Can one use Earth's magnetic axial dipole field intensity to predict reversals?

K. Gwirtz*, M. Morzfeld*, A. Fournierº, G. Hulotº

SEDI 2020

*Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego;
ºUniversité de Paris, Institut de Physique du Globe de Paris, CNRS, F-75005 Paris, France.

Predictions based on dipole intensity

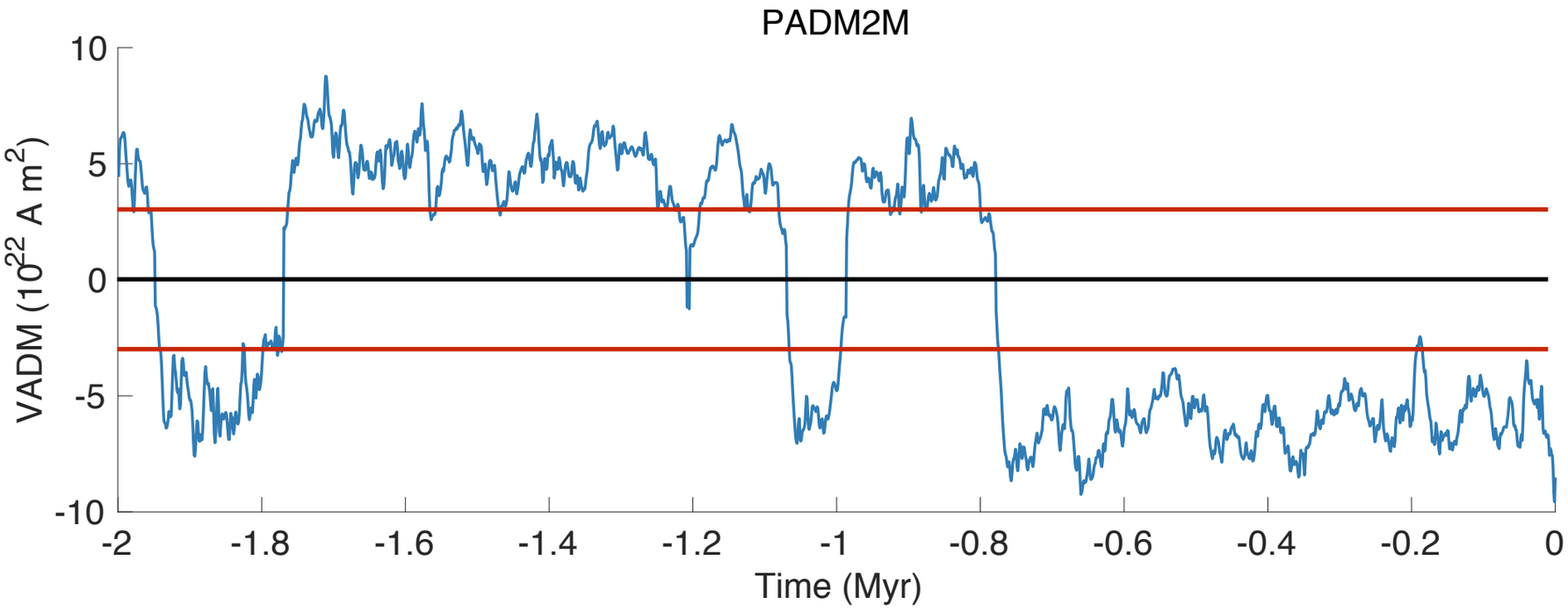


Figure 1. Axial-dipole intensity over the last 2 Myr with the sign indicating polarity (negative corresponding to the present day). We ask, is there an axial dipole intensity below which a reversal is likely to occur within a reasonable amount of time? For example, is the intensity dropping below the level indicated by the red lines a good indicator of an imminent reversal?

- We begin by asking: can one use the intensity of the axial dipole to predict reversals? This immediately leads to questions about the approach to be taken.

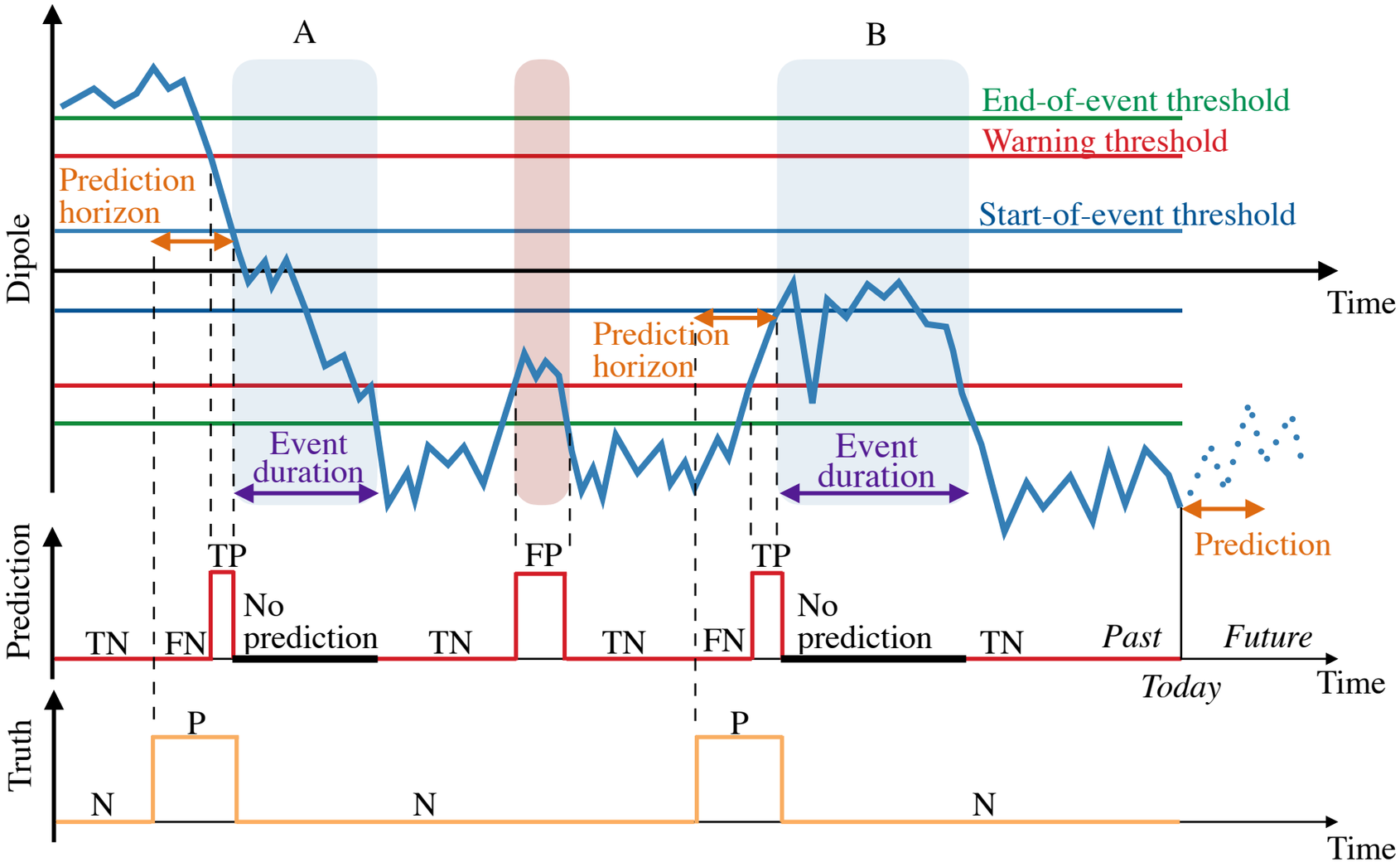
  - **Question:** How do we address events like the Cobb Mountain Subchron? Specifically, how do we view events with multiple, successive changes in the sign of the axial dipole?
    Answer: A rigorous definition of the low-dipole events we wish to predict and a strategy for predicting them.

  - **Question:** How do we explore this question given the limited data?
    Answer: A hierarchy of models.

  - **Question:** How do we evaluate how useful axial-dipole intensity is as a predictor of these events?
    Answer: A well-established scoring system appropriate for predictions of this type.

- The three following slides address the above questions in more detail.

Definitions & prediction strategy



- Start-of-event threshold: An event begins when the dipole intensity drops below this threshold.
- End-of-event threshold: An event ends when the dipole intensity exceeds this threshold.
- Event duration: Time between first decrease below the start-of-event threshold and exceeding the end-of-event threshold.
- Prediction horizon: How far ahead we wish to predict events.
- Warning threshold: We predict an event will happen within the prediction horizon when the dipole intensity drops below this value.
- Positive (P): An event occurs with the prediction horizon.
- Negative (N): An event does not occur within the prediction horizon.
- Every prediction is either a true positive (TP), true negative (TN), false positive (FP) or false negative (FN).

Figure 2. Illustration of definitions and the prediction strategy. *Top:* dipole (solid blue) as a function of time. The thin blue, green and red horizontal lines represent the start-of-event, the end-of-event and the warning thresholds. Two low-dipole events are labeled A (reversal) and B (excursion), and we indicate their event durations. Highlighted in red is a period of low intensity, which is not a low-dipole event, but where the low intensity causes false positives (FP). Towards the right, we illustrate a prediction over a given prediction horizon, which will lead to true negatives (TN). The prediction horizon also defines the true labels, (see bottom panel). *Center:* prediction as a function of time. The red line at zero corresponds to the prediction "no low-dipole event occurs during the prediction horizon," and the red line at one corresponds to the prediction "a low-dipole event occurs during the prediction horizon." The thick black line segments correspond to periods during which no prediction is made. For events A and B, we first observe TNs, followed by false negatives (FN), caused by the warning threshold being small; then we observe TPs followed by a period during which no prediction is made. *Bottom:* true occurrences of low-dipole events within the prediction horizon. The orange line at zero corresponds to negatives (N), i.e., "no low-dipole event occurs during the prediction horizon." The orange line at one corresponds to positives (T), i.e., "a low-dipole event occurs during the prediction horizon."

Hierarchy of models

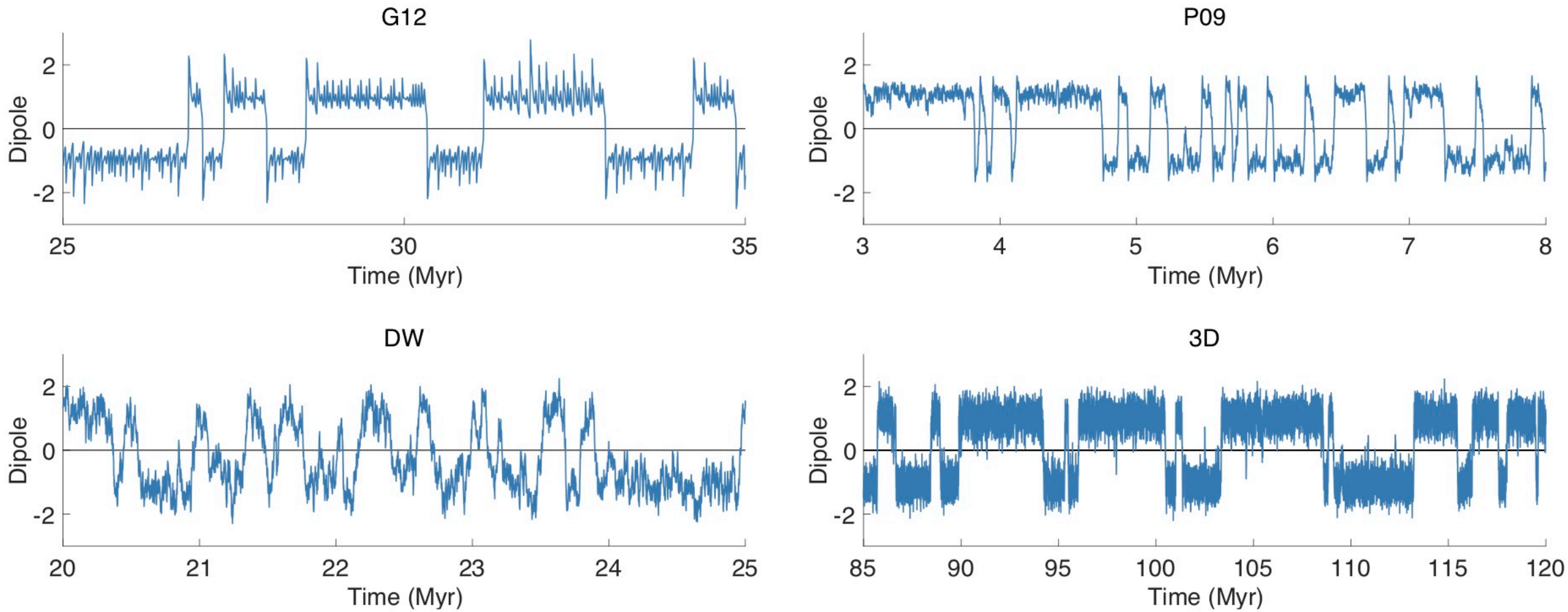


Figure 3. The hierarchy of models. *Upper left:* G12 (Gissinger 2012) consisting of a system of three ODEs. *Upper right:* P09 (Pétrélis et al. 2009) defined by a stochastic differential equation. *Lower left:* DW (Buffett el al. 2013) so named because of its construction using a double-well potential in a stochastic differential equation. *Lower right:* 3D (Fournier et al., unpublished) given by a full three-dimensional numerical dynamo model.

- Given the limited data, we experiment with the hierarchy of models shown in figure 3.
- To ensure results are comparable, the models are rescaled:
  - The dipole is scaled to the average intensity of the respective model.
  - Time is scaled by the average event duration for each model.
- As a result of the intensity rescaling, thresholds are expressed as portions of the average dipole intensity. For the definition of a low-dipole event we use a start-of-event threshold of 10% and an end-of-event threshold of 80% but similar results are obtained with reasonably similar values.
- As a result of the time rescaling, the prediction horizon is proportional to the average event duration. We use a prediction horizon equal to one average event duration but similar results are obtained with reasonably similar values.

Evaluating performance

- We wish to evaluate the performance of various warning thresholds in predicting low-dipole events.
- Because low-dipole events are rare, high accuracy (*ACC*) is easy to obtain by alway predicting no event will occur and hence, is not an ideal measure of performance.

$$ACC = \frac{TP + TN}{P + N}$$

- The Matthews correlation coefficient (MCC) is robust to imbalances in the occurrence of events.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- We "train" on a data set by applying various warning thresholds and selecting the one which maximizes the MCC. This warning threshold can later be applied to an independent verification data set
- Our choice of optimal warning threshold is restricted to be between the start-of-event threshold (10%) and end-of-event threshold (80%) to avoid a scenario where events are never predicted (warning threshold < start-of-event threshold) or always predicted immediately after the conclusion of an event (warning threshold > end-of-event threshold).

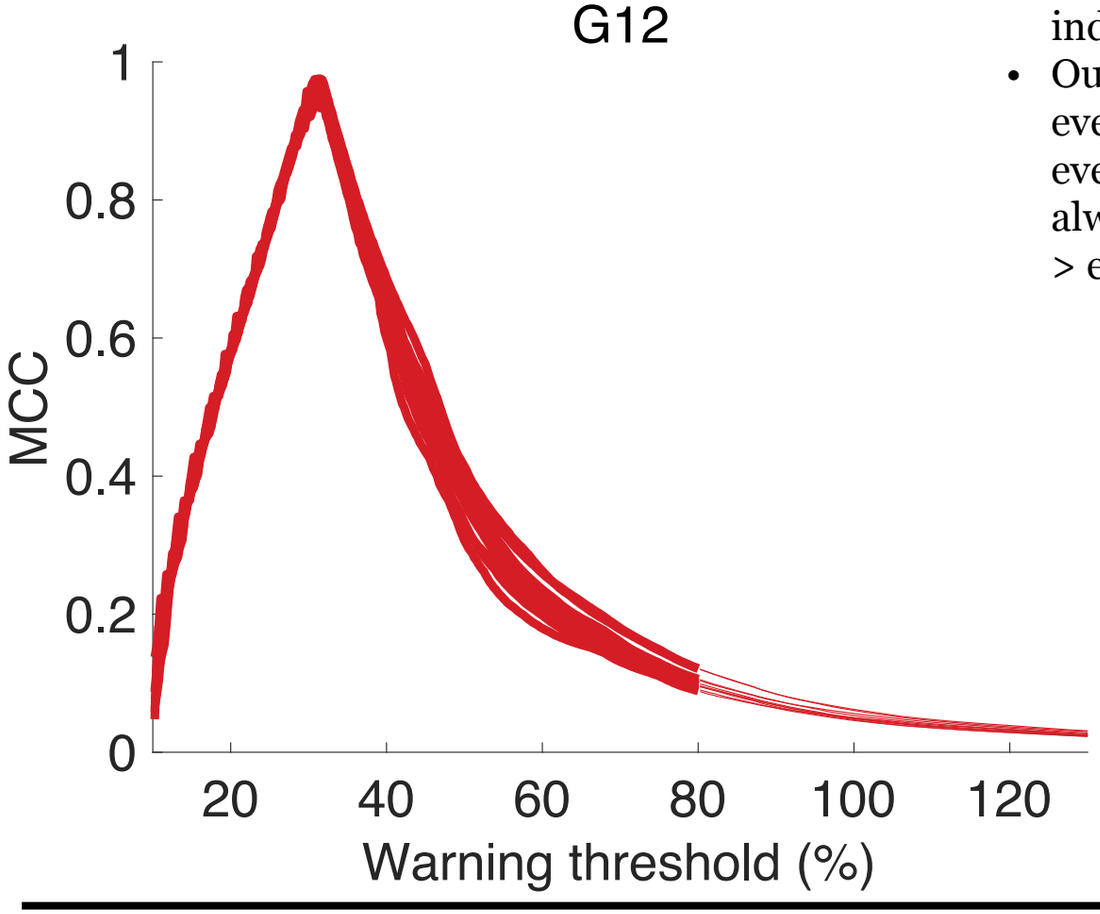


Figure 4. MCC vs. warning threshold for ten different training sets of G12. The thin portion of each line corresponds to warning thresholds which are greater than the end-of-event threshold and so, not considered. This training indicates an optimal warning threshold of around 30%.

Performance among model hierarchy

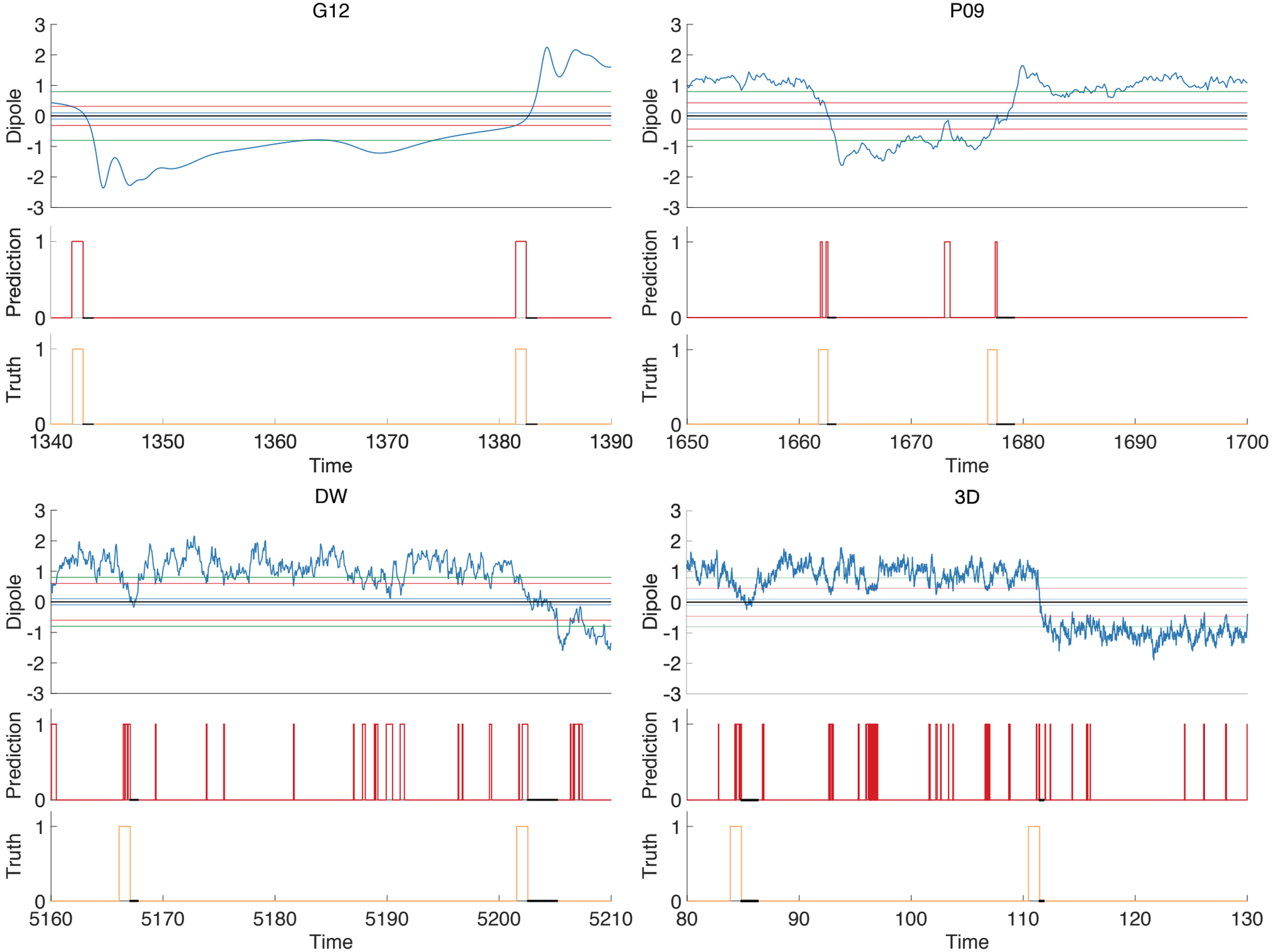


Figure 5. Dipole, prediction and truth plots, mirroring the illustration of figure 2, for G12 (upper left), P09 (upper right), DW (lower left) and 3D (lower right) all using the optimal warning threshold (according to MCC). It is found that G12 is the easiest to predict followed by P09, DW and finally 3D.

Decay time

- One may be inclined to assume that G12 and P09 are found to be more predictable by a threshold strategy solely as the result of the frequent false positives in predictions of DW & 3D however their rapid decay in dipole intensity preceding events leads to frequent false negatives.
- We define the *decay time* to be the time between the beginning of an event and the last previous instance at which the dipole exceeded the end-of-event threshold.
- It is found that the predictability of a model corresponds to the ratio of the average decay time and event duration (see the bottom row of figure 6).

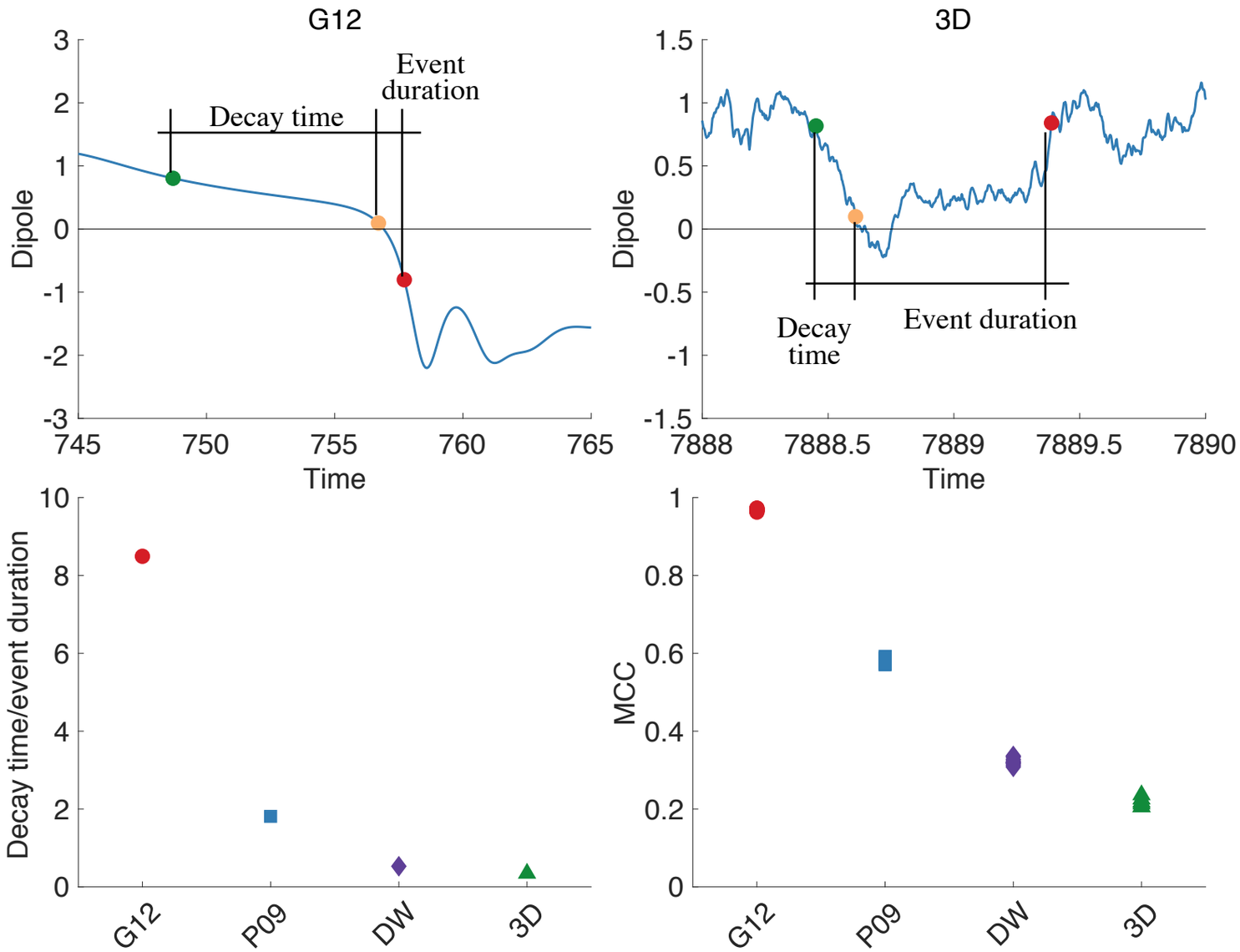


Figure 6. *Top row:* illustration of decay time and event duration for G12 (left) and 3D (right). The dipole strength is scaled by the model's average intensity and time by each model's average event duration. In both plots, the yellow and red markers indicate the beginning and end of an event. The green markers indicate the last time, prior to the start of the event, at which the intensity was above the end-of-event threshold (80%). G12 decays to a low-dipole event much more slowly than DW or 3D making it easier to anticipate a low-dipole event by a warning-threshold strategy. *Bottom row:* Ratio of average decay time to average event duration for each model (left) and the MCC achieved from multiple independent rounds of training and verification (right). The ranking of the models according to their decay time/event duration ratios reflects the ordering according to the typical skill score (MCC) of predictions.

Data needed for training

- A warning threshold that is optimal (in the sense of maximizing the MCC) can be found on a training set with as few as five events however, the reason for this varies across the hierarchy of models.

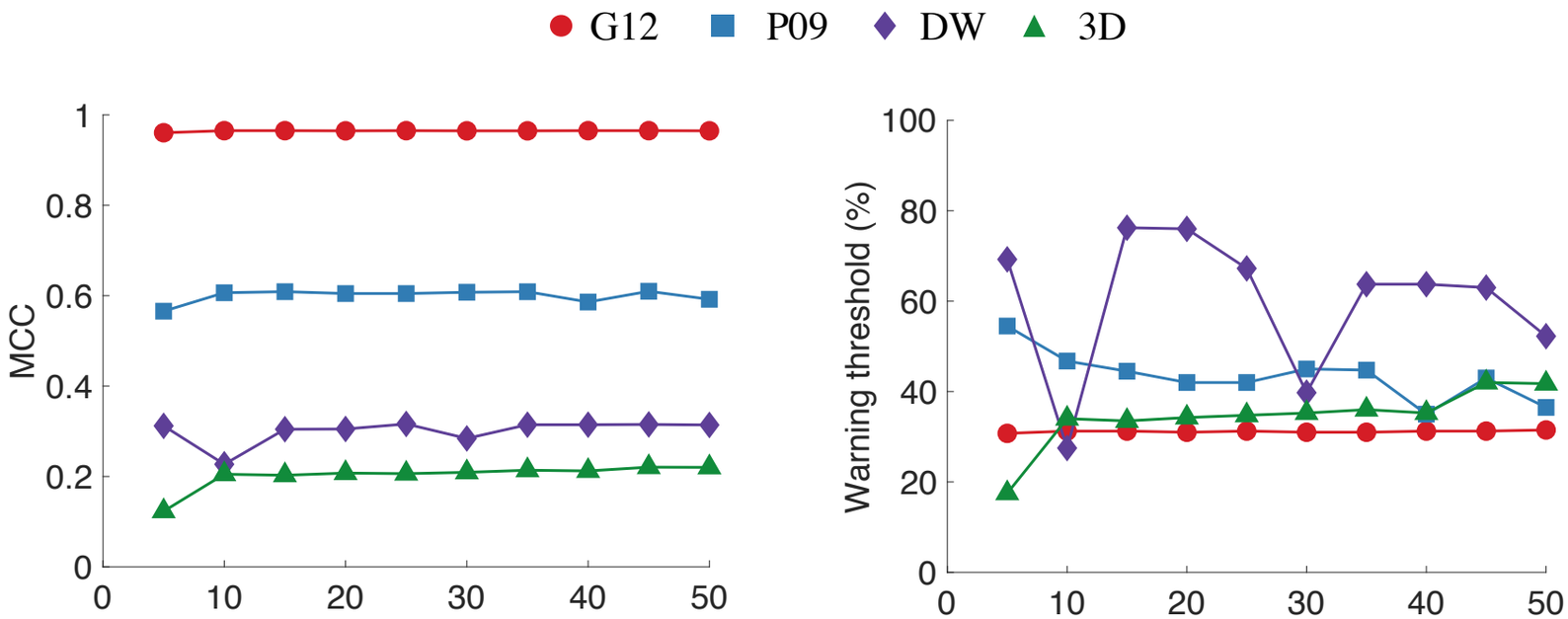


Figure 7. *Left:* MCC (vertical axis) achieved on validation data after training on data containing various numbers of events (horizontal axis). *Right:* optimal warning threshold (vertical axis) determined from training containing various numbers of events (horizontal axis). Notice that while the skill scores (left plot) are steady the warning thresholds learned from training (right plot) for some models, e.g., DW, can vary widely.

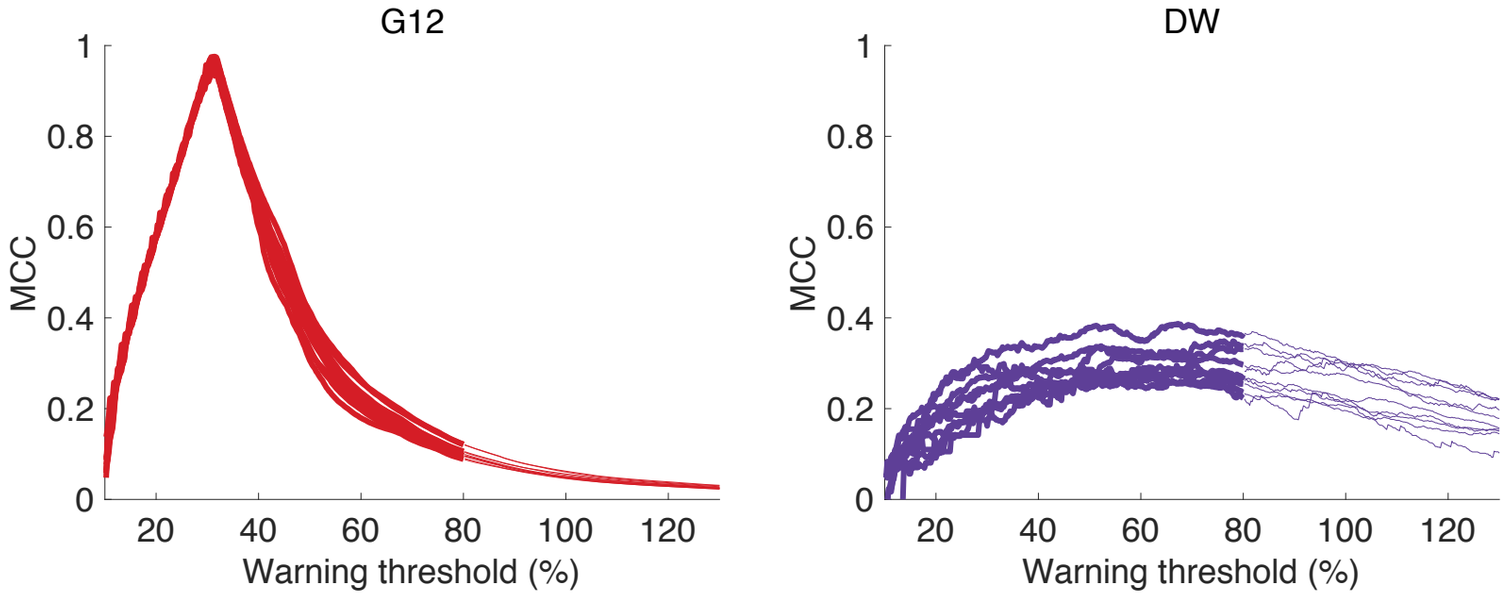


Figure 8. MCC vs. warning threshold for the ten training data sets of G12 (left) and DW (right) used in figure 7. All curves for G12 are sharply peaked and thus an optimal warning threshold is easily identified with minimal training data (containing only five events). All curves of DW are fairly flat, indicating a wide range of warning thresholds will achieve similar (but low) skill scores.

Paleomagnetic reconstructions

- We apply the threshold prediction strategy to the paleomagnetic reconstructions of PADM2M and Sint-2000 and find performance most similar to the P09 model.
- Training is carried out for the period up to 1.05 Myr in the past with the remaining portion used for verification (see magnified portion of figure 9).

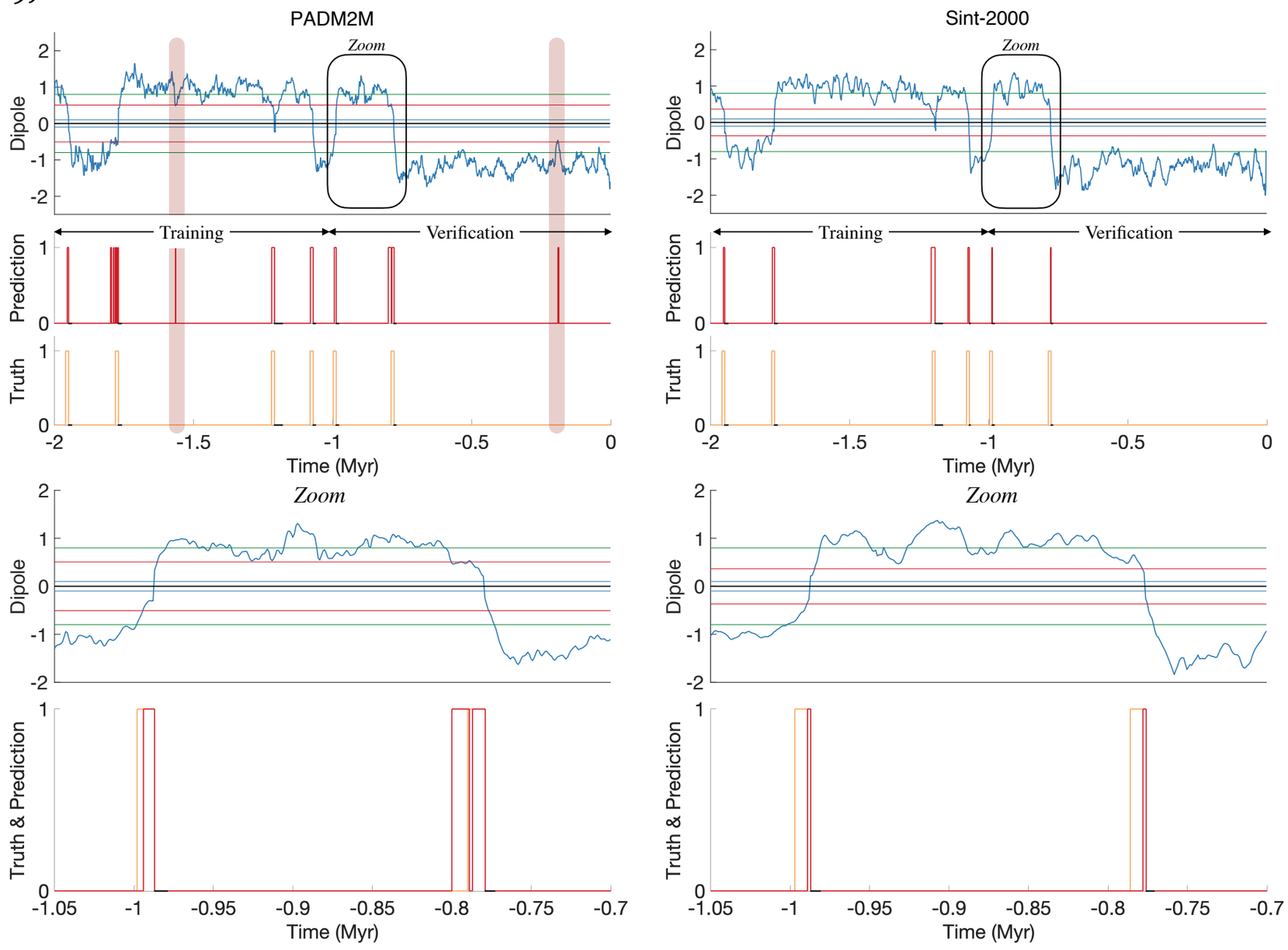


Figure 9. Dipole, prediction and truth plots, mirroring the illustration of figure 2, for PADM2M (left) and Sint-2000 (right). The warning threshold determined for PADM2M of 50.75% (2.70×10^{22} Am^2) results in occasional false positives two of which are highlighted in red. The warning threshold of 36.75% (2.14×10^{22} Am^2) determined for Sint-2000, results in no false positives but tends to predicts events late (see lower right truth and prediction magnification).

Comparison of reconstructions and models

- The performance of the prediction strategy on the paleomagnetic record is most similar to P09.

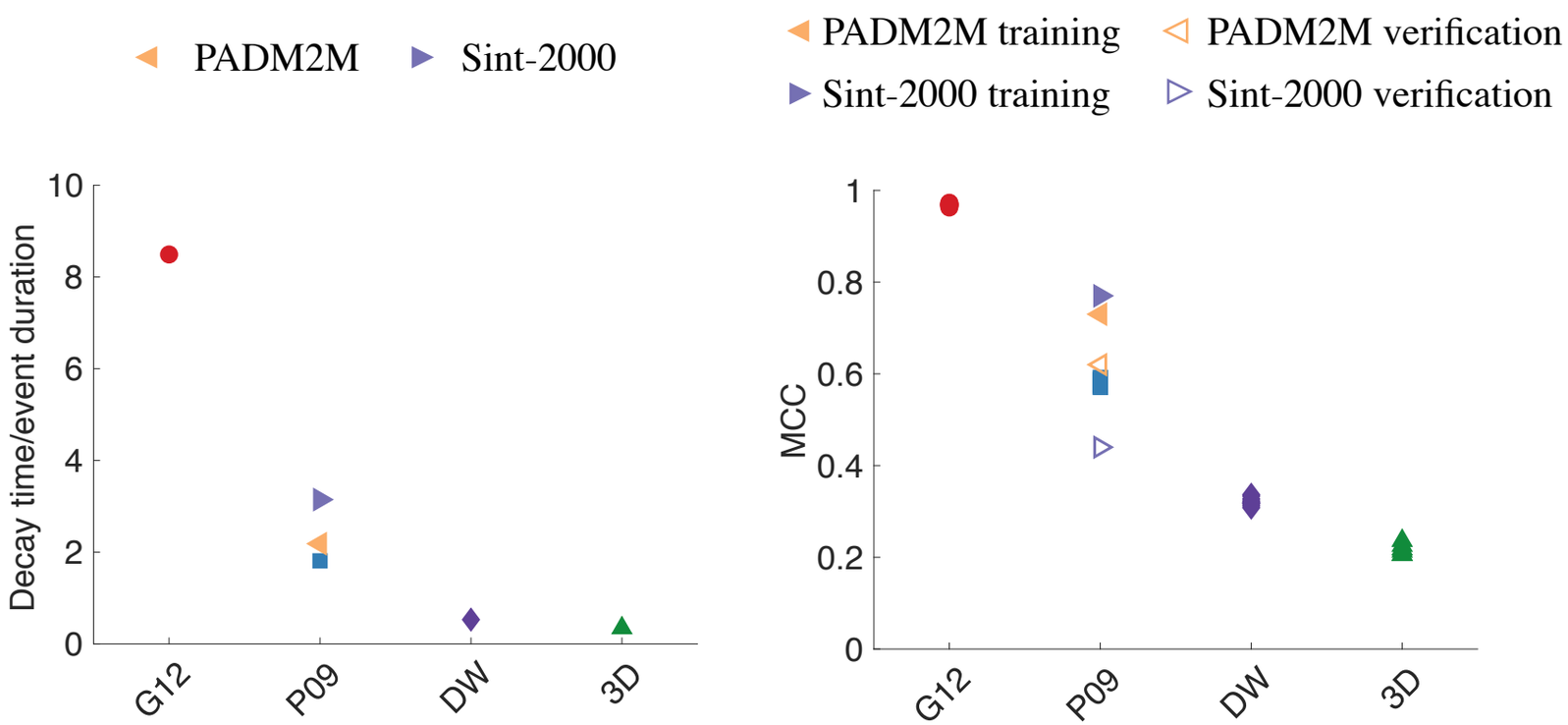


Figure 10. Decay time to event duration ratios (left) and MCC values (right) previously shown (bottom row of figure 6) with results for PADM2M and Sint-2000 added. In both plots, the paleomagnetic reconstructions are seen to be most similar to P09.

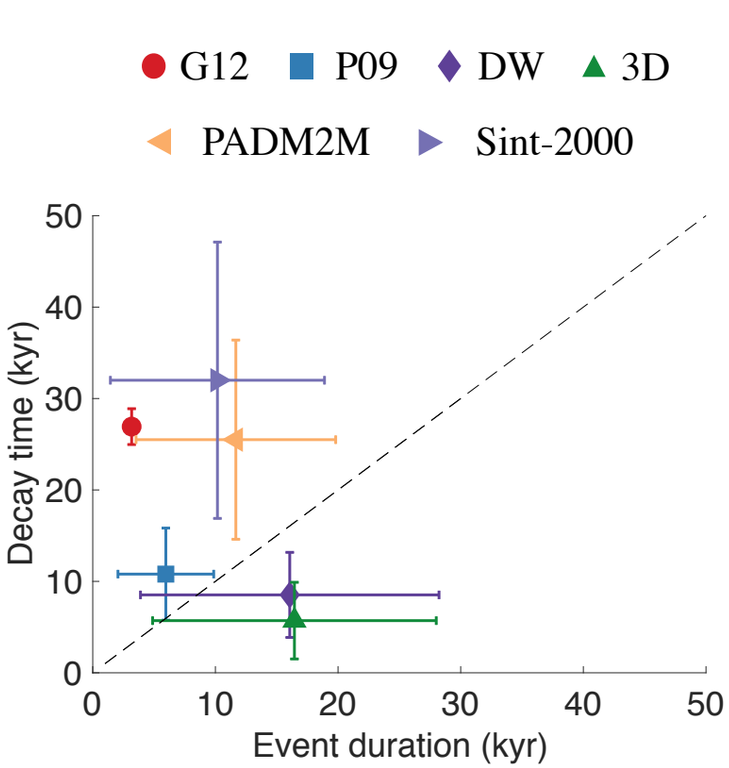
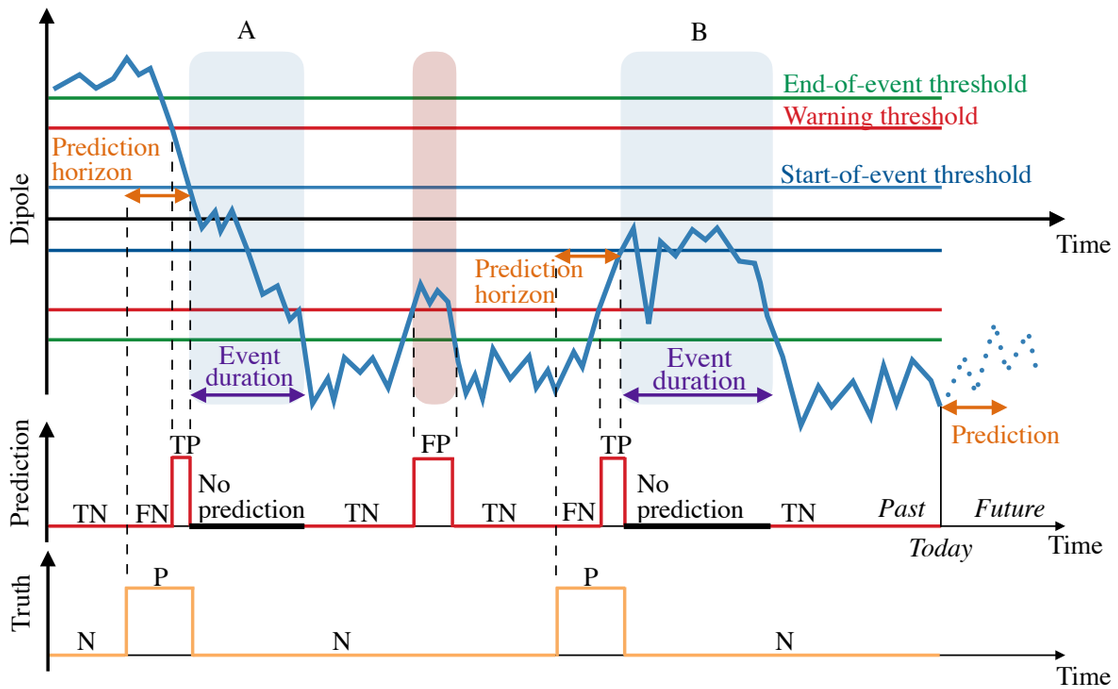
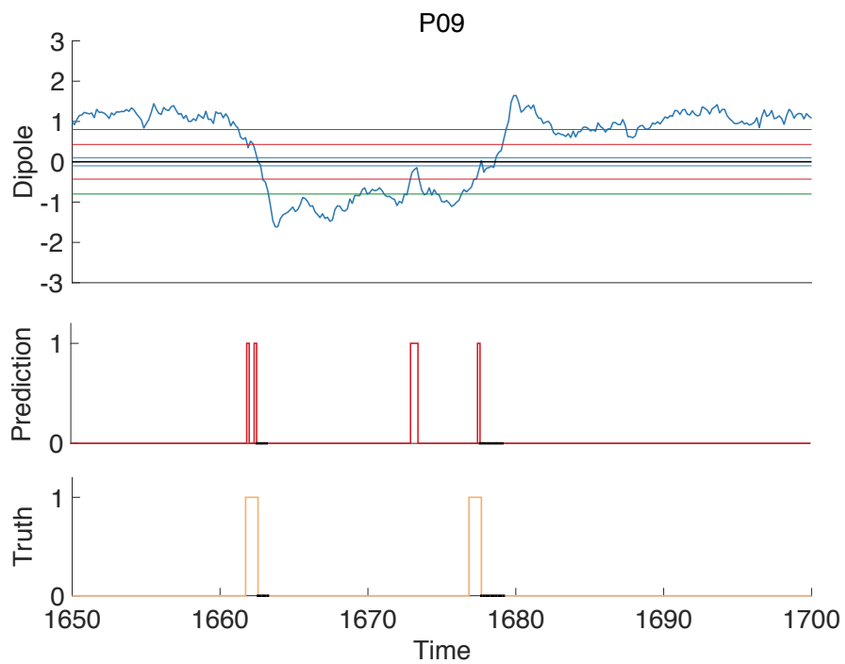


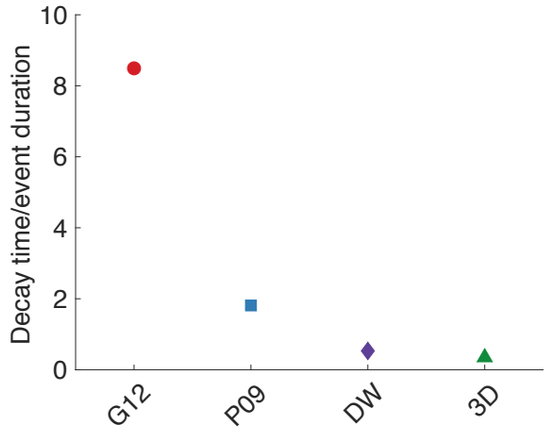
Figure 11. Decay time plotted as a function of the event duration for the four models and the paleomagnetic reconstructions. Shown are the mean and error bars based on one standard deviation. In the case of the G12 model, the standard deviation of the event duration is too small to be visible as an error bar. Also shown is a 45° line that separates models or data for which Decay time > Event duration from models for which Decay time < Event duration.

Concluding remarks





- The primary purpose of this study was to test whether a low value of the axial dipole intensity can be used to predict an upcoming reversal.
- Implementing and evaluating such a prediction strategy in a robust way requires a carefully constructed methodology.



- The performance of threshold-based predictions varies widely across the hierarchy of models.
- Overall, threshold based predictions appear to be of only limited value.
- The speed at which dipole intensity decays plays a significant role in the performance of threshold based predictions.
- We specifically note the asymmetry between decay and recovery time (event duration) which has previously been discussed (see, e.g., Ziegler & Constable 2011; Avery et al. 2017).
- The performance of threshold-based predictions is a discriminating way of testing the Earth-like nature of the axial dipole field of a model.

- Approaches considering the dynamics of the axial dipole field or the thresholding of a combination of features should be investigated.
- Consideration should be given to machine-learning techniques, in particular in the search for precursors to reversals in 3D dynamo simulations.
- Utilizing data assimilation for predicting reversals has been explored (Morzfeld et al. 2017) where it was found that the way in which the G12 model approaches reversals is more similar to Earth than P09. This, in light of the finding that P09 possesses threshold based prediction properties closest to the paleomagnetic reconstructions, suggests the possibility of a useful low-dimensional model with properties intermediate between G12 and P09.

References

Gissinger, C., 2012. A new deterministic model for chaotic reversals, Eur. Phys. J. B., 85, 137.

Pétrélis, F., Fauve, S., Dormy, E., & Valet, J.-P., 2009. Simple mechanism for reversals of Earth's magnetic field, Phys. Rev. Lett., 102, 144503.

Buffett, B., Ziegler, L., & Constable, C., 2013. A stochastic model for paleomagnetic field variations, Geophys. J. Int., 195(1), 86–97.

Ziegler, L. & Constable, C., 2011. Asymmetry in growth and decay of the geomagnetic dipole, Earth planet. Sci. Lett., 312(3), 300–304.

Avery, M. S., Gee, J. S., & Constable, C. G., 2017. Asymmetry in growth and decay of the geomagnetic dipole revealed in seafloor magnetization, Earth planet. Sci. Lett., 467, 79 – 88.

Morzfeld, M., Fournier, A., & Hulot, G., 2017. Coarse predictions of dipole reversals by low- dimensional modeling and data assimilation, Phys. Earth planet. Inter., 262, 8–27.