A panel test read climate driving tree rings, and its own rule still failed it

A preregistered second round on 161 held-out boreal ring-width sites, with planted forward and reverse worlds, time reversal, a shared-trend world and 199 far-site pairings

Zain Dana Harper

Version 1.0, 9 November 2026. Finding F675 in the findings ledger. Honest null (tier D).

Abstract

Tree rings and summer temperature give a causal-direction test with a known answer: warmth shapes the ring, and the ring has no meaningful effect on the weather. An earlier round on 135 European sites found that a site-by-site transfer entropy estimator could not tell the two directions apart. In this second round we read lagged Granger tests across a panel of 161 held-out ring-width chronologies at or above 55 degrees north, under a plan hash-locked before any ring series was set against climate. On the high-pass series the panel direction index was +1.206 (95% interval +0.871 to +1.591), against a far-pairing 95th percentile of +0.213. Reversing time flipped the index to -0.894. Rings predicted the next summer at 6.8 percent of sites, inside the no-coupling world's 7.8 percent. Two binding first-difference checks failed by less than 0.02: the planted reverse world was called correctly at 58.4 percent of sites against a bar of 60 percent, and the shared-trend world's reverse rate was 0.117 against a limit of 0.10. The kill condition written before the run therefore fired, and the instrument earned no standing in this round. The real-data asymmetry is reported as an exploratory reading.

1 What this means for a reader

A method that points the right way on real data has not yet shown it can be trusted, because it must also pass tests on built data where we choose the answer. Here a simple panel test pointed from summer to ring by a wide margin and reversed when time ran backwards, and still missed two of the checks we had fixed in advance, by less than two percentage points each. We report that as a failure, because a near miss promoted after the fact is how weak methods earn trust they have not earned.

2 Introduction

Methods that read causal direction from two time series are used across climate science and ecology, and Runge and colleagues [1] set out the conditions under which such readings can be trusted in the Earth sciences. A known-answer test asks a method to recover a direction that is fixed in advance. Summer warmth shapes ring width at cold sites, and a ring cannot set the next summer, so annual rings and summer temperature give one such test on public data.

The first round of this study, finding F132, ran a nearest-neighbour transfer entropy estimator site by site on 135 European chronologies and failed: the two directions were detected at similar rates, time reversal made no difference, and a planted direction was misread at about half the sites. Its own account was low power at about 78 years of data and a ring-temperature correlation near 0.2.

This round changes that assumption. It reads the direction across a panel of sites at once, where F132 asked each site for its own verdict. The tool is the Granger test, the lagged regression test Granger introduced in 1969. Panel versions exist in econometrics: Dumitrescu and Hurlin [2] average per-unit Wald statistics into one

panel test of non-causality. Our index is simpler and was written from that idea, not copied from their statistic: a mean difference of per-site z-scores, with its null drawn from far-site pairings.

Two parts of the literature bear on the design. Preprocessing can create or hide Granger causality. Florin and colleagues [3] reported spurious and missed causalities after filtering, Barnett and Seth [4] showed that Granger causality is unchanged by filtering in theory and that filtering in practice raises both error rates, and Anderson and colleagues [5] gave conditions under which filtering does and does not distort it. That is why the plan ran three preprocessing arms and why their disagreement matters. Second, the lag structure of boreal conifer growth is not simple. Tumajer and Lehejcek [6] found in white spruce a positive effect of summer temperature in the year of ring formation and a negative effect of the summer one year before, and Lloyd and colleagues [7] reported that the previous growing season's temperature and precipitation control growth at most of their interior Alaska sites. Our searches, run on 9 November 2026, found no panel Granger test of ring width against climate with planted worlds of known direction. That is a search result, not proof that none exists.

3 Preregistration and design

The plan was written as a preregistration and hash-locked on 2 November 2026 at 09:42:37 UTC. Before the lock, the only facts read from the new data were outcome-free: the site count, the regions, the blocks, the series lengths, the distance to the nearest climate cell and the species. No ring series had been correlated, regressed or tested against any climate series. The estimator had been tested on synthetic autoregressive series only.

The plan named the assumption it changed from F132 and split the instrument's job into two readings scored apart. The first is reverse-route exclusion: past rings carry no lagged route into next year's summer temperature at the strength of the observed same-year coupling. The second is the forward reading: past summers add to next year's ring beyond the ring's own past. The plan forecast that the forward reading would lack power.

The kill condition was written without room for judgement: if any binding check fails, or the reverse-route exclusion fails on real data, the instrument gets no standing and F132's block on direction readings stands. Exclusion status needed every binding check and the exclusion rule to pass. Forward status needed, in addition, the planted forward world at real strength and the forward rule on real data.

One deviation was recorded before any result was seen. The first run stopped on a division by zero inside the precipitation arm, which is descriptive and decides no verdict. CRU precipitation at many of these cells is filled with climatology where stations are absent, and at four sites the June to August total is constant over the whole window. The fix returns a p value of 1 when the target's residual sum of squares is zero. No temperature series can reach that branch: temperature repeats for at most 2 years at any site. The plan, the verdicts and the seed were unchanged.

4 Data and provenance

Rings: standard ring-width chronologies from the International Tree-Ring Data Bank [8], held by NOAA's National Centers for Environmental Information as public data with a request to cite the contributors. The selection rule was fixed in the build script before the lock: standard chronologies with no earlywood, latewood or density keyword, latitude at or above 55 degrees north, outside longitudes 12 W to 45 E so that no site falls in F132's European region, years 1901 to 2025 with at least 5 samples, and a longest gap-free run of at least 60 years. This gave 161 sites: 100 in Canada, 45 in Alaska and 16 in Asia, in 77 two-degree blocks, with series of 60 to 115 years and a median of 89. The species are mostly white spruce, Picea glauca.

Climate: June to August mean temperature of the ring's year from the nearest land cell of CRU TS 4.10 [9] [10], released under the Open Database License 1.0 with attribution to the Climatic Research Unit, University of East Anglia. The nearest cell lay a median of 16.8 km from the site and at most 31 km. June to August precipitation was run as a descriptive secondary target.

5 Methods

The Granger pair at each site

For ring width R_t and summer temperature C_t at one site, the climate-to-ring test compares the full regression

$$R_t = c + a_1 R_{t-1} + a_2 R_{t-2} + b_1 C_{t-1} + b_2 C_{t-2} + e_t \quad (1)$$

with the same regression without the two climate terms, by an F test on the same rows. The ring-to-climate test is the mirror, with the roles of R and C exchanged. Each p value becomes a z-score, $z = Q(1 - p)$, where Q is the quantile function of the standard normal distribution, capped at 6. The site's direction statistic is

$$d_i = z_{f,i} - z_{r,i} \quad (2)$$

where $z_{f,i}$ is the forward (climate to ring) score and $z_{r,i}$ the reverse one.

The panel index

The panel direction index is the mean over the $N = 161$ sites:

$$D = \frac{1}{N} \sum_i d_i \quad (3)$$

Its 95% interval is a block bootstrap over the 77 two-degree blocks with 2,000 resamples. The detection rate in each direction is the share of sites with p 0.05 or below. The same-year correlation between ring and temperature is reported and gives no direction.

Three preprocessing arms

Each pair of series was analysed three ways: through a high-pass filter (the primary arm, as in F132), as first differences, and raw, scaled only. The raw arm is descriptive and shows what a shared trend does.

Synthetic worlds

Each world is built per site from iAAFT surrogates of that site's own series, five draws per site. In the planted forward world a surrogate S_t of the ring series receives the site's own climate:

$$R_t = S_t + aC_t + 0.4aC_{t-1} \quad (4)$$

At real strength a is the larger of the site's absolute same-year correlation on the high-pass arm and 0.1. In the planted reverse world a surrogate of the climate series receives the ring:

$$C_t = S_t + aR_t + 0.4aR_{t-1} \quad (5)$$

with the same a. The strong worlds use coefficients 0.6 and 0.3. In the no-coupling world both series are independent surrogates. In the common-driver world an autoregressive driver with coefficient 0.5 enters both series in the same year with weight $\sqrt{0.2}$. In the shared-trend world independent surrogates receive a common line and random walk before preprocessing.

For a world with a planted direction, direction accuracy is

$$A = \frac{m}{N} \quad (6)$$

where m is the number of sites whose d_i has the planted sign.

Far-site pairing and multiple testing

The far-site pairing pairs each ring with the summer temperature of a random site at least 1,000 km away, 199

times, on the high-pass and first-difference arms. Every site had at least 65 partners at that distance. Its p value is the share of pairings at or above the real index:

$$p = \frac{k + 1}{n + 1} \tag{7}$$

with n = 199, so the floor is 0.005. That floor sits under 0.0083, the smallest threshold the Benjamini-Hochberg correction at 0.05 over six tests can need, a lesson from F132, where 20 pairings left a floor too high to correct. The six family p values are the two pairing tests, the real reverse rate on each of two arms against a binomial 5 percent, and two sign tests on the planted worlds at real strength.

6 Controls and planted checks

Every control and synthetic world ran in the same run as the real data. The planted checks were split by role in the plan.

Binding checks, any failure fatal:

- Planted reverse world seen, on the high-pass and first-difference arms: the upper end of its index interval below 0, and accuracy above 0.6.
- No-coupling world, on both arms: forward and reverse detection rates at or below 0.10, twice the nominal 0.05.
- Shared-trend world on first differences: both rates at or below 0.10.

Power and limit checks, recorded but not binding:

- Strong planted forward world seen on high-pass: interval lower end above 0, accuracy above 0.55.
- Planted forward world at real strength seen on high-pass: interval lower end above 0. Forecast to fail.
- Shared-trend world on high-pass: both rates at or below 0.10. Forecast to fail, since F132 saw 0.126 and 0.170.
- Common-driver world on high-pass: its index interval contains 0.

Real-data rules:

- Exclusion: real reverse rate at or below 0.10 on both arms, and the real index above the planted reverse world's upper interval end on both arms.
- Forward: real high-pass index above the pairing 95th percentile, with its interval lower end above 0. The same comparison on first differences.

The time-reversed real series were reported beside the rules and decide nothing.

7 Results

Table 1 gives the panel index and both detection rates in every world, and Figure 1 draws the index with its interval on the two binding arms. Synthetic worlds are averaged over five draws per site.

On the high-pass series the panel direction index was +1.206 (95% interval +0.871 to +1.591), against a far-pairing 95th percentile of +0.213. The pairing mean was +0.047 and its 5th percentile -0.099, and no pairing reached the real value, so p is 0.005, the floor at 199 pairings. First differences gave +1.107 against a pairing 95th percentile of +0.186 (p 0.005), and the raw series +1.242 (+0.873 to +1.655). Climate predicted next year's ring at 39.8 percent of sites. Rings predicted the next summer at 6.8 percent of sites, inside the no-coupling world's 7.8 percent. The mean same-year correlation on the high-pass arm was 0.160, positive at 87.0 percent of sites.

Reversing time flipped the index to -0.894. On first differences it flipped to -1.092 and on the raw series to -1.041. The shared trend pushed the high-pass index negative (-0.317) and the raw index further (-0.744, with

Table 1. Panel direction index with block-bootstrap 95% interval, and forward and reverse detection rates, 161 sites.

World	High-pass index	Rates	First-difference index	Rates
Real data	+1.206 (+0.871 to +1.591)	0.398, 0.068	+1.107 (+0.729 to +1.496)	0.360, 0.062
Real, time reversed	-0.894 (-1.326 to -0.541)	0.137, 0.391	-1.092 (-1.528 to -0.720)	0.081, 0.385
No coupling	-0.011 (-0.095 to +0.073)	0.055, 0.078	-0.055 (-0.164 to +0.048)	0.066, 0.072
Planted forward, real strength	+0.180 (+0.073 to +0.296)	0.104, 0.052	+0.155 (+0.045 to +0.285)	0.123, 0.096
Planted forward, strong	+1.274 (+1.079 to +1.465)	0.402, 0.061	+0.789 (+0.512 to +1.088)	0.404, 0.176
Planted reverse, real strength	-0.429 (-0.556 to -0.303)	0.058, 0.154	-0.157 (-0.279 to -0.042)	0.073, 0.123
Planted reverse, strong	-2.230 (-2.458 to -1.978)	0.086, 0.740	-1.448 (-1.662 to -1.218)	0.096, 0.507
Common driver	-0.090 (-0.193 to +0.023)	0.061, 0.067	-0.082 (-0.181 to +0.019)	0.075, 0.086
Shared trend	-0.317 (-0.426 to -0.199)	0.135, 0.221	-0.070 (-0.171 to +0.038)	0.092, 0.117
Precipitation, descriptive	+0.344 (+0.059 to +0.664)	0.174, 0.043	+0.071 (-0.182 to +0.333)	0.143, 0.081

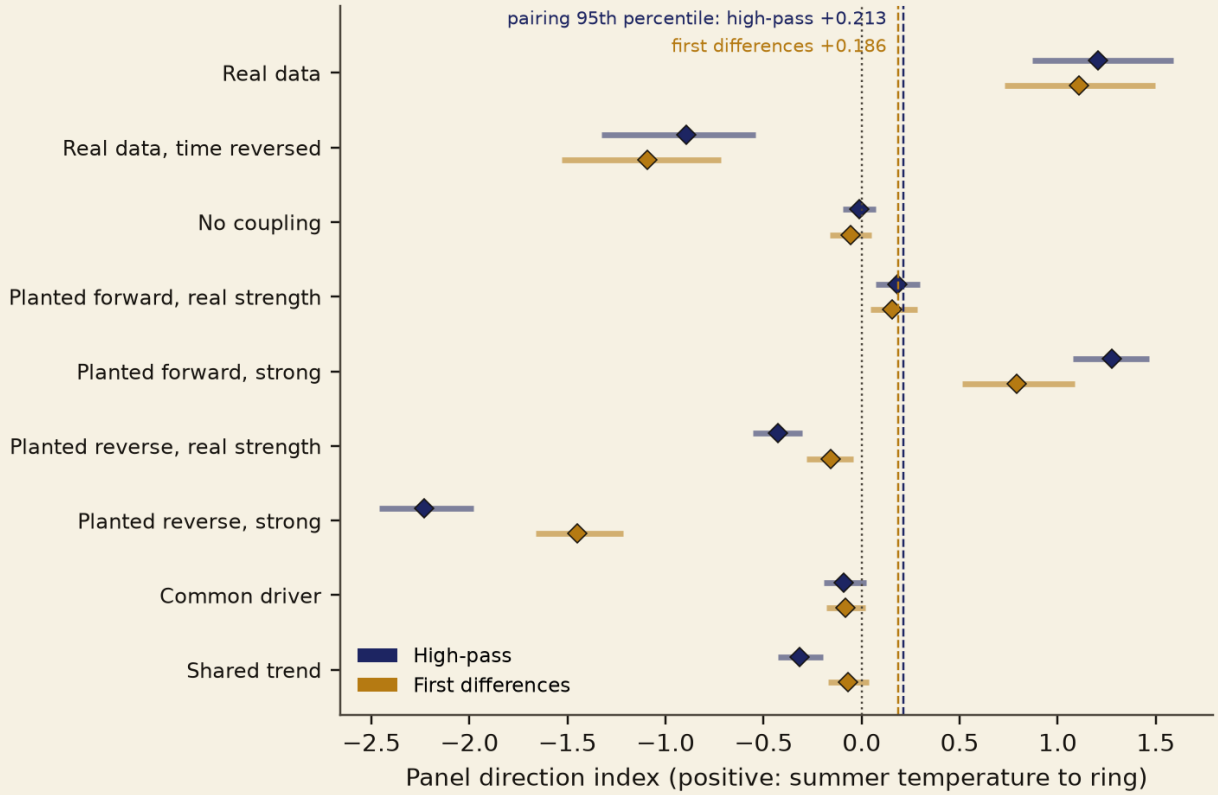


Figure 1. Panel direction index with block-bootstrap 95% intervals, high-pass (navy) and first differences (ochre). Dashed lines are the 95th percentiles of 199 far-site pairings; the dotted line is zero. Drawn from the run's result file.

rates 0.335 and 0.594), and the common driver gave -0.090 with an interval across zero. Neither confound produced a positive index like the real one.

Figure 2 sets every binding rule against its bar.

Two binding first-difference checks failed by less than 0.02: the planted reverse world was called correctly at 58.4 percent of sites against a bar of 60 percent, and the shared-trend world's reverse rate was 0.117 against a limit of 0.10. Every high-pass binding check passed: the planted reverse world was called at 71.4 percent with its interval upper end at -0.303, and the no-coupling rates were 0.055 and 0.078. Every real-data rule passed, the power checks passed, including the planted forward world at real strength (interval lower end +0.073), and the shared-trend limit on high-pass failed as forecast (0.135 and 0.221). Under the kill condition one binding

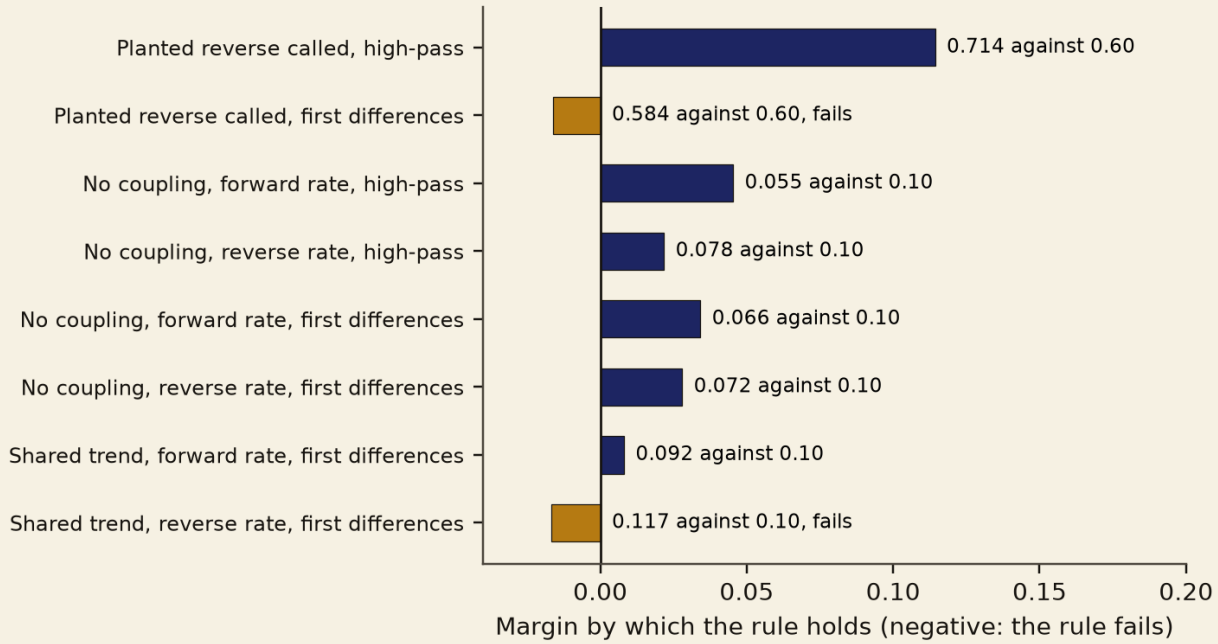


Figure 2. Binding checks as the margin by which each rule holds. Navy holds, ochre fails. The planted reverse rows show the accuracy rule; the interval rule on both arms held. Drawn from the verdicts in the run's result file.

failure is enough. The instrument gets no standing from this round.

The *p* values over the six-test family, corrected by Benjamini and Hochberg, rejected both pairing tests (q 0.0086 each), the reverse-world sign test (q 1.6e-7) and the forward-world sign test (q 0.0086). The real reverse rates against 5 percent were not rejected (q 0.22 on high-pass, 0.29 on first differences). The binomial tests treat sites as independent; the block intervals carry the spatial dependence.

Table 3 in Appendix A sets the forecasts written before the run against the outcomes. The forecasts for the forward signal were far off. They leaned on F132's forward rate of 0.089 in its variant without the same-year term, and on synthetic smoke tests in which a positive planted prior-summer term was absorbed by the ring's own past.

8 Discussion

On the high-pass arm the panel test did what F132's estimator could not. It pointed from summer to ring on sites the first round never touched, reversed with time, kept the reverse rate at the no-coupling level, and stayed clear of what a shared trend or a common driver produce. Read alone, that is a working instrument. It is not credited here, because the plan made two first-difference checks binding and both failed. This is an honest null about the instrument as locked, and the real-data asymmetry stays exploratory.

The failures say something about first differences. Differencing weights the fastest changes, and in the planted reverse world it shrank the index from -0.429 to -0.157 and the accuracy from 0.714 to 0.584. Smoke tests before the lock had already shown the first-difference arm weaker against the reverse world, and the plan made it binding anyway. That is a design error of ours. It agrees in kind with the filtering literature, where preprocessing can raise both error rates of a Granger test [4] [3] [5].

A look after the run, which is exploratory and not a result, explains why the forward signal was so much larger than forecast. Regressing ring on same-year and prior-year summer temperature on the high-pass arm, without the ring's own past, gave a same-year slope with median +0.170, positive at 88.8 percent of sites, and a prior-summer coefficient with median -0.127. The planted forward effect carried a positive prior-summer term, and

the real prior-summer coefficient was negative at most sites (median -0.127, positive at 21.7 percent). Its size was a median of 0.76 times the same-year term, and with the ring's own past added the median grew to -0.173. The inference, not tested here, is that a negative lag-one term adds to the Granger signal where a positive one is absorbed by the ring's own past. The planted world at real strength was then not at the real lag structure, and the same error sat in F132. A negative effect of the previous summer in white spruce is reported in the literature [6] [7].

The real index (+1.206) sits close to the strong planted forward world (+1.274) and far from the one at real strength (+0.180), which fits that reading.

A third round, finding F676, ran the same day on 145 further held-out sites with the planted prior-summer term fitted to the measured sign and only the high-pass arm binding. It gave a panel index of +1.204 against a pairing 95th percentile of +0.461, passed every binding check but one, and its kill condition did not fire, so the panel method gained a limited direction status at that coupling strength. That later result does not change this round's verdict.

9 Limitations

- The planted forward and reverse worlds used a positive prior-summer term, the wrong sign for most of these stands. The pass of the planted forward world at real strength is a pass on a world unlike the real one.
- Both failed checks are ours: we made the first-difference arm binding when smoke tests had shown it weaker.
- CRU temperature at these latitudes rests on few stations before about 1950, a recalled fact held with high confidence and not checked for this paper. CRU precipitation is filled with climatology at many cells, so the precipitation arm says little.
- 145 of the 161 sites are in Alaska and Canada, in 77 blocks. The block intervals carry that clustering; the binomial p values do not.
- Lag tests cannot see a same-year reverse effect. The known answer's reverse half, that a ring does not change the next summer, is taken as given.
- Results concern annual ring width at high-latitude conifer sites with lags of 1 and 2 years. A Granger pass says that one series improves the prediction of the other, and makes no claim about physical cause.
- Scoring was done by our own code, the second reading of the results was done by the same reviewer who ran the analysis, and the analysis was not blind.

10 Verified

Verified: on 9 November 2026 the site, region, block and length counts, every index, interval and rate in Table 1, the raw-arm values, the pairing mean and percentiles, the pairing p values, every binding and non-binding rule with its value and bar, and the same-year correlation were re-read from the run's result file. The family q values, the forecasts against outcomes and the post hoc coefficients were re-read from the study's results report and its post hoc output file, and the forecasts again from the preregistration. They match the findings ledger. The SHA-256 of the result file matches the hash in the run's manifest and the abbreviated hash in the ledger, and the SHA-256 of the preregistration matches the hash recorded in the result file and its lock. Both figures were drawn by a script that reads only the result file. Three constants were re-derived for the study without its code (one summer mean from the CRU file, one Granger p value and one site distance) and matched; that check was not repeated for this paper. The F676 values in the discussion were read from the findings ledger; its result file was not opened for this paper. The study was not re-run. The literature was found by five searches run on 9 November 2026 and read at the level of abstracts, publisher pages and search summaries; the full texts were not read.

11 Corrections

- 2 November 2026: the preregistration gave a smoke-test range of -0.31 to +0.12 for the planted forward index at real strength, which mixed high-pass and first-difference rows; the high-pass range alone is -0.14 to +0.12. Stated in the results report; the preregistration is unchanged.
- 9 November 2026: first published version.

12 Reproduction

- Data: ITRDB standard ring-width chronologies at or above 55 N, outside 12 W to 45 E, years 1901 to 2025 with at least 5 samples and a gap-free run of at least 60 years; June to August temperature from the nearest CRU TS 4.10 land cell.
- Preprocess both series three ways: high-pass with a 10-year Gaussian and scaling, first differences, and scaling only.
- At each site run eq. 1 both ways with lags 1 and 2, convert the p values to z-scores capped at 6, and form eq. 2; average by eq. 3 with a block bootstrap over two-degree blocks.
- Build the synthetic worlds of eqs. 4 and 5 and the no-coupling, common-driver and shared-trend worlds from five iAAFT draws per site; score eq. 6; run 199 far-site pairings and eq. 7.
- Preregistration SHA-256: a2f1dd47e803bf92b5d078f7845554df1f75e9ca8274805cc2049a5d099319ae. Plan SHA-256: 784d61f1c10e5d428ce7ad318d856f1aa9f591c24436a6d9b643e8791eb2e2c8. Run 20261102T094346Z. Result file SHA-256: 61ea2ec9829044fa56dde1110302cd4f7714471194e57a90 d4ea04f6dd2f3d93.
- Cost: the analysis took 131 seconds and the two pairing controls 31 and 33 seconds on one desktop CPU.

13 Appendix A. The preregistered verdicts and forecasts

Each verdict is a set of rules written before the run. A verdict passes only when every rule holds. Values are read from the run's result file.

Table 2. Preregistered verdicts, their role and the rules that decided them.

Verdict	Role	Deciding rule	Value, holds
Planted reverse, high-pass	binding	index interval upper end below 0; accuracy above 0.6	-0.303 yes; 0.714 yes
Planted reverse, first differences	binding	index interval upper end below 0; accuracy above 0.6	-0.042 yes; 0.584 no
No coupling, high-pass	binding	forward and reverse rates at most 0.10	0.055 and 0.078, yes
No coupling, first differences	binding	forward and reverse rates at most 0.10	0.066 and 0.072, yes
Shared trend, first differences	binding	forward and reverse rates at most 0.10	0.092 yes; 0.117 no
Strong planted forward, high-pass	power	index interval lower end above 0; accuracy above 0.55	+1.079 yes; 0.894 yes
Planted forward at real strength	power	index interval lower end above 0	+0.073, yes
Shared trend, high-pass	limit	forward and reverse rates at most 0.10	0.135 and 0.221, no, as forecast
Common driver, high-pass	limit	index interval contains 0	-0.193 to +0.023, yes
Exclusion	real data	reverse rate at most 0.10 on both arms; real index above planted reverse upper end on both arms	0.068 and 0.062; +1.206 and +1.107, yes
Forward, high-pass	real data	index above pairing 95th percentile (+0.213); interval lower end above 0	+1.206; +0.871, yes
Forward, first differences	real data	index above pairing 95th percentile (+0.186)	+1.107, yes
Decision	kill condition	any binding check failing means no standing	two failed; no standing

Table 3 sets the forecasts fixed in the preregistration against the outcomes.

Table 3. Forecasts fixed in the preregistration against what the run gave. Ranges are 80 percent ranges.

Quantity	Forecast	Outcome
Mean same-year correlation, high-pass	0.15 (0.05 to 0.25)	0.160
Share of sites with positive correlation	0.70	0.870
Real forward rate, high-pass	0.09 (0.05 to 0.15)	0.398
Real reverse rate, high-pass	0.07 (0.03 to 0.11)	0.068
Real index, high-pass	-0.05 (-0.30 to +0.20)	+1.206
Planted reverse index, high-pass	-0.6 (-1.2 to -0.3)	-0.429
Planted reverse accuracy, high-pass	0.68	0.714
Pairing 95th percentile, high-pass	+0.15	+0.213
Both planted reverse checks pass	probability 0.85	failed
Shared trend, first differences, passes	probability 0.75	failed
Exclusion rule passes	probability 0.60	passed
Forward rule passes	probability 0.15	passed
Planted forward at real strength seen	probability 0.10	passed
Exclusion status granted	probability 0.30	not granted
Forward status granted	probability 0.03	not granted

14 Glossary

Each term is defined where it is first used in the text. The same definitions, with the field each comes from and

what each excludes, are collected here.

- **known-answer test**. A test of a method on a case where the true answer is fixed by physics or by construction, so that any wrong reading is a fault of the method. Field: measurement and method validation. Excludes agreement with another method, which can share the same fault.
- **panel**. A set of units, here tree-ring sites, each observed over time and analysed together so that one statistic summarises all of them. Field: statistics and econometrics. Excludes a verdict reached site by site, and a pooled regression that treats all years of all sites as one series.
- **Granger test**. A regression test of whether the past values of one series improve the prediction of another beyond that series' own past; here an F test on lags 1 and 2. Field: econometrics and time-series statistics. Excludes a test of same-year effects, and proof of a physical cause.
- **preregistration**. A dated written plan for a study, fixed and hash-locked before any result is seen: the question, the data, the statistic, the controls, the forecasts and the kill conditions. Field: research method (metascience). Excludes a plan revised after looking at the outcome, and a protocol written only for the record after the run.
- **kill condition**. A rule written in the preregistration that, when met, decides against the claim with no room for judgement after the run. Field: research method (our usage, after Popper's falsification). Excludes reasons found after the run for setting a result aside.
- **binding check**. A preregistered rule on a synthetic world that the instrument must pass before any reading of the real data counts; under the kill condition a single failure removes all standing. Field: research method, our usage. Excludes power checks and limit checks, which were recorded but could not remove standing.
- **z-score of a p value**. The point of the standard normal distribution that leaves the p value above it, so that a smaller p gives a larger z; here capped at 6. Field: statistics. Excludes a standardised value of the data themselves.
- **panel direction index**. The mean over sites of d, the z-score of the climate-to-ring Granger test minus that of the ring-to-climate test; positive values point from summer temperature to ring width. Field: time-series statistics, our usage. Excludes a test of significance at any one site.
- **block bootstrap**. An interval made by resampling whole 2 degree blocks of sites with replacement, here 2,000 times, so that neighbouring sites that share climate move together and the interval is not too narrow. Field: statistics. Excludes a bootstrap over single sites, and a binomial interval that treats sites as independent.
- **detection rate**. The share of sites at which the Granger test in one direction gives p 0.05 or below; for synthetic worlds it is averaged over the draws. Field: statistics, our usage. Excludes a rate corrected for testing 161 sites.
- **high-pass filter**. A filter that keeps fast variation and removes slow trends; here each series minus a Gaussian smooth of width 10 years, then scaled to unit variance. Field: signal processing. Excludes first differences, which are reported as a separate arm.
- **first differences**. Each year's value minus the previous year's, a second way to remove slow trends that weights the fastest changes most. Field: time-series statistics. Excludes the high-pass filter, which removes trends by smoothing.
- **iAAFT surrogate**. A copy of a series built by repeated rounds of Fourier phase randomisation and rank reordering, so that it keeps both the original spectrum, and with it the autocorrelation, and the original set of values, while its timing against any other series is lost. Field: time-series statistics, surrogate data. Excludes a shuffle of single years, which destroys the autocorrelation.
- **direction accuracy**. The share of sites at which the sign of d matches the direction planted in a synthetic world; 0.5 is what a coin would give. Field: method validation, our usage. Excludes whether the test at any site is significant.

- **far-site pairing**. A control in which each ring series is paired with the summer temperature of a random site at least 1,000 km away, so that real coupling is removed while both series keep their own form; repeated 199 times. Field: statistics, our usage. Excludes pairings with nearby sites, which share regional climate.
- **Benjamini-Hochberg correction**. A rule that adjusts a set of p values so that, among the tests called significant, the expected share of false calls stays at or below a chosen rate, here 0.05; the adjusted value is called q. Field: statistics, multiple testing. Excludes the stricter Bonferroni rule, which bounds the chance of any false call.
- **control**. A version of the data with the structure under test removed (for example shuffled or pooled), scored the same way as the real data, so the real score has something to beat. Field: statistics and experimental design. Excludes a comparison with a different dataset or with a published number.
- **planted check**. Data built so that the right answer is known, run through the same analysis before the real data, to show that the analysis can see the effect when it is there. Field: research method; called a positive control in experimental science. Excludes any check made on the real data, which cannot show what the analysis misses.
- **time reversal**. The same test run on both series read backwards in time; a method that reads direction from lag order should flip. Field: time-series analysis. Excludes a control on coupling strength.
- **honest null**. A working test that found no effect beyond a stated bound, a forecast that failed, or an error of our own found and fixed, reported with the same weight as a positive result. Field: research method (our usage). Excludes a void run, and a result left unpublished because it was negative.
- **prior-summer coefficient**. The slope on the previous year's summer temperature in a regression of this year's ring width on this year's and last year's summer temperature. Field: dendroclimatology. Excludes the same-year slope, and any effect of the ring's own past unless it is added to the regression.
- **verdict**. The pass or fail of one preregistered rule set against the run's numbers, read off by code. Field: research method (our usage). Excludes an interpretation written after reading the numbers.

15 References

[1] Runge, J., Bathiany, S., Bollt, E. and others (2019). Inferring causation from time series in Earth system sciences. Nature Communications 10, 2553. doi:10.1038/s41467-019-10105-3

[2] Dumitrescu, E.-I. and Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. Economic Modelling 29(4), 1450-1460. doi:10.1016/j.econmod.2012.02.014

[3] Florin and colleagues (2010). The effect of filtering on Granger causality based multivariate causality measures. NeuroImage.

[4] Barnett, L. and Seth, A. K. (2011). Behaviour of Granger causality under filtering: theoretical invariance and practical application. Journal of Neuroscience Methods 201(2), 404-419. doi:10.1016/j.jneumeth.2011.08.010

[5] Anderson and colleagues (2019). On the sensitivity of Granger causality to errors-in-variables, linear transformations and subsampling. Journal of Time Series Analysis. Open access, PubMed Central PMC7814891.

[6] Tumajer, J. and Lehejcek, J. (2019). Boreal tree-rings are influenced by temperature up to two years prior to their formation: a trade-off between growth and reproduction? Environmental Research Letters 14(12), 124024. doi:10.1088/1748-9326/ab5134

[7] Lloyd and colleagues (2013). Nonlinear responses of white spruce growth to climate variability in interior Alaska. Reprint held by the USDA Forest Service, Pacific Northwest Research Station.

[8] International Tree-Ring Data Bank. NOAA National Centers for Environmental Information, World Data Service for Paleoclimatology. Ring-width chronologies. https://www.ncei.noaa.gov/pub/data/paleo/treering/

[9] Harris, I., Osborn, T. J., Jones, P. and Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Scientific Data 7, 109. doi:10.1038/s41597-020-0453-3

[10] Climatic Research Unit, University of East Anglia. CRU TS 4.10 monthly temperature and precipitation, 0.5 degree. Open Database License 1.0. https://crudata.uea.ac.uk/cru/data/hrg/