

Bijux Pollenomics: scientific evidence

What the methods estimate, what the receipts establish, and what remains open

Bijux Pollenomics

11 October 2026

Source-bound research briefing

Define the reconstruction terms before interpreting results

- **REVEALS**: Regional Estimates of VEgetation Abundance from Large Sites; estimates regional vegetation fractions from large-site pollen.
- **LOVE**: LOcal VEgetation Estimates; uses regional background and small-site pollen to estimate local fractions.
- The **Landscape Reconstruction Algorithm (LRA)** combines them. **PPE** means relative pollen productivity estimate; production, fall speed and transport connect grains to cover.

A standard error (SE) describes the declared model calculation. It is not automatically a confidence interval or a measured-cover error.

Evidence: `reports/science/reveals-theory-paper.md`; `reports/science/love-theory-paper.md`

Four quantities must not become one accuracy claim

Quantity	Denominator and unit	What tests it here?
Observed pollen	Named grains / eligible grain sum	Withheld-site pollen score
Regional REVEALS	Modeled taxon cover / regional target area	Author-code numerical replay
Local LOVE	Modeled taxon cover / chosen local area	Declared R/native fixture checks
Measured vegetation	Surveyed cover / matched landscape area	No admitted Nordic pair

A low pollen error, a small R/native residual, and a small modeled SE answer different questions.

Evidence: `reports/science/model-use.md`

LRA has a regional step and a local step

Large-site pollen + PPE + transport + basin

↓ REVEALS: regional cover over a declared area

Regional background + small-site pollen + local ring

↓ LOVE: modeled cover within that ring

- Sugita (2007a) tests regional REVEALS in POLLSCAPE simulations; the accessible abstract stresses sufficiently large, distributed lakes.
- Sugita (2007b) tests background subtraction and local LOVE in simulations; the relevant local source area remains hard to infer for a past landscape.

The two 2007 simulation papers are not reproduced by the later LandClim author example.

Evidence: `reports/science/reveals-theory-paper.md`; `reports/science/love-theory-paper.md`

A physical pollen-to-cover model needs more than counts

Input	Scientific role and failure if unspecified
Age model and window	Assigns a dated depth to an interval; a bin label is not a sample date.
Taxon mapping and count sum	Defines which grains enter the response denominator.
PPE and fall speed	Correct unequal production and distance-dependent transport.
Basin, atmosphere and ring	Set deposition geometry, regional reach and local estimand.
Regional cohort	Supplies LOVE background; shared target sites can defeat independence.

Parameter parity under one profile does not qualify those inputs for a different core.

Evidence: [reports/science/landclim-paper.md](#)

WHH05: one count, age and denominator trace

Source step	Exact retained value
Physical site	White Horse Hill 2005 lake; grid 512; 150 m radius.
Raw depth 16	1320 cal BP; 301 grains across 30 named fields.
Author preparation	Four named one-grain fields excluded; 297 grains on a 31-target vector; ten taxa positive.
Age assignment	1320 lies in [1250,1750) cal BP; output interval labelled 1500 BP.

The 297-grain denominator belongs to this prepared example row; the 1500 BP label is not a new date.

Evidence: `reports/science/publication-correspondence/author-example-site-replay.md`

WHH05: a modeled site value is not measured cover

1500 BP site estimate	Cover fraction	Model SE, fraction
<i>Calluna vulgaris</i>	0.9248830793	0.0468069771
Poaceae	0.0574176519	0.0154854693

- Author R and native code agree for the 31 WHH05 means and SEs: maxima 4.77×10^{-15} and 8.90×10^{-16} fraction.
- The nine site vectors enter grid 512's regional calculation; WHH05's site output is not the archived regional vector.

About 92.49% modeled *Calluna* cover is not 92.49% of the 297 counted grains.

Evidence: `reports/science/landclim-paper.md`

The LandClim paper publishes regional modeled history

- Githumbi et al. (2022) present 31 taxa, stated as 12 plant functional types (PFTs) and three land-cover types (LCTs) over 25 Holocene windows and European one-degree cells.
- Its 1,128 selected records yield 16,711 aggregated site/window counts, not 16,711 physical sediment depths.
- The authors' selected-window maps show modeled open-land increase from the mid-Holocene and declining modeled summer-green tree extent. These maps are not a measured-vegetation score or a single-cause attribution.

Original paper result: regional cover estimates with reported model uncertainty; our replay is a separate result.

Evidence: `reports/science/landclim-paper.md`

Author-code parity closes a bounded calculation

Bounded input population	Observed numerical correspondence
17 example lakes; two grids; 25 bins; 31 taxa Raw-to-prepared source replay	184/425 possible site-windows supported. 17 files byte-identical; 871 raw rows; 15,028 prepared cells checked.
Independent author-R/native execution	14,074 supported mean/SE values; max residuals 1.84×10^{-13} and 7.15×10^{-14} fractions.

This checks formatting and arithmetic for the released examples, not the paper's 1,128 historical inputs or vegetation accuracy.

Evidence: `reports/science/method-correspondence.md`

The archive differs on its only exactly aligned window

TW2, 100–350 BP; author R vs archive	Grid 512	Grid 562
Largest taxon mean gap, percentage points	9.19787	10.55270
Largest taxon SE gap, percentage points	3.21615	4.25814
31-taxon total variation, cover fraction	0.092406	0.110704

- Comparison: 31 means and 31 SEs per grid, author fractions converted to archive percentage-cover units; 124 paired values.
- Other author and archive intervals have different bounds, so they are not treated as identical-time rows.

These are two modeled products. Historical source revisions, preparation and chronology remain unverified; no causal diagnosis follows.

Evidence: `reports/science/publication-correspondence/published-window-provenance.md`

An SE envelope is not a validation interval

- At TW2, corresponding mean $\pm$ one-SE envelopes overlap for all 31 taxa in both grids. Unknown dependence and source identity prevent a significance interpretation.
- Archived ISTS and TBS *mean* sums close over 10,500 grid/window rows. The paper states 12 PFTs, deposit prose 11 and archive header 13 positions.
- Grouped SEs require their covariance/delta-method route; the archive-producing PFT/LCT code, ISTS meaning and deciduous-*Quercus* mapping are not authenticated.

A matched R residual, a grouped mean identity and a reported SE leave biological interpretation open.

Evidence: `reports/science/publication-correspondence/pft-crosswalk/README.md`

The 2024 LOVE paper and our fixtures have different sites

- **Original 2024 paper.** Eight artificial ponds in a simulated German landscape. LOVEr and LOVEoptim yield local-cover curves; the optimized method is slightly better for inner rings. The Figure 2 layout, generated pollen and random draws are unavailable here.
- The authenticated public publisher PDF supplies 151 numeric table cells through a source-bound extraction; this closes printed constants, not the simulation inputs.
- **Our later DISCOVER fixture.** One local/regional pair and 36 rings. The printed parameters match 30/30 numeric cells by row and 9/10 names exactly. The packaged BÜK300 example is not the paper's BÜK200 simulation.

The paper's 100–500 m tendency belongs to its simulation; no Nordic source radius is selected.

Evidence: `reports/science/loveoptim-paper.md`

Finite LOVE output can be invalid vegetation

- Constructed interface: 1,000 REVEALS draws $\times$ one 200–300 m ring $\times$ 31 taxa yield 31,000 finite LOVE cells; 8,312 are negative (minimum -0.46953).
- Separate packaged example: only 3/10 first rings satisfy its basin boundary; 16/30 valid taxon/ring cells are negative and six exceed one.
- The constructed local site also enters its regional cohort, so neither calculation is a held-out ecological comparison.

Negative or greater-than-one fractions remain visible as method diagnostics, never silently normalized into cover.

Evidence: `reports/science/love-theory-paper.md`

English forward pollen: direct baselines are lower here

14 paired English sites, 30 taxa Method	MAE, fraction		Hellinger	
	Direct	LRA forward	Direct	LRA forward
Training mean	.020496	.026273	.282082	.360671
Nearest site	.019552	.021132	.255767	.286001
Inverse distance squared (IDW)	.018579	.021088	.243673	.285415

MAE is mean absolute pollen-fraction error; Hellinger is composition distance from 0 (identical) to 1. Both average equally across sites; lower is better. Direct used all 14 training pollen sites; LRA used a regional anchor and 13 local reconstruction predictors. One of 15 outer folds was refused after three training fits failed. Predictions were frozen before target scoring, but sites were previously explored.

The score is withheld *pollen* composition, not vegetation reconstruction accuracy or a fair equal-information solver ranking.

Evidence: `reports/science/pollen-prediction.md`

Nordic observed-pollen transfer has support refusals

31-target country exclusion	Physical sites	Scored/attempted bins	Refused bins
Denmark	1	24/24	0
Finland	28	435/435	0
Norway	52	1,084/1,178	94
Sweden	3	51/54	3

Four prior-exposed country cohorts; these are closed observed-pollen subcompositions, with zero Nordic LRA admissions. Denmark has one target site.

Evidence: `reports/science/source-cohorts.md`

The 16-label and 31-target scores have different denominators

- **31 author taxa.** The Swedish country cohort assigns 221,800/240,966 native grains. Error is measured within this closed author target set.
- **16 literal labels.** Sämbofsjön covers 224,291/393,822 native grains; Lilla Gloppsjön covers 55,638/127,027. Error is measured within each lake's exact-label subcomposition.

A lower MAE on one axis does not establish a better model on the other; excluded and unmapped grains are not zero.

Evidence: `reports/science/pollen-prediction.md`

Independent Nordic vegetation validation is still empty

- EMPD2–NFI screen: 16 nearby leads, zero admitted matched measured-vegetation pairs.
- Three Småland bogs have published historical maps and pollen anchors, but our retained packet lacks the full distance-weighted map-class fractions needed to score our prediction.
- A southern Swedish published figure yields 5.11 versus 17.33 percentage-point mean absolute differences on *nine rounded printed labels*; it does not reproduce the original 26-taxon study or score this repository.
- LandClim archived modeled cover cannot serve as independent measured truth for a related input lake.

A validation pair needs one physical site, matched time, taxa, measured cover and spatial support.

Evidence: `reports/science/vegetation-reference.md`

Resolve source identity, then ask for qualified review

- ① For grids 512/562 at TW2, request exact historical sequence revisions, ages, count denominator, sites/basins, parameters and producing transformations. Confirm other window definitions and reuse terms.
- ② For one proposed Nordic core, bind selected chronology, dated basin geometry, taxon/PPE/fall-speed version and eligible count mass before LRA admission.
- ③ Ask a palynology/landscape reviewer to assess the physical assumptions; ask a reference owner whether measured cover matches the same site, time, taxa and area.
- ④ Freeze a new qualified cohort before target access; retain refusals and both direct-pollen and LRA-forward baselines.

Requests are prepared but unsent. Human scientific reading and independent ecological endorsement remain open.

Evidence: `reports/science/landclim-paper.md`; `reports/science/model-use.md`

Technical appendix: four-country source mass

31-author-target country	Samples	Assigned/native grains	Scored/attempted bins
Denmark	91	122,998/129,768	24/24
Finland	1,797	1,229,019/1,260,281	435/435
Norway	3,544	2,018,295/2,286,020	1,084/1,178
Sweden	228	221,800/240,966	51/54

NISP means number of identified specimens (grains here); counts and site-bin outcomes use different units. Assignment is a declared subcomposition, not a complete native flora.

Evidence: `reports/science/source-cohorts.md`

Technical appendix: membership is not prediction support

- Neotoma join: 12,388 source sample IDs; 5,660 selected across 84 sites, 6,218 outside selected sites, 233 at selected sites but unselected, and 277 without assigned source country.
- All 233 same-site unselected samples fail the source pollen screen; six selected Swedish negative-BP samples also fail their source age screen and are not scored.
- SEAD: 25,109 chronology claims at 853 assigned-country sites. All claim propagation is refused; these are neither pollen observations nor ecological events.

A source row, selected sample, attempted site-bin and admitted prediction are distinct denominators.

Evidence: `reports/science/neotoma-study-membership/README.md`; `reports/science/source-cohorts.md`

Technical appendix: three further Swedish lakes

- Giltjärnen, Lake Avegol and Flinkasjön were metadata-only additions outside the original three Swedish targets.
- 268/273 source samples satisfy the nonnegative-BP rule; five negative-age samples retain their native observations outside this study.
- 231,115/238,797 selected native grains map to 31 author targets; 63/63 site-bins yield 189 fixed predictions.
- Equal-site MAE: IDW 0.017334, nearest 0.017932, training mean 0.020682. Nearest wins at Lake Avegol.

Three prior-exposed lake groups; measured target is pollen, with no LRA or independent vegetation score.

Evidence: `reports/science/sweden-lake-transfer.md`

Conditional scenarios from recovered pollen counts

- Four Swedish cores supply explicitly complete small taxon panels: 261/273 source samples in 77 core/windows. Missing selected observations remain refusals.
- Wider parameter-transfer alternatives retain 267/273 samples in 79 core/windows under an explicit assumed-zero reporting rule and surrogate parameter choices.
- The Rootsi EKE workbook's 42 rows partition into 25 positive-feature and 17 zero-feature rows. This reconciles its row-count claim; 24 positive rows contain noninteger features.

Conditional modeled compositions and row accounting do not authenticate historical LandClim inputs, grain units, independent vegetation accuracy or scientific approval.

Evidence: `reports/science/landclim-paper.md`; `reports/science/publication-correspondence.md`

Technical appendix: Swedish training and source support

Matched 31-target pollen comparison	IDW equal-site MAE
Two Swedish training sites vs same two + 81 DK/FI/NO; 39 shared bins	0.032846 vs 0.023226
Same-source policy vs excluding centers within 100 km; 52 shared bins	0.023151 vs 0.029171

- Sweden-only refuses 15/54 bins; four-country predicts 54/54.
- Support-policy study produced 473/486 queries and refused 13; one-bin age shifts and nearest-five cap had smaller, site-dependent effects.

These are three exposed Swedish sites, not a spatial population or calibrated source radius.

Evidence: `reports/science/pollen-prediction.md`

Technical appendix: alternative age support changes eligibility

- **ABB / DALMUT LegacyAge.** Joined 107/107 and 71/72 depths; DALMUT 4.5 cm is unmatched. ABB has eight strict samples in four bins; paired IDW MAE moves .060058 to .059762. DALMUT has two strict samples and 20/22 bins unsupported.
- **Solsø LegacyAge.** Only 15/91 comparable samples fall in eight of 24 bins. On 16 labels, nearest MAE .032310 versus .032104 with parent-bin counts; a separate 31-target result is .025673.
- **Swedish 1,000 BP source block.** All 54/54 frozen target bins are scored; matched-bin IDW MAE moves .023647 to .024023.

Point ages, marginal ranges and nominal bins are not a joint age posterior or historical LandClim admission.

Evidence: `reports/science/pollen-prediction.md`

Technical appendix: two new Swedish pollen targets

- **Sämbosjön, 16 labels.** All 22/22 bins scored; IDW equal-bin MAE .022991. The 21 interval-contained bins yield .023054 versus original-count .022881. Labels contain 224,291/393,822 native grains; 53/111 samples are interval-contained. The score was inspected before protocol annotation correction.
- **Lilla Gloppsjön, 16 labels.** All 23/23 bins scored; mean MAE .016272 versus IDW .017295. Nearest Hellinger .228724 versus IDW .237576. Labels contain 55,638/127,027 grains; two samples are age-refused. The target partition was parsed in a prior validation screen.

Two distinct one-site development results; different age support and no untouched confirmation.

Evidence: `reports/science/sambosjon-direct-pollen.md`; `reports/science/lilla-gloppsjon-direct-pollen.md`

Technical appendix: one-window R lineage

- Original 2016 REVEALS function and pinned later Petr Kuneš source ran the same 225 BP input in four site configurations: one lake, unequal pair, six sites, nine-site grid.
- All 124 paired cover values and 124 SEs agree at retained CSV precision; full-grid bootstrap site indices were identical.
- Native historical profile agrees with later R on those fixed indices within declared tolerances.

One conditional lineage; distinct R function bodies and one window do not certify corrected uncertainty or ecological validity.

Evidence: `reports/science/method-correspondence.md`

Technical appendix: LOVE numerical correspondence

Declared source test	Result and limit
Original/current R and Python	43/43 core and 23/23 source-domain cases pass; 97 numeric CSVs byte-match native R; 48 original-R shared vectors agree.
One fixed ten-taxon LOVEexample	Eight ring/repeat candidate selections agree; 40 finite cover and 40 pollen cells match exactly; max ring-factor/objective gaps $5.11 \times 10^{-15}/4.67 \times 10^{-15}$.
Three shared optimized objectives	Both attain objective gaps $< 10^{-4}$ from independent optima; cross-evaluated candidate gap $< 10^{-10}$. Search paths and vectors need not match.

Declared arithmetic and objective quality, not the 2024 paper's eight-pond Figure 2 or real vegetation accuracy.

Evidence: `reports/science/method-correspondence.md`

Technical appendix: English LRA routes and refusals

Route	Population	Retained disposition
Analytical LOVE	31 taxa; 15 English folds	15/15 invalid compositions; all 54 local ring rows include negative cover.
Differential evolution	30 taxa; one pair, two requests	Both refused convergence; no held-out score.
Constrained SLSQP	30 taxa; 15 folds, 195 training items	192/195 fits complete; BECKHAM outer fold refused; 14 folds yield 42 scored predictions.

SLSQP is sequential least-squares programming. The selected 1250–1750 BP profile excludes 159 *Filipendula* grains from 20,732 source NISP; routes do not form a paired solver ranking.

Evidence: [reports/science/english-lra-pollen-prediction.md](#)

Technical appendix: target basin radius is a forward sensitivity

English target radius factor	80%	100%	120%
IDW-vegetation forward pollen MAE	.021032	.021088	.021457

- All 42 baseline LRA pollen fractions replay exactly before scoring; 126 radius projections use the same 14 target sites.
- Vegetation, regional cover and PPE are fixed, so only the forward consumer's assumed lake radius changes.

The shift exposes geometry sensitivity; it does not calibrate a physical radius or admit a Nordic basin.

Evidence: `reports/science/english-lra-pollen-prediction.md`

Technical appendix: tuning cannot be promoted across axes

Development search	Matched result
17-site radial basis function (RBF) vs fixed IDW, 31 taxa	174 common bins: MAE .017774 vs .018948; RBF covers 174/184 attempted vs IDW 178/184.
Country-excluded 16-label inner-fold RBF selection	Finland selected .034871 vs IDW .034375; Norway .033942 vs IDW .027012; Denmark selected IDW .033022.
Two-case synthetic PPE/transport sweep	Winner reflects a transport-law mismatch, not a real-core PPE recommendation.

Coverage, response axis and target exposure govern model use; no universal LRA parameter follows.

Evidence: `reports/science/model-use.md`

Technical appendix: context cannot fill the reference gap

- HYDE modeled cropland: 84 country/year/scenario grid sums match retained source totals within $8.51 \times 10^{-11} \text{ km}^2$; rights/version and historical spatial semantics remain open.
- LandClim Nordic map: 1,875 normalized published-cell features are modeled estimates, not independent field vegetation.
- Named core gaps: Abborrtjärnen needs selected chronology/basin/denominator/PPE; Solsø, Dalmutladdo and Laihalampi have source links but lack complete time-matched LRA input closure.
- The next ecological test still requires a separately governed measured-cover pair and qualified review.

A clean map export or historical context layer does not add an admitted vegetation truth label.

Evidence: `reports/science/source-cohorts.md`; `reports/science/model-use.md`

Technical appendix: country inputs still missing

Country	Named-core acceptance still needed
Denmark	Solsø: original 79 input rows, selected chronology and dated basin.
Finland	Laihalampi: selected age vector and historical basin for the 150-sample lead.
Norway	Dalmutladdo: historical basin and qualified count/parameter/use closure despite matched source controls.
Sweden	ABB/ABBO, BUS and BOKE: one jointly qualified core/window packet, including selected ages and basin history.

The 31-target country studies score 1,594/1,691 bins with 97 refusals. They admit zero LRA samples; no independent Nordic measured-cover benchmark or untouched confirmation follows.

Evidence: `reports/science/country-evidence-status.md`