Supplementary Tables

Use of Spatial Embeddings in Genosoil Identification

Table S1. Bootstrap stability of UK genosoil centroids for each pedogenon (B = 2000 resamples, 90 % CI width). The reported CI width summarises the dispersion of cosine similarity between bootstrap centroids and the full centroid, and therefore reflects centroid directional stability. For P2932, bootstrapping was performed on a representative sample of 500,000 pixels from the full genosoil population; the reported CI width is therefore a conservative upper bound.

Pedogenon	Embedding model	Genosoil pixels	Embedding cohesion	Bootstrap 90% CI width
2932	AlphaEarth	35278615	0.888	4.967e-08†
2932	Tessera	36064279	0.8678	4.211e-08†
1564	AlphaEarth	1890	0.8946	0.0001932
1564	Tessera	2017	0.8186	0.0001999
845	AlphaEarth	305	0.926	0.0006407
845	Tessera	294	0.9225	0.0005493

† Representative 500,000-pixel bootstrap sample; reported width is a conservative upper bound.

Table S2. Global genosoil similarity by country to the UK. Global pedogenon areas: P845 = 1,766,312 ha; P1564 = 3,870,136 ha; P2932 = 1,064,246 ha. Cohesion and HMI-bin silhouette are computed from pixels with HMI ≤ 0.10 within each country. Countries with fewer than 20 pixels in any HMI bin are shown without an HMI-bin silhouette value.

Pedogenon ID	Embedding model	Country	Pedogenon area in country (ha)	Genosoil area (ha)	Embedding cohesion	HMI-bin silhouette	Cosine similarity to UK centroid
845	AE 2022	United Kingdom	88.6	3.0	0.926	-0.830	1.000
845	AE 2022	Ireland	2898.9	854.7	0.878	-0.961	0.907
845	AE 2022	Georgia	10.7	10.8	0.884	-	0.307
845	AE 2022	USA	611573.8	329485.1	0.844	-0.927	0.219
845	AE 2022	Japan	111021.9	44259.8	0.817	-0.714	0.181
845	AE 2022	New Caledonia	281.0	268.5	0.858	0.564	0.108
845	AE 2022	China	561985.6	504286.8	0.837	-0.888	0.091
845	AE 2022	Bolivia	16.4	16.4	0.993	-	0.090
845	AE 2022	Australia	15590.8	13725.8	0.916	-0.421	0.065
845	AE 2022	New Zealand	90267.0	65953.8	0.832	-0.470	0.053
845	AE 2022	Brazil	12.7	12.7	0.952	-	0.037
845	AE 2022	Peru	176.5	163.5	0.866	0.568	0.029
845	AE 2022	Nepal	64150.5	54357.8	0.863	-0.791	0.026
845	AE 2022	Thailand	400.3	401.1	0.918	-	0.006
845	AE 2022	India	7031.7	2654.8	0.848	-0.621	-0.010
845	AE 2022	Mexico	16563.8	12389.9	0.882	-0.887	-0.015
845	AE 2022	Myanmar	1418.2	1385.3	0.872	-0.617	-0.038
845	AE 2022	Central African Rep.	12.8	0.9	0.997	0.445	-0.048
845	Tessera 2024	United Kingdom	88.6	2.9	0.923	-0.560	1.000
845	Tessera 2024	Ireland	2898.9	874.5	0.888	-0.963	0.889
845	Tessera 2024	China	561985.6	519329.1	0.820	-0.940	0.546
845	Tessera 2024	Japan	111021.9	44866.8	0.810	-0.754	0.524
845	Tessera 2024	USA	611573.8	343590.7	0.829	-0.941	0.501
845	Tessera 2024	Georgia	11.7	11.7	0.889	-	0.494
845	Tessera 2024	Peru	176.5	164.9	0.853	-0.102	0.486
845	Tessera 2024	New Zealand	90267.0	67948.5	0.830	-0.799	0.454
845	Tessera 2024	New Caledonia	281.0	269.1	0.859	0.617	0.449
845	Tessera 2024	Central African Rep.	12.8	0.9	0.972	-0.624	0.404
845	Tessera 2024	Mexico	16563.8	12476.4	0.835	-0.815	0.385

845	Tessera 2024	Brazil	13.1	13.1	0.919	-	0.382
845	Tessera 2024	India	7031.7	2697.0	0.852	-0.788	0.382
845	Tessera 2024	Myanmar	1418.2	1391.9	0.849	-0.745	0.375
845	Tessera 2024	Australia	15590.8	13917.0	0.893	-0.708	0.370
845	Tessera 2024	Thailand	401.1	400.5	0.902	-	0.361
845	Tessera 2024	Nepal	64150.5	55493.5	0.856	-0.824	0.351
845	Tessera 2024	Bolivia	16.4	16.4	0.954	-	0.337
1564	AE 2022	United Kingdom	72413.7	18.9	0.895	-0.648	1.000
1564	AE 2022	Germany	18747.9	130.5	0.901	-0.659	0.931
1564	AE 2022	Hungary	5078.3	498.6	0.850	-0.959	0.787
1564	AE 2022	Lithuania	258.7	13.2	0.919	-0.001	0.784
1564	AE 2022	Ireland	10101.6	14.2	0.923	-0.330	0.753
1564	AE 2022	Sweden	4713.3	43.2	0.904	-0.530	0.747
1564	AE 2022	Austria	1777.3	20.0	0.910	-0.879	0.731
1564	AE 2022	Ukraine	34898.9	4034.6	0.902	-0.481	0.657
1564	AE 2022	Czechia	5509.9	4.3	0.942	-0.252	0.638
1564	AE 2022	Estonia	3854.9	1644.9	0.920	-0.056	0.602
1564	AE 2022	Slovakia	36.6	27.7	0.970	-0.696	0.600
1564	AE 2022	Belarus	1260.8	188.8	0.892	-0.856	0.585
1564	AE 2022	Bulgaria	28023.1	240.7	0.896	-0.445	0.520
1564	AE 2022	Spain	16879.0	54.1	0.862	-0.944	0.473
1564	AE 2022	Romania	1272.0	478.8	0.853	-0.971	0.454
1564	AE 2022	USA	11280.6	8744.4	0.864	-0.660	0.385
1564	AE 2022	Russia	1775353.6	831885.5	0.881	-0.612	0.318
1564	AE 2022	Greece	3462.0	384.0	0.835	-0.852	0.284
1564	AE 2022	Morocco	42.1	3.0	0.939	-0.045	0.255
1564	AE 2022	North Macedonia	716.9	2.8	0.799	-0.335	0.204
1564	AE 2022	Ecuador	804.1	567.1	0.937	-0.601	0.122
1564	AE 2022	China	259494.8	45072.0	0.835	-0.025	0.108
1564	AE 2022	Canada	1769.4	1395.2	0.912	-0.476	0.070
1564	AE 2022	Mexico	1269.3	72.4	0.948	0.323	0.041
1564	AE 2022	Chile	150.9	148.5	0.936	-0.766	-0.019
1564	AE 2022	Argentina	47.4	47.6	0.945	-	-0.020
1564	AE 2022	Australia	2780.4	249.8	0.928	-0.949	-0.029
1564	AE 2022	New Zealand	1349.8	17.6	0.879	-0.334	-0.056
1564	Tessera 2024	United Kingdom	72413.7	20.2	0.819	-0.876	1.000
1564	Tessera 2024	Spain	16879.0	56.9	0.845	-0.961	0.916
1564	Tessera 2024	Germany	18747.9	133.6	0.844	-0.866	0.910
1564	Tessera 2024	Lithuania	258.7	13.9	0.861	-0.733	0.910

1564	Tessera 2024	Hungary	5078.3	516.3	0.828	-0.972	0.895
1564	Tessera 2024	Austria	1777.3	20.8	0.850	-0.905	0.866
1564	Tessera 2024	Ireland	10101.6	15.3	0.852	-0.569	0.866
1564	Tessera 2024	Sweden	4713.3	44.2	0.849	-0.825	0.863
1564	Tessera 2024	Ukraine	34898.9	4266.8	0.835	-0.759	0.838
1564	Tessera 2024	Czechia	5509.9	4.3	0.863	-0.759	0.836
1564	Tessera 2024	Slovakia	36.6	28.3	0.914	-0.879	0.799
1564	Tessera 2024	Belarus	1260.8	202.3	0.823	-0.937	0.782
1564	Tessera 2024	Romania	1272.0	493.3	0.851	-0.982	0.762
1564	Tessera 2024	Estonia	3854.9	1672.4	0.843	-0.538	0.753
1564	Tessera 2024	Bulgaria	28023.1	252.7	0.843	-0.684	0.674
1564	Tessera 2024	USA	11280.6	9316.3	0.839	-0.866	0.662
1564	Tessera 2024	Russia	1775353.6	863162.5	0.783	-0.857	0.603
1564	Tessera 2024	Greece	3462.0	405.0	0.861	-0.919	0.580
1564	Tessera 2024	New Zealand	1349.8	18.9	0.818	-0.576	0.573
1564	Tessera 2024	Canada	1769.4	1477.4	0.756	-0.563	0.544
1564	Tessera 2024	Morocco	42.1	3.1	0.943	-0.764	0.489
1564	Tessera 2024	Mexico	1269.3	73.2	0.903	-0.291	0.485
1564	Tessera 2024	Argentina	49.1	49.1	0.902	-	0.467
1564	Tessera 2024	China	259494.8	47946.1	0.812	-0.230	0.467
1564	Tessera 2024	Chile	150.9	149.9	0.886	-0.472	0.458
1564	Tessera 2024	North Macedonia	716.9	2.9	0.904	0.286	0.440
1564	Tessera 2024	Ecuador	804.1	568.1	0.937	-0.752	0.433
1564	Tessera 2024	Australia	2780.4	258.8	0.868	-0.989	0.408
2932	AE 2022	United Kingdom	574595.6	362201.4	0.865	-0.936	1.000
2932	AE 2022	Norway	2.5	1.5	0.907	0.775	0.622
2932	AE 2022	USA	7622.4	1318.9	0.640	-0.453	0.329
2932	AE 2022	Russia	30.6	1.5	0.972	0.386	0.221
2932	AE 2022	Canada	377.0	362.5	0.930	0.744	0.165
2932	AE 2022	Switzerland	92.4	2.5	0.954	-0.330	0.157
2932	AE 2022	Angola	8.8	8.8	0.989	-	0.052
2932	Tessera 2024	United Kingdom	574595.6	360642.8	0.868	-0.931	1.000
2932	Tessera 2024	Norway	2.5	1.7	0.909	0.627	0.789
2932	Tessera 2024	USA	7622.4	1325.7	0.769	-0.768	0.704
2932	Tessera 2024	Switzerland	92.4	2.3	0.829	-0.797	0.649
2932	Tessera 2024	Canada	377.0	367.9	0.832	0.525	0.613
2932	Tessera 2024	Russia	30.6	1.6	0.901	0.315	0.591
2932	Tessera 2024	Angola	8.6	8.6	0.964	-	0.341

Table S3. Tessera UK genosoil temporal stability between 2022 and 2024. Inter-annual silhouette measures the separation between year-specific pixel populations relative to the within-year spread, with values closer to 1 indicating a stronger separation between years, and values closer to 0 indicating greater overlap and thus higher temporal stability. The centroid cosine similarity is calculated between the UK genosoil centroids of 2022 and 2024.

Pedogenon ID	Genosoil pixels in 2022	Genosoil pixels in 2024	Cohesion in 2022	Cohesion in 2024	Cohesion change (2024-2022)	Inter-annual silhouette	Centroid cosine similarity (2022 vs 2024)
845	294	294	0.916	0.923	0.007	0.651	0.858
1564	2017	2017	0.820	0.819	-0.002	0.109	0.978
2932	36136094	36064279	0.874	0.868	-0.007	0.122	0.982

Table S4. World Reference Base (WRB) raster composition for each pedogenon within the UK, derived from SoilGrids 250 m WRB predictions (Poggio et al., 2021).

Pedogenon	WRB reference soil group	Area (km²)	Share of pedogenon area (%)
845	Acrisols	8301.9	46.73
845	Cambisols	4860.3	27.36
845	Alisols	2791.1	15.71
845	Andosols	997.7	5.62
845	Leptosols	436.4	2.46
845	Luvisols	143.2	0.81
845	Regosols	118.2	0.67
845	Nitisols	34.8	0.2
845	Fluvisols	28.1	0.16
845	Podzols	25.4	0.14
845	Gleysols	20.3	0.11
845	Ferralsols	4.2	0.02
845	Phaeozems	2.4	0.01
845	Technosols	0.3	0.0
845	Arenosols	0.1	0.0
1564	Phaeozems	16236.3	41.72
1564	Albeluvisols	13607.8	34.97
1564	Cambisols	4014.8	10.32
1564	Luvisols	3375.9	8.67
1564	Chernozems	537.2	1.38
1564	Gleysols	369.1	0.95
1564	Fluvisols	340.4	0.87
1564	Histosols	159.0	0.41
1564	Podzols	80.1	0.21
1564	Kastanozems	47.1	0.12
1564	Alisols	41.6	0.11
1564	Technosols	36.3	0.09
1564	Arenosols	24.5	0.06
1564	Cryosols	19.9	0.05
1564	Retisols	14.6	0.04
1564	Ferralsols	5.2	0.01
1564	Andosols	4.3	0.01
1564	Calcisols	1.8	0.0
1564	Solonchaks	0.6	0.0
1564	Planosols	0.4	0.0
1564	Acrisols	0.3	0.0
1564	Regosols	0.2	0.0
1564	Leptosols	0.1	0.0
1564	Lixisols	0.1	0.0
2932	Podzols	7621.6	74.42
2932	Cambisols	1990.5	19.44
2932	Gleysols	618.4	6.04
2932	Leptosols	7.9	0.08
2932	Luvisols	2.2	0.02
2932	Andosols	0.2	0.00

Table S5. Great Group and associated landform summary for the United States occurrences of the three study pedogenons. Great Group composition is derived from component-weighted gSSURGO taxonomy outputs, and landforms are derived from the NRCS `cogeomordesc` table linked at the component level. Great Group shares are reported as percentages of total pedogenon component weight in the United States subset, whereas landform shares are reported as percentages within each Great Group. Great Groups contributing at least 5 % are shown together with their three most abundant non-zero associated landforms. The Great Groups shown account for 77.9-92.3 % of pedogenon composition across the three case studies.

Pedogenon	Great Group	Great Group share of pedogenon (%)	Landform 1	Landform 1 share within Great Group (%)	Landform 2	Landform 2 share within Great Group (%)	Landform 3	Landform 3 share within Great Group (%)
845	Hapludults	67.29	hillslopes	29.4	interfluves	29.3	ridges	21.1
845	Dystrudepts	14.07	hillslopes	45.6	ridges	32.9	escarpments	9.8
845	Dystrochrepts	10.92	hillslopes	51.1	ridges	34.9	coves	14.1
1564	Udipsamments	69.55	deltas	67.9	moraines	12.1	outwash plains	8.7
1564	Haplosaprists	8.82	depressions	86.1	swales	5.8	deltas	5.7
1564	Endoaquods	5.43	deltas	61.4	lake plains	28.2	lake terraces	10.4
2932	Quartzipsamments	42.42	moraines	99.3	outwash terraces	0.7	-	-
2932	Udipsamments	16.56	outwash plains	56.2	dikes	22.8	outwash terraces	11.1
2932	Haplosaprists	13.76	depressions	94.3	swamps	5.7	-	-
2932	Xerochrepts	5.17	mountain slopes	100	-	-	-	-

Table S6. World Reference Base (WRB) soil profile records located within the global occurrences of the three study pedogenons (Hengl and Gupta, 2025).

Profile ID	Pedogenon	Assigned WRB profile class	WRB grouping label	Source database	Longitude	Latitude
207839	845	Endoleptic Luvisols Abruptic	Abruptic	{MX-INEGI}	-95.50109289	16.8669264
64904	845	Haplic Alisols	Alisols	{CN-SOTER}	118.11187744	30.14492989
64904	845	Ferric Alisols	Alisols	{CN-SOTER}	118.11187744	30.14492989
SIM51744	845	Haplic Acrisols	Acrisols	HWSDv2	-84.09094357638483	35.32400275978988
SIM87807	845	Dystric Regosols	Regosols	HWSDv2	118.59019265096369	30.383268697132515
2742	1564	Eutric Planosols	Planosols	BZE_LW	12.560489622196622	51.18087491096414
607393	1564	Haplic Chernozems	Chernozems	{RU-SPDB}	36.29999924	51.29999924
607393	1564	Luvic Chernozems	Chernozems	{RU-SPDB}	36.29999924	51.29999924
63778	1564	Calcaric Cambisols	Cambisols	{WD-ISIS}	13.41305556	55.54805556
63778	1564	Eutric Cambisols	Cambisols	{WD-ISIS}	13.41305556	55.54805556
65099	1564	Dystric Gleysols	Gleysols	{CN-SOTER}	122.07846832	41.41205597
65099	1564	Eutric Gleysols	Gleysols	{CN-SOTER}	122.07846832	41.41205597
70364	1564	Eutric Planosols	Planosols	{EU-SOTER,EU-SPADE,WD-WISE}	27.2	57.84972222
70364	1564	Dystric Planosols	Planosols	{EU-SOTER,EU-SPADE,WD-WISE}	27.2	57.84972222
SIM13684	1564	Cambic Umbrisols	Umbrisols	HWSDv2	-5.194847363232742	43.4662615190105
SIM61025	1564	Greyzemic Phaeozems	Phaeozems	HWSDv2	49.73152185915074	55.944649991583
612391	2932	Cambic Umbrisols	Umbrisols	{WD-eSOTER}	-4.86389017	57.99440002
612391	2932	Acric Umbrisols	Umbrisols	{WD-eSOTER}	-4.86389017	57.99440002
612393	2932	Rustic Podzols	Podzols	{WD-eSOTER}	-3.89582992	57.20560074
612393	2932	Albic Podzols	Podzols	{WD-eSOTER}	-3.89582992	57.20560074
612470	2932	Fibric Histosols	Histosols	{WD-eSOTER}	-4.75	58.03609848
612470	2932	Sapric Histosols	Histosols	{WD-eSOTER}	-4.75	58.03609848
66794	2932	Rustic Podzols	Podzols	{WD-WISE}	-3.625	57.20833333
66794	2932	Albic Podzols	Podzols	{WD-WISE}	-3.625	57.20833333

66798	2932	Rustic Podzols	Podzols	{WD-WISE}	-3.67777778	57.23888889
66798	2932	Albic Podzols	Podzols	{WD-WISE}	-3.67777778	57.23888889
66799	2932	Rustic Podzols	Podzols	{WD-WISE}	-4.10416667	57.20555556
66799	2932	Albic Podzols	Podzols	{WD-WISE}	-4.10416667	57.20555556