Review paper: egusphere-2025-1896

General View of the Paper:

This paper introduces and applies a pattern-based evaluation approach for Digital Soil Mapping (DSM) products, built around the concept of letting "the map speak for itself." Through three distinct case studies—BIS-4D in the Netherlands, the global SoilGrids v2.0 product, and SOLUS100 in the USA—the authors explore how aggregation, segmentation, and clustering methods can be used to evaluate whether DSM products reflect meaningful soil landscape patterns. This approach complements traditional point-based accuracy assessments by focusing instead on spatial coherence and landscape-level pattern recognition.

One of the strengths of this paper is that it goes beyond traditional point-based accuracy metrics and puts emphasis on evaluating the **spatial realism** of DSM outputs. This aspect is often neglected in many DSM studies, where maps are validated statistically, but not checked visually or structurally to see if they actually reflect real soil landscape patterns.

The case studies are carefully chosen to illustrate the strengths and limitations of the methods under different mapping conditions: one that performs well (BIS-4D), one global product with mixed success (SoilGrids), and one that fails to capture the landscape structure (SOLUS). This comparative structure strengthens the paper and offers valuable insights for practitioners working with DSM.

That said, the paper can be quite dense and difficult to follow in places due to the high level of technical detail. Long descriptive passages filled with numeric results might be better supported by well-structured summary tables. For example, in the Netherlands case study, a table showing different properties, segmentation scales, and the resulting number or size of segments would allow readers to quickly scan and compare outcomes across scenarios. Presenting complex results visually or in tabular form would improve readability and accessibility.

Figures are generally effective, Figure 11 (left), for instance, is particularly compelling, showing detailed patch delineation that appears well aligned with conventional maps. This kind of visual confirmation is rarely provided in DSM studies, yet it plays a crucial role in demonstrating the practical utility of such products. The paper succeeds in making a strong case for including spatial pattern analysis in DSM evaluation workflows.

In summary, the study is a valuable contribution to the DSM literature, highlighting a needed shift toward landscape-aware evaluation. It will be especially useful for practitioners seeking to validate their maps beyond traditional metrics and foster broader acceptance of DSM by bridging the gap between digital outputs and expert expectations from conventional mapping.

I would recommend this paper for publication in *EGU Sphere*, pending minor adjustments. Please find my detailed comments below:

Introduction

General terminology: Readers outside soil geography may be unfamiliar with specialized terms such as *supercells, consociation*, and *SCORPAN*. Please provide a brief parenthetical definition (or a short glossary) on first mention of each term.

Line 40: The phrase *"catena of Milne"* could confuse readers who do not know the historical context. Add the publication year to anchor the historical reference.

Consider wording such as:

"…the classic example is the *catena*, as defined by Milne (1935), meaning 'a sequence of distinct but pedogenetically related soils consistently located on specific slope facets…' (Borden et al., 2020)."

This makes it explicit that "Milne" refers to Geoffrey Milne's original definition of a catena.

Line 50: "At increasingly detailed scales and with increasingly fine distinctions in the definition of soil bodies, increasingly finer patterns are revealed" This sentence is conceptually clear but stylistically repetitive. The repeated use of "increasingly" and "fine" makes it feel redundant and slightly difficult to follow. I suggest rephrasing it for clarity and flow.

Line 97: The objectives paragraph states that the study will derive soil-landscape units via aggregation, segmentation, and clustering, and that these units can be used for routine DSM evaluation. However, it is unclear **how** the derived patterns will be incorporated into the evaluation itself (e.g., spatial correspondence to legacy polygon maps, pattern-matching indices, error aggregation by unit). A brief statement of the specific pattern-based metrics or comparison framework you intend to use would help readers see exactly how the new units feed into DSM evaluation.

Materials and Methods:

Line 105: Typo – "Aggregnation" should be corrected to aggregation.

Line 109: Typo – Remove the duplicate "in" in "defined in in multivariate space."

Line 109: Remove the stray period after the Lin (1991) citation.

Line 125: The sentence "Unlike supercells, segments must have rectilinear borders" could benefit from a brief explanation of the cartographic rationale behind this constraint. Why is this a requirement?

Figure 2: Please spell out the abbreviations "gpat_gridhis" and "gpat_gridts" in the caption (and first mention in the text), so readers immediately understand what each tool does within the workflow.

Line 143: For smoother readability, rephrase to: "Two important thresholds for joining grid cells into segments are:"

Line 146: "Spatial" does not need to be capitalized here.

Lines 175 - 177: Avoid using the vague term "better." Clarify what "better" means in this context, does it refer to lower inhomogeneity, higher isolation, or both? Clearly define the direction of desired improvement for each metric to avoid ambiguity.

Figure 3: Again, ensure that abbreviations like "gpat_polygons" and "gpat_distmtx" are spelled out in the caption, especially for readers unfamiliar with the GeoPAT suite.

Consistency issue: You shift between past and present tense (“Here we used…” vs. “Here we use…”). Choose either past or present tense and apply it consistently throughout the **Methods** section.

Case studies:

Case study 1 – BIS-4D (Netherlands)

Line 205: 54 maps (7 properties, 6 layers): fine, though “depth layers” may be clearer than “layers”.

Line 211: minarea was to 1 600 25 m x 25 m pixels” → typo → should be “minarea was set to 1,600 pixels (25 m × 25 m).

Line 221: classification of soil property maps for example: pH (0.1), clay (5 %), etc., are arbitrary to the reader and not clear enough how you choose these values. Provide a brief justification (measurement precision, agronomic relevance, histogram breakpoints).

Line 264: isn’t it 54 layers?

Line 265: It is not clear which depth for which soil properties has been used, rewrite the whole sentence: for example, pH, clay, silt and SOM for the depth of 0-5 cm, clay and bulk density for the depth of 15-30 cm and continue. Otherwise, the reader would be confused.

Line 281: As you mentioned it, segmentation is greatly affected by two thresholds, it doesn’t seem that default thresholds and liberal threshold would produce comparable maps. As I don’t see the segments capture the distinct pattern in figure 11 (right) with more liberal thresholds.

Section 3.3 Clustering: First paragraph: It is not clear how to follow cluster numbers? And how you link each cluster to the soil landscape components? But also state how the number of clusters is selected for this case.

Line 305: “due to the extremely high quality”, it would be nice to add the quality of source data for better comparison.

In general, it would strengthen the paper to include basic descriptive statistics and accuracy/error metrics for the DSM products or baseline maps used in each case study. For example, if the maps are described as having “good accuracy,” providing supporting metrics—such as RMSE, R^2, or standard deviation for selected properties (e.g., pH)—would give readers a clearer understanding of the underlying data quality. These statistics could be included in the Supplementary Materials if space is a concern, but they would offer helpful context when interpreting the segmentation and aggregation results.

Case study 2 – SoilGrids v2.0 (Global)

Line 338: mineara” → typo → should be **minarea**

Line 356: “many segments seem to be of a single class” → vague. Maybe clarify in this way: “...many segments contain only one SOC class, offering little internal variability.”

Figure 17: the highest standard deviation appears to be approximately **6.1**, but in the text (Line 346), it is stated that the standard deviation ranges from 0.34 to **6.08**. Please double-check this value, there may be a small discrepancy between the figure and the text.

Section 4.4: While Section 4.4 offers a qualitative summary of the results from aggregation, segmentation, and clustering, it remains unclear how these steps contribute to the evaluation of DSM products, which is a central stated goal of the paper. The discussion is primarily descriptive and does not explain whether the derived spatial units are being used to validate DSM predictions, identify mapping artifacts, support field design, or quantify model accuracy. To strengthen the study's impact, I recommend adding a brief clarification of the intended **evaluation framework**—for example, whether it involves comparison to legacy soil maps, pattern-based quality metrics, or expert validation. This would help readers understand how these methods serve as tools for DSM assessment rather than just spatial analysis outputs.

Case study 3 – SOLUS100 (USA)

Line 419: The mean is area 5.30 ha -> the mean area is

Line 424: The statement "From this we conclude that SOLUS in no way represents the actual soil pattern*"* may be too strong without formal pattern-matching metrics or more comprehensive property comparisons. Consider softening this claim or supporting it with a visual or statistical metric beyond visual inspection.

The authors clearly demonstrate the challenges of applying the DSM approach of SOLUS100 in a glaciated landscape with fine-scale patterning. Despite methodical application of aggregation and segmentation, the results did not reflect landscape features accurately, suggesting a mismatch between the DSM product's input data and the local soil-forming processes.

Discussion:

The authors acknowledge limitations in parameter selection, but what guidance can they offer for users who are not yet experts? This study suggests that appropriate parameter selection is crucial for success, otherwise, the methods may fail to produce meaningful results. For instance, the text states, "There is no objective way to adjust compactness and supercell number parameters." It would be helpful if the authors could recommend whether any metrics (e.g., internal variability, boundary length, or other summary statistics) might serve as semi-objective tools for parameter tuning, even as part of a sensitivity analysis.

The discussion refers to traditional mapping units such as polypedons, consociations, complexes, and associations. It would be helpful if the authors clarified whether their derived units are intended to correspond to these traditional concepts, or whether they serve a different function within the context of DSM evaluation. As the authors note, traditional surveys themselves can be inconsistent—particularly when comparing products like gSSURGO in the U.S. and the INEGI map in Mexico. This raises the question of which map should be treated as the baseline for comparison, especially when these traditional sources vary in quality and methodology.

Conclusion:

The conclusion effectively summarizes the intent and broader value of the methods, particularly the shift toward spatial pattern-based evaluation of DSM products. However, the claim that the map is allowed to "speak for itself" might be reconsidered or clarified, since the process still depends heavily on analyst-defined parameters.