Panovska, S., Korte, M., Finlay, C. C., Constable, C. G. (2015): Limitations in paleomagnetic data and modelling techniques and their impact on Holocene geomagnetic field models. - *Geophysical Journal International*, *202*, 1, p. 402-418.

DOI: http://doi.org/10.1093/gji/ggv137

Geophysical Journal International

Geophys. J. Int. (2015) **202,** 402–418 doi: 10.1093/gji/ggv137

GJI Geomagnetism, rock magnetism and palaeomagnetism

Limitations in paleomagnetic data and modelling techniques and their impact on Holocene geomagnetic field models

S. Panovska,[1] M. Korte,[2] C. C. Finlay[3] and C. G. Constable[1]

[1]*Institute for Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, CA, USA. E-mail: spanovska@ucsd.edu*

[2]*Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum GFZ, Germany*

[3]*DTU Space, Technical University of Denmark, DK-2800, Kongens Lyngby, Denmark*

Accepted 2015 March 19. Received 2015 March 18; in original form 2014 October 24

SUMMARY

Characterization of geomagnetic field behaviour on timescales of centuries to millennia is necessary to understand the mechanisms that sustain the geodynamo and drive its evolution. As Holocene paleomagnetic and archeomagnetic data have become more abundant, strategies for regularized inversion of modern field data have been adapted to produce numerous time-varying global field models. We evaluate the effectiveness of several approaches to inversion and data handling, by assessing both global and regional properties of the resulting models. Global Holocene field models cannot resolve Southern hemisphere regional field variations without the use of sediments. A standard data set is used to construct multiple models using two different strategies for relative paleointensity calibration and declination orientation and a selection of starting models in the inversion procedure. When data uncertainties are considered, the results are similar overall regardless of whether we use iterative calibration and reorientation, or co-estimation of the calibration and orientation parameters as part of the inversion procedure. In each case the quality of the starting model used for initial relative paleointensity calibration and declination orientation is crucial and must be based on the best absolute information available. Without adequate initial calibration the morphology of dipole moment variations can be recovered but its absolute value will be correlated with the initial intensity calibrations, an effect that might be mitigated by ensuring an appropriate fit to enough high quality absolute intensity data with low uncertainties. The declination reorientation mainly impacts regional field structure and in the presence of non-zonal fields will result in a non-zero local average. The importance of declination orientation is highlighted by inconsistencies in the West Pacific and Australian sediment records in CALS10k.1b model. Great care must also be taken to assess uncertainties associated with both paleomagnetic and age data and to evaluate the effects of poor data distribution. New consistently allocated uncertainty estimates for sediment paleomagnetic records highlight the importance of adequate uncertainties in the inversion process, as they determine the relative weighting among the data and overall normalized misfit levels which in turn influence the complexity of the inferred field models. Residual distributions suggest that the most appropriate misfit measure is the L_1 norm (minimum absolute deviation) rather than L_2 (least squares), but this seems to have relatively minor impact on the overall results. For future Holocene field modelling we see a need for comprehensive methods to assess uncertainty in individual archeomagnetic data so that these data or models derived from them can be used for reliable initial relative paleointensity calibration and declination orientation in sediments. More work will be needed to assess whether co-estimation or an iterative approach to inversion is more efficient overall. This would be facilitated by realistic and globally consistent data and age uncertainties from the paleomagnetic community.

Key words: Archaeomagnetism; Magnetic field; Palaeointensity; Palaeomagnetic secular variation.

1 INTRODUCTION

Global geomagnetic field models spanning the past four hundred years (Jackson *et al.* 2000) have proven to be a powerful tool, enabling the mapping of evolving field structure at the surface of Earth's core, and are also often used for analyses of global and regional field variations. Such models have been extended to millennial time scales using compilations of field directions and intensity obtained from archaeological artefacts, lavas and sediments. In comparison to modern or historical observational data, however, archeo- and paleomagnetic data contain rather large and often not well understood uncertainties not only in the field values but also in their ages, and their global distribution is very heterogeneous. Several modelling methodologies have been attempted to assess the influences of these limitations. The CALSx series of global geomagnetic models (Korte & Constable 2003, 2005; Korte *et al.* 2009; Korte & Constable 2011; Korte *et al.* 2011) covering the past 3, 7 and 10 kyr use spherical harmonic expansions and cubic spline temporal basis functions together with regularization methods to find the models with minimum complexity that fit the data to the desired accuracy. Licht *et al.* (2013) presented three ensembles of low spherical harmonic degree field models (truncated at degree and order 5) that cover the past three millennia, built from archeomagnetic, volcanic and sedimentary data sets. They argued that only low degree features can be resolved with the available data sets and introduced a modelling error to account for unmodelled higher order Gauss coefficients. They increased the weights of archeomagnetic compared to sediment data, arguing that sediment data should be treated with care, mainly due to timing errors. Considering the dating issue, Nilsson *et al.* (2014) published a new set of geomagnetic field models covering the last 9000 years employing a new modelling strategy with temporally evenly resampled sediment records and iterative adjustments of their timescales, taking the age uncertainties into account. Recently, Pavón-Carrasco *et al.* (2014b) proposed a new Holocene geomagnetic field model (SHA.DIF.14k) spanning the past 14 kyr based only on archeomagnetic and lava flow data, avoiding the use of sedimentary records.

Uncertainty estimates are important prior information for global field modelling because they strongly influence the relative weighting of different data sources relative to one another. Several strategies have been employed to estimate uncertainties for modelling purposes. Korte *et al.* (2005) assigned minimum uncertainties for sediment records of 3.5° in inclination, 5.0° in declination and 5 μT in intensity, based on the comparisons with the historical model *gufm1* (Jackson *et al.* 2000). For the more recent CALSxk models, minimum uncertainties for sediment directional data have been expressed in terms of an α_{95} of 6° (Donadini *et al.* 2009; Korte *et al.* 2009). To take into account both data and age uncertainties, recent models CALS3k.4b and CALS10k.1b were obtained by averaging multiple models built with resampling of the data sets across distributions representing their estimated data and age uncertainties (Korte *et al.* 2011). Licht *et al.* (2013) assigned error estimates to each data point by combining in quadrature a modelling error of 5 μT for intensity and $\alpha_{95} = 3.4°$ for directions and the published measurement error. Uncertainty estimates for the Holocene sediment magnetic records were obtained by Panovska *et al.* (2012) via comparisons with archeomagnetic estimates and from the data variance of individual records about robust smoothing spline models. The resulting error estimates demonstrated a diversity in the quality of the sediment paleomagnetic records, with a much wider range of uncertainties than used so far in global modelling of sediment records.

Archeointensity data are of great importance for Holocene geomagnetic field modelling because of their use for scaling the global magnitude of the geomagnetic field. Potential sources of the uncertainty in archeomagnetic data include: (1) dating errors; (2) cooling rate differences; (3) magnetic anisotropy (Lanos *et al.* 2005); as well as (4) uncertainties related to the thermal remanence acquisition, local magnetic field anomalies, and errors during sampling (Constable *et al.* 2000). Data from archeomagnetic and lava flows are often obtained by averaging several measurements from individual samples. Based on a comparison between archeomagnetic data and the historical geomagnetic field model *gufm1* for overlapping times, α_{95} for archeomagnetic directional data have been assigned a minimum value of 4.3° and 5 μT minimum uncertainty for archeointensity (Donadini *et al.* 2009; Korte & Constable 2011; Korte *et al.* 2011). Previously used percentage errors for archeomagnetic or lava flow intensity data rather than absolute uncertainty values have been found to result in a biased inference of global intensity (Suttie *et al.* 2011). Nilsson *et al.* (2014) assigned minimum errors depending on the number of samples/specimens used to calculate the mean direction or intensity, while Pavón-Carrasco *et al.* (2014a) assessed the reliability of European archeointensity data by assigning them in different quality categories based on paleomagnetic criteria. Another treatment of measurement uncertainties has recently been employed in the construction of CALS3k.4b and CALS10k.1b models, with bootstrap archeomagnetic samples obtained from a normal distribution centred on the magnetic field value with a standard deviation equal to the data uncertainty estimate.

When constructing millennial time scale field models it is often assumed that the noise inherent in the measurements may be described by a Gaussian distribution; consequently the L_2 norm measure of misfit is employed. In many geophysical scenarios however, when longer tailed distributions are found empirically, a Laplacian distribution of residuals is a more suitable description and the L_1 maximum likelihood estimate of parameters is desired (Claerbout & Muir 1973; Constable 1988; Walker & Jackson 2000). Use of the L_1 measure of misfits is less sensitive to the influence of outliers and yields more stable model estimates (e.g. Claerbout & Muir 1973; Scales *et al.* 1988; Farquharson & Oldenburg 1998; Tarantola 2005). The L_1 norm measure of misfit has been successfully employed in geomagnetic field modelling of recent and historical data (Walker & Jackson 2000; Lesur *et al.* 2008; Finlay *et al.* 2012) and regional archeomagnetic field models (Pavón-Carrasco *et al.* 2009), but it has not previously been tested in the construction of global field models on millennial time scales.

A serious complication of using sediment paleomagnetic data for field modelling is the fact that intensity and declination records are usually relative in nature, rather than absolute. Relative paleointensity (RPI) records used in the CALS3k models (Korte *et al.* 2009; Donadini *et al.* 2009) were calibrated using comparisons to a previous model as described by Korte & Constable (2006), and by comparison with archeomagnetic data from nearby locations. Relative declination records were compared to the predictions from the previous models (e.g. CALS3K.1), and unoriented records were adjusted according to their overlap with the historical *gufm1* model by Jackson *et al.* (2000) before being used in construction of the earliest CALSxk models (Korte & Constable 2005). However, in the most recent models, CALS3K.4b and CALS10k.1b (Korte & Constable 2011; Korte *et al.* 2011), the calibration strategy was improved by implementation of iterative re-calibration of RPI combined with outlier rejection. Nilsson *et al.* (2014) adjusted each declination and RPI record using a comparison with a prior tilted

dipole field model or archeomagnetic data, when enough data points are available.

Here, we present a new set of spherical harmonic geomagnetic field models covering the Holocene based on sediment records and archeomagnetic data (including lava flows) designed to test a number of new modelling strategies. In particular, we test the impact of (i) weighting the sediment paleomagnetic records by new uncertainty estimates, (ii) calibration of RPI and determination of declination offset within the inversion rather than using prior adjustment, (iii) the starting models for such calibrations, and (iv) investigating the influence of L_1 versus L_2 measures of misfit. Models are compared among each other and to previously published models in terms of the evolution of spatial and temporal norms and misfit, spherical harmonic power spectra, time variations in dipole and quadrupole Gauss coefficients, and some regional and global model predictions at the Earth's surface and the core-mantle boundary.

The structure of the paper is as follows. Section 2 gives an overview of the available data compilation of sediment magnetic records and archeomagnetic data and their uncertainty estimates. Section 3 presents technical details about the modelling, inversion procedure and a new method of implementation for non-absolute quantities (relative declination and RPI). Section 4 collates the results: global properties of the models, similarities and differences in model parameters, regional variations and comparison of the time-averaged non-axial dipole field structure. Sections 5 and 6 present a discussion of results and conclusions, respectively.

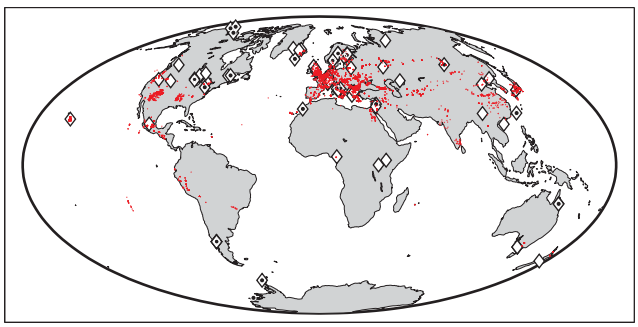

Figure 1. The global spatial distribution of Holocene sediment paleomagnetic records used in this study, directional (declination or inclination) data (white diamonds), RPI (black circles) and archeomagnetic data and lava flows (red dots). Only inclination data are available for the records AD1 (Adriatic Sea), PEP (Lake Pepin, USA), TUR (Lake Turkana, Kenya), and WPA (West Pacific). See supplementary table (Table S1) for codes, full names, locations, coordinates and references of all sediment records. (A colour version of this figure is available in the online version.)

2 DATA AND UNCERTAINTY ESTIMATES

Paleomagnetic data spanning the past 10 kyr used to construct the CALS10k.1b model by Korte *et al.* (2011) are the basis for the new field models presented here. The data set has comparatively poor coverage of Southern hemisphere (see Fig. 1 for spatial distribution of the data), and also the type of data (sediment vs. archeomagnetic) and elements are known to produce different results (Donadini *et al.* 2009). The whole data set comprises 85 500 data, of which about 4 per cent are archeomagnetic declination data, 6 per cent are archeomagnetic inclination and 5 per cent are absolute archeomagnetic intensity, while the sediment paleomagnetic data contribute 35 per cent relative declination, 37 per cent inclination and 13 per cent RPI (Table 1). Code name, location, coordinates and references of each sediment record are listed in supplementary table (Table S1). Initial data sets including the uncertainties used are available online from the EarthRef Digital Archive (ERDA) at http://earthref.org/ERDA/2101.

Korte & Constable (2011) previously found that the RPI record from Lake Pepin, USA (Brachfeld & Banerjee 2000), has a suspicious drop in amplitude from 1800 AD to 2000 AD and these unreliable data caused a reverse flux patch at the core-mantle boundary (CMB) over North America in the CALS3k.3 model which does not agree with historical field information provided by *gufm1* (Jackson *et al.* 2000). Therefore, this part of the Lake Pepin record was discarded before carrying out the modelling. The Lake Biwa record (BIW) (Ali *et al.* 1999) has been replaced with a new record from the same lake (BI2) (Hayashida *et al.* 2007), while both records were used in the models CALS10k.1b (Korte *et al.* 2011) and pfm9k (Nilsson *et al.* 2014). To avoid the problems of heterogeneous sediment data set in building the pfm9k models, Nilsson *et al.* (2014) resampled all sedimentary records in 50-yr bins and obtained uncertainty estimates based on the dispersion of the data within each bin. We instead used the raw data with improved uncertainty estimates determined as follows.

Revised uncertainty estimates for the sediment paleomagnetic records were obtained by Panovska *et al.* (2012) with the aim of providing a more consistent weighting of the various sediment records and these have been used for most of the models presented here (see Section 4, Table 1). These uncertainties are obtained by fitting smoothing splines to each record using a cross-validation

Table 1. Data sets used to construct the HFM.* and CALS10k.* models. GAD denotes a geocentric axial dipole model with $g_1^0 = 30\,\mu$T. The final number of data after iterative rejection and the percentage of rejected data are reported. The original number of data is 85 500 (12 830 archeomagnetic and 72 670 sediment data).

Model	Measure of misfit	Starting model for Gauss coefficients	Calibration of rel. D and scaling of RPI	Number of final data (reject.)	rms
HFM.OL1	L_1	GAD	GAD[a]	85 000 (0.58 per cent)	1.096
HFM.OL2	L_2	GAD	GAD[a]	85 132 (0.43 per cent)	1.052
HFM.OL1c	L_1	CALS10k.1b	CALS10k.1b[a]	85 019 (0.56 per cent)	1.085
CALS10k.OL2c	L_2	GAD	CALS3k.3[b]	82 389 (3.64 per cent)	1.030
CALS10k.OL2	L_2	GAD	GAD[b]	82 308 (3.73 per cent)	0.994
CALS10k.OL2w*	L_2	GAD	GAD[b]	82 719 (3.25 per cent)	1.603

[a]Models used to calculate initial scaling factor for RPI and offsets for rel. declination.
[b]Field models employed to calibrate RPI and rel. declination prior to the inversion.
*CALS10k.OL2w model differs in the uncertainty estimates of sediment data.

approach, based on an absolute deviation (L_1) measure of misfit. The procedure is bounded by a minimum smoothing time derived from the sedimentation rate and an assumed fixed lock-in depth. Departures from the spline models provide information concerning the random variability in each record. The results are combined with comparisons among the sediment magnetic records and archeomagnetic data and with predictions from the global historical and archeomagnetic field models to obtain individual uncertainty estimates for each sediment record. The resulting error estimates are mostly higher and more variable amongst the records than the error estimates that have been used in previous global field models. Uncertainty estimates obtained in this way for the RPI and inclination of the West Pacific (WPA) record (Richter *et al.* 2006), however, appeared unrealistically small. Initial modelling tests showed a very strong influence of this dense record on the models including even the dipole moment (DM). In accordance with Nilsson *et al.* (2014), who considered WPA to show an incompatible long-term trend over the Holocene with the nearby records, we omitted the WPA record from our models.

The archeomagnetic data set used is the same as that employed in the construction of CALS10k.1b model by Korte *et al.* (2011) and we adopted the same uncertainty estimates for archeomagnetic data. Equations for conversion of α_{95} to σ_I and σ_D are given in Donadini *et al.* (2009). Based on the average deviation of archeomagnetic data and historical *gufm1* model, minimum uncertainty estimates of 5 μT and a minimum α_{95} of 4.3° are applied to archeomagnetic data (Korte *et al.* 2009; Korte & Constable 2011; Korte *et al.* 2011). Such absolute thresholds address the criticism raised by Suttie *et al.* (2011) that percentages of intensities as error thresholds will later result in a biased inference of global intensity.

3 FIELD MODELLING METHODOLOGY

3.1 Forward modelling and parametrization

The solution of the forward problem of geomagnetic field modelling may be expressed in terms of spherical harmonics as

$$V(r,\theta,\phi) = a\sum_{l=1}^{L}\sum_{m=0}^{l}\left(\frac{a}{r}\right)^{l+1}\left[g_l^m\cos\ m\phi + h_l^m\sin\ m\phi\right]P_l^m(\theta) \tag{1}$$

where $a = 6371.2$ km is the Earth's mean radius, r, θ, ϕ are the geocentric spherical coordinates, radius, colatitude and longitude, and P_l^m are the Schmidt quasi-normalized associated Legendre functions of degree l and order m (e.g. Chapman & Bartels 1940; Langel 1987).

Modelling the temporal evolution of the geomagnetic field requires a further expansion of the Gauss coefficients g_l^m and h_l^m in time. For this purpose, a cubic B-splines basis is adopted

$$g_l^m(t) = \sum_{k}^{N_{\rm spl}} g_l^{mk}B_k(t) \tag{2}$$

and the same for h_l^m, where $N_{\rm spl}$ is the number of B-splines. The time-dependent Gauss coefficients are thus linear combinations of the spline coefficients and the piecewise polynomial functions $B_k(t)$ of degree 3 (order 4).

In the new models presented here we use a maximum spherical harmonic degree of $L = 10$ and a temporal knot point spacing of 40 years. This is the same parameterization as used by Korte *et al.* (2011). The chosen maximum degree of the spherical harmonic expansion and knot point spacing allow for more spatial and temporal structure than is expected to be resolved by the data. Overfitting of the data is avoided by spatial and temporal regularization as described in Section 3.2. This ensures that the resolution of the model is determined by the information in the data and not by an arbitrary choice of truncation of spherical harmonic expansion or knot point spacing of the splines.

Following Korte *et al.* (2011) our new models span the past 12 kyr, but are considered valid only for the past 10 kyr, with the first 2000 years (10 010 BC to 8000 BC) included in an attempt to mitigate undesirable spline endpoint effects. At the recent end, some of the models are forced to agree with the *gufm1* model (Jackson *et al.* 2000) in the same way as the CALS10k.1b model (Korte *et al.* 2011) while others employ no further endpoint constraints (see Section 4 and Table 1).

3.2 Inversion procedure

Finding time-dependent field model coefficients from Holocene magnetic data is an example of a nonlinear inverse problem (e.g. Parker 1994; Gubbins 2004). The relation between the observed data $\mathbf{d}$ and model parameters $\mathbf{m}$ may be expressed in vector notation as

$$\mathbf{d} = f(\mathbf{m}) + \mathbf{e}, \tag{3}$$

where f is the non-linear functional relating the data vector and the model vector, and $\mathbf{e}$ is a vector of the differences between the model predictions and the observations. The solution to such an inverse problem involving noisy observations is not-unique (Parker 1994). It is, however, possible to find a suitable solution by minimization of an objective functional $\Phi(\mathbf{m})$ containing two terms, the first $Q(\mathbf{m})$ that corresponds to some measure of the misfit between the observations and the model predictions ($\mathbf{d} - f(\mathbf{m})$), and the second $R(\mathbf{m})$ that measures the complexity of the model (e.g. Shure *et al.* 1982; Gubbins & Bloxham 1985; Gubbins 2004):

$$\Phi(\mathbf{m}) = Q(\mathbf{m}) + R(\mathbf{m}). \tag{4}$$

Archeo- and paleomagnetic data are obtained using a variety of laboratory methods. No standard procedure to obtain uncertainty estimates exists, making it difficult to characterize the data errors in a statistical sense. Residuals between the initial and final data sets and all derived models suggest that the data errors are described better by a Laplacian rather than a Gaussian distribution (Fig. 2). Consequently, the L_1 measure of misfit seems more appropriate than the traditional L_2 measure of misfit. The heavy tails strongly impact the L_2 norm due to the quadratic contribution to the misfit measure. We explore differences among models obtained using both these measures of misfit here. The L_2 measure of misfit is given by

$$Q_2(\mathbf{m}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{d_i - f(\mathbf{m})_i}{\sigma_i}\right]^2} \tag{5}$$

and the L_1 measure of misfit by

$$Q_1(\mathbf{m}) = \frac{\sqrt{2}}{N}\sum_{i=1}^{N}\left|\frac{d_i - f(\mathbf{m})_i}{\sigma_i}\right| \tag{6}$$

where σ_i are the *a priori* estimated errors on the data, and N is number of data. Here, the L_1 norm is implemented via the technique of iteratively reweighted least squares (IRLS; e.g. Schlossmacher 1973; Constable 1988; Farquharson & Oldenburg 1998; Walker & Jackson 2000).

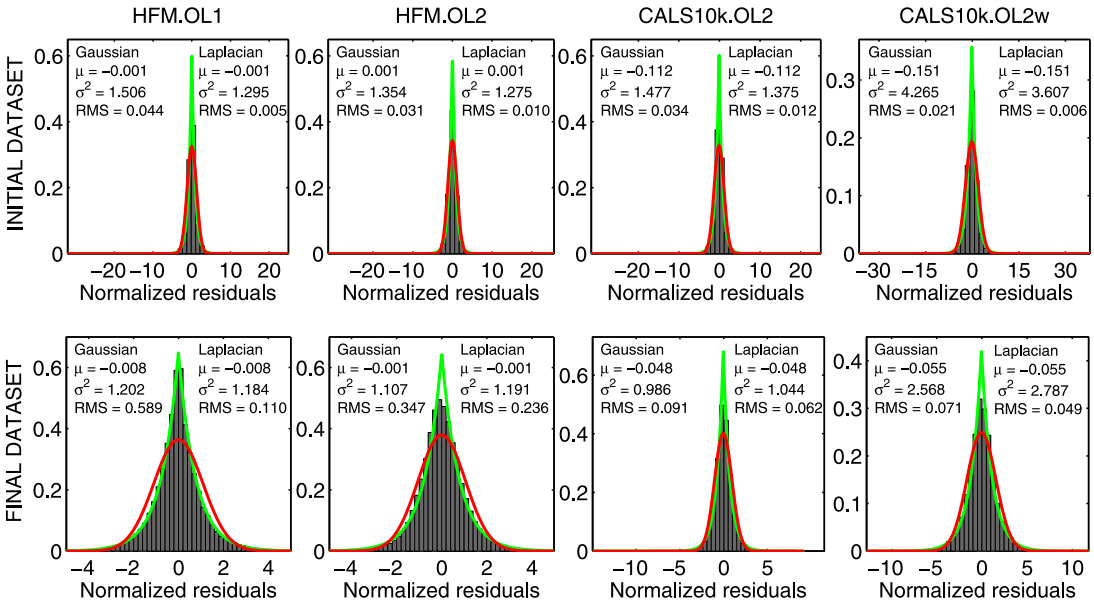

Figure 2. Histograms of normalized residuals to the initial (top) and final (bottom) data sets for the following models: HFM.OL1, HFM.OL2, CALS10k.OL2 and CALS10k.OL2w (from left to right). Details about how these models were built are given in Section 4. Histograms are normalized to unit area and a Gaussian distribution (red line) and a Laplacian distribution (green line) calculated from the residuals and their parameters are listed in each subplot. The rms error values show how well these distributions fit the residuals. (A colour version of this figure is available in the online version.)

Measures of the spatial and temporal complexity of a model were combined in a single regularization term of eq. (4). This can be expanded as

$$R(\mathbf{m}) = \lambda_S R_S(\mathbf{m}) + \lambda_T R_T(\mathbf{m}), \tag{7}$$

where λ_S and λ_T are damping parameters which describe the trade-off between the misfit and the norms measuring the complexity of a model. Two options commonly employed for choosing the damping parameters were used: a trade-off curve (e.g. Gubbins 2004) and the comparison of main field and secular variation power spectra (e.g. Korte *et al.* 2011). Large values of damping parameters put more emphasis on model smoothness, while smaller values promote a better fit to the data. Various choices of regularization norm in time and space have been explored by different authors (e.g. Bloxham & Jackson 1992; Korte & Constable 2005; Jackson *et al.* 2007; Gillet *et al.* 2007; Korte & Holme 2010; Finlay *et al.* 2012). In this study we used the Ohmic dissipation norm at the CMB (Gubbins & Bloxham 1985; Jackson *et al.* 2000) for the spatial regularization:

$$\begin{aligned} R_S(\mathbf{m}) &= \int_{\mathrm{CMB}} \frac{(\nabla \times \mathbf{B})^2}{\mu_0} \mathrm{d}V \\ &= 4\pi \sum_{l=1}^{L} \left(\frac{a}{c}\right)^{2l+4} \frac{(l+1)(2l+1)(2l+3)}{l} \\ &\quad \times \sum_{m=0}^{l} \left[\left(g_l^m\right)^2 + \left(h_l^m\right)^2 \right], \end{aligned} \tag{8}$$

where μ_0 is magnetic permeability and $c = 3485$ km is the radius of the core–mantle boundary. Insufficient data in the early part of the Holocene result in an underestimation of the field intensity at these times if the entire field at the core surface is spatially regularized. To mitigate this effect, we follow Korte *et al.* (2009) and exclude the dipole terms ($l = 1$, $m = 0$, 1) from the spatial regularization and the summation over l starts at 2 instead of 1 in eq. (8).

To measure the temporal complexity of the model we follow Bloxham & Jackson (1992), Jackson *et al.* (2000) and Korte & Constable (2005) and adopt a norm based on the second time derivative of the radial magnetic field integrated over the CMB and over the time span of the model:

$$R_T(\mathbf{m}) = \frac{1}{t_\mathrm{e} - t_\mathrm{s}} \int_{t_\mathrm{s}}^{t_\mathrm{e}} \int_{\mathrm{CMB}} \left(\frac{\partial^2 B_r}{\partial t^2} \right)^2 \mathrm{d}\Omega \mathrm{d}t \tag{9}$$

where t_s is the model's start time, t_e is the model's end time and the integration is over the CMB. This choice is optimal if one wishes to reconstruct a smoothly varying function using a cubic spline basis (e.g. De Boor 2001). The same regularization norms were used in the previous CALSxk models and the pfm9k models recently constructed by Nilsson *et al.* (2014).

We carry out the minimization of the objective function (eq. 4), using an iterative Newton type algorithm (e.g. Gubbins & Bloxham 1985; Tarantola 2005). This algorithm requires knowledge of the Fréchet derivatives at the current model iterate $\mathbf{m}_j$ for the observed field elements

$$\mathbf{A} = \left. \frac{\partial \mathbf{f}(m)}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_j}. \tag{10}$$

These derivatives are obtained by summing the derivatives of the observed field elements with respect to all model parameters. Following previous workers, the elements of the matrix **A** for the non-linear geomagnetic components are derived with help of the linear components using the chain rule.

3.3 Inclusion of relative paleointensity and relative declination observations in field modelling

Two of the three components derived from the sediment records are available only in relative form. Relative declination and

paleointensity are considered to be related to absolute observations through a multiplication by a constant scaling factor (γ_F) for the RPI and addition of a constant offset value (γ_D) for the relative declination for each record:

$$F_i^{\mathrm{rel}}(\theta,\phi,r,t) = \gamma_F \cdot F_i(\theta,\phi,r,t) \tag{11}$$

and

$$D_i^{\mathrm{rel}}(\theta,\phi,r,t) = \gamma_D + D_i(\theta,\phi,r,t). \tag{12}$$

In order to obtain absolute values, following the previous equations, RPI records have to be divided by the scaling factor γ_F while the offset γ_D is subtracted from relative declination records. We consider $N_F = 26$ RPI and $N_D = 67$ relative declination records in this study. We co-estimate the scaling factors for relative intensity and the offset values for declination by extending the model vector to include $N_F + N_D$ additional calibration parameters γ_F and γ_D. Each corresponds to a different sediment record and is solved for during the inversion. The total number of model parameters is now $N = N_{\mathrm{spl}} \cdot L(L+2) + N_F + N_D$.

$$\mathbf{m} = \{g_1^0(t), g_1^1(t), h_1^1(t), \ldots, h_{10}^{10}(t), \gamma_{F_1}, \gamma_{F_2}, \ldots, \gamma_{F_{N_F}}, \gamma_{D_1}, \gamma_{D_2}, \ldots, \gamma_{D_{N_D}}\}. \tag{13}$$

Co-estimation of the relative model parameters requires a modification of the matrix of Fréchet derivatives $\mathbf{A}$ (eq. 10), which now also includes derivatives with respect to γ_F and γ_D as follows:

$$\begin{aligned}\mathbf{A}^{F_i^{\mathrm{rel}}} &= \frac{\partial \mathbf{f}(m)}{\partial \mathbf{m}}\Big|_{\mathbf{m}=\mathbf{m}_j} \\ &= \left(\frac{\partial F_i^{\mathrm{rel}}}{\partial g_1^0(t)} \; \frac{\partial F_i^{\mathrm{rel}}}{\partial g_1^1(t)} \; \frac{\partial F_i^{\mathrm{rel}}}{\partial h_1^1(t)} \cdots \frac{\partial F_i^{\mathrm{rel}}}{\partial \gamma_{F_1}} \; \frac{\partial F_i^{\mathrm{rel}}}{\partial \gamma_{F_2}} \right. \\ &\quad \left. \cdots \frac{\partial F_i^{\mathrm{rel}}}{\partial \gamma_{F_i}} \cdots \frac{\partial F_i^{\mathrm{rel}}}{\partial \gamma_{D_{N_D}}}\right) \\ &= \gamma_{F_i}\left(\frac{\partial F_i}{\partial g_1^0(t)} \; \frac{\partial F_i}{\partial g_1^1(t)} \; \frac{\partial F_i}{\partial h_1^1(t)} \; \ldots \; 0 \; 0 \; \ldots \; \frac{F_i}{\gamma_{F_i}} \; \ldots \; 0\right)\end{aligned} \tag{14}$$

for the relative intensity records, and

$$\begin{aligned}\mathbf{A}^{D_i^{\mathrm{rel}}} &= \frac{\partial \mathbf{f}(m)}{\partial \mathbf{m}}\Big|_{\mathbf{m}=\mathbf{m}_j} \\ &= \left(\frac{\partial D_i^{\mathrm{rel}}}{\partial g_1^0(t)} \; \frac{\partial D_i^{\mathrm{rel}}}{\partial g_1^1(t)} \; \frac{\partial D_i^{\mathrm{rel}}}{\partial h_1^1(t)} \cdots \frac{\partial D_i^{\mathrm{rel}}}{\partial \gamma_{F_1}} \cdots \frac{\partial D_i^{\mathrm{rel}}}{\partial \gamma_{D_1}} \right. \\ &\quad \left. \cdots \frac{\partial D_i^{\mathrm{rel}}}{\partial \gamma_{D_i}} \cdots \frac{\partial D_i^{\mathrm{rel}}}{\partial \gamma_{D_{N_D}}}\right) \\ &= \left(\frac{\partial D_i}{\partial g_1^0(t)} \; \frac{\partial D_i}{\partial g_1^1(t)} \; \frac{\partial D_i}{\partial h_1^1(t)} \; \ldots \; 0 \; \ldots \; 0 \; \ldots \; 1 \; \ldots \; 0\right)\end{aligned} \tag{15}$$

for the relative declination records. Here i represents the ith location which has associated RPI scaling or relative declination offset parameters γ_{F_i} or γ_{D_i}. Initial RPI scaling factors and declination offsets were chosen by comparing the relative records with the starting model predictions. We employed a median estimator for the calibration of RPI records, but mean for the re-orientation of relative declination. RPI and zero-mean declination records are used as input data in the models, without prior calibration to absolute values. We find that for the co-estimation the RPI calibration factor and declination offset value show large changes in the first four to five iterations and the increments in magnitude drop rapidly after the initial adjustment of calibration coefficients. After ten iterations, the values remain almost constant. Examples of some RPI scaling factors and declination offset values for the first 30 iterations for one of the models are given in Fig. 3. Similar convergence is achieved for all models derived using the co-estimation strategy.

In the CALS10k models, rather than co-estimation the RPI calibration factors and declination offsets were determined sequentially by comparison to a starting model, which was not necessarily the starting model used for the inversion. They were iteratively updated by comparison to the derived model using the calibrated values (Korte & Constable 2011). The scheme converged quickly and did not change much after 3 or 4 steps of iterative re-calibration (where each model is obtained by several iteration steps for linearization).

Note that in order to facilitate comparisons with previous studies, the values for the calibration parameters presented in our figures and tables follow the convention of Korte & Constable (2006) and are

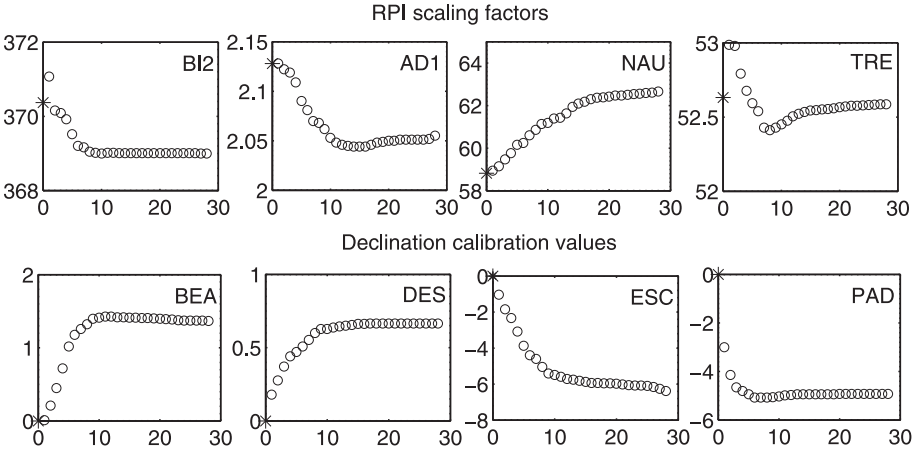

Figure 3. Examples of the convergence of the RPI scaling factors (Γ_F) and relative declination calibration values (Γ_D) for the HFM.OL1 model. Number of iteration is shown on the x-axis. The starting coefficient is shown with a star. Sediment records: BI2 - Lake Biwa, Japan (Hayashida *et al.* 2007); AD1 - Adriatic Sea, Italy (Vigliotti 2006); NAU - Nautajärvi, Finland (Ojala & Saarinen 2002; Snowball *et al.* 2007); TRE - Laguna El Trébol, Argentina (Gogorza *et al.* 2006; Irurzun *et al.* 2006); BEA - Beaufort Sea, Arctic Ocean (Barletta *et al.* 2008); DES - Dead Sea, Israel (Frank *et al.* 2007); ESC - Lake Escondido, Argentina (Gogorza *et al.* 2002, 2004) and PAD - Palmer Deep, Antarctic Pen (Brachfeld *et al.* 2000).

$\Gamma_F = 1/\gamma_F$ for the RPI scaling factor and $\Gamma_D = -\gamma_D$ for the relative declination offset factor, respectively. The alternative definitions given in eqs (14) and (15) lead to simpler algebra in the expressions for the Fréchet derivatives in the co-estimation scheme.

4 RESULTS: HOLOCENE FIELD MODELS

We refer to the new models that use the improved sediment uncertainty estimates and the co-estimation of RPI calibration and declination offset in the inversion as HFM.*, where HFM stands for Holocene Field Model and * describes the choice of regularization norm and misfit. Preliminary versions of the HFM models were obtained by Panovska (2012). In addition, we built three models that use outlier rejection and re-calibration of RPI in the same manner as CALS10k.1b, but now including an evaluation of declination offsets at each iteration. These models are named CALS10k.*. Two of the CALS10k models also use the revised sediment uncertainty estimates. Major differences in modelling strategy between the HFM and the new CALS10k models are the selection of appropriate regularization factors, estimation of RPI scaling and declination offset, and outlier rejection. An overview of the differences in the data and modelling choices is provided in Table 1. The models are available online from the EarthRef Digital Archive (ERDA) at http://earthref.org/ERDA/2102.

All models presented use the Ohmic heating spatial norm and the second time derivative for spatial and temporal regularization. But models HFM.OL1 and HFM.OL1c use an L_1 measure of misfit, while all other models use an L_2 measure of misfit. All models use a constant axial dipole with $g_1^0 = 30\mu$T as starting point for the iterative linearization required in our inversion method, except for model HFM.OL1c, which starts from CALS10k.1b.

CALS10k.OL2w is the only new model that does not use the revised sediment uncertainty estimates of Panovska *et al.* (2012). Instead it employs the previous sediment uncertainty estimates of CALS10k.1b (Korte *et al.* 2011). Comparing CALS10k.OL2 and CALS10k.OL2w provides a direct test of the impact of the revised sediment uncertainty estimates of Panovska *et al.* (2012). The initial RPI calibration and declination orientation are based on a constant axial dipole model with $g_1^0 = 30\mu$T for all models except HFM.OL1c and CALS10k.OL2c. For these models (the 'c' indicates a CALSxk type starting model) the initial calibrations are based on CALS10k.1b and CALS3k.3, respectively.

In the HFM models, the RPI calibration factors and declination offsets were co-estimated during the inversion at each iteration step. Outlier rejection at the level of five standard deviations was performed as a part of each iteration except for the first two iterations (these were performed using the whole data set). Convergence was reached quickly and the 10th iteration was taken as the final model. The five standard deviation level (higher than for the CALSK models) was adopted because the L_1 measure of misfit is less sensitive to outliers. Regularization parameters λ_S and λ_T were determined from trade-off curves. First, λ_S is chosen from a trade-off curve keeping λ_T constant. In the second step, we investigate the trade-off curve of the temporal norm against the misfit, with λ_S fixed to the value found in the first step.

For the CALS10k models median RPI calibration factors and mean declination offsets for each sediment record were determined in comparison to a starting model following the procedure described by Korte & Constable (2006) and as used in recent CALSxk type models (Korte & Constable 2011; Korte *et al.* 2011). Final models were derived in a two-step iterative process: 20 iterations were performed for linearization with each fixed data set, and then outlier rejection at the level of three standard deviations and re-calibration of RPI and declination was performed. This two-step process was repeated four times to obtain the final models. Regularization parameters λ_S and λ_T for the CALS10k models were obtained at each of these four steps by visual comparison of main field and secular variation power spectra with those of the *gufm1* historical field model (Korte *et al.* 2009).

Due to differences in the outlier rejection the final models are constrained by different numbers of data points as listed in Table 1. In general, only about 1 per cent of the data were rejected in the HFM models compared to about 3.5 per cent of data rejected in the CALS10k models. Summaries of root mean square and average of residuals calculated for different data types (sediments and archeo) and components for all models are given in supplementary tables S2 and S3, and values of RPI scaling factors and declination offsets in Tables S4 and S5.

4.1 Global properties of the models

The time-averaged main field and secular variation spatial spectra at the core-mantle boundary (Fig. 4), and evolution of spatial and temporal complexity and normalized root mean square (rms) misfit to the data (Fig. 5) allow an assessment of temporal and spatial resolution in the various models. In Fig. 4, spectra for model *gufm-sat-E3* (Finlay *et al.* 2012), constrained by satellite and observatory observations between 2000.0 and 2010.0 and for *gufm1*, using all available historical and modern observations from 1590 to 1990, are also shown for reference. Model CALS10k.1b is strongly smoothed by bootstrap averaging and consequently shows lowest spatial power and temporal resolution both on average and over the whole time interval together with the highest misfits to the data. The HFM and CALS models show generally similar time-average spectra for the large-scale coefficients. The comparison to modern-day model *gufm-sat-E3* (Finlay *et al.* 2012) and historical model *gufm1* shows the effect of both the time-averaging and the limited resolution of millennial models in quadrupole and higher degrees, which is, however, comparable to the power in the historical field model *gufm1* up to degree and order 4. Models obtained using stronger damping (i.e. HFM.OL1 and HFM.OL1c) as expected present lower spatial complexity and a more rapid drop-off in secular variation spectra above degree 4.

Model pfm9k.1a shows a surprisingly low SV spectrum at low degree that may be because the sediment age scales have been adjusted for optimal internal consistency. This highlights the fact that we still struggle to understand and treat the age uncertainties in the diverse data basis correctly. Model CALS10k.OL2w with less variable and in general higher sediment data weights due to lower uncertainty estimates than used in all other models shows the highest SV spectrum and temporal norm. The revised error estimates used in CALS10k.OL2 and CALS10k.OL2c allow the construction of simpler models with less complexity (smaller temporal norm and SV spectrum) while giving an improved fit (in terms of normalized rms misfit) to the data.

In general, all models display a trend towards higher spatial and temporal complexity (eqs 8 and 9) in the recent 3 millennia when numbers of data are highest, aside from two intervals of somewhat high spatial complexity at about 4000 BC and 7000 BC and rather homogeneous temporal variability (Fig. 5). Note that all CALS10k models show high variability starting about 1500 AD due to the *gufm1* end constraint. Fig. 5 reveals that L1 norm

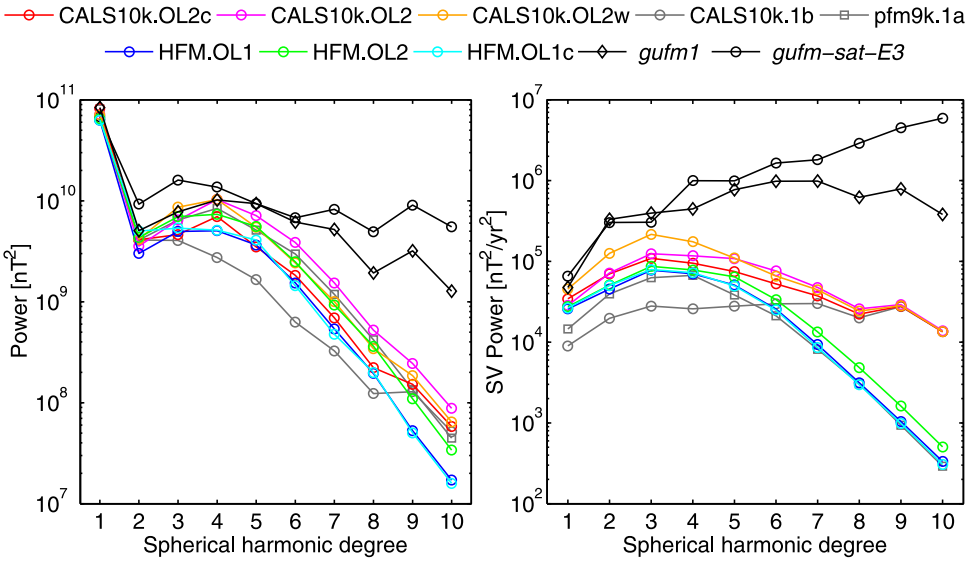

Figure 4. Comparison of the time-averaged spherical harmonic spectra of the geomagnetic field and its secular variation at the CMB for all models. Spectra of the models *gufm1* (L = 14) and *gufm-sat-E3* (L = 24) are plotted only to degree 10 for comparison purposes. (A colour version of this figure is available in the online version.)

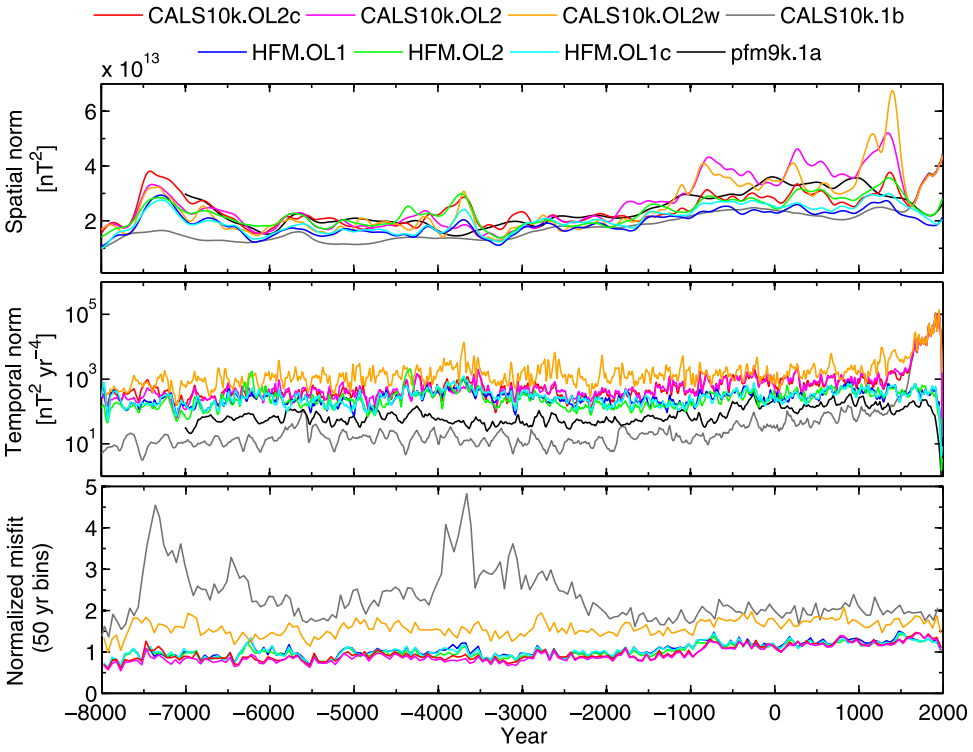

Figure 5. Comparison of the evolution of the spatial (Ohmic dissipation) and temporal norms and normalized misfit for all models. The misfits are calculated to each final data set (after outlier rejection), except for the CALS10k.1b model where the misfit corresponds to the initial data set. (A colour version of this figure is available in the online version.)

models (HFM.OL1 and HFM.OL1c) in particular show lower spatial complexity over the most recent 3000 years, comparable to the strongly smoothed averaged CALS10k.1b model, when the number of archeomagnetic data is highest. The normalized misfits in Fig. 5 are not absolutely comparable. The misfits of all models are calculated to each final data set, except for the CALS10k.1b model where the misfit to the initial data set (before outlier rejection) is shown. Model CALS10k.1b shows very high values of misfits due to a combination of several effects: it is the smoothest model, the underlying data set is more inconsistent as it includes the dense WPA record rejected from all other models shown here, the relative nature of the declination records was not taken into account, and uncertainty estimates for sediment records (used for normalizing the rms misfit here) are lower than in all but the CALS10k.OL2w model. The latter also explains the higher misfit of CALS10k.OL2w compared to CALS10k.OL2 and CALS10k.OL2c. The rms misfit

between the HFM and CALS10k models furthermore differ slightly due to the different levels of outlier rejection. It is obvious that all new models show a more homogeneous fit to the data over time than CALS10k.1b, indicating that the treatment of declination records as relative information and the *a priori* rejection of record WPA led to internally more consistent data sets. All five models using the new sediment data uncertainty estimates (HFM.OL1, HFM.OL1c, HFM.OL2, CALS10k.OL2, CALS10k.OL2c) show slightly higher normalized rms misfits during the most recent 3000 years when high numbers of archeomagnetic data are available. The influence of the uncertainty estimates is such that the archeomagnetic data are given more weight in the modelling. However, these data have not been subject to the same systematic assessment of uncertainty as the sediment records, and inconsistencies in the assigned uncertainties are likely.

4.2 Differences in model parameters

Figs 6 and 7 show a comparison of the evolution of the dipole and quadrupole coefficients for the HFM and CALS10k models, together with CALS10k.1b and pfm9k.1a for reference. The largest dispersion is seen among the g_1^0 coefficients and this reflects the differences in RPI scaling factors discussed in details below. No clear systematic and persistent differences among the models are visible in the equatorial dipole and quadrupole coefficients. The only exception is the g_2^2 coefficient of CALS10k.1b, which suffers from the rather strong influence of the incorrectly oriented Australian records in combination with internally inconsistent South-East-Asian records and the artificially high impact of the West Pacific (WPA) record. Centennial and longer term trends appear rather robustly resolved in all models. The HFM models and the CALS10k models, respectively, agree more closely on details within each group than between the two groups. Stronger deviations of details, however, are mostly seen in the pfm9k.1a model (e.g. g_1^1 at about 800 AD) and CALS10k.OL2w (e.g. g_2^0 over the past 3.5 kyr). Differences in the temporal evolution of the CALS10k.OL2w dipole and quadrupole coefficients compared to all the other models indicates that the weighting of the data plays a more important role than the differences in modelling strategy between the HFM and CALS10k models.

Now we turn to an investigation of the scaling of RPI records and declination orientation in sediments. Figs 8 and 9 show the final RPI scaling factors and declination offsets, respectively, ordered by geographic latitude. The RPI scaling factors are normalized relative to those used in CALS10k.1b as the raw values differ greatly from record to record. Differences of up to 40 per cent can be seen in RPI scaling factors, indicating that the different calibration strategies do not always give consistent results. The largest dispersion among all model scaling factors occurs in the Arctic. One might have expected that robust scaling factors would be obtained in regions with plenty of archeomagnetic data such as Europe, while larger dispersion would be evident in regions devoid of such information. However, this is not clearly the case, as for example records Laguna El Trébol (TRE) and Lake Escondido (ESC) from Argentina find similar Γ_F in all models despite the fact that there are no nearby archeomagnetic data, whereas for example Lago di Mezzano (MEZ), Italy, and

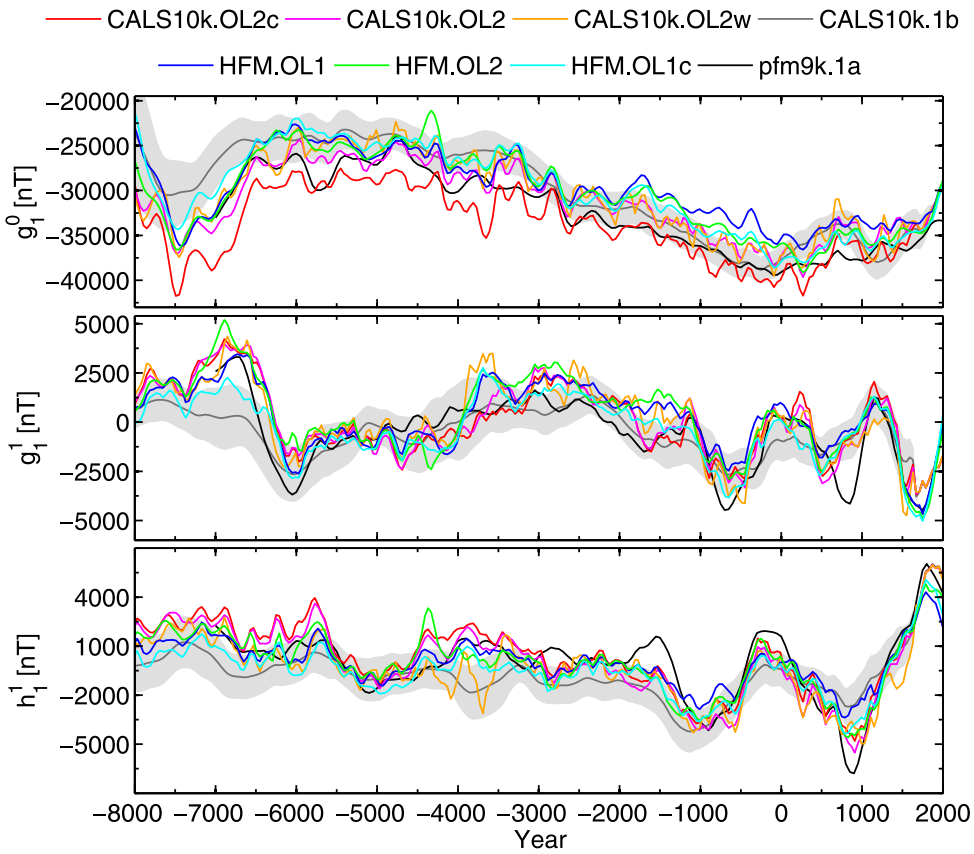

Figure 6. Evolution of dipole coefficients [in units of nT] of all models. The g_1^0 is the axial dipole component parallel to rotation axis, and g_1^1 and h_1^1 are the equatorial dipole terms. These three terms g_1^0, g_1^1 and h_1^1 represent a tilted dipole field. The grey area shows the bootstrap uncertainty estimates for the CALS10k.1b coefficients. Note that some lines overlap. (A colour version of this figure is available in the online version.)

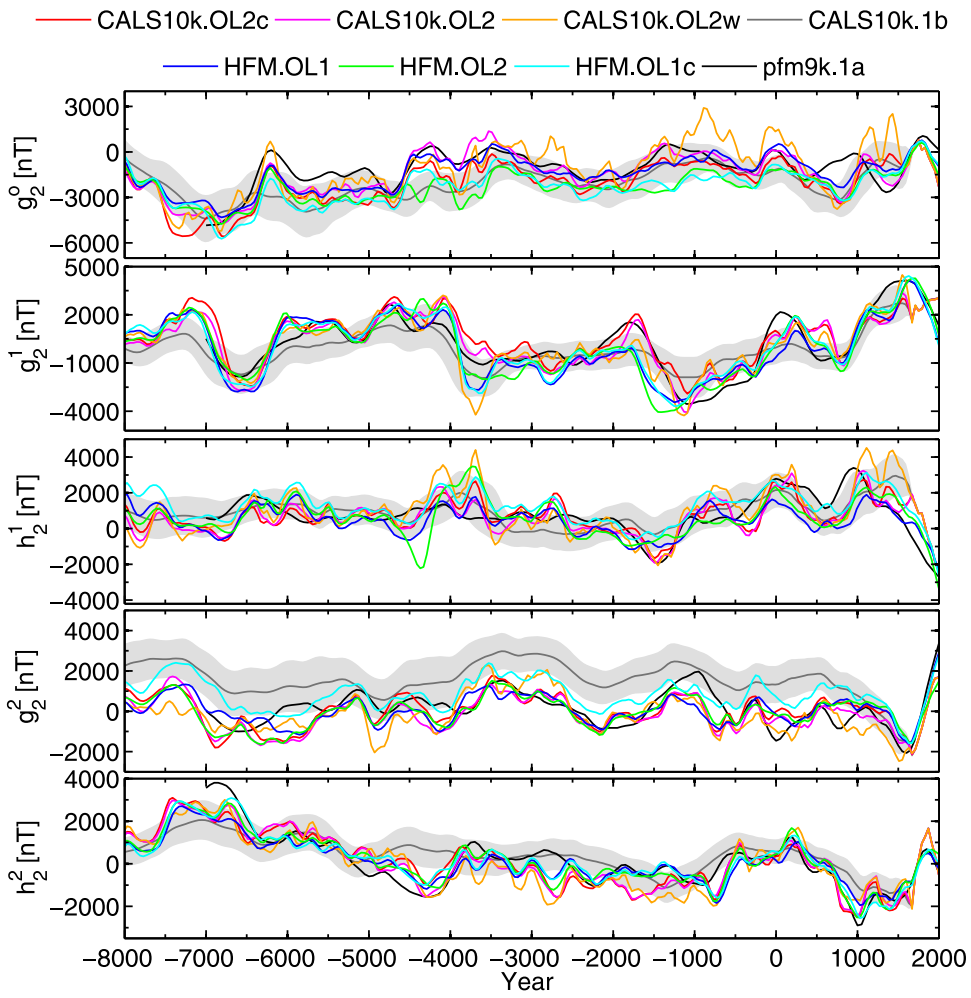

Figure 7. Evolution of quadrupole coefficients [in units of nT] of all models. The grey area shows the bootstrap uncertainty estimates for the CALS10k.1b coefficients. (A colour version of this figure is available in the online version.)

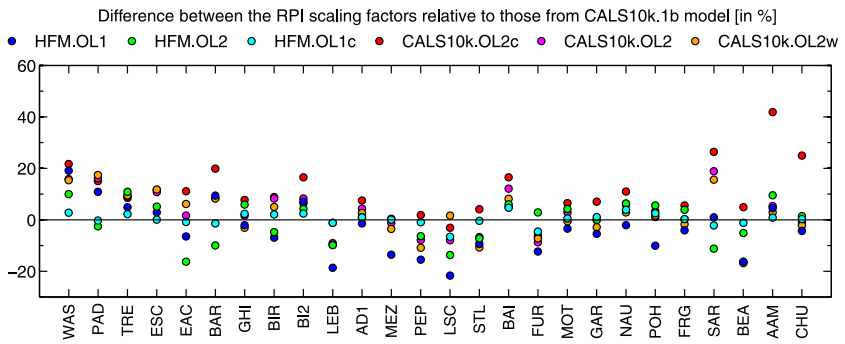

Figure 8. Distribution of the RPI scaling coefficients (Γ_F) relative to those from CALS10k.1b model [in per cent] for each RPI sediment record. The records are ordered south to north from left to right. See supplementary table (Table S1) for full names of the records, locations, coordinates and references. (A colour version of this figure is available in the online version.)

Birkat Ram (BIR), Israel, show large differences in Γ_F despite nearby absolute archeointensity data.

Overall, the HFM models tend to scale the RPIs lower than CALS10k.1b, while the new CALS10k models tend to scale them higher. This is confirmed by Table 2, where globally and temporally averaged DM, virtual axial dipole moment (VADM) values and their standard deviation (SD) as a measure of variability for the models are listed. The mean VADM for the Holocene calculated from all models listed in Table 2 is $8.1 \pm 0.4 \times 10^{22}\mathrm{Am}^2$. While all models show similar morphology of dipole evolution, the absolute values of DM clearly correlate with the strength of the RPI scaling factors and number of archeointensity in the final data set (Table 2). The smallest SD of the RPI scaling factors (2.44) seen for the HFM.OL1c model shows that its calibration records are closest to those of CALS10k.1b which was the starting model for scaling and linearization in this case. CALS10k.OL2c model has the

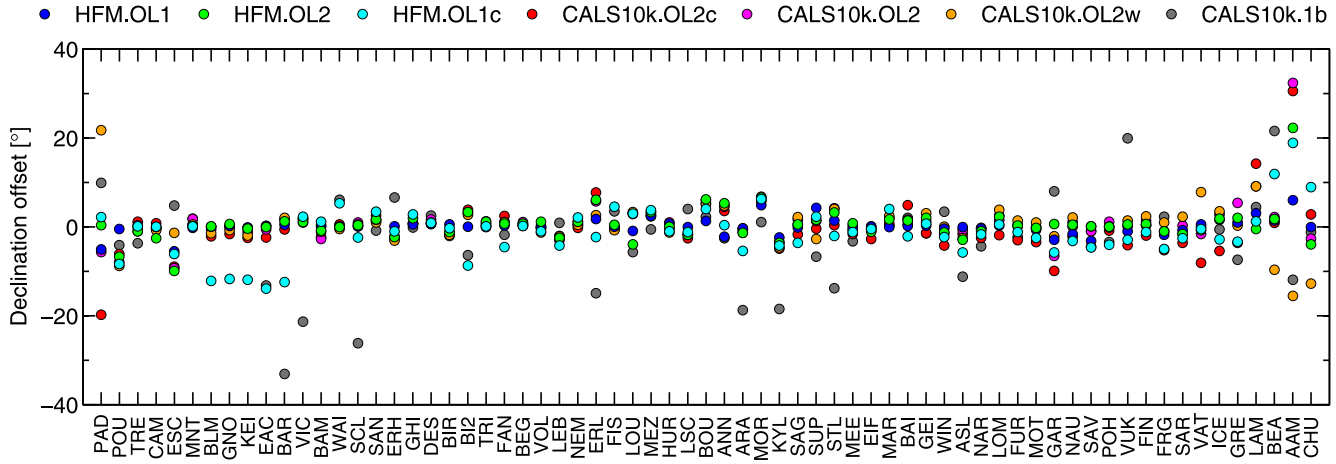

Figure 9. Distribution of the declination offsets from zero mean (Γ_D) for each declination sediment record. The records are ordered south to north from left to right. Declination offsets for the CALS10k.1b are calculated from the model predictions at each sediment location as offsets from the zero mean values. See supplementary table (Table S1) for full names of the records, locations, coordinates and references. (A colour version of this figure is available in the online version.)

Table 2. Table of globally and temporally averaged DM in $\times 10^{22}$ Am^2; VADM and its standard deviation (SD) in $\times 10^{22}$ Am^2; mean and SD of the RPI scaling factors, expressed as percentage difference from the CALS10k.1b model, and relative declination offsets as plotted in Figs 8 and 9; number of archeointensity data (NF_a) in the final data set; and mean F_a residuals in unit of μT and normalized by their individual uncertainty estimates. Models are ordered by decreasing mean VADM value.

Model	VADM		DM	RPI scaling f.		dec. offsets		N F_a	mean F_a res.	
	mean	SD	mean	mean	SD	mean	SD		μT	norm.
CALS10k.OL2c	8.85	0.60	8.72	10.48	10.44	−0.57	5.87	4079	0.15	0.00
pfm9k.1a (9 kyr)	8.45	0.40	8.40							
CALS10k.OL2	8.18	0.43	8.07	2.76	8.76	0.93	4.94	4040	0.85	0.13
CALS10k.OL2w	8.03	0.47	7.91	1.75	8.72	0.40	4.73	4010	1.59	0.26
HFM.OL2	7.90	0.65	7.83	−0.38	7.64	0.55	3.76	4055	1.81	0.26
CALS10k.1b	7.86	0.50	7.81							
HFM.OL1c	7.83	0.70	7.79	0.29	2.44	−1.20	5.42	4039	2.22	0.33
HFM.OL1	7.74	0.41	7.72	−3.42	9.84	0.08	1.79	4017	4.28	0.68
AVERAGE	8.10	0.38	8.03							

highest mean of the RPI scaling factors, but it starts from initially calibrated RPI records with higher amplitude (initial RPI scaling factors for CALS10k.OL2c are on average 10 per cent higher than the CALS10k.OL2 and CALS10k.OL2w initial factors).

We further evaluate our models by looking at the fit to just the absolute intensity information provided by archeomagnetic and volcanic material. All the models show a positive bias of archeointensity residuals (Table 2), that is the predicted intensities are on average lower than the data. The strength of the bias is correlated with the average VADM and DM; the weaker the DM the stronger the bias (see order of the models listed in Table 2). Also, the smallest bias corresponds to the largest number of retained archeointensity data. Assuming that we should expect no bias in the average fit to the absolute intensity information, the models with the smallest archeointensity residual bias resolve the DM and VADM level better. In general, all HFM models show stronger bias than the new CALS10k models. Additional tests showed that the archeointensity uncertainties play a very important role in the RPI scaling and VADM strength. A further HFM test model built with hypothetically smaller archeointensity uncertainties showed increased scaling factors for all RPI records, which in turn produced higher VADM and DM values.

We conducted an additional test to determine how much the iterative RPI scaling depends on the starting model for the first scaling, constructing two more extreme HFM and CALS10k models starting from constant axial dipoles of 20 μT and 40 μT, respectively, instead of 30 μT for the initial RPI scaling and linearization. Both the co-estimation of RPI scaling used in the HFM models and the iterative process as used in the CALS10k models produced RPI scaling factors that were about 30 per cent lower or higher than for the 30 μT starting model, respectively. Globally and temporally averaged VADM results were 20 per cent lower and 12 per cent higher. Apparently there are not enough absolute intensity observations (available only from archeomagnetic and volcanic material) at present to robustly determine the correct RPI scaling. The RPI records for the pfm9k.1a model were determined independently by comparison to archeomagnetic data and an archeomagnetic VADM starting model, and were not iteratively modified further. In the absence of larger amounts of globally better distributed archeomagnetic data it seems that the best strategy is to start from a model based only on archeomagnetic data.

An additional issue regarding RPI scaling mentioned by Nilsson *et al.* (2014) is the fact that despite the best normalization efforts changes in lithology and depositional environment might create

artificial changes in the intensity signal, so that constant RPI scaling factors throughout one record might not always be adequate. Nilsson *et al.* (2014) considered two 'jumps' in the RPI signal in records Lake St.Croix (LSC), USA, and Laguna El Trébol (TRE), Argentina, and scaled these RPI records in two parts. More gradual changes might exist and lead to rather large inconsistencies between sediment RPI records and model predictions. We obviously cannot resolve this issue with current models and underline once more that absolute intensity information from all regions of the world is important to resolve the RPI scaling and dipole strength issues.

4.3 Regional variations

Declination offsets mostly turn out to be more consistent and robust with differences among the models on the order of only a few degrees when starting from an axial dipole (i.e. zero mean declination) starting model (Fig. 9). Note that declination values were not treated as relative in CALS10k.1b. Previously, data included in the CALSxk series had been compared to earlier models (in particular including the historical *gufm1* model for the overlapping period) and in very few cases declination records had been adjusted based on this (Korte & Constable 2003; Korte *et al.* 2005). However, in the case of the Australian record BAR, the record had been retrieved again from original sources for the CALS3k.4 and CALS10k.1b models and used without orientation check and adjustment. This is the origin of the large offset shown by this record in Fig. 9. The same is probably true for some of the other records that show larger offsets in CALS10k.1b, where a check of declination orientation might have been neglected. Larger overall dispersions occur as expected in high latitude regions [Palmer Deep, Antarctic Pen. (PAD), and Beaufort Sea (BEA), Alaskan Margin (AAM) and Chukchi Sea (CHU) in the Arctic Ocean] where large declination values and fast variations naturally occur in proximity to the magnetic poles. Other large dispersions for example at Erlongwan Lake, China (ERL) and Gardar Drift, North Atlantic (GAR) may point to inconsistent data series. In the region of Australia [records Lake Barrine (BAR), Lake Bullenmerri (BLM), Lake Eacham (EAC), Lake Gnotuk (GNO), and Lake Keilambete (KEI)] the initial GAD zero mean offset does not change much in most of the models. Absolute declination information being very sparse for this region we have to assume that this is correct, while being aware that it also might reflect the fact that there is not enough absolute information to change the orientation from the initial assumption. An exception is the HFM.OL1c model, which uses the CALS10k.1b as starting model for initial declination offset and Gauss coefficients. The declination records end up re-oriented to an offset of approximately -12° for all Australian records. In general, we obtain more robust declination offset estimates than RPI scaling factors under the assumption that zero mean is a good starting model. This might also have to do with declination records being more reliable in general than RPI records which despite best normalizing efforts might be influenced by environmental changes. The zero-mean assumption is probably valid for records spanning several millennia. Nevertheless, it seems certainly advisable to check wherever possible against available absolute declination information if the assumption is justified.

5 DISCUSSION

Six models have been constructed, using the different strategies that are outlined above. Comparisons are also made to previously published time-dependent Holocene field model CALS10k.1b (Korte *et al.* 2011) as well as to the recent pfm9k.1a model (Nilsson *et al.* 2014).

Examples of model predictions and their fit to the data are presented in Fig. 10. Time series of directional data with model predictions are shown for locations in central Europe (EIF - Eifel maars, Germany), Hawaii (WAI - Lake Waiau), Australia (GNO - Lake Gnotuk) and China (ERH - Erhai Lake). Nearby archeomagnetic data within 15° radius, which were relocated with an axial dipole correction to the location of the sediment record, are also plotted. All models agree well in central Europe, where both archeomagnetic and sediment data coverage is best. In Hawaii, all models except for CALS10k.1b agree most of the time, with only short time intervals of somewhat larger discrepancies. The WAI declination record was offset by 6° in the CALS10k.1b data set, which shows at times prior to 2000 BC when volcanic absolute declination data become sparse. The inclination data themselves show an offset between volcanic data and the Lake Waiau record in times BC, and the CALS10k.1b model predicts even lower values. We believe that this is an influence of the overweighted and inconsistent WPA record from the West Pacific. In Australia, both HFM.OL1c, which used CALS10k.1b as its starting model and that model itself show a clear offset in declination, which is not compatible with zero mean declination. Inclination is not influenced in this case. Even in China, a region where we suspect influences from some internally inconsistent data sets, the model predictions largely agree rather well, CALS10k.1b again being the exception in inclination. In declination, none of the models predict the long-term trend shown by the sediment paleomagnetic data between 4400 BC and 2000 BC, dismissing this part of the record as incompatible with the rest of the regional and global data.

It is clear that a model cannot be better than the information contained in the underlying data, and we discuss here one further aspect of data selection. Pavón-Carrasco *et al.* (2014b) recently published a 14 kyr model based only on archeomagnetic data and not including any sediment data, arguing that sediment records strongly smooth the secular variation signal and only archeomagnetic and volcanic material can provide high time resolution. This might be true depending on the sedimentation rate and lock-in processes, but one must also keep in mind the limitations imposed by spatial distribution of the available data. We show examples of fits to sediment inclination records from four locations in Europe in Fig. 11. The data come from Eifel maars (EIF), Germany (Stockhausen 1998), Furskogstjärnet (FUR), Sweden (Zillén 2003), Nautajärvi (NAU), Finland (Ojala & Saarinen 2002) and Finnish Lakes (FIN) (Haltia-Hovi *et al.* 2010), and model predictions come from the purely archeomagnetic and volcanic material based SHA.DIF.14k, pfm9k.1a and CALS10k.OL2c. The data are quite consistent and all models agree quite well with features shown by the data for the most recent 4 millennia. In particular, the Finnish Lakes (FIN) record is of no lower resolution than the SHA.DIF.14k model. Between 2000 BC and 6000 BC the data are somewhat inconsistent with respect to age; the models agree well with Eifel Maars (EIF) and Furskogstjärnet (FUR) but not so clearly with Nautajärvi (NAU) and Finnish Lakes (FIN). The very high variability of the SHA.DIF.14k model around 4500 BC is not supported by any of the sediment paleomagnetic records and also appears very fast compared to the younger times, when that model is based on much more data per time interval and region. Around 7000 BC all data series consistently show a clear inclination minimum, which is predicted by pfm9k.1a and CALS10k.OL2c, but not by SHA.DIF.14k. This is because at that time the few available data (only 355 values for the whole interval of 12000 BC to 6000 BC) come from a small area at the western coast

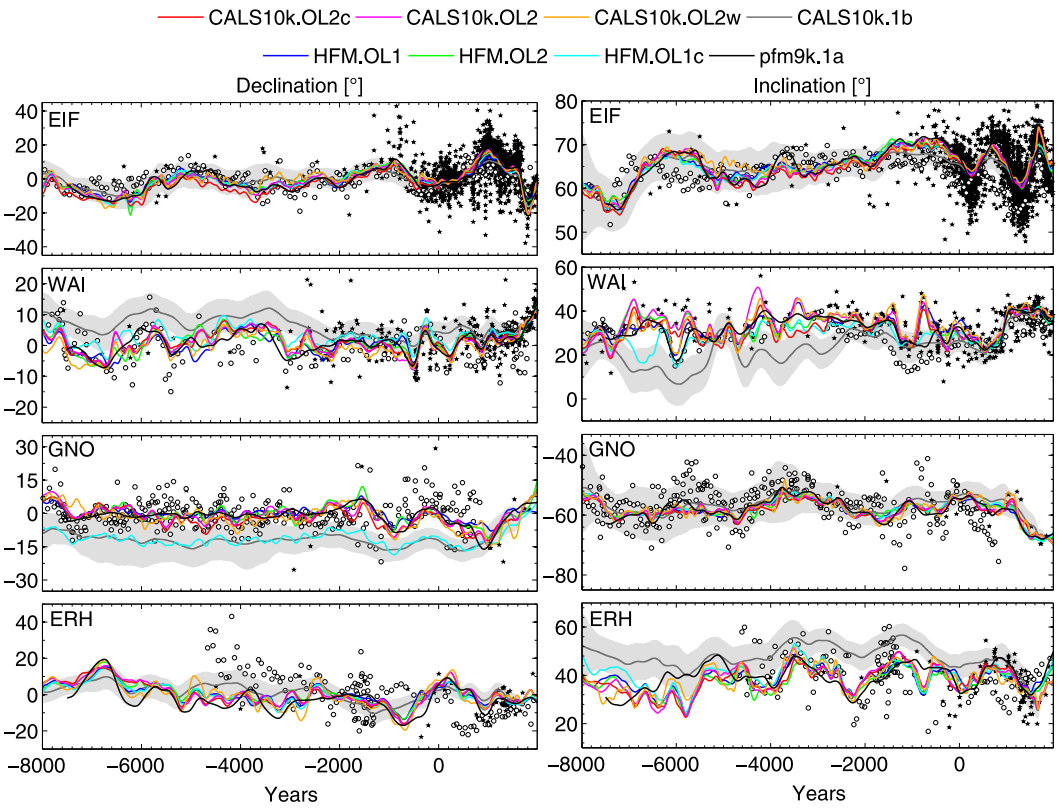

Figure 10. Model predictions of declination (left) and inclination (right) for four sediment paleomagnetic records compared to the sediment data (open circles) and archeomagnetic data (black stars), relocated under the axial dipole assumption. The grey area shows the bootstrap uncertainty estimates for the CALS10k.1b model. Sediment records: EIF - Eifel maars, Germany (Stockhausen 1998); WAI - Lake Waiau, Hawaii (Peng & King 1992); GNO - Lake Gnotuk, Australia (Barton & McElhinny 1981) and ERH - Erhai Lake, China (Hyodo *et al.* 1999). (A colour version of this figure is available in the online version.)

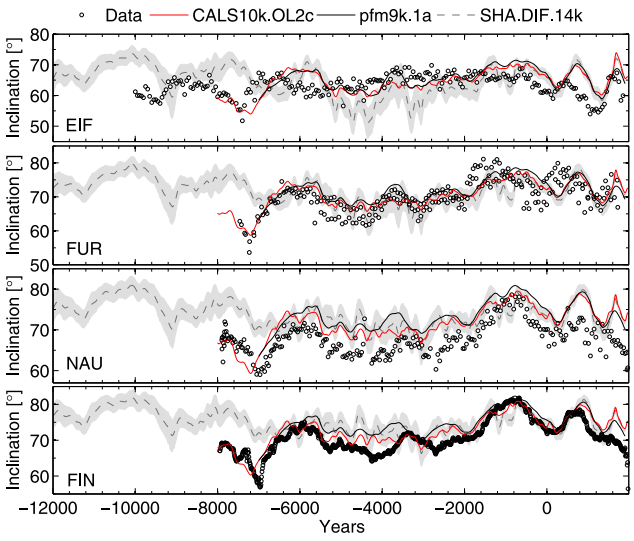

Figure 11. Comparison of sediment inclination data from Europe with global geomagnetic field models. The data (black circles) come from four locations in Europe: EIF - Eifel maars, Germany (Stockhausen 1998); FUR - Furskogstjärnet, Sweden (Zillén 2003); NAU - Nautajärvi, Finland (Ojala & Saarinen 2002); and FIN - Finnish Lakes, Finland (Haltia-Hovi *et al.* 2010). CALS10k.OL2c (red) and pfm9k.1a (black) are models based on sediments and archeomagnetic data together; SHA.DIF.14k (dashed grey) is a model based on archeomagnetic and lava flow data, avoiding the use of sediment paleomagnetic data. The grey area is the 2 standard deviation uncertainty estimates of the SHA.DIF.14k model. (A colour version of this figure is available in the online version.)

of the USA and Hawaii (Pavón-Carrasco *et al.* 2014b). With so few data the SHA.DIF.14k model describes only a time-varying dipole field. All (possibly regional) variation seen in the data are ascribed to dipole variations. Users should be aware that smaller-scale spatial features are not resolved in the earlier half of the SHA.DIF.14k model and dipole tilt predictions might be distorted by northern Atlantic regional field behaviour.

The comparison of time-averaged non-axial dipole (NAD) radial field component when downward-continued to the CMB are presented in Fig. 12 in order to illustrate how much the model differences discussed above influence the CMB structure. For large scales, they mainly agree with each other, giving some confidence in the interpretation of the existence of such structures. Smaller scale details differ more substantially, especially noteworthy is the field behaviour in the Australian/South-East Asian region. CALS10k.1b shows very strong features not present in the other models, which we no longer trust because of the declination orientation problems in the Australian records in addition to the artificially strong influence of the WPA record. The data from this region clearly contain some conflicting information that is interpreted differently by different models. Further high quality studies are needed in this region. Nonetheless, all HFM models compare well with the new CALS10k models except the CALS10k.OL2c, in which the influence of CALS3k.3 model for calibrating relative records is very noticeable. The similarities in structure for the time averages are generally positive NAD radial fields at equatorial and low latitudes and generally negative NAD radial fields at high latitude. The higher magnitude negative NAD field in the Northern polar region in the CALS10k.OL2c model is not present

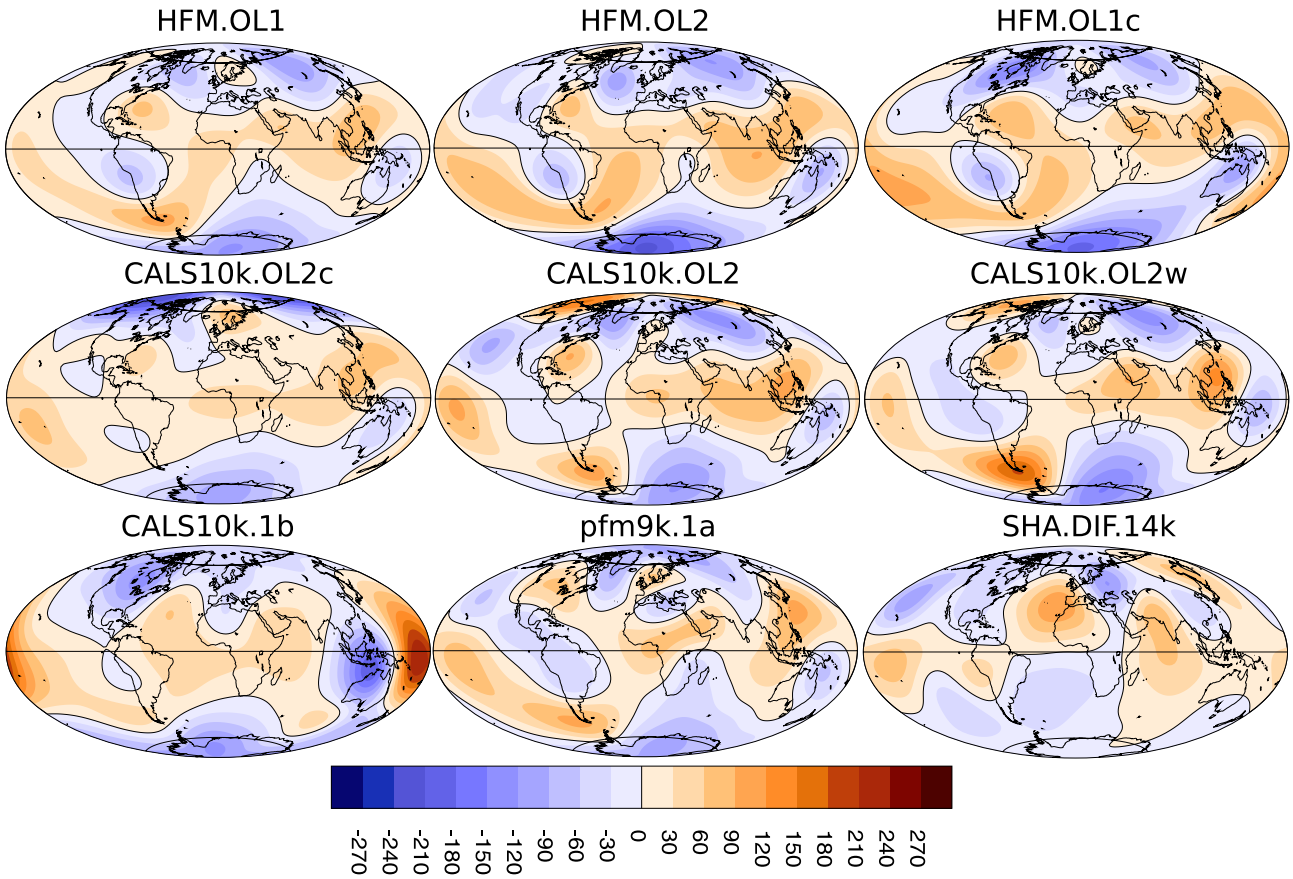

Figure 12. Comparison of the radial component of the time-averaged non-axial dipole (NAD) field at the CMB (in μT) from the models HFM.OL1, HFM.OL2, HFM.OL1c, CALS10k.OL2c, CALS10k.OL2, CALS10k.OL2w, CALS10k.1b, pfm9k.1a and SHA.DIF.14k, for the past 10 kyr, except for the pfm9k.1a model which is averaged over the past 9 kyr. The time-averaged NAD field for the SHA.DIF.14k model averaged over the past 14 kyr shows the same structure with decreased magnitude. Models plotted in the first (HFM) and second row (CALS10k) present the influence of the different modelling techniques. Comparison can be carried out considering different norms (example models HFM.OL1 and HFM.OL2); different data weighting (ex, CALS10k.OL2 and CALS10k.OL2w); different starting models for the Gauss coefficients and relative components (ex, HFM.OL1 and HFM.OL1c) or only different starting calibrations of RPI and rel. declination (ex, CALS10k.OL2 and CALS10k.OL2c). CALS10k.1b, pfm9k.1a and SHA.DIF.14k models are shown for reference. (A colour version of this figure is available in the online version.)

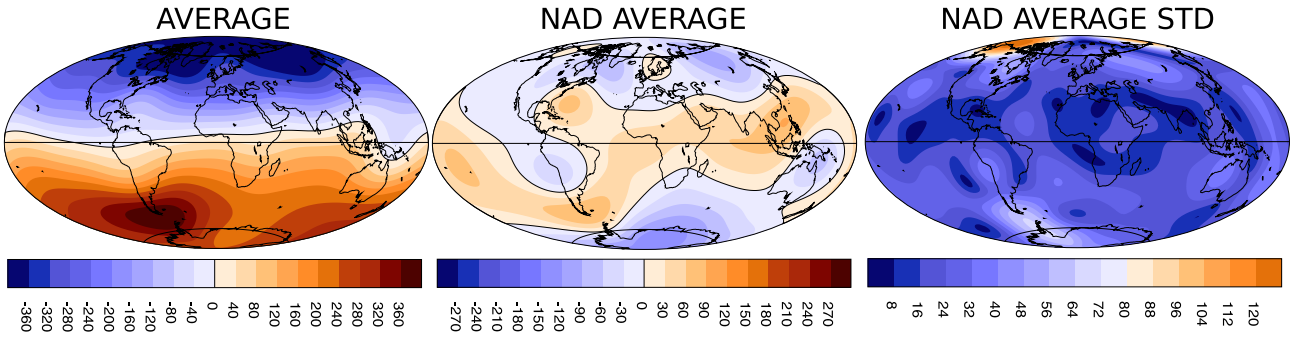

Figure 13. Averages of the six models produced in this study: HFM.OL1, HFM.OL2, HFM.OL1c, CALS10k.OL2c, CALS10k.OL2, and CALS10k.OL2w, made only for comparison purposes. Plots represent the averages of the radial component of the time-averaged field at the CMB (in μT), non-axial dipole (NAD) field average and its standard deviation (NAD AVERAGE STD). Thus, the NAD AVERAGE field is a mean of the six plots presented in the first and second row in Fig. 12. (A colour version of this figure is available in the online version.)

in the other models and contrasts the positive NAD features in CALS10k.OL2 and CALS10k.OL2w in the same region. To determine which time-averaged NAD field features are likely more robust, we averaged the NAD fields of the six new models produced in this study (HFM.OL1, HFM.OL2, HFM.OL1c, CALS10k.OL2, CALS10k.OL2c, and CALS10k.OL2w) in Fig. 13. We also plotted the averaged total field and the standard deviation of the NAD average field. The NAD average (Fig. 13) shows most of the features in the HFM models, and loses the northern polar positive NAD patches present in CALS10k.OL2 and CALS10k.OL2w models. In

general, the NAD average compares well with the reference models pfm9k.1a and CALS10k.1b, except the Australia/Indonesia/West Pacific region. Here, the difference between models with/without use of sediment paleomagnetic records is also evident by comparison of the NAD average with the SHA.DIF.14k model, whose features are not in a good agreement with the NAD structure of the other individual models or with the NAD average. Examples of clear differences are the lack of the negative NAD flux patch in the southern polar region and a strongly positive NAD structure in the western north Atlantic area in the SCHA.DIF.14k model. Greatest variability in the NAD average field over the six models is in the Northern polar region (see Fig. 13, NAD AVERAGE STD). Higher variability is also present in Western Antarctica and Southern South America. Based on the standard deviation plot, NAD fields in the broader equatorial and low latitudes regions agree the best across all the models and they present the most robust features of the different models, preserved in their averages. However, readers should keep in mind that the lowest standard deviation across regularized models may also be associated with data poor regions (*cf.* Fig. 1).

6 CONCLUSIONS

Six new geomagnetic field models approximately covering the Holocene, based on paleomagnetic sediment and archeomagnetic data, have been constructed using a variety of modelling strategies in order to test the robustness and reliability of millennial time scale field models. Previously published models CALS10k.1b and pfm9k.1a were included in the comparisons. Our analyses show that the field models are in general more sensitive to changes in relative weighting of data and the scaling of relative intensity and declination offsets than to differences in modelling methodology.

Surprisingly, one of the largest challenges in reconstructing the past geomagnetic field remains the recovery of the absolute strength of the DM. One might expect that the largest-scale feature should be easiest to resolve, but in fact it depends critically on absolute intensity information. When these are sparse or strongly geographically biased, as indeed is the case for the available archeomagnetic and volcanic data, robust estimates of the DM are a major challenge. Sediment records are indispensable to describe the (regional) field variability at least prior to about 1000 BC, but they require scaling of RPI and orientation adjustment of declination information.

Absolute intensity information are as yet too few to provide robust results regarding RPI scaling factors, independent of the initial starting model. This is particularly true for the co-estimation during the modelling, which seems to underestimate scaling factors and consequently the resulting DM more than the iterative re-scaling used in the CALS10k models. Currently, the best strategy seems to be to perform the initial RPI scaling based on a starting model using as much absolute information as possible, for example as done by Nilsson *et al.* (2014) with the combined VADM and VGP starting model, or by using an archeomagnetic and volcanic data only model as the starting model for the Gauss coefficients and for the estimation of the initial values of the RPI scaling factors and declination offsets. We should also ask whether archeomagnetic intensities might be biased and whether we therefore should expect a biased residual distribution. There could be effects that could produce such a bias. The procedures used to account for the effects of thermoremanent magnetization (TRM) including magnetic alterations, TRM anisotropy and cooling rate corrections, are not considered in a systematic way by all authors. This aggravates the problem of obtaining unbiased archeointensities and reliable uncertainty estimates.

Treating sediment declination records as relative variations with zero mean starting model and iterative adjustment in the inversion works well with the available absolute declination information. However, we have to be aware that orientation adjustments might be biased by the starting GAD assumption in larger areas which lack sufficient archeomagnetic and volcanic declination data. Note that this was not done when deriving the CALS10k.1b model, which as a consequence shows unrealistic variations in the Australian - South-East Asian region. It also included the strongly weighted, but probably inconsistent, West Pacific record (WPA). Additional data or a rigorous assessment of reliable and inconsistent features seen in data series from that region is required to resolve the apparently complex field evolution there.

Data uncertainty estimates are used to weight the data fit during field modelling. They are thus vital in the determination of reliable field structures and variations in field models. We find that internally consistent error estimates for sediment records based on information regarding temporal resolution and sedimentation rate as provided by Panovska *et al.* (2012) help in obtaining consistent normalized rms misfits over the entire modelling period.

The proper treatment of age uncertainties in particular for the stratigraphically correlated sediment records remains a challenge. Nilsson *et al.* (2014) applied a bootstrap optimization intended to reduce inconsistency of age scales across sedimentary records used in their pfm9k.1a model. As a result, this model shows considerably less temporal variability than the models we have constructed. The variability in this model can be seen as a lower limit of temporal change required by the available data, as it is possible that some real variations have been eliminated by the age shifting process. Total least squares solutions that co-estimate age offsets in the inversion still seem infeasible due to the large number of additional free parameters compared to the available robust data information.

With regard to modelling strategy, the choice of regularization parameters in combination with the weighting of the data determines how much spatial and temporal structure the models show. The choice of the regularization factors that trade off a good fit to the data against simple, smooth models is a somewhat subjective task. Between the two different ways of choosing the damping factors, we derived simpler models by finding the knee of the trade-off curves than by using comparisons of main field and secular variation spectra. The amount of spatial and temporal structure shown by these models is no more than required and justified by the available data. The residual distributions suggest that the L_1 norm is a better measure of misfit than L_2. However, we found no striking differences between resulting models when applying the L_1 norm with outlier rejection at 5 standard deviations or the L_2 norm with a more restrictive outlier rejection at 3 standard deviations. However, using the L_2 norm with 5 standard deviation residual rejection produces models with much more variability than with 3 standard deviation rejection. Therefore, 3 standard deviation rejection is advisable when generating models with the L_2 measure of misfit.

In summary, we find that the data quality and the availability of realistic uncertainty estimates play a more important role than the details of field modelling choices in the current generation of millennial timescale field models. Although the number of paleomagnetic data spanning the Holocene is constantly increasing, more data are still needed to address the serious deficiencies of coverage in the Southern hemisphere, lack of data at earlier times, and to overcome ambiguities resulting from regionally inconsistent data.

ACKNOWLEDGEMENTS

S. Panovska and C.G. Constable acknowledge support from the US NSF grant EAR 1246826, and C.G.C. additionally thanks the Alexander von Humboldt Foundation for supporting collaborations with GFZ. S. Panovska gratefully acknowledges support from the Swiss National Science Foundation grant PBEZP2 -142912 and the CHIRP1 project of ETH (CH1-02-08-2) for providing the PhD scholarship for the early stage of this work. M. Korte acknowledges support from DFG grant KO2870/4-1 within SPP1488.

REFERENCES

Ali, M., Oda, H., Hayashida, A., Takemura, K. & Torii, M., 1999. Holocene palaeomagnetic secular variation at Lake Biwa, central Japan, *Geophys. J. Int.,* **136,** 218–228.

Barletta, F., St-Onge, G., Channell, J.E.T., Rochon, A., Polyak, L. & Darby, D., 2008. High–resolution paleomagnetic secular variation and relative paleointensity records from the western Canadian Arctic: implication for Holocene stratigraphy and geomagnetic field behaviour, *Can. J. Earth Sci.,* **45,** 1265–1281.

Barton, C.E. & McElhinny, M.W., 1981. A 10000 years geomagnetic secular variation record from three Australian maars, *Geophys. J. R. astr. Soc.,* **67,** 465–485.

Bloxham, J. & Jackson, A., 1992. Time-dependent mapping of the magnetic field at the core-mantle boundary, *J. geophys. Res.,* **97,** 19 537–19 563.

Brachfeld, S.A. & Banerjee, S.K., 2000. A new high-resolution geomagnetic relative paleointensity record for the North American Holocene: a comparison of sedimentary and absolute intensity data, *J. geophys. Res.,* **105**(B1), 821–834.

Brachfeld, S., Acton, G.D., Guyodo, Y. & Banerjee, S.K., 2000. High-resolution paleomagnetic records from Holocene sediments from the Palmer Deep, Western Antartic Peninsula, *Earth planet. Sci. Lett.,* **181,** 429–441.

Chapman, S. & Bartels, J., 1940. *Geomagnetism,* Clarendon.

Claerbout, J.F. & Muir, F., 1973. Robust modeling with erratic data, *Geophysics,* **38,** 826–844.

Constable, C.G., 1988. Parameter-estimation in non-gaussian noise, *Geophys. J.,* **94,** 131–142.

Constable, C.G., Johnson, C.L. & Lund, S.P., 2000. Global geomagnetic field models for the past 3000 years: transient or permanent flux lobes, *Phil. Trans. R. Soc. Lond. A,* **358,** 991–1008.

De Boor, C., 2001. *A Practical Guide to Splines,* Springer.

Donadini, F., Korte, M. & Constable, C.G., 2009. Geomagnetic field for 0-3 ka: 1. New data sets for global modelling, *Geochem. Geophys. Geosyst.,* **10,** Q06007, doi:10.1029/2008GC002295.

Farquharson, C.G. & Oldenburg, D.W., 1998. Non-linear inversion using general measures of data misfit and model structure, *Geophys. J. Int.,* **134,** 213–227.

Finlay, C.C., Jackson, A., Gillet, N. & Olsen, N., 2012. Core surface magnetic field evolution 2000-2010, *Geophys. J. Int.,* **189,** 761–781.

Frank, U., Nowaczyk, N.R. & Negendank, J.F.W., 2007. Palaeomagnetism of greigite bearing sediments from the Dead Sea, Israel, *Geophys. J. Int.,* **168,** 904–920.

Gillet, N., Jackson, A. & Finlay, C.C., 2007. Maximum entropy regularization of time-dependent geomagnetic field models, *Geophys. J. Int.,* **171,** 1005–1016.

Gogorza, C.S.G., Sinito, A.M., Lirio, J.M., Nuñez, H., Chaparro, M. & Vilas, J.F., 2002. Paleosecular variations 0–19,000 years recorded by sediments from Escondido Lake (Argentina), *Phys. Earth planet. Inter.,* **133,** 35–55.

Gogorza, C.S.G., Lirio, J.M., Nuñez, H., Chaparro, M., Bertorello, H.R. & Sinito, A.M., 2004. Paleointensity studies on Holocene–Pleistocene sediments from Lake Escondido, Argentina, *Phys. Earth planet. Inter.,* **145,** 219–238.

Gogorza, C.S.G., Irurzun, M.A., Chaparro, M.A.E., Lirio, J.M., Nuñez, H., Bercoff, P.G. & Sinito, A.M., 2006. Relative paleointensity of the geomagnetic field over the last 21,000 years BP from sediment cores, Lake El Trébol (Patagonia, Argentina), *Earth planet. Space,* **58,** 1323–1332.

Gubbins, D., 2004. *Time Series Analysis and Inverse Theory for Geophysicists,* Cambridge University Press.

Gubbins, D. & Bloxham, J., 1985. Geomagnetic field analysis - III. Magnetic fields on the core-mantle boundary, *Geophys. J. R. astron. Soc.,* **80,** 695–713.

Haltia-Hovi, E., Nowaczyk, N. & Saarinen, T., 2010. Holocene palaeomagnetic secular variation recorded in multiple lake sediment cores from eastern Finland, *Geophys. J. Int.,* **180,** 609–622.

Hayashida, A., Ali, M., Kuniko, Y., Kitagawa, H., Torii, M. & Takemura, K., 2007. Environmental magnetic record and paleosecular variation data for the last 40 kyrs from the Lake Biwa sediments, Central Japan, *Earth planet. Space,* **59,** 807–814.

Hyodo, M. *et al.*, 1999. A Late Holocene geomagnetic secular variation record from Erhai Lake, southwest China, *Geophys. J. Int.,* **136,** 784–790.

Irurzun, M.A., Gogorza, C.S.G., Chaparro, M.A.E., Lirio, J.M., Nuñez, H., Vilas, J.F. & Sinito, A.M., 2006. Paleosecular variations recorded by Holocene-Pleistocene sediments from Lake El Trébol (Patagonia, Argentina), *Phys. Earth planet. Inter.,* **154,** 1–17.

Jackson, A., Jonkers, A.R.T. & Walker, M.R., 2000. Four centuries of geomagnetic secular variation from historical records, *Phil. Trans. R. Soc. Lond. A,* **358,** 957–990.

Jackson, A., Constable, C.G. & Gillet, N., 2007. Maximum entropy regularization of the geomagnetic core field inverse problem, *Geophys. J. Int.,* **171,** 995–1004.

Korte, M. & Constable, C.G., 2003. Continuous global geomagnetic field models for the past 3000 years, *Phys. Earth planet. Inter.,* **140,** 73–89.

Korte, M. & Constable, C.G., 2005. Continuous geomagnetic field models for the past 7 millennia: 2. CALS7K, *Geochem. Geophys. Geosyst.,* **6,** Q02H16, doi:10.1029/2004GC000801.

Korte, M. & Constable, C.G., 2006. On the use of calibrated relative paleointensity records to improve millennial–scale geomagnetic field models, *Geochem. Geophys. Geosyst.,* **7,** Q09004, doi:10.1029/2006GC001368.

Korte, M. & Constable, C.G., 2011. Improving geomagnetic field reconstructions for 0-3 ka, *Phys. Earth planet. Inter.,* **188,** 247–259.

Korte, M. & Holme, R., 2010. On the persistence of geomagnetic flux lobes in global Holocene field models, *Phys. Earth planet. Inter.,* **182,** 179–186.

Korte, M., Constable, C.G., Donadini, F. & Holme, R., 2011. Reconstructing the Holocene geomagnetic field, *Earth planet. Sci. Lett.,* **312,** 497–505.

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,** Q02H15, doi:10.1029/2004GC000800.

Korte, M., Donadini, F. & Constable, C.G., 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.

Langel, R.A., 1987. *The Main Field,* p. 249–512. Academic Press.

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

Lesur, V., Wardinski, I., Rother, M. & Mandea, M., 2008. GRIMM: the GFZ Reference Internal Magnetic Model based on vector satellite and observatory data, *Geophys. J. Int.,* **173,** 382–394.

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.

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**(1), 229–248.

Ojala, A. & Saarinen, T., 2002. Palaeosecular variation of the Earth's magnetic field during the last 10000 years based on the annually laminated sediment of Lake Nautajärvi, central Finland, *Holocene,* **12**(4), 391–400.

Panovska, S., 2012. Modelling Holocene geomagnetic field evolution. *PhD thesis,* ETH, Zürich.

Panovska, S., Finlay, C.C., Donadini, F. & Hirt, A.M., 2012. Spline analysis of Holocene sediment magnetic records: Uncertainty estimates for field modelling, *J. geophys. Res.,* **117,** B02101, doi:10.1029/2011JB008813.

Parker, R.L., 1994. *Geophysical Inverse Theory,* Princeton University Press.

Pavón-Carrasco, F.J., Osete, M.L., Torta, J.M. & Gaya-Piqué, L.R., 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.

Pavón-Carrasco, F.J., Gómez-Paccard, M., Hervé, G., Osete, M.L. & Chauvin, A., 2014a. Intensity of the geomagnetic field in Europe for the last 3 ka: Influence of data quality on geomagnetic field modeling, *Geochem. Geophys. Geosyst.,* **15,** doi:10.1002/2014GC005311.

Pavón-Carrasco, F.J., Osete, M.L., Torta, J.M. & De Santis, A., 2014b. A geomagnetic field model for the Holocene based on archaeomagnetic and lava flow data, *Earth planet. Sci. Lett.,* **388,** 98–109.

Peng, L. & King, J.W., 1992. A late Quaternary geomagnetic secular variation record from Lake Waiau, Hawaii, and the question of the Pacific nondipole low, *J. geophys. Res.,* **97**(B4), 4407–4424.

Richter, C., Venuti, A., Verosub, K.L. & Wei, K.-Y., 2006. Variations of the geomagnetic field during the Holocene: relative palaeointensity and inclination record from the West Pacific (ODP Hole 1202B), *Phys. Earth planet. Inter.,* **156,** 179–193.

Scales, J.A., Gersztenkorn, A. & Treitel, S., 1988. Fast l_p solution of large, sparse, linear systems: application to seismic travel time tomography, *J. Comput. Phys.,* **75**(2), 314–333.

Schlossmacher, E.J., 1973. An iterative technique for absolute deviations curve fitting, *J. Am. Stat. Assoc.,* **68,** 857–859.

Shure, L., Parker, R.L. & Backus, G.E., 1982. Harmonic splines for geomagnetic modelling, *Phys. Earth planet. Inter.,* **28**(3), 215–229.

Snowball, I., Zillén, L., Ojala, A., Saarinen, T. & Sandgren, P., 2007. FENNOSTACK and FENNORPIS: Varve dated Holocene palaeomagnetic secular variation and relative palaeointensity stacks for Fennoscandia, *Earth planet. Sci. Lett.,* **255,** 106–116.

Stockhausen, H., 1998. Geomagnetic paleosecular variation (0–13 000 yr BP) as recorded in sediments from the three maar lakes from the West Eifel (Germany), *Geophys. J. Int.,* **135,** 898–910.

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.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Soc. Ind. Appl. Math.

Vigliotti, L., 2006. Secular variation record of the Earth's magnetic field in Italy during the Holocene: Constraints for the construction of a master curve, *Geophys. J. Int.,* **165,** 414–429.

Walker, M.R. & Jackson, A., 2000. Robust modelling of the Earth's magnetic field, *Geophys. J. Int.,* **143,** 799–808.

Zillén, L., 2003. *Setting the Holocene Clock Using Varved Lake Lediments in Sweden. PhD thesis,* Lund University, Sweden.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. List of sediment paleomagnetic records. Table provides the codes, location, coordinates and references for each record.
Table S2. Summary of average and RMS of residuals calculated by type of data (sediments and archeo) and by component for all models. Residuals are normalized by the data uncertainty estimates.
Table S3. Summary of average and RMS of residuals calculated by type of data (sediments and archeo) and by component for all models. Values are in absolute units of μT for the intensity and degrees for declination and inclination.
Table S4. Summary of RPI calibration coefficients (Γ_F) for all the models in this study. Using these values RPI records are calibrated in units of μT.
Table S5. Summary of relative declination calibration coefficients (Γ_D) in units of degree (http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggv137/-/DC1).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.