increases. This is especially the case for the Aitken protocol.
(c) The original Thellier protocol produces results which are, on average, more affected by MD effects than results produced by experiments employing the Coe protocol. Experiments employing the Aitken protocol can produce more or less non-ideal behaviour than the Coe and Original Thellier protocols depending on the value of $\theta$ used.
(d) Despite the fact that no sample alteration was simulated in these experiments (i.e. the characteristic function remained constant through all the simulations), the simulated results commonly exhibit pTRM check failure. The implication of this is that pTRM checks may fail in real experiments purely as a consequence of MD effects (and not just because of alteration as commonly assumed).

A further observation which can be made from Fig. 1 but which has not yet been experimentally verified is:

(e) The introduction of pTRM and pTRM checks into experiments performed using the perpendicular protocol may enhance the non-linearity of the Arai plot slightly and can also produce highly discrepant results in the checks themselves (e.g. pTRM checks are much larger than the original pTRM measurements in the “Perpendicular + checks” simulation).

Other relevant observations which have been made (Biggin, 2006) from the results of experiments simulated by this phenomenological model (but which are not illustrated in Fig. 1 and which have not yet been explicitly tested using empirical data) are:

(f) Variation of the ratio $H_{lab}/H_{nrm}$ in either direction about unity progressively increases the degree of non-ideal behaviour exhibited.
(g) Increasing the number of temperature steps used and/or modifying the temperatures so that points on the Arai plot become more clustered can exaggerate non-ideal behaviour.

3. A comparison of microwave and thermal palaeointensity measurements

The starting material for this study is a total of 20 published papers which allow 13 direct comparisons between microwave and thermal palaeointensity experiments to be made. The selected studies were those in which results from both types of experiments had been carried out on samples taken from the same rock unit (lava flow or intrusion) or from the same standard (1 in. diameter) palaeomagnetic drill core. In Table 2 they are listed together with their references, a summary of the findings of Sections 3 and 4, and a three character code which will be used throughout this paper to refer to them individually. The results of two recent studies (Hill et al., 2008; Zhu et al., 2008) describing experiments performed on samples from Inner Mongolia which would otherwise
Table 2
Summary of datasets used in this study. For detailed reviews, see supplementary information. The PI column gives the discrepancies as percentages of the microwave mean. The group number in the final column is explained in Section 4.1 of the text.

<table>
<thead>
<tr>
<th>Code</th>
<th>Dataset and references</th>
<th>Age</th>
<th>N (core samples)</th>
<th>N (rock units)</th>
<th>Microwave protocol(s)</th>
<th>Thermal protocol(s)</th>
<th>PI (μT)</th>
<th>β (microwave/thermal)</th>
<th>Success rates (microwave/thermal)</th>
<th>Concave-up Arai plots reported?</th>
<th>Notes</th>
<th>Likely cause of discrepancy</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>XIT</td>
<td>Xitle lava, Mexico (Böhnel et al., 1997, 2003)</td>
<td>1670 CE</td>
<td>4</td>
<td>1</td>
<td>Perpendicular and parallel Aitken ((H_{lab} = 40–70 \mu T))</td>
<td>Coe* ((H_{lab} = 50 \mu T))</td>
<td>58.3/67.5</td>
<td>+16%</td>
<td>0.022/0.052</td>
<td>89%/38%</td>
<td>√</td>
<td></td>
<td>MD 1</td>
</tr>
<tr>
<td>ICE</td>
<td>R3-N3 Iceland lavas (Goguitchaichvili et al., 1999; Brown et al., 2006)</td>
<td>ca. 2.1 Ma</td>
<td>20</td>
<td>10</td>
<td>Perpendicular ((H_{lab} = 8–70 \mu T))</td>
<td>Original Thellier+ ((H_{lab} = 40 \mu T))</td>
<td>15.4/26.6</td>
<td>+73%</td>
<td>0.042/0.050</td>
<td>31%/37%</td>
<td>√</td>
<td></td>
<td>MD 1</td>
</tr>
<tr>
<td>SOH</td>
<td>SOH-1 Hawaii lavas (Teanby et al., 2002; Gratton et al., 2005a)</td>
<td>0–45 ka</td>
<td>0</td>
<td>92</td>
<td>Perpendicular (+) and Coe ((H_{lab} = 8–60 \mu T))</td>
<td>Original Thellier+ ((H_{lab} = 40 \mu T))</td>
<td>25.1/33.5</td>
<td>+33%</td>
<td>0.027/0.031</td>
<td>50%/54%</td>
<td>√</td>
<td></td>
<td>MD 1</td>
</tr>
<tr>
<td>WAI</td>
<td>Waianae lavas, Hawaii (Hill et al., 2005, 2006; Herrero-Bervera and Valet, 2005)</td>
<td>3.3–3.6 Ma</td>
<td>17</td>
<td>12</td>
<td>Perpendicular (+) and Coe (+) ((H_{lab} = 5–60 \mu T))</td>
<td>Aitken+ ((H_{lab} = 45 \mu T))</td>
<td>23.4/37.4</td>
<td>+60%</td>
<td>0.033/0.078</td>
<td>68%/14%</td>
<td>√</td>
<td>Microwave–thermal results mostly coherent but some thermal results several times as high</td>
<td>MD 1</td>
</tr>
<tr>
<td>GWS</td>
<td>Great Whin Sill, UK (Garcia et al., 2006)</td>
<td>270–301 Ma</td>
<td>13</td>
<td>2</td>
<td>Perpendicular+ and Coe+ ((H_{lab} = 25 \mu T))</td>
<td>Coe+ ((H_{lab} = 25 \mu T))</td>
<td>9.2/14.2</td>
<td>+54%</td>
<td>0.037/0.061</td>
<td>60%/69%</td>
<td></td>
<td>Alt, MD 2</td>
<td></td>
</tr>
<tr>
<td>NIU</td>
<td>Niutoushan lava, Mongolia (Shi et al., 2005)</td>
<td>106 Ma</td>
<td>7</td>
<td>1</td>
<td>Perpendicular+ ((H_{lab} = 15 \mu T))</td>
<td>Coe+ ((H_{lab} = 30 \mu T))</td>
<td>15.6/23.9</td>
<td>+53%</td>
<td>0.073/0.112</td>
<td>77%</td>
<td></td>
<td>Standard deviation of thermal results ((8.0 \mu T)) +that for microwave ((3.2 \mu T))</td>
<td>Alt, MD 2</td>
</tr>
<tr>
<td>H60</td>
<td>1960 flow, Hawaii (Hill and Shaw, 2000; Yamamoto et al., 2003)</td>
<td>1960 CE</td>
<td>0</td>
<td>1</td>
<td>Perpendicular ((H_{lab} = 36 \mu T))</td>
<td>Coe+ ((H_{lab} = 30 \mu T))</td>
<td>33.9/49</td>
<td>+45%</td>
<td>0.030/0.038</td>
<td>99%/89%</td>
<td>√</td>
<td>Expected PI = 36 μT</td>
<td>FL, MD 2</td>
</tr>
<tr>
<td>E20</td>
<td>20th Century lavas, Mt Etna, Sicily (Biggin et al., 2007b)</td>
<td>1950–1983 CE</td>
<td>5</td>
<td>3</td>
<td>Various (+) ((H_{lab} = 25–60 \mu T))</td>
<td>Coe+ and QP ((H_{lab} = 50 \mu T))</td>
<td>38.4/40.0</td>
<td>+4%</td>
<td>0.032/0.035</td>
<td>100%/100%</td>
<td>√</td>
<td>Flows grouped by eruption year in Fig. 2. Expected PI = 43–44 μT</td>
<td>MD 3</td>
</tr>
<tr>
<td>Location</td>
<td>Age/Event</td>
<td>SOH Type</td>
<td>Parameters</td>
<td>Measurements</td>
<td>Success Rate</td>
<td>Comment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
<td>------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIG</td>
<td>Holocene surface flows, Big Island, Hawaii (Pressling et al., 2006, 2007)</td>
<td>0–20 ka</td>
<td>16 15 Perpendicular (+) (H_{lab} = 25–75 \mu T)</td>
<td>Original Thellier+ (H_{lab} = 40 \mu T)</td>
<td>63.2/57.9 −8%</td>
<td>0.023/0.039 16%/53%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAR</td>
<td>Barrington Tops lavas, Australia (Hill et al., 2002a)</td>
<td>53 Ma</td>
<td>0 0 Perpendicular and parallel Aitken (H_{lab} = 15 \mu T)</td>
<td>Coe+ (H_{lab} = 50 \mu T)</td>
<td>11/−</td>
<td>0.037/− 58%/0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRB</td>
<td>Peat’s Ridge Basalt, Australia (Thomas et al., 2004)</td>
<td>ca. 49 Ma</td>
<td>20 1 Perpendicular+ (H_{lab} = 5–30 \mu T)</td>
<td>Coe+ (H_{lab} = 25 \mu T)</td>
<td>27.4/27.1 −1%</td>
<td>0.044/0.054 82%/68%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIM</td>
<td>Simulated Experiments on Lava Samples (Hill et al., 2002b)</td>
<td>3</td>
<td>1 Perpendicular (H_{lab} = 50 \mu T)</td>
<td>Coe+ (H_{lab} = 50 \mu T)</td>
<td>53.2/48.0 −10(^c)</td>
<td>0.024/0.038 100%/100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOR</td>
<td>Jorulla volcano, Mexico (Gratton et al., 2005b)</td>
<td>1759–1774 CE</td>
<td>3</td>
<td>1 Perpendicular+ and Coe+ (H_{lab} = 45 \mu T)</td>
<td>Coe+ (H_{lab} = 30 \mu T)</td>
<td>38.6/51.6 +14(^c)</td>
<td>0.033/0.032 100%/100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **SOH samples were all from the same long drill core.**
- * symbol indicates checks were also applied (for only some experiments if in parantheses).
- c Both results were within 10% of the correct answer and within errors of one another.
- d MD = multidomain, Alt = sample alteration in the experiment, Fl = studies from different locations in the lava flow, and Ov = unremoved overprint in the microwave experiments.
- e MD effects were clearly present in both sets of results but the lack of discrepancy reflects the unique experimental conditions (see Section 4.4 and supplementary section S8).
- f The discrepancy in these case refers to the success rates rather than the PI measurements.
have qualified for inclusion in this study are not featured as they were published after the meta-analysis was already completed.

This section will describe a statistical meta-analysis of collated data taken from all of these studies allowing primary first-order differences to be observed. To make best use of the diverse data in the literature and to ensure that our findings are robust, these comparisons will be made at three levels. Where paired data from microwave and thermal experiments performed on specimens from the same palaeomagnetic core sample are available, these individual specimen data are used and a comparison at the ‘core sample’ level is made. Where data is available from one or more specimens from the same lava flow or intrusive unit measured using microwave and thermal experiments, these will be averaged and used for comparisons at the ‘rock unit’ level. Finally, these rock unit data will themselves be averaged to produce single microwave and thermal results for the study or group of studies from which they are derived. The final type of comparison will then be made at the study level. The entire dataset used in this analysis is provided in the online appendix and includes only data published directly in journals or as supplementary information. These had generally already been filtered according to variable sets of criteria (given in Table 1) which is a limitation of the meta-analysis (ideally, the raw data would be processed identically).

The number of data pairs at each of the core sample and rock unit levels from each study is given in Table 2. Note that since the SOH data were taken from a single deep drill core, these were not suitable for comparing at the core sample level. Also note that although the BAR dataset does not contain any thermal PI results (since none were accepted by Hill et al., 2002b), it was nonetheless included in this study because of the strong disparity of the microwave and thermal success rates which was of interest. An individual review of each of the sets of studies analysed here that ‘fleshes out’ the information given in Table 2 is provided in the online supplementary information.

Some form of statistical test is required to measure the significance of the difference between paired microwave and thermal data. For this we use the Wilcoxon signed-rank test (see e.g. the online textbook http://faculty.vassar.edu/lhowry/webtext.html) which is a nonparametric version of the correlated Student’s t-test. It estimates the probability (P) that the median difference between N pairs of data (where N ≥ 5) is zero. It will be used in this study to provide the likelihood that the use of the microwave approach on specimens from the same core sample/rock unit/suite of rock units results in a value that is systematically offset from that obtained with the thermal approach.

In addition to comparing the palaeointensity estimates produced by microwave and thermal experiments, the values of two associated parameters (f and β) will also be compared. These were chosen firstly because they reflect important characteristics of the associated palaeointensity measurement (the fraction of NRM and
the amount of Arai plot scatter, respectively; see Section 2.1) and secondly because they can be calculated from the details published with every palaeointensity measurement used in this study.

The top row of Fig. 2 compares the palaeointensity estimates produced by microwave and thermal experiments at the three different levels. While there is a reasonably good correlation in each of these, it can also be seen that there is a highly significant tendency for palaeointensity estimates produced by thermal experiments to be higher than those produced by microwave experiments. At the core sample level, the thermal results are higher than their microwave counterparts more than twice as often as the other way around. Taking the mean (median) of the difference between the two, the thermal results are, on average, 38% (22%) higher than the microwave results. At the rock unit level, the inclusion of the large SOH dataset makes the difference even more extreme: the thermally-derived mean estimate is higher than its microwave estimate in 88% of the cases and the mean (median) of the differences is 39% (33%). At the study level, the thermal palaeointensity estimate was higher than the microwave estimate in 9 of the 12 cases and though the mean (median) difference has fallen to 28% (22%), this is still significant at the 95% confidence level.

It is interesting and rather surprising to see that there is no correlation between the parameters measured in the thermal and microwave experiments (middle row of Fig. 2). There is also little evidence at the core sample and study level for any systematic difference between the two approaches. At the rock unit level, the difference is significant at the 99% confidence level but this is due solely to a single group of studies. If the SOH dataset (92 paired data) is excluded, a systematic difference is no longer visible.

At the core sample and rock unit levels, the correlation is also very poor between the parameters measured in the thermal and microwave experiments (bottom row of Fig. 2). The correlation at the study level is statistically significant but heavily dependent on a single study (NIU). However, at all three levels there is a clear and highly significant tendency for the parameter to be higher (that is the Arai plot more concave-up) in the results of the thermal experiments than in the microwave experiments. Note that the parameter is a measure of relative scatter and that its increased expression in the microwave results, therefore presents itself as an obvious candidate for causing at least some of the discrepancies outlined in the previous section.

MD effects in the thermal experiments, or conversely, their suppression in the microwave results, may in some cases stem from the fact that different selection criteria were applied to the two sets of results (Table 1). However, even in those studies where the same criteria were applied to both types of data (XIT, GWS, JOR, NIU, PRB), no recurring pattern is evident.

In summary, the findings that can be inferred from Fig. 2 are that thermal palaeointensity experiments tend to produce results which are 20–40% higher on average than those produced by microwave samples on the same materials. The average value of the parameter is approximately the same (~0.6) in the results of both types of experiments indicating that there is no systematic tendency to use a larger or smaller fraction of the NRM in producing the estimates. However, the technical quality of the estimates does differ in one important respect: the thermally-derived results tend to have slightly higher values, i.e. their points tend to be more scattered around the best-fit lines on the Arai plots. Other observed differences were in the protocols used to produce the measurements (dominantly perpendicular for the microwave experiments and almost exclusively double-treatment for the thermal experiments; Fig. 3) and in the shape of the Arai plots (reported to be more frequently concave-up in the thermal experiments).

4. Interpretation of the comparison results

4.1. Candidate explanations for the discrepancies

It is the conclusion of this study that MD effects (Section 2.2), exaggerated in the thermal experiments over their microwave counterparts, are very likely the single most dominant (though not unique) cause of the discrepancies outlined in the previous section.

Hysteresis data was measured from representative samples in all of the studies considered here and the vast majority of this data suggests a dominant PSD grain size or a mixture of SD and MD grains (i.e. $M_s/M_H$ ratio between 0.05 and 0.4). This is typical of samples taken from subaerial lavas or high-level intrusive rocks and is also typical of samples that have previously been argued to exhibit non-ideal MD behaviour in PI experiments (e.g. Carlut and Kent, 2002; Biggin and Thomas, 2003; Coe et al., 2004; Michalk et al., 2008). The hysteresis data associated with the studies in Table 2 will be discussed in more detail in Section 5.2, but are mentioned here to demonstrate that MD effects are not ruled out in any of the datasets by accompanying rock magnetic data.

MD effects tend to make Arai plots concave-up instead of linear (Section 2.2); if the low-temperature portion of this Arai plot is then used to produce the PI estimate, it will be too high. As discussed in Section 3 (and in greater detail in the supplementary information), concave-up Arai plots appear to be more common in the results of the thermal experiments than in their microwave counterparts. Furthermore, in such cases, the PI estimate was generally obtained from the low-temperature portion of the plot. The exaggeration of MD effects in the thermal experiments, or conversely, their suppression in the microwave results, therefore presents itself as an obvious candidate for causing at least some of the discrepancies observed here.

Palaeointensity results affected in this way might be expected not only to be higher, but also to be associated with higher $\beta$ and lower $f$ values (since shorter, only quasi-linear portions of the Arai plot are available). The first of these conjectures is generally true of
Fig. 2. One-to-one plots comparing parameters associated with microwave and thermal palaeointensity results at three different levels. The core sample level (left) compares the results from the same palaeomagnetic drill core (averaging multiple results if necessary). The rock unit level (middle) compares the averages of all results produced by each of the approaches for a given lava flow or intrusion. Finally, the study mean is an average of the means used at the rock unit level. \( N \) is the number of pairs of data, \( P \) is the probability of the median of the difference between the two types of data being equal to zero (i.e., no systematic discrepancy) calculated using the Wilcoxon signed-rank test. \( R^2 \) is the coefficient of determination. See Table 2 for study codes. All data in this figure is provided in online appendix. The percentages given next to the arrows indicate the proportion of data either side of the line defining equality (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

The thermal results in this study (Fig. 2) but the second is not. However, this second prediction presupposes that the \( f \) values obtained in the microwave experiments are not themselves suppressed by any factors and this may not be the case. Firstly, in experiments performed using the perpendicular protocol (as the majority of microwave experiments were), it is necessary to be entirely sure of having removed any secondary magnetisations using demagnetisation treatments prior to applying the laboratory field. This tends to result in an over-cautious approach whereby part of the low blocking power spectrum that might contribute to the PI measurement in
a double-treatment experiment is not used. Secondly, microwave PI experiments are frequently halted at some point when the NRM of the sample is only partially replaced with a laboratory TMRM. This is most commonly because further increases in the microwave power or application time are not noticeably affecting the remanence of the sample. However, it may also be because the sample has become detached from the holder or even because a melt-spot has appeared on the sample.

If the disagreement between the PI estimates made using the two types of experiments is dominantly caused by MD effects in the thermal experiments then, regardless of the $f$ value obtained in the microwave experiments, we would expect larger discrepancies to be associated with smaller $f$ values in the thermal experiments. This is because larger $f$ values would tend to be associated with Arai plots that displayed less curvature and also because PI estimates with large $f$ values would tend to average more of this curvature out.

Fig. 4 shows that this relationship clearly holds and indicates that, at both core sample and rock unit level, the largest disagreements between the microwave and thermal PI estimates tend to
occur when the thermal measurements are obtained from smaller portions of the Arai plot. This is to be expected if the discrepancies are a consequence of thermal PI estimates being based on quasi-linear low-temperature portions of Arai plots made concave-up by MD effects. It is surprising that significant discrepancies remain even for high f values (e.g. \( f > 0.70 \) in the SOH and WAI datasets) but this could be related to the fact that these f values are calculated presumably using the scalar intensity of the NRM (and are therefore prone to being too high) and not the vector difference sum as advocated by Tauxe and Staudigel (2004).

In addition to the observational evidence for its occurrence, there is an obvious mechanism for causing MD behaviour to be exaggerated in the thermal experiments. As already discussed, microwave studies dominantly used the single-treatment perpendicular protocol while thermal studies almost uniquely used double-treatment protocols. Fig. 1 indicates that on this basis alone,
if the samples in question exhibit MD behaviour, we should expect more concave-up Arai plots in the thermal than in the microwave results.

Before moving on and considering, case-by-case, how well this interpretation applies, some potential rival explanations will also be considered. In all, seven mechanisms presented themselves as possible alternative causes of the discrepancies and these will be discussed in turn. The list given below may not be exhaustive, but it does make some attempt to include all of the most obvious candidates and a few of the less obvious.

1. A full microwave-induced thermoremanence (TMRM) is intrinsically stronger/weaker than its conventional equivalent.
2. The different rates at which the samples are cooled in the thermal and microwave experiments cause them to acquire magnetisations of different intensities.
3. The small sample size in the microwave experiments (ca. 5 mm × 2 mm discs are used in place of standard 25.4 mm cylinders) causes magnetic anisotropy effects to be enhanced.
4. There is a systematic source of error introduced to the microwave data by the apparatus used to make the measurements.
5. Magnetomineralogical sample alteration is enhanced in the microwave experiments.
6. Magnetomineralogical sample alteration is enhanced in the thermal experiments.
7. The (un)blocking behaviour of a sample in a microwave experiment is not the same as in a conventional thermal experiment.

Theoretically, we would not expect point 1 to be valid: a TMRM should be identical to a TRM in every sense, except that the magnons cause the remagnetisation are generated directly by microwave photons instead of indirectly by phonons (Walton et al., 1992, 1993). Furthermore, there is empirical evidence from ceramic (Shaw et al., 1999) and lava (Hill et al., 2002a) samples given a full TRM in the laboratory, which supports an equivalence. In particular, Hill et al. (2002a; SIM in this study) performed thermal Thellier experiments on samples carrying a TMRM and microwave experiments on samples carrying a TRM and obtained the expected result in both cases.

The smaller size of the samples used in the microwave experiments implies a shorter cooling time relative to thermal experiments (point 2). Where cooling times have been measured after microwave treatments (Suttie et al., 2010), they were on the order of tens of seconds and, in general, we would not expect them to exceed a few minutes in the longest cases. This implies differences in the cooling rate between microwave and thermal experiments of up to two orders of magnitude. Could these differences be responsible for the discrepancies observed between their results? In single domain grains, the TRM intensity is proportional to the product of the grain size (in microns) and the saturation magnetic moment of the magnetic mineral (Perrin, 1998) and therefore seems incapable of explaining offsets of the magnitude observed here. The effect of cooling rate on TRM held in pseudo-single domain (PSD) grains is poorly known, but given that Shaw et al. (1999) and Hill et al. (2002a) produced indistinguishable remanences in samples with PSD hysteresis properties using both microwave and thermal magnetisation, this is also unlikely to be large.

Point 3 may be a significant factor in some studies, but cannot explain most of the discrepancies outlined in Table 2 (where the microwave results are lower than their thermal counterparts). If the direction of a sample’s NRM and/or the laboratory field in a microwave experiment is random then an increased level of magnetic anisotropy would tend to have a random biasing effect on the palaeointensity measured. If the NRM direction tended towards the easy axis of magnetisation (as would be expected) and the laboratory field was applied orthogonal to the NRM (as was the case in the majority of the microwave experiments), then a bias to high values would be expected—the opposite of what is observed.

Point 4 is worth considering, because while the thermal results discussed were measured at a variety of laboratories around the world (and would therefore be expected to have averaged any systematic errors), the microwave results were all produced at the University of Liverpool. However, a total of three different microwave systems have been developed at Liverpool and, since every component (magnetometer, field coils, cavity, etc.) as well as the overall configuration in each of these is unique, it seems highly unlikely that systematic errors will be the same in multiple systems. Results produced on the first two of these systems (the 8.2 GHz and the first 14 GHz system) are used in this study. The study of Hawaiian lavas by Gratton et al. (2005b, Section 3.3) used both of these systems and demonstrated that there was no observable difference in their results. This strongly suggests that the systematic error introduced by both systems is negligible.

Point 5 is unlikely to be a significant factor in the majority of cases. While microwave energy can sometimes become focussed on an irregularity on the surface of a sample causing local melting to occur, this is always made immediately apparent in the experimental results. In any case, the maximum size of a sample typically used in thermal experiments is 25.4 mm diameter, so this factor is unlikely to be significant.

Point 6 is therefore difficult to dismiss as a potential contributor to at least some of the discrepancies shown in Table 2. However, it also seems to be an unlikely candidate for the single most dominant cause of the systematic offsets. In the first instance, there is no known reason why enhanced alteration should almost invariably cause the measured palaeointensity to be biased high. Thermally-induced alteration might be expected, at least in some cases, to cause bias in the opposite sense. However, none of the thermal study means shown in Table 2 are more than 10% lower than their microwave counterparts. In the second instance, pTRM checks were employed in practically all of the thermal experiments and were positive (passing criteria outlined in Table 1) for the portion of the Arai plot used to produce the PI estimate. Finally, and most convincingly, it is very difficult to explain the common association of higher thermal palaeointensity measurements with low–temperature portions of concave-up Arai plots in terms of enhanced alteration in these thermal experiments. It would require both that the alteration began at low temperatures (<200 °C) in many of studies and that it somehow changed its sign (to make the sample more rather than less capable of acquiring pTRM) at higher temperatures. While such a pattern of alteration is not entirely implausible, it seems very unlikely that it would occur...
with sufficient frequency to account for the discrepancies observed in this study.

Point 7 is a subtler variant of point 1 that is, in this case, supported by some evidence. Hill et al. (2002a) tested the equivalence of unblocking behaviour in the two approaches by performing microwave demagnetisation on six samples containing two perpendicular thermal pTRMs. They observed that in five cases, the two pTRMs unblocked cleanly whereas in one sample, their unblocking power spectra overlapped so that the high-temperature pTRM was difficult to isolate. This last case illustrates a potential problem with the microwave approach when applied to some samples: namely, that all magnetic grains might not be heated equally in an experiment and that certain grains may be demagnetised after others through the conduction of heat from them. If these secondary-heated grains have lower blocking temperatures than the primary-heated ones, then the component structure of a sample’s remanence as revealed by microwave demagnetisation will differ from that revealed by thermal demagnetisation. This process would tend to smear components of magnetisation out across the unblocking power spectrum in the microwave experiments so that their Zijderveld and Arai plots would be curved (or even linear if the components entirely overlapped) as opposed to comprising two (or more) distinct straight lines in the thermal experiments. This, in itself, is consistent with the greater occurrence of concave-up Arai plots in the results of thermal experiments. However, this explanation would require that the first of these two slopes in the thermal Arai plot correspond to a magnetic overprint. This seems highly unlikely given that the associated Zijderveld plot is usually reported to be univectorial, that the first slope often extends to high temperatures (>450 °C) regardless of the rock’s age, and that it is generally this first slope which is used to produce the palaeointensity measurement. We therefore expect that point 7 is the cause of the discrepancy between microwave and thermal measurements in only a single study considered here (JOR, see Section 4.5).

The factors numbered 1–4 in the list above appear, alone or in combination, entirely unable to explain any of the large systematic discrepancies described in Table 2 and will therefore not be considered in the discussion to follow in the rest of Section 4. Factors 5–7 are worth considering in certain cases although they are also unable to explain the most common type of discrepancy whereby the thermally-derived results are higher and associated with noisier, frequently concave-up Arai plots. It is obviously impossible to rule out some other, as yet unidentified, source of bias but for the present time, the only identified candidate remaining to explain the broad differences is the enhancement of MD effects in the thermal experiments. Nevertheless, the careful study-by-study analysis that will be described below reveals the situation to be more complex than might be expected from the discussion so far. Specifically, some of the alternative explanations identified above cannot be ruled out in a number of cases (indeed, in one highly discrepant dataset, an alternative is favoured). Conversely, in other cases it appears that MD effects have a significant impact on the results of the PI experiments while not causing any systematic disagreement in the PI estimates.

To formalise this complicated situation, four groups are introduced and each of the sets of studies are uniquely assigned to one of them (as shown in Table 2). Group 1 datasets contain a significant microwave–thermal PI estimate discrepancy that can be reasonably attributed to the process of MD effects being enhanced in the thermal experiments, but not to any of the other processes listed earlier in this section. Group 2 datasets contain significant microwave–thermal discrepancies that could potentially be explained (partly or wholly) by enhanced MD effects in the thermal experiments, but which could also be explained plausibly by other identified processes. Group 3 datasets are those which, for very different reasons, do not contain significant microwave–thermal PI estimates, but which have nonetheless produced results that are potentially affected to some degree by MD effects. Finally, Group 4 datasets display no evidence for MD effects having any significant impact on either their thermal or microwave results. Each of the 13 datasets will be discussed briefly in the next four sections and in more depth in supplementary information.

4.2. Group 1 datasets (XIT, ICE, SOH, WAI)

In all four of the original articles reporting microwave results associated with the Group 1 datasets (Böhnel et al., 2003; Gratton et al., 2005b; Brown et al., 2006; Hill et al., 2006), it was explicitly reported that higher PI estimates in the thermal experiments tended to be associated with low-temperature portions of concave-up (or ‘two-slope’ or ‘curvilinear’) Arai plots while samples from the same rock unit or core sample produced more linear Arai plots in the microwave experiments. Consequently explanations numbered 1–7 are relatively implausible for reasons already given in Section 4.1. Furthermore, their discrepancies can be qualitatively reproduced by the phenomenological model outlined in Section 2.2.

In the case of XIT (Fig. 5; supplementary section S1) and ICE (supplementary section S2), the choice of protocols in the microwave and thermal experiments can trivially explain the discrepancy providing that grains exhibiting MD behaviour are present in these samples. This is because, as Fig. 1 indicates, the protocols used in the thermal experiments (Coe and original Thellier, respectively) have a tendency to produce more concave Arai plots than the perpendicular protocol, which was uniquely used in the microwave experiments. The situation is not so simple for the SOH and WAI datasets because these microwave studies, while dominantly using the perpendicular protocol, also used the Coe protocol for some experiments. If MD effects were responsible for the concavity of the Arai plots produced in the thermal experiments, then why were they not reported in the results of the double-treatment microwave experiments?

The possibility that microwave experiments may be intrinsically (i.e. regardless of protocol) less prone to producing MD effects than those from thermal experiments cannot be ruled out and will be returned to later. Nonetheless, this explanation may not be required here in any case because of a number of other factors acting alongside the simple issue of single versus double-treatment protocols.

In the case of the SOH dataset (supplementary section S3), these other factors were related to the specific design of the experiment and the different users’ preference in analysing the data. Fig. 6 presents the results of model simulations produced assuming a sample with identical properties subject to three different PI experiments which were designed to mimic those used to produce the SOH dataset. The model’s output shows good qualitative agreement with the example taken from Gratton et al. (2005b) and demonstrates that the choice of protocols, temperature steps, and preference for line-fitting can cause a sample that gives similar results using Coe’s and the perpendicular protocol to produce a distinctive two-slope Arai plot in an experiment using Thellier’s original protocol.

In addition to questioning the absence of two-slope Arai plots in the double-treatment microwave experiments, Gratton et al. also doubted MD effects as the cause of the SOH discrepancies on the grounds that the high-temperature slope in the thermal results often gave a similar PI estimate to that obtained from the single slope measured in the microwave experiment (and not a significantly lower estimate as might be expected). A future study is planned in which the original thermal data will be analysed in detail to evaluate the extent to which this observation discriminates against MD effects. However, for the moment, it is worth noting two potentially mitigating factors that are discussed further.
Fig. 6. (a–c) An example of experimentally-derived Arai plots produced by specimens from the same rock unit measured using different protocols (microwave perpendicular, microwave Coe, and thermal original Thellier, respectively) in the SOH dataset (plots taken from Fig. 9 of Gratton et al., 2005b). Gratton et al. reported that the thermal experiments produced concave-up Arai plots much more frequently than the microwave experiments. (d–f) Simulated Arai plots (produced using the same characteristic function \( \chi \) as Fig. 1) demonstrating the capacity of MD effects (in conjunction with the choice of protocols and the temperature/power steps) to explain some of the differences in these results.

in supplementary section S3. The first is that we cannot rule out that alteration also affects the high-temperature portions of some of the Arai plots produced in the thermal experiments (which could act to bring the high-temperature PI estimates closer to, or further away from, the microwave estimates in certain cases). The second is that, in any case, the high-temperature portions of purely concave-up Arai plots do not always produce PI estimates as low as might be expected from intuition (e.g. in Fig. 6f, the degree of overestimation in the low-temperature slope is much greater than the degree of underestimation in the high-temperature slope). Therefore, while this argument certainly casts some doubt over whether MD effects are the sole (or even primary) cause of the observed discrepancies, it is not considered sufficient by this study to reject the possibility altogether.

In the case of the WAI dataset (supplementary section S4), the few very large discrepancies observed between the microwave and thermal results (examples given in Fig. 7) are expected to have occurred when the value of \( \theta \) approached 180° in the thermal experiments. This is a ‘danger point’ where the Aitken protocol (which was used in these thermal experiments) is most prone to producing non-ideal behaviour but where the pTRM tails’ effects on the samples’ NRM cannot be detected. Under this assumption, the phenomenological model is able to simulate experiments which produce Arai plots and NRM decay plots (Fig. 8) in good qualitative agreement with those shown in Fig. 7.

Unfortunately, it has not been possible to obtain the original data and test the prediction that \( \theta \) was close to 180° in the thermal results in Fig. 7a and b. However, there are other lines of evidence that support the presence of MD effects in the WAI dataset. Firstly, MD effects provide a neat explanation for a combination of characteristics observed in the WAI dataset: a low success rate in the thermal experiments, a dominant coherency between the thermal and microwave results, and the few large discrepancies between them (see S4 for details). Secondly, the microwave perpendicular
results had enhanced Arai plot linearity relative to the microwave Coe results as well as a higher success rate; this also agrees with the broad predictions made in Fig. 1.

Finally, there is an additional line of evidence for MD effects in the WAI, ICE, and SOH datasets that has not been discussed yet. The microwave studies associated with these all used pTRM and pTRM tail checks in conjunction with some or all of their perpendicular protocol experiments. As discussed in Section 2.2, the phenomenological model predicts that modest MD effects, while not causing severe non-linearity in the Arai plots themselves, may cause these checks to fail dramatically. Poorly behaved pTRM and pTRM tail checks are a characteristic of some of these results (Table 1 shows that Gratton et al. (2005a,b) and Brown et al. (2006) relaxed their appropriate criteria to allow for this. Given the generally improved linearity of the microwave results over their thermal counterparts, sample alteration is an unlikely cause of this behaviour. Poor power reproducibility can also effectively be ruled out; this would tend to make the pTRM checks both higher and lower than the reference pTRMs to which they are being compared and yet there is strong evidence in the WAI, ICE and SOH datasets (see sections S2, S3, and S4 in the SI) that the tendency is for them to be higher more often than lower. The phenomenological model predicts that pTRM checks may be significantly higher than their reference values in a perpendicular experiment performed on a sample exhibiting MD effects (but no alteration; see Fig. 1 and Fig. S1). It therefore seems likely that MD effects are, in part, responsible for the check discrepancies observed in the perpendicular experiments associated with these datasets.

4.3. Group 2 datasets: GWS, NIU, H60

These are studies that produced significant microwave-thermal discrepancies that could potentially be explained (partly or wholly) by enhanced MD effects in the thermal experiments but for which other plausible explanations exist too.
The mean palaeointensities of the thermal results for the two intrusions comprising the GWS study (supplementary section S5) were 48% and 59% larger than the corresponding means from the microwave results. In absolute terms, this translates to only a relatively small (~5 μT) difference, but this is nonetheless highly significant on the basis of individual sample comparisons. There is some indirect evidence that high thermal results may have been associated with the low-temperature portions of concave-up Arai plots. However, it has not been possible to establish this and therefore some other explanation (of which enhanced alteration in the thermal experiments is the most obvious) cannot be ruled out as a possible cause of these microwave–thermal discrepancies.

Thermal palaeointensity results derived from the single lava flow analysed in the NIU study (supplementary section S6) were considerably more variable, lower quality, and higher on average than those derived using the microwave approach. In this case, there is no evidence that the thermal results were associated with concave-up plots sensu stricto and, again, enhanced alteration in the thermal experiments cannot be ruled out as the potential cause. However, as discussed in supplementary section S6 and illustrated in Fig. 9, the phenomenological model can duplicate the relative behaviour of the two types of experiments very well, providing the assumption is made that \( \theta \) was not close to zero (something that it has unfortunately not been possible to check with the original data). MD effects, amplified in the thermal experiments through the choice of the protocol, \( H_{lab} \) and the specific temperature steps, therefore provide a likely explanation for the observed discrepancies in this study.

The H60 comparison (supplementary section S7) is made difficult because the samples used for microwave study were taken from a completely different part of the flow than those studied using the thermal approach. This problem is compounded by evidence that this flow is spatially highly variable in terms of its magnetic behaviour (Hill and Shaw, 2000). Evidence for MD effects takes the form of very high thermal Coe protocol results that are clearly associated with low-temperature portions of concave-up
Arai plots (Yamamoto et al., 2003). However, this is countered by evidence that some of the microwave perpendicular experiments performed by Hill and Shaw (2000) also produced concave-up Arai plots (which is not predicted by the phenomenological model). Additionally, low-temperature demagnetisation treatments performed by Yamamoto et al. (2003) had little effect on these samples which would not be the case if they were dominated by truly multidomain grains.

As discussed in supplementary section S7, the balance of evidence neither confirms nor rejects MD behaviour (potentially caused by interacting SD rather than truly MD grains) as a cause of the non-ideal behaviour. It is also quite plausible that some other factor (e.g. primary TCRM acquisition; Yamamoto, 2006) is responsible for the non-ideal Arai plots produced by both sets of experiments and that these occurred more commonly in the thermal results simply because these samples happened to have been taken from a part of the flow that was more prone to it.

4.4. Group 3 datasets: E20, BIG, BAR

These are studies that, for very different reasons, do not produce significant microwave–thermal discrepancies in terms of their PI estimates, but which have produced results that are nonetheless potentially affected to some degree by MD effects.

The E20 dataset (supplementary section S8) may, to some extent, be considered as the ‘exception that proves the rule’ in terms of MD effects being responsible for discrepancies between thermal and microwave PI measurements. There is clear evidence that the remanence in many of these samples is dominantly held in MD (and interacting SD) grains (Biggin et al., 2007a) and yet there is not a significant difference in the average results produced by the microwave and thermal approaches (Table 2). This seemingly anomalous outcome is very likely a result of another anomaly pertaining to E20: that it is the only study in which the thermal experiments were dominantly performed using a single-treatment protocol (Fig. 3). In fact, as discussed in supplementary section S8 and in Biggin et al. (2007a,b), the differences between the individual results of the majority of thermal and microwave experiments can be readily explained in terms of MD effects developed to different extents because of the choice of specific Thellier-type protocol. Fig. 10a shows that double-treatment (Coe) protocol applied in the thermal experiments tended to produce higher PIs than the single-treatment (QP) protocol on the same samples which is as expected from MD behaviour. Similarly, Fig. 11 indicates that samples treated using the same protocols in both thermal and microwave experiments tended to display similar (in some cases non-ideal) behaviour.

In the BIG dataset (supplementary section S9), MD behaviour was probably not manifested to a sufficient degree to cause any significant bias to either the thermal or microwave results. Nonetheless, the failure rate was much higher in the microwave study and this was largely due to failing pTRM and pTRM tail checks, which were blamed on poor microwave power reproducibility by Pressling et al. (2007). Similar to three of the Group 1 datasets (SOH, ICE, and WAI, see Section 4.2), however, this explanation seems incompatible with the observation that the sign of the pTRM check discrepancy (i.e. the number of times that they were higher and lower than their original pTRM measurement counterpart) was heavily biased (184–57) to positive values. The phenomenological model predicts that positive discrepancies will be prevalent and, what is more, that mild MD effects can have a strongly adverse effect on the checks associated with a perpendicular experiment while not simultaneously causing significant bias in the PI result. It also predicts that the original Thellier protocol performed on the same sample can largely escape all of these negative effects (Fig. 12). On this basis, it seems likely that MD effects are primarily responsible for the high check failure observed by Pressling et al. (2007).

In the BAR study (supplementary section S10), Hill et al. (2002b) explained their much higher success rate in the microwave experiments (58% vs. 0% for the thermal results) in terms of the thermal experiments causing more alteration to the samples. In S10, the case for a possible contribution of MD effects to this discrepancy is also made. This is largely based on the very different apparent values of $H_{lab}/H_{nrm}$ used in the two different types of experiment. If even a small amount of the remanence in these samples was held in MD grains, then the very high value of this ratio (4.54 if the microwave results are taken to be correct) would be expected to enhance the non-linearity of the resulting Arai plots dramatically (Fig. 13).

4.5. Group 4 datasets: JOR, PRB, SIM

These are studies which produced both thermal and microwave results that displayed no evidence of MD effects having any significant impact on them.

In the PRB and SIM studies (supplementary sections S11 and S12), there were no significant offsets between the thermal and microwave results and no indications of non-ideal behaviour due to
MD effects in the Arai plots. In the JOR study (supplementary section S13), the familiar pattern of the mean thermal PI results being significantly higher than its microwave counterpart is again evident. However, in this case, the higher thermal results tend to be associated with the high-temperature portions of convex-up shaped Arai plots (as opposed to the more normal case of the low-temperature portions of concave-up Arai plots) and the balance of evidence seems to suggest that some of the microwave PI results may be biased to low values by the influence of an unremoved magnetic overprint.

4.6. Summary of interpretation

In the previous sections, MD effects have been argued to have been potentially impacting on the results of 10 of the 13 studies in Table 2 (all except Group 4). In seven of these (Groups 1 and...
2), MD effects have been argued to have potentially contributed to significant discrepancies in the PI estimates produced by thermal experiments. In four of these (Group 1), no other plausible explanation for the observed microwave–thermal PI estimates appears to exist. In another single study (E20) MD effects are clearly at play and would have caused a significant microwave–thermal discrepancy if the standard practice of employing dominantly double-(single-)treatment protocols in the thermal (microwave) experiments had been followed.

In summary (Table 2), the results of five of the 13 studies were almost certainly significantly affected by MD effects and the results of another five may have been. This is a highly significant finding whose implications will be discussed in the next section. Other potential sources of microwave–thermal discrepancies (in PI estimates, success rates, etc.) were identified as enhanced alteration in the thermal experiments (maximum three studies), an unremoved overprint magnetisation in the microwave results (one study), and the sampling location within a lava flow (one study).

Why should MD effects be exaggerated in the thermal experiments over their microwave counterparts performed on similar samples? The most obvious answer is that the thermal experiments tended to be performed using protocols that intrinsically promote this behaviour. Or rather, the microwave studies chose predominantly to use a protocol (perpendicular) which, by good fortune, tended to suppress this non-ideal behaviour. This hypothesis received strong support from the results of the study which best allows it to be tested (E20; see Section 4.4), however, it is probably not the complete story in all cases. For example, in the
SOH study, the choice of protocol alone was insufficient to explain the discrepancy. At the very least, the concentration of points on the Arai plot and the natural preference of the individual researcher in terms of which points they choose for the best-fit line are also likely influencing factors in this and possibly other cases.

The most effective test of the degree to which the choice of protocol alone causes the microwave–thermal discrepancy would be to simply compare results produced using the same protocol in both the microwave and thermal experiments. The Coe protocol is the only one used to a sufficient degree in both types of experiment to allow some limited statistical comparison to be made. Fig. 10b indicates that at the sample level (ten comparisons from four datasets), there is no evidence for a systematic difference between results produced using the thermal and microwave approaches and the Coe protocol. At the unit level, (seven comparisons from four studies, Fig. 10c) some discrepancy does appear suggesting that the thermal results tend to be larger than their microwave counterparts. A problem with both of these comparisons is that, in two of the four studies providing the data (JOR and PRB), MD effects are not thought to be biasing these data in any case (see Section 4.5). This fact, coupled with the small number of data available makes it difficult to draw conclusions from this particular analysis.

**Fig. 11.** Experimental data produced from Etna lava samples by the study of Biggin et al. (2007b). Thermal and microwave measurements produced from the same core samples are displayed on the same Arai plots. Data were normalised by their relative values of $H_{lab}$ so that an equal palaeointensity would produce parallel lines. The pTRM and pTRM tail checks are omitted for clarity. *This data was translated along the abscissa so that it did not overlap.*

**Fig. 12.** Arai plots produced from four simulations using the same characteristic function $\chi$ as Fig. 1 but with $\lambda = 0.07$. These suggest that the perpendicular protocol produces pTRM check measurements that are more likely to fail through MD behaviour than those produced by experiments using the Original Thellier protocol. They also suggest that the degree of MD behaviour required to cause this pTRM check failure need not have a significant effect on the measured PI in any of the experiments.
Another approach is to look within datasets produced using just one method (i.e., microwave or thermal) and compare results derived using different protocols. A significant number of microwave studies employed both the perpendicular and Coe protocol and comparisons of their results are provided in Fig. 10d and e. These indicate that there is a tendency at the sample level (where there are 39 pairs of results from 3 studies) for the Coe results to be slightly higher than their thermal counterparts but that this entirely disappears at the unit level (where there are 40 pairs of results from 6 studies). There is therefore inconclusive evidence over whether the Coe protocol produces small overestimates in the microwave experiments (as predicted by the phenomenological model, Fig. 1).

Despite the generally good agreement between the model and experimental results in the majority of cases, it cannot yet be ruled out that there is something intrinsic about microwave experiments that suppress MD behaviour relative to identical thermal experiments. The evidence does not conclusively confirm or refute the hypothesis that identical experiments performed using thermal and microwave energy produce identical results when MD effects are present. Further work is clearly required to address this question specifically. One important factor, not discussed yet, that could be responsible for generally improved PI results is that microwave experiments are performed in their entirety, one sample at a time, rather than using batches of samples as in the thermal experiments. In principle, this allows microwave experiments to be progressively modified by the user (changing the power steps used and the applied field intensity and direction) to minimise recurring non-ideal behaviour. It is beyond the remit of this study to test whether this effect was significant in these studies but it is worth bearing in mind as another plausible factor, in addition to the choice of protocol, potentially causing a minimisation of MD effects in the results of the microwave experiments.

5. Discussion

5.1. Implications for published palaeointensity results

In the majority of cases where palaeointensity measurements derived using the thermal and microwave approaches were compared with one another, discrepancies were observed in at least some of their results. This raises some important questions, the first of which is: in such cases where the measurements do not agree, which set of results should be trusted as being accurate? Section 4 argued that, in seven of the eight studies which show significant discrepancies in the measured PI, we would expect the answer to be that the microwave results were more trustworthy because the cause of the discrepancy is an error introduced into the thermal results (Table 2). It is only possible to test this claim in one case: the H60 dataset is the only one which showed significant discrepancies between the thermal and microwave results and where we can know for certain what the correct result should be. In this comparison, the microwave approach produced the much more accurate mean palaeointensity measurements although this may have been largely a result of the sampling location for the two studies (Section 4.3).

It is possible that in some of the cases shown in Table 2, the introduced errors in the thermal results counteract some other source of bias in both sets of results. For example, Biggin et al. (2007a)
demonstrated that enhanced MD behaviour in some palaeointensity results from Etna lavas actually increased their accuracy by countering another source of bias which was ascribed to cooling rate effects. Nonetheless, a dataset which is not shown to be subject to any specific source of bias should generally be considered as more reliable than one that is strongly suspected of being so. Because of this, I argue that for every case where a discrepancy was observed (except JOR), the mean result from the microwave experiments should be considered more reliable than its thermal counterpart.

That eight out of the 13 studies considered here should give significant discrepancies in the palaeointensity results of the microwave and thermal experiments (implying that at least one is incorrect) suggests that inaccurate measurements may be obtained more frequently during the course of normal palaeointensity studies than is generally thought. According to the interpretation made here, some of the thermally-derived results in approximately half of these studies are expected to be significantly biased to high values. There seems little reason to consider the bulk of the rock units listed in Table 2 as out of the ordinary in any way with respect to targets for palaeointensity studies in the literature. Furthermore, the thermal PI and rock magnetic results were also fairly typical of what has been published over the last few decades. As for practically all igneous material (a notable exception being volcanic glass, Pick and Tauxe, 1993), the samples exhibited hysteresis properties typical of PSD grains or of a mixture of MD and SD grains. They tend to be rather young which may have been important in allowing some of the thermal results (those with concave-up Arai plots) to give overestimates. Significant secondary components of magnetisation may have caused the PI estimate to have been reduced or biased in the opposite sense in some of these samples.

Regardless of the cause(s) of the overestimates in the thermal PI results that have been inferred in this study, their frequency in this study suggests that this may be a common problem for thermal Thellier-type studies performed on igneous rocks in general. Is there any evidence for this in published databases?

A recent study (Donadini et al., 2007) compared palaeointensity measurements with ages less than 4 kyr made using lava samples and the (thermal) Thellier method against measurements made using other materials and/or methods. They divided the data into 24 age bins and found that most (18) of the binned averages of the Thellier lava data were higher than the associated binned average of the other data. On the basis of these differences being mostly within 1 standard deviation of one another, they argued that there was no significant offset between the two datasets. However, the observation of 18 similar outcomes from 24 trials is significant at the 95% confidence level. Furthermore, the data shown in their Fig. 11, when subject to a paired t-test, indicate the ’lavas [Thellier]’ bin averages tend to be higher than their ’all data’ counterparts with 95% confidence. Consequently, there is evidence from the GEOMAGIA database for widespread overestimation of the palaeointensity by lavas studied using the thermal Thellier method.

Ziegler et al. (2008) compared the distributions (after temporal resampling) of palaeointensities and archaeointensities from the time range 0–7 ka (also using the GEOMAGIA database) and found the two to be significantly different. Specifically, they observed that the mean of the data derived from the lavas was more than 10% higher than the mean of the data derived from archaeological materials and that the former distribution was more scattered. They cited difference in average cooling rates in the two materials as the most likely explanation for this systematic difference and made a correction to the palaeointensity data assuming they were biased 15% too high.

As discussed in Section 4.1, differences in cooling rate do not appear to offer a viable explanation for the microwave–thermal discrepancies observed in this study. They therefore seem an unlikely candidate for explaining the discrepancies observed in the database by Ziegler et al. and Donadini et al. too. Any cooling rate effect in lavas would be expected to be even larger in microwave PI estimates than in their thermal counterparts (since the difference between natural and laboratory cooling times is even greater). Therefore, if this effect has the tendency to make PI measurements higher, then microwave PI estimates should be higher still than their thermal equivalents.

An alternative interpretation is that it is multidomain effects which are largely responsible for the significant differences observed in the GEOMAGIA database. This is plausible because lavas tend to have a larger average grain size than the bulk of materials used to produce archaeointensity estimates (pottery, ceramics, bricks, etc) and because the Thellier method is the dominant methodology used for both types of measurement (Donadini et al., 2007). This is favoured as the mechanism here because the effects of cooling rate have not been shown to lead to a significant palaeointensity overestimation in lavas as it has in truly single domain material (Fox and Aitken, 1980). Indeed, the influence of cooling rate on palaeointensities measured from lava samples may be negligible or even have the reverse effect (Hill and Shaw, 2000; Biggin et al., 2007a).

Various studies by Japanese groups (e.g. Yamamoto et al., 2003; Mochizuki et al., 2004; Yamamoto and Hoshi, 2008) have argued, on the basis of differences between thermal Thellier and LTD-DHT Shaw results (particular in historic lavas), for consistent overestimation of the palaeointensity by the former method applied to lavas. Yamamoto and Tsukakawa (2005) recently argued that the average dipole moment for the last 5 Myr may be a factor of 2 too high as a consequence of this problem. However, the differences observed by Donadini et al. (2007) and Ziegler et al. (2008) do not amount to anything on this scale.

Could MD behaviour be the dominant cause of the observed discrepancies between thermal Thellier palaeointensities and those measured using the LTD-DHT Shaw method? A detailed comparison is beyond the scope of the present study but it is worth noting a few factors which might favour this hypothesis. Firstly, it is unlikely that MD behaviour would cause any intrinsic bias to LTD-DHT Shaw measurements because this involves cooling the samples down from above the Curie temperature of the magnetic minerals. Therefore, the discrepancies could be due to this problem in the Thellier results. Secondly, many of the discrepant Thellier results are associated with concave-up Arai plots (e.g. Yamamoto et al., 2003) where the lower temperature portion is used for the estimate. Finally, the alternative explanation given in these studies to explain the higher Thellier results is that the lavas acquired a thermochemical remnant magnetisation (TRCM) during primary cooling. However, it is not yet fully clear why this process would produce an overestimate in thermal Thellier results but not in microwave Thellier-type or LTD-DHT Shaw measurements.

5.2. Magnetic Hysteresis Measurements

It is standard practise for palaeointensity studies to present measurements of the hysteresis properties of some representative samples from the studied units. Those displayed in Table 2 are no exception: hysteresis measurements were made for some of the samples from each of these datasets.

Some examples of Day et al. (1977) plots published in the studies summarised in Table 2 are provided in Fig. 14. The overwhelming majority of all the measured samples produced ratios of saturated remanent to saturation magnetisation ($M_r/M_s$) between 0.05 and 0.4 and therefore fall within the region generally designated ’PSD’ or ’SD + MD’ (for titanomagnetites). It has already been shown numerous times (e.g. Carlut and Kent, 2002; Biggin and Thomas, 2003; Coe et al., 2004) that samples with such hysteresis properties may pro-
Fig. 14. Examples of Day et al. (1977) plots for four sets of samples whose PI results were featured in this study. Examples were chosen to demonstrate that, at the study level at least, the degree of MD behaviour indicated in the PI experiments appears not to be well predicted by the behaviour in the hysteresis experiments. Plots are redrawn from (a) Thomas et al. (2004), (b) Biggin et al. (2007a), (c) Gratton et al. (2005b), and (d) Herrero-Bevera and Valet (2005).

Produce PI estimates that are biased by MD effects. Despite this, some PI studies are still published using hysteresis measurements as the sole criterion with respect to defining the suitability of the domain state of the samples.

It has not been possible in this study to perform any quantitative comparison of hysteresis properties with behaviour in the PI experiments because this data was generally not available in a sample-specific form. Nonetheless, it is not necessary to perform more than a cursory examination of the hysteresis data to conclude that they do not tie in particularly well with the specific levels of MD behaviour implied by this study. Four studies (ICE, E20, NIU, and PRB) are notable for producing hysteresis data that are particularly suggestive of the magnetic carriers having a coarse mean grain size (e.g. Fig. 14a and b). Taken at face value, these data suggest that these sets of samples are the most likely to produce non-ideal behaviour related to MD effects in the PI experiments. However, there is nothing in Table 2 (success rates, \( \beta \) values, microwave–thermal discrepancy, etc.) to single these samples out as particularly unsuitable palaeointensity recorders. Indeed it was concluded in Section 4.5 that the PRB samples probably behaved in a manner close to ideal SD grains in the PI experiments. Similarly, SOH samples produced hysteresis properties that were among the most SD-like observed in all the studies (Fig. 14c) and yet were apparently significantly affected by MD effects in the thermal PI experiments (Section 4.2).

A very recent study by Peterson et al. (2009) tested the efficacy of hysteresis parameters as pre-selectors for a large number of PI experiments performed on samples from pyroclastic deposits emplaced during historic eruptions. They found that analyses of both hysteresis loops and, in contrast to the findings of Carvallo et al. (2006), FORC diagrams could not provide adequate pre-selection criteria for their samples. This can, in part, be ascribed to thermally-induced alteration (which is not evident from isothermal hysteresis measurements) affecting the PI experiments and also different phases dominating the high-field and remanent magnetisation of the samples. However, experimental design was also found to be a likely factor in causing failure of the PI experiments. Specifically, a very low ratio of \( H_{\text{lab}}/H_{\text{nrm}} \) (0.18) coupled with modest MD properties probably made a significant contribution to the complete failure of their Mount St Helens samples in the PI experiments.

The present study seems to highlight further that the degree of non-ideal behaviour associated with MD grains that is expressed in a PI experiment may not be well-indicated in advance by hysteresis experiments. That is not to say that these are not useful, but that the extent of MD effects in Thellier-type experiments appears to be sensitive not only to the properties of the studied sample but
also very strongly to the conditions (protocol, applied field intensity, direction, etc) of the individual experiment. It is likely that this sensitivity explains a large part of the apparently poor correlation between hysteresis properties and PI behaviour in the present study.

5.3. Recommendations for future studies

The design of the Thellier-type experiments used in most palaeointensity studies tends to be optimised for detecting alteration and secondary magnetisations (i.e. they use many heating steps with small temperature increments and pTRM checks). However, given the frequency with which it may bias the results of PI studies, it seems apparent that future studies should also take MD behaviour into account when designing their experiments.

There are two broad approaches which can be taken to deal with the problem of MD in PI experiments: minimisation and detection. Techniques which may help minimise MD behaviour include:

- Applying the field \(H_{\text{lab}}\) during both the heating and cooling legs of any remagnetisation treatment (Biggin, 2006).
- Avoiding, where possible, particularly large or small ratios of \(H_{\text{lab}}/H_{\text{anc}}\).
- Applying \(H_{\text{lab}}\) such that \(\theta\) is close to zero (this is especially effective in conjunction with the Aitken protocol but also effective using the Coe, original Thellier, or IZZI protocols; Fig. 1).
- Using the perpendicular or quasi-perpendicular (QP) protocol and avoiding use of the original Thellier protocol in particular (Fig. 1).
- Designing the experiment to use fewer well-spaced temperature steps (Biggin, 2006).
- Abandoning the Thellier-type approach altogether and using the Dekkers–Böhnel (2006) multi-specimen method or the LTD-DHT Shaw method (Yamamoto et al., 2003) instead.

The first two of these steps should be enacted wherever possible as they have no drawbacks and may be a powerful means of improving experimental behaviour (Biggin, 2006; Paterson et al., 2009). A corollary of the second of these is that PI results which retrospectively indicate an extreme high or low ratio of \(H_{\text{lab}}/H_{\text{anc}}\) should be treated with a high degree of caution. Similarly, the use of the original Thellier protocol appears to produce exaggerated concave-up behaviour while offering no obvious benefit over other double-treatment protocols. It is therefore not recommended for use in future studies.

For the reliable detection of MD behaviour in Thellier-type experiments on the other hand, \(H_{\text{lab}}\) should be applied such that \(\theta\) is large (i.e. 90–180°) and then pTRM tail checks should be incorporated into a Coe protocol experiment or the IZZI protocol used. If the IZZI method is used then the large value of \(\theta\) would be expected to produce exaggerated zig-zagging of the Arai plot if significant MD behaviour is exhibited (Yu and Taueke, 2005). If pTRM tail checks are used then these will fail modest criteria.

It is significant that, of the seven thermal studies listed in Table 2 that may have been biased by MD effects, only one employed pTRM tail checks (and none the IZZI protocol). Furthermore, that study (GWS; Garcia et al., 2006) together with several of the microwave studies that employed pTRM tail checks were rather lenient with the criteria they applied in this respect (Table 1). In addition to the relevant criterion being strict enough, another requirement for the pTRM tail check to be useful is that \(H_{\text{lab}}\) is such that \(\theta\) is not close to zero (Yu and Dunlop, 2003). Objectively-based pTRM tail check criteria which take into account the angle between the applied field and the NRM are provided by Leonhardt et al. (2004).

Palaeointensity studies employing the Aitken protocol have no use for the pTRM tail check: careful monitoring of the direction of the NRM in sample coordinates after each demagnetisation treatment allows a check for MD behaviour. However, in this case, the check is only valid when the laboratory field is not parallel or anti-parallel to the NRM of the sample. In the former instance the MD behaviour will be minimised in any case, but in the latter, the MD behaviour may be intensely exaggerated so that a strongly biased measurement is made (Fig. 4 and Section 4.2).

Obviously, an approach optimised to minimise MD effects may render it very difficult to detect if these effects are nonetheless still able to bias the PI estimate. Extra precautions must therefore be taken to ensure that this is not the case. For example, if \(H_{\text{lab}}\) is chosen so that \(\theta\) is close to zero (rendering pTRM tail checks relatively useless), then the resulting palaeointensity estimate can only be considered free from MD bias if it is associated with a sufficiently large linear portion of the Arai plot. The question of what constitutes ‘sufficiently large’ remains an open one. Fig. 7c shows the median discrepancy dropping dramatically as \(f\) increases from 0.5 to 0.6 whereas Fig. 7d shows it to be still significant at \(f > 0.6\). However, it is possible that some of these \(f\) values have been artificially increased by the presence of secondary magnetisations; an effect that may be avoided in future studies through the replacement of the conventional \(f\) value with \(f_{\text{vds}}\) (Tauxe and Staudigel, 2004). A criterion requiring \(f_{\text{vds}} > 0.5\) is consistent with what certain previous studies (e.g. Biggin and Thomas, 2003; Chauvin et al., 2005), have suggested as being necessary to avoid biased from MD effects. However, some of the observations made in this study (Figs. 1.6 and 7) suggest that this may not be high enough and that \(f_{\text{vds}} > 0.7\) is required to be absolutely certain that a PI estimate is free from any bias due to MD effects.

The converse of the argument made above is also true: any approach designed to maximise the likelihood of detecting MD behaviour (e.g. using \(\theta \geq 0\)) runs the risk of exaggerating effects that might otherwise be negligible. Therefore, it may be possible to obtain an ambiguous result with either approach: in the minimisation approach, seemingly good-quality estimates associated with values of \(f_{\text{vds}} < 0.7\) may be biased; in the detection approach, estimates failing even mild selection criteria may have been successful given less exacting experimental conditions.

On balance, it seems in most cases preferable to adopt the detection approach first (e.g. apply the IZZI protocol with \(\theta > 90°\)) and then, if this first set of experiments yields low quality results (e.g. visible zig-zagging of the Arai plot), apply the minimisation approach (e.g. Aitken protocol with \(\theta = 0°\) or the QP protocol) to sister samples. If this second set of experiments cannot produce high quality results with \(f_{\text{vds}}\) values \(> 0.7\), then it must be conceded that the samples in question cannot produce PI estimates which are unambiguously free from biasing by MD effects.

It is the findings of this study that pTRM and pTRM tail checks used in conjunction with the perpendicular protocol can be rendered essentially useless by even slight MD behaviour. This may well be true for the QP protocol too although this remains to be tested. Consequently, a recommendation of this study is that these experiments should generally omit these checks and rely (as does the Dekkers–Böhnel method) on other information to guard against the occurrence of alteration in the PI experiments.

6. Conclusions

1. There is clear, statistically significant evidence at the sample, rock unit, and study level that microwave palaeointensity experiments performed on igneous rocks tend to produce lower palaeointensity results (20–40%) than thermal Thellier-type experiments. Furthermore, thermally-derived results tend to have a higher intrinsic Arai plot scatter (higher \(\beta\) values) associated with them but an \(f\) (fraction) parameter that is comparable.
In at least five out of the eight comparison studies showing significant disagreement between the results, higher thermal PI estimates tended to be associated with the low-temperature portions of concave-up Arai plots.

2. These systematic differences are most likely caused by the enhancement of non-ideal effects associated with ‘multidomain’ grains (also potentially grains in the vortex state and interacting single domain grains) in the results of the thermal experiments relative to their microwave counterparts. This is probably due, in large part, to the use of, predominantly, a single-treatment protocol in the microwave experiments and double-treatment protocols in the thermal experiments. However, it cannot presently be ruled out that there is also something (as yet unidentified) intrinsic to microwave experiments that minimises MD effects.

3. A significant proportion of thermal Thellier-type palaeointensity estimates derived from igneous rocks and published in the literature may be biased high (by tens of percent) by MD effects. This conclusion is arrived at by logical extension of the finding that, in more than half of the studies considered here, the thermally derived results appear to be biased in this way. Evidence that Thellier PI results derived from lavas are frequently biased to higher values is also available in the results of recent analyses of the GEOMAGIAO database (Donadini et al., 2007; Ziegler et al., 2008).

4. Sample pre-selection using rock magnetic techniques is on its own, clearly not an adequate means of avoiding bias caused by MD effects in PI experiments. However, numerous steps can be taken in the Thellier-type experiments themselves to minimise the probability of bias affecting future results. The recommendations made here include the use of the IZZI method and/or pTRM tail checks in conjunction with a large angle between the applied field and the sample NRM so that potential bias can be recognised easily. Alternative approaches (including the use of non-Thellier methods) allow the minimisation of these non-ideal effects but can seldom rule them out entirely and may increase the potential for biasing from other sources.

Acknowledgements

This research was initiated with funding provided by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) and continued during a Natural Environment Research Council (NERC) Advanced Fellowship (NE/F015208/1). Maxwell Brown, Martin Gratton, Mimi Hill, Nicola Pressling, and Yuhji Yamamoto (NERC) Advanced Fellowship (NE/F015208/1). Maxwell Brown, and continued during a Natural Environment Research Council (NWO) Erdlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) Erdlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) advanced fellowship. The Centre for the Study of Earth and Space Science and Engineering (CSESSP) supports the broader goals of this research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.pepi.2010.03.005.

References


