

FULL PAPER

Open Access

# Geomagnetic observatory monthly series, 1930 to 2010: empirical analysis and unmodeled signal estimation

Jiaming Ou<sup>1,3</sup>, Nicolas Gillet<sup>2\*</sup> and Aimin Du<sup>1,3</sup>

## Abstract

Ground-based magnetic observatory series are the main source of information for constructing time-dependent spherical harmonic geomagnetic field models from sub-annual to pluri-decadal time scales. Assessing the reliability of such models requires accurate estimation of the data errors. We propose an analysis of observatory monthly means over the period 1930 to 2010, where we sequentially isolate (i) a stochastic regression for the main field at every site, performed in the framework of Gaussian processes, (ii) a local fit to annual and semiannual signals, (iii) a month by month estimate of global, large length-scale external and induced fields. We then estimate the unmodeled signal level (UMSL, which refers to the instrumental noise plus extra signals not captured by the above data treatment) from the standard deviation of the residuals to the sequential analysis. This may be used to estimate data error covariances in future field modeling studies. Mainly a function of the geomagnetic latitude, the UMSL is larger towards auroral regions and carries the temporal signature of solar activity. While the UMSL shows rather similar magnitudes in all three components in recent epochs (typically a few nT), a significant decrease is found in the downward component of the field around 1960, which correlates with the introduction of proton magnetometers. We detail the geographic distribution of the periodic signals and confirm the variation of their amplitude at pluri-decadal time scales. From the spherical harmonic description of horizontal and vertical fields, we isolate the main patterns of the inducing field in  $Z$ . These are dominated by a zonal structure of degree 1 (and to a lesser extent, of degree 3) in dipole coordinates. We nevertheless isolate secondary, non-zonal sources that are most active during the 1960s and around 1990, periods of particularly large solar activity, denoting an unusual morphology of the inducing field.

**Keywords:** Geomagnetic field modeling; Observatory data; Error estimates

## Background

Dissociating external (ionospheric, magnetospheric) and internal (induced, crustal, core dynamo) magnetic sources is a major limitation to the spherical harmonic global modeling of geomagnetic records (e.g. Olsen et al. 2010a). This has repercussions on studies concerning both core physics and external fields. Indeed, the recovery of core motions at periods shorter than a few years is hindered by the domination of external sources towards high frequencies, and the reconstruction of changes in the magnetospheric ring current over decadal periods (e.g. Gillet et al. 2013; McLeod 1996; Sabaka et al. 1997) is made

ambiguous because of non-global coverage of observatory sites.

Observatory series are the only records that permit the analysis of the geomagnetic field from sub-annual to pluri-decadal periods. Several global modeling avenues may be employed using such data. In a comprehensive approach, Sabaka et al. (2002, 2004) simultaneously inverted all sources from hourly values spanning 1962 onward. Alternatively, the complex signature from external fields is considered as contributing to the data error budget when recovering main field changes from the annual means series starting in 1840 (Jackson et al. 2000). Intermediate strategies considering a simple external field parameterization (Gillet et al. 2013), or a sequential elimination of the sources, may also be used.

\*Correspondence: nicolas.gillet@ujf-grenoble.fr

<sup>2</sup>University Grenoble Alpes, ISTERre, CNRS, F-38041 Grenoble, France  
Full list of author information is available at the end of the article

However, assessing the robustness of the output models is made difficult by the crude estimates of the observation errors. Accurate magnitudes of the data error covariances are required in order to provide realistic *a posteriori* confidence levels for the field models. For instance, by allocating constant error bars on observatory annual means series, Jackson et al. (2000) and Gillet et al. (2013) have relatively underweighted the information contained in the most accurate data when building, respectively, the *gufm1* and *COV-OBS* field models. With such an approach, *a posteriori* uncertainties tend to be overestimated towards recent epochs (when the data are of better quality), and time changes in the variance of some unmodeled signals are ignored (when they should be accounted for through the data error covariance matrix).

One could, in principle, consider the best documented period (the satellite era) to characterize unmodeled sources at ancient epochs. However, this approach is limited by the existence of external field variations for periods longer than a decade. Furthermore, there is no consensus on the variance one should consider to characterize satellite data errors (Finlay et al. 2012; Lesur et al. 2010; Olsen et al. 2010b). In a situation where there exists a frequency overlap within the different sources (given the recent estimates of the mantle conductivity, there is possibly a signature in magnetic series of core processes at periods shorter than a year, see Velínski and Finlay (2011)), it is tempting to analyze the observatory monthly mean (OMM) series to try to better isolate the several contributions at periods ranging from a few months to a few decades.

When allocating error variances to observatory series, residuals from smooth fits to individual series, or from an initial time-dependent global field model prediction, have sometimes been considered. As an example, the generalized cross-validation method was applied to individual observatory annual mean series (see Bloxham and Jackson 1992) for the construction of the *gufm1* and *COV-OBS* models, and Wardinski and Holme (2006) and Wardinski and Lesur (2012) calculated error cross-covariances between residuals from three-component OMM series in constructing the  $C^3FM$  models from 1957 onward. This pragmatic avenue nevertheless implicitly relies on the modeling assumptions, through a regularization process involving free damping parameters and specific choices of temporal basis functions. Those govern the temporal smoothness through their differentiability properties and density and thus implicitly state the cutoff between the noise and the signal.

Using principal component analysis, Wardinski and Holme (2011) found large correlations between the residuals from a main field model and the *Dst* magnetic index. This led them to remove from the data series a statistical proxy for unmodeled signals of external origin. Such

an approach is promising in the quest for detailed secular variation patterns at observatories. However, indices such as *Dst* do not necessarily represent well the external activity worldwide (especially at high latitudes). Furthermore, it may be biased at periods longer than a year because of the time changes of the *Dst* baseline (see for instance Lühr and Maus 2010; Olsen et al. 2005b). The alternative RC-index (Olsen et al. 2014) has recently been introduced to avoid this specific issue.

In an ideal situation, in the context of global spherical harmonic modeling, one should follow a comprehensive approach where all signal sources are simultaneously inverted for, in order to see the unmodeled signals vanish as much as possible. This is unfortunately not possible when looking back to the early observatory era before 1960, as common measurement protocols had not yet been adopted via the INTERMAGNET global network of magnetic observatories (e.g., Love and Chulliat 2013), and the measurement quality was poorer (e.g., Matzka et al. 2010). In this configuration, modeling errors have to be considered along with measurement uncertainties.

The purpose of the present work is to present an empirical analysis of OMM series, by which we provide an estimate of the unmodeled signal level (UMSL). It relies on a sequential cleaning of the contributions from several magnetic sources in the monthly series of the northward (*X*), eastward (*Y*), and downward (*Z*) components of the geomagnetic field. These compose the magnetic field vector **B** in geographic coordinates. We isolate three main contributions as follows:

- 1) First, a stochastic fit for the main field signal is obtained separately for each observatory using a Gaussian process regression that includes our prior knowledge of the core field temporal cross-covariances.
- 2) Next, signals with annual and semiannual periods are fit to the residuals between the OMM series and the main field regression, performed separately for each observatory.
- 3) Finally, we perform a month-by-month large length-scale global inversion of the remaining residuals. This approximates the main external source together with its associated induced fields.

Using a monthly sampling rate, we have to consider the large signals of periods of 1 year and half a year. If these have been isolated in observatory series for more than a century (e.g., Chapman and Bartels 1940), their mechanisms are still under debate. For this reason and because of their complex geographic distribution, we must fit them separately from the global, large length-scale external fields.

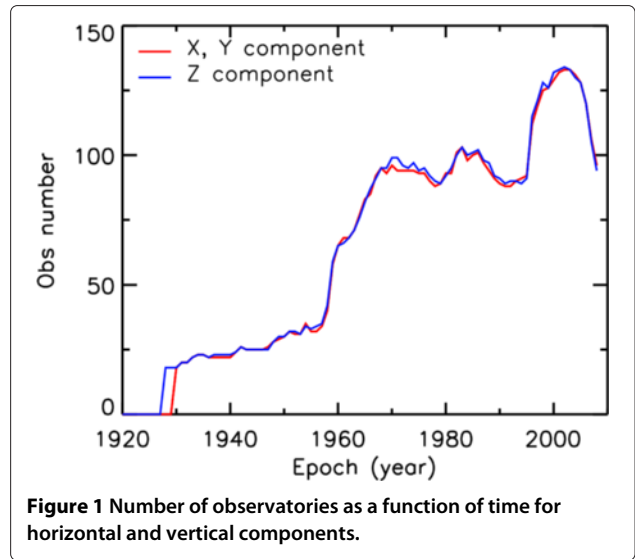
After presenting the OMM series used throughout this study (Section ‘Data’), we detail in Section ‘Methods’ the three steps of our sequential analysis. The several contributions to the OMM series are presented in Section ‘Results and discussion’, together with the variations of the UMSL in space and time. Finally, we summarize our analysis in Section ‘Conclusions’.

**Data**

We present in this section the OMM database employed throughout this study. We refer for instance to Matzka et al. (2010) for a more complete description of observatory data and the several signals they contain. We benefit from the strategy of Chulliat and Telali (2007), who built the worldwide OMM archive from hourly mean values provided by the world data center (WDC), checking for discontinuities (baseline jumps) in magnetic series. We build our own OMM series from hourly means provided by the Edinburgh WDC Catalog, for two main reasons:

- 1) Some series for which hourly means are available are not included in the Chulliat and Telali (2007) database (either because of lack of continuity, too low quality, or the presence of incompatibilities with the annual means provided by the WDC). Instead, we consider that those data may be helpful to infer some information on the core signal at ancient epochs and find it interesting to estimate their associated UMSL.
- 2) We recalculate the OMM in a slightly different way. In anticipation of a future purpose of main field modeling, we use only the midnight 3-h data in order to avoid the diurnal variation from the ionospheric *Sq* current system, which is large (weak) in the local daytime (midnight). We are aware that induced fields, with amplitudes of a few nT, are associated with *Sq* variations (Olsen et al. 2005a), but these essentially consist of large length-scale patterns that should be accounted for through the global large length-scale fit. To retrieve the UMSL of the OMM, we thus do not consider the impact from *Sq* variations.

We paid much attention to the correction for discontinuities in the monthly series. Since a large number of these ‘jumps’ are neither documented nor integrated in the WDC archives, we follow the strategy of Chulliat and Telali (2007) to diagnose and fix the baseline change issues in the series of daily values of midnight to 3-h means. It is unfortunately not possible to recognize all these offsets, especially the small ones, which will be hidden by the signature of natural rapid variations from external or induced origin. We thus consider that this type of undiscovered discontinuities contribute to the unmodeled signals.



**Figure 1** Number of observatories as a function of time for horizontal and vertical components.

We finally end up with a total of 197 OMM series. The total number of sites changes with time (see Figure 1). Additionally, we introduce some monthly data from Chinese observatories for the period 1995 to 2008 that are not published in the WDC and are not part of INTERMAGNET. These data are also calculated from daily values and are collected and administrated by the Institute of Geology and Geophysics, Chinese Academy of Sciences. All observatories used throughout this study are listed in Table 1.

**Methods**

**Local stochastic regression of the main field**

We wish to avoid the use of a previously built time-dependent geomagnetic field model when removing the internal contribution to the observed series, because most spherical harmonic internal models rely on temporal basis functions presenting specific properties that are not justified in a physical or statistical manner. Along the same lines, we wish to avoid the use of global (e.g., polynomial) or local (e.g., splines) functions to directly fit the series. We thus employ the Gaussian processes modeling framework (e.g., Rasmussen and William 2006), where the *a posteriori* expected value  $\mathbf{p}$  of the process, calculated at epochs stored in a vector  $\mathbf{t}_p$  (of size  $N_p$ ), is entirely determined by the knowledge of (i) an observation vector  $\mathbf{d}$  recorded at epochs stored in a vector  $\mathbf{t}_d$  (of size  $N_d$ ), (ii) the covariance matrix  $C_{ee}$  that describes the statistics of the observation errors (of size  $N_d \times N_d$ ), and (iii) the *a priori* covariance function  $C(\tau) = E(\varphi(t)\varphi(t + \tau))$  for the series  $\varphi(t)$  of time  $t$  (where  $E$  stands for ‘expectation’). It comes down to calculating the best linear unbiased estimate (BLUE):

$$\hat{\mathbf{p}} = C_{pd} (C_{dd} + C_{ee})^{-1} \mathbf{d}. \tag{1}$$

**Table 1 List of the observatories used throughout this study**

Observatory	Code	$\lambda$	$\phi$	Period	Observatory	Code	$\lambda$	$\phi$	Period
Abinger	ABN	51.2	359.6	1925-1965	La Quiaca	LQA	-22.1	294.4	1963-1981
Abisko	ABK	68.4	18.8	1950-2010	Lanzhou	LZH	36.1	103.9	1980-2006
Addis Ababa	AAE	9.0	38.8	1959-1995	Las Mesas	TEN	28.5	343.7	1961-1992
Agincourt	AGN	43.8	280.7	1931-1970	Learmonth	LRM	-22.2	114.1	1990-2010
Alert	ALE	82.5	297.6	1962-2005	Leirvogur	LRV	64.2	338.3	1962-2010
Alibag	ABG	18.6	72.9	1922-2010	Lerwick	LER	60.1	358.8	1925-2010
Alice Springs	ASP	-23.8	133.9	1991-2010	Livingston Island	LIV	-62.7	299.6	1996-2010
Alma Ata	AAA	43.2	76.9	1962-2010	Loparskaya	MMK	68.2	33.1	1960-1990
Annamalainagar	ANN	11.4	79.7	1967-1994	Lovo	LOV	59.3	17.8	1930-2004
Apia	API	-13.8	188.2	1933-2010	Lunping	LNP	25.0	121.2	1980-2000
Arctowski	ARC	-62.2	301.5	1978-1995	Lvov	LVV	49.9	23.7	1957-2010
Arti	ARS	56.4	58.6	1972-2010	MBour	MBO	14.4	343.0	1952-2010
Ascension Island	ASC	-7.9	345.6	1993-2010	Macquarie Island	MCQ	-54.5	158.9	1957-2010
Ashkhabad	ASH	37.9	58.1	1960-1990	Manhay	MAB	50.3	5.7	1995-2010
Baker Lake	BLC	64.3	264.0	1951-2010	Manzhouli	MZL	49.6	117.4	1995-2007
Bangui	BNG	4.3	18.6	1955-2010	Martin de Vivies	AMS	-37.8	77.6	1981-2010
Bar Gyora	BGY	31.7	35.1	1990-2010	Mawson	MAW	-67.6	62.9	1963-2010
Barrow	BRW	71.3	203.4	1964-2010	Meanook	MEA	54.6	246.7	1931-2010
Bear Island	BJN	74.5	19.2	1987-2006	Memambetsu	MMB	43.9	144.2	1958-2010
Beijing	BJI	40.0	116.2	1957-2005	Mirny	MIR	-66.6	93.0	1956-1997
Beijing Ming Tombs	BMT	40.3	116.2	1996-2007	Mizusawa	MIZ	39.1	141.2	1980-2010
Belsk	BEL	51.8	20.8	1966-2010	Molodezhnaya	MOL	-67.7	45.8	1967-1977
Bereznyaki	KGD	49.8	73.1	1965-1976	Mould Bay	MBC	76.3	240.6	1962-1996
Borok	BOX	58.1	38.2	1980-2010	Muntinlupa	MUT	14.4	121.0	1963-1972
Boulder	BOU	40.1	254.8	1972-2010	Nagpur	NGP	21.1	79.0	1995-2009
Brorfelde	BFE	55.6	11.7	1981-2009	Nagycenk	NCK	47.6	16.7	1992-2010
Budkov	BDV	49.1	14.0	1994-2010	Narsarsuaq	NAQ	61.2	314.6	1968-2009
Cambridge Bay	CBB	69.1	255.0	1972-2010	Newport	NEW	48.3	242.9	1966-2010
Canberra	CNB	-35.3	149.4	1981-2010	Niemegk	NGK	52.1	12.7	1901-2010
Cape Chelyuskin	CCS	77.7	104.3	1956-1985	Novolazarevskaya	NVL	-70.8	11.8	1961-1978
Cape Wellen	CWE	66.2	190.2	1956-1987	Novosibirsk	NVS	54.8	83.2	1967-2010
Casey	CSY	-66.3	110.5	1989-2005	Nurmijarvi	NUR	60.5	24.7	1953-2009
Chambon-la-Foret	CLF	48.0	2.3	1936-2010	Ottawa	OTT	45.4	284.4	1968-2010

**Table 1 List of the observatories used throughout this study (Continued)**

Observatory	Code	$\lambda$	$\phi$	Period	Observatory	Code	$\lambda$	$\phi$	Period
Changchun	CNH	43.8	125.3	1995-2007	Pamatai	PPT	-17.6	210.4	1967-2010
Changli	CHL	39.7	119.0	1995-2007	Panagjurishte	PAG	42.5	24.2	1963-2010
Charters Towers	CTA	-20.1	146.3	1990-2010	Paratunka	PET	53.0	158.2	1969-1995
Cheltenham	CLH	8.7	283.2	1901-1956	Phuthuy	PHU	21.0	105.9	1996-2010
Chengdu	CDP	31.0	103.7	1995-2007	Pilar	PIL	-31.7	296.1	1941-1980
Chongqing	COQ	29.4	106.6	1999-2006	Pleshentzi	MNK	54.5	27.9	1961-1995
Chichijima	CBI	27.1	142.2	1991-2010	Pondicherry	PND	11.9	79.9	1995-2008
College	CMO	64.9	212.1	1948-2010	Port Alfred	CZT	-46.4	51.9	1974-2009
Dalian	DLG	39.0	121.5	2000-2007	Port Moresby	PMG	-9.4	147.2	1957-1993
Dallas	DAL	33.0	263.2	1964-1974	Port Stanley	PST	-51.7	302.1	1994-2009
De Bilt	DBN	52.1	5.2	1906-1938	Port-aux-Francais	PAF	-49.4	70.3	1957-2009
Dedu	DED	48.6	126.1	1995-2005	Poste-de-la-Baleine	PBQ	55.3	282.3	1984-2007
Del Rio	DLR	29.5	259.1	1982-2009	Qaanaaq	THL	77.5	290.8	1947-2010
Dikson	DIK	73.5	80.6	1957-1986	Qeqertarsuaq	GDH	69.3	306.5	1946-2009
Dombas	DOB	62.1	9.1	1957-2009	Qianling	QIX	34.6	108.2	1995-2006
Dourbes	DOU	50.1	4.6	1957-2010	Resolute Bay	RES	74.7	265.1	1956-2010
Dumont d'Urville	DRV	-66.7	140.0	1957-2010	Rude Skov	RSV	55.8	12.4	1927-1981
Dusheti	TFS	42.1	44.7	1956-2001	Sable Island	SBL	43.9	300.0	1999-2010
Dymer	KIV	50.7	30.3	1958-1991	St. Johns	STJ	47.6	307.3	1969-2010
Ebro	EBR	40.8	0.5	1995-2010	San Fernando	SFS	36.7	354.1	1996-2010
Eilat	ELT	29.7	34.9	1998-2010	San Juan	SJG	18.1	293.8	1921-2010
Ekaterinburg	EKT	56.8	60.6	1901-1925	San Pablo-Toledo	SPT	39.5	355.6	1996-2010
Esashi	ESA	39.2	141.4	1997-2010	Sanae III	SNA	-70.3	357.6	1961-1990
Eskdalemuir	ESK	55.3	356.8	1911-2010	Scott Base	SBA	-77.8	166.8	1957-2010
Eyrewell	EYR	-43.4	172.4	1980-2010	Sheshan	SSH	31.1	121.2	1932-2010
Faraday Islands	AIA	-65.2	295.7	1957-2010	Simosato	SSO	33.6	135.9	1957-1975
Fort Churchill	FCC	58.8	265.9	1957-2010	Sitka	SIT	57.1	224.7	1904-2010
Fredericksburg	FRD	38.2	282.6	1956-2010	Sodankyla	SOD	67.7	26.6	1913-2009
Fresno	FRN	37.1	240.3	1982-2010	Stekolnyy	MGD	60.1	151.0	1966-1990
Fuquene	FUQ	5.5	286.3	1955-2010	Stennis Space Centre	BSL	30.3	270.4	1986-2010
Furstenfeldbruck	FUR	48.2	11.3	1940-2010	Stepanovka	ODE	46.8	30.9	1957-1991
Glenlea	GLN	49.6	262.9	1984-1997	Surlari	SUA	44.7	26.3	1957-2009
Gnangara	GNA	-31.8	115.9	1957-2010	Tamanrasset	TAM	22.8	5.5	1953-2010

**Table 1 List of the observatories used throughout this study (Continued)**

Observatory	Code	$\lambda$	$\phi$	Period	Observatory	Code	$\lambda$	$\phi$	Period
Golmud	GLM	36.4	94.9	1995-2007	Tianshui	TSY	34.6	105.9	1995-2007
Gornotayezhnaya	VLA	43.7	132.2	1957-1990	Tihany	THY	46.9	17.9	1957-2009
Great Whale River	GWC	55.3	282.2	1965-1981	Tirunelveli	TIR	8.7	77.8	1999-2008
Guam	GUA	13.6	144.9	1957-2010	Tonghai	THJ	24.0	102.7	1995-2007
Guangzhou	GZH	23.1	113.3	1960-2009	Trelew	TRW	-43.3	294.6	1957-2010
Guimar	GUI	28.3	343.6	1993-2010	Trivandrum 2	TRD	8.5	76.9	1957-1999
Guiyang	GYX	26.6	106.8	1995-2005	Tromso	TRO	69.7	18.9	1957-2009
Halley Bay	HBA	-75.5	333.4	1957-1967	Tsumeb	TSU	-19.2	17.6	1964-2009
Hangzhou	HZC	30.2	120.1	1995-2007	Tucson	TUC	32.2	249.3	1910-2010
Hartebeesthoek	HBK	-25.9	7.7	1972-2009	Ujjain	UJJ	23.2	75.8	1980-2003
Hartland	HAD	51.0	355.5	1957-2010	Urumqi	WMQ	43.8	87.7	1995-2007
Hatizyo	HTY	33.1	139.8	1986-2009	Val Joyeux	VLJ	48.8	2.0	1901-1936
Heiss Island	HIS	80.6	58.0	1960-1970	Valentia	VAL	51.9	349.7	1957-2010
Hel	HLP	54.6	18.8	1966-2010	Vassouras	VSS	-22.4	316.3	1915-2010
Hermanus	HER	-34.4	19.2	1941-2010	Victoria	VIC	48.5	236.6	1964-2010
Honolulu	HON	21.3	202.0	1905-2010	Vieques	VQS	18.1	294.5	1905-1924
Hornsund	HRN	77.0	15.5	1978-2010	Visakhapatnam	VSK	17.7	83.3	1995-2009
Huancayo	HUA	-12.0	284.7	1921-2010	Voeikovo	LNN	59.9	30.7	1947-1988
Hurbanovo	HRB	47.9	18.2	1906-2010	Vostok	VOS	-78.4	106.9	1958-1979
Iqaluit	IQA	63.8	291.5	1995-2010	Vysokaya Dubrava	SVD	56.7	61.1	1930-1974
Irkutsk	IRT	52.2	104.4	1957-2010	Watheroo	WAT	-30.3	115.9	1919-1959
Istanbul-Kandilli	ISK	41.1	29.1	1950-2000	Wien Kobenzl	WIK	48.3	16.3	1957-2005
Jaipur	JAI	26.9	75.8	1980-209	Wingst	WNG	53.7	9.1	1942-2010
Jinghai	JIH	38.9	116.9	1995-2007	Witteveen	WIT	52.8	6.7	1938-1987
Juyongguan	JYG	39.8	98.2	1998-2007	Wuhan	WHN	30.5	114.6	1995-2007
Kakadu	KDU	-12.7	132.5	1995-2010	Yakutsk	YAK	62.0	129.7	1957-2009
Kakioka	KAK	36.2	140.2	1913-2010	Yangi Bazar	TKT	41.3	69.6	1957-1981
Kanoya	KNY	31.4	130.9	1958-2010	Yellowknife	YKC	62.5	245.5	1977-2010
Kanozan	KNZ	35.3	140.0	1961-2010	Yinchuan	YCB	38.5	106.3	1995-2007
Kiruna	KIR	67.8	20.4	2001-2010	Yongning	YON	22.8	108.5	1996-2006
Kourou	KOU	5.2	307.3	1996-2010	Yuzhno Sakhalinsk	YSS	46.9	142.7	1957-1988
Krasnaya Pakhra	MOS	55.5	37.3	1957-2005	Zaymishche	KZN	55.8	48.8	1964-1989
L'Aquila	AQU	42.4	13.3	1960-2010					

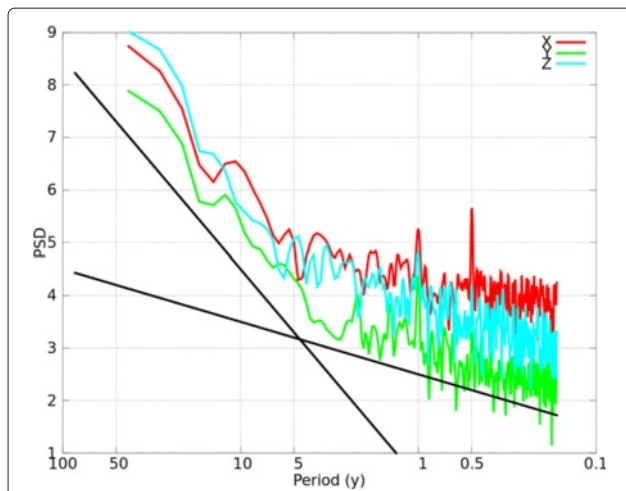
Together with their code, position (latitude  $\lambda$  and longitude  $\phi$ , in degrees), and period over which their data were employed.

Here, the element in the  $i$ th row and  $j$ th column of the matrix  $C_{pd} = E(\mathbf{pd}^T)$ , of size  $N_p \times N_d$ , is  $C_{pdij} = C(t_{p_i} - t_{d_j})$ ; and similarly for  $C_{dd} = E(\mathbf{dd}^T)$ , of size  $N_d \times N_d$ , we have  $C_{ddij} = C(t_{d_i} - t_{d_j})$ .

Our choice for  $C(\tau)$  is motivated by the analysis of observatory series. The auto-covariance function of a series defines its power spectrum  $S(f)$ , where  $f$  is the frequency. As illustrated in Figure 2, that of observatory records evolves approximately as  $S(f) \propto f^{-4}$  at periods longer than 5 years (Currie 1968; De Santis et al. 2003). As highlighted by Gillet et al. (2013), it is convenient to model such series in the framework of stochastic differential equations (SDEs), for which a  $-4$  slope spectrum at high frequencies corresponds to an SDE of order 2. If there is no evidence for the continuation of the  $-4$  spectrum at periods as short as 1 month, Gillet et al. (2013) nevertheless show that there is no need for a higher order process (and thus a steeper spectrum slope) at such short periods. It is indeed difficult to define the *a priori* information about the core field for interannual to monthly periods, due to the ambiguity between internal and external magnetic sources.

We choose the particular *a priori* auto-covariance function

$$C(\tau) = \sigma^2 \left( 1 + \frac{\tau}{\tau_c} \right) \exp\left(-\frac{\tau}{\tau_c}\right) \quad (2)$$



**Figure 2 Power spectral density ( $\log_{10}$ , in  $\text{nT}^2/\text{yr}^{-1}$ ) obtained from OMM series.** For the three components of the magnetic field at the Kakioka observatory, superimposed with  $-4$  and  $-1$  slopes (black). The series have been corrected for referenced jumps (Chulliat and Telali 2007). A multitaper analysis has been performed. For each sub-series, we removed the end-to-end linear trend before applying a Hanning window (see De Santis et al. 2003). Generating many realizations of 100-year synthetic AR-2 series (for which we know the theoretical  $-4$  slope of the spectrum), we retrieve with the above method a slope of  $-4.2 \pm 0.3$  (i.e., a typical error of about 5% to 10%) in the period band from 2 to 25 years.

within the Matérn family of functions. It is characteristic of the second-order statistics for the stationary, auto-regressive (AR) process of order 2 (Yaglom 1962)

$$\frac{d^2\varphi}{dt^2} + \frac{2}{\tau_c} \frac{d\varphi}{dt} + \frac{\varphi}{\tau_c^2} = \epsilon(t), \quad (3)$$

where  $\epsilon(t)$  is a white noise process<sup>a</sup>. We use the correlation time  $\tau_c = 1,730$  years and a typical standard deviation of  $\sigma = 40,000$  nT. This corresponds to the *a priori* function used for the dipole only series by Gillet et al. (2013) - see their Equations 14 to 16. Of course, observatory records contain more than the dipole signal. However, we wish here to keep the definition of  $C(\tau)$  as simple as possible, still allowing the predicted series to potentially display the same continuity properties at monthly periods as they do at interannual periods. We performed a test using a more complex *a priori* function (such as a sum of AR-2 processes, to account for higher order spherical harmonic coefficients), and found that it did not significantly change the local estimate of residuals to the AR-2 fit (this is because we consider all observatories separately, and not all together through a global model). We estimate, at all sites, the BLUE (1) separately for all three components of  $\mathbf{B}$ . We note  $\mathbf{B}_{AR2}$  the stochastic regression of the observed series  $\mathbf{B}$ . We discuss in Section ‘On the sensitivity of the residuals to the Gaussian process regression’ the sensitivity of the variances of the residuals ( $\mathbf{B} - \mathbf{B}_{AR2}$ ) to the *a priori* error variances entering  $C_{ee}$ .

### Local analysis of annual and semiannual signals

Annual and semiannual signals clearly appear in the OMM series and indices (e.g., Lyatsky and Tan 2003). This is illustrated by the peaks in the power spectral density shown in Figure 2. They essentially result from ionospheric and magnetospheric fields (plus their induced counterparts) and exist worldwide in observatory records (see for instance Wardinski and Mandea 2006). Their short latitudinal extent at high latitude nevertheless prevents them from projecting entirely onto the large length-scale model estimated in Section ‘Spherical harmonic modeling of remaining large length-scale fields’, which is why we treat them separately at this stage of our empirical analysis.

In the construction of comprehensive field models, periodic signals originating from ionospheric fields are accounted for through a fit of global functions in quasi-dipole coordinates, calibrated *a priori* in time on the solar radiation flux index  $F_{10.7}$  (see Sabaka et al. 2002). However, this index does not account for the whole modulation of periodic signals. Indeed, the underlying mechanisms responsible for annual variations are subject to discussion and may vary depending on the geographical site (Malin and Winch 1996). Malin and Isikara (1976) suggested that

they are generated by seasonal changes in the morphology of the auroral electrojet and the ring current system and imply variations of conductivity in the ionosphere (see for instance Liu et al. 2009).

Variations with a period of half a year are related to a larger occurrence and intensity of geomagnetic storms and substorms at equinoxes than at solstices (e.g., Cliver et al. 2000). For a long time, semiannual variations have been linked to the largest geomagnetic activity, occurring when the southward component of interplanetary magnetic activity is at a maximum in geocentric solar magnetospheric coordinates (Russell and McPherron 1973). A more complex picture has emerged since then. The latter mechanism is now believed to explain only part of the observed signal, and a more important role is given to the equinoxial hypothesis (see for instance Cliver et al. 2000; Lyatsky et al. 2001), by which the angle between the Earth dipole axis and the solar wind direction governs the efficiency in the response of the magnetosphere to solar wind flow.

Given the complex origin of periodic signals, we favor an empirical approach where the amplitude and phase of the periodic signal are kept completely free, along the lines of the studies by Le Mouél et al. (2004a,b). These authors calculated the Fourier coefficients of annual and semiannual periodic signals from band-pass filtered observatory hourly means. Their analysis indicates that both the amplitude and phase of these signals are changing with time. A similar result was found from wavelet analysis of the semiannual signal in the *aa* magnetic index (Elias et al. 2011).

Here, we fit simple sine/cosine functions with periods of 6 and 12 months to the remaining residuals  $\delta\mathbf{B} = \mathbf{B} - \mathbf{B}_{AR2}$ . Since neither the amplitude nor the phase are constant, these periodic signals are fitted using a sliding window with a restricted length  $\delta t = 5$  years. Minimizing

$$\sum_{|t_j - t_m| \leq \frac{\delta t}{2}} \left[ \delta X(t_j) - \hat{X}_1(t_m) \cos(\omega_1 t_j + \phi_1(t_m)) - \hat{X}_2(t_m) \cos(\omega_2 t_j + \phi_2(t_m)) \right]^2, \quad (4)$$

we obtain the slowly varying functions  $\phi_1(t)$ ,  $\phi_2(t)$ ,  $\hat{X}_1(t)$ , and  $\hat{X}_2(t)$ , with  $\omega_1 = \omega_2/2 = 2\pi$  rad/year (we use similar notations for the *Y* and *Z* components). This defines the annual and semiannual signals  $\mathbf{B}_1(t)$  and  $\mathbf{B}_2(t)$ , whose characteristics are presented in Section ‘Worldwide description of annual and semiannual signals’.

### Spherical harmonic modeling of remaining large length-scale fields

Some of the remaining residuals are due to resolvable external and induced magnetic fields at large length-scales. We approximate them by means of global modeling, constructing a discrete set of monthly models. These

are considered independent from one epoch to another. Indeed, the power spectrum observed for external contributions, excluding annual and semiannual peaks, is only slightly red (with a slope about  $-1$ , see Figure 2).

In principle, external and internal sources can be separated from vector observations on a sphere. However, this separation is ambiguous because of the non-global coverage of observatory sites. Nevertheless, horizontal and vertical components of the field do probe complementary combinations of internal and external sources. Indeed, let us assume that the remaining large length-scale magnetic field  $\mathbf{B}_{LLS}$  derives from a potential,

$$V_{LLS}(r, \theta, \phi, t) = a \sum_{n=1}^N \sum_{m=0}^n \left( \left( \frac{a}{r} \right)^{n+1} g_n^m(t) + \left( \frac{r}{a} \right)^n q_n^m(t) \right) \times P_n^m(\theta) e^{im\phi} + c. c., \quad (5)$$

where  $g_n^m$  and  $q_n^m$  are the complex internal and external Gauss coefficients,  $N$  is the truncation level,  $P_n^m$  is the Schmidt semi-normalized Legendre polynomials of degree  $n$  and order  $m$ , and ‘*c. c.*’ stands for complex conjugate. Then at the Earth’s surface, of radius  $r = a$ , one has (see Section 2.1 in Olsen et al. 2010a)

$$\begin{cases} X_{LLS} = \sum_{n,m} h_n^m(t) \frac{dP_n^m}{d\theta} e^{im\phi} + c. c. \\ Y_{LLS} = \sum_{n,m} h_n^m(t) \frac{im}{\sin\theta} P_n^m e^{im\phi} + c. c. \\ Z_{LLS} = \sum_{n,m} v_n^m(t) P_n^m e^{im\phi} + c. c. \end{cases}, \quad (6)$$

with the complex coefficients

$$\begin{cases} h_n^m(t) = g_n^m(t) + q_n^m(t) \\ v_n^m(t) = (n+1)g_n^m(t) - nq_n^m(t) \end{cases}. \quad (7)$$

Below,  $h_n^{mc}$  and  $h_n^{ms}$  stand for the real and imaginary part of the complex coefficient  $h_n^m$  (with similar notation for  $v_n^{mc}$  and  $v_n^{ms}$ ). We will thus model *X* and *Y* data on the one hand, and *Z* data on the other, resorting to a least-squares fit. We emphasize that in this context, what will be isolated is a combination of global-scale fields that may arise from sources both external (e.g., magnetospheric) and internal (e.g., induced or remaining core signals) to the observation sites.

At each epoch  $t$  spanning 1930 to 2010 every month, we store residuals  $\mathbf{B} - (\mathbf{B}_{AR2} + \mathbf{B}_1 + \mathbf{B}_2)$  at all available observatories in two data vectors  $\mathbf{y}_h(t) = \{X(t), Y(t)\}$  and  $\mathbf{y}_v(t) = \{Z(t)\}$ . The number of available observatories ( $N_x, N_y, N_z$ ) for the three components are a function of time (cf. Figure 1). Then, from Equations 5 and 6, we write at every epoch  $t$ :

$$\begin{cases} \mathbf{y}_h = \mathbf{A}_h \mathbf{h} + \mathbf{e}_h \\ \mathbf{y}_v = \mathbf{A}_v \mathbf{v} + \mathbf{e}_v \end{cases}. \quad (8)$$



The unknown vectors  $\mathbf{h}(t)$  and  $\mathbf{v}(t)$ , of size  $P=N(N+2)$ , contain, respectively, the coefficients  $h_n^m(t)$  and  $v_n^m(t)$ , whose statistics are defined by the *a priori* model covariance matrices  $R_h = E(\mathbf{h}\mathbf{h}^T)$  and  $R_v = E(\mathbf{v}\mathbf{v}^T)$ . Matrices  $A_h$  and  $A_v$ , of size, respectively,  $(N_x + N_y) \times P$  and  $N_z \times P$ , translate into matrix form the linear relation (Equation 6) between the data vectors ( $\mathbf{y}_h$  and  $\mathbf{y}_v$ ) and the vectors of model coefficients ( $\mathbf{h}$  and  $\mathbf{v}$ ). Vectors  $\mathbf{e}_h$  and  $\mathbf{e}_v$  stand for the errors in the horizontal and vertical components, whose statistics are given by the covariance matrices  $C_h = E(\mathbf{e}_h\mathbf{e}_h^T)$  and  $C_v = E(\mathbf{e}_v\mathbf{e}_v^T)$ . The Bayesian average solution is then

$$\hat{\mathbf{h}} = H_h A_h^T C_h^{-1} \mathbf{y}_h, \tag{9}$$

with  $H_h = [A_h^T C_h^{-1} A_h + R_h^{-1}]^{-1}$  the covariance matrix defining the *a posteriori* error statistics of the model  $\mathbf{h}$  (with similar definitions for  $\mathbf{v}$ ). From  $\hat{\mathbf{h}}$  and  $\hat{\mathbf{v}}$ , we build the large length-scale field vector  $\mathbf{B}_{LLS}$ .

The quantities entering the *a priori* data error and model error covariance matrices are *a priori* unknown. For the sake of simplicity, we consider uncorrelated errors with stationary statistics, i.e.,  $C_h = e_h^2 I_h$  at every epoch ( $I_h$  stands for the identity matrix of rank  $N_x + N_y$ ), with similar notation for  $C_v$ . Of course, errors are probably increasing towards the past, when records were less accurate. However, we tested several values for  $e_h^2$  and  $e_v^2$  and found that within a reasonable range, the final estimate of  $\mathbf{B}_{LLS}$ , and then of the UMSL, was not significantly affected. The results presented in Sections ‘Main patterns of remaining external and induced fields’ to ‘The UMSL throughout 1930 to 2010’ are obtained using  $e_h = e_v = 6$  nT. In practice, we fix the truncation level to  $N = 3$ , which results in 15 unknowns per epoch. This is much smaller than the number of observatories after 1960 but not necessarily in the first half of the twentieth century.

If we were retrieving  $\mathbf{B}_{LLS}$  from Equation 8 in geographic coordinates, the model cross-covariances would vary with time depending on the dipole orientation. Thus, in order to solve Equation 8 using stationary variance properties, we rotate at each epoch the forward equations in dipole coordinates (for which we use the *COV-OBS* internal dipole field). Thus, we build matrices  $R_h$  and  $R_v$  from the variances of the coefficient series  $h_n^m(t)$  and  $v_n^m(t)$  in dipole coordinates, first obtained over the period 1960 to 2010 for which the model is weakly sensitive to the *a priori* matrices (we would underestimate the variances of the model parameters by considering the era before 1960, where fewer, less accurate data are available). Coefficients are then rotated back into geographic coordinates.

To finally obtain the UMSL presented in Section ‘The UMSL throughout 1930 to 2010’, we consider (i) the residuals  $\delta\mathbf{y}_h = \mathbf{y}_h - A_h \hat{\mathbf{h}}$  and  $\delta\mathbf{y}_v = \mathbf{y}_v - A_v \hat{\mathbf{v}}$ , i.e., the final

residual series  $\mathbf{B} - (\mathbf{B}_{AR2} + \mathbf{B}_{LLS} + \mathbf{B}_1 + \mathbf{B}_2)$ , and (ii) the modeling errors arising from the uncertainties of the large length-scale field parameters. We thus estimate, for each epoch  $t$ , the UMSL at all observatories as the square root of the diagonal elements of the covariance matrices

$$K_h(t) = A_h(t) H_h(t) A_h(t)^T + \delta\mathbf{y}_h(t) \delta\mathbf{y}_h(t)^T \tag{10}$$

and  $K_v(t)$  (with similar notation).

## Results and discussion

### On the sensitivity of the residuals to the Gaussian process regression

In this section, we test the sensitivity of the variance of the residuals  $\delta\mathbf{B} = \mathbf{B} - \mathbf{B}_{AR2}$ , obtained using the BLUE (Section ‘Local stochastic regression of the main field’), to the *a priori* noise amplitude that enters the covariance matrix  $C_{ee}$  in Equation 1. We first study the case of synthetic data, before we analyze the geophysical series.

We generate monthly values of a noisy AR-2 process over a time window  $\Delta T = 50$  years. We consider first the case of stationary white noise, varying the *a priori* noise variance that enters  $C_{ee}$  from 1/100 to 100 times the true noise variance. We tested two values for the dimensional r.m.s. of the true noise, 10 and 100 nT, corresponding to a noise over signal ratio of 0.33% and 0.033%, respectively. The ratio between the magnitudes of the residuals to the BLUE and of the true noise, summarized in Table 2, remains close to unity in all cases. Given the range of *a priori* values, we conclude that we are able to retrieve well the true noise amplitude in this synthetic case. We checked that this result is not sensitive to the window length  $\Delta T$ . We also consider a situation that mimics a quality jump (such as the advent of proton magnetometers), where the noise magnitude is ten times larger over the first half of the window (case C in Table 2). Again, we correctly retrieve the magnitude of the true noise, over both halves of the time window.

**Table 2 Ratio between the true and retrieved noise magnitude when varying the prior noise variance**

Case	Prior / true noise variance ratio				
	1/100	1/10	1	10	100
A	0.91	0.95	0.98	1.01	1.16
B	0.97	0.98	0.99	1.00	1.04
C <sub>a</sub>	0.94	0.97	0.99	1.00	1.00
C <sub>b</sub>	0.94	0.97	0.99	1.01	1.11

From 1/100 to 100 times the true noise variance, for 50 years of synthetic, noisy AR-2 series with magnitude 30,000 nT (result obtained after averaging over 20 realizations). In cases A and B, the magnitudes of the true noise are 10 and 100 nT, respectively. In case C, it is 100 nT over the first half of the time window, then 10 nT (there, the *a priori* over true ratio are that of the second half, i.e., the noise has a *a priori* the same magnitude over the whole window); the *a posteriori* ratios are calculated separately over the two halves, labeled C<sub>a</sub> and C<sub>b</sub>, respectively.

We now apply the same type of fit to the geophysical series. Examples of residual magnitudes obtained for the  $X$ ,  $Y$ , and  $Z$  components for several observatories are displayed in Table 3. We used values for the *a priori* error variance  $\eta$  (the diagonal elements of the matrix  $C_{ee}$ ) of 10, 100, and 1,000 nT<sup>2</sup>. Given the standard error bars generally associated with observatory series, by using the smaller (larger) values of  $\eta$ , we supposedly under- (over-) estimate the signal that cannot enter the AR-2 fit. We observe that the *a posteriori* standard deviation of the residuals is only weakly sensitive to the *a priori* choice for  $\eta$  in the range of 10 to 1,000 nT<sup>2</sup>. In comparison, when using a polynomial fit instead of the BLUE, we obtain (not shown) residual variances that are much more sensitive to the cutoff frequency (the number of polynomials) for both synthetic and geophysical series. We conclude that the method presented in Section ‘Local stochastic regression of the main field’ is relatively robust for estimating the variance of the residuals of the main field regression.

As expected, the amplitude of the residuals is weaker for the  $Y$  component, which is less affected by the magnetospheric ring current. It is also larger for  $X$  in the equatorial area (see for instance Alibag or Guam in Table 3). We observe larger values towards higher geomagnetic colatitudes (for instance at Sitka), where the magnetic fields of ionospheric origin associated with the auroral ring currents are known to produce signals of particularly large amplitude. From now on, we use an *a priori* noise variance  $\eta = 36$  nT<sup>2</sup>. Examples of series of the residuals between the observed field and the AR2 fit are presented in Figures 3, 4, and 5 for all three components at equatorial, mid-, and high latitude sites.

### Worldwide description of annual and semiannual signals

We concentrate first on the time evolution of the amplitude and phase of the periodic signals. These are displayed in Figure 6, where we filter out the effect of the solar cycle

**Table 3 Dimensional r.m.s. of the residuals to an AR-2 fit at several observatory sites**

$\sqrt{\eta}$	(nT)	3.3	10.0	33.3	$\sqrt{\eta}$	(nT)	3.3	10.0	33.3
	X	9.3	10.1	11.9	X	8.1	8.6	9.6	
KAK	Y	2.5	2.9	4.3	HER	Y	1.8	2.6	4.5
	Z	7.3	8.2	10.4	Z	3.2	3.4	6.6	
	X	8.1	8.4	9.1	X	13.7	14.3	15.0	
ESK	Y	4.1	4.5	5.3	SIT	Y	6.1	6.4	7.4
	Z	9.4	11.0	13.1	Z	13.5	14.9	16.4	
	X	11.0	12.5	15.2	X	10.6	11.2	12.7	
ABG	Y	4.4	4.9	6.4	GUA	Y	1.9	2.2	3.5
	Z	8.7	9.7	13.0	Z	3.4	4.3	8.4	

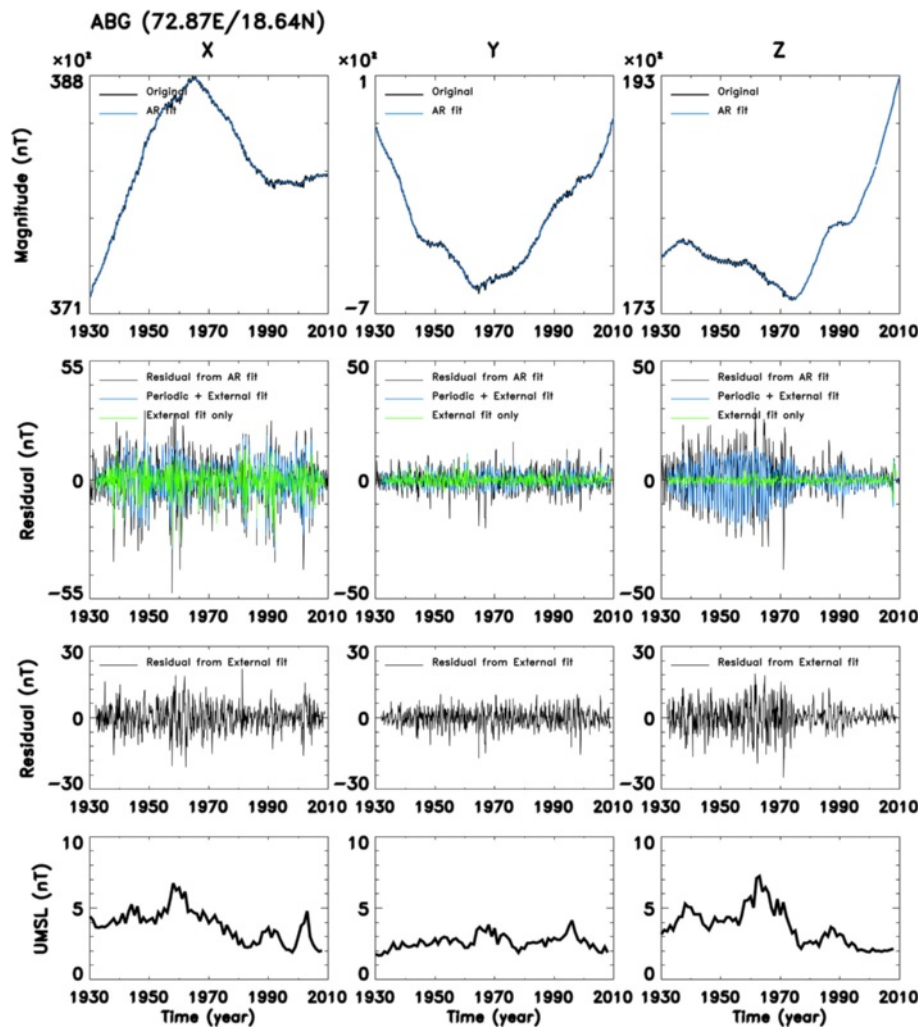
To the  $X$ ,  $Y$ , and  $Z$  components (in nT), for several values  $\eta$  of the *a priori* noise variance.

using an 11-year running mean. We find trends similar to those computed from a Fourier analysis of filtered hourly means (see Figures four to five and eight to nine in Le Mouël et al. 2004a). In particular, we retrieve the two large variations, at pluri-decadal time scales, in the amplitude of the semiannual signal. The epochs at which these extrema occur are similar to those found from the analysis of the *aa* index by Elias et al. (2011). These variations are in phase at all observatories. The phase of the semiannual signal also shows decadal changes, in agreement with those found from *aa* (see Figure seven in Elias et al. 2011), with small delays from one site to another.

We also see decadal changes in the phase and amplitude of the annual signal. These variations vary depending on the location, the phase showing, for instance, shifts as large as a couple of months. Several origins have been proposed for annual changes, including local effects of induction in the ocean, temperature dependency of the soil magnetization around sensors (Mishima et al. 2013), or spurious temperature effects in the  $Z$  component before the introduction of fluxgate sensors (Le Mouël et al. 2004b). Some effects related to measurement issues cannot be discarded from early epochs. Indeed, Figure 6 indicates that annual signals display less dispersion among observatories towards more recent periods.

We also observe differences when comparing our results to those of Le Mouël et al. (2004a), which can be attributed to several factors. We present in Figure 7 the impact of data selection (night only versus all-hours values) and of the window length  $\delta t$ , for the Sitka and Tucson observatories (some differences may also be related to the filtering strategy used by Le Mouël et al. 2004a). At high geomagnetic latitudes, the amplitude of semiannual oscillations (and to a lesser extent, of annual ones) is much increased by the use of a shorter window length  $\delta t$  (with weak impact on the phase), whereas this parameter has a minor impact at low to mid-latitude sites. Selecting nighttime data generates a downward (upward), almost constant bias of about 2 nT for the 6 (12) months signal amplitude at mid-latitudes, and a positive phase delay of about 15 days for the annual signal, when the phase of the semiannual signal is not much affected. Compared to all-hours data, using nighttime data also enhances the amplitude of the peaks in the 6-month variation at high latitudes and generates a time-dependent phase shift between zero and about ten days. It also shifts the amplitude of the annual signal by a few nT, due to aliasing, while its phase is not significantly affected.

The geographical dependence of the amplitude of annual and semiannual periodic signals can essentially be summarized as a function of geomagnetic colatitude, as shown in Figure 8. The amplitude of semiannual changes shows a large maximum at geomagnetic latitudes between 60° and 70° in all three components. It is particularly



**Figure 3** Residuals to OMM series and the UMSL: example of a low-latitude site. Top to bottom: OMM series superimposed with AR-2 fit at the Alibag observatory; residuals between OMM series and the AR-2 proxy, superimposed with the LLS field model with and without periodic signals; final residuals after removing the LLS field; our estimate of the UMSL. Left to right: X, Y, and Z components.

strong in X, which also displays a secondary maximum towards the geomagnetic equator. In the polar regions, the semiannual signal is larger in Z and weaker in X (note that the amplitude in Y can be significantly larger than that found in X at some sites). Such features could be associated with an increase of the ring currents and the westward auroral electrojet at equinoxes (see Lyatsky and Tan 2003). The latitudinal distributions observed in the X and Z components recall those found by Fujii and Schultz (2002) for the residuals between observatory data and the induction response to a  $P_1^0$  source at periods ranging from a few days to a few months (see their Figure nine). These authors associate such features mainly with the complementary effect of auroral ring currents (cf. their Figure fourteen).

The annual signal is generally larger in X, and its amplitude strongly increases at latitudes above 60°. A secondary

maximum clearly appears in X at mid-latitudes. Also at mid-latitudes, non-zonal contributions are observed in the X and Y components, which may be associated with regional induction effects, as suggested by Wardinski and Mandea (2006). We note some discrepancies between the latitudinal dependence of annual and semiannual signals (see for instance in X). This may indicate that the source fields responsible for annual changes are different from those proposed by Fujii and Schultz (2002).

**Main patterns of remaining external and induced fields**

Examples of the contributions from periodic and large length-scale fields to the three components of observatory series are shown, from low to high latitudes, in Figures 3,4 and 5. In the horizontal components, modulations with the solar cycle are observed for the large length-scale field at mid- to low latitudes (more obvious in the X

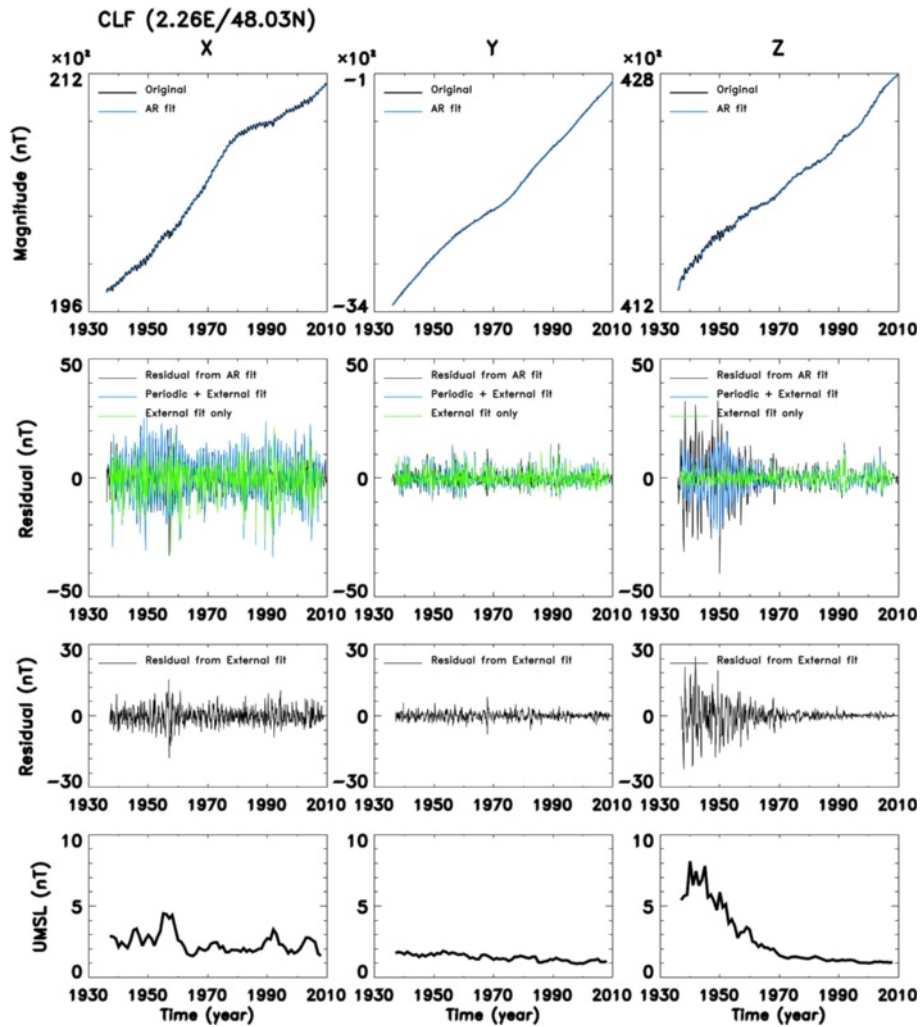


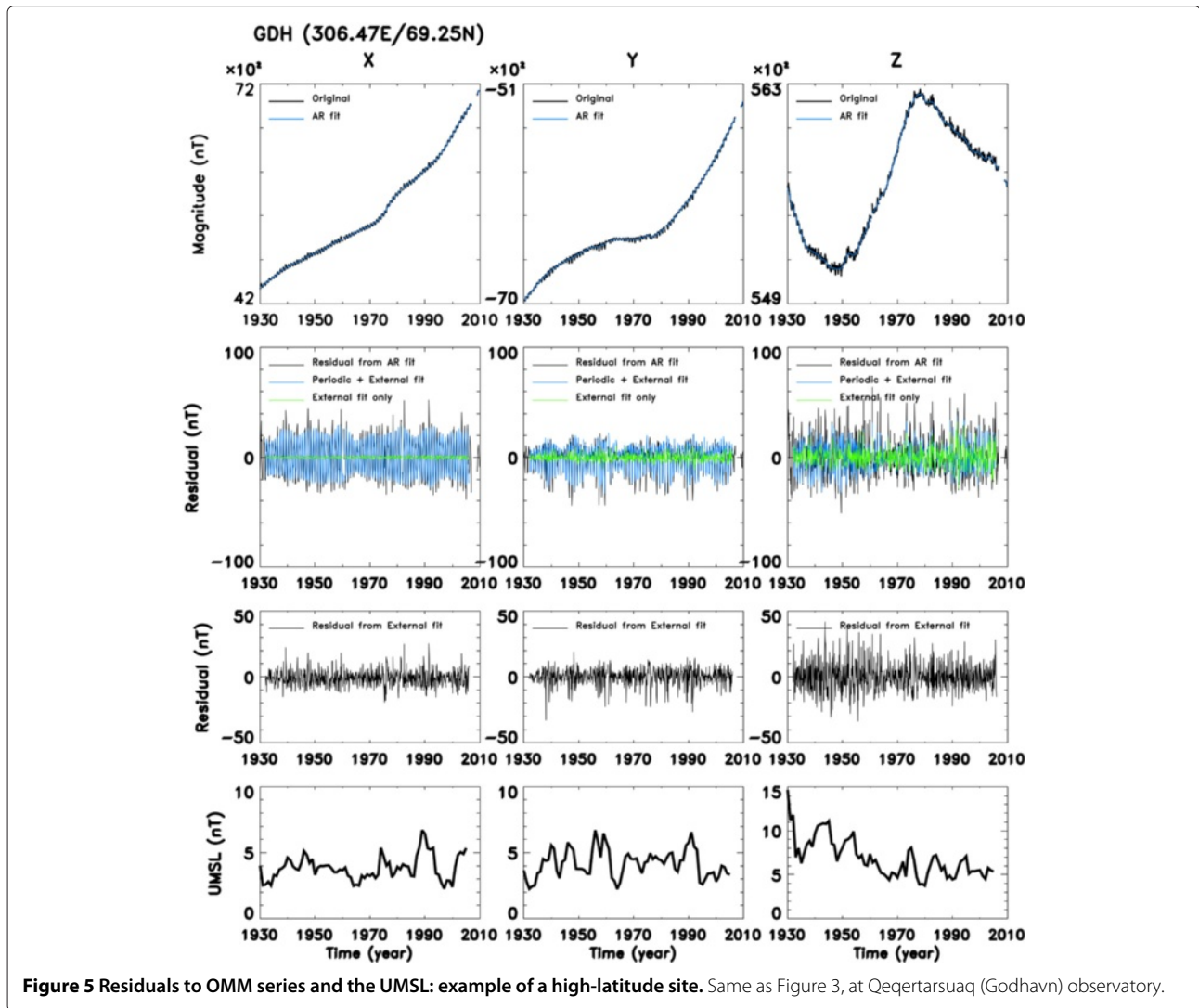
Figure 4 Residuals to OMM series and the UMSL: example of a mid-latitude site. Same as Figure 3, at Chambon-la-forêt observatory.

direction), and for periodic signals, in horizontal components at high latitudes. The modulation of periodic signals at longer time scales, highlighted in the previous section, also clearly appears.

We use below the notations  $\tilde{h}_n^m$  and  $\tilde{v}_n^m$  for the coefficients in dipole coordinates. We present in Figure 9 their associated (normalized) series, together with the standard deviation obtained for each of the coefficients. In association with the magnetospheric ring current and its induced counterpart, the largest contribution to the X and Y components comes from  $\tilde{h}_1^0$ , while to Z, it is shared between  $\tilde{v}_1^0$  and  $\tilde{v}_3^0$ . The important contribution of  $\tilde{v}_3^0$  to the vertical component is associated with the well-known larger impact of externally induced fields on Z (see for instance the discussion of Figure ten in Hulot et al. 2007). Zonal coefficients are generally better resolved than the non-zonal ones, and sectorial coefficients are associated with relatively larger error bars (see the  $\tilde{h}_2^2$  and  $\tilde{v}_3^3$  series). We

also notice weaker uncertainties from 1960 onward (for instance  $\tilde{h}_2^{1c}$ ,  $\tilde{v}_1^{1s}$ ,  $\tilde{v}_2^{1s}$ , and  $\tilde{v}_3^{2s}$ ). The *a posteriori* errors we obtain exclude the occurrence of large variations for some coefficients at particular epochs (see the  $\tilde{h}_n^0$ ,  $\tilde{h}_1^1$ , or  $\tilde{h}_3^1$  coefficients in the early 1930s, a period of relatively weak solar activity).

The signature of the solar activity is clear for most coefficients (see the  $K_p$  index in Figure 10). In the 1960s and around 1990, periods of particularly intense solar cycles, we see unusually large fluctuations of coefficients, showing at other epochs a relatively modest dynamics (see the  $\tilde{v}_1^1$  and  $\tilde{v}_2^1$  coefficients after 1960). This suggests a particular response of the external (and then induced) fields to the solar flux at those epochs. We note that many observatories started with the International Geophysical Year 1957 to 1958 (cf. Figure 1), some of them having displayed unstable magnetic series. We corrected as much as possible for anomalous data during this period as well



**Figure 5** Residuals to OMM series and the UMSL: example of a high-latitude site. Same as Figure 3, at Qeqertarsuaq (Godhavn) observatory.

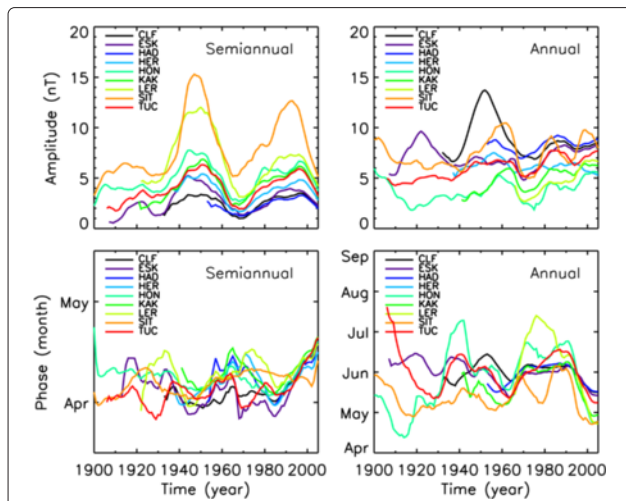
(see Section ‘Data’). However, if the remaining outliers partly enter the UMSL, they may also induce biases in the global model coefficients of the least-squares fit (a drawback which may be reduced by using a Huber norm for the measure of residuals, see e.g., Olsen 2002).

The combination of  $\tilde{h}_n^m$  and  $\tilde{v}_n^m$  series is used to separate internal and external contributions; we obtain from Equation 7 the complex coefficients

$$\begin{cases} \tilde{s}_n^m = \frac{\tilde{h}_n^m + \tilde{v}_n^m}{2n+1} \\ \tilde{q}_n^m = \frac{(n+1)\tilde{h}_n^m - \tilde{v}_n^m}{2n+1} \end{cases} \quad (11)$$

(here in dipole coordinates). Their analysis as a function of frequency may be used to constrain C responses and the mantle conductivity profile (see e.g., Olsen 1999), which is out of the scope of the present study. We concentrate below on the description of the inducing field. From Equations 6 and 7, we obtain the coefficients describing the vertical component Z from external origin,

$\tilde{z}_n^m = -n\tilde{q}_n^m$ , which compose the time-dependent vector  $\mathbf{z}(t)$ . From  $S = [\mathbf{z}(1930.0) \dots \mathbf{z}(2010.0)]$ , we build the matrix  $M = SS^T$  and perform on it a principal component analysis. We show in Figure 11 the vertical component obtained at the Earth’s surface for the first four eigen modes  $\mathcal{Z}_1$  to  $\mathcal{Z}_4$ . These account for, respectively, 70.6%, 9.4%, 5.2%, and 4.5% of the variance of  $\mathbf{z}$ . Most of the variability is carried by  $\mathcal{Z}_1$ , which is associated with the axial dipole field (in dipole coordinates), although it is also composed of some non-axisymmetric features.  $\mathcal{Z}_2$  essentially summarizes the role played by the zonal coefficient of degree 3. We present in Figure 10 the projection of the  $\tilde{z}_n^m$  coefficients onto the first four eigen modes.  $\mathcal{Z}_3$  (with large contributions from  $\tilde{z}_1^{1s}$ ,  $\tilde{z}_2^{0c}$ , and  $\tilde{z}_2^{1s}$  coefficients) and  $\mathcal{Z}_4$  (dominated by a sectorial structure of degree 2) present unusual time changes in the 1960s and around 1990, in connection with the above observation made about the original  $\tilde{v}_n^m$  and  $\tilde{h}_n^m$  series. This illustrates that the structure of the inducing field is more complex



**Figure 6** Time changes of the amplitude and phase of periodic signals at various sites. Amplitude (top) and phase (bottom) obtained for the annual (right) and semiannual (left) periodic signals for the X component at several observatories, for a window length of  $\delta t = 5$  years, after filtering with an 11-year running mean. The phase of the periodic signal (originally in radians) is translated into the date at which occurs the first maximum of the annual (or semiannual) fit to the data (for each epoch, a phase of zero would correspond to January 1).

than the ring current, as discussed for instance by Olsen (1998) for periods shorter than 30 days, or by Fujii and Schultz (2002) for periods ranging from 5 to 106.7 days.

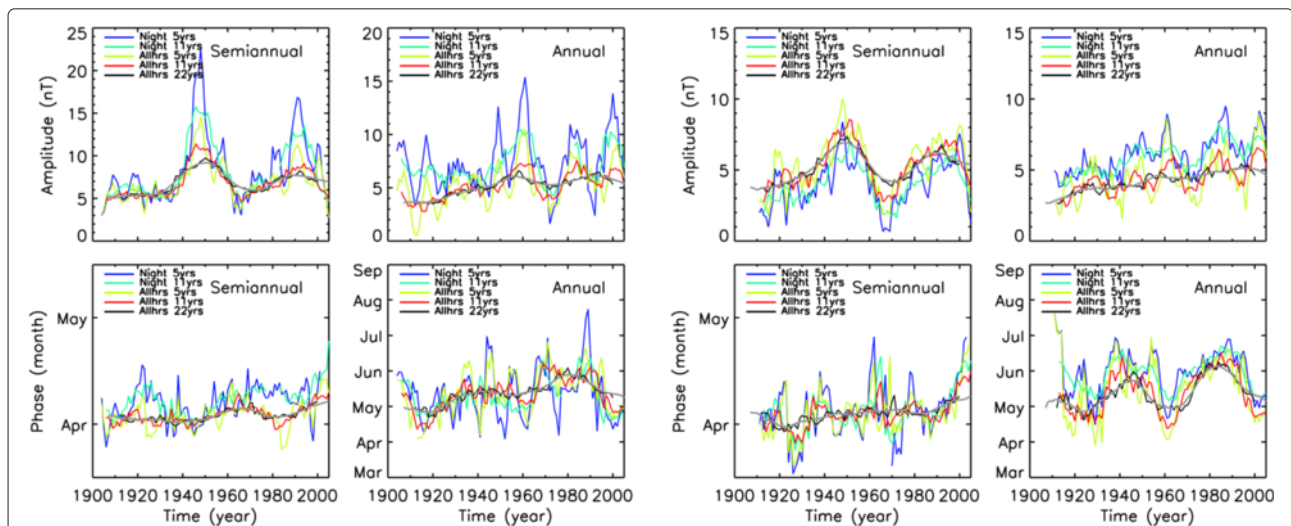
**The UMSL throughout 1930 to 2010**

We finally focus on the evolution, in space and time, of the UMSL as defined by Equation 10. Two main sources

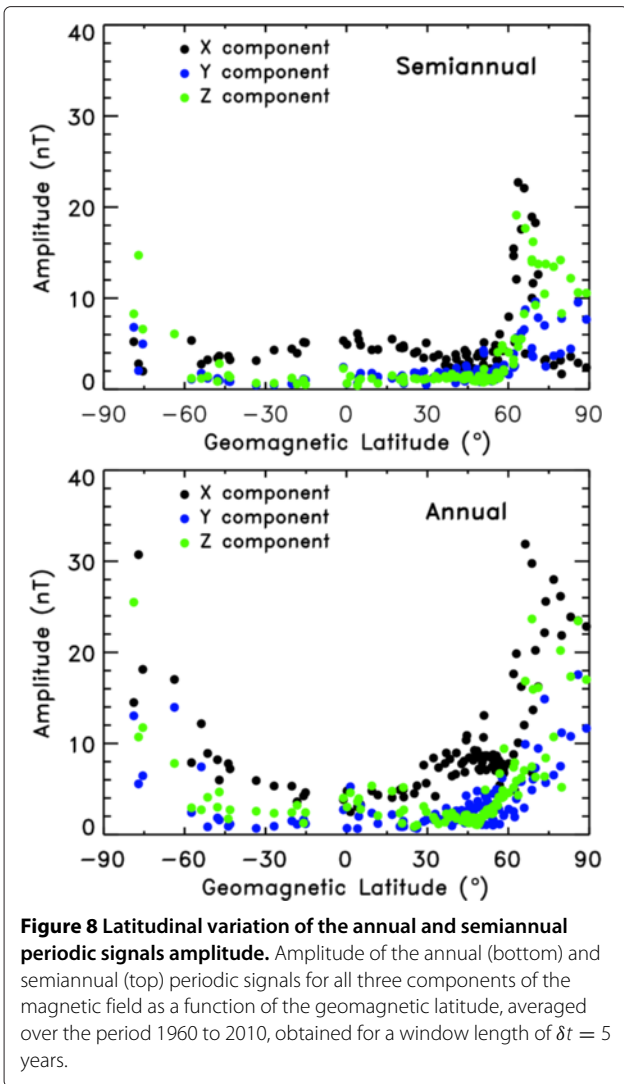
contribute to the UMSL: instrumental noise and unmodeled regional signals. Since the number of available observatories becomes limited for retrieving large-scale external coefficients from the very early twentieth century, we present our UMSL estimate from 1930 onward. Examples of the final residuals and of the obtained UMSL series are presented in Figures 3,4 and 5.

In several observatories (see for instance the series for Alibag or Chambon-la-Forêt in Figures 3 and 4), periodic signals in the Z component tend to decrease towards recent epochs, which coincides with a decrease of the UMSL in Z. A link to instrumental effects is suspected, such as the introduction of proton magnetometers for absolute measurements after 1960 (Turner et al. 2007) and of fluxgate sensors for vector variometry, as suggested by Le Mouël et al. (2004b). Indeed, measurement improvements have been recorded by Chulliat and Telali (2007) at this particular time at Alibag and Chambon-la-Forêt. This is also illustrated by Figure four in Gillet et al. (2010) and Figure fourteen in Gillet et al. (2009), where an important decrease is observed for the scatter in the difference between two Z series at nearby observatories (the difference removing the effect of external fields that are coherent at the two sites).

The three components of the UMSL are presented in Figure 12 as a function of geomagnetic colatitude in 1996 (low solar activity) and 2001 (high solar activity). The most important pattern is the higher UMSL values found in (geomagnetic) polar regions, with a sharp transition between moderate and large values around 55° geomagnetic latitude. We notice only a weak dependence with longitude. The global maps of the UMSL in 1996 and



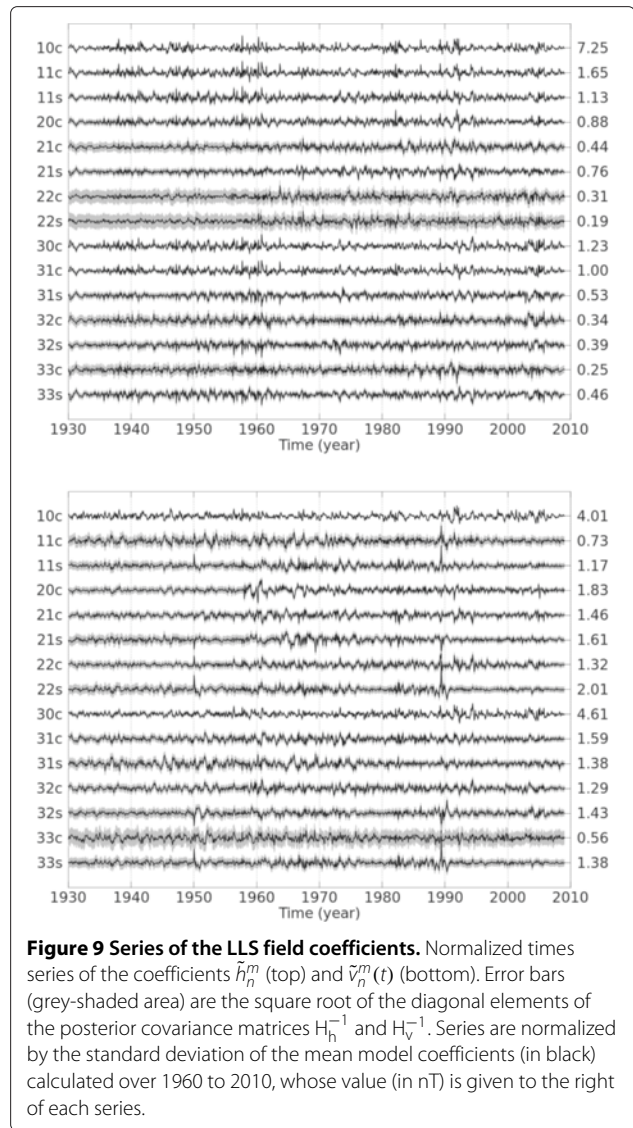
**Figure 7** Sensitivity to the data treatment of the amplitude and phase of periodic signals. Amplitude and phase obtained for annual (right) and semiannual (left) periodic signals for the X component at Sitka (57°N, 225°E, four top panels) and Tucson (32°N, 249°E, four bottom panels), for different data selection (OMM constructed from night only data or from all-hours values) and window lengths  $\delta t$ . The phase of the periodic signal (originally in radians) is translated into the date at which occurs the first maximum of the annual (or semiannual) fit to the data (for each epoch, a phase of zero would correspond to January 1).



2001, presented in Figure 13, illustrate with more detail the geographical distribution of the final residual level.

We encapsulate in Figure 14 the UMSL obtained for the three components at all observatories throughout the selected time span, separately for sites with geomagnetic latitudes lower and higher than  $60^\circ$ . We see in the general trend a clear signature of the solar cycle at high latitudes. It is also there, to a lesser extent, at lower latitudes. On average, the UMSL from the low to mid-latitude sites is similar for all three geomagnetic components (a couple of nT towards the most recent epochs), even though perturbations may be larger at some periods and locations for Z. Depending on the observatory, it ranges between 1 and 10 (6) nT for Z (X and Y) data. We notice a slight decrease of the median UMSL in Z after 1960, while the UMSL in X and Y is rather stationary.

At latitudes higher than  $60^\circ$ , a significant reduction in the UMSL after 1960 is found for Z data (on average, from

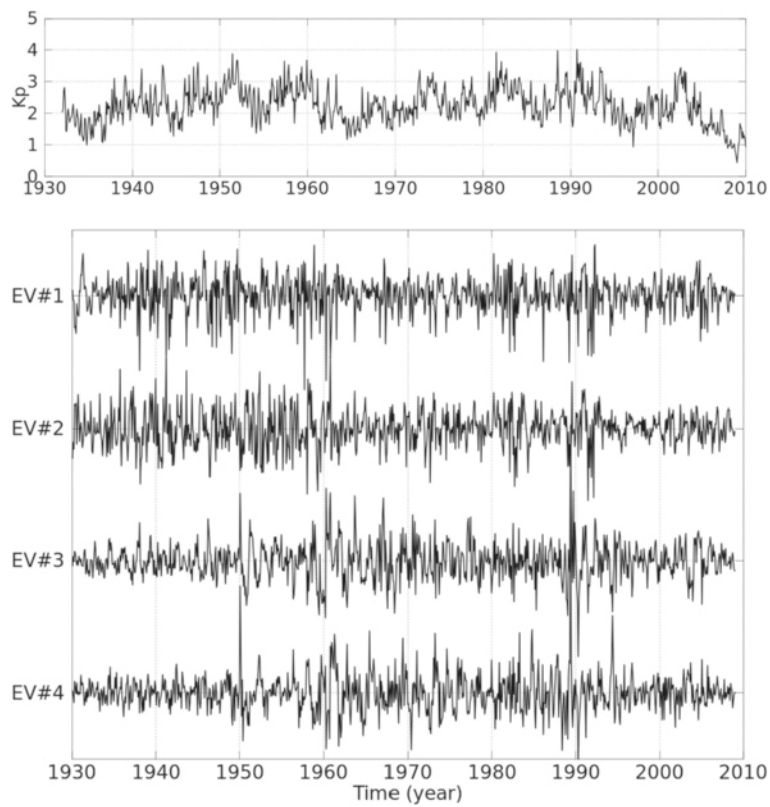


10 to 6 nT). In the series of the horizontal components, the median UMSL is more stationary (around 7 and 4 nT for X and Y, respectively). The decrease observed on average before 1950 is essentially due to the fact that we only have access to a restricted group of sites from this period, which does not include the observatories that display the higher UMSL values later on. Note that some sites display a UMSL as large as 20 nT in X at periods of maximum solar activity, a value much larger than found for Y and Z at recent epochs.

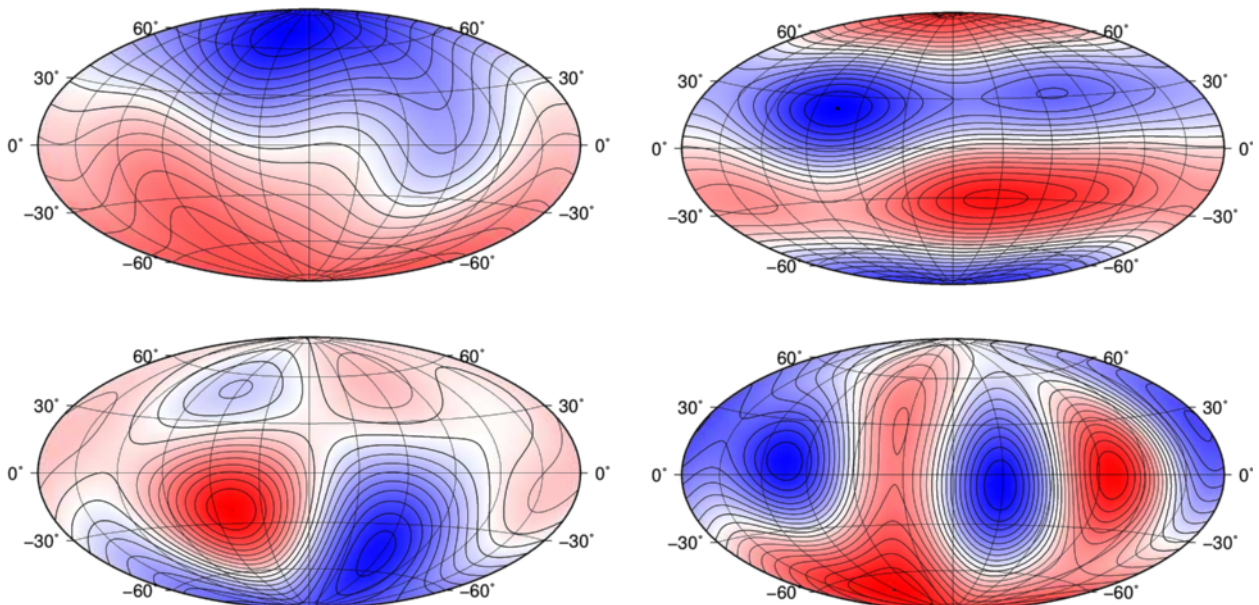
### Conclusions

We present in this study a sequential analysis of OMM series, where we isolated three contributions:

1. A local, stochastic regression for the main field, constructed in the framework of Gaussian processes.

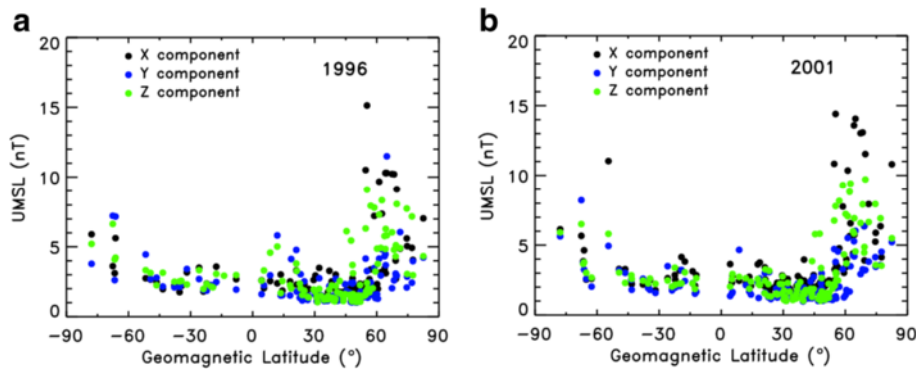


**Figure 10** Projection of the coefficients  $\mathbf{z}(t)$  on the first eigen modes, and comparison with the  $K_p$  index. Top: monthly averages of the  $K_p$  index (see e.g., Menvielle and Berthelier 1991). Bottom: normalized projection of the coefficient vectors  $\bar{z}_n^m$  on the first four eigen modes  $\mathcal{Z}_1$  to  $\mathcal{Z}_4$ , which represent 70.6% (top left), 9.4% (top right), 5.2% (bottom left), and 4.5% (bottom right) of the  $\bar{z}_n^m$  variances.



**Figure 11** Maps of the Z component at the Earth's surface for the first eigen modes. Maps (centered on the Greenwich meridian) of the Z component at the Earth's surface for the first four normalized eigen modes,  $\mathcal{Z}_1$  to  $\mathcal{Z}_4$ , obtained from the  $\bar{z}_n^m$  coefficient series in dipole coordinates. They represent 70.6% (top left), 9.4% (top right), 5.2% (bottom left), and 4.5% (bottom right) of the  $\bar{z}_n^m$  variances. For each mode, the color scale is saturated at the maximum value of the mode.





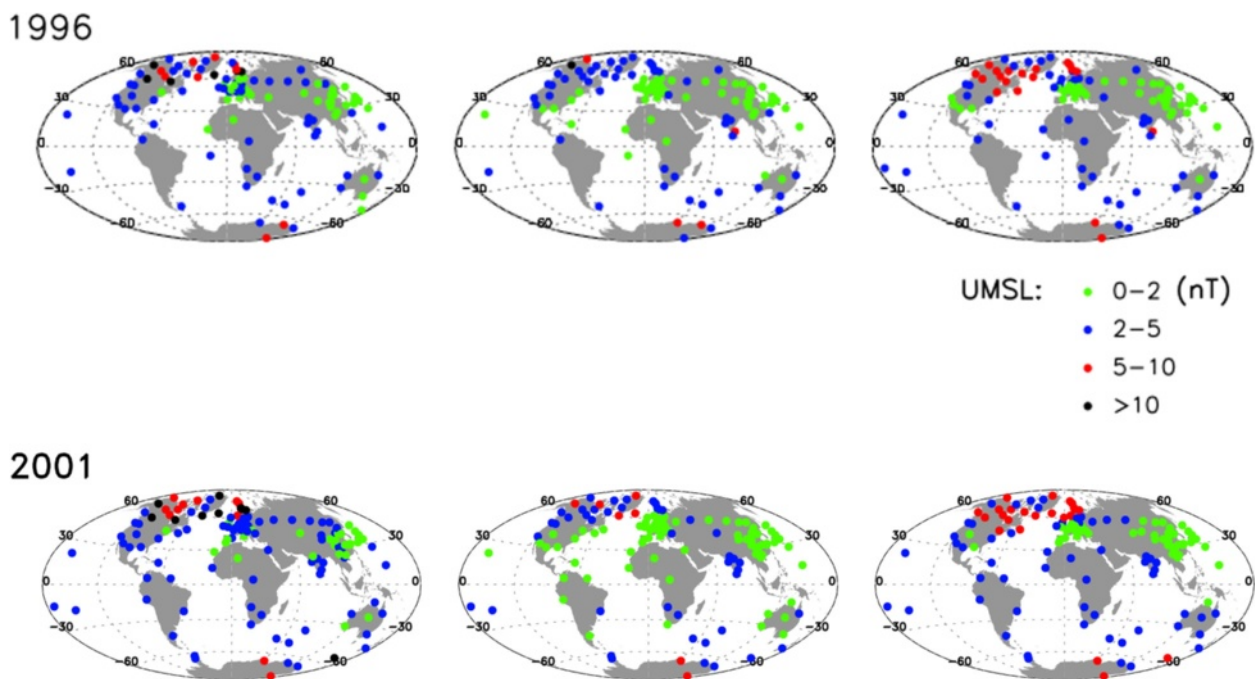
**Figure 12 Latitudinal dependence of the UMSL.** Amplitude of the UMSL in 1996 (a, solar minimum) and 2001 (b, solar maximum) for all three components of the magnetic field, as a function of geomagnetic latitude.

We use an *a priori* function based on a  $-4$  slope temporal power spectral density, observed at periods longer than 5 years and extrapolated for shorter periods. This method appears robust, in the sense that it is only weakly sensitive to the *a priori* data errors that enter the definition of the BLUE in Equation 1.

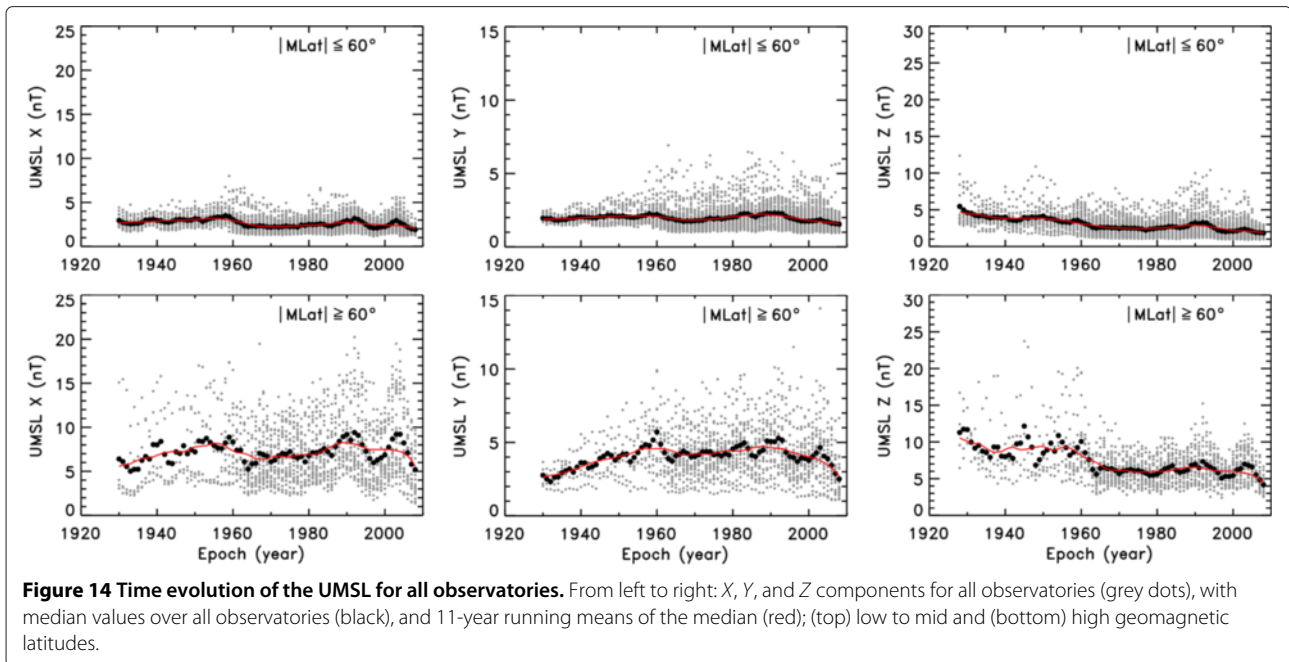
2. A local analysis of annual and semiannual signals. We retrieve their modulation over long time scales, previously put forward by Le Mouél et al. (2004a,b). We suspect that part of these slow variations in amplitude is instrumental in origin, especially in the Z component. The geographic distribution of the

amplitude of periodic signals is essentially zonal in geomagnetic coordinates, presenting a sharp gradient in the auroral zone. We see only a little longitudinal dependence, possibly due to regional induction effects.

3. A global spherical harmonic analysis of the remaining large length-scale fields originating from external and induced sources. In order to avoid the ambiguity between internal and external sources, we fit horizontal and vertical components separately, before reconstructing the internal and external coefficients. We then describe the main patterns of the inducing field in Z, showing a non-negligible role



**Figure 13 Global distribution of the UMSL at observatories.** From left to right: X, Y and Z components, in geographic frame; at solar minimum (1996, top) and at solar maximum (2001, bottom).



**Figure 14** Time evolution of the UMSL for all observatories. From left to right: X, Y, and Z components for all observatories (grey dots), with median values over all observatories (black), and 11-year running means of the median (red); (top) low to mid and (bottom) high geomagnetic latitudes.

of components that are not carried by  $P_1^0$ . We observe an unusual morphology of this field in the 1960s and around 1990, periods of particularly intense solar activity.

We end up with the final residual series, from which we define the UMSL. At recent epochs, it is mainly a function of the geomagnetic colatitude, and its magnitude is rather similar for all three components (at least at mid- to low latitudes). This reflects the standards required to enter the INTERMAGNET network. If some UMSL variations denote changes in the quality of the instrumentation, as clearly seen in the Z component around 1960, part of the UMSL fluctuations is related to the solar activity. This highlights the difficulty of modeling globally the complex magnetic fields at high latitudes from monthly values. A strategy similar to that presented here could be followed for hourly values if one were to also model  $Sq$  variations (see Stening and Winch 2013). This would not only have implications for the analysis of annual and semiannual signals; indeed, it would also require accounting for periodic fields at higher frequencies (see Love and Rigler 2014).

The UMSL could enter the data error covariance matrix in future field modeling studies if one wishes to invert jointly the internal, induced, and external signals at periods shorter than a few years. Indeed, modeling internal and external coefficients through the  $\tilde{h}_n^m$  and  $\tilde{v}_n^m$  associated, respectively, with horizontal and vertical data, it is possible to constrain the mantle conductivity through C responses (Olsen 1999). Furthermore, it seems possible that some core signals may reach the Earth's surface at

short periods (Velínský and Finlay 2011), making the joint inversion of core and induced signals relevant. Since the signature of core flows are relatively weak at short periods, accurate estimates of error covariances are required to assess the reliability of the reconstructed motions. In this situation, it may be of importance to consider the UMSL time changes at every site, since accurate error statistics are required to best extract the information contained in OMM series. For this purpose, one may consider the strategy put forward by Haines (1993) when building data error covariance matrices for differentiated series.

### Endnote

<sup>a</sup>Note that the SDE (3) differs from that defined by Equation (12) in (Gillet et al. 2013). Both SDEs define processes with the same auto-covariance function (2), but their Equation (12), contrary to our Equation (3), corresponds to a non-stationary process. This does not affect their results that have been computed from the covariance function and not the SDE.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

JO carried out all the cleaning of geomagnetic series from hourly values. He performed numerically the several steps of the sequential analysis and proposed to consider separately periodic signals. NG supervised the work of JO at ISTERre. He proposed to use the Gaussian processes framework and the global modeling of residuals in terms of horizontal/vertical components to estimate external fields coefficients. NG also carried out the principal components analysis on the external field coefficients and led the writing of the paper. AD supervised the work of JO at IGGCAS, introduced him to geomagnetic observatory records, and made geomagnetic data available for this study. All authors read and approved the final manuscript.

### Acknowledgements

We thank the national institutions that support ground magnetic observatories and INTERMAGNET for promoting high standards of practice and making the data available. Jiaming Ou's two visits to France were financed by the Institute of Geology and Geophysics, Chinese Academy of Sciences. Discussions with Dominique Jault, Nils Olsen, Chris Finlay, and Jurgen Matzka have helped us throughout the evolution of the present study. We thank the two anonymous referees, whose comments helped improve the quality of the manuscript. This work was supported in part by the French Centre National d'Etudes Spatiales (CNES) for the preparation of the Swarm mission of ESA. This work has been partially supported by the French 'Agence Nationale de la Recherche' under grant ANR-11-B556-011. This work is also supported by the National Basic Research Program of China (2014CB845903 and 2012CB825604) and the National Natural Science Foundation of China (41174122, 41031066, 41104093).

### Author details

<sup>1</sup>Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China. <sup>2</sup>University Grenoble Alpes, ISTerre, CNRS, F-38041 Grenoble, France. <sup>3</sup>Key Laboratory of Earth and Planetary Physics, Chinese Academy of Sciences, Beijing, China.

Received: 18 July 2014 Accepted: 15 December 2014

Published online: 30 January 2015

### References

- Bloxham J, Jackson A (1992) Time-dependent mapping of the magnetic field at the core-mantle boundary. *J Geophys Res* 97:19,537–63
- Chapman S, Bartels J (1940) *Geomagnetism*, Vol. 2. Clarendon Press, Oxford
- Chulliat A, Telali K (2007) World monthly means database project. *Publ Inst Geophys Pol Acad Sc C-99(398)*:537–552
- Cliwer E, Kamide Y, Ling A (2000) Mountains versus valleys: semiannual variation of geomagnetic activity. *J Geophys Res* 105(A2):2413–2424
- Currie RG (1968) Geomagnetic spectrum of internal origin and lower mantle conductivity. *J Geophys Res* 73:2779–2786
- De Santis A, Barraclough D, Tozzi R (2003) Spatial and temporal spectra of the geomagnetic field and their scaling properties. *Phys Earth Planet Int* 135(2):125–134
- Elias AG, Silbergleit VM, de Gonzalez ALC (2011) Long-term variation of the semi-annual amplitude in the aa index. *J Atmos Sol Terr Phys* 73:1492–1499
- Finlay CC, Jackson A, Gillet N, Olsen N (2012) Core surface magnetic field evolution 2000–2010. *Geophys J Int* 189:761–781. doi:10.1111/j.1365-246X.2012.05395.x.
- Fujii I, Schultz A (2002) The 3d electromagnetic response of the earth to ring current and auroral oval excitation. *Geophys J Int* 151(3):689–709
- Gillet N, Pais MA, Jault D (2009) Ensemble inversion of time-dependent core flow models. *Geochem Geophys Geosyst* 10(6):Q06,004. doi:10.1029/2008GC002,290.
- Gillet N, Lesur V, Olsen N (2010) Geomagnetic core field secular variation models. *Space Sci Rev* 155:129–145. doi:10.1007/s11,214-009-9586-6.
- Gillet N, Jault D, Finlay CC, Olsen N (2013) Stochastic modelling of the earth's magnetic field: inversion for covariances over the observatory era. *Geochem Geophys Geosyst* 14(4):766–786. doi:10.1029/2012GC004355.
- Haines GV (1993) Modelling geomagnetic secular variation by main-field differences. *Geophys J Int* 114(3):490–500. doi:10.1111/j.1365-246X.1993.tb06982.x
- Hulot G, Olsen N, Sabaka TJ (2007) The present field. In: In: Kono M, Schubert G (eds) *Treatise in Geophysics, Geomagnetism*. Elsevier, New York Vol. 5. pp 33–75
- Jackson A, Jonkers ART, Walker MR (2000) Four centuries of geomagnetic secular variation from historical records. *Phil Trans R Soc Lond A* 358:957–990
- Le Mouél J, Blanter E, Chulliat A, Shnirman M (2004a) On the semiannual and annual variations of geomagnetic activity and components. *Ann Geophys* 22:3583–3588
- Le Mouél J, Blanter E, Shnirman M (2004b) The six-month line in geomagnetic long series. *Ann Geophys* 22:985–992
- Lesur V, Wardinski I, Asari S, Minchev B, Manda M (2010) Modelling the Earth's core magnetic field under flow constraints. *Earth Planets Space* 62:503–516
- Liu L, Zhao B, Wan W, Ning B, Zhang ML, He M (2009) Seasonal variations of the ionospheric electron densities retrieved from constellation observing system for meteorology, ionosphere, and climate mission radio occultation measurements. *J Geophys Res: Space Phys* 114(A02302). doi:10.1029/2008JA013819.
- Love JJ, Chulliat A (2013) An international network of magnetic observatories. *Eos, Trans Am Geophys Union* 94(42):373–374
- Love JJ, Rigler EJ (2014) The magnetic tides of Honolulu. *Geophys J Int* 197(3):1335–1353
- Lühr H, Maus S (2010) Solar cycle dependence of quiet-time magnetospheric currents and a model of their near-Earth magnetic field. *Earth Planets Space* 62:843–848
- Lyatsky W, Tan A (2003) Latitudinal effect in semiannual variation of geomagnetic activity. *J Geophys Res: Space Phys* 108(A5):1978–2012
- Lyatsky W, Newell P, Hamza A (2001) Solar illumination as cause of the equinoctial preference for geomagnetic activity. *Geophys Res Lett* 28(12):2353–2356
- Malin S, Isikara AM (1976) Annual variation of the geomagnetic field. *Geophys J R Astron Soc* 47(3):445–457
- Malin S, Winch D (1996) Annual variation of the geomagnetic field. *Geophys J Int* 124(1):170–174
- Matzka J, Chulliat A, Manda M, Finlay C, Qamili E (2010) Geomagnetic observations for main field studies: from ground to space. *Space Sci Rev* 155:29–64. doi:10.1007/s11214-010-9693-4
- McLeod MG (1996) Spatial and temporal power spectra of the geomagnetic field. *J Geophys Res* 101:2745–2763
- Menvielle M, Berthelier A (1991) The k-derived planetary indices: description and availability. *Rev Geophys* 29(3):415–432
- Mishima T, Owada T, Moriyama T, Ishida N, Takahashi K, Nagamachi S, Yoshitake Y, Minamoto Y, Muromatsu F, Toyodome S (2013) Relevance of magnetic properties of soil in the magnetic observatories to geomagnetic observation. *Earth Planets Space* 65(4):337–342
- Olsen N (1998) The electrical conductivity of the mantle beneath Europe derived from c-responses from 3 to 720 hr. *Geophys J Int* 133(2):298–308
- Olsen N (1999) Long-period (30 days–1 year) electromagnetic sounding and the electrical conductivity of the lower mantle beneath Europe. *Geophys J Int* 138(1):179–187
- Olsen N (2002) A model of the geomagnetic field and its secular variation for epoch 2000 estimated from ørsted data. *Geophys J Int* 149(2):454–462
- Olsen N, Lowes F, Sabaka TJ (2005a) Ionospheric and induced field leakage in geomagnetic field models, and derivation of candidate models for DGRF 1995 and DGRF 2000. *Earth Planets Space* 57:1191–1196
- Olsen N, Sabaka TJ, Lowes F (2005b) New parameterization of external and induced fields in geomagnetic field modeling, and a candidate model for igrf 2005. *Earth Planets Space* 57(12):1141–1149
- Olsen N, Glassmeier KH, Jia X (2010a) Separation of the magnetic field into external and internal parts. *Space Sci Rev* 152(1-4):135–157
- Olsen N, Manda M, Sabaka T (2010b) Tøffner-Clausen L. The CHAOS-3 geomagnetic field model and candidates for the 11th generation IGRF. *Earth Planets Space* 62:719–727
- Olsen N, Lühr H, Finlay CC, Sabaka TJ, Michaelis I, Rauberg J (2014) Tøffner-Clausen, L. The CHAOS-4 geomagnetic field model. *Geophys J Int* 197(2):815–827
- Rasmussen CE, Williams CKI (2006) *Gaussian processes for machine learning*. The MIT Press, Cambridge
- Russell C, McPherron R (1973) Semiannual variation of geomagnetic activity. *J Geophys Res* 78(1):92–108
- Sabaka T, Langel R, Baldwin R, Conrad J (1997) The geomagnetic field 1900-1995, including the large-scale field from magnetospheric sources, and the NASA candidate models for the 1995 revision of the IGRF. *J Geomag Geoelectr* 49:157–206
- Sabaka TJ, Olsen N, Langel RA (2002) A comprehensive model of the quiet-time, near-earth magnetic field: phase 3. *Geophys J Int* 151(1):32–68. doi:10.1046/j.1365-246X.2002.01774.x
- Sabaka TJ, Olsen N, Purucker ME (2004) Extending comprehensive models of the Earth's magnetic field with Oersted and CHAMP data. *Geophys J Int* 159:521–547
- Stening R, Winch D (2013) The ionospheric sq current system obtained by spherical harmonic analysis. *J Geophys Res: Space Phys* 118(3):1288–1297

- Turner GM, Rasson JL, Reeves CV (2007) Observation and measurement techniques. In: Kono M, Schubert G (eds) *Treatise in Geophysics, Geomagnetism* Vol. 5. pp 93–146
- Velínský J, Finlay CC (2011) Effect of a metallic core on transient geomagnetic induction. *Geochem Geophys Geosyst* 12(5):Q05,011. doi:10.1029/2011GC003557.
- Wardinski I, Holme R (2006) A time-dependent model of the Earth's magnetic field and its secular variation for the period 1980–2000. *J Geophys Res* 111. doi:10.1029/2006JB004,401.
- Wardinski I, Holme R (2011) Signal from noise in geomagnetic field modelling: denoising data for secular variation studies. *Geophys J Int* 185:653–662
- Wardinski I, Lesur V (2012) An extended version of the C<sup>3</sup>FM geomagnetic field model – application of a continuous frozen-flux constraint. *Geophys J Int* 189:1409–1429. doi:10.1111/j.1365-246X.2012.05384.x.
- Wardinski I, Mandea M (2006) Annual and semi-annual variations of the geomagnetic field components analysed by the multi-taper method. *Earth Planes Space* 58(6):785
- Yaglom AM (1962) *An introduction to the theory of stationary random functions*. Prentice-Hall, Upper Saddle River

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---