A candidate secular variation model for IGRF-14 based on a recurrent neural network method with an extended Kalman filter

Sato, S.[1], S. Nakano[2,5], T. Minami[3], M. Matsushima[4] and H. Toh[1]

[1] *Graduate School of Science, Kyoto University*
[2] *The Institute of Statistical Mathematics*
[3] *Graduate School of Science, Kobe University*
[4] *School of Science, Institute of Science Tokyo*
[5] *Center for Data Assimilation Research and Applications, Joint Support Center for Data Science Research*

This document briefly describes how our candidate secular variation model for the next generation of International Geomagnetic Reference Field (IGRF-14) has been constructed by a machine learning method. As for 'data', we used a series of MCM field models evaluated/provided by a French group (Ropp et al., 2020; Ropp and Lesur, 2023). The task force chairs, therefore, are advised to refer to the associated document with the MCM models to be submitted to IGRF Task Force and the references therein for, e.g., what kind of data were used in the field modeling, their selection/rejection criteria and spatial coverage, the inversion schemes (starting models, number of iterations etc.) and their results. Because our French colleagues kindly sent us their very latest version up to Epoch 2024.37, we think the difference between the final set of the French field model submitted to the task force and what we used is quite minor, if any.

1. Introduction

There are various approaches to construct IGRF candidate models for the geomagnetic secular variation. At the time of IGRF-13, a few groups adopted data assimilation methods to derive the geomagnetic secular variation estimates for Epoch 2020 (e.g., Sanchez et al., 2020; Tangborn et al., 2021). Our group also did use a data assimilation method based on En4dVar (Minami et al., 2020) together with a non-linear geodynamo simulation code (Takahashi, 2012; 2014), in which we had to work with an ensemble of 960 geodynamo models at the same time. Because the size of a state vector for each MHD dynamo run is of the order of 10^6, any data assimilation in search for reliable geomagnetic secular variation estimates tends to be very computer intensive.

On the other hand, machine learning methods require us much lighter calculation in order to estimate the geomagnetic secular variation, although they are also data-driven systems like data assimilation. Some of the authors of this report have made the very first attempt for short-term prediction of the geomagnetic secular variation (Nakano et al., 2024) using an echo state network (ESN; Jaeger and Haas, 2004)) method. However, the ESN-based machine learning methods are very simplified version of normal deep

learning methods such as recurrent neural network (RNN; Elman, 1990) and classified as reservoir computing rather than deep learning. We, therefore, adopted an RNN method with an extended Kalman filter (EKF) in its training process here. Even though RNN methods are often used with back propagation in time-series forecast (e.g., Werbos, 1990), we newly formulated an EKF-RNN method so as to circumvent overfitting problems known to appear RNNs with back propagation.

2. Formulation

RNN consists of input $\mathbf{x}(t) \in \mathbb{R}^{D(in)}$, output $\mathbf{z}(t) \in \mathbb{R}^{D(out)}$ and hidden layers $\mathbf{h}(t) \in \mathbb{R}^{D(rec)}$ between $\mathbf{x}(t)$ and $\mathbf{z}(t)$ as shown in Fig. 1. Adopting the hyperbolic tangent function as an activation function, the hidden layers are characterized iteratively by

$$\mathbf{h}(t_i) = \tanh (\mathbf{W}^{(in)} \cdot \mathbf{x}(t_i) + \mathbf{W}^{(rec)} \cdot \mathbf{h}(t_{i-1}) + \mathbf{b}^{(rec)}) \tag{1}$$

to give output $\mathbf{z}(t_i)$

$$\mathbf{z}(t_i) = \mathbf{W}^{(out)} \cdot \mathbf{h}(t_i) + \mathbf{b}^{(out)} \tag{2}$$

where $\mathbf{W}^{(in)}$ and $\mathbf{W}^{(out)}$ are weights applied respectively to $\mathbf{x}(t)$ and $\mathbf{z}(t)$, while $\mathbf{W}^{(rec)}$ are those among the hidden layers. $\mathbf{b}^{(rec)}$ and $\mathbf{b}^{(out)}$ are bias vectors to be optimized in the course of the whole RNN run as well as $\mathbf{W}^{(in)}$, $\mathbf{W}^{(rec)}$, and $\mathbf{W}^{(out)}$. The size of input and output vectors are determined by the shape of training data. The size of $\mathbf{h}(t)$, which corresponds to the number of hidden units $D^{(rec)}$, is left as a free parameter that determines the whole network architecture.

One problem here is how to train RNN so as to yield the best weights and biases $\mathbf{w}$ defined by

$$\mathbf{w} = (\text{vec}(\mathbf{W}^{(in)})^T, \text{vec}(\mathbf{W}^{(rec)})^T, \mathbf{b}^{(rec)\,T}, \text{vec}(\mathbf{W}^{(out)})^T, \mathbf{b}^{(out)\,T})^T \tag{3}$$

by comparing the RNN output $\mathbf{z}(t)$ with given observed time-series $\mathbf{y}^o(t)$ and their data covariance matrix $\mathbf{R}$. The observation $\mathbf{y}^o$ could be the internal Gauss coefficient vector (g_1^0, g_1^1, h_1^1, … , g_{13}^{13}, h_{13}^{13}) in this case.

Here we adopted a training algorithm with an extended Kalman filter by Puskorius and Feldkamp (1994) as follows.

Defining a Jacobian $\mathbf{H}(t)$ as $\partial \mathbf{z}/\partial \mathbf{w}$, we used the following recursive equations with a Kalman gain matrix $\mathbf{K}(t)$,

$$\mathbf{w}^a(t) = \mathbf{w}^f(t) + \mathbf{K}(t)(\mathbf{y}^o(t) - \mathbf{z}(t)) \tag{4}$$

and

$$\mathbf{P}^a(t) = (\mathbf{I} - \mathbf{K}(t)\mathbf{H}(t)) \, \mathbf{P}^f(t) \tag{5}$$

where we set:

$$\mathbf{K}(t) = \mathbf{P}^f(t)\mathbf{H}(t)^T \, (\, \mathbf{H}(t)\mathbf{P}^f(t)\mathbf{H}(t)^T + \mathbf{R}(t) \,)^{-1}, \tag{6}$$
$$\mathbf{w}^f(t_i) = \mathbf{w}^a(t_{i-1}), \tag{7}$$

and

$$\mathbf{P}^f(t_i) = \alpha\mathbf{P}^f(t_{i-1}) + (1-\alpha)\mathbf{P}^a(t_{i-1}). \tag{8}$$

We also set $\alpha = 0.5$. Eq. (8) corresponds to the relaxation to prior spread (Whitaker and Hamill, 2012).

One advantage of this recursion is that it enables construction of the forecast error covariance matrix $\mathbf{P}$ from the data covariance matrix $\mathbf{R}$ in addition to its sufficient protection against the overfitting problem associated with RNNs with back propagation.

The remaining problem is how to determine the best initial weights and biases $\mathbf{w}_0$ as well as the number of hidden units $D^{(rec)}$.

3. Hindcast experiment

We created a training dataset spanning from 2004.87 to 2014.62 using the internal Gauss coefficients given by MCM (Ropp et al., 2020) to yield the geomagnetic secular variation estimates for 2014.75 through 2019.50, which were compared with those by Minami et al. (2020). We worked with second time derivatives of the internal Gauss coefficients rather than the coefficients themselves in order to eliminate trends in the observed time-series longer than the training period (10 years in our case). Time-series of Gauss coefficients were recovered by integrating the obtained second derivatives twice with time. Figure 2 shows the observed time-series of Gauss coefficients and their second derivatives for both training and validation periods.

Figures 3 and 4 summarize the hindcast results, which show the maximum forecast error of 80.79 nT after 5 years. Following Whaler and Beggan (2015), we used the following metrics to evaluate our EKF-RNN performance:

$$\text{Core field:} \quad \sqrt{dP_n} = \sqrt{(n+1)\sum_m\{(g^m_{n,obs} - g^m_{n,RNN})^2 + (h^m_{n,obs} - h^m_{n,RNN})^2\}} \tag{9}$$
$$\sqrt{dP} = \sqrt{\sum_n dP_n}. \tag{10}$$

and

$$\text{SV:} \quad \sqrt{d\dot{P}_n} = \sqrt{(n+1)\sum_m\{(\dot{g}^m_{n,obs} - \dot{g}^m_{n,RNN})^2 + (\dot{h}^m_{n,obs} - \dot{h}^m_{n,RNN})^2\}} \tag{11}$$
$$\sqrt{d\dot{P}} = \sqrt{\sum_n d\dot{P}_n}. \tag{12}$$

D$^{(rec)}$ and $\mathbf{w}_0$ required to determine the RNN architecture and initiate our EKF recursion were determined by a numerical experiment. We prepared 32 patterns of initial state $\mathbf{w}_0$' for each EKF-RNN run and changed D$^{(rec)}$ from 1 through 40. The experiment results were trained in the period covering from 2004.87 to 2022.37 and evaluated in terms of $\sqrt{d\dot{P}}$ at t = 2023.00. The optimal number of hidden units was determined as D$^{(rec)}$ = 34 because it achieved the minimum variation of $\sqrt{d\dot{P}}$ within the 32 EKF-RNN runs among all D$^{(rec)}$ candidates ranging from 1 to 40. As for $\mathbf{w}_0$, the best initial pattern that gave the minimum $\sqrt{d\dot{P}}$ for the optimal D$^{(rec)}$ among the 32 runs was used as the initial $\mathbf{w_0}$ in both the hindcast experiment and the forecast model described in the next section.

4. Geomagnetic Secular Variation Forecast for 2025

The actual forecast for Year 2025 was conducted by creating a new training dataset from 2004.87 through 2024.37, while the forecast by our EKF-RNN scheme was done from 2024.62 through 2025.12. The SV forecast at 2025.00 was obtained as follows:

$$\dot{g}^m_{n,\text{RNN}}(t = 2025.00) = \frac{1}{\Delta t}\left(g^m_{n,RNN}(t = 2025.12) - g^m_{n,\text{RNN}}(t = 2024.87)\right) \quad (13)$$

The forecast results were summarized in Table 1.

5. Summary

We newly developed an RNN method to estimate the geomagnetic secular variation for IGRF-14 with an EKF as a training interface. This reduced considerable amount of calculation compared with data assimilation methods. Time-series forecast using RNNs is normally realized by back propagation (Elman, 1990), which is always associated with overfitting problems. On the other hand, our new method was found free from overfitting problems. Furthermore, our method returned forecast error covariance matrices stemming from the given data covariance matrices.

Hindcast experiment for Epoch 2020 using our method yielded a 5-year forecast error less than 100 nT. This is better than our own result using data assimilation based on 4dEnVar (Minami et al., 2020).

Forecast of the geomagnetic secular variation for Epoch 2025 was finally conducted using a 20-year training dataset and summarized in Table 1 together with error estimates at each spherical harmonic mode.

REFERENCES

Elman, J.L. (1990). "Finding Structure in Time". Cognitive Science. 14 (2): 179–211. https://doi.org/10.1207/s15516709cog1402_1.

Hulot, G., and J.-L. LeMouël (1994), A statistical approach to the Earth's main magnetic field, Phys. Earth Planet. Inter., 82, 167–183.

Jaeger, H. and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science, 304, 78–80 (2004). https://www.science.org/doi/10.1126/science.1091277

Minami, T., Nakano, S. Y., Lesur, V., Takahashi, F., Matsushima, M., Shimizu, H., Nakashima, R., Taniguchi, H., Toh, H. (2020) A candidate secular variation model for IGRF-13 based on MHD dynamo simulation and 4DEnVar data assimilation. Earth, Planets and Space, 72(1), 1-24. https://doi.org/10.1186/s40623-020-01253-8.

Nakano, S., Sato, S. & Toh, H. Short-term prediction of geomagnetic secular variation with an echo state network. Earth Planets Space 76, 121 (2024). https://doi.org/10.1186/s40623-024-02064-x

Puskorius, G.V. and Feldkamp, L.A., "Neurocontrol of nonlinear dynamical systems with Kalman filter trained recurrent networks," in IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 279-297, March 1994, doi: 10.1109/72.279191.

Ropp, G., Lesur, V., Baerenzung, J., & Holschneider, M., 2020. Sequential modelling of the Earth's core magnetic field, Earth, Planets and Space, 72(1), 153, doi: 10.1186/s40623-020-01230-1.

Ropp, G. and V Lesur, Mid-latitude and equatorial core surface flow variations derived from observatory and satellite magnetic data, Geophysical Journal International, Volume 234, Issue 2, August 2023, Pages 1191–1204, https://doi.org/10.1093/gji/ggad113

Sanchez, S., Wicht, J. & Bärenzung, J. Predictions of the geomagnetic secular variation based on the ensemble sequential assimilation of geomagnetic field models by dynamo simulations. Earth Planets Space, 72, 157 (2020). https://doi.org/10.1186/s40623-020-01279-y.

Takahashi, F. (2012) Implementation of a high-order combined compact difference scheme in problems of thermally driven convection and dynamo in rotating spherical shells. Geophys. Astrophys. Fluid Dyn., doi:10.1080/03091929.2011.

Takahashi, F. (2014) Double diffusive convection in the Earth's core and the morphology of the geomagnetic field. Phys. Earth Planet. Inter., 226, 83-87. https://doi.org/10.1016/j.pepi.2013.11.006

Tangborn, A., Kuang, W., Sabaka, T.J. et al. Geomagnetic secular variation forecast using the NASA GEMS ensemble Kalman filter: A candidate SV model for IGRF-13.

Earth Planets Space 73, 47 (2021). https://doi.org/10.1186/s40623-020-01324-w.

Werbos, P.J. "Backpropagation through time: what it does and how to do it," in Proceedings of the IEEE, vol. 78, no. 10, pp. 1550-1560, Oct. 1990, doi: 10.1109/5.58337.

Whaler, K. A., & Beggan, C. D. (2015). Derivation and use of core surface flows for forecasting secular variation. Journal of Geophysical Research: Solid Earth, 120(3), 1400-1414. https://doi.org/10.1002/2014JB011697

Whitaker, J. S & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. Monthly Weather Review, 140, 3078–3089. https://doi.org/10.1175/MWR-D-11-00276.1

Table 1 Geomagnetic secular variation coefficients and their errors for 2025

| n	m	$\dot{g}$ [nT/y]	$\dot{h}$ [nT/y]	$\delta\dot{g}$ [nT/y]	$\delta\dot{h}$ [nT/y]
1	0	12.55	0.00	4.72	0.00
1	1	9.73	-20.57	7.65	8.72
2	0	-13.55	0.00	3.83	0.00
2	1	-4.90	-27.35	4.99	7.71
2	2	-9.75	-10.07	4.64	6.79
3	0	-1.65	0.00	3.12	0.00
3	1	-4.96	3.34	4.04	4.02
3	2	0.25	0.60	4.17	3.27
3	3	-15.96	-3.90	4.12	5.94
4	0	-1.58	0.00	3.22	0.00
4	1	-1.91	-1.67	3.76	2.72
4	2	-5.75	3.95	3.09	2.45
4	3	5.44	1.35	2.86	3.01
4	4	-7.32	-4.03	3.56	4.16
5	0	0.88	0.00	1.84	0.00
5	1	1.04	-0.92	2.71	2.74
5	2	0.04	2.35	2.02	1.94
5	3	0.62	0.43	2.04	2.74
5	4	2.24	1.42	2.64	2.21
5	5	1.31	2.17	4.12	2.90
6	0	0.37	0.00	2.63	0.00
6	1	-0.28	0.01	2.31	2.67
6	2	0.83	-1.49	2.01	1.80
6	3	1.20	-0.31	1.59	1.63
6	4	-0.94	0.83	1.72	1.78
6	5	0.45	0.73	2.00	2.07
6	6	0.96	0.87	2.55	2.68
7	0	0.25	0.00	2.24	0.00
7	1	-0.10	0.29	1.64	1.69
7	2	-0.33	0.51	1.63	1.57
7	3	0.46	-0.63	1.44	1.42
7	4	-0.13	0.10	1.73	1.46
7	5	-0.75	-0.97	1.60	1.51
7	6	-0.82	0.68	1.97	1.98
7	7	0.89	-0.33	2.32	2.60
8	0	0.18	0.00	1.34	0.00
8	1	0.27	-0.39	1.28	1.59
8	2	-0.02	0.46	1.28	1.21
8	3	0.49	-0.38	1.24	1.25
8	4	-0.15	0.35	1.26	1.24
8	5	0.26	-0.58	1.21	1.50
8	6	0.07	-0.65	1.32	1.29
8	7	-0.02	0.25	1.41	1.27
8	8	0.21	0.19	1.38	2.07
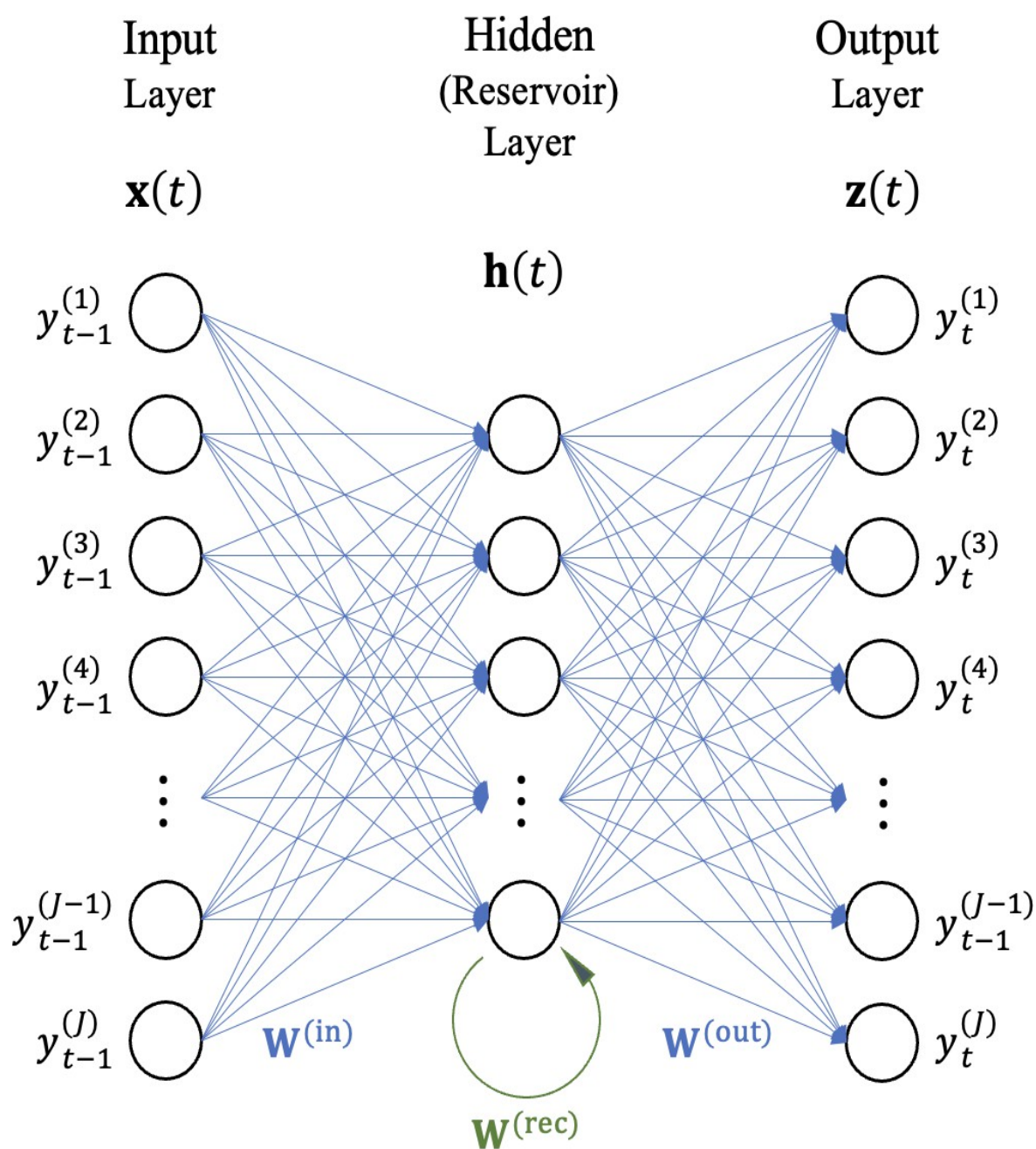
Fig. 1 Schematic cartoon of RNN. Unlike ESN, all the weights, $\mathbf{W}^{(in)}$, $\mathbf{W}^{(rec)}$ and $\mathbf{W}^{(out)}$ are updated over time t in RNN.
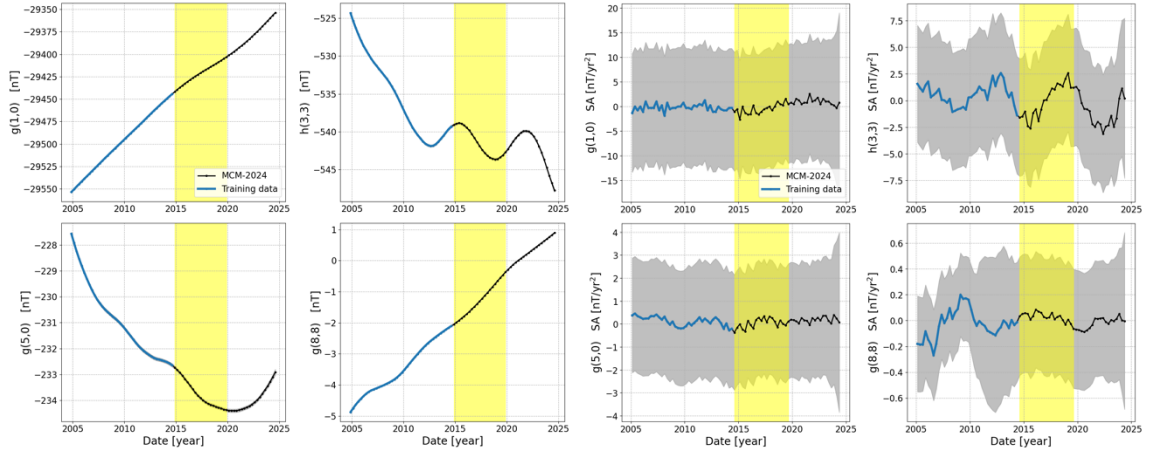
Fig. 2 The observed Gauss coefficients (left) and their second time derivatives (right) from 2004.87 through 2024.37. The zones shaded by yellow indicate the validation period in the hindcast experiment. The zones in gray of the second time derivatives show the 2-σ intervals of observation errors, respectively.
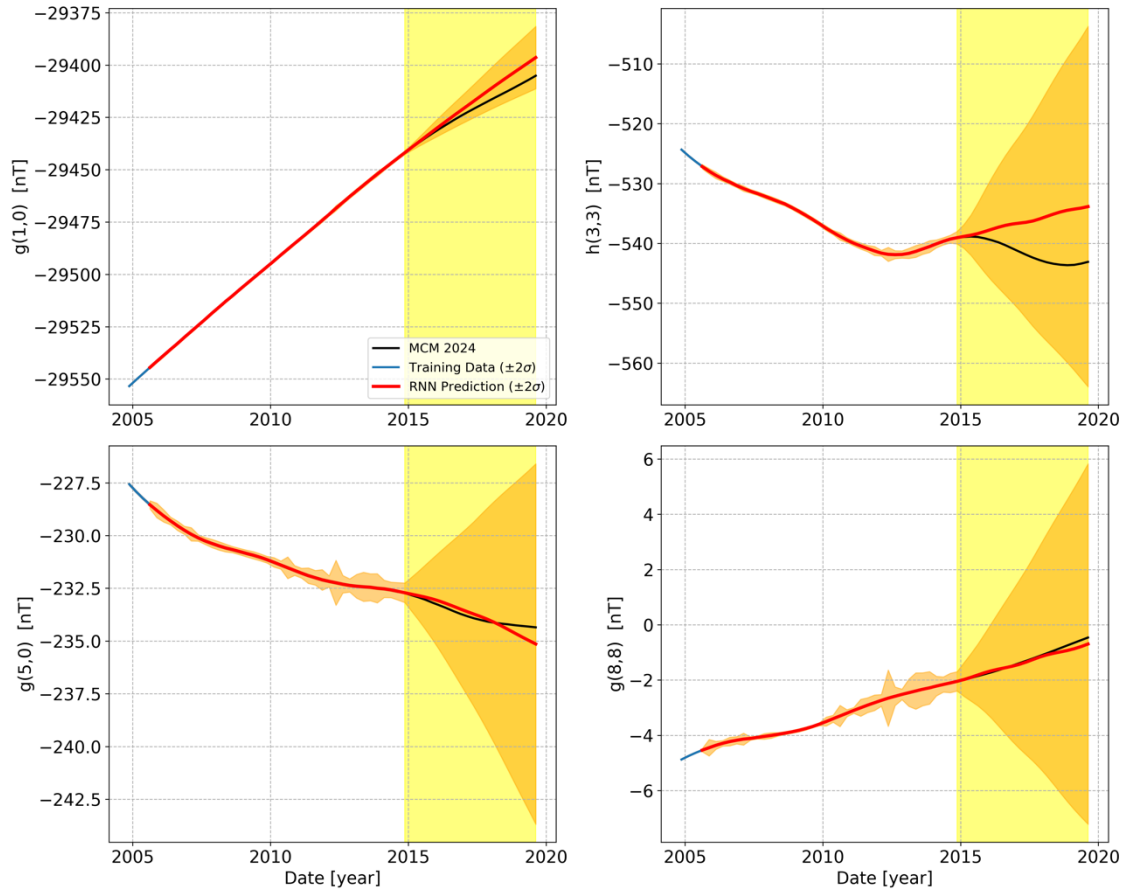

Fig. 3 Temporal evolution of g_1^0, h_3^3, g_5^0, g_8^8 coefficients. Observation and EKF-RNN prediction are shown by blue and red lines, respectively. The orange zones represent 2-σ intervals of forecast errors.
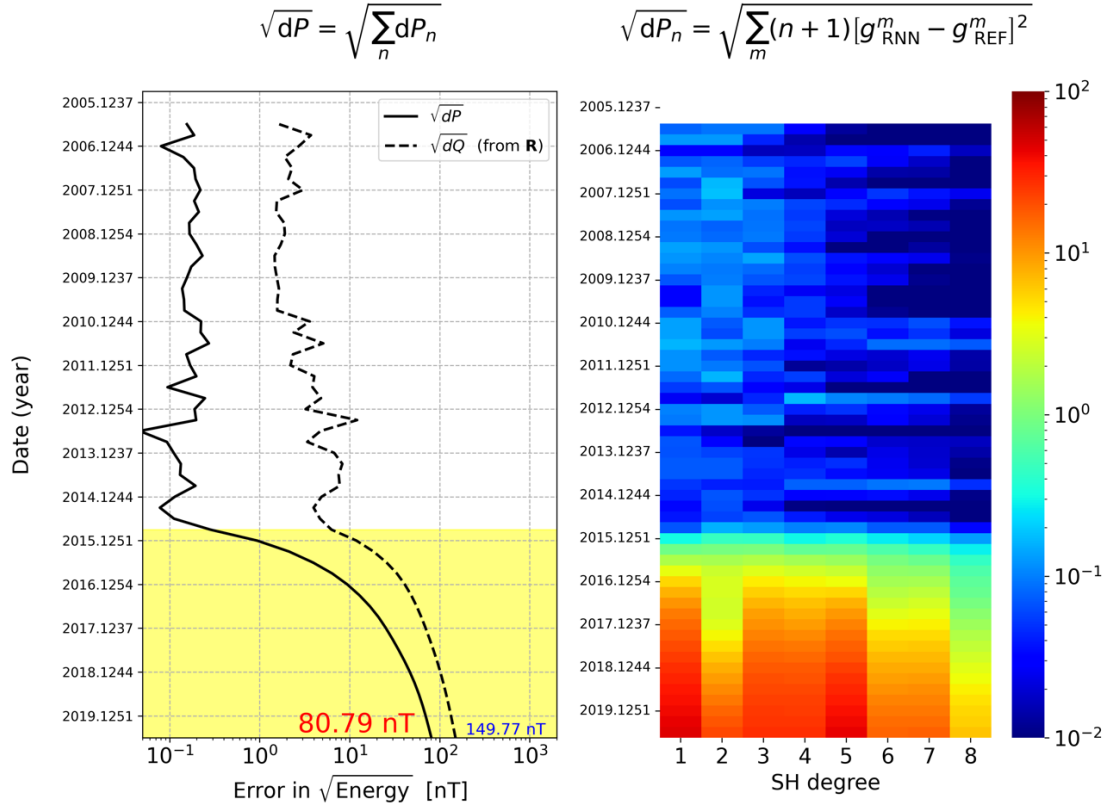
Fig. 4 Increase in the forecast error (left) and the degree dependence of the forecast error (right) over all the training and validation period. It is noteworthy that the larger the degree, the smaller the forecast error over time except for Degree 2. This may be because small-scale (i.e., large-degree) core fields tend to exhibit short-term variations (Hulot and LeMouël, 1994).