https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

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 & School of Life and Environmental Sciences, The University of Sydney, Eveleigh, NSW 2015, Australia

Correspondence to: Julio César Pachón Maldonado (julio.pachon@sydney.edu.au)

Abstract

Genosoils are minimally disturbed reference states within pedogenons, that is, soil units shaped by similar pedogenic processes within the Soil Security framework. They are central to assessing human impacts on soil functions, services, and resistance to threats. At present, genosoil delineation relies on the Human Modification Index (HMI), yet in intensively managed landscapes HMI thresholds may exclude all local pixels, leaving no local reference state available. Because the same pedogenon may occur across geographically distant regions, non-local occurrences may provide an alternative source of reference information. Using the United Kingdom as a case study, we tested whether satellite-derived spatial embeddings can detect genosoil signatures at 10 m resolution and whether these signatures can be transferred to regions with limited or absent local low-human-modification examples. We evaluated two satellite foundation-model embedding products, AlphaEarth and Tessera, across three contrasting pedogenons selected from the Global Pedogenon Map. Within each pedogenon, pixels with lower HMI values were generally more similar to the genosoil reference, indicating that the embeddings capture a reproducible low-modification surface-state signal. At the global scale, similarity to the UK genosoil was largely confined to biogeographically coherent regions. Cross-border substitution of local UK genosoil delineation was mostly limited, with meaningful partial recovery observed primarily in the highly modified agricultural pedogenon. These results indicate that satellite foundation-model embeddings can support higher-resolution genosoil delineation than is currently possible from global human modification products alone, extending the operational framework from 90 m to 10 m. They also suggest a pathway towards future genosoil identification frameworks that rely less on coarse disturbance proxies and more on validated surface-state similarity.

Keywords: Soil Security; pedogenon; genosoil; spatial embeddings; AlphaEarth; Tessera; human modification index

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

1. Introduction

For over a century, the study of soil as a natural body within the landscape, pedology, has sought not only to describe soil properties but to explain their genesis and dynamics through the interaction of soil-forming factors (Richter & Yaalon, 2012; Richter et al., 2020). A contemporary manifestation of this genetic perspective is the concept of the "pedogenon," a soil taxon defined not solely by morphology but also by quantitative representations of soil-forming factor proxies at a specific reference time (Román Dobarco et al., 2021a). In practice, a pedogenon groups all locations on Earth where the same combination of climate, parent material, topography, organisms, and time has shaped the soil in broadly similar ways. The pedogenon framework presupposes that relatively homogeneous soil-forming processes, operating over shared temporal scales, give rise to distinct soil entities with characteristic capacities and functional limits.

The development of the pedogenon reflects the broader evolution of pedology from 19th-century conceptions of soil as weathered rock or plant substrate to its modern understanding as an interdependent Earth system component emerging from interactions among the atmosphere, biosphere, lithosphere, and hydrosphere (Huggett, 2021). During the 20th century, the anthroposphere increasingly became a dominant soil-forming sphere (Richter, 2007; Dror et al., 2021), necessitating explicit consideration of human–soil interactions and their implications for soil function.

The Soil Security Assessment Framework (SSAF) was developed to quantify the impact humans have on the functions and services provided by soil, as well as changes in the ability to withstand threats to these functions and services. These measurements have been regularly performed across the natural sciences, but SSAF expands the analysis to include socio-economic and cultural factors as well (Evangelista et al., 2023). Central to SSAF is the distinction between pedogenon, genosoil, and phenosoil (Román Dobarco et al., 2021b). A pedogenon represents soils that share relatively uniform environmental covariates acting as indicators for soil-forming factors. A genosoil indicates the "least disturbed" or "reference" state within a pedogenon, reflecting its inherent functional potential and long-term trajectory, whereas phenosoils are anthropogenically altered expressions of the same soil-forming environment caused by different land management practices (Jang et al., 2023). In other words, a genosoil represents the soil as it would exist without significant human alteration, whereas a phenosoil reflects the same soil-forming setting after modification by agriculture, urbanisation, or other land uses. These distinctions are essential for understanding anthropogenic impacts on soil, and SSAF suggests that such studies first create a pedogenon map and then classify areas within each pedogenon as either geno- or phenosoil (Román Dobarco et al., 2023).

Beyond academic endeavours, the urgency of identifying genosoils is emphasised by the ongoing loss of rare and endemic soils (Bockheim, 2005). Amundson et al. (2003) defined rare soil in the United States as any soil series covering less than 1,000 ha,and deemed it endangered if over half of its area has been lost to land disturbance. The unique information and functional memory found in these soils, along with their irreplaceable biodiversity, are gradually and irreversibly disappearing (Ibáñez et al., 2012; Dazzi & Papa, 2019). Given the universality of the anthropogenic driver, classifying pedogenons into their geno- and phenosoil components at the scales required for management remains a current research challenge.

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

Previous studies have delineated genosoils utilising disturbance proxies such as the Human Modification Index (HMI) and the Habitat Condition Assessment System (Román Dobarco et al., 2023; Francos et al., 2025a). These threshold-based methodologies offer a pragmatic means of identifying candidate minimally disturbed areas within pedogenons. In Australia, however, thirty-two pedogenon classes lacked genosoils, and an additional two hundred and eighteen contained less than five percent of their area as genosoils (Román Dobarco et al., 2023). In Europe, Francos et al. (2025a) identified genosoils across all pedogenons employing the fifth percentile of HMI values within each pedogenon, acknowledging that small or heavily modified pedogenons may lack sufficient observations to establish a valid reference; in such cases, utilising the genosoil of the most similar pedogenon was proposed as an alternative. Building upon this foundation, embedding-based similarity analysis facilitates the characterisation of internal coherence among candidate genosoil areas and supports the spatial extension of representative reference signatures across the entire pedogenon extent.

Recent advances in geospatial artificial intelligence provide a complementary approach to surface characterisation. Foundation models are large neural networks pre-trained on extensive archives of satellite imagery that can process the spectral, temporal, and contextual information in Earth observation data and compress it into a fixed-length numerical vector, termed a spatial embedding, for each pixel on the ground. Pixels with similar embedding vectors share similar land-surface characteristics as observed by satellites, irrespective of their geographic separation. Unlike traditional spectral indices such as NDVI, which summarise a single vegetation property, spatial embeddings encode a broad range of surface attributes simultaneously. The specific content of each embedding depends on satellite inputs and the learning objectives of the foundation model. Embeddings do not directly measure soil properties; rather, they represent the observable surface state, primarily vegetation structure and land cover, at a given location and time.

Currently, two publicly accessible foundational models are AlphaEarth Foundations and Tessera. AlphaEarth produces 64-dimensional embeddings at a 10-meter resolution from multimodal inputs, including optical and radar satellite imagery, elevation models, climate data, and other Earth system variables (Brown et al., 2025). Tessera generates 128-dimensional embeddings at a 10-meter resolution through self-supervised learning on Sentinel-1 and Sentinel-2 time series, explicitly preserving spectral-temporal signals pertinent to land surface dynamics (Feng et al., 2025). Both products aim to provide annual global coverage, facilitating consistent surface-dynamic characterisation across regions and temporal scales, although Tessera's temporal coverage remains incomplete at the time of writing.

Spatial embeddings mainly represent vegetation communities and land cover observable in satellite imagery; they do not directly encode pedogenesis. It is hypothesised that within a pedogenon, genosoil areas share a characteristic embedding signature not because embeddings encode pedogenesis, but because they capture an aboveground vegetation community that develops on minimally disturbed soil potentially sharing characteristic features represented by the embeddings. A pedogenon with minimal human modification should support a vegetation assemblage characteristic of that soil-climate combination, resulting in a coherent and reproducible embedding signature. This signature can subsequently be employed to identify

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

genosoil candidates at higher resolutions across the extent of the pedogenon, within the same temporal range. When a pedogenon lacks genosoil within a given region, embedding similarity may facilitate the cross-border transfer of genosoil references from regions where the pedogenon remains better preserved. However, such transfer assumes that the vegetation-soil equilibrium remains consistent across the targeted geographic area. This concept has been operationalised as a straightforward workflow wherein low-HMI pixels within each pedogenon define a reference embedding, and similarity to this reference is subsequently mapped throughout the pedogenon extent (see Fig. 1).

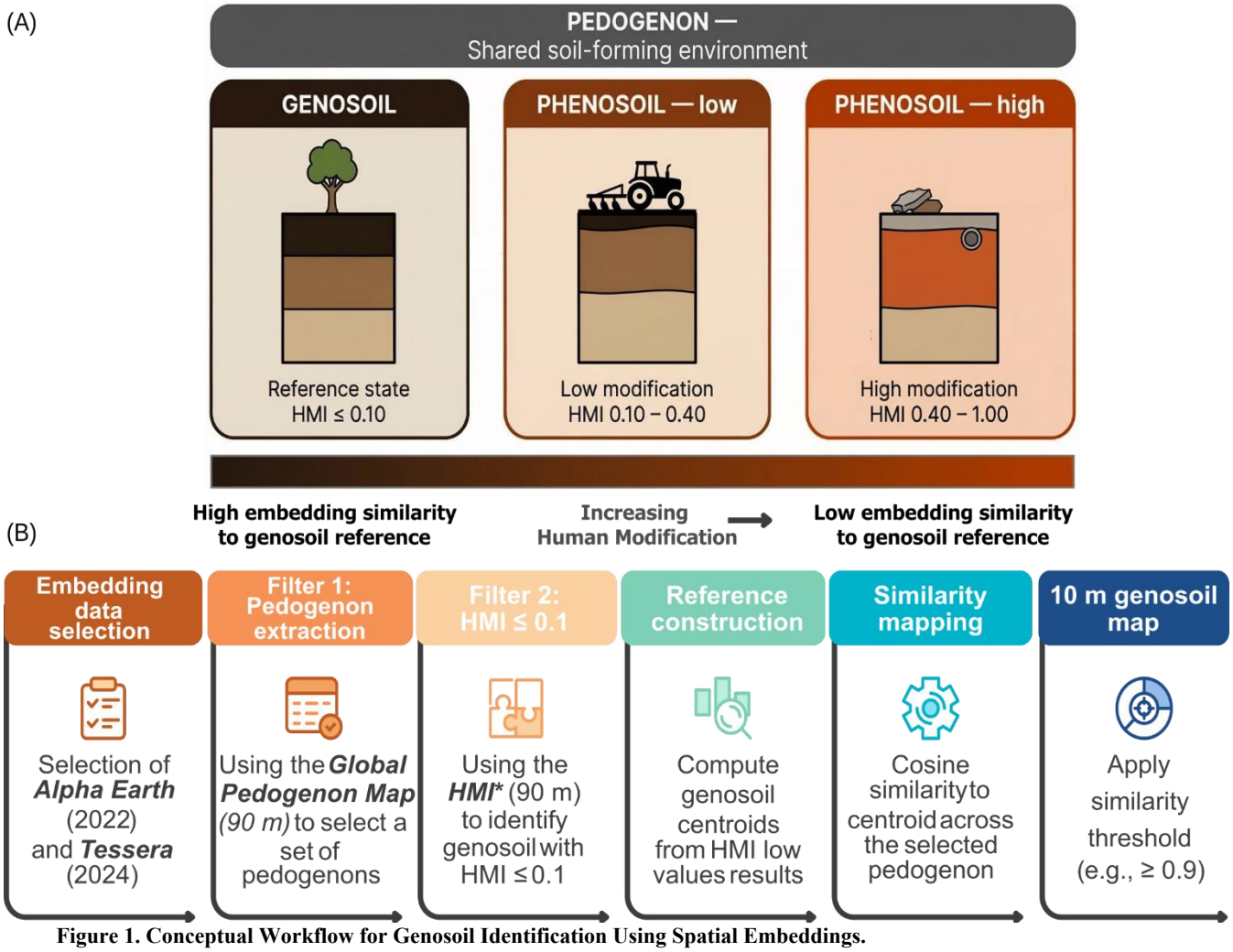

Figure 1. Conceptual Workflow for Genosoil Identification Using Spatial Embeddings.

This paper examines whether spatial embeddings are capable of capturing human modifications constrained by the pedogenon, thereby enabling the identification of genosoils at a high resolution. Additionally, it investigates whether reference signatures can be transferred across geographic boundaries when local genosoil data is scarce or unavailable. The specific objectives are as follows:

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

1. Quantify whether genosoil pixels within a pedogenon demonstrate greater embedding similarity to each other than to phenosoil pixels.
2. Determine whether a representative genosoil embedding can accurately identify analogous regions within the entire pedogenon extent.
3. Test whether genosoil reference signatures derived in other countries can substitute for sparse local references and still identify plausible genosoil candidates.
4. Compare AlphaEarth and Tessera with regard to genosoil discrimination and the transferability of genosoils on a global scale.

2. Materials and Methods

2.1. Study Pedogenons

Three pedogenons were selected from the Global Pedogenon Map (GPM; Francos et al., 2025b; 6,499 pedogenons globally at 90 m resolution) to exemplify contrasting pedogenic environments within the United Kingdom (Table 1; Fig. 2). The United Kingdom was selected as the case study due to the comprehensive temporal coverage (2015-2024) provided by Tessera at the commencement of the study.

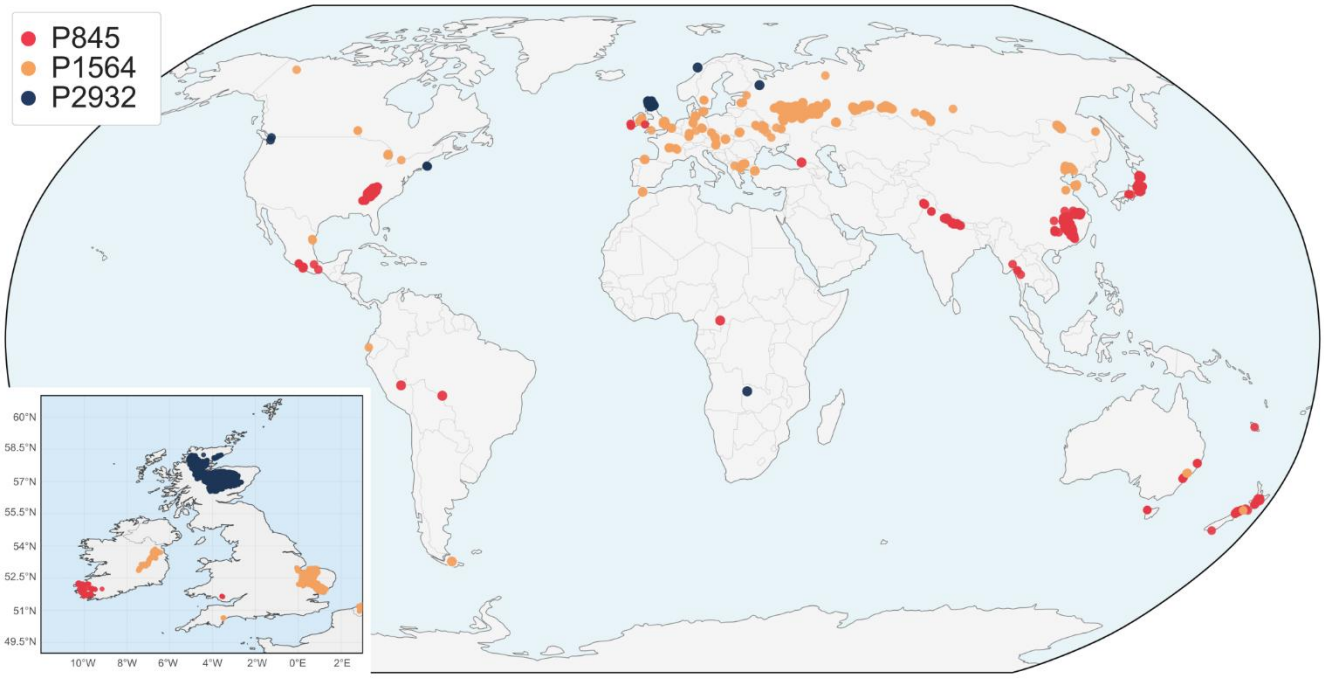

Figure 2. Spatial location of the pedogenons selected around the world and in the UK (inset lower left).

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

Pedogenon 2932 (P2932, Scottish Highlands) is predominantly composed of Podzols (74.4%) and Cambisols (19.4%), displaying limited pedogenic diversity with a Shannon H′ value of 0.72 based on the World Reference Base classification (WRB). It represents the most geographically restricted pedogenon among the three, encompassing 10,642 km² globally and 5,739.73 km² within the UK). Located in Southeastern UK, P1564 is the most widely distributed pedogenon studied, with a total area of 38,701 km² globally and 730.69 km² within the UK) and is predominantly characterized by Phaeozems (41.7%) and Albeluvisols (35.0%), with a Shannon H′ value of 1.40. This pedogenon is the most extensively modified by agriculture, with only 0.17 km² (0.02% of the UK P1564 extent) exhibiting HMI values of 0.10 or less at the time of analysis. Located in Southern Wales, P845 is associated with Acrisols (47%) at the summits and shoulders of hills. Its total global extent is 17,663 km², while its extent within the UK is 0.95 km².

Table 1. Characteristics of the three study pedogenons. WRB classifications from SoilGrids 250 m raster. USDA taxonomy from gSSURGO, which covers the US occurrence of each pedogenon label. Soil classifications for description purposes only.

P*	Location	WRB Major Soil Group	USDA Soil Taxonomy Great Group	Global	UK	UK genosoil Area (HMI ≤ 0.10)
		- % -	- % -	- km² -	- km² -	- % -
2932	Scottish Highlands	Podzols (74), Cambisols (19)	Quartzipsamments (39), Udipsamments (17)	10,642	5,740	61.4
1564	Southeastern UK	Phaeozems (42), Albeluvisols (35)	Udipsamments (60), Haplosaprists (9)	38,701	731	0.02
845	S. Wales	Acrisol (47), Cambisols (27)	Hapludults (65), Dystrudepts (16)	17,663	0.95	2.4

P*= Pedogenon

2.2. Spatial Embeddings

Two publicly accessible spatial embeddings were utilised: AlphaEarth (Brown et al., 2025) and Tessera (Feng et al., 2025).

AlphaEarth consists of 64-dimensional embeddings at a 10-meter resolution, obtained from multimodal inputs such as optical and radar imagery, topography, climate, and additional variables. Annual composites for the year 2022 were acquired through Google Earth Engine. Recent interpretability analyses of AlphaEarth embeddings across the contiguous United States (Rahman, 2026) have demonstrated that the 64 dimensions capture a wide range of environmental variables, with temperature, elevation, and vegetation indices accounting for over 90% of the variance ($R^2 > 0.90$). Conversely, anthropogenic variables such as impervious surfaces and nightlights were less prominently represented ($R^2 \approx 0.58$). The same study reported an average inter-annual dimension stability of $\bar{r} = 0.963$ across all 64 dimensions (2017–2023), with 51 out of 64 dimensions exceeding $\bar{r}$

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

= 0.95. This suggests that the embeddings encode land-surface properties that are temporally stable, rather than transient seasonal signals.

Tessera provides 128-dimensional embeddings at a 10-meter resolution, derived from Sentinel-1 and Sentinel-2 time series through self-supervised learning employing a Barlow Twins objective (Feng et al., 2025). Tessera's training methodology enforces temporal sampling invariance: two views of the same pixel are extracted from different random time windows within the same year, thereby promoting the model to learn representations that maintain stability across seasons.

All analyses used 2022 embeddings where available in order to maintain temporal alignment with the Human Modification Index, for which 2022 was the most recent year used in this study (Theobald et al., 2025). This applied to the UK-based delineation analyses and the UK reference centroids. The exception was the Tessera cross-country comparison: because non-UK Tessera coverage was not available for 2022, the globally available 2024 product was used instead. To estimate the temporal shift introduced by this substitution, UK Tessera genosoil centroids were compared between 2022 and 2024 across all three study pedogenons. AlphaEarth temporal variation was not assessed in this study.

Embeddings are not explicitly trained to represent soil properties. They capture the spectral and temporal co-evolution of biophysical land-surface traits over a calendar year. The pedological interpretation developed here is that, within a pedogenon and under low human modification, the characteristic vegetation–soil equilibrium produces a reproducible surface signal that the embedding encodes.

2.3. Genosoil Delineation and Reference Embedding

The 2022 Human Modification Index (HMI; Theobald et al., 2020, 2025) quantifies cumulative human modification from 0 (unmodified) to 1 (fully modified) at 90 m resolution. For each of the three pedogenons, genosoil candidate pixels were defined as 10 m pixels falling within 90 m HMI cells where HMI ≤ 0.10, following Francos et al. (2025a).

The genosoil reference centroid of pedogenon i (P_gi) was calculated for each country as the L2-mean of all qualifying pixel embeddings:

$$P_g\mathrm{i} = \frac{\mu_i}{|\mu_i|}, \quad \text{where} \quad \mu_i = \frac{1}{n}\sum_{j=1}^{n} z_j \tag{1}$$

with z_j being the L2-normalised embedding vector of the j-th qualifying pixel and n the total number of genosoil pixels in the reference country. This approach retains all genosoil pixels, regardless of their extent, making it suitable for pedogenons with a sparse extent, such as P1564 in the UK. Genosoil internal coherence, $\mathrm{Cohesion}(\mathrm{P_{g_i}})$, was quantified by cohesion, the mean cosine similarity of genosoil pixels to the group centroid, which indicates how representative the reference embedding is of the reference population:

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

$$\text{Cohesion}(P_{g_i}) = \frac{1}{n}\sum_{j=1}^{n} z_j \cdot P_{g_i} \quad (2)$$

Since all embedding vectors are L2-normalised to unit length prior to analysis, the cosine similarity between any two vectors reduces to their dot product. Cohesion measures range from 0 (indicating no coherence) to 1 (indicating complete pixel uniformity with the centroid). Bootstrap confidence intervals, with B = 2000 resamples at the pixel level, were computed for each reference embedding to assess the statistical reproducibility of the centroid's orientation (Table S1).

Similarity maps were generated by calculating the cosine similarity between the genosoil reference embedding and each pixel within the pedogenon extent. Given that all vectors are normalised to unit length, cosine similarity reduces to the dot product. Pixels exhibiting a similarity ≥ 0.90 were considered highly congruent with the genosoil reference, representing a rigorous threshold indicative of near-identical surface conditions. A comprehensive per-country analysis was performed to compute genosoil centroids and related metrics for all countries where the pedogenon is present, thereby facilitating a systematic comparison of genosoil signature coherence and UK similarity across the entire global distribution. Cosine distance is defined as 1 minus the cosine similarity.

Spatial overlap between genosoil masks was quantified using the Jaccard index (intersection over union, IoU):

$$\text{IoU} = J(A,B) = \frac{A \cap B}{A \cup B} \quad (3)$$

Where A and B are areas of the genosoils being compared.

2.4. Global Similarity Analysis

For each country in which a study pedogenon occurred, we calculated a genosoil centroid and its cohesion. Cohesion describes the internal consistency of a country's genosoil population, whereas the centroid provides its representative embedding direction.

Country-level transferability was then assessed by comparing each country's genosoil centroid with that of the UK using cosine similarity (cosine-to-UK). Higher cosine-to-UK values indicate that a country's genosoil signature is more similar to the UK reference for the same pedogenon.

To quantify how well genosoil pixels were separated from a comparison group, we used the silhouette coefficient (Rousseeuw, 1987). Given two predefined clusters A and B, the silhouette score for each pixel z_j in cluster A is:

$$s(z_j) = \frac{b(z_j) - a(z_j)}{max\left(b(z_j,), a(z_j)\right)} \quad (4)$$

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

where $a(z_j)$ is the mean cosine distance from z_j to all other members of cluster A, and $b(z_j)$ is the mean cosine distance from z_j to all members of cluster B. Values range from −1 to 1 where positive values indicate that z_j is more similar to its own cluster than to the comparison group, zero indicates an equal boundary, and negative values indicate greater similarity to the comparison group. We report the median silhouette and 95% confidence interval (2.5th–97.5th percentiles) to characterise this distribution.

Two cluster comparisons were employed. The HMI silhouette (s_HMI) defines cluster A as the genosoil population of $P_g i$ and cluster B as the HMI-modified pixels within the same country, grouped into three bins (moderate-low 0.10–0.20, moderate-mid 0.20–0.40, moderate-high 0.40–1.00). Positive median s_HMI values indicate that genosoil pixels are, on the whole, more internally cohesive than they are similar to modified pixels. The inter-country silhouette (s_inter) defines cluster A as one country's genosoil population for $P_g i$ and cluster B as another country's genosoil population for the same pedogenon, thereby assessing whether the two countries' genosoil signatures are separable or overlapping.

For AlphaEarth, the cross-country analysis used 2022 embeddings. For Tessera, the cross-country analysis used 2024 embeddings because non-UK Tessera coverage was not available for 2022. To assess the effect of this year difference, UK Tessera genosoil centroids were also compared between 2022 and 2024 across the three study pedogenons.

For the cross-border substitution analysis, the most similar non-UK country identified from the centroid comparison was used to generate a second UK genosoil mask. This foreign-derived mask was then compared with the UK-derived mask using IoU.

2.5. Describing Pedogenons

The soil taxonomic composition of each study pedogenon was characterised using two sources. WRB Reference Soil Groups (IUSS Working Group WRB, 2022) were summarised from the SoilGrids 250 m global classification raster (Poggio et al., 2021), with pixel counts weighted by area to determine percentage cover per pedogenon. USDA Soil Taxonomy suborders and great groups were derived from the Gridded Soil Survey Geographic database (gSSURGO; Soil Survey Staff, 2023), which provides coverage for the contiguous United States. Shannon diversity indices (H') were calculated based on WRB Reference Soil Group membership to assess pedogenic homogeneity within each pedogenon globally. WRB raster summaries were used as descriptive spatial context rather than independent validation, as SoilGrids products share covariate inputs with the pedogenon map.

3. Results and Discussion

3.1 Mapping genosoil candidates at 10 m

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

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

reference. Pixels with cosine similarity ≥ 0.90 were treated as highly congruent with the genosoil reference, yielding a 10 m map of genosoils per pedogenon (Fig. 3, Table 2).

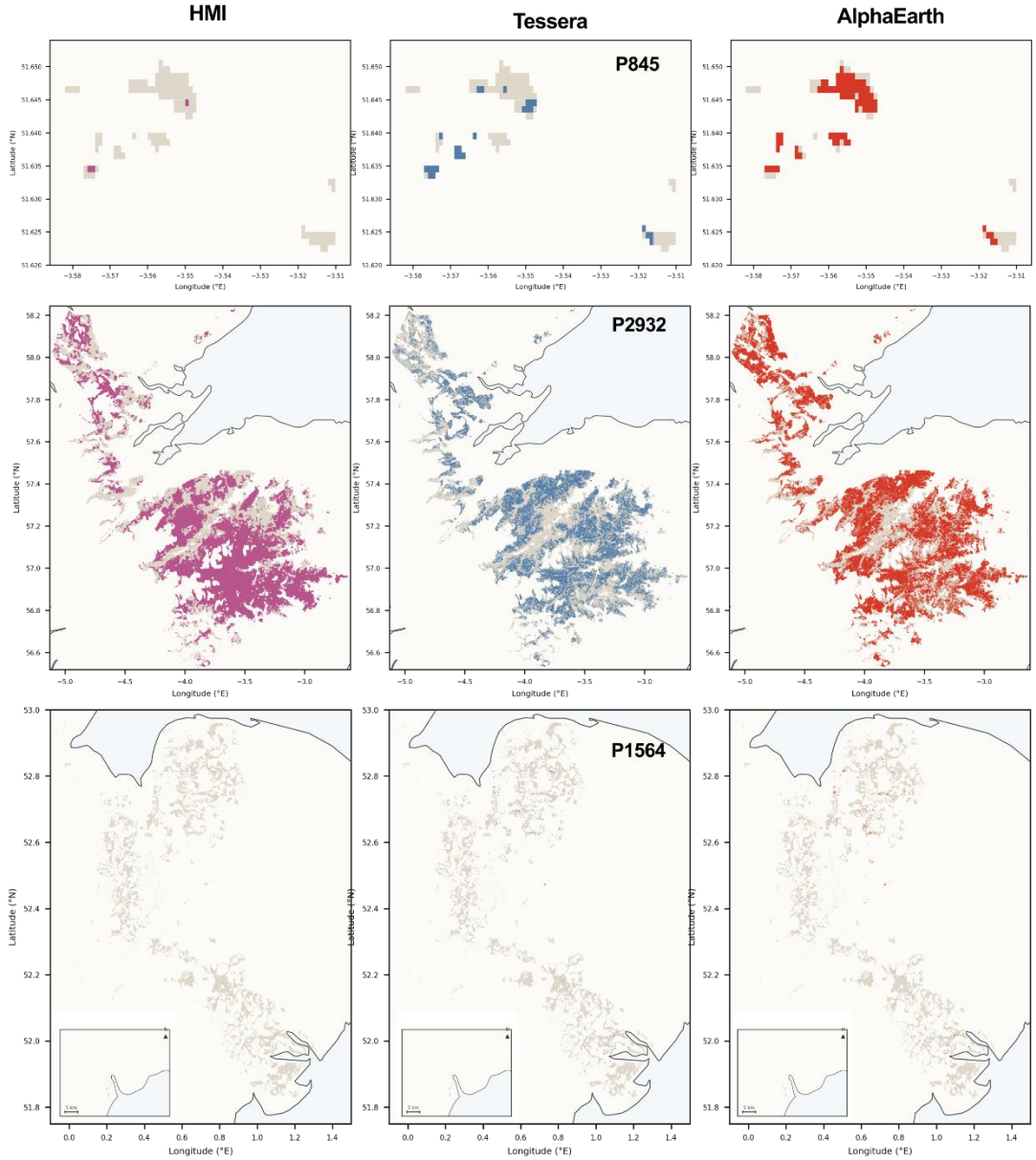

Figure 3. Genosoil maps for pedogenon 845 (top row), 2932 (middle row), and 1564 (bottom row) using the 90 m Human Modification Index Map (left column), Tessera 10 m centroid embeddings (middle column), AlphaEarth 10 m centroid embeddings (right column).

The three pedogenons encompass the entire spectrum of genosoil availability, ranging from extensively preserved to nearly extinct. The degree of concordance among delineation methods varied accordingly (see Table 2).

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

P2932 (Scottish Highlands) showed the broadest agreement between the disturbance proxy and the embedding-derived genosoil area extents. Within its 5,739.73 km² UK extent, 3,532.08 km² (61.4%) was classified as genosoil according to the HMI criterion (HMI ≤ 0.10), compared to 3,446.92 km² (60.1%) for AlphaEarth and 2,450.29 km² (42.7%) for Tessera (Table 2). Consequently, AlphaEarth closely replicated the overall extent defined by HMI, whereas Tessera delineated a more conservative and smaller subset. The highest pairwise overlap was observed between the two embedding products (IoU = 0.540), with Tessera predominantly nested within the AlphaEarth genosoil area extent (84.4% of Tessera genosoil area also classified as genosoil by AlphaEarth). All three methodologies converged on a common core of 1,342.09 km² (23.3% of the pedogenon), while a total of 4,912.20 km² (85.6%) was identified as genosoil by at least one method. The remaining discrepancies suggest that a low HMI value does not necessarily guarantee a genosoil-like embedding signature, and vice versa, a genosoil-like surface may be present in cells exceeding the conventional HMI threshold.

P1564 (Southeastern UK) presented the opposite case. Only 0.17 km² (0.02%) of its 731 km² territory in the UK qualified as genosoil according to the HMI criterion. This resulted in an exceedingly limited reference population from which to develop a genosoil embedding. AlphaEarth identified 7.68 km² (1.1%), while Tessera identified 7.70 km² (1.1%) as genosoil, each surpassing forty times the HMI input area (see Table 2). However, these masks were almost entirely non-overlapping: their pairwise IoU was 0.067, with only 12.5% of each mask intersecting with the other, and all three methods concurred on merely 0.02 km². Consequently, AlphaEarth and Tessera identified comparable amounts of potential genosoil area but in distinct locations. This divergence may be attributed to the multimodal inputs utilised by AlphaEarth in contrast to the Sentinel time series employed by Tessera. Despite this spatial disagreement, both models consistently prioritised lower-HMI pixels as more akin to the genosoil reference (see Section 3.4).

P845 (Southern Wales) occupied the smallest extent within the UK, measuring 0.95 km², with merely 0.02 km² (2.4%) at HMI ≤ 0.10. AlphaEarth classified 0.58 km² (61.0%) and Tessera 0.20 km² (21.1%) as genosoil, with Tessera again encompassed within the AlphaEarth genosoil extent (80.8% inclusion; Table 2). Cohesion was the highest among the three pedogenons, with AlphaEarth exhibiting a cohesion of 0.926 and Tessera 0.923, indicative of the narrow environmental envelope this pedogenon inhabits. Nonetheless, these areas corresponded to 75 and 26 pixels, respectively, thereby limiting the reliability of spatial overlap metrics at this scale. Although the area was small, it resulted in high internal coherence. The limited extent of the genosoil using HMI implies that a comprehensive characterisation of genosoil necessitates evidence from sources beyond the United Kingdom (Section 3.3).

Across the three pedogenons, AlphaEarth generated a wider genosoil area extent than Tessera in two of the three pedogenons, whose delineation was typically encompassed within AlphaEarth's boundaries. In regions where genosoil was prevalent, such as P2932, all three methods converged on a significant common core. Conversely, in areas where it was scarce, such as P1564, the embedding models extended the delineation well beyond the HMI input, exhibiting divergence in their spatial placement.

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

Table 2. Genosoil delineation overlap for the three study pedogenons within the UK. Three independent methods are compared: HMI threshold (≤ 0.10 at 90 m), AlphaEarth embedding similarity (≥ 0.90), and Tessera embedding similarity (≥ 0.90). IoU = Jaccard index (intersection over union).

	P2932	**P1564**	**P845**
Pedogenon extent (km²)	5,739.73	731	0.95
Genosoil area km² (% of pedogenon extent)			
HMI ≤ 0.10	3,532.08 (61.4%)	0.17 (0.02%)	0.02 (2.4%)
AlphaEarth ≥ 0.90	3,446.92 (60.1%)	7.68 (1.1%)	0.58 (61.0%)
Tessera ≥ 0.90	2,450.29 (42.7%)	7.70 (1.1%)	0.20 (21.1%)
All three agree	1,342.09 (23.3%)	0.02 (< 0.01%)	0.02 (2.4%)
Any method	4,912.20 (85.6%)	14.47 (2.0%)	0.62 (65.0%)
Pairwise IoU			
HMI - AE	0.445	0.013	0.04
HMI - Tess	0.377	0.005	0.115
AE - Tess	0.54	0.067	0.263
Inclusion (%)			
HMI captured by AE	60.9	56.5	100
HMI captured by Tess	46.4	21.7	100
Tess inside AE	84.4	12.5	80.8

3.2. Modification gradient and embedding comparison

Across the global occurrences of the three study pedogenons, increasing human modification was generally associated with greater embedding distance from each country's genosoil centroid (Fig. 4). Across all country profiles with at least one modified HMI bin, 82 of 94 showed a higher distance in the highest available modified class than in the lowest modified class, and 68 of 94 were non-decreasing across the available bins. This indicates that, within a given pedogenon and country, the low-HMI genosoil centroid defines a consistent local reference state in embedding space, from which progressively modified pixels tend to diverge.

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

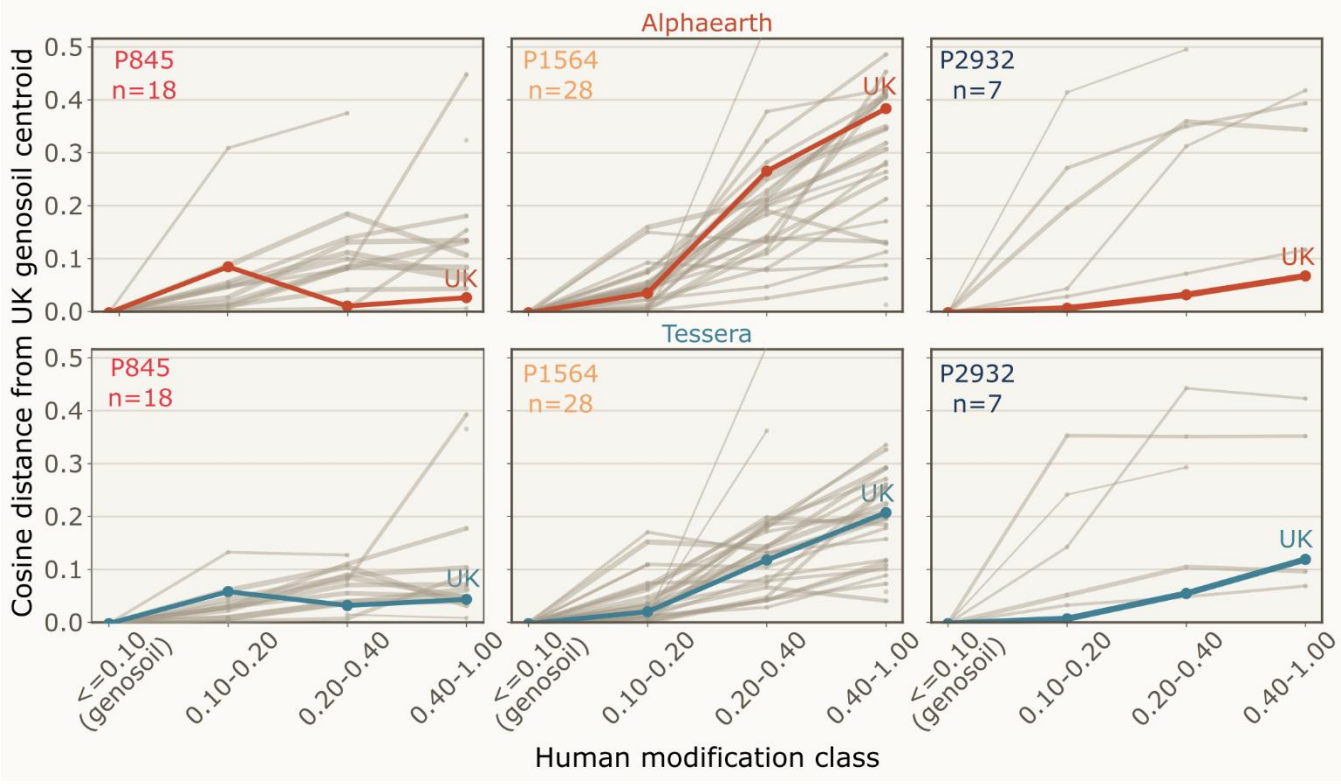

Figure 4. Cosine distance from each country's genosoil centroid across human modification classes for all countries in which each pedogenon occurs. Line width scales with log10(pedogenon area + 1).

P1564 is present in 28 countries and is the most extensively modified among the three studied pedogenons globally. Within the UK, only 0.17 km² of its 731 km² extent was classified as genosoil according to the HMI threshold, rendering it the most depleted local reference in the study (Table S2). Twenty-one nations have more than 1,000 ha of P1564, and the distance profiles demonstrate the clearest and most consistent pattern across countries. In AlphaEarth and Tessera, 25 out of 27 country profiles with modified-bin data exhibited greater distances in the highest available modified class compared to the lowest modified class. In the UK, the cosine distance in AlphaEarth increased from 0.183 to 0.484 ($\Delta = 0.301$) from genosoil to the most modified class, whereas in Tessera it increased from 0.181 to 0.408 ($\Delta = 0.227$), indicating stronger separation of modification levels in AlphaEarth for this pedogenon.

P2932 occurs in seven countries, but only the UK and the USA exceed 1,000 ha in pedogenon area (Table S2). Its distance gradient is weaker than that of P1564, which is consistent with soil properties that reduce development pressure and with the high proportion of UK extent remaining at HMI ≤ 0.10 (61%; Table 2). For AlphaEarth, the cosine distance increased from 0.135 in the genosoil class to 0.221 in the most modified class; for Tessera, it increased from 0.132 to 0.282. Both models

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

preserved the same rank ordering in the UK; however, Tessera demonstrated a larger overall dynamic range, indicating a more pronounced differentiation among the levels of modification within this pedogenon.

P845 occurs in 18 countries, nine of which exceed 10 ha in pedogenon area (Table S2). Among the three study pedogenons, it exhibits the greatest variability in distance profiles across different countries, and the pattern observed is correspondingly less consistent than that of P1564. Within the UK, its very limited extent results in minimal internal gradient discernible, although centroid cohesion remains high (0.926 for AlphaEarth; 0.923 for Tessera).

Together, these cross-country profiles demonstrate that the genosoil embedding constitutes a reproducible low-modification reference state within pedogenons. The pixels in this state are generally more similar to each other than to the increasingly modified phenosoil pixels across most countries.

3.3. Cross-border genosoil embedding signal

Once the UK reference signatures were established, we compared them with genosoil populations of the same pedogenon in other countries (Fig. 5). The purpose was not to redefine the UK reference but to identify external analogues, countries where the pedogenon may persist in a less modified state or where a larger genosoil population provides independent corroboration of the reference signature.

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

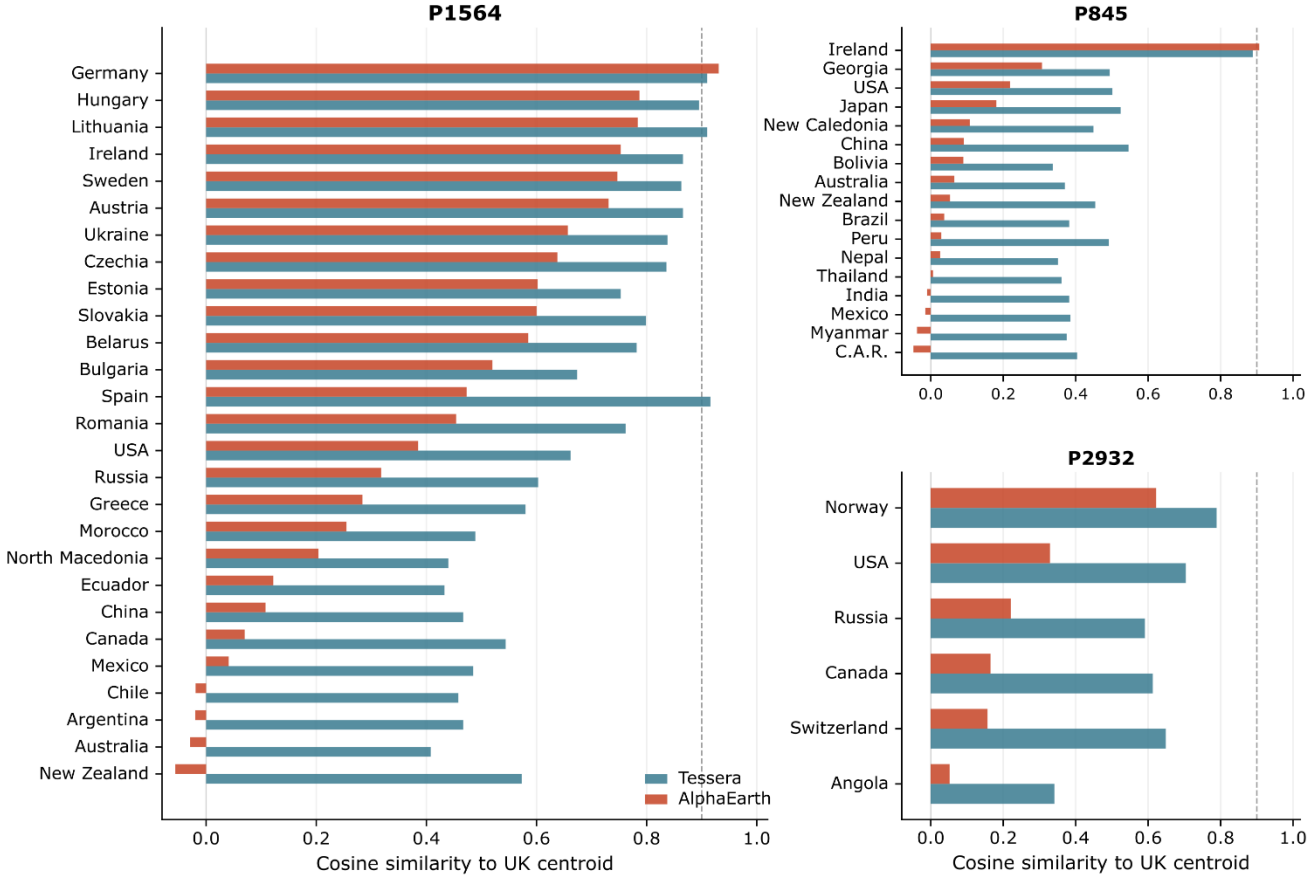

Figure 5. Comparison of country-level genosoil similarity to the United Kingdom reference centroid for the three study pedogenons under AlphaEarth (2022) and Tessera (2024). The bars represent the cosine similarity between each country's genosoil centroid and the UK genosoil centroid for the same pedogenon.

For P1564, this comparison produced the most practically valuable result. Despite the limited UK reference (0.17 km² at HMI ≤ 0.10), the embedding signature pointed clearly toward central and eastern Europe. Germany was identified as the most analogous AlphaEarth region (cosine similarity to UK = 0.931), followed by Hungary (0.787) and Lithuania (0.784). Under Tessera, Spain exhibited the highest similarity (0.916), with Germany and Lithuania closely following (both 0.910), and Hungary trailing slightly behind (0.895). The embedding signature predominantly pointed toward temperate Europe, with the strongest similarities concentrated in central, eastern, and certain western or southwestern European nations, depending on the model. Even in instances where UK genosoil is depleted, the embedding signature can point at countries potentially containing a more preserved analogue of the same pedogenon.

For P2932, the closest analogue was Norway (cosine-to-UK = 0.622 AlphaEarth, 0.789 Tessera), aligning with the shared boreal–Atlantic environmental setting. However, similarity dropped beyond Norway. Under AlphaEarth, the subsequent

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

closest was the United States at 0.329, followed by Russia at 0.221, Canada at 0.165, Switzerland at 0.157, and Angola at 0.052. Tessera demonstrated a considerably higher similarity threshold, with the United States at 0.704, Switzerland at 0.649, Canada at 0.613, Russia at 0.591, and Angola at 0.341.

Regarding P845, Ireland was identified as the most evident external analogue (cosine-to-UK = 0.907 AlphaEarth, 0.889 Tessera). Under AlphaEarth, the next most similar countries were Georgia (0.307), the United States (0.219), and Japan (0.181), demonstrating a significant decline in similarity beyond Ireland. Under Tessera, the differences were considerably less pronounced, with China (0.546), Japan (0.524), the United States (0.501), and Georgia (0.494) all maintaining moderate resemblance to the UK reference.

In summary, these comparisons demonstrate that cross-border transfer functions serve primarily as an analogue-search tool rather than a comprehensive global classifier. The efficacy of transfer correlates with biogeographic coherence: P1564 exhibits broad transfer within temperate Europe, P2932 is limited to boreal Scandinavia, and P845 is restricted to Atlantic Ireland. Additionally, the country rankings reveal significant differences in selectivity between the two embedding products: AlphaEarth exhibits steeper declines beyond the nearest analogue, whereas Tessera maintains higher similarity across a wider geographic range.

3.4. Comparing AlphaEarth and Tessera

The cross-border analysis showed a complementary pattern: Tessera consistently returned higher absolute similarity values between countries, compressing the difference between true analogues and distant non-analogues. This elevated global similarity floor means Tessera tolerates greater biogeographic distance for analogue searches but provides less discrimination between neighbouring countries. AlphaEarth offered sharper boundaries at the cost of lower absolute similarity at distance.

A plausible explanation for this contrast lies in the training objectives. During training, Tessera obtained two views of the same pixel drawn from different random time windows within the same calendar year (Feng et al., 2025). This design explicitly forces the encoder to suppress seasonal and phenological variations, producing representations that are inherently more similar across time. This enhances similarity across biogeographically similar yet geographically distant regions. Conversely, 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.

Because the cross-country Tessera comparison used the globally available 2024 product rather than 2022, temporal stability was assessed by comparing UK Tessera genosoil centroids between 2022 and 2024 (Table 3; Table S3). The cosine similarity of Tessera centroids across these years was 0.978 for P1564 and 0.982 for P2932, with cohesion variations remaining below 0.01 in both instances. P845 demonstrated lower inter-annual stability, with a centroid cosine similarity of 0.858, consistent with its relatively small sample size, where seasonal vegetation fluctuations can induce measurable shifts in the centroid

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

position. Inter-annual silhouette indices, which assess the separation between the 2022 and 2024 genosoil populations, yielded low values for P1564 (+0.109) and P2932 (+0.122), but were notably higher for P845 (+0.651). These outcomes suggest that the reference remained stable for adequately sampled pedogenons, yet was susceptible to annual variation when the sample size was considerably limited.

Table 3. Stability of Tessera embeddings at UK genosoil locations from 2022 to 2024 for pixels intersecting HMI ≤ 0.10. Centroid cosine similarity compares the 2022 and 2024 UK genosoil centroids. Cohesion change is the difference in cohesion between 2024 and 2022. Inter-annual silhouette measures the separability of the two annual embedding snapshots at those pixel locations; values range from -1 to 1, so the small positive values indicate detectable but modest drift between years.

Pedogenon	Centroid cosine similarity (2022 vs 2024)	Cohesion change	Inter-annual silhouette
P1564	0.978	< 0.01	0.109
P2932	0.982	< 0.01	0.122
P845	0.858	0.007	0.651

3.5. Cross-border genosoil centroid

The preceding sections established two points. Firstly, embedding-derived genosoil maps can be generated from local low-HMI reference populations within each pedogenon. Secondly, the global comparison identified a limited number of countries whose genosoil centroids were significantly more similar to the UK reference than the rest of the pedogenon's worldwide distribution. The subsequent inquiry concerns whether these non-UK centroids can serve as practical substitutes in instances where the UK reference data are sparse or absent. This directly addresses the third objective of the study: assessing whether genosoil signatures can be transferred across political boundaries when local reference data are limited.

Using the strongest non-UK analogue identified for each pedogenon, we then tested whether the foreign-derived centroid reproduced the UK delineation (Table 4).

Table 4. Cross-border substitution results using the strongest non-UK analogue for each pedogenon.

	P1564	P845	P2932
AlphaEarth			
Strongest non-UK analogue	Germany	Ireland	Norway
Non-UK genosoil centroid cosine similarity UK genosoil centroid	0.931	0.907	0.622
UK local mask from UK centroid (km²)	7.68	0.58	3446.92
UK foreign-derived mask from non-UK centroid (km²)	2.75	0.0000	0.0000
Intersection (km²)	2.63	0.0000	0.00
Union (km²)	7.80	0.58	3446.92
AlphaEarth IoU	0.3376	0.0000	0.0000
Tessera			

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

Strongest non-UK analogue	Spain	Ireland	Norway
Non-UK genosoil centroid cosine similarity UK genosoil centroid	0.916	0.889	0.789
UK local mask from UK centroid (km^2)	6.24	0.28	2141.51
UK foreign-derived mask from non-UK centroid (km^2)	0.20	0.0000	0.0000
Intersection (km^2)	0.11	0.0000	0.0000
Union (km^2)	6.32	0.28	2141.51
Tessera 2024 IoU	0.0179	0.0000	0.0000

These results indicate that substitution was limited rather than general. Two of the three pedogenons showed no meaningful transferability under either model, and the single informative positive case was P1564.

P1564 provides the clearest positive case because it is the most extensively modified agricultural pedogenon in the UK and retains almost no HMI-defined genosoil locally. Under AlphaEarth, Germany was the strongest external analogue (cosine-to-UK = 0.931), and the Germany-derived genosoil showed moderate overlap with the UK-derived extent (IoU = 0.338), recovering 2.63 km² of overlapping candidate area despite the severe depletion of the UK reference population. Under Tessera, Spain rather than Germany was the strongest external analogue, although the difference was minimal (0.916 vs 0.910). Substitution performance nevertheless remained weak: the Spain-derived Tessera centroid produced 0.20 km² of UK genosoil extent, of which 0.11 km² overlapped the UK-derived 2024 Tessera delineation (IoU = 0.0179). P1564 therefore supports a cautious conclusion: where local genosoil has been strongly depleted, a biogeographically coherent external analogue may partially recover the local pattern.

P845 did not produce a viable substitute under either model. Although Ireland was the closest and most plausible external analogue, there was no overlap with the UK-derived delineation for both AlphaEarth and Tessera.

P2932 also failed as a transfer case. Norway was the most comparable non-UK analogue; however, neither model produced a meaningful delineation of the UK substitute from the foreign centroid. For both AlphaEarth and Tessera, the foreign-derived mask within the UK was effectively absent, resulting in no meaningful overlap with the local UK delineation.

In summary, these findings indicate that cross-border substitution may be advantageous under specific conditions; however, it did not function as a widely applicable method within the current workflow and thresholds. Its value lies not in the complete replacement of local references but in the possibility of partial recovery in cases of severe local depletion, as demonstrated by P1564.

3.6. Interpretation and methodological limits

The Global Pedogenon Map should be regarded as an initial global framework for pedogenon-based soil assessment. The ideal foundation for genosoil delineation would be a global inventory of field-validated soil series or equivalent fine-grained soil bodies; however, such a dataset is currently unavailable at the necessary spatial coverage and resolution. The GPM addresses

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

this gap by providing a covariate-driven, process-oriented spatial framework at 90 m resolution, without which the present analysis would not be possible. At the same time, our results indicate that 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.

The three case studies exemplify the relationship between taxonomic homogeneity and embedding behaviour. P2932 demonstrates the lowest WRB diversity (H′ = 0.72; Podzols 74%, Cambisols 19%), effectively representing a single-soil-body pedogenon, with its genosoil signal transferred solely to Scandinavia, aligning with a narrow, pedologically consistent class. Conversely, P1564 is taxonomically heterogeneous (H′ = 1.40; Phaeozems 42%, Albeluvisols 35%, encompassing 21 Reference Soil Groups globally), and its genosoil signature is broadly transferred within temperate Europe but diminishes beyond this region. This heterogeneity is further emphasised by a classification discrepancy: SoilGrids maps the global extent of P1564 predominantly as Phaeozems (Table S4), whereas gSSURGO classifies the US occurrence as Typic Udipsamments (Entisols; Table S5), which are minimally developed sandy soils representing a fundamentally different pedogenic pathway from the humus-rich Phaeozems of central Europe. P845 (H′ = 1.37; Acrisols 47%, Cambisols 27%) illustrates the most notable case: the UK–Ireland genosoil relationship appears plausible given their shared Atlantic, mesic-thermic Acrisol environment; however, its global distribution extends to tropical and subtropical regions such as Japan, Nepal, China, and Peru, which together account for over 60% of the global P845 extent. In these regions, Acrisols develop under monsoon and subtropical regimes, characterised by distinct weathering histories, vegetation assemblages, and land-use pressures. The embedding results corroborate this disjunction: most non-Atlantic countries showed low AlphaEarth cosine-to-UK values, although Georgia (0.307) and the United States (0.219) were above 0.20 (Table S2), indicating that the shared pedogenon label does not necessarily imply a transferable genosoil reference.

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.

A further limitation is that the present workflow still depends on HMI to define the initial local reference population. This is operationally necessary because HMI remains the only globally available disturbance framework for genosoil delineation, but it also constrains the analysis to the assumptions and spatial grain of that proxy. The embeddings therefore are not yet being used to define genosoil independently; rather, they are being used to refine and extend an HMI-anchored reference. Future work should test whether embedding clustering, combined with ground validation and pedological interpretation, can identify coherent genosoil populations without relying exclusively on fixed HMI thresholds.

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

4. Conclusion

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 critically limited, the reference centroids remained stable, and within reference regions, the pixels with the lowest HMI values were typically the most similar to the 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, was not generally successful, with two of the three pedogenons showing no meaningful replacement of the local delineation by non-local references.

Our findings suggest that spatial embeddings are promising high-resolution tools for identifying potential genosoils, although their application warrants further investigation and validation. This work constitutes a step towards future genosoil identification frameworks that depend less on coarse disturbance proxies and more on validated surface-state similarity.

Code Availability

All analysis scripts used in this study are archived at: 10.5281/zenodo.19424156

Data Availability

AlphaEarth spatial embeddings: Brown et al. (2025). Tessera spatial embeddings: Feng et al. (2025). Global Pedogenon Map data: Francos et al. (2025b); data available at https://doi.org/10.5281/zenodo.15678953. Human Modification Index: Theobald et al. (2020, 2025).

Author contributions

Julio Cesar Pachon Maldonado: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Visualisation, Writing – original draft, and Writing – review & editing. **Jose Padarian:**

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

Conceptualisation, Methodology, Supervision, and Writing – review & editing. **Quentin Styc:** Writing – review & editing. **Alex McBratney:** Funding acquisition, Resources, Supervision, and Writing – review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

We gratefully acknowledge the use of Google Earth Engine for geospatial data processing and analysis. We thank the University of Sydney for the institutional and computing support that made this work possible. We also thank the Soil Security team for their support, discussion, and for fostering the collaborative research environment in which this study was developed.

Claude Opus 4.6 was used for coding and workflow assistance. All scientific analyses, interpretations, claims, and final manuscript text were produced, verified, and approved by the authors.

Financial support

This research was supported by the Australian Research Council Laureate Fellowship (FL210100054) on Soil Security, titled A Calculable Approach to Securing Australia's Soils.

References

Amundson, R., Guo, Y., and Gong, P.: Soil Diversity and Land Use in the United States, Ecosystems, 6, 470–482, https://doi.org/10.1007/s10021-002-0160-2, 2003.

Bockheim, J. G.: Soil endemism and its relation to soil formation theory, Geoderma, 129, 109–124, https://doi.org/10.1016/j.geoderma.2004.12.044, 2005.

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., and Kohli, P.: AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data, arXiv [preprint], arXiv:2507.22291, 29 July 2025.

Dazzi, C., and Papa, G. L.: Soil genetic erosion: New conceptual developments in soil security, Int. Soil Water Conserv. Res., 7, 317–324, https://doi.org/10.1016/j.iswcr.2019.08.001, 2019.

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

Dror, I., Yaron, B., and Berkowitz, B.: The human impact on all soil-forming factors during the anthropocene, ACS Environ. Au, 2, 11–19, https://doi.org/10.1021/acsenvironau.1c00010, 2021.

Evangelista, S. J., Field, D. J., McBratney, A. B., Minasny, B., Ng, W., Padarian, J., Román Dobarco, M., and Wadoux, A. M. J.-C.: A proposal for the assessment of soil security: Soil functions, soil services and threats to soil, Soil Security, 10, 100086, https://doi.org/10.1016/j.soisec.2023.100086, 2023.

Feng, Z., Atzberger, C., Jaffer, S., Knezevic, J., Sormunen, S., Young, R., Lisaius, M. C., Immitzer, M., Jackson, T., Ball, J., Coomes, D. A., Madhavapeddy, A., Blake, A., and Keshav, S.: TESSERA: Temporal Embeddings of Surface Spectra for Earth Representation and Analysis, arXiv [preprint], arXiv:2506.20380, 25 June 2025.

Francos, N., Jayasekara, T., Ng, W., Sharififar, A., Styc, Q., Watt, D. J., and McBratney, A.: Delineating Genosoils and Phenosoils in Europe Using the Global Pedogenon Map, Eur. J. Soil Sci., 76, e70249, https://doi.org/10.1111/ejss.70249, 2025a.

Francos, N., Sharififar, A., Flynn, T., Styc, Q., Evangelista, S. J., Padarian, J., Ng, W., Field, D. J., Minasny, B., and McBratney, A. B.: The global pedogenon map: Combining and spatialising the factors of soil formation, Geoderma, 460, 117462, https://doi.org/10.1016/j.geoderma.2025.117462, 2025b.

Huggett, R. J.: Soil as a System, in: Hydrogeology, Chemical Weathering, and Soil Formation, edited by: Hunt, A., Egli, M., and Faybishenko, B., Wiley, Hoboken, NJ, 1–20, https://doi.org/10.1002/9781119563952.ch1, 2021.

Hengl, T. and Gupta, S.: An Open Compendium of Soil Datasets: Soil Observations and Measurements (V1.0.1), Zenodo [data set], https://doi.org/10.5281/zenodo.15593990, 2025.

Ibáñez, J. J., Krasilnikov, P. V., and Saldaña, A.: Archive and refugia of soil organisms: applying a pedodiversity framework for the conservation of biological and non-biological heritages, J. Appl. Ecol., 49, 1267–1277, https://doi.org/10.1111/j.1365-2664.2012.02213.x, 2012.

IUSS Working Group WRB: World Reference Base for Soil Resources – International soil classification system for naming soils and creating legends for soil maps, 4th edn., International Union of Soil Sciences (IUSS), Vienna, Austria, 2022.

Jang, H. J., Dobarco, M. R., Minasny, B., Campusano, J. P., and McBratney, A.: Assessing human impacts on soil organic carbon change in the Lower Namoi Valley, Australia, Anthropocene, 43, 100393, https://doi.org/10.1016/j.ancene.2023.100393, 2023.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

https://doi.org/10.5194/egusphere-2026-1944
Preprint. Discussion started: 20 April 2026

Rahman, M.: Physically Interpretable AlphaEarth Foundation Model Embeddings Enable LLM-Based Land Surface Intelligence, arXiv [preprint], arXiv:2602.10354, 10 February 2026.

Richter, D. deB.: Humanity's Transformation of Earth's Soil: Pedology's New Frontier, Soil Sci., 172, 957, 2007.

Richter, D. D., Eppes, M. C., Austin, J. C., Bacon, A. R., Billings, S. A., Brecheisen, Z., Ferguson, T. A., Markewitz, D., Pachon, J., Schroeder, P. A., and Wade, A. M.: Soil production and the soil geomorphology legacy of Grove Karl Gilbert, Soil Sci. Soc. Am. J., 84, 1–20, https://doi.org/10.1002/saj2.20030, 2020.

Richter, D. deB., and Yaalon, D. H.: "The Changing Model of Soil" Revisited. Soil Sci. Soc. Am. J., 76, 766–778. https://doi.org/10.2136/sssaj2011.0407, 2012.

Román Dobarco, M., McBratney, A., Minasny, B., and Malone, B.: A modelling framework for pedogenon mapping, Geoderma, 393, 115012, https://doi.org/10.1016/j.geoderma.2021.115012, 2021a.

Román Dobarco, M., McBratney, A., Minasny, B., and Malone, B.: A framework to assess changes in soil condition and capability over large areas, Soil Security, 4, 100011, https://doi.org/10.1016/j.soisec.2021.100011, 2021b.

Román Dobarco, M., Padarian Campusano, J., McBratney, A. B., Malone, B., and Minasny, B.: Genosoil and phenosoil mapping in continental Australia is essential for soil security, Soil Security, 13, 100108, https://doi.org/10.1016/j.soisec.2023.100108, 2023.

Rousseeuw, P. J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math., 20, 53–65, https://doi.org/10.1016/0377-0427(87)90125-7, 1987.

Soil Survey Staff: Gridded Soil Survey Geographic (gSSURGO) Database, United States Department of Agriculture, Natural Resources Conservation Service [data set], https://www.nrcs.usda.gov/resources/data-and-reports/gridded-soil-survey-geographic-gssurgo-database (last access: 5 April 2026), 2023.

Theobald, D. M., Kennedy, C., Chen, B., Oakleaf, J., Baruch-Mordo, S., and Kiesecker, J.: Earth transformed: detailed mapping of global human modification from 1990 to 2017, Earth Syst. Sci. Data, 12, 1953–1972, https://doi.org/10.5194/essd-12-1953-2020, 2020.

Theobald, D. M., Oakleaf, J. R., Moncrieff, G., Voigt, M., Kiesecker, J., and Kennedy, C. M.: Global extent and change in human modification of terrestrial ecosystems from 1990 to 2022, Sci. Data, 12, 606, https://doi.org/10.1038/s41597-025-04892-2, 2025.