Author's response
Use of Spatial Embeddings in Genosoil Identification

Julio César Pachón Maldonado[1], José Padarian[1], Quentin Styc[1], and Alex McBratney[1]

[1]Sydney Institute of Agriculture and School of Life and Environmental Sciences, The University of Sydney, Eveleigh, NSW 2015, Australia

Dear Professor Rossiter,

Thank you and the two reviewers for the careful and constructive review of our manuscript egusphere-2026-1944, "Use of Spatial Embeddings in Genosoil Identification". We have addressed every comment.

We added a sensitivity analysis showing how candidate areas respond to the cosine threshold (Table S7). Land-cover composition and Sentinel-2 imagery show the present cover of the HMI-defined reference areas (Table S8, Fig. S1); we present these as supporting context, not validation. The complete WRB class list with Shannon diversity documents taxonomic heterogeneity within each pedogenon (Table S4). Every metric is now defined in one table (Table 2), and the figures were revised in response to the comments. Discussion section 3.6 has an expanded discussion on the quality of the input datasets, taxonomic heterogeneity, time as a soil-forming factor and land-use history, and the Conclusion was rewritten to state the most productive use of the embeddings and the next steps.

Additionally, the 10 m candidate genosoil maps have been added on Zenodo under the reserved DOI; a preview link is given in our reply to Reviewer 1, Minor Comment 15.

Yours sincerely,

Julio César Pachón Maldonado

on behalf of all authors

RELEVANT CHANGES

1. Due to an issue with the external drive where data was housed, we completed a reanalysis which led to minor improvements to ensure pedogenon, HMI and country are sampled at the centre of each 10 m embedding pixel; overlapping AlphaEarth tiles are counted once; reference centroids are rebuilt from the raw reference pixels; and areas are computed on the WGS84 ellipsoid. This last one is an improvement from planar global extents previously reported based simply on pixel size without latitude correction. This recomputation led to small differences in values but not to the conclusions derived from those values. The 10 m results are also now available in the Zenodo companion.

2. Based on the review responses, we expanded the analyses. Cosine-threshold sensitivity (0.85, 0.90, 0.95) for both products and all three pedogenons (Table S7); land-cover composition of the HMI-defined reference areas and of more modified areas (Table S8) with 2022 Sentinel-2 imagery of the P1564 reference areas (Fig. S1); the WRB class list with Shannon diversity (Table S4).

POINT-BY-POINT RESPONSES

Comment R1-1.1: *The HMI dataset is used to identify undisturbed genosoils. However, especially for pedogenon 1564, I have doubts whether this dataset actually shows minimally disturbed soils. Their minimal presence and highly scattered occurrence (Table 1, Fig. 3) raise the question whether they are actually genosoils, or just wrongly classified phenosoils in the HMI dataset. An evaluation of Table S2 shows that the most similar countries to the UK also have a very small percentage of genosoils in their pedogenon, while countries with much higher percentage of genosoils show lower similarity. This makes me wonder if the HMI actually identifies genosoils for P1564, or whether wrongly classified agricultural soils or built-up areas are compared with each other, which could also be an explanation for their similarity in European context. A thorough evaluation of the accuracy of the HMI dataset is essential before spatial embeddings of derived genosoils can be reliably compared.*

Response: We followed previous publications where HMI is used as an operational proxy following Roman Dobarco et al. (2023) and Francos et al. (2025a). However we now describe all HMI-derived populations as candidate genosoils. We tested the reviewer's hypothesis that the reference areas are misclassified agricultural or built-up land, we compared them with ESA WorldCover 2021 and CORINE Land Cover 2018 (Table S8) and inspected 2022 Sentinel-2 imagery of every P1564 reference area (Fig. S1).

Changes in manuscript:

1. New Table S8 and Fig. S1.

- Table S8: "Land-cover composition of HMI defined candidate genosoil areas (HMI $\leqslant$ 0.10) and areas with higher human modification (HMI > 0.10) within the UK for the pedogenons studied, using ESA WorldCover 2021 v200 (Zanaga et al., 2022) and CORINE Land Cover 2018 (European Environment Agency, 2019)."
- Fig. S1: "(a) Extent of pedogenon 1564 in the UK, with black squares showing the locations of the all the HMI derived candidate genosoils (18.88 ha) in the pedogenon. (b) Median composites of Sentinel-2 imagery from May–September 2022 … Red outlines mark boundaries of HMI derived candidate genosoils."

2. Revised Sect. 2.3, p. 8, L176–180: "For each of the three pedogenons, reference pixels were selected from valid 10 m embedding pixels whose centres corresponded to HMI $\leqslant$ 0.10 in the 90 m HMI map, following Styc et al. (2025) who combined "very low" and "low" modification categories; HMI $\leqslant$ 0.01 and 0.01 < HMI $\leqslant$ 0.10 respectively (Theobald et al., 2025). Stricter thresholds of HMI $\leqslant$ 0.01 and $\leqslant$ 0.05 left no reference pixels for either P845 or P1564 within the UK." Previously: "genosoil candidate

pixels were defined as 10 m pixels falling within 90 m HMI cells where HMI $\leqslant$ 0.10, following Francos et al. (2025a)."

3. Added in Sect. 3.2, p. 15, L302–305: "Across the 92 country–model profiles with at least two available modified HMI classes, 82 showed a higher distance in the highest available modified class than in the lowest, and 65 were non-decreasing across the available modified classes (Fig. 5). This indicates that, within a given pedogenon and country, the low-HMI candidate genosoil centroid defines a consistent local reference state in embedding space, from which progressively modified pixels tend to diverge."

4. Added in Sect. 3.6, p. 22, L447–p. 23, L451 (land-use history): "Limited transferability does not necessarily indicate a limitation of the GPM since these results may be due to land-use history, which can alter the present condition of a soil within the same pedogenon. Previous cultivation, drainage or plantation forestry can leave persistent changes in soil properties that are not apparent from present vegetation cover or a low HMI value. The land-cover comparison (Table S8) and satellite imagery (Fig. S1) provide supporting context for selecting candidate reference areas, but historical information and field characterisation remain necessary to assess their suitability as genosoil references."

5. Revised terminology throughout: HMI-derived pixels, populations, centroids and maps are "candidate genosoils".

Comment R1-1.2: *Another concern comes from the uniformity of the pedogenons. Figure 5 shows occurrences of the same pedogenon in countries with a wide variety of climatic and topographic conditions, and the pedogenons sometimes contain soils that are formed under contrasting climatic conditions (Table S4). This heterogeneity will make it difficult to define one genosoil type with corresponding spatial embedding for each pedogenon. This point is addressed in Section 3.6, but in my opinion with insufficient detaiLI would like to see a more extensive discussion of how the quality of the used datasets could influence the outcomes of this study, and also see this reflected in the conclusions.*

***Response*:** Section 3.6 was expanded to discuss how the quality and nature of each dataset affect the results: the GPM as a covariate-based framework rather than an inventory of field-validated soil bodies, the missing time factor, the indicative quality of the SoilGrids WRB map, taxonomic heterogeneity within each pedogenon (Topic Editor comment 2), land-use history, and the dependence on HMI. The Conclusion reflects these points through the limits of cross-border substitution and the field-verification and time-covariate next steps. WRB class areas and the Shannon diversity table are deposited with the code that produces them, and Table S4 now lists every mapped class.

Changes in manuscript:

1. Revised and expanded Sect. 3.6:

L421: "Soil properties can differ considerably even over short distances." Previously: "The Global Pedogenon Map should be regarded as an initial global framework for pedogenon-based soil assessment."

L424–432: "At the same time, our results indicate that a shared pedogenon label may still encompass very disparate soils, and fieldwork is needed to determine whether this reflects limitations of the pedogenon map, such as too few classes, need for more variables, or instead reflects anthropogenic modification that may be reversible if the pressure subsides but currently leads to very disparate soils." Previously: "… a shared pedogenon label may not, by itself, ensure pedological equivalence across disparate regions, and that the internal heterogeneity within a pedogenon influences both the coherence and transferability of its genosoil signature. Time, one of the soil-forming factors, spans from extremely

recent events such as landslides, to glaciation events tens of thousands of years ago, to highly stable soils that have developed over millions of years. Román Dobarco et al. (2021a) included the Weathering Intensity Index from airborne gamma-ray spectrometry across Australia to account for time; however, these data are not yet available globally. The pedogenic variable of time is therefore an important area of research to improve pedogenon maps."

L433–446: "The three case studies illustrate differences in mapped taxonomic diversity and embedding behaviour. P2932 demonstrates the lowest WRB diversity ($H' = 0.73$; Podzols 73 %, Cambisols 21 %; Table S4), although its US reference populations were less cohesive than the UK populations (Table S2). Conversely, P1564 is taxonomically heterogeneous ($H' = 1.44$; Table S4), and the similarity of its reference centroid to the UK is higher within temperate Europe than beyond this region." … "This interpretation requires caution because Hengl et al. (2017) reported map purity below 50 % for Phaeozems, something further emphasised by a classification discrepancy between SoilGrids and gSSURGO, which classifies the US occurrence of P1564 primarily as Udipsamments (Table S5)." … "… its global distribution extends to tropical and subtropical regions such as Japan, Nepal, China, and Peru, which together account for nearly 50 % of the global P845 extent. … The embedding results corroborate this disjunction: most non-Atlantic countries showed low AlphaEarth cosine-to-UK values, although Georgia (0.31) and the United States (0.22) were above 0.20 (Table S2), indicating that the shared pedogenon label does not necessarily imply a transferable genosoil reference." Previously: "The three case studies exemplify the relationship between taxonomic homogeneity and embedding behaviour. P2932 … effectively representing a single-soil-body pedogenon, with its genosoil signal transferred solely to Scandinavia …" Removed: the statement that the US Typic Udipsamments of P1564 represent "a fundamentally different pedogenic pathway from the humus-rich Phaeozems of central Europe".

2. Rewritten Conclusion; it now reads:

"The delineation of genosoil in this study was based on the current globally available operational reference, the HMI. However, the results demonstrate that spatial embeddings can extend this framework beyond the limitations of HMI alone. At a resolution of 10 m, the embeddings did not merely reflect HMI classes; rather, they captured a reproducible surface-state signal within each pedogenon that was generally associated with lower levels of human modification and consistent local reference conditions. Even where genosoil availability was limited, the reference populations remained internally cohesive (Table S1), and within reference regions, the pixels with the lowest HMI values were typically the most similar to the candidate genosoil centroid. This suggests that spatial embeddings are more useful for identifying and ranking plausible genosoil candidates than for exact classification. Across the three case studies, similarity patterns followed biogeographic and pedological coherence more closely than pedogenon labels alone: P2932 remained largely boreal, P845 was concentrated in nearby European analogue regions, and P1564 linked the UK most strongly with temperate European analogue regions. The clearest positive example was P1564, in which severe local depletion still permitted partial recovery of the UK pattern, more distinctly under AlphaEarth than under Tessera. Conversely, cross-border substitution, as implemented here within the proposed workflow and thresholds, needs improvement, with two of the three pedogenons showing no meaningful replacement of the local delineation by non-local references.

The country analysis, in Table S2, also showed how scarce reference soils have become: P1564 retained no HMI-defined candidate genosoil in Belgium, France or Turkey, and less than 1 % of its area in nine further countries, including the UK and Germany. For such pedogenons, reference populations identified with embeddings, locally or in other countries, may offer the only practical reference state, underlining the importance of this research.

Two steps follow directly from these results. First, the reference populations identified here should be checked in the field: the method selects candidates, and only profile description can confirm that a low-modification surface signature corresponds to a minimally disturbed soiLSecond, pedogenon products

would benefit from integrating a covariate for time such as the Weathering Intensity Index; however, these datasets are not yet available globally. Operationally, the approach can identify candidate reference sites at 10 m needed to understand soil condition, in which soil measurements are interpreted against a minimally disturbed soil of similar pedogenesis."

3. Added in Data availability, L484–488: complete WRB class areas and Table S4 diversity summary at https://doi.org/10.5281/zenodo.19424156; Table S4 now lists every mapped class.

Zenodo doi is reserved but not registered until publication is accepted. Preview link is https://zenodo.org/records/19424156?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjU4NzA5MTQwLTk1Y2EtNGZmNS05MmMwLTcyNmNjMTZhODk5ZCIsImRhdGEiOnt9LCJyYW5kb20iOiIzYTcwNzg3YTQwNWIyODQwZjJkYWVhNTRhNDY5ZmNmNSJ9.6Dwt8tRhJQox3evJN0T63nh-wan6--UBnE6Ut8VptmMqf_b8CGBWJ-2PsFL31t5ks2U6TYhwcICYf0gZ5hYZmQ

Comment R1-2.1: *The authors remark that spatial embeddings "do not directly encode pedogenesis" (lines 83-84). However, many of the land surface properties of the spatial embeddings correspond to the soil forming factors, where Tessera seems to mainly focus on organisms through land cover from Sentinel, while AlphaEarth seems to represent topography and climate as well (lines 76-82). I think it would benefit the paper to frame the spatial embeddings in the context of soil forming factors, or a comparable model, to connect to more familiar frameworks in soil science. It would also be interesting to see a discussion on how the lack of representation of the other factors, especially time, could have an influence on the outcomes.*

Response: Section 2.2 now states how each product relates to the soil-forming factors: AlphaEarth is generated from satellite imagery only but trained to reconstruct environmental variables including elevation and climate, so its embedding space is organised to be predictive of relief and climate; Tessera uses no climate or topographic data at any stage and represents the stable spectral–temporal signature of the surface. Section 3.4 uses this distinction to explain why AlphaEarth separates distant countries more sharply than Tessera. Time is now discussed in Sect. 3.6.

Changes in manuscript:

1. Revised Sect. 2.2, L142–148: "AlphaEarth consists of 64-dimensional embeddings at 10 m resolution derived from satellite imagery (Sentinel-1, Sentinel-2, and Landsat 8/9) and trained to reconstruct a diverse set of environmental data sources, including elevation, climate, and land cover (see Brown et al., 2025, Table S1, for the complete list)." … "… the 64 dimensions capture a wide range of environmental variables, reproducing temperature (R^2 = 0.97) and elevation (R^2 = 0.96), along with reconstructing 12 of 26 variables at $R^2 > 0.90$. Conversely, anthropogenic variables such as nightlights and population density were less prominently reconstructed (R^2 = 0.58 and R^2 = 0.56 respectively)." Added L154–155: "Unlike AlphaEarth, no climate or topographic data were used to develop Tessera either as input or reconstruction target."

2. Revised Sect. 3.4, L372–L375: "AlphaEarth's lower similarity to distant countries may reflect stronger discrimination of climate and relief, which were included as reconstruction targets (Brown et al., 2025); Tessera, trained on Sentinel-1 and Sentinel-2 time series (Feng et al., 2025), retained higher similarities across countries." Previously: "… AlphaEarth, trained on a broader set of multimodal inputs including climate variables and topography (Brown et al., 2025), retains a greater degree of local environmental context, thereby yielding sharper spatial discrimination but reduced cross-border similarity."

3. Added in Sect. 3.6, L428–432: "Time, one of the soil-forming factors, spans from extremely recent events such as landslides, to glaciation events tens of thousands of years ago, to highly stable soils that

have developed over millions of years. Román Dobarco et al. (2021a) included the Weathering Intensity Index from airborne gamma-ray spectrometry across Australia to account for time; however, these data are not yet available globally. The pedogenic variable of time is therefore an important area of research to improve pedogenon maps."

Comment R1-2.2**:** *Next to that, despite the remark that spatial embeddings "do not directly encode pedogenesis" (lines 83-84), the authors actually suggest to use the spatial embeddings to "flag pedogenons whose global extent may conflate distinct soil-forming environments" (lines 409-410). I think these two statements contradict each other and need revising.*

Response: As explained in online response, the first statement is mechanistic and the second diagnostic. To avoid the apparent contradiction, the paragraph containing lines 409–410 was removed; the statement at lines 83–84 is retained (now L83–84).

Changes in manuscript:

1. Removed paragraph from former Sect. 3.6: "These observations suggest that embedding-based genosoil analysis can serve a dual diagnostic role: it identifies analogue regions within the operational range of the GPM label, and simultaneously flags pedogenons whose global extent may conflate distinct soil-forming environments. Future iterations of the pedogenon framework may benefit from incorporating such feedback to refine class boundaries, particularly for pedogenons with high taxonomic diversity and low cross-border embedding coherence."

Comment R1-2.3: *Statistical terminology. The manuscript introduces various statistical concepts and metrics that are not consistently named and referenced throughout the manuscript. For example, lines 164 – 166 uses the terms "internal coherence", "cohesion" and "cosine similarity" to describe the representativeness of a reference embedding to its population. Throughout the rest of the manuscript, these terms are used interchangeably. Other important metrics are not defined by equation, but as a quick mention in the main text (e.g. cosine distance, lines 176-177). I think that the paper can benefit from more strict use of statistical terminology and a clear overview of their definitions and descriptions, for example in a table.*

Response: We added Table 2, which defines every metric with its equation and range, and aligned the terminology of the text and equations with it: "cohesion" is the mean cosine similarity of a population to its centroid, "cosine similarity" compares two vectors, and "cosine distance" is 1 minus the cosine similarity.

Changes in manuscript:

1. New Table 2, L238–240: "Definitions of the similarity, separation and overlap metrics used throughout. Cohesion and the silhouette summarise the embeddings; the Jaccard index and the inclusion fraction summarise the spatial delineations those embeddings produce and are computed on an area basis." Rows: L2-normalisation, cosine similarity, cosine distance, reference centroid, cohesion, silhouette, Jaccard index (IoU) and inclusion fraction, each with definition, equation and range.

2. Revised Sect. 2.3, L171, and Eqs. (1), (2) and (4): "Table 2 summarises the metrics used in this study."; the equations now use the notation of Table 2.

Comment R1: Most Tables and Figures present results for three pedogenons and two spatial embeddings. Their order of presentation is however not consistent, where some Figures present the pedogenons as rows (Fig. 3) and others as columns (Fig. 4). The order of presentation of the pedogenons

also varies between Figures, Tables and their description in the text. I think the manuscript will be much clearer when the pedogenons and spatial embeddings are consequently presented and described in the same order throughout the Methods and Results.

Response: Figures and tables revised. Pedogenons are presented in the order P845, P1564, P2932, and products in the order AlphaEarth, Tessera, throughout the text, tables and figures.

Changes in manuscript:

1. Revised Sect. 3.1 (L256–285), Sect. 3.2 (L306–323), Sect. 3.3 (L343–360), Table 3, Fig. 5, Fig. 6 and Table 5 to follow the order P845, P1564, P2932.

2. Revised Fig. 4 caption, L277–278: "Candidate genosoil maps for pedogenon 845 (top row), 1564 (middle row), and 2932 (bottom row) … AlphaEarth reference embeddings (middle column) and Tessera reference embeddings (right column)." Previously: "… 845 (top row), 2932 (middle row), and 1564 (bottom row) … Tessera 10 m centroid embeddings (middle column), AlphaEarth 10 m centroid embeddings (right column)."

R1, minor comments

Minor Comment 1: *The abstract contains a lot of technical terms which will not be understandable without reading the manuscript first (e.g. spatial embeddings , line 12; reproducible low-modification surface-state signal, line 16-17). Please make sure that the abstract is in itself understandable to the audience of SOIL.*

***Response*:** We defined spatial embeddings on first use, replaced the surface-state phrase with plain language, and removed the final sentence.

***Changes in manuscript*:**

1. Revised Abstract, L13–19: "… we tested whether spatial embeddings (vectorised representations of land surface properties derived from satellite imagery) can detect genosoil signatures at 10 m resolution …" Previously: "… whether satellite-derived spatial embeddings can detect …"; "… indicating that the embeddings capture a consistent reference signature in areas with low human modification." Previously: "… capture a reproducible low-modification surface-state signal." Removed final sentence: "They also suggest a pathway towards future genosoil identification frameworks that rely less on coarse disturbance proxies and more on validated surface-state similarity."

***Minor Comment 2*:** *Figure 1. The Caption should better describe what is visible in the two panes. Pane B would fit better in the Methodology than in the Introduction.*

***Response*:** Panel B was moved to the Methods as Fig. 3, and Fig. 1 now has a caption that describes the HMI classes and the expected similarity gradient.

Changes in manuscript:

1. New Fig. 3 (former Fig. 1b), Sect. 2.3, L173–174: "Workflow for genosoil identification using spatial embeddings in this study. HMI = Human Modification Index."

2. Revised Fig. 1 caption, L94–99: "Conceptual representation of how a single pedogenon, soil sharing similar soil-forming covariates, can display very different properties based on human management. Classes based on the Human Modification Index (HMI): genosoil combines very low and low-

modification classes (HMI ≤ 0.10), while phenosoils span moderate (0.10 < HMI ≤ 0.40) and high/very high modification (0.40 < HMI ≤ 1.00). Highly modified soils will have embeddings that differ from the reference state, thus displaying lower similarity." Previously: "Conceptual Workflow for Genosoil Identification Using Spatial Embeddings."

Minor Comment 3: *Section 2.1. Another argument for selecting these pedogenons is that they have different levels of occurrence and modification (lines 224-225), which could be mentioned here as well.*

***Response*:** Adopted.

Changes in manuscript:

1. Revised Sect. 2.1, L115–117: "Three pedogenons were selected from the Global Pedogenon Map (GPM; Francos et al., 2025b; 6,500 pedogenons globally) to exemplify contrasting pedogenic environments within the United Kingdom; as well as the spectrum of genosoil availability, ranging from extensively preserved to nearly extinct (Table 1; Fig. 2)."

***Minor Comment 4*:** *I was wondering whether the cohesions derived from AlphaEarth and Tessera are comparable, as the spatial embeddings have different dimensions and are based on different datasets. I can imagine that Tessera has less internal variation as it is based on less variable data, which could lead to a higher cohesion. Could you add a remark about this in the manuscript?*

***Response*:** We added this caveat to the Methods and to the cohesion row of Table 2.

Changes in manuscript:

1. New Table 2, cohesion row: "(Cohesion values are not directly comparable between AlphaEarth and Tessera, which differ in dimensionality and training data.)"

2. Added in Sect. 2.3, L192–193: "Since AlphaEarth and Tessera differ in dimensionality (64 and 128 dimensions) and in training data and objectives, cohesion values are not directly comparable between them."

Minor Comment 5: *Line 160, 167. The terms L2-mean and L2-normalization need explanation.*

***Response*:** The term "L2-mean" was replaced by a description of the operation, and both terms are defined in Table 2.

Changes in manuscript:

1. New Table 2: L2-normalisation, "Scaling each embedding vector to unit length. On the resulting unit hypersphere only direction matters."; reference centroid, "The mean direction of the L2-normalised embeddings on the hypersphere."

2. Revised Sect. 2.3, L181–183: "… was calculated for each country from the embedding vectors of its reference pixels (HMI ≤ 0.10): each vector was scaled to unit length, the vectors were averaged, and the mean was rescaled to unit length". Previously: "… calculated for each country as the L2-mean of all qualifying pixel embeddings".

Minor Comment 6: Eq. 3. Can you explain more extensively what the Jaccard Index calculates and represents?

***Response*:** Table 2 explains the Jaccard index

Changes in manuscript:

1. New Table 2: Jaccard index, "Spatial overlap of two delineations. The area delineated by both A and B over the total area delineated by either A or B.", range "0 (disjoint) to 1 (identical)"; inclusion fraction, "Fraction of the area of delineation B also included in A."

Minor Comment 7: *Line 192. Silhouette confidence intervals are mentioned here but not provided in the results. Can you use uniform confidence-interval widths for silhouettes and bootstrapping?*

***Response*:** All reported intervals now use one method: 95 % intervals from 2,000 bootstrap resamples of GPM cells, keeping all reference pixels of a selected cell together; where more than 256 cells were available, 256 were drawn at random. Point estimates are computed on the complete population and reported separately. This applies to cohesion (Table S1) and to the silhouette (Table 4).

Changes in manuscript:

1. Revised Sect. 2.3, L191–192: "Uncertainty in cohesion was assessed using 2,000 bootstrap resamples of GPM cells, retaining all reference pixels within each resampled cell (Table S1)." Previously: "Bootstrap confidence intervals, with B = 2000 resamples at the pixel level, were computed for each reference embedding …"

2. Added in Sect. 2.4, L221–223: "The relationship between human modification and the candidate genosoil signature was assessed by comparing the reference centroid with the centroids of three progressively modified HMI classes within each country (0.1–0.2, 0.2–0.4, 0.4–1.0)." Added L232–236: "Confidence intervals were estimated using 2,000 bootstrap resamples of the GPM cells containing reference pixels in the Global Pedogenon Map. All embedding pixels within a selected cell were retained together, including their observations from both years. Where more than 256 GPM cells were available, 256 were selected at random; otherwise, all cells were included. The mean silhouette for the complete reference population is reported separately from the mean and 95 % confidence interval obtained from the sampled cells (Table 4)." Removed: "We report the median silhouette and 95% confidence interval (2.5th–97.5th percentiles) to characterise this distribution." and the paragraph defining the HMI and inter-country silhouettes ("Two cluster comparisons were employed …").

3. Revised Table 4, which now gives the 95 % bootstrap interval of the inter-annual silhouette; revised Table S1 caption: "The 95 % confidence intervals were estimated from 2,000 bootstrap resamples of GPM cells, retaining all reference pixels within each resampled cell. For P2932, the intervals describe subsets containing 10,614 AlphaEarth and 11,104 Tessera reference pixels …"

Minor Comment 8: Lines 216-219: Should be moved to methods, and Fig 1b should be referenced here.

***Response*:** Moved to the start of Sect. 2.3 with a reference to the workflow figure (now Fig. 3).

Changes in manuscript:

1. Moved to Sect. 2.3, L169–171: "We applied the workflow in a simple sequence: identify low-HMI candidate pixels within each pedogenon, compute a representative embedding for that low-modification state, and map cosine similarity from every pedogenon pixel to that reference. Pixels with cosine

similarity ≥ 0.90 were treated as highly congruent with the candidate genosoil reference, yielding a 10 m map of candidate genosoils per pedogenon (Fig. 3)."

Minor Comment 9: *Figure 3. This Figure needs some modification: Axis descriptions and labels are not readable due to their size and because they sometimes overlap with the Figures. Can you add the pedogenon code as a row label instead of as an inset in one of the panes? Can you indicate where each area is located within the UK? The details on the maps of P1564 are barely visible. Is there a way to improve the readability of these maps?*

***Response*:** The figure (now Fig. 4) was redrawn: larger, non-overlapping axis labels; pedogenon codes as row labels; a UK overview map locating each detailed view; and, for P1564, hollow circles marking candidate locations and an inset for its south-western occurrence.

Changes in manuscript:

1. Revised Fig. 4 and caption, L277–281: "… Southwestern P1564 is shown in an inset; the UK overview locates the detailed views. Coloured footprints show pedogenon extent and dark grey shows candidate genosoil. Hollow circles indicate candidate locations for P1564."

Minor Comment 10: *Table 2. Can you add values for coherence / cohesion, as these are mentioned in the text as well (e.g. lines 247-248, 250)?*

***Response*:** Cohesion was added to Table 3 (formerly Table 2) and is reported for every country and product in Table S2, with confidence intervals for the UK in Table S1.

Changes in manuscript:

1. Added row "Reference cohesion (AlphaEarth / Tessera)" to Table 3, and column "Reference cohesion" to Table S2 for every country and product; revised Table S1: cohesion of the UK reference populations with 95 % confidence intervals.

***Minor Comment 11*:** *Figure 4. Shouldn't the Y-axis label be "cosine distance from each country's genosoil centroid" instead of "distance from UK genosoil centroid"? Could you add a legend for the line widths instead of mentioning the scaling in the caption? You could, for example, group lines based on certain thresholds of pedogenon area. Also make the line width of the UK pedogenon consistent with the other line widths.*

***Response*:** Fixed. The axis now shows cosine distance and the caption states that it is measured from each country's own centroid. Countries are grouped by quartile of pedogenon area in a legend. Online response mentioned line widths; we used line styles instead, because they remain distinguishable where lines overlap. The United Kingdom profile is shown as a black line.

Changes in manuscript:

1. Revised Fig. 5 and caption, L327–330: y axis "Cosine distance"; legend of line styles by pedogenon-area quartile (Q1–Q4); United Kingdom in black. Caption: "Cosine distance between each modified HMI-class centroid and the corresponding country's candidate genosoil centroid. Panel n counts countries with a reference and at least one modified class. Countries listed in Table S2 but not shown either had no modified HMI classes (P845: Bolivia, Brazil, Georgia, Thailand; P1564: Argentina; P2932: Angola) or no HMI ≤ 0.10 reference (P1564: Belgium, France, Turkey)." Previously: "Cosine distance from each country's genosoil centroid across human modification classes for all countries in which each pedogenon occurs. Line width scales with log10(pedogenon area + 1)."

Minor Comment 12: *Figure 5. Can you use colors that are distinguishable for colorblind people and for black-and-white prints?*

***Response*:** As specified In the author guidelines, the figures had already been tested for colorblind friendliness. Now, figures now also tested for black and white prints. Fig. 6 distinguishes the two products by hatching as well as colour, so that they can be told apart in greyscale.

Changes in manuscript:

1. Revised Fig. 6: AlphaEarth bars hatched, Tessera bars solid; legend shows both.

Minor Comment 13: *Line 337. You state that AlphaEarth shows a greater degree of local environmental context, which would be relevant for understanding soil formation. Yet, this spatial embedding scores systematically lower than Tessera. Could you discuss the reasons for this in the context of soil formation, data quality and used statistics? See also comment 7.*

***Response*:** We replaced the "local environmental context" wording with an explanation tied to how each product was trained, and clarified that centroid similarity alone does not establish transfer.

Changes in manuscript:

1. Added in Sect. 3.3, L362–363: "Centroid similarity alone does not establish successful cross-border substitution, which is evaluated in Sect. 3.5."

2. Revised Sect. 3.4, L367–375: "The cross-border analysis showed a complementary pattern: Tessera generally returned higher absolute similarity values between countries, compressing the difference between true analogues and distant non-analogues. … AlphaEarth offered sharper boundaries at the cost of lower absolute similarity at distance." … "AlphaEarth's lower similarity to distant countries may reflect stronger discrimination of climate and relief, which were included as reconstruction targets (Brown et al., 2025); Tessera, trained on Sentinel-1 and Sentinel-2 time series (Feng et al., 2025), retained higher similarities across countries." Previously: "… AlphaEarth, trained on a broader set of multimodal inputs including climate variables and topography (Brown et al., 2025), retains a greater degree of local environmental context, thereby yielding sharper spatial discrimination but reduced cross-border similarity."

Minor Comment 14: *Table 4. Could you use consistent rounding of the decimals?*

***Response*:** Corrected.

Changes in manuscript:

1. Revised Tables 3, 4 and 5 (former Table 4), S1–S3 and S7: two-decimal rounding throughout.

Minor Comment 15: *Code availability. The referenced Zenodo page does not exist.*

***Response*:** DOI referenced is the reserved DOI which will be registered and published once paper is accepted. A preview code had been made available to reviewers. Both of those conditions remain the same:

Preview link is:
https://zenodo.org/records/19424156?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjU4NzA5MTQwLTk1Y2EtNGZmNS05MmMwLTcyNmNjMTZhODk5ZCIsImRhdGEiOnt9LCJyYW5kb20iOiIzYTcwNzg3YTQwNWIyODQwZjJkYWVhNTRhNDY5ZmNmNSJ9.6Dwt8tRhJQox3evJN0T63nh-wan6--UBnE6Ut8VptmMqf_b8CGBWJ-2PsFL31t5ks2U6TYhwcICYf0gZ5hYZmQ

Changes in manuscript:

1. Revised Code availability, L484–485: "The analysis code used in this study is archived on Zenodo and available at https://doi.org/10.5281/zenodo.19424156 (Pachón Maldonado et al., 2026)."; added Data availability, L487–488: "Candidate genosoil 10 m products available on Zenodo (Pachón Maldonado et al., 2026; https://doi.org/10.5281/zenodo.19424156)."; added reference: "Pachón Maldonado, J. C., Padarian, J., Styc, Q., and McBratney, A.: Use of Spatial Embeddings in Genosoil Identification: analysis code and UK candidate genosoil maps, Zenodo [code and data set], https://doi.org/10.5281/zenodo.19424156, 2026."

Zenodo doi is reserved but not registered until publication is accepted. Preview link is https://zenodo.org/records/19424156?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjU4NzA5MTQwLTk1Y2EtNGZmNS05MmMwLTcyNmNjMTZhODk5ZCIsImRhdGEiOnt9LCJyYW5kb20iOiIzYTcwNzg3YTQwNWIyODQwZjJkYWVhNTRhNDY5ZmNmNSJ9.6Dwt8tRhJQox3evJN0T63nh-wan6--UBnE6Ut8VptmMqf_b8CGBWJ-2PsFL31t5ks2U6TYhwcICYf0gZ5hYZmQ

Minor Comment 16: *Table S2. Why are the surface areas of pedogenons and genosoils different for each embedding model? Aren't these based on the independent HMI?*

***Response*:** Table S2 now reports one pedogenon area and one HMI-defined candidate area per country, both from a single census of the native GPM grid and independent of the embedding product; only the embedding candidate area, cohesion and similarity to the UK differ between products.

Changes in manuscript:

1. Added in Sect. 2.3, L205–207: "Pedogenon extent and HMI-defined candidate area were calculated using the GPM grid, sampling HMI values at cell centres. Areas were calculated on the WGS84 ellipsoid."

2. Revised Table S2 caption: "Global geographic distribution of the pedogenons studied, candidate genosoil areas and similarity of reference centroid to the United Kingdom's reference centroid. AlphaEarth uses 2022 data while Tessera uses 2024 data."; one "Pedogenon area" and one "HMI candidate area" value per country.

Minor Comment 17: *Table S6 is not referenced in the text.*

***Response*:** Table S6 is now referenced in Sect. 2.5.

Changes in manuscript:

1. Added in Sect. 2.5, L249–250: "The WRB soil profile records within the global occurrences of each pedogenon were also compiled to provide complementary taxonomic context (Table S6)."

Referee 2

R2: *Among the three pedogenons tested in current study, there are very unique characteristics. It's good; meanwhile it raises my further question. How about two pedogenons with similar characteristics such as much closer distance on geography or taxonomy? How would be the differentiation behavior from either AlphaEarth or Tessera?*

Response: The three pedogenons were chosen to contrast. Part of the question is answered within each pedogenon: occurrences of the same pedogenon in different countries share pedogenesis but differ in their similarity to the UK (Fig. 6, Sect. 3.3). Close analogues were found only in a few countries (Ireland for P845; central and eastern Europe for P1564; Norway for P2932), and AlphaEarth separated distant occurrences more sharply than Tessera (Sect. 3.4).

Changes in manuscript:

1. No change to the manuscript. Relevant results: Sect. 3.3, L332–365; Sect. 3.4, L366–375.

Comment R2: *Another minor comment is on Figure 3: the bottom subfigures are difficult to read clearly.*

***Response*:** The figure (now Fig. 4) was redrawn. The P1564 panels, which were hardest to read, now mark candidate locations with hollow circles and show the south-western occurrence in an inset. AC2 stated that 0.02 % of the UK extent of P1564 meets HMI ≤ 0.10; the revised value is 0.03 %.

Changes in manuscript:

1. Revised Sect. 2.1, L126–127: "P1564 is greatly modified by agriculture, with only 0.19 km² (0.03 % of the UK P1564 extent) exhibiting HMI values of 0.10 or less at the time of analysis."

2. Revised Fig. 4 and caption, L277–281: "… Southwestern P1564 is shown in an inset; the UK overview locates the detailed views. Coloured footprints show pedogenon extent and dark grey shows candidate genosoil. Hollow circles indicate candidate locations for P1564."

Topic Editor

Editor comment 1: *In Table S4 the reference to SoilGrids v2 (Poggio et al.) is wrong in that the WRB map in SG2 is still the 2017 map from Tom Hengl, the priority for SG2 was properties and not classes, and at the time ISRIC management did not understand the value of classes. The Poggio et al. paper does not mention classes at all, although the product includes them. So the 2017 paper should be cited. Second the map is not particularly good (acknowledged in the 2017 paper) although best in areas with many points, especially USA — although the correlation ST -> WRB was not exact — so some disclaimers are needed. It's really too bad that the SG250 product is the best world WRB map… the authors' use of it is operationally correct but based on a source which is exploratory/preliminary. That needs to be explained in the revised text, when Table S4 is presented. From the 2017 paper "at the level of WRB soil groups, map purity jumps to 60%… WRB soil groups with map purity < 50% are Planosols, Phaeozems and Ferrasols" but "for some classes, such as Cambisols, Luvisols, Fluvisols and Planosols in the WRB system, the confusion of predictions with other classes is still relatively high." So Table S4 is only indicative.*

***Response*:** We agree and thank the editor for the correction. The WRB classes are now attributed to the SoilGrids250m most-probable class map of Hengl et al. (2017), and the citation of Poggio et al. (2021) was removed. The Methods, the Table 1 caption and the Table S4 caption now state that the mapped classes are indicative context rather than validation, and that SoilGrids shares covariate inputs with the pedogenon map. Section 3.6 notes the map purity below 50 % that Hengl et al. (2017) report for Phaeozems, the dominant mapped group of P1564, and the discrepancy with gSSURGO for its US occurrence. While revising Table S4 we also found and corrected a legend error in our extraction: the

class names now follow the official 30-class SoilGrids legend. Areas and diversity values were unaffected.

Changes in manuscript:

1. Revised Table 1 caption, L132–136: "Last two columns are used for description purposes only. World Reference Base classes from the SoilGrids 250 m most-probable class raster (Hengl et al., 2017) across each pedogenon's global extent and are indicative rather than validated. USDA Soil Taxonomy Great Groups are from gSSURGO (Soil Survey Staff, 2026), which covers only the United States occurrence of each pedogenon."

2. Revised Sect. 2.5, L243–244: "WRB Reference Soil Groups were summarised from the SoilGrids 250 m global classification raster (Hengl et al., 2017; Table S4)." Previously: "WRB Reference Soil Groups (IUSS Working Group WRB, 2022) were summarised from the SoilGrids 250 m global classification raster (Poggio et al., 2021) …"

3. Added in Sect. 3.6, L436–439: "This interpretation requires caution because Hengl et al. (2017) reported map purity below 50 % for Phaeozems, something further emphasised by a classification discrepancy between SoilGrids and gSSURGO, which classifies the US occurrence of P1564 primarily as Udipsamments (Table S5)."

4. Revised Table S4 caption: "Predicted soil classes are from the SoilGrids250m WRB class map (Hengl et al., 2017), sampled at Global Pedogenon Map cell centres. These mapped predictions provide taxonomic context, rather than independent validation." Corrected class names (legend codes 25–29) to the official SoilGrids legend; the complete class list and diversity table are deposited (https://doi.org/10.5281/zenodo.19424156).

5. Added Hengl et al. (2017) to references; removed Poggio et al. (2021) and IUSS Working Group WRB (2022).

Editor comment 2: *Tables S4/5 are presented in the text but more can be said. In particular, an explanation of how this heterogeneity affects centroid construction, transferability, and the meaning of a failed cross-border match. This discussion might also discuss how taxonomic distance between WRB groups and between ST Great Groups might affect the actual heterogeneity. In fact, Table S5 has a direct synonym due to changes in ST and partially unrevised gSSURGO from which the authors extracted the series name and linked it to a GG via the Official Series Description: Most Dystrochrepts are a previous name for most of the Dystrudepts -- not a perfect correspondence but probably 90% overlap. There are no more "Dystrochrepts", nor "Xerochrepts" -- these latter are now mostly Dystroxerepts. […] In the WRB case, Acrisols/Alisols are more similar than the Andosols and Leptosols in Pedogeneon 845. It's not the purpose of the paper to incorporate taxonomic distance but it does bear on apparent vs. true heterogeneity, so could be mentioned. See for example […] Minasny, B., McBratney, A. B., & Hartemink, A. E. (2010). Global pedodiversity, taxonomic distance, and the World Reference Base. Geoderma, 155(3–4), 132–139.*

***Response*:** Section 3.6 now relates taxonomic heterogeneity to the embedding results for each pedogenon. P2932 has the lowest WRB diversity, yet its US reference population is less cohesive than the UK population. P1564 is taxonomically heterogeneous, and its reference centroid is most similar to the UK within temperate Europe. For P845, the UK–Ireland relationship is plausible given their shared Atlantic Acrisol environment, but nearly half of the global P845 extent lies in Japan, Nepal, China and Peru, where Acrisols form under monsoon and subtropical regimes. The low AlphaEarth cosine-to-UK values for these countries show that a shared pedogenon label does not necessarily imply a transferable genosoil reference. Land-use history can alter the present condition of soils within the same pedogenon,

and therefore ground validation is needed. Shannon diversity is computed on the complete WRB class areas (Table S4).

Changes in manuscript:

1. Revised Sect. 2.5, L246–247: “Shannon diversity indices (H') were calculated based on WRB Reference Soil Group membership to assess pedogenic homogeneity within each pedogenon globally (Minasny et al., 2010).” Table S4 now lists every mapped class.

2. Revised and expanded Sect. 3.6:

> L433–446: “The three case studies illustrate differences in mapped taxonomic diversity and embedding behaviour. P2932 demonstrates the lowest WRB diversity (H′ = 0.73; Podzols 73 %, Cambisols 21 %; Table S4), although its US reference populations were less cohesive than the UK populations (Table S2). Conversely, P1564 is taxonomically heterogeneous (H′ = 1.44; Table S4), and the similarity of its reference centroid to the UK is higher within temperate Europe than beyond this region.” … “This interpretation requires caution because Hengl et al. (2017) reported map purity below 50 % for Phaeozems, something further emphasised by a classification discrepancy between SoilGrids and gSSURGO, which classifies the US occurrence of P1564 primarily as Udipsamments (Table S5).” … “… its global distribution extends to tropical and subtropical regions such as Japan, Nepal, China, and Peru, which together account for nearly 50 % of the global P845 extent. … The embedding results corroborate this disjunction: most non-Atlantic countries showed low AlphaEarth cosine-to-UK values, although Georgia (0.31) and the United States (0.22) were above 0.20 (Table S2), indicating that the shared pedogenon label does not necessarily imply a transferable genosoil reference.” Previously: “The three case studies exemplify the relationship between taxonomic homogeneity and embedding behaviour. P2932 … effectively representing a single-soil-body pedogenon, with its genosoil signal transferred solely to Scandinavia …” Removed: the statement that the US Typic Udipsamments of P1564 represent “a fundamentally different pedogenic pathway from the humus-rich Phaeozems of central Europe”.
> Added L447–449: “Limited transferability does not necessarily indicate a limitation of the GPM since these results may be due to land-use history, which can alter the present condition of a soil within the same pedogenon.”

3. Added Table S5 note: “Legacy Dystrochrepts are pooled with Dystrudepts for presentation. Legacy Xerochrepts series were reassigned to Haploxerepts or Dystroxerepts according to their current USDA Official Series Descriptions. All other names are as recorded in gSSURGO.”

Editor comment 3: *The missing role of time as a factor of soil formation is mentioned but can be re-emphasized. The whole concept of Genosoils depends on pedogenons, which are "shaped by similar pedogenic processes" -- these surely operate over time. The only way time is taken into account is perhaps correlation with some geomorphic surfaces in some contexts. The use of the word "candidate" is good but the omission of time is a deeper problem. […] In the context of this study, some failed transfers may reflect incomplete representation of soil-forming history rather than weakness of the embedding products alone. Could some of the highlighted pedogenons be interpreted this way?*

Response**:** We agree. Section 3.6 now discusses time explicitly.

Changes in manuscript:

1. Revised Sect. 3.6, L424–432: “At the same time, our results indicate that a shared pedogenon label may still encompass very disparate soils, and fieldwork is needed to determine whether this reflects limitations of the pedogenon map, such as too few classes, need for more variables, or instead reflects anthropogenic modification that may be reversible if the pressure subsides but currently leads to very

disparate soils." Previously: "… a shared pedogenon label may not, by itself, ensure pedological equivalence across disparate regions, and that the internal heterogeneity within a pedogenon influences both the coherence and transferability of its genosoil signature." Added: "Time, one of the soil-forming factors, spans from extremely recent events such as landslides, to glaciation events tens of thousands of years ago, to highly stable soils that have developed over millions of years. Román Dobarco et al. (2021a) included the Weathering Intensity Index from airborne gamma-ray spectrometry across Australia to account for time; however, these data are not yet available globally. The pedogenic variable of time is therefore an important area of research to improve pedogenon maps."

2. Added in Conclusion, L479–480: "Second, pedogenon products would benefit from integrating a covariate for time such as the Weathering Intensity Index; however, these datasets are not yet available globally."

Editor comment 4: *The Conclusion can be tightened. It has vague language such as "although their application warrants further investigation and validation." — so please, specify some ways you or others might proceed. What would be a good followup? Other methods? Other embedding sources? Other source maps? etc. What do you conclude from the study about the most productive use of embeddings? What would be operational uses of these products -- i.e., what would be the most productive uses for a soil geographer/pedologist? Perhaps when pedogenon classes are internally coherent or environmentally heterogeneous.*

***Response*:** The conclusion was rewritten and the vague sentence removed. It now states that embeddings are more useful for identifying and ranking plausible candidate genosoils than for exact classification, and that cross-border substitution, as implemented, needs improvement because two of the three pedogenons showed no meaningful replacement of the local delineation. It names two follow-ups: field checking of the identified reference populations by profile description, and integrating a time covariate such as the Weathering Intensity Index into pedogenon products. It also names the operational use: identifying candidate reference sites at 10 m against which soil measurements can be interpreted. Section 3.6 adds that future work should test whether embedding clustering, combined with ground validation, can identify coherent genosoil populations without relying on fixed HMI thresholds.

Changes in manuscript:

1. Rewritten Conclusion; it now reads:

> "The delineation of genosoil in this study was based on the current globally available operational reference, the HMI. However, the results demonstrate that spatial embeddings can extend this framework beyond the limitations of HMI alone. At a resolution of 10 m, the embeddings did not merely reflect HMI classes; rather, they captured a reproducible surface-state signal within each pedogenon that was generally associated with lower levels of human modification and consistent local reference conditions. Even where genosoil availability was limited, the reference populations remained internally cohesive (Table S1), and within reference regions, the pixels with the lowest HMI values were typically the most similar to the candidate genosoil centroid. This suggests that spatial embeddings are more useful for identifying and ranking plausible genosoil candidates than for exact classification. Across the three case studies, similarity patterns followed biogeographic and pedological coherence more closely than pedogenon labels alone: P2932 remained largely boreal, P845 was concentrated in nearby European analogue regions, and P1564 linked the UK most strongly with temperate European analogue regions. The clearest positive example was P1564, in which severe local depletion still permitted partial recovery of the UK pattern, more distinctly under AlphaEarth than under Tessera. Conversely, cross-border substitution, as implemented here within the

proposed workflow and thresholds, needs improvement, with two of the three pedogenons showing no meaningful replacement of the local delineation by non-local references.
The country analysis, in Table S2, also showed how scarce reference soils have become: P1564 retained no HMI-defined candidate genosoil in Belgium, France or Turkey, and less than 1 % of its area in nine further countries, including the UK and Germany. For such pedogenons, reference populations identified with embeddings, locally or in other countries, may offer the only practical reference state, underlining the importance of this research.
Two steps follow directly from these results. First, the reference populations identified here should be checked in the field: the method selects candidates, and only profile description can confirm that a low-modification surface signature corresponds to a minimally disturbed soil. Second, pedogenon products would benefit from integrating a covariate for time such as the Weathering Intensity Index; however, these datasets are not yet available globally. Operationally, the approach can identify candidate reference sites at 10 m needed to understand soil condition, in which soil measurements are interpreted against a minimally disturbed soil of similar pedogenesis."

Editor comment 5: *In response to R1, the HMI threshold is discussion and justified with a sensitivity analysis, but this is not promised for the equally important cosine similarity threshold. Is the chosen threshold robust so that the overlap areas don't change much? Or is this something to be investigated in a future study (see comments on Conclusion, above)?*

***Response*:** We added a sensitivity analysis at cosine thresholds of 0.85, 0.90 and 0.95 for both products and all three pedogenons (Table S7). Candidate areas, and their overlap with the HMI-defined area, change with the threshold, so the delineations are threshold-sensitive. For P2932, AlphaEarth delineated 98 % of the HMI-defined candidate area at 0.90, compared with 128 % at 0.85 and 13 % at 0.95; Tessera delineated 70 %, 118 % and 2 %, respectively. We therefore describe the 0.90 maps as a conservative prioritisation of candidates that would benefit from calibration against independently characterised soil profiles and land-use histories.

Changes in manuscript:

1. New Table S7: "Sensitivity of embedding defined candidate genosoil areas to the cosine similarity threshold. IoU is the intersection divided by the union of the HMI defined and embedding defined candidate areas for each model. Both embedding products use UK reference populations from 2022."

2. Added in Sect. 2.3, L196–198: "No cosine threshold is universally established and 0.90 was set as a conservative cutoff. Nonetheless, we also tested 0.85, 0.90 and 0.95 for both models and all three pedogenons." Added in Sect. 3.1, L291–296: "Genosoil candidate areas changed across the tested thresholds for both embedding products and the three pedogenons, indicating that delineations are sensitive to the choice of cosine similarity threshold (Table S7). For P2932, AlphaEarth delineated 98 % of the HMI-defined candidate area at 0.90, compared with 128 % at 0.85 and 13 % at 0.95. Tessera delineated 70 % of the HMI-defined candidate area at 0.90, compared with 118 % at 0.85 and 2 % at 0.95. At the 0.90 threshold, the maps are therefore a conservative prioritisation of candidates that would benefit from calibration against independently characterised soil profiles and land-use histories."

Editor comment 6: *The landuse maps (WorldCover, CORINE) can not fully validate the HMI-derived candidate genosoils. This is because apparently "natural" classes could have been extensively modified by drainage, plantations of different forest species, atmospheric deposition, or previous management. […] So some caution is needed. […] some cautious language about this issue is needed.*

***Response*:** We agree. The land-cover comparison and the satellite imagery are now presented explicitly as supporting context for selecting candidate reference areas, not as validation, and Sect. 3.6 states that

past cultivation, drainage or plantation forestry can leave persistent changes that neither present cover nor a low HMI value reveals.

Changes in manuscript:

1. New Table S8 and Fig. S1.

> Table S8: "Land-cover composition of HMI defined candidate genosoil areas (HMI ≤ 0.10) and areas with higher human modification (HMI > 0.10) within the UK for the pedogenons studied, using ESA WorldCover 2021 v200 (Zanaga et al., 2022) and CORINE Land Cover 2018 (European Environment Agency, 2019)."
> Fig. S1: "(a) Extent of pedogenon 1564 in the UK, with black squares showing the locations of the all the HMI derived candidate genosoils (18.88 ha) in the pedogenon. (b) Median composites of Sentinel-2 imagery from May–September 2022 … Red outlines mark boundaries of HMI derived candidate genosoils."

2. Added in Sect. 3.6, L447–451: "Limited transferability does not necessarily indicate a limitation of the GPM since these results may be due to land-use history, which can alter the present condition of a soil within the same pedogenon. Previous cultivation, drainage or plantation forestry can leave persistent changes in soil properties that are not apparent from present vegetation cover or a low HMI value. The land-cover comparison (Table S8) and satellite imagery (Fig. S1) provide supporting context for selecting candidate reference areas, but historical information and field characterisation remain necessary to assess their suitability as genosoil references."