

REVEALS validation against regional vegetation in southern Sweden

Citation and question. Hellman, S., Gaillard, M.-J., Broström, A. and Sugita, S. (2008), “[The REVEALS model, a new tool to estimate past regional plant abundance from pollen data in large lakes: validation in southern Sweden](#),” *Journal of Quaternary Science* **23**(1), 21–42. The journal issue is dated 2008, although the publisher records first online publication in July 2007. This is the validation paper called **Hellman et al. (2008a)** in [Gaillard et al. \(2010\)](#). It asks whether a pollen-based model can estimate present regional plant cover more faithfully than uncorrected pollen percentages. We use it as published empirical support for the method, not as a validation of predictions made by this repository. [Publisher citation and abstract](#); [our vegetation-reference decision](#).

The accessible publisher abstract reports good agreement for broad vegetation groups and reasonable estimates for named individual taxa. Our separate, narrower calculation from a later published figure finds a mean absolute difference of **5.11 percentage points** for REVEALS against observed cover, versus **17.33 points** for raw pollen, across **nine printed region/taxon labels**. Neither number is an error rate for all 26 taxa in Hellman et al., and neither scores our own model. [Exact retained rows](#); [calculation and selection](#).

What the comparison means

Pollen counted in a lake’s surface sediment is a share of *grains* in an assemblage. A plant-cover percentage is a share of *vegetated area* in a defined landscape. Those denominators differ because plant taxa release different numbers of grains and their pollen travels differently. **REVEALS** (Regional Estimates of VEgetation Abundance from Large Sites) converts large-site pollen assemblages into estimated regional vegetation abundance using taxon-specific pollen productivity estimates (**PPEs**) and dispersal assumptions. Fall speed and basin geometry affect dispersal and therefore the inferred source area. A correct model comparison must match taxon definitions, spatial support and observation time; it cannot treat raw grain shares as measured land cover. [Our method profile](#); [published Figure 6 explanation](#).

surface-sediment pollen from large lakes
+ taxon productivity and dispersal parameters
→ REVEALS regional cover estimates
↔ independently compiled regional vegetation cover

The publisher describes **26 key taxa**, surface sediments from **10 large lakes**, and modern vegetation in **Skåne and Småland**. It gives a regional model scale of **10⁴–10⁵ km²** and a compiled vegetation reference within **10⁴ km²**, drawing on remote sensing, inventories and agricultural statistics. The 26 taxa are not independent landscapes. The abstract omits lake membership by region, collection dates and numerical vectors. [Publisher abstract](#).

The authors report good estimates for broad groups (trees, herbs, conifers and broad-leaved trees) and reasonable estimates for several individual taxa. The abstract gives no per-taxon error table or threshold for “reasonable”; we assign no numerical error to its full cohort. [Publisher abstract](#).

One printed result, and its limited denominator

[Gaillard et al. \(2010\)](#), [Figure 6](#), p. 490 later displays southern Swedish pollen, REVEALS and compiled vegetation percentages for 26 taxa, naming taxa at or above its 2% display threshold and referring to both Hellman papers. In the **Skåne Pinus** labels, pollen is **13%**, REVEALS is **3%**, and observed vegetation is **3%**. The absolute differences against the printed observed value are $|13 - 3| = 10$ and $|3 - 3| =$

0 percentage points. This illustrates the correction on one rounded displayed label; it is not an original unrounded Hellman data row or an independently rerun model. [Exact transcription](#).

Our rule includes a region/taxon cell only when its name and whole-percentage value appear in all three panels. It admits **five Skåne labels and four Småland labels**, each given equal weight in the nine-row pooled mean. The nine rows are selected labels from **two regional compositions**, not nine sites; their rounded values and display selection cannot recover accuracy for the original 26 taxa. No standard error or confidence interval follows from our nine-label arithmetic. [Selection and result](#); [producer](#).

Evidence boundary and use

The publisher page exposes the citation and abstract; the full original paper and its input tables are not in the retained lawful packet examined for this account. The [Gaillard figure account](#) and our [calculation](#) verify a published *figure transcription*, not original ten-lake data or the Hellman model run. The later [sampling-design paper](#) asks a different question about sample selection, survey size and parameter choices. A difference between our nine-label mean and any original reported metric cannot be diagnosed without matching their units, taxon set, survey boundaries, parameters and rounding.

To repeat the bounded figure calculation with lawful access to the Gaillard PDF, run from the repository root:

```
sh
```

```
PYTHONDONTWRITEBYTECODE=1 .venv/bin/python \  
  reports/science/vegetation-reference/southern-sweden-figure/compare.py \  
  --source-pdf /path/to/gaillard-2010.pdf \  
  --output artifacts/jobs/southern-sweden-figure/reproduction
```

The producer checks the PDF against SHA-256 46bd92497d03169735b9a4665d71214c5dd8ff52ceddf39c577f0a4495fe73bd before writing the selected-label table and summary. [Retained data and method](#). It cannot rerun Hellman et al. (2008a).

The precise **unsent source request** is for a lawful copy of the full 2008a article and any supplementary tables; the 10 lake identifiers, basin sizes and region memberships; surface-sediment sample dates and pollen counts; the full 26-taxon mapping and unrounded REVEALS and vegetation-cover vectors; reference-map boundaries and compilation dates; PPE, fall-speed and dispersal versions; uncertainty inputs; and permitted reuse terms for those materials. A **qualified reviewer request**, also unsent, is to check the model-to-survey spatial and temporal match, the selected Figure 6 transcription, and whether proposed new vegetation data actually provide independent cover for the same estimand. No independent review or reader exercise is recorded here.