A new set of qualitative reliability criteria to aid inferences on palaeomagnetic dipole moment variations through geological time

Andrew J. Biggin[1]* and Greig A. Paterson[2]

[1] Geomagnetism Laboratory, Oliver Lodge Laboratories, School of Environmental Sciences, University of Liverpool, Liverpool, UK
[2] Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China

Records of reversal frequency support forcing of the geodynamo over geological timescales but obtaining these for earlier times (e.g., the Precambrian) is a major challenge. Changes in the measured virtual (axial) dipole moment of the Earth, averaged over several millions of years or longer, also have the potential to constrain core and mantle evolution through deep time. There have been a wealth of recent innovations in palaeointensity methods, but there is, as yet, no comprehensive means for assessing the reliability of new and existing dipole moment data. Here we present a new set of largely qualitative reliability criteria for palaeointensity results at the site mean level, which we term Q_{PI} in reference to the long-standing Q criteria used for assessing palaeomagnetic poles. These represent the first attempt to capture the range of biasing agents applicable to palaeointensity measurements and to recognize the various approaches employed to obviate them. A total of 8 criteria are proposed and applied to 312 dipole moment estimates recently incorporated into the PINT global database. The number of these criteria fulfilled by a single dipole moment estimate (the Q_{PI} value) varies between 1 and 6 in the examined dataset and has a median of 3. Success rates for each of the criteria are highly variable, but each criterion was met by at least a few results. The new criteria will be useful for future studies as a means of gauging the reliability of new and published dipole moment estimates.

Keywords: palaeointensity, virtual dipole moment, geomagnetic field, palaeomagnetism, rock magnetism

INTRODUCTION

The strength of the Earth's magnetic field varies on a wide range of spatial and temporal scales. These fluctuations are important in determining the degree of magnetic shielding that Earth's surface and near-Earth space receives from solar wind radiation. They also provide insight into the dynamo process occurring in Earth's outer core, which are responsible for generating most of the field observable at the surface.

We are concerned here with global-scale variations of the geomagnetic intensity occurring on the timescale of tens of millions to billions of years. Such variations may reflect changes in the forcing of the geodynamo in the outer core related to planetary evolution (Aubert et al., 2009) and mantle convection (Biggin et al., 2012). In potentially providing direct constraints on deep Earth dynamics back into deep geological time, palaeo-geomagnetic variations are geophysically unique.

The spatially and temporally inhomogeneous nature of palaeomagnetic records implies that describing variations in geomagnetic parameters on geological timescales is a non-trivial task. Nevertheless, this is essentially what has been done for decades by studies using measurements of palaeomagnetic directions to describe apparent polar wander. In such studies, it is standard practice to calculate a time-averaged *palaeomagnetic pole*, which,

under the geocentric axial dipole hypothesis, is assumed to be co-located with the geographic pole in the reference frame of the tectonic unit in question.

Palaeointensity measurements are frequently converted into *virtual dipole moments* (VDMs) or *virtual axial dipole moments* (VADMs) to normalize for the expected spatial variability from a dipolar or axial dipolar field (Merrill et al., 1983). No consistent means to normalize for the spontaneous temporal variability expected from normal dynamo operation (i.e., secular variation) yet exists however. Some attempt was made by McFadden and McElhinny (1982) who introduced the term *palaeomagnetic dipole moment* (PDM) and used it in a specific statistical sense (the mode of a collection of *true dipole moments,* themselves estimated from VDMs). The term *palaeomagnetic dipole moment* was further adopted by Tarduno et al. (2001) to refer to values of average *virtual dipole moments* (VDMs) taken from a collection of rocks whose palaeomagnetic directions suggested that secular variation had been averaged.

In this and future studies, we will not make any assumptions concerning underlying distributions of dipole moment values nor will we assume that secular variation as expressed in directional and intensity data are equivalent. Rather, we will use the term palaeomagnetic dipole moment in pure analogy

to the well-established concept of a palaeomagnetic pole. That is, it refers to an estimate of the dipole moment that is sufficiently representative of spatio-temporal variability in VDMs or VADMs to average spontaneous secular variation. Since we are implicitly assuming that such an average does exist for a dataset spanning not more than a few million years, it directly follows that any observed changes in the palaeomagnetic dipole moment must reflect external forcing of the geodynamo. Prime candidates for producing such forcing are firstly, changes in core-mantle heat flow resulting from mantle convection, and secondly, inner core nucleation and growth caused by secular changes in core temperature.

Two primary problems exist with isolating a signal of palaeomagnetic dipole moment variability:

(1) There is considerable variability in the quality of published palaeointensity estimates and significant uncertainty regarding the reliability of each of these.
(2) Even assuming their reliability, the number and spatio-temporal distribution of V(A)DM estimates required to infer a change in palaeomagnetic dipole moment is highly uncertain.

Here we will introduce a new set of reliability criteria aimed at helping to overcome the first of these major challenges (a future study will focus on the second of them). Palaeointensity measurements are generally presented alongside a swathe of statistical parameters associated with different technical aspects of the result (see http://www.paleomag.net/SPD/ for a comprehensive overview). The necessary values for these parameters, especially in relation to Thellier-type experiments (Thellier and Thellier, 1959) is an area of active and ongoing research (Paterson et al., 2012, 2014; Shaar and Tauxe, 2013). Here we are approaching the problem of palaeointensity reliability from a different angle using criteria that are largely qualitative and wider in scope.

The new criteria take their lead from a set of seven reliability quality (Q) criteria for palaeomagnetic poles that were introduced by Van der Voo (1990) and have since proved very useful for palaeomagnetic and palaeogeographical studies. The effectiveness of the Q criteria stems from the fact that they are comprehensive, straightforward to apply, and relatively uncontroversial (although several modifications have been suggested since the original publication (see e.g., Deenen et al., 2011). For each criterion that is met, the Q score is increased by one to a maximum of 7. Authors of palaeomagnetic studies frequently rate their own poles in terms of their Q scores or they are assessed afterwards by those interested to put them to further use. Subsequent users of global or regional datasets may then choose their own minimum Q score that published data must meet in order to be included in their study. It seems likely that the Q criteria have helped improve the general quality of palaeomagnetic poles by encouraging practitioners to seek out additional evidence for reliability so as to score as highly as possible.

Here we will develop and apply a new set of criteria that attempt to replicate the usefulness of the van der Voo criteria but for absolute palaeointensity estimates. Specifically, we are focusing on the problem of inferring changes in the palaeomagnetic dipole moment through time. We therefore stress that the criteria set out in Section *Q_{PI}—a New Set of Qualitative Reliability Criteria for Palaeointensity Estimates* are not intended to be applied to archeointensity data or palaeointensity data where they are used in the context of other problems where a much higher degree of precision (in result and or associated age) is required. We point out, that for palaeointensity data younger than 50 kyr, a broadly similar approach was incorporated into the GEOMAGIA50 database search interface (Korhonen et al., 2008), but no such system currently exists for older data. Section *Q_{PI} Values From a Selection of Recent Palaeointensity Studies* will report on the application of the new criteria to 312 estimates from 25 studies recently incorporated in the PINT database (older than 50 kyr).

Q_{PI}—A NEW SET OF QUALITATIVE RELIABILITY CRITERIA FOR PALAEOINTENSITY ESTIMATES

Table 1 outlines a proposed set of eight criteria for palaeointensity studies based on the same fundamental principles as the van der Voo Q criteria. These "Q_{PI}" criteria are intended to be used in precisely the same manner as the original Q criteria with one important exception. While the Q criteria are applied to the pole generated from the average direction of many palaeomagnetic sampling sites (usually at the formation or study level), the Q_{PI} criteria are designed to be applied at the level of the individual cooling unit or palaeomagnetic sampling site. Here, the average of individual specimen measurements is used to produce a VDM or VADM that is, in the case of rapidly cooled lavas and shallow intrusive bodies, a geologically instantaneous spot-reading of spatially-normalized geomagnetic intensity. In deep, larger intrusions, a single sampling site or even specimen may average many thousands of years of cooling. Nevertheless, low frequency secular variation implies that these are still unlikely to represent a palaeomagnetic dipole moment and therefore, they will not be treated differently to a V(A)DM from a lava here.

The reason that the Q_{PI} criteria are to be applied at the individual site mean level is that such palaeointensity data may be usefully employed in isolation to a far greater degree than individual site mean directions. The former may be gainfully assimilated into a global compilation (then used to estimate the PDM or infer a change in it) whereas, for the latter, multiple sites (or direction groups) must be accepted or rejected *en masse* based on whether the palaeomagnetic pole they comprise is judged reliable. This is not simply palaeointensity studies being more lenient than directional ones, but rather a difference in how the data are utilized—the nature of tectonic and palaeogeographical studies requires averaging of directional secular variation at the formation level to be of any use at all.

Palaeointensity reliability is a controversial issue so we have attempted to be as inclusive as possible in defining these criteria. Following discussions with numerous members of the international community, we feel that they sit in the middle ground of opinions incorporating as wide a range as possible. We stress the indicative and necessarily imperfect nature of these criteria. Each criterion is not equal and some will carry particular weight in certain instances. There may be some clear-cut cases where precision

Table 1 | Summary of proposed Q$_{PI}$ criteria.

	Code	Criterion
1	AGE	A reliable (if approximate) age and palaeomagnetic behavior consistent with PI derived from a primary component of remanence.
2	STAT	A minimum of 5 individual sample estimates per unit with low dispersion (true SD/mean $\leq 25\%$; Paterson et al., 2010).
3	TRM	Reasonable independent (e.g., microscopic) evidence that the component of remanence in the bulk of samples is likely a thermoremanent magnetization (TRM).
4	ALT	Reasonable evidence (e.g., pTRM checks or rock mag) that the final estimate was not significantly biased by alteration occurring during the experiment.
5	MD	Reasonable evidence (e.g., high f or pTRM tail checks) that the final estimate was not significantly biased by multidomain behavior during the experiment.
6	ACN	Reasonable evidence that the final estimate was not significantly biased by anisotropy of TRM, cooling rate effects, and non-linear TRM effects.
7	TECH	Estimate is an average of results from more than one palaeointensity technique.
8	LITH	Estimate is an average of results from more than one lithology or from samples from the same lithology showing significantly different unblocking behavior.

or even accuracy is uncorrelated with the Q$_{PI}$ value. Where we hope least controversy exists, is in the argument that additional analyses that raise a Q$_{PI}$ value can only help improve the reliability of a result.

To avoid situations whereby demonstrably unreliable palaeointensity data have unjustifiably high Q$_{PI}$ values, we propose a special condition in certain cases. When a new palaeointensity estimate is published as wholly unreliable (say for example because a unit demonstrably does not retain a primary TRM), then the Q$_{PI}$ should be automatically set at zero. Similarly, if unequivocal new evidence comes to light that a published palaeointensity estimate is likely to be grossly biased, and this is itself published in the peer-reviewed literature, then, again, Q$_{PI}$ for the original estimate should be set to zero. The authors have set up a website (http://qpi.wikispaces.com/) for the purpose of allowing members of the community to view, post, confirm, discuss, dispute, and update individual Q$_{PI}$ values. The intention is for the values given on this website to be the most up-to-date available and therefore to supercede those given in previous publications (including this one).

The rationale and practical application of each of the individual criteria in turn are briefly set out below. In general, there are several ways that each criterion could be met and the discussion below is not intended to be exhaustive. In Section *Q$_{PI}$ Values From a Selection of Recent Palaeointensity Studies*, some examples of real results that pass and fail each of the criteria will be given.

The AGE criterion assesses whether the associated absolute age estimate, remanence component structure, and palaeomagnetic direction (if it exists) are consistent with a reliable and useful palaeointensity. For the vast majority of results, radiometric or stratigraphic age estimates will be sufficiently precise for tracking PDM variations between geological periods requiring only a resolution of several tens of millions of years. If measured, palaeomagnetic directions should be consistent with the palaeointensity being derived from a component of remanence that is the same age as the rock itself. Similarly, the component structure should be sensible for the age of the rock and coherent with sister units. Results that do not have palaeomagnetic directions directly associated with them (e.g., unoriented ocean cores) can still meet this criterion so long as the other requirements are met.

STAT is the sole quantitative criterion and is based on the premise that there is no good (i.e., geomagnetic) reason for specimens from the same cooling unit or sampling site to produce inconsistent palaeointensity estimates. Furthermore, there is a need to have a sufficient number of specimens to test whether this consistency is reasonable and to have a moderate precision. Note that the maximum cutoff for the measured standard deviation ratio is N dependent and somewhat lower than 25% (e.g., 16% for $N = 5$). The values can be read off Table 3 in Paterson et al. (2010) or determined precisely using the equation for $\delta B_N(\%)$ in SPD (http://www.paleomag.net/SPD/). Note also that we choose to specify only a relative cutoff for the standard deviation and not an absolute value (e.g., $5\,\mu T$) as others have occasionally done (Shaar and Tauxe, 2013). Use of an absolute cut-off increases leniency for weak palaeointensities and is unwarranted unless the laboratory field is set inappropriately high. The values chosen imply that, for a Gaussian population of palaeointensity measurements distributed about the true value (i.e., suitably screened for biasing factors), an estimate just meeting STAT will have a 90% probability of being within 20% of the true value and a 60% probability of being within 10%.

The TRM criterion primarily recognizes the usefulness of a microscopic analysis in practically all palaeointensity studies. Petrography can rarely uniquely discriminate between remanence type, but can effectively rule out obvious sources of non-thermal remanence acquisition (e.g., hydrothermal precipitation or alteration of opaques and low temperature oxidation). Evidence should be presented for a primary igneous texture with opaque grains attributable to a process of formation that likely occurred at high temperature and not subject to significant post-cooling alteration. Note that this criterion can be met at a formation level (from microscopic analysis of representative and consistent units) and then applied to each result from that formation. Note also that there are other approaches than microscopy to meeting this criterion and some of these will be discussed in the next section. The key is that reasonably convincing evidence is presented that remanence is of thermal origin.

Heating induced alteration is a major threat to the accuracy of palaeointensity measurements (Thellier and Thellier, 1959) that is exceptionally difficult to demonstrably avoid even using a controlled atmosphere (Zhang et al., 2012; Zhao et al., 2014). The ALT criterion recognizes reasonable efforts to guard against significant bias from it. In Thellier-type experiments,

this is often achieved with integrated repeat partial thermoremanent magnetization (pTRM) checks. Other techniques (e.g., the Multispecimen parallel differential pTRM and Shaw methods; Shaw, 1974; Dekkers and Böhnel, 2006), generally rely on comparison of separate measurements of sample magnetic properties (e.g., anhysteretic remanent magnetization) made before and after heating the specimen or its sister specimen. Either of these approaches will satisfy this criterion and any technique that involved no heating of the sample whatsoever (e.g., Muxworthy and Heslop, 2011) would qualify automatically. Note that we make no attempt to designate a numerical requirement for the allowed level of disagreement in each case except that some cut-off must be specified and this should not lie well outside of the range of published values. Studies that appear to successfully correct for alteration using peer-reviewed techniques (e.g., Rolph and Shaw, 1985; McClelland et al., 1996) will also automatically meet this criterion.

The presence of multidomain particles and magnetostatic interactions is well-known to be deleterious to Thellier-type palaeointensity experiments in particular (Levi, 1977). The MD criterion is passed if the result is derived from an experimental approach that is considered to be domain-state independent (e.g., the domain-state-corrected multispecimen, LTD-DHT Shaw, and Wilson methods; Yamamoto et al., 2003; Fabian and Leonhardt, 2010; Muxworthy, 2010) or where the absence of domain-state bias is demonstrated. In Thellier-type experiments, this is often achieved through obtaining positive pTRM tail checks (Riisager and Riisager, 2001), quantitative identification of Arai plot curvature (Paterson, 2011), or by obtaining IZZI protocol results (Tauxe and Staudigel, 2004) that are not zigzagged. Note that bulk hysteresis measurements that fall in the "pseudo-single domain" portion of a Day et al. plot should not be considered sufficient to meet this criterion. This is because there is now a substantial body of evidence that supports that such samples may still be biased by domain-state related issues (Calvo et al., 2002; Carlut and Kent, 2002; Biggin and Thomas, 2003; Böhnel et al., 2003; Coe et al., 2004; Chauvin et al., 2005; Biggin et al., 2007; Michalk et al., 2008, 2010).

The criterion ACN is intended to encompass three separate issues each with the potential to cause significant bias to a palaeointensity result, but representing threats that can be corrected for and that are far from universal in the context of palaeointensity studies typically used to describe PDM variations. Unlike for archeointensity studies, each of these issues may commonly be dealt with in a relatively trivial manner supporting their combination into a single criterion. The criterion is judged to have been met if all three issues are dealt with and failed if one or more of them is not.

The first of these issues is anisotropy of TRM which rarely affects igneous targets, but is generally much more of a consideration for archeomagnetic materials, which these criteria are not designed for. Practically all palaeointensity methods involve the impartation of a partial or full TRM. In the case where the directions of these components of magnetization are observed to be sub-parallel to the applied field direction, this first part of the criterion has been met. A simple, quantitative check that can be determined from all studies is the use of γ, the angle between the

pTRM acquired at the last step used for the best-fit segment and the applied field direction (see the SPD for details). If this is not the case, then a correction made using the calculated anisotropy tensor would be necessary to address this issue (Veitch et al., 1984; Selkin et al., 2000).

The second component of the ACN criterion pertains to the effect of cooling rate on TRM intensity and the bias this can produce to palaeointensity estimates (Fox and Aitken, 1980). Recent studies have strongly suggested that this is likely to only be a serious source of bias for materials containing assemblages of near ideal single domain grains (Yu, 2011; Biggin et al., 2013; Ferk et al., 2014) or else for large intrusions that cool over many hundreds of kyr. In such cases, an appropriate cooling rate correction should be applied; otherwise, this issue is considered to have been addressed automatically.

The non-linear dependence of TRM on applied field has been demonstrated to be a potential significant source of bias for certain assemblages of ferromagnetic grains (Selkin et al., 2007). The effects are minimal when the laboratory and ancient field strengths are approximately equal (Selkin et al., 2007; Paterson, 2013). Such results are considered to automatically have dealt with this issue. For most typical geological materials (i.e., lavas) if the two fields are within a multiple of ~1.5 times each other, then the influence of non-linear TRM is likely to be minimal (the uncorrected average result should be within ~20% of the expected value; Paterson, 2013) and the issue is viewed as having been dealt with. If, however, the material used has the potential for a strong magnetic fabric (e.g., through metamorphic deformation) it is likely affected by non-linear TRM and should be corrected. Selkin et al. (2007) outlined a correction to be performed where bias might exist, but even this may be unnecessary if the study can demonstrate a lack of apparent sensitivity of the measured PI to applied field strength (i.e., sister specimens from experiments with different laboratory fields yielding the same palaeointensities).

This general approach can also be applied to the other two components of this criterion; that is, if palaeointensities from sister samples can be demonstrated to exhibit no dependence on applied field direction and/or cooling rate in different experiments then the respective issue can also be considered to have been dealt with without recourse to any deliberate checks.

The TECH criterion recognizes the fact that while there may be considerable disagreement between different groups over preferred experimental techniques, few would disagree that the addition of complementary results from a different technique helps to increase the reliability even further. For this purpose, independent "techniques" should be substantially different from one another and not simply different protocols of the same method. For example, the Coe-modified version of the Thellier method is not considered independent of the original Thellier version of the experiment. We point out that this criterion does not require any level of agreement between the results of the different experiments as this element of reliability is dealt with by the STAT criterion.

The final criterion is arguably the most difficult to achieve. LITH is met in those cases where materials with substantially different magnetomineralogy and/or domain state record the

same cooling event. In such cases, even disagreement can help to produce a more reliable outcome (Biggin et al., 2007). Some examples of where this criterion may be met include baked sediments vs. igneous rocks and chilled margins of intrusive rocks vs. massive interiors. To qualify, the magnetic mineralogy or the normalized unblocking behavior of the magnetization of the end-members should differ substantially (i.e., by several tens of percent at a single temperature).

TECH and LITH could occasionally lead to a situation whereby a study is rewarded with a higher Q_{PI} value for including data that is apparently of lower technical quality (e.g., failing ALT and MD), but from a diverse source. This is not necessarily problematic as the correlation between accuracy and technical quality is seldom clear-cut (Biggin et al., 2007). In other cases, while adding these specific criteria, inclusion of poorer quality data may lead to the sacrifice of others (e.g., STAT or ALT) thus helping to produce a more balanced and overall representative Q_{PI} value.

Q_{PI} VALUES FROM A SELECTION OF RECENT PALAEOINTENSITY STUDIES

Table 2 summarizes the results of applying the Q_{PI} criteria to the 312 palaeointensity estimates reported in the 25 publications recently used to update the PINT palaeointensity database (updates 2012.08 and 2014.01). The entire database is publicly available at http://earth.liv.ac.uk/pint/, the list of references for the **Table 2** are given in Supplementary Table 1, Supplementary Table 2 gives the complete breakdown of Q_{PI} values by estimate and will be shortly uploaded to http://qpi.wikispaces.com for commenting and modification by the community.

The Q_{PI} values in **Table 2** do not give any indication of the importance or rigor of the study—only the reliability of the data, as presented, according to the criteria set out in Section QPI—a New Set of Qualitative Reliability Criteria for Palaeointensity Estimates. Every attempt was made to be as objective and consistent as possible in deciding whether a given estimate fulfilled each criterion.

Table 2 | Breakdown of criteria fulfillment and Q_{PI} values determined for results published in 25 studies recently added to the absolute palaeointensity (PINT) database.

PINT REF	N	Ages (Ma)	Method	AGE	STAT	TRM	ALT	MD	ACN	TECH	LITH	Q_{PI}
669	1	14.6	T+	1	1	1	1	1	0	0	1	6
670	8	1786	T+	1/0	0/1	0/1	1	1	0	0	0	3/2/5
671	5	0.485	T+/ MSPDp/W	1	0/1	0	1	1/0	1/0	1/0	0	5/4
672	3	119.3	T+	1	0/1	0	1	1	0	0	0	3/4
673	7	50–117	T+/W	1/0	0/1	1	1	1	0	1	0	5/4
674	3	0.15–0.3	T−	1	0	0	0	0	0	0	0	1
675	6	8.3–10.95	T+/W	1	1/0	0	1	1	0	1	0	5/4
676	9	1.99–2.7	T+	1	0/1	0	1	1/0	0	0	0	3/2/4
677	5	132	T+	1	0	0	1	0	0	0	0	2
678	18	0.771	LTD-DHT-S	1	0/1	0	1	1	0	0	0	3/4
679	3	250	T+	1	0/1	0	1	1	0	0	0	3/4
680	112	0.0514–0.4031	T+	1	0	0	1	0/1	0	0	0	2/3
681	26	110–130	T+/W	1	0/1	0	1	1/0	0	1/0	0	4/2/3/5
682	21	0.78–0.862	T+	1	0/1	1	1	0/1	0	0	0	4/3/5
683	3	141	T+	1	0/1	0	1	1/0	0	0	0	2/3/4
684	8	3.5–3.73	T+	1	0	0	1	1	0	0	0	3
685	4	0.1438–0.2925	T+	1	0	0	1	1	0	0	0	3
686	1	211	T+	1	0	1	1	1	0	0	1	5
687	30	1087	T+	1	0/1	1	1	1	0	0	0	4/5
688	2	0.0634–0.0696	T+	1	0	0	1	0	0	0	0	2
689	4	0.089–0.266	LTD-DHT-S	1	1	1	1	1	0	0	1/0	6/5
690	1	2784	T+	1	1	1	1	1	0	0	0	5
691	22	250	T+/W	1	0/1	1	1	1	0	1/0	0	5/6/4
692	6	167	T+	1	0/1	1	1	1	0	0	0	4/5
693	4	187	T+	1	0	0	1	0/1	0	0	0	2/3

Note that this table is for the summary purposes only—Q_{PI} values are associated with individual estimates and may be subject to change (see http://qpi.wikispaces.com/). PINT REF relate to the list given in Supplementary Table 1 and also at http://earth.liv.ac.uk/pint/ with hyperlinks to those studies in the MagIC database. The complete breakdown of Q_{PI} values by estimate is given in Supplementary Table 2. N refers to the number of site mean estimates in PINT. Techniques are: Thellier-type methods with or without pTRM checks (T+, T−; Thellier and Thellier, 1959), multispecimen parallel differential pTRM method (MSPDp; Dekkers and Böhnel, 2006), modified Shaw method with low temperature demagnetization (LTD-DHT-S; Yamamoto et al., 2003), and the Wilson method (W; Wilson, 1962). A "1" indicates the given criterion was judged to have been met while a "0" indicates failure. Where both appear, this indicates that some estimates passed and others failed with the proportion indicated by the order. Multiple Q_{PI} values are also given in the order that reflects the relative proportions.

Figure 1 summarizes the breakdown of Q_{PI} values by criterion. The values fell in the range 1–6 out of a possible 8 and the median value was 3. Success rates were highly variable between criteria: the most frequently met criteria were AGE and ALT (both 99%) whereas the least frequently met were ACN (1%) and LITH (2%).

In the few cases where AGE was not met, this was either because the associated palaeomagnetic direction was anomalous and therefore suspected to reflect a remagnetization (Donadini et al., 2011; PINT REF 670) or where the age constraints were unusually weak (dating based on petrological similarities to distant intrusive units; e.g., Shcherbakova et al., 2012; PINT REF 673). The addition of reliable dates in the future could lead to such values being increased.

Only 29% of the site mean data were derived from fewer than 5 individual sample estimates and approximately half of these exhibited the required amount of precision to pass the STAT criterion. This suggests that a significant increase in the number of estimates meeting this criterion could be achieved if studies simply measured more samples per site or avoided the use of overly strict selection criteria (Paterson et al., 2012). The main barrier to achieving this is the generally high failure rate of palaeointensity experiments and the very time consuming nature of the popular Thellier method. Where overprint magnetizations are small, a

suggested work-around to this problem is to perform the MSPD protocol (Dekkers and Böhnel, 2006) or another rapid multi-specimen technique (e.g., Hoffman and Biggin, 2005) alongside the favored method.
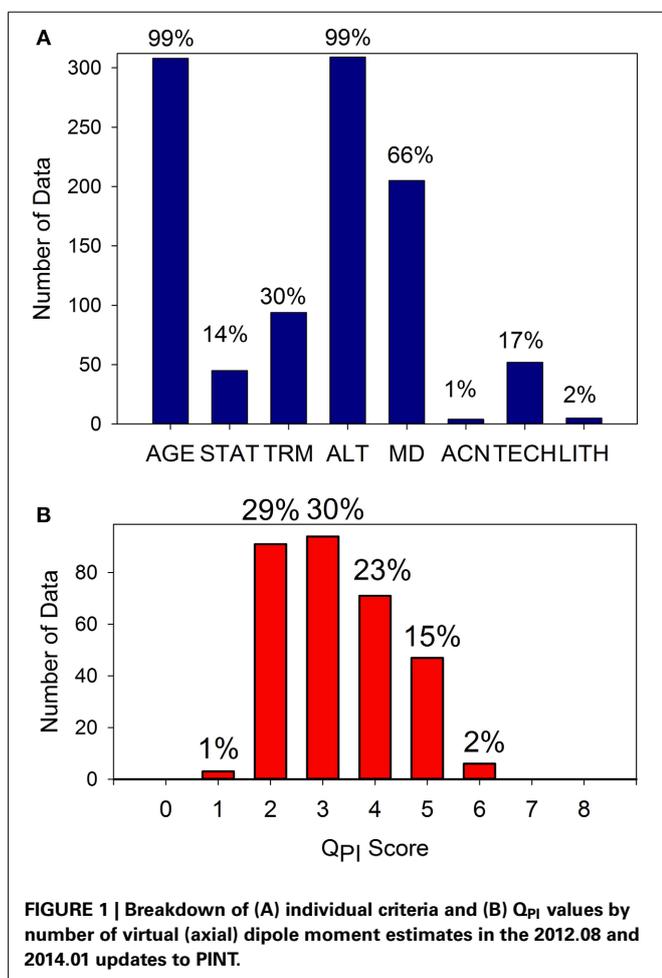
The relatively low proportion of estimates (30%) meeting the TRM criteria largely reflects the absence of microscopic analyses in most studies. In one study (Muxworthy et al., 2013; PINT REF 690), the criterion was judged to be met on the basis of rock magnetic analyses alone because the remanence was convincingly shown to be held by inclusions in silicate minerals, which, in general, will very likely have acquired a primary TRM. In another case (Donadini et al., 2011; PINT REF 670), a microscope analysis was performed, but this revealed the existence of some secondary iron-oxide grains leading to failure of this criterion. Finally, the study of Shcherbakova et al. (2012; PINT REF 673) only provided microscopy information for one locality, but argued for a thermal origin of remanence for all localities on the basis of strong similarities in the measured shapes of continuous thermal demagnetization and remagnetization curves. This is not as strong a constraint as dedicated microscopy analyses, but nevertheless was judged sufficient to fulfill the TRM criterion.

The high success rate in the ALT criterion reflects a widespread recognition of the importance of pTRM checks or some other alteration test in modern palaeointensity studies. A large number of legacy palaeointensity data in PINT would not meet this criterion because, prior to the 1990s, pTRM checks were not routinely included in Thellier-type analyses.

The MD criterion was met by the majority of results (66%). In most cases, this reflected the inclusion of pTRM tail checks or the use of the domain state independent Wilson technique (Muxworthy, 2010). However, in other cases where the Thellier-type methods were used without specific checks, the criterion was met in some cases by the Arai plot of some estimates being linear over a demonstrably large ($\geq$70%) fraction of the NRM unblocking range (Morales et al., 2003; Calvo-Rathert et al., 2011; Laj et al., 2011; Shcherbakova et al., 2011, 2012; PINT REFs 676, 680, 693, 681, 673, respectively). Whenever the MD criterion was fulfilled by Thellier results with large fraction values or by Wilson results, these comprised at least half of the total number of sample estimates for that site mean or else produced measurements within the range of the other sample estimates for that site mean.

The low success rate with respect to the ACN criterion reflects that it is not standard practice for anisotropy of TRM to even be mentioned in palaeointensity publications except where it has been recognized as a problem to be dealt with (Selkin et al., 2000). This percentage could be dramatically increased in the future simply by studies documenting that the angular deviation between the applied field direction and the resulting remanence (γ) is not significant. This information is generally not provided, even for the few representative samples typically shown in figures, but is available in all studies.

The TECH criterion was met by only 17% of studies reflecting a predominance of the Thellier-type methods used in isolation. In addition to increasing the likelihood of meeting STAT, undertaking rapid multi-specimen techniques, where appropriate, in conjunction with this approach could also dramatically improve success in this criterion.

FIGURE 1 | Breakdown of (A) individual criteria and (B) Q_{PI} values by number of virtual (axial) dipole moment estimates in the 2012.08 and 2014.01 updates to PINT.

Finally, although only 5 estimates (2%) met the demanding LITH criterion, these were drawn from 3 different studies. Two of these (Koch et al., 2012; Eitel et al., 2014; PINT REF 669, 686) were based on single meteorite craters whose large areas afforded a variety of distal materials recording the same cooling event. A third study (Mochizuki et al., 2013; PINT REF 689) combined consistent results from glass and crystalline samples from the same cooling units.

DISCUSSION AND CONCLUSIONS

It is encouraging that every criterion outlined in Section Q_{PI} *Values from a Selection of Recent Palaeointensity Studies* is met by several of the studies included in the latest database update. Furthermore, while a comprehensive assessment of the entire contents of PINT has not yet been performed, it is apparent that the recent update contains a great deal more data meeting certain criteria (ALT in particular, but also probably MD, TRM, and others) than older studies do. This is also encouraging and indicates that, as our understanding of possible sources of bias to palaeointensity determination has advanced, so has the inclusion of techniques to detect these. We hope that the formal recognition of these approaches in Q_{PI} will help accelerate this process. Such techniques may not be foolproof but their incorporation into any study generally does increase confidence.

It is also clear that there is great potential for many of these criteria to be met much more frequently in the future if researchers are convinced of the value of doing so. The criteria STAT, TRM, MD, ACN, and TECH all lend themselves to being increasingly fulfilled by future studies that incorporate further analyses or even just report certain aspects of the analysis they have already performed.

It is now more than 75 years since the first absolute palaeointensity measurements were made from geological materials (Königsberger, 1938), but the reliability of virtually all data, new and old, remains controversial. Here we have outlined and applied a modified version of an approach that has helped reduce similar problems in the study of palaeomagnetic poles. In forthcoming studies, these criteria will be applied by the authors to assess the reliability of both their own data and selections from databases used to infer whether and when changes in the PDM occurred. The authors plan to publish these new Q_{PI} values alongside the studies in question and hope that others in the community will publish values that they have themselves calculated for new and existing datasets.

In deriving Q_{PI}, we have attempted to find the path of minimal controversy and maximal usefulness while acknowledging the imperfection of any such set of criteria. We repeat that published Q_{PI} values are not intended to reflect the value of a study and, in some cases, may do a poor job of describing its reliability, but they do allow for a more balanced comparison of a diverse range of studies. We would welcome any future attempt to improve on Q_{PI} in providing an objectively based means of comprehensively assessing reliability of palaeointensity data.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/feart.2014.00024/abstract

REFERENCES

Aubert, J., Labrosse, S., and Poitou, C. (2009). Modelling the palaeo-evolution of the geodynamo. *Geophys. J. Int.* 179, 1414–1428. doi: 10.1111/j.1365-246X.2009.04361.x

Biggin, A. J., Badejo, S., Hodgson, E., Muxworthy, A. R., Shaw, J., and Dekkers, M. J. (2013). The effect of cooling rate on the intensity of thermoremanent magnetization (TRM) acquired by assemblages of pseudo-single domain, multidomain and interacting single-domain grains. *Geophys. J. Int.* 193, 1239–1249. doi: 10.1093/gji/ggt078

Biggin, A. J., Perrin, M., and Dekkers, M. J. (2007). A reliable absolute palaeointensity determination obtained from a non-ideal recorder. *Earth Planet. Sci. Lett.* 257, 545–563. doi: 10.1016/j.epsl.2007.03.017

Biggin, A. J., Steinberger, B., Aubert, J., Suttie, N., Holme, R., Torsvik, T. H., et al. (2012). Possible links between long-term geomagnetic variations and whole-mantle convection processes. *Nat. Geosci.* 5, 526–533. doi: 10.1038/ngeo1521

Biggin, A. J., and Thomas, D. N. (2003). The application of acceptance criteria to results of Thellier palaeointensity experiments performed on samples with pseudo-single-domain-like characteristics. *Phys. Earth Planet. Inter.* 138, 279–287. doi: 10.1016/S0031-9201(03)00127-4

Böhnel, H., Biggin, A. J., Walton, D., Shaw, J., and Share, J. A. (2003). Microwave palaeointensities from a recent Mexican lava flow, baked sediments and reheated pottery. *Earth Planet. Sci. Lett.* 214, 221–236. doi: 10.1016/S0012-821X(03)00370-4

Calvo, M., Prevot, M., Perrin, M., and Riisager, J. (2002). Investigating the reasons for the failure of palaeointensity experiments: a study on historical lava flows from Mt. Etna (Italy). *Geophys. J. Int.* 149, 44–63. doi: 10.1046/j.1365-246X.2002.01619.x

Calvo-Rathert, M., Goguitchaichvili, A., Bogalo, M. F., Vegas-Tubia, N., Carrancho, A., and Sologashvili, J. (2011). A paleomagnetic and paleointensity study on Pleistocene and Pliocene basaltic flows from the Djavakheti Highland (Southern Georgia, Caucasus). *Phys. Earth Planet. Inter.* 187, 212–224. doi: 10.1016/j.pepi.2011.03.008

Carlut, J., and Kent, D. V. (2002). Grain-size-dependent paleointensity results from very recent mid-oceanic ridge basalts. *J. Geophys. Res.* 107, EPM 2-1–EPM 2-12. doi: 10.1029/2001JB000439

Chauvin, A., Roperch, P., and Levi, S. (2005). Reliability of geomagnetic paleointensity data: the effects of the NRM fraction and concave-up behavior on paleointensity determinations by the Thellier method. *Phys. Earth Planet. Inter.* 150, 265–286. doi: 10.1016/j.pepi.2004.11.008

Coe, R. S., Riisager, J., Plenier, G., Leonhardt, R., and Krasa, D. (2004). Multidomain behavior during Thellier paleointensity experiments: results from the 1915 Mt. Lassen flow. *Phys. Earth Planet. Inter.* 147, 141–153. doi: 10.1016/j.pepi.2004.01.010

Deenen, M. H. L., Langereis, C. G., van Hinsbergen, D. J. J., and Biggin, A. J. (2011). Geomagnetic secular variation and the statistics of palaeomagnetic directions. *Geophys. J. Int.* 186, 509–520. doi: 10.1111/j.1365-246X.2011.05050.x

Dekkers, M. J., and Böhnel, H. N. (2006). Reliable absolute palaeointensities independent of magnetic domain state. *Earth Planet. Sci. Lett.* 248, 508–517. doi: 10.1016/j.epsl.2006.05.040

Donadini, F., Elming, S. A., Tauxe, L., and Halenius, U. (2011). Paleointensity determination on a 1.786 Ga old gabbro from Hoting, Central Sweden. *Earth Planet. Sci. Lett.* 309, 234–248. doi: 10.1016/j.epsl.2011.07.005

Eitel, M., Gilder, S. A., Kunzmann, T., and Pohl, J. (2014). Rochechouart impact crater melt breccias record no geomagnetic field reversal. *Earth Planet. Sci. Lett.* 387, 97–106. doi: 10.1016/j.epsl.2013.11.021

Fabian, K., and Leonhardt, R. (2010). Multiple-specimen absolute paleointensity determination: an optimal protocol including pTRM normalization, domain-state correction, and alteration test. *Earth Planet. Sci. Lett.* 297, 84–94. doi: 10.1016/j.epsl.2010.06.006

Ferk, A., Leonhardt, R., Hess, K. U., Koch, S., Egli, R., Krasa, D., et al. (2014). Influence of cooling rate on thermoremanence of magnetite grains: identifying the role of different magnetic domain states. *J. Geophys. Res. Solid Earth* 119, 1599–1606. doi: 10.1002/2013JB010845

Fox, J. M. W., and Aitken, M. J. (1980). Cooling-rate dependence of thermoremanent magnetisation. *Nature* 283, 462–463. doi: 10.1038/283462a0

Hoffman, K. A., and Biggin, A. J. (2005). A rapid multiple-sample approach to the determination of absolute paleointensity. *J. Geophys. Res. Solid Earth* 110:B12108. doi: 10.1029/2005JB003646

Koch, S. A., Gilder, S. A., Pohl, J., and Trepmann, C. (2012). Geomagnetic field intensity recorded after impact in the Ries meteorite crater, Germany. *Geophys. J. Int.* 189, 383–390. doi: 10.1111/j.1365-246X.2012.05399.x

Königsberger, J. G. (1938). Natural resisidual magnetism of eruptive rocks (part 2). *Terr. Magn. Atmosph. Electric.* 43, 119–130. doi: 10.1029/TE043i002p00119

Korhonen, K., Donadini, F., Riisager, P., and Pesonen, L. J. (2008). GEOMAGIA50: an archeointensity database with PHP and MySQL. *Geochem. Geophys. Geosyst.* 9, Q04029. doi: 10.1029/2007GC001893

Laj, C., Kissel, C., Davies, C., and Gubbins, D. (2011). Geomagnetic field intensity and inclination records from Hawaii and the Reunion Island: geomagnetic implications. *Phys. Earth Planet. Inter.* 187, 170–187. doi: 10.1016/j.pepi.2011.05.007

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

McClelland, E., Muxworthy, A. R., and Thomas, R. M. (1996). Magnetic properties of the stable fraction of remanence in large multidomain (MD) magnetite grains: single-domain or MD? *Geophys. Res. Lett.* 23, 2831–2834. doi: 10.1029/96GL01492

McFadden, P. L., and McElhinny, M. W. (1982). Variations in the geomagnetic Dipole-2 - statistical-analysis of Vdms for the past 5 million years. *J. Geomagn. Geoelectr.* 34, 163–189. doi: 10.5636/jgg.34.163

Merrill, R. T., McElhinny, M. W., and McFadden, P. (1983). *The Magnetic Field of the Earth*. London: Academic Press.

Michalk, D. M., Biggin, A., Knudsen, M. F., Böhnel, H. N., Nowaczyk, N., Ownby, S., et al. (2010). Application of the multispecimen palaeointensity method to Pleistocene lava flows from the Trans-Mexican Volcanic Belt. *Phys. Earth Planet. Inter.* 179, 139–156. doi: 10.1016/j.pepi.2010.01.005

Michalk, D. M., Muxworthy, A. R., Bohnel, H. N., Maclennan, J., and Nowaczyk, N. (2008). Evaluation of the multispecimen parallel differential pTRM method: a test on historical lavas from Iceland and Mexico. *Geophys. J. Int.* 173, 409–420. doi: 10.1111/j.1365-246X.2008.03740.x

Mochizuki, N., Maruuchi, T., Yamamoto, Y., and Shibuya, H. (2013). Multi-level consistency tests in paleointensity determinations from the welded tuffs of the Aso pyroclastic-flow deposits. *Phys. Earth Planet. Inter.* 223, 40–54. doi: 10.1016/j.pepi.2013.05.001

Morales, J., Goguitchaichvili, A., Alva-Valdivia, L. M., and Urrutia-Fucugauchi, J. (2003). Absolute paleointensity of the Earth's magnetic field during Jurassic: case study of La Negra Formation (northern Chile). *C. R. Geosci.* 335, 661–670. doi: 10.1016/S1631-0713(03)00116-0

Muxworthy, A. R. (2010). Revisiting a domain-state independent method of palaeointensity determination. *Phys. Earth Planet. Inter.* 179, 21–31. doi: 10.1016/j.pepi.2010.01.003

Muxworthy, A. R., Evans, M. E., Scourfield, S. J., and King, J. G. (2013). Paleointensity results from the late-Archaean Modipe Gabbro of Botswana. *Geochem. Geophys. Geosyst.* 14, 2198–2205. doi: 10.1002/ggge.20142

Muxworthy, A. R., and Heslop, D. (2011). A Preisach method for estimating absolute paleofield intensity under the constraint of using only isothermal measurements: 1. Theoretical framework. *J. Geophys. Res. Solid Earth* 116:B04102. doi: 10.1029/2010JB007844

Paterson, G. A. (2011). A simple test for the presence of multidomain behavior during paleointensity experiments. *J. Geophys. Res. Solid Earth* 116:B10104. doi: 10.1029/2011JB008369

Paterson, G. A. (2013). The effects of anisotropic and non-linear thermoremanent magnetizations on Thellier-type paleointensity data. *Geophys. J. Int.* 193, 694–710. doi: 10.1093/gji/ggt033

Paterson, G. A., Biggin, A. J., Yamamoto, Y., and Pan, Y. X. (2012). Towards the robust selection of Thellier-type paleointensity data: the influence of experimental noise. *Geochem. Geophys. Geosyst.* 13:Q05Z43. doi: 10.1029/2012GC004046

Paterson, G. A., Heslop, D., and Muxworthy, A. R. (2010). Deriving confidence in paleointensity estimates. *Geochem. Geophys. Geosyst.* 11:Q07Z18. doi: 10.1029/2010GC003071

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

Riisager, P., and Riisager, J. (2001). Detecting multidomain magnetic grains in Thellier palaeointensity experiments. *Phys. Earth Planet. Inter.* 125, 111–117. doi: 10.1016/S0031-9201(01)00236-9

Rolph, T. C., and Shaw, J. (1985). A new method of paleofield magnitude correction for thermally altered samples and its application to Lower Carboniferous lavas. *Geophys. J. Roy. Astr. Soc.* 80, 773–781. doi: 10.1111/j.1365-246X.1985.tb05124.x

Selkin, P. A., Gee, J. S., and Tauxe, L. (2007). Nonlinear thermoremanence acquisition and implications for paleointensity data. *Earth Planet. Sci. Lett.* 256, 81–89. doi: 10.1016/j.epsl.2007.01.017

Selkin, P. A., Gee, J. S., Tauxe, L., Meurer, W. P., and Newell, A. J. (2000). The effect of remanence anisotropy on paleointensity estimates: a case study from the Archean Stillwater Complex. *Earth Planet. Sci. Lett.* 183, 403–416. doi: 10.1016/S0012-821X(00)00292-2

Shaar, R., and Tauxe, L. (2013). Thellier GUI: an integrated tool for analyzing paleointensity data from Thellier-type experiments. *Geochem. Geophys. Geosyst.* 14, 677–692. doi: 10.1002/ggge.20062

Shaw, J. (1974). A new method of determining the magnitude of the palaeomagnetic field, application to five historic lavas and five archaeological samples. *Geophys. J. Roy. Astr. Soc.* 39, 133–141. doi: 10.1111/j.1365-246X.1974.tb05443.x

Shcherbakova, V. V., Bakhmutov, V. G., Shcherbakov, V. P., Zhidkov, G. V., and Shpyra, V. V. (2012). Palaeointensity and palaeomagnetic study of Cretaceous and Palaeocene rocks from Western Antarctica. *Geophys. J. Int.* 189, 204–228. doi: 10.1111/j.1365-246X.2012.05357.x

Shcherbakova, V. V., Kovalenko, D. V., Shcherbakov, V. P., and Zhidkov, G. V. (2011). Paleointensity of the geomagnetic field in the Cretaceous (from Cretaceous rocks of Mongolia). *Izvestiya Phys. Solid Earth* 47, 775–791. doi: 10.1134/S1069351311070081

Tarduno, J. A., Cottrell, R. D., and Smirnov, A. V. (2001). High geomagnetic intensity during the mid-Cretaceous from Thellier analyses of single plagioclase crystals. *Science* 291, 1779–1783. doi: 10.1126/science.1057519

Tauxe, L., and Staudigel, H. (2004). Strength of the geomagnetic field in the Cretaceous Normal Superchron: new data from submarine basaltic glass of the Troodos Ophiolite. *Geochem. Geophys. Geosyst.* 5, Art. No. Q02H06. doi: 10.1029/2003GC000635

Thellier, E., and Thellier, O. (1959). Sur l'intensité du champ magnétique terrestre dans la passé historique et géologique. *Ann. Géophys.* 15, 285–376.

Van der Voo, R. (1990). The reliability of paleomagnetic data. *Tectonophysics* 184, 1–9. doi: 10.1016/0040-1951(90)90116-P

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

Wilson, R. L. (1962). An instrument for measuring vector magnetization at high temperatures. *Geophys. J. Roy. Astr. Soc.* 7, 125–130. doi: 10.1111/j.1365-246X.1962.tb02257.x

Yamamoto, Y., Tsunakawa, H., and Shibuya, H. (2003). Palaeointensity study of the Hawaiian 1960 lava: implications for possible causes of erroneously high intensities. *Geophys. J. Int.* 153, 263–276. doi: 10.1046/j.1365-246X.2003.01909.x

Yu, Y. (2011). Importance of cooling rate dependence of thermoremanence in paleointensity determination. *J. Geophys. Res. Solid Earth* 116:B09101. doi: 10.1029/2011JB008388

Zhang, C., Paterson, G. A., and Liu, Q. (2012). A new mechanism for the magnetic enhancement of hematite during heating: the role of clay minerals. *Stud. Geophys. Geodaet.* 56, 845–860. doi: 10.1007/s11200-011-9018-4

Zhao, X. Y., Liu, Q. S., Paterson, G. A., Qin, H. F., Cai, S. H., Yu, Y., et al. (2014). The effects of secondary mineral formation on Coe-type paleointensity determinations: theory and simulation. *Geochem. Geophy. Geosyst.* 15, 1215–1234. doi: 10.1002/2013GC005165