Schild et al. “Spatial Downscaling of Pollen-Based Vegetation Reconstructions into Paleo Land Cover Maps”

General comments:

The manuscript outlines a comprehensive three-step framework combining Elastic Net regression, the ordination-based “Pixel Flow” spatial method, and an autoregressive U-Net to reconstruct 20,000 years of continuous land cover across Alaska and western Canada at 300 m resolution. Transforming sparse pollen data into detailed spatial grids tackles a key challenge in paleoecology, and the authors thoughtfully maintain data provenance by including empirical source-area maps with the modeled grids.

However, the pipeline relies heavily on heuristic hyperparameters, scaling factors, and loss weights, with no systematic sensitivity testing, obscuring whether reconstructed patterns reflect genuine ecological dynamics or algorithmic priors. Evaluating accuracy primarily on training pixels conflates internal fitting with spatial predictive skill, leaving model performance in proxy-void regions unvalidated. This is compounded by a fundamental scale mismatch: downscaling regional pollen signals (which integrate over tens to hundreds of kilometers) to a 300 m grid risks implying unsupported micro-scale precision. Furthermore, omitting dynamic glacial constraints allows vegetation to be predicted beneath the Laurentide and Cordilleran Ice Sheets during the Last Glacial Maximum (LGM), which is physically untenable.

Establishing this framework as a defensible paleoecological benchmark requires rigorous spatial out-of-sample validation, systematic ablation and sensitivity analyses, dynamic ice masking, and a candid discussion of proxy resolution limits. The discussion should focus on these foundational methodological constraints rather than speculative downstream applications.

Major Scientific Concerns:

- Evaluating model performance across non-overlapping tiles yields overly optimistic metrics (U-Net model accuracy ranging from 0.75-0.90 and 0.80-0.90 in Sections 3.2 and 3.3) because strong spatial autocorrelation in climatic and topographic predictors persists across tile boundaries. To determine genuine predictive skill across proxy-void regions, the framework requires a spatial block or leave-one-site-out cross-validation scheme where geographic clusters of pollen records and their surrounding areas are systematically withheld during training.
- The architectural complexity of a U-Net requires quantitative justification against standard spatial interpolation baselines, such as spatial Random Forests, Gaussian Process Regression, or Inverse Distance Weighting. The authors must also include an ablation study that systematically removes the modern land-cover prior (13^{th} Channel), the autoregressive predecessor (14^{th} Channel), and the temporal consistency loss ($\lambda_{\text{temporal}}$) to confirm that the model captures authentic paleoecological dynamics rather than relying on temporal inertia and modern spatial patterns.
- Omitting dynamic ice-sheet boundaries allows the model to predict vegetation beneath the Laurentide and Cordilleran Ice Sheets during the LGM and the late-glacial transition (~20–12 ka BP), rendering early time slices physically and ecologically invalid. Rather than treating this in Sections 4.1 and 4.4 as an optional future improvement, the authors must apply an established time-transgressive glacial mask (e.g., ICE-6G, ICE-7G, or GLAC-1D) to exclude ice-covered areas before presenting the late-glacial reconstructions as valid scientific products.

- Linearly interpolating raw pollen records with millennial-scale sampling (often at coarse temporal resolution) into fixed 500-year slices creates synthetic time steps that are statistically collinear with adjacent points. Training sequential U-Nets on these slices optimizes the network on interpolated pseudoreplicates rather than on empirical observations. The authors should report the raw temporal-resolution distribution across the 119 time slices and, if possible, assess how temporal autocorrelation in the training labels affects the reported accuracy.
- Averaging five mathematically disparate metrics (U_{Reg}, U_{RDA}, U_{temp}, U_{Error}, U_{unc}) into an unweighted arithmetic mean in Equation 10 obscures the physical interpretation of the resulting uncertainty surface. The framework should instead provide disaggregated raster layers that separate input calibration error, spatial allocation ambiguity, and model epistemic and aleatoric uncertainty. Any single composite index must be supported by formal variance propagation or by statistical weighting.
- Converting discrete pollen counts into soft probability targets via $\mathrm{softmax}(-D/\tau)$ in Equation 3 risks causing class collapse, where distinct transitional classes (e.g., lichen and moss) are smoothed into dominant forest classes before training. The manuscript must provide the mathematical formulation, symmetry constraints, and normalization of matrix *D*, clarifying how it penalizes structural versus compositional dissimilarity.
- Key hyperparameters, including the ecological smoothing temperature (τ), Dirichlet concentration parameter (α_0), discrete channel weights (0.35), temporal loss weight ($\lambda_{\text{temporal}} = 0.18$), and the confidence gate threshold (0.60), are presented without tuning criteria. A sensitivity analysis of the start equation tau and lambda sub temporal, alongside a documented optimization protocol, is essential for methodological reproducibility.
- Inferring treeline shifts from the 90th percentile of forest elevation without controlling for basin hypsometry (like the case study in Section 2.3 and Section 3.4) confounds true ecotone migration with geometric area constraints and patch-size fluctuations. The reconstructed elevation distributions must be normalized against the underlying digital elevation model's hypsometric curve and validated against independent local proxy evidence, such as macrofossils or regional lapse-rate constraints.
- Applying a hard, winner-take-all selection based on edge-weighted confidence to stitch overlapping 256 × 256 tiles risks introducing linear seam artifacts across continuous spatial domains. The inference pipeline should compute a distance-weighted blend (e.g., linear or Gaussian feathering) across overlapping probability or logit tensors before argmax classification.

Specific/minor comments:

Abstract (if space allows within the character limit)

Lines 3–4. Briefly explain how this conversion is performed.

Lines 4–5. List the key environmental covariates and address whether non-analog conditions violate the assumption of ecological niche stationarity.

Lines 5–6. Specify the predictive drivers fed into the U-Net and explain how the architecture prevents edge artifacts or spatial hallucination across large proxy gaps.

Line 8. Replace or pair the computational volume (2.56 billion pixels) with the number of underlying empirical pollen sites to ground the dataset's actual empirical support.

Lines 8~. Report quantitative validation metrics (e.g., out-of-sample accuracy) alongside the qualitative uncertainty trends.

Lines 11–12. Clarify whether uncertainty is formally propagated through all three pipeline stages or computed as an empirical post-hoc heuristic.

Introduction

Lines 48–49. The sentence “Pollen data … of approaches” breaks the narrative momentum. Paragraph 3 focuses on the limitations of pollen and contrasts it with modern satellite data. Jumping back to a generic statement about reconstruction approaches at the very end creates an awkward transition into Paragraph 4.

Lines 64–66. Explicitly justify why a convolutional architecture (U-Net), designed for regular, dense grids, is superior to continuous geostatistical frameworks (e.g., Gaussian Process Regression, Spatiotemporal Kriging, or Hierarchical Bayesian Spatial Models) for interpolating sparse, irregular point proxies.

Lines 68–69. Pollen records measure taxon relative abundance (composition), whereas land cover represents physical surface properties (e.g., canopy closure, bare ground, shrub height, wetland fraction). Moving from plant composition to categorical/fractional land cover classes introduces a major conceptual conversion step that the authors gloss over. The Introduction should briefly clarify how taxonomical signals are mapped to structural surface cover.

Lines 73–74. Address the risk of structural circularity: explain how training the network on modern landscape patterns avoids projecting modern spatial autocorrelation onto non-analog paleo-vegetation.

Line 75. Explicitly acknowledge major late-Quaternary geomorphic boundary conditions (decay of the Laurentide and Cordilleran Ice Sheets, flooding of the Bering Land Bridge) rather than treating the region as a static, continuous landmass.

Methods & Results

Line 248. Section 2.2 states that missing values outside source areas are filled using spatial weighted means. Imputing proxy uncertainty via spatial averaging in data-void regions artificially suppresses uncertainty where empirical data are absent; please defend this formulation or provide an alternative.

Lines 347–351: When proxy density declines sharply prior to 10 ka BP, aggregate trends become dominated by the few surviving sites. Demonstrate that pre-10 ka transitions reflect genuine regional ecological shifts rather than sampling-network geometry (e.g., by evaluating a core subset spanning the full 20 ka sequence).

Lines 358–359: The text cuts off mid-thought: “As expected, accuracy is lower where pollen records overlap because multiple observers (Fig. 3e)”. The explanation for why multiple observers lower accuracy is missing.

Line 388. Assuming static regression and RDA uncertainties across deep time is unrealistic; modern transfer functions lose skill in non-analog late-glacial assemblages. Discuss how assuming constant early-stage error affects total downstream uncertainty.

Line 397. The elevational trajectory description lacks a figure reference.

Discussion

Lines 418~. Reconcile Section 4.1’s framing of the 300 m resolution with Section 4.3’s scale acknowledgment: state explicitly that 300 m pixels represent micro-topographic downscaling of regional pollen signals, not direct empirical detection of local patches.

Lines 431–433. Rather than treating this as optional future work, the authors should apply an existing dynamic ice sheet mask (e.g., ICE-6G, ICE-7G, or GLAC-1D) to mask out glaciated pixels before presenting 20–12 ka BP maps as valid scientific products.

Lines 531–533. Expand the discussion on spatial circularity to address how modern environmental space (13th Channel) limits the model's ability to reconstruct extinct or non-analog biome architectures.

Lines 534–536. Revise statements equating multi-site agreement in high-density areas with broad spatial accuracy; high performance in dense clusters does not validate model skill in proxy-void interiors.

Lines 547–549. Without dispersal constraints or cost-distance friction masks, plant functional types can "teleport" across insurmountable geographic barriers (e.g., mountain ranges, unglaciated corridors, or ocean channels) between 500-year time steps whenever environmental suitability changes. The Discussion needs to temper claims about "habitat connectivity" and "migration pathways" (Section 4.2) when the underlying modeling engine explicitly lacks dispersal physics.

Lines 612~. Substantially shorten forward-looking speculation on downstream applications (carbon budgets, fire regimes) in Sections 4.2 and 4.4 to prioritize the discussion of proxy resolution, error propagation, and spatial interpolation limits.