Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cycle-Scale Forecasting of Galactic Cosmic Ray Intensity: A Skill Assessment and The Role of Recurrent Solar Wind Structure

V. K. Mishra

vkmishra74@yahoo.com

Atal Bihari Vajpayee Hindi Vishwavidyalaya

Praveen Tyagi

Atal Bihari Vajpayee Hindi Vishwavidyalaya

Research Article

Keywords: cosmic rays, solar modulation, space weather forecasting, heliospheric current sheet, corotating interaction regions, solar wind

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11453123/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Cycle-Scale Forecasting of Galactic Cosmic Ray Intensity: A Skill Assessment and The Role of Recurrent Solar Wind Structure

V. K. Mishra*, Praveen Tyagi
Department of Physics, Atal Bihari Vajpayee Hindi Vishwavidyalaya, Bhopal 462003, India
*Email : vkmishra74@yahoo.com

Abstract

Long-term forecasting of Galactic Cosmic Ray (GCR) intensity is conventionally assessed by how closely a model reproduces the observed record rather than by whether it outperforms a naive alternative. Using monthly Oulu and Moscow neutron monitor data for 1976 to 2025, together with heliospheric current sheet (HCS) tilt, Interplanetary Magnetic Field (IMF) magnitude, solar wind speed and the aa geomagnetic index, we evaluate cycle-scale Cosmic Ray Intensity (CRI) forecasting against a persistence benchmark under strict rolling-origin validation. Here all smoothed predictors constructed causally so that no information from after the forecast issue time enters any forecast. A conventional regression on CRI level, evaluated at matched forecast horizons, attains out-of-sample correlations as high as 0.91 yet fails to beat persistence at every horizon. It acquires a systematic bias of approximately -4 % when extrapolated from the calibration period into a later heliospheric state characterised by weaker IMF. Reformulating the target as the change in CRI over the forecast horizon largely removes this bias. The point-estimate skill is negative at six months, where persistence is difficult to beat, and turns positive from twelve months onward, reaching 0.36 at twenty-four months. All 95% bootstrap intervals nevertheless include zero and Diebold-Mariano tests do not reject equal predictive accuracy with persistence. A 27-day recurrence index derived from the aa index, hitherto known as a descriptive correlate of cosmic ray flux, contributes negligible aggregate skill at twelve months despite being individually significant there, and raises the aggregate point-estimate skill at eighteen and twenty-four months, with a standardised coefficient rising monotonically with lead time from 0.09 to 1.20. The orthogonal component of the index, after removal of the variance shared with the mean-field predictors, retains most of that contribution. No statistically robust incremental forecast skill from aggregate transient injection was detected, on either the Richardson and Cane near-Earth Interplanetary Coronal Mass Ejection (ICME) list or the CDAW SOHO/LASCO catalogue, under an acceptance criterion fixed in advance. Calibrating the neutron monitor network against AMS-02 daily proton fluxes accepts seven of twenty-two station and detector combinations and yields median response rigidities from 10 to 72 GV, with Thule, Oulu and Moscow degenerate near 10 GV despite cutoffs differing by a factor of eight, and with atmospheric depth rather than cutoff governing the response. Forecast skill is highest and most variable among monitors near 10 GV, falls to a lower level near 19 GV, and falls essentially to zero at 72 GV, consistent with useful cycle-scale skill being concentrated below approximately 20 GV. Nested cross-validation confirms that adaptive choice of the predictor specification yields no improvement over the fixed model, while retaining the recurrence index in every training fold at the longer horizons. No skill

score attains significance at 95%. For the present validation design the estimated detectability threshold is about 0.4, and the observed 0.36 falls just below it, so the results do not provide sufficient evidence to reject the persistence benchmark. We conclude that the results provide suggestive evidence that recurrent solar wind organisation contains information beyond the mean interplanetary field at horizons of eighteen to twenty-four months, that no comparable evidence is found for aggregate transient injection, and that the available record is too short to establish differences of this magnitude at conventional confidence.

Keywords: cosmic rays, solar modulation, space weather forecasting, heliospheric current sheet, corotating interaction regions, solar wind

1. Introduction

Galactic cosmic rays (GCRs) entering the heliosphere are modulated by the outflowing solar wind and its embedded magnetic field, producing an intensity variation in antiphase with the solar activity cycle (Mishra & Tiwari, 2003; Gupta et.al,2006a; Gupta et.al,2006b; Usoskin, 2017; Mishra & Mishra, 2018; Mishra, 2025). The practical consequences are substantial. GCRs dominate the chronic radiation dose received by crew on interplanetary trajectories, drive single-event upsets in spacecraft and avionics electronics, and contribute to atmospheric ionisation. Mission planning and component qualification both require estimates of the cosmic ray environment months to years in advance, and the required lead times are frequently longer than the horizons at which any forecast has been formally validated. Throughout this paper the forecast horizon denotes the interval between the time at which a forecast is issued and the time to which it refers, so that a twelve-month forecast issued in January 2019 refers to January 2020.

The physical framework is well established. Parker (1965) formulated the transport of energetic charged particles in the expanding solar wind in terms of outward convection, diffusion along and across the mean magnetic field, adiabatic energy loss in the expanding flow, and drift. Jokipii et al. (1977) showed that gradient and curvature drifts in the large-scale heliospheric field introduce a dependence on the sign of the solar polar magnetic field. In epochs conventionally labelled $A > 0$, when the northern polar field is directed outward, positively charged particles reach the inner heliosphere preferentially over the polar regions and are comparatively insensitive to the configuration of the Heliospheric Current Sheet (HCS) whereas in $A < 0$ epochs where they arrive along the current sheet and are correspondingly more sensitive to its tilt. The observational signature is the alternation between flat-topped cosmic ray maxima in $A > 0$ and sharply peaked maxima in $A < 0$, and the associated 22-year Hale-cycle periodicity superposed on the 11-year Schwabe modulation.

Because the modulating medium requires time to propagate to the outer heliosphere, Cosmic Ray Intensity (CRI) responds to solar activity with a delay. Usoskin et al. (1998) established the cross-correlation approach now standard for estimating this lag, and subsequent work has consistently recovered delays of several months (Tomassetti et al. 2017; Koldobskiy et al. 2022). The estimates vary widely between cycles, however, and the variation appears systematic rather than random:

lags of one to four months in even cycles contrast with more than a year in odd cycles (Ross and Chaplin 2019), transport modelling yields delay times of 18 months in Cycle 21 against 12 months in Cycle 23 (Siluszyk and Iskra 2020), and the lag between sunspot number and neutron monitor rates changes with the polarity cycle (Koldobskiy et al. 2022). Mishra and Mishra (2018) examined a broader set of solar and heliospheric parameters and identified HCS tilt as the dominant single modulator. The resulting hysteresis or modulation loops, in which the relation between CRI and any solar proxy is double-valued, are a persistent feature of the literature (Fiandrini et al. 2021) and are wider in $A < 0$ than in $A > 0$.

Two physically distinct channels carry solar variability into the modulating medium. The first is quasi-steady and recurrent. Long-lived coronal holes emit high-speed streams which, interacting with slower wind, form corotating interaction regions that sweep past Earth at the 27-day synodic rotation period. Richardson et al. (1999) found that the recurrent cosmic ray depressions associated with these structures are approximately 50 % larger when $A > 0$ than when $A < 0$, and interpreted this as a polarity dependence of the transport coefficients, specifically a reduction of the radial diffusion coefficient enhancing the effect of solar wind convection in $A > 0$. Kóta and Jokipii (2001) subsequently reproduced this polarity dependence in a three-dimensional transport model. The result is not unchallenged: Gil et al. (2012) found the amplitude of the 27-day CRI variation in the $A < 0$ minimum of 2007 to 2009 to be comparable with that in the $A > 0$ minimum of 1995 to 1997, contradicting the earlier finding, and Gil and Mursula (2017) proposed a combination of convection and drift effects as the explanation. A long-baseline measure of this recurrent organisation already exists. Sargent (1986) constructed the 27-day recurrence index R27 from the autocorrelation of consecutive 27-day intervals of the aa geomagnetic index (Mayaud 1972), as a measure of the cycle-to-cycle stability of high-speed solar wind structure. A revised version, strengthened by pre-smoothing the aa input, has since been circulated as a preprint (Sargent 2021). Chapman et al. (2021) re-engineered the index directly from the autocovariance of aa, mapped it onto a Hale-cycle phase clock, and reported among other results that cosmic ray flux rises in step with the recurrence parameter and then remains high. That relationship has, to our knowledge, been described but never exploited predictively.

The second channel is transient. Coronal Mass Ejections (CMEs) inject magnetised plasma into the heliosphere, and at large heliocentric distances successive ejecta merge to form global merged interaction regions which produce step-like decreases in CRI. Near Earth, Richardson and Cane (2011) quantified the GCR response to individual Interplanetary Coronal Mass Ejections (ICMEs) and magnetic clouds. Whether the integrated transient injection contributes independently to cycle-scale modulation, over and above what the mean interplanetary field already registers, is a separate question and is one we address directly below.

Numerous models have been developed to predict the evolution of GCR intensity, and a forecasting framework has recently been set out by Pelosi et al. (2025), which splits the record into phases of constant polarity while excluding the magnetic reversal phase according to the criteria of Sun et al. (2015). He applies a time lag to address the double-valued behaviour of the correlation

function. What such studies characteristically report is agreement between model output and observed flux. What they do not report is skill relative to a naive baseline. By skill we mean the fractional reduction in mean-square forecast error achieved relative to a stated benchmark, and by the naive baseline we mean the persistence forecast, which asserts simply that the quantity of interest will be unchanged over the interval concerned. This distinction has practical consequences. CRI is a smooth, strongly autocorrelated quantity; a forecast which simply asserts that the intensity will not change over the coming year already achieves a low Root-Mean-Square Error (RMSE). Unless a model is shown to improve upon that assertion, its reported agreement with data does not establish that it carries useful predictive information.

This paper addresses that gap. We evaluate cycle-scale CRI forecasting against an explicit persistence benchmark. Validation is by expanding-window rolling origin, in which no observation at or after a forecast target time enters the fit that produced it. Every predictor is evaluated at the time the forecast is issued, and all smoothing windows are trailing. The conventional levels formulation fails against persistence at every horizon tested, and we diagnose the reason. A differenced formulation yields positive point-estimate skill from twelve months onward. The recurrence index, used as a predictor rather than as a descriptor, raises that skill at eighteen and twenty-four months but not at twelve. Aggregate transient injection, tested under a criterion fixed before the data were examined, yields no detectable contribution at any horizon. Two further contributions are methodological. We introduce a continuous polarity index that retains the reversal epochs conventionally discarded, and we calibrate the response rigidity of each neutron monitor against AMS-02 daily proton fluxes in order to ask how forecast skill varies with the energy to which a station responds.

The study differs from existing work in three specific respects. Reikard (2024) applies persistence benchmarking to the solar modulation potential; we apply it to neutron-monitor intensity and add an explicit decomposition into recurrent and transient heliospheric channels. Pelosi et al. (2025, 2026) solve the Parker transport equation and report reconstruction error against observed flux; we ask the complementary empirical question of whether observed heliospheric quantities carry incremental predictive information relative to a naive baseline. Chapman et al. (2021) described the association between the recurrence index and cosmic ray flux; we measure whether that association yields forecast skill, and at which horizons.

2. Data

Table 1 summarises the observational inputs. CRI is taken from the Oulu neutron monitor, with a geomagnetic cutoff rigidity of 0.8 GV, and independently from the Moscow neutron monitor at 2.4 GV, which serves as a replication at different rigidity. Both series are pressure- and efficiency-corrected monthly means. Count rates are expressed as a percentage of the mean level of 1965, a solar minimum epoch, so that CRI = 100 % denotes the 1965 baseline. Months recorded as zero in the Moscow series are treated as missing.

HCS tilt angles are taken from the Wilcox Solar Observatory (WSO) potential field source surface computations, in both the classic line-of-sight and radial boundary formulations. Each Carrington rotation is expanded to daily values and averaged to calendar months. The WSO tilt series available to us terminates at Carrington rotation 2302, which begins on 9 September 2025 and extends into the following month. October 2025 is therefore the last calendar month for which a complete monthly tilt average can be formed, whereas all other inputs extend into 2026. This truncation is propagated through the analysis and is stated explicitly wherever it constrains a result.

Interplanetary parameters are drawn from the OMNI compilation at 27-day Bartels rotation resolution, comprising the magnetic field magnitude (the root-mean-square of the field magnitude), solar wind bulk speed, and the Kp and ap geomagnetic indices. Bartels rotations are expanded to daily values and averaged to calendar months in the same manner as the tilt series. Standard OMNI fill values are set to missing throughout. Sunspot numbers are the version 2 monthly and 13-month smoothed series from WDC-SILSO (Clette and Lefevre 2016).

The aa geomagnetic index (Mayaud 1972) is taken at three-hourly resolution from ISGI for 1 January 1927 to 10 September 2026, comprising the definitive series to 2024 and provisional values thereafter. Daily means are formed from the eight three-hourly values before any derived index is computed.

Transient activity is characterised from two separate catalogues. The near-Earth ICME list of Richardson and Cane (2010), updated from Cane and Richardson (2003) which supplies 643 events between May 1996 and July 2026 with onset and end times, mean field strength, ICME speed and minimum Dst. The CDAW SOHO/LASCO CME catalogue (Gopalswamy et al. 2009, 2025; Yashiro et al. 2004), in its version 2 release, supplies 43,139 events between January 1996 and May 2026 with central position angle, angular width, linear speed and, where available, mass and kinetic energy, together with LASCO/C2 downtime files used for duty-cycle normalisation.

For the rigidity analysis of Section 4.8 we additionally use five neutron monitors spanning a wide range of geomagnetic cutoff: Thule (0.30 GV), McMurdo (0.01 GV), Mexico City (8.2 GV), Tsumeb (9.29 GV) and Doi Inthanon, the Princess Sirindhorn monitor (16.8 GV), all obtained from NMDB. For Tsumeb the NMDB series is adopted in preference to the WDC-CR archive at Nagoya: the two agree to $r = 0.998$ over 476 overlapping months, but the archive carries an uncalibrated scale change of a factor 1.67 at the start of 2019 with no 2017 data to bridge it, whereas the NMDB series is continuous. The NMDB series in turn shows a 16 % step between November 2019 and January 2020, so Tsumeb is used only to November 2019. Space-borne reference fluxes are the AMS-02 daily proton measurements of Aguilar et al. (2021), comprising 2824 daily spectra from 20 May 2011 to 29 October 2019 in 30 rigidity bins between 1 and 100 GV.

Source	Quantity	Span	Resolution
Oulu NM	Corrected count rate, Rc = 0.8 GV	1964-04 to 2026-08	monthly
NMDB	Thule, McMurdo, Mexico City, Tsumeb, Doi Inthanon	1957 onward	monthly
AMS-02	Daily proton flux, 1-100 GV, 30 bins	2011-05 to 2019-10	daily
Moscow NM	Corrected count rate, Rc = 2.4 GV	1964-04 to 2026-08	monthly
WSO	HCS tilt, classic and radial models	CR 1642 to CR 2302	Carrington rotation
WSO	Polar field, filtered N, S and mean	1976-05 onward	10-day
WDC-SILSO	Sunspot number v2, monthly and smoothed	1749 onward	monthly
OMNI	|B|, RMS |B|, Vsw, Kp, ap	1963-11 onward	27-day
ISGI	aa geomagnetic index	1927-01 to 2026-09	3-hourly
Richardson & Cane	Near-Earth ICME list, 643 events	1996-05 to 2026-07	per event
CDAW / LASCO v2	CME catalogue, 43 139 events	1996-01 to 2026-05	per event

Table 1. Observational inputs, spans and native resolutions. All series are aggregated to a common monthly grid.

3. Methods

3.1 A continuous polarity index

The sign of the solar polar field, A, is conventionally treated as a binary state. Months near a reversal, where the sign is ill defined, are then excluded from analysis, following criteria such as those of Sun et al. (2015); this is the approach adopted by Pelosi et al. (2025, 2026). Exclusion carries two costs. It discards several years of record in each cycle, and it imposes a discontinuity at the reversal, which is the point at which the drift pattern is in fact reorganising gradually rather than abruptly. We therefore replace the binary state by a continuous index. The 20 nHz filtered WSO north polar field is first smoothed over a trailing 25-month window, which suppresses the

annual modulation arising from the Earth's heliographic latitude. The smoothed field is then divided by a scale s and passed through a hyperbolic tangent.

The scale determines how rapidly the index saturates. It is set to 61 microtesla, the 60th percentile of the absolute smoothed field over the record, so that the field strength at which the index reaches about three quarters of saturation is one that is exceeded in roughly 40 per cent of months. Months well inside a polarity epoch are therefore close to plus or minus one, while the index varies smoothly through zero across a reversal. Defining the scale as a percentile of the record rather than as a fixed value in microtesla makes it self-calibrating, so that the construction can be applied unchanged to a different polar field series. The choice is not critical: the hyperbolic tangent is monotonic, so the sign and ordering of the index are unaffected by the scale, and the predictors are in any case standardised within each training fold.The window is strictly trailing rather than centred: the smoothed value at any month uses only that month and the preceding twenty-four, so that when the index is later evaluated at a forecast issue time t - h, meaning the month in which the forecast is made, it contains no information from after t - h. A centred window is the more usual choice for smoothing a geophysical series, but it is inadmissible here: a centred twenty-five-month window would place up to twelve months of information beyond the nominal date into the smoothed value, and that information would not have been available when the forecast was issued. The effect is concentrated at the reversal transitions, where the two constructions disagree in sign in a substantial fraction of months. All results reported below use the trailing construction. The index saturates near ± 0.85 well inside each epoch and passes smoothly through zero at reversal, and is defined for every month from 1977 onward, the first twenty-four months of the record being required to initialise the trailing window.

Applying a sign criterion to this index recovers the conventional epoch boundaries (Table 2). Because the smoothing window is trailing, a reversal is registered only after it has been in progress for several months, so the boundaries in Table 2 lag the true reversal dates by up to about a year; this is a direct and expected consequence of causality and is preferable to a centred window that would anticipate the reversal. These boundaries are thus derived from the same data used in the analysis rather than adopted from external determinations.

Epoch	**Polarity**	**Duration**
1977-05 to 1980-09	A > 0	41 months
1981-02 to 1990-10	A < 0	117 months
1991-04 to 2001-02	A > 0	119 months
2001-06 to 2012-11	A < 0	138 months
2016-03 to 2024-04	A > 0	98 months
2025-02 to 2026-08	A < 0	19 months

Table 2. Magnetic polarity epochs derived from the sign of the smoothed WSO north polar field. Intervals between listed epochs are reversal transitions, over which the continuous index passes through zero.

3.2 The recurrence index

Following the autocorrelation construction of Chapman et al. (2021), which re-engineered the original 27-day recurrence index of Sargent (1986), we compute the lag-27-day autocorrelation of daily aa within a trailing 81-day window, equal to three solar rotations, ending on the day in question, and average the result to calendar months. The aa series is used without prior smoothing. The index therefore corresponds to the original R27 construction as re-engineered by Chapman et al. (2021), and not to the revised index of Sargent (2021), which differs chiefly in pre-smoothing the aa input. As with the polarity index, the window is trailing rather than centred, since a centred 81-day window would admit roughly forty days of information postdating the nominal date. The two constructions correlate at 0.78 across the full record and diverge most where the recurrence level is changing rapidly, so the choice is not cosmetic. Explicitly, with a(d) the daily mean aa index and W(d) the set of 81 consecutive days ending on day d, the daily index is the Pearson correlation between a(i) and a(i - 27) taken over i in W(d), evaluated on those days for which both values are present. A minimum of 50 valid pairs is required and days failing that condition are left undefined. Missing aa values are excluded pairwise rather than interpolated, and the monthly value is the unweighted mean of the daily indices falling within the calendar month. Because the window ends on day d, no observation later than d enters the index. We denote the resulting monthly series aa(r27). High values indicate a solar wind organised into long-lived corotating streams; low values indicate a disordered wind dominated by transients or by structures that do not survive a full rotation. The index requires daily resolution and cannot be constructed from monthly averages. This is why it is built from the three-hourly aa series rather than from the interplanetary data of Section 2, which are used here at 27-day Bartels resolution and therefore cannot resolve recurrence at the rotation period.

3.3 Forecast formulation

For a forecast horizon h the target variable is the change in CRI over the horizon,

$$dCRI(t) = CRI(t) - CRI(t - h)$$

Every predictor is evaluated at time t - h, so that the model constitutes a genuine h-month forecast requiring no information unavailable at the time the forecast is issued. This constraint is more restrictive than it may appear and is, in our view, the point at which many published models lose their claim to being forecasts. A regression which includes a contemporaneous interplanetary field measurement is a diagnostic relation, not a prediction, because that measurement does not exist in advance.

The predictor set comprises the CRI at t - h and its rate of change over the preceding 12 months; the HCS tilt, its square, and its 12-month rate of change; the interplanetary field magnitude and its

12-month rate of change; the solar wind speed; the continuous polarity index together with its interactions with tilt and with field magnitude; and, in the fuller specification, the recurrence index. We designate the model without the recurrence index as Model C and the model including it as Model D.

Predictors are standardised using training-set statistics only, so that no information from the validation period enters through the scaling. Coefficient inference uses the heteroskedasticity- and autocorrelation-consistent covariance estimator of Newey and West (1987) with 12 lags, which is necessary because monthly overlapping differences guarantee strongly autocorrelated residuals and Ordinary Least Squares (OLS) standard errors would be substantially understated.

3.4 Validation and skill measurement

Validation follows an expanding-window rolling-origin scheme, in which the period used for fitting grows as the origin advances through the record. The model is fitted on all data up to an origin. The set of observations so used being termed the training fold and the observations on which the resulting forecasts are judged lying outside it. The frozen fit is then used to generate one forecast for each of the twelve calendar months following the origin, each using predictors evaluated at that month's own t - h; the origin is then advanced by one year and the model refitted. This yields twelve forecasts per annual refit and, over the validation span used here, 360 forecasts at each horizon. No observation at or after a forecast target time enters the fit which produced that forecast, and no predictor value used to generate an individual forecast is drawn from after that forecast's own issue time (t - h); refitting only once per year is a computational convenience and does not relax either constraint, since the model is a fixed function evaluated at twelve subsequent, still-unseen predictor vectors. Random k-fold cross-validation, which is occasionally applied to time series, is inadmissible here because it permits future observations to inform past predictions and inflates apparent skill.

Table 2a gives a worked example of the indexing for a single forecast, to make the data availability explicit. Skill is measured against persistence. For the differenced target, persistence corresponds to the null forecast dCRI = 0, that is, the assertion that CRI will be unchanged over the horizon. The skill score is given by Equation (4), so that positive values indicate improvement upon that assertion and zero indicates parity with it.

Item	Value for the example
Forecast target month	January 2020
Forecast horizon h	12 months
Forecast issue month (t - h)	January 2019
Target quantity	CRI(Jan 2020) - CRI(Jan 2019)
Latest predictor observation permitted	January 2019

Item	Value for the example
Rate-of-change predictors span	January 2018 to January 2019
Recurrence index window	81 days ending January 2019
Polarity index window	25 months ending January 2019
Training origin for this forecast	December 2018
Latest training observation	target month December 2018, issued December 2017
Standardisation statistics	computed on training observations only
Observations from 2019 or 2020 used in fitting	none

Table 2a. Worked example of the forecast indexing for a single twelve-month forecast. Each of the twelve target months following an origin is forecast separately from its own predictor vector; the model is refitted annually rather than monthly to limit computation, and the fit for a given origin uses no observation whose target month falls at or after the first forecast month.

3.5 Pre-specified test of the transient channel

Transient activity admits many candidate monthly aggregates, and testing several against a single dataset invites selection of whichever happens to reach significance. To guard against this, an acceptance criterion was pre-specified, that is, fixed before the transient catalogues were examined. A variable is retained only if it adds at least 0.02 to the rolling-origin skill score at h = 18, has a Heteroscedasticity and Autocorrelation Consistent (HAC) p-value below 0.05, and has a variance inflation factor (Equation 9), below 5. Any variable satisfying these conditions is further required to exhibit a coherent dependence on horizon across h = 12, 18 and 24, and to replicate on an independent station. The rationale for the horizon requirement is that a genuine physical predictor should not contribute at one horizon and detract at neighbouring ones. Moscow is used for the replication because the transient test spans 1996 to 2025, and of the accepted stations only Thule, Oulu and Moscow have continuous records across that interval; Moscow is the one of the three that differs most from Oulu in cutoff rigidity, location and calibration history. As Section 4.8 shows, all three nevertheless share an effective response rigidity near 10 GV, so the replication tests for a station-specific artefact rather than for a dependence on particle energy.

3.6 Rigidity response calibration

Geomagnetic cutoff rigidity is an imperfect abscissa for comparing neutron monitors, because the atmospheric yield function, the probability that a primary particle of given rigidity produces a count in the detector, rises steeply with energy and the detector response therefore extends far above the cutoff, by an amount that also depends on atmospheric depth. We calibrate the response empirically against AMS-02.

AMS-02 fractional flux variations at different rigidities are almost perfectly correlated in time and differ essentially only in amplitude. A fit with a free overall gain is consequently degenerate, since any median rigidity can be made to reproduce a given station by adjusting the gain. We therefore fix the gain at unity, so that the station amplitude alone determines the response, and define the median response rigidity R50 operationally as the rigidity at which the AMS-02 fractional variation amplitude equals that of the station over the common window. R50 is an empirical effective response rigidity obtained by amplitude matching against a primary proton spectrum. It should not be read as the median of the neutron monitor yield function since the yield function is broad, depends on atmospheric depth, and describes the response to secondaries produced in the atmosphere rather than to primaries directly. R50 is used here only as a physically ordered abscissa for cross-station comparison, in place of the geomagnetic cutoff.

Each station is first screened by its maximum correlation with the AMS-02 fractional variation over any rigidity bin. A sound monitor must track the space-borne flux at some rigidity; a station failing this test carries instrumental drift rather than modulation. We adopt a threshold of 0.90.

3.7 Statistical measures and inference

This subsection sets out the definitions of the forecast-evaluation and inference quantities used throughout the paper. Each equation is followed by a definition of the terms appearing in it, and the numbered equations are referred to by number in the Results and Discussion. Subscripts identify the quantity concerned, the time index or the predictor whereas superscripts in parentheses identify the forecast horizon or the bootstrap replicate.

For forecast horizon h, the target is the change in CRI across the horizon,

$$y_t^h = CRI_t - CRI_{\{t-h\}}, \tag{1}$$

in which CRI_t, written here with the month as a subscript, is the monthly Oulu intensity in month t expressed as a percentage of the 1965 baseline; the subscript t − h denotes the same quantity h months earlier; and the superscript (h) on the target records the horizon to which it refers. Writing the model forecast of that change as $\hat{y}$ with the same subscript and superscript, the forecast error is

$$e_t^h = y_t^h - \hat{y}_t^h \tag{2}$$

the observed change minus the forecast change in month t. The persistence benchmark is the null forecast $\hat{y} = 0$, that is, the assertion that intensity will be unchanged over the horizon; its error is therefore y_t^h itself, and where the two must be distinguished the errors carry the further subscripts model and pers.

Averaging the squared error over the N_h forecasts available at horizon h gives the mean-square error,

$$MSE(h) = \left(\frac{1}{N_h}\right) \sum e_t^2, \tag{3}$$

in which the subscript h on N records that the number of forecasts depends on the horizon, the summation subscript t indicates that the sum runs over all those forecasts, and the superscript 2 is an exponent rather than an index. The RMSE is the square root of this quantity. Forecast skill relative to persistence is

$$S(h) = 1 - \frac{MSE_{model}(h)}{MSE_{pers}(h)}, \tag{4}$$

where the subscripts model and pers (persistence) identify which forecast the mean-square error belongs to, both being computed over the same set of forecasts. A value $S > 0$ denotes improvement upon persistence, $S = 0$ parity with it, and $S < 0$ inferior performance.

Predictors enter the regression after standardisation within each training fold,

$$x^*_{j,t} = \frac{x_{j,t} - \bar{x}_{j,train}}{s_{j,train}}, \tag{5}$$

in which the paired subscript j,t identifies the j^{th} predictor in month t, so that $x_{j,t}$ is its raw value and $x^*_{j,t}$ the standardised value. The subscript j,train on the mean and on the standard deviation is the essential one: it indicates that both are computed over the training observations of the current fold only, not over the whole record and not over the validation period. As a worked example, for the training fold ending in December 1995 the HCS tilt has a training mean of 45.94 degrees and a training standard deviation of 19.53 degrees. A forecast for June 1996 at h = 24 is issued in June 1994, when the tilt was 29.47 degrees, so the standardised value entering the regression is (29.47 − 45.94) / 19.53 = −0.843. The negative sign and magnitude mean that the current sheet was then 0.843 training standard deviations flatter than the training-period average, as expected in the declining phase of Cycle 22. Note that the training statistics are used even though the target month lies outside the training period. The fitted forecasting equation is then

$$y^h_t = b_0 + \sum_j b_j x^*_{j,t-h} + \varepsilon_t, \tag{6}$$

where the b_0 denotes the intercept, the subscript j the coefficient on the j^{th} standardised predictor, the summation subscript j indicates that the sum runs over the predictors of the specification concerned, and ε_t is the residual in month t. The subscript t − h on the predictor is the point of the construction: every predictor is evaluated at the forecast issue time rather than at the target time.

Because the target is constructed from overlapping h-month intervals, the residuals are serially correlated and OLS standard errors understate coefficient uncertainty. The covariance matrix is therefore estimated by the heteroskedasticity- and autocorrelation-consistent sandwich estimator of Newey and West (1987),

$$Var_{HAC}(\hat{b}) = (X'X)^{-1} X' \Omega_{HAC} X \, (X'X)^{-1}, \tag{7}$$

in which the subscript HAC on the variance and on Ω marks them as the heteroskedasticity- and autocorrelation-consistent forms; $\hat{b}$ is the vector of fitted coefficients; X is the design matrix of standardised predictors with one row per training observation and X′ its transpose and Ω_{HAC} is the

estimated residual covariance matrix, formed with a truncation lag of 12 months. The reported coefficient statistic is

$$t_j = \frac{\hat{b}_j}{SE_{HAC}(\hat{b}_j)} \ , \tag{8}$$

in which the subscript j on both sides identifies the predictor concerned, and the subscript HAC on the standard error indicates that it is the square root of the corresponding diagonal element of Equation (7) rather than an ordinary least-squares standard error. The associated p-value is the two-sided probability of obtaining a statistic at least as large in absolute value as the one observed, were the coefficient in truth zero; small values therefore indicate that the coefficient is unlikely to have arisen by chance alone. The same convention applies to the p-values reported for the forecast-comparison tests in Section 4.9.

Collinearity among predictors is diagnosed by the variance inflation factor,

$$VIF_j = \frac{1}{1 - R_j^2} \ , \tag{9}$$

in which the subscript j identifies the predictor, and R_j^2 is the coefficient of determination, squared by convention in its symbol, obtained by regressing predictor j on all the remaining predictors of the specification. A value near unity indicates a predictor close to orthogonal to the others. For the recurrence index this auxiliary regression gives $R_j^2 = 0.323$, so VIF = 1 / (1 − 0.323) = 1.48; the same auxiliary regression underlies the orthogonalisation test reported in Section 4.4. Goodness of fit is summarised by

$$R^2 = 1 - \frac{SSE}{SST} \ , \tag{10}$$

where SSE is the residual sum of squares and SST the total sum of squares of the target about its mean. The degrees-of-freedom adjusted form is used where the number of predictors differs between specifications. Residual autocorrelation is measured by the Durbin-Watson statistic,

$$DW = \frac{\sum_t (e_t - e_{(t-1)})^2}{\sum_t e_t^2}, \tag{11}$$

in which the subscripts t and (t − 1) denote successive residuals, so that the numerator accumulates squared first differences and the denominator the squared residuals themselves. A value near 2 indicates uncorrelated residuals; the values near 0.55 reported below indicate strong positive autocorrelation, which the HAC estimator accommodates in the inference but does not remove from the specification.

Uncertainty in the skill score is quantified by a moving-block bootstrap applied to the paired out-of-sample squared-error series, the term denoting a resampling scheme that draws contiguous blocks of months rather than individual months, so that serial dependence survives the resampling. Blocks of L = 24 consecutive months are drawn with replacement and concatenated until the original validation length is recovered, the model and persistence series being resampled jointly

so that their month-by-month pairing is preserved. For bootstrap replicate b the skill is recomputed as

$$S^{*(b)} = 1 - \frac{MSE^{*(b)}_{model}}{MSE^{*(b)}_{pers}}, \tag{12}$$

in which the superscript (b) in parentheses identifies the bootstrap replicate rather than denoting a power, the asterisk marks a quantity computed on resampled rather than observed data, and the subscripts model and pers carry the same meaning as in Equation (4). The 95 % interval is taken from the 2.5th and 97.5th percentiles of B = 4000 such replicates.

Predictive accuracy is tested separately using the Diebold-Mariano framework. Under squared-error loss the loss differential is

$$d_t = \left(e_{model,t}\right)^2 - \left(e_{pers,t}\right)^2, \tag{13}$$

in which the paired subscripts (model,t) and (pers,t) identify whose error is being squared and in which month, so that d_t is the excess squared error of the model over that of persistence in month t; negative values indicate months in which the model was the more accurate. The null hypothesis of equal predictive accuracy is

$$H_0: E[d_t] = 0, \tag{14}$$

in which the H_0 marks the null hypothesis and E denotes expectation. The test statistic is formed from the sample mean loss differential $\bar{d}$, the average of d_t over the N forecasts, and its Long-Run Variance (LRV), meaning the variance of that mean computed with allowance for autocorrelation rather than on an assumption of independence,

$$LRV(\bar{d}) = \left(\frac{1}{N}\right)[\gamma_0 + 2\sum_{k=1}^{Z} w_k \gamma_k], \tag{15}$$

in which γ_0 is the sample variance of d_t, the subscript 0 denoting lag zero; γ_k is its sample autocovariance at lag k; w_k is the Newey-West weight applied at that lag; and the summation subscript and superscript indicate that k runs from 1 to the truncation lag Z. We set Z = h − 1, because a forecast at horizon h shares h − 1 months of the underlying intensity record with its neighbour: at h = 24, for instance, the targets ($CRI_t - CRI_{t-24}$) and ($CRI_{t+1} - CRI_{t-23}$) have 23 months in common. The statistic (Diebold-Mariano test) is

$$DM = \frac{\bar{d}}{\sqrt{[LRV(\bar{d})]}}, \tag{16}$$

the mean loss differential divided by the square root of its long-run variance. It is corrected by the Harvey-Leybourne-Newbold small-sample adjustment and referred to the t distribution on N − 1 degrees of freedom. A negative value favours the model, which then has the smaller mean-square forecast error. Model D is compared directly with Model C by the same construction, with d_t redefined as the difference between their squared errors.

The bootstrap and the Diebold-Mariano test address different aspects of uncertainty. The bootstrap quantifies uncertainty in the estimated skill itself, whereas the Diebold-Mariano test evaluates whether model and benchmark have equal expected predictive loss. Both accommodate serial dependence but rest on different inferential constructions, so agreement between them carries more weight than either alone.

Finally, the lag analysis of Section 4.1 uses the cross-correlation between two monthly series x and y at lag k,

$$r(k) = \sum_t \frac{(x_t - \bar{x})(y_{t-k} - \bar{y})}{\sqrt{[\sum_t (x_t - \bar{x})^2 \sum_t (y_{t-k} - \bar{y})^2]}}, \quad (17)$$

where x_t denotes the CRI in month t, y_{t-k} denotes the candidate driver k months earlier, and $\bar{x}$ and $\bar{y}$ their respective means over the interval of overlap; $\sum_t$ indicates that the sums run over all months for which both values are present. The lag maximising the absolute value of r(k) is adopted as the optimum lag.

3.8 Assessment of skill differences

A skill score is a point estimate and requires an accompanying statement of uncertainty. Because the differenced target at horizon h is constructed from overlapping intervals, consecutive monthly forecasts are strongly dependent and the nominal count of forecasts overstates the information available. We therefore assess skill differences in two ways. Confidence intervals are obtained by a moving-block bootstrap with a block length of 24 months and 4000 replicates, resampling the paired model and persistence squared-error series jointly in 24-month moving blocks. Therefore both the within-series temporal dependence and the month-by-month pairing between the two series are preserved. Forecast-error differences against persistence are tested with the Diebold-Mariano (DM) statistic, using a Newey-West (NW) variance with h - 1 lags to accommodate the overlap and the Harvey-Leybourne-Newbold small-sample correction. We additionally report, for each horizon, the smallest skill score that would attain significance at 95 % given the observed bootstrap dispersion, since a non-significant result is interpretable only against the sensitivity of the test that produced it.

The contribution of individual predictor groups is assessed by ablation, the removal of one predictor group at a time in order to gauge its contribution, both as a nested build-up from persistence and by removing each group in turn from the full specification. In each case re-running the complete rolling-origin protocol rather than re-reading coefficients from a single fit. Ablation identifies which groups contribute but cannot be used to choose a specification, because a specification selected by inspecting validation skill is no longer validated out of sample. To test whether adaptive specification is worthwhile we therefore use nested cross-validation. The outer loop is the rolling-origin scheme already described. Within each outer training set, and using only data from inside it, an identical expanding-window procedure evaluates nine candidate specifications and selects the one with the highest inner skill. That specification is then refitted on the whole outer training set and applied to the outer test year. The candidates range from the cosmic

ray history alone through the full Model D, and include variants omitting the cosmic ray history, the polarity terms, or both. No information from the validation period enters the choice, so the resulting skill remains genuinely out of sample.

4. Results

The results are presented in the order in which the argument develops. Section 4.1 establishes the lag structure linking the heliospheric drivers to CRI and shows that it is polarity dependent, which motivates the interaction terms carried by every subsequent model. Section 4.2 then demonstrates that a conventional levels formulation, despite reproducing the observed record closely, fails against the persistence benchmark at every forecast horizon. Section 4.3 introduces the differenced formulation that corrects this failure, and Section 4.4 isolates the contribution of the recurrence index within it. Section 4.5 reports the pre-specified test of the transient channel. Sections 4.6 and 4.7 give the forward forecast and the conditional Cycle 26 projection respectively, Section 4.8 the rigidity calibration and the dependence of skill upon response rigidity, and Sections 4.9 and 4.10 the statistical assessment of the skill differences together with the ablation and nested cross-validation analyses.

Three conventions apply throughout. CRI is expressed as a percentage of the mean 1965 level, so that a skill score or a bias quoted in % refers to that normalisation. Every predictor is strictly causal and evaluated at the forecast issue time t - h, as set out in Section 3.3 and illustrated in Table 2a, and the polarity and recurrence indices are constructed on trailing windows so that no value postdates its own timestamp. All skill scores are computed by Equation (4) against the persistence benchmark under the expanding-window rolling-origin scheme of Section 3.4, and are point estimates; the uncertainty attaching to them is deferred to Section 4.9 and should be borne in mind when reading the intervening sections.

4.1 Lag structure and its polarity dependence

Cross-correlation of monthly CRI against each driver, Equation (17), over 1976 to 2025 yields a pooled optimum lag of 8 months for HCS tilt, with r = -0.704, and 10 months for sunspot number, with r = -0.862. The Moscow monitor reproduces the tilt result almost exactly, at lag 8 with r = -0.706, confirming that the relationship is not an artefact of a single station or cutoff rigidity.

Separating the record by polarity epoch demonstrates why the pooled figure understates the underlying relationship (Table 3). Within each epoch the correlation lies between 0.77 and 0.94 in absolute value, well above the pooled 0.70 in every case but one. The pooled value is degraded not because the relationship is weak but because the lag is not constant across epochs, so that a single shift cannot align the whole record. This is consistent with the variation in reported lags noted by Koldobskiy et al. (2022) and by Ross and Chaplin (2019), and it constitutes the empirical justification for the polarity interaction terms introduced in Section 3.3.

The fourteen-month lag obtained for the 2001 to 2012 epoch stands out, both for its length and for its comparatively low correlation of 0.77 against 0.87 to 0.94 elsewhere. That epoch spans the long, weak solar cycle 23 to 24 minimum, during which the cosmic ray maximum was unusually broad and flat; the correlation surface over such an interval is shallow, so the lag estimate is poorly

constrained. We report it for completeness but do not interpret it as evidence for a physically longer delay specific to that epoch.

Epoch	Polarity	Tilt lag (months)	r
1977-05 to 1980-09	A > 0	4	-0.874
1981-02 to 1990-10	A < 0	3	-0.941
1991-04 to 2001-02	A > 0	11	-0.928
2001-06 to 2012-11	A < 0	14	-0.768
2016-03 to 2024-04	A > 0	6	-0.930
Pooled, 1976 to 2025	all	8	-0.704

Table 3. Polarity-separated cross-correlation lags between HCS tilt (classic model) and Oulu CRI. Within-epoch correlations substantially exceed the pooled value.

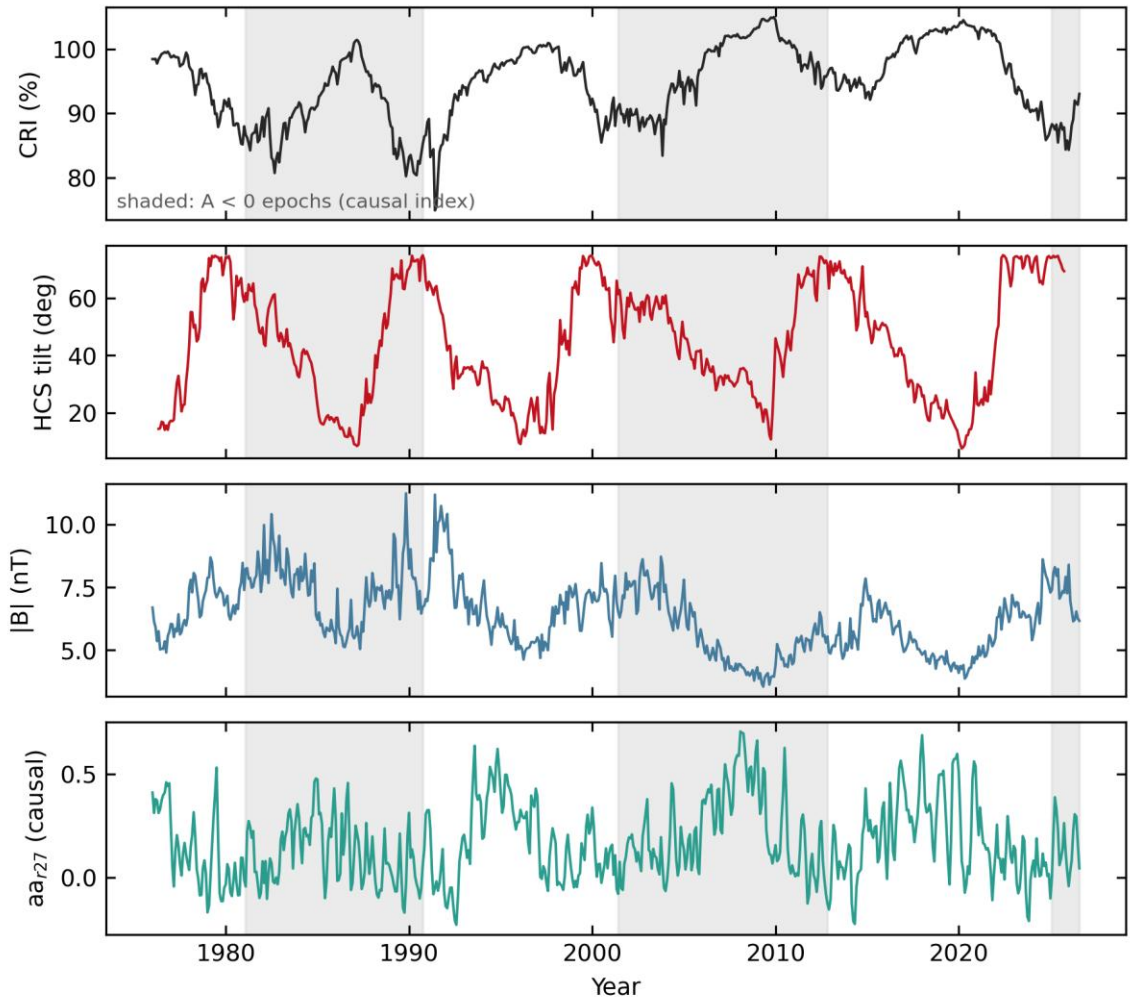


Figure 1. Monthly CRI, HCS tilt, interplanetary field magnitude and the causal aa recurrence index, 1976 to 2026. Shaded intervals mark A < 0 epochs as determined in Table 2.

Figure 1 shows the CRI together with the three principal heliospheric predictors over the full analysis period. The antiphase relation between intensity and both tilt and field magnitude is evident throughout, as is the alternation between the broad, flat intensity maxima of A > 0 epochs and the sharper maxima of A < 0. The recurrence index in the lowest panel varies on a visibly different timescale from the other three, which anticipates its low collinearity with them reported in Section 4.4.

4.2 A levels formulation and persistence

A regression on the level of CRI, fitted on 1976 to 2013 and tested on 2014 to 2025, reproduces the shape of the variation well but fails against persistence at every genuine forecast horizon (Table 4). At h = 12 months its RMSE is 4.38 % against 3.24 % for persistence, giving a skill score of -0.83. Comparison is made at matched horizons, with every predictor evaluated at the forecast issue time, so that the model and the persistence benchmark have access to the same information; a levels model using contemporaneous interplanetary measurements would not be a forecast and could not be compared with persistence on equal terms.

Horizon	Out-of-sample r	Mean bias (%)	Model RMSE (%)	Persistence RMSE (%)	Skill
12 months	0.909	-3.67	4.38	3.24	-0.83
18 months	0.825	-3.80	4.93	4.39	-0.26
24 months	0.618	-4.22	6.03	5.55	-0.18

Table 4. Levels-formulation regression evaluated at matched forecast horizons, trained on 1976-2013 and tested on 2014-2025. Skill against persistence is negative at every horizon.

Diagnosis shows the failure to be a level offset rather than a shape error. The mean bias is -4.34 %; removing it reduces the root-mean-square error to 3.33 %, close to persistence. Mean CRI over the training period is 94.1 % and over the test period 97.6 %. The model is being asked to extrapolate into a heliospheric state lying outside the range represented in its training data, characterised by weaker interplanetary fields and correspondingly higher cosmic ray intensities. Durbin-Watson statistics, Equation (11), between 0.35 and 0.55 confirm that the levels specification is misspecified in the presence of this non-stationarity.

We interpret this bias primarily as evidence of non-stationarity in the relationship between CRI and heliospheric conditions, rather than as a direct quantitative measure of heliospheric weakening. The training period has a lower mean CRI (94.1 %) than the test period (97.6 %), while the latter also corresponds to weaker Interplanetary Magnetic Fields (IMF), so a model calibrated on the earlier state systematically underestimates the later level. The offset of approximately 4 % is

therefore a measure of the extrapolation error incurred under a changed heliospheric regime, not an independent estimate of the magnitude of that change.

4.3 The differenced formulation and point-estimation skill (Twelve Months)

Reformulating the target as the change in CRI over the horizon, Equation (1), and evaluating skill by Equation (4) under expanding-window rolling origin, produces 360 genuinely out-of-sample forecasts at each horizon (Table 4a). Skill is negative at six months: over so short an interval the intensity changes little, persistence is close to optimal, and the added predictors introduce more estimation noise than they remove. Skill turns positive at twelve months, rises to a maximum at twenty-four months, and the point-estimate improvement over persistence there reaches 36 %.

This horizon structure follows directly from the lag physics set out in Section 4.1. A forecast model can add value only once sufficient time has elapsed that the present intensity ceases to be the best available estimate of the future intensity. Below that threshold the autocorrelation of the cosmic ray record dominates; above it, the lagged solar and heliospheric information carries the greater share of the predictable variation.

Horizon	Persistence RMSE (%)	Model C RMSE (%)	Model D RMSE (%)	Skill C	Skill D
6 months	2.22	2.49	2.50	-0.26	-0.26
12 months	3.30	3.10	3.10	+0.12	+0.12
18 months	4.26	3.63	3.51	+0.28	+0.32
24 months	5.22	4.34	4.17	+0.31	+0.36

Table 4a. Rolling-origin forecast skill against persistence. Model C comprises the drift and diffusion predictors; Model D adds the aa recurrence index. 360 out-of-sample forecasts per horizon. All predictors are strictly causal.

The arithmetic of Equation (4) may be followed directly from the table. At h = 24 the 360 forecasts give a persistence mean-square error of 27.283 and a Model D mean-square error of 17.369, both in units of squared percentage points, so that S(24) = 1 - 17.369 / 27.283 = 0.363, equivalent to a fall in RMSE from 5.223 to 4.168 %. The corresponding Model C mean-square error of 18.865 gives S = 0.309, so the recurrence index accounts for a skill difference of 0.363 - 0.309 = 0.055 at that horizon. The behaviour of the twelve-month forecast across the validation period is shown in Figure 2.

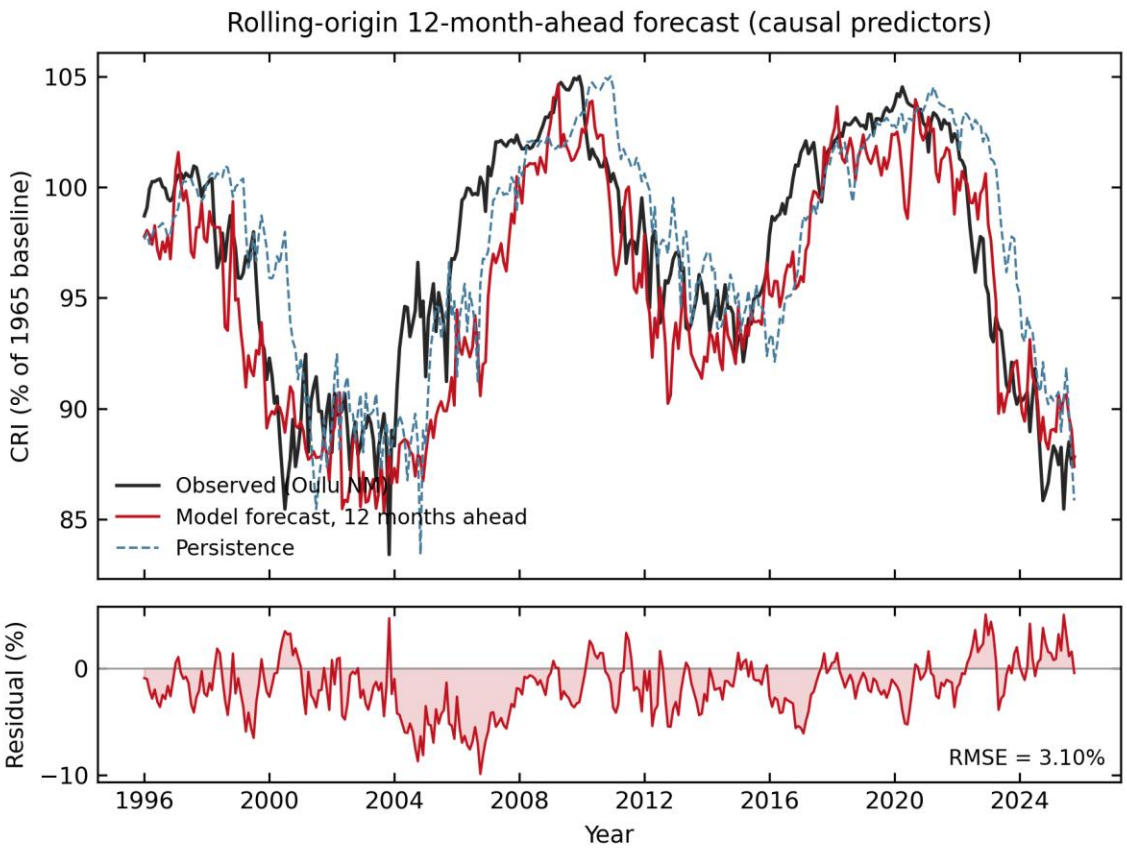


Figure 2. Rolling-origin twelve-month-ahead forecast against observed Oulu CRI, with residuals in the lower panel. Each forecast uses only data available at the time it was issued.

The forecast tracks the observed intensity through both the 2009 and the 2019 minima and through the intervening maximum, with the residual panel showing no systematic drift across the period. The largest residuals occur during the rapid transitions of the declining phases, where the intensity changes fastest and the twelve-month lead is least adequate.

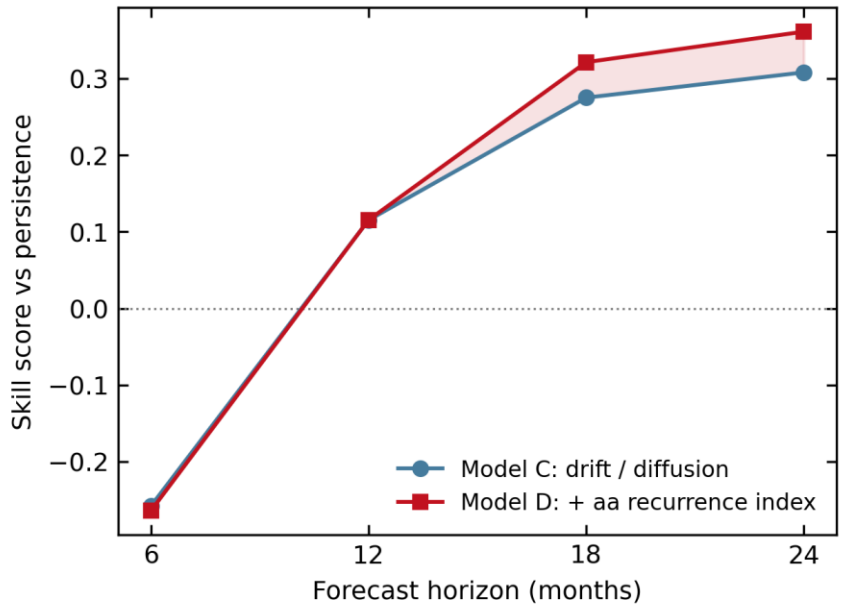


Figure 3. Skill score against persistence as a function of forecast horizon. Shading marks the contribution attributable to the recurrence index.

Figure 3 summarises the horizon dependence of Table 4. Skill is negative at six months, crosses zero between six and twelve, and rises thereafter; the shaded band between the two curves, negligible below twelve months and widening beyond eighteen, is the contribution of the recurrence index discussed in Section 4.4.

Adoption of the continuous polarity index in place of the conventional categorical treatment increases the number of usable forecasts from 306 to 360, a gain of eighteen %, by retaining the reversal epochs that the conventional treatment discards. This comes with a trade-off rather than an unambiguous gain in accuracy: point skill at h = 24 is 0.363 for the continuous index against 0.469 for the categorical treatment restricted to non-reversal months, so the additional forecasts are somewhat noisier on average, as would be expected since they include the physically more difficult reversal transitions. We regard the larger, gap-free forecast set as the more useful basis for an operational forecast, since a model that cannot be run during reversal years is of limited practical value, but we report both figures rather than presenting the continuous index as a strict improvement.

At h = 12 the strongest terms in the continuous-index specification are the rate of change of tilt, the solar wind speed, the interaction of polarity with field magnitude, and the cosmic ray level, all significant at conventional levels under the HAC covariance estimator. That the rate of change of tilt contributes alongside its level is physically informative. The approach to maximum, and not only the instantaneous current sheet configuration, carries information about modulation over the following year. The negative coefficient on the cosmic ray level represents mean reversion, an already-high intensity having less room to rise further.

4.4 The recurrence index contribution to different horizons

The standardised coefficient of the recurrence index rises monotonically with lead time, from a value indistinguishable from zero at six months to 1.20 at twenty-four months (Table 5, Figure 4). Its variance inflation factor, Equation (9), is approximately 1.5 at every horizon, so it is not acting as a proxy for another predictor already in the model. At twelve months the coefficient is already significant ($p = 0.004$), yet Table 4 shows essentially no difference in aggregate skill between Model C and Model D at that horizon, 0.114 against 0.114. This is not a contradiction: a term can carry statistically reliable information in a multivariate fit without moving an aggregate out-of-sample error metric that several other terms also contribute to, particularly when its own coefficient is still small relative to the dominant predictors at that horizon. The recurrence index becomes more influential in the aggregate point-estimate skill from eighteen months onward, when its coefficient is larger and the difference between Models C and D becomes measurable in the error metric. The corresponding skill gains at eighteen and twenty-four months are +0.046 and +0.055; Section 4.9 shows that these gains, although consistent across several complementary diagnostics, do not attain significance on the present sample.

Horizon	Standardised coefficient	HAC standard error	t statistic	p value
6 months	+0.09	0.20	0.45	0.65
12 months	+0.55	0.19	2.89	0.004
18 months	+0.86	0.22	3.88	0.0001
24 months	+1.20	0.25	4.91	<0.0001

Table 5. Coefficient of the aa recurrence index in Model D as a function of forecast horizon. The standardised coefficient is that of Equation (6), expressed in units of one training-period standard deviation of the predictor. HAC denotes the heteroskedasticity- and autocorrelation-consistent standard error of Newey and West (1987), computed with a truncation lag of 12 months, as given by Equation (7). The statistic t is the ratio of coefficient to standard error, Equation (8), and p is the corresponding two-sided probability under the null hypothesis of a zero coefficient.

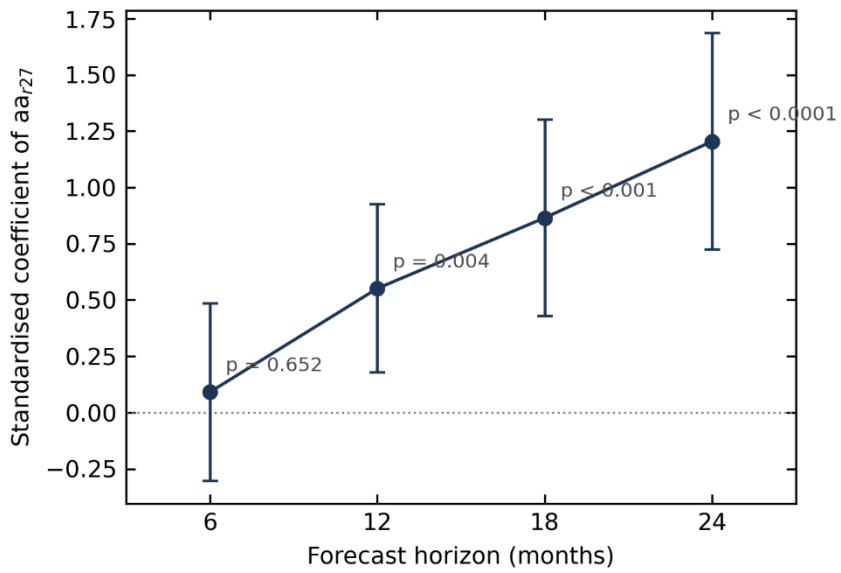


Figure 4. Standardised coefficient of the recurrence index against forecast horizon, with 95 % confidence intervals. The monotonic rise is consistent with a physically interpretable contribution from recurrent solar wind organisation, although it does not by itself establish incremental forecast skill.

The natural interpretation is that at short horizons the contemporaneous interplanetary field magnitude, which remains available at (t - h) when h is small, absorbs the relevant information. As the horizon lengthens, the field magnitude measured at the issue time becomes progressively less informative about conditions at the target time, because the field at 1 AU decorrelates within a few solar rotations. The degree of corotating organisation decorrelates more slowly, since the coronal holes that produce recurrent streams can persist for many rotations. It is therefore the quantity that retains useful information at the horizons where the recurrence index contributes.

The contribution is also polarity dependent: the correlation between the recurrence index and CRI is 0.63 in $A < 0$ epochs against 0.47 in $A > 0$. Drift theory offers a reading of this, although the present analysis does not test it. In $A < 0$ epochs positively charged particles reach the inner heliosphere predominantly along the heliospheric current sheet, so their access path samples the corotating structure that the recurrence index measures; in $A > 0$ they arrive preferentially over the poles, where that structure is weaker. A greater sensitivity to recurrence in $A < 0$ is what this picture would predict.

4.5 Transient injection and forecast skill

Six candidate variables were constructed from the near-Earth ICME list of Richardson and Cane (2010) and tested under the criterion of Section 3.5 (Table 6). All six failed outright at h = 18, with skill gains ranging from -0.045 to +0.014 and every p-value above 0.06; none approaches the required gain of 0.02 combined with p below 0.05. This outcome is cleaner than an earlier screen

conducted during the development of the analysis, and the difference is instructive. That screen used the centred smoothing windows described in Section 3.1, before the look-ahead they introduce had been identified. Under those non-causal constructions one candidate, the twelve-month mean of ICME hours, passed both the skill-gain and the significance thresholds, and was eliminated only by the two follow-up requirements of horizon coherence and replication at Moscow. Once the predictors were rebuilt on trailing windows, the same candidate failed the initial criterion outright. Removing the look-ahead therefore removed an apparent transient signal, which is a further indication that the look-ahead was material rather than incidental. Under the causal constructions adopted throughout this study, that candidate does not pass even the initial screen.

It is worth recording that the variable expected a priori to succeed also failed. The residual of the ICME rate relative to the sunspot-predicted rate is orthogonal to sunspot number by construction, with r = 0.02, and captures the documented difference in CME production per unit sunspot number between Cycles 23 and 24. It carried no information about modulation, with p = 0.58.

Candidate variable	Skill gain at h = 18	p	VIF	Outcome
ICME count	-0.023	0.80	1.5	reject
ICME hours	-0.028	0.46	1.6	reject
Sum of B times duration	-0.030	0.55	1.8	reject
Sum of BV times duration, 12-month mean	+0.007	0.24	4.5	reject
ICME hours, 12-month mean	+0.003	0.067	3.8	reject
Rate residual over SSN, 12-month mean	-0.043	0.58	1.4	reject
CME sum of v squared (CDAW)	-0.066	0.06	17.7	reject

Table 6. Pre-specified test of the transient channel. Acceptance required a skill gain of at least 0.02, p below 0.05 and VIF below 5, together with coherent horizon dependence and replication on the Moscow monitor. The CME row is evaluated at h = 18 for consistency with the ICME candidates; see text for its dependence on horizon.

The corresponding test on global CME injection could not be carried out in the form specified. The pre-registered variable was the twelve-month integral of summed CME kinetic energy, chosen because the formation of global merged interaction regions in the outer heliosphere responds to accumulated injected energy rather than to event counts. Mass and kinetic energy are, however, absent from the CDAW catalogue for 2020 to 2023 and for 2025 to 2026, and where present the covered fraction varies between 0.30 and 0.84 from year to year. Any sum of kinetic energy therefore carries a trend set by processing coverage rather than by physics. We note additionally that catalogue masses and kinetic energies are themselves lower limits, being subject to plane-of-

sky projection, with kinetic energies underestimated by up to a factor of eight according to Vourlidas et al. (2010).

A mass-free analogue was tested in its place: the duty-cycle-normalised sum of the square of linear speed over CMEs of angular width at least 30 degrees, averaged over twelve months. Skill gain is negligible at twelve months and negative at eighteen and twenty-four months, by 0.066 and 0.096 respectively, with variance inflation factors between 17 and 18 throughout. These reflect strong correlation with sunspot number, tilt and field magnitude. The variable is a noisy restatement of predictors already present, and its inclusion degrades the model at every horizon where it has any detectable effect.

A subsidiary finding concerns the homogeneity of the CDAW catalogue itself (Table 7). Annual event counts do not track solar activity monotonically: 2018 records 476 events against 647 in 2020 despite lower sunspot number, and 2008 records 863 against 476 in 2018 despite being a deeper minimum. Restriction to widths of at least 30 degrees reduces but does not remove the drift, which reflects changes in image cadence, operator practice and the version 2 reprocessing. We note this because raw CDAW counts are sometimes used as a homogeneous solar activity proxy in long-term studies, and Table 7 indicates that this is not safe without correction.

Year	All CMEs	Width at least 30 degrees
1997	888	505
2000	2024	1236
2003	1267	796
2008	863	163
2012	2178	1365
2014	2478	1634
2018	476	177
2020	647	289
2022	1539	1033
2024	2153	1641

Table 7. Annual CDAW CME counts for ten representative years spanning Cycles 23, 24 and 25. The catalogue is continuous from 1996 to 2026; the years shown are selected to contrast comparable phases of different cycles. The non-monotonic relation to solar activity indicates drift in detection sensitivity across the catalogue.

4.6 Forward forecast

The last complete predictor set is October 2025 for specifications including HCS tilt and August 2026 for tilt-free specifications; the observed cosmic ray record ends in August 2026 at 93.0 %. Both formulations indicate continued recovery as Cycle 25 declines (Table 8, Figure 5).

Within the first year the tilt-free specification is to be preferred notwithstanding its lower nominal skill, because the tilt-bearing model must operate at a horizon of at least eleven months owing to the truncated tilt series, and a weaker model at short horizon outperforms a stronger model forced to extrapolate. Beyond approximately fifteen months the ranking reverses. This is a practical consequence of data latency rather than of model quality, and it would be removed by a current tilt series. No predictor is extrapolated at any point. For each future target month the horizon is set to the smallest value of h for which every predictor at t - h falls within the observed record, so the tilt-bearing model is simply unavailable at short horizons rather than being driven by assumed tilt values. Table 8 reports a target as beyond reach, the required predictor vector does not exist.

Table 8 gives the forecast at four representative target months. The two columns correspond to the two specifications of Section 4.3, differing only in whether the HCS tilt terms are included. Both are issued from observed data alone; neither extrapolates any predictor. For each target month the horizon is set to the smallest value of h for which every predictor at t − h lies within the observed record, so the two columns operate at different horizons for the same target, and the tilt-bearing model is unavailable once the required predictor vector no longer exists.

Target month	Tilt-free model (%)	Model with HCS tilt (%)
2026-12	92.9 (91.0 to 94.8)	91.8 (88.7 to 95.0)
2027-06	93.3 (90.4 to 96.2)	95.1 (91.5 to 98.7)
2027-12	94.7 (91.2 to 98.2)	97.3 (93.1 to 101.6)
2028-06	95.7 (91.4 to 100.0)	beyond reach

Table 8. Forward forecast of Oulu CRI with plus or minus one standard deviation intervals derived from the rolling-origin RMSE at the relevant horizon.

The last observed value is 93.0 per cent in August 2026. Both specifications indicate continued recovery, reaching 95 to 97 per cent by the end of 2027, consistent with the declining phase of Cycle 25 and the approach to the minimum expected near 2031. The intervals widen with target month, from plus or minus 1.9 per cent in December 2026 to plus or minus 4.3 per cent in June 2028 for the tilt-free specification, because a more distant target requires a longer horizon and the rolling-origin RMSE grows accordingly.

The two specifications diverge in the near term, giving 92.9 and 91.8 per cent for December 2026, and converge thereafter. This reflects the horizons at which each operates rather than a disagreement about the physics: the tilt-bearing model must work at h = 14 for that target because

its last complete predictor vector is October 2025, whereas the tilt-free model works at h = 4. The intervals overlap substantially at every common target month, so the two are not in conflict.

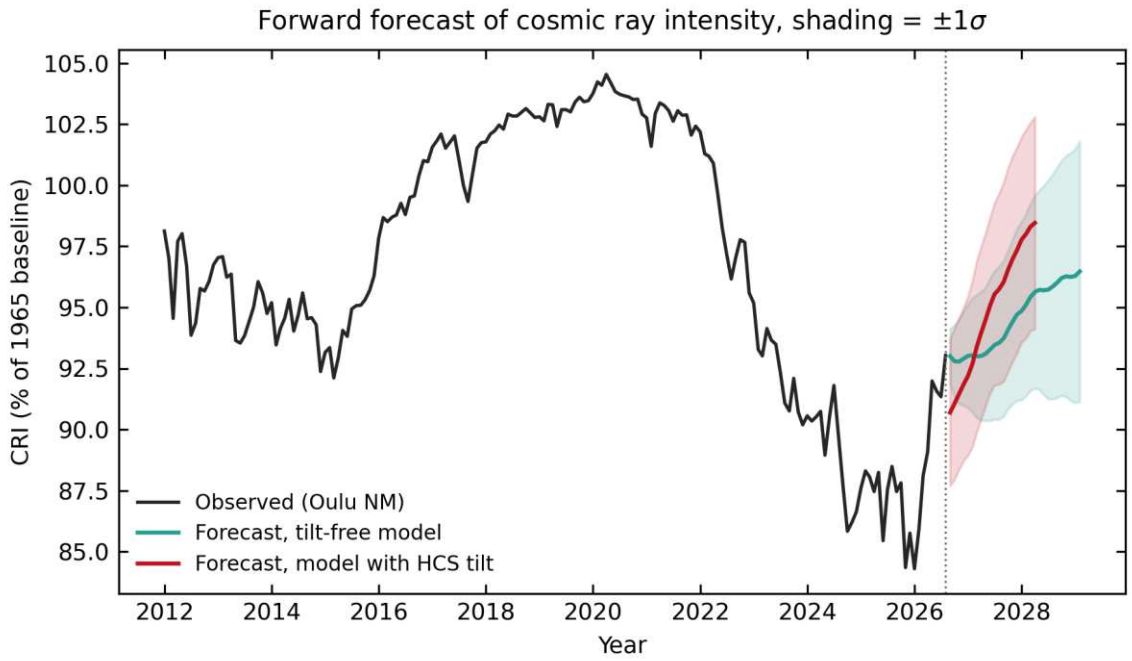


Figure 5. Forward forecast of Oulu cosmic ray intensity from September 2026, extending to April 2028 for the tilt-bearing specification and to February 2029 for the tilt-free specification. Shading indicates $\pm 1\sigma$.

Figure 5 places the forecast against the observed record since 2012. The recovery from the 2024 to 2025 minimum in intensity is already under way in the observations, and both forecast curves continue it at a comparable rate. The widening of the shaded bands with time is the dominant visual feature and is the point of the figure: by early 2029 the tilt-free interval spans roughly nine percentage points, which is a substantial fraction of the full range of intensity variation over a solar cycle. The figure should therefore be read as indicating the direction and approximate rate of the recovery rather than a specific level at a specific date.

4.7 Conditional projection for Cycle 26

Cycle 26 lies beyond the reach of the forecast model, whose predictors must be observed rather than assumed. Extending the forecast to that horizon would require forecasting the predictors themselves, compounding the uncertainty of a solar cycle prediction with that of the modulation model. We therefore present a conditional projection using a reduced-form specification driven by smoothed sunspot number at ten-month lag together with polarity, which attains an adjusted

coefficient of determination of 0.858 with residual standard deviation 2.27 % and leave-one-cycle-out root-mean-square errors between 2.18 % and 3.80 %.

The polar field precursor relation obtained from the present data is a linear dependence of cycle maximum sunspot number on the absolute polar field strength at the preceding minimum, with slope 1.280 and intercept 69.1, giving r = 0.842 on five points. With so few points the relation is indicative rather than predictive. More importantly, it cannot yet be applied to Cycle 26: the relevant minimum falls around 2030 and the polar field, at 24 microtesla in 2026, is still rebuilding following the 2024 reversal. The three amplitude scenarios of Table 9 are therefore imposed rather than derived.

Scenario	Projected minimum CRI (%)	Epoch of minimum
Weak, SSN max 80	93.7	2035-05
Moderate, SSN max 130	89.3	2035-04
Strong, SSN max 180	85.9	2035-03

Table 9. Conditional Cycle 26 projections. All three scenarios reach 101 to 102 % near the 2031 minimum.

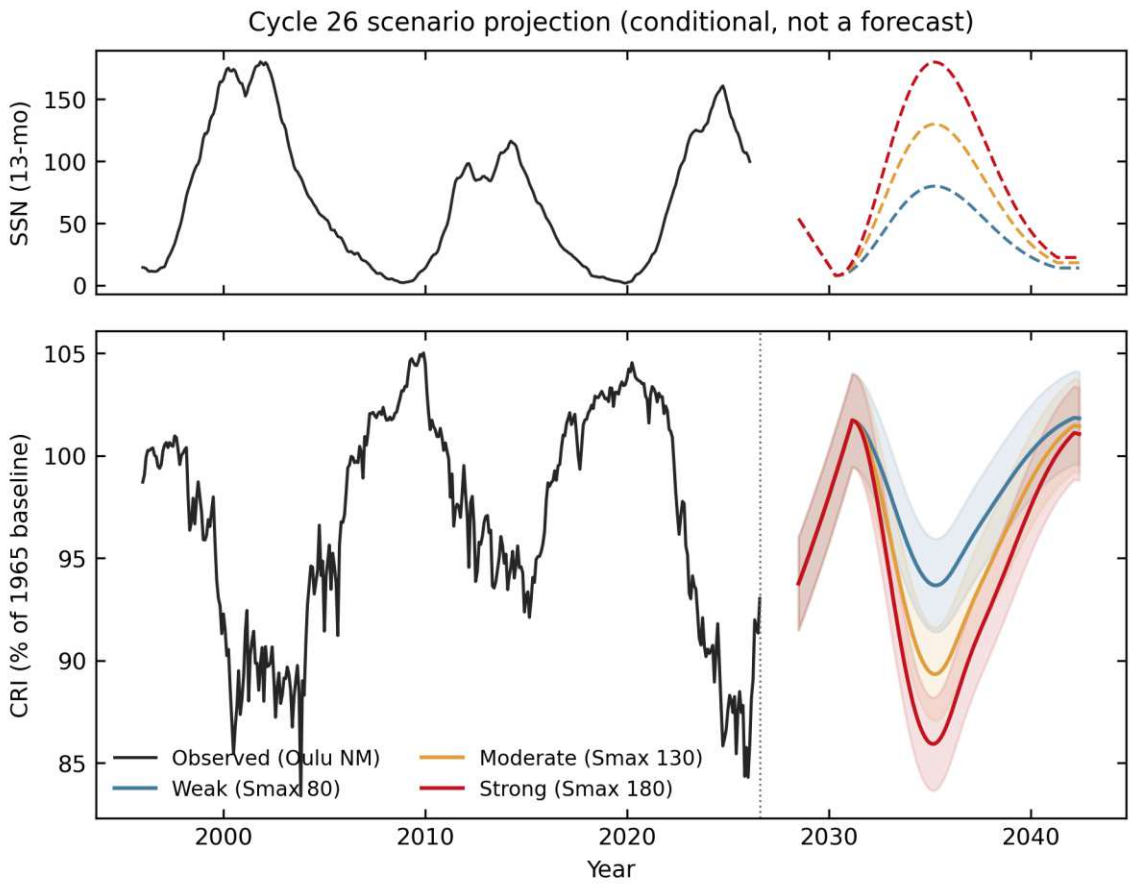


Figure 6. Conditional Cycle 26 projection. Upper panel, imposed sunspot number scenarios; lower panel, resulting CRI trajectories with plus or minus one standard deviation of the reduced-form model.

The three trajectories are shown in Figure 6, together with the sunspot number scenarios that generate them. These trajectories are indistinguishable through the approach to the 2031 minimum, where intensity is governed by the decline of Cycle 25 rather than by the amplitude of Cycle 26, and separate only as the new cycle rises. The uncertainty bands overlap throughout the rising phase and are well separated only near Cycle 26 maximum.

The spread at Cycle 26 maximum is 7.8 percentage points, several times the residual scatter of the reduced-form model. Projection uncertainty at that horizon is therefore dominated by the unknown cycle amplitude rather than by the cosmic ray model, which indicates where effort is best directed. Four assumptions require explicit statement: a minimum in June 2030, an eleven-year cycle length, an idealised asymmetric cycle profile, and a polarity reversal imposed at Cycle 26 maximum.

4.8 Forecast skill declines with rigidity

Of twenty-two station and detector combinations examined, seven passed the AMS-02 screen of Section 3.6 (Table 10). Four failures are instructive. Jungfraujoch shows a 42 % step between 2012

and 2019 with isolated outliers; Rome declines while CRI rises toward the 2019 minimum; and Mexico City passes when taken from NMDB but fails when taken from the archived fixed-width series for the same instrument and the same years, a concrete illustration of the source dependence documented by Väisänen et al. (2021). The fourth, Mt Hermon, is the most instructive of all. Its annual means track the solar cycle correctly and it correlates with Oulu at r = 0.83 over 1999-2018, yet its residual scatter about that relation is 1.54 % at monthly resolution, against an expected Poisson contribution of 0.005 % for a monitor counting 11 000 per minute. The instrumental scatter is therefore comparable to the 1.5 to 1.9 % modulation amplitude expected at its rigidity, and no filtering recovers the signal. Athens fails in the same way. This is the characteristic difficulty of high-cutoff monitors: the modulation signal is small, so stability rather than counting statistics sets the limit, and only large, well-maintained instruments such as the 18NM64 at Tsumeb succeed at these rigidities.

Station	**Rc (GV)**	**Amplitude (%)**	**max r vs AMS-02**	**R50 (GV)**	**Screen**
McMurdo NM64	0.01	2.54	0.985	12.6	accept
Thule	0.30	3.48	0.972	10.1	accept
Oulu	0.81	3.38	0.995	10.3	accept
Moscow	2.43	3.22	0.989	10.6	accept
Jungfraujoch IGY	4.49	25.5	0.496	-	reject
Rome	6.27	8.04	0.443	-	reject
Mexico City (NMDB)	8.20	1.50	0.960	18.9	accept
Athens	8.53	0.95	0.741	-	reject
Tsumeb (NMDB)	9.29	1.24	0.970	22.7	accept
Mt Hermon (ESOI)	10.80	1.91	0.529	-	reject
Doi Inthanon	16.80	0.65	0.941	72.1	accept

Table 10. Rigidity response calibration and quality screen against AMS-02 daily proton fluxes over 2011-2019. Seven of twenty-two station and detector combinations examined were accepted; the four informative rejections are listed. R50 is the median response rigidity defined in Section 3.6. The calibration is displayed in Figure 7.

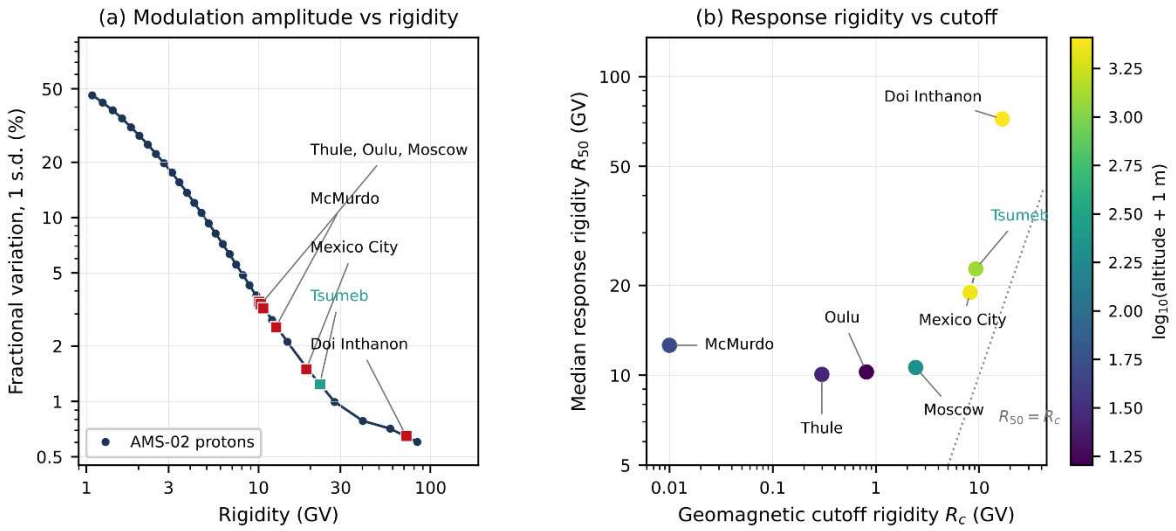


Figure 7. (a) AMS-02 fractional modulation amplitude as a function of rigidity, with each of the seven accepted stations placed at the rigidity matching its own amplitude; Tsumeb, which fills the former gap between 19 and 72 GV, is marked in green. Thule, Oulu and Moscow coincide to within five per cent in response rigidity and are therefore drawn as a single labelled group. (b) Median response rigidity against geomagnetic cutoff rigidity, coloured by altitude. The dashed connector links Mexico City and Tsumeb, whose near-equal cutoffs but differing altitudes separate their response rigidities.

The left panel shows the steep fall in modulation amplitude with rigidity against which each station is placed, and the clustering of the three low-cutoff monitors at a common amplitude. The right panel shows that every station lies well above the line R50 = Rc, by factors between two and thirty-four, and that the departure is largest for the lowest cutoffs.

Two features of Table 10 are notable before any skill comparison. First, the median response rigidity of the low-cutoff stations is approximately 10 GV irrespective of cutoff: Thule at 0.30 GV, Oulu at 0.81 GV and Moscow at 2.43 GV differ in cutoff by a factor of eight but in response rigidity by five %. For modulation purposes these are effectively the same detector, and they are treated below as replicate measurements at a single rigidity rather than as three independent points. Second, response rigidity exceeds cutoff by factors between two and thirty-four, confirming that cutoff alone is not an adequate abscissa for cross-station comparison. Mexico City and Tsumeb make the point directly: their cutoffs differ by only 13 %, at 8.20 and 9.29 GV, but their altitudes differ by a factor of 1.8, at 2274 and 1240 m, and their response rigidities differ by a factor of 1.2, at 18.9 and 22.7 GV, the lower-altitude station giving the harder response as the atmospheric yield function requires. Atmospheric depth, not cutoff alone, governs the response.

Forecast skill was then evaluated exactly as in Section 4.3, separately for each station, over two matched windows. Matching is essential, because skill depends strongly on which part of the solar cycle the validation covers; comparing stations over different periods measures the period rather

than the station. The calibration and the skill comparison necessarily use different station sets. Every accepted station with a response rigidity between 19 and 72 GV has a record ending before 2020: Tsumeb is usable only to November 2019 and Mexico City in the archive form to 2016. Extending the skill comparison into that rigidity range would therefore require truncating every station to 2019, and because the skill is concentrated in the period after 2020 such truncation collapses the skill for all stations, including Oulu, from 0.26 to -0.02 at h = 18. The skill comparison is consequently restricted to the three rigidities for which records extend to 2025, while the calibration uses all seven accepted stations. Set A trains from 1990 and validates over 2003-2025 with 274 forecasts per horizon. Set B trains from 2009 and validates over 2017-2025 with 106 forecasts, and serves only to extend the rigidity range to 72 GV.

Set	**Station**	**R50 (GV)**	**Skill h=12**	**h=18**	**h=24**	**aa r27 gain, h=24**
A	Thule	10.1	+0.03	+0.35	+0.41	+0.03
A	Oulu	10.3	+0.06	+0.26	+0.39	+0.03
A	Moscow	10.6	-0.09	+0.17	+0.30	+0.01
A	Mexico City	18.9	-0.00	+0.10	+0.10	+0.06
B	Oulu	10.3	-0.05	+0.50	+0.17	+0.02
B	Moscow	10.6	+0.14	+0.54	+0.29	+0.01
B	Mexico City	18.9	-0.50	+0.22	+0.28	+0.02
B	Doi Inthanon	72.1	-0.21	+0.07	+0.03	-0.01

Table 11. Rolling-origin forecast skill against persistence by station over matched validation windows. Set A: 274 forecasts per horizon, 2003-2025. Set B: 106 forecasts, 2017-2025. Figure 8 displays these results against response rigidity.

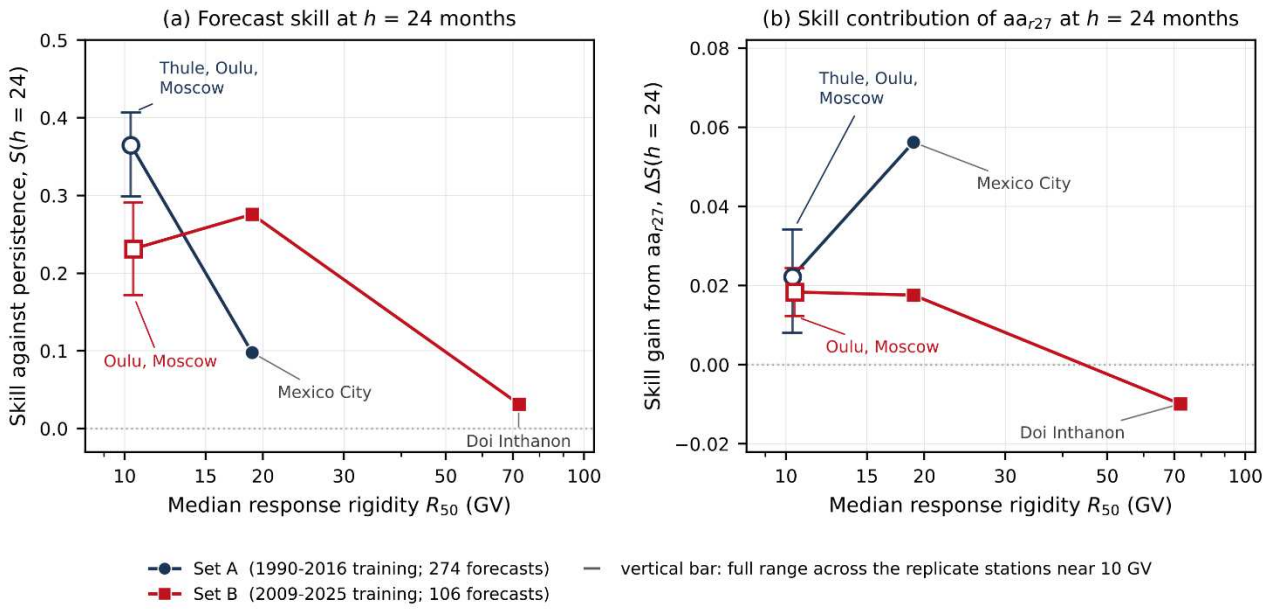


Figure 8. (a) Rolling-origin forecast skill against persistence at h = 24 months as a function of median response rigidity. (b) The skill contribution of the recurrence index aa_r27 at the same horizon. Thule, Oulu and Moscow are degenerate in response rigidity and are plotted at their group mean, the vertical bar giving the full range across those stations. Skill in (a) falls monotonically with rigidity in Set A and collapses to 0.03 at 72.1 GV in Set B. The recurrence contribution in (b) is small and positive at every rigidity in Set A, and in Set B falls from positive near 10 GV to slightly negative at 72.1 GV.

The two sets agree in slope but not in level, the shorter Set B lying above Set A at the lower rigidities, which reflects the different validation periods rather than any property of the stations. The three points near 10 GV should be read as replicate measurements at a single rigidity, so the trend rests on the separation between that cluster, Mexico City and Doi Inthanon.

Skill shows a broadly declining trend with response rigidity. In Set A at h = 24 months it falls monotonically from 0.41 at 10.1 GV through 0.39 and 0.30 to 0.10 at 18.9 GV, and the same ordering holds at h = 18. Set B extends the range, giving 0.03 at 72.1 GV, where the model retains essentially no advantage over assuming no change. In Set B the fall is not monotonic: skill is comparable at 10 and 19 GV, at 0.23 and 0.28, and declines only at 72.1 GV, so the Set B evidence rests on the high-rigidity end rather than on a gradient across the whole range. These results provide evidence for a decline in forecastability with effective response rigidity rather than a precisely determined relationship.

The spread among the three degenerate stations near 10 GV provides an estimate of station-to-station measurement noise, and its size relative to the rigidity gradient depends on the horizon. At h = 24 the three stations span 0.30 to 0.41, a width of 0.11, against a fall of 0.27 between 10 and 19 GV; the decline is therefore about two and a half times the replicate spread and is reasonably well resolved. At h = 18 the stations span 0.16 to 0.35, a width of 0.19, against a fall of 0.16 over

the same interval; at that horizon the two are comparable and the decline is suggestive rather than decisive. In both cases it is the extension to 72 GV, where skill falls essentially to zero, that carries the weight of the evidence. Because no accepted station lies between 19 and 72 GV, the behaviour across that interval is unsampled. Set B rests on only 106 forecasts and the Mexico City skill differs markedly between the two sets, so Set A is taken as primary and Set B as evidence that the decline continues rather than as a precise measurement.

4.9 Statistical assessment of the skill differences

The skill scores reported above are point estimates. Because the differenced targets overlap in time, the 360 monthly forecasts at each horizon are not 360 independent observations; the validation span comprises approximately thirty annual forecast blocks, treated as dependence-preserving units rather than as strictly independent observations. Table 12 reports moving-block bootstrap confidence intervals (Equation 12), and Diebold-Mariano tests against persistence, Equations (13) to (16).

Horizon	Model	Skill	95 % bootstrap interval	DM*	p	Minimum detectable skill
6 months	C	-0.260	-0.730 to +0.129	+1.55	0.123	0.429
6 months	D	-0.267	-0.732 to +0.144	+1.56	0.120	0.438
12 months	C	+0.114	-0.448 to +0.488	-0.52	0.606	0.468
12 months	D	+0.114	-0.451 to +0.481	-0.52	0.604	0.466
18 months	C	+0.268	-0.217 to +0.604	-1.11	0.269	0.411
18 months	D	+0.315	-0.173 to +0.630	-1.32	0.188	0.401
24 months	C	+0.309	-0.227 to +0.620	-1.13	0.258	0.424
24 months	D	+0.363	-0.130 to +0.626	-1.41	0.160	0.378

Table 12. Moving-block bootstrap intervals (block length 24 months, 4000 replicates) and Harvey-Leybourne-Newbold-corrected Diebold-Mariano statistics against persistence; a negative DM* favours the model. The final column gives the smallest skill score that would attain significance at 95 % given the bootstrap dispersion.

No skill score attains significance at 95 %. Every bootstrap interval includes zero, and the Diebold-Mariano p-values lie between 0.12 and 0.61. A direct comparison of Model D against Model C gives p = 0.11 at h = 18 and p = 0.19 at h = 24, so the incremental contribution of the recurrence index is likewise not significant as a forecast-error comparison.

The construction may be followed through for one case. At h = 24 for Model D the mean loss differential of Equation (13) is $\bar{d} = -9.914$, in units of the fourth power of percentage points

because it is a difference of squared errors, and the long-run variance of Equation (15) with truncation K = 23 gives a standard error of 6.580. The ratio in Equation (16) is therefore -9.914 / 6.580 = -1.507, which the Harvey-Leybourne-Newbold factor of 0.935 reduces to DM* = -1.408, corresponding to p = 0.160 on 359 degrees of freedom. The negative sign indicates that the model has the smaller mean-square error, but not by a margin the test can distinguish from zero.

The final column of Table 12 places these results in context. For the present validation span and the adopted moving-block bootstrap, the estimated 95% detectability threshold lies between approximately 0.38 and 0.47 depending on horizon. These thresholds follow from the bootstrap dispersion: at h = 24 the interval spans -0.130 to +0.626, a width of 0.756, which treated as ±1.96 standard errors implies a standard error of 0.756 / 3.92 = 0.193 and hence a smallest distinguishable skill of 1.96 x 0.193 = 0.378. The observed value at h = 24 for Model D is 0.363, just below that threshold. The results therefore do not provide sufficient evidence to reject the persistence benchmark, although the long-horizon point estimates lie close to the estimated threshold rather than near zero. These thresholds are properties of this particular validation design and bootstrap configuration rather than universal detectability limits, but they indicate that a record substantially longer than the present one would be required to establish a forecast improvement of this magnitude at conventional confidence.

Four diagnostics nevertheless point in the same direction, and they are not repetitions of a single test. The skill is negative at six months and rises monotonically to twenty-four; the standardised recurrence coefficient rises monotonically with lead time and is significant under HAC inference from twelve months (Table 5). Removing the recurrence group from Model D reduces skill at eighteen and twenty-four months but not at twelve (Table 13); and forecast skill shows a broadly declining pattern with response rigidity across the stations (Table 11). These convergent patterns are mutually consistent with a contribution from recurrent structure and motivate the physical interpretation advanced in Section 5.2, but they do not establish a statistically significant incremental forecast contribution on the present sample. None of the four is a formal test of incremental predictive accuracy, and all four rest on the same cosmic ray record and overlapping validation periods, so they are complementary rather than statistically independent.

4.10 Ablation analysis

Table 13 reports a nested build-up of the predictor set and a leave-one-component-out decomposition of Model D. Both are evaluated under the same rolling-origin protocol as Section 4.3.

Nested model	h = 12	h = 18	h = 24
M0 persistence	0.000	0.000	0.000
M1 CRI history only	-0.027	+0.009	0.000
M2 + HCS tilt terms	+0.170	+0.233	+0.226

Nested model	h = 12	h = 18	h = 24
M3 + field and wind	+0.307	+0.378	+0.368
M4 + polarity (= Model C)	+0.114	+0.268	+0.309
M5 + recurrence (= Model D)	+0.114	+0.315	+0.363

Table 13a. Nested build-up of the predictor set, skill against persistence, Equation (4), evaluated under the rolling-origin protocol of Section 4.3. Each specification retains every predictor of the one above it and adds a further group. M0 is the persistence benchmark itself, which forecasts no change and therefore scores zero by construction. M1 adds the cosmic ray history, namely the intensity at t - h and its rate of change over the preceding 12 months. M2 adds the HCS tilt terms: the tilt, its square and its 12-month rate of change. M3 adds the interplanetary field magnitude with its 12-month rate of change, together with the solar wind speed. M4 adds the continuous polarity index and its interactions with tilt and with field magnitude, and is identical to Model C. M5 adds the aa recurrence index and is identical to Model D. All predictors are strictly causal.

Component removed from Model D	h = 12	h = 18	h = 24
CRI history	+0.233	+0.217	+0.173
HCS tilt terms	-0.023	-0.083	-0.140
Interplanetary field	-0.117	-0.056	-0.018
Solar wind speed	-0.169	-0.106	-0.089
Polarity terms	+0.195	+0.109	+0.052
Recurrence index	0.000	-0.046	-0.055

Table 13b. Change in skill when each predictor group is removed from Model D. A positive entry means that removing the group improves the forecast. Each interaction term is removed with its parent group, so the tilt row also removes the polarity-tilt interaction and the field row the polarity-field interaction; those interactions are removed again in the polarity row, and the entries are therefore not additive. All specifications are evaluated on the common sample of months for which every Model D predictor is available.

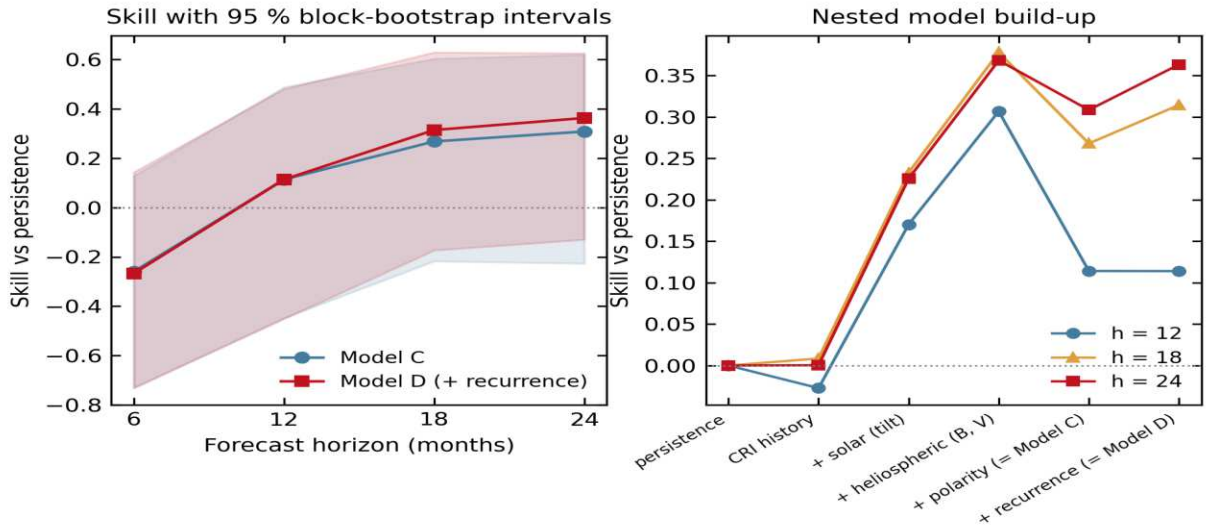


Figure 9. Left: skill against forecast horizon with 95 % moving-block bootstrap intervals. Right: nested build-up of the predictor set.

The left panel shows that the bootstrap intervals include zero at every horizon while the point estimates rise with lead time, so the apparent gain in skill is not resolved by the available sample. The right panel shows the gain accumulating through the tilt, field and wind terms, the fall at the polarity step, and the partial recovery when the recurrence index is added. Three groups behave as intended. Removing the HCS tilt terms, the interplanetary field or the solar wind speed lowers the point-estimate skill at every horizon, and removing the recurrence index lowers it at eighteen and twenty-four months while leaving the twelve-month forecast unchanged, exactly the horizon pattern reported in Section 4.4. The entries of Table 13b are differences of skill scores: at h = 24, for instance, removing the recurrence group raises the mean-square error from 17.369 to 18.865 and so lowers the skill from 0.363 to 0.309, the change of -0.055 entered in the table.

Two groups do not. Removing the cosmic ray history improves skill by between 0.17 and 0.23, and removing the polarity terms improves it by between 0.05 and 0.20, with the largest gains at the shortest horizon. The nested build-up shows the same effect from the other direction: the specification M3, comprising cosmic ray history with tilt, field and wind but without polarity, attains 0.378 at h = 18 against 0.315 for the full Model D. The polarity terms and their interactions, which are physically motivated and statistically significant under HAC inference, nevertheless cost out-of-sample accuracy.

The apparent advantage of M3 cannot be acted upon directly. Selecting a predictor set by inspecting validation skill would make that skill no longer out of sample, which is precisely the failure of assessment this paper is concerned with. We therefore tested adaptive specification properly, by nested cross-validation (Section 3.8), with the result given in Table 14.

Horizon	Nested cross-validation	95 % bootstrap interval	Fixed Model D	Fixed M3
12 months	+0.186	-0.300 to +0.500	+0.114	+0.307
18 months	+0.318	-0.123 to +0.623	+0.315	+0.378
24 months	+0.351	-0.110 to +0.622	+0.363	+0.368

Table 14. Skill obtained when the predictor specification is selected inside each training fold by nested cross-validation, compared with the two fixed specifications. 360 out-of-sample forecasts per horizon.

Adaptive specification yields skill indistinguishable from the fixed Model D: 0.318 against 0.315 at eighteen months, 0.351 against 0.363 at twenty-four, and 0.186 against 0.114 at twelve, the differences varying in sign. Crucially, the nested procedure does not attain the 0.378 recorded for M3 in Table 13a. The inner loop, which cannot see the outer test year, selects M3 or its variants only intermittently and never realises the advantage that M3 appears to hold when evaluated on the same data used to identify it. The superiority of the reduced specification is therefore a selection effect rather than a property that can be exploited in forecasting, and the fixed Model D is retained throughout this paper.

The pattern of specifications chosen by the inner loop is itself informative. The recurrence group is retained in every one of the thirty folds at eighteen and at twenty-four months, and in twenty-five of thirty at twelve months, despite four of the nine candidates omitting it. A procedure with no access to the validation data and free to discard the recurrence index almost never does so at the horizons where Section 4.4 reports a contribution. This provides an additional diagnostic consistent with the contribution of the recurrence index, alongside the coefficient test, the leave-one-out decomposition and the ordering by response rigidity, although it is not itself a formal test of incremental forecast accuracy.

A further check addresses whether the recurrence index is acting as an indirect proxy for the mean heliospheric field rather than carrying information of its own. Regressing the index on HCS tilt, field magnitude, solar wind speed and sunspot number over the 591 months from May 1976 to October 2025 gives a coefficient of determination of 0.323, so approximately two thirds of its variance is not accounted for by the other predictors. Replacing the index by the residual from that regression, which is orthogonal to them by construction, retains most of the contribution: the skill gain at eighteen months falls from +0.046 to +0.034 and at twenty-four months from +0.055 to +0.043, while remaining negligible at twelve months. The contribution therefore does not arise from the component of the index shared with the mean field. This does not establish that the index is a direct measure of corotating interaction regions; the aa index responds to several solar wind drivers, and the recurrence measure is best described as an indirect indicator of persistent solar wind organisation.

The polarity and cosmic ray history groups behave oppositely. Eighteen of thirty folds exclude the polarity terms at eighteen months while twenty-five of thirty retain them at twelve; nineteen of thirty exclude the cosmic ray history at twelve months while it is retained at longer horizons. The inner selection cannot settle on either group. Taken with the bootstrap dispersion of Table 12, this indicates that with approximately thirty annual forecast blocks, treated as dependence-preserving units rather than as strictly independent observations, the contribution of a marginal predictor group is not determined even in sign, which reinforces rather than contradicts the interpretation of Section 4.9.

5. Discussion

5.1 The case for persistence benchmarking

Section 4.2 provides a concrete demonstration of a general point. A model attaining out-of-sample correlations between 0.62 and 0.91 with the observed record would, presented in the conventional manner, appear to be a successful description of cosmic ray modulation. Plotted against observations it tracks the cycle faithfully. Yet it is materially worse than the assertion that nothing will change, because it carries a systematic offset that correlation is insensitive to.

A further consequence follows from Section 4.9. For the present validation span and bootstrap design, the smallest skill improvement that could be distinguished from zero at 95 % is approximately 0.4. A study using a comparable record and a similar dependence-preserving inferential procedure would face the same constraint. Reported agreement between model and observation therefore cannot substitute for a skill measurement, and a skill measurement without an accompanying interval cannot establish that a model carries predictive information. Persistence benchmarking is not itself unprecedented in this field: Reikard (2024) applies it to forecasts of the solar modulation potential and concludes that time-series forecasting skill is limited to proximate horizons. What has not previously been done, to our knowledge, is to apply it to direct cycle-scale forecasting of neutron-monitor intensity together with an explicit decomposition of recurrent and transient heliospheric predictors. By contrast, the physics-based frameworks of Pelosi et al. (2025, 2026), which solve the Parker transport equation and calibrate against PAMELA, AMS-02 and ACE, report reconstruction error relative to observed flux and relative to their own earlier model, but not relative to a naive baseline. CRI is unusually vulnerable to this failure of assessment, being smooth, strongly autocorrelated and slowly varying. The persistence forecast is correspondingly strong, and any evaluation which does not measure against it cannot establish that a model carries predictive information. We suggest that skill scores against persistence, and against climatology, should accompany reported correlations and errors as a matter of routine in this literature. The cost of doing so is negligible and the interpretive gain is substantial.

5.2 An asymmetry between the recurrent and transient channels

The central empirical result of this work is an asymmetry in the point-estimate forecast diagnostics. Recurrent solar wind organisation, quantified through the aa recurrence index, produces larger long-horizon point-estimate skill when added to the mean-field predictors, and does so specifically at horizons beyond one year. Transient injection, quantified through two separate catalogues and seven candidate constructions, shows no comparable contribution at any horizon tested. The forecast-error comparison does not, however, establish a statistically significant incremental advantage for the recurrence index on the present sample.

A plausible account is that the two channels differ in how completely the scalar field magnitude at 1 AU already registers them. Transient ejecta are, by construction, intervals of enhanced field, and their integrated contribution over a month or a year is largely reflected in the monthly mean field

magnitude which the model already uses. Adding a count of them supplies little that the field magnitude has not already supplied, which is consistent with the variance inflation factors of 16 to 18 obtained for the CDAW variable. Recurrence, by contrast, is not a measure of amplitude but of temporal organisation. Two months with identical mean field and identical mean speed may differ substantially in whether that wind is arranged into persistent corotating structures, and it is this distinction that the recurrence index captures and the field magnitude cannot.

That the recurrent channel should matter more at longer horizons follows naturally. Corotating structures associated with long-lived coronal holes persist over many rotations, so a measurement of their organisation retains relevance over a year or more, whereas an instantaneous field measurement decorrelates faster. The horizon dependence in Table 5 is thus not merely a statistical curiosity but the expected signature of a quantity whose autocorrelation time exceeds that of the alternatives.

We emphasise that the transient result is a null and should be read as such. It states that no skill was detected under the specified test on the interval 1996 to 2025, not that transient injection is physically irrelevant to modulation, which the literature on Forbush decreases and on merged interaction regions makes implausible. Richardson and Cane (2011) established a clear cosmic ray response to individual ICMEs. Our finding concerns the aggregate contribution at cycle-scale forecast horizons, which is a different question.

5.3 Reconciliation with the 27-day variation literature

An apparent tension requires explicit treatment. Richardson et. al, (1999) found the amplitude of recurrent cosmic ray depressions to be roughly 50 % larger in $A > 0$ than in $A < 0$. We find the correlation between solar wind recurrence and the cosmic ray level to be stronger in $A < 0$, at 0.63, than in $A > 0$, at 0.47. Read carelessly these are contradictory. They are not, because they concern different quantities. The earlier result describes the amplitude of the oscillation in CRI at the 27-day period, that is, how deeply each corotating structure depresses the flux as it passes. Our result describes the strength of association between a long-term measure of how organised the wind is and the slowly varying level of CRI, averaged over many rotations. A polarity epoch may exhibit shallow individual depressions while the overall degree of corotating organisation nonetheless tracks the modulation level closely, and vice versa. The two measurements are not merely different in magnitude but different in kind: one is an amplitude at a period, the other a correlation between two slowly varying envelopes.

It is relevant that the earlier result is itself not settled. Gil et. al, (2012) found the 27-day intensity amplitude in the $A < 0$ minimum of 2007 to 2009 to be comparable with that in the $A > 0$ minimum of 1995 to 1997, explicitly contradicting the polarity dependence reported by Richardson et al. (1999) and by Gil and Alania (2001), even while the corresponding anisotropy amplitude behaved as drift theory predicts. Gil and Mursula (2017) proposed a combination of convection and drift effects. The polarity dependence of the 27-day variation therefore remains an open question, and our result, being a different measurement, neither resolves nor is resolved by it.

5.4 Relation to the Sargent index and prior use

The recurrence index used here is not new. Sargent (1986) introduced R27 as a measure of the cycle-to-cycle stability of high-speed solar wind structure, and has since circulated a revised version as a preprint (Sargent 2021). The index used here follows the original construction as re-engineered by Chapman et al. (2021), computed without pre-smoothing the aa input, and so should not be identified with the revised R27. Chapman et al. (2021) re-engineered it from the autocovariance of aa and, in mapping it onto a Hale-cycle clock, observed that cosmic ray flux rises in step with the recurrence parameter and thereafter remains elevated. The index has also been used in reconstructions of near-Earth solar wind conditions, where its rationale is that fast streams raise recurrence through the formation of corotating interaction regions.

The contribution of the present work is therefore not the index but its use. To our knowledge this is the first assessment of the recurrence index as a predictor within a validated forecast model, and the first measurement of the horizon at which it contributes. The distinction matters: a variable may co-vary with CRI without improving a forecast, as several of our transient candidates demonstrate, and only an out-of-sample skill measurement can distinguish the two cases.

5.5 Non-stationarity and the limits of calibration

The bias identified in Section 4.2 has implications beyond the specific model that exhibited it. A relation calibrated over 1976 to 2013 underpredicts CRI over 2014 to 2025 by 4.3 %, because the later heliosphere lies outside the range of states represented in the calibration period. Any model whose parameters are estimated once on a historical interval and then applied forward inherits this exposure.

Two mitigations follow from the present analysis. The first is the differenced formulation, which removes the level dependence and thereby the principal channel through which the drift enters. The second is the rolling-origin refitting scheme, which allows the most recent cycle to inform each forecast rather than freezing the parameters at a historical epoch. Both are adopted here, and their combination is what converts the negatively skilled levels formulation into a model attaining positive out-of-sample skill from twelve months onward, reaching 0.36 at twenty-four months.

5.6 A rigidity ceiling on cosmic ray forecasting

The decline of forecast skill with response rigidity reported in Section 4.8 has a direct physical reading. Solar modulation is a low-rigidity phenomenon: the AMS-02 fluxes used for calibration vary by a factor of 4.5 between solar maximum and minimum at 1.1 GV, by 1.17 at 8.9 GV, and by only 1.03 at 83 GV. At high rigidity there is progressively less modulation signal to predict, and the residual variation is increasingly dominated by counting statistics and instrumental drift rather than by heliospheric conditions.

The consequence is a ceiling on what cycle-scale forecasting can achieve, expressed in the variable that governs it. The available measurements indicate that useful cycle-scale forecast skill is

concentrated at response rigidities below approximately 20 GV, while at 72 GV the model retains essentially no advantage over persistence. The calibration now samples the intermediate range, Tsumeb sitting at 22.7 GV, but its record ends in 2019 and so cannot enter the skill comparison; a precise rigidity threshold therefore still cannot be established from these data. Conversely the skill is greatest at the low-rigidity end, which is also where the particles most relevant to radiation dose and to single-event upsets reside. This is a favourable conclusion for the operational application, and it identifies polar and low-cutoff monitors as the appropriate targets for forecasting effort.

The parallel decline of the recurrence-index contribution is a consistency check of some value. A predictor that entered the model through a statistical accident would have no reason to fade with rigidity in step with the modulation signal itself. That it does so supports the physical interpretation advanced in Section 5.2.

5.7 Limitations

Several limitations qualify the results. The WSO tilt series available to us terminates in October 2025, which truncates the forecast reach of any tilt-bearing specification and forces the near-term forecast onto a weaker tilt-free model. The transient null rests on 180 rolling-origin forecasts spanning two polarity epochs, and a null on so restricted a sample cannot exclude a small effect. It establishes that no effect large enough to matter at this sample size was present. Residual autocorrelation persists in the differenced model, with a Durbin-Watson statistic of 0.554, which the HAC covariance estimator accommodates in the inference but does not eliminate from the specification. The polar field precursor relation rests on five points and is reported as indicative only. Finally, the Cycle 26 projection is conditional on imposed assumptions regarding cycle timing, length and profile, and should not be read as a prediction.

6. Conclusions

1. Conventional levels-based regression forecasting of CRI does not outperform persistence out of sample at any horizon tested. It fails through a systematic bias of approximately -4 % when extrapolated beyond the range of heliospheric states represented in its calibration period, despite out-of-sample correlations as high as 0.91.
2. A differenced formulation, in which the target is the change in intensity over the forecast horizon and all predictors are strictly causal and evaluated at the forecast issue time, yields positive point-estimate skill against persistence from twelve months onward, reaching 0.36 at twenty-four months under expanding-window rolling-origin validation. At six months the model does not outperform persistence.
3. The 27-day recurrence index constructed from the aa geomagnetic index contributes no measurable skill at horizons of twelve months or less, and contributes at eighteen and twenty-four months, with a standardised coefficient rising monotonically with lead time from 0.09 to 1.20 and a variance inflation factor of 1.5.

4. No statistically robust incremental forecast skill from transient injection was detected on either the near-Earth ICME list or the global CME catalogue, tested under an acceptance criterion fixed in advance across seven candidate constructions.
5. Neither the skill of the differenced model against persistence nor the incremental contribution of the recurrence index reaches the 95 % significance level under Diebold-Mariano testing or moving-block bootstrap. The twenty-four-month Model D point estimate is 0.363, with a 95 % bootstrap interval spanning -0.130 to +0.626 and a Diebold-Mariano comparison against persistence giving $p = 0.160$. The results therefore do not establish a statistically significant forecast improvement, although the long-horizon point estimates and four complementary diagnostics provide suggestive evidence warranting further investigation with longer records.
6. Ablation shows that the HCS tilt, interplanetary field, solar wind and recurrence groups each raise the point-estimate skill, while the cosmic ray history and polarity groups reduce it despite being statistically significant under HAC inference. Nested cross-validation, in which the specification is chosen inside each training fold, yields skill indistinguishable from the fixed model and does not realise the apparent advantage of the reduced specification, which is therefore a selection effect. The inner selection retains the recurrence group in every fold at eighteen and twenty-four months, providing an additional diagnostic consistent with its contribution.
7. A continuous polarity index derived from the smoothed solar polar field retains the reversal epochs that are conventionally discarded and increases the number of usable forecasts from 306 to 360. It does not improve point skill: the categorical treatment attains 0.469 at twenty-four months against 0.363 for the continuous index, so the index is offered as a methodological proposal that widens coverage rather than as an improvement in accuracy.
8. Calibration against AMS-02 daily proton fluxes gives median response rigidities of approximately 10 GV for all low-cutoff neutron monitors irrespective of cutoff, 18.9 GV for Mexico City, 22.7 GV for Tsumeb and 72.1 GV for Doi Inthanon. Comparison of Mexico City with Tsumeb, whose cutoffs differ by 13 % but whose altitudes differ by a factor of 1.8, shows that atmospheric depth rather than cutoff governs the response. Forecast skill shows a broadly declining trend across this range and falls essentially to zero at the highest rigidity, indicating that useful cycle-scale forecast skill is concentrated below approximately 20 GV.
9. Annual CDAW CME counts are not stationary in detection sensitivity and should not be used as a homogeneous solar activity proxy in long-term studies without correction.

Acknowledgements

This work was carried out under a research project sponsored by the Madhya Pradesh Council of Science and Technology (MPCST), Bhopal [A/RD/RP-2/413], and the author gratefully acknowledges that support. Neutron monitor data were obtained from the Neutron Monitor

Database (NMDB, www.nmdb.eu), Heliospheric current sheet tilt angles and polar field measurements were obtained from the Wilcox Solar Observatory via wso.stanford.edu, courtesy of J. T. Hoeksema; the Wilcox Solar Observatory is currently supported by NASA. Interplanetary magnetic field and solar wind data were obtained from the OMNI dataset through NASA Goddard Space Flight Center's Space Physics Data Facility OMNIWeb service. Sunspot numbers were obtained from the World Data Center SILSO, Royal Observatory of Belgium, Brussels. The aa geomagnetic index was provided by the International Service of Geomagnetic Indices (isgi.unistra.fr) and the collaborating institutes that operate the contributing observatories.

Proton flux measurements used for the rigidity response calibration of Section 3.6 were taken from the published daily spectra of the AMS-02 collaboration, and the author thanks the collaboration for making those data available. The near-Earth interplanetary coronal mass ejection catalogue compiled by I. G. Richardson and H. V. Cane was used for the transient-channel test of Section 4.5. Coronal mass ejection statistics were taken from the SOHO/LASCO CME catalogue, which is generated and maintained at the CDAW Data Center by NASA and the Catholic University of America in cooperation with the Naval Research Laboratory; SOHO is a project of international cooperation between ESA and NASA.

Data availability

Neutron monitor data are available from the Oulu cosmic ray station and from the Neutron Monitor Database. HCS tilt angles and polar field measurements are provided by the Wilcox Solar Observatory. Sunspot numbers are from WDC-SILSO, Royal Observatory of Belgium. Interplanetary parameters are from the OMNIWeb service of the NASA Space Physics Data Facility. The aa index is provided by the International Service of Geomagnetic Indices. The near-Earth ICME list is maintained by I. G. Richardson and H. V. Cane. The SOHO/LASCO CME catalogue is generated and maintained at the CDAW Data Center by NASA and the Catholic University of America in cooperation with the Naval Research Laboratory.

References

Aguilar, M., et al. (AMS Collaboration): 2021, Periodicities in the daily proton fluxes from 2011 to 2019 measured by the Alpha Magnetic Spectrometer on the International Space Station from 1 to 100 GV. Phys. Rev. Lett. 127, 271102.

Cane, H.V., Richardson, I.G.: 2003, Interplanetary coronal mass ejections in the near-Earth solar wind during 1996-2002. J. Geophys. Res. 108(A4), 1156.

Chapman, S.C., McIntosh, S.W., Leamon, R.J., Watkins, N.W.: 2021, The Sun’s magnetic (Hale) cycle and 27 day recurrences in the aa geomagnetic index. Astrophys. J. 917, 54. DOI 10.3847/1538-4357/ac069e.

Clette, F., Lefèvre, L.: 2016, The new sunspot number: assembling all corrections. Solar Phys. 291, 2629.

Fiandrini, E., Tomassetti, N., Bertucci, B., Donnini, F., Graziani, M., Khiali, B., Reina Conde, A.: 2021, Numerical modeling of cosmic rays in the heliosphere: solar modulation of protons. Phys. Rev. D 104, 023012.

Gil, A., Alania, M.V.: 2001, 27-day variation of cosmic rays for the minima epochs of solar activity: experimental and 3-D drift modeling results. Proc. 27th ICRC, Hamburg, 9, 3725.

Gil, A., Modzelewska, R., Alania, M.V.: 2012, 27-day variations of the galactic cosmic ray intensity and anisotropy in the minimum of the 23rd solar activity cycle. Adv. Space Res. 50, 712.

Gil, A., Mursula, K.: 2017, Hale cycle and long-term trend in variation of galactic cosmic rays related to solar rotation. Astron. Astrophys. 599, A112.

Gopalswamy, N., Yashiro, S., Michalek, G., Stenborg, G., Vourlidas, A., Freeland, S., Howard, R.: 2009, The SOHO/LASCO CME catalog. Earth Moon Planets 104, 295. DOI 10.1007/s11038-008-9282-7.

Gopalswamy, N., Yashiro, S., Akiyama, S., Mäkelä, P.: 2025, The SOHO/LASCO CME catalog - Version 2. Proc. IAU. DOI 10.1017/S1743921325000559.

Gupta, M., Mishra, V.K., Mishra, A.P., 2006a. Study of cosmic ray intensity variations in relation to solar activity for sunspot cycles 19-23. Indian Journal of Radio & Space Physics, 35, 167-173.

Gupta, M., Mishra, V.K., Mishra, A.P., 2006b. Correlation of the long-term cosmic ray intensity variations with sunspot numbers and tilt angle. Indian Journal of Radio & Space Physics, 35, 387-395.

Jokipii, J.R., Levy, E.H., Hubbard, W.B.: 1977, Effects of particle drift on cosmic-ray transport. I. General properties, application to solar modulation. Astrophys. J. 213, 861.

Koldobskiy, S.A., Kähkönen, R., Hofer, B., Krivova, N.A., Kovaltsov, G.A., Usoskin, I.G.: 2022, Time lag between cosmic-ray and solar variability: sunspot numbers and open solar magnetic flux. Solar Phys. 297, 38.

Kóta, J., Jokipii, J.R.: 2001, Recurrent depressions of galactic cosmic rays in the heliosphere. Proc. 27th ICRC, Hamburg.

Mayaud, P.N.: 1972, The aa indices: a 100-year series characterizing the magnetic activity. J. Geophys. Res. 77, 6870.

Mishra, V.K., Tiwari, D.P., 2003. Long-term association of solar activity and cosmic ray intensity for solar cycles 21 and 22. Indian Journal of Radio & Space Physics, 32, 65-69.

Mishra, V.K., Mishra, A.P.: 2018, Long-term modulation of cosmic-ray intensity and correlation analysis using solar and heliospheric parameters. Solar Phys. 293, 141.

Mishra, V.K., 2025. Statistical Comparison of Solar Activity and Cosmic Ray Modulation during Solar Cycles 24 and 25. Vigyan Prakash, 23, 17-27. https://doi.org/10.5281/zenodo.18442719

Newey, W.K., West, K.D.: 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703.

Parker, E.N.: 1965, The passage of energetic charged particles through interplanetary space. Planet. Space Sci. 13, 9.

Pelosi, D., Barao, F., Bertucci, B., Fiandrini, E., Orcinha, M., Tomassetti, N.: 2025, A forecasting framework for galactic cosmic ray flux in space weather applications. Adv. Space Res. 76, 5700. DOI 10.1016/j.asr.2025.08.022.

Pelosi, D., Barao, F., Bertucci, B., Fiandrini, E., Orcinha, M., Tomassetti, N.: 2026, New model for long-term forecasting of galactic cosmic rays. Phys. Rev. D 114, 063005. DOI 10.1103/npfk-lwy9.

Reikard, G.: 2024, Forecasting the solar modulation potential: tests of time series models. J. Atmos. Solar-Terr. Phys. 262, 106326. DOI 10.1016/j.jastp.2024.106326.

Richardson, I.G., Cane, H.V., Wibberenz, G.: 1999, A 22-year dependence in the size of near-ecliptic corotating cosmic ray depressions during five solar minima. J. Geophys. Res. 104, 12549.

Richardson, I.G., Cane, H.V.: 2010, Near-Earth interplanetary coronal mass ejections during solar cycle 23 (1996-2009): catalog and summary of properties. Solar Phys. 264, 189. DOI 10.1007/s11207-010-9568-6.

Richardson, I.G., Cane, H.V.: 2011, Galactic cosmic ray intensity response to interplanetary coronal mass ejections/magnetic clouds in 1995-2009. Solar Phys. 270, 609.

Ross, E., Chaplin, W.J.: 2019, The behaviour of galactic cosmic-ray intensity during solar activity cycle 24. Solar Phys. 294, 8.

Sargent, H. H., “The 27-DAY Recurrence Index”, in *Solar Wind Magnetosphere Coupling*, 1986, vol. 126, p. 143. doi:10.1007/978-90-277-2303-1_11.

Sargent, H.H.: 2021, A revised 27-day recurrence index. Preprint, arXiv:2101.02155.

Siluszyk, M., Iskra, K.: 2020, Modeling the time delay problem of galactic cosmic ray flux in solar cycles 21 and 23. Solar Phys. 295, 68. DOI 10.1007/s11207-020-01628-w.

Sun, X., Hoeksema, J.T., Liu, Y., Zhao, J.: 2015, On polar magnetic field reversal and surface flux transport during solar cycle 24. Astrophys. J. 798, 114.

Tomassetti, N., Orcinha, M., Barão, F., Bertucci, B.: 2017, Evidence for a time lag in solar modulation of galactic cosmic rays. Astrophys. J. Lett. 849, L32.

Usoskin, I.G., Kananen, H., Mursula, K., Tanskanen, P., Kovaltsov, G.A.: 1998, Correlative study of solar activity and cosmic ray intensity. J. Geophys. Res. 103, 9567.

Usoskin, I.G.: 2017, A history of solar activity over millennia. Living Rev. Solar Phys. 14, 3.

Väisänen, P., Usoskin, I., Mursula, K.: 2021, Seven decades of neutron monitors (1951-2019): overview and evaluation of data sources. J. Geophys. Res. Space Phys. 126, e2020JA028941.

Vourlidas, A., Howard, R.A., Esfandiari, E., Patsourakos, S., Yashiro, S., Michalek, G.: 2010, Comprehensive analysis of coronal mass ejection mass and energy properties over a full solar cycle. Astrophys. J. 722, 1522.

Yashiro, S., Gopalswamy, N., Michalek, G., St. Cyr, O.C., Plunkett, S.P., Rich, N.B., Howard, R.A.: 2004, A catalog of white light coronal mass ejections observed by the SOHO spacecraft. J. Geophys. Res. 109, A07105.