DOI: 10.1029/2008GC002297

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

The Geomagnetic Field for 0-3ka, Part II: A New Series of Time-Varying Global Models

M. Korte[1], F. Donadini[2], C. G. Constable[2]

M. Korte, Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum - GFZ, Telegrafenberg, 14473 Potsdam, Germany. (e-mail: monika@gfz-potsdam.de)

F. Donadini and C. Constable, Institute for Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0225, USA. (e-mail cconstable@ucsd.edu)

[1]Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum - GFZ, Germany

[2]Institute of Geophysics and Planetary Physics, University of California, San Diego, USA

Abstract. Steadily increasing numbers of archeo- and paleomagnetic data for the Holocene have allowed development of temporally continuous global spherical harmonic models of the geomagnetic field extending present and historical global descriptions of magnetic field evolution. The current work uses various subsets of improved data compilations (details in a companion paper by Donadini et al. , 2009, doi: 10.1029/2008GC002295) and minor modifications of standard modeling strategies (using temporally and spatially regularized inversion of the data and cubic spline parametrizations for temporal variations) to produce five models with enhanced spatial and temporal resolution for 0–3 ka. Spurious end effects present in earlier models are eliminated by enforcing large scale agreement with the *gufm1* historical model for AD 1650–1990 and by extending the model range to accommodate data older than 3 ka. Age errors are not considered as a contribution to data uncertainties but are included along with data uncertainties in an investigation of statistical uncertainty estimates for the models using parametric bootstrap resampling techniques. We find common features, but also significant differences among the various models, indicating intrinsic uncertainties in global models based on the currently available Holocene data. Model CALS3k.3 based on all available archeomagnetic and sediment data, without a priori quality selection, currently constitutes the best global representation of the past field. The new models have slightly higher dipole moments than our previous models. Virtual axial dipole moments (VADMs) calculated directly from the data are in good agreement with all corresponding model predictions of

VADMs. These are always higher than the spherical harmonic dipole moment, indicating the limitations of using VADMs as a measure of geomagnetic dipole moments.

1. Introduction

The past evolution of the geomagnetic field is of interest not only to study the underlying processes in the Earth's core, but also for studies where the shielding effect of the geomagnetic field plays a role, e.g. in past cosmogenic nuclide production rates. The change of the dominating dipole contribution is often estimated from archeointensity data by means of virtual axial dipole moment descriptions (VADM) [e.g. *McElhinny and Senanayake*, 1982; *Yang et al.*, 2000; *Genevey et al.*, 2008; *Knudsen et al.*, 2008]. The amount of archeomagnetic and high-resolution lake sediment data covering several millennia, however, also allows for global modeling attempts. First efforts at spherical harmonic models on millennial time-scales were limited to very low degrees [e.g. *Braginskiy and Burlatskaya*, 1979; *Sakai*, 1979; *Ohno and Hamano*, 1993; *Hongre et al.*, 1998]. Starting with a series of snapshot models [*Constable et al.*, 2000], the spherical harmonic descriptions were expanded to higher degrees, with regularization techniques used to suppress spurious structure. Continuous models by the names of CALS3K.1 [*Korte and Constable*, 2003] and CALS3K.2 and CALS7K.2 [*Korte and Constable*, 2005a] have been developed for the past 3 and 7 kyrs, respectively. The name stands for "Continuous model from Archeomagnetic and Lake Sediment data".

The CALSxK models have been widely used for different purposes, like the investigation of core dynamics [*Dumberry and Bloxham*, 2006; *Dumberry and Finlay*, 2007; *Wardinski and Korte*, 2008] or to take into account the shielding effect of the magnetic field for galactic cosmic rays and its influence on the production of cosmogenic isotopes [*Lifton et al.*, 2008; *Selesnick et al.*, 2007; *Usoskin et al.*, 2006, 2008]. However, millennial scale models have significant limitations compared to models from directly measured field data for recent and historical times. The limited spatial and temporal resolution compared to recent field models is inherent to the available data and can only be overcome by a significantly larger number of accurate data with much better distribution over the globe. The number of available data has increased notably since the development of CALS7K.2. *Genevey et al.* [2008] and *Donadini et al.* [2006, 2007] independently improved and significantly enlarged existing collections of archeointensity data, including the important meta-data necessary to evaluate the data quality which are not reported in the compilation used for CALS7K.2 [*Korte et al.*, 2005]. The GEOMAGIA50 intensity database by *Donadini et al.* [2006] has now been expanded [*Donadini et al.*, 2009] and updated and contains all archeomagnetic intensity and directional data that are known to us. Moreover, *Korte and Constable* [2006] demonstrated that suitably calibrated relative intensity records from lake sediments can improve global models. For the time interval since 1000 B.C., we now have 29980 values (11077 declination, 13204 inclination, 5699 intensity) compared to only 19376 (7596 declination, 9464 inclination, 2316 intensity) used for CALS3K.2, thus suggesting models of higher resolution are feasible.

The previous models were developed without consideration of direct magnetic field observations or models thereof, in order to have an independent comparison to assess the reliability of archeo-/paleomagnetic field models. This leads to a discontinuity in the transition from the CALSxK models to recent models or to the 400 year model *gufm1* [*Jackson et al.*, 2000], which is based on recent and historical observations from AD 1590 to 1990 and thus is of higher resolution and reliability. The disagreement is aggravated by edge effects of the splines used as the temporal basis [*Korte and Constable*, 2008]. These problems can be overcome to a certain degree by applying suitable end conditions in the modeling.

A question to be resolved in this work is a systematic difference observed between VADM results and the CALS7K.2 dipole moment. VADMs are simply a geometric transformation of intensity and cannot take into account non-axial-dipole contributions, any higher degree field parts are mapped into the VADM of a single location. Higher degree contributions are assumed to cancel out if individual VADM values are averaged over space and time. However, if most data come from a location with field intensity higher or lower than the average dipole intensity over the considered time interval, then even an averaged VADM is biased high or low, respectively. On the other hand, in spherical harmonic models all the intensity and directional observations are described by the spherical harmonic functions. The directional information also has an influence on how the power is distributed between the spherical harmonic degrees. For data with high uncertainties it is possible that power that in fact belongs to the dipole contribution is mapped into higher degrees, and the regularization, although its influence is stronger on higher degrees, might damp even the dipole moment. Based on investigations of present field data and VADMs from model predictions we proposed that about half of the difference can be explained by a systematic bias of the VADMs due to the geographical data distribution [*Korte and Constable*, 2005b, 2006]. However, *Genevey et al.* [2008] and *Knudsen et al.* [2008] averaged regionally binned VADM results and conclude that the geographical distribution seems to have little influence on the averages. *Valet et al.* [2008] obtained a dipole moment similar to that from averaged VADMs by fitting a large amount of archeomagnetic data by a tilted dipole. The misfit to the data is slightly worse than for CALS7K.2. *Valet et al.* [2008] argue that the difference in misfit is insignificant and the presently available data do not require more complex models. However, their model seems to produce a satisfactory fit to the data mainly in Europe and Asia, where the majority of data come from, and a worse fit in the rest of the world. We suppose that by not allowing a model to include influences of higher field complexity the danger is high that power of such structure is mapped into the dipole, overestimating that field contribution.

Here and in a companion paper by *Donadini et al.* [2009], we consider a large number of recently published data in addition to the previous global data set to develop new regularized spherical harmonic models for the time interval 1000 B.C. to AD 1990. The updates and improvements to the data set are briefly summarized in section 2 and described in detail by *Donadini et al.* [2009]. Archeomagnetic and lake sediment data have different characteristics and an uneven global distribution. In order to investigate the influence of different data types and gain a better understanding of our modeling technique we developed five individual models based on different data sets for the same time interval. We test whether the performance of the models improves when only high quality data, according to pre-assigned data and dating uncertainties, are taken into account. Differences and similarities among the five models illustrate the reliability of certain features of millennial scale global models. All are derived by the same modeling method outlined in section 2 and are presented individually in section 3. Uncertainty estimates for coefficients and model predictions have been obtained by statistical methods described in section 4. Finally, differences among the models are discussed in section 5.

2. Data and Modeling method

Significant improvements to the data set used to reconstruct the past magnetic field have been carried out since our earlier work. The details are given in the accompanying article by *Donadini et al.* [2009], and only a brief summary is given here. A large number of newly published data have been included, both archeomagnetic results and sediment time series, increasing the number of data for the past 3000yrs by 55% compared to our earlier model CALS3K.2. All data have been carefully checked again and some previous errors were corrected. Minimum values for uncertainty estimates assigned for our modeling purposes have been revised, particularly the intensity uncertainty minimum has been increased and a minimum α_{95} is used instead of independent minimum uncertainty estimates for declination and inclination. Relative intensities from sediment cores have been calibrated by a model based on archeomagnetic data or by using archeomagnetic data from nearby locations where available, and have subsequently been used together with the sediment directional records. To gain a better understanding of the influence of the various data, five different datasets have been compiled. Two of them (ARCH3kcst_dat0/1 and CALS3kcst_dat0/1) comprise only data considered to be the most reliable and selected on the basis of age and data uncertainties provided by the authors of the data for the archeomagnetic values and according to regional consistency for sediment data.

The temporally continuous inverse modeling method based on spherical harmonics in space and cubic B-splines for the Gauss coefficients in time was originally described and used for historical field models by *Bloxham and Jackson* [1992] and *Jackson et al.* [2000]. In the millennial scale context the same regularized methodology has been outlined and used by *Korte and Constable* [2003, 2005a] and *Korte and Constable* [2008]. With the approximation of an insulating mantle the time-dependent geomagnetic main field, $\mathbf{B}(t)$ is described as the negative gradient of a scalar potential $V(t)$,

$$\mathbf{B}(t) = -\nabla V(t), \tag{1}$$

which can be expanded as

$$V(r,\theta,\phi,t) = a \sum_{l=1}^{L} \sum_{m=0}^{l} \sum_{k=1}^{K} \left(\frac{a}{r}\right)^{l+1} [g_l^{m,k} \cos(m\phi) + h_l^{m,k} \sin(m\phi)] P_l^m(\cos\theta) M_k(t) \tag{2}$$

where (r, θ, ϕ) are spherical polar coordinates and $a = 6371.2$km is the mean radius of the Earth's surface. The $P_l^m(cos\theta)$ are the Schmidt quasi-normalized associated Legendre functions of degree l and order m. The coefficients $\{g_l^{m,k}, h_l^{m,k}\}$ are related to the standard Gauss coefficients $\{g_l^m, h_l^m\}$ for a single epoch t by a series of cubic B-splines, M,

$$g_l^m(t) = \sum_{k=1}^{K} g_l^{m,k} M_k(t) \tag{3}$$

and the same for $h_l^m(t)$.

We generally follow our modeling strategy from the earlier CALSxK models, where the maximum degree of the spherical harmonics and the knot-point spacing of the splines are chosen to allow for higher resolution than we can expect from the data. The spatial basis is expanded up to spherical harmonic degree and order 10. The number of splines has been increased to provide a knot-point spacing of 10 years instead of the previous value of 55 years, to accommodate the possibility of higher temporal resolution. A physically motivated quadratic norm regularization is used to find the smoothest, simplest model that satisfactorily fits the data. The regularization minimizes a lower bound on Ohmic dissipation [*Gubbins*, 1975] at the core-mantle boundary $(r = c)$, given by

$$\Psi = \frac{4\pi}{t_e - t_s} \int_{t_s}^{t_e} f(B_r) dt \tag{4}$$

with

$$f(B_r) = \sum_{l=1}^{L} \frac{(l+1)(2l+1)(2l+3)}{l} \left(\frac{a}{c}\right)^{2l+3} \sum_{m=0}^{l} \left[(g_l^m)^2 + (h_l^m)^2 \right] \tag{5}$$

for spatial smoothness and also minimizes a norm defined in terms of the second time derivative of the field, i.e. the integral

$$\Phi = \frac{1}{t_e - t_s} \int_{t_s}^{t_e} \oint_{CMB} \left(\partial_t^2 B_r\right)^2 \mathrm{d}\Omega \mathrm{d}t \tag{6}$$

for temporal simplicity over the whole time interval $[t_s, t_e]$. Following *Bloxham and Jackson* [1992] some constant factors have been omitted in the spatial regularization (eq. 5, namely a factor of $a/(\mu_0^2\sigma)$ with μ_0 the magnetic constant and σ the electrical conductivity of the core fluid. Note, that according to *Gubbins* [1975] the involved relation of radii is $a(a/c)^{(2l+3)}$ and not $(a/c)^{(2l+4)}$ as given by *Bloxham and Jackson* [1992]; *Jackson et al.* [2000] and *Korte and Constable* [2003]. The earlier CALSxK models were in fact derived using a radii factor of $(a/c)^{(2l+3)}$ while the spatial norm value in table 1 of *Jackson et al.* [2000] suggests that $(a/c)^{(2l+4)}$ was indeed used for the historical model. Moreover, contrary to the published descriptions not all of the constant factors had been omitted in deriving the earlier CALSxK models. Consequently, none of the norm values reported by *Korte and Constable* [2003] and *Korte and Constable* [2005a] are directly comparable to those presented by *Jackson et al.* [2000] or to those described below. The spatial complexity norm for *gufm1* shown in several figures in section 3 had been recomputed using eq. 4) to permit direct comparison.

The constants used to control the balance between model complexity and misfit to the data are labeled λ for the spatial and τ for the temporal regularization. The resulting objective function to be minimized is

$$(\gamma - \mathbf{fm})^{\mathbf{T}} \mathbf{C_e^{-1}} (\gamma - \mathbf{fm}) + \lambda \mathbf{\Psi} + \tau \mathbf{\Phi}, \tag{7}$$

where $(\gamma - \mathbf{fm})$ is the error vector given by the difference between data γ and the prediction of the model $\mathbf{m}$ and $\mathbf{f}$ is the operator relating the data vector to the model according to eq. 2. $\mathbf{C_e}$ is the data error covariance matrix. Our earlier CALS7K.2 model showed a small bias in intensity residuals and seemed to underestimate the dipole moment slightly. Arguing that a tilted dipole should be well resolved by the available observations and that it is a better smooth field assumption than a zero field, we now exclude the dipole terms from the spatial regularization, i.e. the summation over l starts at degree 2 instead of 1 in eq. 5 in this case. The dipole clearly stands out in terms of power in the spherical harmonic description. On the other hand, the Ohmic dissipation regularization has a stronger effect on higher SH degrees, but also damps the dipole if strong regularization is applied. By excluding the dipole from our spatial regularization norm we try to avoid any damping of the dipole moment and tilt as a by-product of the strong regularization required to suppress unrealistic small-scale structure. In some earlier models the dipole terms were additionally penalized, and increased weight given to intensity data. We did not do this in the current work.

The solution to an inverse problem as given here is non-unique. Particularly with the large and often not well known data errors of the archeo- and paleomagnetic dataset a large range of models will provide acceptable solutions. Choosing the regularization parameters in order to get a preferred solution which might be considered closest to reality is a difficult and somewhat subjective task. We assume that a reasonable solution does not show more spatial and temporal complexity on average than present field models. Even if the field was more complex at times in the past, we cannot expect to resolve such structure with the available data. However, none of our datasets can be fit within the estimated data uncertainties (ignoring age uncertainties) under this assumption. Therefore, we use a comparison of average main field and secular variation power spectra to those of a current field model (the International Geomagnetic Reference Field IGRF for epoch 2000 [e.g. *Maus et al.*, 2005]) and of the time-averaged historical field model *gufm1* as a criterion to choose the regularization factors. The chosen regularization norms result in a damping of power in main field and secular variation that increases for higher SH degrees (i.e. small scale / short term structure) with higher factors of λ and τ, respectively. By simple visual comparison of the resulting spectra to current field spectra we therefore aim for our new models to show a comparable average amount of structure as given by the power in the first three to four degrees and definitely no more power in the higher degrees. The regularization parameters used for each model are given in the following section.

Directional and intensity data are related non-linearly to the Gauss coefficients, so the solution has to be found iteratively from linearized equations. We use a constant axial dipole of $g_1^0 = 30\mu$T as starting model. The strongest variance reduction is achieved in the first two or three iteration steps and convergence is reached quickly. We always chose the 5$^{\text{th}}$ iteration as the final model. An iterative rejection of data outliers was applied, discarding all data lying more than three standard deviations in data uncertainty of predictions from a preliminary model (model version 0 in the following) and building the final model from the new dataset. This rejection at the 99% confidence level is less restrictive than that used in CALS7k, where all data lying more than two standard deviations from a preliminary model were rejected. Another difference from CALS7K.2 is that we no longer map the age uncertainty into a corresponding uncertainty in the magnetic observations. In the previous approach used for CALS7K.2 we used very rough categories for increasing the data uncertainty depending on the age uncertainty, while in fact the influence of the age uncertainty depends strongly on the variability of the field. Therefore we decided to consider the age errors only in the determination of statistical uncertainties for the new models (see section 4). The final error estimates thus being smaller, the new models consequently have larger rms misfits when normalized with the uncertainty estimates than CALS7K.2 or CALS3K.2 even though they have higher spatial and temporal resolution, seen in the main field geomagnetic power and secular variation spectra, and in absolute terms the fit to the data in fact is better.

The recent end of our models has been penalized for agreement with the *gufm1* historical model. This makes the model more reliable in the recent past and also overcomes the spline end effects described by *Korte and Constable* [2008]. Agreement with the *gufm1* model [*Jackson et al.*, 2000], which describes the field based on historical and recent magnetic data, has been implemented as an additional term in the objective function (eq. 7) through a penalty to minimize the difference between the model coefficients for the time span AD 1650 to 1990. We exclude the earliest epochs of *gufm1*, because an increase in the spatial norm of that model with age prior to AD 1650 suggests spurious spline end effects. The axial dipole coefficient is further excluded from this penalty for the time before AD 1840, because this coefficient is extrapolated and not determined by data prior to that time in the *gufm1* model. The factor governing the closeness of the fit to the *gufm1* coefficients, named *gufm const.* in the table of parameters in the following section, is chosen so that a close agreement is given for the low degree coefficients without fitting too closely very short term variations or high degree details that cannot in general be resolved by the archeomagnetic data.

We have no way to penalize the model by a-priori information at the early end. However, we expanded the modeling time span beyond the time of interest. The models in fact start at 2000 B.C., so that any end effects can be assumed to have decayed within the millennium outside the validity range of the models.

3. Five new models

Five new models based on different data sets have been obtained. Comparative information on the number of data and the root mean square (rms) misfit of a constant axial dipole of 30μT and the models to the data is given in table 1 for all models, both before and after the rejection of outliers (versions 0 and 1, respectively). All rms misfit valus are normalized with the data uncertainty estimates used for weighting in the modeling. The value of

30μT for the constant axial dipole, close to the present day value, is rather arbitrary but seemed more reasonable for comparison than rms against zero field. Note that this value only influences the intensity rms, while the directional rms is the same for any strength of axial dipole. Although the number of rejected data is only of order 1.4%, it is obvious that the fit to the data is improved while at the same time the amount of spatial and temporal structure required to fit the data has been decreased by the rejection of outliers. The spatial (λ) and temporal (τ) regularization parameters, corresponding values of the norms (Ψ and Φ, respectively) measuring the amount of structure, and the strength of the end penalty ($gufm$ const.) are listed in table 2. In the following we describe the final models after rejection of outliers.

3.1. ARCH3k.1

The first model, ARCH3k.1, is based only on archeomagnetic data, without any a priori data selection. We expect to achieve a higher spatial and temporal resolution in such a model compared with when sediment data are included. However, this model is certain to be more reliable for the northern than the southern hemisphere, as archeomagnetic data from the southern hemisphere are extremely sparse (only 261 data, and mostly intensity only, compared to 9589 data in the northern hemisphere). The fit to the data and the model norms over time are shown in Fig. 1.

The normalized rms misfit lies between 1.1 and 1.6 for all components (Table 1) and the variance reduction between the fit to a constant dipole and the final model is 68%. The comparison of spatial and temporal norm between ARCH3k.0 (Fig.1a) and ARCH3k.1 (Fig.1b) shows how the rejection of ouliers leads to slightly less variability in these two quantities. This behavior is quite representative for all the models. After some outliers have been removed by the rejection procedure, the spatial complexity of this model remains at roughly the same level throughout, with a significant drop due to decreasing number of data only towards the extra millennium added to accommodate any edge effects. The complexity is about the same as for the early part of *gufm1*. The temporal complexity is in general rather variable in this kind of model. This is partly due to the changes in spatio-temporal data coverage, but likely also reflects complexities in how the geomagnetic field varies.

3.2. ARCH3k_cst.1

A second model is also based only on archeomagnetic data, but in the case of ARCH3k_cst.1 the data set is constrained a priori to only include data fulfilling certain quality requirements [*Donadini et al.*, 2009]). This model may serve to test whether the uncertainty estimates are internally consistent and a better model can be obtained from data with small uncertainty estimates. Misfit and norms of this model over time are shown in Fig. 2. While the spatial complexity in general is less variable and slightly lower than in ARCH3k.1, a clear maximum appears in the model between 0 and AD 500. The lower temporal complexity of ARCH3k_cst.1 up to AD 500 is attributed to the sparsity of data in the constrained data set in this time interval. The selection criteria led to a rejection of a similar number of data at all times, so the effect on the model is stronger at times when the overall amount of data is smaller. The rms misfit is indeed somewhat smaller and the variance reduction is slightly higher (74%) for the constrained data set, confirming a higher internal consistency of the selected data.

3.3. SED3k.1

A model based only on sedimentary data was developed for comparison and named SED3k.1. We expect lower resolution, but a globally more homogeneous model due to the more evenly spaced data distribution. The ARCH3k.1 model has been used for calibration of the relative intensity records. Fig. 3 shows the characteristics of this model.

According to the generally lower quality of lake sediment data, the normalized misfit is clearly higher on average and the spatial complexity is more variable than in the models based on archeomagnetic data only. The variance reduction reaches only 41%. Times of minimum or maximum complexity are different from those of the archeomagnetic models, but because there is no decrease in amount of data there also is no drop in complexity at the earliest epochs. The temporal complexity is more variable than in the archeomagnetic models. Clearly, and not surprisingly, SED3k.1 is significantly different from ARCH3k.1 and ARCH3k_cst.1.

3.4. CALS3k.3

The fourth model is the one most directly comparable to the earlier CALSxK models and therefore has been named CALS3k.3 (with CALS3k.3.0 as its version prior to outlier rejection). It is based on all available data and presents the compromise between good global data coverage provided only by the sediment records and maximum possible resolution achievable from the supposedly higher quality archeomagnetic data. The sediment intensity values have again been calibrated by ARCH3k.1. Characteristics are shown in Fig. 4.

The variance reduction in CALS3k.3 is better than in SED3k.1 (50%) but the average misfit in all components also is relatively high. The spatial complexity lies between that of the purely archeomagnetic and sediment data only models, but shows strong influences from the sediment data. There is no drop in complexity in the earliest millennium, but rather strong, short-term variations in model roughness occur in the AD time span. This interval in CALS3k.3 is also characterized by relatively high temporal complexity. A comparison with Fig. 3 indicates that most of this influence seems to come from the sediment data.

3.5. CALS3k_cst.1

Finally, we developed a model based on the constrained archeomagnetic data set and a selection of lake sediment data considered to be the most reliable, CALS3k_cst.1. The selection of lake sediment data was not straightforward, because uncertainty estimates on direction or relative intensity are rarely published with the data and mostly fixed values have been assumed for all these records. The relative intensity records in this case have been calibrated by comparison to nearby archeomagnetic data where possible, and by ARCH3k.1 in the other cases [see *Donadini et al.*, 2009]. Note that the difference between calibration by ARCH3k.1 or ARCH3k_cst.1 is insignificant. The selection procedure is given by *Donadini et al.* [2009]. Model misfit and norms over time are shown in Fig. 5. The misfit of this model is comparable to ARCH3k_cst.1, and the variance reduction reaches 63%. Note, however, that a large number of lake sediment data have been rejected by the constraining procedure, and the total number of data used for this model is smaller than that of SED3k.1, the model based purely on lake sediments.

3.6. Brief comparison of model characteristics

Comparing Figs. 1 to 5 shows that for all models the fit to the data is rather uniform over time. The fact that the rms misfit of individual components are in general roughly the same and no significant systematic biases are seen in the residuals (see Table 6 of *Donadini et al.* [2009]) suggests a reasonable relative weighting among all the data. Apparent systematic differences, like a slightly better fit to declination than inclination in the archeomagnetic data only models and vice versa in SED3k.1 might be an indication that the error estimates used for weighting are not yet optimal. The resolution of all the models as represented by the spatial norm is similar over large parts of the time interval (Fig. 6a) and consistently higher than for the previous CALS7K.2 model (re-calculated according to eq. 4). The spatial roughness of all models is about the same as for *gufm1* for the earliest century. A slightly higher amount of spatial structure in the archaeomagnetic models up to AD 1700 might be justified by the relative sparsity of historical data. Apart from declination, only 51 inclination values and no intensity had been available for *gufm1* up to that time. The increase in the *gufm1* spatial norm prior to AD 1650 likely has to be attributed to an end effect in the model. However, note that the models which include lake sediment data show a similar variation in required spatial structure at that time. For the recent two millennia SED3k.1 and CALS3k.3 have slightly less structure on average than the other models, with exceptions between AD 1000 and 1500. The significant changes in spatial complexity require a high temporal complexity in these models during that time interval. It is difficult to understand the influence of specific data on the resulting model and the source of this feature in the models is still under investigation. Strong differences among all models complexities occur between 0 and AD 500, while the fit to the data does not differ significantly from the average in any of the cases. Here, we see a clear difference beteween the models based on unconstrained or constrained data, the latter showing more spatial structure. Perhaps in this time interval the constrained data actually are more internally consistent, allowing more detailed structure to be resolved reliably.

A maximum in temporal complexity (Fig. 6b) around 1650 can be linked to the *gufm1* penalty taking effect. The reason could be a certain degree of incompatibility between the historical model and the archeomagnetic data. It seems understandable that this effect is stronger for models that include sediment data, where the penalty had to be stronger to achieve a similar degree of agreement with *gufm1* (see Table 2) and due to the fact that data from the top of sediment cores might be more disturbed or are lacking, but it is not clear why this effect is strongest in CALS3k_cst.1.

4. Estimation of model uncertainties

We investigated the uncertainties caused by the combined effects of poor data distribution and large data and age uncertainties using two statistical sampling methods and a combination thereof. For each of the three final models ARCH3k.1, SED3k.1 and CALS3k.3 a large number of additional models were created by simulating statistical variations of the underlying datasets ARCH3k_dat1, SED3k_dat1 and CALS3k.3_dat1, using each of the statistical methods with fixed modeling parameters. Uncertainties in the final model coefficients were then determined as standard deviation of the coefficients from 2000 models for each final model and each method. The number of models necessary to reach a stable value for the standard deviation varies with coefficients and time. Figure 7 demonstrates for a few example coefficients that the number of 2000 models is enough to reach convergence.

In the first test, which we call the magnetic values and age bootstrap (MA), for each of the 2000 bootstrap samples we simulate the same number of data as used in creating the final model. The simulated data at each location are generated by independent sampling from two normal distributions: one is centered on the value of the magnetic element with a standard deviation corresponding to the uncertainty estimate assigned for our modeling purposes and the other is centered on the age estimate, and uses its respective standard error. For the sediment data, the age resampling was constrained to shift whole records in time, rather than individual ages, in order not to mix up older and younger results as determined by the stratigraphy. In this case we simply used a representative value for the standard error in the age for each sediment record. Variations in the range of the data uncertainties were in this case carried out with fixed values within time intervals of the length by which the record was shifted. Directions and intensities were not treated separately for the time variation.

In the second test, which we call the spatial and temporal distribution bootstrap (ST), we performed bootstraps on the datasets. For the archeomagnetic data the number of data locations was fixed at the original value of 6337 and values were picked uniformly randomly distributed from the dataset. In contrast to our earlier exercise on CALS7K.2 [*Korte and Constable*, 2008] we kept the number of sediment records fixed (61 records) and included or excluded whole time series. Again the complete vector was used. We expect this change in strategy to produce higher, but also more realistic uncertainty estimates because only eliminating a couple of arbitrary points and amplifying the influence of some others from a somewhat correlated time series does not simulate the uncertainty caused if a whole interval (or even the whole time series) is influenced by orientation or normalization problems. Therefore the number of data, both overall and the distribution between the components, varies somewhat among these bootstrap models.

The results from both methods differ somewhat, varying over time and space. These differences are more pronounced when using datasets

SED3k_dat1 and CALS3k.3_dat1, with somewhat higher uncertainties resulting from the ST bootstrap, where whole sediment records are ignored or are considered more than once. In general, however, the uncertainties obtained from both methods are of the same order of magnitude. A combination of sequentially applying MA and ST resampling to a dataset takes into account the influence of uncertainties in both magnetic elements and ages as well as the unsatisfactory data distribution. We adopted this combination of MA and ST (MAST) to produce 2000 statistically variable models from each dataset (ARCH3k_dat1, SED3k_dat1 and CALS3k.3_dat1) as our preferred method to derive model uncertainties and all uncertainties presented for the three final models in the following are based on it. All uncertainty estimates are small for the time span AD 1650 to 1990, where the *gufm1* penalty dominates the fit to the data. Note, however, that all uncertainty estimates will necessarily be unsatisfactory in large regions devoid of data, where bias from the complete absence of data cannot be estimated or varied. Moreover, the statistical variation was carried out with fixed regularization parameters. Slightly different parameters will also result in acceptable models and the uncertainty estimates for the individual models might be somewhat optimistic.

The uncertainties in the coefficients obtained by the MAST method are given in the electronic model files provided as supplemental material. Error estimates for model predictions of the individual components can be obtained by applying error propagation rules, if we assume that the errors in the

Gauss coefficients are uncorrelated. This certainly is not strictly true, but also not unreasonable for our purpose, given how poorly determined the true uncertainties on data and dating are in the first place. Details of the error propagation are given in the appendix.

The quality of the propagated error estimates can be checked by a comparison to the standard deviation resulting from averaging the model predictions of all the models produced with the statistical method. The errors in general are roughly of the same order of magnitude. The error propagation from the coefficient uncertainties tends to give errors that are larger than those estimated directly from model predictions in areas with plenty of data (e.g. Europe) and smaller ones in areas with sparse data coverage (e.g. the whole southern hemisphere in ARCH3k.1). Two examples are shown in Fig. 8. In areas with plenty of data, a good fit to different variations of these values can obviously be obtained by rather different combinations of coefficients, leading to similar model predictions with small standard deviations, but larger standard deviations when averaged coefficients are considered. In all cases, the uncertainty estimates become very small for the times where the models are penalized by *gufm1*. Note some small differences among the three models in Europe during this time, which are due to the fact that agreement in small-scale features and fast variations of the historical model is traded off against fit to the archeomagnetic and sediment data.

Averaging model predictions might give a more realistic estimate of the uncertainties, but in order to make the model and prediction code publicly available the error propagation from the coefficients is more practical and not unreasonably different. This notion is supported by the fact that the average predictions from all MAST models (left panels) in general show somewhat less short-period temporal variation than the models based on our original datasets (right panels), so that the predictions from our models occasionally lie rather at the borders of the MAST component prediction error estimates. This can for example be seen when comparing ARCH3k.1 declination around AD 900 or CALS3k.3 intensity between AD 650 and 900 in Figs. 8c and d. Note that CALS3k.3 and SED3k.1 in general agree within the error estimates in both examples, whereas ARCH3k.1 is significantly different for several time intervals in the south Atlantic, particularly in inclination. There are no data constraints for ARCH3k.1 in this region. The fact that the regularization has not smoothed out this strong variation suggests that it is ascribed to dipole or similar large-scale field changes by ARCH3k.1, which is not compatible with the southern hemisphere sediment data. In summary we can say that the uncertainty estimates as given by the coefficient uncertainties and available from the supplemental material tend to be pessimistic in regions well covered by data, optimistic in regions devoid of data, and most realistic in regions with medium data coverage, where the uncertainties based directly on model predictions are largest.

A general idea about the average size and distribution of field prediction uncertainties is given by Fig. 9, where temporal averages of the standard deviations of the field components at the Earth's surface over the 3kyrs are shown. Keep in mind that some temporal changes of the uncertainties exist, as demonstrated by the examples in Fig. 8. The uncertainties are largest in the sediment only model (SED3k.1) and smallest in the archeomagnetic only model (ARCH3k.1), according to the data characteristics. Rather small errors in the southern hemisphere of ARCH3k.1 compared to the other models result from the fact that there basically are no data to vary by our statistical method, rather than an absolute higher accuracy of the model in this region. As expected, uncertainties in general are smallest in the areas best covered by data and their surroundings.

5. Discussion of differences among models

5.1. Spatial and temporal resolution

The average main field and secular variation spectra, shown in Fig. 10, were used as a criterion to choose the regularization factors, so it is not surprising that the spectra of all our models look similar. The SED3k.1 spectra fall off faster with higher spherical harmonic degrees, reflecting the lower resolution obtainable from the sediment data only. In all our models the distribution of main field power within degrees 2 to 5 is significantly different from the present field, which shows more power in octupole than in quadrupole and higher degrees. The archeomagnetic only models show less or equal power in the octupole, and all models have the highest power (after the dipole) in degree 4. The differences among our new models suggest that this might be a consequence of some broad incompatibilities among the data, mapping power from large-scale into smaller-scale structure depending on the data selection, rather than from a significantly different average field structure, but even the constrained data models CALS3k_cst.1 and ARCH3k_cst.1 show this characteristic. The spatial resolution drops significantly for all models beyond SH degree 5, the temporal one as given by the secular variation spectrum even lacks power starting from degree 4. Note that all our models show somewhat more secular variation power in the dipole than the recent field models do, and somewhat less even in quadrupole and octupole. We could not find any combination of parameters in the modeling which provided a stronger contrast between dipole and higher degree secular variation power while maintaining a reasonable spatial power distribution.

5.2. Dipole moment and dipole tilt

Large-scale features like the dipole contribution should be the most robust feature of millennial scale models. Nevertheless significant differences are seen even for dipole moment (Fig. 11) and dipole tilt (Fig. 12).

The dipole moments of the new models are somewhat higher than for CALS7K.2 and mostly lie between that prediction and archeointensity VADM estimates, even for models based only on archeomagnetic data. Moreover, VADMs calculated as averages from model predictions with the distribution of the underlying intensity data, respectively, agree closely with the real data VADMs averaged in the same way for all five models. Fig. 11b shows the example for CALS3k.3 with 500 year average VADMs shown every 250 years. Some of these average VADMs, particularly the values between AD 500 and 1000, are slightly lower than the ones shown in Fig. 11a, because the calibrated lake sediment intensity data used for CALS3k.3 have been taken into account in calculating these data VADMs.

For most of the time the different dipole moment predictions in Fig. 11a agree within the uncertainty estimates. From 900 B.C. to 250 B.C. and AD 250 to AD 750 the models show larger differences with ARCH3k.1 and ARCH3k_cst.1 predicting a larger dipole moment, whereas between AD 1000 and AD 1500 all the models that considered sediment data show higher values and a fast variation of the dipole moment.

The dipole tilt shows some clear differences between the models only from archeomagnetic data and the models containing sediment record information. Two explanations seem possible. On the one hand, regional field variations might have leaked into the dipole in ARCH3k.1 and ARCH3k_cst.1 due to the void of data in the southern hemisphere. Indeed these two models show significant differences to the other models in the southern hemisphere (see Fig. 8). On the other hand, the often rather inconsistent sediment records might smooth out some strong, large-scale variations. The maximum dipole tilt in ARCH3k.1 does not exceed today's values and it does not seem obvious why the tilt should have been significantly smaller than today over all of the past few millennia. However, the comparatively strong dipole moment throughout the studied time interval might play a role. The previous CALS7K.2 model suggests that low dipole tilt might be related to a strong dipole moment even though no rigorous correlation was found [*Korte and Mandea*, 2008]. This agrees with paleomagnetic observations over 0-5 Ma for which large virtual geomagnetic pole (VGP) dispersion is compatible with lower VADMs [*Love*, 2000]. Moreover, the dipole tilt of the model which considers only consistent sediment data, CALS3k_cst.1, agrees with the other models showing a lower dipole tilt. The uncertainty of the dipole tilt is large and lies nearly in the order of some of the strongest differences (Fig. 12b). The averaged model from our statistical MAST uncertainty estimate approach shows less pronounced variations for the archeomagnetic model and slightly better agreement with the CALS3k.3 predictions.

The longitudinal change of the dipole axis agrees reasonably well for all models from 1000 B.C. to 100 B.C. and AD 500 to AD 1000, but significant changes are seen among nearly all models between those time intervals. Note that the axes predicted by ARCH3k.1 and the old model CALS7K.2 reach nearly the same longitudes in AD 350, but through movements in the opposite direction. Contrary to all the other models, SED3k.1 shows westward movement of the axis after AD 1000 and until it is constrained by the *gufm1* model.

5.3. Regional differences between models

The five models show some regional discrepancies and these are illustrated by the temporal averages of various field components in Fig. 13. The radial magnetic field component, B_r, is shown for the core-mantle boundary (CMB). The same component, but with the axial dipole contribution removed (B_rNAD) is given both at the CMB and Earth's surface. Inclination anomaly, declination and field intensity are displayed for Earth's surface.

B_r is quite similar for all models over the northern hemisphere but shows significant differences between the models excluding and including sediment data, practically the only source of southern hemisphere data. B_rNAD, not surprisingly, reveals much clearer differences in many regions. The most significant difference is a negative flux patch over the South-East Asian region which is present only, but consistently in all the models including sediment data. Another region of negative flux in the southern hemisphere, contrary to the dipole field direction, is present in all averaged models close to the location of the present field South Atlantic Anomaly, perhaps indicating the longevity of this feature. However, interestingly this feature is least strong in the models with southern hemisphere data coverage, as can also clearly be seen in the figures of field intensity.

The directional averages of the models also differ significantly, but models based on similar global distributions of data clearly show some similar, robust features. Some further insight into regional differences of the models is given by *Donadini et al.* [2009] and a more comprehensive study on regional differences and regional fit to the data is in preparation.

6. Conclusions

We have presented five new models describing the geomagnetic field behavior over the past 3 millennia. The models are based on significantly different datasets in order to gain a better understanding of the reliability and the limitations of global spherical harmonic models based on the presently available archeo- and paleomagnetic data with very inhomogeneous global distribution and large uncertainties. The overall number of data has been increased by about 55% compared to our earlier 3kyr model CALS3K.2 by newly available archeomagnetic data (20% increase), sediment directional records (53% increase) and the inclusion of calibrated relative sediment intensity records. We have aimed at highest possible spatial and temporal resolution of the models by choosing the regularization factors in comparison with recent field spectra. The average main field power of the new models is of the same order as the historical field for SH degrees up to 5 while the average secular variation power is comparable up to SH degree 4. Moreover, we have applied statistical techniques to estimate the effects of the uneven data distribution as well as data and dating uncertainties in terms of error bars for the model coefficients and predictions.

Our results show that the distribution of power among the low degree spherical harmonic coefficients cannot be completely resolved by the presently available data. This is even true for the largest-scale features like the dipole contribution. Nevertheless VADM estimates from the intensity data and from corresponding model predictions now agree in all cases, while all VADMs are systematically higher than the dipole moments of the models. This is in good agreement with our studies of the difference between the CALS7K.2 dipole moment and archeointensity VADMs [*Korte and Constable*, 2005b] and confirms a systematic bias of VADMs compared to the SH dipole moment. Models based only on archeomagnetic data, which come nearly exclusively from the northern hemisphere, predict stronger dipole tilts than models including sediment data. The archeomagnetic data are supposed to be more accurate. Nevertheless, we suspect that the lower dipole tilts are more reliable because all the different models that include southern hemisphere data give rather consistent predictions and even the model where the lake sediments have been constrained to the most consistent data (CALS3k_cst.1) gives this result. This would mean, however, that the very recent strong dipole tilt of more than 10° is rather exceptional, perhaps related to the recent decrease in dipole moment.

All the models predict a surprisingly high dipole moment for the time interval AD 1590 to 1840. It is higher than the prediction from CALS7K.2 and the linear result obtained by *Gubbins et al.* [2006] for that time interval by estimating the axial dipole strength from the CALS7K.2 archeointensity data and directional information from the *gufm1* model. A recent study by *Finlay* [2008], who included the same archeointensity data in new field models based on historical data also suggest nearly constant axial dipole strength as maximum likelihood solution for that time interval under linearity constraints. The predictions from our new models in contrast are close to the extrapolation used in *gufm1*, although that contribution had been removed from penalizing the departure from the historical model between 1650 and 1840. Part of the discrepancies might come from differences in the used datasets. Further work to look in detail at the transition between directly measured data and archeomagnetic results seems advisable.

The different models presented here could be useful for different purposes. In general we still consider models with the maximum amount of data most appropriate. The comparisons between ARCH3k.1 and ARCH3k_cst.1 or CALS3k.3 and CALS3k_cst.1 have shown that a selection of what is considered high quality data based on the data error estimates much of the time has rather small effects on the resulting models (see also *Donadini et al.* [2009]). In particular, it is not obvious that the version of the model based on the constrained dataset is really more reliable. ARCH3k.1 probably has a slightly higher resolution and may be more reliable for parts of the northern hemisphere than the models including sediment records, but it can only be recommended for regional work using model predictions for that hemisphere. For global studies and investigations using model coefficients the CALS3k.3 model is our preferred choice. Except for the dipole strength, which indeed seems to be slightly underestimated by CALS7K.2, this new, more detailed 3kyr version and the previous 7kyr model are rather similar in their large-scale features, which is reassuring for the longer time span.

Acknowledgments. This study was supported by NSF grant EAR 0537986 and FD acknowledges additional support from the Academy of Finland.We very much appreciate the efforts of Roman Leonhardt and two anonymous reviewers whose pertinent and detailed comments led to significant improvements on the original manuscript.

References

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

Braginskiy, S. I., and S. P. Burlatskaya (1979), Spherical analysis of the geomagnetic field based on archaeomagnetic data, *Izv. Earth Phys.*, *15*, 891–895.

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

Donadini, F., K. Korhonen, P. Riisager, and L. J. Pesonen (2006), Database for Holoecene geomagnetic intensity information, *EOS Trans. Am. Geophys. Soc.*, *87(14)*, 92–93.

Donadini, F., P. Riisager, K. Korhonen, K. Kahma, L. J. Pesonen, and I. Snowball (2007), Holocene geomagnetic paleointensities: A blind test of absolute paleointensity techniques and materials, *Phys. Earth Planet. Interiors*, *161*, 19–35.

Donadini, F., M. Korte, and C. Constable (2009), The geomagnetic field for 0-3ka, part I: New data sets for revised global models, *Geochem. Geophys. Geosys.*, *submitted*, this issue.

Dumberry, M., and J. Bloxham (2006), Azimuthal flows in the Earth's core and changes in length of day at millennial timescales, *Geophys. J. Int.*, *165*, 32–46.

Dumberry, M., and C. C. Finlay (2007), Eastward and westward drift of the Earth's magnetic field for the last three millennia, *Earth Planet. Sci. Lett.*, *254*, 146–157.

Finlay, C. C. (2008), Historical variations of the geomagnetic axial dipole, *Phys. Earth Planet. Inter.*, *170*, 1–14.

Genevey, A., Y. Gallet, C. Constable, M. Korte, and G. Hulot (2008), ArcheoInt: An upgraded compilation of geomagnetic field intensity data for the past ten millennia and its application to the recovery of the past dipole moment, *Geochem. Geophys. Geosys.*, *9,Q04038*, doi:10.1029/2007GC001881.

Gubbins, D. (1975), Can the Earth's magnetic field be sustained by core oscillations?, *Geophys. Res. Lett.*, *2*, 409–412.

Gubbins, D., A. L. Jones, and C. C. Finlay (2006), Fall in Earth's magnetic field is erratic, *Science*, *312*, 900–902.

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

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

Knudsen, M. F., P. Riisager, F. Donadini, I. Snowball, R. Muscheler, K. Korhonen, L. J. Pesonen, and B. H. Jacobsen (2008), Variations in the geomagnetic dipole moment during the Holocene and the past 50 kyr, *Earth*

Planet. Sci. Lett., *272*, 319–329.

Korte, M., and C. G. Constable (2003), Continuous global geomagnetic field models for the past 3000 years, *Phys. Earth Planet. Interiors*, *140*, 73–89.

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

Korte, M., and C. G. Constable (2005b), The geomagnetic dipole moment over the last 7000 years - new results from a global model, *Earth Planet. Sci. Lett.*, *236*, 348–358.

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

Korte, M., and C. G. Constable (2008), Spatial and temporal resolution of millennial scale geomagnetic field models, *J. Adv. Space Res.*, *41*, 57–69.

Korte, M., and M. Mandea (2008), Magnetic poles and dipole tilt variation over the past decades to millennia, *Earth Planets Space*, *60*, 937–948.

Korte, M., A. Genevey, C. G. Constable, U. Frank, and E. Schnepp (2005), Continuous geomagnetic field models for the past 7 millennia: 1. A new global data compilation., *Geochem., Geophys., Geosys.*, *6, Q02H15*, doi:10.1029/2004GC000800.

Lifton, N., D. F. Smart, and M. A. Shea (2008), Scaling time-integrated in situ cosmogenic nuclide production rates using a continuous geomagnetic model, *Earth Planet. Sci. Lett.*, *268*, 190–201.

Love, J. J. (2000), On the anisotropy of secular variation deduced from paleomagnetic volcanic data, *J. Geophys. Res.*, *105*, 5799–5816.

Maus, S., S. Macmillan, T. Chernova, S. Choi, D. Dater, V. Golovkov, V. Lesur, F. Lowes, H. Lühr, W. Mai, S. McLean, N. Olsen, M. Rother, T. Sabaka, A. Thomson, and T. Zvereva (2005), The 10th generation Internations Geomagnetic Reference Field, *Geophys. J. Int.*, *161*, 561–565.

McElhinny, M. W., and W. E. Senanayake (1982), Variations in the geomagnetic dipole: I. The past 50 000 years, *J. Geomag. Geoelectr.*, *34*, 39–51.

Ohno, M., and Y. Hamano (1993), Spherical harmonic analysis of paleomagnetic secular variation curves, *Central Core of the Earth*, *3*, 205–212.

Sakai, H. (1979), Spherical harmonic analysis of the geomagnetic field during the last 2,000 years, *Rock Mag. Paleogeophys.*, *7*, 8–15.

Selesnick, R. S., M. D. Looper, and R. A. Mewaldt (2007), A theoretical model of the inner proton radiation belt, *Space Weather*, *5, S04003*, doi:10.1029/2006SW000275.

Usoskin, I., M. Korte, and G. Kovaltsov (2008), Role of centennial geomagnetic changes in local atmospheric ionization, *Geophys. Res. Lett.*, *35, L05811*, doi:10.1029/2007GL033040.

Usoskin, I. G., S. K. Solanki, and M. Korte (2006), Solar activity reconstructed over the last 7000 years: the influence of geomagnetic field changes, *Geophys. Res. Lett.*, *33, L08103*, doi:10.1029/2006GL025921.

Valet, J.-P., E. Herrero-Bervera, J.-L. LeMoüel, and G. Plenier (2008), Secular variation of the geomagnetic dipole during the past 2000 years, *Geochem. Geophys. Geosys.*, *9, Q01008*, doi:10.1029/2007GC001728.

Wardinski, I., and M. Korte (2008), The evolution of the core-surface flow and changes in the length of day over the last seven thousand years, *J. Geophys. Res.*, *113, B05101*, doi:10.1029/2007JB005024.

Yang, S., H. Odah, and J. Shaw (2000), Variations in the geomagnetic dipole moment over the last 12000 years, *Geophys. J. Int.*, *140*, 158–162.

Appendix

In order to obtain error estimates for model predictions of the individual components we assume that the errors in the Gauss coefficients are uncorrelated. The general rule for error propagation gives the error ΔA for a value $A(x_i)$ depending on $i = 1, .., N$ variables x_i with the errors Δx_i as

$$(\Delta A)^2 = \sum_{i=1}^{N} \left(\frac{\partial A}{\partial x_i} \right)^2 (\Delta x_i)^2. \tag{8}$$

In the linear case

$$A = \sum_{i=1}^{N} c_i x_i, \tag{9}$$

with factors c_i assumed to be error-free, the error propagation rule simply turns into

$$(\Delta A)^2 = \sum_{i=1}^{N} (c_i \Delta x_i)^2. \tag{10}$$

The geocentric magnetic field components B_r, B_θ and B_ϕ are given by

$$B_r = \sum_{l=1}^{l_{max}} \sum_{m=0}^{l} (l+1) \left(\frac{R_E}{r} \right)^{l+2} [g_l^m \cos(m\phi) + h_l^m \sin(m\phi)] P_l^m(\theta) \tag{11}$$

$$B_\theta = - \sum_{l=1}^{l_{max}} \sum_{m=0}^{l} \left(\frac{R_E}{r} \right)^{l+2} [g_l^m \cos(m\phi) + h_l^m \sin(m\phi)] \frac{dP_l^m(\theta)}{d\theta} \tag{12}$$

$$B_\phi = \frac{1}{\sin\theta}\sum_{l=1}^{l_{max}}\sum_{m=0}^{l} m\left(\frac{R_E}{r}\right)^{l+2}[g_l^m \sin(m\phi) - h_l^m \cos(m\phi)]P_l^m(\theta). \quad (13)$$

Geodetic north (X), east (Y) and vertical (Z) component are then obtained as

$$X = -B_\theta \cos\psi - B_r \sin\psi, \quad (14)$$

$$Y = B_\phi \quad (15)$$

$$Z = B_\theta \sin\psi - B_r \cos\psi \quad (16)$$

with

$$\sin\psi = \sin\alpha \sin\theta - \cos\alpha \cos\theta, \quad (17)$$

where α is the geodetic (geographic) latitude. For all these steps error propagation according to eq. 9 is applied. The non-linear components intensity (F), declination (D) and inclination (I) are given by

$$F = \sqrt{X^2 + Y^2 + Z^2}, \quad (18)$$

$$D = \arctan\left(\frac{Y}{X}\right) \quad (19)$$

$$I = \arctan\left(\frac{Z}{H}\right), \quad (20)$$

with horizontal intensity

$$H = \sqrt{X^2 + Y^2}. \quad (21)$$

Using eq. 8 the errors are

$$(\Delta F)^2 = \frac{1}{F^2}\left[(X\Delta X)^2 + (Y\Delta Y)^2 + (Z\Delta Z)^2\right], \quad (22)$$

$$(\Delta H)^2 = \frac{1}{H^2}\left[(X\Delta X)^2 + (Y\Delta Y)^2\right], \quad (23)$$

$$(\Delta D)^2 = \left(\frac{1}{1+(\frac{Y}{X})^2}\right)^2 \left[\left(\frac{\Delta Y}{X}\right)^2 + \left(\frac{Y\Delta X}{X^2}\right)^2\right] \tag{24}$$

$$(\Delta I)^2 = \left(\frac{1}{1+(\frac{Z}{H})^2}\right)^2 \left[\left(\frac{\Delta Z}{H}\right)^2 + \left(\frac{Z\Delta H}{H^2}\right)^2\right]. \tag{25}$$

Table 1. Number of data and misfit for time interval 1000 B.C. to AD 1990. rms_i is the normalized root mean square misfit to a constant axial dipole ($30\mu T$) and rms_f to final model. Version 0 and 1 are before and after rejection of outliers, respectively.

Data type	N	rms_i	rms_f	N	rms_i	rms_f
		ARCH3k.0			ARCH3k.1	
All data	9605	2.66	1.62	9483	2.54	1.40
Inclination	4174	2.69	1.81	4129	2.52	1.59
Declination	2761	2.17	1.27	2715	2.11	1.13
Intensity	2670	3.04	1.62	2639	2.95	1.49
		ARCH3k_cst.0			ARCH3k_cst.1	
All data	6211	2.57	1.46	6122	2.45	1.23
Inclination	2969	2.62	1.69	2929	2.45	1.42
Declination	1942	2.17	1.14	1911	2.10	1.00
Intensity	1300	2.96	1.33	1282	2.90	1.23
		SED3k.0			SED3k.1	
All data	20375	2.87	2.08	20090	2.49	1.98
Inclination	9030	2.01	1.77	8919	1.90	1.63
Declination	8316	3.22	2.58	8174	2.91	2.22
Intensity	3029	2.78	1.44	2997	3.71	1.84
		CALS3k.3.0			CALS3k.3	
All data	29980	2.80	2.09	29585	2.62	1.83
Inclination	13204	2.25	1.87	13055	2.13	1.71
Declination	11077	2.99	2.39	10892	2.71	2.07
Intensity	5699	3.48	1.94	5638	3.38	1.78
		CALS3k_cst.0			CALS3k_cst.1	
All data	19908	2.23	1.40	19687	2.15	1.27
Inclination	10415	2.00	1.53	10335	1.93	1.42
Declination	6149	1.96	1.19	6039	1.85	1.11
Intensity	3344	3.19	1.30	3313	3.14	1.22

Table 2. Parameters, spatial (Ψ) and temporal (Φ) norms for time interval 1000 B.C. to AD 1990 for all models.

Model	$\lambda(\mathrm{nT}^{-2})$	$\Psi(\mathrm{nT}^2)$	$\tau(\mathrm{nT}^{-2}\mathrm{yr}^4)$	$\Phi(\mathrm{nT}^2\mathrm{yr}^{-4})$	*gufm* const.
ARCH3k.0	3×10^{-14}	177×10^{11}	1×10^{-3}	234	5×10^{-3}
ARCH3k.1	3×10^{-14}	170×10^{11}	1×10^{-3}	203	5×10^{-3}
ARCH3k_cst.0	2×10^{-14}	169×10^{11}	1×10^{-3}	145	5×10^{-3}
ARCH3k_cst.1	2×10^{-14}	164×10^{11}	1×10^{-3}	125	5×10^{-3}
SED3k.0	2×10^{-13}	165×10^{11}	2×10^{-3}	260	1×10^{-2}
SED3k.1	2×10^{-13}	162×10^{11}	2×10^{-3}	202	1×10^{-2}
CALS3k.3.0	2×10^{-13}	176×10^{11}	2×10^{-3}	304	1×10^{-2}
CALS3k.3	2×10^{-13}	172×10^{11}	2×10^{-3}	243	1×10^{-2}
CALS3k_cst.0	8×10^{-14}	177×10^{11}	8×10^{-4}	374	1×10^{-2}
CALS3k_cst.1	8×10^{-14}	170×10^{11}	8×10^{-4}	337	1×10^{-2}

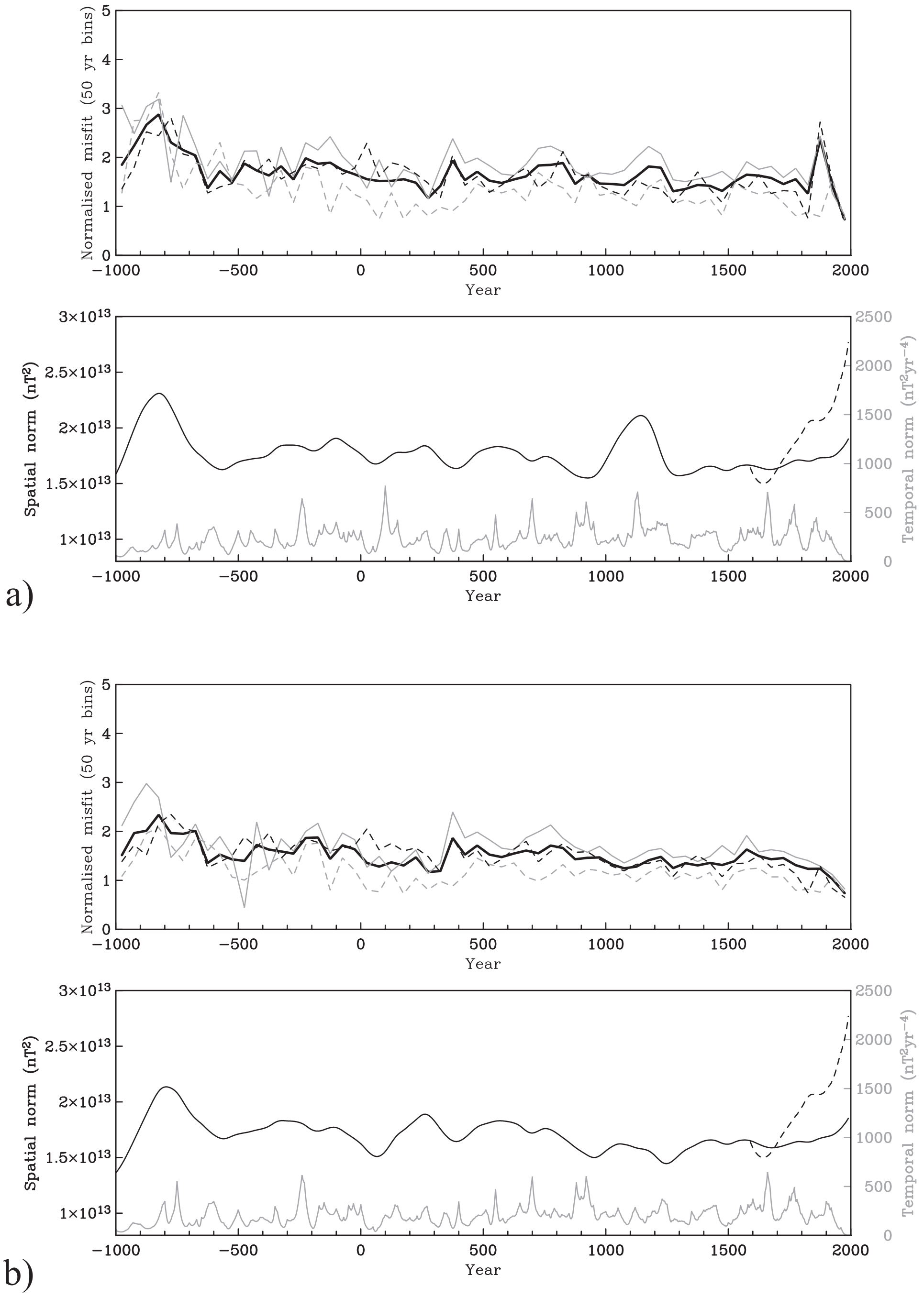
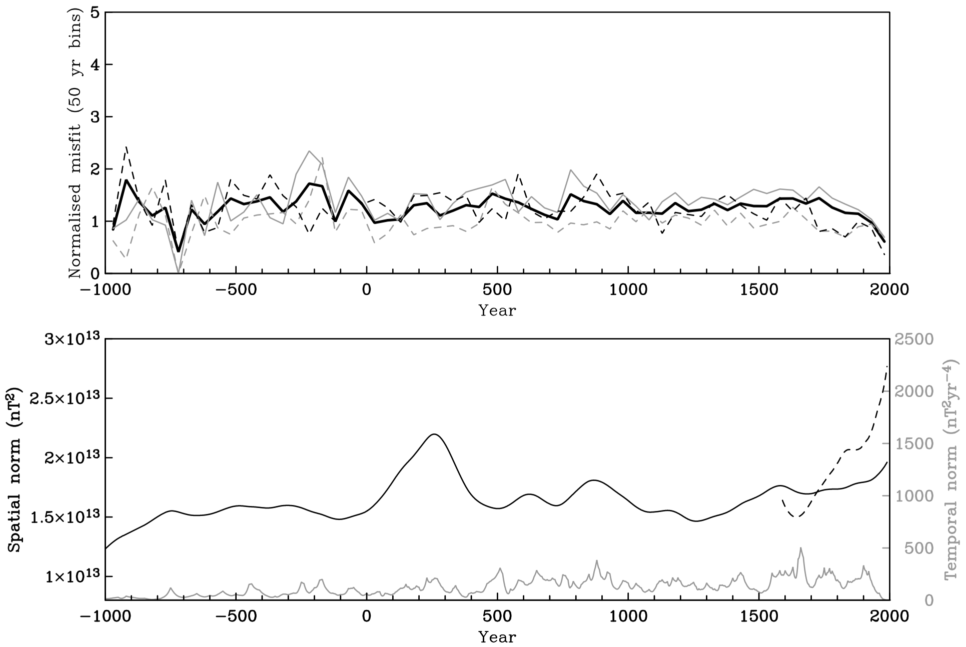


Figure 1. a) Normalized misfit (top) between model ARCH3k.0 (i.e. before outlier rejection) and underlying dataset Arch3k_dat0. All data (thick black line) and inclination (gray), declination (dashed gray) and intensity (dashed black), respectively. Spatial (black) and temporal (gray) roughness of ARCH3k.0 with time (bottom). The dashed black line is the spatial roughness of *gufm1*. b) The same for final model ARCH3k.1 and respective dataset.



q

Figure 2. Normalized misfit (top) between model ARCH3k_cst.1 and underlying dataset Arch3kcst_dat1. All data (thick black line) and inclination (gray), declination (dashed gray) and intensity (dashed black), respectively. Spatial (black) and temporal (gray) roughness of the model with time (bottom). The dashed black line is the spatial roughness of *gufm1*.

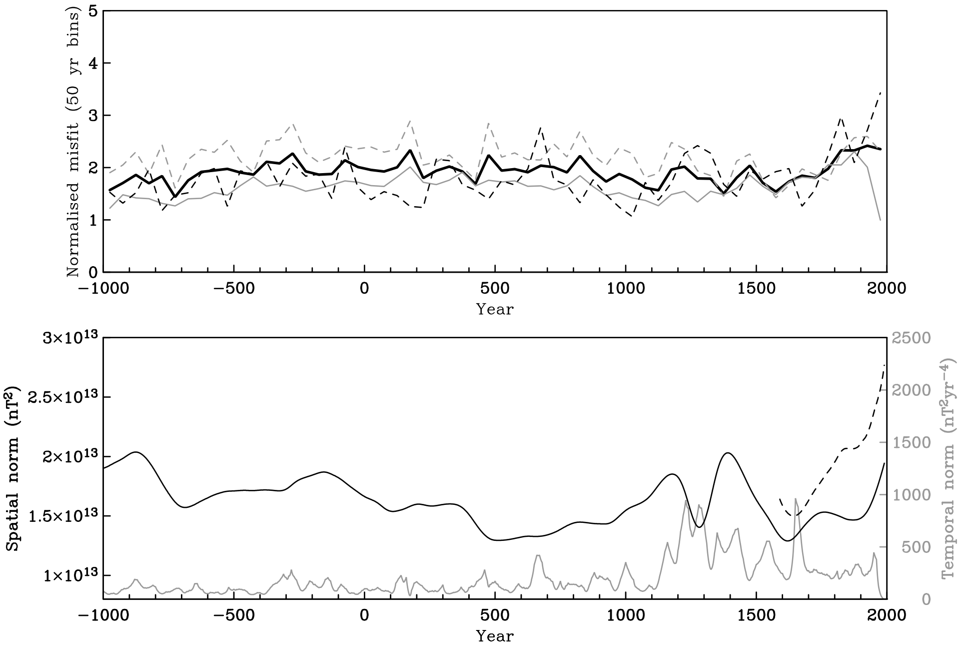


Figure 3. Normalized misfit (top) between model SED3k.1 and underlying dataset Sed3k_dat1. All data (thick black line) and inclination (gray), declination (dashed gray) and intensity (dashed black), respectively. Spatial (black) and temporal (gray) roughness of the model with time (bottom). The dashed black line is the spatial roughness of *gufm1*.

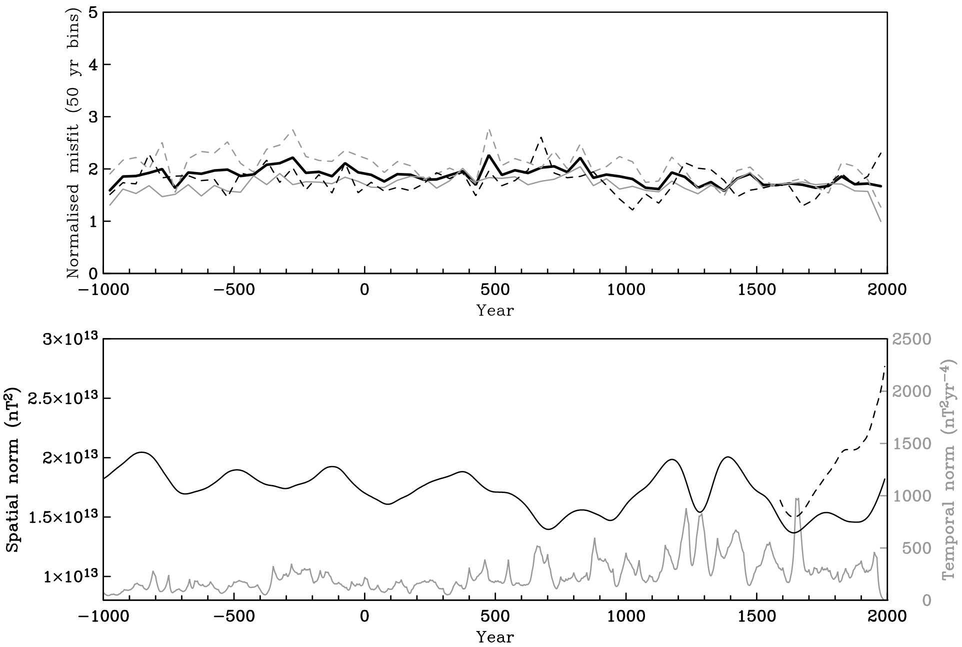


Figure 4. Normalized misfit (top) between model CALS3k.3 and underlying dataset Cals3k_dat1. All data (thick black line) and inclination (gray), declination (dashed gray) and intensity (dashed black), respectively. Spatial (black) and temporal (gray) roughness of the model with time (bottom). The dashed black line is the spatial roughness of *gufm1*.

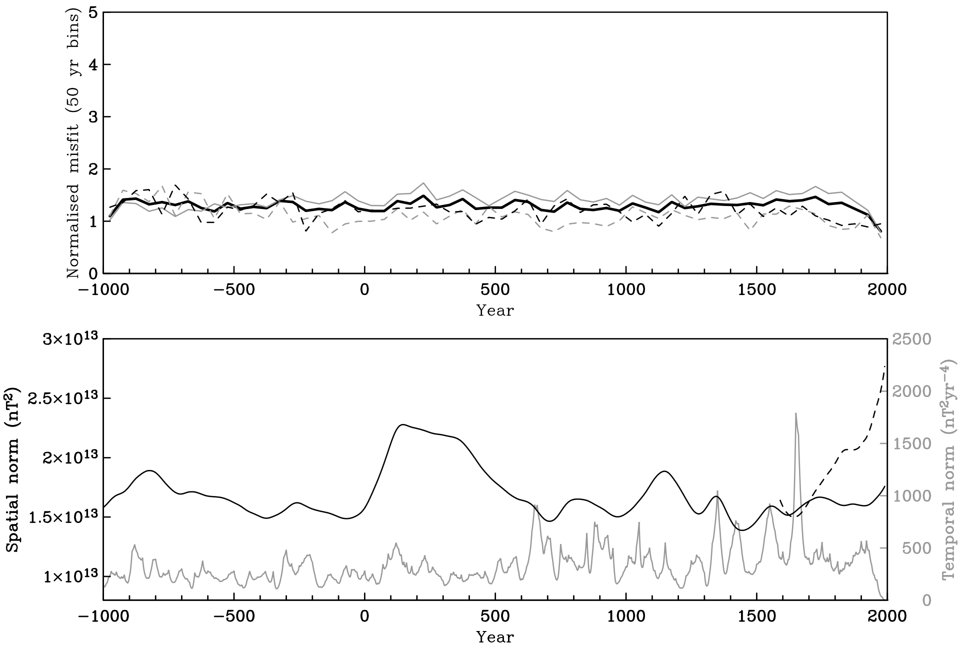


Figure 5. Normalized misfit (top) between model CALS3k_cst.1 and underlying dataset Cals3kcst_dat1. All data (thick black line) and inclination (gray), declination (dashed gray) and intensity (dashed black), respectively. Spatial (black) and temporal (gray) roughness of the model with time (bottom). The dashed black line is the spatial roughness of *gufm1*.

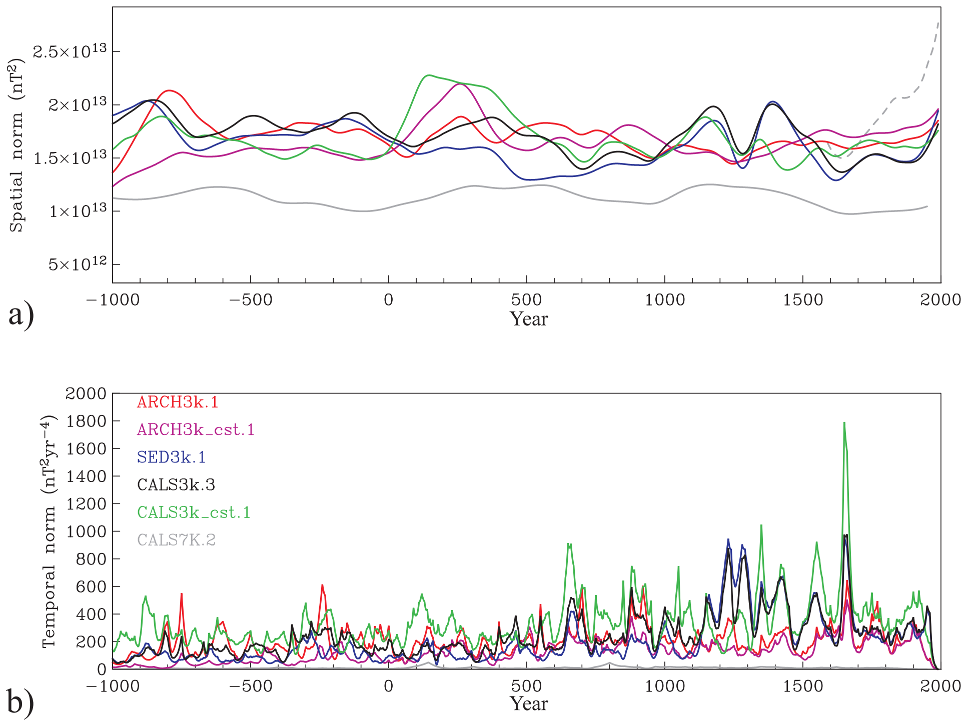


Figure 6. Comparison of a) spatial and b) temporal roughness of all models from Fig. 1b to Fig. 5. Here and in the subsequent figures the models are ARCH3k.1 (red), ARCH3k_cst.1 (magenta), SED3k.1 (blue), CALS3k.3 (black), CALS3k_cst.1 (green) and the previous CALS7K.2 (gray). The dashed gray line is the spatial roughness of *gufm1*. Temporal roughness of *gufm1* (not shown) and CALS7K.2 (hardly visible at this scale) are significantly higher and lower, respectively, due to the different knot-point spacing of the spline basis.

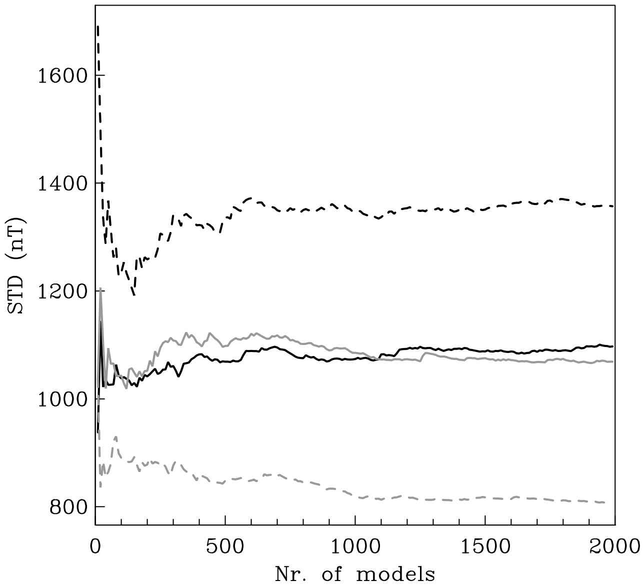


Figure 7. Standard deviation of coefficients g_1^1 (black), h_1^1 (dashed black), g_2^2 (solid gray) and h_2^2 (dashed gray) plotted against number of statistically varied models (MAST method, see text) of model CALS3k.3 for time AD 300 as example demonstrating the convergence.

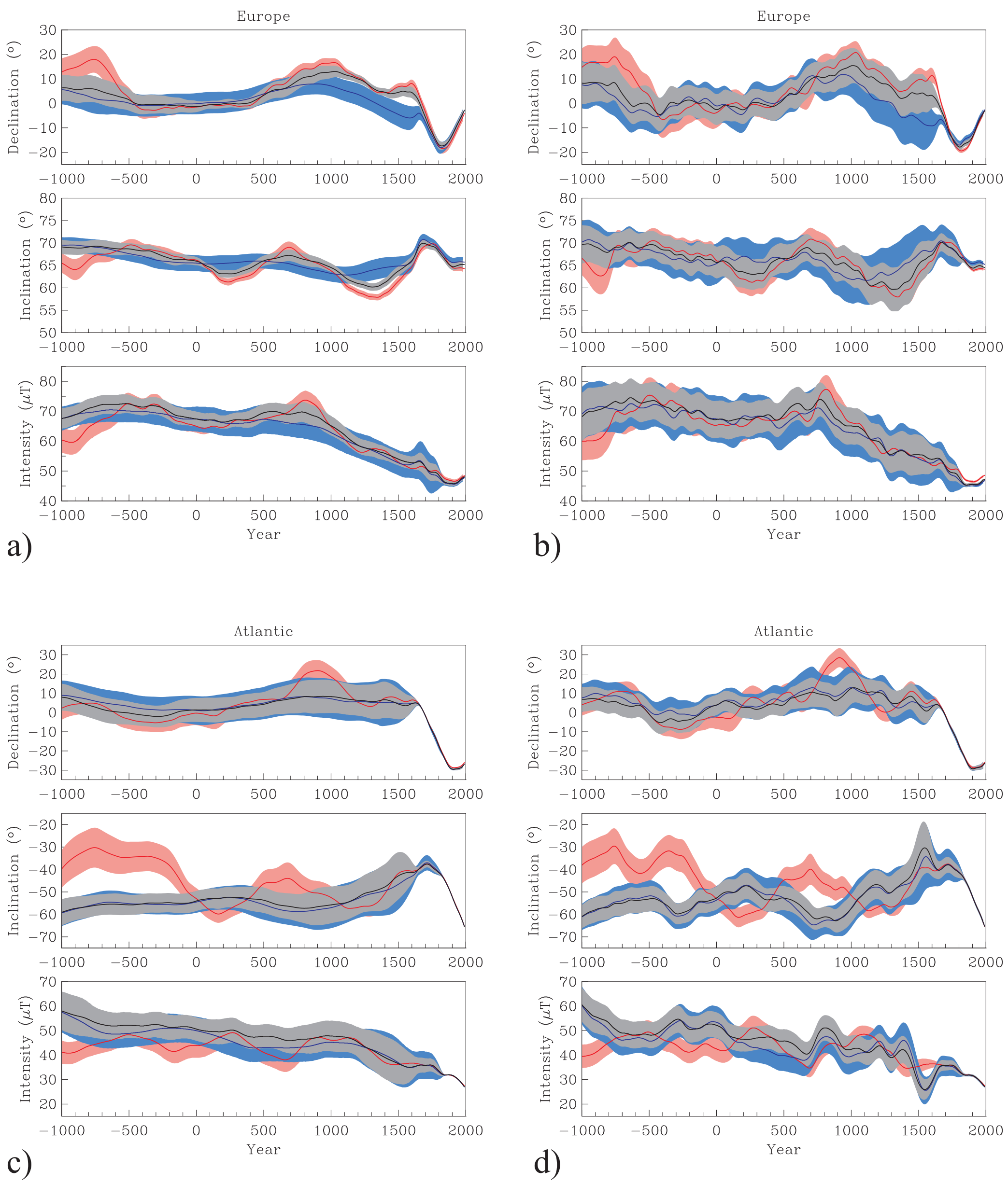


Figure 8. Examples of model prediction uncertainties for a location in central Europe (a,b, location 50°N, 5°E) and in the south Atlantic (c,d, location 35°S, 0°E) for the models ARCH3k.1 (red), SED3k.1 (blue) and CALS3k.3 (black) with shaded uncertainty estimates. On the left side (a,c) average model prediction and uncertainties have been estimated as standard deviation of the component predictions of the statistically varied models, on the right side (b,d) error propagation of the coefficient uncertainties has been used together with the original model coefficients. See text for details.

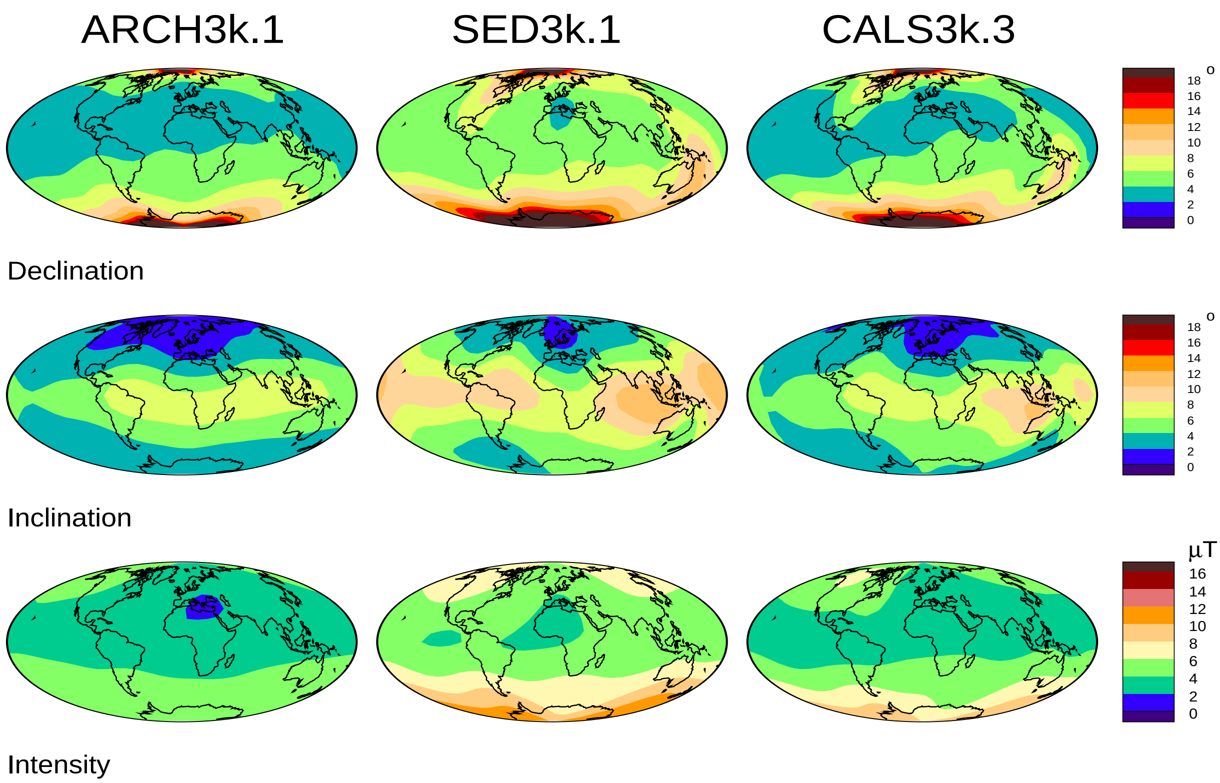


Figure 9. Mean standard deviation in model predictions of declination (top), inclination (middle) and intensity (bottom) for ARCH3k.1 (left), SED3k.1 (middle) and CALS3k.3 (right) at the Earth's surface. Uncertainties in the southern hemisphere of ARCH3k.1 are unrealistically small, due to the lack of data to vary in our statistical estimation.

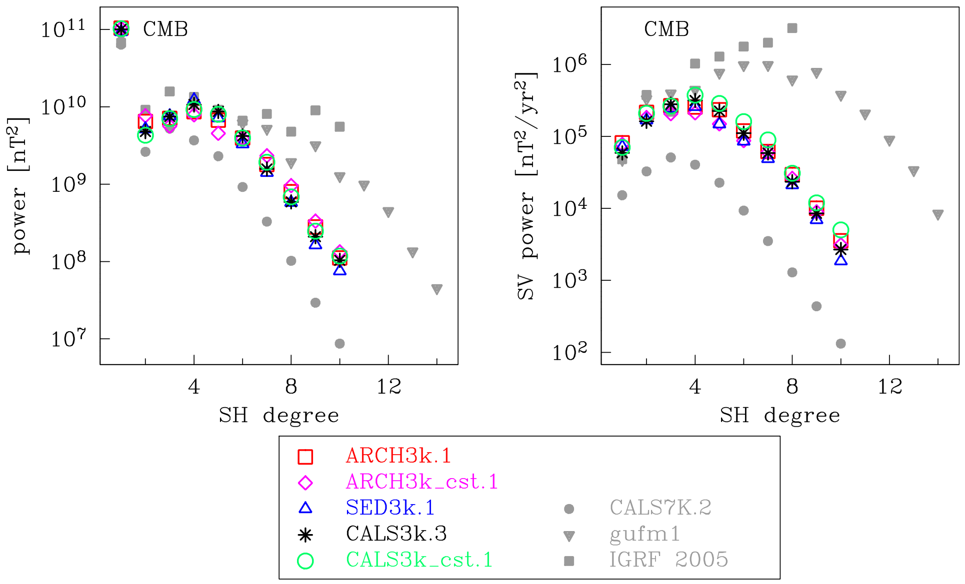


Figure 10. Geomagnetic power spectrum and secular variation power spectrum, all at core-mantle boundary, of different field models.

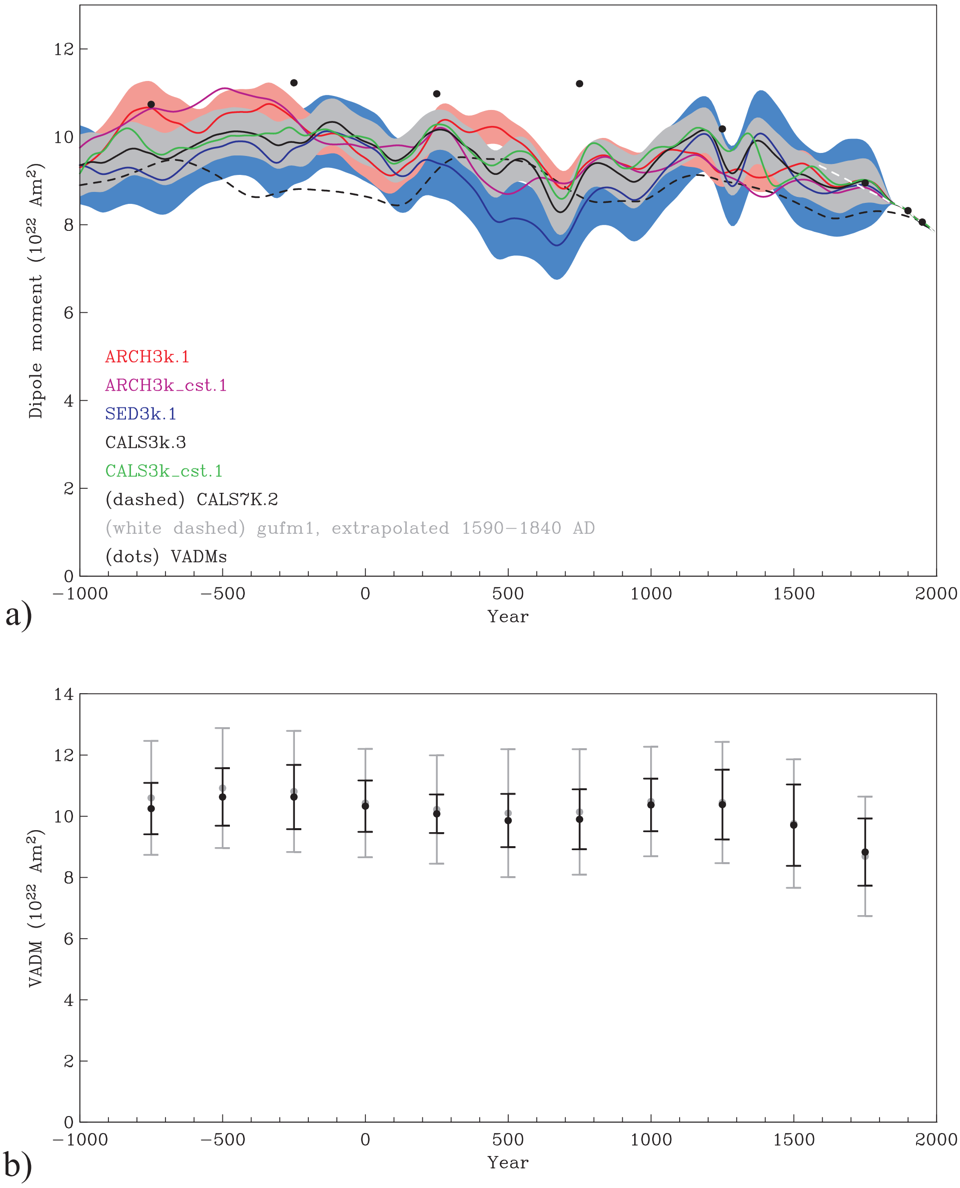
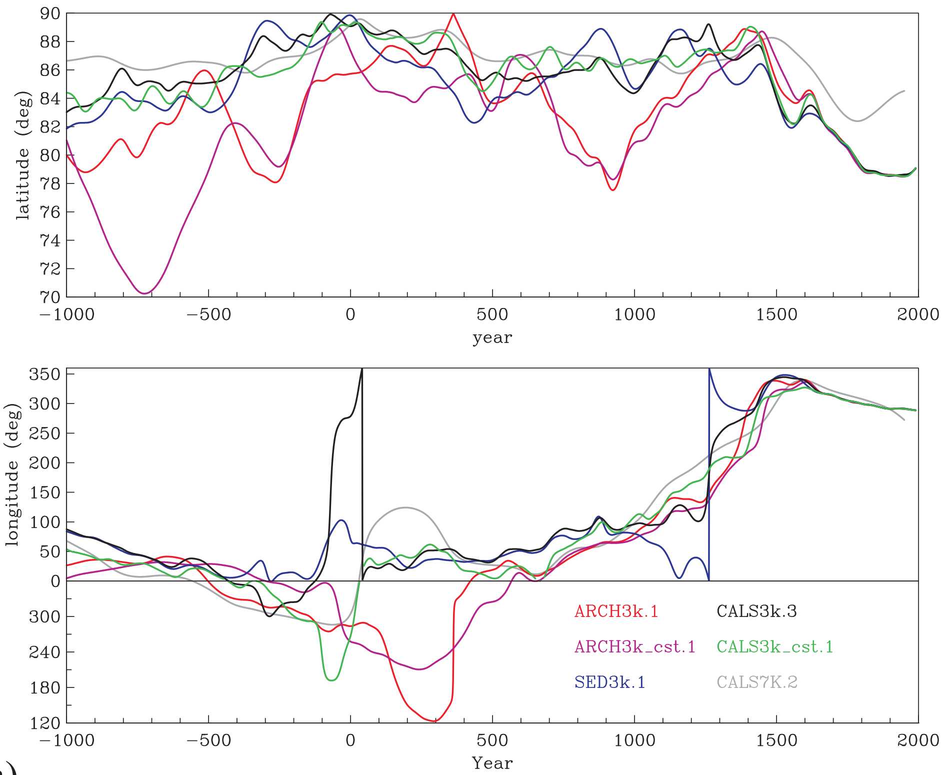
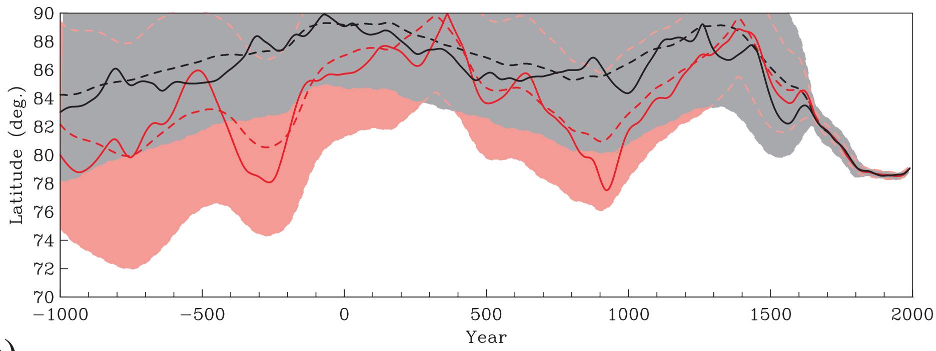


Figure 11. a) Dipole moment of different field models. For ARCH3k.1, SED3k.1 and CALS3k.3 uncertainty estimate ranges are shown in light red, blue and gray, respectively. Black dots are archeointensity VADMs. b) VADMs calculated from the intensity data (gray) used for CALS3k.3 (i.e. including calibrated sediment intensities) and from the model predictions at the same times and locations (black), averaged in 500 year intervals. Both with standard deviations of the distribution.



a)



b)

Figure 12. a) Latitude (tilt) and longitude of the geomagnetic dipole axis according to different field models. The longitude is shown for more than one full circle range in order to avoid too many "jumps" as still seen in CALS3k.3 and SED3k.1 when the axis crosses the zero meridian. b) Latitude of the geomagnetic dipole axis according to ARCH3k.1 (solid red) and CALS3k.3 (solid black) together with error estimates (light red and gray, respectively) and the averages of the statistical method to obtain these error estimates (dashed red and dashed black, respectively).

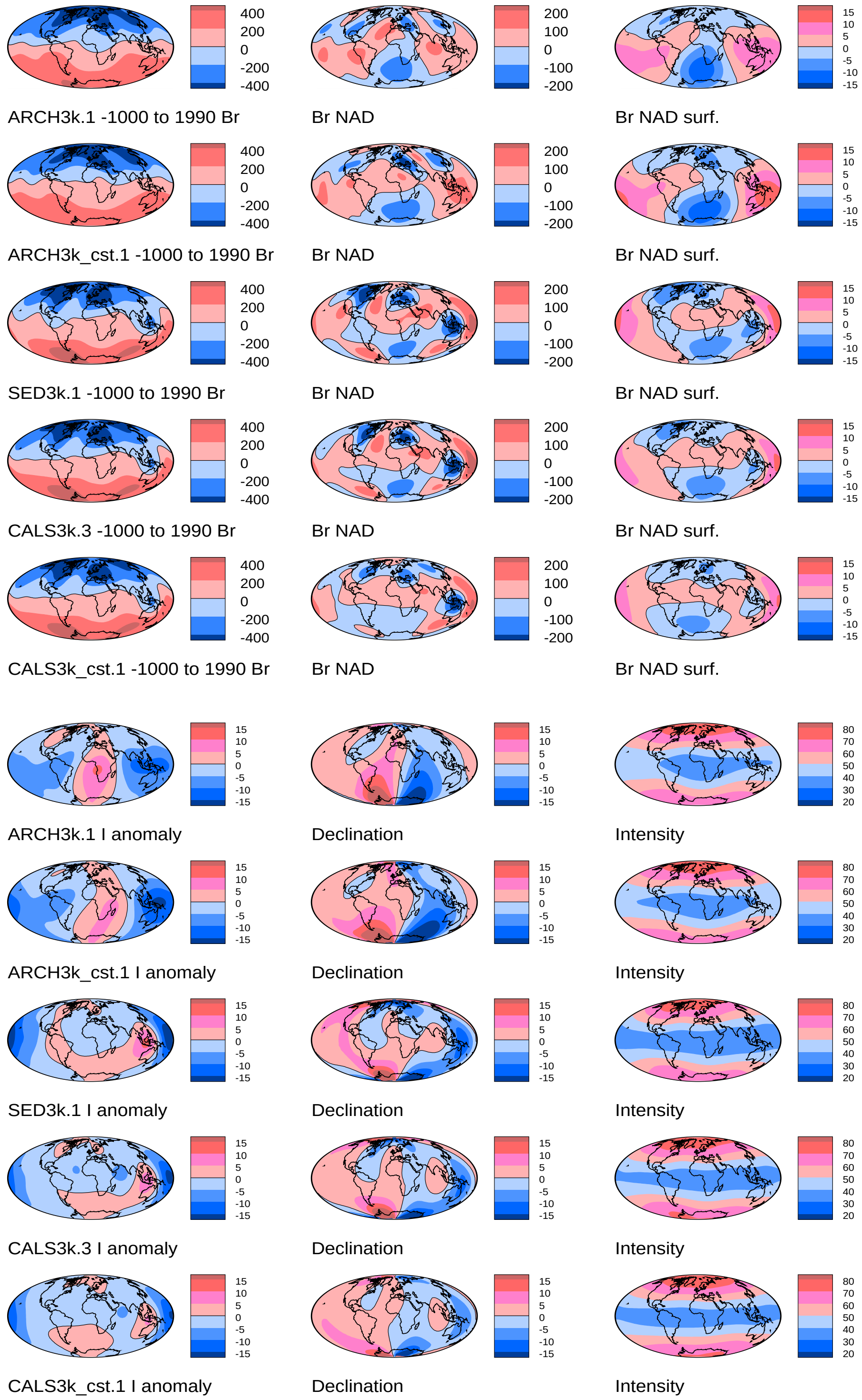


Figure 13. Comparison of averages of different field components of the five models for the whole 3kyrs: radial component (Br) and Br with axial dipole contribution subtracted (Br NAD), both at the core-mantle boundary (top left and middle), and Br NAD at the surface (top right). Inclination (I) anomaly, declination (D) and intensity (F) at the Earth's surface (bottom left to right). Units are degrees for D and I anomaly and μT for other components.