The first exact-likelihood, held-out test of cross-domain crisis entanglement

Paper A of the Fenocrisis founding program (PROMPTS.md P-2.1; Arc 2). Aimed at the polycrisis field's home venue — Global Sustainability — with the discrimination-atlas power result carried in the statistics supplement (§S).

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

Status. Complete draft, gated on owner review before submission (Owner queue). Every quantitative statement traces to a committed, regenerable record under `out/` (record v0 = the atlas; record v1 = the founding tournament; record v3 = the **variance-maximal shared-modulation absorption bound**, called "the attention control" in an earlier draft — see §5.3 for why that name is withdrawn). This is a working draft, not a submission.

Abstract

The polycrisis field's founding empirical claim is that crises are *causally entangled* across domains — that wars, disasters, outbreaks, defaults, and coups interact to produce behaviour beyond the sum of the elemental crises. The claim is drawn from causal-loop diagrams, cross-impact balance matrices, and expert elicitation; it has never been written as an estimable generative model, and no paper in the field has ever tested it with a held-out likelihood. We supply the missing test. We fit the no-interaction null the field never wrote down — independent per-domain inhomogeneous Poisson processes with declared, flexible reporting baselines — against three coupled alternatives (multivariate Hawkes cross-excitation; a shared Markov-switching-multifractal latent cascade in a Cox intensity; and the nested hybrid), with identical exposure and baseline for every arm, on two crisis corpora: the harmonized six-domain country-year shock corpus of Delannoy et al. (2025) as published (Corpus-A), and a country-month rebuild from primary conflict and epidemic sources (Corpus-B). The headline is never that a coupled arm beat the null — a raw branching ratio on clustered data is not evidence of contagion (Filimonov & Sornette 2015) — but the **cross-domain excess** of that advantage over a trend-robust circular-shift surrogate, gated by a bidirectional recovery check and cross-referenced to a discrimination-power atlas. On Corpus-A the coupled arms' large advantage is reproduced in full by the surrogate: it is within-domain clustering, not entanglement, and the founding claim does not survive its first likelihood test. On Corpus-B a cross-domain excess does survive the surrogate — but it is **fragile to the declared reporting-baseline family** (significant under three of four baselines, absorbed by the flexible spline) and is then **absorbed below significance by a single shared contemporaneous attention covariate the per-stream baseline family cannot represent**. The audited verdict is therefore a claim-weakening negative: under the field's own corpus, tested honestly, the cross-domain entanglement signal is either absent (Corpus-A) or does not survive the reporting-artifact channel (Corpus-B). Every number here is a retrospective model-comparison on public, aggregate data; nothing is a forecast,

a warning, or a country score. **Scope: this tests entanglement in the *timing* of crisis onset. It does not test entanglement in crisis *severity or magnitude*, which is the more plausible form of polycrisis and a separate deliverable (marked-process tournament, forthcoming); the onset-time result does not speak to it.**

---

1. The gap, in the field's own words

The polycrisis literature is conceptual, elicited, or descriptive. Its causal-mechanism synthesis (Lawrence, Homer-Dixon, Janzwood, Rockström, Renn & Donges 2024, *Global Sustainability* 7, "Global polycrisis: the causal mechanisms of crisis entanglement") develops a research agenda for the emerging polycrisis program but writes down no estimable model, and its companion consensus roadmap (Lawrence et al. 2024, *Polycrisis Research and Action Roadmap*, Cascade Institute) states plainly that "polycrisis analysis also needs formal models of crisis interaction and early-warning indicators of crisis emergence and escalation." A 2025 systematic review (Rakowski et al., *Annual Review of Environment and Resources* 50) found **one dynamic model among 59 analyzed polycrisis publications**. The single harmonized cross-domain shock corpus in the field (Delannoy et al. 2025, *Global Sustainability* 8 — six domains × 173 countries × 1970–2019) has been analyzed only with counts and co-occurrence plots, with event magnitudes deliberately excluded by its authors.

Meanwhile the point-process literature owns exactly the machinery the field needs — self-exciting Hawkes processes (Hawkes 1971; Ogata 1988), latent-regime Cox processes, and the identification apparatus that separates them — but has never taken it across crisis domains. One near-neighbour must be named and differentiated: Szymanski, Lin, Asztalos & Sreenivasan (2015, *Scientific Reports* 5:10998) fit a maximum-likelihood contagion model over the World Economic Forum's global risks across five domains, with a likelihood-ratio test against a no-contagion null. That programme is the closest prior art, and we engage it directly in §7. Its data, however, are **expert-survey risk activations** with observed (not latent) states and state-conditioned (not kernel-based) excitation; ours are **observational event streams** with a latent regime and an explicit artifact channel. The contribution here is the first exact-likelihood, held-out test of cross-domain crisis entanglement *on observational event data*, with the reporting artifacts the field has never corrected for measured and swept.

Either verdict is a deliverable. A confirmation would be the first quantitative evidence of crisis entanglement; a failure would be the first honest falsification of the field's central claim under its first real test. We report the failure plainly. The house precedent is a sibling programme's honest home-turf loss (a 0/6 null result published as a deliverable), and this programme's own μ(t) audit, which weakened its own headline.

---

2. The claim as the field states it — and the null it never wrote down

The claim. Crises interact across domains through common stresses, domino effects, and inter-systemic feedbacks, producing emergent, entangled dynamics (Lawrence et al. 2024; the Cascade Institute's cross-impact-balance models; Gambhir et al. 2025). In the field's causal-loop and cross-impact formalisms the interactions are asserted and expert-weighted; they are never estimated from event data against an alternative in which the domains do not interact at all.

The null. The alternative the claim implicitly rejects is that each domain's event stream is generated *independently*, with only its own within-domain dynamics and its own reporting trend — no cross-domain coupling of any kind. We write it down explicitly, because the field has not: a per-domain inhomogeneous Poisson process,

$$\lambda_{c,d,t} = E_{c,t} \cdot \mu_d(t; \theta) \cdot \exp(\beta_d' z_{c,t}),$$

with exposure E, a *declared, flexible* reporting baseline $\mu_d(t)$ (a spline or changepoint family — F4 makes it mandatory, not optional), and covariates z. Independence across d is the whole content of the null. Entanglement, if it is real, is what a coupled arm captures *over and above* this — and over and above the within-domain clustering that a raw coupling estimate confounds with cross-domain coupling.

---

3. The two corpora

The corpus is assembled twice, so that a native-resolution result and a finer primary-source rebuild can disagree in the open.

Corpus-A — the field's own corpus, as published. The Delannoy et al. (2025) harmonized shock corpus at native country-year resolution: six domains (CLIMATIC, CONFLICTS, ECOLOGICAL, ECONOMIC, GEOPHYSICAL, TECHNOLOGICAL), 173 countries, 1970–2019 (T = 50), **232,378 events**. Loaded verbatim, with the per-component F4 artifact declarations attached (the corpus inherits EM-DAT relief-reporting channels, UCDP/GTD newswire coding, and expert-dated financial chronologies; magnitudes are absent by the authors' design). Release: `github:LouisD-KVA/Polycrisis-shocks@main`, content-pinned by SHA-256 (`fb3ef0be…`), retrieved 2026-08-17.

Corpus-B — a country-month rebuild from primary sources. A finer panel: UCDP Georeferenced Event Dataset v24.1 conflict events (Sundberg & Melander 2013) and WHO Disease Outbreak News epidemic onsets via the Carlson et al. (2023) machine-readable derivative — two domains (conflict, epidemic), 192 countries, 1990–2024 (T = 420), **350,218 events**. Releases: UCDP GED `ged241`; WHO-DON `github:cghss/dons@master`, content-pinned by SHA-256 (`442cdd63…`), retrieved 2026-08-17. The disaster (EM-DAT, registered), economic (chronologies), and coup (Powell–Thyne) components are recorded *pending* here — Corpus-B in record v1 builds from whatever open primary streams load, and the two-domain cut is the honest committed version; the grown four-domain panel is a later record (v4) and does not enter this test.

Every re-resolution and harmonization choice is logged in `docs/corpus-choices.md`; no restricted or registered rows enter any committed artifact (the license gate makes this impossible-by-construction, not merely discouraged — SPECS §6, Standing Rule 7).

---

4. The tournament

The roster. Four arms sharing one panel observation layer, fit to each corpus: `null` (independent per-domain inhomogeneous Poisson); `hawkes` (multivariate discrete-time cross-excitation); `msm_cox` (a shared k-level Markov-switching-multifractal latent cascade modulating a Cox intensity, exact 2^k-state filter); and the nested `hybrid` (MSM-Cox baseline + Hawkes excitation). The latent cascade and the point-process estimators are inherited, vendored verbatim with provenance headers and pinned commits (Standing Rule 4); the panel observation layer is new work (SPECS §4).

Fairness rules (verbatim from the sibling tournament protocol). Identical exposure and baseline covariates for every arm; identical held-out splits (temporal blocks *and* country folds); BIC (in-sample, penalized) and held-out conditional log-loss both reported; parametric bootstrap on every verdict; depth-honest penalties (ladder depth k = 2, a dense transition, at or below the SPECS §4 identifiability ceiling for a single global monthly series).

Why BIC carries the verdict, and held-out is corroborative. The held-out log-loss *magnitudes* are inflated by the parameter-driven null's trend extrapolation on the large-count domains (a 50% rate error on a ~2,000/month conflict stream costs hundreds of nats per cell). They are reported for completeness; the verdict is read off BIC plus the F2 surrogate excess, which are robust to that artefact (`out/v1/x1/RECORD.md`, honest caveats).

The excess, not the margin (F2). The raw margin by which a coupled arm beats the null is large on both corpora and is never the finding. A raw branching ratio on clustered data is a Filimonov–Sornette artefact: regime-switching Poisson data with *no* self-excitation at all is fit by a Hawkes arm as a near-critical branching ratio $\hat{\eta}$ = 0.80 (reproduced across N; §S, `filimonov_sornette.png`). The finding is the **cross-domain excess**: the coupled advantage over a *trend-robust circular-shift surrogate*, in which each domain's multiplicative fluctuation is shifted around its fitted baseline so that cross-domain alignment is destroyed while the trend and the marginal autocorrelation are preserved. What the surrogate reproduces is within-domain clustering; what survives the surrogate is the cross-domain part. Before any of it, a bidirectional parametric-bootstrap recovery check confirms the generating arm is recoverable at the actual configuration (F6), and the operative cell is cross-referenced to the atlas (record v0). The surrogate machinery is itself recovery-gated: on synthetic independent (no-coupling) clustered-and-trended panels the excess is ≈ 0, and on shared-regime panels it is positive (`tests/test_entangle.py`).

5. Results

All figures below are committed to `out/v1/x1/`; the results table is `x1_results_table.png` and the excess decomposition is `x1_excess_decomposition.png`. Seed `20260901`, profile `default`; 120 surrogate draws per channel; the numeric record is `out/v1/x1/x1.json`, bit-reproducible from the seed.

5.1 Corpus-A — no cross-domain excess survives

The recovery gate passes (the configuration is decidable: winner recovered in 0.85 of bootstrap replications, null recovered in 1.0; threshold 0.8), so this is a verdict, not undecidability. The coupled arms beat the null by a large raw margin — the winner is the hybrid, raw ΔBIC = **48,772** over the null (parametric-bootstrap winner-margin 95% CI [−259, 56,758], vs-null p = 0.016). But the trend-robust surrogate reproduces essentially all of it. The cross-domain excess, by channel:

channel	observed ΔBIC	within-domain (surrogate)	excess	surrogate-tail p
hawkes	44,197	33,024	**11,173**	0.231
msm_cox	46,713	40,912	**5,801**	0.355
hybrid	48,772	38,778	**9,994**	0.207

No channel's excess clears the surrogate ensemble (all p > 0.2). With cross-domain alignment destroyed, the coupled advantage is undiminished: **the advantage is within-domain clustering, not cross-domain entanglement.** The field's founding claim does not survive on the field's own corpus — a committed negative result.

5.2 Corpus-B — an excess survives the surrogate, then does not survive the audit

The recovery gate passes (winner recovered 1.0, null 1.0). The coupled arms again beat the null by a large raw margin (hybrid raw ΔBIC = **93,935** over the null; winner-margin 95% CI [153,003, 380,859], vs-null p = 0.016), and here a cross-domain excess *does* survive the surrogate — strongest in the hawkes channel:

channel	observed ΔBIC	within-domain (surrogate)	excess	surrogate-tail p
hawkes	86,420	59,163	**27,257**	0.008
msm_cox	73,091	58,757	**14,334**	0.025
hybrid	93,935	66,028	**27,907**	0.008

Roughly **32%** of the raw likelihood gain is genuinely cross-domain; the rest is within-domain clustering. Read at this stage alone, a modest but non-zero cross-domain coupling survives the field's first real likelihood test on Corpus-B. It is far short of the raw margins, and — per F2 and F4 — it is not yet a claim: it must survive the reporting-baseline audit first.

5.3 The audit-adjusted verdict (the headline number, F2)

The μ(t) reporting-baseline sweep (record v1 / x3). Every headline is recomputed under the whole declared baseline family — constant, linear, spline, and changepoint (the last with the streams' declared structural break dates dummied) — and a claim survives only if it holds across the family (F4). It does not:

- **Corpus-A:** the cross-domain excess is significant under only **2 of 4** baselines — constant +40,995 ($p = 0.012$), linear +11,530 ($p = 0.160$), spline +21,555 ($p = 0.012$), changepoint +6,371 ($p = 0.086$) — range [6,371, 40,995] ΔBIC. It does not survive the declared (spline + changepoint) family; **fragile, not a claim.**
- **Corpus-B:** the cross-domain excess is significant under only **3 of 4** baselines — constant +5,344 ($p = 0.012$), linear +25,208 ($p = 0.012$), spline +6,289 ($p = 0.346$), changepoint +23,747 ($p = 0.025$) — range [5,344, 25,208] ΔBIC. The flexible spline absorbs it; it does not survive the declared family; **fragile, not a claim.**

The excess *is* invariant to a constant under-detection (a coverage artifact cannot manufacture it; max relative deviation 0.005 on Corpus-A, 0.015 on Corpus-B), which rules out the crudest reporting confound but not the reporting-baseline choice itself.

The shared-modulation absorption bound (record v3) — and why it is not an attention measurement. The single most important missing cross-domain detection control is a shared media-attention shock that could co-detect events across domains and mimic coupling. The per-stream, smooth μ(t) family structurally *cannot* represent such a shock (it is per-stream and smooth; the confound is shared, spiky, and contemporaneous — measured: the covariate is 0.986 unspanned by the μ(t) family on Corpus-B). Adding a shared contemporaneous covariate to the model on Corpus-B **absorbs the entire remaining excess, driving it to or below zero, below significance.** (On the Corpus-A primary baseline there was no significant excess for the control to test; the μ(t) audit already carried that verdict.)

An earlier draft called record v3 "the attention control" and read the absorption as evidence that media attention explains the excess. **Both are withdrawn**, and the object is renamed here to what it is: a **variance-maximal shared-modulation absorption bound** — an *upper bound* on how much of the cross-domain excess any single shared modulation could absorb.

- **No attention data enters it.** `shared_modulation_covariate` (`fenocrisis/controls.py`) fits the no-interaction null, forms each domain's standardized log residual, and takes the **leading right singular vector of the D×T standardized-residual matrix** — by construction the single time series that maximizes shared cross-domain residual variance. Real country-month GDELT over 1990–2024 is unavailable to a determinism-gated run; the per-country loader is built as code and the live run is the gated next step (P-5.2/P-5.4).
- **And the placebo does not establish specificity.** An earlier draft argued that a placebo — the same covariate with its cross-domain timing destroyed — leaves the excess standing, "confirming that the absorption is specifically the covariate's cross-domain contemporaneous alignment". That inference is **circular** and is withdrawn: the

direction is computed on the same residuals whose cross-domain covariance *is* the excess under test, so a large absorption is close to guaranteed, and the placebo compares the variance-maximal shared mode against a scrambled version of itself, which must favour the former.

What survives unchanged is what the bound is for: it is real and tight, it was pre-committed by name, and it is **downside-only** — it can only shrink the claim, never revive it — and the Corpus-B excess was already found fragile by this program's own μ(t) audit before it ran. All three pending controls (shared-modulation absorption, capture–recapture detection, real-time vintage) share that property, and each was pre-committed by name before its data was in hand. An out-of-fold construction — fit the leading component on one half of the months and apply it to the other — is the honest strengthening, and is queued. Record v5's stated motivation should be read the same way: the redesign answers the strongest shared-modulation confound the data admit, not an observed attention series.

The audited verdict. Under the field's own corpus and a finer primary-source rebuild, tested with an exact likelihood and held out: the cross-domain entanglement signal in crisis *timing* is absent on Corpus-A, and on Corpus-B is fragile to the reporting baseline and absorbed below significance by a shared attention channel the field has never controlled for. **We report no robust cross-domain onset-timing entanglement.** This is a claim-weakening negative, stated plainly (F2/F4).

---

6. What this does *not* say

Three scope limits are load-bearing.

1. **Onset timing, not severity.** The entire test concerns whether the *timing* of crisis onsets is entangled across domains. It says nothing about whether the *severity or magnitude* of crises propagates — which is arguably the more plausible form of polycrisis, and the home domain of the multiplicative cascade (volatility of magnitude). That is a separate, marked-process tournament (forthcoming), whose decidability must be mapped before any severity claim; the onset-time result does not license or foreclose it.
2. **A negative under this apparatus, at this N and resolution.** Corpus-A's country-year, magnitude-free resolution and Corpus-B's two open domains are what the honest committed corpus supplies. The atlas (§S) says where the tournament *can* separate the arms; a negative in a decidable cell is a verdict, but the cross-section — ~200 countries the global aggregation throws away — carries power a directional country-panel redesign is built to exploit (a later record). This paper reports the aggregated-panel result.
3. **No oracle (F7).** Nothing here forecasts, warns, ranks, or scores a country. Every number is a retrospective, dataset-level model-comparison on public, aggregate data.

Posterior regime paths, where shown, are scientific record with uncertainty, never rankable scores.

---

7. Engaging Szymanski et al.

Szymanski, Lin, Asztalos & Sreenivasan (2015) and successors (Lin et al. 2017; Brissette et al. 2021) are the closest formal prior art: a maximum-likelihood contagion model over the ~50 WEF global risks across five domains, with a likelihood-ratio test against a no-contagion null. We share their instinct — that the entanglement claim demands a likelihood, not a diagram — and we adopt it. Three things change when the same instinct is carried to observational event streams:

- **Observed vs latent states.** Their risk activations are elicited from experts and treated as observed; a crisis event stream has a *latent* regime whose unsounded states are real (the metacrisis, in the owner's definition, is literally a latent-regime model). A shared latent regime and cross-excitation are second-order-equivalent (Hawkes 1971), so the entanglement question cannot be answered by a contagion arm alone — it requires racing the contagion arm against the latent-regime arm, which is the whole of Q2 (Paper C) and the reason the excess here is decomposed, not attributed.
- **Elicited activations vs a reporting channel.** Expert activations do not have a newswire. Observational crisis streams do, and their detection is plausibly self-exciting (crises attract journalists) and partly response-defined (states of emergency, appeals). A contagion signal on such data is confounded with the detection channel unless that channel is declared and swept — which is why the shared-modulation absorption bound (§5.3), not the contagion likelihood, carries the audited verdict here.
- **State-conditioned vs kernel excitation, and no multifractal structure.** Their excitation is state-conditioned; ours is a kernel Hawkes term raced against a multiplicative-cascade latent regime with an exact filter. The comparison is not a refutation of their programme — it is what their programme implies once the states go latent and the data carry a photon channel.

The respectful engagement is the price of credibility; the contribution is carrying the likelihood test to observational event data with latent-regime machinery and an audited artifact channel, and reporting the negative it yields.

---

8. Limitations

- **The artifact channel (F4).** The audited verdict rests on a *variance-maximal shared-modulation* covariate — an absorption **bound** constructed from the residuals themselves, not an attention measurement and not a live per-country GDELT series over the full window (unavailable to a determinism-gated run; DOC-API coverage begins ~2017). Because the bound is computed on the same residuals whose covariance is the excess, it cannot attribute the absorption to attention; the live GDELT run is the named next step. The ACLED capture–recapture detection re-weighting is registered-access and pending. These are downside-only controls; a live run can only push the Corpus-B claim further down, not revive it.

- **Vintage (F5).** All Corpus splits are retrospective-vintage — the corpora are edited with post-split knowledge — and are labelled as such, never called out-of-sample. The one true out-of-sample check (the UCDP Candidate real-time monthly slice) is wired but pending; the current loader fetches only the annually-revised GED (`ged241`).
- **Resolution and coverage.** Corpus-A is magnitude-free country-year; Corpus-B is two open domains. The severity channel and the grown multi-domain panel are later deliverables and are out of scope here.
- **k, honestly bounded.** The fitted ladder depth (k = 2) is bounded by its identifiability ceiling for a single global monthly series and is never reported as "the depth of the metacrisis."

---

9. Data and code availability

All streams are open-tier, fetched at run time into a git-ignored cache and never committed beyond license-clean derived summaries (Standing Rule 7). The entanglement test regenerates bit-for-bit from the committed record:

```
python3 -m venv venv && venv/bin/pip install -r requirements.txt
venv/bin/python x0_verify.py            # the F1 anchor gate — must print X0 PASS
./fetch_data.sh                         # open-tier data (Delannoy, UCDP GED, WHO-DON)
venv/bin/python x0_atlas.py             # record v0, the atlas this test cites (if
absent)
venv/bin/python x1_entanglement.py --profile default --seed 20260901   # → out/v1/x1/
```

Record: `out/v1/x1/` (RECORD.md with seeds, environment, and regeneration commands; `x1.json` the machine-readable numeric record; figures). Data releases: Delannoy `github:LouisD-KVA/Polycrisis-shocks@main` (2026-08-17, SHA-256 `fb3ef0be…`); UCDP GED v24.1; WHO-DON `github:cghss/dons@master` (2026-08-17, SHA-256 `442cdd63…`). Environment: Python 3.11.15, NumPy 2.4.6, SciPy 1.17.1, Matplotlib 3.11.1. The audit that adjusts the verdict is `out/v1/x3/` (μ(t) sweep) and `out/v3/` (the variance-maximal shared-modulation absorption bound).

Deposit DOIs (published 2026-08-31). This paper and the records it reads are deposited separately.

- This paper — `10.5281/zenodo.22180505`.
- Record `out/v0` (the discrimination atlas) — `10.5281/zenodo.22180515`.
- Record `out/v2` (atlas v2, the near-critical power-law degenerate cell) — `10.5281/zenodo.22180517`.

Each DOI above is **published**: the deposits went live on Zenodo on **2026-08-31** and every identifier resolves. The source repository is private and will remain so, so each deposit is standalone and carries its own reproduction materials; deposit metadata deliberately omits repository links. *(Corrected 2026-08-31 — publication day: until today this paragraph read "reserved on Zenodo and its deposit is a private draft … the record is not yet published", with "(Zenodo DOI reserved; publication pending)" wherever a DOI appeared. That staging state ended when the owner pressed Publish on 2026-08-31.)*

S. Statistics supplement — the discrimination atlas (record v0)

No published power curve for Hawkes-vs-Cox discrimination exists anywhere; the atlas is a standalone methods contribution (SPECS §4) and the gate every verdict in §5 cites its cell of. Simulated known-truth only; no real data enters (F6).

Each arm of the roster (poisson, hmm, hawkes, msm_cox, bblocks) generates streams across a pinned grid — event count N ∈ {100, 300, 1,000, 10,000, 100,000}, Hawkes branching ratio η ∈ {0.2, 0.5, 0.8, 0.9}, MSM-Cox ladder depth k ∈ {2, 3, 4} — and every arm is refit to every stream. A cell is *decidable* when the true generating arm is recovered (selected by BIC) in ≥ 80% of replications, with held-out log-loss and the mean BIC margin reported as cross-checks. Seeds `20260817` (grid), `20260824` (panel), `20260828` (anchor); the numeric record `out/v0/atlas.json` reproduces bit-for-bit.

The decidable region. Of 45 single-stream cells, **26 are decidable and 19 are not**. Discrimination is a story about sample size; the onset (smallest N at which the generating arm is recovered) is: flat Poisson N ≈ 100, MSM-Cox N ≈ 300, self-exciting Hawkes N ≈ 1,000, regime-switching HMM N ≈ 10,000. Below its onset a mechanism is undecidable, and any in-sample "winner" there is an artefact of sample size, not a verdict (F6).

The mandatory anchor (Filimonov–Sornette). A regime-switching Poisson process with true η = 0 — no self-excitation at all — fit by a Hawkes arm infers a near-critical branching ratio **$\hat{\eta}$ = 0.80** (reproduced across N; `filimonov_sornette.png`). Its mirror, a genuine Hawkes truth (η = 0.5) at N = 3,000, is preferred over MSM-Cox by BIC in 8/8 replications: the discrimination runs the right way when self-excitation is real. This is why F2 requires coupling to be reported only as an excess and F6 requires the recovery gate — a raw branching ratio is never a finding on its own.

The panel-design gain. A shared latent regime is identifiable where a single stream is not: with one short stream the shared-cascade model beats the independent null with probability 0.25 (undecidable), and the recovery rises with the number of parallel streams sharing that regime (`panel_gain.png`). The statistical power lives in the cross-section, not the length of any one series — the lesson a later directional country-panel redesign is built to exploit. *How many* streams it takes is not a bare count; see the correction immediately below.

Dated correction (2026-08-25). This paragraph previously read "with **D = 3 parallel streams** of the same length it crosses the 80% decidable threshold". That reading is retracted by this program's own record `out/v4`, which rebuilt the panel gain at **R = 8** replications — a deliberate 2× over v0's R = 4, so recovery is no longer a quarter-granular fraction — and with per-domain heterogeneity in loadings and reporting baselines. The measured thresholds (smallest D at which the shared regime is recovered in ≥ 80% of replications): **homogeneous replicate streams onset at D = 4** — v0's "D = 3" was the quarter-granular R = 4 artifact, since at R = 8 that cell reads recovery **0.75**, just below the threshold; **heterogeneous-mixed** domains (per-domain loading spread over [0.2, 1.0], mean 0.6) onset *earlier*, at **D = 2**, because the strongly-coupled domains concentrate the identification; and **weakly coupled** domains (mean loading 0.3) are **not recovered anywhere in the D grid**. So **D-licensing is coupling-strength-conditioned**, not a domain count: the cross-section is still where the power lives, but how much of it a given D buys depends on how strongly the domains load the shared regime. Record `out/v4`.

Atlas figures (`out/v0/`). `decidability_atlas.png` (the recoverable region, BIC + held-out); `selection_matrix_N300.png` and `selection_matrix_N100000.png` (confusion matrices at small and largest N); `filimonov_sornette.png` (the spurious-inflation anchor across N); `panel_gain.png` (the shared-regime gain vs number of streams). The atlas is queryable — `decidable(atlas, N, eta=…, k=…, design=…) → {decidable, recovery, margin}` — so each verdict in §5 names the cell it stands in.

References

(Verified against the actual papers before release — Standing Rule 11. Volume and page details confirmed against source or flagged where unverifiable.)

- Brissette, C., Lin, X., et al. 2021. (Successor to Szymanski et al. 2015 — contagion dynamics of the WEF global-risk network.)
- Calvet, L. E., & Fisher, A. J. 2001. Forecasting multifractal volatility. *Journal of Econometrics* 105(1), 27–58; 2004. How to forecast long-run volatility: regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics* 2(1), 49–83.
- Carlson, C. J., et al. 2023. The World Health Organization's Disease Outbreak News: a retrospective database. *PLOS Global Public Health* 3(1), e0001083. doi:10.1371/journal.pgph.0001083. (The `cghss/dons` machine-readable derivative.)
- Delannoy, L., Verzier, A., Bastien-Olvera, B. A., Benra, F., Nyström, M., & Søgaard Jørgensen, P. 2025. Dynamics of the polycrisis: temporal trends, spatial distribution, and co-occurrences of national shocks (1970–2019). *Global Sustainability* 8, e24. doi:10.1017/sus.2025.10008.
- Filimonov, V., & Sornette, D. 2015. Apparent criticality and calibration issues in the Hawkes self-excited point process model. *Quantitative Finance* 15(8), 1293–1314.

- Gambhir, A., et al. 2025. A systemic risk assessment methodological framework for the global polycrisis. *Nature Communications* 16, 7382.
- Hawkes, A. G. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1), 83–90.
- Homer-Dixon, T., et al. 2022. A call for an international research program on the risk of a global polycrisis. (Cascade Institute, Technical Paper 2022-3.)
- Lawrence, M., Homer-Dixon, T., Janzwood, S., Rockström, J., Renn, O., & Donges, J. F. 2024. Global polycrisis: the causal mechanisms of crisis entanglement. *Global Sustainability* 7, e6. doi:10.1017/sus.2024.1. (And Lawrence et al. 2024, *Polycrisis Research and Action Roadmap*, Cascade Institute — the source of the "formal models of crisis interaction and early-warning indicators" call.)
- Ogata, Y. 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* 83(401), 9–27.
- Rakowski, J. J., et al. 2025. Characterizing the global polycrisis: a systematic review of recent literature. *Annual Review of Environment and Resources* 50, 159–183. (One dynamic model among 59 analyzed publications.)
- Sundberg, R., & Melander, E. 2013. Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50(4), 523–532.
- Szymanski, B. K., Lin, X., Asztalos, A., & Sreenivasan, S. 2015. Failure dynamics of the global risk network. *Scientific Reports* 5, 10998. (With Lin et al. 2017; Brissette et al. 2021.)

Full program reference list: SPECS.md §10.