Archeomagnetic intensity variations during the era of geomagnetic spikes in the Levant

Philip W Livermore, Yves Gallet, Alexandre Fournier

Physics of the Earth and Planetary Interiors 312 (2021) 106657

Contents lists available at ScienceDirect

Physics of the Earth and Planetary Interiors

journal homepage: www.elsevier.com/locate/pepi

Archeomagnetic intensity variations during the era of geomagnetic spikes in the Levant

Philip W. Livermore [a,*], Yves Gallet [b], Alexandre Fournier [b]

[a] School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK
[b] Université de Paris, Institut de Physique du Globe de Paris, CNRS, Paris, France

ARTICLE INFO

Keywords:
Archeomagnetic intensity variations
Geomagnetic spikes
Uncertainty quantification
Geodynamo
Bayesian

ABSTRACT

Observational records of rapidly varying magnetic fields strongly constrain our understanding of core flow dynamics and Earth's dynamo. Archeomagnetic analyses of densely sampled artefacts from the Near-East have suggested that the intensity variation during the first millennium BCE was punctuated with two geomagnetic spikes with rates of change of intensity exceeding 1 μT/yr, whose extreme behaviour is challenging to explain from a geodynamo perspective. By applying a new transdimensional Bayesian method designed to capture variations on both long and short timescales, we show that the data considered only at the fragment (thermal-unit) level require a complex intensity variation with no less than six spikes, each with an approximate duration of between 30 and 100 years. However, the nature of the inferred intensity evolution and the number of spikes detected are fragile and highly dependent on the specific treatment of the archeomagnetic data. No spikes are observed when the data are considered only at the level of a group of fragments from the same archeological context, with a minimum of three different artefacts per context. Furthermore, the number of spikes decreases to zero when increasing the error budget for the intensity at the fragment level within reasonable levels of 3–6 μT and the data age uncertainty up to 50 years. Of the six spikes found, the most resilient when increasing the error budget was dated at ~970 BCE. However, we show that even this spike sensitively depends on the age model proposed for data from the Levant archeological site Timna-30 and disappears when considering a single Gaussian age prior distribution for these data and a moderate minimum intensity error. Thus, depending on the choices made, the Near-Eastern data are compatible with a broad range of time-dependence, from six spikes at one extreme to zero spikes on the other. An error of 6 μT at the fragment level produces a spikeless model with strong similarity with the reconstruction from the SHAWQ-Iron Age global model with rates of change of ~0.2–0.3 μT/yr.

1. Introduction

Anomalous behaviour of the past geomagnetic field, captured by materials carrying an ancient magnetization, can have profound consequences for our understanding of the Earth's magnetic field and how it is generated. Our focus here is on archeomagnetic data from the Near East during the first half of the 1st millennium BCE, which suggest a new type of geomagnetic event, the geomagnetic spike (Ben-Yosef et al., 2009; Shaar et al., 2011) characterised by extreme rates of intensity change with estimates ranging $\sim 0.75 - 4$ μT/yr (Livermore et al., 2014; Shaar et al., 2016; Ben-Yosef et al., 2017; Shaar et al., 2011), many times larger than the maximum rate of change in the present day (~0.12 μT/yr, see below). Importantly, these spikes have no analogue in the historical and modern day geomagnetic field whose structure is much better resolved both in space and time; furthermore the extreme behaviour of the spikes is not reproducible in numerical models (Davies and Constable, 2017, 2018; Troyano et al., 2020). If the geomagnetic spikes stand up to scrutiny, their explanation may therefore require new theories of geomagnetic field evolution. Yet tracing the details of any rapid variations in the geomagnetic field that may have occurred is a challenging exercise, despite the Near Eastern record being temporally well-sampled in archeomagnetic terms, with an average of ~20 analyzed fragments from independent archeological artefacts per century during the first half of the 1st millennium BCE. A primary reason for this is the inherent difficulty of assembling the dataset of intensity and age values of artefacts having experienced heating, along with estimates

* Corresponding author.
E-mail address: p.w.livermore@leeds.ac.uk (P.W. Livermore).

https://doi.org/10.1016/j.pepi.2021.106657
Received 25 September 2020; Received in revised form 13 January 2021; Accepted 27 January 2021
Available online 4 February 2021

of their uncertainties. A second difficulty is in the analysis technique: there is the need for a method to fit an intensity variation curve through the dataset with quantification of its uncertainty, which takes adequate account of both intensity and age uncertainties. In this study, we apply the recently developed transdimensional Bayesian method (Livermore et al., 2018) for the calculations of regional geomagnetic intensity variation curves, which relies on minimal a priori information and innately takes account of uncertainties in both intensity and age, and importantly here makes no assumption on inherent time-scales through any form of explicit regularisation. This method is therefore particularly well suited to examine the extreme fluctuations in the geomagnetic field proposed in the Near East. It should be noted that other geomagnetic spikes have been proposed in other regions of the world (e.g. Texas, China, New Zealand: Bourne et al. (2016); Cai et al. (2014, 2017); Turner et al. (2020), respectively), some at dates different from the 1st millennium BCE, but they remain beyond the scope of this study.

Thanks to the richness of its archeology, numerous archeomagnetic studies have been carried out in the Near East, mainly focused on the variations in geomagnetic field intensity over the past few millennia. All these studies agree that the first half of the 1st millennium BCE was marked by high geomagnetic field intensity values, probably the highest known so far throughout the Holocene (e.g. Genevey et al., 2003, 2008; Ben-Yosef et al., 2009). However, the published data do not show a smooth albeit high intensity trend, but rather a complex picture with both extremely high intensity values of more than 100 μT (Ben-Yosef et al., 2008, 2009), interspersed with moderately high values (of $70-80$ μT). Many of the data came from metallurgical residues (slag), whose stratigraphy place strong constraints in the form of small relative ages of the studied fragments. To describe the data, Ben-Yosef et al. (2009) coined the term of geomagnetic spike to describe, on the one hand, the extreme values of archeointensity and, on the other, the high variation rates (also extreme) implied by the intensities and ages reported. These extreme intensity peaks, short-lived over only a few decades, are quite different to the much longer and broader intensity peaks that have been associated with (i) archeomagnetic jerks, one of which was proposed at the beginning of the 1st millennium BCE Gallet et al. (2003, 2006) as well as (ii) the peaks observed in Western Europe during the 1st millennium BCE which show much lower intensity variation rates on the order of $0.10-0.25$ μT/yr (Hervé et al., 2017; Osete et al., 2020).

Following Ben-Yosef et al. (2009), Shaar et al. (2011) proposed the occurrence of two geomagnetic spikes at the beginning of the 1st millennium BCE, at ~980 BCE and ~890 BCE, from the study of a chronological sequence of slag layers (archeological site Timna-30 in Israel). These geomagnetic spikes are characterised by intensity peaks, which converted in VADM (Virtual Axial Dipole Moment) reach values in excess of 20×10^{22} Am^2 and variation amplitudes in the order of $2-3 \times 10^{22}$ Am^2, within a total duration of about $20-30$ years (see Fig. 7 in Shaar et al. (2011)). More recently, Shaar et al. (2016) synthesised new data from pottery excavated in Tel Megiddo and Tel Hazor in Israel, with a revision of the archeointensity data previously acquired by Ben-Yosef et al. (2008, 2009) and Shaar et al. (2011) using more stringent data acceptance criteria than they hitherto had adopted. Their study still proposed the spike at the very beginning of the 1st millennium BCE (~980 BCE) while suggesting a new geomagnetic spike during the 8th century BCE (see also Ben-Yosef et al. (2017)). However the spike dated at ~890 BCE was no longer supported by their revised data. Shaar et al. (2016) (see also Shaar et al. (2017)) now characterise the intensity variation as follows: a strong positive magnetic anomaly, which they call the Levantine Iron Age Anomaly (LIAA) with VADM values of the order of 14×10^{22} Am^2 (almost twice the current value), existed in the Near East for about 350 years, between the 11th and 8th centuries BCE. This anomaly was punctuated by two short-lived spikes, both mentioned above, with values reaching $16-18.5 \times 10^{22}$ Am^2, implying rapid rates of change and large deviations in the direction of the dipole field. These rapid changes stand in contrast to reconstructions given by global field models, which tend to be much smoother compared to those of regional curves due to the regularisation required (e.g. Genevey et al., 2016; Hervé et al., 2019; Campuzano et al., 2019; Osete et al., 2020; Licht et al., 2013; Nilsson et al., 2014; Constable et al., 2016). One recent model is of particular interest, that of SHAWQ-Iron Age (Osete et al., 2020), which adopts enhanced weighting of data judged to be of higher quality (see also Campuzano et al. (2019)), and shows increased temporal variations in the spikes era compared with other global models. Notably this model shows in the Levant two intensity peaks at ~970 BCE and ~750 BCE, almost coeval with the two spikes discussed above, but it nevertheless exhibits rates of change not exceeding 0.15 μT/yr and does not reproduce the extreme spike-like behaviour.

Rapid variations in the surface magnetic field, ultimately an expression of the field created deep within Earth's core, are challenging to explain from a geodynamo perspective. At the core surface, there are two dynamical processes that can individually or mutually cause changes in the observed magnetic field: advection by the flow of conducting fluid or diffusion (e.g. Finlay et al., 2010). From basic considerations of the associated timescales, the more rapid of these two is likely advection, which led Livermore et al. (2014) to consider purely (non-diffusive) advection: they placed an upper bound on the rate of change of ~0.6 μT/yr assuming a stable layer at the top of the core in which the flow is horizontal, or 1.2 μT/yr for an arbitrary flow structure. Another possibility of the dynamics is flux expulsion, describing radial advection of deep-rooted magnetic field to the surface followed by its diffusion. However, Troyano et al. (2020) showed that such a mechanism was unlikely to produce extreme intensity changes but instead is more consistent with evolution over multi-centennial timescales. Although diffusion by itself can explain short sections of arbitrary geomagnetic field change (Metman et al., 2019), this assumes an unphysical interior field structure. The notion that advection is dominant when explaining rapid changes is supported also by Davies and Constable (2018), who showed the importance of the role of the flow in intensifying flux patches in the search for geomagnetic spikes within numerical geodynamo models. Korte and Constable (2018) further asserted that the Near-Eastern spikes may have been caused by intense flux patches that grew and decayed in situ during a period of high and rapidly varying dipole field moments, a mechanism also suggested by Osete et al. (2020) together with possible westward expansion. In terms of geomagnetic spikes, based on the archeointensity data obtained by Ben-Yosef et al. (2009) and Shaar et al. (2011), Livermore et al. (2014) estimated that the rates of intensity variations involved could be as high as $4-5$ μT/year, although subsequent studies indicated a somewhat lower rate of change from ~0.75 μT/yr to ~2.5 μT/yr for the geomagnetic spike dated to ~980 BCE (Shaar et al., 2016), and the spike identified during the 8th century BCE characterised by rates of change between 0.75 and 1.5 μT/yr (Shaar et al., 2016; Ben-Yosef et al., 2017). All these estimates remain above the upper bounds of Livermore et al. (2014) which quantifies the limit on what can be explained from a geodynamo perspective, yet because none of the estimates has any quantification of its associated uncertainty it is not possible to rule out the possibility that the data are in fact consistent with lower rates of change. The Bayesian method developed by Livermore et al. (2018) that we adopt in this study should therefore allow us to study in more detail the complexity of the intensity variations associated with the LIAA and provide more insight into the rates of change implied by the archeomagnetic data reported so far in the Levant.

2. The Near East archeointensity dataset

2.1. Available data: provenance and definition

In recent years a significant number of archeological artefacts, obtained in the Near East, have been analyzed for their thermal remanent magnetisation. All the resulting data meet modern quality criteria, which include testing for the stability of the magnetic mineralogy of the samples during the experimental (mainly thermal) treatment and, in

most cases, taking into account the anisotropy and cooling rate effects on thermoremanent magnetization acquisition; for a discussion of these effects, see Brown et al. (2021); Genevey and Gallet (2002); Herve et al. (2019); Genevey et al. (2008). In this study focusing on the period of geomagnetic spikes, which we take to span 1200 BCE to 500 BCE, the dataset has been restricted to three geographical areas: the Levant (Israel and Jordan), Syria and Turkey. In the Levant, the data come from the archeological sites of Tel Hazor and Tel Megiddo (Israel) (Shaar et al., 2016), Timna-30 (Israel; data from Shaar et al. (2011) revised by Shaar et al. (2016)) and Khirbat en-Nahas (Jordan; data from Ben-Yosef et al. (2009) revised by Shaar et al. (2016)), as well as from a study of Judean stamped jar handles (Ben-Yosef et al., 2017). For Syria, the data come from the sites of Qatna (Tell Mishrifeh, western facade of modern Syria), Tell Masaikh, Tell Mashtale and Sheikh Hamad (Genevey et al., 2003; Gallet et al., 2006; Gallet and Al-Maqdissi, 2010, and this study), the latter sites being located further east in the country. The Turkish data are from the archeological sites of Arslantepe (Ertepinar et al., 2012), Kilise Tepe and Tell Tayinat (Ertepinar et al., 2020). All these data are distributed in a circular region with a radius of ~600 km centred on Palmyra (Syria). It should be further noted that a series of data from Cyprus dating from the middle of the 1st millennium BCE has not been included because of their large age uncertainties >150 years, which provide poor constraints on possible rapid variations in regional geomagnetic intensities (Shaar et al., 2015). Details of the experimental procedures and available archeological constraints can be found in the publications referenced above. It is beyond the scope of our paper to discuss here the selection criteria that were used in these studies, as well as the notion of reliable data or "high-quality" data (for a general discussion, see for instance Genevey et al. (2008); Paterson et al. (2014); Hervé et al. (2019); Campuzano et al. (2019), see also Brown et al. (2021)). We simply note that the data considered in the present study would likely be retained in most, if not all datasets assembled using a selection scheme. Furthermore, it should be noted that our approach essentially consists of considering the data as published by the authors (i.e. the same data that led to the observation of spikes), avoiding, for example, the implementation of any form of weighting of the data either at the fragment level or at the level of group of fragments (see below).

In addition to the above published data, we also consider new archeointensity results obtained in Qatna by analysing fragments of large jars found crushed to the ground in a palatial storage complex likely destroyed in ~720 BCE when the Assyrian troops of Sargon II devastated and conquered the Hamat Kingdom (Northwestern Syria; e.g. Al-Maqdissi (2003); Al-Maqdissi and Bonacossi (2005); Baaklini (2019); see below, and Fig. S1 and Table S1 in supplementary material).

The available archeomagnetic data are heterogeneous in the sense that their definition differs from one study to another. For the most part, the data are considered at the fragment level corresponding to a thermal unit (e.g. Levant, Turkey), i.e. a shard, a brick, or a kiln. By extension, a set of mud bricks burnt in a fire also belongs to this category (e.g Ertepinar et al., 2012, 2020); see below and also discussion in Genevey et al. (2008). An alternative way of presenting archeomagnetic data is at the level of groups of several fragments (fragments of baked bricks or potsherds) found in the same archeological context, and therefore assumed to be of the same age (e.g. Genevey et al., 2003; Gallet et al., 2006, 2014, 2020). This approach aims to limit the effects that could, for example, be due to an object whose actual age is outside its presumed age range, to a biased experimental determination, or more generally to take into account the fact that, individually, for other reasons each fragment may not give an accurate record of the geomagnetic field intensity. Averaging at the level of the group of fragments should therefore give greater statistical robustness to the intensity value obtained at the age of interest.

Thus, there are disparities in the way archeointensity data are reported: the unequal statistical equivalence of fragment and group level data will pose a problem when estimating a geomagnetic intensity variation curve. There are two approaches we might consider to mitigate the heterogeneity. Within the set of all published data, we could weight the subset of group-averaged data more heavily than the subset of fragment level data, but the weighting would be subjective. A simpler alternative that we follow here is that of homogenization, by gathering the data into sets of only a single type: single thermal units (a shard, a brick, a furnace, a fire) or an archeological ensemble (or context) comprising several fragments (thermal units) of the same age as defined by archeologists. In our study, we therefore distinguish several levels of data as described below and as illustrated in Fig. 1.

(i). Heterogeneous mixed level, which includes data obtained at both the fragment and fragment-group levels (Fig. 1a). This is the one most often used to represent the available results (Table S2 in supplementary material) (e.g. Shaar et al., 2016; Ben-Yosef et al., 2017; Ertepinar et al., 2020);

(ii). Homogeneous fragment level, which is equivalent to the thermal unit level (Fig. 1b; Table S3 in supplementary material);

(iii). Homogeneous group of fragments level, which includes N artefacts found in the same archeological ensemble (Fig. 1c,d; Tables S4 and S5 in supplementary material). As explained in section 2.3, we consider two cases, $N \geq 2$ (Fig. 1c) and $N \geq 3$ (Fig. 1d).

For comparison, in Fig. 1a we also show the intensity variation in Palmyra (Syria) according to the model SHAWQ-Iron Age (Osete et al., 2020). Importantly, the data with the highest intensities we consider at the fragment level are outside the 95% threshold of the model (two standard deviations) highlighting their anomalously high intensity. All computations were carried out after transferring the intensity values to the latitude of Palmyra, Syria (geodetic latitude 34.560°N; Tables S2-S5 in supplementary material). The corresponding Virtual Axial Dipole Moments are also reported in these tables.

2.2. Data at the fragment level

Data at the fragment level come from two sources: those studies which specifically reported the fragment data, and those studies that originally displayed the intensity values only at the fragment group level; for the latter case, we have considered the individual data (i.e. per fragment) provided in the related publications (Genevey et al., 2003; Gallet and Le Goff, 2006; Gallet et al., 2006; Gallet and Al-Maqdissi, 2010). When several specimens were analyzed per fragment, a mean intensity value was estimated with an experimental uncertainty that is either the half-range when two specimens were analyzed or the standard deviation when the number of specimens is at least three. In a number of cases, however, only one specimen was analyzed per fragment. We have chosen to arbitrarily assign reasonable experimental uncertainties of 5% of the corresponding intensity values (data marked with * in Table S3 in the supplementary material). Note that this 5% threshold corresponds to one of the selection criteria used in related publications (Genevey et al., 2003; Gallet et al., 2006, 2014; Gallet and Butterlin, 2015).

The ages with their uncertainties have been used as reported in the original studies and are either historically/archeologically dated or dated based on an age model. For the sake of simplicity we consider each age to be uniformly distributed, although the algorithm can take into account a Gaussian (i.e. a normal) distribution. There are three data points that were published with an exact date (i.e. with dating uncertainties of 0 years). These are the data from Khirbat en-Nahas (Ben-Yosef et al., 2009), as revised by Shaar et al. (2016) and from Tell Tayinat (TT1; Ertepinar et al., 2020). Because such a precision is rather anomalous, we have assigned an age uncertainty of ± 10 years to these data. In addition, to be more consistent with the approximate age uncertainties of ~70 years (800–732 BCE) made by Shaar et al. (2016) of the shards found in the destruction layers of Tel Megiddo and Tel Hazor dated 732 BCE, we have extended to ±50 years the age uncertainties of the fragments of jars associated with the destruction of the storage

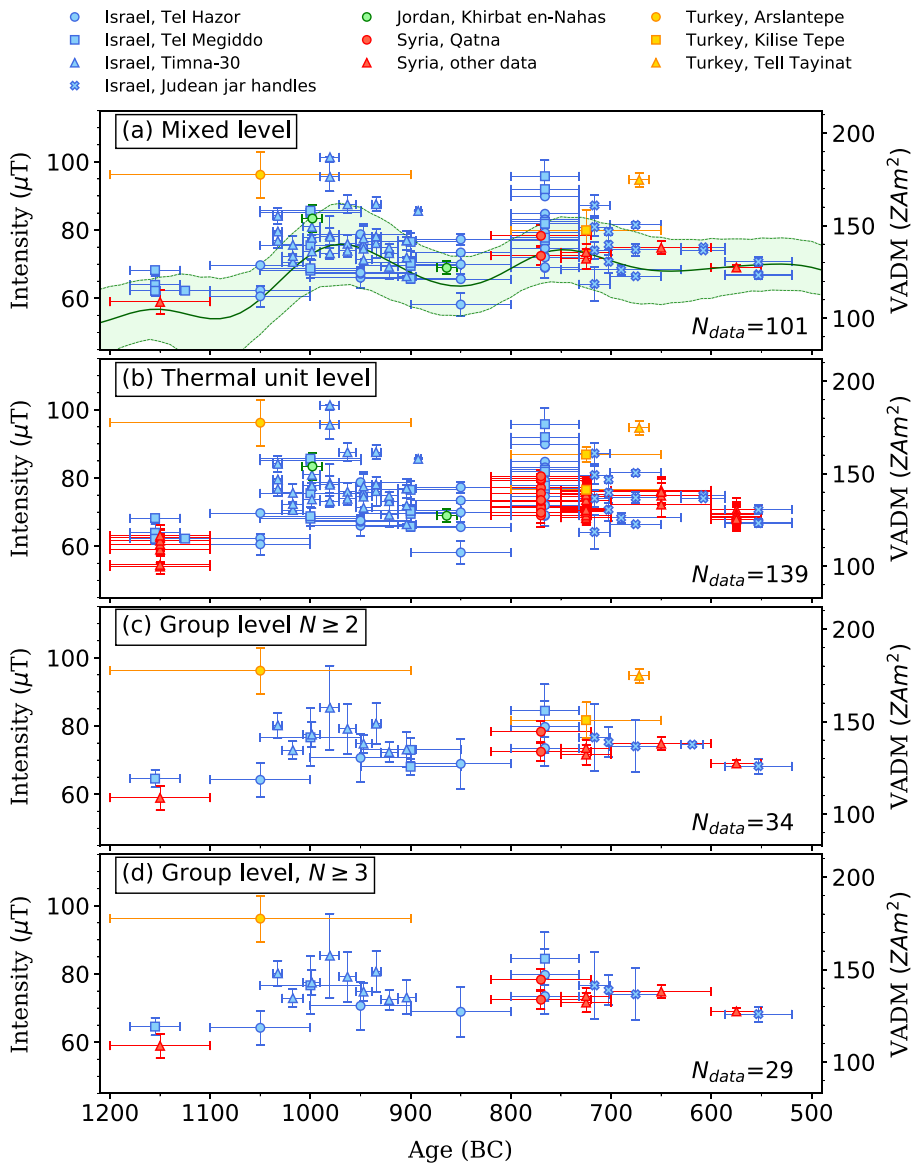

Fig. 1. The Near Eastern archeointensity dataset, grouped into data of the same level: (a) mixed, (b) fragment and (c,d) group-averaged with N individual fragments. All data were transferred to the latitude of Palmyra. The provenance of the data is shown by the legend at the top. Error bars indicate one standard deviation of the assumed normally distributed experimentally-derived intensity, and the total range for the uniformly distributed ages. In (a), we show in green the intensity variation for Palmyra (Syria) with two standard deviations of uncertainty from the SHAWQ-IronAge model (Osete et al., 2020). The size of each data set, N_{data}, is indicated in each part figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

complex at Qatna in ~720 BCE (i.e. to between 820 and 720 BCE). This takes into account the probable life span of these jars before the destruction of the building, which is certainly longer than that of the common pottery studied by Shaar et al. (2016). Note that the age interval considered by Gallet and Al-Maqdissi (2010) for this context was only of 20 years (740–720 BCE).

One dataset of particular importance is that from the Timna-30 archeological site, consisting of an accumulation of metallurgical residues (slags), whose narrow age ranges significantly constrain any model of time-dependent intensity variation. These ages were derived from an age model, itself a combination of three independent sources of information: (i) the stratification across the sequence of slag deposits imposing a time order relationship between ten superimposed layers; (ii) individual layers dated through radiocarbon analysis of short lived material, such as seeds, wood bark and twigs, and (iii) approximations made on the accumulation rates of each slag layer that determine the age of the different layers and how the relative stratigraphic position of individual fragments within each layer relates to their age. Here, motivated by the lack of firm constraints on the accumulation mechanism of each layer, we simplify the ages ascribed to the data by considering a single age range for each layer. Thus all fragments within the same layer have the same age range; time-order constraints only apply between the different layers. Our interpretation of the ages therefore differs (and is a less-constrained version) compared to that of Shaar et al. (2011); the time-ordered Timna-30 dataset nevertheless has a significant impact on intensity evolution reconstruction which is discussed in section 5.2.

In this study, we also include new archeointensity data, hitherto unpublished, obtained from four new fragments of jars (site referred to as SY03) found in the same archeological context as site SY01 (Gallet and Al-Maqdissi, 2010), and analyzed using the experimental procedure developed for the Triaxe magnetometer (Le Goff and Gallet, 2004) with the very same selection criteria as in all our previous studies (Fig. S1, Table S1). The latter archeomagnetic site contained fragments of Aramean-type jars ~1.2-m high without handles, whereas SY03 comprises fragments of jars of the same size but with four handles each. Archeologically, SY01 and SY03 have exactly the same context and are of the same age; however, it seems possible that the presence or absence of handles indicates that these different jars were installed at slightly different times and/or possibly for a different use (but still for storage purposes).

In total, the homogenized fragment dataset has 170 entries, of which 139 data are within the range 1200 BCE to 500 BCE (Fig1b, Table S3 in supplementary material).

2.3. Data at the level of groups of fragments

For the group level, we have averaged the individual (fragment) data available within the same archeological context. These are mainly data obtained in the Levant, and thanks to the information provided in the articles concerned (Shaar et al., 2011, 2016; Ben-Yosef et al., 2017), it was easy to compute these averages and, in some cases, to maintain a chronological order relationship between these data. This last aspect mainly concerns the results obtained from slags in Timna-30 (Shaar et al., 2011), with data revised by Shaar et al. (2016).

The calculations were performed considering groups defined by an average of N individual (fragment) data, where $N \geq 2$ (Fig. 1c; Table S4 in supplementary material) and $N \geq 3$ (Fig. 1d; Table S5 in supplementary material). Note that three is the minimum considered to define an average intensity value in all the studies carried out in Syria. The difference between the resulting datasets is that $N \geq 2$ contains five additional data points (two in Tel Megiddo, one for the stamped Judean handles and two in Turkey). It is worth mentioning the particular case of the intensity value referred to as AT1 obtained in Arslantepe from a set of adobe brick fragments that were burnt during a violent fire (Ertepinar et al., 2012). This datum appears both when the data are considered at the fragment level, as this context corresponds to a single thermal unit, but we also chose to consider it when the data are considered at the group level, as this data point is obtained from an average of nine different fragments collected on the field. The context is similar for the data obtained at Tell Tayinat (TT1; Ertepinar et al., 2020). Being defined by only two fragments (for a total of three specimens) it is included in Table S4 but not in Table S5. In total, Table S4 contains 42 entries (Fig. 1c) and Table S5 contains 37 entries (Fig. 1d), of which respectively 34 and 29 are in the range 1200 to 500 BCE.

3. Method

3.1. Finding the posterior intensity variation curve

Fitting a time-dependent intensity variation to the datasets and quantifying its uncertainty is not straightforward due to the scatter of the data, the fact that some data obey strict time-ordering constraints, and the fact that all data have uncertainty in both intensity and age. Here, we use an algorithm based on the transdimensional Bayesian method of Livermore et al. (2018) but with added functionality, designed to determine the posterior probability distribution of variation curves from datasets with precisely these challenges. Using a Monte-Carlo Markov Chain algorithm the method produces a large ensemble of models whose statistics converge to those of the joint posterior distribution of the unknown model vector $\mathbf{m} = (\mathbf{f}, k, \mathbf{a})$; where $\mathbf{f}$ is a discretisation of the intensity variation, k is its complexity, and $\mathbf{a}$ are the sample ages. By marginalisation it is straightforward to produce the posterior probabilities of any of these quantities individually: in this study we focus on the intensity variation, but posterior distributions of sample ages are a useful product for archeomagnetic dating (Livermore et al., 2018; Gallet et al., 2020; Genevey et al., 2021; Shaar et al., 2020).

The intensity variation associated with any individual ensemble member is modelled as piecewise linear between a series of interior k knot points; however diagnostics of the ensemble such as the mean or median vary smoothly in time. The number k of the knot points and their associated ages are self-selected by the data and are not fixed: they are in fact an output of the model. Of additional note is that the method does not have any subjective smoothing parameter but instead relies on the innate parsimony of Bayesian methods to show preference for simple models. Thus the method will produce rapid changes (described by more complex models) only when required by the data. Any rapid changes localised in time will not affect the preference for simple time-dependence elsewhere (to be contrasted with the use of global regularisation where smoothing due to rapid variation at one time will impact the entire time series). This is very important because of the likely range of timescales that may describe the intensity evolution, ranging from the proposed decadal duration of a spike, to the typical centennial timescale of secular variation.

Including the sample ages into the model vector that we seek allows us to take account of not only the uncertainty but any constraints on the temporal order that may exist between fragments or groups of fragments of the same archeological sequence. In our datasets, this mainly concerns the slag data from the Timna-30 site (Shaar et al., 2011), revised by Shaar et al. (2016), where the fragments were obtained from ten different superimposed layers whose formation ages are uncertain. Between any two consecutive layers, fragments within the lower layer must all have an age strictly greater than any of the fragments in the upper layer. Within each of the layers there is often no order relationship between the fragments due to mixing. It is also important to specify that for fragments collected within the same archeological ensemble, the algorithm allows for the distinction between independent fragments that can be of any age within the age interval of the archeological ensemble, as in the case of the slag data mentioned above, and fragments whose age is identical within the age interval of the ensemble. The latter case relates to bricks of the same pavement sampled at Tell Masaikh in Syria (TM01; Genevey et al., 2003; Gallet and Le Goff, 2006). Here, it is reasonable to assume that they were fired at the same time and most probably in the same kiln, and therefore the age of the different fragments analyzed from these different pavement bricks cannot be dissociated and are taken to have the same age.

The posterior distribution relies on choices for the likelihood and prior distributions. The likelihood is based on an assumed normally distributed intensity with a mean and standard deviation derived by laboratory analysis. We adopt uniform priors in the range [40,120] μT for the intensity of the internal vertices that define the piecewise linear dependence, a uniform prior in the range [0,150] for the number of internal vertices, and uniform prior distributions in data age. Each ensemble member is defined over the time period 1250 to 450 BCE, marginally extending the data interval of 1200 to 500 BCE by 50 years at each end. To characterise the intensity variation, we report the median of the marginalised intensity distribution discretised over the 801 yearly intervals from 1250 to 450 BCE. The distribution mean has a similar time dependence to the median, but we report only the median as it is less sensitive to extreme values of the distribution. We use 50,000 iterations as burn-in with a down-sampled total chain length of 200,000,000 which ensures converged sampling of each parameter within the model vector. The setup of the inversion remains the same for all cases, although some of the sampling parameters are dataset-dependent in order to speed up convergence. The code, data and input files required by the software are freely available.

3.2. Defining and detecting geomagnetic spikes

Despite multiple studies describing the possibility of geomagnetic spikes in the Near East and elsewhere, there is no currently accepted definition of the defining characteristics of a spike. Ben-Yosef et al. (2009) and Shaar et al. (2011) defined spikes to be "short episodes of exceptionally high field intensity in excess of 200 ZA m^2", "accompanied with high field fluctuation". To broaden the concept of geomagnetic spikes, Cai et al. (2014) proposed the following definition: "a large increase in intensity beyond twice the present value (equivalent to $\sim 16 \times 10^{22}$ Am^2) in less than 500 years". Yet this terminology does not capture the extreme intensity variation rates required by the two geomagnetic spikes proposed by Shaar et al. (2011) which are an order of magnitude greater than the current maximum rates of intensity change and which go well beyond the rates currently explicable by known dynamo processes (Livermore et al., 2014; Troyano et al., 2020).

Geomagnetic spikes are undoubtedly significant local maxima in the intensity variation, but from a geodynamo perspective it is the rate of change which makes the time-dependence most anomalous. Here we propose a definition that incorporates both the magnitude and rate of

change, that is sufficiently generic to detect spikes over periods other than the 1st millennium BCE. Given a time series of intensity variation, we consider each local maximum of intensity, at age t_{max}, in turn. In order to judge whether it is a spike, we first need to define start (t^-) and end (t^+) points for the period of anomalous behaviour. We do this by searching both backwards and forwards in age from t_{max}, seeking the first occurrence of an age that marks the transition between typical and anomalous behaviour both in the rate of change, and the intensity difference from that at the peak. The two end points must lie within the range bounded by the two neighbouring local minima. A local maximum of intensity at t_{max} defines an anomalous peak (which may, or may not, be a spike) over the age range t^- to t^+ if:

(a). the intensity at t_{max} exceeds by 5 μT the intensity at t^- and t^+
(b). at t^- (t^+) the rate of change marks a transition from (to) typical values $|dF/dt| \leq 0.12\mu$T/yr to (from) anomalous values of $|dF/dt| > 0.12\mu$T/yr and t^- and t^+ are the closest such values to t_{max}.

We introduce a further criterion that discriminates between an anomalous peak and a spike:

(c). the anomalous peak is a spike over $[t^-, t^+]$ if at some time within its age range $|dF/dt| > 0.6\mu$T/yr.

The threshold in (a) of 5 μT is a conservative value, allowing the definition to be useful elsewhere in the world at any latitude and especially at different periods (in particular when the background intensity level is much lower). This is the minimum jump we can reasonably expect to detect using archeointensity data because of their typical uncertainties (e.g. Korte et al., 2005; Licht et al., 2013). For reference, the reported Near-Eastern spikes have intensity jumps of more than 10 μT which fall well inside our definition. The criterion (b) defines the transition from typical to anomalous behaviour, which we take to be the approximate current maximum value of 0.12 μT/yr derived from the IGRF-13 model (Alken et al., 2020). The final criterion that selects only those peaks with extreme rate of change has a threshold of 0.6 μT/yr, five times the approximate current maximum value; it is

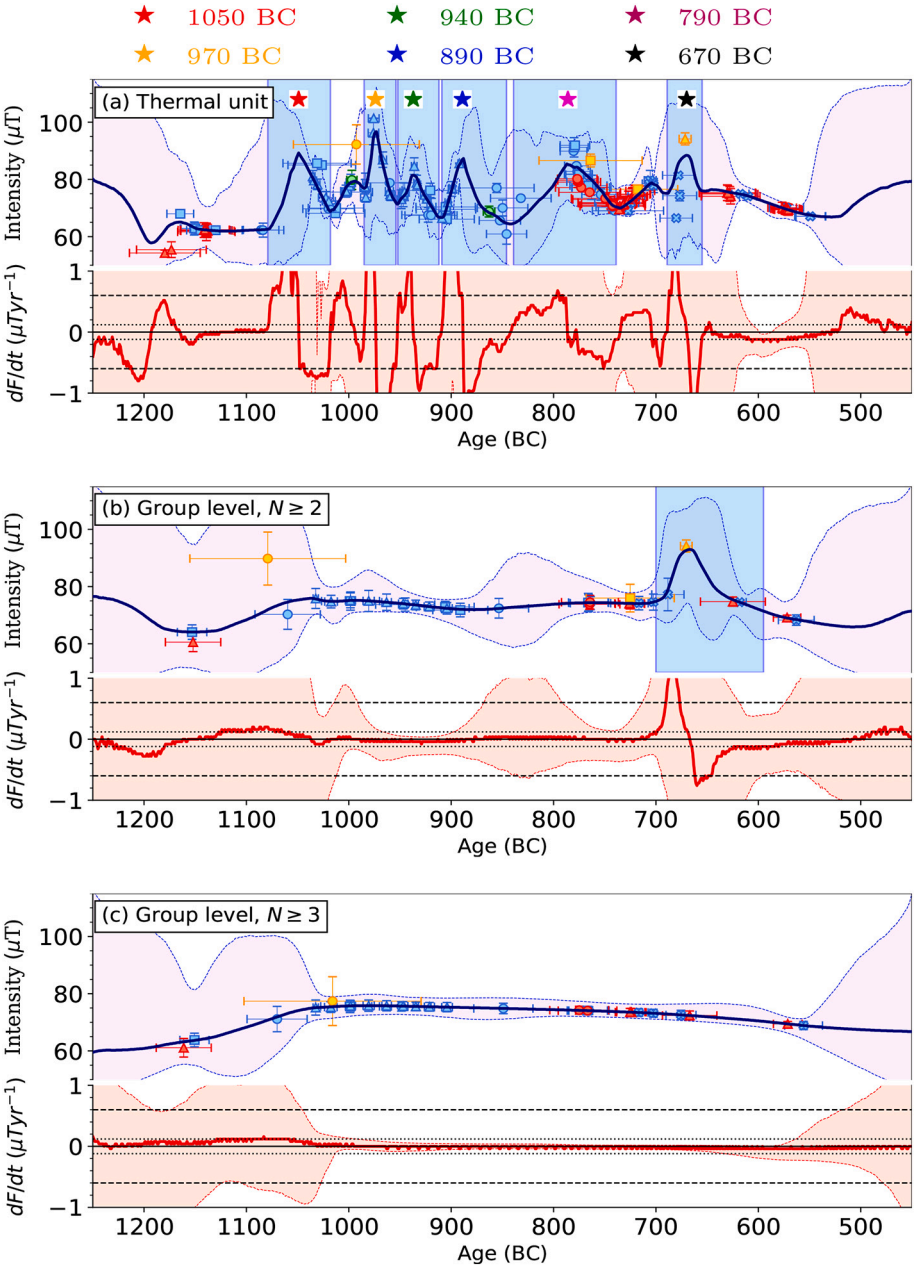

Fig. 2. The posterior intensity variation during the era of geomagnetic spikes associated with different datasets a-c: thermal unit (i.e. fragment level) data, group level $N \geq 2$ and $N \geq 3$. Each upper panel shows the median (dark blue) and 95% credible intervals (light purple), and each lower panel shows the rate of change of intensity characterised by its median (red) and 95% credible intervals (orange) with threshold levels of $\pm 0.12\mu$T/yr and $\pm 0.6\mu$T/yr depicted respectively as dotted and dashed lines. Geomagnetic spikes are shown by the light blue boxes, and in (a) are labelled by coloured stars with their approximate ages at the top of the figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

also the same as the upper bound of Livermore et al. (2014) that assumes a horizontal core-flow, and is also consistent with the more general upper bound of 1.2 μT/yr but when assuming that only 50% of the energy of the core is available to drive the spike. Because this value is not well constrained, in the next section we report the sensitivity of our results to this threshold, by testing the effect of taking the higher value of 1.2 μT/yr. Before applying these definitions, we first smooth the intensity variations by a running five-year average whose window length does not influence the anticipated decadal duration of the geomagnetic spikes.

4. Geomagnetic intensity variations: a multitude of spikes

We now apply our methodology to find the intensity variation curve along with our spike detection algorithm to the various datasets described. The results are summarised in Fig. 2, which shows the evolution of both intensity and its rate of change for the thermal unit (i.e. fragment level) dataset (top), group averaged datasets with $N \geq 2$, $N \geq 3$ (middle, bottom) by the ensemble median and its uncertainty quantified by 95% credible intervals. We also show each dataset by the individual posterior median intensity and age along with error bounds given by the posterior standard deviation in both intensity and age; this is distinct from Fig. 1 that shows the data by their a priori values. We note that the posterior and a priori ages can differ by up to 60 years, depending on the error budget for the data. We omit showing the results from the mixed dataset as they are almost identical to those from the thermal unit dataset because the number of group level data at the time of spikes is relatively small (in fact this involves only three data around the spike of the 8th century BCE; see below). We therefore do not consider further the mixed level dataset and we only focus on the homogenized datasets.

Fig. 2a shows that despite the scattered nature of the thermal unit dataset (Fig. 1) there is a smooth intensity evolution consistent with the data and their uncertainties, characterised by eight local maxima in intensity between ~1050 BCE and ~650 BCE of which six are prominent at ~1050 BCE, ~970 BCE, ~940 BCE, ~890 BCE, ~790 BCE and ~670 BCE and fulfill our definition of geomagnetic spikes. These spikes along with their duration are highlighted in light blue boxes, and labelled with coloured stars. We will discuss the duration of the spikes below, but for now we simply note that the longest spike, at ~790 BCE, has a duration of about a century (including both rise and fall of intensity). A particularly important result is thus that the available data require a much more complex evolution of geomagnetic intensity than the quasi-steady behaviour punctuated with only two spikes that has been proposed so far. Indeed, of the six spikes that we find, two of them (ca. 970 and 790 BCE) have a similar age to those proposed at 980 BCE (Shaar et al., 2011) and during the 8th century BCE (Shaar et al., 2016).

We note that the identification of more than two spikes consistent with the thermal unit data does not depend on the threshold criterion of our spikes definition: even by doubling the rate of change threshold to 1.2 μT/yr, significantly narrowing the criteria that the spikes must satisfy, four spikes (at around 1050 BCE, 970 BCE, 890 BCE and 670 BCE) are identified, leaving this key conclusion unaltered.

When the raw archeointensity data are averaged according to their archeological context (or layer), the situation changes radically. For the group-level data defined with $N \geq 2$, only one spike remains during the 7th century BCE, lasting slightly less than a century (Fig. 2b), clearly generated by a single data point obtained in Turkey (TT1). On the other hand, the high intensity value obtained at Arslantepe (Turkey; Ertepinar et al., 2012) does not require a significant fluctuation, probably due to the relatively large age and intensity uncertainties associated with this data point, in combination with the other archeointensity values available over the time interval concerned. When the groups are defined with $N \geq 3$, removing the TT1 data point, the evolution in intensity shows no spikes but instead a rather simple multi-centennial slow increase then decrease (Fig. 2c). Overall, depending on the choice of the dataset (fragments or groups of fragments), the intensity evolution alters markedly from a highly oscillatory curve with six spikes (fragment level) to one with no spike at all (group-level $N \geq 3$).

The intensity variation required by the fragment level data depends significantly on their associated error budgets: small uncertainties of some data trigger a complex time-dependence. We briefly illustrate the impact on our results of what Shaar et al. (2016) term as "extended errors", which are based on the extremes of the possible intensity values for all the specimens of the same fragment, also including the standard deviation error. These extended errors apply only to the datasets provided or revised in Shaar et al. (2016), for which the authors consider that they give a better expression of the true experimental uncertainty of the intensity obtained at the fragment level. Although these values characterise the breadth of the error distribution, its shape is not known. We interpret their results in terms of a uniform distribution (Fig. S2a; Table S6) and a Gaussian distribution (Fig. S2b; Table S7) with, in the latter case, a standard deviation being given by 1/4 of the extended error (i.e. assuming that 95% of the possible values are within the extended error interval). For the uniform distribution we recover 6 spikes as in Fig. 2a, whereas for the Gaussian distribution only 5 spikes are detected although the spike in the 8th century is only very slightly below the threshold limit of 0.6 μT/yr (Fig. S2b). It therefore appears that adopting either of our two interpretations of the errors proposed by Shaar et al. (2016) does not modify the observations previously made.

5. Sensitivity of the spikes

5.1. The effect of an increased error budget

Although the extended errors of Shaar et al. (2016) did not affect the results in any significant way, we now explore the impact of further increasing the error budgets both for intensity and age, with the general idea that the accuracy of the intensity values, as well as the date of the artefacts or contexts studied, could be overestimated (e.g. Korte et al., 2005; Suttie et al., 2011; Davies and Constable, 2017). There are many factors that could lead to under-estimation of errors; for instance, a fragment may move in a stratigraphic sequence (see for instance discussion in Gallet et al., 2020; Gimatzidis and Weninger, 2020) or the experimental approach used may not be adequate enough to detect (and correct) slightly biased intensity values (which, we note, does not mean that the data are any less reliable according to the selection criteria used). It should also be noted that an intensity determination generally does not take into account the diversity of all interpretations that are equally possible or acceptable (although, in this respect, the approach used by Shaar et al. (2016) is a step towards this). We arbitrarily impose a sequence of minimum values on the intensity error: 0 (i.e. no change), 3, 4, 5, 6 μT, (with laboratory-determined values greater than these minima being unaffected), anticipating that larger uncertainties will require fewer fluctuations in intensity. We adopt an absolute rather than a relative minimum error in order not to bias the analysis towards data with lower intensity (see discussion in Suttie et al., 2011). Similarly, we successively place a minimum accuracy on the dating of artefacts or archeological contexts of 0 yr (i.e. no change), $\pm$25 and $\pm$50 years. This latter modification mainly affects the Timna-30 data, which were published with a high precision of less than $\pm$10 years and for which we keep the time-order relationship (Shaar et al., 2011, 2016). We recognize that this is a somewhat arbitrary and blind modification, but it is intended to assess the extent to which the detection of geomagnetic peaks depends on the accuracy of the available archeomagnetic data. We focused these calculations on the data at the thermal unit level showing the greatest number of spikes (Fig. 2a). Fig. 3 shows a subset of the 15 cases (see Fig. S3 for all cases), chosen to highlight the important effects, which incorporates minimum intensity errors of 3–6 μT and minimum age uncertainties of 0 (left column) and 25 years (right column).

When we impose minimum errors of 3 μT on the intensity, affecting 93 of the 139 data, we find that the resulting intensity variation has only five of the six spikes of Fig. 2a, the missing spike being that of the 8th

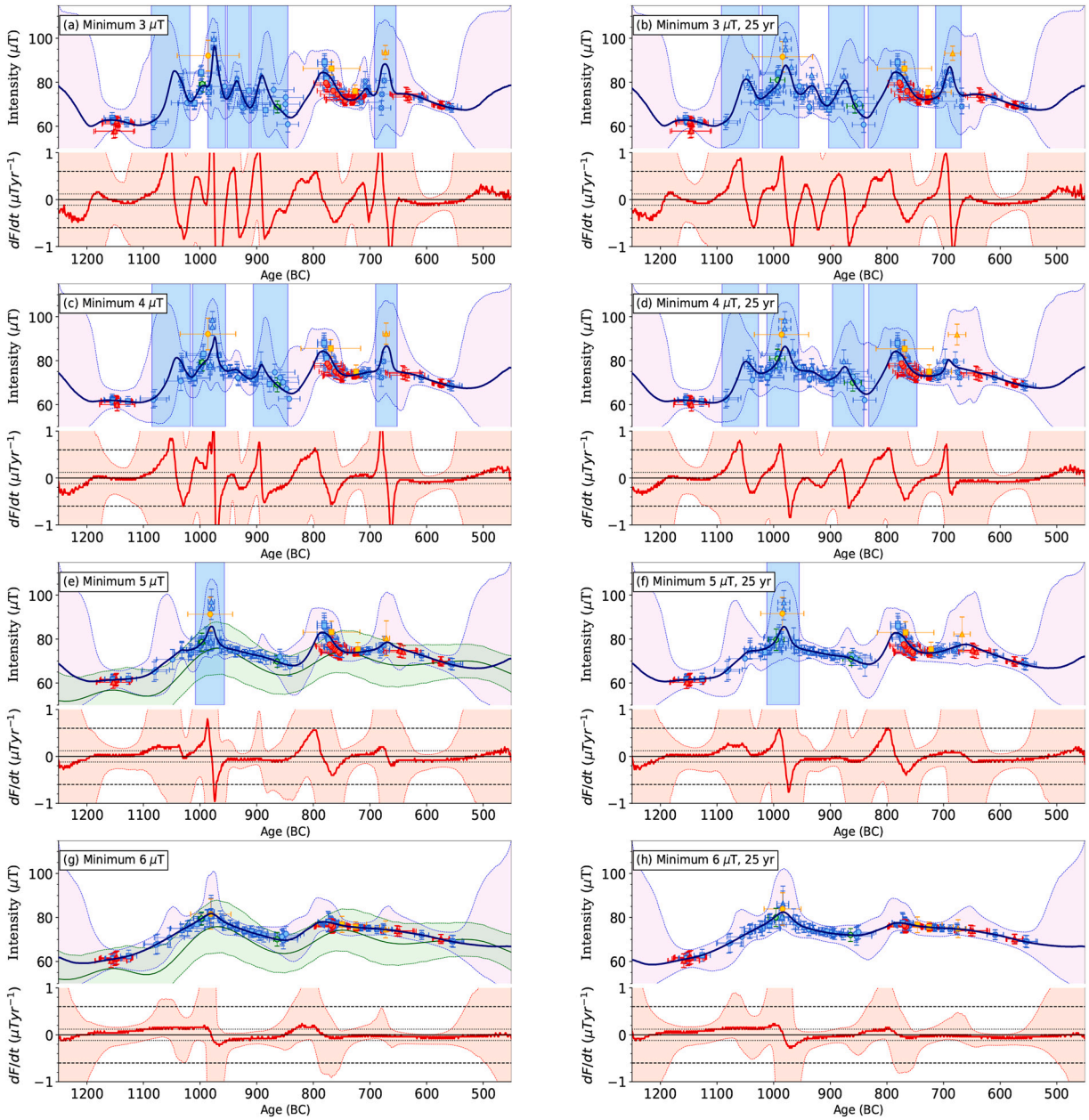

Fig. 3. The effect of a series of minimum error budgets in both intensity and age on the posterior intensity variation based on the thermal unit dataset. For each sample in the dataset, the posterior median intensity and age is plotted with the error bars showing the posterior standard deviation. Left column shows the intensity variation using the a priori ages but with a minimum intensity error of 3–6 μT. Right column shows the variation additionally assuming a minimum age error of 25 years. The green region in (e) and (g) shows the SHAWQ-Iron Age model with 2 standard deviations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

century BCE when the maximum dF/dt fell just below the threshold at 0.6 μT/yr (Fig. 3a). Compared to Fig. 3a, increasing the minimum intensity errors to 4 μT results in the elimination of the spike around 940 BCE (thus leaving four spikes; Fig. 3c). Alteration of the minimum intensity error to 5 μT is an important case since this threshold is also considered by other authors (Korte and Constable, 2011; Davies and Constable, 2017); it affects 135 of the 139 data and significantly only one spike remains, that defined by the slag data around 970 BCE (Fig. 3e) (Shaar et al., 2011, 2016). A very strong fluctuation is also observed during the 8th century BCE but the associated *dF/dt* curve falls marginally short of the threshold of 0.6 μT/yr. It appears that this case leads to a situation closest to that suggested by Shaar et al. (2016). In all cases, placing a minimum age uncertainty of 25 years (Fig. 3b,d,e,h) has only a minor effect by reducing slightly the amplitude of both the intensity and dF/dt. In each case, the number of spikes remains the same although there is some alteration of which peaks are classified as spikes among those peaks very close to the thresholds we define.

When the minimum intensity error is slightly increased to 6 μT, this threshold representing ~7.5% of the average intensity values, no more spikes remain; of particular note is that the spike at 970 BCE, so far the most robust to increases in intensity error budget, also vanishes. Only two moderate (albeit strong compared to the current field behaviour) fluctuations are present around 1000 BCE and during the 8th century BCE, marked by a maximum of dF/dt of about 0.2 μT/yr.

The minimum 6 μT case shows a striking similarity to the reconstruction from the SHAWQ-Iron Age model (Fig. 3g) in terms of its general structure of a dual broad peak with fairly similar rates of change. Note, however, that the peaks defined by the Bayesian median curve are

not only slightly offset by a few decades earlier, but have higher values, than those defined by the SHAWQ-Iron Age model.

Thus as the error budget is increased from the a priori values, the strongly varying intensity becomes rather flat and devoid of remarkable structure, with rates of variations close to those known for the present-day field.

5.2. The effect of the Timna-30 age model

Very accurate intensity and/or dating results have a strong influence on the calculation of intensity variation models. We now highlight the importance of the dataset from the Timna-30 archeological site whose very precise dating places severe temporal constraints on any intensity evolution curve. Indeed, Fig. 2 shows that four of the six spikes inferred from the fragment level dataset occurred during the period 1109–836 BCE of the Timna-30 data.

There are two aspects to the temporal constraints of the Timna-30 data: (i) the narrow age-range proposed for each layer and (ii) time-order constraints between layers. Following the philosophy of the previous section, we now investigate whether a softening of these stringent constraints has an effect on the inferred posterior intensity evolution. The stratification at Timna-30 was established with great care in Shaar et al. (2011) (see their Fig. 3), and challenging this stratification does not seem to be a reasonable option. In contrast, the assumptions that lie behind the proposed age range of each layer are probably more questionable, in part because they rely on poorly constrained layer accumulation mechanisms. We therefore consider the following option: the stratification between layers is preserved but the authors' narrow age ranges for each layer are replaced by a single Gaussian distribution, whose 2σ limits are taken to be the range 1109–836 BCE that spans the whole dataset (Shaar et al., 2011). In the context of the method, the prior distribution of each of the Timna-30 data then are defined by the same Gaussian distribution, and the stratification is taken into account by the posterior distribution. In particular, the marginal posterior ages necessarily obey any time-ordering constraints.

Fig. 4 shows four calculations based on the redefined prior age distributions for the Timna-30 data, the remainder of the thermal-unit dataset remaining unchanged. Fig. 4(a) shows that the modified dataset itself is consistent with three spikes, only one of which (~ 970 BCE) is prior to 850 BCE. Note that the double-peak at ~1070 BCE is not a spike as it does not meet the first criterion of section 3.2. Thus, for the period prior to 850 BCE, our broadening of the prior ages of the Timna-30 data reduces four spikes to one. The spikes ca. 790 BCE and 670 BCE are similar to those of Fig. 2. On allowing for a greater error budget for the intensity errors (Fig 4b,d) to either 3 μT or using the extended (uniformly distributed) errors, the spike at ~970 BCE disappears. Relaxing the intensity error budget still further to 5 μT results in an intensity evolution with only modest maxima around 970 BCE, 790 BCE and 670 BCE which are not sufficiently extreme to be classified as spikes, and indeed which shows an evolution consistent with that of SHAWQ-Iron Age (Osete et al., 2020).

In the previous section we noted that the spike at 970 BCE was the most resilient, when increasing the intensity errors. We see now that this spike is heavily dependent on the published stringent age ranges for the Timna-30 dataset: a relaxation of the age ranges results in the disappearance of this spike.

6. Discussion

6.1. Characterisation of the geomagnetic spikes

The observations made so far are now complemented by a more quantitative characterisation of the Near-Eastern geomagnetic spikes (of Fig. 2) derived from the thermal unit dataset.

Fig. 5a shows a comparison of the shape and duration of the spikes of Fig. 2a. The total duration of the geomagnetic spikes ranges from ~30 years for the two spikes dated around 970 BCE and ~670 BCE (orange and black curves, respectively) to ~100 years for the spike of the 8th century BCE (purple curve). Of particular note is that the two spikes (respectively orange and purple) that were proposed at ~980 BCE and during the 8th century BCE (Shaar et al., 2016) have durations ~30 years and ~100 years, i.e. at the two extremes. The duration between a spike's ascending and descending branches can be significantly different, with clearly shorter ascending branches for the spikes dating from ~970 BCE (orange curve) and ~890 BCE (blue curve). The spike around ~670 BCE (black curve) is particularly symmetric with respect to the maximum intensity. The ~30-yr duration for the spike at ~970 BCE is consistent with the timescale originally proposed by Shaar et al.

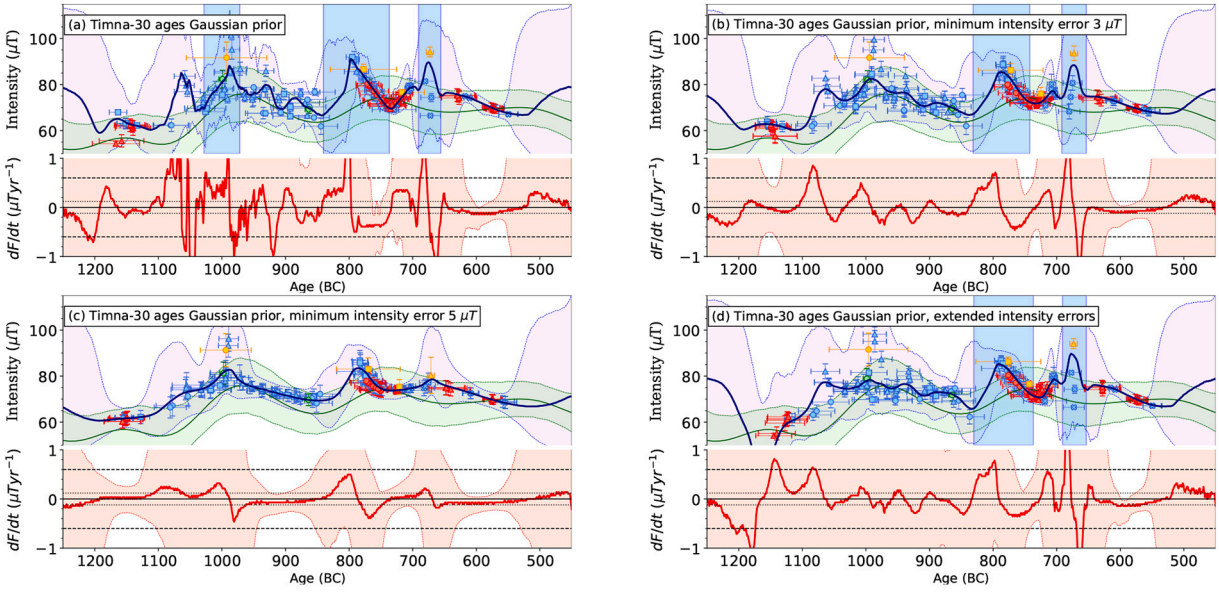

Fig. 4. Panel (a) shows the evolution of the posterior distribution based on a modified fragment level dataset in which the age model for the Timna-30 data is replaced with a Gaussian prior distribution with the 2σ limits of 1109 and 836 BCE. Panels (b-d) show respectively similar calculations when assuming minimum intensity errors for all data of 3 μT and 5 μT, and the extended error case assuming a uniform distribution. The individual data are shown by their posterior median age and intensity with the error bars indicating the posterior standard deviation.

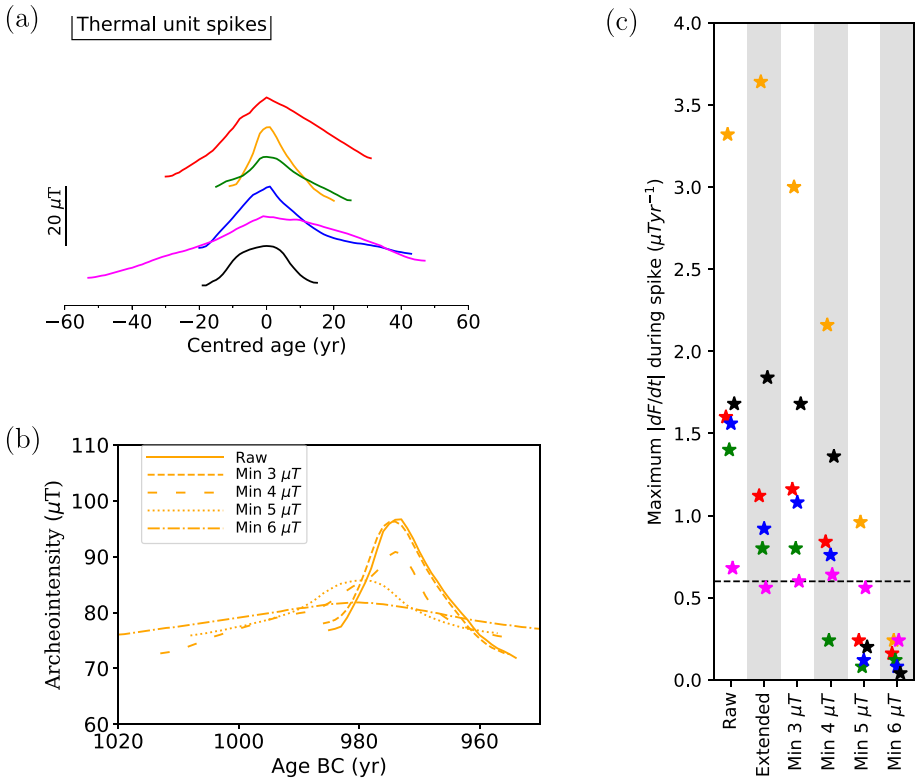

Fig. 5. Characterisation of the geomagnetic spikes of the thermal unit dataset with colours given by the stars in Fig. 2a. (a) The spikes are shifted in age and intensity to have a common centred age and intensity maxima offset by 10 μT; the spike duration is determined by our criteria. (b) The geomagnetic spike of ca. 970 BCE as a function of minimum intensity error imposed; when imposing a minimum error of 6 μT the spike disappears and here we show the intensity variation over the entire window. (c) The maximum absolute rate of change of intensity at the times of the spikes based on the raw dataset and their dependence on the minimum intensity error imposed; the dashed line marks the threshold of 0.6 μT/yr used to define a spike.

(2011).

Fig. 5b illustrates how the shape of the spike dated ~970 BCE changes with minimum intensity error, showing how the intensity curve is flattened by an increasingly generous error budget. For a minimum error of 3 μT the duration of this spike remains ~30 years. For higher minimum errors of 4 and 5 μT the duration increases to ~55–60 years and finally the spike disappears at 6 μT. On the basis of the available data, therefore, it appears that this spike is intrinsically shorter than the spike of the 8th century BCE. Interestingly, the peak of the spike increases slightly in age (by about 10 years) between the minimum error cases of 3 μT and 5 μT.

Finally, Fig. 5c shows in a quantitative way how the extreme nature of geomagnetic spikes, defined by the maximum *dF/dt* for each spike, evolves as a function of the error budget attributed to the data (same colour code as before). Note that only the intensity errors are modified, and we include also the case of the extended errors (Fig. S2); thus, the dating uncertainties, sometimes very small, remain those assigned by the authors (Fig. 1). All geomagnetic spikes retain their respective exceptional character up to a minimum error of 3 and 4 μT. Above this value (e.g. at 5 μT), the max *dF/dt* is strongly suppressed and only two intensity fluctuations maintain maximum *dF/dt* values higher or very close to the threshold of 0.6 μT/yr. With a minimum error of 6 μT, all maximum *dF/dt* values are well below this threshold (maximum ~0.25 μT/yr) and the observed events then become compatible with the current understanding of the geodynamo (Livermore et al., 2014). As noted above, they also become similar to the rates of change of other intensity peaks mentioned in the literature for the Near East and Western Europe (Hervé et al., 2017; Yutsis-Akimova et al., 2018; Gallet et al., 2020; Osete et al., 2020). Of the two spikes proposed by Shaar et al. (2016), we note that the spike dated ca. ~970 BCE is the most persistent of all the spikes in terms of an increasing intensity error budget. The 8th century BCE peak is remarkably stable up to experimental errors of 5 μT, even though it is not always classified as a spike due to rate of changes that fluctuate around the threshold. When increasing the error budget to 6 μT, only relatively modest intensity variations remain which we do not classify as spikes.

6.2. Concluding remarks

We have illustrated that the observation and very existence of geomagnetic spikes in the Near East, as based on the set of archeointensity data currently available, is crucially dependent on the definition of the data that are used, as well as the error budget allocated to these data. The analysis presented in this work was made possible by adopting the transdimensional Bayesian method recently developed by Livermore et al. (2018), which is particularly well adapted to the detection of rapid and extreme intensity variations. In the case of considering all data at the fragment level, we showed that the available dataset is compatible with not just two spikes (as proposed by Shaar et al. (2011) and others), but with six spikes, which all meet the same definition criteria (mainly considering a threshold of 0.6 μT/yr for the rate of change associated with each event), and whose duration varies from ~30 years to ~100 years. To argue for a number of spikes less than six would require a demonstration that some data are less reliable than others, or, as we have shown, that either (or both) of the intensity error budget or age constraints associated with the data are softened. However, the fact remains that even if the quality criteria imposed are considered particularly severe compared to those prevailing for much of the archeointensity results obtained worldwide (see Genevey et al. (2008); Brown et al. (2021) and the references associated with the datasets selected in the present study), the data available in the Near East remain fairly scattered which obscures the observation of ultra-fast archeointensity variations.

We also showed that the number of spikes is a function of the error budget allocated to the data when considered at the fragment level. Note that in our numerical experiments (and for the dataset considered), minimum intensity errors seem to have more influence than minimum uncertainties on dating. For the specific choice of minimum intensity errors of 5 μT, which has been considered for global field reconstructions (Korte and Constable, 2011; Davies and Constable, 2017), we find two

(and only two) significant intensity peaks, the first around 970 BCE which we classified as a spike of duration ~55 years, and the second event (which did not meet the criteria for a spike) during the 8th century of longer duration (~1 century), both of which have been widely discussed in the literature (e.g. Shaar et al., 2016). We also showed that the situation prior to 850 BCE radically changes if the narrow age range considered for each slag layer of the Timna-30 data is widened, assuming instead a single age interval for all these data covering the entire sequence of deposits, while preserving the time-order relationship as provided by Shaar et al. (2011). The spike at 970 BCE, seemingly persistent with respect to enhanced intensity errors, appears particularly fragile if the layer-age constraints for the Timna-30 dataset are relaxed. If the minimum intensity error is slightly increased to 6 μT, then although the two broad peaks remain and the curve becomes remarkably consistent with the SHAWQ-Iron Age model Osete et al. (2020), all spikes disappear and the rates of change (with maximum values at or below 0.25 μT/yr) become close to the rates determined for other intensity peaks of different ages in the Near East and Western Europe (Hervé et al., 2017; Yutsis-Akimova et al., 2018; Gallet et al., 2020; Osete et al., 2020).

We briefly comment on the how our choice of method has shaped the study and the conclusions. Underpinning the identification of spikes is the ability to model multiple timescales, of paramount interest here are the interspersed periods of rapid and slow evolution of the intensity (see, for example, Fig. 3). Our approach with minimal prior information and no smoothing is ideally suited to this situation, where the data themselves identify any periods of temporally-localised rapid changes. Other methods (e.g. Lanos, 2004; Hellio et al., 2014; Thébault and Gallet, 2010) which all involve additional constraints may not be able reproduce the behaviour we find, in particular, the existence of six spikes consistent with the thermal unit dataset. However, for the most generous error budget, the resulting smooth evolution characterised by a single timescale is likely reproducible by other methods.

The strong sensitivity of intensity variation inferred from a specific dataset has been widely discussed in the literature (e.g. Genevey et al., 2016; Hervé et al., 2019). Here this is also seen when we compared fragment to group level data. In the literature, the definition of archeointensity data varies from one author to another (Brown et al., 2021; Genevey et al., 2008). This difference is almost philosophical as it depends on the authors' confidence in their own individual results, even beyond the uncertainty estimates experimentally found for the intensity or provided by archeologists for the dating for their data. Geomagnetic spikes appear if the data are considered at the level of the thermal unit, i.e. a pottery fragment, a brick or a kiln. Crucially, each fragment is assumed reliable within its error bars. If the data are defined on the level of a group of fragments of the same age and collected in the same archeological context (with a minimum of three independent fragments per context), in order to reduce, for example, a possible bias linked to a badly dated fragment in relation to the others, then the spikes disappear from this reduced homogenized dataset.

Our principal conclusion is that there is currently no indisputable evidence of geomagnetic spikes which have been proposed in the Near East during the first half of the 1st millennium BCE, whose very existence appears rather fragile. Instead, we cannot exclude the possibility that the Near-Eastern geomagnetic intensity variation of the 1st millennium was characterised by intensity peaks with much more modest rates of change (~0.2–0.3 μT/yr) as shown also by the recent SHAWQ-Iron Age model, possibly higher than the modern field, yet consistent with secular variation as known for other ages and in other regions. Our inability to offer unequivocal evidence of a spike hinges on the fact that the true uncertainties in both the intensity and ages of the data are difficult to estimate reliably, particularly when working at the fragment/thermal unit level. One thing is certain: spikes bring us to the boundaries of accurate experimentally-determined archeomagnetic intensity, while also pushing archeological practice to its limits in terms of gathering temporally homogeneous artefacts from the same archeological context and dating precision. Beyond their dramatic implications for geomagnetism, geomagnetic spikes present practitioners in archaeomagnetism with an unprecedented challenge.

Availability of codes

The code and datafiles that can reproduce all results and figures in this work can be found at https://github.com/plivermore/AH-RJMCMC1

Author statement

All authors contributed equally.

Declaration of Competing Interest

None.

Acknowledgements

PWL would like to thank the invitation program of the Université de Paris that funded a visit to the Institut de physique du Globe de Paris (2020) during which much of this work was conducted. PWL was partially supported by NERC grant NE/G014043/1. YG would like to thank Maria Grazia Masetti-Rouault and Adonice-Ackad Baaklini for helpful discussion on Qatna. Marisa Osete and Saioa Campuzano are thanked for providing the SHAWQ-Iron Age intensity variation of Palmyra. F. Javier Pavón-Carrasco and an anonymous reviewer are thanked for providing constructive comments on the manuscript. This study was partly supported by the INSU-CNRS program PNP. This is IPGP contribution no. 4193.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pepi.2021.106657.

References

Alken, P., Thébault, E., Beggan, C.D., Amit, H., Aubert, J., Baerenzung, J., Bondar, T.N., Brown, W., Califf, S., Chambodut, A., Chulliat, A., Cox, G., Finlay, C.C., Fournier, A., Gillet, N., Grayver, A., Hammer, M.D., Holschneider, M., Huder, L., Hulot, G., Jager, T., Kloss, C., Korte, M., Kuang, W., Kuvshinov, A., Langlais, B., Léger, J.M., Lesur, V., Livermore, P.W., Lowes, F.J., Macmillan, S., Magnes, W., Mandea, M., Marsal, S., Matzka, J., Metman, M.C., Minami, T., Morschhauser, A., Mound, J.E., Nair, M., Nakano, S., Olsen, N., PavÃ3n-Carrasco, F.J., Petrov, V.G., Ropp, G., Rother, M., Sabaka, T.J., Sanchez, S., Saturnino, D., Schnepf, N.R., Shen, X., Stolle, C., Tangborn, A., Tøffner-Clausen, L., Toh, H., Torta, J.M., Varner, J., Vervelidou, F., Vigneron, P., Wardinski, I., Wicht, J., Woods, A., Yang, Y., Zeren, Z., Zhou, B., 2020. International Geomagnetic Reference Field: the 13th generation. Earth, Planets and Space. https://doi.org/10.1186/s40623-020-01288-x.

Al-Maqdissi, M., 2003. Recherches archéologiques syriennes à Mishirfeh-Qatna au nord-est de Homs (émèse). In: Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres, 147, pp. 1487–1515.

Al-Maqdissi, M., Bonacossi, D.M., 2005. The Metropolis of the Orontes: Art and Archaeology from the Ancient Kingdom of Qatna: Seven Years of Syrian-Italian Collaboration at Mishrifeh-Qatna (Damascus).

Baaklini, A.A., 2019. Présence et influence assyriennes dans le royaume de Hamat. Ph.D. thesis. Sorbonne université.

Ben-Yosef, E., Ron, H., Tauxe, L., Agnon, A., Genevey, A., Levy, T.E., Avner, U., Najjar, M., 2008. Application of copper slag in geomagnetic archaeointensity research. J. Geophys. Res. 113 https://doi.org/10.1029/2007JB005235.

Ben-Yosef, E., Tauxe, L., Levy, T.E., Shaar, R., Ron, H., Najjar, M., 2009. Geomagnetic intensity spike recorded in high resolution slag deposit in southern Jordan. Earth Planet. Sci. Lett. 287, 529–539.

Ben-Yosef, E., Millman, M., Shaar, R., Tauxe, L., Lipschits, O., 2017. Six centuries of geomagnetic intensity variations recorded by royal judean stamped jar handles. Proc. Natl. Acad. Sci. 114, 2160–2165.

Bourne, M.D., Feinberg, J.M., Stafford Jr., T.W., Waters, M.R., Lundelius Jr., E., Forman, S.L., 2016. High-intensity geomagnetic field 'spike' observed at ca. 3000 cal BP in Texas, USA. Earth Planet. Sci. Lett. 442, 80–92.

Brown, M., Herve, G., Korte, M., Genevey, A., 2021. Global archaeomagnetic data: the state-of-the-art and future challenges. Phys. Earth Planet. Inter. in review. this volume.

Cai, S., Tauxe, L., Deng, C., Pan, Y., Jin, G., Zheng, J., Xie, F., Qin, H., Zhu, R., 2014. Geomagnetic intensity variations for the past 8 kyr: new archaeointensity results from eastern China. Earth Planet. Sci. Lett. 392, 217–229.
Cai, S., Jin, G., Tauxe, L., Deng, C., Qin, H., Pan, Y., Zhu, R., 2017. Archaeointensity results spanning the past 6 kiloyears from eastern China and implications for extreme behaviors of the geomagnetic field. Proc. Natl. Acad. Sci. 114, 39–44.
Campuzano, S.A., Gómez-Paccard, M., Pavón-Carrasco, F.J., Osete, M.L., 2019. Emergence and evolution of the South Atlantic anomaly revealed by the new paleomagnetic reconstruction shawq2k. Earth Planet. Sci. Lett. 512, 17–26.
Constable, C., Korte, M., Panovska, S., 2016. Persistent high paleosecular variation activity in southern hemisphere for at least 10 000 years. Earth Planet. Sci. Lett. 453, 78–86.
Davies, C., Constable, C., 2017. Geomagnetic spikes on the core-mantle boundary. Nat. Commun. 8, 15593.
Davies, C.J., Constable, C.G., 2018. Searching for geomagnetic spikes in numerical dynamo simulations. Earth Planet. Sci. Lett. 504, 72–83.
Ertepinar, P., Langereis, C.G., Biggin, A.J., Frangipane, M., Matney, T., Ökse, T., Engin, A., 2012. Archaeomagnetic study of five mounds from upper Mesopotamia between 2500 and 700 BCE: further evidence for an extremely strong geomagnetic field ca. 3000 years ago. Earth Planet. Sci. Lett. 357, 84–98.
Ertepinar, P., Hammond, M.L., Hill, M.J., Biggin, A.J., Langereis, C.G., Herries, A., Yener, K., Akar, M., Gates, M.H., Harrison, T., Greaves, A.M., Frankel, D., Webb, J. M., Ozgen, I., Yazicioglu, G.B., 2020. Extreme geomagnetic field variability indicated by eastern Mediterranean full-vector archaeomagnetic records. Earth Planet. Sci. Lett. 531, 115979.
Finlay, C.C., Dumberry, M., Chulliat, A., Pais, M.A., 2010. Short timescale core dynamics: theory and observations. Space Sci. Rev. 155, 177–218. https://doi.org/10.1007/s11214-010-9691-6.
Gallet, Y., Al-Maqdissi, M., 2010. Archéomagnétisme à mishirfeh-qatna: Nouvelles données sur l'évolution de l'intensité du champ magnétique terrestre au moyen-orient durant les derniers millénaires. Akkadica 131, 29–46.
Gallet, Y., Butterlin, P., 2015. Archaeological and geomagnetic implications of new archaeomagnetic intensity data from the early bronze high terrace 'massif rouge' at Mari (tell Hariri, Syria). Archaeometry 57, 263–276.
Gallet, Y., Le Goff, M., 2006. High-temperature archeointensity measurements from Mesopotamia. Earth Planet. Sci. Lett. 241, 159–173.
Gallet, Y., Genevey, A., Courtillot, V., 2003. On the possible occurrence of 'archaeomagnetic jerks' in the geomagnetic field over the past three millennia. Earth Planet. Sci. Lett. 214, 237–242.
Gallet, Y., Genevey, A., Le Goff, M., Fluteau, F., Ali Eshraghi, S., 2006. Possible impact of the Earth's magnetic field on the history of ancient civilizations. Earth Planet. Sci. Lett. 246, 17–26.
Gallet, Y., D'Andrea, M., Genevey, A., Pinnock, F., Le Goff, M., Matthiae, P., 2014. Archeomagnetism at Ebla (tell Mardikh, Syria). New data on geomagnetic field intensity variations in the Near East during the Bronze Age. J. Archaeol. Sci. 42, 295–304.
Gallet, Y., Fortin, M., Fournier, A., Le Goff, M., Livermore, P., 2020. Analysis of geomagnetic field intensity variations in Mesopotamia during the third millennium BC with archeological implications. Earth Planet. Sci. Lett. 537, 116183.
Genevey, A., Gallet, Y., 2002. Intensity of the geomagnetic field in western Europe over the past 2000 years: New data from ancient French pottery. J Geophys Res-Sol Ea 107. EPM 1–1–EPM 1–18.
Genevey, A., Gallet, Y., Margueron, J.C., 2003. Eight thousand years of geomagnetic field intensity variations in the eastern mediterranean. J. Geophys. Res. 108, 2228.
Genevey, A., Gallet, Y., Constable, C., Korte, M., Hulot, G., 2008. ArcheoInt: an upgraded compilation of geomagnetic field intensity data for the past ten millennia and its application to the recovery of the past dipole moment. Geochem. Geophys. Geosyst. 9, Q04038 https://doi.org/10.1029/2007GC001881.
Genevey, A., Gallet, Y., Jesset, S., Thébault, E., Bouillon, J., Lefèvre, A., Goff, M.L., 2016. New archeointensity data from French early medieval pottery production (6th-10th century AD). Tracing 1500 years of geomagnetic field intensity variations in western europe. Phys. Earth Planet. Inter. 257, 205–219. https://doi.org/10.1016/j.pepi.2016.06.001.
Genevey, A., Gallet, Y., Thèbault, E., Livermore, P., Fournier, A., Jesset, S., LefÃvre, A., Mahé-Hourlier, N., Marot, E., Regnard, S., 2021. Archeomagnetic intensity investigations of French Medieval ceramic workshops: Contribution to regional field modeling and archeointensity-based dating. Phys. Earth Planet. Inter. in review. this volume.
Gimatzidis, S., Weninger, B., 2020. Radiocarbon dating the greek protogeometric and geometric periods: the evidence of sindos. PLoS One 15, e0232906.
Hellio, G., Gillet, N., Bouligand, C., Jault, D., 2014. Stochastic modelling of regional archaeomagnetic series. Geophys. J. Int. 199, 931–943.
Hervé, G., Faβbinder, J., Gilder, S.A., Metzner-Nebelsick, C., Gallet, Y., Genevey, A., Schnepp, E., Geisweid, L., Pütz, A., Reuβ, S., Wittenborn, F., Flontas, A., Linke, R., Riedel, G., Walter, F., Westhausen, I., 2017. Fast geomagnetic field intensity variations between 1400 and 400 BCE: new archaeointensity data from Germany. Phys. Earth Planet. Inter. 270, 143–156.
Herve, G., Chauvin, A., Lanos, P., Rochette, P., Perrin, M., Perron d'Arc, M., 2019. Cooling rate effect on thermoremanent magnetization in archaeological baked clays: an experimental study on modern bricks. Geophys. J. Int. 217, 1413–1424.
Hervé, G., Perrin, M., Alva-Valdivia, L., Tchibinda, B.M., Rodriguez-Trejo, A., Hernandez-Cardona, A., Tello, M.C., Rodriguez, C.M., 2019. Critical analysis of the holocene palaeointensity database in Central America: impact on geomagnetic modelling. Phys. Earth Planet. Inter. 289, 1–10.
Korte, M., Constable, C., 2011. Improving geomagnetic field reconstructions for 0–3ka. Phys. Earth Planet. Inter. 188, 247–259.
Korte, M., Constable, C.G., 2018. Archeomagnetic intensity spikes: global or regional geomagnetic field features? Front. Earth Sci. 6, 17. https://doi.org/10.3389/feart.2018.00017.
Korte, M., Genevey, A., Constable, C.G., Frank, U., Schnepp, E., 2005. Continuous geomagnetic field models for the past 7 millennia: 1. A new global data compilation. Geochem. Geophys. Geosyst. 6.
Lanos, P., 2004. Bayesian inference of calibration curves: application to archaeomagnetism. In: Tools for Constructing Chronologies. Springer, pp. 43–82.
Le Goff, M., Gallet, Y., 2004. A new three-axis vibrating sample magnetometer for continuous high-temperature magnetization measurements: applications to paleo- and archeo-intensity determinations. Earth Planet. Sci. Lett. 229, 31–43.
Licht, A., Hulot, G., Gallet, Y., Thébault, E., 2013. Ensembles of low degree archeomagnetic field models for the past three millennia. Phys. Earth Planet. Inter. 224, 38–67. https://doi.org/10.1016/j.pepi.2013.08.007.
Livermore, P.W., Fournier, A., Gallet, Y., 2014. Core-flow constraints on extreme archeomagnetic intensity changes. Earth Planet. Sci. Lett. 387, 145–156. https://doi.org/10.1016/j.epsl.2013.11.020.
Livermore, P.W., Fournier, A., Gallet, Y., Bodin, T., 2018. Transdimensional inference of archeomagnetic intensity change. Geophys. J. Int. 215, 2008–2034.
Metman, M.C., Livermore, P.W., Mound, J.E., Beggan, C.D., 2019. Modelling decadal secular variation with only magnetic diffusion. Geophys. J. Int. 219, S58–S82.
Nilsson, A., Holme, R., Korte, M., Suttie, N., Hill, M., 2014. Reconstructing Holocene geomagnetic field variation: new methods, models and implications. Geophys. J. Int. 198, 229–248.
Osete, M.L., Molina-Cardín, A., Campuzano, S.A., Aguilella-Arzo, G., Barrachina-IbaÃ±ez, A., Falomir-Granell, F., Foix, A.O., Gómez-Paccard, M., Martiín-Hernández, F., Palencia-Ortas, A., Pavón-Carrasco, F.J., Rivero-Montero, M., 2020. Two archaeomagnetic intensity maxima and rapid directional variation rates during the early iron age observed at iberian coordinates. implications on the evolution of the levantine iron age anomaly. Earth Planet. Sci. Lett. 533, 116047. URL. http://www.sciencedirect.com/science/article/pii/S0012821X19307423. https://doi.org/10.1016/j.epsl.2019.116047.
Paterson, G.A., Tauxe, L., Biggin, A.J., Shaar, R., Jonestrask, L.C., 2014. On improving the selection of Thellier-type paleointensity data. Geochem. Geophys. Geosyst. 15, 1180–1192.
Shaar, R., Ben-Yosef, E., Ron, H., Tauxe, L., Agnon, A., Kessel, R., 2011. Geomagnetic field intensity: how high can it get? How fast can it change? Constraints from Iron age copper slag. Earth Planet. Sci. Lett. 301, 297–306.
Shaar, R., Tauxe, L., Ben-Yosef, E., Kassianidou, V., Lorentzen, B., Feinberg, J.M., Levy, T.E., 2015. Decadal-scale variations in geomagnetic field intensity from ancient Cypriot slag mounds. Geochem. Geophys. Geosyst. 16, 195–214.
Shaar, R., Tauxe, L., Ron, H., Ebert, Y., Zuckerman, S., Finkelstein, I., Agnon, A., 2016. Large geomagnetic field anomalies revealed in bronze to Iron age archeomagnetic data from Tel Megiddo and Tel Hazor, Israel. Earth Planet. Sci. Lett. 442, 173–185.
Shaar, R., Tauxe, L., Goguitchaichvili, A., Devidze, M., Licheli, V., 2017. Further evidence of the Levantine Iron age geomagnetic anomaly from Georgian pottery. Geophys. Res. Lett. 44, 2229–2236.
Shaar, R., Bechar, S., Finkelstein, I., Gallet, Y., Martin, M., Ebert, Y., Keinan, J., Gonen, L., 2020. Synchronizing geomagnetic field intensity records in the Levant between the 23rd and 15th centuries bce: chronological and methodological implications. Geochem. Geophys. Geosyst. 21 (in press).
Suttie, N., Holme, R., Hill, M.J., Shaw, J., 2011. Consistent treatment of errors in archaeointensity implies rapid decay of the dipole prior to 1840. Earth Planet. Sci. Lett. 304, 13–21. https://doi.org/10.1016/j.epsl.2011.02.010.
Thébault, E., Gallet, Y., 2010. A bootstrap algorithm for deriving the archeomagnetic field intensity variation curve in the middle east over the past 4 millennia bc. Geophys. Res. Lett. 37 https://doi.org/10.1029/2010GL044788.
Troyano, M., Fournier, A., Gallet, Y., Finlay, C.C., 2020. Imprint of magnetic flux expulsion at the core–mantle boundary on geomagnetic field intensity variations. Geophys. J. Int. 221, 1984–2009.
Turner, G.M., Kinger, R., McFadgen, B., Gevers, M., 2020. The first archaeointensity records from new zealand: evidence for a fifteenth century ad archaeomagnetic 'spike' in the sw pacific region? Geol. Soc. Lond., Spec. Publ. 497. URL: https://sp.lyellcollection.org/content/early/2020/04/01/SP497-2019-71. https://doi.org/10.1144/SP497-2019-71. arXiv:https://sp.lyellcollection.org/content/early/2020/04/01/SP497-2019-71.full.pdf.
Yutsis-Akimova, S., Gallet, Y., Amirov, S., 2018. Rapid geomagnetic field intensity variations in the near east during the 6th millennium BC: new archeointensity data from Halafian site Yarim Tepe II (northern Iraq). Earth Planet. Sci. Lett. 482, 201–212.