AnySpec: Response-Conditioned Spectral Reconstruction Across Sensors

Liam Chalcroft, Eduardo Camarillo Abad, Peter Christopher, Thomas Albrow-Owen, Oliver Burton
Prospectral, Cambridge, UK
{liam, eduardo, peter, tom, oli}@prospectral.tech

***Abstract*—Spectral reconstruction estimates a dense spectrum from a few broad-band measurements, such as camera RGB or satellite multispectral bands. Trained reconstructors are usually tied to one sensor: the input layer expects a fixed set of channels, and the weights absorb the spectral response functions used to render the training data. We study AnySpec, a reconstructor that takes the response functions as an input. Each measured channel becomes a token pairing its value with an embedding of its response curve, so within a dataset's spectral grid one set of weights serves sensors with different responses and channel counts, and a classical consistency projection makes the output reproduce the measurement. On ARAD-1K, AnySpec reaches a mean relative absolute error (MRAE) of 0.0883 ± 0.0013 on the training camera and 0.0879 ± 0.0008 on 8 camera responses held out of training, where a model trained on one fixed response reaches 0.3359 ± 0.2794. Used as the input stem of the MST++ spatial network, the same encoder lowers held-out MRAE of the network output from 0.0837 to 0.0767, with lower error on 8 of 8 cameras; after the consistency projection the two are close. On HySpecNet-11k satellite data, a projected linear minimum mean-square error (LMMSE) estimator is slightly more accurate on clean measurements and AnySpec is more accurate under added noise. On two satellite instrument families absent from training, the clean-data gap to LMMSE widens within the measured spectral range, while AnySpec stays ahead on noisy MODIS data.**

***Index Terms*—spectral reconstruction, sensor-conditioned models, measurement consistency, hyperspectral, EnMAP, LMMSE**

I. Introduction

Hyperspectral images resolve narrow bands for material identification, vegetation and water analysis, and colour-accurate imaging. Their sensors can be costly, slow or spatially coarse; spectral reconstruction instead estimates dense spectra from inexpensive RGB or multispectral sensors [1], [2].

A spectral reconstructor is usually tied to its camera. The input layer takes a fixed number of channels in a fixed order, the training data are rendered through one table of spectral response functions (SRFs), and the weights encode that table. A change in sensor response can therefore degrade reconstruction accuracy, while each sensor leaves a different null space of unmeasured spectral variation (Fig. 1). Sensor-conditioned representations address this by describing the instrument to the network: STARS encodes sensor metadata into a universal spectral representation [3], and a diffusion model can swap its responsivity matrix without retraining at the cost of iterative sampling [4].

We combine response-conditioned representations [3] and classical measurement-consistency constraints [5], [6] in a single-pass reconstructor shared across sensors. Tokens pairing measurements with responses accept variable channel counts, and a projection reproduces the measurement (Fig. 2). A companion paper analyses what a measurement determines and what the projection can repair [7].

Response tokens and the consistency projection both exist in prior work; our contribution is to combine them in one single-pass reconstructor with weights shared across sensors and to characterise it empirically: (i) transfer to camera responses held out of training on ARAD-1K; (ii) the same encoder as the input stem of a spatial backbone, the multi-stage spectral-wise transformer MST++ [8]; and (iii) satellite instruments on HySpecNet-11k against a linear minimum mean-square error (LMMSE) estimator projected onto the same output constraint.

II. Related Work

The NTIRE 2022 spectral reconstruction challenge and its ARAD-1K dataset of 1,000 31-band scenes (400–700 nm) fixed the protocol most RGB-to-spectrum work reports [2]; MST++ is a widely used architecture on that track [8]. Those methods condition on one camera. Lin and Finlayson require a recovered spectrum to re-integrate to the measured colour [5], and Matrix-R theory replaces the fundamental metamer [9] of any reconstruction with the one fixed by the camera, which cannot increase squared spectral error for a noiseless measurement [10]; PhySpec builds the same range-null split into a fixed RGB forward model [11], as does compressive imaging [12]. We take the operator as an input rather than a property of one camera. SenPa-MAE, Panopticon and CARL write responses into embeddings for analysis [13], [14], [15]; we reconstruct, and the measured component is reproduced by a classical projection.

III. Method

Problem. Let $s \in \mathbb{R}^L$ be a spectrum sampled on the dataset's L-band wavelength grid and $R \in \mathbb{R}^{C\times L}$ the sensor's response matrix, whose row c is the SRF of channel c. The sensor records $y = Rs+\epsilon$, where ϵ is measurement noise, and the reconstructor receives y and R and returns an estimate $\hat{s}$. Because $C \ll L$, many spectra produce the same y: in the noiseless case they form the affine set $\{s : Rs = y\}$, whose directions make up the null space $\ker R$ of spectral changes the sensor cannot see; R^+ is the Moore–Penrose

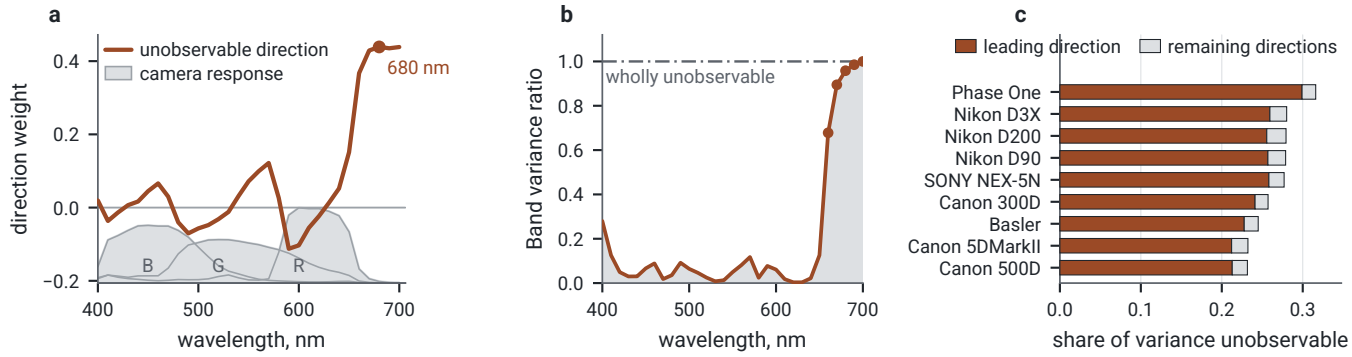


Fig. 1. Spectral covariance in the measurement null space, estimated from ARAD-1K training spectra. (a) Leading eigenvector of $Q\Sigma Q$ for the Basler response (sign arbitrary), with the three responses as silhouettes. (b) Per-band ratio $\mathrm{diag}(Q\Sigma Q)_j/\mathrm{diag}(\Sigma)_j$. (c) Fraction $\mathrm{tr}(Q\Sigma Q)/\mathrm{tr}(\Sigma)$ per camera, leading mode and remainder.

pseudoinverse. Each SRF is discretised onto the grid by trapezoid quadrature and normalised to unit mass, so each entry of y is a weighted average of s and rows from different instruments are comparable. A channel whose support falls entirely in an unsampled grid hole is dropped.

Architecture. The evaluated model is a pre-norm transformer of width 128, with 4 heads, 2 encoder layers and 2 decoder layers (0.96 M parameters). Each channel token concatenates the measurement with an embedding of its response row: real and imaginary parts of the row's characteristic function at 32 log-spaced frequencies, on coordinates $(\lambda - 350)/2150$ relative to a 350–2500 nm span, capped at one cycle per 40 nm. A decoder cross-attends one query per output wavelength; query coordinates use the same normalisation with 16 log-spaced frequencies from 1 to 430 cycles over that span. A linear head emits a raw spectrum f.

Consistency projection. With $S = \mathrm{row}(R)$ and $\Pi_S = R^+R$, the network's raw spectrum f is mapped to $\hat{s} = R^+y + (I - \Pi_S)f$, which satisfies $R\hat{s} = y$ in the clean undamped case. Training adds a relative Gram floor of 10^{-6} times the mean diagonal of $RR^\top$, and evaluation is undamped. An auxiliary loss, the relative mean squared measurement residual $\|Rf - y\|^2/\|y\|^2$ with weight 1, supervises the raw components the projection overwrites; in an ablation without it, validation MRAE rose from 0.0887 to 0.0996.

Null-space spectral variation. Let Σ be the covariance of the 3,596,000 training spectra and $Q = I - \Pi_S$. For the Basler response, 24% of the spectral variance lies in $\ker R$, and the leading eigenvector of $Q\Sigma Q$ carries 93% of it, peaking at 680 nm where the three responses have fallen away (Fig. 1). Across the nine camera responses the fraction ranges from 23% to 32% (Phase One). This variation is not fixed by the measurement, though correlation with the measured components can still make it predictable.

Training. Camera and satellite checkpoints are trained separately for 30 000 AdamW steps (learning rate 0.002, weight decay 0.01, 500-step warmup, unit gradient clip) and the last step is kept. The camera runs use batch 4096; the satellite

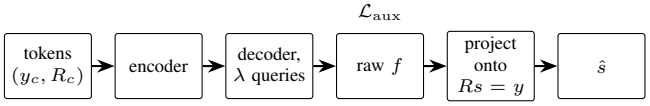


Fig. 2. AnySpec architecture. Channel measurements and response embeddings are encoded as a set of tokens. Wavelength queries decode the raw spectral estimate f, which is projected onto the measurement constraint to give $\hat{s}$. An auxiliary loss $\mathcal{L}_{\mathrm{aux}}$ supervises the measurement consistency of f.

runs use 2048. Response-conditioned runs use full precision. Each conditioned training step draws a published sensor or, with probability 0.5, a procedural sensor whose randomly generated response curves vary in number, shape, width and position. All per-pixel runs add Gaussian noise with standard deviation σ uniform on $[0, 0.01]$. The conditioned objective is MRAE plus mean absolute error on the corrected spectrum, plus the auxiliary term above. Exposure augmentation, where used, feeds the network gy with g log-uniform on $[0.25, 4]$ but keeps the unscaled y for the projection and auxiliary loss; unit-gain training fixes $g = 1$.

Classical comparators. Mean-consistent projection, $\hat{s} = m + R^+(y - Rm)$ with m the training-bank mean, is the minimal measurement-consistent reference. Projected LMMSE uses the training-set mean μ and covariance Σ: the LMMSE estimate (the Gaussian-model posterior mean [6]) is projected onto the same y,

$$\hat{s}_{\mathrm{L}} = R^+y + Q\mu_y, \quad \mu_y = \mu + \Sigma R^\top (R\Sigma R^\top + \sigma^2 I)^+ (y - R\mu). \tag{1}$$

IV. Experiments

The mean relative absolute error (MRAE) is the mean of $|\mathrm{pred} - \mathrm{gt}|/\mathrm{gt}$ over entries with $\mathrm{gt} > 0$; we also report peak signal-to-noise ratio (PSNR, dB) and spectral angle mapper (SAM, degrees). HySpecNet scores use valid pixels only and weight each patch equally.

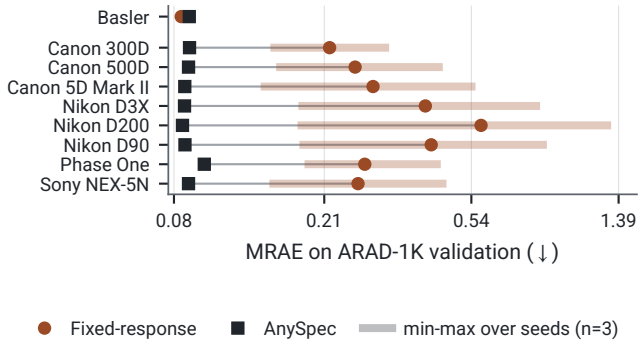


Fig. 3. Camera transfer on ARAD-1K validation, one row per camera. Basler (top) is seen in training; the 8 cameras below are held out. Circles: fixed-response AnySpec. Squares: AnySpec. Bars span the minimum to maximum over training seeds, in each model's colour.

A. Camera transfer on ARAD-1K

We render measurements from ARAD-1K cubes through public response tables. These simulated measurements differ from the challenge's fixed-camera JPEG track, so published challenge scores are not comparable. We retrain the NTIRE 2022 winning architecture, MST++, under this cross-sensor protocol (Sec. IV-B); the splits, response tables and scripts will be released. Both models train on per-pixel spectra and are scored on the same validation cubes at a clean noiseless render. Fixed-response AnySpec sees only the Basler table used to render ARAD-1K RGB. AnySpec sees procedural responses and real camera tables, with 8 cameras from a published database [16] held out of training. The fixed-response reference omits the auxiliary loss; two seeds use mixed precision and one full precision.

AnySpec reaches 0.0883 $\pm$ 0.0013 MRAE on the training camera and 0.0879 $\pm$ 0.0008 over the 8 held-out responses (per-camera means 0.0845 $\pm$ 0.0012–0.0972 $\pm$ 0.0015; mean $\pm$ s.d. over 3 independent seeds). Fixed-response AnySpec reaches 0.0838 $\pm$ 0.0004 on its training camera and 0.3359 $\pm$ 0.2794 on the same held-out cameras. Held-out MRAE thus stays close to training-camera MRAE, while PSNR is lower and SAM higher on the held-out cameras (Table I). Mean-consistent projection, the minimal measurement-consistent reference, reaches 0.8900 seen and 1.0131 held out (Table I, Figs. 3 and 4).

B. Response conditioning in a spatial backbone

AnySpec reconstructs each pixel independently, whereas competitive RGB-to-spectrum networks also use spatial context. To evaluate response conditioning in a spatial reconstructor, we replace MST++'s input convolution [8] with the AnySpec channel encoder. The encoded channel tokens are averaged at each pixel and mapped to 31 latent features; the spatial body and output convolution are unchanged, so the 31 features are an interface to that body rather than a preliminary band estimate. Both variants draw from the same pool of published three-channel camera tables, excluding the

TABLE I

ARAD-1K VALIDATION, CLEAN MEASUREMENT ($\sigma = 0$). SEEN IS THE BASLER RESPONSE BOTH MODELS SAW IN TRAINING; HELD-OUT IS THE MEAN OVER THE 8 CAMERA RESPONSES NEITHER MODEL SAW, AVERAGED WITHIN A SEED AND THEN ACROSS SEEDS. MEAN PROJ. IS THE TRAINING-BANK MEAN PROJECTED ONTO THE MEASUREMENT. MEAN $\pm$ STANDARD DEVIATION OVER 3 SEEDS.

Model, responses	MRAE ↓	PSNR ↑	SAM ↓
Fixed-response, seen	0.0838 $\pm$ 0.0004	38.31 $\pm$ 0.03	4.97 $\pm$ 0.01
Fixed-response, held-out	0.3359 $\pm$ 0.2794	32.95 $\pm$ 3.56	12.01 $\pm$ 7.33
AnySpec, seen	0.0883 $\pm$ 0.0013	37.61 $\pm$ 0.38	5.27 $\pm$ 0.20
AnySpec, held-out	0.0879 $\pm$ 0.0008	36.47 $\pm$ 0.27	5.56 $\pm$ 0.27
Mean proj., seen	0.8900	23.86	22.56
Mean proj., held-out	1.0131	23.35	22.91

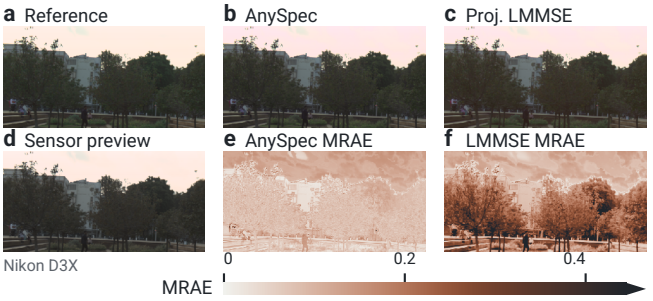


Fig. 4. Spectral reconstruction of one ARAD-1K crop under the Nikon D3X response. (a)–(c) 650/550/450 nm displays of the reference cube, AnySpec and projected LMMSE. (d) The rendered measurement. (e)–(f) Per-pixel MRAE for the two reconstructions on a shared scale. Noiseless measurement; one trained model of each kind.

8 test responses and containing no procedural operators, and see identical sampled training examples. They optimise native-output MRAE for 30,000 Adam updates at an initial learning rate of 4×10^{-4} on a cosine schedule, using batches of eight 128×128 patches drawn from 899 cubes, keeping the final checkpoint. A Basler-only run trains vanilla MST++ on a single response.

The conditioned stem reduces mean held-out native MRAE from 0.0837 to 0.0767, an 8.4% reduction, with lower error on 8 of 8 cameras, and reduces training-camera MRAE from 0.0846 to 0.0677. After projection the corresponding held-out means are 0.0726 and 0.0711, a 2.0% reduction with lower error on 6 of 8 cameras (Table II, Fig. 5). Resampling the 50 validation cubes (10,000 paired resamples) gives 95% intervals of 3.7–13.4% for the native reduction and -2.6 to 5.8% after projection, where the interval includes zero.

Trained on Basler alone, MST++ reaches 0.0624 native MRAE on that camera but 0.1245 on the held-out responses, so pooled-response training gives vanilla MST++ lower held-out error than single-response training. The two stems share a spatial body and training stream; the conditioned stem has 2.04 M parameters on its forward path against 1.62 M, 26% more. Both pooled variants also reach lower held-out MRAE than per-pixel AnySpec, which reaches 0.0955 $\pm$ 0.0018 native and 0.0879 $\pm$ 0.0008 projected and differs from them in objective, response pool, spatial context and training exposure

TABLE II

SPECTRAL RECONSTRUCTION ERROR ON ARAD-1K VALIDATION (MRAE, CLEAN MEASUREMENT, STRIDE 1, 50 CUBES). NATIVE IS THE ESTIMATE BEFORE CONSISTENCY PROJECTION AND PROJ. THE SAME ESTIMATE AFTER IT. SEEN IS THE BASLER RESPONSE; HELD-OUT IS THE MEAN OVER 8 JIANG CAMERAS EXCLUDED FROM EVERY TRAINING POOL. THE TWO POOLED MST++ VARIANTS SHARE A SPATIAL BODY AND IDENTICAL SAMPLED TRAINING EXAMPLES, AND DIFFER ONLY IN THEIR INPUT STEM; MST++ (BASLER ONLY) TRAINS ON A SINGLE RESPONSE. MST++ ROWS ARE SINGLE RUNS AND THE ANYSPEC ROW IS MEAN $\pm$ S.D. OVER 3 SEEDS.

Model	Seen native	Seen proj.	Held-out native	Held-out proj.
MST++ (pooled)	0.0846	0.0728	0.0837	0.0726
MST++ + response conditioning (pooled)	0.0677	0.0648	0.0767	0.0711
MST++ (Basler only)	0.0624	0.0610	0.1245	0.0967
AnySpec (per-pixel)	0.0978 ± 0.0019	0.0883 ± 0.0013	0.0955 ± 0.0018	0.0879 ± 0.0008
Nearest-neighbour mean ($k = 64$)	0.0833	0.0824	0.0885	0.0876

($30{,}000 \times 8 \times 128^2$ against $30{,}000 \times 4096$ pixel updates). A nearest-neighbour mean over the $k = 64$ nearest training-bank measurements, rendered through the same response (k chosen by leave-scene-out validation) and projected onto y, reaches 0.0876 held-out projected MRAE, similar to AnySpec.

C. Satellite instruments on HySpecNet-11k

HySpecNet-11k [17] is a benchmark of image patches from the Environmental Mapping and Analysis Program (EnMAP) satellite with 202 valid bands. Five instruments are discretised onto that grid: Sentinel-2A/B MSI (S2A, S2B) and Landsat 8 OLI (L8 OLI) in training; Sentinel-2C MSI (S2C) and Landsat 9 OLI-2 held out as near-sibling transfer tests.

On 300 matched test patches from 7 EnMAP L2A parent products across 3 paired seeds, projected LMMSE reaches 0.150–0.166 full-spectrum MRAE at $\sigma = 0$ and unit-gain AnySpec reaches 0.163–0.187, a paired ratio of 0.99–1.24 for AnySpec over LMMSE (Fig. 6): on clean satellite measurements projected LMMSE is slightly more accurate.

At $\sigma = 0.005$, where LMMSE uses the true noise variance and both estimators reproduce the same noisy y, unit-gain AnySpec reaches 0.224–0.248 against 0.257–0.274. Unit-gain training reduces error by 22–52% relative to exposure augmentation across seeds, sensors and both noise conditions.

The training covariance helps describe this difference in behaviour. The 8-band Landsat and 12-band Sentinel-2 operators leave 60–67% of the spectral variance in the null space, more than the cameras' 23–32%. Under each dataset's training covariance, the noiseless LMMSE estimate leaves residual variance of 0.45–0.63% of the total on satellites and 2.44–4.04% on cameras. These are in-sample squared-error diagnostics on different spectral grids; most null-space variation is linearly predictable on the satellite training bank. On the ARAD validation cubes, AnySpec has lower MRAE than projected LMMSE (Fig. 4): on every eighth row and column it reaches 0.089 against 0.124 on the training camera and 0.088 against 0.136 on held-out cameras.

Without retraining, we also scored the satellite models on two families absent from training, Terra MODIS (18 reflective or 7 land bands) and Sentinel-3 OLCI (19 bands centred below 1020 nm). On clean MODIS data the seed-mean AnySpec/LMMSE MRAE ratio grows to 1.29–1.59, from 1.07–1.16 on the near-siblings; at $\sigma = 0.005$ AnySpec stays ahead (0.93–0.96). On clean OLCI, AnySpec's lower MRAE (0.34 against 0.66) comes from wavelengths above 1000 nm (0.44 against 1.24); below, LMMSE is more accurate (0.05 against 0.24).

V. CONCLUSION AND LIMITATIONS

Trained across many responses with the response table as input, one set of weights transfers to unseen camera responses with little change in MRAE, and the same encoder improves a spatial network as its input stem. On satellites, fitted covariance gives a strong clean-data reference, including unseen MODIS responses and OLCI targets below 1000 nm. AnySpec is more accurate under added noise on the tested Sentinel-2, Landsat and MODIS responses.

AnySpec is per-pixel and deterministic and assumes a calibrated measurement scale. The stem comparison of Section IV-B rests on one run per variant and is not capacity matched; its interval reflects scene sampling, and the other spreads are standard deviations over training seeds. Noise is simulated and additive; the cross-family comparisons use one matched patch population from seven EnMAP products, and each grid comes from a single dataset. Every response table used is public: the Jiang camera database [16] and the Sentinel-2 and Landsat workbooks.

REFERENCES

[1] J. He, Q. Yuan, J. Li, Y. Xiao, D. Liu, H. Shen, and L. Zhang, "Spectral super-resolution meets deep learning: Achievements and challenges," *Information Fusion*, vol. 97, p. 101812, 2023.

[2] B. Arad, R. Timofte, R. Yahel, N. Morag, A. Bernat *et al.*, "NTIRE 2022 spectral recovery challenge and data set," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022.

[3] E. King, J. Rodriguez, D. Llanes, T. J. Doster, T. Emerson, and J. Koch, "STARS: Sensor-agnostic transformer architecture for remote sensing," in *IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2024.

[4] A. Siyavashi, J. Schlipf, S. Reiter, I. Fischer, C. Wenger, and C. Herglotz, "Sensor-adaptive infrared spectral reconstruction with plug-and-play diffusion priors," *arXiv preprint arXiv:2607.05636*, 2026.

[5] Y.-T. Lin and G. D. Finlayson, "Physically plausible spectral reconstruction from RGB images," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020.

[6] P. Morovič and H. Haneishi, "Estimating reflectances from multi-spectral camera responses," in *14th Color Imaging Conference*, 2006, pp. 131–137.

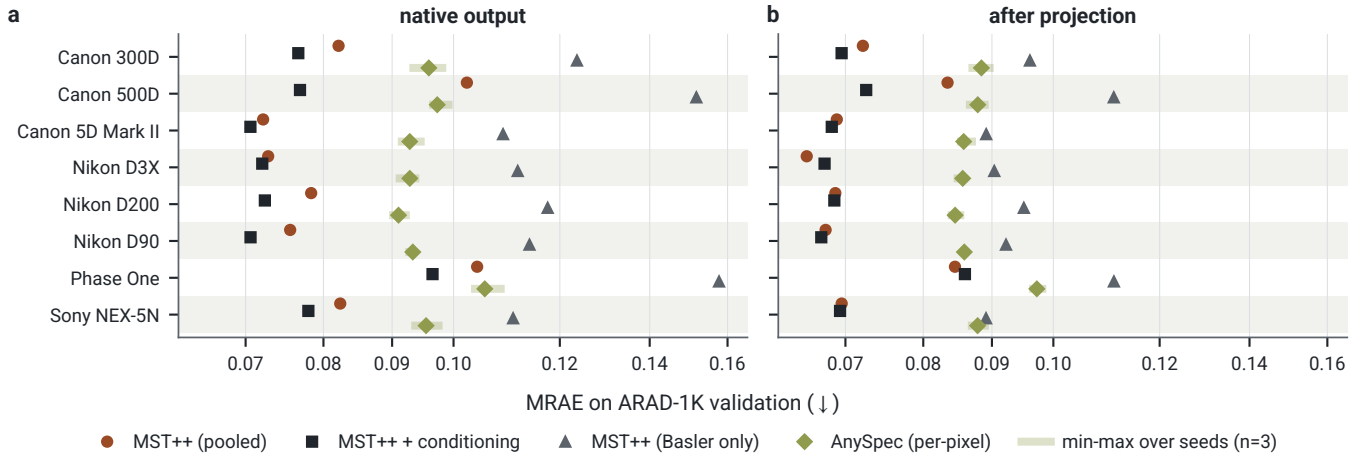


Fig. 5. Effect of response conditioning in MST++ before and after consistency projection. Per-camera held-out MRAE for (a) native output and (b) the projected output, on a shared logarithmic axis. Bars: range over 3 AnySpec seeds; MST++ variants are single runs.

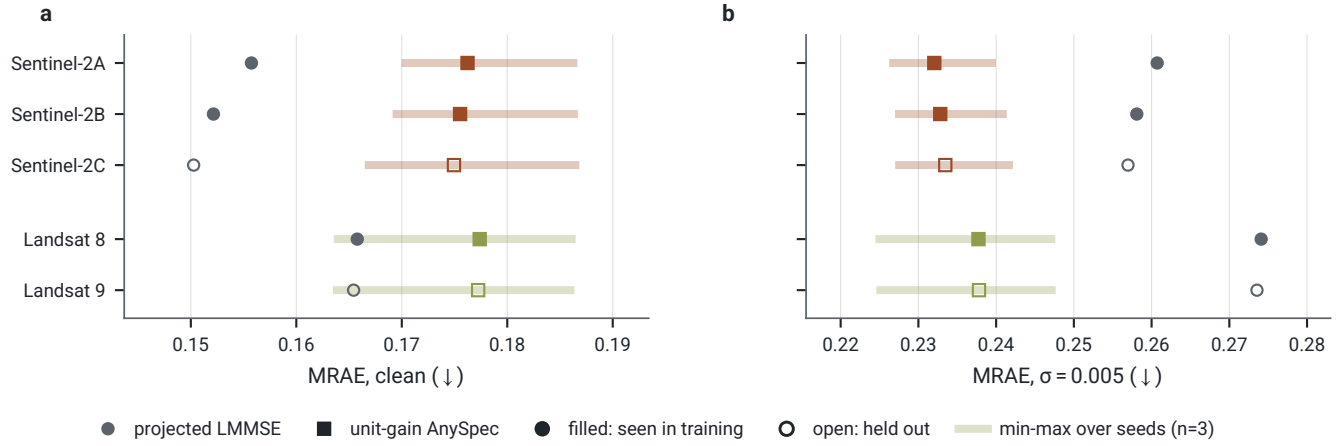


Fig. 6. Satellite reconstruction with unit-gain training on 300 matched HySpecNet-11k patches. (a) Clean MRAE; (b) MRAE at $\sigma = 0.005$. Squares: AnySpec, mean over 3 seeds with bars spanning the seed range. Circles: projected LMMSE. Filled markers are instruments seen in training, open markers the held-out near-siblings.

[7] L. Chalcroft, E. Camarillo Abad, P. Christopher, T. Albrow-Owen, and O. Burton, "What a multispectral measurement determines: Response summaries and consistency repair," in *Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Glasgow, UK, 2026.

[8] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, and L. Van Gool, "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022.

[9] J. B. Cohen and W. E. Kappauf, "Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks," *The American Journal of Psychology*, vol. 95, no. 4, pp. 537–564, 1982.

[10] G. D. Finlayson, Y.-T. Lin, and A. Kucuk, "Matrix-R theory: A simple generic method to improve RGB-guided spectral recovery algorithms," *Sensors*, vol. 25, no. 24, p. 7662, 2025.

[11] X. Yang, J. Chen, and Z. Yang, "PhySpec: Physically consistent spectral reconstruction via orthogonal subspace decomposition and self-supervised meta-auxiliary learning," in *International Conference on Machine Learning (ICML)*, 2025. [Online]. Available: https://openreview.net/forum?id=WISfJyOA6M

[12] J. Wang, S. Wang, R. Zhang, Z. Zheng, W. Liu, and X. Wang, "A range-null space decomposition approach for fast and flexible spectral compressive imaging," *arXiv preprint arXiv:2305.09746*, 2023.

[13] J. Prexl and M. Schmitt, "SenPa-MAE: Sensor parameter aware masked autoencoder for multi-satellite self-supervised pretraining," in *DAGM German Conference on Pattern Recognition (GCPR)*, 2024.

[14] L. Waldmann, A. Shah, Y. Wang, N. Lehmann, A. J. Stewart, Z. Xiong, X. X. Zhu, S. Bauer, and J. Chuang, "Panopticon: Advancing any-sensor foundation models for earth observation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (EarthVision)*, 2025.

[15] A. Baumann, L. Ayala, S. Seidlitz, J. Sellner, A. Studier-Fischer, B. Özdemir, L. Maier-Hein, and S. Ilic, "CARL: Camera-agnostic representation learning for spectral image analysis," *arXiv preprint arXiv:2504.19223*, 2025.

[16] J. Jiang, D. Liu, J. Gu, and S. Süsstrunk, "What is the space of spectral sensitivity functions for digital color cameras?" in *IEEE Workshop on Applications of Computer Vision (WACV)*, 2013.

[17] M. H. P. Fuchs and B. Demir, "HySpecNet-11k: A large-scale hyperspectral dataset for benchmarking learning-based hyperspectral image compression methods," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2023, pp. 1779–1782.