A bootstrap algorithm for deriving the archeomagnetic field intensity variation curve in the Middle East over the past 4 millennia BC

E. Thébault, Y. Gallet

A bootstrap algorithm for deriving the archeomagnetic field intensity variation curve in the Middle East over the past 4 millennia BC

E. Thébault[1] and Y. Gallet[1]

Received 26 July 2010; revised 28 September 2010; accepted 4 October 2010; published 18 November 2010.

[1] Recent compilations of archeomagnetic intensity data have shown the existence of outliers. Acknowledging this problem, we propose a new way of generating and studying the uniqueness of regional archeomagnetic master curves bypassing the need to assume a normal data error distribution. Our approach lessens the weight of outliers by applying an iteratively re-weighted least-squares method combined with a bootstrap algorithm. Given a particular set of archeomagnetic data associated with experimental and dating uncertainties, we produce an ensemble of curves that are designed to sample the conditional probability distribution of the 'true' master curve. Using this technique, we propose an archeointensity master curve with its probability density for the Middle East region that covers the past four millennia BC.
Citation: Thébault, E., and Y. Gallet (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*, L22303, doi:10.1029/2010GL044788.

1. Introduction

[2] The recovery of the shape and strength of the ancient Earth's magnetic field has important implications for dynamo simulation, core flow investigation and dating purposes [e.g., *Constable*, 2007]. Prior to instrumental era and during archeological times, directional and intensity geomagnetic field variations can be inferred from the remanent magnetization imprinted in volcanic deposits, sediments or in human artefacts such as kilns, bricks, tiles or pottery. The measurements can then be interpolated in space and time at global or at continental scales [e.g., *Hongre et al.*, 1998; *Korte et al.*, 2009; *Pavón-Carrasco et al.*, 2009] but the models still suffer from the relatively poor spatial and temporal resolutions of the available data.

[3] At a regional scale, the data can be transferred to a common site under the Geomagnetic Axial Dipole (GAD) assumption and combined to directional and intensity master curves [*Le Goff et al.*, 2002; *Lanos et al.*, 2005]. This approach usually considers normally distributed errors and the existence of a unique solution. However, these hypotheses appear poorly verified because the large scatter of regional archeointensity datasets betrays the presence of outliers [e.g., *Genevey et al.*, 2008]. For this reason, we present a bootstrap algorithm allowing to mitigate the effect of outliers and to estimate the probability density function (*pdf*) of a regional master curve without making an explicit assumption on its form. Our algorithm is tested against synthetic data and then applied to real measurements to retrieve the geomagnetic field intensity variations in the Middle East between 4000 BC and 0 AD.

2. Fundamentals

[4] We have a discrete set of data $\mathbf{f} = \{f_i\}$, $i = 1..N$, associated with the epochs $\mathbf{t} = \{t_i\}$ and two sets of *prior* information $\epsilon^{\mathbf{f}} = \{\epsilon_i^f\}$ and $\epsilon^{\mathbf{t}} = \{\epsilon_i^t\}$ concerning the measurement and dating uncertainties. We construct a curve $\widehat{\mathbf{f}}(t)$ whose first time derivative is continuous and express it in terms of cubic B-splines with knots evenly spaced [*de Boor*, 2001]. The forward problem is

$$\widehat{\mathbf{f}}(t) = S(\alpha, t), \tag{1}$$

where $\alpha = \{\alpha_p\}$, $p = 1..P$, is the vector of parameters defining the function $S(\alpha, t)$. For over-determined problems ($N \geq P$) with *prior* information on measurement uncertainties, we solve a regularized weighted least-squares (RWLS) inverse problem [*Menke*, 1989]

$$\widehat{\alpha} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A} + \lambda\mathbf{D}\right)^{-1}\mathbf{A}^T\mathbf{W}^T\mathbf{f}, \tag{2}$$

where $\mathbf{A}$ is the design matrix of B–splines, $\mathbf{W}$ the diagonal matrix of weights $w_i = (1/\epsilon_i^f)^2$ and $\mathbf{D}$ a regularization matrix with λ a scalar determining the trade-off between misfit and regularization. In the absence of a more plausible physical model driving the intensity evolution, we assume relatively smooth temporal variations and we penalize the second time derivative of the solution as follows

$$\mathbf{D} = \int_t \partial_t^2\widehat{\mathbf{f}}(t)dt. \tag{3}$$

The marginal parameter uncertainties $\delta\widehat{\alpha}$ are then traditionally estimated by the diagonal entries of the covariance matrix

$$\delta\widehat{\alpha} = diag\left(\left[\mathbf{A}^T\mathbf{W}\mathbf{A} + \lambda\mathbf{D}\right]^{-1}\right). \tag{4}$$

For linear regression analyses with normal errors and $P \ll N$, the least-squares method provides the best estimator in the maximum likelihood sense. This approach assumes that there is only one solution and that the *a priori* and *a posteriori* data uncertainties are correctly estimated. However, the large dispersion in regional archeointensity data, such as in Western Europe [*Genevey et al.*, 2009], suggests that some data violate the assumption regarding the normal error distribution. We thus seek a way to verify the extent to which the

[1]Institut de Physique du Globe de Paris, Sorbonne Paris Cité, INSU, CNRS, Paris, France.

0094-8276/10/2010GL044788

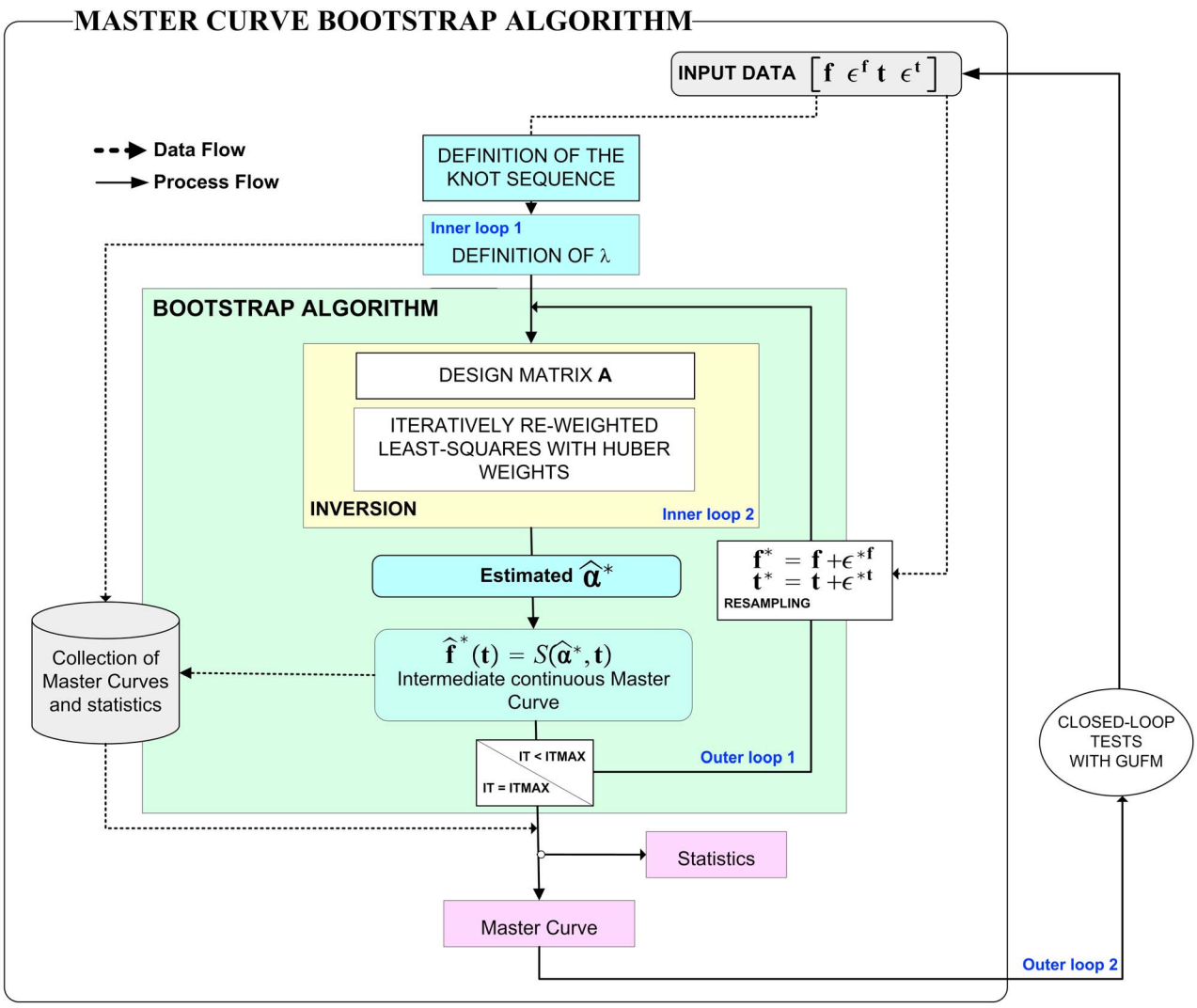


Figure 1. Schematic description of the master curve bootstrap algorithm (see text for further explanation). Solid and dashed lines are process and data flows, respectively.

solution is unique. We proceed numerically because the computation of a realistic error distribution is not feasible analytically since the measurement and dating errors do not necessarily follow the same distribution (generally Gaussian for the measurements but uniform for the dating) and that the scarcity of observations introduces possible non-linear interactions between the data errors and the model parameters.

[5] In our approach, we approximate a $L1$-norm solution to minify the detrimental effect of outliers. We apply an iteratively re-weighted least-squares method using Huber's weights (H-IRWLS) that considers the error distribution as normal within one standard deviation but longer tailed outside [*Huber*, 1981]. Then, we develop a bootstrap algorithm for investigating the non-uniqueness and variability of the solution. This technique aims at estimating a parameter either from a series of random data subsets, or by resampling the data with values modified thanks to *prior* information [*Davison and Hinkley*, 1999]. We apply an hybrid protocol because we have too few data to consider random subsets. Nevertheless, we remove each data once (the so-called *Jackknife* [*Efron and Tibshirani*, 1993]) to investigate its effect on the master curve solution, then we resample the data values and epochs using $\epsilon^{\mathbf{f}}$ and $\epsilon^{\mathbf{t}}$. This procedure, also implemented by *Korte et al.* [2009], does not formally take the dating error into account but only shows its influence.

3. Bootstrap Algorithm for Master Curves

[6] The algorithm is sketched in Figure 1. First, a sequence of knots evenly spaced between the minimum and the maximum epochs is defined by computing the mean time interval between two consecutive data. Then, λ is found automatically by trials: starting from a negligible value, the program records the evolution of the normalized misfit (*nrms*) with increasing λ (inner loop 1 in Figure 1). The first value of λ allowing fitting the data within tolerance (corresponding to *nrms* ~ 1) is selected and remains fixed thereafter. This automatic setting is suboptimal in two cases requiring user decision: when the data are clustered around some epochs and when *nrms* > 1 for all values of λ (i.e., if the *a priori* data uncertainties are severely underestimated).

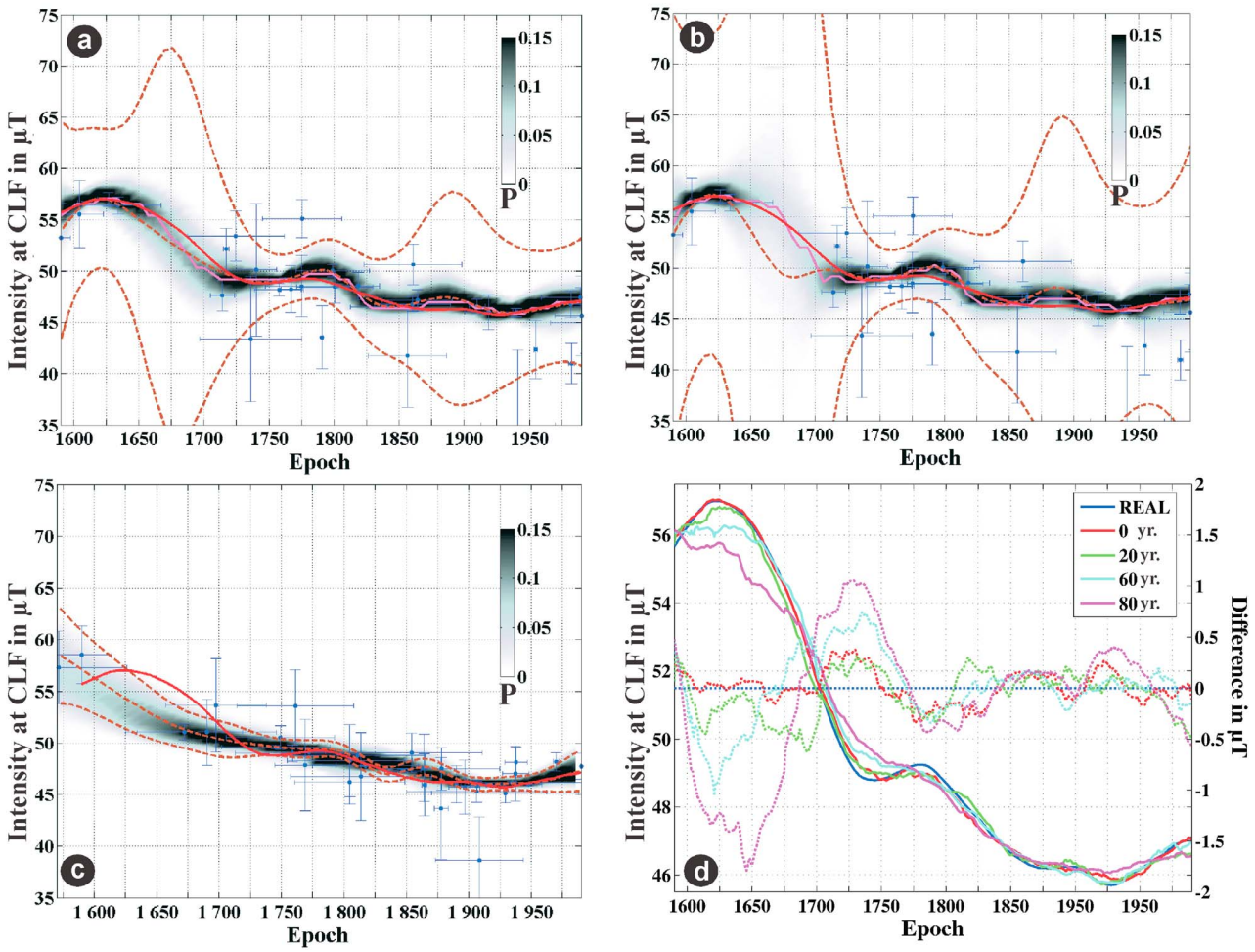


Figure 2. Synthetic curves obtained with a dataset simulated from the GUFM model. The brown dashed lines correspond to one single solution with 95% error bars obtained by Weighted Least-Squares inversion. The pink curve is the maximum of the *pdf* (represented by the grey shaded tones) found using the bootstrap algorithm. The red curve shows the "true" GUFM intensity values. The master curve is obtained considering a maximum dating uncertainty of ±40 yr with a (a) manual and (b) automatic setting of the regularization. (c) Same as in Figure 2b with a dating uncertainty of ±60 years. (d) Maximum of the *pdf* when the algorithm is bootstrapped using different dating errors. Left ordinates show the intensity variations of the maximum likelihood (solid curves); right ordinates show the difference with increasing dating uncertainties (legend in years) between the GUFM data and the solutions (dashed curves).

[7] The bootstrap procedure removes one data (*the jackknife*) and iteratively resamples and replaces the measurement values and ages. A new value is defined by $f_i^* = f_i + \epsilon_i^{*,f}$, where $\epsilon_i^{*,f}$ is the error sampled from the Gaussian distribution $\mathcal{N}(0, \epsilon_i^f)$, while a new date is defined by $t_i^* = t_i + \epsilon_i^{*,t}$, with $\epsilon_i^{*,t}$ sampled from the uniform distribution in the interval $t_i \pm \epsilon_i^t$.

[8] The inverse problem is solved by the H-IRWLS method and the parameters are estimated at each iteration (outer loop 1) by

$$\widehat{\alpha_{n+1}^*} = \widehat{\alpha_n^*} + \left(\mathbf{A}^{*,T}\mathbf{W}_n\mathbf{A}^* + \lambda\mathbf{D}\right)^{-1}\mathbf{A}^{*,T}\mathbf{W}_n^T(\mathbf{f}^* - S(\widehat{\alpha_n}, \mathbf{t}^*). \tag{5}$$

Starting from $n = 1$ and $w_i^1 = 1/\epsilon_i^f$, $\mathbf{W}_n$ is recomputed at each loop (inner loop 2) with $w_i^n = w_i^{n-1}/1.35$ (the constant of 1.35 is defined by *Huber* [1981]) when $|S(\widehat{\alpha_n}, t_i^*) - f_i^*| > \epsilon_i^f$ and $w_i^n = w_i^{n-1}$ otherwise. The symbol * indicates that data values $\mathbf{f}^*$ and epochs $\mathbf{t}^*$ are re-sampled and the matrix $\mathbf{A}^*$ recomputed. After convergence, the function $S(\widehat{\alpha^*}, t)$ is stored and the program proceeds until it reaches the final iteration.

[9] The program displays the master curve *pdf* and its temporal resolution varying with time. This latter information is important since it allows one to verify whether the solution is constrained by the data or dominated by the regularization. We define the temporal resolution $\tau(t)$ as the half-width of $\widetilde{\delta} = \mathbf{R}\delta(t)$ with

$$\mathbf{R} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A} + \lambda\mathbf{D}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{A}, \tag{6}$$

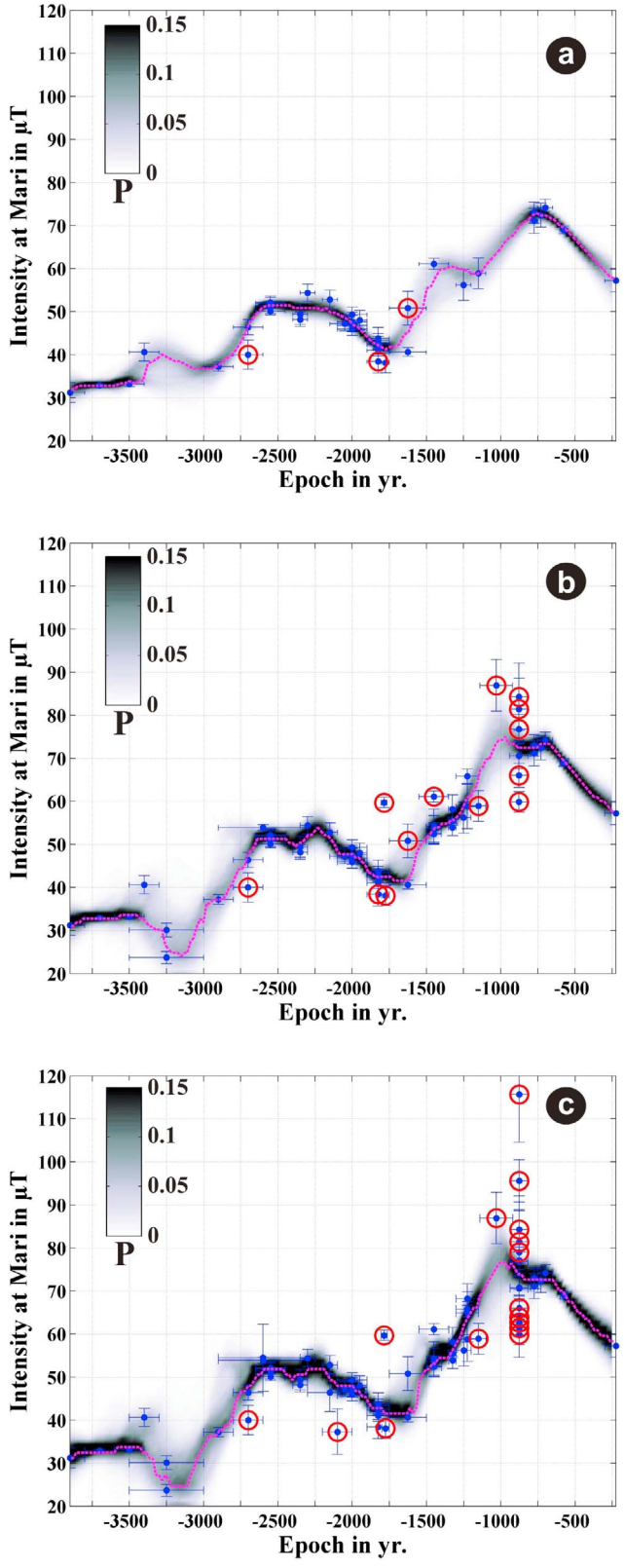
the resolution matrix, and $\delta(t)$ the Dirac function. This resolution time should be used qualitatively and with extreme

caution [*Parker*, 1994, chap. 4.02], especially near the interval boundaries where edge effects occur.

4. Closed-Loop Test

[10] We tested the algorithm against synthetic datasets that mimic the characteristics of a regional archeomagnetic data collection. A discrete series of magnetic field intensities was computed for every year at the location of Chambon-La-Forêt, France, between 1590 AD and 1990 AD with the GUFM model [*Jackson et al.*, 2000]. We removed randomly 90% of these perfect data and imposed that at least one 50-yr time interval, selected randomly, was empty. This simulates data sparsity and gaps. Then, we defined randomly the vectors of intensity ϵ^f and dating ϵ^t uncertainties. The maximum element of $\epsilon^{\mathbf{f}}$ (the standard deviations) does not exceed the arbitrary value of 3 μT and the maximum dating uncertainty does not exceed $\epsilon^t_{\max}$. Intensity and dating errors for each data were then selected from the normal and the uniform distributions using ϵ^f_i and $\pm\epsilon^t_i$. Finally, we simulated outliers by shifting randomly 10% of the remaining experimental intensity values outside their 95% confidence interval. Thus, starting from an original sample of 400 synthetic data equally distributed in time we ended with 30–40 noisy data unevenly distributed in time.

[11] Figure 2 displays two simulated datasets with their uncertainties. In Figure 2a, the damping parameter λ was set manually (10^{-8}) with $\epsilon^t_{\max}$ = 40 yr. The estimated master curve behaves reasonably well, encompassing the GUFM model, and replicating its oscillations. Figure 2b shows the curve estimated with an automatic setting of λ (10^{-10}). The curve suggests that the inverse problem is unstable and justifies setting λ manually. It also illustrates one remarkable property of the bootstrap algorithm: the maximum likelihood converges to the same value, which is not the case of the solution obtained with the RWLS solution (brown dashed curve). A second test was performed with an automatic setting and a new random dataset generated with $\epsilon^t_{\max}$ = 60 yr. This case was found to induce a noticeable loss of time resolution (Figure 2c). Figure 2d generalizes this result; for different fixed $\epsilon^t_{\max}$, we applied the algorithm 20 times on different random datasets (outer loop 2 in Figure 1), recorded at each iteration only the maximum likelihood (pink curve in Figure 2a), and finally computed the median of this population. For $\epsilon^t_{\max}$ = 0, the solution should (and does) converge towards the GUFM time series (within a modelling error incurred by the regularization and the knot spacing). In all cases, the dating uncertainty invariably leads to smooth estimates of the magnetic field temporal variations.



5. Application to the Geomagnetic Field Intensity Variations in the Middle East Between 4000 BC and 0 AD

[12] Thanks to recent archeomagnetic studies carried out in Syria [*Genevey et al.*, 2003; *Gallet et al.*, 2006, 2008; *Gallet and Al Maqdissi*, 2010], 39 independent dated archeointensity data covering the past four millennia BC are now available for this country. This dataset is homogenous considering that all results were obtained by the same team using up-to-date

Figure 3. Geomagnetic field intensity variations in the Middle East between 4000 BC and 0 AD (see Figure 2 for the color codes). (a) Archeointensity master curve computed with its *pdf* obtained at Mari/Tell Hariri (34° N33′, 40°E53′) using Syrian data (see text for references). Same as Figure 3a with an expanded dataset using (b) strict and (c) loose criteria discussed in the text. Circles indicate data that are considered as outliers by the algorithm.

experimental methods with a careful evaluation of possible occurrence of magnetomineralogical alteration during thermal treatment, and systematically taking into account the anisotropy and cooling rate effects on thermoremanent magnetization (TRM) acquisition [e.g., *Genevey et al.*, 2008]. All archeointensity data have a common definition, each being the average of results obtained from at least three independent baked clay fragments found from the same archeological context, and being dated from archeological and historical constraints.

[13] We report in Figure 3a the obtained Syrian master curve. The data were transferred to the archeological site of Mari/Tell Hariri, Syria using the GAD approximation. The curve, derived automatically, has a 154-yr knot spacing. It exhibits two intensity maxima, the first between 2500 and 2000 BC, the second around 750 BC. *Gallet et al.* [2008] and *Gallet and Al Maqdissi* [2010] discussed the possible occurrence of more rapid variations, in particular during the second half of the third millennium BC with two successive relative maxima, and another peak around the middle of the second millennium BC. The *pdf*, however, indicates that such behavior could only be assessed by further data acquisition. Interestingly, the *pdf* is bimodal between 3500 and 3000 BC, and between 1500 and 1300 BC, which shows that the data are not informative enough to ascertain the significance of the peaks found in the maximum likelihood sense around 3250 and 1400 BC.

[14] We next expanded the previous dataset with archeointensity data obtained from the southern Levantine region [*Ben-Yosef et al.*, 2008, 2009] and Iran [*Gallet et al.*, 2006]. The entire dataset is now less homogeneous since the former data were mostly obtained from a different type of archeological objects (metallurgical -slag- residues) dated by radiocarbon and with an experimental procedure that involved no systematic TRM anisotropy correction and assumed negligible cooling rate effect. Furthermore, those data correspond to mean intensity values computed at the fragment level. Using the MagIC database (http://earthref.org/MAGIC), we selected the best-dated data applying "strict" versus "loose" criteria: mean intensity defined by at least three specimens per fragment (following *Ben-Yosef et al.* [2008]) versus two, and with a standard deviation of less than 10% of the mean (as considered for the Syrian and Iranian data) versus 15%. This double selection yielded subsets of 13 and 27 data, to which we added 7 data from Iran. For simplicity, the algorithm resampling procedure considered all age uncertainties drawn from a uniform *pdf* although we deal in some cases with radiocarbon dating. The two cases call for user decision because $nrms > 1$, hence revealing a significant number of underestimated data errors. For better comparison, we computed the curves with the same knot spacing (107 yr) and $\lambda = 10^{-6}$. The incorporation of the additional data provides master curves very similar to that derived from the Syrian data only. Expanding the data collection from 39 to 59 or 73 data (and increasing the temporal resolution of the curve, see auxiliary material) only reduces the ambiguity due to the bimodal *pdf* discussed previously but does not help to resolve more details concerning the intensity evolution in the Middle East region.[1]

[15] In all three cases, the curve after 1250 BC is mostly constrained by the regularization and the data errors follow a double exponential distribution (see auxiliary material). Our algorithm highlights values that are down-weighted in more than 68% of the iterative process (encircled data in Figure 3). These data lie outside the *pdf* distribution estimated over their age range and can be considered as outliers regarding the available data collection. This concerns 3 out of 39 Syrian data, and 13 out of 59 and 16 out of 73 data, when applying strict and loose selection criteria, which would probably benefit from further experimental analyses and discussion with archeologists for better age determination.

6. Concluding Remarks

[16] The iterative inverse method that we propose for constructing regional intensity master curves presents several merits, among which the possibility to study the uniqueness of the solutions in the presence of outliers, to avoid making stringent assumption on the *pdf* associated with the master curves, and to visualize the consequences of data age uncertainties. The application of our approach on the archeointensity data recently obtained from the Middle East allows one to verify both the consistency of apparent short-term variations and which data, given a particular dataset, can be considered as outliers. This information is important if the master curve is to be used for archeomagnetic dating purposes, for deciphering the possible occurrence of regional non-dipole features, or for advocating possible correlations with other records such as climatic ones.

[17] **Acknowledgments.** We thank C. Finlay, A. Genevey, G. Hulot, J. J Schott, and two anonymous reviewers for their comments. This study was partly financed by the INSU-CNRS program SYSTER. This is IPGP contribution 3075.

References

Ben-Yosef, E., H. Ron, L. Tauxe, A. Agnon, A. Genevey, T. E. Levy, U. Avner, and M. Najjar (2008), Application of copper slag in geomagnetic archaeointensity research, *J. Geophys. Res.*, *113*, B08101, doi:10.1029/2007JB005235.

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

Constable, C. G. (2007), Archeomagnetic and paleomagnetic studies of centennial to millennial-scale geomagnetic field variations, in *Treatise on Geophysics*, vol. 5, *Geomagnetism*, edited by G. Schubert, chap. 9, pp. 337–372, Elsevier, Amsterdam.

Davison, A. C., and D. V. Hinkley (1999), *Bootstrap Methods and Their Application*, 592 pp., Cambridge Univ. Press, Cambridge, U. K.

de Boor, C. (2001), *A Practical Guide to Splines*, 368 pp., Springer, New York.

Efron, B., and R. Tibshirani (1993), *An Introduction to the Bootstrap*, 456 pp., Chapman and Hall, New York.

Gallet, Y., and M. Al Maqdissi (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., A. Genevey, M. Le Goff, F. Fluteau, and S. A. Eshraghi (2006), Possible impact of the Earth's magnetic field on the history of ancient civilizations, *Earth Planet. Sci. Lett.*, *246*, 17–26.

Gallet, Y., M. Le Goff, A. Genevey, J. Margueron, and P. Matthiae (2008), Geomagnetic field intensity behavior in the Middle East between ~3000 BC and ~1500 BC, *Geophys. Res. Lett.*, *35*, L02307, doi:10.1029/2007GL031991.

Genevey, A., Y. Gallet, and J.-C. Margueron (2003), Eight thousand years of geomagnetic field intensity variations in the eastern Mediterranean, *J. Geophys. Res.*, *108*(B5), 2228, doi:10.1029/2001JB001612.

Genevey, A., Y. Gallet, C. G. Constable, M. Korte, and G. Hulot (2008), ArcheoInt: An upgraded compilation of geomagnetic field intensity data

[1] Auxiliary materials are available in the HTML. doi:10.1029/2010GL044788.

for the past ten millennia and its application to the recovery of the past dipole moment, *Geochem. Geophys. Geosyst.*, *9*, Q04038, doi:10.1029/2007GC001881.

Genevey, A., Y. Gallet, J. Rosen, and M. Le Goff (2009), Evidence for rapid geomagnetic field intensity variations in Western Europe over the past 800 years from new archeointensity French data, *Earth Planet Sci. Lett.*, *284*, 132–143.

Hongre, L., G. Hulot, and A. Khokhlov (1998), An analysis of the geomagnetic field over the past 2000 years, *Phys. Earth Planet. Inter.*, *106*, 311–335.

Huber, P. J. (1981), *Robust Statistics*, 308 pp., Wiley, New York.

Jackson, A., A. R. T. Jonkers, and M. R. Walker (2000), Four centuries of geomagnetic secular variation from historical records, *Philos. Trans. R. Soc. A*, *358*, 957–990.

Korte, M., F. Donadini, and C. G. Constable (2009), Geomagnetic field for 0–3 ka: 2. A new series of time-varying global models, *Geochem. Geophys. Geosyst.*, *10*, Q06008, doi:10.1029/2008GC002297.

Lanos, P., M. Le Goff, M. Kovacheva, and E. Schnepp (2005), Hierarchical modelling of archeomagnetic data and curve estimation by moving average technique, *Geophys. J. Int.*, *160*, 440–476.

Le Goff, M., Y. Gallet, A. Genevey, and N. Warmé (2002), On archeomagnetic secular variation curves and archeomagnetic dating, *Phys. Earth. Planet. Inter.*, *134*, 201–203.

Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, 289 pp., Academic, San Diego, Calif.

Parker, R. L. (1994), *Geophysical Inverse Theory*, 1st ed., 408 pp., Princeton Univ. Press, Princeton, N. J.

Pavón-Carrasco, F. J., M. L. Osete, J. M. Torta, and L. R. Gaya-Piqué (2009), A regional archeomagnetic model for Europe for the last 3000 years, SCHA.DIF.3K: Applications to archeomagnetic dating, *Geochem. Geophys. Geosyst.*, *10*, Q03013, doi:10.1029/2008GC002244.

Y. Gallet and E. Thébault, Institut de Physique du Globe de Paris, Sorbonne Paris Cité, INSU, CNRS, 1 rue Jussieu, F-75005 Paris CEDEX, France. (gallet@ipgp.fr; ethebault@ipgp.fr)