

Why large-site pollen can estimate regional vegetation

Paper and question. Sugita, S. (2007), “[Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition](#),” *The Holocene* **17**(2), 229–241. This is the first Landscape Reconstruction Algorithm (LRA) theory paper, commonly cited as **Sugita (2007a)**. It asks when pollen from large depositional sites can estimate a region’s plant-cover composition despite heterogeneous vegetation. The distinction matters here because numerical agreement with a REVEALS program says nothing by itself about whether its inferred cover represents the vegetation around a particular core. [Indexed publisher citation and abstract](#); [our source-specific method account](#).

Scoped conclusion. In Sugita’s **POLLSCAPE simulations**, the publisher abstract reports accurate regional estimates with relatively small standard errors from multiple lakes at least approximately **100–500 ha** in size; much smaller lakes can vary strongly between sites. Its target region is **10⁴–10⁵ km²**. These simulated findings do not measure accuracy at a Nordic core. Our separate author-R/native check covers **two later LandClim II example grids**. [Publisher abstract](#); [two-grid execution](#).

The modeled quantity and source roles

A pollen percentage divides the counted grains assigned to a taxon by a defined grain sum. It is not that taxon’s share of land cover: taxa produce unequal pollen quantities, grains disperse differently, and a basin receives pollen from more than its immediate shore. **REVEALS** (“Regional Estimates of VEgetation Abundance from Large Sites”) uses large-site pollen and taxon-specific relative pollen productivity estimates (**PPEs**, called RPP in some sources), with dispersal and deposition assumptions, to estimate **regional cover fractions**. Pollen fall speed, lake or bog type and basin size influence the distance weighting. These input roles are explicit in the [later LandClim author-method profile](#); the accessible 2007 abstract establishes the model’s regional question and large-lake/multiple-site conditions but does not expose its equations or parameter tables. [Publisher abstract](#).

regional vegetation in a simulated or observed landscape
→ taxon-specific production and transport → large-site pollen counts
→ REVEALS with declared productivity, fall speed and basin assumptions
→ estimated regional taxon-cover composition

The original test used **POLLSCAPE simulations** across landscapes and atmospheric conditions. Its abstract gives no site or taxon count, scenario ledger, or numeric error distribution. Large lakes are the regional evidence unit; the companion [LOVE theory](#) uses that regional estimate to address background pollen at small sites. [Publisher abstract](#).

An exact later source-to-result trace

The trace below tests an **installed calculation**, not a 2007 observation. The released **LandClim II author example** identifies WHH05 as a grid-512 lake with a **150 m radius**. Its raw source sample at **depth 16, 1320 cal BP**, contains **301 counted grains** across 30 named fields; the author-prepared **1500 BP** interval vector has **297 grains** across ten positive taxa, with the four-grain difference accounted for by four named, one-grain excluded fields. The age falls inside [1250, 1750) cal BP; 1500 BP is the interval label, not the sample’s measured age. The prepared model uses 31 taxa, and the source parameter table has 31 PPE means and a 31 × 31 covariance axis. [Authenticated source-to-input trace](#).

The author R release and installed native engine then produce a **WHH05 site-level REVEALS** estimate for that 1500 BP window. In the retained paired row, *Calluna vulgaris* has author-R mean **0.9248830793**

and standard error (SE) **0.0468069771**, both fractional-cover units. The row is one modeled taxon fraction and its model SE, not a measured cover fraction. Across the 31 means the composition sums to one to numerical precision. The maximum R/native residual among the window's means is 4.77×10^{-15} , and among SEs 8.90×10^{-16} . WHH05 contributes **eight of 25** supported windows in this example. These exact source identities and denominators make a calculation check possible while leaving ecological accuracy untested. [WHH05 account](#); [paired rows](#).

The wider comparison covers **17 prepared lake sites** in two grids, **25 bins**, **31 taxa**, and **184 of 425** possible site-bins with available support. It matches **14,074 declared regional and site-level mean/SE values** within the recorded numerical tolerances. The released source is a later LandClim II author example using a 2016 REVEALS implementation; it neither reproduces Sugita's POLLSCAPE scenarios nor proves the 2007 paper's simulated error claim for historical vegetation. [Execution report](#); [method profile](#).

What remains unresolved and how to use this account

The indexed publisher abstract is the direct 2007 evidence available here. Direct full-article access did not yield a lawful text for this account, and the retained source packet examined contains no Sugita 2007 paper PDF or original POLLSCAPE scenario tables. Therefore original numeric scenario errors, standard-error formula, exact atmospheric regimes, taxon list and parameter values are **not evaluated**. For our later example, the 17 sites are all lakes, so its agreement does not check bog transport; its historical PPE covariance propagation and 1,128-sequence archive provenance have separate limitations. Even a small R/native residual cannot resolve those source and physical-interpretation questions. [Source-specific profile](#); [publication boundary](#).

Readers can inspect the [retained WHH05 trace](#) and [two-grid result](#) without private input bytes. To verify the two-grid governed execution report from the repository root:

sh

```
PYTHONDONTWRITEBYTECODE=1 PYTHONPATH=packages/bijux-pollenomics-dev/src \
.venv/bin/python -m bijux_pollenomics_dev.essd_reproduction.example_reexecution \
--verify-report docs/report/publication-replay/landclim-example-reexecution/report.json
```

A fresh R/native reexecution requires the separately identified lawful author snapshot and pinned R dependency, as specified in the [reexecution instructions](#). It cannot rerun the 2007 simulation.

The **unsent source request** is for a lawful full Sugita (2007a) paper, its original POLLSCAPE landscape and atmospheric scenario definitions, basin sizes and positions, taxon/PPE/fall-speed inputs, regional truth vectors, REVEALS outputs and error/SE calculations, plus usable source permissions. The **unsent qualified-review request** is to assess whether a proposed pollen cohort has sufficiently large and appropriately distributed sites, matched taxa and support, and a defensible parameter/uncertainty profile for its target region. No independent scientific endorsement, original figure reproduction or colleague reading outcome is recorded here.