SpecBPP: A Self-Supervised Learning Approach for Hyperspectral Representation and Soil Organic Carbon Estimation

Daniel La'ah Ayuba[1]
d.ayuba@surrey.ac.uk
Jean-Yves Guillemaut[1]
j.guillemaut@surrey.ac.uk
Belen Marti-Cardona[2]
b.marti-cardona@surrey.ac.uk
Oscar Mendez Maldonado[1]
o.mendez@surrey.ac.uk

[1] Center for Vision, Speech, and Signal Processing
University Of Surrey
Guildford, UK

[2] Centre for Environmental Health and Engineering
University Of Surrey
Guildford, UK

Abstract

Self-supervised learning has revolutionized representation learning in vision and language, but remains underexplored for hyperspectral imagery (HSI), where the sequential structure of spectral bands offers unique opportunities. In this work, we propose *Spectral Band Permutation Prediction (SpecBPP)*, a novel self-supervised learning framework that leverages the inherent spectral continuity in HSI. Instead of reconstructing masked bands, SpecBPP challenges a model to recover the correct order of shuffled spectral segments, encouraging global spectral understanding. We implement a curriculum-based training strategy that progressively increases permutation difficulty to manage the factorial complexity of the permutation space. Applied to Soil Organic Carbon (SOC) estimation using EnMAP satellite data, our method achieves state-of-the-art results, outperforming both masked autoencoder (MAE) and joint-embedding predictive (JEPA) baselines. Fine-tuned on limited labeled samples, our model yields an R^2 *of 0.9456, RMSE of 1.1053%*, and *RPD of 4.19*, significantly surpassing traditional and self-supervised benchmarks. Our results demonstrate that spectral order prediction is a powerful pretext task for hyperspectral understanding, opening new avenues for scientific representation learning in remote sensing and beyond.

1 Introduction

Hyperspectral imagery (HSI) provides rich spectral information across hundreds of contiguous wavelength bands, offering powerful capabilities for material characterization on Earth's surface. A vital application is estimating soil organic carbon (SOC) content, a key indicator of soil health and a significant component of the global carbon cycle [53]. Despite the data-rich nature of HSI, obtaining ground-truth SOC labels remains costly and labor-intensive, creating a label-scarce domain where unlabeled hyperspectral data are plentiful but labeled examples

are rare [38]. This gap motivates self-supervised learning approaches to leverage abundant unlabeled HSI for robust representation learning.

Self-supervised learning (SSL) has transformed representation learning by devising *pretext tasks* that require no manual labels. Current approaches include masked image modeling, where models reconstruct masked regions [25], and contrastive/joint embedding methods that learn invariant representations [2, 3, 10]. However, applying these techniques directly to hyperspectral data has limitations: masked reconstruction primarily captures local spectral correlations, while contrastive methods require careful augmentation design for HSI. Unlike RGB images with well-established augmentations (color jitter, cropping), HSI augmentations must preserve spectral fidelity and avoid disrupting absorption features critical for material identification [26, 40]. Existing methods fail to explicitly leverage the sequential structure [20] of hyperspectral data, where bands follow the electromagnetic spectrum's natural ordering and exhibit strong correlations with adjacent wavelengths.

We propose SpecBPP, a novel self-supervised approach tailored to hyperspectral data. SpecBPP treats each spectrum as a sequence, partitioning it down into multiple contiguous segments, shuffling them, and training the model to predict the correct original order. This "spectral jigsaw puzzle" requires the model to learn the natural ordering [42] of wavelength regions and capture long-range dependencies across the spectrum. Unlike masked autoencoding, which focuses on local smoothness, our permutation prediction task encourages global understanding of spectral signatures, requiring networks to identify and place characteristic absorption features in the appropriate spectral context.

To address the factorial growth of possible permutations, we implement a curriculum learning strategy [4, 23, 43], that begins with simpler permutation tasks (fewer segments) and progressively increases complexity as the model learns. This staged approach helps the network gradually master spectral ordering from coarse to fine-grained patterns, making the learning process more efficient and stable.

Empirically, SpecBPP provides outstanding results for SOC prediction. Using our self-supervised task to pretrain a hyperspectral encoder on unlabeled HSI data and fine-tuning on limited SOC-labeled samples, we achieve representations that significantly outperform conventional SSL methods and supervised baselines, particularly in low-label scenarios. Our contributions include (1) a self-supervised framework that uses spectral ordering for hyperspectral representation learning, (2) robust SOC estimation with R^2 of 0.9456, and (3) improved generalization for ML and environmental science applications.

2 Related Work

Self-supervised learning has emerged as a powerful paradigm for learning visual features from unlabeled data. Contrastive approaches like SimCLR [10] and MoCo [24] learn representations by maximizing agreement between augmented views of the same image while distancing different images. Negative-free methods such as BYOL [21] and SimSiam [11] mitigate representational collapse using asymmetric networks and stop-gradient operations. Another successful approach is masked prediction: Masked Autoencoders (MAE) [25] reconstruct randomly masked image patches, forcing models to capture contextual information. More recently, I-JEPA [2] operates in feature space by predicting representations of masked regions, avoiding pixel-level reconstruction while maintaining competitive performance.

Several self-supervised methods leverage intrinsic data ordering as a supervisory signal. In computer vision, solving jigsaw puzzles [37] and predicting relative positions of image

regions [16] enable learning of spatial context. For sequential data like video, models learn temporal structure by verifying frame order [36] or detecting shuffled sequences [17]. Similarly, NLP models like BERT [15] and XLNet [50] incorporate order-based pretraining tasks such as next-sentence prediction and permutation language modeling. These methods share a common principle: shuffling natural data ordering to create self-supervised tasks that require understanding of high-level structure. Our work extends this paradigm to the spectral dimension of hyperspectral imagery, leveraging the natural ordering [42] of the electromagnetic spectrum.

Recent work has begun adapting self-supervised techniques to hyperspectral image (HSI) analysis, where labeled data is often scarce. Contrastive learning approaches [22, 28, 30, 51] for HSI define positive pairs through spectral and spatial augmentations, while masked modeling methods [32, 34, 35, 48] reconstruct masked portions of the hyperspectral cube. Hybrid approaches combine contrastive and reconstruction objectives to leverage the unique characteristics of HSI data [6, 7, 31]. These methods explore various pretext tasks from spectral clustering to band prediction, aimed at learning transferable spectral-spatial features. Our approach contributes to this emerging field by introducing spectral permutation prediction, which explicitly leverages the sequential structure [20] of spectral bands in a way previous HSI self-supervised methods have not explored.

Soil Organic Carbon (SOC) estimation using hyperspectral data has traditionally relied on supervised calibration models trained on field-collected samples with laboratory-measured SOC values. Studies have demonstrated the feasibility of mapping SOC from proximal, airborne, and satellite hyperspectral imagery [44, 45]. Traditional approaches use methods like partial least squares regression (PLSR) [19], while recent work explores spaceborne hyperspectral sensors like EnMAP [5] and PRISMA [39] for regional SOC mapping. Advanced techniques including deep learning have shown promise [13], but supervised approaches remain limited by the need for extensive soil sampling and struggle with generalization across diverse soil conditions. This creates motivation for self-supervised methods that can leverage unlabeled hyperspectral data to learn generalizable spectral features. Our SpecBPP approach addresses this need by learning soil-relevant representations through a novel spectral permutation task, bridging hyperspectral self-supervised learning and practical soil property estimation.

3 Spectral Band Permutation Prediction (SpecBPP)

We introduce **Spectral Band Permutation Prediction (SpecBPP)**, a novel self-supervised framework that leverages the sequential structure of electromagnetic spectra [20, 42]. This section formalizes our task definition, details the model architecture, specifies training objectives, and explains our curriculum learning strategy for managing permutation complexity.

3.1 Task Definition

We formalize **Spectral Band Permutation Prediction** as a self-supervised ordering task. Let $x \in \mathbb{R}^B$ be a spectral signature with B bands (EnMAP: $B = 224$). We partition x into N disjoint segments $x^{(1)}, x^{(2)}, \ldots, x^{(N)}$, each spanning $\ell = B/N$ bands. A random permutation $\pi \in S_N$ (where S_N denotes the symmetric group of all permutations of N elements) reorders these segments to produce a permuted signature:

$$\hat{x} = \pi(x) = \left[x^{(\pi(1))}, x^{(\pi(2))}, \ldots, x^{(\pi(N))}\right] \tag{1}$$

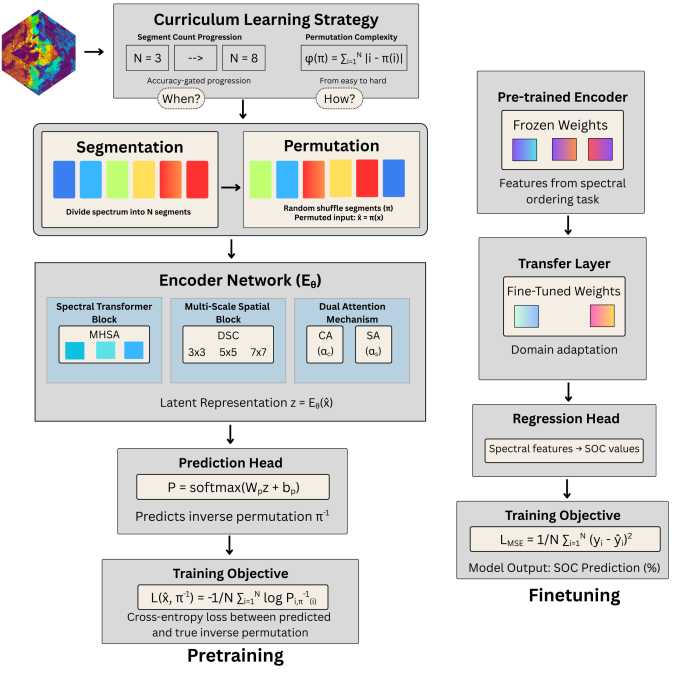


Figure 1: Overview of the **SpecBPP** pretraining–finetuning pipeline. **Pretraining (left):** An input spectrum x is divided into N contiguous segments and randomly permuted by π to form $\pi(x)$. The encoder E_θ produces latent representation z, which feeds into a prediction head that outputs probability matrix P over segment positions. The model predicts the inverse permutation π^{-1} to restore original ordering. Loss is cross-entropy: $\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} \log P_{i,\pi^{-1}(i)}$, where $\pi^{-1}(i)$ is the original position of the segment at shuffled position i. Training follows a curriculum strategy increasing segment count (N:3→8) with accuracy-gated progression. **Finetuning (right):** The pretrained encoder weights are transferred and combined with a regression head for SOC prediction, optimized via MSE loss.

where $\pi(i)$ gives the original segment index at position i in $\hat{x}$. Since π is a bijection, its inverse π^{-1} recovers the original ordering: $\pi^{-1}(\hat{x}) = x$.

In pretraining, the model receives $\hat{x}$ and must predict π^{-1} to reconstruct the original sequence. This prediction corresponds to one of $N!$ possible permutations. By completing this task, the model learns the natural spectral ordering and continuity characteristics found in real hyperspectral signatures.

3.2 Model Architecture

The SpecBPP architecture comprises an encoder network that extracts features from permuted spectral data and a prediction head that estimates the inverse permutation, as illustrated in Fig. 1.

3.2.1 Encoder Network

Our encoder E_θ processes the permuted spectral signature $\hat{x}$ to produce a latent representation $z = E_\theta(\hat{x}) \in \mathbb{R}^d$, where d is the embedding dimension. As shown in Fig. 1(left), the encoder consists of three principal components working in concert:

The **Spectral Transformer Block** employs a multi-head self-attention mechanism to capture dependencies across spectral bands [27, 47]. Formally, given input features $F \in \mathbb{R}^{H\times W\times B}$ (for spatial-spectral data with height H and width W), the self-attention operation is defined as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where Q, K, and V are query, key, and value projections. To incorporate domain knowledge about spectral relationships, we introduce a band weighting mechanism $\alpha \in \mathbb{R}^B$ that adaptively emphasizes spectral regions known to correlate with soil properties. The mechanism applies element-wise multiplication to scale reflectance values by band importance:

$$F' = F \odot \alpha \tag{3}$$

where $\odot$ represents element-wise multiplication broadcast along the spatial dimensions. The **Multi-Scale Spatial Block** captures spatial context at multiple scales through efficient depthwise separable convolutions [12, 29]. For each scale $s \in \{3,5,7\}$ (representing kernel sizes), we compute:

$$F_s = \text{DSConv}_s(F') = \text{PWConv}(\text{DWConv}_s(F')) \tag{4}$$

where DWConv_s is a depthwise convolution with kernel size s, and PWConv is a pointwise (1×1) convolution. The multi-scale features are combined via:

$$F_{\text{MS}} = \text{Concat}(F_3, F_5, F_7)W + b \tag{5}$$

where W and b are learnable parameters.

The **Dual Attention Mechanism** integrates spatial and channel attention to focus on informative regions [18, 49]. The channel attention α_c and spatial attention α_s are computed as:

$$\alpha_c = \sigma(W_2\delta(W_1\text{GAP}(F_{\text{MS}}))) \tag{6}$$

$$\alpha_s = \sigma(f_{7\times7}(\text{MaxPool}(F_{\text{MS}}), \text{AvgPool}(F_{\text{MS}}))) \tag{7}$$

where σ is the sigmoid function, δ is ReLU, GAP is global average pooling, and $f_{7\times7}$ is a 7×7 convolution. The final encoder output integrates these attention mechanisms:

$$z = \text{GAP}(F_{\text{MS}} \odot \alpha_c \odot \alpha_s). \tag{8}$$

3.2.2 Permutation Prediction Head

The prediction head maps encoded representation z to a distribution over possible inverse permutations. To handle the factorial growth of permutation space ($N!$), we adopt a factorized approach treating prediction as N separate classification problems [1, 14].

The head outputs an $N \times N$ matrix P, where P_{ij} represents the probability that segment at position i originated from position j:

$$P = \text{softmax}(W_p z + b_p) \tag{9}$$

where $W_p \in \mathbb{R}^{N\times N\times d}$, $b_p \in \mathbb{R}^{N\times N}$ are learnable parameters with softmax applied row-wise. The predicted inverse permutation is:

$$\hat{\pi}^{-1}(i) = \underset{j\in\{1,2,\ldots,N\}}{\operatorname{argmax}} P_{ij}. \tag{10}$$

This approach efficiently decomposes the exponentially large permutation space into a tractable form.

3.3 Training Objectives

We train the model to minimize the cross-entropy loss between the predicted and true inverse permutation. Given a permuted spectrum $\hat{x} = \pi(x)$ and its inverse permutation π^{-1}, the model predicts which original segment belongs at each position. $\pi^{-1}(i)$ represents the original position of the segment currently at shuffled position i, which is the ground truth target for our prediction. The correct segment index for position i in the sorted spectrum is $\pi^{-1}(i)$. For example, if a segment originally at position 3 is shuffled to position 1, then $\pi^{-1}(1) = 3$, and the model should output high probability $P_{1,3}$. The loss is:

$$\mathcal{L}(\hat{x}, \pi^{-1}) = -\frac{1}{N}\sum_{i=1}^{N} \log P_{i,\pi^{-1}(i)} \tag{11}$$

where $P_{i,\pi^{-1}(i)}$ is the predicted probability that the segment at position i originated from position $\pi^{-1}(i)$.

This objective forces the model to learn the spectral continuity patterns necessary to unscramble the permuted segments. By correctly identifying the original positions, the model must develop representations that capture physically meaningful relationships between different regions of the electromagnetic spectrum.

3.4 Curriculum Learning Strategy

To address the factorial explosion of permutation space ($N!$ possible permutations), we implement a curriculum learning strategy along two dimensions: segment count and permutation complexity.

For segment count progression, we define an adaptive six-phase curriculum from 3 to 8 segments:

$$N_t \;=\; 3 + \sum_{i=1}^{5} \mathbf{1}\big(\text{val\_acc} \geq \alpha_i\big) \tag{12}$$

where $\mathbf{1}(\cdot)$ is the indicator function and α_i are validation accuracy thresholds (99%). This ensures mastery at each level before progressing to permutation spaces of increasing cardinality: 6, 24, 120, 720, 5 040, and 40 320.

Within each phase, we control permutation difficulty using function $\phi(\pi) = \sum_{i=1}^{N} |i - \pi(i)|$ that measures distance from identity [11]. We bias sampling toward "easier" permutations initially, gradually increasing complexity:

$$p(\pi) \propto \exp\left(-\frac{\phi(\pi)}{T_s(t)}\right) \tag{13}$$

where $T_s(t)$ increases with training progress.

Empirically, direct training with $N = 8$ fails to converge, while our curriculum achieves 100% validation accuracy for $N \leq 7$ and 84.2% for $N = 8$, demonstrating its effectiveness for stable learning of spectral structure.

4 Experiments

We evaluate our SpecBPP approach on the task of Soil Organic Carbon (SOC) estimation using hyperspectral imagery. This section describes our experimental setup, presents comparative results against state-of-the-art methods, and provides analysis of the learned representations.

4.1 Experimental Setup

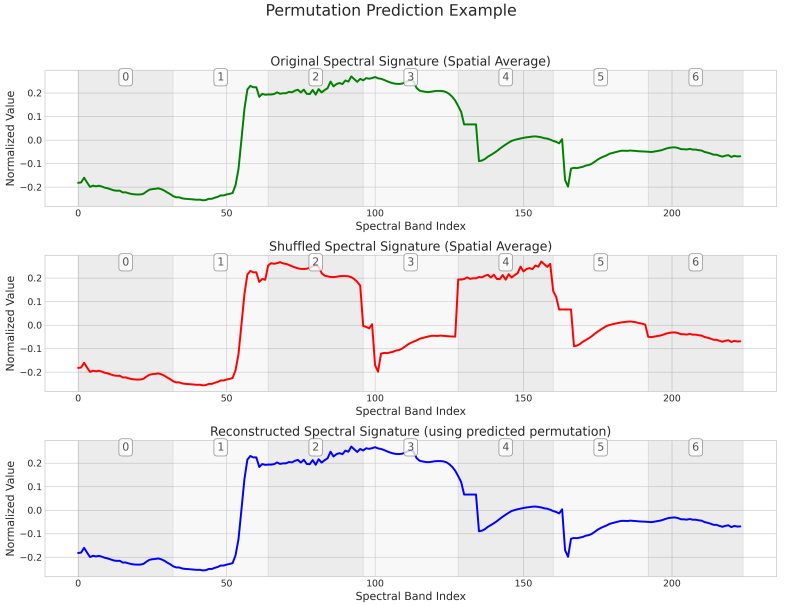
Datasets:
We use EnMAP satellite imagery (224 bands, 420-2450 nm, 30m resolution) [8, 46]. For pretraining, we extract 196 875 non-overlapping 64×64 patches from 1 000 scenes. For fine-tuning, we use patches corresponding to 1 540 soil samples with laboratory-measured SOC (0.5-23.8%), split into training (70%), validation (15%), and test (15%) sets with stratified sampling.
Baseline Methods:
We compare against: (1) traditional methods like Partial Least Squares Regression (PLSR) [19], Random Forest (RF), and Support Vector Regression (SVR); (2) *Supervised*: identical architecture trained only on labeled data; (3) *MAE*: masked autoencoder for HSI [25]; (4) *I-JEPA*: joint-embedding architecture [2]; and (5) *SimCLR*: contrastive learning [10].
Evaluation Metrics:
We use standard regression metrics with explicit formulations for clarity. The coefficient of determination measures explained variance: $R^2 = 1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2$. Root mean square error quantifies prediction magnitude: $\text{RMSE} = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2}$. Mean absolute error provides robust error measurement: $\text{MAE} = \frac{1}{n}\sum_i |y_i - \hat{y}_i|$. The ratio of performance to deviation offers scale-independent assessment: $\text{RPD} = \text{SD}(\mathbf{y})/\text{RMSE}$, where $\text{SD}(\mathbf{y})$ is the standard deviation of reference values. RPD values above 3.0 indicate excellent performance for agricultural applications [9].
Implementation:
Models are implemented in PyTorch and trained on A100 GPUs with batch size 256. We train SpecBPP for 200 epochs (SGD, initial lr=1×10^{-3}, cosine decay) and fine-tune for 150 epochs (lr=5×10^{-4}) with early stopping on validation R^2.

4.2 Results

Pretraining Performance:
Table 1 compares permutation prediction accuracy between direct training and our curriculum strategy across different segment counts. Our curriculum approach achieves 100% accuracy on increasingly complex tasks (3→7 segments) by progressively advancing through segment counts upon mastery. For the most challenging case with 8 segments (40,320 possible permutations), the curriculum strategy still maintains strong performance with 84.2% accuracy, significantly outperforming direct training which only achieves 4.9% after the same number

Table 1: Permutation prediction accuracy (%) on validation data for different segment counts and training strategies.

Training Strategy	Number of Segments (N)					
	3	4	5	6	7	8
Direct (Epoch 30)	100.0	94.8	67.4	35.7	9.2	0.3
Direct (Epoch 100)	100.0	100.0	99.5	70.2	32.5	1.7
Direct (Epoch 200)	100.0	100.0	100.0	87.6	68.3	4.9
Curriculum (Epoch 30)	100.0	-	-	-	-	-
Curriculum (Epoch 50)	-	100.0	-	-	-	-
Curriculum (Epoch 90)	-	-	100.0	-	-	-
Curriculum (Epoch 140)	-	-	-	100.0	-	-
Curriculum (Epoch 170)	-	-	-	-	100.0	-
Curriculum (Epoch 200)	-	-	-	-	-	84.2



	Segment 0	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
True	0	1	3	5	2	4	6
Pred	0	1	3	5	2	4	6

Figure 2: Each panel shows: (top) the original spectral signature with 7 segments, (middle) the shuffled signature after permutation, and (bottom) the reconstructed signature after applying the predicted inverse permutation. The table shows the true vs. predicted segment positions.

of epochs. This performance gap validates the effectiveness of our curriculum approach for complex permutation tasks.

Figure 2 demonstrates the spectral permutation prediction process. An original spectral signature with 7 segments (top) is randomly shuffled (middle), disrupting characteristic absorption features. The model correctly predicts the inverse permutation, reconstructing the original signature (bottom) with 100% segment placement accuracy. The preservation of spectral features like the prominent absorption peak around band 100 and the characteristic shape transitions between segments indicates that the model has learned to recognize and spatially

Table 2: SOC estimation performance comparison. Bold values indicate best performance

Method	Pre-train	R^2 (↑)	RMSE (%) (↓)	MAE (%) (↓)	RPD (↑)
PLSR	✗	0.4823	3.8126	2.9843	1.3903
RF	✗	0.5010	2.7944	2.0871	1.4156
SVR	✗	0.4952	2.8103	2.1067	1.3794
Supervised	✗	0.6231	2.3792	1.3654	2.2276
MAE	✓	0.8145	1.7218	1.0354	3.0783
I-JEPA	✓	0.8630	1.6543	0.9876	3.2037
SimCLR	✓	0.8073	1.7864	1.0892	2.9670
SpecBPP ($N = 3$)	✓	0.6028	2.5629	1.9203	1.6908
SpecBPP ($N = 4$)	✓	0.7236	2.1347	1.4829	2.4651
SpecBPP ($N = 5$)	✓	0.8345	1.6694	1.0147	3.1045
SpecBPP ($N = 6$)	✓	0.8978	1.3421	0.8576	3.7329
SpecBPP ($N = 7$)	✓	**0.9456**	**1.1053**	**0.7394**	**4.1886**
SpecBPP ($N = 8$)	✓	0.9033	1.2987	0.8102	3.8452

Table 3: Ablation study of key components in SpecBPP ($N = 7$).

Model Configuration	R^2 (↑)	RMSE (%) (↓)	RPD (↑)
SpecBPP (Full)	**0.9456**	**1.1053**	**4.1886**
SpecBPP w/o CL	0.9021	1.6632	3.1865
SpecBPP w/o SW	0.9318	1.3892	3.8151
SpecBPP w/o DA	0.9402	1.2147	3.8223

organize distinct spectral patterns rather than relying on low-level statistical correlations. This example illustrates how SpecBPP forces the network to develop a global understanding of spectral signatures that extends beyond local band-to-band relationships.

SOC Estimation Performance:

Table 2 presents SOC estimation results after fine-tuning. SpecBPP performance improves consistently with segment count up to $N = 7$ (R^2=0.9456, RMSE=1.1053%, MAE=0.7394%, RPD=4.1886), which significantly outperforms all baselines. Notably, performance decreases at $N = 8$ (R^2=0.9033), suggesting an optimal segment count beyond which the factorial complexity becomes prohibitive.

Ablation Studies:

Table 3 presents ablation results for key components of SpecBPP. Removing curriculum learning (CL) produces the largest performance drop (R^2 decreases by 0.0435), confirming its crucial role in managing permutation complexity and demonstrating that this training strategy is a fundamental enabler rather than merely a convergence accelerator. The moderate impact of spectral band weighting (SW) removal (R^2 decreases to 0.9318) and minimal effect of dual attention (DA) removal (R^2=0.9402) indicate that while these components provide measurable benefits, SpecBPP's superior performance primarily derives from the novel spectral permutation prediction objective combined with curriculum learning.

5 Conclusion

We have presented Spectral Band Permutation Prediction (SpecBPP), a novel self-supervised framework that leverages the natural ordering of hyperspectral bands to learn globally coherent spectral representations. Our curriculum learning strategy overcomes the factorial complexity of segment ordering, yielding rich embeddings without manual labels. When fine-tuned on limited soil organic carbon samples, our approach achieves state-of-the-art performance (R^2 = 0.9456, RMSE = 1.1053%, RPD = 4.19), substantially surpassing masked autoencoders, contrastive methods, and traditional regressors. SpecBPP opens new avenues

for self-supervised learning in remote sensing by explicitly encoding spectral continuity and long-range dependencies.

5.1 Limitations and Future Work

Despite achieving state-of-the-art performance, SpecBPP has several limitations that present opportunities for future development. The factorial growth of permutation space ($N!$) imposes computational constraints, with accuracy decreasing from 100% at $N = 7$ to 84.2% at $N = 8$, though our curriculum learning strategy successfully manages this complexity within the effective operating range. Our uniform spectral segmentation approach, while ensuring domain-agnostic applicability, may inadvertently split meaningful absorption features across segment boundaries, suggesting that adaptive segmentation strategies could further improve performance.

Our validation focuses on EnMAP imagery to establish the fundamental effectiveness of spectral permutation prediction within a consistent sensor framework, avoiding confounding variables that would obscure the core methodological contribution. While this rigorous evaluation demonstrates the robustness of our approach, several extensions merit investigation. First, explicit wavelength-based positional encoding (analogous to transformer positional encodings) could enhance cross-sensor transfer by enabling interpolation of learned features across different spectral configurations. Second, ordinal regression formulations that exploit natural spectral ordering to penalize nearby position errors less than distant ones could improve sample efficiency, particularly for larger segment counts. Third, adaptive spectral segmentation aligned with physical absorption features, rather than uniform division, may prevent splitting meaningful spectral patterns across boundaries.

Acknowledgement: The authors would like to thank Abayomi Awobukun for funding this research work.

References

[1] Ryan Prescott Adams and Richard S Zemel. Ranking via sinkhorn propagation. *arXiv preprint arXiv:1106.1925*, 2011.

[2] Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15619–15629, 2023.

[3] Daniel La'ah Ayuba, Jean-Yves Guillemaut, Belen Marti-Cardona, and Oscar Mendez. Hyperkon: A self-supervised contrastive network for hyperspectral image analysis. *Remote Sensing*, 16(18), 2024. ISSN 2072-4292. URL https://www.mdpi.com/2072-4292/16/18/3399.

[4] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 41–48, 2009. doi: 10.1145/1553374.1553380.

[5] Yassine Bouslihim, Abderrazak Rochdi, Nicolas Baghdadi, Carole Sinfort, Amine El-Alem, Anass Chakir, and Said Lahssini. Using enmap satellite and machine learning for soil organic carbon assessment: A case study in semi-arid region. *Remote Sensing*, 16 (3):387, 2024.

[6] Xianghai Cao, Haifeng Lin, Shuaixu Guo, Tao Xiong, and Licheng Jiao. Transformer-based masked autoencoder with contrastive loss for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12, 2023.

[7] Xianghai Cao, Jiayu Yu, Ruijie Xu, Jiaxuan Wei, and Licheng Jiao. Mask-enhanced contrastive learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[8] Sabine Chabrillat, Saskia Foerster, Karl Segl, Alison Beamish, Maximilian Brell, Saeid Asadzadeh, Robert Milewski, Kathrin J Ward, Arlena Brosinsky, Katrin Koch, et al. The enmap spaceborne imaging spectroscopy mission: Initial scientific results two years after launch. *Remote Sensing of Environment*, 315:114379, 2024.

[9] Cheng-Wen Chang, David A. Laird, Maurice J. Mausbach, and Charles R. Hurburgh. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 65(2):480–490, 2001. doi: 10.2136/sssaj2001.652480x.

[10] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 2020.

[11] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15750–15758, 2021. doi: 10.1109/CVPR46437.2021.01549.

[12] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1251–1258, 2017. doi: 10.1109/CVPR.2017.195.

[13] Dristi Datta, Manoranjan Paul, Manzur Murshed, Shyh Wei Teng, and Leigh Schmidtke. Soil moisture, organic carbon, and nitrogen content prediction with hyperspectral data using regression models. *Sensors*, 22(20):7998, 2022.

[14] Aryan Deshwal, Syrine Belakaria, Janardhan Rao Doppa, and Dae Hyun Kim. Bayesian optimization over permutation spaces. In *Proceedings of the AAAI conference on artificial intelligence*, pages 6515–6523, 2022.

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1:4171–4186, 2019. doi: 10.18653/v1/N19-1423.

[16] Carl Doersch, Abhinav Gupta, and Alexei A. Efros. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1422–1430, 2015. doi: 10.1109/ICCV.2015.167.

[17] Basura Fernando, Hakan Bilen, Efstratios Gavves, and Stephen Gould. Self-supervised video representation learning with odd-one-out networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3636–3645, 2017.

[18] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3146–3154, 2019.

[19] A. Gholizadeh, M. Saberioon, L. Bořůvka, A. Wayayok, and M.A.M. Soom. Leaf chlorophyll and nitrogen dynamics and their relationship to lowland rice yield for site-specific paddy management. *Information Processing in Agriculture*, 4(4):259–268, 2017.

[20] Alexander F. H. Goetz. Three decades of hyperspectral remote sensing of the earth: A personal view. *Remote Sensing of Environment*, 113:S5–S16, 2009. doi: 10.1016/j.rse.2007.12.014.

[21] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Ávila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, 33:21271–21284, 2020.

[22] Peiyan Guan and Edmund Y Lam. Cross-domain contrastive learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

[23] Guy Hacohen and Daphna Weinshall. On the power of curriculum learning in training deep networks. In *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 2535–2544, 2019.

[24] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9729–9738, 2020. doi: 10.1109/CVPR42600.2020.00975.

[25] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16000–16009, 2022. doi: 10.1109/CVPR52688.2022.01553.

[26] Danfeng Hong, Lianru Gao, Jing Yao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. Graph convolutional networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):5966–5978, 2020.

[27] Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. doi: 10.1109/TGRS.2021.3130716.

[28] Sikang Hou, Hongye Shi, Xianghai Cao, Xiaohua Zhang, and Licheng Jiao. Hyperspectral imagery classification based on contrastive learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2021.

[29] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[30] Xiang Hu, Teng Li, Tong Zhou, Yu Liu, and Yuanxi Peng. Contrastive learning based on transformer for hyperspectral image classification. *Applied Sciences*, 11(18):8670, 2021.

[31] Lingbo Huang, Yushi Chen, and Xin He. Spectral–spatial masked transformer with supervised and contrastive learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–18, 2023.

[32] Weili Kong, Baisen Liu, Xiaojun Bi, Jiaming Pei, and Zheng Chen. Instructional mask autoencoder: A scalable learner for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:1348–1362, 2023.

[33] Rattan Lal. Soil carbon sequestration impacts on global climate change and food security. *Science*, 304(5677):1623–1627, 2004. doi: 10.1126/science.1097396.

[34] Ziyu Li, Zhaohui Xue, Mingming Jia, Xiangyu Nie, Hao Wu, Mengxue Zhang, and Hongjun Su. Demae: Diffusion enhanced masked autoencoder for hyperspectral image classification with few labeled samples. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[35] Weiwei Liu, Kai Liu, Weiwei Sun, Gang Yang, Kai Ren, Xiangchao Meng, and Jiangtao Peng. Self-supervised feature learning based on spectral masking for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023.

[36] Ishan Misra, C. Lawrence Zitnick, and Martial Hebert. Shuffle and learn: Unsupervised learning using temporal order verification. In *Computer Vision – ECCV 2016*, pages 527–544, 2016. doi: 10.1007/978-3-319-46448-0_32.

[37] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *Computer Vision – ECCV 2016*, pages 69–84, 2016. doi: 10.1007/978-3-319-46466-4_5.

[38] Mercedes E. Paoletti, Juan M. Haut, Javier Plaza, and Antonio Plaza. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:279–317, 2019. doi: 10.1016/j.isprsjprs.2019.09.006.

[39] Bonthu Sandeep Reddy and HR Shwetha. Integrating soil spectral library and prisma data to estimate soil organic carbon in crop lands. *IEEE Geoscience and Remote Sensing Letters*, 2024.

[40] Swalpa Kumar Roy, Suvojit Manna, Tiecheng Song, and Lorenzo Bruzzone. Attention-based adaptive spectral–spatial kernel resnet for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7831–7843, 2020.

[41] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7374–7383, 2019.

[42] Gary A. Shaw and Heidi-Marie K. Burke. Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14(1):3–28, 2003.

[43] Petru Soviany, Radu Tudor Ardei, Radu Tudor Ionescu, and Marius Leordeanu. Curriculum learning: A survey. *International Journal of Computer Vision*, 130:1526–1565, 2022. doi: 10.1007/s11263-022-01611-x.

[44] Bo Stenberg, Raphael A. Viscarra Rossel, Abdul M. Mouazen, and Johanna Wetterlind. Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 107: 163–215, 2010. doi: 10.1016/S0065-2113(10)07005-7.

[45] Antoine Stevens, Marco Nocita, Gergely Tóth, Luca Montanarella, and Bas van Wesemael. Prediction of soil organic carbon at the european scale by visible and near infrared reflectance spectroscopy. *PLOS ONE*, 8(6):e66409, 2013. doi: 10.1371/journal.pone.0066409.

[46] Tobias Storch, Hans-Peter Honold, Sabine Chabrillat, Martin Habermeyer, Paul Tucker, Maximilian Brell, Andreas Ohndorf, Katrin Wirth, Matthias Betz, Michael Kuchler, et al. The enmap imaging spectroscopy mission towards operations. *Remote Sensing of Environment*, 294:113632, 2023.

[47] Lei Sun, Guang Zhao, Yingqian Zheng, and Zebin Wu. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[48] Yue Wang, Ming Wen, Hailiang Zhang, Jinyu Sun, Qiong Yang, Zhimin Zhang, and Hongmei Lu. Hsimae: A unified masked autoencoder with large-scale pre-training for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[49] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19, 2018.

[50] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

[51] Suhua Zhang, Zhikui Chen, Dan Wang, and Z Jane Wang. Cross-domain few-shot contrastive learning for hyperspectral images classification. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.