Hidden multi-state structure versus the piecewise-Poisson flare rate on 50 years of GOES/XRS solar flares

Several latent switching layers beat one; the geometric ladder is not what the record prefers — a threshold-dependent boundary, decomposed and measured

Evan Tabak Atlas (ORCID 0009-0007-7374-2338)

Draft — result paper. Companion to the methods paper `papers/01-correspondence-metrology.md` (the MSM-Cox model, its closed forms, and the photon-channel metrology) and to the X-ray-binary result paper `papers/02-tournament-grs1915-cygx1.md` (the same tournament at continuous-flux altitude). Target venue: ApJ, where the incumbent description this paper races is resident; Solar Physics is the in-hand alternative. Zenodo DOI first (it needs no endorsement and establishes the priority date), then arXiv (astro-ph.SR / physics.space-ph). Every quantitative claim traces to a named, immutable results record under `out/` — the real-record tournament `v25` and its now-cast `v26`; the sealed synthetic two-world calibration `v16` / `v17`; the event-altitude estimability floor `v20` and its selection-vs-recovery split `v32`; the real-stream event-time surrogate audit `v29`; and the cross-arena contrasts §6.4 reads against — `v13` / `v14` (X-ray binaries) and `v31` (stellar flares, synthetic). Figures regenerate with `python -m fenocosm.paper_figures`. GOES/XRS flare event lists: NOAA National Centers for Environmental Information and the Space Weather Prediction Center.

Priority note (updated 2026-08-31; this note previously read “no DOI and no priority date yet”): the Zenodo deposit is published — 10.5281/zenodo.22180473 went live 2026-08-31 and establishes the priority date. arXiv posting and the choice of journal remain owner actions, as does the domain-collaborator pass §10 asks for.

Abstract

The statistics of solar flare occurrence have a mature incumbent description: a **time-dependent (nonstationary) Poisson process** whose rate varies with the solar cycle and with active-region evolution, usually represented by a piecewise-constant rate. No *estimable hidden-state* alternative — an exact hierarchical regime filter with a likelihood — has been entered against that description. We enter one, and report what it settles and what it does not.

On the full ~50-year GOES/XRS soft-X-ray flare record (**8,393** M1.0+ and **64,996** C1.0+ flares, 1975–2025, era-stitched from the NCEI legacy reports and the GOES-R L2 flare summary), we run a fair likelihood tournament on daily counts. The entries are an exact 2^k-state Markov-switching multifractal cascade (MSM-Cox, four free parameters at every depth), a homogeneous Poisson process, the piecewise-constant Poisson of a Bayesian-blocks partition — the incumbent — and a **free-transition Poisson hidden Markov model** with an unconstrained transition matrix. Every model carries its parameter count, the satellite-era recalibration travels as a per-bin covariate, and every entrant is scored both in-sample (BIC) and out-of-sample (held-out log-likelihood on a contiguous final block).

(i) Several latent switching layers beat one, decisively, at both thresholds and at fixed parameter count. The depth gain from one switching layer to six is ΔBIC **+3,175** (M1.0) and **+10,195** (C1.0), and **+460** / **+1,436** nats out of sample, with the parameter count unchanged across the scan. Read against the sealed two-world calibration this is 11× the no-hierarchy world’s depth gain per flare at M1.0 and 2.8× at C1.0; a pure-Poisson parametric bootstrap

places the cascade's edge over *homogeneous* Poisson beyond every replicate (**p = 0.0099**, the ensemble floor at N = 100 — run at k = 3, so it rules out only the weakest entrant); and a residual-specificity cell separates a genuine cascade from a smooth modulated rate at **81×** where the naive statistic separates them at **1.4×**, at C1.0, where that cell is clean.

(ii) Against the field's piecewise-constant incumbent the cascade wins on parsimony, not on fit — and we decompose it. ΔBIC is **+3,102** (M1.0) and **+6,172** (C1.0), but the Bayesian-blocks partition carries 531 and 1,423 parameters against the cascade's four and *fits the record better in raw likelihood*, by **+1,035** and **+3,876** nats. The entire BIC margin is the parameter penalty net of that fit deficit. What the comparison establishes is that a four-parameter hidden-state description comes within about a thousand nats of a 531-parameter partition on 18,262 days — not that it describes the record better.

(iii) The specific *geometric* ladder buys no descriptive advantage over a free transition matrix at these sample sizes, and the threshold dependence is consistent with a resolving-power effect rather than evidence about the Sun. Against the free HMM the cascade leads at M1.0 by only **+119** BIC — of which **+117.8** is the parameter-count penalty on twelve extra parameters and only **+1.7** is fit (a **+0.83**-nat log-likelihood difference over 18,262 days) — and it loses by **−1,475** BIC at C1.0. (That lead is over the HMM state counts we ran, $J \in \{2, 4\}$, and not over the free-transition family: the J = 2 → J = 4 step alone gains 2,640 BIC, and an extended-J run is queued — Limitations.) Out of sample the M1.0 margin is **+2.17** nats over a 5,479-day held-out block (a tie) and the C1.0 margin is **−310.83** nats (a loss). The sealed calibration shows this axis cannot be read as a hierarchy test: in the synthetic *no-hierarchy* world the same comparison still favours the cascade (**+147** BIC, **+10.3** nats), and the real record sits below that no-hierarchy world at both thresholds.

(iv) A causal, filter-only 24-h M-class now-cast carries real skill from the event record alone, with a Brier Skill Score of **+0.589** (95 % CI [0.521, 0.667]) against the training-period Poisson climatology — a reference that over-forecasts the quieter verification period by 2.9×. Rescored against the verification period's own sample climatology the same forecasts give **+0.338**, and the ordering is unchanged: the filter leads the cycle-timescale rate baseline (**+0.507** → **+0.205**) and edges recency persistence (**+0.576** → **+0.317**) on point estimates, with no paired interval claimed either way, while sitting behind it on peak TSS (**0.633** vs **0.621**). The forecast over-predicts in all ten reliability bins, which we trace to a train/test activity shift rather than to a generic model pathology, and the honest ceiling stands: this filter never sees a magnetogram.

The claim we make is narrow: the first estimable hidden-state description of the flare-rate record, and the fair tournament placing several latent switching layers against one and against the piecewise-constant description. We do **not** claim a first arbitration of self-organized criticality against turbulence, we do not claim a physical mechanism, and — on the evidence of (iii) — we do not claim that the geometric parameterisation is the one the record prefers.

---

1. Introduction

1.1 The incumbent description of flare timing

Solar flares arrive in a manner that has been described, tested, and re-tested for a quarter of a century as a **time-dependent Poisson process**. Wheatland (2000) showed that the observed distribution of waiting times between GOES flares — a power-law tail over a broad range, often read as evidence of a scale-free or self-organized-critical process — follows analytically from a *nonstationary Poisson* process in which the fraction of time the Sun spends at a given flaring rate is exponentially distributed, with no memory required beyond the varying rate. Wheatland (2004) turned that reading into an operational Bayesian flare forecaster. Scargle et al. (2013) supplied the representation the field now uses for such a rate: a **Bayesian-blocks** piecewise-constant partition, optimal under an explicit fitness function and change-point prior. Later work has probed the picture from both sides. Snelling et al. (2020) found excess mutual information in GOES flare waiting times at a timescale of ≈ 30 hr, measured against Bayesian-block-generated nonstationary-Poisson surrogates — which *rejects* the pure nonstationary-Poisson null at that lag, and is the closest prior result to the one reported here. Aschwanden & Johnson (2021) measured solar memory across timescales from hours to the dynamo rise time, and Aschwanden, Johnson & Nurhan (2021) derived the waiting-time power-law slope analytically from a Poisson process with a time-varying rate, $\alpha = 2 + 1/p$.

The incumbent is therefore not a straw model. It is a well-tested, parsimonious, physically motivated description with an operational forecaster attached. What it asserts is that, conditional on a rate, flares are independent — that all of the structure lives in $\lambda(t)$, and that $\lambda(t)$ is adequately described by a slow modulation with a piecewise-constant representation.

1.2 The gap: an estimable hidden-state alternative has never been entered

What has *not* been placed against that description is a **hidden-state** model: one in which the rate is generated by unobserved discrete regimes with an exact likelihood, so that the two descriptions can be raced on the same data with the same accounting. The obstacle has been estimability. Latent multi-timescale descriptions of solar activity are usually written as unobservable fields and summarised by a power spectrum; they are rarely *fit*, because their likelihood is intractable.

The companion methods paper (`papers/01-correspondence-metrology.md`) supplies the estimable member of that class: the **Markov-switching multifractal** (Calvet & Fisher 2001, 2004), a product of *k* independent two-state layers whose switching rates lie on a geometric ladder, coupled to a Poisson counting channel (**MSM-Cox**). Its likelihood is exact — a Kronecker-factored forward pass over the 2^k latent states (Baum et al. 1970; Rabiner 1989) — and it carries a *fixed* set of four free parameters at every depth *k*. It is therefore a proper state-space model that can be entered into a BIC and held-out tournament against the incumbent without a Bayesian apparatus, an MCMC, or an approximation.

1.3 What this paper claims, and what it does not

The claim is narrow and we hold it narrowly: **the first estimable hidden-state description of the flare-rate record, and the fair tournament placing several latent switching layers against one and against the piecewise-constant description.** It is a model-comparison result.

Three things it is not. We do **not** claim a first arbitration of self-organized criticality against turbulence; the tournament does not adjudicate that question and was not built to. We do not claim a physical mechanism for the fitted layers. And — the point §6 is about — we do not claim that the *geometric* form of the ladder is what the record prefers. On the contrary: we measure that it is not, and we decompose why. One further scope note belongs here rather than in a footnote: the margin over the piecewise-constant incumbent is a **parsimony** margin, not a fit margin (§4.2), and we state it that way throughout.

1.4 Contributions

1. **The tournament, on the real record.** The committed flare tournament, reused byte-for-byte from its sealed synthetic calibration, run on the full ~50-year GOES/XRS event record with the satellite-era recalibration carried as a covariate (§§2–4; record `v25`).
2. **The depth result, with its null stated in absolute terms.** The depth gain over the single switch, in sample and out, at fixed parameter count — quoted beside the depth gain the sealed *no-hierarchy* world produces under the same scan, because that world's gain is large and positive and a bare depth gain is therefore not diagnostic (§4.1, §4.4).
3. **The incumbent margin, decomposed.** The BIC margin over the Bayesian-blocks partition separated into its parameter-penalty and its fit components, which run in *opposite* directions: the incumbent fits better and the cascade is preferred on parsimony. We hold the claim to what that decomposition supports (§4.2).
4. **Significance and a specificity control.** A pure-Poisson parametric bootstrap (which rules out the weakest entrant and no more), and a residual-specificity cell that separates a genuine cascade from a smooth modulated rate — the artefact any hidden-structure claim on a cycle-modulated record has to clear first, and which this cell clears at C1.0 only (§5).
5. **The threshold-dependence finding, decomposed — a first-class result, not a caveat.** The cascade's margin over a free-transition Poisson HMM separated into penalty and fit, read on held-out data at both thresholds, and placed against the sealed calibration, which shows that this comparison **cannot** be read as a hierarchy test (§6).
6. **The fitted ladder, quoted under the program's degeneracy discipline.** Span, multiplier and rung timescales reported only as excesses over raced rivals, with the selection-vs-recovery split made explicit and the boundary-solution flag stated (§7).
7. **A causal now-cast, its reference forecast audited, and an honest ceiling.** A walk-forward, filter-only 24-h M-class probability forecast with block-bootstrap intervals, rescored against the verification period's own climatology as well as the training period's, its reliability diagram, and the near-tie against recency persistence reported plainly (§8; record `v26`).
8. **A degenerate Hawkes reported as a result.** The event-times arena runs to its estimator's ceiling at both thresholds, and a zero-self-excitation smooth-cycle null runs to the same ceiling — the Filimonov–Sornette confound, on real solar data, with the ceiling itself named (§9; record `v29`).

---

2. Data

2.1 The era-stitched record

We use the **GOES/XRS soft-X-ray flare event lists**, era-stitched from two public NOAA products into one continuous daily-binned count series over 1975–2025 (**18,262** daily bins on the protocol's 1975.0–2025.0 decimal-year grid):

- the **NCEI legacy** fixed-width flare reports (`goes-xrs-report_YYYY.txt`), 1975–2017, which carry the operational SWPC-scaled long-channel flux; and
- the **GOES-R L2 XRS flare summary** (`xrsf-l2-flsum_science`, GOES-16), 2017–present, which reports **true (unscaled)** 1–8 Å flux.

The two are stitched at decimal year **2017.107** — the flsum start, so there is no overlap by construction. For both thresholds the class **letter** is the exact discriminant (peak flux $\geq 10^{-5}$ W m^{-2} ⟺ class ∈ {M, X}; $\geq 10^{-6}$ ⟺ {C, M, X}), which is robust even in the early era where the reported significand code is `0`. The flux ↔ class convention was validated against the live SWPC JSON feed before the archival fetch: **76 of 77** live flares reproduce it, the single miss being class-boundary rounding. The live feed post-dates the archival flsum, so this checks the assignment logic the whole fetch depends on rather than the archival values themselves.

The record contains **8,393** flares at M1.0+ (the primary threshold) and **64,996** at C1.0+ (the sensitivity threshold; a completeness choice from the literature — Snelling et al. 2020, Aschwanden & Johnson 2021 — not a NOAA product).

Table 1 — The record by satellite era (record `out/v25/`; counts are C1.0+).

era	years	flares (C+)	C / M / X	`era_calibration_factor`
0	1975–1986	16,938	14,605 / 2,167 / 166	0.90
1	1986–1996	14,724	12,532 / 2,038 / 154	0.92
2	1996–2009	14,747	13,178 / 1,443 / 126	0.97
3	2009–2020	8,720	7,931 / 738 / 51	1.00
4	2020–2025	9,867	8,357 / 1,439 / 71	1.38

Era totals reconcile to the whole exactly (16,938 + 14,724 + 14,747 + 8,720 + 9,867 = 64,996; the M and X columns sum to 8,393).

2.2 The recalibration, carried as a covariate — and what the covariate is not

GOES-8 through GOES-15 carried operational SWPC scaling factors on the archived irradiances (0.70 on the 1–8 Å long channel, 0.85 on the short channel, per the NOAA XRS documentation); the GOES-R series reports true irradiances with no scaling. Hudson et al. (2024) re-reduced the whole ≥ C1 record and give the correction as a uniform factor of **1.43** on operational GOES-1 through GOES-15 irradiances, so at a fixed class threshold the modern era admits correspondingly more events, and the discontinuity lives **entirely in the class labels**. The loader carries it as a per-bin covariate (`era_calibration_factor`, 1.38 in the ≥ 2020 era) and divides it out, so no entrant is asked to explain an instrumental step change as physics.

That is the honest treatment available in this record, and it is a coarse one. Four limits travel with it, and we state them rather than claiming them away.

- **The covariate is not Hudson's factor.** Hudson's correction is *uniform* across GOES-1…15; the loader instead applies a five-step ladder (0.90 / 0.92 / 0.97 / 1.00 / 1.38 across the five eras of Table 1). The 1.38 modern step is close to, but not equal to, the 1.43 of the published re-reduction, and the three sub-unity factors inside Hudson's uniform range are the loader's own and are not sourced to a published recalibration. What the covariate implements is therefore a *stepwise era normalisation informed by* the published recalibration, not the recalibration itself.
- **A factor on flux is not a factor on counts.** At a fixed class threshold, with an occurrence distribution N(>F) ∝ F^(1−α) and α ≈ 2, a flux factor r maps to a count factor r^(α−1), which is threshold-dependent. A single per-era count factor is applied at both M1.0 and C1.0.

- **The era boundary is not the stitch date, and it is near a cycle minimum.** GOES-R reports true flux from the 2017.107 stitch, but the covariate places the recalibration boundary at 2020 (when GOES-16 became the operational primary), so the ~573 GOES-R C+ flares (63 of them M1.0+) in [2017.107, 2020) — near solar minimum, few flares — carry factor 1.00 and are *under*-corrected. More generally, the four era boundaries (1986, 1996, 2009, 2020) all fall within about a year of a tabulated cycle minimum, so the covariate is structurally close to a per-cycle amplitude rescaling and cannot be cleanly separated from one.
- **The covariate under-corrects the measured flux shift.** The sealed provenance records the median true GOES-R 1–8 Å flux per era; the era-4/era-3 ratio of those medians is **1.73**, against a covariate ratio of 1.38. Part of that difference is genuine solar-cycle amplitude rather than calibration, and no attempt is made here to separate them.

What the covariate is *not* asked to do is carry the verdict. Its measured effect on the fit is small (§7.3), and the entrant that would be most sensitive to a mis-specified era step — the piecewise-constant partition, which must spend change points to absorb one — is in fact fitted **without** the covariate (§3.2), which we flag there rather than here because it bears on the comparison, not on the data.

2.3 What is not in this analysis

Per-active-region flare streams, magnetograms, and any solar index other than the documented cycle envelope are outside this paper. The filter sees the flare event list and nothing else. That restriction is deliberate — it is what makes the now-cast of §8 a clean statement about the information in the event record — and it is also this paper's principal limitation (§8.4).

---

3. Models and tournament design

3.1 Two arenas, kept apart

The tournament runs in two likelihood **arenas**, and results are never carried across them. The **counts arena** works on the daily count series and penalises with n = 18,262 daily bins; the **event-times arena** works on the flare times as a point process and penalises with n = the number of events. A BIC computed against one sample size is not comparable to a BIC computed against the other, and we do not compare them. The counts arena carries the verdict of this paper; the event-times arena appears in §9, where its own result is that it is degenerate.

3.2 The entries (counts arena)

- **Homogeneous Poisson** (1 parameter) — a constant rate; the floor.
- **Piecewise-constant Poisson** — the Bayesian-blocks partition (Scargle et al. 2013) at `p0 = 0.05`, fitted on the same series. This is **the incumbent**: it is the field's standard representation of a time-dependent flare rate. It selects 266 blocks at M1.0 and 712 at C1.0.
- **Free-transition Poisson HMM** ($J \in \{2, 4\}$; J^2 parameters — J emission rates plus J(J − 1) free transition entries — so 4 and 16) — a hidden Markov model with *J* Poisson emission rates and an **unconstrained** J × J transition matrix. This is the adversary that matters: it is discrete and multi-state like the cascade, but it is free where the cascade is constrained. Whatever the cascade can do that generic discreteness cannot must show up here.

- **MSM-Cox cascade, depth k** (4 parameters at every k) — *k* independent two-state layers whose switching rates lie on a geometric ladder spanning a factor *S*, with mirror-balanced multipliers m_hi and 1/m_hi; the product modulates a Poisson rate λ_0. Free parameters (m_hi, f_slow, S, λ_0); κ is fixed at 1 for identifiability. Scanned over k = 1…6, with a BIC-only deep extension to k = 7, 8 at the primary threshold. **k = 1 is the single switch** — a two-state regime model — so the k = 1 → k *margin* is* the hierarchy-vs-single-switch test, measured within one model family at fixed parameter count.[1]

The per-bin era covariate multiplies the rate for the cascade, the homogeneous Poisson entrant and the Poisson HMM. **It is not passed to the Bayesian-blocks entrant**, whose implementation takes no covariate: the blocks partition therefore has to absorb the era steps by spending change points, each costing 2·ln(18,262) = 19.6 BIC. That asymmetry runs *against* the incumbent, and it is one of two accounting effects in this comparison that do (the other is §3.3's selection-versus-scoring penalty mismatch). We name both rather than asserting that every accounting error favours the incumbent.

3.3 Two criteria, and why both are reported

Every entrant is scored two ways: **BIC** (Schwarz 1978) on the full series, and **held-out log-likelihood** on a contiguous final block. Both are reported for every comparison in this paper, and where they disagree we say so.

This is not decoration. The cascade's parameter set is *fixed in k*, so BIC never charges it for the growing flexibility of a 2^k-state latent space; and the free HMM carries twelve more parameters than the cascade at J = 4, so a BIC comparison between them is dominated by a penalty term of 12 · ln(18,262) = **+117.8** before a single likelihood is evaluated. A BIC margin of that order between these two entrants is therefore not evidence of anything about fit. The program's own depth-honest discipline (records `v13`, `v14`) reached the same conclusion at continuous-flux altitude and settled on out-of-sample scoring as the arbiter. §6 follows it.

Three consequences must be flagged before the tables, and two of them cut against the incumbent.

- **The blocks parameter count understates its complexity.** 2·n_blocks − 1 counts block rates and change-point *locations* but cannot count the search over partitions. This runs in the cascade's favour.
- **The partition is selected under a softer penalty than it is scored by.** The Bayesian-blocks dynamic program buys change points at Scargle et al.'s `ncp_prior` — 7.39 nats each at p_0 = 0.05 and N = 18,262 — while the tournament then charges 2·ln(18,262) = 19.6 BIC per extra block. Selection is therefore 2.7× more permissive than scoring, so the partition deliberately carries more change points than the scoring rule wants. This runs against the incumbent, and it is not a small effect: §4.2 shows the entire margin is the penalty term. A p_0 / `ncp_prior` sensitivity sweep, and a variant partition selected under `ncp_prior = ln n` so that selection and scoring agree, are the obvious controls and are not in this record; we name them in Appendix B.
- **The blocks held-out column is not reportable from this record, and the repair alone will not make it comparable.** A zero-rate guard in the blocks entrant sat on an unreachable branch under the protocol version this record was sealed at (`p6`), so an empty final training block could score a held-out event under a predicted rate of zero; the defect was found and repaired at protocol `p7` (record `v32`), and records v0–v31 predate the repair and are not edited. But the entrant's held-out *predictor* is also the terminal training block's constant rate, extrapolated flat across fifteen years of an 11-year-cycle series, while the cascade and the HMM run their forward filters through the held-out block. No practitioner forecasts with the terminal

Bayesian block; the fair form is an online or expanding-window change-point predictor, which this record does not contain. We therefore compare the cascade to the blocks incumbent **on BIC only**, say so at every table, and do not present the missing column as a gap a guard repair would close.

3.4 The calibration, sealed before the real data

The tournament was not built on the real record. It was built and frozen on a deterministic **two-world synthetic calibration** (record `v16`, with its now-cast companion `v17`), both worlds riding the same solar-cycle envelope and the same era calibration so that the *only* difference between them is the presence of a faster multiplicative hierarchy:

- the **cycle world** — a smooth, strongly modulated Poisson rate with **no** hierarchy: the piecewise-Poisson/Wheatland picture made synthetic. This is the **specificity arm**: the cascade must not manufacture a hierarchy here.
- the **cascade world** — the same envelope with a genuine six-layer cascade underneath. This is the **positive control**.

Both are sealed and immutable; this paper cites them and does not recompute them. The driver that produced record v25 calls the v16 tournament, bootstrap and smoother byte-for-byte — only the light curve differs. Every real-record number below is therefore read against a calibration that was committed before the real data were touched, which is what makes the specificity statements in §5 and §6 statements at all.

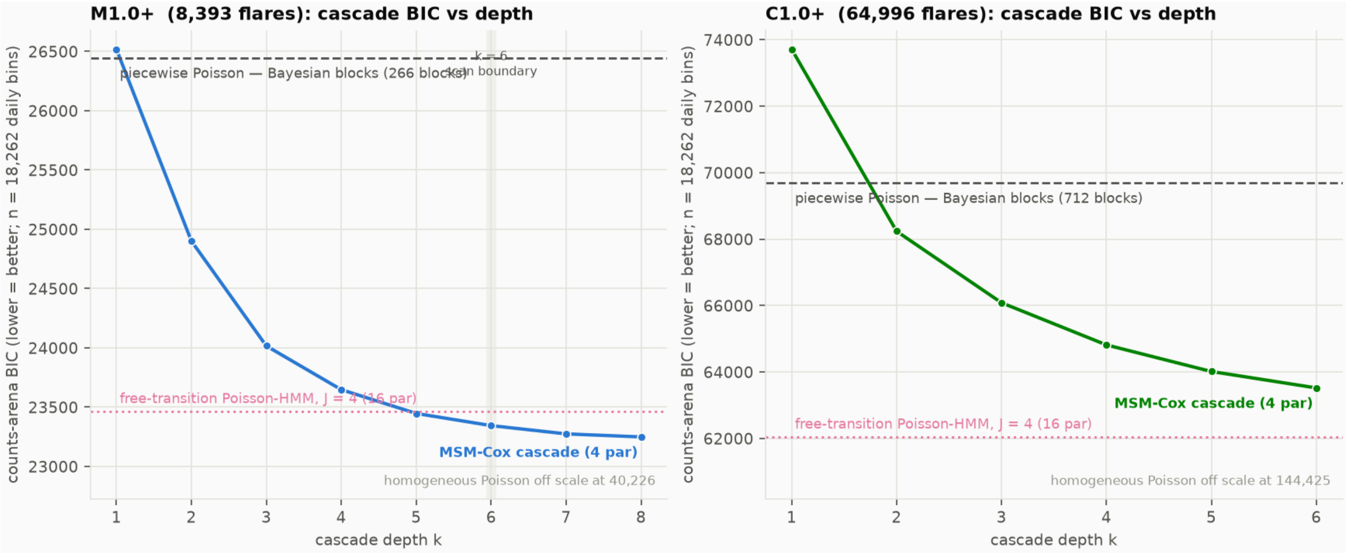
Two limits of that calibration must be stated where it is introduced, because every “the real record sits above/below the control” comparison in this paper inherits them.

- **The cycle world is a single-timescale envelope, not the fitted incumbent.** It is a smooth raised-cosine activity pulse per cycle with a fixed contrast, identical in amplitude from cycle to cycle — genuinely one slow timescale. The incumbent the tournament actually races carries 266 change points at M1.0 and 712 at C1.0, i.e. structure down to weeks, and the real Sun’s cycles differ in amplitude by more than a factor of two. So the cycle world is a *clean negative control for a smooth modulated rate*, and it is not “the piecewise-Poisson picture made synthetic”. The control that would carry that name — a parametric bootstrap simulating counts from the *fitted* 266-/712-block rate and re-running the whole tournament — is not in this record, and §5.1’s bootstrap is against homogeneous Poisson only. We name it in Appendix B as the specificity control §4 most wants.
- **The positive control is a self-recovery test.** The cascade world is generated by the same vendored module that supplies the fitter’s transition and emission machinery, so it certifies that the filter recovers *its own* generative process, not that it behaves well under some other multi-timescale mechanism. And the synthetic cells are not N-matched to the real ones (M1.0: 4,885 control / 5,559 cycle against 8,393 real; C1.0: 57,460 / 63,371 against 64,996), while ΔBIC scales with N — which is why §4.4 normalises the depth gain per flare before comparing cells.

---

4. Results I: the tournament on the real record

(Record `out/v25/`; Figure `figures/fig_flare_tournament.png`



figures/fig_flare_tournament.png

.)

4.1 The counts-arena table, and the depth result

Every entry is fit on the same 18,262-day series with the same era covariate (except the blocks entrant — §3.2). BIC is computed on the full series; the held-out log-likelihood is scored on the contiguous trailing 30 % (**5,479 days, about 15 years**) after refitting on the leading 70 %, from the model's own stationary prior.

Table 2 — The counts-arena tournament, real GOES/XRS record (record v25; BIC lower is better, held-out log-likelihood higher is better; parameter counts carried).

model	params	M1.0+ BIC	M1.0+ held-out	C1.0+ BIC	C1.0+ held-out
homogeneous Poisson	1	40,226	−5,842	144,425	−21,838
single switch (cascade, k = 1)	4	26,519	−3,623	73,715	−10,856
piecewise Poisson — Bayesian blocks	531 / 1,423 †	**26,446**	*(not reportable)* ‡	**69,693**	*(not reportable)* ‡
free-transition Poisson HMM, J = 2	4	26,104	−3,549	73,691	−10,849
free-transition Poisson HMM, J = 4	16	**23,464**	**−3,165**	**62,046**	**−9,109**
MSM-Cox cascade, k = 6	**4**	**23,344**	**−3,162**	**63,521**	**−9,420**
MSM-Cox cascade, k = 7 (deep, BIC only)	4	23,274	—	—	—
MSM-Cox cascade, k = 8 (deep, BIC only)	4	23,249	—	—	—

† 2·n_blocks − 1 for 266 blocks (M1.0) and 712 blocks (C1.0). ‡ Not reportable from this record, and not repairable by the `p7` guard fix alone — see §3.3. Every blocks comparison in this paper is on BIC alone.

The depth result. Within the cascade family, at a **fixed** four parameters, going from one switching layer to six improves BIC by ΔBIC **+3,175** (M1.0) and **+10,195** (C1.0), and improves the *out-of-sample* log-likelihood by **+460** and **+1,436** nats. Because the parameter count does not change with k, this margin is fit rather than flexibility bought with parameters, and it survives the out-of-sample test. (Nominally fixed: the k = 1 entrant carries one *inactive* parameter — see the footnote in §3.2 — so it is over-penalised by ln n = 9.81 BIC and these depth gains are lower bounds by that amount.)

That is not, by itself, evidence of hierarchy, and the sealed calibration is explicit about why. Run on the **no-hierarchy cycle world** — a single slow timescale, by construction — the same scan returns a depth gain of **+190** BIC and **+36** nats at M1.0 and **+3,607** BIC and **+648** nats at C1.0, and at C1.0 it selects k* = 6 at the scan boundary exactly as the real record does. A low-order cascade is a flexible, parsimonious model of *any* nonstationary Poisson rate; the discriminator is the **magnitude**, read against that null. §4.4 does that reading.

4.2 The incumbent margin, decomposed: parsimony, not fit

Against Bayesian blocks the cascade leads by ΔBIC **+3,102** at M1.0 and **+6,172** at C1.0. §6.1 refuses to quote a BIC margin between models of very different parameter count without decomposing it, and the same discipline applies here — to a margin twenty-six times larger.

Table 3 — The Bayesian-blocks margin, decomposed (record v25; n = 18,262 bins, so the BIC charge is ln n = 9.813 per parameter).

	M1.0+	C1.0+
blocks parameters (2·n_blocks − 1)	531	1,423
cascade parameters	4	4
BIC parameter penalty on the difference	**+5,171**	**+13,924**
twice the log-likelihood difference, 2(ℓ_cascade − ℓ_blocks)	**−2,069**	**−7,752**
total ΔBIC (cascade over blocks)	**+3,102**	**+6,172**
ℓ_blocks (nats)	−10,618	−27,865
ℓ_cascade (nats)	−11,652	−31,741
the blocks partition fits better by	**+1,035** nats	**+3,876** nats

The two terms run in opposite directions, and the fit term runs *against* the cascade. The Bayesian-blocks partition describes the record better in raw likelihood at both thresholds; the cascade is preferred because it does so with four parameters instead of 531 and 1,423, and BIC charges the difference.

We therefore do not say the cascade "out-describes" the incumbent, and the reader should not. What the comparison establishes is worth stating precisely and no more strongly: **a four-parameter hidden-state model comes within about a thousand nats of a 531-parameter piecewise-constant partition over 18,262 days, and BIC prefers it by a wide margin.** A parsimony ratio of that size is a real result — it says the record's rate variation is compressible into a handful of latent switching layers — but it is a statement about description length, not about goodness of fit. It is also softened further by the two accounting effects of §3.2 and §3.3 that run against the incumbent, and it rests on BIC alone, since the incumbent's held-out column is unusable in this record.

4.3 Depth: the scan is right-censored, and the span drifts with it

The BIC is still descending at the edge of the scan. At M1.0 the deep extension gives k = 7 at 23,274 and k = 8 at 23,249, both below the k* = 6 value of 23,344 — so **k* = 6 is a scan boundary, not an optimum**, and the ladder the tournament reports is a lower bound on the depth the record supports. The held-out score is computed only over k = 1…6 and is still improving at k = 6 as well, though by a decelerating amount (+38 nats from k = 4 to k = 5, +18 from k = 5 to k = 6).

The censoring has a consequence for §7 that we flag here: the *fitted span* drifts upward with depth, S = 143.6 at k = 6 → 161.3 at k = 7 → 270.4 at k = 8. A ladder span read off a right-censored scan is a depth-conditional quantity, and we quote it as one.

4.4 The real record read against both synthetic worlds

Every claim in §§4–6 is a comparison between the real record and two sealed worlds, so the honest way to present it is one table with all six cells. ΔBIC scales with N and the cells are not N-matched (§3.4), so the depth gain is also given per flare.

Table 4 — The six calibration cells (records v16 and v25, counts arena; each margin positive in the cascade's favour; the blocks and HMM margins are BIC).

cell	N flares	k*	depth gain BIC	per flare	depth gain held-out	over blocks	over free HMM (BIC)	over free HMM (held-out)	fitted span *S*
M1.0 cycle world (no hierarchy)	5,559	4	+190	0.034	+36	+152	**+147**	**+10.3**	**1.00**
M1.0 cascade world (control)	4,885	6	+1,304	0.267	+326	+1,595	**+221**	**+38.8**	219.55
M1.0 REAL	8,393	6	**+3,175**	**0.378**	**+460**	**+3,102**	**+119**	**+2.17**	**S ≈ 144**
C1.0 cycle world (no hierarchy)	63,371	6	+3,607	0.057	+648	+200	**+461**	**+68.1**	1.00
C1.0 cascade world (control)	57,460	6	+23,373	0.407	+4,068	+11,538	+3,892	**+561.7**	136.77
C1.0 REAL	64,996	6	**+10,195**	**0.157**	**+1,436**	**+6,172**	**−1,475**	**−310.83**	102.24

Reading the table across, and no further:

- **On the depth axis the real record is far above the no-hierarchy world at both thresholds** — 11× its per-flare depth gain at M1.0 and 2.8× at C1.0 — and at M1.0 it is above the positive control as well (0.378 against 0.267 BIC per flare). At C1.0 it sits between the two worlds: 2.8× the null's per-flare gain, and 0.39× the positive control's. Well clear of the null, and well short of the control.
- **On the blocks axis the same ordering holds**: +3,102 against a null-world +152 at M1.0 (above the control's +1,595), and +6,172 against a null-world +200 at C1.0 (below the control's +11,538). Given §4.2 this is a parsimony ordering, and it separates the worlds cleanly.
- **On the span axis the real record sits between the worlds at both thresholds** — 144 and 102 against a collapsed 1.00 and a control 220 / 137. Far from the null, below the control. §7 says what that does and does not license.
- **On the free-HMM axis the real record sits *below the no-hierarchy world* at both thresholds and on both criteria.** That axis is §6, and it is the one place in this table where the real record fails against a synthetic world with no hierarchy in it at all.
- **k* = 6 is selected in five of six cells**, including the C1.0 no-hierarchy world. Depth selection alone carries no information here.

4.5 What the tournament settles

Two things, stated at the strength the table supports. **Several latent switching layers beat one**, decisively, at both thresholds, in-sample and out-of-sample, at fixed parameter count, and by a margin that is an order of magnitude above what a smooth single-timescale rate produces under the same scan. And **a four-parameter hidden-state description is preferred by BIC over the field's piecewise-constant representation** — on parsimony, by §4.2, not on fit, and on BIC alone.

What the tournament does **not** settle is the *form* of that latent structure. §6 is about exactly that, and the answer there is negative.

5. Results II: significance, specificity, and the recovered cycle

5.1 The pure-Poisson parametric bootstrap

(Record `out/v25/`.)

A ΔBIC needs a null distribution. We simulate 100 replicate count series from a **pure homogeneous Poisson** process matched to the real record, fit both the Poisson null and the cascade to every replicate at k = 3, and build the null distribution of ΔBIC(Poisson − cascade). Under the null the cascade is *penalised*, as it should be: the null mean is −28.9 and the 95th percentile −26.4. The observed statistic on the real record at the same depth is **+16,209**, and no replicate comes near it, so the empirical **p = 0.0099** — the `(1 + Σ)/(n + 1)` floor at N = 100, which is the smallest honest statement an ensemble of that size can make, not a literal zero.

Two scope notes travel with this, and the second is the important one. The bootstrap is run at k = 3, so its observed statistic is the k = 3 depth gain over homogeneous Poisson, not the k = 6 one (16,881). And it rules out *homogeneous* Poisson, which is the weakest entrant in the field — it establishes that the cascade's edge is not an artefact of the exact filter, and nothing more. The bootstrap this section would most like to report is the one against the *fitted piecewise-constant rate* — simulate Poisson counts from the fitted 266-/712-block λ(t), refit both entrants, and build the null distribution of the depth gain and of the blocks margin. That control is not in this record. It is named in Appendix B, and until it exists §4.4's two-world contrast carries the specificity claim on its own.

5.2 The specificity control: a smooth cycle can fake volatility clustering

A naive test for multiplicative structure — the autocorrelation of the magnitude of log-count residuals — **false-positives on a purely smooth, strongly modulated Poisson rate** when it is scored against a shuffle null, because rate heteroscedasticity alone produces residual-magnitude correlation. The ~11-year solar cycle is precisely that kind of modulated rate, so this is not a hypothetical hazard for flare statistics; it is the first thing a hidden-structure claim on this record has to clear.

The sealed cell measures it on both synthetic worlds and on the real record, at both thresholds, against two nulls: a shuffle null (the naive detector's own null) and a rate-matched smooth-Poisson null (the correct one).

Table 5 — The residual-specificity cell (record v25; *S* is the mean autocorrelation of |residual| over lags 1–30; each *p* from 300 surrogates, so 0.003 is the ensemble floor; ✗ = does not detect, ✓ = detects).

threshold	world	naive S	naive vs shuffle	naive vs smooth-Poisson	Pearson S	Pearson vs shuffle	Pearson vs smooth-Poisson
C1.0	cycle (no hierarchy)	0.124	**✓ false positive**	✗ (p = 0.887)	0.0023	✗ (p = 0.056)	✗ (p = 0.080)
C1.0	cascade (hierarchy)	0.170	✓	✓	0.183	✓	✓
C1.0	**real record**	0.293	✓	✓	**0.208**	**✓**	**✓ (p = 0.003)**
M1.0	cycle (no hierarchy)	0.553	**✓ false positive**	✗ (p = 0.824)	0.041	**✓ false positive**	✗ (p = 0.269)
M1.0	cascade (hierarchy)	0.639	✓	✓	0.100	✓	✓
M1.0	**real record**	0.492	✓	✓	0.185	✓	✓

Specificity ratios *S*(cascade)/*S*(cycle): at C1.0, **1.4×** naive against **81×** Pearson; at M1.0, 1.2× naive against 2.4× Pearson.

Three readings, in the order the table supports them.

- **Half the fix is the null, not the statistic.** Scored against the *correct* smooth-Poisson null, the naive log-count residual already returns the right negative on the cycle world at both thresholds (p = 0.887 and 0.824). It is only against the shuffle null — the null a practitioner reaches for first — that it fires with no cascade present. Any report of residual volatility clustering on a cycle-modulated flare record that uses a shuffle null is reading heteroscedasticity.
- **The other half is the statistic, and it buys separation only at C1.0.** The Pearson residual collapses to near-white on the cycle world and retains the clustering on the cascade world: a specificity ratio of **81×** at C1.0 against the naive statistic's **1.4×**. At the sparse M1.0 rate (~0.46 counts/day, mostly zeros and ones) residual quantisation leaves the ratio at 2.4× — and the Pearson residual itself false-positives on the cycle world under the shuffle null there. **This cell is clean only at the sensitivity threshold**, and we do not claim it at the primary one.
- **On the real record the Pearson clustering survives the correct null** (p = 0.003 against rate-matched smooth Poisson at both thresholds): there is structure beyond the smooth-cycle heteroscedasticity.

Two further qualifications. The real-record cell uses the *documented* solar-cycle envelope as its smooth model rather than a rate fitted to the data, so what survives is "structure beyond the documented envelope" — corroborative, not a self-contained proof; the controlled two-world cells carry the specificity claim. And **the tournament reads neither residual**: the counts-arena verdict is computed from the exact 2^k-state likelihood with the era covariate, and no raw log-count residual enters it at any point. The cell is reported because the artefact is real and a reader is entitled to know it was checked, not because the verdict depends on it.

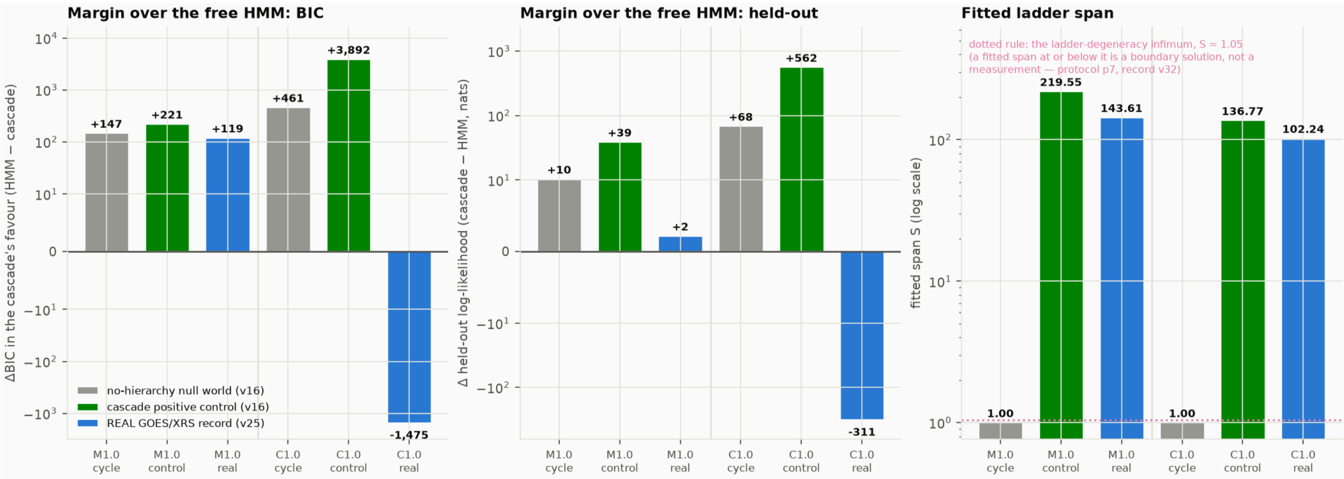
5.3 The smoother recovers the solar cycle — a sanity check, and only that

The posterior activity path — the smoothed expectation of the latent ladder over the 50-year record — carries adjusted mutual information AMI **0.323** with solar cycle phase, exceeding all 200 time-shuffle draws (the reported p = 0.005 is the 1/201 ensemble floor, not a measured value). The hidden state the filter infers from flare counts alone tracks the cycle it was never told about.

We report this as a sanity check and give it no evidential weight, for two reasons that a reader should have in front of them. The null is an **i.i.d. permutation** of the cycle-phase labels, and both series are strongly autocorrelated, so the null AMI is ≈ 0 by construction (null mean -1×10^{-5}, s.d. 6×10^{-5}) and the test cannot fail for any autocorrelated pair; an autocorrelation-preserving null — a circular shift or a block permutation — is what would make it a test. And "cycle phase" here is the *documented* raised-cosine envelope built from tabulated minima and maxima, not an observed index, so the statement is that tertiles of the posterior track a smooth function of time built from ten dates. Re-running against a circular-shift null and against an observed index (sunspot number or F10.7) is named in Appendix B. The value of the check as it stands is negative: a filter that recovered *nothing* recognisable would have failed here, and it did not.

6. Results III: the honest boundary — the free-transition Poisson HMM

(Records `out/v25/` , `out/v16/` , `out/v20/` *; Figure* `figures/fig_flare_boundary.png`



figures/fig_flare_boundary.png

.)

The cascade's structure is a *constraint*: k independent two-state layers, mirror multipliers, switching rates on a geometric ladder. The free-transition Poisson HMM has the same discreteness with none of the constraint — J Poisson rates and an unconstrained transition matrix. If the geometric ladder is the right description of the flare rate, the constraint should pay: the cascade should describe the record about as well with far fewer parameters, and should *generalise better*, because a correct constraint is exactly what stops a model over-fitting.

It does not pay. This section says so precisely, and then says precisely how little that means.

6.1 The M1.0 margin is parsimony, not fit

At the primary threshold the cascade leads the best free HMM (J = 4) by **+119** BIC. Decomposed at n = 18,262 bins:

component	value
BIC parameter penalty on the HMM's 12 extra parameters, (16 − 4)·ln(18,262)	**+117.8**
twice the in-sample log-likelihood difference, 2·(ℓ_cascade − ℓ_HMM)	**+1.7**
total ΔBIC in the cascade's favour	**+119.4**

The in-sample log-likelihood difference is **+0.83** nats — over 18,262 days carrying 8,393 flares. The two models fit the record *indistinguishably*; the entire visible margin is the parameter count. Quoting +119 without this decomposition would overstate the case, and we do not. (This is the same decomposition §4.2 performs on the incumbent margin, where it comes out the other way.)

6.2 Read out of sample, at both thresholds

The honest criterion here is the held-out score, which charges flexibility where BIC's fixed penalty cannot. On the trailing 5,479 days:

Table 6 — Cascade against the free-transition Poisson HMM (J = 4) (record v25; every Δ positive in the cascade's favour, so the BIC column is BIC_HMM − BIC_cascade and the log-likelihood columns are ℓ_cascade − ℓ_HMM).

threshold	flares	ΔBIC in the cascade's favour	Δ in-sample log-lik (nats)	Δ held-out log-lik (nats)
M1.0+	8,393	**+119.4**	+0.83	**+2.17**
C1.0+	64,996	**−1,475.1**	−796.4	**−310.83**

At M1.0 the cascade's out-of-sample margin over a model with four times its parameters is **+2.17** nats across fifteen years of held-out days: a tie by any reading. At C1.0, where the record carries 7.7× more flares and correspondingly more resolving power, the free HMM wins on both criteria and by a wide margin — the cascade loses by **−1,475** BIC and by **−310.83** nats out of sample.

So at M1.0 the constraint buys nothing and at C1.0 it costs a great deal. Read strictly as a description contest at these sample sizes — which is all this axis supports, for the reason §6.3 gives — **the geometric parameterisation is not the better description of these data.**

6.3 This axis cannot be read as a hierarchy test

It would be easy, and wrong, to read the C1.0 loss as evidence against hierarchy in the flare rate. One measurement forbids that reading outright, and a second suggests the sign is sample-size-dependent without settling it.

(i) The comparison fires positive where there is no hierarchy at all. In the synthetic **cycle world** — a smooth modulated Poisson rate with no cascade underneath — the cascade still beats the free HMM, by ΔBIC **+147** (M1.0) and **+461** (C1.0), and out of sample by **+10.3** and **+68.1** nats. A test that returns a positive result on its own negative control is not a test of what it appears to test; it is measuring parsimony against a flexible rival on smooth data. The positive control's margins are larger (**+221** BIC and **+38.8** nats at M1.0; +3,892 and **+561.7** at C1.0), so the *magnitude* carries some information — but the axis has no zero, and Table 4 shows the real record sitting **below the no-hierarchy world at both thresholds on both criteria**. Whatever that ordering means, it is not "hierarchy absent"; it is "the geometric parameterisation is a worse description of these data than a free one, more so than on smooth synthetic data."

(ii) The sign of this comparison can reverse with N even on genuine-cascade truth — sometimes. The sealed event-altitude estimability grid (record `v20`) injects a genuine six-layer ladder and sweeps N from 50 to 5,000 events at six seeds each. On its canonical cascade track at uniform exposure the cascade's rate of out-BIC-ing the free HMM sits between 0.50 and 1.00 across N = 50–800 and then **falls to 0/6** at N = 1,600, 3,200 and 5,000 — on data whose truth is a cascade by construction.

We report that as a caution and not as a defence, because the record does not support the stronger reading and three things in it cut against generalising:

- **It is one track of four.** On the *sessions*-exposure track of the same injected truth the rate at those same N is 0.50 / 0.67 / 0.50, and on the strong-amplitude track 0.50 / 0.50 / 0.50 — no reversal on either. Only the weak-amplitude track also collapses. With six seeds a rung, none of these rates is well resolved.
- **The mechanism v20 records is absent from the real cells.** v20 attributes its reversal to a *collapsed-ladder under-fit*: on two of the three reversal rows the fitted span sits on its infimum (S = 1.000 at N = 3,200 and 5,000; S = 2.34 at N = 1,600), in every case far below the injected 40, so the cascade under-fits and the flexible HMM competes in-sample. *(Record v35 identifies what that under-fit is: in the selection race the cascade is deliberately not given the generator's 16-bin activity window — the periodic-window entrant is the rival being raced — so its ladder is spent absorbing a deterministic comb instead of the injected hierarchy. That is the correct design for the race, and it is why the fitted span collapses there; it is also why those collapsed spans are not estimability statements. See §7.1 rule 2.)* The real cells are interior (143.6 and 102.2, §7.2), so the pathology that produces v20's 0/6 does not obtain here. At the one sealed anchor where a broad ladder *was* recovered — a fast repeater stream at N ≈ 1,641 with S ≈ 1,972 — the cascade does beat the HMM.
- **It is a different arena, and §3.1 forbids carrying results across arenas.** v20 is the sparse-event altitude, with its own injected operating point, its own null field, its own exposure and its own N → span mapping; it shares nothing with the daily-counts flare tournament but the shape of the comparison.

So what (ii) establishes is only this: **a free-HMM win at large N is a thing this comparison does on data that is genuinely a cascade, at least under some conditions — so a free-HMM win is not by itself evidence about the generating process.** The mechanism is not mysterious, and Table 6 exhibits it directly on the real record: at M1.0 the HMM's twelve extra parameters buy essentially no fit (+0.83 nats), while at C1.0, with 7.7× the flares, they buy 796 — enough to pay their penalty several times over. At large N a free transition matrix has the data to afford its parameters, and the geometric constraint stops being worth its parsimony. What would settle the question is a counts-arena, cycle-modulated, era-covariate grid at the real cells' N and operating point, with non-degenerate fitted ladders. That grid does not exist; we name it in Appendix B.

Put together, and no more strongly: **the threshold dependence of the cascade-vs-HMM verdict is consistent with a resolving-power effect and is not evidence about the Sun.** At M1.0 the record cannot tell the two apart; at C1.0 it can, and it prefers the free transition matrix.

6.4 Why the direction is the interesting part

The program has raced this same adversary at two other altitudes, one on real data and one synthetic, and the flare record does not behave like the first.

- **Daily X-ray binaries, real Swift/BAT** (record `v14`): a free 5-state Gaussian HMM (30 parameters) beats the cascade *in-sample* on Cyg X-1 by ΔBIC 1,851, and a Markov-switching DRW beats it on GRS 1915+105 by 1,536 — yet the cascade beats both **out of sample** (+296 and +184 nats against the free

HMM). There, the constraint generalised: generic discreteness over-fit and the ladder did not. Note that even in that arena the fitted ladder is a boundary solution on Cyg X-1 (S = 1.0000012, on the infimum), so the win is not accompanied by a measured hierarchy there either.
- **Stellar flares, synthetic, calibrated to GJ 1243** (record `v31`): on data *generated by* a published flat three-state hidden Markov model, the geometric cascade returns 0/12 false cascade-wins on held-out log-loss above a decidability floor of ≈ 1,600 flares. This is a specificity result on a synthetic control world — the flat HMM is the better description *by construction of that world* — not a measurement on a real M dwarf, and we weight it as such.
- **Solar flares, real, here**: the cascade neither generalises better (M1.0) nor describes better (C1.0).

What organises these is not whether a stream is multi-state — each of them is — but whether a *geometric ladder* is the right parameterisation of that multi-state structure. A reading, and we mark it as one: a geometric ladder may be the better idealisation where fluctuations are injected across a continuum of scales, and a free transition matrix the better one where the generator is a discrete population of active regions emerging, evolving and decaying. Nothing here measures either mechanism, and §1.3 forbids claiming one. What *is* a finding is the model-comparison result — a free transition matrix over a handful of activity states is at least as good as the cascade at M1.0 and better at C1.0 — and it points at the model a solar audience should probably be fitting.

6.5 What survives

Nothing in §6 changes a number in §4: the depth gain is measured *within* the cascade family at fixed parameter count, and the incumbent margin is measured against Bayesian blocks. Table 4 shows how the real record sits against both synthetic worlds on each axis, and the pattern is consistent — far above the no-hierarchy world on depth, blocks and span at both thresholds, above the positive control on depth and blocks at M1.0, between the worlds elsewhere, and below the no-hierarchy world only on the free-HMM axis, which §6.3 has just established is not a hierarchy test.

§6 does bear on the *name*. What §4 establishes is that several latent switching layers beat one and that a four-parameter latent description beats a piecewise-constant one on parsimony. At C1.0 the entrant that describes the record best is a free four-state HMM, which is multi-state without being a timescale hierarchy. **“Multi-state latent structure” is therefore the safe reading of §4, and “geometric hierarchy” is not** — which is why this paper’s title says what it says.

7. Results IV: the fitted ladder, under the degeneracy discipline

(Records `out/v25/`, `out/v20/`, `out/v32/`; Figure `figures/fig_flare_boundary.png`, right panel.)

7.1 The discipline, stated before the numbers

A fitted ladder is the most quotable and the most abusable thing this model produces, and this program has measured three separate ways it misleads. We state the rules first and then obey them.

1. **A span is only ever an excess over its raced rivals.** An MSM-Cox ladder fitted to single-mechanism near-critical self-exciting truth — data with no hierarchy whatsoever — returns a span of 1.14 decades, 87 % of a genuine multi-scale cascade’s 1.31, demonstrably off its initialisation. A span is therefore never evidence of scale hierarchy until the self-excitation alternative has been raced at the actual N, and it is quoted in the same sentence as that race.

2. **Selection is not recovery — and a recovery failure is a statement about the likelihood, not about N.** The event-altitude floor certifies that regime *structure exists* from **N*_selection ≈ 800** events at m_hi = 1.8 (≈ 400 at 2.2 on the real-anchored grid; off-grid at 1.5). That is a **model-selection** floor and it is what this paper uses. It certifies nothing on its own about parameter recovery, and the ladder is treated here as a selection quantity throughout.

This rule was rewritten on 2026-08-30, and the rewrite removes a claim rather than adding one. Earlier drafts stated, on record v32, that the ladder is recovered to truth *nowhere in the grid* — that the fitted multiplier runs "+45 % to +86 %" high with a bias that never shrinks with N, and that **61 of 72** sealed cascade track-rows return a span on its own parameter infimum. Record **v35** (`out/v35/`) shows that reading was a **model-misspecification artifact**. The sealed grids generate the cascade world on `exposure × activity_window` — a hard-gated 16-bin comb, exactly zero in ~70 % of bins — but fit every cascade entrant with `covariate = exposure` alone. That omission is *correct* for the selection race, where the periodic-window entrant is the rival being raced, and *wrong* for a truth-referenced recovery readout, where it forces the six-layer ladder to absorb the comb. Refitting the same sealed seeds with the generator's own covariate recovers the injected multiplier **inside** the grid — on the canonical track from N = **100**, an order of magnitude *below* N*_selection, with **6/6** seeds recovering it at N = 1,600 where the covariate-free fits recovered 0/6. **So `N*_recovery is above the grid` is withdrawn**, and with it the transported multiplier-bias figure; the `61 of 72` census stands only as a property of the covariate-free fits, not as a statement about estimability.

The withdrawal is scoped, and the scope is part of the claim. It is withdrawn **over the three re-analysed tracks** — `v20 | trivial` (N ≤ 1,600), `v28 | transit` and `v28 | trivial` (N ≤ 800), **canonical injected amplitude only**. The weak and strong amplitude brackets and the `v20 | sessions` track were **not** re-fit, and their v32 rows stand as read under the covariate-free likelihood. Nothing here claims a number for a cell record v35 did not fit, and this paper quotes none.

Two readings of the corrected floor, and this paper reports both (added 2026-08-31). The floors above count every fitted cell. But rule 3's boundary flag — which v35 retains — says that on S ≤ 1.05 the fitted (m_hi, S, f_slow) are **jointly** not reportable, and m_hi is one of them, so a flagged cell cannot be counted as having recovered the multiplier either. Excluding the flagged fits from both numerator and denominator and asking the same ≥ 5/6 question of what remains gives, beside the headline **100 / 200 / 140** (v20 trivial, v28 transit, v28 trivial), a degeneracy-conditioned **100 / 400 / 140** at first crossing, or **200 / 400 / 400** if the rule is additionally required to hold at every larger re-analysed N. Two conditioned columns because the conditioned series is not monotone: v20 | trivial meets the rule at 100 (3/3 of its non-degenerate fits) and loses it at 140 (3/4), and v28 | trivial meets it at 140 (4/4) and loses it at 200 (2/3). The single unambiguous move is **v28 | transit, 200 → 400**, whose headline 5/6 at N = 200 is 4/5 once its one flagged fit — itself one of the five successes — is set aside. **The decisive cell is clean either way:** the canonical N = 1,600 cell carries zero flagged fits, so its 6/6 is 6/6 under both readings, and even the worst conditioned floor (400) remains well below N*_selection ≈ 800. Nothing this paper uses turns on the choice; the conditioned column is quoted because the conservative floor is the one a reader should carry.

What survives, and what this paper still obeys: `N*_selection` is untouched; the boundary flag of rule 3 is untouched; and the ladder is *still* a selection quantity here — for the independent reason §6.5 gives, that a free transition matrix describes these data at least as well, so the geometric parameterisation is not the preferred description whatever its estimator can recover. 3. **A boundary solution is not a measurement.** The span is fitted through S = 1 + exp(u), so S ≤ 1.05 is the u → −∞ boundary, and m_hi ≥ 10^4 is the flat $\lambda_0 \cdot m^k$ ridge. On either boundary the fitted (m_hi, S, f_slow) are not reportable; the flag is surfaced instead of the numbers.

7.2 What the real record's ladder is, and what it licenses

Both real cells are **interior**: the ladder-degeneracy flag does not fire at either threshold (M1.0, S ≈ 144; C1.0, S = 102.2), unlike the program's real repeater-FRB cell, where a fitted S = 1.00 is a boundary solution and is reported as one. So the fitted spans here are quotable — as fitted values, under rule 2.

Table 7 — Fitted ladders (real rows: record v25, counts arena, k = 6; control rows: record v16, each at its own selected k* — the cascade worlds at k = 6, the M1.0 cycle world at k = 4).

cell	N flares	k*	m̂_hi	fitted span *S*	slowest rung	fastest rung	degeneracy flag
real M1.0+	8,393	6	2.11	**S ≈ 144**	699.7 d	4.87 d	no
real C1.0+	64,996	6	1.63	102.2	564.8 d	5.52 d	no
v16 M1.0 cascade world (control)	4,885	6	1.82	219.5	796.0 d	3.63 d	no
v16 M1.0 cycle world (no hierarchy)	5,559	4	1.29	**1.00**	—	—	on the infimum
v16 C1.0 cascade world (control)	57,460	6	1.69	136.8	461.6 d	3.37 d	no
v16 C1.0 cycle world (no hierarchy)	63,371	6	1.20	1.00	—	—	on the infimum

Read under the rules:

- **Against the calibration, the fit does not collapse.** The no-hierarchy cycle world returns S = 1.00 at both thresholds — on the infimum, i.e. the fit declines to spread at all — while the real record returns 144 and 102 and the positive control returns 220 (M1.0) and 137 (C1.0). This is the axis on which the two synthetic worlds are furthest apart, and the real fit sits between them, far from the collapse. We stop there, for two reasons. Rule 1 requires the span to be quoted beside its raced self-excitation alternative, and §9 records that this alternative is degenerate on this record — so the race the rule demands does not exist, and the span's weight rests entirely on the counts-arena rivals of §§4 and 6.
- **Against the selection floor, both cells clear it on N — and here is the constant's uncertainty, which no earlier draft attached.** The sealed grid reaches its model-selection floor at **N*_selection ≈ 800** events for the canonical injected amplitude. That number is the first grid point at which a **six-seed** panel crosses the ≥ 5/6 rule, on a rate curve that is strongly non-monotone (0.67 at N = 100, falling to 0.17 at N = 140, back to 0.67 at 200) and whose crossing turns on a single seed flip. The Clopper–Pearson 95 % interval on 5/6 is **[0.36, 1.00]**; the adjacent grid points read 0.50 at N = 400 and 1.00 at N = 1,600, so the honest statement is a **grid bracket, N* ∈ [400, 1,600]**, and a deeper panel could move the point estimate a grid step either way. With 8,393 and 64,996 flares both real cells are one to two decades past the *top* of that bracket, so the clearance does not depend on where inside it the constant falls.
- **The clearance is carried by N, not by the multiplier estimate.** Earlier drafts also placed the two cells on the grid's *amplitude* axis, by de-biasing m̂_hi with a "+45 % to +86 %" figure read off the sparse event-altitude grid, and concluded that both sit at or below the weak track whose floor is off-grid. **That de-biasing is withdrawn (2026-08-30)**, for two independent reasons. *First, wrong calibration:* the transported bias was measured under a different observation process — a sparse event stream through a hard-gated 16-bin activity window — while these cells are dense daily counts. The observation-process-matched calibration is the v16 cascade world, run through the same fitter, the same κ = 1 and the same k-scan as these cells, and there the fitted multiplier is essentially **unbiased**: **1.819** against an injected **1.700** (**+7.0 %**) at M1.0, and **1.694** against 1.700 (**−0.4 %**) at C1.0. A +45–86 % correction there is a six- to twelve-fold over-correction. *Second, the bias itself does not stand:* per rule 2 and record `out/v35/` it is a model-misspecification artifact, not a measured property of the estimator at N. Read with the matched calibration

the de-biased multipliers are ≈ **1.97** (M1.0) and ≈ **1.64** (C1.0) — at or *above* the canonical 1.8 track rather than below the weak one. We do not lean on that either. **The floor clearance rests on N alone**, which is what the bullet above already carries and all this paper needs.

- **And clearing that floor licenses existence, not the ladder.** What N*_selection ≈ 800 certifies is that regime *structure* is decidable at this N. So what each cell delivers is a statement that latent structure is decidable here — and not a measurement of the ladder.
- **The span is still a model-selection quantity, not a recovered parameter** — now on §6.5's ground rather than on rule 2's withdrawn census. A free transition matrix describes these data at least as well, so the geometric parameterisation is not the description this record prefers: "S ≈ 144" means *the ladder this likelihood prefers*, not *the Sun's hierarchy spans a factor 144*, and $\hat{m}$_hi = 2.11 is not read as a physical contrast at all.
- **And the span is depth-conditional** (§4.3): 143.6 at k = 6, 161.3 at k = 7, 270.4 at k = 8, on a scan whose BIC is still descending. The quoted value is the one at the scan's operating point.

7.3 The fit does not turn on the era covariate

Refitting without the era covariate moves the span from **143.6** to **143.0** and the multiplier from 2.110 to 2.108. The model-selection outcome is therefore essentially covariate-invariant.

That is a robustness statement about the fit, and it is not evidence about what the latent structure is made of — the span is a selection quantity (rule 2), and rule 1 forbids reading it as hierarchy evidence without its raced rival. It is also an ablation of one known step rather than a race against a step-artefact generator, and §2.2 has already conceded that the covariate under-corrects the measured 1.73 flux-median ratio, so "with versus without" brackets neither the true correction nor zero. What it does establish is bounded and worth having: the ~38 % modern-era step, the one instrumental feature with the size and sharpness to matter, does not move the fit. The covariate stays in regardless — dividing out a known instrumental step is the honest treatment whether or not it changes the answer.

7.4 The rung timescales, and what is deliberately not claimed

Under the model the k rungs are geometric, $f_i = f_{slow} \cdot S^{(i/(k-1))}$, so the M1.0 ladder's six rungs sit at **699.7, 259.1, 95.9, 35.5, 13.2 and 4.87 days**.

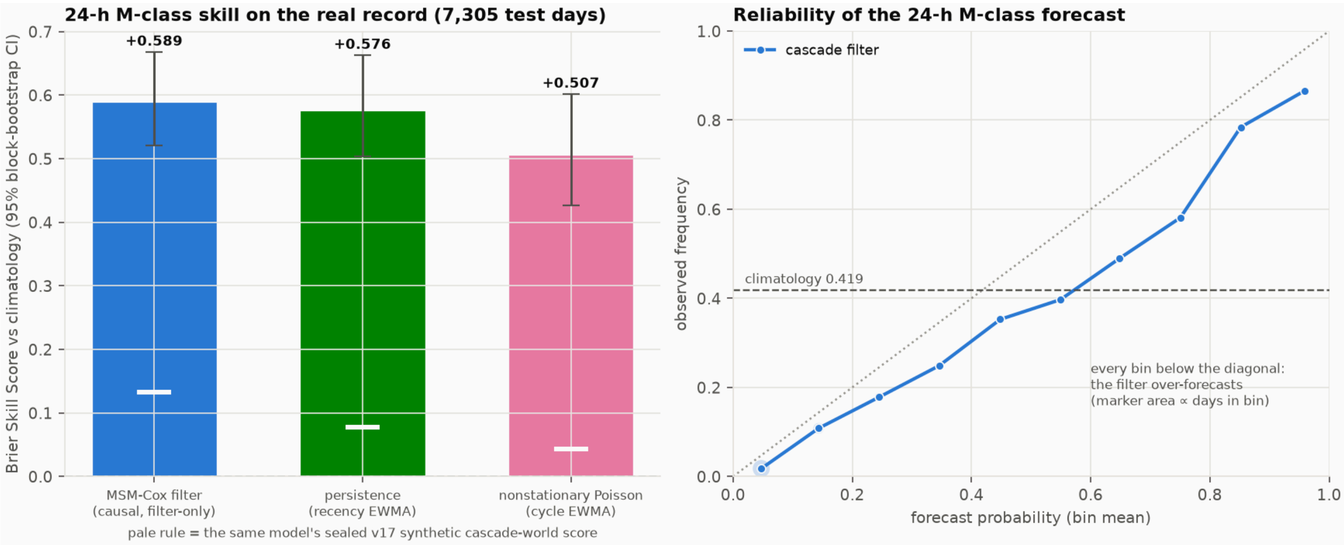
Those numbers land, unprompted, in the neighbourhood of timescales solar physics already names: active-region emergence and the days-to-weeks life of small regions, the 30–60-day disappearance of spots and the multi-month lifetimes of large regions (van Driel-Gesztelyi & Green 2015); activity complexes maintained over three to six solar rotations by repeated flux emergence in the same longitude belt (Gaizauskas et al. 1983); and, at the top of the ladder, the quasi-biennial band of solar activity, whose period varies over roughly 0.6–4 years and which is a ubiquitous feature of solar activity distinct from the 11-year cycle (Bazilevskaya et al. 2014).

We do not claim that correspondence, and the reasons are unchanged by the 2026-08-30 correction to rule 2 — none of them was ever the recovery claim. The rungs are geometric **by construction rather than by measurement**; the interior rungs are interpolated by the model's own parameterisation and not separately estimated; and §6 has just established that a free transition matrix describes these data at least as well — which means the rung positions are not even the preferred description's parameters. (What *has* changed: earlier drafts led this list with "the ladder is not recovered to truth anywhere in the calibration grid", read off record v32. That reading is withdrawn — see rule 2 and record `out/v35/` — so it is removed here rather than left standing as a reason it no longer is. The three reasons that remain are sufficient on their own, and the conclusion is the same one.) Reading physics off this ladder would require an experiment this paper does not run: a fit with the rung rates *freed*, scored out of sample against the geometric constraint, on per-active-region streams where the candidate mechanisms are separable. We name that as the measurement that would settle it, and we register it as follow-up rather than reporting it as a result. Note also that the fitted ladder tops out near two years — well short of the ~11-year cycle, which the filter nevertheless tracks in its posterior (§5.3) because the cycle enters as slow amplitude modulation rather than as a rung.

---

8. Results V: the causal now-cast, its reference forecast, and its ceiling

(Record `out/v26/`, against the sealed synthetic calibration `out/v17/`; Figure `figures/fig_flare_nowcast.png`



figures/fig_flare_nowcast.png

.)

8.1 Design, and what the reference forecast is

The fitted filter is turned into a **causal, walk-forward, filter-only** 24-hour M-class flare probability: the forecast issued for day t uses only counts strictly before t, propagated through the filter's own predictive distribution. There is no smoother anywhere in it — the posterior path of §5.3 is explicitly *not* used, since it conditions on the future. The filter is trained on the leading 60 % of the record and scored on the trailing **7,305 days** (twenty years), with 95 % intervals from a 1,000-replicate block bootstrap at 90-day blocks. The metrics are the standard ones — Brier Skill Score against a reference forecast (Brier 1950; Jolliffe & Stephenson 2012) and the True Skill Statistic (Woodcock 1976; Bloomfield et al. 2012).

The reference forecast is a training-period climatology, and it is badly calibrated on the verification period. The record's "climatology" is the constant probability $1 - \exp(-\bar{\lambda}\_train) = \mathbf{0.419}$, built from the *training* block's base rate of 0.542 M-class flares/day and held fixed across the test block. The verification period's own frequency of days carrying at least one M-class flare is **0.144** — the training-period reference over-forecasts the period it is scored on by 2.9×. Two mechanisms compound to produce that gap. The first is solar: the leading 60 % of the record spans cycles 21–23 and the trailing 40 % spans the much weaker cycles 24 and 25. The second is the Poisson conversion itself: for any mixed-Poisson or Cox process, $P(N \geq 1) = 1 - E[e^{(-\Lambda)}] \leq 1 - e^{(-E[\Lambda])}$ by Jensen's inequality, so $1 - \exp(-\bar{\lambda})$ is a strict upper bound on the frequency of ≥ 1-flare days whenever the rate is clustered — which is exactly what this paper measures it to be. The reference would over-forecast on the training block too.

That matters for how the headline number reads, so we give both. Every Brier score below is the sealed one; only the reference changes.

Two baselines are scored on the identical test set: **recency persistence** (a short EWMA of recent flare counts) and a **cycle-timescale rate** (a 365-day EWMA). The second is the paper's proxy for the nonstationary-Poisson picture, and a scope note belongs with it: Wheatland's (2004) event-statistics forecaster estimates the *current* rate from an active region's recent flares, so it is structurally closer to the recency baseline than to the 365-day EWMA. We therefore call the 365-day EWMA what it is — a cycle-timescale climatological rate — and do not present a win over it as a win over the published forecaster.

8.2 Skill from the event record alone

Table 8 — 24-h M-class now-cast on the real record (record v26; 7,305 test days; Brier scores sealed; BSS against two references — the training-period Poisson climatology the record uses, and the verification period's own sample climatology, derived here from the sealed reliability bins).

model	Brier score	BSS vs training climatology	95 % CI (training ref.)	BSS vs verification climatology	peak TSS
MSM-Cox filter (causal)	0.0816	**+0.589**	[0.521, 0.667]	**+0.338**	0.621
persistence (recency EWMA)	0.0842	**+0.576**	[0.504, 0.662]	**+0.317**	**0.633**
cycle-timescale rate (365-d EWMA)	0.0980	**+0.507**	[0.427, 0.601]	**+0.205**	0.526
reference (training climatology, p = 0.419)	0.1987	0.000	—	−0.611	0.018
reference (verification climatology, p = 0.144)	0.1234	+0.379	—	0.000	—

- **The filter carries real skill from the event list alone.** Against the reference the record uses it is BSS **+0.589**; against the verification period's own climatology — the reference a space-weather reader will assume — it is **+0.338**. Both are far from zero; the second is the one to quote in a forecasting context, and the first is skill against a forecaster frozen at the training rate.
- **The ordering is identical under both references**, and so is every qualitative conclusion below. The reference choice moves all three entrants by the same additive-in-Brier amount; it does not reorder them.
- **It leads the cycle-timescale rate baseline** by +0.083 BSS under the training reference and +0.133 under the verification one (+0.589 vs **+0.507**; +0.338 vs **+0.205**).
- **Against recency persistence it is a near-tie**, and we report it as one: a point margin of +0.013 BSS under the training reference (+0.021 under the verification one), and persistence has the *higher* peak TSS (**0.633** against **0.621**). On the real record the flare rate is strongly short-autocorrelated — that is the genuine clustering §4 fits — so a recency rate is a strong baseline.

- **No CI separation is claimed against either baseline.** The record bootstraps each model's BSS separately and never bootstraps the *paired difference*, which is the statistic both comparisons need; the marginal intervals overlap against persistence and against the 365-day baseline alike. The resamples are drawn under a common seed, so the paired interval is available from the stored replicates at no cost — we name it in Appendix B and do not reason from marginal overlap here.

The synthetic calibration this is measured against produced far smaller numbers (v17 cascade world: filter +0.133, persistence +0.077, cycle-rate +0.043), which is expected — the real record is far more strongly clustered than the synthetic worlds — and it is precisely why persistence closes the gap here and did not there.

8.3 Reliability: the forecast over-predicts, and we can say why

All ten forecast bins lie below the diagonal. When the filter says 0.55, the observed frequency is 0.40; at 0.75 it is 0.58; at 0.96 it is 0.87. Weighted by bin occupancy the mean forecast probability is 0.195 against an observed frequency of 0.144 — a 1.35× over-forecast across the whole range.

The dominant cause is not a generic model pathology; it is the same train/test activity shift that miscalibrates the reference forecast (§8.1). Parameters fitted to cycles 21–23 are applied to the much quieter cycles 24–25, and a filter fitted to an active era over-forecasts a quiet one by construction. The reference's miscalibration is larger (2.9×) than the filter's (1.35×), which is the sense in which the filter is doing work: it tracks the decline partway, and not all the way.

Operational verification has reported over-forecasting in issued flare probabilities too (Sharpe & Murray 2017), but we do not lean on that parallel — the mechanism here is identified and specific. Two things would settle it: a rolling-origin split in which training and verification have matched activity, and a recalibration step (isotonic or Platt, fitted on a validation split). Neither is applied here, because doing so would change a sealed measurement rather than report it; both are named in Appendix B. A Brier decomposition into reliability, resolution and uncertainty for all three entrants is the cheap diagnostic that would show whether the ranking survives recalibration, and it is named there too.

8.4 The information ceiling, stated plainly

This forecast never sees a magnetogram. Its entire input is the history of flare occurrence. The operational and research effort in flare forecasting is overwhelmingly directed at photospheric vector-magnetogram parameters — an SVM on 25 SHARP parameters reaches TSS ≈ 0.76 in an operational-style split, though at about 42 % precision and in the configuration tuned to maximise TSS (Bobra & Couvidat 2015) — and none of that information is available to this filter.

The correct reading of Table 8 is therefore: *this is the skill available in the event record alone*, and it is a floor for what a magnetogram-informed model should achieve, not a competitor to one. We note, without leaning on it, that comparison studies of operational systems have repeatedly found no single method substantially better than the others or than climatology for M-class and above, and that which method scores best depends decisively on the event definition and the metric (Barnes et al. 2016; Leka et al. 2019) — so a like-for-like ranking against those systems is not a comparison this paper can make. The peak-TSS figures in Table 8 are maximised over threshold on the test set and are an optimistic readout that should not be set beside operational numbers.

8.5 The X-class row is a synthetic-label diagnostic

The record also carries an X-class BSS of +0.082. **It must not be read as a real-data result.** The now-cast harness constructs its X-class occurrence series by binomial thinning of the M-class daily counts at a fixed 6 % rate, so the "152 X-flares in the test tail" are thinned draws, not observations — the real record contains 568 genuine X flares (Table 1). Note that record v26's own summary reads this row as "a small but positive skill on the rare, high-impact events"; that reading is superseded here, and the record is not edited, because records are append-only. Re-running the X-class arm against the real labels is a straightforward measurement and it will be a **new record**. We make no X-class claim.

8.6 One asymmetry in the comparison

The cascade filter receives the per-bin era covariate; the two EWMA baselines do not. The test tail spans the 2020 era boundary, so the filter alone gets an exogenous 1.38× rate multiplier over the most active stretch of the verification period — and §2.2 has already conceded that this multiplier is partly genuine solar amplitude rather than calibration. Giving the baselines the same covariate (or scoring the cascade without it) is a one-line change and is named in Appendix B. It is also worth noting that a per-era calibration factor is metadata a real-time forecaster would not have for the current era; the forecast is causal in the flare counts, not in the instrument metadata.

9. The event-times arena: a saturated Hawkes, reported as a result

(Records `out/v25/`, `out/v29/`.)

The second arena races the flare *times* directly: an exponential-kernel Hawkes self-exciting process against the MSM-Cox point-process likelihood and homogeneous Poisson. On the real record it does not produce a comparison. At **both** thresholds the Hawkes maximum-likelihood fit returns **η = 0.999**, and the record flags it degenerate and keeps η as a diagnostic only, setting the arena's ΔBIC to null.

The 0.999 is the estimator's ceiling, not a measured value. The vendored exponential-kernel fitter clips the branching ratio at 0.999 inside its objective and again on return, so no fit of this estimator can report a higher number. What the record establishes is therefore that the optimiser runs *to that ceiling* on this record, which is a statement about the estimator's behaviour on strongly cycle-modulated data rather than a measurement of a near-critical branching ratio. We correct the sealed record's framing on this point: the likelihood does not diverge; the parameter saturates.

The saturation is a confound, and the sealed null-audit shows it. A zero-self-excitation null — a smooth solar-cycle-modulated Poisson process with no contagion whatsoever, drawn to preserve the real record's intensity — is fitted with the same estimator and returns a band whose median is **0.999** at both thresholds, with the observed value covered. So a process with no self-excitation in it at all saturates the same estimator on the same record's intensity profile. That is the Filimonov–Sornette effect, in which concatenated Poisson processes of differing rates are read back as critical ($\hat{n} \approx 1$ for a true $n = 0$), with residual diagnostics unable to reject the misspecified fit (Filimonov & Sornette 2015) — here on 50 years of real solar data.

Two qualifications belong with that, and the second is the one the sealed record supplies and we would otherwise be accused of omitting.

- **A saturated null covers a saturated observation trivially.** The intensity-preserving band is a degenerate point at the clip — [0.999, 0.999] over all 100 replicates at both thresholds, zero spread. "Covered" is not a distributional statement when both arms are pinned at the same ceiling. Note that the *synthetic* calibration this re-derives did **not** saturate (its smooth-cycle world returned a median $\eta \approx 0.956$), so the real-data arm is qualitatively different from its own calibration. Re-running with an unclipped parameterisation, or profiling the likelihood in η over a grid, would give the null a distribution and turn this into a measurement; it is named in Appendix B.
- **A second null in the same record points the other way, and we report it.** The local-interval-shuffle null — which preserves the local waiting-time marginal and destroys longer-range gap ordering — gives a *lower* band (median 0.886 at M1.0, and 0.976 at C1.0) that the observed η exceeds at $p = 0.0099$ at both thresholds — and because the clip censors the observation only from above, that exceedance is conservative: an unclipped MLE could only sit higher. So there *is* some short-range gap structure in the flare times. What the smooth-cycle null establishes is only that the **near-critical value** is not evidence of it.

The consequence for §7 is direct and we take it. The span-mirror rule requires a fitted span to be quoted beside its *raced* self-excitation alternative; here that alternative saturates, so it supplies no clean comparison. The evidential weight of $S \approx 144$ therefore rests entirely on the rivals the counts arena *did* race cleanly — the piecewise-constant Poisson (beaten on parsimony, §4.2), the single switch (beaten decisively), and the free-transition HMM (beaten narrowly at M1.0 and lost at C1.0, §6) — plus the two-world contrast of Table 4, the residual-specificity cell and the pure-Poisson bootstrap. A span is never the evidence; the raced rivals are.

10. Discussion

What the result says to the flare-statistics literature. The nonstationary- Poisson description of flare timing is not wrong, but as usually implemented it is under-specified. Represented as a piecewise-constant rate — the Bayesian-blocks form the field standardised on — it needs 531 parameters at M1.0 and 1,423 at C1.0, and a four-parameter hidden-state model comes within **+1,035** nats of the first and **+3,876** nats of the second, with BIC preferring the latter by thousands at both thresholds. Within the hidden-state family, several latent switching layers beat one decisively, at fixed parameter count, in sample and out. The rate is not *compactly* summarised by a slow modulation with change points — the piecewise-constant form spends hundreds to over a thousand parameters on what four latent parameters nearly match — and the structure it is spending them on is estimable multi-state latent structure.

What it says about the model to fit. §6 is the useful half for a practitioner. A free-transition Poisson hidden Markov model over a handful of activity states describes the flare-rate record at least as well as the geometric cascade at M1.0 and materially better at C1.0, while remaining exactly as estimable. If the goal is description or forecasting rather than a claim about scale hierarchy, the free HMM is the model this record prefers, and it is a strict generalisation of the single switch the field already uses implicitly whenever it speaks of "active" and "quiet" periods. The cascade's contribution is not that it wins; it is that it made the comparison possible, that it establishes the depth at fixed parameter count, and that its constrained form gives the free HMM something to be measured against.

Why the geometric constraint fails here and holds elsewhere. In the accreting compact-object arena the same tournament finds the cascade generalising better than a free discrete-state model out of sample, which is what a *correct* constraint does. Here it does not, and a synthetic stellar-flare control reports the same sign. The natural reading — offered as a reading, not a measurement, and forbidden as a claim by §1.3 — is that a geometric ladder is the right idealisation for fluctuations injected across a continuum of scales, and the wrong one for a discrete population of active regions whose emergence and decay do not lay themselves out on a log-uniform grid. Testing that means freeing the rung rates and scoring the freed ladder against the geometric one out of sample; we name it in §7.4 and do not run it.

Why the branching-ratio arena matters even though it produced no number. The flare literature has a live thread that reads clustering in flare times as evidence of triggering or sympathetic flaring. §9 shows that on this record an exponential-kernel Hawkes MLE saturates its own ceiling at $\eta = 0.999$ at both thresholds, and that an intensity-preserving, self-excitation-free smooth-cycle null saturates it too — while a local-shuffle null gives a lower band the observed value does exceed. On this estimator and this binning, therefore, a near-critical η carries no evidence of contagion until the regime/background alternative has been raced. We would extend that as a caution to comparable fits on flare event lists; we do not claim it as a universal, since it rests on one estimator, one binning and one record.

Relation to the companion papers. The model, its closed forms and its photon-channel metrology are in `papers/01-correspondence-metrology.md`; the same tournament at continuous-flux altitude on 21 years of Swift/BAT X-ray binaries is in `papers/02-tournament-grs1915-cygx1.md`. This paper is the event-altitude member of that set, and the three together are what make §6.4's cross-arena comparison possible.

Limitations

Scope and claim.

- **The incumbent margin is parsimony, not fit** (§4.2): the Bayesian-blocks partition fits the record better in raw likelihood at both thresholds, and the BIC margin is the parameter penalty net of that deficit. The claim is held at that strength throughout, and it rests on BIC alone (§3.3).
- **The cascade-vs-free-HMM comparison has no zero.** It returns a positive result on the no-hierarchy control world (§6.3), so we use it only as a description contest at these sample sizes and say so.
- **"Hierarchy" is the wrong name for what §4 establishes.** Several latent switching layers beat one; at C1.0 the best description is a free four-state HMM, which is multi-state without being a timescale hierarchy (§6.5).
- **The fitted ladder is a selection quantity, not a recovered one** — on §6.5's ground, that a free transition matrix describes these data at least as well, so the geometric parameterisation is not the preferred description. Nothing physical is inferred from the rung positions (§7.4). *(Restated 2026-08-30: this bullet previously rested on record v32's "no track in the sealed estimability grid recovers a ladder to truth" and on a multiplier bias that "does not shrink with N". Record v35 shows both were misspecification artifacts and they are withdrawn — see §7.1 rule 2. The bullet's conclusion is unchanged; its reason is now the one that survives.)*

Apparatus and accounting.

- **The scan is right-censored.** BIC is still descending at $k = 8$ at the primary threshold, so $k^* = 6$ is a scan boundary and the reported depth is a lower bound; the fitted span drifts upward with k (§4.3).
- **The blocks entrant is fitted without the era covariate** (§3.2) and its partition is selected under a penalty 2.7× softer than the one it is scored by (§3.3). Both run against the incumbent. A p_0 / `ncp_prior` sensitivity sweep and a selection-scoring-matched partition variant are not in this record.

- **The blocks held-out column is unusable, and the guard repair alone will not fix it** (§3.3): its predictor is the terminal training block's constant rate extrapolated across fifteen years.
- **The HMM adversary is a two-point grid** (`protocol.GOES_HMM_STATES = [2, 4]`). The entrant this paper calls "the adversary that matters" was scanned over $J \in \{2, 4\}$ only — no J = 3, 5 or 6 entrant was run at either threshold. The observed spacing on that axis is large next to the margin being reported: at M1.0 the J = 2 → J = 4 step alone gains 2,640 BIC (26,104 → 23,464) while the cascade's reported lead over the best HMM is **+119**, and a J = 5 or J = 6 entrant would cost only $9 \cdot \ln n = 88$ and $20 \cdot \ln n = 196$ BIC in extra penalty. **So the +119 is a lead over J ≤ 4, not over the free-transition Poisson HMM family**, and Abstract (iii) and Table 2's M1.0 row should be read at that strength. The extended-J re-run ($J \in \{2, 3, 4, 5, 6\}$, Baum–Welch on 18,262 bins — cheap) is queued and is not in this record; it cannot be run from the committed tree, because the real GOES/XRS event record is fetched, never committed (`data/**` is git-ignored; §12). The direction is conservative for this paper's thesis — a larger J winning would *strengthen* (iii)'s "the honest boundary is a description contest" reading — but the stated margin would be wrong.
- **The k = 1 entrant carries a dead parameter** (§3.2 footnote): the span *S* is structurally inactive at a single layer, so the single-switch baseline is charged ln n = 9.81 BIC more than its effective dimension, and §4.1's depth gains are lower bounds by that amount.
- **The positive control is a self-recovery test** and the synthetic cells are not N-matched to the real ones (§3.4); the cycle world is a single-timescale envelope, not the fitted incumbent made synthetic.
- **There is not one uncertainty interval on a fitted parameter or a model-comparison margin in this paper.** Fitted ladders, ΔBIC values and held-out margins are point estimates from a multi-start numerical optimisation; §7.1 rule 2 addresses *bias* and says nothing about *variance*. Profile-likelihood or parametric-bootstrap intervals on ($\hat{m}$_hi, S, f_slow, λ_0) and paired intervals on every margin are what a referee should ask for, and we ask for them in Appendix B.
- **The event-times arena produced no comparison** (§9): the Hawkes rival saturates its estimator's ceiling, so the span-mirror rule is satisfied only by the counts-arena rivals.
- **κ is fixed at 1** for identifiability, so only the contrast enters; and the maximum-likelihood fit is multi-start numerical optimisation, a local-optimum risk mitigated by the documented protocol rather than eliminated.

Data and instrument.

- **The era covariate is a coarse instrument** (§2.2): its five steps are not Hudson's uniform factor, a flux factor is not a count factor at fixed class threshold, its boundaries sit near cycle minima, and it under-corrects the measured 1.73 flux-median ratio. Its measured effect on the fit is negligible (§7.3), but the correction is not exact.
- **Detection completeness against cycle phase is not quantified here.** The soft-X-ray background rises by more than an order of magnitude from minimum to maximum, so above-background flare identification loses small events preferentially near maximum — a cycle-phase-dependent, multiplicative detection efficiency, which is the same shape as the signal being fitted. At C1.0, where the mean rate is 3.56 per day and reaches tens per day at maximum, overlapping flares are also merged into single catalogued events, producing rate-dependent censoring. Both act most strongly at C1.0, where this paper's largest numbers live. We state the systematic and its expected sign; we do not correct for it.
- **The daily binning is a choice this paper does not defend by experiment.** The incumbent's strongest evidence (waiting-time analyses) and the one published result that rejects pure nonstationary Poisson (the ≈ 30-hr mutual-information excess) both live at or below this resolution, and all sub-daily clustering is

aliased into daily counts as overdispersion. The entrant roster is conditionally-Poisson-given-rate throughout, so it contains no model that can absorb within-day burstiness directly: "multi-state rate structure" and "daily counts are not conditionally Poisson" are not separated by this tournament.
- **The specificity cell is clean only at C1.0** (§5.2), and at M1.0 the Pearson statistic itself false-positives on the no-hierarchy world under the shuffle null.
- **The smoother's cycle recovery is a sanity check with a null that cannot fail** (§5.3), against a documented envelope rather than an observed index.

Now-cast.

- **The headline BSS depends on which climatology is the reference** (§8.1): +0.589 against the training-period one, +0.338 against the verification period's own. The ordering of the entrants is identical under both.
- **The forecast over-predicts in all ten reliability bins** (§8.3) and is not recalibrated; the cause is a train/test activity shift, not a generic pathology.
- **No paired bootstrap of the BSS difference was run** (§8.2), so no CI separation is claimed against either baseline.
- **The cascade receives the era covariate and the baselines do not** (§8.6).
- **The X-class row is a synthetic-label diagnostic** (§8.5), and record v26's own summary reads it otherwise.

And one that is not a limitation of the analysis.

- **No collaborator pass yet.** The solar-physics reading of §7.4 in particular is written to be resisted, and a domain referee should resist it.

---

11. Conclusions

1. **Several latent switching layers beat one, decisively, on 50 years of real GOES/XRS flares**, at both thresholds, in sample and out, at fixed parameter count: ΔBIC **+3,175** / **+10,195** over one switching layer, and **+460** / **+1,436** nats held-out. Read against the sealed no-hierarchy world — which produces a large positive depth gain of its own, and selects the same k* at C1.0 — the real record's per-flare depth gain is 11× the null's at M1.0 and 2.8× at C1.0.
2. **Against the field's piecewise-constant incumbent the cascade wins on parsimony, not on fit.** ΔBIC **+3,102** / **+6,172**, but the Bayesian-blocks partition fits better in raw likelihood by **+1,035** / **+3,876** nats with 531 and 1,423 parameters against four. A four-parameter hidden-state description comes within about a thousand nats of a 531-parameter partition over 18,262 days. That is a statement about description length, and we do not upgrade it.
3. **The edge over a pure-Poisson null is significant, and the smooth-rate artefact is ruled out.** The parametric bootstrap places the observed statistic beyond every replicate (**p = 0.0099**, the ensemble floor at N = 100); the residual-specificity cell separates a genuine cascade from a smooth modulated rate at **81×** where the naive statistic separates them at **1.4×**, and the real record's clustering survives the correct smooth-Poisson null — at the C1.0 threshold, where that cell is clean.
4. **The specific geometric ladder is not what the record prefers, and this is a measurement rather than a caveat.** The cascade's M1.0 margin over a free-transition Poisson HMM is **+119** BIC, of which **+117.8** is the parameter penalty and only **+1.7** is fit; out of sample it is **+2.17** nats over fifteen held-out years — a tie

— and at C1.0 the cascade loses by **−1,475** BIC and by **−310.83** nats out of sample. The sealed calibration shows the axis firing positive with no hierarchy present (**+147** BIC, **+10.3** nats in the cycle world), so this is a description contest at these sample sizes and not a hierarchy test.
5. **A causal, filter-only 24-h M-class now-cast carries real skill from the event record alone**: BSS **+0.589** against the training-period climatology (CI [0.521, 0.667]) and **+0.338** against the verification period's own, leading the cycle-timescale baseline (**+0.507** / **+0.205**) and edging recency persistence (**+0.576** / **+0.317**) on point estimates while sitting behind it on peak TSS (**0.633** vs **0.621**), with no paired interval claimed either way. It over-forecasts in every reliability bin, which we trace to a train/test activity shift, and it never sees a magnetogram.
6. **A near-critical Hawkes branching ratio on this record is a confound and a ceiling.** η = **0.999** at both thresholds is the estimator's clip, and a zero-self-excitation smooth-cycle null saturates it too — the Filimonov–Sornette effect on 50 years of real solar data — while a local-shuffle null gives a lower band the observation does exceed.

The claim we make is the narrow one: the first estimable hidden-state description of the flare-rate record, and the fair tournament placing several latent switching layers against one and against the piecewise-constant description. The result is ready for owner review and a domain collaborator pass, and is fully reproducible from the named records behind the apparatus gate.

---

12. Reproducibility

Everything in this paper regenerates from the named records behind the repository's apparatus gate, which is data-free: it exercises the archival and residual-specificity paths on deterministic synthetic series, so the repository verifies from a fresh clone with no network. The real numbers need the NOAA fetch.

```
python3 -m venv venv && venv/bin/pip install -r requirements-dev.txt
./verify.sh                                # apparatus gate, data-free
venv/bin/python -m fenocosm.goes_fetch     # the real NOAA fetch (network + h5py)
venv/bin/python -m fenocosm.p93_flare_real     # the tournament + now-cast (~25 min)
venv/bin/python -m fenocosm.paper_figures      # -> papers/figures/*.png
```

Every **bolded in-repo** number in this paper is additionally bound to a sealed record's JSON by the repository's paper-anchor gate, which renders the value from the record and asserts it appears verbatim here; a re-cut that moves a number fails the gate until the paper is swept in the same commit. Appendix A is the map.

---

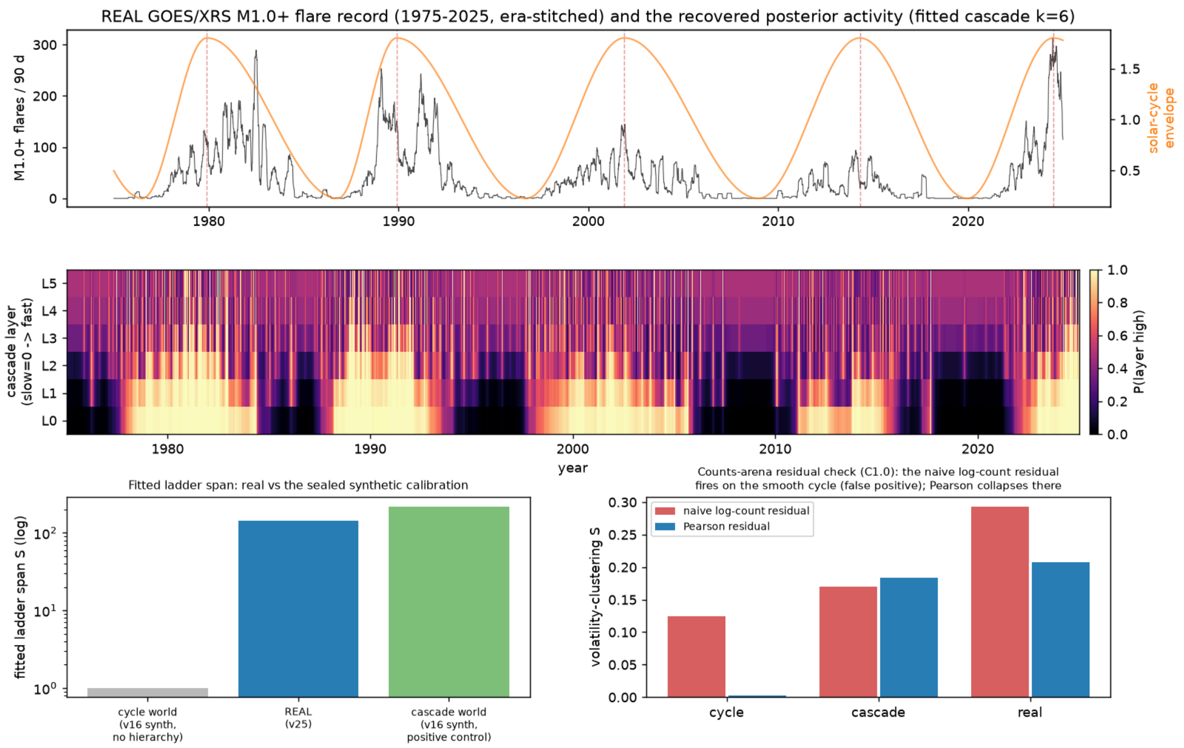
13. Figures

All figures regenerate from the named records via `python -m fenocosm.paper_figures`:

- `figures/fig_flare_tournament.png` — counts-arena BIC against cascade depth on the real record at both thresholds, with the Bayesian-blocks and free-transition-HMM reference lines and the k = 6 scan boundary marked (record v25; §4).

- `figures/fig_flare_boundary.png` — the cascade-versus-free-HMM margin on BIC (left) and on held-out log-likelihood (centre) across the sealed two-world calibration and the real record at both thresholds, showing that the axis fires positive with no hierarchy present and that the real record sits below the no-hierarchy world; and (right) the fitted ladder spans on the same six cells, beside the ladder-degeneracy infimum (records v16, v25, v32; §§6–7). Every Δ in the two left panels is positive in the cascade's favour, as in Table 6.
- `figures/fig_flare_nowcast.png` — the 24-h M-class Brier Skill Score against the training-period climatology, with block-bootstrap intervals, for the causal filter and its two baselines, each beside its sealed synthetic cascade-world value; and the reliability diagram, every bin below the diagonal (records v26, v17; §8).

The record's own overview figure (`out/v25/p93_flare_real.png`



out/v25/p93_flare_real.png

) additionally shows the 90-day rolling M1.0+ flare rate against the cycle envelope, the per-layer posterior activity heatmap over 1975–2025, and the residual-specificity cell.

---

14. References

Every entry below was checked against a primary source (publisher page, Crossref registry record, or arXiv abstract) in the pass that produced this draft.

- **Aschwanden, M. J., & Johnson, J. R. 2021**, "The Solar Memory from Hours to Decades," *ApJ*, 921, 82. doi:10.3847/1538-4357/ac2a29

- **Aschwanden, M. J., Johnson, J. R., & Nurhan, Y. I. 2021**, "The Poissonian Origin of Power Laws in Solar Flare Waiting Time Distributions," *ApJ*, 921, 166. doi:10.3847/1538-4357/ac19a9
- **Barnes, G., Leka, K. D., Schrijver, C. J., et al. 2016**, "A Comparison of Flare Forecasting Methods. I. Results from the 'All-Clear' Workshop," *ApJ*, 829, 89. doi:10.3847/0004-637X/829/2/89 *(Paper I of the comparison series; §8.4.)*
- **Baum, L. E., Petrie, T., Soules, G., & Weiss, N. 1970**, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annals of Mathematical Statistics*, 41, 164–171. doi:10.1214/aoms/1177697196 *(the forward–backward machinery the* 2^k*-state likelihood uses).*
- **Bazilevskaya, G., Broomhall, A.-M., Elsworth, Y., & Nakariakov, V. M. 2014**, "A Combined Analysis of the Observational Aspects of the Quasi-biennial Oscillation in Solar Magnetic Activity," *Space Science Reviews*, 186, 359–386. doi:10.1007/s11214-014-0068-0 *(quasi-biennial band, periods 0.6–4 yr; cited for context in §7.4 and not claimed).*
- **Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. 2012**, "Toward Reliable Benchmarking of Solar Flare Forecasting Methods," *ApJ*, 747, L41. doi:10.1088/2041-8205/747/2/L41 *(the TSS standard in flare forecasting).*
- **Bobra, M. G., & Couvidat, S. 2015**, "Solar Flare Prediction Using SDO/HMI Vector Magnetic Field Data with a Machine-learning Algorithm," *ApJ*, 798, 135. doi:10.1088/0004-637X/798/2/135 *(cited in §8.4 for magnetogram-based TSS in a research setting, not for an information-content comparison, which it does not make).*
- **Brier, G. W. 1950**, "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review*, 78, 1–3. doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2
- **Calvet, L. E., & Fisher, A. J. 2001**, "Forecasting multifractal volatility," *Journal of Econometrics*, 105, 27–58. doi:10.1016/S0304-4076(01)00069-0 *(the continuous-time Poisson multifractal, the direct precursor).*
- **Calvet, L. E., & Fisher, A. J. 2004**, "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes," *Journal of Financial Econometrics*, 2, 49–83. doi:10.1093/jjfinec/nbh003 *(the discrete-time MSM and its maximum-likelihood estimation).*
- **Filimonov, V., & Sornette, D. 2015**, "Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data," *Quantitative Finance*, 15, 1293–1314. doi:10.1080/14697688.2015.1032544 *(concatenated Poisson processes read back as critical; §9).*
- **Gaizauskas, V., Harvey, K. L., Harvey, J. W., & Zwaan, C. 1983**, "Large-scale patterns formed by solar active regions during the ascending phase of cycle 21," *ApJ*, 265, 1056–1065. doi:10.1086/160747 *(activity complexes maintained over 3–6 solar rotations; §7.4, not claimed).*
- **Hudson, H., Cliver, E., White, S., et al. 2024**, "The Greatest GOES Soft X-ray Flares: Saturation and Recalibration over Two Hale Cycles," *Solar Physics*, 299, 39. doi:10.1007/s11207-024-02287-x (arXiv:2310.11457) *(the source of the 1.43 recalibration factor. The individual SWPC long/short divisors, 0.70 and 0.85, are from the NOAA/NCEI GOES XRS documentation and the GOES-R XRS L2 Users' Guide, not from this paper).*
- **Jolliffe, I. T., & Stephenson, D. B. (eds.) 2012**, *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed. (Chichester: Wiley). doi:10.1002/9781119960003
- **Leka, K. D., Park, S.-H., Kusano, K., et al. 2019**, "A Comparison of Flare Forecasting Methods. II. Benchmarks, Metrics, and Performance Results for Operational Solar Flare Forecasting Systems," *ApJS*, 243, 36. doi:10.3847/1538-4365/ab2e12 *(Paper II; the operational-systems comparison, §8.4).*
- **Rabiner, L. R. 1989**, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77, 257–286. doi:10.1109/5.18626

- **Scargle, J. D., Norris, J. P., Jackson, B., & Chiang, J. 2013**, “Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations,” *ApJ*, 764, 167. doi:10.1088/0004-637X/764/2/167 *(the piecewise-constant incumbent, and the `ncp_prior` calibration §3.3 discusses).*
- **Schwarz, G. 1978**, “Estimating the Dimension of a Model,” *The Annals of Statistics*, 6, 461–464. doi:10.1214/aos/1176344136
- **Sharpe, M. A., & Murray, S. A. 2017**, “Verification of Space Weather Forecasts Issued by the Met Office Space Weather Operations Centre,” *Space Weather*, 15, 1383–1395. doi:10.1002/2017SW001683
- **Snelling, J. M., Johnson, J. R., Willard, J., Nurhan, Y., Homan, J., & Wing, S. 2020**, “Information Theoretical Approach to Understanding Flare Waiting Times,” *ApJ*, 899, 148. doi:10.3847/1538-4357/aba7b9 *(the ≈ 30-hr mutual-information excess against Bayesian-block nonstationary-Poisson surrogates — the closest prior result to this one).*
- **van Driel-Gesztelyi, L., & Green, L. M. 2015**, “Evolution of Active Regions,” *Living Reviews in Solar Physics*, 12, 1. doi:10.1007/lrsp-2015-1 *(emergence ≈ 60 hr, spot decay 30–60 d, region lifetimes to months; §7.4, not claimed).*
- **Wheatland, M. S. 2000**, “The Origin of the Solar Flare Waiting-Time Distribution,” *ApJ*, 536, L109–L112. doi:10.1086/312739
- **Wheatland, M. S. 2004**, “A Bayesian Approach to Solar Flare Prediction,” *ApJ*, 609, 1134–1139. doi:10.1086/421261 *(the event-statistics forecaster; §8.1 notes it is a recency estimator, so the 365-day EWMA baseline is not it).*
- **Woodcock, F. 1976**, “The Evaluation of Yes/No Forecasts for Scientific and Administrative Purposes,” *Monthly Weather Review*, 104, 1209–1214. doi:10.1175/1520-0493(1976)104<1209:TEOYFF>2.0.CO;2

Instrument and data. GOES/XRS flare event lists: NOAA National Centers for Environmental Information (NCEI) legacy reports, 1975–2017, and the GOES-R L2 XRS flare summary product (`xrsf-l2-flsum_science`, GOES-16), 2017–present.

The MSM-Cox model definition, its closed forms, and the photon-channel metrology are in the companion methods paper, `papers/01-correspondence-metrology.md`.

Data availability & acknowledgments

The flare event lists are the public NOAA products named above, fetched at run time by `fenocosm.goes_fetch` and **not committed** to the repository; the frozen summary statistics, per-era tallies, fetch provenance and SWPC-validation result live in the sealed record `out/v25/`, and the analysis is reproducible from the documented NOAA products (the NCEI archive and the GOES-R flare summary are static or append-only, so a later re-fetch returns at least the flares frozen here). The reproduction commands are in §12.

The records, by DOI. Each sealed record this paper reads is deposited as its own citable object. The source repository is **private and remains private**, so records are cited by DOI, never by repository path; the DOIs below were **published on Zenodo 2026-08-31** and resolve as of that date.

record	what it carries here	DOI
`out/v25/`	the real GOES/XRS counts-arena tournament (§§2–7, §9)	10.5281/zenodo.22180493 *(published 2026-08-31)*
`out/v26/`	the causal now-cast and its reference forecast (§8)	10.5281/zenodo.22180495 *(published 2026-08-31)*

Two dated notes travel with those deposits and are part of the record as deposited: `out/v25`'s, which marks its JSON verdict strings `ladder_reads_as` and `cascade_beats_piecewise_and_hmm` as superseded by §4.2/§6.5 (parsimony-not-fit; selection-not-recovery), and `out/v26`'s, which marks its X-class row as a synthetic-label diagnostic per §8.5.

The analysis package is deposited as a software snapshot, 10.5281/zenodo.22180503 *(published 2026-08-31)*.

GOES/XRS soft-X-ray flare event lists courtesy of NOAA's National Centers for Environmental Information and Space Weather Prediction Center. The C1.0 completeness threshold is a literature choice (Snelling et al. 2020; Aschwanden & Johnson 2021), not a NOAA product; the era recalibration follows Hudson et al. (2024) and the GOES-R XRS L2 Users' Guide, with the loader's departures from it stated in §2.2.

All tournament code — the era-stitching fetch, the Cox counting channel, the Kronecker forward pass, the Bayesian-blocks and free-transition-HMM entrants, the bootstrap, the smoother, the residual-specificity cell and the now-cast — is new Fenocosm work built on the vendored CC BY 4.0 MSM-Cox generator and exact forward algorithm (see PROVENANCE.md).

Appendix A — Provenance table (every headline number → its sealed record)

Every **bolded in-repo** number in this paper is bound to a sealed record's JSON by the repository's paper-anchor gate (`fenocosm.verify.check_paper_anchors`), which renders the value from the record and asserts it appears verbatim here; a re-cut that moves a number fails the gate until the paper is swept in the same commit. External numbers — literature values, and anything not produced by a sealed record in this repository — are **not** anchored and are marked *external* below. Quantities marked *derived* are computed here from sealed values by the formula given; the gate binds them through the same renderer.

number (as used)	record	file / field
8,393 M1.0+ and **64,996** C1.0+ flares; the five-era table	v25	`p93_flare_real.json` → `fetch_provenance.totals`, `.era_tallies`
SWPC live-feed validation **76 of 77**	v25	`fetch_provenance.swpc_validation`
18,262 daily bins; era stitch at **2017.107**; the ~573 GOES-R C+ (63 M1.0+) flares under-corrected in [2017.107, 2020)	v25	`meta.T_bins`; `fetch_provenance.stitch_decyear`, `.era_boundary_caveat`
era-4/era-3 true-flux median ratio **1.73** vs covariate 1.38	v25 *(derived)*	ratio of `fetch_provenance.era_tallies[4,3].goesr_true_flux_median_wm2`
the counts-arena table (Table 2), every BIC and held-out value, both thresholds	v25	`real_tournament.cells.{M1.0,C1.0}.counts_arena`
held-out block **5,479 days, about 15 years**	v25 *(derived)*	`meta.T_bins` at `GOES_HOLDOUT_FRAC = 0.3`
depth gain over the single switch **+3,175** / **+10,195**; held-out **+460** / **+1,436**	v25	`...counts_arena.dbic_cascade_over_single_switch`, `.depth_gain_heldout`
margin over piecewise Poisson **+3,102** / **+6,172**; blocks parameters 531 / 1,423	v25	`...counts_arena.dbic_cascade_over_piecewise_poisson`, `.bayesian_blocks.n_blocks`
the Table 3 decomposition: penalty **+5,171** / **+13,924**, fit **−2,069** / **−7,752**, blocks fits better by **+1,035** / **+3,876** nats	v25 *(derived)*	from `...cascade.bic` / `.n_params` and `...bayesian_blocks.bic` / `n_blocks` at n = `meta.T_bins`
Bayesian-blocks `ncp_prior` = 7.39 nats vs the BIC charge 2·ln n = 19.6	*derived from code*	`event_models.bayesian_blocks_counts` (Scargle eq. 21 form) at p_0 = 0.05, N = 18,262
margin over homogeneous Poisson +16,881 / +80,904	v25	`...counts_arena.dbic_cascade_over_homog`
deep extension k = 7 (23,274) and k = 8 (23,249); spans 161.3 / 270.4	v25	`...counts_arena.deep_extension`
bootstrap observed **+16,209** at k = 3, null mean −28.9, p95 −26.4, **p = 0.0099**	v25	`real_tournament.bootstrap_significance`
smoother AMI **0.323**; null mean and s.d.; *p* = 0.005 = 1/201	v25	`real_tournament.smoother_activity`, `.time_shuffle_null`
the residual-specificity cell (Table 5), all six world × threshold rows; ratios **81×**, **1.4×**, **2.4×**; naive-vs-smooth-Poisson **p = 0.887** / **p = 0.824**	v25	`counts_arena_residual_check`
era sensitivity **143.6** with covariate vs **143.0** without	v25	`real_tournament.era_sensitivity`
fitted ladders and rung timescales (Table 7, real rows)	v25	`...counts_arena.cascade.ladder`; interior rungs derived geometrically as f_i = f_slow·S^(i/(k−1))
Hawkes **η = 0.999**, flagged degenerate at both thresholds	v25	`...event_times_arena.hawkes_eta`, `.hawkes_degenerate`
the 0.999 is the estimator's clip, not a fitted value	*derived from code*	`vendor/msm_cox_hawkes/models.py`, `hawkes_exp_mle` (`eta = min(eta, 0.999)`)
cascade vs free HMM: by **+119** BIC at M1.0, loses by **−1,475** BIC at C1.0; held-out **+2.17** / **−310.83**; **+119.4** / **−1,475.1**	v25	`...counts_arena.dbic_cascade_over_hmm`, `.dheldout_cascade_over_best_non_cascade`

number (as used)	record	file / field
the **+117.8** parameter-penalty term, the **+1.7** fit term and the **+0.83**-nat log-likelihood difference	v25 *(derived)*	from `...cascade.bic` / `.n_params` and `...poisson_hmm.panel.4` at n = `meta.T_bins`
Table 4's six calibration cells: every depth gain, blocks margin, HMM margin (BIC and held-out) and span for the two synthetic worlds	v16	`p81_flare_tournament.json` → `cells.{M1.0,C1.0}:{cycle,cascade}.counts_arena`, `verdict`
cycle-world HMM margins **+147** BIC / **+10.3** nats (M1.0) and **+461** / **+68.1** (C1.0); cascade-world **+221** / **+38.8** and +3,892 / **+561.7**	v16	`cells.*.counts_arena.dbic_cascade_over_hmm` and the cascade/HMM `heldout_ll` pair
cycle-world span **1.00**, cascade-world **220** (M1.0) and 136.8 (C1.0)	v16	`verdict.cycle_world_ladder_span_S`, `.cascade_world_ladder_span_S`, `cells.C1.0:cascade...ladder.S`
per-flare depth gains (Table 4)	v16 / v25 *(derived)*	depth gain ÷ `n_flares` per cell
the cascade beats the free HMM **0/6** at N ≥ 1,600 on its canonical cascade track (i.e. the free HMM wins every seed there), and 0.50–0.67 on the sessions and strong-amplitude tracks at the same N; N* ≈ 800 at the canonical amplitude, ≈ 140 at the strong one, above the grid at the weak one; `S_hat_median` ≈ 1.0 on the reversal rows	v20	`p20_event_altitude.json` → `tracks.*.rows[*].cascade_beats_hmm_rate`, `.S_hat_median`, `.floor`, `headline`
selection ≠ recovery, **as corrected 2026-08-30** (scope clause and degeneracy-conditioned floors added 2026-08-31): `N*_selection ≈ 800` stands (quoted in §7.2 with its Clopper–Pearson interval on 5/6, [0.36, 1.00], and its grid bracket [400, 1,600]); the `S ≤ 1.05` / `m_hi ≥ 10⁴` ladder-degeneracy flag stands, and both real cells are interior. **Withdrawn, over the three re-analysed tracks only** (`v20 | trivial` N ≤ 1,600; `v28 | transit` and `v28 | trivial` N ≤ 800; canonical amplitude only — the weak/strong amplitude brackets and the `v20 | sessions` track were not re-fit and their v32 rows stand as read): v32's "recovery above the grid everywhere" and its "+45 % to +86 %" multiplier bias — a model-misspecification artifact (the sealed fits omit the generator's activity window), refuted by record v35. The **61 of 72** span-infimum census is retained only as a property of those covariate-free fits. The corrected floors are quoted in §7.1 rule 2 under both readings — headline 100 / 200 / 140, degeneracy-conditioned 100 / 400 / 140, or 200 / 400 / 400 requiring the ≥ 5/6 rule to hold at every larger re-analysed N as well	v32 (10.5281/zenodo.22180493 is v25; v32 travels with the Paper F record set) / **v35**	`p96_floor_estimability.json` → `headline`; `p97_recovery_readout.json` → `headline`, and `cells[*].{ladder_degenerate, multiplier_recovered_truth}` for the conditioned floors (recomputed by `verify._rr_floor_triple`)

number (as used)	record	file / field
the de-biasing calibration §7.2 actually uses: fitted multiplier **1.819** vs injected **1.700** (**+7.0 %**) at M1.0 and **1.694** vs 1.700 (**−0.4 %**) at C1.0, on the observation-process-matched cascade world	v16	`p81_flare_tournament.json` → `cells.{M1.0,C1.0}:cascade.counts_arena.cascade.ladder.m_hi`; injected truth `goesdata.GOES_META.cascade_world.m_hi`
the Bayesian-blocks held-out zero-rate guard, repaired at protocol `p7`	v32	`out/v32/README.md`, "The two vacuous diagnostics, repaired"
now-cast: Brier scores, BSS **+0.589** [0.521, 0.667], **+0.576**, **+0.507**; peak TSS **0.621** / **0.633**; **7,305 days**; training base rate 0.542 and climatology 0.419; the ten reliability bins	v26	`p93_flare_nowcast_real.json` → `real_record.M_class`, `.n_test`, `verdict`
verification-period frequency **0.144**, its climatology Brier 0.1234, and the rescored BSS **+0.338** / **+0.317** / **+0.205**; mean forecast probability 0.195 and the 1.35× over-forecast	v26 *(derived)*	Σn and Σ n·obs_freq over `real_record.M_class.reliability_cascade`
the X-class row (+0.082, "152 events") is a **synthetic-label** diagnostic — 6 % binomial thinning of M-class counts	v26	`real_record.X_class`; `meta.x_rate_fraction = 0.06`
synthetic now-cast calibration: cascade world +0.133 / +0.077 / +0.043	v17	`p82_flare_nowcast.json` → `worlds.cascade.M_class`
the real Hawkes η covered by a zero-self-excitation smooth-cycle null, band median **0.999** with a degenerate [0.999, 0.999] interval; the local-shuffle null medians 0.886 / 0.976 with the observation exceeding at p = 0.0099	v29	`p95_surrogates_real.json` → `eta_verdict`, `eta_goes_real.*.bands`
XRB cross-arena contrast: free 5-state HMM in-sample ΔBIC 1,851 (Cyg X-1) / ms-DRW 1,536 (GRS 1915+105); cascade held-out +296 / +184 over the free HMM; Cyg X-1's fitted span 1.0000012, on the infimum	v14 / v13	`out/v14/README.md`; `p72_discrete.json`
stellar-flare cross-arena contrast (**synthetic**, calibrated to GJ 1243): 0/12 false cascade-wins on held-out log-loss above ≈ 1,600 flares	v31	`out/v31/README.md`
the span-mirror datum: span 1.14 decades on single-mechanism near-critical truth, 87 % of a true cascade's 1.31	*external* (Fenocrisis record v2)	named in Fenopan `CONSTITUTION.md` §6.6
the **1.43** recalibration factor and the 0.70 / 0.85 SWPC divisors	*external*	Hudson et al. 2024 (1.43); NOAA/NCEI GOES XRS documentation (the divisors)
every other literature value in §§1, 2, 7.4, 8 and 9	*external*	see §14; each verified against a primary source

Appendix B — Companion work, named and out of scope

The measurements named here are for the record and are **not** part of this paper's claim. None of them changes a number in §§1–11; several of them would change how firmly a number can be read, and we say which.

Controls this paper wants and does not have.

- **The blocks-rate parametric bootstrap.** §5.1's bootstrap is against homogeneous Poisson, the weakest entrant. The control §4 most wants simulates Poisson counts from the *fitted* 266-/712-block λ(t), refits both entrants, and builds the null distribution of the depth gain and of the blocks margin. Until it exists, the two-world contrast of Table 4 carries the specificity claim alone.
- **A `p₀` / `ncp_prior` sensitivity sweep, and a selection-scoring-matched partition.** §3.3 records that the blocks partition is selected at 7.39 nats per change point and scored at 19.6. A variant selected under `ncp_prior = ln n`, and a sweep of the blocks margin in p_0, would show how much of §4.2's margin is that mismatch.
- **A misspecified positive control.** §3.4 records that the cascade world is generated by the fitter's own module, so it is a self-recovery test. A multi-timescale rate that is *not* an MSM cascade — superposed Ornstein–Uhlenbeck or shot-noise components at the same rung timescales, or a superposition of active-region birth–decay processes — is the control that would make the positive arm informative about mechanisms other than the model's own.
- **A counts-arena reversal grid.** §6.3(ii) imports its N-dependence caution from a different arena with collapsed fitted ladders. The grid that would settle it is cycle-modulated, era-covariate, counts-arena, at the real cells' N, with non-degenerate ladders.
- **Uncertainty intervals.** There is not one on a fitted parameter or a model-comparison margin in this paper. Profile-likelihood or parametric-bootstrap intervals on (m̂_hi, S, f_slow, λ_0) at both thresholds, and paired block-bootstrap intervals on every ΔBIC and Δheld-out margin, are what a referee should ask for. The now-cast's paired interval is nearly free: the sealed bootstrap uses a common seed across models, so the difference distribution is available from the stored replicates.

Re-cuts that would supply a missing column.

- **An online blocks forecaster.** §3.3 records that the blocks held-out predictor is the terminal training block's constant rate. The fair form is an expanding-window or online change-point predictor; the `p7` guard repair alone does not make the column comparable.
- **A `p7` re-cut of the tournament.** Would carry the ladder-degeneracy flag inside the sealed ladder dicts and the repaired guard. A new record beside v25, never an edit to it.
- **The X-class now-cast against real labels.** §8.5's row is a synthetic-label diagnostic. The real record carries 568 X flares; scoring the X-class arm against them is one run.
- **The now-cast, with matched information and a matched split.** Give the EWMA baselines the same era covariate the filter gets (§8.6), or score the filter without it; add a rolling-origin split in which training and verification have matched activity (§8.3); add a Brier decomposition into reliability, resolution and uncertainty for all three entrants; and re-run the smoother's AMI against a circular-shift null and an observed solar index (§5.3).

Experiments that would change what can be claimed.

- **The freed-ladder experiment.** §7.4 names the measurement that would license a physical reading of the rung timescales: fit the rung rates *freely* and score the freed ladder against the geometric constraint out of sample. Given §6, the expected outcome is that the constraint is not preferred — which is worth measuring rather than assuming.
- **An overdispersed emission entrant, and a sub-daily cut.** The current roster is conditionally Poisson given the rate, so within-day burstiness enters as overdispersion with nowhere to go but the latent rate. A negative-binomial blocks entrant and a negative-binomial HMM would separate "multi-state rate structure" from "daily counts are not conditionally Poisson"; a 6-h or 12-h cut would let the analysis engage the waiting-time literature it cites.
- **A completeness audit against cycle phase.** The Limitations name the systematic — background-dependent detection efficiency and high-rate event merging, both strongest at C1.0. Quantifying it, and showing the depth gain and blocks margin survive restriction to a completeness-safe subsample, is the measurement a solar referee will ask for first.
- **Per-active-region streams and magnetograms.** The information-ceiling statement of §8.4 is a bound, not a measurement of what magnetograms add. Closing it means a different instrument and a different paper.

Two further companions belong to the wider program rather than to this paper: the sub-daily / cross-band behaviour of the same tournament (`papers/04-cross-band-cascade-seed.md`) and the repeating-FRB event-altitude arm (`papers/07-repeater-cascade-seed.md`).

A note on this file's name. This paper was grown from the seed `papers/06-flare-rate-cascade-seed.md`, which the sealed records v16, v17 and v25 name by that path. Those records are append-only and are not edited; this file is the live surface, and the mapping is recorded here and in `papers/README.md`.

---

1. At k = 1 the ladder span *S* is structurally inactive — a single layer has no span to spread — but the fit still emits it as a free parameter and BIC charges it. The single-switch entrant is therefore penalised by exactly ln(18,262) = 9.81 BIC more than its *effective* dimension warrants, so "fixed parameter count" is a nominal rather than an effective statement, and the depth gains quoted in §4.1 are **lower bounds by ln n**. At +3,175 and +10,195 the correction is immaterial; we state it because a referee checking the parameterisation will find it. (Source: `vendor/msm_corpus_fit/models.py` `hz_ladder_probs`, which ignores `S` at k = 1, while `param_specs` still emits it.) ↩