Comparison of Thermal and Microwave Paleointensity Estimates in Specimens Displaying NonIdeal Behavior in ThellierStyle Paleointensity Experiments

Grappone J Michael (Orcid ID: 0000-0001-5004-8561)

Biggin Andrew J. (Orcid ID: 0000-0003-4164-5924)

Barrett Thomas, J (Orcid ID: 0000-0002-5351-457X)

Comparison of thermal and microwave paleointensity estimates in specimens displaying non-ideal behavior in Thellier-style paleointensity experiments

J. Michael Grappone [a,*], Andrew J. Biggin [a], Thomas J. Barrett [a,b], Mimi J. Hill [a], Courtney J. Sprain [a,c]

[a] *Geomagnetism Laboratory, School of Environmental Sciences, University of Liverpool, Liverpool L69 7ZE, UK*

[b] School of Physical Sciences, The Open University, Walton Hall, Milton Keynes, MK7 6AA UK

[c] Department of Geological Sciences, The University of Florida, Gainesville, FL 32611

* Corresponding author. Email address: grappone@liv.ac.uk (J. M. Grappone)

Key Points:

- Causes for significant inconsistencies in important paleointensity records over the 0 – 45ka time period in Hawaii are investigated.
- Disagreement is related to both the different demagnetization mechanisms and paleointensity protocols employed.
- Both previous records are likely biased with the true values expected to lie intermediate between them.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2020JB019802

Abstract

Determining the strength of the ancient geomagnetic field is vital to our understanding of the core and geodynamo but obtaining reliable measurements of the paleointensity is fraught with difficulties. Over a quarter of magnetic field strength estimates within the global paleointensity database from 0-5 Ma come from Hawai'i. Two previous studies on the SOH1 drill core gave inconsistent, apparently method-dependent paleointensity estimates, with an average difference of 30%. The paleointensity methods employed in the two studies differed both in demagnetization mechanism (thermal or microwave radiation) and Thellier-style protocol (perpendicular and Original Thellier protocols) – both variables that could cause the strong differences in the estimates obtained. Paleointensity experiments have therefore been conducted on 79 specimens using the previously untested combinations of Thermal-Perpendicular and Microwave-Original Thellier methods to analyze the effects of demagnetization mechanism and protocol in isolation. We find that, individually, neither demagnetization mechanism nor protocol entirely explains the differences in paleointensity estimates. Specifically, we found that non-ideal multi-domain-like effects are enhanced using the Original Thellier protocol (independent of demagnetization mechanism), often resulting in paleointensity overestimation. However, we also find evidence, supporting recent findings from the 1960 Kilauea lava flow, that Microwave-Perpendicular experiments performed without pTRM checks can produce underestimates of the paleointensity due to unaccounted-for sample alteration at higher microwave powers. Together, these findings support that the true paleointensities fall between the estimates previously published and emphasize the need for future studies (thermal or microwave) to use protocols with both pTRM checks and a means of detecting non-ideal grain effects.

1. Motivation

The Pacific Ocean covers 30% of the Earth's surface but has few islands, with the Hawaiian Islands being some of the most easily accessed. Volcanic islands, like Hawai'i, contain some of the best records of the temporal variation in Earth's magnetic field over the last few Myr. Paleosecular variation timescales of this length are necessary to better understand long-term variations in geomagnetic behavior, as well as crust, mantle, and core interactions (e.g. McElhinny and Merrill (1975)). Accurate paleointensity data in the 0 – 5 Ma period is crucial because only in this interval is there enough spatial and temporal global coverage of data to characterize long term (Myr) variations. In order to obtain accurate paleointensity data, an appropriate paleointensity method for the mineral magnetic characteristics of any particular site must be used.

Many paleomagnetic studies over the last 60 years (e.g. Coe et al. (1978), Cromwell et al. (2018), de Groot et al. (2013), Doell and Cox (1963), Doell and Cox (1965), and Teanby et al. (2002)) have found the Hawaiian Islands ideal for studying magnetic field variations in the central Pacific Ocean over the past hundred to few million years. Hawaiian absolute paleointensities (PI) have substantial temporal coverage and comprise 28% of the global paleointensity (PINT) database in this interval (Biggin et al., 2015) and are therefore important to study to understand long-term field behavior over this time interval. Numerous studies have taken advantage of the drill core from the Scientific Observation Hole (SOH) and Hawai'i Scientific Drilling Project (HSDP) projects to extract the required paleointensity data over the last 45 kyr (e.g. Cai et al. (2017), Gratton et al. (2005), and Teanby et al. (2002)), but the data have proven to be inconsistent and thus of potentially limited use.

The paleomagnetism of the SOH1 core was studied twice previously- once by Teanby et al. (2002) using thermal PI experiments and again by Gratton et al. (2005) using microwave PI experiments and a different PI protocol. Teanby et al. (2002) and Gratton et al. (2005) each sampled the core independently and extracted paleointensities from 83 common flows. Teanby et al. (2002) additionally reported a new inclination record and dated the flows from 0 - 45 ka. The mean paleointensities reported by the two studies, 33.5 μT (Teanby et al., 2002) and 25.1 μT (Gratton et al., 2005), differ by approximately 33%. Both studies reported mean uncertainty estimates of approximately 10%, which implies the possibility for a resolvable difference between them.

At this stage it is useful to introduce our nomenclature that a given paleointensity "method" is composed of a combination of a specific "demagnetization mechanism" and a specific "protocol". Teanby et al. (2002) used the conventional thermal demagnetization mechanism paired with the Original Thellier protocol (Thellier and Thellier, 1959), while Gratton et al. (2005) used the microwave demagnetization mechanism paired (predominantly) with the Perpendicular protocol (Kono and Ueno, 1977). A third study, Laj et al. (2011), used the raw data from Teanby et al. (2002) (reanalyzed with their updated selection criteria to give an SOH1 average of 29.7 μT) combined with additional (non-SOH1) data (acquired using the same method) from the SOH4 and HSDP1 drill cores, which cover additional flows, to create a more complete and statistically rigorous Hawaiian paleointensity record. Even after the Laj et al. (2011) reassessment, there remains nearly a 20% discrepancy between the thermal mechanism Original Thellier protocol and the microwave mechanism Perpendicular protocol results.

To investigate discrepancies between paleointensity results obtained using different demagnetization mechanisms (i.e. thermal and microwave), a meta-analysis of 13 paired studies (including that of the SOH1 drill core) was undertaken by Biggin (2010). It is important to note that the studies assessed in the Biggin (2010) analysis differed not only in demagnetization mechanism but also in protocol. Biggin (2010) concluded, firstly, that systematic differences existed between paleointensity estimates derived from thermal and microwave experiments performed on the same rocks, with the former tending to be significantly higher than the latter (at the 95% confidence level).

We carried out a further analysis (available in Supporting Information A), which showed this as well at the flow level. Secondly, Biggin (2010) suggested that the most plausible explanation for the bulk of these discrepancies resided in unrecognized biasing from multi-domain-like effects being more prevalent in the thermal results. Importantly it was suggested that the discrepancies were more or less entirely due to the difference in the *protocols* used rather than in the *demagnetization mechanism*. With respect to the SOH1 case, Biggin (2010) hypothesized that the thermal results being higher than the microwave results was likely unrelated to the choice of thermal or microwave energy for demagnetization. Rather, the discrepancy was due to Teanby et al. (2002) employing the Original Thellier protocol while Gratton et al. (2005) employed the Perpendicular protocol.

Since our initial flow level analysis of the SOH1 data confirmed the results of Biggin (2010) the aim of the present study is to test the hypothesis that the differences found in paleointensity results from the SOH1 core are entirely due to protocol and not due to demagnetization mechanism. We hypothesize that if the demagnetization mechanism-protocol pairs are reversed, such that microwave demagnetization is paired with the Original Thellier protocol and thermal demagnetization is paired with the Perpendicular protocol, then the sense of the discrepancy between the microwave and thermal results should reverse such that the former should give higher estimates than the latter. This study will explicitly test this hypothesis using new experiments performed on 24 of the same SOH1 flows as studied originally by both Teanby et al. (2002) and Gratton et al. (2005).

The results of this study are important on several levels. First, they provide improved insight into the strength of the magnetic field at the time of emplacement of the 241 flows sampled by the SOH1 drill core dataset. Second, they have implications for how the swathes of paleointensity estimates obtained by similar methods from rocks elsewhere in the world should be interpreted. As such, they expand the results of a recent restudy of the 1960 Kilauea lava flow by Grappone et al. (2019) to more Hawaiian lava flows. Lastly, they can provide guidance on how future paleointensity experiments should be performed and analyzed in order to maximize their reliability.

2. Drill core geology and sampling

The samples used in this study are from the SOH1 drill core, which was drilled from the Kilauea volcano on Hawai'i Island between 1989 and 1991 at 19°29'N, 154°54'W, to a total depth of 1685 m. The Hawai'i Institute of Geophysics and Planetology and the Hawai'i Natural Energy Institute drilled the borehole to assess the viability of using geothermal energy in the area (Quane et al., 2000). The core consists primarily of a'a (~ 66%) and pahoehoe (~ 22%) lavas from 241 aerial, subaerial, and submarine flows with thicknesses varying from 0.3 m to 17.4 m (Gratton et al., 2005; Teanby et al., 2002). The remaining ~11% of the core consists of dyke intrusions. Samples were taken from 196 lava flows from the upper 779 m of the core to avoid the increasing number of dyke intrusions and apparent alteration at greater depths (Gratton et al., 2005). This portion of the core has been modelled with an age range of 0 – 45 ka (Teanby et al., 2002).

The portions of each drilled 2.5 cm diameter core that were saved by Gratton et al. (2005) from their SOH1 study were retrieved from the University of Liverpool archive and subsequently used in this study. We cut 120 new specimens for the restudy from the material remaining from 24 flows that span the range of flows sampled in Gratton et al. (2005). Although the specimens used in this study are sister specimens from Gratton et al. (2005), we cannot rule out the possibility that the flows may be inhomogeneous even on a cm scale. Flows were selected based on the degree of disagreement in paleointensity (PI) estimates between the Gratton et al. (2005) and Teanby et al. (2002) studies, number of samples previously studied, and the availability of specimens. For the flows selected for

this study, the Teanby et al. (2002) PI estimates, at the flow level, ranged from 21% lower to 53% higher (with a mean of 27% higher) than the Gratton et al. (2005) estimates.

3. Rock magnetism

Gratton et al. (2005) undertook an extensive survey of the hysteresis loop parameters of the SOH1 borehole. The raw data from their rock magnetic survey were reanalyzed here using *Hystlab's* automatic hysteresis loop processing program (Paterson et al., 2018) and are replotted in Figure 1, with the flows investigated herein highlighted. The bulk domain stability (BDS) trendline from Paterson et al. (2017) lies below the main sequence of hysteresis loop parameters for SOH1 flows, which indicates the data have a mixture of magnetic domain types, potentially also including superparamagnetic grains. The specimens used in this study sample the main sequence of SOH1 data observed in Figure 1. The ratio of magnetic remanence to saturation magnetization (Mrs/Ms) values of the main sequence of SOH1 data lie consistently above the values that would be expected for multi-domain grains of magnetite (Mrs/Ms > 0.1), which implies the presence of single domain and non-single domain grains. Gratton et al. (2005) also determined thermomagnetic behavior for all the flows in their study. They found highly reversible thermomagnetic curves with 98% of Curie temperatures falling in the range from 520 – 600 °C, with a mean of 561 °C and median of 570 °C. 22% of flows also contained a secondary ferrimagnetic phase with Curie temperatures below 340 °C.

The hysteresis and thermomagnetic parameters are typical of low-Ti magnetite-rich basaltic lavas found on Hawai'i and which have been the focus of other Hawaiian paleointensity surveys (e.g. Cai et al. (2017), Cromwell et al. (2018), and Hill and Shaw (2000)). Additional, new rock magnetic information (FORC and SEM analysis) from the main sequence can be found in Supporting Information B.

4. Methods

All new thermal tests were carried out in air using a Magnetic Measurements MMTD-80 thermal demagnetizer, and specimens were cooled quickly using a built-in cooling fan. The specimens were then measured on the University of Liverpool Geomagnetism Laboratory's 2G SQUID Magnetometer and RAPID system. All microwave tests were run on the 14 GHz Tristan Microwave System (Hill et al., 2008), also at the University of Liverpool. The goal of this study was to replicate the experimental conditions of the previous studies as closely as possible to properly isolate each variable of interest.

In the Original Thellier (OT) protocol (Thellier and Thellier, 1959), each specimen is heated to a given temperature, T_N, in a known, non-zero intensity magnetic field and then cooled to room temperature and measured. The polarity of the magnetic field is then reversed, and the sample is taken to T_N again. The protocol can also be referred to as infield-infield, or 'II'. The process is then repeated at T_{N+1}. Partial thermal remanent magnetization (pTRM) checks were included after every 2nd step (i.e. from T_{N+1} to T_{N-1}). For consistency with the PINT database (Biggin et al., 2009), when used with thermal energy, this protocol (with pTRM checks) will be referred to as Th-OT+, to acknowledge the pTRM check addition. The microwave (MW) variant of the OT protocol simply replaces the heat with microwave power integrals and will be referred to as MW-OT+. We used powers ranging from 5 to 40 W in 3 to 5 W steps, applied for 5 to 15 s and assumed that any power not reflected was absorbed by the specimen-cavity coupled system. For MW-OT+, we used a steady magnetic field applied parallel/anti-parallel to the natural remanent magnetization (NRM), which

Biggin (2010) predicted to be the least affected by non-ideal Arai plot behavior. For Th-OT+, we used an applied field with an inclination of $\pm$ 90° in specimen coordinates.

The Perpendicular (Perp) protocol is a modification of the Original Thellier protocol that only requires a single thermal or microwave treatment (Kono and Ueno, 1977). Samples first get stepwise demagnetized in a zero field to remove any soft magnetic overprints. Once the primary component of magnetization is identified as beginning at some T_P, the sample is then heated to T_{P+1} in a magnetic field applied in the direction perpendicular to the characteristic component (the remaining NRM). The process is then repeated for T_{P+2} and higher. Gratton et al. (2005) determined that many samples had a second ferrimagnetic phase with an unblocking temperature around 300 °C, which they interpreted to be the result of oxyexsolution into Ti-rich and Ti-poor lamellae during extrusion. Therefore, samples undergoing Thermal Perpendicular (Th-Perp) were first step-wise demagnetized in 40-50 °C steps from 100 °C to 300-340 °C to ensure that the perpendicular field was applied only to the high temperature ferrimagnetic phase. After successfully finding the characteristic direction, the field in the oven was switched on. The steps continued to 590 °C in 10-30 °C increments. The process is the same using the microwave system (MW-Perp), but with power integral steps instead of temperature steps until a consistent magnetic direction is obtained (Hill and Shaw, 2007). We did not include any pTRM checks in our new perpendicular experiments in order to replicate the methods used by Gratton et al. (2005). All data were analyzed using the methods described in Hill and Shaw (2007).

Laboratory field strengths were selected that were as close to the original analyses as possible. All experiments in Teanby et al. (2002) were carried out using an applied field of 40 μT, as were the new MW-OT+ experiments completed herein. Microwave studies are carried out one sample at a time, so the field strength often varied specimen to specimen within a flow in Gratton et al. (2005). For the specimens we selected, the mean applied field used in Gratton et al. (2005) for microwave treatments was 31 ± 1.3 μT, so we used a field of 31 μT in our Th-Perp experiments.

A summary of the experiments carried out in this study is given in Table 1. We ran 79 specimens using either MW-OT+ or Th-Perp. Additionally, 19 specimens were tested using Th-OT+ and 22 were tested using MW-Perp, replicating the original studies in order to confirm the previous results. The median number of specimens tested from each of the 24 flows we examined was 3, with a range of 1-11.

Our selection criteria are based on the MC-CRIT.C1 selection criteria, without tail checks, from Paterson et al. (2015). These were also used successfully by Grappone et al. (2019) to study the 1960 Kilauea lava flow. We relaxed the FRAC/f criterion (Shaar and Tauxe, 2013) (a measure of the proportion of the NRM used to determine the result) from 0.45 (used in Grappone et al., 2019) to 0.35 because of extensive alteration observed at higher temperatures/power integrals and because of difficulties in demagnetizing the specimens using microwaves. Relaxing this criterion yielded 7 additional specimen-level estimates most notably from MW-Perp experiments, without changing any flow-level estimates in a statistically significant manner. The selection criteria are detailed in Table 2.

For our analysis, we use two statistical tests: the T-test and the Wilcoxon signed rank test (see for example, Klugh (1986)). For the study-level data, which are normally distributed (see the failure to reject the null hypothesis in Kolmogorov-Smirnov test in Supporting Information A), we use a two-sample, unpaired T-test to test if the method-level means are equal. A paired difference test is more appropriate for the flow-level data, which have flow-level data pairings. The paired data do not visually appear to be normally distributed (a weak assumption required for a paired T-test), so we should choose a nonparametric test. Biggin (2010) used the Wilcoxon signed rank test in their analysis of the SOH1 dataset so for consistency we also use that for flow level data.

5. Paleomagnetic results and analysis

5.1 New data

A summary of all experiments run herein is given in Supporting Information C, and these new experiments are described in this subsection. Paleointensity estimates that passed the selection criteria were obtained from experiments performed on 73 specimens from 20 flows. A pass rate of >50% at the specimen-level was achieved for every experimental method tested. The most common reason for failure of the new MW-Perp and MW-OT+ experiments was low FRAC/f, as the microwave often could not fully demagnetize each specimen. The most common reason for failure for the new Th-OT+ experiments was high DRAT (i.e. pTRM check failures), and for Th-Perp, it was high scatter around the best fit line (β). The only clear difference in Arai plot shape observed between the different methods is that the new Th-OT+ data often show two slopes where for some specimens both slopes passed the selection criteria. In these cases, the low temperature slope was selected, as they had the higher FRAC. These new Th-OT+ data additionally show the highest mean curvature values (as defined by $|\vec{K}'|$) at 0.30, compared to those of the Th-Perp (0.22), MW-OT+ (0.21), or MW-Perp (0.093). All new measurement data can be found on the MagIC database. Examples of passed MW-OT+ and Th-Perp data can be found in Figure 2 (Aii, Aiii, Bii, Biii, Cii, Ciii). Examples of passed Th-OT+ and MW-Perp data can be found in Supporting Information C.

Flow mean results are detailed by paleointensity method in Supporting Information C. The new Th-OT+ and MW-OT+ flow-level mean PI estimates tend to yield higher values with 44.0 ± 16 μT across 5 flows and 29.5 ± 9.2 μT, across 19 flows, respectively. New Th-Perp and MW-Perp experiments tend to yield lower mean PI estimates with estimates of 27.8 ± 8.1 μT, across 11 flows, and 18.5 ± 10 μT, across 5 flows, respectively. If we assume our new PI estimates are normally distributed, which is noted (see Supporting Information A), we can use a two-sample T-test to determine whether observed differences between experiments are statistically significant. From the new data, the Th-OT+ mean estimate (44.0 μT) is higher than the MW-OT+ mean (29.5 μT), the Th-Perp mean (27.8 μT), and the MW-Perp mean (18.5 μT), at the 95% confidence level, with p-values of 0.0147, 0.0176, and 0.0001, respectively. The new data's Th-Perp mean estimate is not statistically distinct from the new data's MW-OT+ mean estimate at the 95% confidence level (p = 0.62). The new data's MW-Perp flow-level mean PI estimate of 18.5 μT is lower than the new data's MW-OT+ mean estimate at the 95% confidence level, with a p-value of 0.0015. The new data's Th-Perp and MW-Perp estimates are not statistically different at the 95% confidence interval, with a p-value of 0.0676.

5.2 Incorporation of existing datasets

The new Th-OT+ results reported here are broadly consistent with their original counterparts reported by Teanby et al. (2002) and Laj et al. (2011). Both have the same two-slope (concave-up) behavior (see Figure 2 for Teanby et al. (2002) results and Supporting Information C for a direct comparison). When considering only the four flows tested in both the present study and by Teanby et al. (2002), the mean PI values (43.5 ± 19 μT and 30.8 ± 14 μT) have overlapping uncertainty bounds and are not statistically distinct from each other at the 95% confidence level (p = 0.3232).

The new MW-Perp data reported here also broadly replicate the lower estimates reported by Gratton et al. (2005). The Arai plots appear single-sloped (see Figure 2 for Gratton et al. (2005) results and Supporting Information C for a direct comparison) and the mean PI values for the nine flows tested here and in Gratton et al. (2005) are 18.2 ± 10 μT and 18.6 ± 9.8 μT, respectively. These mean values are not statistically distinct at the 95% confidence level (p = 0.9462). This result supports Grappone et al. (2019)'s finding that the different generations of the microwave systems give equivalent results.

The consistency of our newly obtained results with those from the previous studies enables us to conclude that we may reasonably combine our new Th-OT+ data with the Teanby et al. (2002) data and our new MW-Perp data with the Gratton et al. (2005) data. All the data is therefore combined to create one dataset which forms the basis for discussion in the following section. The combined SOH1 dataset, consisting of data from this study, Teanby et al. (2002), and Gratton et al. (2005) is summarized in Table 3.

5.3 Analysis of combined dataset

Figures 3 - 5 display a series of one-to-one comparisons of the PI data produced by different methods at the flow level, utilizing data from the combined dataset, which allows the influence of demagnetization technique and PI protocol to be scrutinized. From visual inspection, the data do not appear to be symmetrically random about the 1:1 line nor cluster close to it. Thus the PI estimate data pairs do not visually appear to be consistently normally distributed about the 1:1 line, which indicates that a two-sample T-test may be insufficient. We instead use the Wilcoxon signed rank test (see Biggin, 2010) to examine if the respective deviation of the datasets from the 1:1 line is significant at the 95% ($\alpha = 0.05$) confidence interval. The null hypothesis is that the data scatter about the 1:1 line is random.

In keeping with Biggin (2010), we first confirm that the Th-OT+ data are consistently higher than the MW-Perp data (Figure 3A). The Wilcoxon signed rank test gives W = 1, which corresponds to a p-value of 0.0008 for 15 data points, so we can reject the null hypothesis that the deviation from the 1:1 line and hence the data scatter is random.

Next we test the hypothesis of Biggin (2010) that the primary cause for the Th-OT+ data being consistently higher than the MW-Perp data is due to the differing PI protocol (OT+ vs Perp) and not demagnetization mechanism (MW vs. thermal). We do this by comparing the MW-OT+ and Th-Perp data (i.e. the inverse combination) to see if the OT+ protocol continues to yield systematically higher values than the Perp protocol. It can be seen in Figure 3B that in fact, MW-OT+ data are not consistently higher than Th-Perp data; the paired results are significantly closer to and fall on either side of the 1:1 line. The Wilcoxon signed rank test gives W = 18, which corresponds to a p-value of 0.33 for 10 data points, so we cannot reject the null hypothesis that the deviation from the 1:1 line is random. Having failed to support the simple hypothesis that the protocols are entirely responsible for the differences in PI results, we now examine demagnetization mechanisms and protocols in isolation to probe deeper into these specimens' behavior.

If the cause of the discrepancy between the Th-OT+ and MW-Perp data were purely due to demagnetization mechanism, then we would expect that estimates from MW-OT+ and MW-Perp would be similar and would cluster around the 1:1 line. Similarly, estimates from Th-OT+ and Th-Perp, would also be similar, clustering around their 1:1 line. These cases are plotted in Figure 4. For the flows they have in common (Figure 4A), the mean PI estimate for the MW-OT+ data is 30 ± 8.8 μT and for the MW-Perp data is 26.1 ± 7.7 μT. The Wilcoxon signed rank test gives W = 23, which corresponds to a p-value of 0.0065 for 18 data points, so we can reject the null hypothesis that the deviation from the 1:1 line is random. Next, we investigate the Th-OT+ and Th-Perp data (Figure 4B). For the flows they have in common, the mean PI estimate for the Th-OT+ data is 38.3 ± 14 μT and 27.5 ± 8.3 μT for the Th-Perp data. The Wilcoxon signed rank test for Th-OT+ vs Th-Perp gives W = 3, which gives a p-value of 0.05 for 7 data points, therefore also rejecting the null hypothesis that the deviation from the 1:1 line is random. Therefore, changing the protocol to OT+ from Perp does indeed cause higher paleointensities to be measured, but this is not the entire explanation for the discrepancy illustrated in Figure 3A.

For the flows that have both Th-OT+ and MW-OT+ data (Figure 5A), we observe mean PI estimates of 40.7 ± 13 μT and 28.8 ± 10 μT, respectively. The Th-Perp and MW-Perp data give more similar estimates (Figure 5B). For the flows they have in common, the mean PI estimates are 28 ± 8.6 and 26.6 ± 5.9, respectively. For Th-OT+ vs MW-OT+, the Wilcoxon signed rank test gives W = 1, which corresponds to a p-value of 0.0012 for 14 data points. We can therefore reject the null hypothesis that the deviation from the 1:1 line is random, which indicates Th-OT+ data are higher than the MW-OT+ data. For the Th-Perp and MW-Perp data, the Wilcoxon signed rank test gives W = 25, which corresponds to a p-value of 0.24 for 11 data points. Thus, we cannot reject the null hypothesis that the Th-Perp and MW-Perp data's deviation is random, which suggests that the Th-Perp data are not higher than the MW-Perp data. This result agrees with the T-test in section 5.1.

We therefore observe that neither the demagnetization mechanism nor the protocol used can, in isolation, fully explain the differences in paleointensity estimates observed. We additionally find that the Th-OT+ method yields results that are consistently (and statistically) higher than the other three methods used.

6. Discussion

6.1 Paleointensity methodology differences

Double-heating Thellier protocols have long been known to have problems with multi-domain components causing non-linear Arai plots (Levi, 1977). Hodgson et al. (2018) showed that this can apply to non-single domain oxyexsolved titanomagnetite grains as well, which are common in basaltic lavas such as the Hawaiian lavas of the SOH1 drill core. These non-single domain components can lead to concave up (two-slope) Arai plots. If the low blocking temperature (power integral) portion is used, the data give PI overestimates and conversely, underestimates are obtained if the high blocking temperature (power integral) portion is used (Levi, 1977; Smirnov et al., 2017; Thomas, 1993; Xu and Dunlop, 2004). Grappone et al. (2019) studied samples from the 1960 Kilauea lava flow, another Hawaiian lava flow. They showed that the high estimates often found using thermal double treatment methods (of which Thermal-Original Thellier is one) (e.g. Yamamoto et al. (2003)) are not replicated when using the microwave demagnetization mechanism. Our data here confirm this conclusion, as the Thermal-Original Thellier method yields the highest estimates, but the Microwave-Original Thellier data align much more closely with the Perpendicular datasets obtained using either demagnetization mechanism.

A recent study by Cromwell et al. (2018) (using the Thermal-IZZI method and strict selection criteria) reports PI estimates from 22 surface lava flows across the island of Hawai'i. They found estimates consistent with geomagnetic field models in the 270 – 10,000 yr range. However, the study was unable to reproduce the high estimates found in Teanby et al. (2002) and Laj et al. (2011). Cai et al. (2017) additionally found systematically lower estimates than Laj et al. (2011) using subaerial glassy basaltic margins on older flows from the HSDP2 core, using the same techniques as Cromwell et al. (2018). Based on the data herein and the other studies, the thermal Original Thellier method appears to give higher PI estimates than other PI methods and is likely an overestimate of the true palaeointensity due to exaggerated multi-domain-like effects and the lack (in these experiments) of any mechanism to detect these.

The original microwave study for the 1960 Kilauea lava flow, Hill and Shaw (2000), gave site-level estimates up to 20% lower than the expected value of 36.5 μT. In both the Hill and Shaw (2000) and the Gratton et al. (2005) studies, an older MW system was used with the Perpendicular protocol (Kono and Ueno, 1977), which uses only a single treatment per step and no pTRM checks. Grappone

et al. (2019) showed that changing the microwave experimental protocol to a more common double-treatment technique including alteration (pTRM) checks (in their case the IZZI protocol (Yu et al., 2004)) gave the correct answer of 36.5 μT for the 1960 Kilauea lava flow. Although the Original Thellier protocol's additional treatments cause more thermochemical alteration than the perpendicular protocol, it is likely that the SOH1 samples behave in a similar way to the 1960 Kilauea flow and that the Perpendicular datasets give low estimates due to undetected alteration.

The Microwave-Perpendicular and the Thermal-Perpendicular datasets are not statistically distinct implying that it is not something inherent to the microwave causing the differences between the original Teanby et al. (2002) and Gratton et al. (2005) datasets. It has been shown that for well-behaved (SD) grain-containing ceramics, there is a detectable difference in paleointensity estimates due to differences in cooling rate (Poletti et al., 2013). Since the Microwave-Perpendicular and the Thermal-Perpendicular datasets are not statistically distinguishable, any cooling rate effect would be relatively minimal. Further, the Microwave-Original Thellier data have a faster cooling rate but give lower estimates than the Thermal-Original Thellier data, which is the opposite of the expected cooling rate effect (Poletti et al., 2013).

Additionally, the perpendicular protocol having less than half the number of treatments as in the Original-Thellier protocol introduces another complication if these specimens undergo stress relaxation during the experiment. If even mild stress relaxation affects the specimens, the perpendicular data will appear too low because pTRMs would be gained more efficiently than NRM would be lost (Kosterov and Prevot, 1998).

That the Microwave-Original Thellier results are lower than their Thermal-Original Thellier counterparts suggests that something inherent to the microwave demagnetization mechanism dampens the effects on the data from non-ideal vortex-state grains. If this finding were the result of the microwave not causing the same magnitude of stress relaxation as in thermal experiments, the Microwave-Perpendicular data would be expected to be even lower than the Thermal-Perpendicular data (Kosterov and Prevot, 1998). This does appear to be the case here (Figure 5B), as the differences are not statistically significant.

Based on the above discussion, explanations therefore exist both for why the thermal original Thellier experiments overestimate the true value (multi-domain-like effects) and for why perpendicular experiments (both thermal and microwave) underestimate it (unrecognized alteration), which implies that the best estimate should fall between these two extremes.

The inherent uncertainty in the data and techniques used to obtain paleointensity estimates can and should be expressed in the stated uncertainties associated with a flow-level mean (for example, averaging the Th-OT+ and MW-Perp and providing the resulting large std. deviation). Until new, high-quality data exist for all the flows studied here from the SOH1 drill core, the authors propose that for each flow, the available flow-level Th-OT+ and MW-Perp datasets, which are the largest two, be averaged. Supporting Information D provides an example of this averaging for the flows containing at least two specimens in each Th-OT+ and MW-Perp dataset (i.e. the flows studied in the meta-analysis described in Supporting Information A). This method gives a mean paleointensity estimate of 29.2 ± 12 μT. This mean value is virtually identical to the mean MW-OT+ estimate of our newly collected data, 29.5 ± 9.2 μT, but higher than the equivalents measured by Gratton et al. (2005) and lower than those in Teanby et al. (2002). Flows characterized by data produced from only a single method (more than half; 97/181) across these studies (and the Laj et al. (2011) reanalysis) should not be considered accurate to within stated uncertainties.

6.2 Implications for future experiments

The perpendicular protocol has largely fallen out of use due to the lack of pTRM checks, but the Original Thellier protocol lives on as do the IZZI (Yu et al., 2004), Coe (ZI) (Coe, 1967), and Aitken (IZ) (Aitken et al., 1988) protocols. Modeling done in Biggin (2010) carries the implication that the Original Thellier protocol has the potential to exaggerate multi-domain behavior compared to other Thellier-style double treatment experiments, which is largely consistent with the study carried out here. We therefore suggest that any future experiments performed with the Original-Thellier protocol (or indeed the Coe or Aitken protocols), should incorporate pTRM tail checks alongside pTRM checks such that multi-domain-like effects may be detected. The IZZI protocol has the built-in advantage of allowing detection of non-ideal behavior via zig-zags in the Arai plot.

This study is consistent with previous studies (e.g. Grappone et al. (2019)) that have shown that while different methods can give seemingly reliable data, non-ideal effects (multidomain behavior, stress relaxation, undetected alteration) can be a biasing influence. We therefore concur with the finding in Grappone et al. (2019) that at least pilot specimens should be run using at least 2 Thellier-style protocols (or the IZZI protocol) as a 1st order check for non-ideal paleointensity behavior. Better still, different methods (e.g. Thermal Thellier, Microwave Thellier, or Multi-Specimen) should be employed to produce reliable multi-method paleointensity estimates (de Groot et al., 2013), which, ideally, should be internally consistent. Biggin and Paterson (2014) provided a set of quantitative criteria for evaluating the reliability of paleointensity estimates. Such paleointensity estimates are required to satisfy the TECH QPI criterion of Biggin and Paterson (2014), which requires paleointensity estimates to come from multiple techniques.

Conclusions

In this paper, we have sought to identify the cause for systematic discrepancies between previously published paleointensity studies on the SOH1 drill core. New paleointensity data confirm the systematic offset observed from previous studies when using the same methods; namely, Thermal-Original Thellier estimates were ~30% higher than Microwave-perpendicular estimates. For the first time, Thermal-Perpendicular experiments and Microwave-Original Thellier experiments were undertaken on these rocks. Our results confirm that Thermal-Original Thellier data can be too high in the presence of magnetic carriers that do not behave as non-interacting SD grains. We further confirm that perpendicular data, which lack pTRM checks for alteration, can be too low due to undetected thermochemical alteration.

Until new measurements are made using reliable methods, results previously obtained from the SOH1 drill core using different methods should be combined at the flow level. The resulting enhanced standard deviation will accurately reflect the intrinsic uncertainty associated with the mean. The potential for biasing in those flows only represented by estimates produced by only one of the previously applied methods should be recognized.

Future studies undertaken using the thermal and/or microwave demagnetization mechanisms should avoid any protocols which do not contain within them checks for both lab-induced alteration and non-ideal multi-domain- like behavior. The IZZI protocol with pTRM checks satisfies both of these criteria.

Acknowledgments and data availability

This study was carried out by J. Michael Grappone as part of a University of Liverpool match-funded studentship, with support from the Duncan Norman Research Scholarship. J Michael Grappone acknowledges support from the NERC EAO Doctoral Training Partnership, grant NE/L002469/1 and

NERC studentship 1793213. Andrew J Biggin and Courtney J Sprain acknowledge support from NERC standard grant NE/P00170X/1. Mimi J Hill acknowledges NERC grant NE/I013873/1. Courtney J Sprain thanks the Institute for Rock Magnetism at the University of Minnesota for use of their facilities to gather FORC data. The authors thank Nick Teanby for supplying the raw data from Teanby et al. (2002) and Martin Gratton for his systematic data archiving, from which we obtained the raw data from Gratton et al. (2005). All new raw data collected in this study can be found on MagIC at earthref.org/MagIC/16664. The authors also acknowledge the extensive, helpful comments of the reviewers, associate editor, and editors of JGR – Solid Earth.

References

Aitken, M. J., Allsop, A. L., Bussell, G. D., and Winter, M. B., 1988, Determination of the intensity of the earths magnetic-field during archaeological times - reliability of the thellier technique: Reviews of Geophysics, v. 26, no. 1, p. 3-12.

Biggin, A. J., 2010, Are systematic differences between thermal and microwave Thellier-type palaeointensity estimates a consequence of multidomain bias in the thermal results? : Physics of the Earth and Planetary Interiors, v. 180, no. 1-2, p. 16-40.

Biggin, A. J., and Paterson, G. A., 2014, A new set of qualitative reliability criteria to aid inferences on palaeomagnetic dipole moment variations through geological time: Frontiers in Earth Science, v. 2, no. 24.

Biggin, A. J., Piispa, E. J., Pesonen, L. J., Holme, R., Paterson, G. A., Veikkolainen, T., and Tauxe, L., 2015, Palaeomagnetic field intensity variations suggest Mesoproterozoic inner-core nucleation: Nature, v. 526, no. 7572, p. 245-248.

Biggin, A. J., Strik, G., and Langereis, C. G., 2009, The intensity of the geomagnetic field in the late-Archaean: new measurements and an analysis of the updated IAGA palaeointensity database: Earth Planets and Space, v. 61, no. 1, p. 9-22.

Cai, S., Tauxe, L., and Cromwell, G., 2017, Paleointensity From Subaerial Basaltic Glasses From the Second Hawaii Scientific Drilling Project (HSDP2) Core and Implications for Possible Bias in Data From Lava Flow Interiors: Journal of Geophysical Research: Solid Earth, v. 122, no. 11, p. 8664-8674.

Coe, R. S., 1967, Determination of paleo-intensities of earths magnetic field with emphasis on mechanisms which could cause non-ideal behavior in thelliers method: Journal of Geomagnetism and Geoelectricity, v. 19, no. 3, p. 157-179.

Coe, R. S., Gromme, S., and Mankinen, E. A., 1978, Geomagnetic paleointensities from radiocarbon-dated lava flows on hawaii and question of pacific nondipole low: Journal of Geophysical Research, v. 83, no. NB4, p. 1740-1756.

Cromwell, G., Trusdell, F., Tauxe, L., Staudigel, H., and Ron, H., 2018, Holocene Paleointensity of the Island of Hawai'i From Glassy Volcanics: Geochemistry, Geophysics, Geosystems, p. 3224-3245.

de Groot, L. V., Biggin, A. J., Dekkers, M. J., Langereis, C. G., and Herrero-Bervera, E., 2013, Rapid regional perturbations to the recent global geomagnetic decay revealed by a new Hawaiian record: Nature Communications, v. 4 Article No. 2727.

Doell, R. R., and Cox, A., 1963, The accuracy of paleomagnetic method as evaluated from historic Hawaiian lava flows: Journal of Geophysical Research, v. 68, no. 7, p. 1997-2009.

-, 1965, Paleomagnetism of Hawaiian lava flows: Journal of Geophysical Research, v. 70, no. 14, p. 3377-3405.

Grappone, J. M., Biggin, A. J., and Hill, M. J., 2019, Solving the mystery of the 1960 Hawaiian lava flow: implications for estimating Earth's magnetic field: Geophysical Journal International, v. 218, no. 3, p. 1796-1806.

Gratton, M. N., Shaw, J., and Herrero-Bervera, E., 2005, An absolute palaeointensity record from SOH1 lava core, Hawaii using the microwave technique: Physics of the Earth and Planetary Interiors, v. 148, no. 2-4, p. 193-214.

Hill, M. J., Pan, Y. X., and Davies, C. J., 2008, An assessment of the reliability of palaeointensity results obtained from the Cretaceous aged Suhongtu section, Inner Mongolia, China: Physics of the Earth and Planetary Interiors, v. 169, no. 1-4, p. 76-88.

Hill, M. J., and Shaw, J., 2000, Magnetic field intensity study of the 1960 Kilauea lava flow, Hawaii, using the microwave palaeointensity technique: Geophysical Journal International, v. 142, no. 2, p. 487-504.

-, 2007, The use of the 'Kono perpendicular applied field method' in microwave palaeointensity experiments: Earth Planets and Space, v. 59, no. 7, p. 711-716.

Hodgson, E., Grappone, J. M., Biggin, A. J., Hill, M. J., and Dekkers, M. J., 2018, Thermoremanent Behavior in Synthetic Samples Containing Natural Oxyexsolved Titanomagnetite: Geochemistry Geophysics Geosystems, v. 19, no. 6, p. 1751-1766.

Klugh, H. E., 1986, Statistics : The Essentials for Research, London, UNITED KINGDOM, Taylor & Francis Group.

Kono, M., and Ueno, N., 1977, Paleointensity determination by a modified thellier method: Physics of the Earth and Planetary Interiors, v. 13, no. 4, p. 305-314.

Kosterov, A. A., and Prevot, M., 1998, Possible mechanisms causing failure of Thellier palaeointensity experiments in some basalts: Geophysical Journal International, v. 134, no. 2, p. 554-572.

Laj, C., Kissel, C., Davies, C., and Gubbins, D., 2011, Geomagnetic field intensity and inclination records from Hawaii and the Reunion Island: Geomagnetic implications: Physics of the Earth and Planetary Interiors, v. 187, no. 3-4, p. 170-187.

Levi, S., 1977, Effect of magnetite particle-size on paleointensity determinations of geomagnetic-field: Physics of the Earth and Planetary Interiors, v. 13, no. 4, p. 245-259.

McElhinny, M. W., and Merrill, R. T., 1975, Geomagnetic secular variation over the past 5 m.y: Reviews of Geophysics, v. 13, no. 5, p. 687-708.

Paterson, G. A., Biggin, A. J., Hodgson, E., and Hill, M. J., 2015, Thellier-type paleointensity data from multidomain specimens: Physics of the Earth and Planetary Interiors, v. 245, p. 117-133.

Paterson, G. A., Muxworthy, A. R., Yamamoto, Y., and Pan, Y. X., 2017, Bulk magnetic domain stability controls paleointensity fidelity: Proceedings of the National Academy of Sciences of the United States of America, v. 114, no. 50, p. 13120-13125.

Paterson, G. A., Zhao, X., Jackson, M., and Heslop, D., 2018, Measuring, Processing, and Analyzing Hysteresis Data: Geochemistry Geophysics Geosystems, v. 19, no. 7, p. 1925-1945.

Poletti, W., Hartmann, G. A., Hill, M. J., Biggin, A. J., and Trindade, R. I. F., 2013, The cooling-rate effect on microwave archeointensity estimates: Geophysical Research Letters, v. 40, no. 15, p. 3847-3852.

Quane, S. L., Garcia, M. O., Guillou, H., and Hulsebosch, T. P., 2000, Magmatic history of the East Rift Zone of Kilauea Volcano, Hawaii based on drill core from SOH 1: Journal of Volcanology and Geothermal Research, v. 102, no. 3-4, p. 319-338.

Shaar, R., and Tauxe, L., 2013, Thellier GUI: An integrated tool for analyzing paleointensity data from Thellier-type experiments: Geochemistry Geophysics Geosystems, v. 14, no. 3, p. 677-692.

Smirnov, A. V., Kulakov, E. V., Foucher, M. S., and Bristol, K. E., 2017, Intrinsic paleointensity bias and the long-term history of the geodynamo: Science Advances, v. 3, no. 2, e1602306.

Teanby, N., Laj, C., Gubbins, D., and Pringle, M., 2002, A detailed palaeointensity and inclination record from drill core SOH1 on Hawaii: Physics of the Earth and Planetary Interiors, v. 131, no. 2, p. 101-140.

Thellier, E., and Thellier, O., 1959, Sur l'intensité du champ magnétique terrestre dans le passé historique et géologique: Ann. Géophys., v. 15, p. 285-376.

Thomas, N., 1993, An integrated rock magnetic approach to the selection or rejection of ancient basalt samples for paleointensity experiments: Physics of the Earth and Planetary Interiors, v. 75, no. 4, p. 329-342.

Xu, S., and Dunlop, D. J., 2004, Thellier paleointensity theory and experiments for multidomain grains: Journal of Geophysical Research-Solid Earth, v. 109, B07103.

Yamamoto, Y., Tsunakawa, H., and Shibuya, H., 2003, Palaeointensity study of the Hawaiian 1960 lava: implications for possible causes of erroneously high intensities: Geophysical Journal International, v. 153, no. 1, p. 263-276.

Yu, Y. J., Tauxe, L., and Genevey, A., 2004, Toward an optimal geomagnetic field intensity determination technique: Geochemistry Geophysics Geosystems, v. 5, Q02H07.

References from Supporting Information

Cisowski, S., 1981, Interacting vs non-interacting single domain behavior in natural and synthetic samples: Physics of the Earth and Planetary Interiors, v. 26, no. 1-2, p. 56-62.

Harrison, R. J., and Feinberg, J. M., 2008, FORCinel: An improved algorithm for calculating first-order reversal curve distributions using locally weighted regression smoothing: Geochemistry Geophysics Geosystems, v. 9.

Paterson, G. A., Muxworthy, A. R., Roberts, A. P., and Mac Niocaill, C., 2010, Assessment of the usefulness of lithic clasts from pyroclastic deposits for paleointensity determination: Journal of Geophysical Research-Solid Earth, v. 115.

Pike, C. R., Roberts, A. P., and Verosub, K. L., 2001, First-order reversal curve diagrams and thermal relaxation effects in magnetic particles: Geophysical Journal International, v. 145, no. 3, p. 721-730.

Roberts, A. P., Pike, C. R., and Verosub, K. L., 2000, First-order reversal curve diagrams: A new tool for characterizing the magnetic properties of natural samples: Journal of Geophysical Research-Solid Earth, v. 105, no. B12, p. 28461-28475.

Table 1 Summary information of new experiments carried out in this study

Method	Number of specimens	Number of accepted estimates	Success rate	Lab field (μT)
Microwave-OT+	48	33	69%	40
Microwave-Perp	22	13	59%	25-50
Thermal-OT+	19	10	52%	40
Thermal-Perp	31	17	55%	31
Total	120	73	61%	25-50

Number of specimens run for each method is given along with the number of specimens that gave acceptable results (number of passes) and the associated success rate. The final column gives the applied Lab fields used during the experiment for each method.

Table 2 Selection criteria

N	FRAC/f*	β	q	$\lvert\vec{K}'\rvert$	MAD_{ANC}**	α**	DRAT**	Δθ***
≥ 4	≥ 0.35	≤ 0.1	≥ 1	≤ 0.480	≤ 10	≤ 10	≤ 10%	≤ 1°

N is the number of data points, FRAC/f are measures of the NRM used, β is a measure of scatter around the best-fit line, q is a measure of the data quality, $\lvert\vec{K}'\rvert$ is a measure of Arai plot curvature, MAD and α determine the scatter of the specimen's paleodirection. Δθ is the change in the $\theta_1 + \theta_2$ value, an indication of the perpendicularity between the NRM and TRM directions (Hill and Shaw, 2007) for the perpendicular experiment. For further details the reader is referred to Paterson et al. (2015).

*FRAC is used for OT+ experiments and f for Perp experiments.
**OT+ techniques use these criteria, but Perp does not, for technical reasons
*** Used only for Perpendicular experiments

Table 3 Paleointensity results, broken down by flow and method for the combined dataset, consisting of new data, data from Teanby et al. (2002) and data from Gratton et al. (2005).

Flow	Th-OT+ (μT)	N_{pass}/ N_{tested}	MW-OT+ (μT)	N_{pass}/ N_{tested}	Th-Perp (μT)	N_{pass}/ N_{tested}	MW-Perp (μT)	N_{pass}/ N_{tested}
2	58.6	1/2	22.5 ± 4.7	2/2	29.4 ± 5.1	3/3	21.3 ± 2.0	4/4
6	50.4 ± 4.5	2/3	45.1 ± 2.8	2/2			36.5 ± 0.6	3/3
7		0/1			39.2	1/1	35.8 ± 1.6	4/4
8	59.0 ± 3.2	3/3	43.7 ± 4.9	2/2		0/1	36.5 ± 5.9	8/10
22		0/1	37.2 ± 2.6	3/3	26.3	1/2	21.3 ± 9.5	2/2
23	51.4 ± 0.9	2/2	36.7 ± 2.2	3/5	27.7 ± 0.1	2/2	28.8 ± 1.0	2/2
24			25.9	1/2		0/2	18.5 ± 4.1	2/4
26		0/1	32.5	1/2	21.2	1/3	33.1 ± 2.0	5/5
36		0/1	32.7	1/1		0/1	34.8 ± 2.8	2/2
37	54.2 ± 6.0	4/5	33.4 ± 1.6	2/2			24.8 ± 6.0	3/4
48			29.4 ± 3.4	2/2	26.4 ± 4.1	3/3	28.0 ± 0.5	2/2
87	32.2 ± 2.3	2/4		0/1			14.6 ± 1.9	4/4
91	23.8 ± 0.8	2/3		0/1			11.8 ± 0.9	4/4
163	45.9 ± 11	3/3	25.7 ± 1.5	2/3			18.6	1/3
176	26.3 ± 13	3/4				0/1		
186	25.1 ± 4.0	2/6				0/1	18.2 ± 2.6	2/2
189	27.7 ± 0.2	2/3	22.1	1/1	25.6	1/1	19.8 ± 1.0	2/2
193	31.7	1/2	18.5	1/1		0/1	15.0 ± 0.1	2/3
196	24.7 ± 1.3	2/2	15	1/1			12.6	1/3
206	40.1 ± 29	2/5	18.7	1/3	26.2	1/1		0/3
220	46.4 ± 7.9	4/6	37.2 ± 6.1	2/3	33.8	1/1	26.0 ± 4.7	7/12
221	40.9 ± 2.1	3/3	36.7	1/4	39.5 ± 2.8	2/4	31.4 ± 2.8	4/5
222	39.8 ± 3.3	3/3	33.7	1/2		0/1	37.8 ± 1.6	3/3
237	25.5 ± 8.2	3/9	14.3 ± 8.5	4/5	10.4	1/2	19.1 ± 0.2	2/4
Mean	39.0 ± 12	18/22	29.5 ± 9.2	19/21	27.8 ± 8.1	11/18	24.8 ± 8.5	22/23

Note: N_{pass} is the combined number of specimens that passed the PI selection criteria. N_{tested} is the combined number of specimens that were tested from a given flow with a given method. Empty cells indicate no experiments were attempted because of a lack of material. For the mean row, N_{pass} and N_{tested} reference the number of flows.

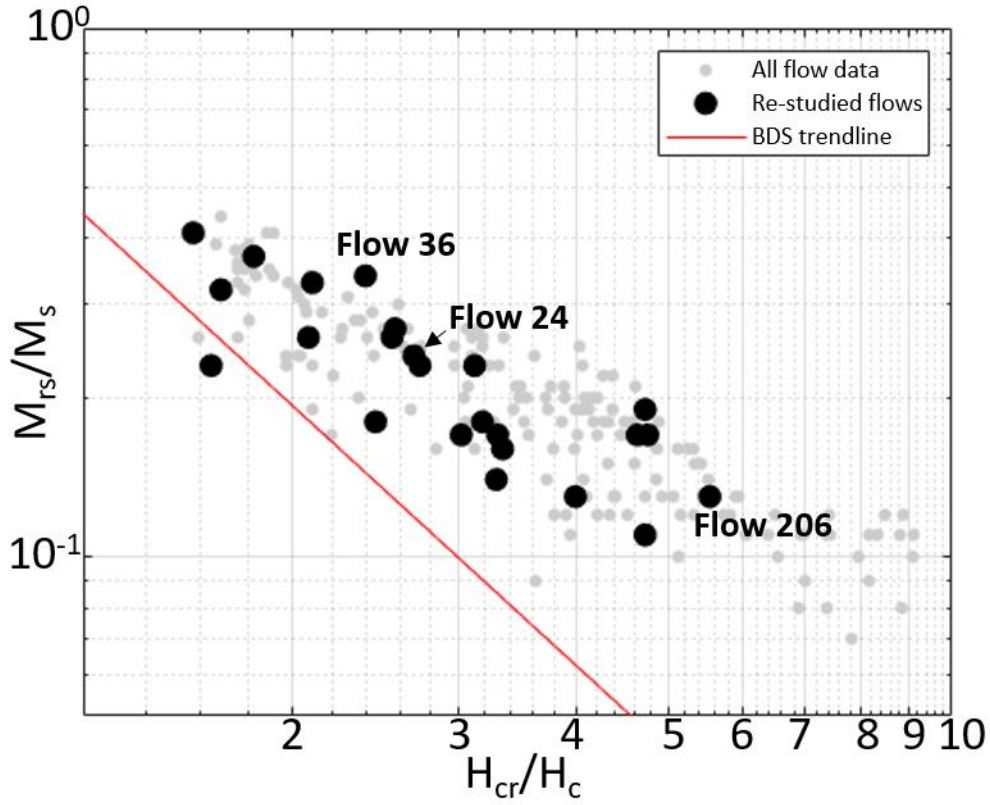

Figure 1 Hysteresis parameters plot for all the flows in the SOH1 borehole (as studied by Gratton et al., 2005), highlighting which flows were studied herein for new PI estimates. M_{rs}/M_s refers to the ratio of remanent saturation magnetization to the saturation magnetization and H_{CR}/H_C refers to the ratio of coercivity of remanence to coercivity. The Bulk Domain Stability trendline derives from the results of Paterson et al. (2017). The three named flows have new PI estimates, as well as FORC and SEM analysis in Supporting Information B.

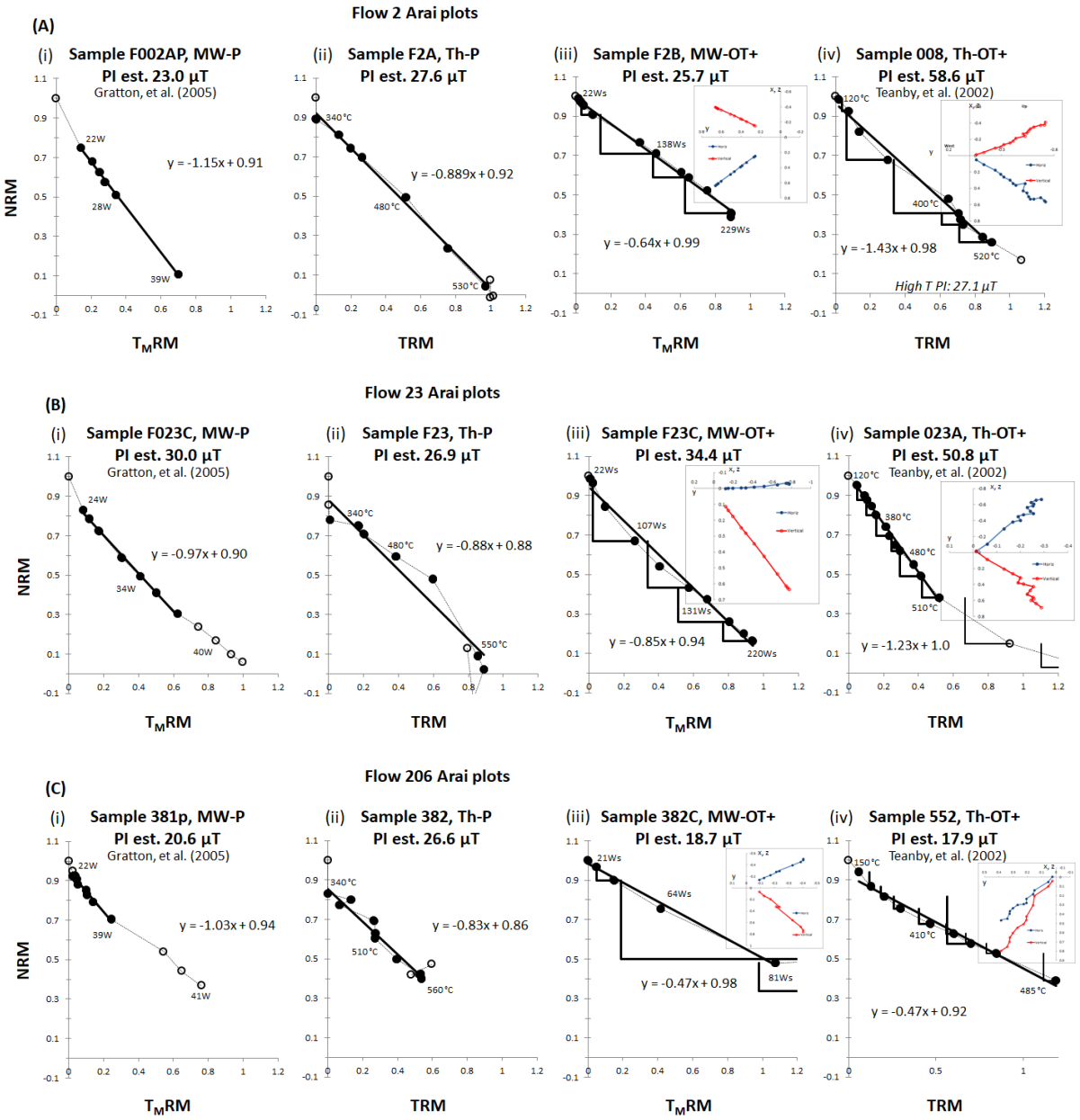

Figure 2 Normalized Arai plot examples comparing the different methods used. A) Flow 2 (503 cm thickness), B) Flow 23 (137 cm thickness), and C) Flow 206 (290 cm thickness). The data are normalized by NRM_0. The filled circles are accepted data points, with the solid black line being the best-fit line. Open circles are rejected data points. The black right-angles lines are $pT_{(M)}RM$ checks, which are only present in OT+ data. Orthogonal vector plots are provided in core coordinates for the OT+ data. All specimens presented pass their original study's selection criteria. The microwave data are visually more linear than the thermal data, and the OT+ data are often two-sloped. The powers given for the Gratton et al. (2005) MW data are power applied and the power integrals given in the new data are (inferred) energy absorbed.

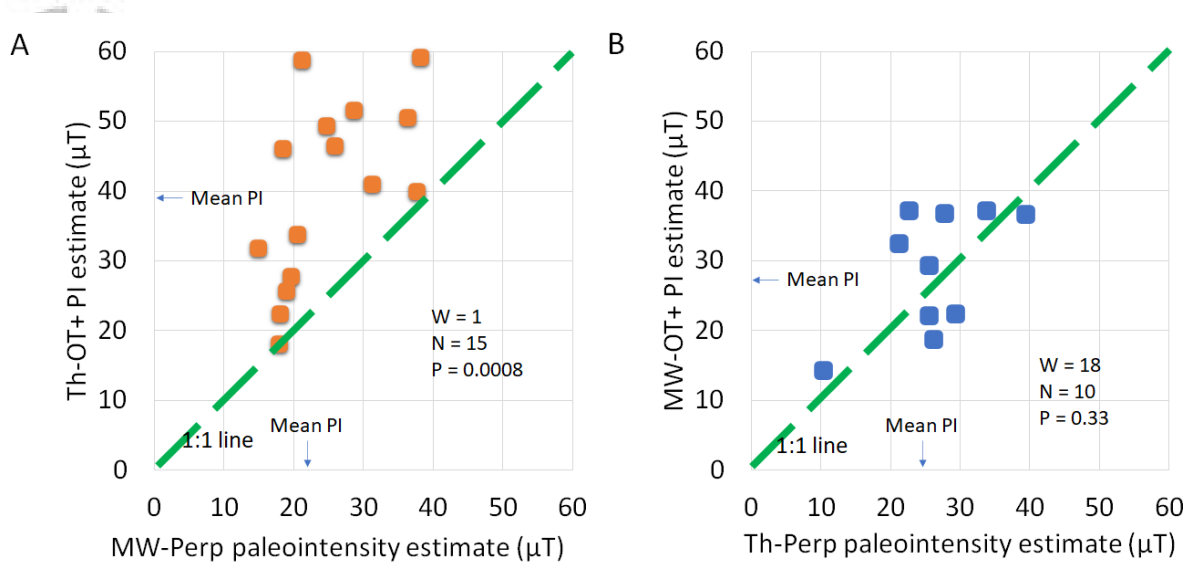

Figure 3 Testing Biggin (2010)'s hypothesis that only paleointensity protocol affects PI estimate. Flow-level paleointensity estimates are plotted against each other for completely distinct PI methods (no shared protocol or demagnetization mechanism). A: Confirming that the Th-OT+ data are higher than the MW-Perp data. B: Checking if MW-OT+ data are higher than Th-Perp. The mean PIs listed are for the flows the methods have in common. N is the number of data points, W and p are the statistics from the Wilcoxon signed rank test.

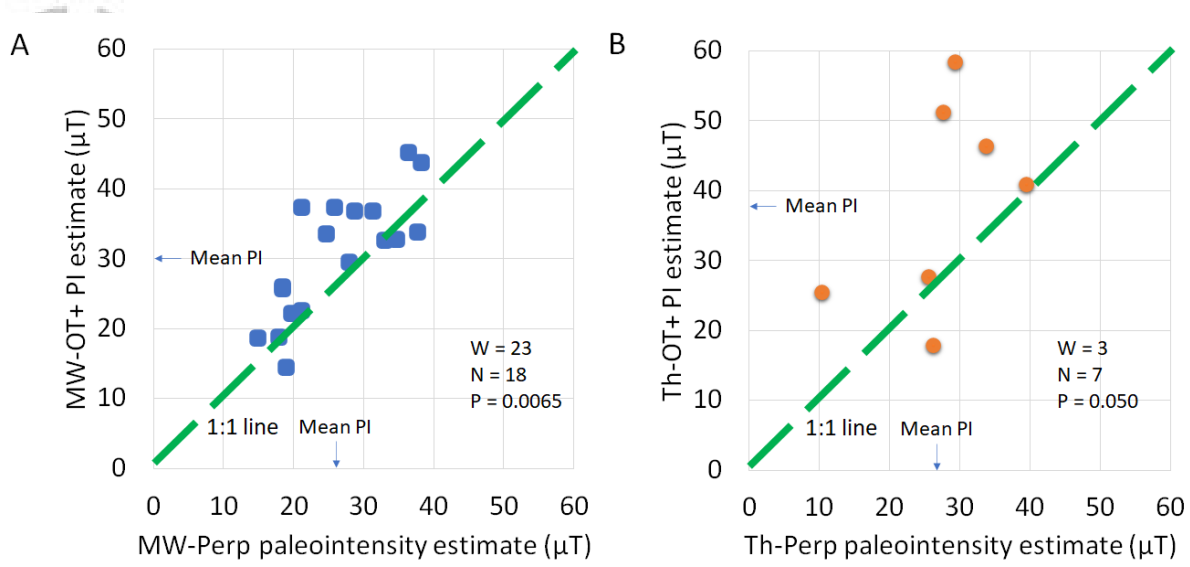

Figure 4 Testing the hypothesis that only paleointensity protocol affects PI estimate, control. Flow-level paleointensity estimates are plotted against each other for different PI protocols, separated by demagnetization mechanism. A: Microwave data; B: Thermal data. The mean PIs listed are for the flows the methods have in common. N is the number of data points, W and p are the statistics from the Wilcoxon signed rank test.

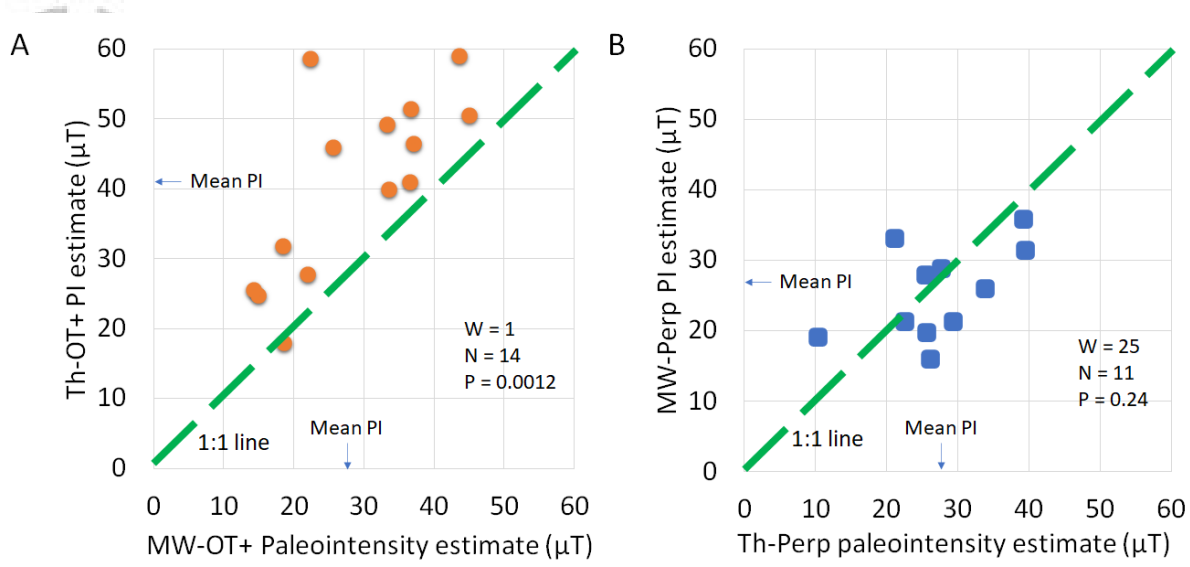

Figure 5 Testing the hypothesis that only demagnetization mechanism matters. Flow-level paleointensity estimates are plotted against each other for different PI methods, separated by protocol. A: Original Thellier data; B: perpendicular data. The mean PIs listed are for the flows the methods have in common. N is the number of data points, W and p are the statistics from the Wilcoxon signed rank test.