Earth, Planets and Space

FULL PAPER Open Access

A novel Bayesian approach for disentangling solar and geomagnetic field influences on the radionuclide production rates

Long Nguyen*, Neil Suttie, Andreas Nilsson and Raimund Muscheler

Abstract

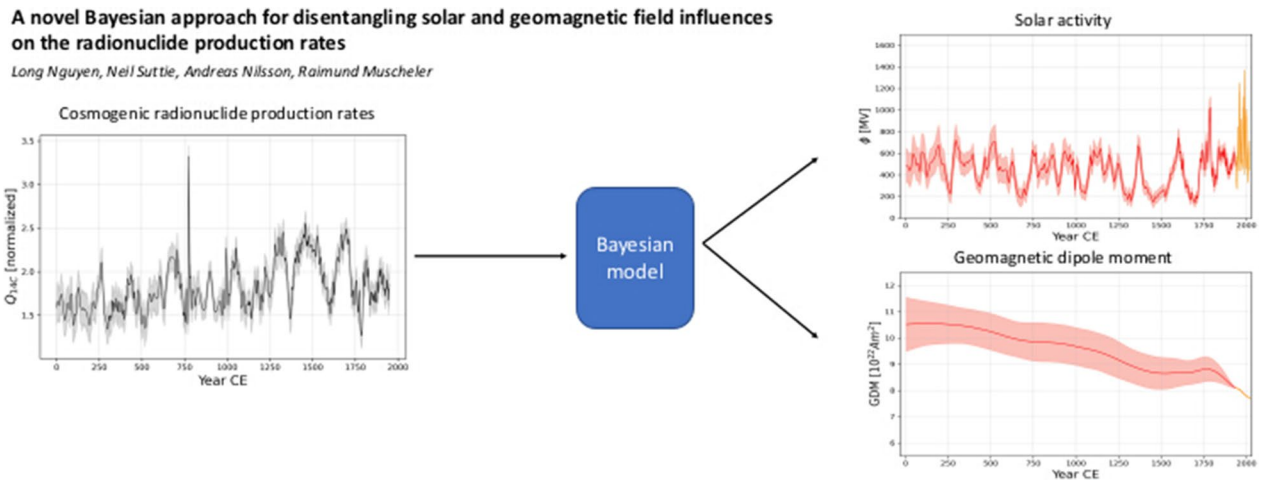
Cosmogenic radionuclide records (e.g., ^{10}Be and ^{14}C) contain information on past geomagnetic dipole moment and solar activity changes. Disentangling these signals is challenging, but can be achieved by using independent reconstructions of the geomagnetic dipole moment. Consequently, solar activity reconstructions are directly influenced by the dipole moment uncertainties. Alternatively, the known differences in the rates of change of these two processes can be utilized to separate the signals in the radionuclide data. Previously, frequency filters have been used to separate the effects of the two processes based on the assumption that millennial-scale variations in the radionuclide records are dominated by geomagnetic dipole moment variations, while decadal-to-centennial variations can be attributed to solar activity variations. However, the influences of the two processes likely overlap on centennial timescales and possibly millennial timescales as well, making a simple frequency cut problematic. Here, we present a new Bayesian model that utilizes the knowledge of solar and geomagnetic field variability to reconstruct both solar activity and geomagnetic dipole moment from the radionuclide data at the same time. This method allows for the possibility that solar activity and geomagnetic dipole moment exhibit variations on overlapping timescales. The model was tested and evaluated using synthetic data with realistic noise and then used to reconstruct solar activity and the geomagnetic dipole moment from the ^{14}C production record over the last two millennia. The results agree with reconstructions based on independent geomagnetic field models and with solar activity inferred from the Group Sunspot number. Our Bayesian model also has the potential to be developed further by including additional confounding factors, such as climate influences on the radionuclide records.

Keywords: Solar activity, Paleomagnetism, Cosmogenic radionuclide, Holocene, ^{14}C

*Correspondence: hoang_long.nguyen@geol.lu.se

Department of Geology, Faculty of Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden

Graphical Abstract



Introduction

Cosmogenic radionuclides such as ^{10}Be and ^{14}C are the best proxies for solar activity reconstructions prior to the period of direct solar observations (Beer et al. 1988; Muscheler et al. 2007) such as the record of group sunspot number (GSN) starting in 1610 CE (Svalgaard and Schatten 2016), and the extended neutron monitor data back to 1939 CE (Herbst et al. 2017). Cosmogenic radionuclide records are vital for studies of long-term changes of the Sun and Sun–climate linkages far back in time. The radionuclides are continuously produced from the interaction between high-energy galactic cosmic rays (GCRs) coming to the Earth (and secondary particles) and atoms in the atmosphere. Their production rates correlate with the flux of GCRs reaching the Earth's atmosphere, which is modulated by solar and geomagnetic shielding (Beer et al. 2012). Thus, reconstructions of solar activity from cosmogenic radionuclide records require correction for the geomagnetic field influence, usually based on independent reconstructions of the geomagnetic dipole moment (GDM). Consequently, solar activity reconstructions are directly influenced by the GDM uncertainties. The uncertainties of GDM reconstructions depend on the choice of the underlying data and the different methods to build global geomagnetic field models from them. Moreover, differences in GDM reconstructions directly lead to discrepancies in solar activity reconstructions. Additional uncertainties in solar reconstruction arise from cosmogenic radionuclide measurement uncertainties and possible systematic biases such as uncorrected climatic influences on ^{10}Be transport and deposition processes or carbon cycle influences on atmospheric ^{14}C concentrations (Muscheler et al. 2007).

An alternative approach to directly disentangle solar and GDM influences on radionuclide records is to utilize the known differences on the rates of change and the GCR shielding effects of these two processes. In general, millennial variations in the radionuclide records are assumed to be dominated by GDM variations, while decadal-to-centennial variations can be mainly attributed to solar activity variations (Snowball and Muscheler 2007). Utilizing this prior knowledge could not only eliminate the need for independent GDM estimates for solar activity reconstructions, but also provide the possibility to infer GDM variations from radionuclide records only. Those radionuclide-based GDM reconstructions provide valuable information on the past GDM as they are dominated by the global dipole field, in contrast to the information of magnetic field directions and intensity stored in, for example, archeological artifacts, igneous rocks and sediment records that only provide local readings of the past geomagnetic field. Long-term changes in GDM have been reconstructed from radionuclide data by removing variations on timescales shorter than 3000 years (Muscheler et al. 2005; Zheng et al. 2021). Similarly, the removal of long-term variations from the radionuclide data has been conducted in some studies to minimize the geomagnetic field influence in order to assess past changes in solar activity (e.g., Adolphi et al. 2014). Until now, these approaches have used simple frequency filtering with a hard cut-off frequency depending on the targets of the studies. However, Snowball and Muscheler (2007) showed that solar and GDM

variability likely overlap making a simple frequency distinction problematic. Quasi-periodicities of ~ 200 years to up to ~ 2400 years in solar variations have been shown in previous studies of long-term radionuclide records (e.g., Bond et al. 2001; Wagner et al. 2001; Snowball and Muscheler 2007; Adolphi et al. 2014; Usoskin et al. 2016; Dergachev and Vasiliev 2019). Meanwhile, GDM can have short-term variation on timescales between ~ 60 and ~ 200 years but with a relatively low power (Hellio and Gillet 2018; Huder et al. 2020). Significant influence of GDM on the radionuclide records begins on timescales between 300 and 500 years as suggested from the correlation between ^{14}C production rate and reconstructed GDM over the last 10,000 years (Snowball and Muscheler 2007). Therefore, the overlap of solar and GDM influences on timescales longer than 300 years is very challenging for separating their effect on cosmogenic radionuclide production rates. Thus, simple frequency filters likely fail in separating the signals and, in addition, often lead to unreliable end effects. These effects make it difficult to connect the reconstructions to present day values of solar activity or GDM intensity inferred from instrumental data. These difficulties motivated this study to improve methods to disentangle solar and geomagnetic influences for the reconstructions of solar activity and GDM from radionuclide records.

Here, we address the challenges discussed above by incorporating prior knowledge of solar and GDM variability and their influence on radionuclide production rates into a Bayesian framework. We present a new Bayesian method to separate solar activity and GDM variations from the radionuclides data inspired by recently developed methods to reconstruct geomagnetic field variations using paleomagnetic data (Hellio and Gillet 2018; Nilsson and Suttie 2021). Our goal is to develop a model that can incorporate and utilize the knowledge of solar and geomagnetic field variations to reconstruct both solar activity and GDM from the radionuclides data at the same time. This method allows for the possibility that solar activity and GDM exhibit variations on overlapping timescales. Different datasets used for the model development and testing are presented in the next section. In Sect. 3, we outline the Bayesian framework and our setup of prior information on the model parameters. The model is tested in Sect. 4 using synthetic ^{14}C data, before being applied to nearly 2000 years of ^{14}C production data inferred from IntCal20 (Reimer et al. 2020).

Data

Observation-based solar activity and geomagnetic field data for model calibration

The reconstruction of solar activity with ^{14}C data requires calibration with the instrumentally measured cosmic-ray flux record during the present period (Muscheler et al. 2016). The solar modulation of the cosmic-ray flux is usually quantified using the solar modulation potential, ϕ [MV], (also known as the solar modulation parameter). Particularly, the parameter approximates the adiabatic energy loss of GCRs in the heliosphere due to solar shielding (Vonmoos et al. 2006; Herbst et al. 2017). In this study, we used the monthly record of the solar modulation potential ϕ_{HE16} from 1939 to 2017 CE published by Herbst et al. (2017). The annual average of ϕ_{HE16} is shown in Fig. 1. The modulation potential depends on the assumed local interstellar spectrum (LIS) which is the flux of GCRs outside the heliosphere. We used the recent LIS model from Herbst et al. (2017).

The GSN record by Svalgaard and Schatten (2016) is the longest and most recent compilation of direct telescope-based solar observations. It contains the yearly average (average over all months of the year) of the GSN from 1610 CE to present. The GSN record is a relatively robust solar proxy which is not affected by the shielding effects of the geomagnetic field and climate influences as in the case of ^{10}Be records or by carbon cycle effects in the case of ^{14}C records. However, there are often gaps in the GSN data due to the discontinuity of observations. Consequently, the GSN record was compiled by many observers and, therefore, exhibits uncertainties from the process of calibration and combination of different datasets, especially during the data-poor times (Svalgaard and Schatten 2016). Moreover, the constant improvement of the observation technique through time adds more challenges to the calibration of the older data. We

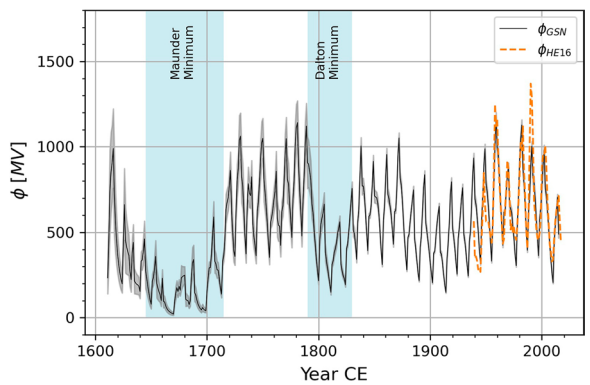


Fig. 1 Solar modulation potential since 1610 CE inferred from the GSN and neutron monitor data. The black line and the grey shading indicate the mean and 2-sigma uncertainty of the solar modulation potential inferred from the GSN (Svalgaard and Schatten 2016). The Grand Solar Minima are indicated by name and blue shadings. The solar modulation potential estimated from neutron monitor data (Herbst et al. 2017) from 1939 onwards is shown as an orange dashed line

inferred the solar modulation potential from the GSN, i.e., ϕ_{GSN}, following the method proposed by Usoskin et al. (2002). A Monte Carlo approach was used to assess the uncertainty of the inference, i.e., 1000 realizations of the solar modulation potential were generated consistent with the GSN record and its uncertainties. The mean and 2-sigma uncertainty of these 1000 realizations are shown in Fig. 1. The record clearly shows the 11-year Schwabe cycle (Schwabe 1844). Centennial variations of solar activity are also represented such as the so-called "Grand Solar Minima". These are extended periods of relatively low solar activity such as the Maunder Minimum from 1645 to 1715 CE and the Dalton Minimum from 1790 to 1820 CE (highlighted in Fig. 1). ϕ_{GSN} is the longest observation-based solar modulation record containing the known typical solar variability, i.e., the 11-year variation and the centennial variation, and therefore we consider it suitable for model parameterization. Thus, we used ϕ_{GSN} to assess the typical behavior of solar activity which was then included as prior knowledge in our Bayesian model.

On the other hand, for the GDM calibration we used the recent model COV-OBS. × 2 (Huder et al. 2020). The GDM was constrained by observational data from both ground-based stations and satellites as well as older surveys over the period 1840–2020 CE.

Radiocarbon data

We inferred the ^{14}C production rate for the period 1–1950 CE from the IntCal20 Northern Hemisphere ^{14}C calibration curve (Reimer et al. 2020). IntCal20 was compiled by the IntCal Working Group for improving the ^{14}C age calibration and it can be used to reconstruct fluctuations in past atmospheric ^{14}C concentrations. We calculated the ^{14}C production rate from the atmospheric ^{14}C fluctuations using a box-diffusion carbon cycle model, which includes atmosphere, biosphere, an upper ocean mixed layer and 42 deep-sea layers. The option for direct ventilation of the deep ocean was turned off (Siegenthaler, 1983). The uncertainty in the IntCal20 calibration curve was quantified using 100 posterior realizations of possible atmospheric ^{14}C curves obtained via fitting Bayesian splines to the ^{14}C data underlying IntCal20 (Heaton et al. 2020; Reimer et al. 2020). The average of the annual ^{14}C production rates inferred from these 100 realizations is depicted in Fig. 2. The ^{14}C production rates were normalized to have the pre-industrial mean of one. The surge in atmospheric CO_2 from 1850 CE due to fossil-fuel burning was included in the calculation to account for the dilution of ^{14}C in relation to ^{12}C (Muscheler et al. 2007).

Short-term solar proton events (SPEs) such as the 774/775 event and the 993/994 event (Miyake et al. 2012, 2013; Mekhaldi et al. 2015; Reimer et al. 2020) can be observed in the annual production rate as rapid rises in the production rate. During these events, the Sun released significant amounts of energetic particles which resulted in short-term radionuclide production enhancements (Mekhaldi et al. 2015). These SPEs are outside the scope of this study and, therefore, they were not included in our model. The 774/775 event resulted in an extreme peak in the production rate and, therefore, it needs to be excluded from the model data. We removed the extreme values during the event and replaced them using a spline interpolation. In addition, we decreased the sampling resolution of the ^{14}C data prior to 1600 CE by computing the mean ^{14}C production rate every 10 years, as the data underlying IntCal20 decreases in amount and resolution further back in time. Figure 2 shows the mean and the 2-sigma uncertainty of the processed ^{14}C production rate. The mix of different temporal resolutions allowed us to assess our model performance at both long (prior to 1600 CE) and short timescales (after 1600 CE) and, at the same time, it helped to save computational costs via reducing the number of data points. This approach was further motivated by the fact that the ^{14}C data underlying IntCal20 is continuously annual only for the last 1000 years. In fact, the processed ^{14}C data retain variations similar to the unprocessed data prior to 1000 CE, except for the periods with annual resolution such as from 300 to 400 CE and around the SPE spikes.

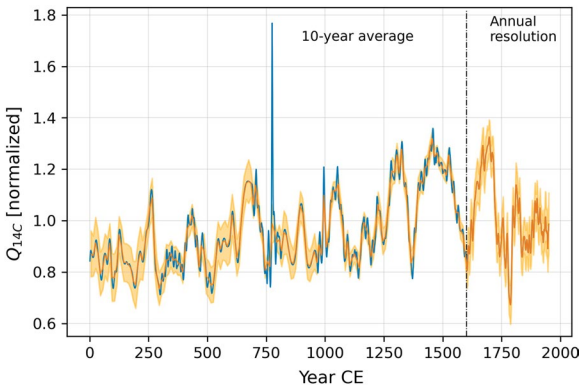


Fig. 2 ^{14}C production rate for the last 2000 years inferred from IntCal20. The blue line indicates the mean of a hundred realizations of annual ^{14}C production rate inferred from IntCal20 and normalized to have the pre-industrial mean of one. The SPEs during 774–775 CE and 993–994 CE can be observed as rapid rises in the ^{14}C production rate. The orange line and orange shading indicate the mean and 2-sigma uncertainty of 100 realizations of the processed ^{14}C data (see text). The vertical dash-dotted line indicates the year where the process data were split into two temporal resolutions, i.e., 10-year average before 1600 CE and annual resolution after 1600 CE

Modeling method

The global production rate of cosmogenic radionuclides

The global production rate of cosmogenic radionuclides depends on the cosmic-ray flux coming to the Earth which is modulated by solar activity (i.e., solar shielding of GCRs in the heliosphere) and the shielding effects of the geomagnetic field (Masarik and Beer 1999). Concentrations of stable and radioactive nuclides in meteorites and terrestrial archives suggest that, on time scales of about 0.5 million years, the GCR flux outside the heliosphere has remained constant within ± 10% over the past ∼ 10 million years (Wieler et al. 2013). Therefore, the assumption of a constant local interstellar GCR spectrum outside the heliosphere is usually made for any reconstruction of solar and geomagnetic field activity using cosmogenic radionuclides. The global production rate of cosmogenic radionuclides (Q [atoms cm^{-2} s^{-1}]) can then be modeled as a function of solar modulation potential (ϕ [MV]) and GDM denoted as M [10^{22} A m^2]:

$$Q(t) = f(\phi(t), M(t)), \tag{1}$$

where t represents time. In this study, we employed the tabulated function established in Kovaltsov et al. (2012) for the ^{14}C production rate (Fig. 3). Since the tabulated function was published with discrete data, we approximated it with a polynomial function (details in section 1 of the Additional file 1). This allowed us to work with continuous values of ϕ and GDM.

The function established by Kovaltsov et al. (2012) was based on the LIS model from Usoskin et al. (2005). According to Herbst et al. (2017), ϕ from one LIS can be converted to ϕ in another LIS by means of linear regression functions. We then used the following regression function to convert from ϕ_{LIS05} to the more recent LIS model by Herbst et al. (2017), i.e., ϕ_{HE16}:

$$\phi_{\mathrm{HE16}} = 1.025 \times \phi_{\mathrm{US05}} + 24.18\mathrm{MV}. \tag{2}$$

Theoretical ^{14}C production rates and the normalized ^{14}C data need to be connected via a normalizing constant. This constant can be estimated through a comparison of the ^{14}C production rate data with theoretically expected ^{14}C production rates inferred from independent records of solar modulation potential and GDM. As mentioned above, the GDM from 1840 CE provided by COV-OBS. × 2 is suitable for our purpose. On the other hand, ϕ_{HE16} provides a record of solar activity inferred from instrumental data. However, it only overlaps with the radiocarbon data for a short period from 1939 to 1950 CE. Therefore, we rely on the solar modulation record inferred from the GSNs as an alternative.

Figure 4 shows the global ^{14}C production rate estimated by combining the mean of ϕ_{GSN} and the mean GDM predicted by COV-OBS. × 2 from 1840 to 1950 CE. The significant 11-year solar variation can be observed in the estimated mean production rate, while the variation has a lower amplitude in the processed ^{14}C data (Fig. 4). The main reason is the smoothening which occurred during the construction of the average IntCal20 ^{14}C record (Heaton et al. 2020; Reimer et al. 2020) and the fact that the 11-year cycle variability is strongly dampened and close

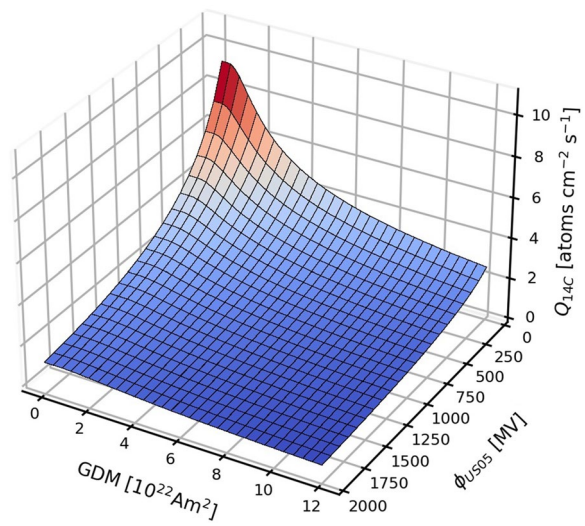


Fig. 3 Global production rate of ^{14}C as a function of solar modulation potential and GDM. The graph is based on the model of Kovaltsov et al. (2012)

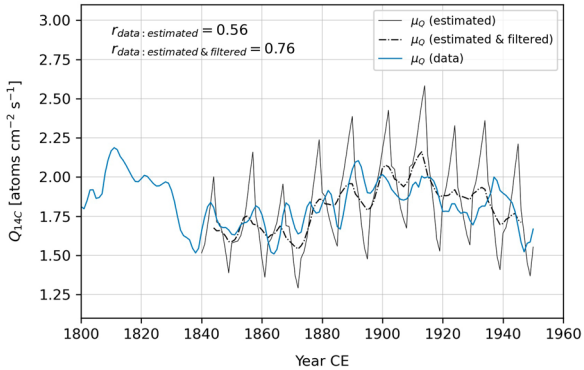


Fig. 4 A comparison between the ^{14}C production rate data and theoretically expected ^{14}C production rates. The theoretical estimate of the mean global ^{14}C production rate (black line) for the period 1840–1950 CE was based on φ_{GSN} and COV-OBS. × 2. The dash-dotted line depicts the estimated ^{14}C production rate after being filtered with a 9-year moving average filter. The ^{14}C production rate based on IntCal20 (blue line) was normalized to have the same mean as the filtered ^{14}C production rate for the overlapping period. The correlation coefficients between the ^{14}C production rate modeled from the data and the filtered or non-filtered estimates of ^{14}C production rate are shown

to the detection limit in tree-ring based reconstructions of the atmospheric ^{14}C concentration (Brehm et al. 2021). The smoothening of the ^{14}C data will have consequences for our modeling approach later on. Assessing this smoothening with a moving average filter showed that a 9-year moving average version of the theoretical ^{14}C production rate (Fig. 4) approximates best the variations in the ^{14}C data ($r=0.76$). A normalizing constant was then computed by comparing the averages of the estimated ^{14}C production rate filtered with a 9-year moving average filter, and the averages of the ^{14}C production data for the period of overlap. The model uses this normalizing constant to connect the input ^{14}C data to the global production rates generated with Eq. 1.

A Bayesian approach for sampling the past solar modulation potential and geomagnetic dipole moment

Consider a model with a set of parameters (θ) and observed data (y); a set of parameters that fit with the observed data can be found using Bayes' theorem (Gelman et al. 2004):

$$p(\theta|y) \propto p(y|\theta) \times p(\theta). \tag{3}$$

Within the Bayes' framework, $p(\theta|y)$ is the unnormalized posterior distribution of the parameters after considering the observed data. $p(y|\theta)$ is the distribution of the data conditional on θ. It is also called the likelihood function if it is treated as a function of θ. $p(\theta)$ is the prior distribution of the model parameters before any observation. Although the posterior distribution $p(\theta|y)$ cannot always be solved analytically, Eq. 3 allows for an approximation by generating samples from the posterior distribution via different sampling methods. This method of statistical inference or Bayesian inference allows us to incorporate our additional knowledge into the prior distribution of the parameters. Moreover, the parameters' distribution could be updated continuously as more observations become available. This is particularly useful in cases where the sample size is not fixed and/or when the users want to incorporate additional information.

We employed the Hamiltonian Monte Carlo (HMC) method which utilizes the Hamiltonian dynamic simulation to efficiently generate samples from the posterior distribution (Neal 2011). We developed and executed our model via Stan (Carpenter et al. 2017), a probabilistic programming language for statistical modeling and high-performance statistical computation. In addition, we used the No-U-Turn Sampler (NUTS), an extension of HMC developed by (Hoffman and Gelman 2014) to sample from the posterior distribution of solar variations and GDM. NUTS provides an auto-tuning of difficult and highly sensitive parameters of the HMC sampler.

If the solar modulation potential (ϕ) and the GDM (M) are considered as parameters in formula 3, the Bayesian approach can be used to find ϕ and M that fit with the observed global production rate (Q):

$$p(\phi, M|Q) \propto p(Q|\phi, M) \times p(\phi, M), \tag{4}$$

and if ϕ and M are independent, the formula can be rewritten as:

$$p(\phi, M|Q) \propto p(Q|\phi, M) \times p(\phi) \times p(M), \tag{5}$$

where the likelihood function $p(Q|\phi,M)$ can be established based on the global production rate function of Kovaltsov et al. (2012) (Fig. 3). This Bayesian method allows us to input our information on the characteristics of the Sun and the geomagnetic field intensity via the prior distribution $p(\phi)$ and $p(M)$. Combining this information with the observed production rate data offers the possibility to reduce the uncertainty and improve the reconstructions.

Setting up the prior distributions

Prior distribution of solar variability

The prior distribution of solar activity was set up using the framework of a Gaussian process (GP) described by Rasmussen and Williams (2006):

$$\phi(t) \sim GP\left(m_\phi(t), k_\phi\left(t, t^{'}\right)\right), \tag{6}$$

where $\phi(t)$ represents the solar modulation potential at time t. t and t' are separated by a time difference r. $m_\phi(t)$ is the mean function and $k_\phi(t,t')$ is the covariance function of $\phi(t)$. The mean function and the covariance function are given by:

$$m_\phi(t) = E[\phi(t)], \tag{7}$$

$$k_\phi\left(t, t^{'}\right) = E\left[\left(\phi(t) - m_\phi(t)\right) \times \left(\phi\left(t^{'}\right) - m_\phi\left(t^{'}\right)\right)\right], \tag{8}$$

where E represents the expected values. The GP is a generalization of the multivariate Gaussian probability distribution (Rasmussen and Williams 2006), where each point t is described by a mean and a joint Gaussian distribution with the surrounding points. Consequently, every point/event in time is influenced by (correlated to) the data before and afterwards. Characteristics of the correlation are determined by the covariance function. The covariance matrix, generated using a specific covariance function (discussed later), is the collection of vectors defining the correlation of every point in time with the surrounding points. In other words, the covariance matrix includes the information on the characteristic timescales and variances of the physical processes that we aim to reconstruct.

The most common covariance function within the machine learning field is the squared exponential (SE) (Rasmussen and Williams 2006). The covariance (k_{SE}) and spectral density (S_{SE}) of the SE covariance function have the forms:

$$k_{SE}(r) = \sigma^2 \times \exp\left(-\frac{r^2}{2l^2}\right), \tag{9}$$

$$S_{SE}(s) = \sigma^2 \times \left(2\pi l^2\right)^{\frac{D}{2}} \times \exp\left(-2\pi^2 l^2 s^2\right), \tag{10}$$

where D is the dimensionality, σ^2 is the signal variance and l is the length-scale that determines how quickly the correlation diminishes with time. s in Eq. 10 represents frequencies. Additional file 1: Fig. S3 shows an example of the covariance k_{SE} *(r)* as a function of the input distance r in time.

In this study, a new covariance function for the solar variations was created by adding two SE covariance functions with different characteristic timescales:

$$k(r) = \sigma^2_{short}\exp\left(-\frac{r^2}{2l^2_{short}}\right) + \sigma^2_{long}\exp\left(-\frac{r^2}{2l^2_{long}}\right). \tag{11}$$

One SE covariance function simulates the observed short-term variations (e.g., 11-year cycle), while the other one simulates the observed centennial variations (e.g., 88–100 years) of the Sun. This combined covariance function was established based on the variations of the solar modulation parameters (i.e., ϕ_{GSN}) which is the longest observational record of solar activity. The record shows solar activity for the last ~ 400 years which was dominated by the 11-year cycle and the centennial variations (Fig. 1). The short length of ϕ_{GSN} is a weakness of the model since the record is not long enough to capture possible quasi-periodicities of ~ 400 years to up to ~ 2400 years in solar variations (e.g., Bond et al. 2001; Snowball and Muscheler 2007; Usoskin et al. 2016; Dergachev and Vasiliev 2019). However, we want to avoid including long-term cycles inferred from the radionuclide records in our prior information which would lead to circular reasoning as the model is applied to radionuclides. Moreover, solar variations inferred from radionuclide records on longer time scales are rather ambiguous since it is challenging to completely eliminate influences from geomagnetic field variations, transport and deposition processes for ^{10}Be and carbon cycle effects on ^{14}C (Vonmoos et al. 2006; Snowball and Muscheler 2007). In addition, we lack a longer direct observational record of solar activity that could help us to obtain a better constrained prior for solar variability on timescales longer than the GSN record. Nevertheless, the GSN record still shows indications of the 200-year cycle such as the Maunder minimum, a period with an almost complete lack of sunspots, which can be considered as an expression of the 200-year cycle (Fig. 1). The millennial-scale cycles can possibly be characterized as bundling of larger-amplitude centennial-scale variations followed by periods of weaker centennial-scale variability, and, in

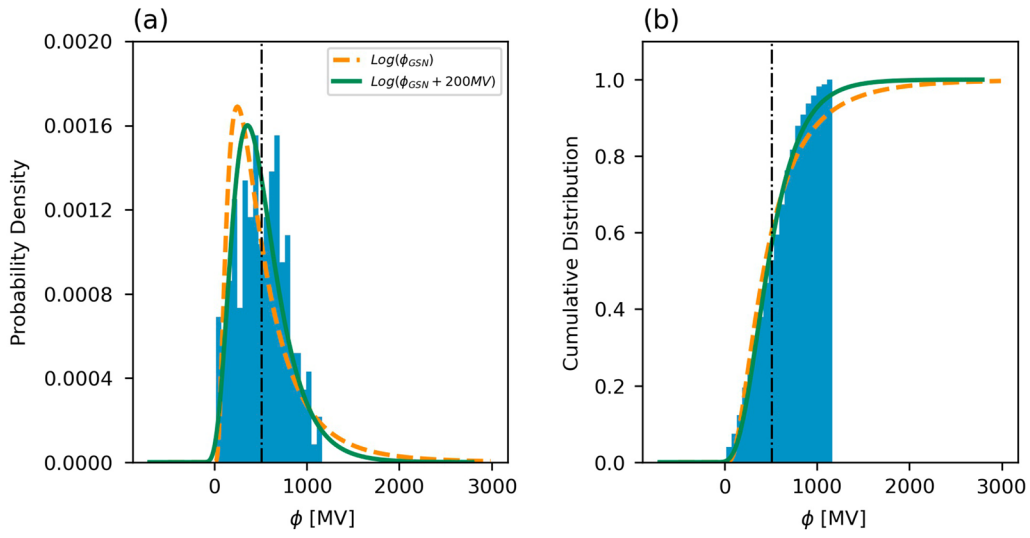


Fig. 5 The distribution of φ_{GSN} approximated with log-normal distributions. The histograms show the density distribution (**a**) and the cumulative distribution (**b**) of φ_{GSN}. The vertical dash-dotted black lines represent the mean of φ_{GSN}. The orange and green lines depict log-normal distributions with the best fit to φ_{GSN} and φ_{GSN} + 200MV, respectively. The green lines in both panels a and b have been shifted to the left (via subtracting **c**) to be able to compare with φ_{GSN}

such cases, the proposed covariance function can reproduce the millennial-scale variations to some extent.

The histograms in Fig. 5 show that ϕ_{GSN} has a positive distribution (i.e., the adiabatic energy loss of GCR cannot be negative) that is skewed toward higher values. Therefore, the symmetrical Gaussian distribution, which also allows for negative values, implied by Eq. 6 is not an appropriate approximation of ϕ. Moreover, the polynomial approximation (Additional file 1: Fig. S2) of the ^{14}C production rate function starts to become unrealistic and produces negative values at ϕ smaller than -362 MV which would be problematic to the modeling process. For ϕ within the range of − 362 to 0 MV, which also implies unphysically negative shielding, the approximation produces unrealistically large values of ^{14}C production rate which would then mostly be rejected by the Bayesian sampler. However, the sampling process would be inefficient if the model has to frequently reject negative values of ϕ. Thus, we considered modeling ϕ using a log-normal distribution which is also a skewed positive distribution:

$$\log(\phi) \sim \mathrm{N}\left(\mu_{\log(\phi)}, \sigma^2_{\log(\phi)}\right), \tag{12}$$

where $\mu_{\log(\phi)}$ represents the mean and $\sigma^2_{\log(\phi)}$ represents the variance of the log transformation of ϕ. A disadvantage of using a log-normal distribution is that the model will occasionally generate extremely high values of ϕ due to the nonsymmetrical characteristic (Fig. 5a). However, the proposed ϕ will be rejected by the model when it is unrealistically high and cannot be explained by the radionuclide production rate. Another problem of the log-normal distribution is a bias toward lower values of ϕ as demonstrated by the histograms. Consequently, values above the mean of ϕ_{GSN} will be generated with a lower probability. A solution for this is to add a constant (c [MV]) to ϕ_{GSN} before fitting with a log-normal distribution:

$$\log(\phi + c) \sim N\left(\mu_{\log(\phi+c)}, \sigma^2_{\log(\phi+c)}\right). \tag{13}$$

After sampling from this distribution, we exponentiate and subtract c to obtain ϕ. This approach allows the model to generate a more flexible distribution agreeing well with the ϕ_{GSN} distribution (Fig. 5). This minimizes the bias toward lower values of ϕ at the cost of allowing for negative values of ϕ (i.e., as low as minus c) with a low prior probability. We assessed the fitted distribution for the case of c equal to 100, 200 and 300 MV (Additional file 1: Fig. S4) and decided to choose 200 MV as this value shows a good balance between the pros and cons. The fitted distribution to $\log(\phi + c)$ with c equal to 200 MV is shown in Fig. 5. Parameters of the fitted log-normal distribution such as mean and variance were assessed using the method of maximum likelihood estimation. In summary, the prior distribution allows for negative values of solar modulation as low as − 200 MV but with a low probability in exchange for a better model performance with less bias toward lower values of ϕ. Negative solar modulation values, while unphysical, could be generated associated with the biases in radionuclide data (e.g., climate impact), extreme spikes/enhancement in the radionuclide records (e.g., SPEs), or simply the data uncertainties. Data uncertainties are included in the model and we assume that climate biases are minor for the ^{14}C production rate. However, SPEs are not included in our model and, therefore, we removed the known SPE production peaks (i.e., the 774/775 and the 993/994 peaks) from the ^{14}C data. In addition, Fig. 5 shows that the prior distribution (green line) still slightly underestimates the probability of solar modulation from around 500 to 1000 MV. This could still lead to a slight bias toward lower values of solar modulation. However, the posterior distribution of solar activity will ultimately be evaluated and selected based on the radionuclide data. Therefore, a small inclination toward low solar modulation values of the prior distribution will not bias the results.

We then replaced Eq. 6 with:

$$log(\phi_c(t)) \sim GP\left(m_{\log(\phi_c)}(t), k_{\log(\phi_c)}\left(t, t^{'}\right)\right), \tag{14}$$

where $\phi_c(t) = \phi(t) + c$. The mean function and the covariance function are adjusted accordingly:

$$m_{\log(\phi_c)}(t) = E[\log(\phi_c(t))], \tag{15}$$

$$k_{\log(\phi_c)}\left(t, t^{'}\right) = E\left[\left(\log(\phi_c(t)) - m_{\log(\phi_c)}(t)\right) \times \left(\log(\phi_c\left(t^{'}\right)) - m_{\log(\phi_c)}\left(t^{'}\right)\right)\right]. \tag{16}$$

The mean of our proposed prior distribution ($m_{\log(\phi c)}$) for the modeling period was equal to the mean of the fitted log-normal distribution.

The short-term and long-term variations of the Sun were defined in Eq. 11 by the characteristic length-scale (l_{short}, l_{long}) and the signal variance (σ^2_{short}, σ^2_{long}). These parameters were determined using ϕ_{GSN}, particularly the power spectrum of $\log(\phi_{GSN})$ (Fig. 6a). In other words, we investigate how $\log(\phi_{GSN})$ behaves in the frequency range and adjust our prior to resemble it. The short-term variations are reflected as a bump and changes in the slope of the power spectrum around the 6- to 16-year period. The centennial variation could be observed for periods longer than 55 years, but the changes in the slope were not as

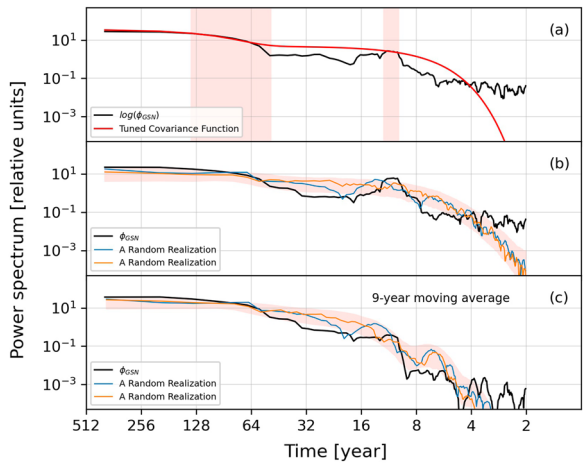


Fig. 6 Power spectrum of log (φ_{GSN}) and φ_{GSN} approximated by our tuned covariance function. **a** Compares the power spectrum of log(φ_{GSN}) (black line, standardized to have mean = 0 and variance = 1) with the theoretical spectral density of the tuned covariance function (red line). The tuning periods are highlighted with light red shading. Panel b compares the power spectrum of φ_{GSN} (black line, standardized to have mean = 0 and variance = 1) with power spectrums of two random realizations (color lines) generated by the tuned covariance function for the period from 1610 to 2015 CE. The light red envelop indicates the estimated 95% confidence interval of the power spectrums from 1000 random realizations. **c** Is similar to panel b, but shows the power spectrum of a 9-year moving average version of φ_{GSN} (black lines) and the power spectra of 1000 random realizations generated by the adapted covariance matrix (see text) for the same time period. All of the power spectra were computed with the Multitaper Method (bandwidth $N\Omega = 8$, tapers $K = 8$)

strong as the short-term variation. First, we determined l_{short} and σ^2_{short} by tuning our covariance function for the period from 10 to 12 years (highlighted in Fig. 6a). This ensures that our prior captures the short-term variations which are most prominent for period lengths from 10 to 12 years. We then tuned l_{long} and σ^2_{long} for the period from 50 to 136 years where the transition in power occurs (highlighted in Fig. 6a). Details of the parameterization process are outlined in section 2 of the Additional file 1.

The spectral density of the combined covariance function with tuned parameters is shown in Fig. 6a. The power decreases stepwise with a constant period between the steps as a result of the combination of our two SE covariance functions. As expected, the combination of the two tuned covariance functions does not fully simulate the bump-like structure in power generated by the short-term variation of the Sun around a 10- to 12-year period. The power was instead raised to a higher level before and after the bump. A fixed periodic signal could be introduced to simulate the bump-like property as a peak in the power spectrum. However, this would also limit the short-term variations to a narrowly defined cycle as, for example, an 11-year cycle. We here chose to apply a more relaxed prior since the short-term variations of the Sun vary around the 11-year timescale rather than being a cycle with constant frequency (Friis-Christensen and Lassen 1991; Solanki et al. 2002; Petrovay 2010; Brehm et al. 2021). Moreover, the short-term variations could have changed further back in time. Overall, despite this drawback, we still chose the SE covariance function because the flexibility in the range of the short-term variations allows for a variable frequency for the "quasi" 11-year cycle. Therefore, the solar cycle can be determined by the data instead of being imposed by the prior.

Figure 6b shows a comparison of the power spectra of ϕ_{GSN} and solar modulation potential realizations that were randomly generated with our tuned covariance function for the same time period. As expected, the covariance function generates short-term variations that are not fixed to, but rather vary around, the 11-year cycle. The transition in the power spectrum of ϕ_{GSN} from 50 to 136 years is well simulated. The covariance function generates curves with higher power than ϕ_{GSN} for periodicities with 16–40 years cycle lengths. This is a drawback of the SE covariance function that cannot be avoided with this rather simple approach, but on the other hand it is necessary to generate high enough power to capture the short-term variations. Variations on timescales shorter than 4 years can be observed in the power spectrum of ϕ_{GSN}. We interpret these variations as not relevant for our radionuclide data interpretation and decided to exclude them in our parameterization. Therefore, it is not a problem when the power spectrum of our generated solar modulation potential diminishes rapidly for periodicities shorter than 4 years.

We generated 1000 realizations of ϕ from 1 to 1938 CE using Eq. 14 to simulate our prior distribution of solar activity. We found that 1000 realizations were enough to capture the main aspect of our covariance matrix, mainly around the diagonal. The covariance matrix and hence the realizations were connected to the present solar activity of ϕ_{HE16}. Every realization was binned (i.e., taking an average of every 10 years) prior to 1600 CE to have the same time resolution as the processed ^{14}C data. Post 1600 CE the realizations were smoothened with a 9-year moving average filter to match the treatment of the ^{14}C data, as discussed in Sect. 3.1. A new adapted covariance matrix for solar variations over the period 1 to 1938 CE was computed from these 1000 processed realizations to account for the lower resolution and smoothening. The final prior distribution of solar activity was estimated via 1000 realizations generated using the adapted covariance matrix (Fig. 7a). Again, we tested and found that 1000

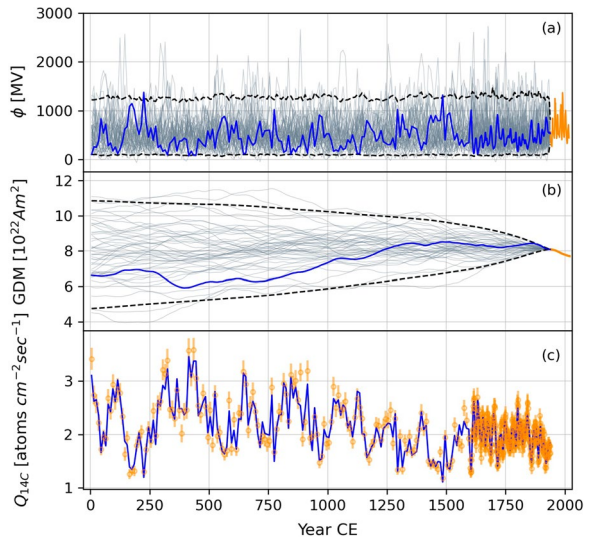


Fig. 7 Prior distributions of φ and GDM, and the synthetic ^{14}C data generated from the distributions. **a** and **b** Show the prior distributions of the solar activity and GDM, respectively. The grey lines depict 50 random realizations and the two black dashed lines indicate 2-sigma uncertainty of a thousand realizations. The orange lines represent the recent observations to which the model is connected, i.e., φ_{HE16} and GDM based on COV-OBS. × 2. The blue lines represent the selected realization used to generate the synthetic ^{14}C production rate. **c** Shows the synthetic ^{14}C production rate (blue line) and the resampled values with 6.7% standard error (orange circles with error bars)

realizations were enough to represent the adapted covariance matrix. The new solar realizations have comparable variations to the 9-year moving average version of ϕ_{GSN} during the same period (Fig. 6c). The approach described here allows us to directly adjust/adapt our covariance matrix without having to adjust/re-parameterize the covariance function. Thus, it will be helpful for modeling radionuclide data with different temporal resolution or smoothening when applied to long-term radionuclide records.

Prior distribution of geomagnetic field intensity

We set up the prior distribution for GDM following Bouligand et al. (2016). We simplified the approach using just the axial component to approximate the GDM. This was justified because the axial dipole is the component that dominates the geomagnetic shielding of galactic cosmic rays (Masarik and Beer 1999). The covariance function for changes in the axial component is given by:

$$k(r) = \frac{\sigma^2}{2\xi} * \left((\chi + \xi) e^{-(\chi-\xi)|r|} - (\chi - \xi) e^{-(\chi+\xi)|r|} \right), \tag{17}$$

with $\xi^2 = \chi^2 - \omega^2$. χ and ω are parameters representing frequencies. σ^2 and r, again, symbolize signal variance and difference in time. The power spectral density of the covariance function is as follows:

$$P(f) = \frac{4\chi\omega^2\sigma^2}{\left(\omega^2 - (2\pi f)^2\right)^2 + (4\pi\chi f)^2}. \tag{18}$$

The power spectrum (P) has an arc shape with the power decreasing with increasing frequency. It is often divided into several frequency ranges which can then be approximated locally with various spectral indices (p) (Bouligand et al. 2016; Hellio and Gillet 2018):

$$P(f) \propto f^{-p}. \tag{19}$$

At very low frequency where $p \cong 0$, the spectrum is almost flat indicating that it has the largest power at long-term periods such as periods longer than 50,000 years as shown in Additional file 1: Fig. S5. The power decreases faster at shorter periods (i.e., larger spectral indices). The decrease in power spectrum is simulated by changes in the slope which are determined by the cut-off frequencies T_s and T_f. The cut-off frequencies indicate the time periods where transitions into steeper slopes occur and their formulas are given by Bouligand et al. (2016):

$$T_s = \frac{2\pi(\chi + \xi)}{\omega^2}, \tag{20}$$

$$T_f = \frac{2\pi(\chi - \xi)}{\omega^2}. \tag{21}$$

χ and ω are then chosen to achieve the desired cut-off frequencies.

A recent geomagnetic field model COV-LAKE spanning the last 3000 years has $T_s \sim 100{,}000$ years and $T_f \sim 60$ years (Hellio and Gillet, 2018). The model is based on measurements of the magnetic field directions and intensity stored in archeological artifacts, igneous rocks and sediment records. The power spectral densities for the axial component of COV-LAKE are shown in Additional file 1: Fig. S5 and the GDM provided by COV-LAKE is shown in Fig. 10. The COV-OBS. × 2 model (Additional file 1: Fig. S5) has the same T_s but higher T_f (~235 years). This earlier transition (at longer time period) in the slope resulted in lower power for variations on timescales shorter than 200 years.

We tested generating a prior distribution of GDM using T_s and T_f from COV-LAKE and COV-OBS. × 2 models (results are shown in Additional file 1: Fig. S6). However, comparisons to the reconstructed GDM based on COV-LAKE suggest the prior distribution is rather

conservative. We also compared our prior distribution with pfm9k.1b another major reconstruction for the last 2000 years (Nilsson et al. 2014; Muscheler et al. 2016). pfm9k.1b provides a low temporal GDM reconstruction (300–400 years) over the Holocene based on almost the same underlying dataset as COV-LAKE (COV-LAKE was extended to include additionally new sediment and archeological intensity data). The GDM reconstructed by pfm9k.1b over the last 2000 is shown in Fig. 10. The GDM indicated by pfm9k.1b was mostly above and outside the prior distribution (Additional file 1: Fig. S6). Too conservative prior distributions could result in a bias toward values of GDM lower than the range seen in the previous GDM models. Therefore, we adjusted the parameters to widen our prior distribution (Additional file 1: Fig. S6c) and allow for more variability in the prior for GDM. We used a lower T_s of 50,000 years, a higher T_f of 433 years and also a larger signal variance. The variations for timescales between 100 and 10,000 years of our prior are larger compared to the prior used for COV-OBS. × 2 (Additional file 1: Fig. S5). For variations shorter than 100 years, our prior agrees with the prior used for COV-OBS. × 2. The parameters of COV-OBS. × 2, COV-LAKE and from this study for the axial component are shown in Additional file 1: Table S1. The variations of the axial dipole are insignificant at timescales shorter than 20 years. Therefore, the prior distribution of the GDM can be used directly without any adjustment (e.g., smoothening) of the covariance matrix. In addition, the prior distribution of GDM was connected to present GDM from 1939 to 2020 CE predicted by COV-OBS. × 2. We generated 1000 realizations of GDM simultaneously with ϕ to estimate the prior distribution (Fig. 7).

Results of palaeomagnetic field models could be used to further constrain the prior distribution. This would reduce the reconstruction uncertainty for periods where past GDM variations are well constrained, but would introduce uncertainties associated with the chosen palaeomagnetic field model. Disagreement in GDM reconstructed by different palaeomagnetic field models would lead to discrepancies in solar activity reconstructions. In addition, this also defeats our purpose of being independent of GDM models and providing a radionuclide-based reconstruction for GDM. Therefore, we did not further constrain our prior distribution based on results of palaeomagnetic field models.

Evaluation of the proposed solar activity and GDM

The samples of ϕ and M drawn from the prior distribution were evaluated via the likelihood function:

$$Q_t \sim N\left(f(\phi_t, M_t), \sigma^2_{Q,t}\right), \tag{22}$$

where Q_t is the data vector of the observed global production rates and $\sigma^2_{Q,t}$ is the vector representing uncertainty (i.e., variance) of the observed data. ϕ_t and M_t are the vectors of the solar modulation potential and GDM, proposed by the Bayesian model as the solution for Q_t. The link between Q_t and ϕ_t and M_t [i.e., $f(\phi_t, M_t)$] was established based on the global production rate function in Kovaltsov et al. (2012) (see Fig. 3 and section 1 in the Additional file 1). In summary, the Bayesian model combines our knowledge about the parameters (i.e., incorporated in the prior distributions) with the additional constraints provided by the observations (here, the global production rate) to yield a new (posterior) distribution of the parameters.

The prior distributions allow for an overlap of geomagnetic field and solar variability, which is the main challenge our method addresses. However, this leads to the fact that the reconstruction for geomagnetic field and solar variability on timescales significantly longer than 200 years is to some extent ambiguous since the prior distribution of solar activity was established based on variations observed in the 400-year-long GSN record. Nevertheless, the prior distribution is based on flexible SE covariance functions so that the choice of timescales does not prevent the model from finding longer periodicities, if the data requires them. The power spectrum of the tuned covariance function in Fig. 6 is essentially flat for longer timescales, but presumably the actual power spectrum decreases at very long periods. The model will compare and decide if the variations in the radionuclide record at timescales longer than 200 years can or cannot be explained by geomagnetic field variations (characterized in the prior distribution) and, if not, the model will likely consider those as solar variations. Overall, the model has larger uncertainties for variations on timescales longer than 200 years, but the prior information allows for some separation of solar and geomagnetic field influences also on these timescales. This is an advantage of the model over a simple band-pass frequency filter in disentangling solar and geomagnetic field variations on timescales longer than 200 years.

We also evaluate the correlation coefficient ($r_{\phi t,Mt}$) between the proposed curves of ϕ and GDM:

$$r_{\phi_t,M_t} \sim N\left(\mu_r, \sigma_r^2\right). \tag{23}$$

We set the mean (μ_r) and standard deviation (σ_r) of the correlation coefficient equal to 0 and 0.01, respectively. This ensures the independence of the reconstructed solar activity and geomagnetic field strength which was the initial assumption of the Bayesian approach (Eq. 6). It is worth mentioning that we tested different values of σ_r and chose 0.01 as the value

providing the best independence constraint. σ_r larger than 0.01 would result in ϕ and GDM correlating more strongly than we expect (we do not expect a link between solar activity and GDM variations), while σ_r lower than 0.01 would be too severe a constraint. The model will then reject the majority of the proposed samples albeit some low correlation can occur just by coincidence. It is also worth mentioning that some chance correlation between ϕ and GDM is more likely if we investigate a short period of time (e.g., shorter than 1000 years), especially if both processes exhibit a long-term trend over the investigated period.

Testing the model with synthetic data

Generating synthetic data

Here we aim at testing how well the model can recover solar activity and GDM from ^{14}C production rates that were calculated from these synthetic solar and geomagnetic field records. The synthetic data were generated from the prior distribution of ϕ and GDM and therefore it contained only variations that were included in the model. Therefore, the model was tested in a control scenario where the data did not contain unknown patterns. This is an important step in model validation before running the model with a real dataset with solar and geomagnetic field variations that are unknown.

Firstly, we randomly generated a realization of solar activity (ϕ') and GDM (M') using the prior distribution. These realizations were considered as the reference (i.e., "true") values which, after adding the assumed uncertainties, the Bayesian model was challenged to reconstruct. Figure 7a and b shows the prior distributions (50 random realizations and a 2-sigma envelope of a thousand realizations), and the reference ϕ' and M'. The differences between solar variations with annual resolution and 10-year resolution are visible in the solar realizations in Fig. 7a. Before 1600 CE, only variability on longer timescales such as centennial variations is left but no 11-year variations as these short-term variations were largely removed due to the low sampling resolution. After 1600 CE, the 11-year solar cycle can be observed. Occasionally, the model generated ϕ values larger than 2000 MV which is significantly larger than the values of ϕ_{HE16}. This is a consequence of the log-normal distribution as discussed above. However, most of the values were below 1300 MV as indicated by the 2-sigma envelope (black dashed line). The prior distribution of GDM (Fig. 7b) shows insignificant short-term variations and larger millennial variations. A synthetic ^{14}C production rate (Fig. 7c) resulting from ϕ' and M' was computed. We then resampled the synthetic production rate (orange circles with associated 1-sigma errors in Fig. 7c) assuming it followed a normal distribution with a standard deviation similar to the standard error of our inferred ^{14}C production rate from IntCal20. This allows us to test our model with realistic levels of uncertainty. The standard error of the ^{14}C production rate (from 1 to 1950 CE) ranges from 2.1% to 6.7%. We assumed the worst-case scenario and used 6.7% uncertainty to resample the synthetic ^{14}C production rate.

Assessing model performance

For a given model parameter (θ_i), we estimate the posterior mean ($\widehat{\theta_i}$) and variance ($\widehat{\sigma}^2_{\theta_i}$) directly from the Markov chain Monte Carlo (MCMC) samples:

$$\widehat{\theta_i} = \frac{1}{N}\sum_{j=1}^{N}\theta_i^{(j)}, \tag{24}$$

$$\widehat{\sigma}^2_{\theta_i} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\theta_i^{(j)} - \widehat{\theta_i}\right)^2, \tag{25}$$

with N being the sample population. In this study, we generated a thousand MCMC samples (N=1000) to estimate the posterior distribution of ϕ and M. We assess our model performance using two diagnostics following (Hellio and Gillet 2018; Nilsson and Suttie 2021), the normalized dispersion ($\Phi_{\theta i}$) and the normalized error ($\Psi_{\theta i}$). The normalized dispersion equals to the standard deviation of the posterior divided by the standard deviation of the prior:

$$\Phi_{\theta_i} = \frac{\widehat{\sigma}_{\theta_i}}{\sigma_{\theta_i}}. \tag{26}$$

The squared normalized dispersion ($\Phi^2_{\theta i}$), also known as the shrinkage factor, measures the amount of information contributed by the observations to the prior distribution. $\Phi^2_{\theta i}$ equal to 1 indicates that no information was added. The second parameter, the normalized error, is equal to the absolute difference between the posterior mean and the reference value (θ_i') (i.e., the selected ϕ' and M', blue lines in Fig. 7a and b) normalized by the standard deviation of the posterior:

$$\Psi_{\theta_i} = \frac{\left|\widehat{\theta}_i - \theta_i^{'}\right|}{\widehat{\sigma}_{\theta_i}}. \tag{27}$$

The normalized error measures the accuracy of the prediction with $E[\Psi_{\theta i}]$=1 indicating that the reference values are mainly within the uncertainty of the prediction values. The root mean squared (RMS) values of $\Phi_{\theta i}$ and $\Psi_{\theta i}$ for ϕ and M are shown in Additional file 1: Table S2.

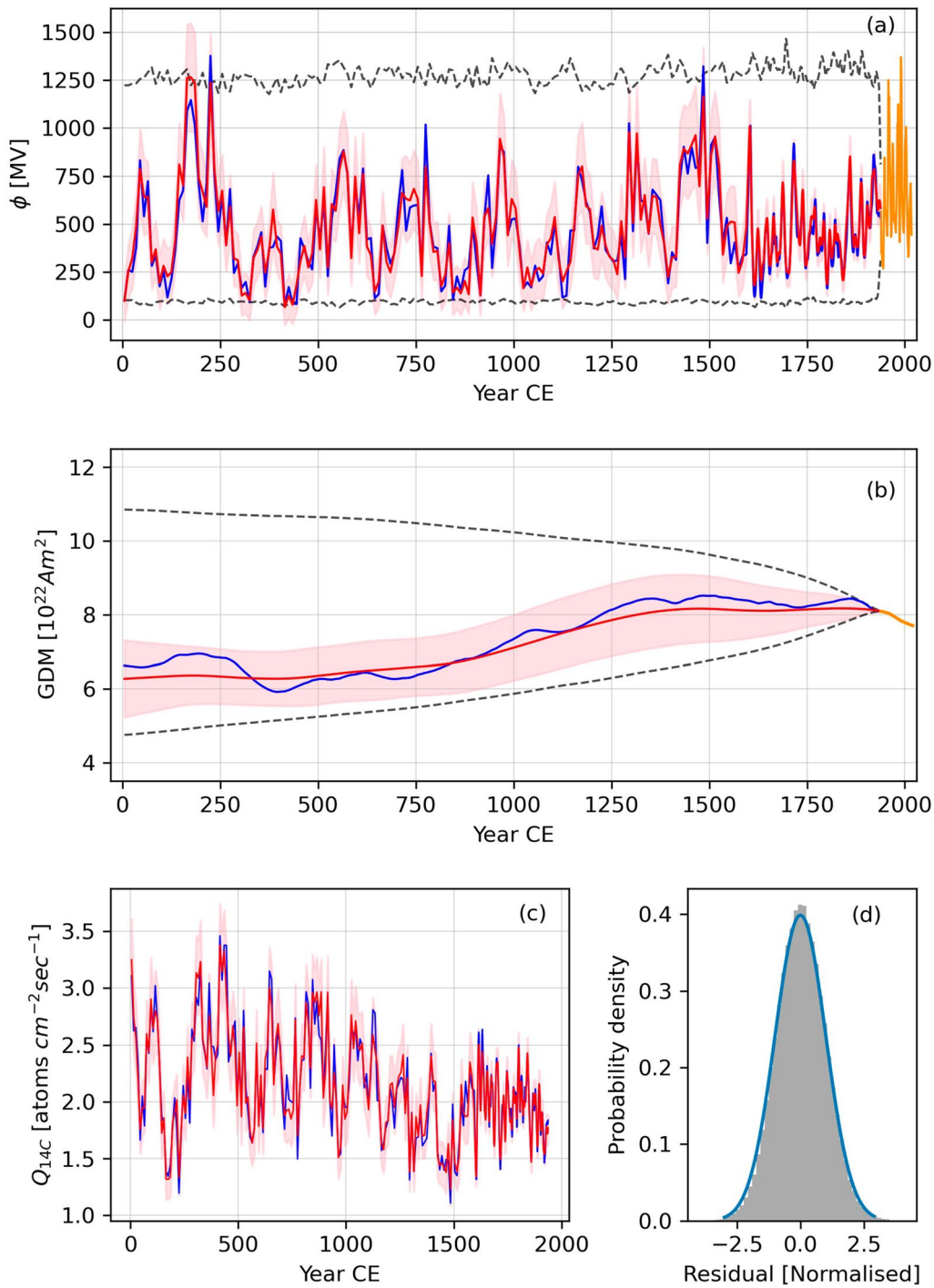


Fig. 8 Assessing solar activity and GDM reconstruction of the Bayesian model with the synthetic dataset. Posterior distributions of solar activity (**a**) and GDM (**b**) depicted with means (red lines) and 2-sigma uncertainty of a thousand realizations (light red shadings). The prior distributions are also depicted with the 2-sigma uncertainty (black dashed lines). The blue lines represent the reference φ' and M' and the orange lines represent the recent values that the model is connected to, i.e., φ_{HE16} and GDM based on COV-OBS. × 2. **c** Shows the mean (red line) and 2-sigma uncertainty (light red shading) of a thousand ^{14}C production rate curves generated by the recovered φ and M. The reference ^{14}C production rate is represented by the blue line. **d** Shows the histogram of the model-data residuals (grey bars) defined as the differences between individual posterior realizations of ^{14}C production rate and the input ^{14}C data, normalized by the data uncertainties. A standard normal distribution is shown as the blue line for comparison

The posterior distribution of ϕ was well estimated and constrained as indicated by a low value of Φ_{RMS} (Additional file 1: Table S2). Figure 8a shows that the posterior distribution of ϕ (i.e., the 2-sigma uncertainty envelope) is significantly smaller than the 2-sigma envelope of the prior. On the other hand, less information was added to the prior distribution of GDM as indicated by Φ_{RMS} equal to 0.73. Figure 8b shows that the posterior distribution of GDM was almost the same as the prior distribution in the recent period, for example after 1500 CE. This suggests that the ^{14}C production rate could not help improve the estimation of GDM for the recent period. The posterior distribution of GDM was smaller than the prior distribution further back in time indicating that GDM was better constrained during the earlier periods. On the other hand, the values of Ψ_{RMS} just below 1.0 (Additional file 1: Table S2) demonstrate that the original solar activity and GDM (Fig. 8a and b) were mainly within the reconstruction uncertainty.

Another important validation is the ^{14}C production rate generated by the recovered ϕ and M. If the model performed well, the recovered ^{14}C production rate will be within uncertainty of the reference ^{14}C production rate. Figure 8c compares the reference ^{14}C production rate with mean of the recovered ^{14}C production rate. The reference values are well within the posterior 2-sigma uncertainty and mostly agree with the recovered production rate. Figure 8d shows the histogram of the model-data residuals (i.e., the differences between individual posterior realizations and the synthetic ^{14}C data, normalized by the data uncertainties). The symmetrical distribution of the model-data residuals resembles a standard normal distribution. This indicates that the model does not lead to a bias toward a low or high value of ^{14}C production rate, and there are no extreme values or outliers in the recovered production rate. Overall, our model performed well with the synthetic data. The solar variations and GDM can be recovered from the ^{14}C production rate corrupted by a realistic level of measurement uncertainty.

We also estimated the upper limit of data uncertainty by corrupting the synthetic ^{14}C production rate with various uncertainty levels from 15 to 70% (Additional file 1: Fig. S7). The reference solar variation (ϕ') was not fully recovered by the model when the data uncertainty was at 20% and larger. An increase in the data uncertainty also resulted in larger Φ_{RMS} and reconstruction uncertainty which indicates that the model's ability to constrain past solar activity decreased. In summary, we conclude that data uncertainty of less than 20% is required for a good model performance.

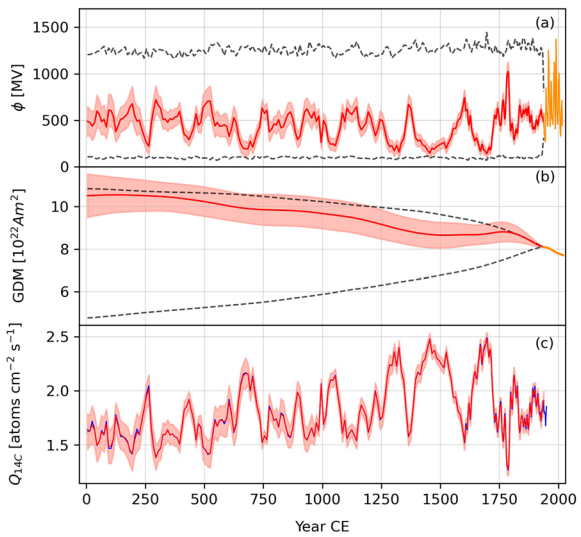


Fig. 9 Solar activity and GDM reconstructed by the Bayesian model for the last 2000 years. Posterior distributions of solar activity (**a**) and GDM (**b**) are depicted with mean (red lines) and 2-sigma uncertainty of a thousand realizations (light red shadings). The 2-sigma uncertainty of a thousand realizations from the prior distributions are indicated by two black dashed lines. Panel c shows the mean (red line) and 2-sigma posterior uncertainty (light red shading) of a thousand curves of ^{14}C production rate generated by the recovered φ and M. The input mean ^{14}C production rate is represented by the blue line. The uncertainty of the input ^{14}C production rate (not shown in the figure) is similar to the posterior uncertainty (light red shading). Figure 2 shows the uncertainty of the input ^{14}C production rate before normalization

Reconstruction of solar and geomagnetic field activity from ^{14}C data

Bayesian reconstruction

We applied the Bayesian model to recover solar activity and GDM variations from the processed ^{14}C production rate data inferred from IntCal20. Results of the Bayesian ^{14}C-based reconstruction are shown in Fig. 9. Short-term variations (from decadal up to ~ 300 years) in the ^{14}C data were mostly attributed to solar variations. For example, the increase and then decrease in ^{14}C production rate between ~ 1350 CE and 1600 CE was interpreted by the model as solar induced. Meanwhile, the long-term increase in the production rate since 1 CE was attributed to the gradual decrease in GDM. The posterior distribution of past geomagnetic field activity was constrained better (i.e., smaller uncertainty range) by the real ^{14}C data than by the synthetic ^{14}C data. One reason was that the synthetic dataset has larger uncertainties since, for testing the model, we included the maximum value of the realistic uncertainty. Figure 9c compares the ^{14}C

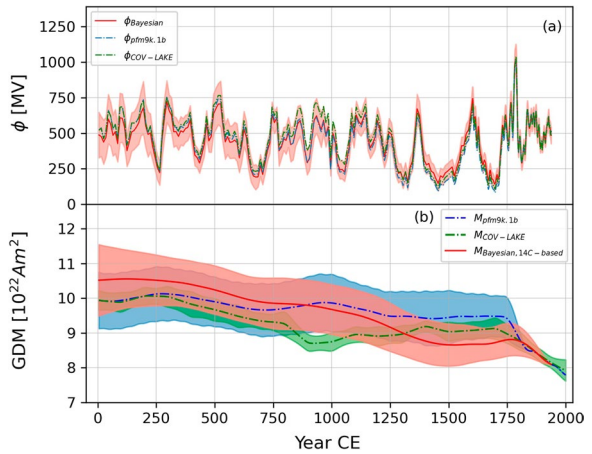


Fig. 10 A comparison between the Bayesian model's results with conventional reconstruction method and independent GDM models. **a** Compares different solar activity reconstructions from the Intcal20 ^{14}C data by the Bayesian model (red line) with the conventional method in which the influences of GDM were removed using the pfm9k.1b (blue dashed line) and the COV-LAKE geomagnetic field model (green dashed line). The light red shading indicates the 2-sigma uncertainty of the Bayesian reconstruction. **b** Compares GDM recovered from the ^{14}C data by the Bayesian model (red color) with GDM from the pfm9k.1b models (blue color) and the COV-LAKE models (green color). The lines depict the mean GDM and the color shadings depict the 2-sigma uncertainties of the models

production rate generated from the reconstructed ϕ and GDM with the input ^{14}C data. They agree well (within the posterior uncertainty) which indicates that the model is able to find the combinations of ϕ and M that are consistent with the ^{14}C data.

Comparing the Bayesian reconstruction with the conventional reconstruction method

Figure 10b compares our Bayesian ^{14}C-based reconstruction of GDM with published reconstructed GDMs based on pfm9k.1b and COV-LAKE. The smaller uncertainty of COV-LAKE during the last 2000 years compared to pfm9k.1b is mainly due to the more conservative assumptions made regarding age uncertainties in pfm9k.1b and to a lesser degree related to the different model strategies. In general, all of the reconstructions demonstrate a decreasing trend over the last 2000 years. Disagreements among the reconstructions can be observed especially from around 850–1750 CE. The COV-LAKE models indicate a small dip in GDM with an average of 8.85 ($\pm$0.12) x10^{22}Am^2 from 850 to 1250 CE. Meanwhile, the pfm9k.1b models suggest a small peak with an average of 9.74 ($\pm$0.38) $\times$ 10^{22}Am^2. Our Bayesian ^{14}C-based reconstruction shows a gradual decrease in GDM from 9.81 ($\pm$0.33) $\times$ 10^{22}Am^2 to 9.23 ($\pm$0.32) $\times$ 10^{22}Am^2 over the same period. From 1250 to 1750 CE, the Bayesian ^{14}C-based reconstruction indicates a mean GDM of around 8.75 ($\pm$0.26) $\times$ 10^{22}Am^2 which is 0.33×10^{22}Am^2 and 0.72×10^{22}Am^2 lower than COV-LAKE and pfm9k.1b, respectively. In addition, the Bayesian ^{14}C-based reconstruction indicates a mean GDM of 10.50 ($\pm$0.40) $\times$ 10^{22}Am^2 prior to 500 CE which is 0.50×10^{22}Am^2 higher than the mean GDMs based on pfm9k.1b and COV-LAKE during the same period.

We now compare solar activity reconstructed from the ^{14}C data via our Bayesian model (ϕ_{Bayesian}) with the conventional reconstruction method described in Muscheler et al. (2016). This conventional method involves conducting a larger number (e.g., often a thousand) of Monte Carlo simulations. In each simulation one of the realizations of ^{14}C production rate is randomly selected and combined with one of the randomly selected realizations from a GDM model via the ^{14}C production function. The mean of a thousand simulations of past solar activity reconstructed from the ^{14}C data and pfm9k.1b models ($\phi_{\text{pfm9k.1b}}$) or COV-LAKE models ($\phi_{\text{COV-LAKE}}$) are shown in Fig. 10a. The solar reconstructions presented here should be the most up-to-date for the last 2000 years based on ^{14}C since we combined the method of Muscheler et al. (2016) with an updated version of ^{14}C data from IntCal20 and the latest geomagnetic field reconstructions.

The different solar reconstructions agree mostly on short-term variations. Disagreements between the long-term solar activity variations can be observed where the reconstructed GDMs start to deviate from each other (Fig. 10b). Before 500 CE, ϕ_{Bayesian} is on average 496 ($\pm$63) MV, which is about 40 MV lower than the other two reconstructions. From 850 to 1250 CE, ϕ_{Bayesian} is 466 ($\pm$44) MV on average which is similar to $\phi_{\text{pfm9k.1b}}$ but about 70 MV lower than $\phi_{\text{COV-LAKE}}$. From 1250 to 1750 CE, the average of ϕ_{Bayesian} is 354 ($\pm$30) MV which is 20–50 MV larger than $\phi_{\text{COV-LAKE}}$ and $\phi_{\text{pfm9k.1b}}$, respectively. This trend is the opposite in the reconstructed GDMs and it illustrates the influences of the selected GDM on the long-term reconstruction of solar activity. However, ϕ_{Bayesian} still agrees with $\phi_{\text{pfm9k.1b}}$ and $\phi_{\text{COV-LAKE}}$ within the reconstruction uncertainties. Moreover, GDM reconstructions based on pfm9k.1b and COV-LAKE are mostly within the reconstruction uncertainty of the Bayesian ^{14}C-based GDM. These results suggest that the reconstructions of solar activity and GDM from ^{14}C data of our Bayesian model are realistic over the last 2000 years.

Figure 11 compares the solar activity reconstructions with the solar modulation potential based on GSN filtered with a 9-year running average filter ($\phi_{\text{GSN,filtered}}$, orange line). The centennial variations of the recovered solar activity from different reconstructions generally agree

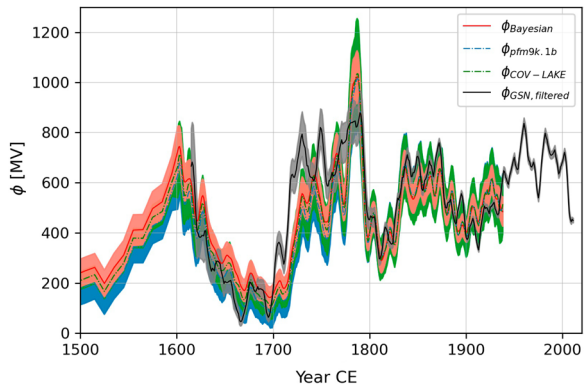


Fig. 11 A close-up comparison between different reconstructions of solar activity with φ_{GSN}. The 9-year running average version of φ_{GSN} is depicted with mean (black solid line) and 2-sigma uncertainty (grey shading). The red line and red shading depict the mean and 2-sigma uncertainty of the Bayesian ^{14}C-based reconstruction. The blue dashed line and blue shading, and the green dashed line and green shading depict the means with their 2-sigma uncertainty of the conventional reconstructions where the influences of GDM were removed using the pfm9k.1b and COV-LAKE models, respectively

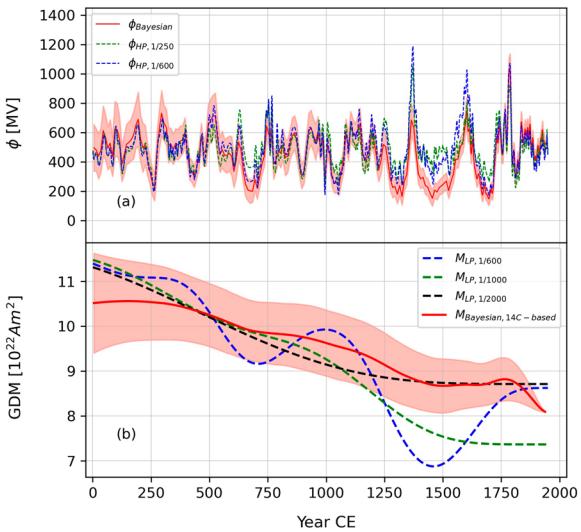


Fig. 12 A comparison between reconstruction by the Bayesian method and reconstruction by a simple frequency filter. Panel a compares solar activity reconstructed by the Bayesian method (red line) and the results using high-pass filters with cut-off frequencies of 1/250 year (green dashed line) and 1/600 year (blue dashed line), respectively. Panel b compares the GDM reconstructed by the Bayesian method (red line) and the results using low-pass filters with cut-off frequencies of 1/600 year (blue dashed line), 1/1000 year (green dashed line) and 1/2000 year (black dashed line), respectively. The light red shadings indicate the 2-sigma uncertainty envelope of the Bayesian reconstruction

well with $\varphi_{GSN,filtered}$. However, differences outside the inferred uncertainty can be observed around 1700–1800 CE. $\varphi_{GSN,filtered}$ suggests solar activity levels, on average, 110–150 MV higher than all of the reconstructions from the ^{14}C data. These disagreements could be due to the limitation of our Bayesian method, as discussed in Sect. 3, such as the small bias toward lower solar modulation values of our prior distribution. However, since our Bayesian reconstruction agrees well with conventional solar reconstructions (i.e., $\varphi_{pfm9k.1b}$ and $\varphi_{COV-LAKE}$), we argue that the disagreements were more likely due to an underestimation of uncertainties of the ^{14}C data and φ_{GSN} inferred from the GSN record. We used the Northern hemisphere ^{14}C curves and there are differences to the Southern hemisphere records (Muscheler et al. 2007). It is also possible that the carbon cycle effects are not fully captured by the ^{14}C production rate calculation with the box-diffusion model. Possible uncertainty in this calculation is hard to quantify but, in general, a good agreement is obtained by calculations with different carbon cycle models (Muscheler et al. 2007). The subtle changes in the carbon cycle that were not fully captured by the box-diffusion model can be explored by adding such a component to future versions of the model and comparing the results from ^{14}C to ^{10}Be data. On the other hand, the uncertainty of the inferred φ_{GSN} was likely also underestimated. The standard error of the GSN data represented only the spread among different counting records (Svalgaard and Schatten 2016).

Figure 11 also shows that $\varphi_{Bayesian}$ has a lower uncertainty compared to the two other reconstructions during the period where the model runs with annual resolution (e.g., after 1600 CE). Although the Bayesian ^{14}C-based GDM has similar or even larger uncertainty than pfm9k.1b and COV-LAKE (Fig. 10b), the reconstruction uncertainty of GDM did not directly affect the reconstruction uncertainty of solar activity in the Bayesian model. On the other hand, the uncertainties of $\varphi_{pfm9k.1b}$ and $\varphi_{COV-LAKE}$ are a direct consequence of the uncertainty in GDM (see also Additional file 1: Fig. S8). This shows that the Bayesian model is able to reduce the solar activity reconstruction uncertainty via utilizing the knowledge of the differences in rates of change between variations of GDM and solar activity. The differences in variations are biggest after 1600 CE since significant short-term solar activity variations are captured by the annual resolution. Consequently, the reconstruction uncertainty was reduced the most.

Comparing the Bayesian reconstruction with reconstruction using frequency filters

In the following, we illustrate the differences between the Bayesian ^{14}C-based reconstruction and reconstruction by applying various frequency filters to separate solar and geomagnetic field influences.

Figure 12a illustrates solar activity reconstructed from variations of the ^{14}C data shorter than 250 years ($\phi_{HP,1/250}$) and 600 years ($\phi_{HP,1/600}$). The short-term variations were extracted using two separate high-pass frequency filters with cut-off frequencies of 1/250 years^{-1} and 1/600 years^{-1} to ensure reconstruction of solar activity on timescales of around 200 years and 500 years, respectively (details in section 5, Additional file 1). It should be noted that we applied the filters on the originally annual ^{14}C data (blue line in Fig. 2) and therefore differences at timescales shorter than 10 years can be observed between $\phi_{Bayesian}$ and $\phi_{HP,1/250}$, and between $\phi_{Bayesian}$ and $\phi_{HP,1/600}$ prior to 1600 CE (Fig. 12a). The deviation of $\phi_{HP,1/250}$ from $\phi_{HP,1/600}$ and $\phi_{Bayesian}$ indicates possible solar variations at timescales between 250 and 600 years. For example, around 630–700 CE and around 1650–1720 CE, $\phi_{HP,1/250}$ is about 100 to 170 MV larger than mean $\phi_{Bayesian}$ and $\phi_{HP,1/600}$. Possible solar variability at timescales significantly longer than 500 years could have caused $\phi_{HP,1/600}$ to be around 185 MV higher than mean $\phi_{Bayesian}$ around 1370–1600 CE. These differences are all outside the reconstruction uncertainty of the Bayesian model. Thus, the Bayesian model indicates long-term solar variations at timescales larger than 200 years and even larger than 500 years that were removed/cut-off by the frequency filters. This result also supports long-term solar variability patterns inferred from the radionuclide records in previous studies (Wagner et al. 2001; Snowball and Muscheler 2007; Adolphi et al. 2014).

Figure 12b shows GDM reconstructions from variations of the ^{14}C data longer than 600 years ($M_{LP,1/600}$), 1000 years ($M_{LP,1/1000}$) and 2000 years ($M_{LP,1/2000}$). We used three separate low-pass frequency filters with cut-off frequencies of 1/600 years^{-1}, 1/1000 years^{-1} and 1/2000 years^{-1} to extract the possible geomagnetic field signal (details in section 5, Additional file 1). The relatively large variations of $M_{LP,1/600}$ for the last 2000 years suggest an uncorrected solar variability influence at timescales from 600 to 1000 years (Fig. 12b). This is also supported by disagreement between solar activity reconstruction by the Bayesian model and the frequency filters from 1250 to 1650 CE (Fig. 12a). In addition, $M_{LP,1/1000}$ suggests a strongly decreasing trend especially after 1250 CE where our Bayesian ^{14}C-based reconstruction as well as pfm9k.1b and COV-LAKE suggest a rather constant and higher value of the GDM. This is potentially due to the fact that $M_{LP,1/1000}$ still contains solar influences at millennial timescale plus influences of the end effects of the frequency filter. It is possible to normalize $M_{LP,1/1000}$ to the average value of the Bayesian ^{14}C-based reconstruction after 1250 CE. However, by doing so it will increase the value of $M_{LP,1/1000}$ prior to 1250 CE to a higher level not supported by independent geomagnetic field models. Figure 12b also demonstrates another practical problem of the frequency filter as the method in general exhibits unreliable end effects. For example, variations before 200 CE and after 1750 CE cannot be recovered with $M_{LP,1/1000}$ and $M_{LP,1/2000}$. For this to be possible, one would need the data to cover much longer periods than the actual reconstruction period. Therefore, it is difficult to connect the reconstructions to the present values.

In summary, Fig. 12 illustrates the advantages of the Bayesian model over simple frequency filters. Frequency filters are useful in general to partially remove GDM influences for studying short-term solar activity variations in radionuclide records and vice versa for studying millennial variations of the GDM. However, they can never completely separate the long-term solar activity from the GDM changes as their variability ranges partly overlap at centennial and possibly millennial timescales. On the other hand, the Bayesian model can separate solar and GDM effects on these overlapping timescales despite the limitation of using the relatively short GSN record to constrain the prior information of solar activity.

Conclusion and outlook

We have introduced a Bayesian model that can separate solar and geomagnetic influences on radionuclide data using prior information on how solar activity and GDM vary through time. Here, we derived prior information on solar variability from a solar modulation reconstruction inferred from the group sunspot number record. The prior distribution of the GDM was adapted from previously proposed priors used in recent geomagnetic field models, i.e., COV-LAKE and COV-OBSx2.

Our model performs well with the synthetic test and can reconstruct the reference solar activity and GDM from a synthetic ^{14}C dataset corrupted with realistic measurement uncertainty. Applying the Bayesian model on the ^{14}C production rate data inferred from the IntCal20 calibration curve resulted in a reconstructed GDM which was gradually decreasing over the period of the last 2000 years. The Bayesian ^{14}C-based GDM agrees mostly with independent reconstructions using the pfm9k.1b and COV-LAKE geomagnetic field models. The solar activity reconstructed by the Bayesian model also agrees with conventional reconstructions where GDM influences were removed using pfm9k.1b and COV-LAKE models. The solar activity reconstructed by the Bayesian model shows similar annual short-term variations as the solar activity inferred from the GSN. There were, however, differences in the long-term variations outside the reconstruction uncertainty. This is probably due to underestimation of the uncertainty in the underlying ^{14}C data (e.g., carbon cycle effect) and GSN data.

We also showed that our Bayesian model outperforms various simple frequency filters. The Bayesian model is able to disentangle solar and GDM influences on the ^{14}C record on timescales where their variability ranges partly overlap. In addition, a comparison between the Bayesian reconstruction with the reconstructions based on the frequency filters indicates that the Bayesian model can recover solar activity on timescales longer than 200 years.

In summary, the Bayesian model allows us to disentangle solar and GDM influences from the radionuclides data. This reduces the dependency of solar activity reconstructions on an independent GDM record and, therefore, can reduce the uncertainties associated with the independent GDM. Moreover, the Bayesian model can provide radionuclide-based GDM reconstructions which are valuable compliments to other GDM reconstructions.

The flexibility of the Bayesian framework outlined in this paper also allows for further improvements in the future. For example, independent GDM reconstructions could be incorporated into the model. This will help with the reconstruction of solar activity during periods where the GDM is well constrained. Moreover, more than one radionuclide dataset can be included into the model such as using several ^{10}Be records from different ice cores in addition to the ^{14}C data, or using a global compilation of ^{10}Be records. Including the different geochemical behavior of these radionuclides might help us to estimate the factors leading to the differences in long-term solar activity reconstructions based on ^{10}Be and ^{14}C as seen in Vonmoos et al. (2006). Prior information on the systematic influences such as changes in climate and carbon cycle could be incorporated into the model so that their signals can also be separated from the radionuclide data. This could further reduce the solar and geomagnetic field reconstruction uncertainties. In addition, the model can also be extended with ^{10}Be records from sediments. Records from these archives often contain non-production signals caused by the local processes and catchment conditions which can be separated via incorporating these processes into the model.

Abbreviations

GCRs: Galactic cosmic rays; GP: Gaussian process; GDM: Geomagnetic dipole moment; GSN: Group sunspot number; HMC: Hamiltonian Monte Carlo; LIS: Local interstellar spectrum; MCMC: Markov chain Monte Carlo; NUTS: No-U-Turn sampler; RMS: Root mean squared; SPE: Solar proton event; SE: Squared exponential.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40623-022-01688-1.

Additional file 1. Supplementary document.

Acknowledgements

We would like to thank Tim Heaton for providing the D14C realisations. We would also like to thank two anonymous reviewers whose constructive comments improved the manuscript, and to thank the editors Takeshi Sagiya and Yuhji Yamamoto for editing and handling the manuscript.

Author contributions

All the authors have made substantial contribution to conception and design of the work. LN constructed the model used in the work and the analysis of the data. All authors interpreted the results of the analysis. LN has drafted the work and the authors NS, AN and RM have revised the work. All authors read and approved the final manuscript.

Funding

Open access funding provided by Lund University. This work was supported by the Swedish Research Council (Grant DNR2013-8421 to R. Muscheler and Grant DNR2020-04813 to A. Nilsson).

Availability of data and materials

The datasets used in this study (e.g., solar modulation potential, group sunspot number, global dipole moments) are available in the corresponding references/publication. The datasets generated during this study including the solar modulation inferred from group sunspot number, the production rate of ^{14}C inferred from IntCal20 and the synthetic dataset are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 7 April 2022 Accepted: 8 August 2022
Published online: 24 August 2022

References

Adolphi F, Muscheler R, Svensson A, Aldahan A, Possnert G, Beer J, Sjolte J, Björck S, Matthes K, Thieblemont R (2014) Persistent link between solar activity and Greenland climate during the Last Glacial Maximum. Nat Geosci 7:662–666. https://doi.org/10.1038/NGEO2225

Beer J, Siegenthaler U, Bonani G, Finkel RC, Oeschger H, Suter M, Wölfli W (1988) Information on past solar activity and geomagnetism from 10Be in the Camp Century ice core. Nature 331:675–679. https://doi.org/10.1038/331675a0

Beer J, McCracken K, von Steiger R (2012) The cosmic radiation near earth. Cosmogenic radionuclides. Springer, Berlin, Heidelberg, pp 19–78

Bond G, Kromer B, Beer J, Muscheler R, Evans MN, Showers W, Hoffmann S, Lotti-Bond R, Hajdas I, Bonani G (2001) Persistent solar influence on north Atlantic climate during the Holocene. Science 294(5549):2130–2136. https://doi.org/10.1126/science.106568

Bouligand C, Gillet N, Jault D, Schaeffer N, Fournier A, Aubert J (2016) Frequency spectrum of the geomagnetic field harmonic coefficients from dynamo simulations. Geophys J Int 207:1142–1157. https://doi.org/10.1093/gji/ggw326

Brehm N, Bayliss A, Christl M, Synal HA, Adolphi F, Beer J, Kromer B, Muscheler R, Solanki SK, Usoskin I, Bleicher N, Bollhalder S, Tyers C, Wacker L (2021) Eleven-year solar cycles over the last millennium revealed by radiocarbon in tree rings. Nat Geosci 14:10–15. https://doi.org/10.1038/s41561-020-00674-0

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker MA, Guo J, Li P, Riddell A (2017) Stan: A probabilistic programming language. J Stat Softw 76(1–32):1. https://doi.org/10.18637/jss.v076.i01

Dergachev VA, Vasiliev SS (2019) Long-term changes in the concentration of radiocarbon and the nature of the Hallstatt cycle. J Atmos Sol-Terr Phys 182:10–24. https://doi.org/10.1016/JJASTP.2018.10.005
Friis-Christensen E, Lassen K (1991) Length of the solar cycle: an indicator of solar activity closely associated with climate. Science 254:698–700. https://doi.org/10.1126/SCIENCE.254.5032.698
Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis, 2nd edn. Chapman & Hall CRC, London
Heaton TJ, Blaauw M, Blackwell PG, Bronk Ramsey C, Reimer PJ, Scott EM (2020) The IntCal20 approach to radiocarbon calibration curve construction: a new methodology using Bayesian splines and errors-in-variables. Radiocarbon 62:821–863. https://doi.org/10.1017/RDC.2020.46
Hellio G, Gillet N (2018) Time-correlation-based regression of the geomagnetic field from archeological and sediment records. Geophys J Int 214:1585–1607. https://doi.org/10.1093/GJI/GGY214
Herbst K, Muscheler R, Heber B (2017) The new local interstellar spectra and their influence on the production rates of the cosmogenic radionuclides 10Be and 14C. J Geophys Res Sp Phys 122:23–34. https://doi.org/10.1002/2016JA023207
Hoffman MD, Gelman A (2014) The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 15:1593–1623
Huder L, Gillet N, Finlay CC, Hammer MD, Tchoungui H (2020) COV-OBS.x2: 180 years of geomagnetic field evolution from ground-based and satellite observations. Earth Planets Sp 72:160. https://doi.org/10.1186/s40623-020-01194-2
Kovaltsov GA, Mishev A, Usoskin IG (2012) A new model of cosmogenic production of radiocarbon 14 C in the atmosphere. Earth Planet Sci Lett 337–338:114–120. https://doi.org/10.1016/j.epsl.2012.05.036
Masarik J, Beer J (1999) Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. J Geophys Res 104:99–112
Mekhaldi F, Muscheler R, Adolphi F, Aldahan A, Beer J, McConnell JR, Possnert G, Sigl M, Svensson A, Synal H-A, Welten KC, Woodruff TE (2015) Multiradionuclide evidence for the solar origin of the cosmic-ray events of AD 774/5 and 993/4. Nat Commun 6:1–8. https://doi.org/10.1038/ncomms9611
Miyake F, Nagaya K, Masuda K, Nakamura T (2012) A signature of cosmic-ray increase in ad 774–775 from tree rings in Japan. Nature 486:240–242. https://doi.org/10.1038/nature11123
Miyake F, Masuda K, Nakamura T (2013) Another rapid event in the carbon-14 content of tree rings. Nat Commun 4:1–6. https://doi.org/10.1038/ncomms2783
Muscheler R, Beer J, Kubik PW, Synal HA (2005) Geomagnetic field intensity during the last 60,000 years based on 10Be and 36Cl from the Summit ice cores and 14C. Quat Sci Rev 24:1849–1860. https://doi.org/10.1016/j.quascirev.2005.01.012
Muscheler R, Joos F, Beer J, Müller SA, Vonmoos M, Snowball I (2007) Solar activity during the last 1000 yr inferred from radionuclide records. Quat Sci Rev 26:82–97. https://doi.org/10.1016/j.quascirev.2006.07.012
Muscheler R, Adolphi F, Herbst K, Nilsson A (2016) The revised sunspot record in comparison to cosmogenic radionuclide-based solar activity reconstructions. Sol Phys 291:3025–3043. https://doi.org/10.1007/s11207-016-0969-z
Neal RM (2011) MCMC using Hamiltonian dynamics. In: Brooks S, Gelman A, Jones G, Meng X-L (eds) Handbook of Markov Chain Monte Carlo. CRC Press, Boca Raton
Nilsson A, Suttie N (2021) Probabilistic approach to geomagnetic field modelling of data with age uncertainties and post-depositional magnetisations. Phys Earth Planet Inter 317:106737. https://doi.org/10.1016/j.pepi.2021.106737
Nilsson A, Holme R, Korte M, Suttie N, Hill M (2014) Reconstructing Holocene geomagnetic field variation: new methods, models and implications. Geophys J Int 198:229–248. https://doi.org/10.1093/gji/ggu120
Petrovay K (2010) Solar cycle prediction. Living Rev Solar Phys 7(1):1–59. https://doi.org/10.12942/lrsp-2010-6
Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT Press, Massachusetts
Reimer PJ, Austin WEN, Bard E, Bayliss A, Blackwell PG, Bronk Ramsey C, Butzin M, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hajdas I, Heaton TJ, Hogg AG, Hughen KA, Kromer B, Manning SW, Muscheler R, Palmer JG, Pearson C, van der Plicht J, Reimer RW, Richards DA, Scott EM, Southon JR, Turney CSM, Wacker L, Adolphi F, Büntgen U, Capano M, Fahrni SM, Fogtmann-Schulz A, Friedrich R, Köhler P, Kudsk S, Miyake F, Olsen J, Reinig F, Sakamoto M, Sookdeo A, Talamo S (2020) The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon 62:725–757. https://doi.org/10.1017/RDC.2020.41
Schwabe SH (1844) Sonnenbeobachtungen im Jahre 1843. Dessau Astron Nachr 21:223
Siegenthaler U (1983) Uptake of excess CO2 by an outcrop-diffusion model of the ocean. J Geophys Res 88:3599–3608. https://doi.org/10.1029/JC088iC06p03599
Snowball I, Muscheler R (2007) Palaeomagnetic intensity data: an Achilles heel of solar activity reconstructions. The Holocene 17:851–859. https://doi.org/10.1177/0959683607080531
Solanki SK, Schüssler M, Fligge M (2002) Secular variation of the Sun's magnetic flux. Astron Astrophys 383(2):706–712. https://doi.org/10.1051/0004-6361:20011790
Svalgaard L, Schatten KH (2016) Reconstruction of the sunspot group number: the backbone method. Sol Phys 291:2653–2684. https://doi.org/10.1007/s11207-015-0815-8
Usoskin IG, Mursula K, Solanki SK, Schüssler M, Kovaltsov GA (2002) A physical reconstruction of cosmic ray intensity since 1610. J Geophys Res Sp Phys 107:13–21. https://doi.org/10.1029/2002JA009343
Usoskin IG, Alanko-Huotari K, Kovaltsov GA, Mursula K (2005) Heliospheric modulation of cosmic rays: monthly reconstruction for 1951–2004. J Geophys Res Sp Phys 110:12108. https://doi.org/10.1029/2005JA011250
Usoskin IG, Gallet Y, Lopes F, Kovaltsov GA, Hulot G (2016) Astrophysics Solar activity during the Holocene: the Hallstatt cycle and its consequence for grand minima and maxima. A&a 587:150. https://doi.org/10.1051/0004-6361/201527295
Vonmoos M, Beer J, Muscheler R (2006) Large variations in holocene solar activity: constraints from 10Be in the greenland ice core project ice core. J Geophys Res Sp Phys 111:A10105. https://doi.org/10.1029/2005JA011500
Wagner G, Beer J, Masarik J, Muscheler R, Kubik PW, Mende W, Laj C, Raisbeck GM, Yiou F (2001) Presence of the Solar de Vries Cycle (∼205 years) during the Last Ice Age. Geophys Res Lett 28(2):303–306. https://doi.org/10.1029/2000GL006116
Wieler R, Beer J, Leya I (2013) The galactic cosmic ray intensity over the past 106–109 years as recorded by cosmogenic nuclides in meteorites and terrestrial samples. Sp Sci Rev 176:351–363. https://doi.org/10.1007/s11214-011-9769-9
Zheng M, Sturevik-Storm A, Nilsson A, Adolphi F, Aldahan A, Possnert G, Muscheler R (2021) Geomagnetic dipole moment variations for the last glacial period inferred from cosmogenic radionuclides in Greenland ice cores via disentangling the climate and production signals. Quat Sci Rev 258:106881. https://doi.org/10.1016/j.quascirev.2021.106881

Publisher's Note