SPECTRAL DEPENDENCE AND REFERENCE PROVENANCE IN HYPERSPECTRAL BENCHMARKS

Marvin Burges & Mounia El Baz

ESA, ESRIN, ϕ-Lab
Frascati
Italy

Marco Celesti & Jens Nieke

ESA, ESTEC
Noordwijk
The Netherlands

ABSTRACT

The Copernicus Hyperspectral Imaging Mission for the Environment (CHIME), one of the Copernicus Sentinel Expansion Missions, will provide routine hyperspectral observations over land and coastal areas through the Copernicus Programme. Its contiguous spectral measurements across the visible to shortwave-infrared domain (400–2500 nm) will enable systematic monitoring of a wide range of surface properties at global scale. These observations could also support pretraining of hyperspectral foundation models, with the potential to derive transferable representations for applications relevant to CHIME. However, assessing this potential requires benchmarks that test whether these models capture useful spectral information and generalize across locations. We therefore audit 63 benchmarks used by 15 foundation models and conduct matched spectral ablations with models trained without pretraining on three tasks. Only five benchmarks use measured references, none spanning the full spectral range of CHIME. The ablations show that spectral benefits vary between benchmarks: compared with a Sentinel-2-like nine-band reduction, the full hyperspectral input improves macro-mIoU by 1.0 points for land-cover mapping (NLCD) and 7.7 for forest-type mapping (BD-Forêt). On the mineral benchmark (OxHyperMinerals), additional spectral information also improves agreement with imagery-derived reference masks. Together with the scarcity of measured-reference benchmarks, this motivates evaluating spectral benefits against target-specific field or laboratory references. To support CHIME's core applications, future benchmarks should prioritize vegetation traits for agriculture, soil properties, and mineral composition for raw-material applications, moving beyond conventional Earth observation tasks with readily available labels.

***Index Terms*—** Hyperspectral imaging; foundation models; spectral ablation; reference provenance

1. INTRODUCTION

The upcoming Copernicus Hyperspectral Imaging Mission for the Environment (CHIME), one of the Copernicus Sentinel Expansion Missions, will provide routine hyperspectral observations over land and coastal areas through the Copernicus Programme. [1, 2]. Its applications in agriculture and vegetation monitoring, soil characterization, and raw-material exploration depend on information about surface properties and composition that detailed spectra can reveal [3]. Hyperspectral foundation models (FMs) could support these tasks. They are machine-learning models initially trained on large image collections to learn patterns in spectra and their spatial context. This pretraining provides a common starting point that can be adapted to different applications using task-specific reference data, rather than training a model from scratch for each task [4]. For CHIME, this could reduce the amount of labeled data needed to develop models for its diverse monitoring applications. Assessing this potential requires benchmarks that test whether these models capture useful spectral information and generalize across locations.

Current evaluations emphasize land-cover and crop-type mapping with readily available thematic labels [4, 5]. However, where broadband information suffices, strong performance on hyperspectral imagery (HSI) alone gives little indication of the added value of imaging spectroscopy over multispectral observations from missions such as Sentinel-2 [6]. Matched comparisons are therefore needed to establish how much the spectral information contributes to performance.

Other benchmarks use targets generated by algorithms applied to the evaluated HSI, such as mineral masks derived from EMIT products [7]. Evaluation then measures how faithfully a model reproduces the reference method, rather than independently testing the inferred material properties [8, 9]. These two concerns motivate examining *spectral dependence*—the performance change under spectral ablation—alongside *reference provenance*, the sources used to construct the targets. For global applications, evaluation must also test whether learned relationships transfer beyond the training locations through geographic holdouts [10].

Earlier studies examine spectral dependence through different controls. Frank, Varga, and Zell [11] compare spatial-only and spectral-only models. Leonardi et al. [12] adapt TerraMind [13] through nearest-band selection or spectral

response function (SRF) grouping. SpectralEarth-FM reports a larger EnMAP-branch advantage on BDForêt than on NLCD [5], although that comparison changes both the pretrained branch and the input representation.

We combine an audit of current evaluation with matched spectral ablations to identify gaps in benchmarking CHIME-relevant capabilities. Section 2 reviews 63 benchmarks used by 15 FMs, examining their targets, reference provenance, spectral coverage, and geographic holdouts. Sections 3 and 4 present controlled ablations on NLCD, BDForêt, and OxHyperMinerals, using matched training without pretraining and an additional pretrained-backbone comparison on the two EnMAP tasks. Section 5 draws on both analyses to recommend larger-scale, geographically diverse evaluation of vegetation traits, soil properties, and mineral composition against target-specific field or laboratory references.

2. AUDIT OF CURRENT EVALUATION

We focus on datasets used for downstream evaluation of 15 HSI-capable FMs, rather than their pretraining corpora. We searched OpenAlex titles and abstracts on 26 August 2026, pairing hyperspectral terms with foundation-model terminology and, in a sensitivity search, terms for large-scale pretraining, masked autoencoders, and cross-dataset models. We include hyperspectral datasets whose wavelength ranges overlap any part of 400–2500 nm. We exclude non-HSI, auxiliary robustness, and compression-only evaluations.

We record each reported model–dataset–task evaluation and group protocol variants, such as different label fractions, under a common benchmark identified by its standardized dataset version and task family. Counting each model–benchmark pair once gives 95 pairs across 63 benchmarks. We trace reference provenance through the model papers and dataset sources, marking assignments as inferred when the evidence supports an origin indirectly.

We classify references by how the target value or identity is established. *Measured* references provide quantitative field or laboratory measurements of a target property, such as soil organic carbon content. *Field label* references instead establish class or target identities through documented field identification, agricultural declarations, or controlled target deployment. Class identity and spatial boundaries can have different sources; EuroCrops belongs to *Field label* because its crop identities come from farmer declarations [14]. For references derived from Earth observation (EO), *Other EO* denotes targets established from observations other than the hyperspectral acquisition used as model input, such as separate satellite imagery or aerial photographs. *Same image* denotes targets generated by interpreting or processing that same hyperspectral acquisition, including through derived products, either alone or alongside other sources. We use *Undocumented* where the checked sources leave the origin unclear.

The table records spectral coverage relative to 400–2500 nm. *Full* reaches within 60 nm of both endpoints; otherwise, *Near* overlaps at least 90% of the range. Of the remaining inputs, *VNIR* ends at or below 1100 nm, while *Partial* extends further.

Most benchmarks use class labels or imagery-derived targets (Table 1). Only five use quantitative field or laboratory measurements (*Measured*). Four of these benchmarks use *VNIR* inputs, while PACE Lake Erie is classified as *Partial*. None therefore combines measured references with *Full* or *Near* coverage of CHIME's 400–2500 nm range. Their narrower spectral ranges may still be appropriate for the individual tasks.

We assess collection scope (Table 2) and, separately, the downstream split protocol for each model–benchmark pair. *Local* denotes one named study area; emphRegional denotes multiple locations within one country or contiguous region; and *Multi-region* denotes multiple countries, a continental collection, or separate broad regions. *Unknown* denotes insufficient source evidence. A geographic holdout requires evaluation on sites, regions, or geographic groups withheld from task-specific training. Note: Separating pixels within one scene is insufficient.

Collection scope differs across reference categories (Table 2): 16 of 17 Field label benchmarks are local, while all 18 Other EO collections are regional or multi-region. Only 8 of 95 model–benchmark pairs have a documented geographically grouped holdout in downstream evaluation. For 47 pairs, the reported protocols use no such holdout; for the remaining 40, the available descriptions are insufficient to determine whether one was used.

3. CONTROLLED SPECTRAL ABLATION

We test how much performance is retained under spectral reduction on two thematic segmentation tasks and one mineral-mapping task.

3.1. Tasks and input representations

NLCD and BDForêt provide land-cover and forest segmentation tasks with 15 and 12 classes, respectively [4]. Both use 202-channel EnMAP L2A surface reflectance with external thematic references (*Other EO*). NLCD labels derive from multispectral products and ancillary data, whereas BDForêt labels come from photo-interpreted aerial and color-infrared imagery. We retain the official train/validation/test splits of 10,125/1,350/2,025 patches for NLCD and 1,912/255/383 for BDForêt. All conditions use the same 128×128 patches on the 30 m EnMAP grid.

OxHyperMinerals [7] extends the comparison to multilabel mineral mapping using 285-channel EMIT radiance without atmospheric correction. Its overlapping goethite, hematite, and kaolinite masks come from EMIT mineral products derived from the evaluated observation (*Same image*). The official split contains 796/198/200 train/validation/test

Table 1. This audit lists 95 model–benchmark pairs from 63 HSI benchmarks reported by FMs. Rows follow task order within each model. Columns report dataset, task, bands, spectral coverage, and reference category. Italic reference categories with a dagger mark inferred reference origins.

Model	HSI Downstream Benchmarks				
	Dataset	**Task**	**Bands**	**Coverage**	**Reference**
HyperSIGMA [15]	HanChuan	Crop type	274	VNIR	*Field label*†
	HongHu	Crop type	270	VNIR	*Field label*†
	Houston	Land cover	144	VNIR	Undocumented
	Indian Pines	Crop type	224	Full	Field label
	Pavia University	Land cover	103	VNIR	Undocumented
	ZY1-02D Yellow River	Land cover	119	Full	*Field label*†
	AVIRIS TD	Target det.	189	Full	*Same image*†
	Bay Area	Change	224	Full	*Same image*†
	Cri AD	Anomaly det.	46	VNIR	*Same image*†
	Farmland	Change	155	Full	*Field label*†
	Hermiston	Change	154	Full	*Field label*†
	Mosaic	Target det.	229	Full	Field label
	Pavia AD	Anomaly det.	102	VNIR	*Same image*†
	Renourban	Target det.	–	Full	Undocumented
	Santa Barbara	Change	224	Full	*Same image*†
	Viareggio	Anomaly det.	127	VNIR	Field label
	Urban	Unmixing	162	Full	Same image
	Houston SR	Reconstruction	48	VNIR	Same image
	Washington DC Mall	Reconstruction	191	Near	Same image
SpectralEarth [4]	DESIS-CDL	Crop type	235	VNIR	Other EO
	EO1-CDL	Crop type	198	Full	Other EO
	EnMAP-BDForet	Tree species	202	Full	Other EO
	EnMAP-BNETD	Land cover	202	Full	Other EO
	EnMAP-CDL	Crop type	202	Full	Other EO
	EnMAP-CORINE	Land cover	202	Full	Other EO
	EnMAP-EuroCrops	Crop type	202	Full	Field label
	EnMAP-NLCD	Land cover	202	Full	Other EO
	EnMAP-TreeMap	Tree species	202	Full	Other EO
	HyBiomass	Biomass	202	Full	Other EO
	Hyperview	Soil property	150	VNIR	Measured
SpectralEarth-FM [5]	DESIS-CDL	Crop type	235	VNIR	Other EO
	EO1-CDL	Crop type	198	Full	Other EO
	EnMAP-BDForet	Tree species	202	Full	Other EO
	EnMAP-BNETD	Land cover	202	Full	Other EO
	EnMAP-CDL	Crop type	202	Full	Other EO
	EnMAP-EuroCrops	Crop type	202	Full	Field label
	EnMAP-NLCD	Land cover	202	Full	Other EO
	EnMAP-TreeMap	Tree species	202	Full	Other EO
	GF-5 Wuhan	Land cover	116	Full	Same image
	OxHyperMinerals	Minerals	285	Full	Same image
SpecAware [16]	AeroRIT	Land cover	372	VNIR	Same image
	HRSSC	Land cover	224	Full	*Same image*†
	Qingpu-HSI	Land cover	251	VNIR	Field label
	WHU-H2SR	Land cover	249	VNIR	Field label
	Bay Area	Change	224	Full	*Same image*†
	Santa Barbara	Change	224	Full	*Same image*†
HSIGene [17]	WHU-Hi-HanChuan	Reconstruction	274	VNIR	Same image
	Xiongan	Reconstruction	250	VNIR	Same image
HyperFM [18]	PACE OCI	Cloud property	291	Partial	Same image
SpecTM [19]	PACE Lake Erie	Water quality	122	Partial	Measured

Model	HSI Downstream Benchmarks				
	Dataset	**Task**	**Bands**	**Coverage**	**Reference**
HyperFree [20]	HanChuan	Crop type	274	VNIR	*Field label*†
	HongHu	Crop type	270	VNIR	*Field label*†
	WHU-Hi HongHu	Land cover	270	VNIR	*Field label*†
	WHU-Hi LongKou	Crop type	270	VNIR	*Field label*†
	XiongAn	Land cover	256	VNIR	Field label
	Airport	Target det.	205	Full	*Same image*†
	Beach1	Anomaly det.	188	VNIR	*Same image*†
	Beach2	Anomaly det.	193	VNIR	*Same image*†
	Cri target detection	Target det.	46	VNIR	*Same image*†
	HOTC 2023	Tracking	56	VNIR	*Same image*†
	Hermiston	Change	154	Full	*Field label*†
	River	Change	198	Full	*Field label*†
	Urban	Unmixing	162	Full	Same image
	Washington DC Mall	Reconstruction	191	Near	Same image
HySens [21]	HanChuan	Crop type	274	VNIR	*Field label*†
	HongHu	Crop type	270	VNIR	*Field label*†
	Indian Pines	Crop type	224	Full	Field label
	Pavia University	Land cover	103	VNIR	Undocumented
	Bay Area	Change	224	Full	*Same image*†
	Hermiston	Change	154	Full	*Field label*†
	Santa Barbara	Change	224	Full	*Same image*†
	Bavaria-SOC	Soil property	157	VNIR	Measured
	GLORIA	Water quality	551	VNIR	Measured
	KarLy	Soil property	125	VNIR	Measured
ChronoEarth [22]	CDL	Crop type	155	Near	Other EO
	CLCD	Land cover	155	Near	Other EO
	CORINE	Land cover	155	Near	Other EO
	ISDASoil	Soil texture	155	Near	Other EO
	NLCD-S	Land cover	155	Near	Other EO
	GFC	Change	155	Near	Other EO
HyperSL [23]	Botswana	Land cover	145	Near	Field label
	Chikusei	Land cover	128	VNIR	Field label
	Indian Pines	Crop type	224	Full	Field label
	Bay Area	Change	224	Full	*Same image*†
	Santa Barbara	Change	224	Full	*Same image*†
LESSViT [24]	EnMAP-BDForet	Tree species	202	Full	Other EO
	EnMAP-BNETD	Land cover	202	Full	Other EO
	EnMAP-CDL	Crop type	202	Full	Other EO
	EnMAP-CORINE	Land cover	202	Full	Other EO
	EnMAP-EuroCrops	Crop type	202	Full	Field label
HyperKD [25]	EnMAP-CDL	Land cover	218	Full	Other EO
	EnMAP-NLCD	Land cover	218	Full	Other EO
	EnMAP-gNATSGO	Soil property	218	Full	Other EO
HyperspectralMAE [26]	Indian Pines	Crop type	224	Full	Field label
Panopticon [27]	EnMAP-CORINE	Land cover	202	Full	Other EO

Table 2. Geographic scope by reference provenance for 63 benchmark collections. Unknown denotes insufficient source evidence.

Reference	Local	Regional	Multi-region	Unknown
Measured	2	2	1	0
Field label	16	0	1	0
Other EO	0	13	5	0
Same image	13	1	2	4
Undocumented	2	0	0	1
Total	33	16	9	5

parent tiles. Each 512×512 tile supplies 16 fixed, non-overlapping 128×128 windows that inherit its split.

We compare five representations to examine the effects of spectral coverage, broad-band integration, and compression. Full HSI retains all supplied channels. S2-9 integrates the input with the version 4.0 Sentinel-2A responses B2–B8, B8A, and B12 [28]. For input channel value x_i, the reduced band is

$$z_j = \frac{\sum_i x_i R_j(\lambda_i) w_i}{\sum_i R_j(\lambda_i) w_i}, \tag{1}$$

where R_j is the dimensionless relative response of band j, λ_i is the listed channel center in nm, and w_i is its quadrature width in nm. We calculate widths within contiguous spectral segments, without interpolation across gaps in the supplied 202-band EnMAP set. B11 overlaps its 1565–1653 nm gap, so we omit B11 and retain the same nine-band definition for EMIT. S2-9 retains each task's original observations and spatial grid, so its comparison with Full HSI measures spectral reduction rather than operational differences between sensors.

Narrow-9 replaces broad integration with the native channel nearest each response center. This compares native-band selection with broad-band integration at matched channel count and approximately matched spectral locations. RGB-3 uses the exact B4, B3, and B2 rows of S2-9, restricting the input to visible bands and reducing the channel count to three.

PCA-9 compresses the full spectrum into nine components, allowing comparison with the nine-band reductions at the same input dimensionality. We fit one IncrementalPCA basis per dataset using valid training pixels, with centering but no per-band variance scaling. Component selection follows input variance rather than task relevance, and bands with larger variance can therefore have greater influence on the basis. Unlike the band-based reductions, these components combine information from all supplied channels.

3.2. Experimental setup and evaluation

We train models to predict pixel-level labels from each input representation. We compare how spectral reduction changes performance in two learning settings: training without pretraining and adaptation of a frozen pretrained backbone.

Training without pretraining. For each task and input representation, we train a separate FCN-ResNet50 segmentation model from random initialization, updating all weights using that task's labeled training set.

Frozen pretrained backbone. On NLCD and BDForêt, we additionally compare Full HSI and S2-9 using the SpectralEarth `spec_rn50` backbone pretrained with DINO [4]. The backbone extracts features for prediction, and its pretrained weights remain fixed during task-specific training. We train only a new input adapter and segmentation head, which respectively map the spectral input into the backbone and its output to pixel-level labels. Because the original adapter cannot accept nine channels, we replace it in both input representations. This tests whether the spectral-dependence pattern persists with a pretrained backbone.

Implementation details. We train for 100 epochs with AdamW, an initial learning rate of 10^{-3}, cosine decay, and weight decay of 10^{-4}. Effective batch sizes are 64, 32, and 16 for NLCD, BDForêt, and OxHyperMinerals, respectively. We apply rotations and flips and calculate normalization statistics from training pixels. Input adapters use 1×1 convolutions to 64 channels without pretraining and to 128 channels followed by batch normalization and ReLU with the pretrained backbone. For the latter, backbone weights and batch-normalization statistics remain fixed, and final-stage downsampling is removed. The EnMAP tasks use multiclass cross-entropy with argmax predictions; OxHyperMinerals uses binary cross-entropy with a 0.5 sigmoid threshold.

Shared protocol and evaluation. Within each task and learning setting, samples, masks, labels, splits, augmentation, and training settings remain fixed across input representations. We repeat each experiment with three random seeds (0–2) and select the checkpoint with the highest validation macro mean intersection over union (mIoU).

We report both test scores and paired performance differences as mean $\pm$ sample standard deviation (SD) across the three seeds. Paired differences are calculated by subtracting scores within each seed before computing their mean and SD. For example, Full$-$S2-9 is the Full-HSI score minus the S2-9 score, expressed in percentage points. Positive differences favor the first-named input representation.

4. RESULTS AND INTERPRETATION

Reducing Full HSI to S2-9 causes a larger performance loss on BDForêt than on NLCD (Table 3). For models trained without pretraining, the paired Full$-$S2-9 differences are 1.0 ± 0.1 percentage points on NLCD and 7.7 ± 0.9 on BDForêt. With the frozen pretrained backbone, the corresponding differences are 3.1 ± 0.1 and 9.5 ± 0.4 points. Thus, BDForêt shows greater spectral dependence under this reduction in both learning settings.

The remaining comparisons use models trained without pretraining. Narrow-9 gives nearly identical mean perfor-

Table 3. Test macro mIoU (%) and paired differences (percentage points), reported as mean ± sample SD over three seeds. OHM denotes OxHyperMinerals.

Input / difference	NLCD	BDForêt	OHM
No pretraining			
Full HSI	41.3 ± 0.2	65.7 ± 0.3	61.9 ± 2.0
S2-9	40.3 ± 0.2	58.0 ± 0.6	54.2 ± 2.0
Narrow-9	40.2 ± 0.8	58.0 ± 0.5	53.4 ± 0.3
RGB-3	34.6 ± 1.4	34.7 ± 1.4	45.2 ± 1.0
PCA-9	41.3 ± 0.2	60.9 ± 0.4	59.7 ± 1.0
Full−S2-9	1.0 ± 0.1	7.7 ± 0.9	7.7 ± 2.6
Narrow−S2-9	-0.1 ± 0.7	0.0 ± 0.8	-0.9 ± 2.3
Full−RGB-3	6.7 ± 1.2	31.0 ± 1.1	16.7 ± 1.0
Full−PCA-9	0.0 ± 0.1	4.8 ± 0.3	2.2 ± 2.9
Frozen pretrained backbone			
Full HSI	37.7 ± 0.1	57.6 ± 0.6	–
S2-9	34.6 ± 0.1	48.1 ± 0.3	–
Full−S2-9	3.1 ± 0.1	9.5 ± 0.4	–

mance to S2-9 on both EnMAP tasks, so selecting native channels at these nine spectral locations offers little observed advantage over broad-band integration. This differs from the nearest-band advantage reported for TerraMind [12] in a different learning setting. Restricting the input to three visible bands with RGB-3 incurs much larger losses: 6.7 ± 1.2 points relative to Full HSI on NLCD and 31.0 ± 1.1 on BDForêt.

PCA-9 nearly matches Full HSI on NLCD, with a paired Full−PCA-9 difference of 0.0 ± 0.1 points. On BDForêt, it recovers part of the Full-HSI advantage over S2-9 but remains 4.8 ± 0.3 points below Full HSI. The different performance of these nine-dimensional representations shows that their construction matters, not just their size. The remaining difference on BDForêt motivates testing whether task-adapted representations from hyperspectral FMs can preserve predictive information more effectively than variance-based PCA at comparable dimensionality.

On OxHyperMinerals, Full HSI exceeds S2-9 by 7.7 ± 2.6 points and RGB-3 by 16.7 ± 1.0 points. PCA-9 retains more performance on average than the band-based reductions, with a paired Full−PCA-9 difference of 2.2 ± 2.9 points. The variability of this difference is substantial relative to its mean.

The meaning of these gains depends on the reference targets. NLCD and BDForêt evaluate agreement with external thematic maps, whereas OxHyperMinerals evaluates agreement with mineral masks derived from the same EMIT observation. Differences in classes, geography, and sample composition prevent isolating the contribution of reference provenance to the observed performance differences. Nevertheless, the mineral experiment shows that additional spectral information can improve reproduction of an existing retrieval product. This motivates complementary evaluation against field or laboratory references that establish material identity independently of the input imagery.

5. DISCUSSION AND RECOMMENDATIONS

The audit identifies an underrepresented combination of target-specific ground references and broad spectral coverage. None of the reviewed measured-reference benchmarks spans nearly all of 400–2500 nm. Geographic coverage alone does not establish performance beyond training locations. Our ablations retain the official dataset splits and test spectral dependence rather than compare geographic split protocols. They show that additional spectral information can improve agreement with imagery-derived products. Together, these findings motivate independent evaluation of target properties and testing whether spectral benefits persist at sites or regions excluded from task-specific training.

Future benchmarks should prioritize vegetation traits for agriculture and vegetation monitoring, soil properties such as soil organic carbon, and mineral identity and composition for raw-material applications. These tasks should combine physical measurements with ground-established species and mineral classes, extending evaluation beyond conventional thematic mapping. Reference records should document origin, date, location, measurement or identification method, processing history, and represented ground area. Evaluation should combine site or region holdouts with matched spectral controls, keeping learners, splits, masks, and training fixed across input conditions and reporting paired performance differences across seeds.

CaRMA [29], SHIFT [30], ABoVE AVIRIS [31], or GEMx [32] could provide candidate airborne/in-situ observations for such a collection. Development should begin by identifying target-specific field or laboratory references, distinguishing these from sensor-calibration records, and matching them to imagery in space and time. Where references require aggregation, the procedure should explain how the measurements represent each pixel's ground area. Compatible observations should then be harmonized through spectral response integration, spatial blur, and resampling to a common CHIME-like 30 m grid within their available wavelength support. The multisensor framework of Koutsantoni et al. [33] and CHEES [34] provide related simulation methods.

A benchmark built around these priorities should test how well hyperspectral foundation models predict surface properties at unseen locations and how much spectral information contributes to that performance. CHIME's application needs, rather than label availability alone, should guide the selection of tasks and references.

Acknowledgments

We used GPT-6 Astra, GPT-6 Pro, and Opus 5 to assist with manuscript drafting and revision, code implementation, and preparation of the initial benchmark audit. Every audit entry was manually checked, and the authors take responsibility for the final manuscript, code, and results.

6. REFERENCES

[1] Jens Nieke, L. Despoisse, Antonio Gabriele, Heidrun Weber, Helene Strese, Nafiseh Ghasemi, Ferran Gascon, Kevin Alonso, Valentina Boccia, Bogdana Tsonevska, Pierre Choukroun, Giuseppe Ottavianelli, and Marco Celesti, "The copernicus hyperspectral imaging mission for the environment (chime): an overview of its mission, system and planning status," 2023.

[2] Michael Rast, Jennifer Adams, Ferran Gascon, Jens Nieke, Cristina Ananasso, Heike Bach, Eyal Ben Dor, Sabine Chabrillat, Roberto Colombo, Jean-Baptiste Feret, Claudia Giardino, Robert O. Green, Luis Guanter, Stuart Marsh, Cindy Ong, Giovanni Rum, Michael Schaepman, Martin Schlerf, Martin Schodlok, Andrew K. Skidmore, Tobias Storch, Peter Strobl, and Mathew Williams, "Copernicus Hyperspectral Imaging Mission for the Environment - Mission Requirements Document," Mission Requirements Document ESA-EOPSM-CHIM-MRD-3216, European Space Agency, 2021, Issue 3.0, 21 January 2021.

[3] Michael E. Schaepman, Susan L. Ustin, Antonio Plaza, T. H. Painter, Jochem Verrelst, and Shunlin Liang, "Earth system science related imaging spectroscopy—an assessment," *Remote Sensing of Environment*, vol. 113, pp. S123–S137, 2009.

[4] Nassim Ait Ali Braham, Conrad M. Albrecht, Julien Mairal, Jocelyn Chanussot, Yi Wang, and Xiao Xiang Zhu, "SpectralEarth: Training hyperspectral foundation models at scale," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 16780–16797, 2025.

[5] Nassim Ait Ali Braham, Aaron Banze, Conrad M. Albrecht, Julien Mairal, Jocelyn Chanussot, and Xiao Xiang Zhu, "SpectralEarth-FM: Bringing hyperspectral imagery into multimodal earth observation pretraining," *arXiv preprint arXiv:2605.21075*, vol. abs/2605.21075, 2026.

[6] European Space Agency, "Sentinel-2 Mission," SentiWiki. `https://sentiwiki.copernicus.eu/web/s2-mission`, Accessed: September 11, 2026.

[7] Vít Ruzicka and Andrew Markham, "Hyperspectralvits: General hyperspectral models for on-board remote sensing," *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, vol. 18, pp. 10241–10253, 2025.

[8] Stephen V. Stehman and Giles M. Foody, "Key issues in rigorous accuracy assessment of land cover products," *Remote Sensing of Environment*, vol. 231, pp. 111199, 2019.

[9] Pontus Olofsson, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder, "Good practices for estimating area and assessing accuracy of land change," *Remote Sensing of Environment*, vol. 148, pp. 42–57, 2014.

[10] Jakub Nalepa, Michal Myller, and Michal Kawulok, "Validating hyperspectral image segmentation," *IEEE Geosci. Remote. Sens. Lett.*, vol. 16, no. 8, pp. 1264–1268, 2019.

[11] Hannah Frank, Leon Amadeus Varga, and Andreas Zell, "Hyperspectral benchmark: Bridging the gap between HSI applications through comprehensive dataset and pretraining," *CoRR*, vol. abs/2309.11122, 2023.

[12] Julia Anna Leonardi, Johannes Jakubik, Paolo Fraccaro, and Maria Antonia Brovelli, "Spectral gaps and spatial priors: Studying hyperspectral downstream adaptation using TerraMind," *arXiv preprint arXiv:2603.06690v2*, 2026, Version 2, 24 March 2026.

[13] Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, Rahul Ramachandran, Paolo Fraccaro, Thomas Brunschwiler, Gabriele Cavallaro, Juan Bernabe-Moreno, and Nicolas Longépé, "TerraMind: Large-scale generative multimodality for earth observation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2025, pp. 7383–7394.

[14] Maja Schneider, Tobias Schelte, Felix Schmitz, and Marco Körner, "EuroCrops: The largest harmonized open crop dataset across the european union," *Scientific Data*, vol. 10, no. 1, pp. 612, 2023.

[15] Di Wang et al., "HyperSIGMA: Hyperspectral intelligence comprehension foundation model," *arXiv preprint arXiv:2406.11519*, 2024.

[16] Renjie Ji, Xue Wang, Chao Niu, Wen Zhang, Yong Mei, and Kun Tan, "SpecAware: A spectral-content aware foundation model for unifying multi-sensor learning in hyperspectral remote sensing mapping," *arXiv preprint arXiv:2510.27219*, vol. abs/2510.27219, 2025.

[17] Li Pang et al., "HSIGene: A foundation model for hyperspectral image generation," *arXiv preprint arXiv:2409.12470*, 2024.

[18] Zahid Hassan Tushar and Sanjay Purushotham, "HyperFM: An efficient hyperspectral foundation model with spectral grouping," *arXiv preprint arXiv:2604.21127*, vol. abs/2604.21127, pp. 6955–6969, 2026.

[19] Syed Usama Imtiaz, Mitra Nasr Azadani, and Nasrin Alamdari, "SpecTM: Spectral targeted masking for trustworthy foundation models," *arXiv preprint arXiv:2603.22097*, 2026.

[20] Jingtao Li et al., "HyperFree: A channel-adaptive and tuning-free foundation model for hyperspectral remote sensing imagery," *arXiv preprint arXiv:2503.21841*, 2025.

[21] Xiangyu Zhao, Zhitong Xiong, and Xiao Xiang Zhu, "HySens: Sensor-agnostic foundation models for hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 64, pp. 5513915, 2026.

[22] Haozhe Si, Yuxuan Wan, Yuqing Wang, Minh Do, and Han Zhao, "ChronoEarth-492K: A large scale and long horizon spatiotemporal hyperspectral earth observation dataset and benchmark," *arXiv preprint arXiv:2605.15666*, vol. abs/2605.15666, 2026.

[23] Weili Kong, Baisen Liu, Xiaojun Bi, Changdong Yu, Xinyao Li, and Yushi Chen, "HyperSL: A spectral foundation model for hyperspectral image interpretation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–19, 2025.

[24] Haozhe Si, Yuxuan Wan, Yuqing Wang, Minh Do, and Han Zhao, "LESSViT: Robust hyperspectral representation learning under spectral configuration shift," *arXiv preprint arXiv:2605.18541*, 2026.

[25] Abdul Matin, Tanjim Bin Faruk, Shrideep Pallickara, and Sangmi Lee Pallickara, "HyperKD: Distilling cross-spectral knowledge in masked autoencoders via inverse domain shift with spatial-aware masking and specialized loss," in *2025 IEEE 12th International Conference on Data Science and Advanced Analytics (DSAA)*, 2025, pp. 1–11.

[26] Wooyoung Jeong, Hyun Jae Park, Seonghun Jeong, Jong Wook Jang, Tae Hoon Lim, and Dae Seoung Kim, "HyperspectralMAE: The hyperspectral imagery classification model using fourier-encoded dual-branch masked autoencoder," *arXiv preprint arXiv:2505.05710*, 2025.

[27] Leonard Waldmann et al., "Panopticon: Advancing any-sensor foundation models for earth observation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2025, pp. 2195–2205.

[28] European Space Agency, "Sentinel-2 spectral response functions 2024," Technical Note COPE-GSEG-EOPG-TN-15-0007, European Space Agency, June 2024, Issue 4.0, Microsoft Excel workbook.

[29] David Marshall Ingram, Tobias Eckert, Uta Heiden, Claas Henning Köhler, Kevin Kühl, Raquel de los Reyes, Gabriel Arden Scheib, Konstantin Torgasin, Belen Franch, Cesar Jose Guerrero Benavent, and Italo Moletto-Lobos, "Surface reflectance validation for multi- and hyperspectral air- and spaceborne sensors using in-situ data from the CARMA campaign in Puch, Germany," Presentation at VH-RODA 2025, 2025.

[30] K. Dana Chadwick et al., "Unlocking ecological insights from sub-seasonal visible-to-shortwave infrared imaging spectroscopy: The SHIFT campaign," *Ecosphere*, vol. 16, no. 3, pp. e70194, 2025.

[31] Charles E. Miller et al., "Airborne imaging spectroscopy surveys of Arctic and boreal Alaska and northwestern Canada 2017–2023," *Scientific Data*, vol. 12, pp. 692, 2025.

[32] James V. Jones, III, Tanya J. Gallegos, Mojisola A. KunleDare, and Charlotte E. Riggs, "USGS 2025 critical minerals review," *Mining Engineering*, vol. 78, no. 5, pp. 42–57, 2026.

[33] Pelagia Koutsantoni, Maria Kremezi, Vassilia Karathanassi, Paola Di Lauro, José Andrés Vargas-Solano, Giulio Ceriola, Antonello Aiello, and Elisabetta Lamboglia, "A multisensor framework for satellite data simulation: Generating representative datasets for future ESA missions—CHIME and LSTM," *Remote Sensing*, vol. 18, no. 9, 2026.

[34] Fernando E. Alemán, Laura Zamora, Claudia Isola, Carolina Tenjo, Karl Segl, Stéphane Guillaso, José Luis García, Jens Nieke, Antonio Gabriele, Adrian Garcia, Ignacio Fernandez Nunez, Miguel Morata, Jochem Verrelst, Adrián Jacinto, Cinzia Panigada, Massimo Musacchio, Stefano Pignatti, Alessandra Nguyen, Andrea Taramelli, Emiliana Valentini, and Giulia Tagliabue, "CHIME's end to end simulator (CHEES): simulating, testing and generating hyperspectral products," in *Sensors, Systems, and Next-Generation Satellites XXIX*, Sachidananda R. Babu, Arnaud Hélière, and Tomoyuki Urabe, Eds. 2025, vol. 13667 of *Proceedings of SPIE*, p. 136670G, SPIE.