Critical analysis of the Holocene palaeointensity database in Central America: Impact on geomagnetic modelling

Gwenaël Hervé, Mireille Perrin, Luis Alva-Valdivia, Brina Madingou Tchibinda, Alejandro Rodriguez-Trejo, Arnaldo Hernandez-Cardona, Mario Córdova Tello, Carolina Meza Rodriguez

To cite this version:

HAL Id: hal-02109315
https://hal.archives-ouvertes.fr/hal-02109315
Submitted on 24 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Critical analysis of the Holocene palaeointensity database in Central America: impact on geomagnetic modelling

Gwenaël Hervé¹, Mireille Perrin¹, Luis Alva-Valdivia², Brina Madingou Tchibinda¹, Alejandro Rodriguez-Trejo², Arnaldo Hernandez-Cardona², Mario Córdova Tello³, Carolina Meza Rodriguez³

¹ Aix Marseille Univ, CNRS, IRD, INRA, Coll France, CEREGE, Aix-en-Provence, France
² Universidad Nacional Autónoma de México: Instituto de Geofísica, Laboratorio de Paleomagnetismo, Mexico City, Mexico.
³ Instituto Nacional de Antropología e Historia, Centro Morelos, Mexico.
Abstract

Thanks to its rich archaeological heritage, Central America is a key region to recover the past secular variation of the geomagnetic field. This article presents 13 new palaeointensity data on Epiclassic (650 – 900 CE) pottery sherds from Central Mexico. Archaeointensities were determined using the Thellier-Thellier protocol with anisotropy and cooling rate corrections. Average results between 25 and 42 µT reveals a fast secular variation in the second half of the first millennium CE but are not in agreement with global geomagnetic models that predict a higher geomagnetic field strength. To check the reasons of this discrepancy, we compiled all intensity data over the last millennia published in Central America. The Bayesian curve calculated from 194 data covering the last 4 millennia highlights a rapid succession of oscillations of the geomagnetic field strength between 20 and 80 µT. But a critical analysis of the dataset shows a large influence of data quality, 74% of them having a poor cooling unit consistency and experimental quality. The small number of specimens per cooling unit and the anisotropy correction absent or incorrectly made increase the scatter between data, whereas the absence of cooling rate correction biases the dataset towards higher palaeointensity. Discarding these data results in a lower secular variation by removing most extreme values and several intensity oscillations. The weaknesses of the dataset are likely the main reason of the limitations of global models in Central America. Pending the acquisition of new high-quality data, archaeomagnetic dating seems premature in Central America.

Keywords

Geomagnetic secular variation; Palaeointensity; Central America; Ceramics
1. Introduction

Ground-based and satellite measurements of the geomagnetic field, together with numerical models of the geodynamo, provide a priceless insight of the flux dynamic in the Earth’s core. However, this knowledge is limited in time because direct absolute measurements cover a short era, at the most the last four centuries. Beyond, secular variation is recovered from archaeological baked clays, volcanic lava flows and marine or lacustrine sediments. The first two materials acquire a thermoremanent magnetization (TRM) during their last high-temperature heating and give an absolute but discrete estimation of the direction and intensity of the past geomagnetic field. Sediments have the advantage to provide a continuous record of the secular variation but palaeointensity and declination are only relative and require a calibration with absolute data (Panovska et al., 2015).

Several global models of the Holocene secular variation have been developed by inversion of data using spherical harmonic analysis in space (e.g. Constable et al., 2016; Hellio & Gillet, 2018; Licht et al., 2013; Nilsson et al., 2014; Pavón-Carrasco et al., 2014a). These models are powerful tools to investigate the evolution of the geomagnetic patterns at the core-mantle boundary (e.g. Constable et al., 2016) and to scale cosmogenic nuclides production (Lifton, 2016). The local predictions of the global models are also more and more used for archaeomagnetic dating purposes (e.g. Goguitchaichvili et al., 2016). However, these global models tend to smooth the amplitude of the secular variation observed in regional master curves (e.g. Tema et al., 2017 in Hawaii; Cai et al., 2016 in China; Hervé et al., 2017 in Western Europe), especially for the intensity variations and even more when models include relative sedimentary data. Models built only with absolute data would better fit the data but their spatial and temporal distribution is very uneven, with ~70% of them coming from West Eurasia and North Africa and ~70% dated in the last two millennia. Data are especially lacking in the south hemisphere.
Another limitation is the unequal quality of the data. The impact of this factor on geomagnetic modelling is often underestimated, although it has been demonstrated with the European model SCHA.DIF.3k (Pavón-Carrasco et al., 2014b). The archaeomagnetic databases, the most complete one being Geomagia50 (Brown et al., 2015), compile all results without a priori assessment on the quality of the data point. However, the quality is clearly heterogeneous, as depicted by the number of samples, the experimental and temporal uncertainties and the laboratory protocols. Here, we first present thirteen new intensity data acquired on potteries from Chalcatzingo in Mexico, before to show the influence of a critical analysis on the knowledge of secular variation of the geomagnetic field with the example of the intensity in Central America.

2. New archaeointensity data

2.1 Archaeological context and sampling

Chalcatzingo (Lat: 18.6766°N, Long: 98.7705°W) is located in the Amatzinac valley (Morelos state) at circa 120 km South-East of Mexico City (Figure 1). The archaeological site has been excavated since the 30s and presently Mario Córdova Tello and Carolina Meza Rodriguez from the Instituto Nacional de Antropología e Historia (INAH) direct the research program. The site was mainly occupied between 800 and 500 BCE during the Preclassic period. The archaeological remains, the most famous being the Olmec style petroglyphs, attest of the importance of the site in the region (e.g. Grove, 1987). The occupation continued later in the Classic (200-650 CE), Epiclassic (650-900 CE) and Postclassic (900-1500 CE) periods of the Central Mexico chronology.

This study focused on the Epiclassic layers from the Terrace 6 “El Cazador”, close to the main square and the pyramid. We sampled 16 sherds of six pottery-types, characteristic of
this period in the Amatzinac valley (Martin Arana, 1987): negro pulido (2 sherds), anaranjado rojo estriado (3), anaranjado pulido (2), anaranjado burdo (3), estriado sin engobe (3) and Coyotlatelco (3).

Figure 1: Spatial distribution of Central America data over the last 4 kyrs. Archaeomagnetic (volcanic) data are plotted in red (blue). The pink star indicates the location of Chalcatzingo. TMVB is the abbreviation of Trans Mexican Volcanic Belt

2.2 Rock magnetism

Thermomagnetic curves were measured on powders from six sherds using Agico MFK1 apparatus. The variation of the susceptibility was measured during heating to 450 °C or 620 °C and subsequent cooling. All thermomagnetic curves were reversible, highlighting the suitability of Chalcatzingo sherds for archaeointensity experiments. We observed two groups of specimens. The types anaranjado rojo estriado, anaranjado pulido and negro pulido showed a single ferromagnetic carrier, identified as a Ti-poor titanomagnetite by the Curie temperature around 550 °C (Figure 2a). The Coyotlatelco, estriado sin engobe and
anaranjado burdo types presented a second phase with a Curie temperature close to 150 °C (Figure 2b), which can be a Ti-rich titanomagnetite or an epsilon iron oxide (ε-Fe₂O₃) (Lopez-Sanchez et al., 2017).

Figure 2: Representative thermomagnetic curves of Epiclassic sherds from Chalcatzingo.

2.3 Archaeointensity study

Archaeointensity experiments were conducted on 0.6 to 3.0 g specimens embedded in 2 cm cubes of amagnetic plaster. We used the classical Thellier-Thellier method (Thellier & Thellier, 1959) that consists to heat and cool twice the specimens at each temperature step with a laboratory field applied in both directions of a specimen axis (+z, -z). We performed partial thermoremanent magnetization (pTRM) checks every two steps to monitor the absence of alteration of the ferromagnetic mineralogy. Heating was performed in an ASC TD-48SC furnace with a 30 or 40 µT laboratory field using 9-12 temperature steps up to 620 °C. In total
we studied 3 to 5 specimens per sherd for a total of 54 specimens. Measurements were realized on a SQUID cryogenic magnetometer (2G Enterprises, model 755R) at CEREGE. The archaeointensities were corrected for the effects of TRM anisotropy and cooling rate. The TRM anisotropy tensor was determined at 550 °C using 6 positions (+x, -x, +y, -y, +z and –z axes) followed by a stability check (Chauvin et al., 2000). The cooling rate procedure of Gómez-Paccard et al. (2006) was carried out at the same temperature with a slow cooling over 5 h. This duration, explained by experimental constraints, may underestimate the archaeological cooling but an incorrect estimation of the duration seems to have a low impact on the accuracy of the archaeointensity (Hervé et al., accepted). The anisotropy and cooling rate corrections were applied at the specimen level. Almost all specimens presented a secondary component of magnetization. The characteristic remanent magnetization (ChRM) was isolated above 150-500 °C. The secondary component could be acquired during the cooking use of the pot or since the excavations in the archaeological repository. Except for one sherd, the fraction of the NRM in the ChRM (f factor) was higher than our acceptance value equal to 0.35. Forty-five specimens presented a linear NRM-TRM diagram with positive pTRM-checks on the temperature interval of the ChRM (Figure 3a-c). The nine other specimens were rejected because of mineralogical changes, as indicated by non-linear NRM-TRM diagram and negative pTRM-checks (Figure 3d). No results could be obtained for the two sherds of the type negro pulido and for one of the type estriado sin engobe. Accepted specimens fulfil up-to-date quality criteria with a quality factor (q) between 5 and 80 (Table 1S, Supplementary Material). Almost all specimens had a ratio of the standard error of the slope to the absolute value of the slope (β) lower than 0.05, a deviation angle (DANG) lower than 5° and a maximum angular deviation (MAD) lower than 5°.
Figure 3: Representative accepted (a-c) and rejected (d) archaeointensity results with NRM-TRM diagrams, orthogonal plots and demagnetization curves.

Table 1 listed the 13 new average archaeointensities. All were calculated with three or four specimens and have an experimental uncertainty between 1.5 and 9.5%. The range of the average values between 25 and 42 µT tend to indicate a fast secular variation during the second half of the first millennium CE. No clear relationship is observed between the archaeointensity and the pottery-type. On Figure 4a, data are relocated to Mexico City using the Virtual Axial Dipole Moment (VADM) correction and compared to the regional curve of Goguitchaichvili et al. (2018a). The curve was computed using a bootstrap approach from 67 data from Mesoamerica and 17 from southwest United States. Our data matches the curve, except CHAL23 and CHAL41 that are below the curve.
<table>
<thead>
<tr>
<th>Type of pottery</th>
<th>Sherd</th>
<th>$N_{\text{acc}}/N_{\text{meas}}$</th>
<th>$F_{\text{ATRM+CR}}$ ± SD (µT)</th>
<th>$F_{\text{Mexico City}}$ (µT)</th>
<th>VADM (10^{22} A.m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaranjado rojo estriado</td>
<td>CHAL01</td>
<td>4/5</td>
<td>35.1 ± 3.0</td>
<td>35.4</td>
<td>7.9 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>CHAL02</td>
<td>3/5</td>
<td>34.5 ± 2.6</td>
<td>34.8</td>
<td>7.8 ± 0.6</td>
</tr>
<tr>
<td></td>
<td>CHAL03</td>
<td>4/5</td>
<td>34.9 ± 1.0</td>
<td>35.2</td>
<td>7.9 ± 0.2</td>
</tr>
<tr>
<td>Coyotlatelco</td>
<td>CHAL04</td>
<td>3/4</td>
<td>31.5 ± 3.0</td>
<td>31.8</td>
<td>7.1 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>CHAL05</td>
<td>4/4</td>
<td>39.5 ± 2.0</td>
<td>39.9</td>
<td>8.9 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>CHAL06</td>
<td>3/4</td>
<td>42.3 ± 0.7</td>
<td>42.7</td>
<td>9.6 ± 0.2</td>
</tr>
<tr>
<td>Estriado sin engobe</td>
<td>CHAL22</td>
<td>4/4</td>
<td>36.4 ± 2.9</td>
<td>36.7</td>
<td>8.2 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>CHAL23</td>
<td>3/3</td>
<td>28.3 ± 0.9</td>
<td>28.6</td>
<td>6.4 ± 0.2</td>
</tr>
<tr>
<td>Anaranjado burdo</td>
<td>CHAL29</td>
<td>4/4</td>
<td>37.8 ± 0.7</td>
<td>38.1</td>
<td>8.5 ± 0.2</td>
</tr>
<tr>
<td></td>
<td>CHAL30</td>
<td>3/3</td>
<td>39.7 ± 2.4</td>
<td>40.1</td>
<td>9.0 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>CHAL31</td>
<td>4/4</td>
<td>36.9 ± 1.8</td>
<td>37.2</td>
<td>8.3 ± 0.4</td>
</tr>
<tr>
<td>Anaranjado pulido</td>
<td>CHAL40</td>
<td>3/3</td>
<td>41.1 ± 2.0</td>
<td>41.5</td>
<td>9.3 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>CHAL41</td>
<td>3/3</td>
<td>24.7 ± 1.4</td>
<td>24.9</td>
<td>5.6 ± 0.3</td>
</tr>
</tbody>
</table>

Table 1: Average archaeointensities of Epiclassic sherds from Chalcatzingo. Columns from left to right: type of potteries, name of the sherd, number of accepted over measured specimens, average archaeointensity corrected for the effects of TRM anisotropy and cooling rate with its standard deviation, average archaeointensity relocated to Mexico City, Virtual Axial Dipole Moment.
Global models CALS10k.2 (Constable et al., 2016), SHA.DIF.14k (Pavón-Carrasco et al., 2014a) and COV-ARCH/COV-LAKE (Hellio & Gillet, 2018) give consistent predictions at Mexico City. SHA.DIF.14k and COV-ARCH were calculated using archaeomagnetic and volcanic data, whereas CALS10k.2 and COV-LAKE also include lacustrine and marine sedimentary data. Models also differ by the inverse method. CALS10k.2 and SHA.DIF.14k imposed regularizations in space and time. The stochastic approach of COV-ARCH and COV-LAKE used the temporal statistics from satellites, ground-based and palaeomagnetic data as prior information through cross-covariance functions. The four models fit only the highest Chalcatzingo values. Their predictions are usually above the regional master curve of Goguitchaichvili et al. (2018a) and predict a slower secular variation.

Figure 4: Comparison of Chalcatzingo data relocated to Mexico City (in pink) with the bootstrap regional curve of Goguitchaichvili et al. (2018a) (in grey) and with the predictions of global models CALS10k.2 (Constable et al. 2016, blue curve), SHA.DIF.14k (Pavon-Carrasco et al, 2014a, orange curve), COV-ARCH and COV-LAKE (Hellio & Gillet, 2018, green and purple dotted curves).
In order to check the reason of these discrepancies, we compiled all intensity data from Central America from the original articles (Figure 5a). We carried out the critical analysis only on the last four millennia because data are much too scarce before (5 data between 5000 and 2000 BCE).

A data point is defined as a cooling unit that acquired TRM at the same time, such as a volcanic lava flow, an archaeological kiln or fireplace and a single pottery. For ceramics, a data point is sometimes defined in the literature as the average of independent pots (e.g. Genevey et al., 2016; Gómez-Paccard et al., 2016; Hervé et al., 2017), because they were discovered in a single homogeneous short-lived archaeological feature. This is not the case in Chalcatzingo and in most studies. Even if ceramics were dated in the same period, the probability of no contemporaneous cooling is high, because sherds came from different archaeological layers or from a single layer associated to a long-lived settlement.

The total number of intensity data in Central America over the last four millennia is 194 (including the 13 new data from Chalcatzingo) out of 40 different papers (Tab. 2S in Supplementary Material). 86% of these studies have been published after the 2000s. Data from 16 studies are not yet included in the Geomagia database (this study; Böhnel et al., 2016; Cifuentes-Nava et al., 2017; Goguitchaichvili et al., 2017, 2018b, 2018c; Herrero-Bervera, 2015; Lopez-Tellez et al., 2008; Mahgoub et al., 2017a, 2017b; Michalk et al., 2010; Morales et al., 2012, 2013, 2015; Rodriguez-Ceja et al., 2012; Terán et al., 2016). With this update, our dataset is almost twice larger than the one used in a recently published compilation of Mesoamerica (Goguitchaichvili et al., 2018a) that contains 106 data with only one study (Böhnel et al., 2016) published in the last five years.

Data are concentrated in Central Mexico around the 20°N parallel and in the Mayan area (Chiapas-Guatemala) (Figure 1). Archaeological baked artefacts (here bricks, burnt walls,
burnt soils and ceramics) constitute 87% of the dataset (168 data), the rest coming from volcanic lava flows of the Trans Mexican Volcanic Belt (TMVB). Only 17 data (9%), all volcanic, are full vector data.

Fifty-four cooling units were dated by radiocarbon. When the uncalibrated age was provided, we updated the date of the archaeomagnetic site using the most recent calibration curve IntCal13 (Reimer et al., 2013). In the case of Xitle lava flow, we used the newest age, 1670±35 uncal. BP, of Siebe (2000). Twelve data come from historical eruptions since the Spaniards conquest, the youngest being the Paricutin in 1943-1952. For the 115 others data (i.e. 58%), the dating was defined by stratigraphy or typology of archaeological artefacts. The temporal density of data is higher in the periods of Late Preclassic (circa 400 BCE – 200 CE), Classic (circa 200 – 650 CE) and the youngest half of Postclassic (circa 1200 – 1500 CE) (Figure 5a). Two important gaps are observed, the first between 800 and 400 BCE and the youngest around 1000 CE.

The intensity values over the last 4 kyrs range between ~15 µT and ~90 µT (Figure 5b). The dataset looks cloudy, especially in the third and four centuries CE with values from 15 µT up to 75 µT. The spread is visually increased by the fact that many data are dated in the same age interval corresponding to an archaeological period, generally wide of several centuries. This tends to divide the dataset in successive time slices, modelled on the archaeological chronologies. The variability of intensities may reflect the secular variation within each archaeological period.

An intensity secular variation curve was computed using a Bayesian framework (Hervé et Lanos, 2018; Lanos, 2004; Schnepp et al., 2015). The curve is given as a smooth continuous curve obtained by averaging cubic splines and took into account experimental and age uncertainties as prior information. The misfit of each data to the curve is minimised by exploring the multidimensional space of probability densities using Monte Carlo Markov
Chains. The curve presents a rapid succession of oscillations of the geomagnetic field strength, especially from 400 BCE to 1500 CE, at the time of the highest data density. The oscillations represent the best fit to the intensity variability within the time slices related to age uncertainties. They can imply a fast secular variation of many dozens of µT over a few centuries, similar to the rates observed in the Levantine area at the beginning of the first millennium BCE (Shaar et al., 2011). But it can also reflect experimental problems and a critical analysis of the database has to be performed, before to evoke a geomagnetic phenomenon.

4. Critical analysis

4.1 Challenges in palaeointensity determination

The linearity between the intensity of the thermoremanent magnetization and the past geomagnetic field strength is the physical basis of the palaeointensity determination. Acquiring reliable data remains a challenging task because several effects can lead to a departure away from the linearity (see Dunlop 2011 for a review). The required reciprocity of blocking and unblocking temperatures is only verified for single-domain grains without interactions. Alteration of the magnetic grains when heated at the laboratory is another cause of failed experiments. Moreover, palaeointensities should be determined only on the Characteristic Remanent Magnetization (ChRM) acquired during the last heating above the Curie point. A calculation including a secondary component of magnetization can differ by a few dozens of µT from the intensity calculated on the sole ChRM (Hervé et al., 2013).

The Thellier-Thellier protocol (Thellier & Thellier, 1959) and its derived versions are admitted as the most reliable to control the respect of these conditions thanks to the
progressive heating steps and the use of pTRM-checks. But the multiple heatings increase the risk of alteration and alternative protocols were developed.

The microwave (MW) technique (Walton, 1991; Walton & Böhnel, 2008) is based on microwave absorption by magnetic grains using progressive frequency steps, which reduces the heating temperature of the sample. This method does not exactly replicate the acquisition process of the initial TRM, as ferromagnetic grains are excited by magnons instead of phonons in the usual thermal techniques.

The multispecimen (MSP) method is performed on a set of specimens from the same sample. All specimens are heated once at the same temperature (usually in the 300-400 °C temperature range) in different laboratory field values applied parallel to the NRM (Dekkers & Böhnel, 2006). The comparison of the demagnetized NRM intensity with the one of the acquired partial-TRM provides the palaeointensity. The Central American MSP data were obtained with the original protocol of Dekkers & Böhnel (2006), which do not correct for the NRM fraction and domain state as advised by Fabian & Leonhardt (2010). They are therefore likely overestimated (Schnepp et al., 2016).

Finally, the Shaw method (Shaw, 1974) compares the alternating field (AF) demagnetization of the NRM with the one of a full laboratory TRM. The scaling factor between the two demagnetizing curves is an estimate of the ratio between ancient and laboratory fields. But the linearity between the magnetization and geomagnetic field cannot be checked and the palaeomagnetic community generally considers this method of lower quality.

It is well known that the cooling rate has an important effect on the palaeointensity of archaeological baked clays. The TRM intensity of single-domain grains increases with the cooling rate duration (e.g. Dodson and McClelland-Brown, 1980). As the laboratory cooling is faster than the archaeological cooling, uncorrected palaeointensities are generally overestimated. In the case of the volcanic lava flows, the correction is usually not considered
necessary, because the coarser mineralogy makes them less sensitive to the cooling rate
effect. For the thick Xitle lava flow, the correction improves the precision of the
palaeointensity but does not change significantly the palaeointensity average (Morales et al.,
2006) (Table 4S).

The anisotropy correction is also crucial depending on the type of material. The preferential
alignment of the iron oxides during the manufacturing process of baked clays results in
distortion of the NRM with respect to the surrounding geomagnetic field (Veitch et al., 1984).
Potteries are generally anisotropic and palaeointensity can sometimes be biased up to a factor
of two (e.g. Hervé et al., 2017; Osete et al., 2016). The effects can also be high in burnt walls
and soils (Palencia-Ortas et al., 2017; Molina-Cardin et al., 2018). Conversely, the influence
of the TRM anisotropy is assumed negligible in lava flows. But the robustness of this
presupposition can be questioned, as the anisotropy of magnetic susceptibility (AMS) is often
significant and used as a proxy for flow direction (e.g. Fanjat et al., 2012).

Two different anisotropy corrections were used in the Central American dataset, first the
determination of the TRM tensor at the specimen level (Chauvin et al., 2000, see section 2.3)
and secondly a method, peculiar to Mexico, that we called in the following Mean(XYZ)
method. Each baked clay fragment is cut in 6 mini-specimens keeping the same orientation.
The laboratory field during the Thellier experiments is applied along +x axis for one
specimen, -x for another one and +y, -y, +z or -z axis for the four others. Averaging of the six
individual intensity estimations is considered to properly correct the TRM anisotropy effect.
Poletti et al. (2016) questioned the reliability of this approach. They calculated the TRM
tensor at two different temperatures and then roughly determined the intensities from the
slope between the two +x TRMs (and -x…) on the NRM-TRM diagram. Their results show
that the Mean(XYZ) method results in imprecision and sometimes in inaccuracies up to 10
µT.
4.2 Validity of the Mean(XYZ) anisotropy correction

To further test the reliability of this correction, we performed the Mean(XYZ) and the TRM tensor methods on five samples. The first two, 11369B-1 and 11369B-11, are small bricks baked in an experimental kiln in Sallèles-d’Aude (southern France). The last three, SAQ48, SAQ51 and SAQ54 are pottery sherds discovered in a sedimentary core from Saqqara (Egypt) and supposed to be from the Pharaonic era. Samples were cut in six identical cubic specimens, each positioned in a different position (+x, -x, +y, -y, +z or –z) in the laboratory furnace during the classical Thellier-Thellier protocol and the cooling rate correction. The tensor of TRM anisotropy was then determined with the same protocol as for Chalcatzingo samples. All specimens provided technically reliable intensity results (Figure 1S and Table 3S in Supplementary Material). The degrees of TRM anisotropy have close values between 1.19 and 1.26 (Table 2).

Table 2 compares the averages intensities of the Mean(XYZ) and TRM anisotropy tensor methods, both being corrected for cooling rate effect. Without the TRM tensor correction, the palaeointensity values at the sample level are distributed over a large range up to 26 µT with systematic differences between axes. The highest palaeointensities values, i.e. the lowest acquired TRMs, are observed for specimens with the laboratory field applied along x axis, perpendicularly to the flattening plane of the brick or the pottery (Veitch et al., 1984). The correction by the TRM tensor groups the palaeointensities, which results in a standard deviation up to five times lower. The Mean(XYZ) average palaeointensity differ of the TRM tensor value by only -5.6 to 3.8% (IE parameter in Table 2). No relationship seems exist between IE and the anisotropy degree or the standard deviation.

The anisotropy correction depends on the NRM direction within the TRM tensor. In our examples, the Mean(XYZ) palaeointensity is close to the TRM tensor value, because the
NRM direction is intermediate between x, y and z axes and because averages are computed from six specimens with an equal contribution of the three axes. Data with two, and of course one, specimens cannot be considered as corrected by the anisotropy.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean(XYZ)</th>
<th>TRM tensor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{\text{min}}$</td>
<td>$F_{\text{max}}$</td>
</tr>
<tr>
<td>11369B1</td>
<td>45.0</td>
<td>52.3</td>
</tr>
<tr>
<td>11369B11</td>
<td>45.0</td>
<td>56.6</td>
</tr>
<tr>
<td>SAQ48</td>
<td>65.8</td>
<td>79.4</td>
</tr>
<tr>
<td>SAQ51</td>
<td>74.8</td>
<td>100.7</td>
</tr>
<tr>
<td>SAQ54</td>
<td>74.9</td>
<td>99.3</td>
</tr>
</tbody>
</table>

Table 2: Comparison of intensities determined with the two corrections of anisotropy. For both methods are indicated the minimal and maximal specimen intensity, the mean intensity and its standard deviation in microTeslas and in percentage. All intensities are corrected for the cooling rate effect. IE is the difference in percentage of the Mean(XYZ) intensity with the TRM tensor intensity. $k_{\text{max}}/k_{\text{min}}$ is the average degree of TRM anisotropy.

A balance between the three axes does not mean that one can be fully confident in the average palaeointensity value. If the NRM is parallel to one of the three axes, the two specimens with the laboratory field on this axis will give the true palaeointensity and the inaccuracy can be larger. In the case of SAQ51, if the NRM were close to x-axis, the true palaeointensity would have been around 102-103 µT, whereas the average of the six specimens is 85.5±11.5 µT (Tab. 3S, Supplementary Material).
For all these reasons, the Mean(X-Y-Z) method enhances noise in a dataset with larger standard deviations and possible high inaccuracies. All Mean(XYZ) palaeointensities can be biased even if the risk is reduced for those calculated with a balanced participation of the three x, y and z axes. Data corrected with the TRM tensor are clearly much more reliable from an experimental point of view.

4.3 Data selection

Estimate a posteriori the quality of a palaeointensity is not straightforward without the raw measurements. Four intensities from Rodriguez-Ceja et al. (2009, 2012) have been defined on a secondary component and clearly have to be removed. Another obvious criterion is the expected internal consistency within a given cooling unit. We classified data in three categories according to the number of specimens and the standard deviation (Figure 5c). Highest quality data (category C1) have at least 3 specimens per cooling unit and a standard deviation (SD) lower than 10%. Those in the second category (C2) have a SD lower than 15%. Other data (C3) represent 38% of the full dataset (i.e. 72 data) in similar proportions for volcanic and archaeomagnetic data (Figure 5c). One can note that most extreme values are included in this category.
Figure 5: Critical analysis of the Central American palaeointensity dataset. (a) Temporal distribution of data over the last 7 kyrs. (b) Compilation of all archaeomagnetic (in red) and volcanic (in blue) data. The grey curve with its 1σ confidence envelope is computed using a Bayesian method. (c) Ranking of the dataset according to the cooling unit consistency. (d) Ranking according to the quality of the intensity protocol. (e) Ranking according to the consideration of the anisotropy effect. For each criterion, data are ranked in three levels C1, C2 and C3 from the highest to the lowest quality. Plots (c-d-e) excluded four data defined on a secondary component from Rodriguez-Ceja et al. (2009, 2012). TH: Thellier-Thellier protocols; Multispecimen-DB: original multispecimen protocol of Dekkers and Böhnel (2006); N: number of data; For data using Mean(XYZ) anisotropy correction, n: number of specimens used in the average intensity.
The criterion of the consistency is useful but cannot screen all problems in the palaeointensity protocol (Figure 5d). For the reasons mentioned above, the most reliable protocol is the Thellier-Thellier protocol with pTRM-checks. This category (C1) is dominant in the dataset (120/190 data). Our mid-quality category (C2) groups microwave data and Thellier-Thellier data without pTRM-checks. We did not assign the latter to the lowest quality category because often it corresponds to the oldest data, as it is the case here (51 data almost all from Bucha et al., 1970 and Lee, 1975), when pTRM-checks were not commonly used. Furthermore, the interpretation of pTRM-checks is not straightforward and experience shows that reliable palaeointensity can be obtained with moderate deviation if the linearity is well maintained over a large NRM proportion. However, the absence of alteration monitoring casts a doubt on the reliability of these data. Finally, the low-quality category (C3) includes the nine multispecimen data without the correction of the NRM fraction (MSP-DB), because they likely overestimate the palaeointensity. The five data acquired with the Shaw protocol also belong to this category.

As archaeological baked clays constitute 86% of the dataset (164/190 data) whose 85% are ceramics (139 data), cooling rate and anisotropy corrections are crucial. For archaeological baked clays, we demonstrated that the anisotropy correction with the TRM tensor is the most reliable (category C1). Two microwave data on potteries associated to the Xitle lava flow are also included in C1 (Böhnel et al., 2003), because the laboratory field was applied parallel to the NRM. Only 27 data are part of this high-quality category (Figure 5e). Both corrections are assumed not mandatory for igneous material. But this is not fully attested and we preferred to assign these data to the mid-quality C2 category. Regarding the Mean(XYZ) correction, the balanced contribution between x, y, and z axes can be attested only for the 25 average intensities calculated with six specimens. We classified them in the C2 category, bearing in
mind that some can be inaccurate especially if the NRM was parallel to a specimen axis.

Another number of specimens gives preponderance to one or two axes, which can result in larger inaccuracies and justifies their attachment to the C3 category. Data calculated with three and four specimens are actually more scattered (Figure 2S, Supplementary Material).

The C3 category finally includes the 61 archaeological data without any anisotropy correction, 11 of them being corrected for the cooling rate effect.

The synthesis of this critical analysis ranks the dataset in three groups, C1, C2 and C3. Data of the high quality group (C1) belong to the three C1 categories of the cooling unit consistency, protocol and anisotropy and cooling rate corrections at the same time. They are only 23, i.e. 12% of the initial dataset, (Chalcatzingo data, this study; Fanjat et al., 2013; Rodriguez-Ceja et al., 2009, 2012 – only intensities determined on the characteristic magnetization), all dated between 400 BCE and 900 CE (Figure 6a).

The group C2 gathers data classified at least in the C2 category for the three criteria. The El Jorullo and Xitle lava flows were duplicated in respectively two and four studies and we computed new averages from the same accepted specimens as the authors (Table 4S, Supplementary Material). In the case of the Xitle, the average (59.8 ± 5.0 μT) takes into account only the palaeointensities corrected for the cooling rate effect on the TRM intensity (Figure 6a).

All other data, those belonging to at least one C3 category, constitute 72% of the dataset (140 data) with almost all the oldest acquired data (Bucha et al., 1970; Lee, 1975). These 50 data on potteries with no anisotropy and cooling rate corrections provide higher palaeointensities than other data (blue circles, Figure 5e). If the weak cooling consistency influences their scatter, the overestimation very likely results from absence of the cooling rate correction.
Figure 6: (a) Secular variation inferred from high-quality data. (b) Comparison of the two Bayesian curves calculated from the total and selected dataset (blue and red curves) with the master curve of Goguitchaichvili et al. (2018a) (grey curve) and with COV-ARCH global model (green dotted curve).

6. Discussion
6.1 Impact of the critical analysis on the knowledge of the secular variation

The large number of palaeointensities available for the last kyrs in Central America may give the impression that the secular variation curve is well-known, with large and fast variations between 20 and 80 µT. However, the critical analysis shows that many of these oscillations
are induced by experimental errors. The main sources of scatter are the absence of cooling unit consistency, the lack or inaccuracy of anisotropy correction and the absence of cooling rate correction on archaeological artefacts. Only 12% of data (23 data) can be considered as high quality (C1 category) that is not enough to compute an intensity secular variation curve. Therefore we also included the 14% of the data (27 data) corresponding to the middle quality (C2 category) (Figure 6a). All data used in the Bayesian curve are plotted per study in Figure 3S (Supplementary Material).

It is worth pointing out that the selection does not take into account the age uncertainties and inaccuracies in age, perhaps affecting some of the selected data. That could be the case for the Xitle lava flow, whose intensity seems more consistent with an age at the turning point of the Christian era than in the IVth and Vth centuries CE (Figure 6a). We decided to use the most recently acquired age 1670±35 uncal. BP (Siebe, 2000) on a charcoal included in the lava flow sampled for the palaeomagnetic studies. But other published radiocarbon ages, 1945±55 uncal BP and 2025±55 uncal BP (Delgado et al., 1998; see also Böhnel et al, 1997), dates the eruption of this monogenetic volcano in a more consistent way with other available quality data.

Selecting data has a strong influence on the secular variation reconstruction, as shown by the comparison of the averages Bayesian curves calculated from the entire and selected datasets (blue and red curves, Figure 6b). After discarding extreme values, the secular variation has smaller amplitude between 20 and 60 μT. The better consistency of the dataset also removes most short-term oscillations of the curve calculated with all data.

The secular variation from ~400 BCE to ~500 CE is rather well constrained by high/middle quality C1 and C2 data. They highlight a fast variation with a ~45 μT maximum in ~400 BCE, a ~25 μT minimum in ~200 BCE, a further ~50 μT maximum circa 0 CE and a ~20 μT
minimum in ~200 CE. Next, the field varied more slightly between 30 and 40 μT from ~400 to ~800 CE that is the period presenting the highest density of quality data.

On the other side, the lack of data and the large age uncertainties make unclear the secular variation between 2000 and 400 BCE. Address this gap is crucial to better understand the highest secular variation and the largest departure from an axial dipolar field observed at this period in Eurasia (Hervé et al., 2017; Shaar et al., 2016).

Considering only high or middle quality data drastically reduces the available data set and one may wonder if we do not miss some geomagnetic variations. For example, there are only five high-quality data in the second millennium CE and it is not surprising to see that the curves differ during this period. If the high geomagnetic field strength around 50 μT is well constrained, the number of maxima, one or two, is unclear. Unselected data indicates a possible minimum ~1300 CE. The timing of the intensity increase is also not clear, either from 1000 CE with C1 and C2 categories or from 800 CE with unselected data.

Finally, the critical analysis of the Central America dataset drastically highlights the need of new high quality data with precise dating for almost all periods. It is clearly a prerequisite to ascertain the secular variation features and to compare the Central American curve to other regional records.

6.2 Comparison with others curves and global models

Our selected curve presents several differences with the bootstrap curve of Goguitchaichvili et al. (2018a), especially between 500 BCE and 500 CE (Figure 6b). The major discrepancy occurs around 200 CE when they have a maximum and us a minimum. Different reasons could be the source of this inconsistency: i) their critical analysis (at least 4 specimens, standard deviation lower than 10 μT, corrections of anisotropy and cooling rate) is pretty
similar to ours except for the inclusion of all Mean(XYZ) data; ii) the dataset of Goguitchaichvili et al. (2018a) duplicated data from the same lava flow (e.g. Xitle, El Jorullo, Ceboruco, Paricutin) and iii) included 17 data from southwest USA but not 28 data from Mexico out of the 50 that we classified in the C1 and C2 categories. Some differences between the curves are also related to the computation technique. The variability of intensity data circa 200 BCE is smoothed by the bootstrap method, whereas the Bayesian of Lanos (2004) interprets it by an intensity minimum. Both updated dataset and Bayesian approach likely provides a finer knowledge of the secular variation curve, even though as mentioned before it has to be confirmed by new reliable data.

The global models such as COV-ARCH are built using the Geomagia database without data selection. They are inconsistent with our curve as well as with Goguitchaichvili et al. (2018a) curve (Figure 6b). First, they predict a smoother secular variation after 500 BCE with intensity varying between 40 and 50 µT, this range corresponding to the average intensity of the total dataset. The model did not succeed to interpret the high data variability and remains blocked in this intensity range. Another inconsistency is the shift to higher intensity values in comparison to our curve. The data of Bucha (1970) and Lee (1975), not corrected for the cooling rate, constitute around 40% of the Central American data in Geomagia. The overestimation emphasizes the need for a selection with the cooling rate correction.

7. Conclusions

The acquisition of 13 new palaeointensities on Epiclassic (650 – 900 CE) sherds from Chalcatzingo completes the Central American database. The up-to-date compilation of 194 available data shows a high variability in palaeointensity during the last 4 millennia. A critical analysis of the database highlights that this variability cannot be only explained by a geomagnetic origin. Selecting data according to the cooling unit consistency and the intensity
protocol discards extreme values and predicts less oscillations of the geomagnetic field strength. The results emphasize the importance of the cooling rate correction. Its absence yields to a shift of the geomagnetic reconstructions towards higher intensity values. Another problem, peculiar to the Central American dataset, is the use of an inaccurate anisotropy correction, here called Mean(XYZ). According to experimental test, this method leads to systematic larger imprecisions and to possible inaccuracies up to 15 µT.

The example of Central America highlights the need of data selection global geomagnetic model. The absence of data selection clearly plays a role in the smoothing and overestimation of the curves inferred from global models compared to our master curve. Considering these problems, the use of current global models for dating purposes in Central America seems inappropriate.

Defining acceptance criteria is not straightforward in global modelling. Our criterion of the cooling unit consistency discards more than 50% of the global dataset and those based on the protocol ~90% (counting on Geomagia50.v3 in October 2018). This low proportion is hardly compatible with the homogeneous data coverage in space and in time that requires the spherical harmonic analysis. Our acceptance limits have to be lowered to better fit with the constraints of global modelling. Priority criteria could focus on parameters biasing the palaeointensity average, such consistency criteria, cooling rate correction for archaeological baked clays and anisotropy correction for potteries.

Acknowledgements

Funding was provided by ANR-CONACYT SVPIntMex project ANR-15-CE31-0011-01 coordinated by M. P. and L. A-V. G.H. also benefited from a support of Campus France PRESTIGE program (PRESTIGE-2017-1-0002). We thank the editor and the reviewer for
their careful work. We are grateful to Gabrielle Hellio (Univ. Nantes) for transmitting the
COV-ARCH and COV-LAKE predictions at Mexico City and to Philippe Lanos for
providing the software to calculate Bayesian local secular variation curves. We acknowledge
Mohamed Hamdan (Giza University) and Ahmed Saleh (NRIAG) for providing the pottery
sherds from Saqqara.

References

Böhnel, H., Morales, J., Caballero Miranda, C., Alva, L., McIntosh, G., Gonzalez, S., &

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

dating of two recent lava flows from Ceboruco volcano, western Mexico. *Geophys. J.

Brown, M. C., Donadini, F., Korte, M., Nilsson, A., Korhonen, K., Lodge, A., Lengyel, S. N.,
& Constable, C. G., 2015. GEOMAGIA50.v3: 1. general structure and modifications to

http://doi.org/10.1186/s40623-015-0232-0

during the past 3000 years in the western hemisphere. *Science*, 168, 111-114.

results spanning the past 6 kiloyears from eastern China and implications for extreme


Fabian, K., & Leonhardt, R., 2010. Multiple-specimen absolute palaeointensity determination: an optimal protocol including pTRM normalization, domain-state


http://doi.org/10.1007/s11200-016-0851-3


http://doi.org/10.1016/j.jasrep.2017.05.030


Hervé, G., & Lanos, P., 2018. Improvements in archaeomagnetic dating in Western Europe
from the Late Bronze to the Late Iron ages: an alternative to the problem of the Hallstattian radiocarbon plateau. *Archaeometry*, 60, 4, 870-883, doi:10.1011/arcm.1234


Lee, S., 1975. *Secular variation of the intensity of the geomagnetic field during the past 3,000 years in North, Central and South America*. University of Oklahoma.


Lopez-Sanchez, J., McIntosh, G., Osete, M. L., del Campo, A., Villalain, J. J., Perez, L.,


Figure 1S: Representative archaeointensity results of specimens used to test the two anisotropy protocols. The laboratory field was fixed to 60 μT. The specimen on the left is a brick from Sallèles-d’Aude and the one on the right a pottery sherd from Saqqara.

Figure 2S: Influence of the number of specimens on the Mean(X-Y-Z) palaeointensities. Data with highest quality protocol are plotted in red. The Bayesian curve, the same as in Figure 6a, is calculated from C1 and C2 data.

Figure 3S: Highest quality data (C1 and C2 categories) per study. Data from Chalcatzingo (this study) are not plotted here.