

What pollen can tell us about past landscapes

Bijux Pollenomics: results, limits and decisions for collaborators

Bijux Pollenomics

11 October 2026

Source-bound research briefing

The question: what grew around a dated sediment core?

- Sediment preserves grains from plants near and far. Their counts record pollen; we want to understand vegetation and landscape change.
- A useful result must identify the physical core, sample age, counted taxa and the landscape area represented. Those decisions determine whether past forest or farming change can be interpreted.
- The project traces those inputs, checks selected published calculations and tests predictions on declared pollen targets.

We currently have observed-pollen prediction and bounded calculation evidence. We have no independent measured-vegetation accuracy score for our Nordic predictions.

Evidence: `reports/science/source-cohorts.md`; `reports/science/vegetation-reference.md`

Grain share and land-cover share are different quantities

What we observe	What a model tries to infer
Pollen fraction: a taxon's grains divided by a declared grain count.	Vegetation fraction: a taxon's cover divided by a declared landscape area.

Plants produce unequal quantities of pollen; grains travel and settle differently. A lake's size and type affect the incoming signal.

Productivity and transport corrections connect these quantities under assumptions. A grain percentage alone is not a measured vegetation percentage.

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

Regional and local reconstruction need different inputs

- **REVEALS**: Regional Estimates of VEgetation Abundance from Large Sites. Large-site pollen informs regional cover under declared transport and basin assumptions.
- **LOVE**: LOfcal VEgetation Estimates. Regional background and small-site pollen inform cover in a defined local area.
- Together they form the **Landscape Reconstruction Algorithm (LRA)**. Relative pollen productivity estimates (PPEs), fall speed, basin geometry and selected ages are inputs, not optional annotations.

A working implementation does not admit a new historical core. Its source identity, parameters, geometry, rights and physical interpretation need evidence.

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

A concrete source-to-result example: one lake sample

- ① White Horse Hill (WHH05), depth 16: source age 1,320 calibrated years before 1950, with 301 grains in named source fields.
- ② Four explicitly excluded grains leave 297 on the author's declared axis. The age enters [1,250, 1,750), labelled 1,500 years BP.
- ③ On this prepared example, author R and our implementation give modeled *Calluna* cover 0.924883 and standard error 0.046807, in fraction units.

The interval label is not a new date. This is a modeled site value, not 92.49% of the grains or measured vegetation. Numerical agreement checks this calculation.

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

Three checks answer three different questions

Check	Compare	Conclusion it can support
Calculation	Our numbers and source R on identical inputs.	Selected numerical methods correspond.
Pollen prediction	Frozen predictions and withheld observed pollen.	Pollen error on the tested cohort.
Vegetation accuracy	Inferred cover and independent measured cover with matched support.	Ecological accuracy on admitted pairs.

The first two exist within bounded scopes. The third remains unavailable for our Nordic predictions; another model's cover is not an independent measurement.

Evidence: [reports/science/method-correspondence.md](#); [reports/science/pollen-prediction.md](#); [reports/science/vegetation-reference.md](#)

Nordic pollen predictions: coverage and error differ

Target	Sites	Samples	Bins	Lowest MAE; method
Denmark	1	91	24/24	0.023242; nearest
Finland	28	1,797	435/435	0.017335; nearest
Norway	52	3,544	1,084/1,178	0.019459; IDW
Sweden	3	228	51/54	0.028740; IDW

Bins show scored/attempted support. IDW weights training sites by inverse squared distance; nearest uses the nearest eligible site. The target country is excluded from training. MAE is mean absolute error in pollen fractions on 31 author targets, aggregated equally across sites; lower is better within this declared comparison. A value 0.02 is two percentage points of mean fraction error.

Previously viewed development outcomes; method order is descriptive. Denmark has one site. Norway has 94 unsupported bins and Sweden three; these do not become zero-error successes.

Evidence: [reports/science/denmark-finland-norway-author-pollen.md](#); [reports/science/sweden-author-pollen.md](#)

Age support can change which observations are scored

- At Sämbošjön, the 16-label observed-pollen study scored 22 bins from 111 samples. Source age intervals cross bin boundaries for 58 samples.
- Keeping only contained intervals leaves 53 samples and 21 bins. On identical predictions, primary MAE is 0.023054 versus 0.022881 for original counts on those same bins.
- One bin has no contained sample and retains refused, null scores. Retained samples contribute 185,150 native grains, including 108,163 on the 16-label response.

One exposed core; the source bounds have no stated confidence level or joint age posterior. This sensitivity does not establish independent confirmation or vegetation accuracy.

Evidence: `reports/science/sambosjon-direct-pollen.md`; `docs/report/regions/nordic/sambosjon-direct-pollen/interval-support/report.json`

English LRA demonstrates a bounded executable workflow

- Start from 15 author sites in the 1,250–1,750 BP interval, with 30 pollen targets after a declared transport exclusion.
- Hold out one site at a time, reconstruct with the remaining sites, and freeze three forward-pollen predictions for that target.
- Fourteen sites produce all three predictions. One fold refuses after three training optimizations fail.

This checks reconstruction-to-pollen prediction on English author examples. No independent measured cover is available, and this does not admit Nordic LRA inputs.

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

On the same English targets, direct pollen error is lower

Method	Pollen MAE		Hellinger distance	
	Direct	LRA forward	Direct	LRA forward
Training mean	.020496	.026273	.282082	.360671
Nearest site	.019552	.021132	.255767	.286001
Inverse distance squared	.018579	.021088	.243673	.285415

Fourteen matched sites, 30 taxa; both metrics average equally across sites. Hellinger compares compositions on a 0–1 scale; zero means identical, and lower is better.

Training information differs between routes and the outcomes are exposed development evidence. Lower pollen error does not establish a better vegetation reconstruction.

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

`docs/report/pollen-reconstruction/author-lra-constrained/direct-pollen/paired-hellinger.json`

Matching author code does not reproduce the whole archive

- Selected LandClim examples: author R and our implementation agree on 14,074 supported means and standard errors. The largest fraction differences are 1.84×10^{-13} and 7.16×10^{-14} .
- Only one of 25 time windows has identical boundaries in example and archive: 100–350 BP. Archived mean cover differs by up to 9.20 and 10.55 percentage points in the two grids.
- The released example has 17 sites; the paper selected 1,128 historical records. Their original input revisions and archive functional-type transformations remain unverified here.

The archive gap is retained, but its cause is not identified. Numerical parity, historical source identity and ecological accuracy are separate questions.

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

Published vegetation references have useful limits

- Historical Småland maps provide coverage and pollen anchors; the numerical class-area fractions needed for a new score are unavailable.
- A southern Swedish published figure supports nine rounded label comparisons. It does not recover all original taxa or score our model.
- Namibian observations join ten soil-pollen sites to 15 vegetation-radius labels. Repeated radius rows are not new pollen samples, and the population differs from Nordic lakes.

Our Nordic plot/pollen screen admits zero of 16 nearby leads. None of these distinct sources supplies independent measured cover for our current Nordic predictions.

Evidence: `reports/science/vegetation-reference.md`; `reports/science/paper-accounts.md`

What collaborators can inspect and decide

- 1 Read the canonical paper account: question, concepts, worked example, original findings and our separate result.
- 2 Follow its exact table and source identities; use the reproduction route with lawful source access.
- 3 Help resolve a named core/age/basin/count/parameter question, historical archive transformation, or matched vegetation reference.
- 4 Review whether assumptions and spatial/time support justify the intended scientific conclusion. Actual reader feedback and qualified decisions remain open.

Local PDFs and source are review material. Scientific review, publication authority and a complete release certificate are not supplied by a successful build.

Evidence: `reports/science/paper-accounts.md`; `reports/needed/README.md`; `docs/public/pollenomics/science/papers.md`