UCM candidate model for IGRF-14

Serrano, Mario[1, 2], Pavón-Carrasco, F. Javier[1, 2], Campuzano, Saioa A.[1, 2], Marsal, Santiago[3], Torta, J. Miquel[3], Tordesillas, J. Manuel[4], and Fiz, Julián[5]

[1]Universidad Complutense de Madrid
[2]Instituto de Geociencias (CSIC-UCM)
[3]Observatori de l'Ebre
[4]Instituto Geográfico Nacional
[5]Real Instituto y Observatorio de la Armada

September 30, 2024

1 Introduction and objectives

In this technical report, we detail the data and methodology used in the development of the 14th generation of the International Geomagnetic Reference Field (IGRF-14) by the leader team from Universidad Complutense de Madrid (UCM) and Instituto de Geociencias (IGEO, CSIC-UCM), which also includes members from the three institutions that currently operate the geomagnetic ground-based observatories in Spain. We present the candidates for the three required products:

1. DGRF-2020: A description of the main geomagnetic field for 2020.0, up to spherical harmonic degree 13.
2. IGRF-2025: A description of the main geomagnetic field for 2025.0, up to spherical harmonic degree 13.
3. SV-2025-2030: The predicted secular variation of the main geomagnetic field for the period 2025.0-2030.0, up to spherical harmonic degree 8.

2 Strategy

To provide these three products, we aim to develop two different one-year parent models that describe both the internal geomagnetic field and the large-scale external field for the following periods:

1. 1st July 2019 – 30th June 2020.
2. 15th September 2023 – 15th September 2024.

From the first one-year parent model (2019-2020), we derive the DGRF-2020 product, and from the second one-year parent model (2023-2024), we extrapolate to obtain the IGRF-2025 and SV-2025-2030 products.

3 Input data and selection criteria

We used data from the three satellites of the Swarm mission of the European Space Agency (ESA) and from geomagnetic ground-based observatories.

3.1 Satellite data

We used the SW_OPER_MAGx_LR_1B data product from the three Swarm satellites (Swarm-A, Swarm-B, and Swarm-C), utilizing the latest available version (baseline 0605). Scalar data were obtained from the Absolute Scalar Magnetometer (ASM), and vector data were obtained from the Vector Field Magnetometer (VFM) in the NEC reference frame (geocentric frame). Since Swarm-C does not provide ASM data, the scalar values were estimated using vector data.

To reduce the influence of external geomagnetic fields, we applied the following selection criteria to the data. The selection was based on several indices:

- The sun is at least 10^{o} below the horizon.
- $Kp \leq 2^{o}$ ($ap \leq 7$).
- $|Dst| \leq 30$ nT & $|dDst/dt| \leq 2$ nT/hr.
- $0 \leq$ IMF_BZ_GSM ≤ 6 nT.
- $-8 \leq$ IMF_BY_GSM ≤ 3 if QDLat $> 0°$ or $-3 \leq$ IMF_BY_GSM ≤ 8 if QDLat $< 0°$.
- Vector data for $|QDLat| < 55°$ and Scalar data for $|QDLat| > 55°$.
- $|F_{\text{VFM}} - F_{\text{ASM}}| < 3$ nT (except for Swarm-C).
- $|F_{\text{data}} - F_{\text{CHAOS-7.18}}| < 100$ nT.

Where IMF_BY_GSM and IMF_BZ_GSM represent the Y and Z components of the interplanetary magnetic field (IMF), Kp, ap, and Dst are different geomagnetic indices that serve as indicators of geomagnetic activity, and $QDLat$ refers to quasi-dipolar latitude.

For the 2019-2020 parent model, we obtained 5,652,033 vector data points (1,883,575 from Swarm-A, 1,883,540 from Swarm-B, and 1,884,918 from Swarm-C) and 3,186,851 scalar data points (1,040,573 from Swarm-A, 1,109,716 from Swarm-B, and 1,036,562 from Swarm-C).

For the 2023-2024 parent model, we obtained 4,801,061 vector data points (1,654,050 from Swarm-A, 1,443,247 from Swarm-B, and 1,703,764 from Swarm-C) and 2,329,668 scalar data points (819,596 from Swarm-A, 680,482 from Swarm-B, and 829,590 from Swarm-C). Fewer data are available for this period since we applied additional selection criteria based on the quality flags provided by the ESA Swarm data product: Flags_B = 0 for the VFM instrument and Flags_F = 0 or 1 for the ASM instrument for Swarm-A and Swarm-B, ensuring high-quality measurements (these flags were not used for Swarm-C due to the lack of ASM information for this satellite). Additionally, we applied Flags_q ≤ 3, characterizing the attitude information, and Flags_Platform ≤ 1, characterizing the platform information, for all three satellites.

Figure 1 shows the spatial and temporal distribution of Swarm data used for the two parent models.

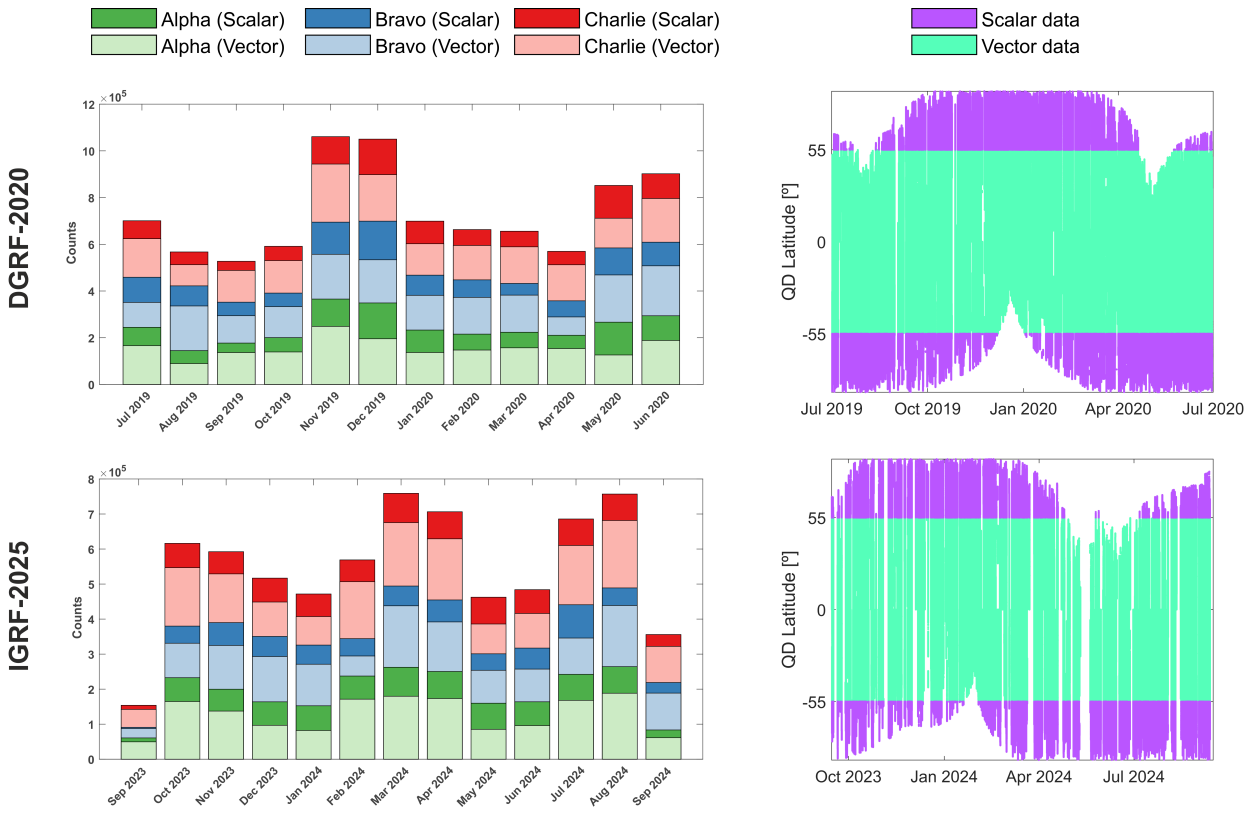

Figure 1: Left: Temporal histogram by months with the number of data (scalar and vector) per Swarm satellite. Right: Temporal and spatial distribution of data for the three Swarm satellites. Upper panels: 2019-2020 parent model. Lower panels: 2023-2024 parent model.

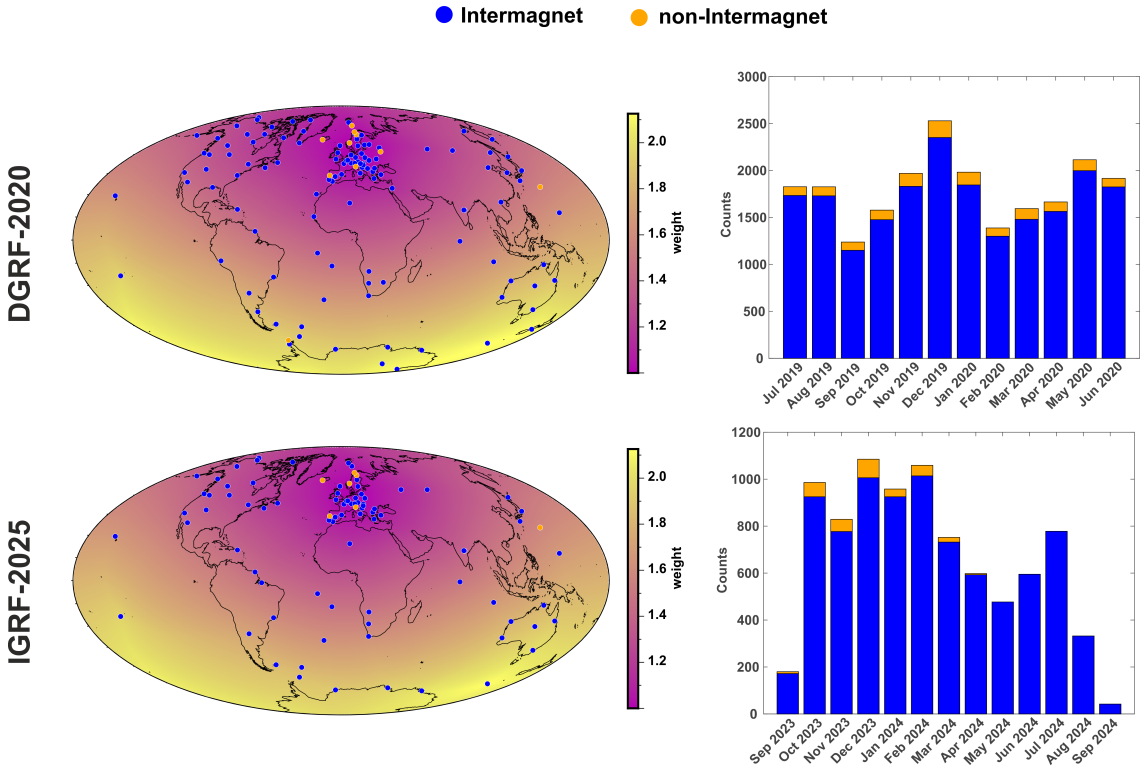

Figure 2: Left: Spatial distribution of the ground observatories. Color scale represents the weighting scheme detailed in section 4.3. Right: Temporal histogram by months with the number of data points. Upper panels: 2019-2020 parent model. Lower panels: 2023-2024 parent model.

3.2 Observatory data

We used hourly mean values of the three geomagnetic field elements, converted from the geodetic reference frame (commonly used by observatories) to the geocentric reference frame. For the 2019-2020 parent model, we directly obtained these values from the SW_OPER_AUX_OBSH2_ product derived from the World Data Centre (WDC) in Edinburgh, primarily using data from INTERMAGNET observatories (as well as some non-INTERMAGNET sources).

For the 2023-2024 parent model, we constructed these hourly mean values from the definitive and quasi-definitive minute values of INTERMAGNET data, complemented with data from some non-INTERMAGNET observatories, also from the SW_OPER_AUX_OBSH2_ product.

We applied the following selection criteria:

- Local Time between 01:00 and 02:00 AM.
- The Sun is at least 10^{o} below the horizon.
- $Kp \leq 1^{+}$ ($ap \leq 5$) for non-polar observatories, or $Kp \leq 0^{+}$ ($ap \leq 2$) for polar observatories.
- $|Dst| \leq 30$ nT.

After applying these criteria, we obtained 21,626 vector data points for the 2019-2020 parent model from 127 observatories (116 INTERMAGNET and 11 non-INTERMAGNET) and 8,671 vector data points for the 2023-2024 parent model from 89 observatories (81 INTERMAGNET and 8 non-INTERMAGNET).

Figure 2 represents the spatial and temporal distribution of ground observatory data, distinguishing between INTERMAGNET and non-INTERMAGNET observatories.

4 Modelling

4.1 Bootstrap approach

In order to manage the large amount of data and to provide not only the required coefficient values but also an estimate of their uncertainty, we applied a bootstrap approach. This involves generating 1000 sub-models for each of the two one-year parent models. Each sub-model was constructed using the following input data:

- 8000 satellite data points, consisting of 2000 data points from Swarm-A, 4000 from Swarm-B, and 2000 from Swarm-C (Swarm-A and Swarm-C fly at approximately the same altitude, while Swarm-B flies in a slightly higher orbit). These data points were selected in a spatially homogeneous manner by randomly sampling them according to a reference grid of 1000 equidistant points across the sphere.
- Approximately 4000 observatory data differences. For each observatory, we randomly selected time-separated pairs of data points to compute differences to better constraint the secular variation of the field. Specifically, we computed 31 differences for each observatory in the 2019-2020 parent model (31 data points per observatory across 127 observatories, yielding approximately 4000 differences) and 44 differences for the 2023-2024 parent model (44 data points per observatory across 89 observatories, also yielding approximately 4000 differences). To ensure meaningful temporal separation, a minimum interval of 15 days between the two selected hourly mean values was imposed for each pair.

4.2 Model parameterization

We parametrize both the internal and external fields in terms of their geomagnetic potentials:

$$V = V^{\text{int}} + V^{\text{ext}} \tag{1}$$

The internal field is expressed in terms of their classical expansion in spherical harmonic functions:

$$V^{\text{int}} = a \sum_{n=1}^{N_{\max}} \sum_{m=0}^{n} \left(g_n^m \cos m\phi + h_n^m \sin m\phi\right) \left(\frac{a}{r}\right)^{n+1} P_n^m \left(\cos\theta\right) \tag{2}$$

where a is the reference radius of the Earth (6371.2 km); n and m are the degree and order respectively of the mathematical expansion; ϕ, θ and r are the longitude, colatitude and radial distance; P_n^m are the Schmidt quasi-normalized associated Legendre functions; and g_n^m, h_n^m are the Gauss coefficients to be estimated. We extended this expansion up to harmonic degree $N_{\max} = 40$ to include not only the core (main) field contributions but also the largest scale lithospheric fields.

For each of the one-year models we considered the Gauss coefficients to be linearly time-dependent up to $n = 10$ and static between $n = 11 - 40$:

$$g_n^m(t) = \begin{cases} g_n^m(t_o) + \dot{g}_n^m \cdot (t - t_o) & \text{,if } n \leq 10 \\ g_n^m & \text{,if } n > 10 \end{cases} \tag{3}$$

For the external field we followed the parametrization of the last CHAOS-7 model (Finlay et al., 2020), aiming to describe the large-scale magnetospheric field sources. We separated it in near (mainly ring-current) and remote (mainly magnetotail and magnetopause) sources:

$$V^{\text{ext}} = V^{\text{near}} + V^{\text{remote}} \tag{4}$$

For the near magnetospheric field we used an expansion in solar magnetic (SM) coordinates up to spherical harmonic degree 2, with special treatment of the $n = 1$ terms to account for the fast varying time dependence:

$$\begin{aligned} V^{\text{near}} = a \sum_{m=0}^{1} \left[q_1^{m,\text{SM}}(t) \cos mT_d + s_1^{m,\text{SM}}(t) \sin mT_d \right] \left(\frac{r}{a} \right) P_1^m(\cos\theta_d) \\ + a \sum_{m=0}^{1} \left[\Delta q_1^{m,\text{SM}}(t) R_{1,c}^{m,\text{SM}}(t, r, \theta, \phi) + \Delta s_1^{m,\text{SM}}(t) R_{1,s}^{m,\text{SM}}(t, r, \theta, \phi) \right] \\ + a \sum_{m=0}^{2} \left[q_2^{m,\text{SM}} R_{2,c}^{m,\text{SM}}(t, r, \theta, \phi) + s_2^{m,\text{SM}} R_{2,s}^{m,\text{SM}}(t, r, \theta, \phi) \right] \end{aligned} \tag{5}$$

where T_d and θ_d are the dipole longitude and colatitude; the first degree terms have the form of $q_1^{m,\text{SM}}(t) = \hat{q}_1^{m,\text{SM}} \left[RC_{\text{external}}(t) + RC_{\text{induced}}(t) \left(\frac{a}{r} \right)^3 \right]$ according to the separation of the RC index in external/induced parts; $\Delta q_n^{m,\text{SM}}$ are baseline corrections to these terms and are estimated in bins of 30 days; and $R_n^{m,\text{SM}}$ are modifications of the solid harmonics in SM coordinates for accounting the associated induced fields according to an Earth conductivity model (same as CHAOS-7).

For the near magnetospheric field we used an expansion in geocentric solar magnetic (GSM) coordinates up to spherical harmonic degree 2, but restricting only to zonal terms ($m = 0$):

$$V^{\text{remote}} = a \sum_{n=1}^{2} q_n^{0,\text{GSM}} R_n^{0,\text{GSM}}(t, r, \theta, \phi) \tag{6}$$

where $R_n^{m,\text{GSM}}$ are similar to $R_n^{m,\text{SM}}$ but in the GSM reference frame.

In total, for each of the one year sub-models we have to estimate 1849 parameters (1680 static internal, 120 time-dependent internal, 8 SM, 39 SM bins and 2 GSM).

4.3 Inversion procedure

Since we are jointly modeling both vector and scalar data, we face a non-linear problem, which we solved by applying an iterative Newton-type algorithm:

$$m_{i+1} = m_i + (A_i' W_i A_i + \lambda)^{-1}(A_i' W_i \delta_i - \lambda m_i) \tag{7}$$

where m_i is a column vector containing all the parameters to estimate at each iteration i. A_i is the matrix of parameters calculated using the Fréchet derivative at iteration i, δ_i is a column vector with the residuals between the input data and the model data at iteration i, W_i is a diagonal weight matrix, and λ is a regularization matrix.

The weight diagonal matrix is composed by the product of three diagonal matrices:

- W_{prior}, which considers the *a priori* variances for each of the input data types. For the satellite data, we used the following variances: $\sigma_X = 2.7$ nT, $\sigma_Y = 2.2$ nT, $\sigma_Z = 1.3$ nT, $\sigma_{Fp} = 2.2$ nT. For the observatory data, we considered: $\sigma_{dXnp} = 2.9$ nT, $\sigma_{dYnp} = 4.1$ nT, $\sigma_{dZnp} = 2.1$ nT, $\sigma_{dXp} = 4.5$ nT, $\sigma_{dYp} = 4.9$ nT, $\sigma_{dZp} = 5.5$ nT; where p and np represent polar or non-polar data, respectively. These *a priori* uncertainties were estimated by analyzing the residuals of the CHAOS-7.18 model.
- W_{spatial}, which accounts for the spatial distribution of the input data, especially for the observatory data (which is spatially inhomogeneous). For each observatory k, we computed a weight according to:

  $$w_k = < \alpha_{jk} > / \alpha_{\min} \tag{8}$$

  where the numerator is the mean distance between observatory k and all other j observatories, and the denominator is the minimum mean distance among all the observatories. This ensures that observatories located in densely covered regions (e.g., Europe) receive less weight compared to those in more remote areas (e.g., the Pacific). See Figure 2 for an illustration of this weight. For the satellite data, all data points have a spatial weight of 1.8, corresponding to a mean angular distance of 90º (homogeneously distributed data over the sphere).
- W_{Huber} , updated at each iteration of the inversion algorithm following the Huber weighting scheme, which handles a mixture of Gaussian and Laplacian noise in the input data (Constable, 1988). For each k-th input data, the weight at an iteration i is assigned according to:

  $$w_{i,k} = \frac{1}{\sigma_k^2} \min\left(\frac{1.5\sigma_k}{|\delta_{i,k}|}, 1\right) \tag{9}$$

  where σ_k is the corresponding *a priori* variance, and $\delta_{i,k}$ is the corresponding residual at iteration i.

The matrix λ is used only to regularize the time evolution of the baseline correction bins. We employed a quadratic form of the time derivative between neighboring bins, which affects only the bin-to-bin variability of the same coefficient. For the 2019-2020 parent model, we used a regularization parameter of 10^4 $(\text{nT}\cdot\text{days}^{-1})^{-2}$, whereas for the 2023-2024 parent model we used 10^2 $(\text{nT}\cdot\text{days}^{-1})^{-2}$. This relaxation in the regularization parameter is motivated by the use of the *Dst* index instead of *Rc* for the last months of this model, as the *Rc* was not available after 1st May 2024.

For each sub-model, 10 iterations of the Newton algorithm were sufficient for convergence,

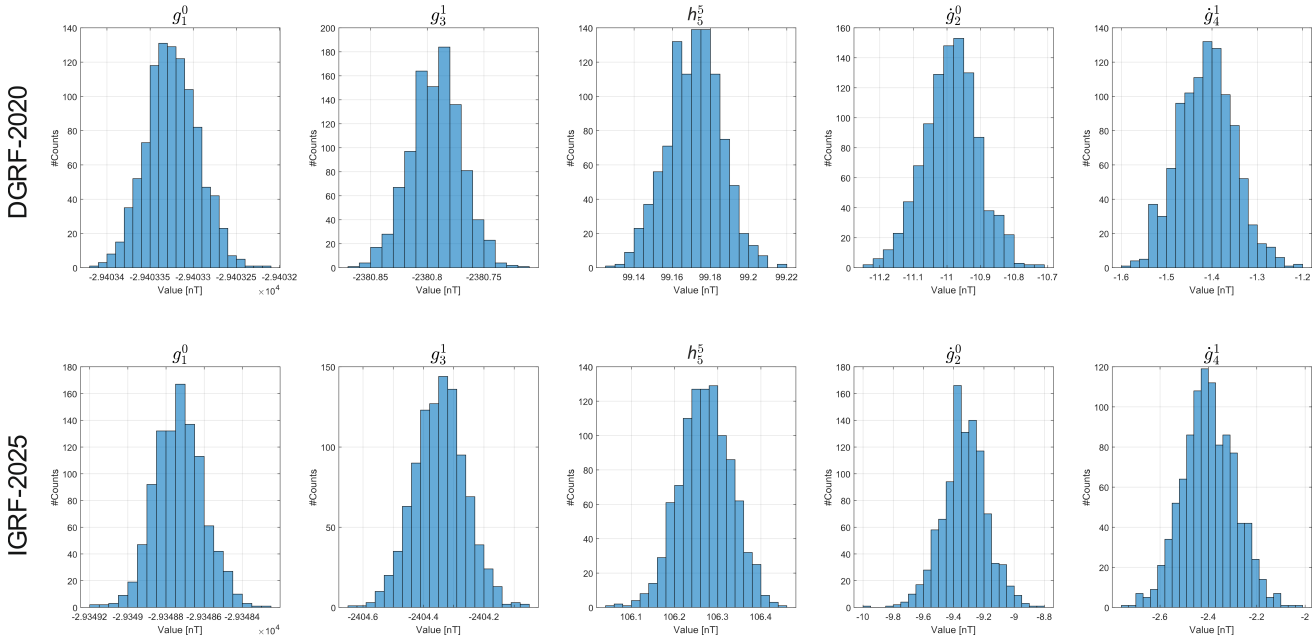

Figure 3: Examples of histograms of some internal coefficients for both the 2019-2020 parent model (upper panels) and the 2023-2024 parent model (lower panels).

starting with an initial model consisting of a static internal axial dipole of $g_1^0 = -30000$ nT and no external field.

4.4 Parent models estimation

Once the 1000 sub-models were obtained for each of the two parent models (2019-2020 and 2023-2024), we estimated the value of their coefficients by taking the mean value of all the corresponding sub-models. This procedure also allows us to derive an estimation of the uncertainty of the parent models' coefficients. The ensemble of 1000 sub-models defines a normal distribution for all the parameters, so we decided to assign their uncertainty as twice their standard deviation (95% of confidence level). Figure 3 shows examples of some histograms illustrating how the parameters obtained with the ensemble of sub-models follow the normal distribution.

4.5 Extraction of IGRF-14 products

With the two parent models constructed and their associated uncertainties, we derive our IGRF-14 products as follows:

- DGRF-2020: We evaluate the internal part of our 2019-2020 parent model at the year 2020.0, truncating the expansion to degree 13 and rounding the coefficients to a 0.01 nT precision.
- IGRF-2025: We linearly extrapolate the internal part of our 2023-2024 parent model to the year 2025.0, truncating the expansion to degree 13 and rounding the coefficients to a 0.01 nT precision.
- SV-2025-2030: We directly take the linear dependence of the internal part of the 2023-2024 parent model, truncate it to degree 8, and round the coefficients to a 0.01 nT precision. This serves as an estimate of the secular variation for this 5-year period.

References

Constable, C. G. (1988). Parameter estimation in non-gaussian noise. *Geophysical Journal International*, 94(1):131–142.

Finlay, C. C., Kloss, C., Olsen, N., Hammer, M. D., Tøffner-Clausen, L., Grayver, A., and Kuvshinov, A. (2020). The chaos-7 geomagnetic field model and observed changes in the south atlantic anomaly. *Earth, Planets and Space*, 72(1).