UC San Diego

UC San Diego Previously Published Works

Title

Thellier GUI: An integrated tool for analyzing paleointensity data from Thellier-type experiments

Permalink

<https://escholarship.org/uc/item/95h4k2gk>

Journal

Geochemistry, Geophysics, Geosystems, 14(3)

ISSN

1525-2027

Authors

Shaar, R
Tauxe, L

Publication Date

2013-03-01

DOI

10.1002/ggge.20062

Peer reviewed

Thellier GUI: An integrated tool for analyzing paleointensity data from Thellier-type experiments

Ron Shaar and Lisa Tauxe

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, 92093-0220, USA (rshaar@ucsd.edu)

[1] Thellier-type experiments are a method used to estimate the intensity of the ancient geomagnetic field from samples carrying thermoremanent magnetization. The analysis of Thellier-type experimental data is conventionally done by manually interpreting data from each specimen individually. The main limitations of this approach are: (1) manual interpretation is highly subjective and can be biased by misleading concepts, (2) the procedure is time consuming, and (3) unless the measurement data are published, the final results cannot be reproduced by readers. These issues compound when trying to combine together paleointensity data from a collection of studies. Here, we address these problems by introducing the *Thellier GUI*: a comprehensive tool for interpreting Thellier-type experimental data. The tool presents a graphical user interface, which allows manual interpretation of the data, but also includes two new interpretation tools: (1) *Thellier Auto Interpreter*: an automatic interpretation procedure based on a given set of experimental requirements, and 2) *Consistency Test*: a self-test for the consistency of the results assuming groups of samples that should have the same paleointensity values. We apply the new tools to data from two case studies. These demonstrate that interpretation of non-ideal Arai plots is nonunique and different selection criteria can lead to significantly different conclusions. Hence, we recommend adopting the automatic interpretation approach, as it allows a more objective interpretation, which can be easily repeated or revised by others. When the analysis is combined with a *Consistency Test*, the credibility of the interpretations is enhanced. We also make the case that published paleointensity studies should include the measurement data (as supplementary files or as a contributions to the MagIC database) so that results based on a particular data set can be reproduced and assessed by others.

Components: 8,900 words, 10 figures.

Keywords: paleointensity; MagIC database; Thellier.

Index Terms: 1521 Geomagnetism and Paleomagnetism: Paleointensity; 1594 Geomagnetism and Paleomagnetism: Instruments and Techniques.

Received 26 October 2012; **Revised** 7 January 2013; **Accepted** 7 January 2013; **Published** 28 March 2013.

Shaar, R., and L. Tauxe (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.

1. Introduction

[2] Ancient materials that carry a thermoremanent magnetization (TRM) can retain information on the absolute intensity of the past geomagnetic field. This information can be retrieved through an experimental procedure known as the Thellier method [Thellier and Thellier, 1959]. Despite the great advantages of the Thellier method, there are some difficulties underlying its application, mainly because it is based on a number of strict assumptions, which are hard to fulfill or even to test. As a result, the interpretation of Thellier-type experimental data is often ambiguous. This issue is compounded when trying to compare or combine different published datasets (which were likely interpreted using different guiding principles) because they often do not include the original measurements. Still, even if the measurement data are available, the conventional approach of manually interpreting each specimen and sample separately may be highly subjective as well as time consuming.

[3] In this study, we address the difficulties in interpreting Thellier-type experimental data and introduce new techniques for analyzing them combined together in the software package, *Thellier GUI*. The main improvements of the new package over the existing approaches are: (1) interpretations are done automatically in a fast, systematic, and consistent fashion, using a given set of experimental requirements, (2) *Thellier GUI* can handle datasets that are virtually unlimited in size, and (3) it includes a consistency test using a subset of samples that are expected to give the same paleointensity value based on field relations. The cross-platform graphical user interface (GUI) allows data visualization and both the conventional (manual) as well as the new automatic interpretation approaches. The GUI has been made part of PmagPy software package (<http://earthref.org/PmagPy/cookbook/>) and is designed to work with the MagIC data format (earthref.org/MAGIC). The MagIC database facilitates the ongoing effort to establish a comprehensive database with global paleointensity measurements.

[4] The outline of this paper is as follows. In section 1, we briefly review the Thellier method: its technical procedure, causes of failure, and the conventional interpretation procedure. In sections 2–4, we introduce the *Thellier GUI* program, which includes two new interpretation tools: *Thellier Auto Interpreter* and *Consistency Test*. We examine two case studies in section 5 using the new interpretation tools. The Supporting Information consists of a description of the paleointensity statistical definitions being used. The Appendix, given as a web supplement, provides

a tutorial for using the GUI, as well as a link for downloading the datasets discussed in this article.

1.1. The Procedure of Thellier-type Experiments

[5] The basic assumption underlying Thellier-type experiments is expressed in equation (1), which can be derived directly from Néel Theory [Néel, 1955] assuming anisotropic TRM [e.g., Selkin et al., 2000]:

$$\text{TRM} = [A'] \tanh(C \cdot B) \quad (1)$$

where TRM is the magnetization vector, C is a scalar, B is the applied field vector, and [A'] is a second-degree anisotropy tensor. Most often, the field B is low enough for the hyperbolic function to be approximated by a linear function [but not always, see Selkin et al., 2007; Shaar et al., 2010]. In these cases, equation (1), reduces to:

$$\text{TRM} = [A] \cdot B \quad (2)$$

[6] In addition, the tensor [A] can often be reasonably approximated by a scalar, but here we will retain the tensor form. Under the assumption of equation (2), once the ancient TRM (TRM_{anc}) of a specimen is measured and the specimen is given a laboratory TRM (TRM_{lab}) in a known lab field (B_{lab}), the intensity ancient field (B_{anc}) can be calculated using:

$$B_{\text{anc}} = |B_{\text{lab}}| \cdot \left(\frac{[A] \hat{r}_{B_{\text{lab}}} \cdot [A]^{-1} \hat{r}_{\text{TRM}_{\text{anc}}}}{[A] \hat{r}_{B_{\text{lab}}} \cdot [A] \hat{r}_{B_{\text{lab}}}} \right) \cdot \frac{|\text{TRM}_{\text{anc}}|}{|\text{TRM}_{\text{lab}}|} \\ = |B_{\text{lab}}| \cdot F_a \cdot |b| \quad (3)$$

where $\hat{r}_{B_{\text{lab}}}$ and $\hat{r}_{\text{TRM}_{\text{anc}}}$ are unit vectors in the direction of B_{lab} and TRM_{anc}, respectively. F_a is termed the anisotropy correction factor. When the hyperbolic function cannot be approximated by a linear function [Selkin et al., 2007; Shaar et al., 2010], B_{anc} is calculated using equations (2)–(3) in Shaar et al. [2010].

[7] Theoretically, the simplest paleointensity experiment could include only two measurements: the NRM and TRM_{lab}, assuming that the NRM is identical to TRM_{anc} and [A] has not changed. As this is very rarely the case, Koenigsberger [1936], followed by Thellier and Thellier [1959], suggested a series of double heating steps at elevated temperatures, instead of a single heating step. In each double heating step, a portion of the NRM is de- or remagnetized (zero-field step, or “Z” or in-field step “I,” respectively) and a partial TRM (pTRM) is acquired (in-field step, or “I”). This experiment can be performed using the original Thellier protocol [Thellier and Thellier, 1959] (a pair of in-field steps with opposing fields, or “II”), or using one of its modifications: “ZI” [Coe, 1967], “IZ” [Aitken et al., 1988], or “IZZI” [Tauxe

and Staudigel, 2004]. For reviews of different protocols, see Yu *et al.* [2004] and Tauxe and Yamazaki [2007].

[8] The results of the Thellier-type experiment are displayed on an Arai plot [Nagata *et al.*, 1963, Figure 1], which is a scatter plot of data points of the NRM (progressively lost) versus pTRM (progressively gained). The slope of a best fit line, b , through the points allows the estimation of B_{anc} from equation (3). The NRM directions measured during the experiment are considered in the analysis and are plotted on orthogonal or Zijderfeld [Zijderfeld, 1967] plots (insets to Figure 1). A successful Thellier-type experiment results in a linear Arai plot (Figure 1a) and a straight-line Zijderfeld plot converging toward the origin (inset to Figure 1a).

1.2. Reasons for Possible Failure of Thellier-type Experiments

[9] A major difficulty in paleointensity research is that frequently very few, if any, specimens in a given study display the ideal behavior shown in Figure 1a. The reasons for nonideal behavior vary and may include experimental noise [Paterson *et al.*, 2012] and/or violation of one of the following basic assumptions, which are built into the Thellier-type experimental procedure:

- (1) The NRM is the TRM_{anc} . The original ancient thermal remanence could be overprinted or replaced by viscous (if the sample contains large fraction of unstable particles), chemical (due to weathering), thermal (due to exposure to high

Figure 1. Different results of Thellier experiments (IZZI protocol) displayed on Arai plots and Zijderfeld plots (insets). Red circles, blue circles, and triangles in the Arai plot are “ZI” steps, “IZ” steps, and pTRM checks, respectively. Blue (red) squares in the Zijderfeld plot are x-y (x-z) projections of the NRMs in the specimen coordinate system, where x axis is rotated to the direction of the NRM in the X-Y plane. ((a) Ideal behavior. (b) A scattered plot resulting with high β and high DRATS. (c) A bilinear Arai plot resulting with low fraction statistic (f_{vds}) for each linear segment. (d) pTRM checks fall far from the curve (high DRATS), and a Zijderfeld plot converging toward the direction of the lab field (high MAD, DANG) indicating alteration during experiment. (e) A curved Zijderfeld plot resulting with high DANG. (f) A zigzagged Arai plot, which results in a high zigzag multidomain statistic (Z).

temperature), thermochemical (due to, for example, high temperature oxidation), or isothermal (exposure to a strong field) remanences.

- (2) The ability of the specimen to acquire and preserve TRM (i.e., the values of the variables in equation (1)) has not changed since the acquisition of the ancient TRM. This ability could alter by physical or chemical weathering, and/or during the repeated heating in the laboratory.
- (3) pTRMs are independent, reciprocal, and additive: These assumptions were formulated by *Thellier* (1938) [see also *Dunlop and Özdemir* (1997), *Yu et al.* (2004), and *Tauxe* (2010) for further discussion] and define the requirement for pTRMs in order to obtain successful results. These requirements are equivalent to demanding that the specimen contain only noninteracting single domain (SD) magnetic particles.

[10] Some of the assumptions of the Thellier method can be partially tested if the experiment is carefully designed. The deviation from linearity in the Arai plot can be used to assess the extent to which the material satisfies the requirement of SD, where a straight line suggests SD or small PSD [*Dunlop and Özdemir*, 1997, *Shaar et al.*, 2010], and other curves such as concave (Figure 1c), convex, and zigzagging (Figure 1f) suggest non-SD particles [e.g., *Levi*, 1977; *Dunlop and Özdemir*, 1997; *Fabian*, 2001; *Coe et al.*, 2004; *Yu et al.*, 2004; *Shaar et al.*, 2011]. Overprints can be detected by inspecting changes in the trends of the Zijderveld plot. Alteration during the experiment can be partially detected using additional “pTRM checks” [*Coe and Gromme*, 1978] which are repeated pTRM acquisition steps performed after temperature steps at higher temperature (triangles in Figure 1). Nonreciprocal pTRMs, typical of multidomain grains, can be partially detected using “tail checks” [*Riisager and Riisager*, 2001], or alternatively using the IZZI method of *Tauxe and Staudigel* [2004] which produces “zigzagged” plots (Figure 1f) when reciprocity is violated [*Yu et al.*, 2004]. For additional methods for detecting violation of the Thellier experimental assumptions, see review by *Tauxe and Yamazaki* [2007] and *Tauxe* [2010].

1.3. Conventional Approaches of Interpretation

[11] The interpretation of the Thellier-type experiment involves two issues. The first is how to determine that the experiment failed and cannot provide reliable paleointensity estimation. In this case, the whole set of measurements of a particular specimen are

discarded. The second issue is how to choose the most appropriate temperature interval for estimating $|b|$. It is not always straightforward how to choose the temperature bounds for the best fit line, especially when the data points are scattered or form a complicated curve. However, this is a critical decision as different choices may lead to different paleointensity estimates.

[12] Paleointensity statistics are a useful tool for evaluating the quality of the interpretation. They are frequently used as acceptance criteria by assigning threshold values to a set of statistics. When adopting this approach, only interpretations that meet the criteria are accepted. Section A1 in the Supporting Information¹ (see also the Appendix to this article) lists the paleointensity statistics that we apply in this contribution (see *Tauxe* [2010] for a more detailed discussion). This is not a complete list of the all the available statistics, and there are many other that are frequently used [e.g., *Leonhardt et al.*, 2004; *Kissel and Laj*, 2004; *Ben-Yosef et al.*, 2008, 2009; *Biggin*, 2010; *Valet et al.*, 2010; *Paterson*, 2011].

[13] Examples of different results that are typical of Thellier-type experiments are shown in Figure 1. The behavior shown in Figure 1a is ideal, with straight-line Arai and Zijderveld plots. The Arai plot in Figure 1b displays a complicated scatter, which results in a high “scatter statistic” [$\sigma_b/|b|$] of *Coe and Gromme* [1978], renamed β by *Selkin and Tauxe* [2000]. Also, the pTRM checks in Figure 1b fall far from the curve, which results in a high “alteration statistic” (e.g., *DRATS* of *Tauxe and Staudigel* [2004]). The Arai plot in Figure 1c is bilinear where each segment is a straight line, but since only a fraction of the magnetization is used, each segment results in a low “fraction statistic” [e.g., f_{vds} of *Tauxe and Staudigel* [2004]]. Alteration during the experiment is apparent in Figure 1d: the pTRM checks above 360°C fall far from the curve, and the Zijderveld plot converges toward the direction of the lab field (-z direction). The pTRM checks in Figure 1d result in a high *DRATS*, and the curved Zijderveld plot results in a high “Maximum Angular Deviation” (*MAD*, of *Kirschvink* [1980]) and high “Deviation of the ANGLE” (*DANG* of *Tauxe and Staudigel* [2004]). Figure 1e is another example of a curved Zijderveld plot, which results in a high *DANG*, caused by a multicomponent magnetization. The zigzagged behavior in Figure 1f is associated with multidomain magnetic grains, which result in a high “zigzag statistic” (*Z* of *Yu and Tauxe* [2005],

¹All Supporting Information may be found in the online version of this article.

revised by Ben-Yosef *et al.* [2008] or *IZZI_MD* of Shaar *et al.* [2011]).

[14] We identify several limitations in the approach normally taken for interpreting paleointensity data.

- (1) The interpretation of specimens that do not display the ideal behavior shown in Figure 1a is not straightforward and requires subjective judgments.
- (2) When interpreting a large number of specimens, it is a challenge to remain consistent and follow the exact guiding principles throughout all the interpretations.
- (3) Manual interpretation is time consuming.
- (4) When adopting paleointensity statistics as acceptance criteria, it is not obvious which statistics and threshold values to use.
- (5) As a reader, reviewing the interpretations in a given published dataset is practically impossible unless the measurement data are given. Yet, even if the measurement data are given, manually reviewing each specimen separately can be laborious for large datasets.
- (6) It is not obvious how to compare or compile together different publications, which were likely interpreted using different guiding principles.

2. *Thellier Auto Interpreter*: Automatic Interpretation of *Thellier* Experiment Data

[15] To address the difficulties in manual interpretation (section 1.3), we introduce a tool for automatic interpretation using a given set of paleointensity statistics. The *Thellier Auto Interpreter* uses three groups of paleointensity statistics:

- (1) Specimen paleointensity statistics (Table A1, Supporting Information, sections 2.1,2.2): a set of statistics that define the acceptance criteria at the specimen level.
- (2) Sample calculation methods (Table A2, Supporting Information, section 2.3): the method by which the sample paleointensity and its corresponding confidence error are estimated.
- (3) Sample paleointensity statistics (Table A3, Supporting Information, section 2.4): a set of statistics that define the acceptance criteria at the sample level.

[16] The procedure works as follows. (1) The program goes through all the specimens in the dataset and inspects each for all possible interpretations of best fitting lines. The interpretations that meet the specimen

acceptance criteria (“accepted interpretations”) are saved into a file, while the others are discarded. (2) The program goes through all the samples in the dataset, and if enough specimens from each sample contain at least one “acceptable interpretation” then the sample paleointensity and its confidence error are calculated, using one of the methods listed in section 2.3 and Table A2, Supporting Information. Samples that do not meet the sample acceptance criteria (Table A3, Supporting Information) are discarded.

2.1. Simplifying Specimen Paleointensity Statistics

[17] To allow an automatic interpretation by *Thellier Auto Interpreter*, we need a set of paleointensity statistics, which is capable of detecting the most likely causes of failure and efficiently identifying unacceptable results. The only consensus in the community is that there are too many ways of characterizing the experimental results (there are at least a half dozen ways of treating pTRM checks alone), yet more statistics are defined every year. Here, we aim to minimize the number of the statistics necessary in order to make the choice of the acceptance criteria easier. We also have detected deficiencies in the existing palette of statistics. For this purpose, we define several new paleointensity statistics as described below.

2.1.1. *FRAC*: Fraction of Remanence

[18] The fraction of remanence used in paleointensity experiments is a critical statistic. Perfect data would allow the use of the entire NRM, but such perfection is rare in actual experiments. There are at least two statistics for estimating the fraction of remanence. The f , defined by Coe *et al.* [1978] is the y-component of the best fit line in the chosen segment divided by the y-intercept of the extrapolated best fit line (Table A1, Supporting Information, Figure 2a). This method of calculation will give misleadingly high values for concaved Arai plot (Figure 2a). To address this problem, Tauxe and Staudigel [2004] introduced f_{vds} , which uses the same numerator as in f , but the sum of all the vector differences lengths between consecutive temperature steps (Vector Difference Sum or *VDS*, see equation (4)), as the denominator. The problem with f_{vds} is that it amplifies the denominator when the Zijdeveld plot is highly scattered or zigzagged (Figure 2b). Therefore, neither f nor f_{vds} is ideal for general use or for the *Thellier Auto Interpreter*. Here, we introduce a new fraction statistic, *FRAC*, which is the *VDS* of the selected component divided by the total *VDS*:

Figure 2. Comparison between different methods of calculating the fraction statistic - f [Coe *et al.*, 1978], f_{vds} [Tauxe and Staudigel, 2004], and $FRAC$ (this study). Temperature bounds and best fit lines are marked in green. (a) The f is calculated by the y-component of the least squares line between the temperature bounds divided by the y-intercept of the least squares line (Table A1, Supporting Information) resulting in f values that may overestimate the quality of a concaved Arai plot. (b) A noisy or zigzagged Zijderveld plot can amplify VDS, resulting in f_{vd} value that can underestimate the quality of the Arai plot.

$$FRAC = \frac{\sum_{i=start}^{i=end-1} |NRM_{i+1} - NRM_i|}{\sum_{i=0}^{i=n_0-1} |NRM_{i+1} - NRM_i| + |NRM_{n_0}|} \quad (4)$$

where n_0 is the total number of data points in the Arai plot, *start* and *end* are the first and the last points in the chosen segment, and NRM_i is the NRM vector at step i . The denominator in equation (4) is the VDS of Tauxe and Staudigel [2004].

2.1.2. SCAT: A Scatter Statistic

[19] In an effort to minimize the number of paleointensity statistics and avoid the need for a complicated combination of many different statistics, we define the scatter statistic, *SCAT*. *SCAT* is a Boolean, which uses the threshold value of β to indicate whether the data points (including the pTRM checks and tail checks) in the chosen segment of the Arai plot are too scattered. *SCAT* is meant to replace *DRATS* (alteration check), *Z* (zigzag statistic), *MD* (tail checks), and any of the other scatter and alteration detection statistics available in the literature. It depends only the threshold value of β , $\beta_{threshold}$.

[20] Figure 3a illustrates graphically the way *SCAT* is determined. First, the least squares line through the chosen segment in the Arai plot (solid line through green diamond) is calculated using York [1966], following Coe *et al.* [1978]. The slope of

this line and its associated standard error are termed b and σ_b , respectively [Coe *et al.*, 1978, Table A1, Supporting Information]. Assuming that β (defined as $\sigma_b/|b|$) is used as a selection criterion, we set the threshold value for the standard error of the slope to be: $\sigma_{threshold} = |b|\beta_{threshold}$. We now use b and $\sigma_{threshold}$ to draw two lines that pass through the center of mass of the chosen segment (green diamond in Figure 3a), with the slopes $b + 2\sigma_{threshold}$ and $b - 2\sigma_{threshold}$ (dashed lines), respectively. The intersections of these lines with the x-axis and the y-axis of the Arai plot define four points that outline a polygon bounded by the blue and the red lines. The gray polygon (the SCAT box) is defined by the red line, the blue line, and two vertical lines passing through the first and the last point in the chosen segment. *SCAT* is *True* if all the data points associated with the chosen segment (including the pTRM checks and the tail checks) fall inside the box. *SCAT* is *False* if any point falls outside the box. The pTRM and tail checks are included in *SCAT* if they check reproducibility of temperature steps within the segment and were measured before reaching a temperature higher than the highest temperature in the segment. Figure 3b illustrates a specimen that passes the SCAT criterion, whereas the specimen in Figure 3c fails.

2.1.3. GAP-MAX: A Maximum Gap Statistic

[21] We define a new statistic for *Theillier Auto Interpreter* routine, *GAP-MAX*, to detect a poorly

Figure 3. The SCAT statistic. (a) Graphical representation of the SCAT calculation (see section 2.1 for details). The SCAT criterion passes if all the data points, including the pTRM checks and the tail checks, fall inside the gray box and fails if one or more point fall outside the box. (b–c) Interpretation using the SCAT statistics. Symbols are as in Figure 1, tail checks are marked with open squares, temperature bounds are marked with green symbols, and the green line is best fit line between the temperature bounds. In both examples, the threshold value for β is 0.10. The interpretation in Figure 3b passes the criterion, while that in Figure 3c fails.

defined Arai plot with two consecutive data points separated by a very large gap. *GAP-MAX* measures the lengths of all the vector differences between consecutive temperature steps in the segment and return the maximum length normalized by the VDS of the component:

$$GAP-MAX = \frac{\max_{i=start}^{i=end-1} |NRM_{i+1} - NRM_i|}{\sum_{i=start}^{i=end-1} |NRM_{i+1} - NRM_i|} \quad (5)$$

where NRM_i , *start* and *end*, are defined in equation (4).

[22] Despite the similarity in names, the *GAP-MAX* is different than the frequently used “gap factor” *g* of Coe *et al.* [1978] in the sense that *g* is the weighted mean of the y-component gaps in the Arai plot (see Table A1, Supporting Information, for the equation), whereas *GAP-MAX* is the maximum gap (measured as vector length).

2.2. Which Statistics to Use at the Specimen Level?

[23] Using the new definitions above, we argue that a set of six statistics is sufficient for a robust performance of *TheLLier Auto Interpreter*: *FRAC*, *GAP-MAX*, n_{ptm} (minimum number of pTRM checks carried out before reaching temperature higher than the upper temperature bound), β , *MAD*, and *DANG*. The first four statistics account for the linearity and the scatter of the Arai plot, and the last two account for the linearity of the Zijdeveld plot and its convergence to the origin

(See Table A1, Supporting Information, for complete definitions).

2.3. Sample Calculation Methods

[24] In the following section, we use the following notation: a sample is a physical object that must have been subjected to the same field conditions during initial cooling, and a specimen is a subsample from this object. Remembering that paleointensity values at the specimen level are not measurements, but estimates based on interpretations of the Arai plots, calculating the paleointensity at the sample level can be nonunique. Some problems in calculating the sample paleointensity include: (1) how to propagate paleointensity estimates from the specimen level to the sample level when there is more than one possible interpretation, each resulting in a different value, (2) how to propagate uncertainties from the specimen level to the sample level, and (3) how to calculate the confidence bounds at the sample level. A full discussion of these issues is beyond the scope of this article. Here, we provide three useful algorithms employed by *TheLLier Auto Interpreter* for calculating the sample paleointensity and its uncertainty. Readers are referred to Paterson *et al.* [2010] and references therein for a thorough discussion of the uncertainties involved in calculating the sample mean.

2.3.1. STDEV-OPT: Optimized Standard Deviation

[25] In the published literature, the most frequent way of calculating the sample paleointensity is to

choose one interpretation from each specimen (sometimes out of many possibilities) and calculate the mean and standard deviation (or standard error) of the chosen specimen interpretations. One approach usually taken (perhaps unconsciously) is to choose interpretations that agree best with each other, leading to a low standard deviation for the sample mean. The *STDEV-OPT* follows this approach. Given a list of specimens, each with a list of “acceptable” interpretations that pass the selection criteria, *STDEV-OPT* chooses values that lead to a minimum standard deviation of the sample mean. For example, if sample S1 includes three specimens with the following “acceptable” interpretations: S1a (40.1, 45.0, and 50.1 μT), S1b (44.5, 45.1, and 46.5 μT), and S1c (42.2, 43.3, and 44.9 μT), then the *STDEV-OPT* algorithm chooses 45.0, 45.1, and 44.9 from specimen a, b, and c, respectively.

2.3.2. Bootstrap

[26] It could be claimed that the *STDEV-OPT* does not provide an accurate assessment of the uncertainty involved in the interpretation as it uses only one “acceptable” interpretation from each specimen and ignores the rest. Following *Shaar et al.* [2011] [see also *Bowles et al.*, 2005] we apply bootstrap statistics [*Efron*, 1981] to calculate the sample paleointensity. Assuming n specimens, each with a list of “acceptable” interpretations, the general bootstrap procedure is as follows. (1) Randomly choose one specimen. (2) Pick the paleointensity estimate for the chosen specimen using one of the methods described below (simple or parametric bootstrap). (3) Repeat (1)–(2) n times (a specimen can be chosen more than one time). (4) Calculate the mean of the n values. (5) Repeat steps (1)–(4) N times (N is a large number, which is 10^4 in the *TheLLier GUI* program by default). (6) Calculate the mean or median of the N bootstrapped means to estimate the sample paleointensity. The interval containing 95% (or 68%) of the bootstrapped means defines the confidence interval. Step #2 of this procedure can use one of the following options:

- **BS (simple bootstrap):** one value is randomly chosen from the list of “acceptable interpretations”.
- **BS-PAR (parametric bootstrap):** The minimum and the maximum values in the list of “acceptable interpretations” define an “accepted interval.” The parametric bootstrap assumes that the true value lies within this interval. Under the assumption that no information is known about the probability function, a value is randomly chosen from a uniform distribution function defined by this interval.

[27] One main advantage in the bootstrap procedure is that it allows users to loosen the selection criteria at the specimen level, and in so doing, obtain paleointensity estimates from MD-like curved Arai plots [*Shaar et al.*, 2011]. Yet caution should be taken when adopting the bootstrap method, as it depends on a large number of specimens per sample, and most studies do not have a sufficient number.

2.4. Which Statistics to Use at the Sample Level?

[28] A list of sample paleointensity statistics is given in Table A3, Supporting Information. We suggest at the minimum using a threshold value for the number of specimens per sample (N_{sample}) and a threshold value for the estimated confidence interval.

[29] Both the sample mean and the confidence interval estimation have an inherent uncertainty that is related to the number of measurements. *Paterson et al.* [2010] provide a detailed discussion of methods for estimating the confidence interval and the uncertainties involved in these estimates. Here, we draw on their conclusions and use the standard deviation of the data (or other estimated confidence interval) as a selection criterion, keeping in mind that the true uncertainty is poorly defined. This selection criterion can be expressed as an absolute value (units of μT) or as a relative value (boundary of confidence divided by the mean in units of %). We suggest using both, and the *TheLLier GUI* combines the two threshold values using a logical OR function. The reason for this is that a low absolute paleointensity value (say 15 μT) may have a standard deviation of several microtesla (say, 5), which could be 30% of the mean, so using only the percentage of the mean will discriminate against what may be experimentally an excellent result. On the other hand, a high absolute value of paleointensity (say 100 μT), may have excellent reproducibility of 10% of the mean (or 10 μT), but this would exceed a strict 5 μT threshold chosen by the investigator.

[30] Another aspect of uncertainty at the sample level is the degrees of freedom in interpreting the Arai plots. If the acceptance criteria at the specimen level are relatively loose (for example $FRAC < 0.6$), then there may be a number of “acceptable” interpretations at the specimens level; each can lead to a significantly different paleointensity estimate (e.g., example Figure 1c). In this case, there are also a number of “acceptable” sample means, which might yield significantly different estimates. In the *TheLLier GUI*, we employ a new statistic (*SAMPLE-INT-INTERVAL*) to account for this issue (see Table A3, Supporting Information, for details).

3. Consistency Test

[31] Choosing acceptance criteria for *Thellier Auto Interpreter* is a central issue, and the final results are strongly influenced by this choice. Ideally, we would want to accept only ideal Arai plots (i.e., Figure 1a) and samples with high N_{sample} and low standard deviation (by choosing very strict threshold values). Yet, very few samples in a realistic dataset, if any, can fulfill nearly ideal requirements (for example $FRAC > 0.9$, $\beta < 0.05$, MAD , $DANG < 5$, $N_{\text{sample}} > 10$, and $\sigma_{\text{sample}} < 5\%$). While investigators have to relax the acceptance criteria, relaxing them too much might result in low reliability of the results.

[32] The purpose of the *Consistency Test* proposed here is to help users assess the reproducibility of results using a given set of acceptance criteria by performing a test for the consistency of the data among samples that can be considered contemporaneous (for example come from the same lava flow, or archaeological horizon). The *Consistency Test* relies on defining groups of samples (termed hereafter *Test Groups*) that are expected to give similar paleointensity values. Examples for *Test Groups* are: archaeological samples that were collected from the same archaeological layer, samples from different parts of the same lava flow, contemporaneous samples dated to the same age, or adjacent samples in a core. In this sense, a *Test Group* is similar to a paleomagnetic site, using the definition from Tauxe (2010): “A site is a single horizon or instant in time and may comprise multiple samples or may be only a single sample, depending on the application. Multiple specimens from a single site are expected to have recorded the same geomagnetic field.”

[33] Following the approach in section 2.2, we found in many “trial and error” tests that the most difficult decisions to make regarding the acceptance criteria are the threshold values for $FRAC$ and β . Hence, we designed the *Consistency Test* to map results using different values for these two statistics.

[34] The *Thellier Consistency Test* requires several inputs (see section A.2, Supporting Information, for details): (1) a list of “*Fixed Criteria*”: threshold values for paleointensity statistics, not including β and $FRAC$, (2) a list of *Test Groups*, (3) an allowable range of values for β and $FRAC$ (for example, β from 0.05 to 0.20 in steps of 0.01, and $FRAC$ from 0.7 to 0.9 in steps of 0.02), and (4) a list of “*Consistency test functions*” (section A.2, Supporting Information).

[35] The procedure of the *Thellier Consistency Test* is as follows. (1) For each β and $FRAC$, an automated interpretation is performed using *Thellier Auto*

Interpreter. The samples that passed the criteria are saved into a file, and *Consistency test functions* are calculated. (2) Upon completing the routine in step (1), the values of the *Consistency Test function* are displayed on a color map. (3) The user employs the color maps to choose an optimal set of β and $FRAC$ that yield minimum scatter in the test-groups but enough samples that pass the criteria. Thus, balancing the “quality” of the interpretation and quantity of the samples. (4) steps (1)–(3) can be repeated multiple times using different “*Fixed Criteria*.”

[36] We demonstrate the application of the *Thellier Consistency Test* in section 5.

4. Thellier GUI

[37] The new procedures described in sections 2–3 are implemented in a GUI called *Thellier GUI*. The *Thellier GUI* is a tool for viewing and analyzing the paleointensity data using the conventional approach as well as the new procedures introduced here. The *Thellier GUI* is designed to work with the MagIC format (earthref.org/MAGIC) and is a contribution to the PmagPy software package (<http://earthref.org/PmagPy/cookbook/>), which is cross-platform and based on the freely available Enthought Python Distribution.

[38] A snapshot of the *Thellier GUI* front panel is shown in Figure 4. The front panel includes graphical displays of the measurements and controls for interpreting the data and viewing the values of the paleointensity statistics. The menu bar includes a number of operations, including setting the preferences for the display, saving plot files, calculating remanence anisotropy tensors, running the *Thellier Auto Interpreter* and the *Consistency Test* routines, and plotting paleointensity curves (field or virtual axial dipole moment versus age). The dialog window for setting paleointensity statistics (Tables A1–A3, Supporting Information) is shown in Figure 5. The dialog window of running the *Consistency Test* (section 3) is shown in Figure 6.

[39] The Appendix provides web links for downloading the program and its tutorial.

5. Case Studies

5.1. Case Study 1: Roman Age Copper Slag Mound, Cyprus

[40] Our first case study uses unpublished paleointensity data from a Roman Age archaeological

Figure 4. Snapshot of the *Thellier* GUI main panel.

Figure 5. Snapshot of acceptance criteria dialog window (section A1, Supporting Information).

mound near Skouriotissa, Cyprus [Ben-Yosef *et al.*, 2011, Figure 3]. The mound was built from layers of slag intermixed with charcoals deposited between the third and the fourth centuries CE. The dataset includes 318 specimens from 70 slag samples, which

were collected from different horizons in the mound. The behavior of the specimens is highly variable: some specimens behave as SD, others suffer from pTRM tails (manifested as zigzagged Arai plots in the IZZI experiment), and some show significant

Figure 6. Snapshot of *Consistency Test* dialog window (Section A2, Supporting Information).

alteration during the experiment. In addition, many specimens exhibit significant anisotropy. In our investigation, we assume that samples which were collected from the same layer in the mound (locus) are close enough in age to have experienced the same geomagnetic field during cooling. Thus, we group the individual samples into *Test Groups* (section 3) using their stratigraphic location within the mound.

[41] After designating test groups, we define a set of “fixed criteria” which all specimens must meet. In this example, following the discussion in section 2.2, we use the following fixed criteria: $n_{ptm} \geq 2$, $GAP-MAX \leq 0.6$, $MAD \leq 5.0$ and $DANG \leq 10.0$ (See Table A1, Supporting Information, for definitions). The choice for the threshold values of MAD and $DANG$ is somewhat arbitrary, and we choose 5.0 and 10.0, respectively, to constrain our analysis to specimens with stable single-component NRM, which are more likely to produce accurate results. At the sample level, we use $STDEV-OPT$ for calculating the sample mean, with $N_{sample} \geq 3$, and σ less than 8% or $3\mu T$. These rather strict values reflect the level of uncertainty we wish to tolerate in this study, as the aim of the study is documenting rapid variation in the geomagnetic field intensity at high temporal

resolution. We run the *Consistency Test* for β s ranging from 0.05 to 0.15 in steps of 0.01, and $FRAC$ from 0.70 to 0.90 in steps of 0.01, using the following *Test Functions* (see Table A4, Supporting Information, for definitions):

- *study_sample_n* (the number of samples from the whole project that pass the criteria): We want this number to be as large as possible.
- *max_group_int_sigma_uT* (The standard deviation of the most scattered *Test Group* in units of μT): We want this number to be as low as possible.
- $((max_group_int_sigma_uT \leq 3) \text{ or } (max_group_int_sigma_perc \leq 8)) \text{ and } study_sample_n$. This function returns the total number of samples that passed the selection criteria, if the condition on the left side of the function is met.

[42] Color maps of the *Consistency Test* are shown in Figure 7. The numbers in Figure 7a are the number of samples that passed the criteria for each $FRAC$ and β combination. As noted above, we want this number to be as high as possible. High $FRAC$ and low β (the “strict” criteria in top right corner of the plot) result in few samples that pass, whereas low $FRAC$ and high β (the “loose” criteria in the bottom left corner of the plot) result in many samples that pass.

Figure 7. Color maps produced by the *Consistency Test* for case study 1 (section 5.1), showing values of “*test functions*” for different combinations of β and $FRAC$. The *test function* in Figure 7a is the number of samples that passed the criteria, and in Figure 7b is the standard deviation of the most scattered group. The expression of the *test function* in Figure 7c returns the number of samples that passed the criteria if the expression in the left side of the equation is met. The top right (bottom left) corners of the map represent relatively “strict” (“loose”) criteria. “Strict” criteria result in many samples that pass criteria, but high scatter of the “worst” *Test Groups*. “Loose” criteria result in low scatter in *Test groups* but only few samples that pass. The *test function* in Figure 7c helps in finding an “optimal” combination.

The standard deviations of the most scattered *Test Group* are shown in Figure 7b. As noted above, we want this number to be as low as possible. We see that the strict $FRAC/\beta$ criteria result in high consistency in *Test Groups* while loose criteria yield low consistency. The function displayed in Figure 7c balances consistency and quantity of acceptable results. There is a region of “optimal” $FRAC$ (0.82) and β (0.10) in Figure 7c that results in a total of 10 samples that pass. We re-run the *Thellier Auto Interpreter* using these selected values. A comparison of the results obtained by using “optimal” criteria, compared with results obtained by “loose” and “strict” criteria, is shown in Figure 8.

[43] One important conclusion we draw from this case study is already known. “Ideal” acceptance criteria net too few results, but the more the acceptance criteria are relaxed, the more the chance that the final result is inaccurate. In realistic datasets, such as this example, the *Consistency Test* is a useful tool for choosing a set of “realistic” acceptance criteria that meets a required uncertainty in the results. Thus, designing a study with enough *Test Groups* allows us to improve the credibility of the final results. We note, that in this example a total set of 10 samples out of 70 passed our “optimal” criteria. While this rate may seem very low, it allows for relatively high precision.

5.2. Case Study 2: DSDP/ODP Submarine Basaltic Glass Collections

[44] Our second case study is a collection of DSDP/ODP submarine basaltic glass 0–160 Ma, compiled and analyzed by Tauxe (2006). The dataset is available from the MagIC database (<http://earthref.org/MAGIC/3474>), and includes 947 specimens from 447 samples obtained from 62 DSDP drill holes. The main purposes of the original investigation were to inspect the long-term behavior of the geomagnetic field intensity, particularly the time-averaged value of the dipole moment, and the variability of the field during stable polarities. Here, we first redo the original interpretation using *Thellier Auto Interpreter*, and then re-interpret the data using the *Consistency Test*, following the approach outlined in section 5.1.

[45] The acceptance criteria in the original publication were: $\beta \leq 0.1$, $f_{vds} \geq 0.2$, $DRATS \leq 30$, $MAD \leq 15$, $DANG \leq 15$. Sample means were calculated using at least two specimens, and the threshold value for the accepted standard deviation was 5 μT or 15%. Also, the original publication restricted the maximum temperature step used in the calculation to be at least 350°C. Here, we do not apply this criterion and run *Thellier Auto Interpreter* using the other original criteria. The results of *Thellier Auto Interpreter* (188 samples) versus the published data (128 samples)

Figure 8. Intensity versus stratigraphic height curves of case study 1 (section 5.1). The data was produced using *Thellier Auto Interpreter* with different values of β and $FRAC$. Figures 8a and 8c demonstrate that too “loose” values result many samples, but the curve is scattered, whereas too “strict” values result with fewer samples, but the curve is smooth. The results for the optimal values, which were found using Figure 7c are shown in Figure 8b.

are shown in Figures 9a and 9b. The advantage of the systematic search of the *Thellier Auto Interpreter* versus the very cumbersome manual one in the original study is clear in that many more samples were found to pass the assumed criteria. Nonetheless, the two datasets are nearly identical (Fig. 9a), a heartening conclusion. We find the advantages of using *Thellier Auto Interpreter* to be twofold. First, the automatic interpretation of such a large dataset is much faster than manual interpretation (the runtime of *Thellier Auto Interpreter* for this dataset is about 1.5 min versus days for the original). Second, it guarantees a consistent interpretation, which is very difficult to achieve in the manual approach.

[46] After an initial assessment with *Thellier Auto Interpreter*, we use the *Thellier Consistency Test* to re-interpret the dataset using the approach demonstrated in section 5.1. We group samples into *Test Groups* according to their relative location in the hole. We assume for this purpose that samples that are less than 1.5 m apart should have recorded

the same geomagnetic field, hence are grouped together. This approach results in 173 samples grouped in 66 “test groups.” We run the *Thellier Consistency Test* using the following “fixed criteria” at the specimen level: $n_{prm} \geq 2$, $GAP-MAX \leq 0.6$, $MAD < 10.0$, and $DANG < 10$. The threshold value for MAD and $DANG$ was chosen after inspecting the overall quality of the results after the *Thellier Auto Interpreter* run. We choose *STDEV-OPT* as sample mean calculation method, using the same selection criteria as the original publication. In addition, we choose to discard samples whose “acceptable” means vary by 100% or by 20 μT (*int_interval* statistic in Table A3, Supporting Information). We run the *Consistency Test* using β s that ranged from 0.04 to 0.24 in steps of 0.02 and $FRAC$ values that ranged from 0.30 to 0.86 in steps of 0.02. We also used the following *Test Functions*:

- *study_sample_n*
- $((\max_group_int_sigma_uT \leq 5) \text{ or } (\max_group_int_sigma_perc \leq 15)) \text{ and } study_sample_n$.

Figure 9. Analysis of DSDP/ODP submarine basaltic glass data [Tauxe, 2006] discussed in case study 2 (section 5.2). (a,b) A comparison between the original manual interpretation (red) and the automatic interpretation done using *Thellier Auto Interpreter*. Both interpretations use the same acceptance criteria and yield nearly identical results. The automatic procedure, however, takes only about 1.5 min and results with more samples. (c) Results of *Thellier Auto Interpreter* using the optimal values found using the *Consistency Test* (Figure 10b).

[47] The color maps from the *Consistency Test* are displayed in Figure 10. “Strict” criteria (top right of Figures 10a and 10b) result in few samples and low scatter, whereas “loose” criteria (bottom left of Figures 10a and 10b) yield more “successful” samples but much higher scatter. The “optimal” values for *FRAC* and β (0.54 and 0.16, respectively), which result in a total of 123 samples that passed the criteria, are shown in Figure 10b. We use these values to re-run *Thellier Auto Interpreter*, and the results are shown in Figure 9c. The new interpretation is not significantly different from the original publication (Figure 9b), and supports its main conclusions, yet the scatter of the new interpretation is slightly lower than in the original publication. The main advantage of using the approach proposed here is that it provides guidelines for robustly choosing the selection criteria and allows the reader to follow the decisions that lead to choosing the criteria. Moreover, it allows others to reproduce the results and test consequences of other choices, assuming that the authors have made the original measurements available.

6. Summary

- A new GUI for analyzing Thellier experimental data is introduced. The GUI allows the interpretation of the measurements using the conventional (manual) approach, and in addition, it includes two new interpretation tools: (1) *Thellier Auto Interpreter*: an automatic procedure for a fast and

consistent interpretation using paleointensity statistics as acceptance criteria, and (2) *Thellier Consistency Test*, a built-in self-test for the consistency of the results, assuming groups of samples that should give similar paleointensity values.

- We introduce three new paleointensity statistics for assessing the quality of the paleointensity interpretation: *FRAC* – a fraction statistic, *SCAT* – a scatter statistic, and *GAP-MAX* – an upper limit for the gap between data points in an Arai plot.
- The following statistics at the specimen level make a complete list of the required statistics for a robust and consistent interpretation: n_{ptm} , *GAP-MAX*, *MAD*, β , *FRAC*, *DANG*. The required statistics at the sample level are: N_{sample} , and estimated confidence interval in units of both μT and %.
- We implement in the GUI three algorithms for calculating the sample paleointensity: arithmetic mean that minimizes the standard deviation (*STDEV-OPT*), simple bootstrap (*BS*), and parametric bootstrap (*BS-PAR*).
- This study demonstrates that published paleointensity interpretations alone are not sufficient to assert robustness of the conclusions, and measurement data should be made available as part of the publication (as supplementary text files or as a contribution to the MagIC database).

[48] Readers are invited to contact R.S for further clarification, technical guidance, support in converting data files to MagIC format, and assistance in uploading data to the MagIC database. On demand,

Figure 10. Color maps produced by The *Consistency Test* for case study 2 (section 5.2), showing “test functions” for different values of β and *FRAC*.

the authors will provide a script for converting different formats of measurement data files to MagIC format. Suggestions and comments are very much appreciated.

Appendix: Downloads, Datasets, and Tutorial

[49] The software can be downloaded as a part of the PmagPy software from: <https://github.com/ltaxe/PmagPy> and <http://earthref.org/PmagPy/cookbook/>.

[50] A tutorial for *Thellier GUI* can be downloaded from: http://earthref.org/PmagPy/pmagpydocs/Thellier_GUI_tutorial.pdf.

[51] A list of paleointensity statistics used in the *Thellier GUI* can be downloaded from: http://earthref.org/PmagPy/pmagpydocs/Thellier_GUI_Paleointensity_Statistics_Definitions.pdf

[52] The datasets used in this article can be downloaded from the PmagPy examples datasets: http://magician.ucsd.edu/Software/PmagPy/Datafiles_2.0.zip (All the data files used here can be found under the folder Thellier GUI).

Acknowledgments

[53] We would like to thank Jeff Gee and Cathy Constable for constructive comments and suggestions. Thorough and helpful reviews by Greig A. Paterson and an anonymous reviewer greatly improved the quality of this manuscript. This work was partially funded by NSF grants EAR1141840 and EAR1225520 to Lisa Tauxe. We also thank the other members of the MagIC database team for constructing and maintaining the database: Anthony Koppers, Rupert Minnett, Cathy Constable, and Nick Jarboe.

References

- Aitken, M. J., A. L. Allsop, G. D. Bussell and M. B. Winter (1988), Determination of the intensity of the Earth's magnetic-field during archaeological times - reliability of the Thellier technique, *Rev. Geophys.*, **26**, 3–12.
- Ben-Yosef, E., H. Ron, L. Tauxe, A. Agnon, A. Genevey, T. E. Levy, U. Avner and M. Najjar (2008), Application of copper slag in geomagnetic archaeointensity research, *J. Geophys. Res.*, **113**, B08101, doi:10.1029/2007JB005235.
- Ben-Yosef, E., L. Tauxe, T. E. Levy, R. Shaar, H. Ron and M. Najjar (2009), Geomagnetic intensity spike recorded in high resolution slag deposit in Southern Jordan, *Earth Planet. Sci. Lett.*, **287**, 529–539, doi:10.1016/j.epsl.2009.09.001
- Ben-Yosef, E., R. Shaar, L. Tauxe, T. E. Levy and V. Kassinidou (2011), Cyprus Archaeomagnetic Project (CAMP): targeting the slag deposits of Cyprus and the Eastern Mediterranean. *Antiquity*, **85**.
- Biggin, A. J. (2010), Are systematic differences between thermal and microwave Thellier-type palaeointensity estimates a consequence of multidomain bias in the thermal results?, *Phys. Earth Planet. Inter.*, **180**, 16–40, doi:10.1016/j.pepi.2010.03.005.
- Bowles, J., J. S. Gee, D. V. Kent, E. Bergmanis, and J. Sinton (2005), Cooling rate effects on paleointensity estimates in submarine basaltic glass and implications for dating young flows, *Geochem. Geophys. Geosyst.*, **6**, Q07002, doi:10.1029/2004GC000900.
- Coe, R. S. (1967), Paleointensities of the Earth's magnetic field determined from Tertiary and Quaternary rocks, *J. Geophys. Res.*, **72**, 3247–3262.
- Coe, R. S., S. Gromme, and E. A. Mankinen (1978), Geomagnetic paleointensities from radiocarbon-dated lava flows on Hawaii and question of Pacific nondipole low, *J. Geophys. Res.*, **83**, 1740–1756.
- Coe, R. S., J. Riisager, G. Plenier, R. Leonhardt and D. Krasa (2004), Multidomain behaviour during Thellier paleointensity experiments: Results from the 1915 Mt. Lassen flow, *Phys. Earth planet. Inter.*, **147**, 141–153.
- Dunlop, D. J. and Ö. Özdemir (1997), *Rock Magnetism: Fundamentals and Frontiers*, Cambridge University Press, New York.
- Efron, B. (1981), Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, **68**, 589–599.
- Fabian, K. (2001), A theoretical treatment of paleointensity determination experiments on rocks containing pseudo-single or multi domain magnetic particles, *Earth Planet. Sci. Lett.*, **188**, 45–58, doi:10.1016/S0012-821X(01)00313-2.
- Kirschvink, J. L. (1980), The least-squares line and plane and the analysis of paleomagnetic data, *Geophys. J. R. Astron. Soc.*, **62**, 699–718.
- Kissel, C. and C. Laj (2004), Improvements in procedure and paleointensity selection criteria (PICRIT 03) for Thellier and Thellier determinations: application to Hawaiian basaltic long cores, *Phys. Earth Planet. Inter.*, **147**, 155–169, doi:10.1016/j.pepi.2004.06.010.
- Koenigsberger, J. (1936), Die abhaengigkeit der naturlichen remanenten magnetisierung bei eruptivgesteinen von deren alter und zusammensetzung, *Beitriige zur Angewandte Geophysik* **5**, 193–246.
- Levi, S. (1977), The effect of magnetite particle size on paleointensity determinations of the geomagnetic field, *Phys. Earth Planet. Inter.*, **13**, 245–259.
- Leonhardt, R., C. Heunemann and D. Krasa (2004), Analyzing absolute paleointensity determinations: Acceptance criteria and the software ThellierTool4.0., *Geochem. Geophys. Geosyst.*, **5**, Q12016, doi:10.1029/2004GC000807
- Nagata, T., Y. Arai and K. Momose (1963), Secular variation of the geomagnetic total force during the last 5000 years, *J. Geophys. Res.*, **68**, 5277–5282.
- Néel, L. (1955), Some theoretical aspects of rock magnetism, *Adv. Phys.*, **4**, 191–243.
- Paterson, G. A., D. Heslop, and A. R. Muxworthy (2010), Deriving confidence in paleointensity estimates, *Geochem. Geophys. Geosyst.*, **11**, Q07Z18, doi:10.1029/2010GC003071.
- Paterson, G. A. (2011), A simple test for the presence of multidomain behaviour during paleointensity experiments, *J. Geophys. Res.*, **116**, B10104, doi:10.1029/2011JB008369.
- Paterson, G. A., A. J. Biggin, Y. Yamamoto, and Y. Pan (2012), Towards the robust selection of Thellier-type paleointensity data: The influence of experimental noise, *Geochem. Geophys. Geosyst.*, **13**, Q05Z43, doi:10.1029/2012GC004046

- Riisager, P. and J. Riisager (2001), Detecting multidomain magnetic grains in Thellier palaeointensity experiments, *Phys. Earth Planet. Inter.*, **125**, 111–117, doi: 10.1016/S0031-9201(01)00236-9
- Selkin, P., and L. Tauxe (2000), Long-term variations in palaeointensity, *Philos. Trans. R. Soc. London A*, **358**, 1065–1088, doi:10.1098/rsta.2000.0574.
- Selkin, P., J.S. Gee, L. Tauxe, W. Meurer, and A. Newell (2000), The effect of remanence anisotropy on paleointensity estimates: A case study from the Archean Stillwater complex, *Earth Planet. Sci. Lett.*, **182**, 403–416, doi:10.1016/S0012-821X(00)00292-2
- Selkin, P. A., J. S. Gee, and L. Tauxe (2007), Nonlinear thermoremanence acquisition and implications for paleointensity data, *Earth Planet. Sci. Lett.*, **256**, 81–89, doi:10.1016/j.epsl.2007.01.017.
- Shaar, R., H. Ron, L. Tauxe, R. Kessel, A. Agnon, E. Ben-Yosef and J. M. Feinberg (2010), Testing the accuracy of absolute intensity estimates of the ancient geomagnetic field using copper slag material, *Earth Planet. Sci. Lett.*, **290**, 201–213, doi:10.1016/j.epsl.2009.12.022
- Shaar, R., H. Ron, L. Tauxe, R. Kessel and A. Agnon (2011), Paleomagnetic field intensity derived from non-SD: Testing the Thellier IZZI technique on MD slag and a new bootstrap procedure, *Earth Planet. Sci. Lett.*, **310**, 213–224, doi:10.1016/j.epsl.2011.08.024.
- Tauxe, L. (2006), Long-term trends in paleointensity: The contribution of DSDP/ODP submarine basaltic glass collections, *Phys. Earth Planet. Inter.*, **156**, 223–241, doi:10.1016/j.pepi.2005.03.022.
- Tauxe, L. (2010). *Essentials of Paleomagnetism*, University of California Press, Berkeley.
- Tauxe, L., and H. Staudigel (2004), Strength of the geomagnetic field in the Cretaceous Normal Superchron: New data from submarine basaltic glass of the Troodos Ophiolite, *Geochem., Geophys., Geosyst.*, **5**, Q02H06, doi:10.1029/2003GC000635.
- Tauxe, L., and T. Yamazaki (2007), Paleointensities, in *Treatise on Geophysics*, vol. 5, Geomagnetism, edited by M. Kono, pp. 509–563, Elsevier, Amsterdam.
- Thellier, E. (1938), Sur l'aimantation des terres cuites et ses applications, *géophysiques, Ann. Inst. Phys. Globe Paris*, **16**, 157–302.
- Thellier, E. and O. Thellier (1959), Sur L'intensité Du Champ Magnétique Terrestre Dans Le Passé Historique Et Géologique. *Ann. Geophys.*, **15**, 285–376.
- Valet, J. P., E. Herrero-Bervera, J. Carlot and D. Kondopoulou (2010), A selective procedure for absolute paleointensity in lava flows, *Geophys. Res. Lett.*, **37**, L16308, doi:10.1029/2010GL044100.
- York, D. (1966), Least-squares fitting of a straight line, *Can. J. Phys.*, **44**(5), 1079–1086.
- Yu, Y. J., L. Tauxe and A. Genevey (2004), Toward an optimal geomagnetic field intensity determination technique, *Geochem., Geophys., Geosyst.*, **5**, Q02H07, doi:10.1029/2003GC000630.
- Yu, Y. J., and L. Tauxe (2005), Testing the IZZI protocol of geomagnetic field intensity determination, *Geochem. Geophys. Geosyst.*, **6**, Q05H17, doi:10.1029/2004GC000840.
- Zijderveld, J. D. A. (1967), A.C. demagnetization of rocks: Analysis of results, in *Methods in Paleomagnetism*, edited by D. W. Collinson, K. M. Creer, and S. K. Runcorn, pp. 254–286, Elsevier, Amsterdam.