Cross-Continental Transfer Learning for Soil Organic Carbon Prediction: Bridging European Spectral Libraries to Chinese Croplands via Multi-Strategy Domain Adaptation

Xionghai Chen, Xiaojun Liu, Yongchao Tian, Yan Zhu, Weixing Cao, Qiang Cao*

National Engineering and Technology Center for Information Agriculture, MOE Engineering and Research Center for Smart Agriculture, MARA Key Laboratory for Crop System Analysis and Decision Making,Jiangsu Key Laboratory for Information Agriculture, Collaborative Innovation Centre for Modern Crop Production Co-sponsored by Province and Ministry, Nanjing Agricultural University, Nanjing 211800, China.

** Corresponding author: qiangcao@njau.edu.cn*

A paper from the Proceedings of the

17th International Conference on Precision Agriculture

13–16 July 2026, Porto Alegre, Brazil

Abstract

Spatially explicit soil organic carbon (SOC) mapping is essential for carbon accounting and precision agriculture, yet field sampling remains prohibitively costly in many regions. Cross-continental transfer learning offers a means to leverage data-rich sources for data-scarce targets, but faces severe domain shift from divergent climate, pedology, and spectral conditions. We present a framework that integrates (i) a four-level (L1–L4) bare-soil composite strategy in Google Earth Engine, combining eight spectral indices across six phenological windows; (ii) a feature-engineering pipeline producing 80 predictors from Sentinel-2 bands, spectral indices, climate variables, and importance-guided interactions; and (iii) systematic benchmarking of eight transfer strategies—DirectTransfer, TargetOnly, CORAL+, TrAdaBoost, WeightedJoint, TabPFN Transfer, Progressive Fine-Tuning, and a proposed Multi-Model Ensemble Fusion (MultiFusion)—with Gaussian-process residual correction across 16 source–target quality-level combinations. Using the European LUCAS database (n = 497) as source and two Chinese sites—Suining (n = 154) and Baima (n = 29)—as targets, MultiFusion consistently outperformed alternatives, achieving $R^2 = 0.618$ (RMSE = 1.32 g kg^{-1}) for Suining and $R^2 = 0.442$ (RMSE = 2.08 g kg^{-1}) for Baima. Domain-shift magnitude (MMD^2) showed a significant negative correlation with predictive accuracy (Spearman $\rho = -0.515$, $p = 0.003$), supporting its use as an operational transfer-feasibility diagnostic. The L2 (Strict) target composite strategy best balanced spectral purity and sample retention. With only 29 target samples, MultiFusion improved R^2 by 0.39 over the local-only baseline, demonstrating the viability of cross-continental SOC transfer for cost-effective digital soil mapping in under-sampled regions.

Keywords: Soil organic carbon, transfer learning, domain adaptation, Sentinel-2, bare-soil composite, digital soil mapping, LUCAS, ensemble learning, Google Earth Engine.

1. Introduction

Soil organic carbon (SOC) constitutes the largest terrestrial carbon pool, with global stocks estimated at 1,500 Pg C in the upper metre of the pedosphere—approximately twice the atmospheric carbon pool (Batjes, 1996; Lal, 2004). Even modest SOC changes have disproportionate impacts on the global carbon cycle, making accurate quantification essential for greenhouse-gas inventories, carbon-credit verification, and precision agriculture (Minasny et al., 2017; Smith et al., 2020). Yet direct SOC measurement through field sampling and laboratory analysis remains costly, producing a highly uneven global distribution of observations: dense networks in Europe and North America, but vast data-sparse regions across China, sub-Saharan Africa, and Central and South-East Asia (Arrouays et al., 2014; Hengl et al., 2017).

Sentinel-2 provides free, open-access multispectral imagery at 10–20 m resolution with a five-day revisit, well suited to detecting soil chromophores associated with organic matter (Castaldi et al., 2019; Vaudour et al., 2019). A prerequisite for reliable spectral–SOC relationships is the isolation of bare-soil pixels, since even modest vegetation cover obscures pedological information (Demattê et al., 2018). Multi-temporal bare-soil compositing has therefore become standard preprocessing (Rogge et al., 2018), but most existing approaches rely on a single NDVI threshold; residual green vegetation, crop stubble, and soil crusting can pass this filter while still contaminating spectra. No study has yet systematically assessed how the strictness of multi-index filtering affects SOC prediction accuracy, particularly in transfer settings.

A second bottleneck is domain shift—models trained in one region fail elsewhere because of systematic differences in spectral response, soil mineralogy, climate, and management (Pan and Yang, 2010; Padarian et al., 2019). For cross-continental SOC transfer (e.g., Europe to China), both covariate shift and concept drift are expected to be severe. Transfer learning offers techniques to mitigate domain shift through feature alignment (e.g., CORAL), instance reweighting (e.g., TrAdaBoost, KMM), and parameter-based fine-tuning (Weiss et al., 2016). In soil science, transfer learning has been applied at national or intra-continental scales (Padarian et al., 2019; Shi et al., 2015), and global VNIR libraries have been assembled (Viscarra Rossel et al., 2016). However, cross-continental SOC transfer using satellite-derived features remains largely unexplored.

Three research gaps are identified: (1) the feasibility of cross-continental SOC transfer learning from a large European database (LUCAS) to small Chinese target sites using Sentinel-2 features remains unknown; (2) no systematic benchmark exists comparing classical domain-adaptation, instance-reweighting, and emerging foundation-model approaches under the same protocol; and (3) the interaction between bare-soil composite quality and transfer performance has not been quantified. To address these gaps, we propose a three-stage framework (Fig. 1) that acquires Sentinel-2 imagery via Google Earth Engine using a novel four-level multi-index bare-soil screening protocol (L1–L4), constructs an 80-feature matrix with domain-shift diagnostics based on Maximum Mean Discrepancy (MMD^2), and implements eight transfer strategies under all 16 source–target quality-level combinations with Gaussian-process residual correction. Our objectives are: (i) to quantify domain shift between LUCAS and two Chinese sites at multiple bare-soil quality levels; (ii) to systematically evaluate eight transfer strategies; (iii) to examine how bare-soil composite quality interacts with transfer performance; and (iv) to assess the relationship between MMD^2 and achievable accuracy as operational guidance.

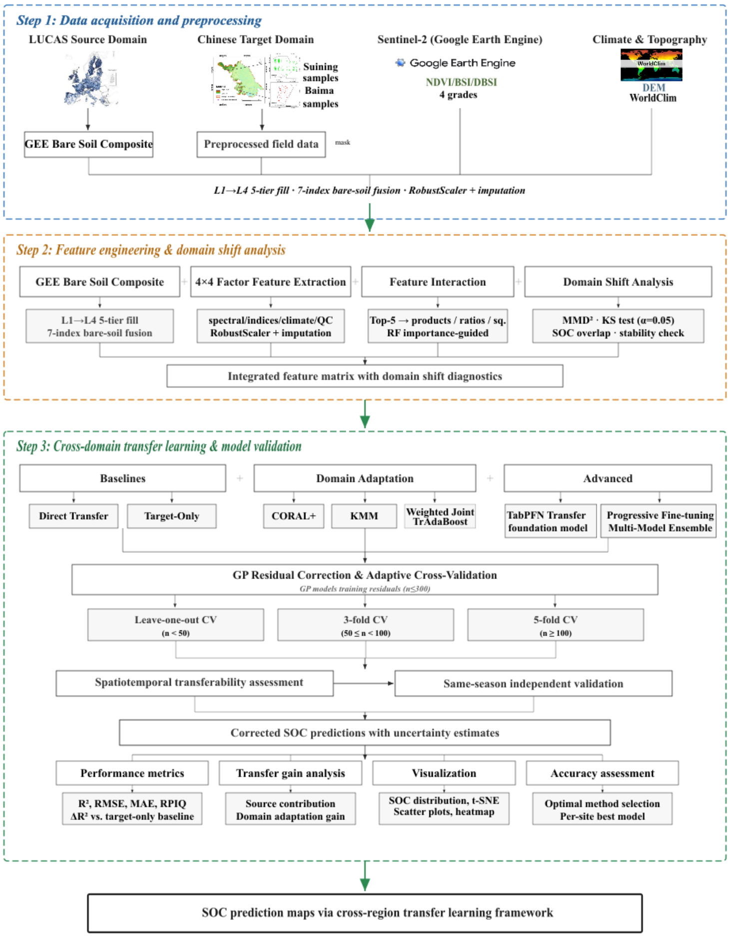

Figure 1. Overall methodological framework for cross-continental transfer learning of soil organic carbon (SOC). Step 1 acquires LUCAS source data, Chinese target data, Sentinel-2 imagery via Google Earth Engine, and climate covariates, with a four-level bare-soil composite. Step 2 performs feature engineering and domain-shift diagnostics. Step 3 implements eight transfer-learning strategies with Gaussian-process residual correction and adaptive cross-validation.

2. Materials and Methods

2.1. Study sites and source database

Two target sites in Jiangsu Province, eastern China, were selected to represent contrasting spatial scales and sample densities within the same warm-temperate agro-climatic zone (Fig. 2; Table 1). Suining County lies on the Huang–Huai–Hai Plain on Quaternary alluvial deposits and represents a regional-scale, intensively managed wheat–maize landscape. Baima Farmland, an experimental farm of Nanjing Agricultural University in the hilly terrain south of the Yangtze River, represents a field-scale paddy–upland cropland on Anthrosols. Sampling at Suining was conducted in June 2020 immediately after wheat harvest; Baima sampling was conducted in November 2021 during a comparable post-harvest period. After quality filtering, 154 and 29 samples were retained for Suining and Baima respectively, with Baima representing an extreme low-sample-size scenario.

The source domain was drawn from the LUCAS 2018 topsoil dataset (Orgiazzi et al., 2018). 497 samples were selected satisfying: (i) agricultural land use; (ii) SOC ranging 2–30 g kg^{-1} to overlap with the Chinese targets; and (iii) valid Sentinel-2 observations in matching bare-soil windows. Selected samples span Mediterranean, Atlantic, Continental and Boreal bioclimatic zones, providing a diverse training set with mean SOC = 12.5 g kg⁻¹ (CV = 49%).

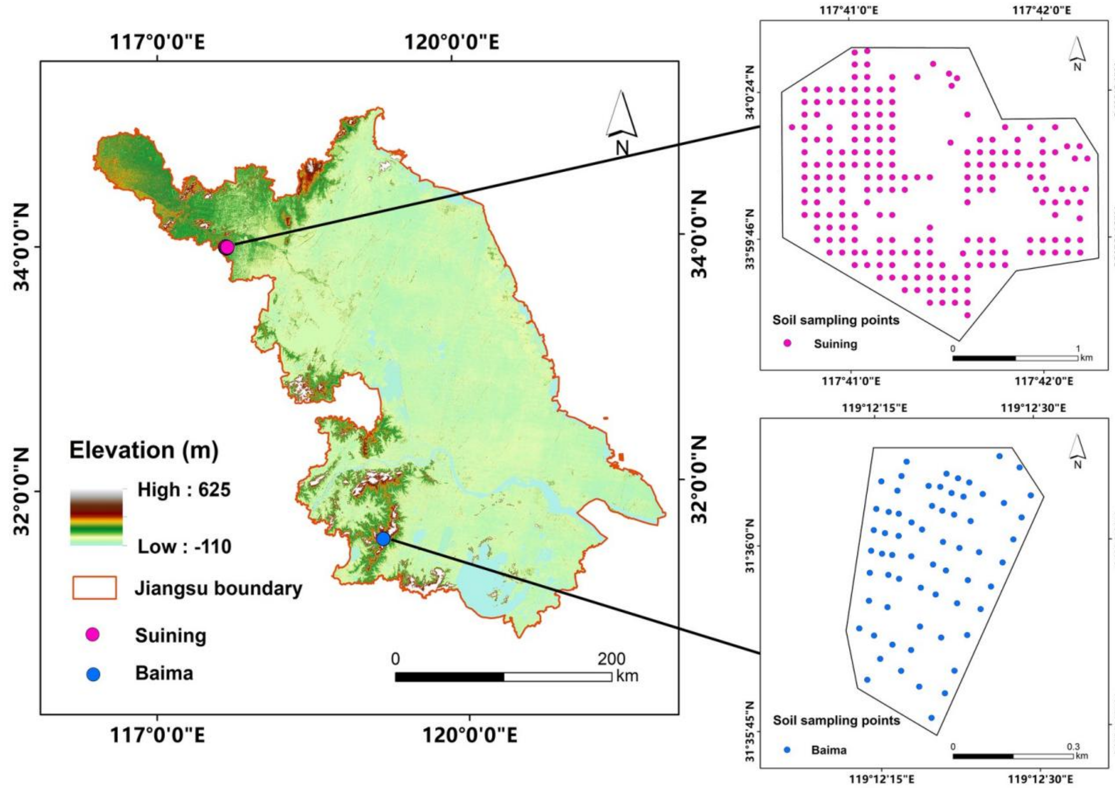

Figure 2. Location of the study areas and soil sampling sites. (a) Geographic overview showing the positions of Suining County and Baima Farmland in Jiangsu, China. (b) Spatial distribution of soil sampling sites in Suining County (n = 154 retained for modelling). (c) Spatial distribution of sampling sites at Baima Farmland (n = 29 retained for modelling).

Table 1. Key characteristics of the two Chinese target sites used in this study.

Attribute	Suining County	Baima Farmland	Notes / units
Coordinates	33.9°N, 117.9°E	31.6°N, 119.2°E	WGS-84
Area	≈ 1,770 km²	≈ 4 km²	Regional vs. field
Soil classification	Fluvisols, Cambisols	Anthrosols	WRB
MAT / MAP	≈ 14.5 °C / 870 mm	≈ 15.4 °C / 1,050 mm	Summer monsoon
Cropping system	Winter wheat–summer maize	Paddy rice / upland	—
Sampling date	June 2020	November 2021	0–20 cm topsoil
Samples (collected/retained)	203 / 154	65 / 29	After QC filtering
SOC range (g kg⁻¹)	5.1–15.4	2.4–11.3	—
SOC mean; CV	9.2 g kg⁻¹; 21%	5.8 g kg⁻¹; 44%	Mean and CV

Note. MAT = mean annual temperature; MAP = mean annual precipitation; CV = coefficient of variation of retained-sample SOC.

2.2. Four-level multi-index bare-soil screening

Sentinel-2 Level-2A surface-reflectance imagery (COPERNICUS/S2_SR_HARMONIZED) was accessed through Google Earth Engine. For each site, all scenes within six phenological windows over a 17-month period (covering pre- and post-harvest, winter fallow, and spring pre-sowing) were filtered for cloud cover (<30%), with masking of cloud shadow, thin cirrus, and water. Pixel-level temporal composites were median values to reduce noise.

A novel four-level multi-index protocol systematically controls composite quality, simultaneously evaluating eight spectral indices (Eqs. 1–8), where ρ denotes Sentinel-2 surface reflectance:

$$NDVI = (\rho\_NIR - \rho\_Red) / (\rho\_NIR + \rho\_Red) \tag{1}$$

$$BSI = [(\rho\_SWIR1 + \rho\_Red) - (\rho\_NIR + \rho\_Blue)] / [(\rho\_SWIR1 + \rho\_Red) + (\rho\_NIR + \rho\_Blue)] \tag{2}$$

$$DBSI = (\rho\_SWIR1 - \rho\_Green) / (\rho\_SWIR1 + \rho\_Green) - NDVI \tag{3}$$

$$NBR2 = (\rho\_SWIR1 - \rho\_SWIR2) / (\rho\_SWIR1 + \rho\_SWIR2) \tag{4}$$

$$EVI = 2.5 \times (\rho\_NIR - \rho\_Red) / (\rho\_NIR + 6\rho\_Red - 7.5\rho\_Blue + 1) \tag{5}$$

$$SAVI = [(\rho\_NIR - \rho\_Red) / (\rho\_NIR + \rho\_Red + L)] \times (1 + L),\ \ L = 0.5 \tag{6}$$

$$NDWI = (\rho\_Green - \rho\_NIR) / (\rho\_Green + \rho\_NIR) \tag{7}$$

$$BI = \sqrt{[(\rho\_Red^2 + \rho\_Green^2 + \rho\_Blue^2) / 3]} \tag{8}$$

NDVI is the primary vegetation metric; BSI is positive over exposed mineral soil; DBSI distinguishes dry bare soil from moist or vegetated surfaces; NBR2 captures crop-residue contamination through SWIR contrasts; EVI and SAVI are vegetation indices for low-biomass conditions; NDWI flags waterlogged pixels; BI excludes deep shadows and anomalously bright surfaces.

Four quality levels were defined by progressively relaxing index thresholds. Level 1 (L1, "Best") applies the strictest criteria (NDVI < 0.15, BSI > 0.10, DBSI > 0, BI > 0.15, NBR2 < 0.05, EVI < 0.10, SAVI < 0.12, NDWI < −0.05). Level 2 (L2, "Strict") relaxes thresholds moderately. Level 3 (L3, "Moderate") permits low vegetation cover. Level 4 (L4, "Relaxed") is a fallback with the most permissive criteria. A five-tier hierarchical gap-filling strategy then assigned each sample location the best available observation across L1 (primary window) → L1 (secondary windows) → L2 → L3 → L4, maximising spatial coverage while preserving quality provenance.

2.3. Feature engineering and domain-shift quantification

For each sample, 80 candidate features were extracted in four thematic groups: spectral bands (10), spectral indices (37+), climate variables (37, from WorldClim V1, ERA5-Land, and TerraClimate), and importance-guided feature interactions (≤15). Preprocessing included missing-value removal, KNN imputation ($k = 5$), low-variance removal, RobustScaler normalisation, and 2 × IQR outlier detection. Multi-method feature selection integrated five ranking methods (Pearson correlation, mutual information, F-regression, XGBoost importance, Random Forest importance) via weighted rank aggregation.

Domain shift was quantified using Maximum Mean Discrepancy (MMD^2) on the standardised feature matrix with a Gaussian kernel (Gretton et al., 2012):

$$MMD^2(X_s, X_t) = (1/n_s^2)\Sigma k(x_{si},x_{sj}) + (1/n_t^2)\Sigma k(x_{ti},x_{tj}) - (2/(n_s n_t))\Sigma k(x_{si},x_{tj}) \tag{9}$$

$$k(x, y) = \exp(-\|x - y\|^2 / (2\sigma^2)) \tag{10}$$

where σ is set to the median pairwise distance. $MMD^2 = 0$ indicates identical distributions; larger values reflect greater divergence. The two-sample Kolmogorov–Smirnov test was applied to SOC distributions, and a per-feature stability score combining distribution overlap, range ratio, and (1 − KS) was computed; features with score ≥ 0.70 were classified as "stable".

2.4. Transfer-learning strategies

Eight transfer strategies were benchmarked, spanning four categories. All except DirectTransfer incorporate a Gaussian-process (GP) residual-correction step (§2.5).

Baselines. DirectTransfer trains a Random Forest (RF; 500 trees, max depth 15) only on source data and applies it directly to the target without adaptation. TargetOnly trains an RF only on target data, providing the local-only reference.

Classical domain adaptation. CORAL+ aligns source and target second-order statistics by computing a linear transformation (Sun et al., 2016): $\hat{X}_s = X_s C_s^{-1/2} C_t^{1/2}$, extended with whitening regularisation ($\lambda = 1.0$) to stabilise high-dimensional feature spaces.

Instance reweighting. TrAdaBoost iteratively down-weights misclassified source instances and up-weights difficult target instances over 50 boosting rounds (Dai et al., 2007). WeightedJoint combines source and target into one training set with weights determined by Kernel Mean Matching (KMM; Huang et al., 2006), so that the weighted source marginal matches the target marginal in the kernel space.

Advanced methods. TabPFN Transfer applies a transformer pre-trained on synthetic tabular datasets (Hollmann et al., 2023). Progressive Fine-Tuning pre-trains a three-layer neural network (128–64–32 hidden units) on the source for 100 epochs, then progressively unfreezes layers from top to bottom over 50 fine-tuning epochs (Yosinski et al., 2014). Multi-Model Ensemble Fusion (MultiFusion), our proposed strategy, operates in three stages: (i) four base models (RF, XGBoost, LightGBM, Ridge) are trained on three representations of the source–target data (raw joint, CORAL-aligned, KMM-reweighted), producing 12 base predictions; (ii) a Ridge meta-learner is trained on target cross-validated base predictions to learn optimal stacking weights, automatically up-weighting models that perform well on the target; and (iii) a GP residual model with a Matérn-5/2 kernel captures any remaining spatially structured bias.

2.5. Gaussian-process residual correction and evaluation

For all strategies except DirectTransfer, a GP regression with a Matérn-5/2 kernel was fitted to training-set residuals (Rasmussen & Williams, 2006):

$$k_{5/2}(r) = \sigma\_f^2 \left(1 + \sqrt{5}\, r/\ell + 5r^2/(3\ell^2)\right) \exp(-\sqrt{5}\, r/\ell), \quad r = \|x - x'\| \tag{11}$$

Hyperparameters were fitted via maximum marginal likelihood under automatic relevance determination. The GP yields both a point correction and $\pm 2\sigma$ posterior predictive intervals.

An adaptive cross-validation (CV) scheme was employed: leave-one-out CV (LOOCV) for $n < 50$ (Baima), 3-fold for $50 \leq n < 100$, and 5-fold for $n \geq 100$ (Suining). Independent held-out test sets were retained for unbiased generalisation: stratified ~50/50 split for Baima ($n = 14$ train, $n = 15$ test); ~62/38 split for Suining ($n = 96$ train, $n = 58$ test). Six metrics were computed: R^2, RMSE, MAE, RPD (= SD(y)/RMSE), Lin's concordance correlation coefficient (CCC), and bias. Transfer gain was computed as $\Delta R^2 = R^2_m - R^2\_\text{baseline}$ against both TargetOnly and DirectTransfer baselines, averaged across the 16 quality-level combinations.

3. Results

3.1. SOC distributions and domain shift

Source and target domains differed markedly in both shape and central tendency of SOC. The LUCAS distribution was broad and positively skewed (CV = 49%), Suining was narrower and shifted toward lower values (CV = 21%), and Baima was even more compressed and left-shifted (CV = 44%). Two-sample KS tests rejected distributional homogeneity for every source–target pair ($p < 0.001$), with Baima showing systematically larger KS statistics than Suining (Table 2)—confirming more severe domain shift at the smaller, lower-SOC site. MMD^2 distances were also consistently higher for LUCAS–Baima (1.08–1.16) than for LUCAS–Suining (0.79–0.99), and the L2 (Strict) source level yielded the lowest MMD^2 for both targets, suggesting that intermediate strictness optimises the spectral-purity–representativeness trade-off.

Table 2. Domain-shift quantification between LUCAS source and target domains for matched source–target quality-level combinations.

Source / Target level	Site	MMD^2	SOC overlap (%)	Sig. features	SOC KS stat.
L1 / L1	Suining	0.988	100.0	61	0.432
L1 / L1	Baima	1.157	88.8	63	0.747
L2 / L2	Suining	0.800	100.0	53	0.553
L2 / L2	Baima	0.922	84.3	54	0.814
L3 / L3	Suining	0.942	100.0	60	0.571
L3 / L3	Baima	1.094	84.3	58	0.810
L4 / L4	Suining	0.928	100.0	63	0.566
L4 / L4	Baima	1.083	92.3	62	0.799

Note. Only diagonal (matched-quality) combinations are shown for compactness; full 32-combination matrix is available in the journal version. SOC overlap = proportion of target SOC range covered by source range. Sig. features = number of features with significant distributional differences ($\alpha = 0.05$). KS stat. = Kolmogorov–Smirnov statistic for SOC distributions.

3.2. Effects of bare-soil quality levels on prediction accuracy

The 16 source–target quality-level combinations produced systematic accuracy patterns (Fig. 3). For Suining, the best R^2 across the eight methods per combination spanned a moderate range (0.43–0.62); Baima models showed a wider and overall lower range (0.00–0.44). At both sites, the L2 (Strict) target quality level yielded the highest column-mean R^2, outperforming the stricter L1—where the filter became over-restrictive, particularly at Baima—and the more permissive L3/L4 levels, where vegetative contamination attenuated the spectral–SOC signal. Source-level effects were weaker and site-specific: relaxed L4 was optimal for Suining; moderate L3 was optimal for Baima, indicating that source and target quality levels may need to be tuned independently. Matched source–target combinations produced a small but consistent advantage over unmatched ones ($\Delta R^2 = +0.005$ at Suining; +0.014 at Baima).

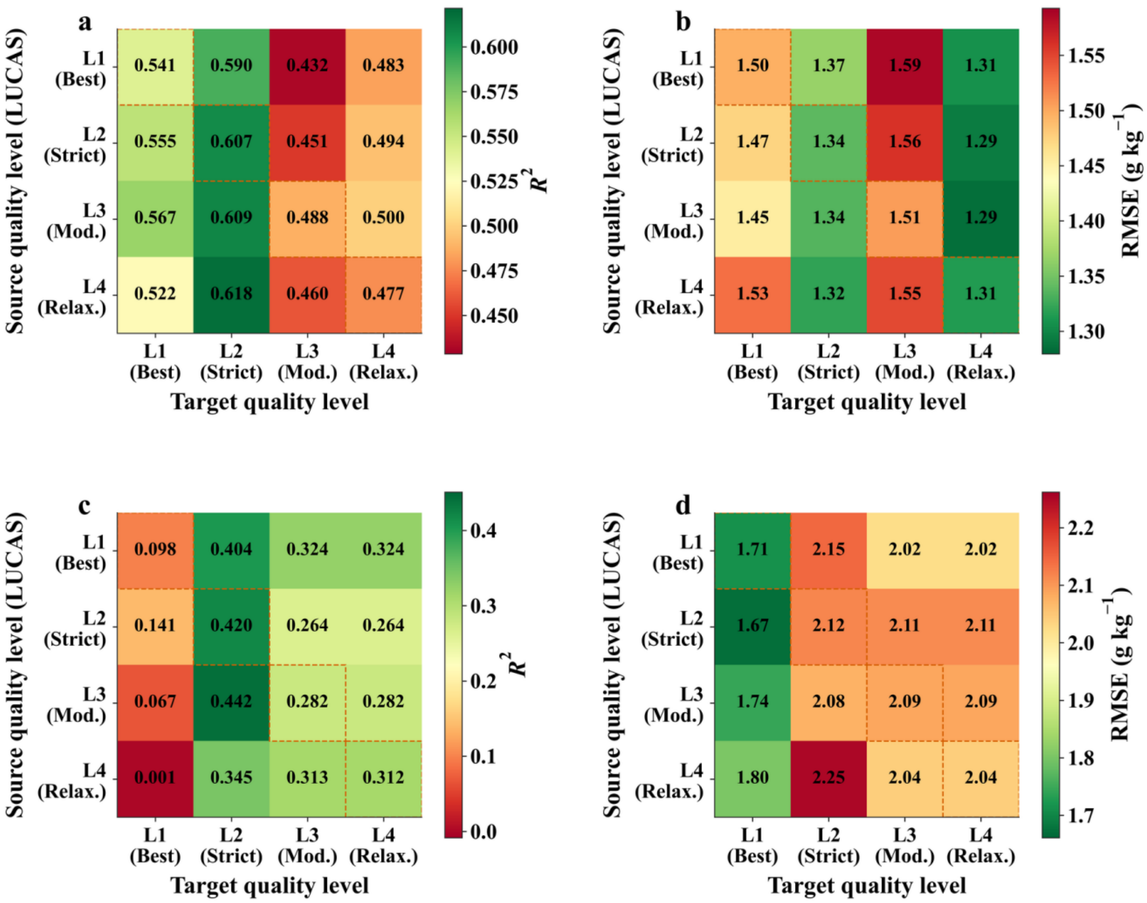

Figure 3. Prediction accuracy heatmaps across all 16 source–target quality-level combinations. Rows: LUCAS source quality level; columns: target quality level. Dashed orange boxes highlight diagonal cells where source and target quality levels are matched. (a, b) R^2; (c, d) RMSE (g kg^{-1}). Upper panels: Suining; lower panels: Baima.

3.3. Comparison of transfer-learning strategies

Method-level performance distributions across all 16 quality-level combinations are summarised in Fig. 4, with optimal-combination test-set results given in Table 3. At Suining (optimal combination L4 → L2, n = 58), MultiFusion ranked first across all six metrics (R^2 = 0.618, RMSE = 1.32 g kg^{-1}, RPD = 1.62, CCC = 0.600), closely followed by WeightedJoint (R^2 = 0.608) and TabPFN Transfer (R^2 = 0.592); CORAL+ was competitive (R^2 = 0.567), while TrAdaBoost and ProgressiveFineTune exhibited clear negative transfer ($R^2 \approx 0$ and −0.220 respectively), and DirectTransfer was severely miscalibrated. The TrAdaBoost failure—CCC and slope of essentially zero—indicates collapse to predicting the sample mean.

At Baima (optimal combination L3 → L2, n = 15), the contrast was even more pronounced. MultiFusion retained moderate but useful accuracy (R^2 = 0.442, RMSE = 2.08 g kg^{-1}, RPD = 1.34), while the TargetOnly baseline dropped below zero (R^2 = −0.041)—predictions worse than the sample mean—demonstrating that local-only modelling fails under extreme sample scarcity. Predicted-versus-measured plots (Fig. 5) confirm that MultiFusion predictions cluster tightly around the 1:1 line at Suining and show moderate, unbiased scatter at Baima. Mean bias was effectively zero at both sites (Suining: 0.00; Baima: −0.00 g kg^{-1}), consistent with the role of the GP residual correction.

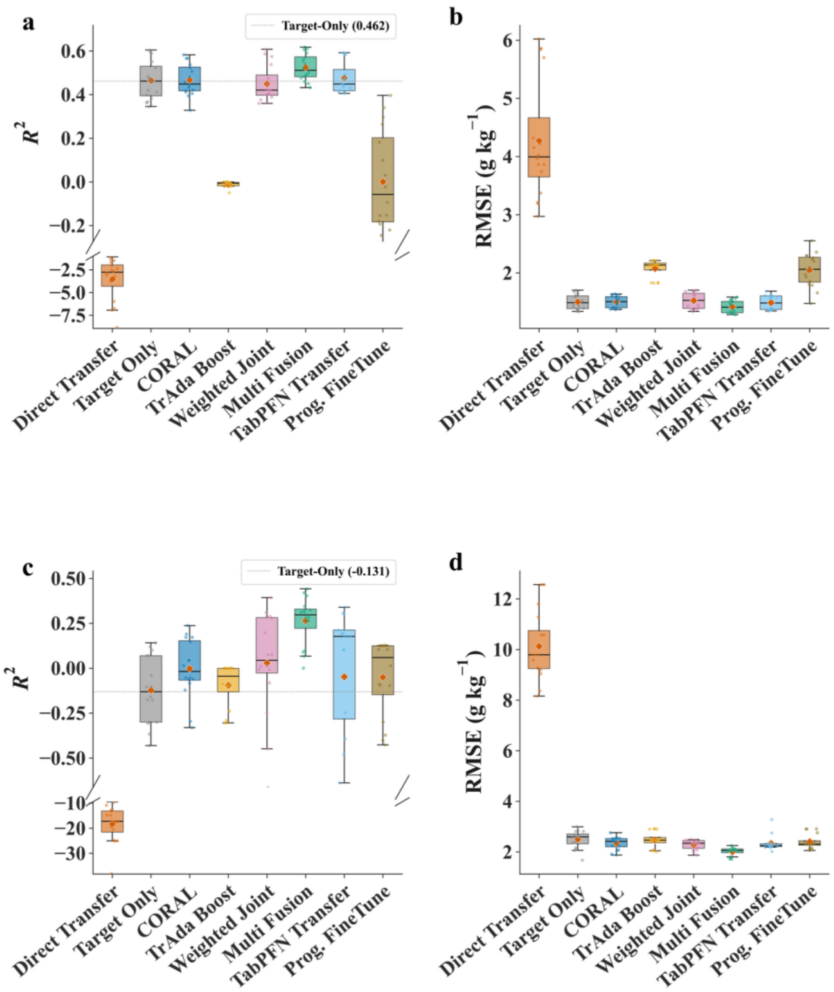

Figure 4. Performance distributions of eight transfer-learning methods across all 16 quality-level combinations. Box plots overlaid with jittered points (one per combination); diamonds indicate cross-combination means. The R^2 panels use a broken y-axis to separate the extreme negative values of DirectTransfer from the compact range of the remaining methods; dotted grey line: TargetOnly median baseline. (a, b) R^2; (c, d) RMSE (g kg^{-1}). Left column: Suining; right column: Baima.

Table 3. Performance of the eight transfer strategies on the held-out test sets at the optimal source–target quality-level combination for each site (L4 → L2 for Suining, L3 → L2 for Baima).

Method	Site	R^2	RMSE	MAE	RPD	CCC	Bias	Slope	*n*
DirectTransfer	Suining	−3.09	4.32	3.76	0.49	0.002	3.76	0.029	58
TargetOnly	Suining	0.604	1.34	1.04	1.59	0.586	−0.08	0.604	58
CORAL+	Suining	0.567	1.41	1.09	1.52	0.549	−0.07	0.579	58
TrAdaBoost	Suining	−0.002	2.14	1.76	1.00	0.000	0.09	0.000	58
WeightedJoint	Suining	0.608	1.34	1.05	1.60	0.585	0.05	0.589	58
MultiFusion	**Suining**	**0.618**	**1.32**	**1.03**	**1.62**	**0.600**	**0.00**	**0.618**	**58**
TabPFN Transfer	Suining	0.592	1.36	1.06	1.57	0.565	−0.02	0.563	58
ProgressiveFineTune	Suining	−0.220	2.36	1.99	0.91	0.239	−1.71	0.275	58
DirectTransfer	Baima	−9.60	9.05	8.68	0.31	0.010	8.68	0.079	15
TargetOnly	Baima	−0.041	2.84	2.38	0.98	0.114	−0.18	0.257	15
CORAL+	Baima	0.147	2.57	2.21	1.08	0.189	0.01	0.308	15

Method	Site	R^2	RMSE	MAE	RPD	CCC	Bias	Slope	*n*
TrAdaBoost	Baima	−0.097	2.91	2.37	0.95	0.000	−0.86	0.000	15
WeightedJoint	Baima	0.280	2.36	2.02	1.18	0.311	0.28	0.433	15
MultiFusion	**Baima**	**0.442**	**2.08**	**1.83**	**1.34**	**0.408**	**−0.00**	**0.442**	**15**
TabPFN Transfer	Baima	0.340	2.26	1.92	1.23	0.272	−0.06	0.233	15
ProgressiveFineTune	Baima	−0.094	2.91	2.36	0.96	0.006	−0.90	0.005	15

Note. Bold orange row indicates the best-performing method (MultiFusion) at each site. RMSE, MAE and Bias in g kg⁻¹. RPD = ratio of performance to deviation. CCC = Lin's concordance correlation coefficient.

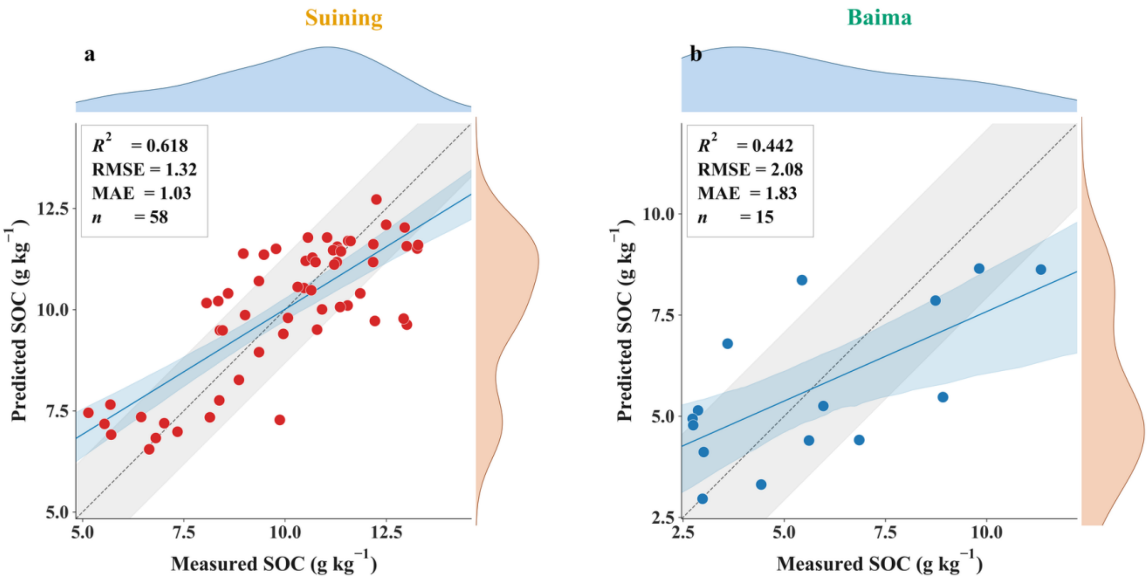

Figure 5. Predicted versus measured SOC for the best-performing model (MultiFusion) under the optimal quality-level combination. Dashed grey line: 1:1 reference; solid blue line: ordinary least-squares regression with shaded 95% confidence band. Insets report R^2, RMSE, MAE and n. Marginal histograms show the distributions of measured (top) and predicted (right) SOC. (a) Suining (L4 → L2, $R^2 = 0.618$, n = 58). (b) Baima (L3 → L2, $R^2 = 0.442$, n = 15).

Transfer-gain analysis revealed the largest positive gain for MultiFusion at both sites, with particularly dramatic improvement at Baima ($\Delta R^2 = +0.390$ over TargetOnly) where the local-only baseline failed. Relative to DirectTransfer, all adaptation methods achieved large positive gains at both sites, with Baima gains an order of magnitude larger because the DirectTransfer baseline was severely miscalibrated there. The uniformly large gains relative to DirectTransfer confirm that domain adaptation is indispensable for cross-continental transfer.

3.4. Domain-shift magnitude and achievable accuracy

Across all 32 quality-level combinations, the best achievable R^2 decreased monotonically with increasing MMD^2 (Spearman $\rho = -0.515$, $p = 0.003$; Fig. 6a). Combinations with $MMD^2 < 0.85$ consistently exceeded $R^2 = 0.50$, whereas combinations with $MMD^2 > 1.10$ fell below $R^2 = 0.30$. This monotonic relationship validates MMD^2 as an a-priori diagnostic of transfer feasibility. Comparison of matched versus unmatched quality-level combinations (Fig. 6b) showed a slight median-R^2 advantage of matching, more pronounced at Baima.

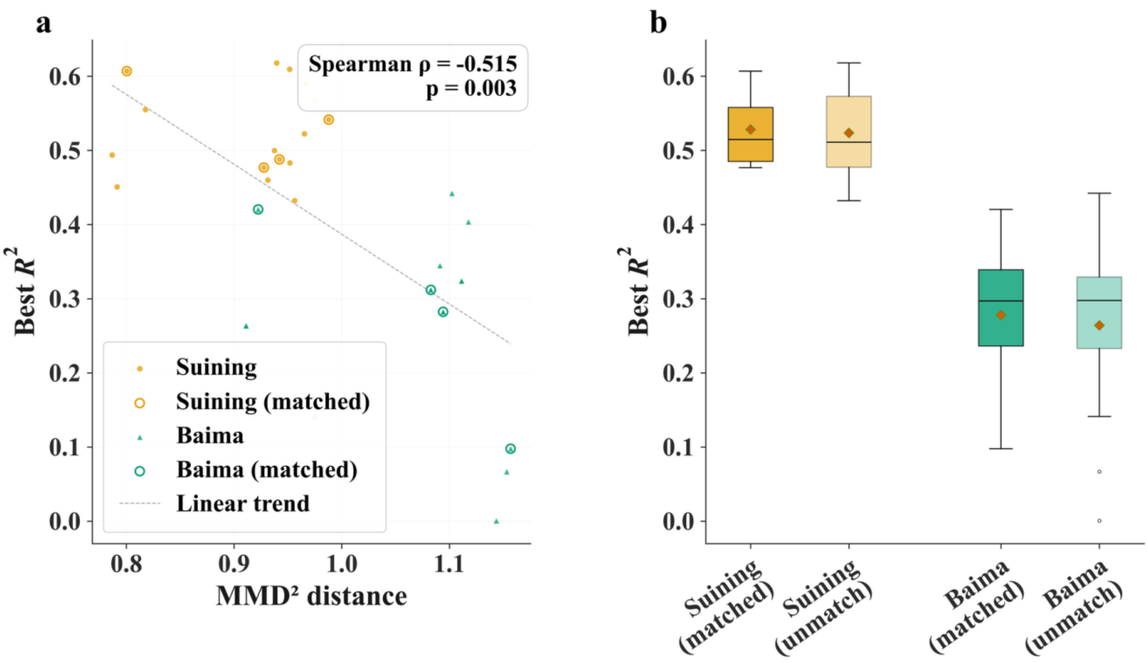

Figure 6. Domain-shift magnitude and quality-level matching effects on prediction performance. (a) Best R^2 as a function of MMD² distance across all 16 quality-level combinations per site; orange circles: Suining; green triangles: Baima; open symbols: matched quality-level combinations; dashed grey line: linear trend. Spearman $\rho = -0.515$, $p = 0.003$. (b) Box plots comparing best R^2 between matched and unmatched quality-level combinations; diamonds indicate the mean.

4. Discussion

4.1. Cross-continental SOC transfer is feasible in data-scarce regions

This study provides the first empirical evidence that cross-continental SOC transfer learning—from LUCAS to Chinese agricultural sites—can substantially outperform local-only modelling when multi-strategy domain adaptation is employed. This challenges the prevailing assumption in digital soil mapping that calibration data must originate from the same pedological province as the prediction target (Viscarra Rossel et al., 2016) and opens a scalable pathway for leveraging open European soil databases to support SOC monitoring in data-scarce regions worldwide.

The Suining performance of MultiFusion ($R^2 = 0.618$) is comparable to—and in places exceeds—values reported for within-region Sentinel-2 SOC models in Europe (Castaldi et al., 2019: $R^2 = 0.45$–0.68; Vaudour et al., 2019: 0.30–0.55), despite extreme environmental and pedological contrasts and without the hundreds of local samples those studies required. The Baima result is arguably of greater practical significance: with only 29 samples the TargetOnly baseline fails outright, while MultiFusion converts this failure into a useful prediction ($R^2 = 0.442$, $\Delta R^2 = +0.390$)—one of the largest documented transfer gains in the SOC literature. The near-zero mean bias and well-behaved residuals confirm that transferred knowledge produces unbiased, physically plausible predictions, providing compelling evidence that transfer learning is most valuable precisely where it is most needed.

4.2. Ensemble diversity as an antidote to negative transfer

MultiFusion's consistent superiority across both sites and all 16 quality-level combinations stems from three complementary mechanisms. First, model and representation diversity provides structural resilience against heterogeneous shift: cross-continental transfer introduces simultaneous covariate shift, concept drift, and label shift, and no single adaptation strategy addresses all three. By ensembling

RF, XGBoost, LightGBM, and Ridge regression across raw, CORAL-aligned, and KMM-reweighted representations, MultiFusion hedges against the failure of any individual assumption.

Second, the target-domain-optimised stacking meta-learner learns adaptive combination weights, implicitly down-weighting methods that exhibit negative transfer. This is critical: TrAdaBoost collapsed to predicting the sample mean at both sites, and ProgressiveFineTune showed negative R^2 at both sites (Table 3). Equal-weight combination would catastrophically degrade the ensemble; the stacker effectively converts potential negative transfer into near-zero contribution, acting as an implicit safeguard (Rosenstein et al., 2005). Third, the GP residual correction captures spatially correlated bias that individual base models leave unresolved—the near-zero mean bias and approximately normal residual distributions (Fig. 5) confirm that this post-hoc correction removes systematic error without introducing artefactual structure.

4.3. Spectral purity vs sample retention: why the purest composites are not the best

A novel contribution is the systematic evaluation of how bare-soil composite quality interacts with transfer performance. The 4 × 4 factorial design (Fig. 3) reveals a paradox: spectrally purest composites do not yield the best predictions. The L1 (Best) filter, while spectrally cleanest, proves overly restrictive at small sites—at Baima it drove R^2 below 0.15—illustrating the bias–variance trade-off: bias reduction from purer spectra is overwhelmed by variance inflation from a drastically reduced sample set. Conversely, permissive L3/L4 filters maximise retention but admit vegetative contamination that attenuates the spectral–SOC signal. The L2 (Strict) target level consistently occupies the sweet spot, extending the single-site European findings of Diek et al. (2017) to a cross-continental context. Practitioners should invest computational effort primarily in optimising target-side quality filtering—the more influential parameter—while setting source quality more permissively to maximise the training pool.

4.4. MMD² as a transfer-feasibility compass

The significant negative correlation between MMD^2 and best achievable R^2 (Fig. 6a) provides a pre-screening tool absent from the digital soil-mapping toolkit. Before committing resources to model development, the MMD^2 between a candidate source database and the target domain can be computed at negligible cost. We propose the following operational thresholds: $MMD^2 < 0.85$ indicates high transfer potential ($R^2 > 0.50$ consistently achievable); $0.85 \leq MMD^2 < 1.05$ indicates moderate potential ($R^2 \approx 0.30$–0.50), where ensemble methods are strongly recommended to mitigate negative-transfer risk; and $MMD^2 \geq 1.10$ indicates limited potential ($R^2 < 0.30$ expected). Critically, MMD^2 captures distributional divergence across the full feature space and is more informative than univariate metrics such as the KS statistic on SOC alone.

4.5. Limitations and future directions

Target-domain sample sizes (154 and 29) are small by machine-learning standards; Baima results in particular should be interpreted with appropriate caution. Validation on larger and more geographically diverse target sites is essential. Temporal mismatch between LUCAS (2018), Suining (2020), and Baima (2021) conflates temporal SOC dynamics with cross-domain adaptation. The framework does not explicitly model spatial autocorrelation, which could positively bias cross-validation estimates. Productive future research includes: (i) validation across diverse agro-climatic zones; (ii) integration of Sentinel-1 SAR and thermal imagery; (iii) deep domain-adaptation architectures such as domain-adversarial networks; (iv) active-learning strategies that optimally select target samples to minimise sampling cost; and (v) extension to other soil properties (pH, clay, N, available P) within the same framework.

5. Conclusions

We presented the first systematic evaluation of cross-continental transfer learning for satellite-based SOC prediction, bridging the European LUCAS database to Chinese farmland sites via Sentinel-2 bare-soil composites and multi-strategy domain adaptation. The principal findings are:

- Cross-continental SOC transfer learning is feasible in data-scarce regions: MultiFusion achieved strong performance at both Suining ($R^2 = 0.618$) and Baima ($R^2 = 0.442$), substantially surpassing both TargetOnly and DirectTransfer baselines.
- Ensemble-based, multi-strategy adaptation is essential. Among eight benchmarked strategies, MultiFusion consistently ranked first owing to its ensemble diversity, target-optimised stacking, and Gaussian-process residual correction.
- Bare-soil composite quality critically modulates transfer accuracy. The L2 (Strict) target level consistently yielded the best predictions by balancing spectral purity against sample retention.
- Domain distance (MMD^2) emerged as a practical indicator of transfer success (Spearman $\rho = -0.515$, $p = 0.003$), providing an operational diagnostic for pre-screening source–target compatibility. Combinations with $MMD^2 < 0.85$ consistently achieved $R^2 > 0.50$.
- With only 29 target samples, transfer learning transformed an otherwise failed local model into a usable predictor ($\Delta R^2 = +0.390$), demonstrating cost-effective digital soil mapping for under-sampled regions globally.

Acknowledgements

The authors gratefully acknowledge the European Soil Data Centre (ESDAC) for providing access to the LUCAS 2018 topsoil database, the European Space Agency for free access to Sentinel-2 imagery, and Google Earth Engine for cloud-computing infrastructure. We thank colleagues at the National Engineering and Technology Center for Information Agriculture (Nanjing Agricultural University) and field staff at Suining and Baima for assistance with sample collection.

References

Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A., & McBratney, A. B. (2014). GlobalSoilMap: Basis of the Global Spatial Soil Information System. CRC Press, London.

Batjes, N. H. (1996). Total carbon and nitrogen in the soils of the world. European Journal of Soil Science, 47(2), 151–163.

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., & van Wesemael, B. (2019). Evaluating the capability of the Sentinel-2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing, 147, 267–282.

Dai, W., Yang, Q., Xue, G.-R., & Yu, Y. (2007). Boosting for transfer learning. In Proceedings of the 24th International Conference on Machine Learning (pp. 193–200).

Demattê, J. A. M., Fongaro, C. T., Rizzo, R., & Safanelli, J. L. (2018). Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sensing of Environment, 212, 161–175.

Diek, S., Fornallaz, F., Schaepman, M. E., & de Jong, R. (2017). Barest pixel composite for agricultural areas using Landsat time series. Remote Sensing, 9(12), 1245.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. Journal of Machine Learning Research, 13(25), 723–773.

Hengl, T., de Jesus, J. M., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotic, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE, 12(2), e0169748.

Hollmann, N., Müller, S., Eggensperger, K., & Hutter, F. (2023). TabPFN: A transformer that solves small tabular classification problems in a second. In Proceedings of the 11th International Conference on Learning Representations.

Huang, J., Smola, A. J., Gretton, A., Borgwardt, K. M., & Schölkopf, B. (2006). Correcting sample selection bias by unlabeled data. In Advances in Neural Information Processing Systems, 19, 601–608.

Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. Science, 304(5677), 1623–1627.

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., et al. (2017). Soil carbon 4 per mille. Geoderma, 292, 59–86.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., & Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: A review. European Journal of Soil Science, 69(1), 140–153.

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Transfer learning to localise a continental soil vis-NIR calibration model. Geoderma, 340, 279–288.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.

Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA.

Rogge, D., Bauer, A., Zeidler, J., Müller, A., Esch, T., & Heiden, U. (2018). Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). Remote Sensing of Environment, 205, 1–17.

Rosenstein, M. T., Marx, Z., Kaelbling, L. P., & Dietterich, T. G. (2005). To transfer or not to transfer. In Proceedings of the NIPS 2005 Workshop on Transfer Learning.

Shi, Z., Ji, W., Viscarra Rossel, R. A., Chen, S., & Zhou, Y. (2015). Prediction of soil organic matter using a nationally distributed vis-NIR spectral library and a local calibration approach. Australian Journal of Soil Research, 53(6), 627–635.

Smith, P., Soussana, J.-F., Angers, D., Schipper, L., Chenu, C., Rasse, D. P., et al. (2020). How to measure, report and verify soil carbon change. Global Change Biology, 26(1), 219–241.

Sun, B., Feng, J., & Saenko, K. (2016). Return of frustratingly easy domain adaptation. In Proceedings of the AAAI Conference on Artificial Intelligence, 30 (pp. 2058–2065).

Vaudour, E., Gilliot, J. M., Bel, L., Lefevre, J., & Chehdi, K. (2019). Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. International Journal of Applied Earth Observation and Geoinformation, 49, 24–38.

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., et al. (2016). A global spectral library to characterize the world's soils. Earth-Science Reviews, 155, 198–230.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big Data, 3, 9.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In Advances in Neural Information Processing Systems, 27, 3320–3328.

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture.

EXAMPLE: Chen, X., Gao, Y., Yuan, F., Karim, S. T. A., Yu, K., Li, W., Liu, X., Tian, Y., Zhu, Y., Cao, W., & Cao, Q. (2026). Cross-continental transfer learning for soil organic carbon prediction. In Proceedings of the 17th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.