SPECTRALFORGE: A PHYSICS-BASED SIMULATOR FOR FLUORESCENCE HYPERSPECTRAL IMAGING

Narek Meloyan[1,2,3], *Narine Sarvazyan*[1,3,4]

[1]Akian College of Science and Engineering, American University of Armenia, Yerevan, Armenia
[2]Engineering Research Center, American University of Armenia, Yerevan, Armenia
[3]L.A. Orbeli Institute of Physiology NAS RA, Yerevan, Armenia
[4]George Washington University, Washington, DC, USA
narek.h.meloyan@gmail.com, phynas@gwu.edu

ABSTRACT

Acquiring multi-excitation fluorescence hyperspectral datasets can be slow and costly. An instrument records a full emission spectrum at each of several excitation wavelengths and for every pixel, fluorescence signals are weak and easily buried in scatter and autofluorescence, and the illumination needed to raise them photobleaches the sample, so a single dataset can consume hours of scarce instrument time. This makes it hard to plan acquisitions, prototype processing methods, or build shared benchmarks. We present **SpectralForge**, an open, physics-based simulator that renders realistic 4D excitation-emission cubes with ground truth from a parametric fluorophore model obeying the standard laws of fluorescence photophysics (Kasha, Vavilov, Franck-Condon, Stokes shift), with optional inner-filter attenuation, reabsorption, Foerster resonance energy transfer (FRET), scatter, photon and read noise, and instrument clutter. Because every rendered cube comes with its concentration maps and true emission peaks, a method can be developed, tested, and compared before any measurement is taken. As a worked example, we study spectral band selection, where the goal is to keep only the informative bands and so acquire or process less. Scenes are built so that band variance is decoupled from class informativeness (measured correlation ≈ -0.1), so the task is not solved by variance ranking, and two controls, a random baseline and the full spectrum, tell a real gain from a chance one. Fitting the model to measured drop and collagen spectra tests it against real data, and running a panel of selectors through the simulator shows it behaving as designed and marks where label-free selection helps and where it does not.

***Index Terms*—** fluorescence hyperspectral imaging, physics-based simulation, synthetic data, excitation-emission, photophysics, band selection.

1. INTRODUCTION

A multi-excitation fluorescence measurement illuminates a sample at several excitation wavelengths and, for each, records an emission spectrum. Every pixel therefore carries an excitation-emission matrix (EEM), and an image is a four-dimensional array over excitation, emission, height, and width (Fig. 1). Fluorophores, tissue states, and species separate in this joint domain, which is why it can be used for biomedical and material imaging, for example label-free imaging of cardiac tissue by its endogenous autofluorescence [9, 10]. Added excitations separate overlapping fluorophores that a single-excitation image cannot, improving tissue and material classification [12–14], but each one multiplies the acquisition by a full emission scan, signals are weak against scatter and autofluorescence, and illumination photobleaches the sample. A single 4D cube can consume hours of instrument time and labeled real cubes remain scarce.

Planning an acquisition, prototyping a processing method, or comparing algorithms all ideally happen before instrument time is committed, yet each needs representative data, and real cubes carry no ground truth about which fluorophores are where or which bands matter. A physically realistic simulator that renders cubes from the laws of fluorescence and returns the underlying concentrations and emission peaks addresses both problems: a method can be built, tested, and benchmarked against an answer key, and only the promising design need reach the instrument.

This paper presents **SpectralForge**, an open simulator of that kind for 4D fluorescence imaging, with an interactive workbench for building scenes by hand and a small library for scripting them. To show the tool at work we take one concrete task, spectral band selection, and run a study through it end to end: the simulator provides the cubes and their ground truth, a controlled evaluation protocol scores the candidate selectors, and the results show both that the tool behaves as designed and where a learned selector actually helps. Band selection is only one use; the same generator supports experiment planning, prototyping, and teaching.

2. THE SPECTRALFORGE FORWARD MODEL

2.1. Photophysics

The generator is built on the physics of fluorescence [1,3] so that "informative" means what it means physically. A fluorophore absorbs a photon, relaxes non-radiatively to the lowest vibrational level of the first excited singlet, and emits (the Jabłoński picture [5]). Four consequences shape the EEM. By *Kasha's rule* [2] emission occurs from that lowest level regardless of the excitation, so the emission shape is independent of excitation wavelength and the EEM is approximately trilinear (an excitation profile times an emission profile times a concentration map), the basis of PARAFAC [4]. By *Vavilov's rule* the quantum yield is, to first order, independent of excitation; only the amount absorbed changes. The *Franck-Condon* principle sets the broadened, roughly mirror-image band envelopes and an optional vibronic shoulder about $1300\,\mathrm{cm}^{-1}$ to the red. The *Stokes shift* red-shifts emission relative to absorption. Together these rules fix the structure of the forward model below; the departures from them that matter in practice (inner-filter attenuation, reabsorption, and FRET) enter as optional, switchable effects (Sec. 2.4), so that each one can be studied in isolation by turning it on or off.

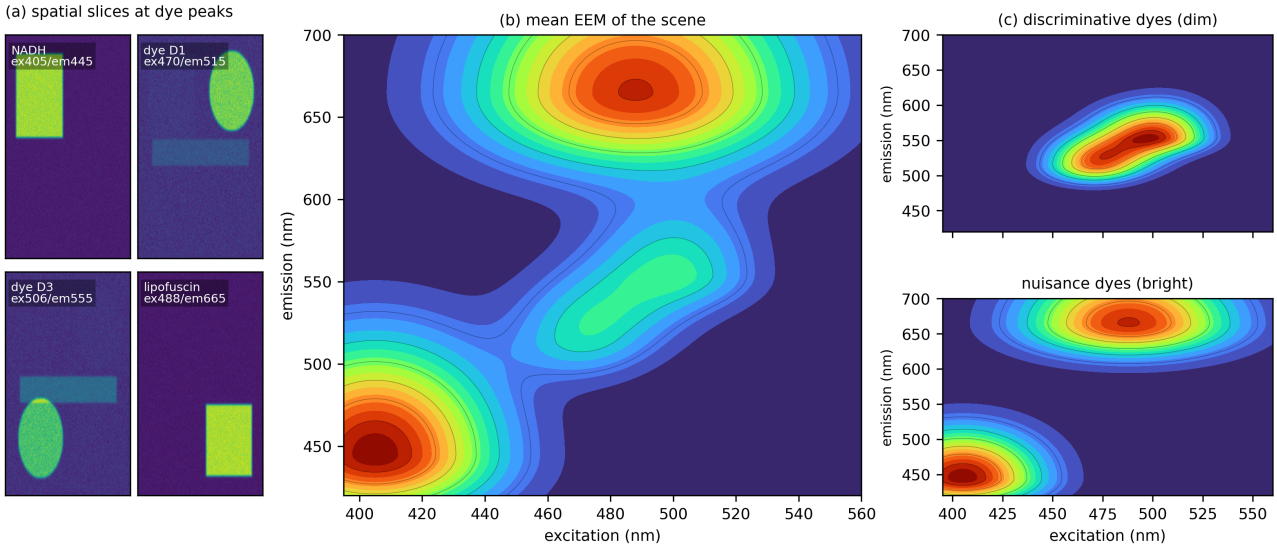


Fig. 1. A rendered 4D fluorescence cube from SpectralForge, an example scene with each dye laid down as a shape. (a) Slices at each dye's emission peak reveal its painted region; every pixel effectively stores an EEM. (b) The mean EEM: each fluorophore is a smooth blob at its excitation and (Stokes-shifted) emission. (c) By role, the discriminative dyes form one dim, overlapping cluster while the class-irrelevant nuisance dyes are bright and well separated, so band variance is dominated by uninformative signal.

2.2. Trilinear render and the additive invariant

Each fluorophore k is a parametric model with a peak-normalized Gaussian excitation profile $a_k(\lambda_{\text{ex}})$ and an area-normalized emission profile $e_k(\lambda_{\text{em}})$, extinction ε_k, quantum yield Φ_k, and a per-pixel concentration map $c_k(x, y)$. The rendered cube is a dilute linear superposition,

$$F(x, y, \lambda_{\text{ex}}, \lambda_{\text{em}}) = g \sum_k c_k(x, y)\, \varepsilon_k\, \Phi_k\, a_k(\lambda_{\text{ex}})\, e_k(\lambda_{\text{em}}), \quad (1)$$

where g folds lamp, exposure, and power. Gaussian widths are $\sigma = \text{FWHM}/(2\sqrt{2\ln 2})$, and brightness is $\varepsilon_k \Phi_k$. With all optional physics and artifacts off the model is *exactly linear*, $\text{render}(A+B) = \text{render}(A)+\text{render}(B)$, an invariant the implementation tests; difficulty is then added in realistic, opt-in increments. A per-pixel class label is the argmax over the discriminative fluorophores. Fig. 2 shows the fluorophore model.

2.3. Variance is decoupled from informativeness

The central design choice is the scene generator. Discriminative dyes are painted with smooth random concentration fields whose per-pixel argmax defines the class; they are dim and spectrally overlapping, so their content lives in low-variance band *shape*, while independent nuisance dyes are bright but class-irrelevant (Fig. 1c). Scatter ridges, bright autofluorescence, and shot noise add high-variance but class-irrelevant signal. On a confounded scene the correlation between per-band variance and per-band discriminability is ≈ -0.1, versus strongly positive on clean data, so the highest-variance bands are uninformative (Fig. 3) and variance ranking fails.

2.4. Artifacts, nonlinear physics, and clutter

On top of the clean cube the generator adds Rayleigh scatter at $\lambda_{\text{em}} = \lambda_{\text{ex}}$ (with a second-order ghost at $2\lambda_{\text{ex}}$) and a water Raman line at a fixed wavenumber offset ($\approx 3400\,\text{cm}^{-1}$), each scaled by a per-pixel turbidity field, then Poisson shot noise and Gaussian read noise. Three optional effects introduce concentration nonlinearity: a primary inner filter $\exp(-s_{\text{ie}} A_{\text{ex}})$; *reabsorption*

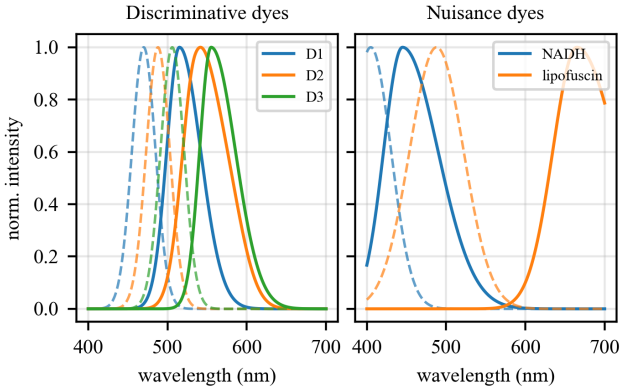


Fig. 2. Parametric fluorophore spectra (dashed, excitation; solid, emission). Left: three dim, overlapping discriminative dyes. Right: two bright nuisance autofluorophores that dominate band variance but carry no class information.

$\exp(-s_{\text{re}} A_{\text{em}}(\lambda))$, whose overlap with the blue edge of emission reshapes the band and relocates discriminative information into band shape; and FRET, where a saturating fraction $E = kc_A/(1 + kc_A)$ of donor energy is transferred to a co-localized acceptor, encoding dye co-localization in a band ratio. Compactly, the attenuated signal and the transfer efficiency are

$$F_{\text{obs}} = F\, e^{-s_{\text{ie}} A_{\text{ex}}}\, e^{-s_{\text{re}} A_{\text{em}}(\lambda_{\text{em}})}, \qquad E = \frac{k\, c_A}{1 + k\, c_A}, \quad (2)$$

with $A_{\text{ex}} = \sum_k \varepsilon_k c_k a_k(\lambda_{\text{ex}})$, dimensionless strength knobs $s_{\text{ie}}, s_{\text{re}}$ (not physical path lengths), reabsorption $A_{\text{em}}(\lambda) = \sum_k \varepsilon_k c_k\, a_k(\lambda)$ (the excitation profile serves as absorption profile at the emission wavelength), and a Raman line at $1/\lambda_R = 1/\lambda_{\text{ex}} - \Delta\bar{\nu}$. Finally, fixed-pattern clutter injects many class-irrelevant low-rank spatial fields on random band subsets, the analogue of illumination and detector artifacts and the dominant real-instrument confound. By dial-

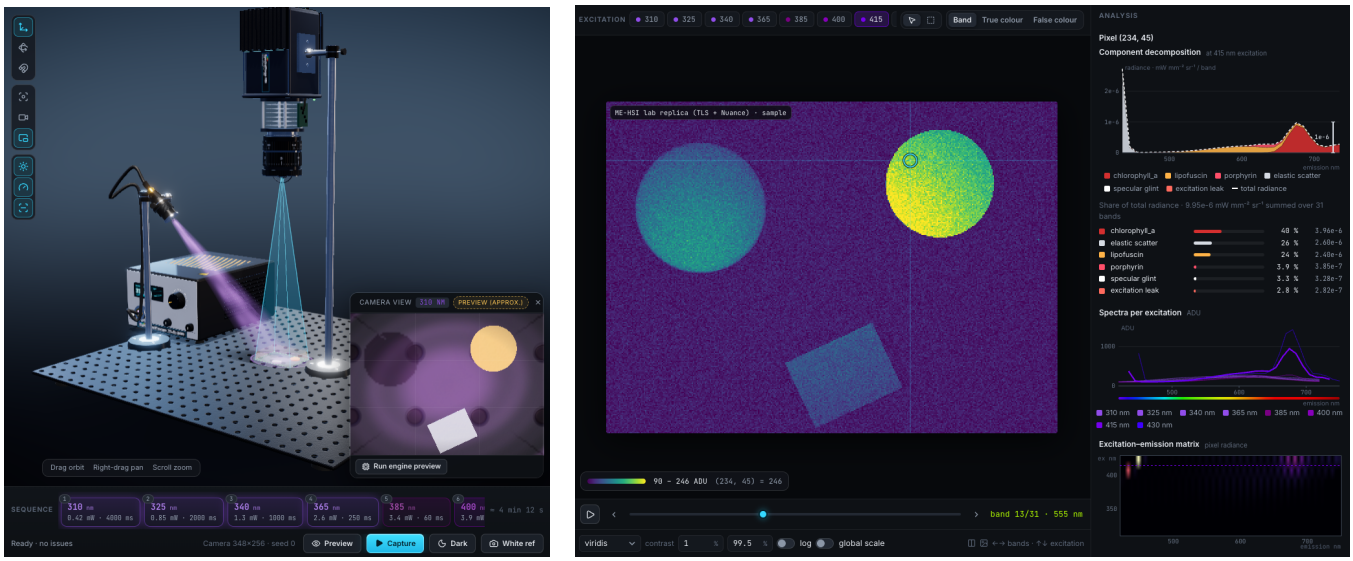


Fig. 4. SpectralForge Studio. (a) Design: lamp, LCTF camera and samples placed in 3D, with the camera view. (b) Explore: a band image and, for the selected pixel, its ground-truth physical components, spectra and EEM. The data analysis shown in Fig. 5 used the core generator of Sec. 2, not Studio's added physics.

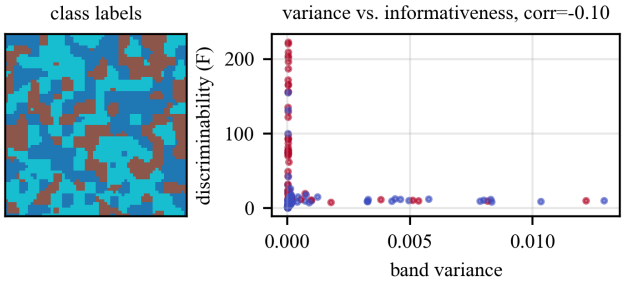


Fig. 3. Variance is decoupled from informativeness. Left: class labels. Right: each point is a band (red: discriminative; blue: nuisance and scatter); the high-variance bands are nuisance or scatter, not discriminative (correlation near zero), so variance ranking fails.

ing these knobs it spans *clean* (variance ≈ informativeness, an idealized baseline real acquisitions never meet), *realistic*, *clutter*, *nonlinear* (reabsorption or FRET), and *spatial-texture* regimes in one call.

2.5. Ground truth, reproducibility, and real fluorophores

Every render returns a ground-truth sidecar with concentration maps, clean cubes, and per-fluorophore emission peaks, which enables a tight *peak-recovery* metric (did a selected band land within a tolerance of a true emission peak; a random selector scores about 0.33) in addition to information-retention metrics. The study scripts (`build_dataset` in `reports/`, not library API) call `render` with explicit knobs and record exact parameters. Parametric dyes can be replaced by measured spectra from a downloaded FPbase payload via a library importer. The 3D workbench, SpectralForge Studio (Fig. 4), lets a user place a lamp, camera, and samples in a 3D scene, render the 4D cube, and inspect it against ground truth. Each Studio scene is a plain JSON document with its seed and capture settings, so any capture can be saved, shared with collaborators and re-rendered exactly from that one file.

Table 1. Experimental setup and the reason for each element.

Element	Setting	Why
Selectors	variance; PCA loadings; AE (label-free); mutual information (labels)	naive to supervised
Regimes	clean; realistic; clutter; reabsorption; FRET; spatial texture	ideal to nonlinear
Noise	5 levels: photon scale 5×10^4 to 200, Rayleigh 0.1 to 0.7, clutter 0 to 4.5	gain under degradation
Classifiers	k-NN, RF, MLP, LR; macro-F1; CV (block CV for texture)	scored by downstream use
Controls	random subsets (95th pct.); full spectrum	chance; ceiling
Seeds	6 (Table 2); 2 (Fig. 5); 3 (texture)	paired CIs
Scene, grid	64×64 px (Table 2: 40×40), 3 classes; 564 or 228 bands (Sec. 3.1)	known labels, peaks

2.6. Validation against real acquisitions

Fitting Eq. (1), through a model of our tunable-lamp/LCTF instrument [12], to 4 excitations (310/340/385/415 nm) of 16 fluorescent drops predicts the 3 held-out excitations without gain refit (median $R^2 = 0.96$, SAM $6.0°$, Fig. 6a). For collagen, whose intensities are incomparable across excitations, only shape is testable: after per-excitation gain and smooth throughput correction, held-out SAM is $8.4 \pm 1.8°$ and shape $R^2 = 0.83 \pm 0.13$ (24 cases; Fig. 6b,c), slightly worse than copying another sample's spectrum ($7.7°$).

3. EVALUATION PROTOCOL

Two questions are kept apart: simulator fidelity, covered by the linearity tests of Sec. 2.2 and the real-data check of Sec. 2.6; and benchmark validity, covered by known ground truth plus random and full-spectrum controls, which need no claim that a simulated cube equals a real one. Table 1 summarizes the setup.

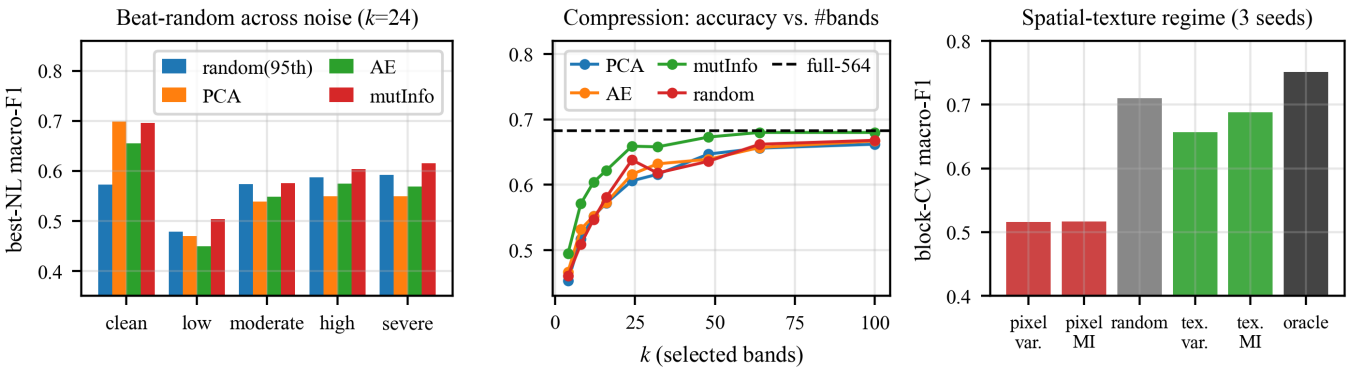


Fig. 5. The benchmark in use. Left: beat-random across five noise levels at $k = 24$ of 564 bands (Sec. 3.1): blind PCA and autoencoder beat the random 95th percentile only on clean data; under clutter only supervised MI does. Middle: compression over 4 to 100 bands; the autoencoder tracks PCA (about 95% of the full spectrum); labels compress about 2.7× better. Right: spatial texture (block CV): per-pixel selectors fall below random and no learned spatial selector beats it.

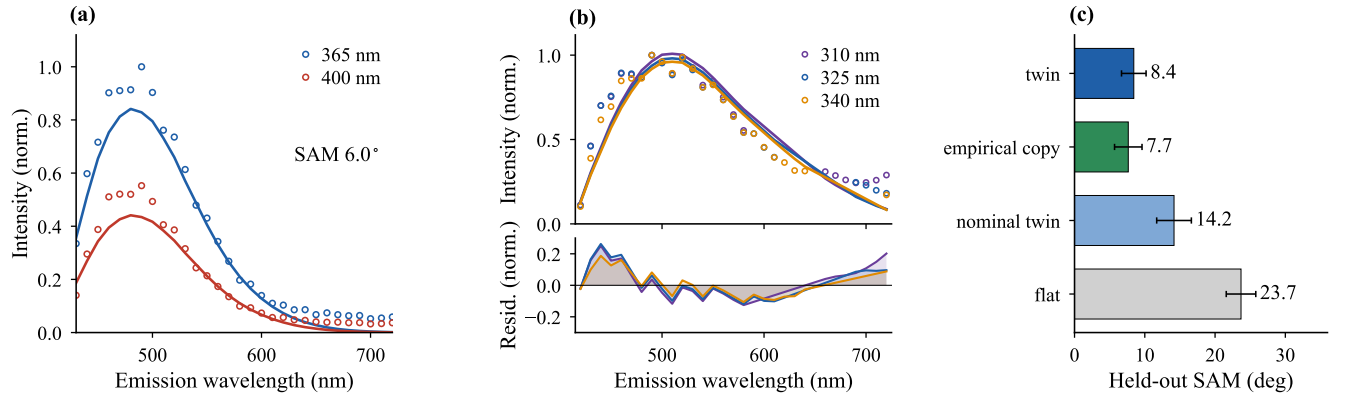


Fig. 6. Real-data validation. (a) The drop with the median SAM (of 16): measured (markers) and predicted (lines) at held-out excitations. (b) The collagen region of interest with the median SAM (of 24): measured, predicted and residual at held-out 310/325/340 nm. (c) Collagen held-out SAM, mean ± SD (24 cases).

3.1. Spectral grid

The study uses 4 excitations (405, 470, 488, 506 nm) and emission from 420 to 700 nm: every 2 nm for Fig. 5 (141 bands, $4 \times 141 = 564$ excitation-emission bands, the "full-564" classifier) and every 5 nm for Table 2 (57 bands, 228 features).

Selection quality is the accuracy of a downstream classifier on the selected bands, with labels used only at evaluation time. We score each subset with a classifier panel (k-nearest neighbors, random forest, MLP, logistic regression), cross-validated, reporting macro-F1 and, on nonlinear regimes, a *nonlinear-only gap* (the accuracy an expressive model recovers beyond a linear one), because a linear metric cannot credit a nonlinear-shape selection. Two controls matter here. A **random** baseline: a method that cannot beat the 95th percentile of random band subsets adds nothing. And the **full spectrum**: the accuracy ceiling a selector tries to approach. Where discrimination is spatial we use spatial-block cross-validation to remove spatial-autocorrelation leakage.

4. USING THE BENCHMARK

Band selection, which removes non-informative bands to shrink a real hyperspectral acquisition [8, 11, 15] (its collagen and lichen data are public [16, 17]), is one thing SpectralForge is built to test, so we exercise it with a panel of standard selectors: variance ranking, PCA loadings [7], supervised mutual information (a labels-using reference), and a fully unsupervised autoencoder [6]-with-latent-perturbation selector, which reconstructs the standardized cube, perturbs each latent dimension, and turns the per-band reconstruction sensitivity into a per-band influence. The question is whether decoupling variance from informativeness makes the benchmark discriminate methods and regimes.

The results show that it does. On confounded scenes a variance ranker does no better than, and usually below, random subsets (e.g. 0.40 vs 0.45 on the realistic regime). PCA loadings is the strongest blind selector only in the idealized *clean* regime, where the informative bands are also the brightest. No real fluorescence acquisition reaches that regime, since real cubes always carry scatter, autofluorescence, and noise; once those enter, on the realistic and FRET regimes, the two selectors tie. Label-free selection marginally overtakes PCA only where the physics places discriminative information in nonlinear band *shape*, as under reabsorption (macro-F1 0.449 vs 0.431, bootstrap 95% CI $[+0.001, +0.029]$, Cohen $d \approx +0.94$, but paired Wilcoxon $p = 0.16$, so the gain is marginal; the AE-MLP subset carries nonlinear-only information $+0.014$, the conv-AE -0.001, against -0.016 for PCA loadings; Table 2). The pattern is a general one: *informative implies has variance implies PCA-selectable*, so a nonlinear *classification* task need not imply a nonlinear *selection* task, and the benchmark separates the two.

Table 2. Label-free AE vs. the strongest blind baseline (PCA loadings): best-of-panel macro-F1, six seeds; bold is better. Clean margin: paired Wilcoxon $p = 0.03$; reabsorption: bootstrap 95% CI excludes zero but $p = 0.16$ (marginal); rest are ties.

Regime	AE	PCA-load	verdict
clean	0.590	**0.635**	linear suffices
realistic	0.527	0.531	tie
reabsorption	**0.449**	0.431	marginal AE gain
FRET	0.635	0.642	tie

The random and full-spectrum controls also mark the limits (Fig. 5). Under fixed-pattern clutter no blind selector beats a random subset, because clutter is high-variance but class-irrelevant; there only label-aware selection helps. No selector beats the full spectrum, so selection buys compression, not accuracy, and blind subsets are unstable band to band (Jaccard about 0.1 to 0.2), which favors selecting spectral regions or consensus sets. And when class is carried by spatial texture, so that every class shares the same per-pixel spectrum, per-pixel selection falls below random (three seeds, 0.52 against a random 0.71), the oracle bands reach 0.75, and none of the spatial selectors we tried (texture variance or MI, 0.66 to 0.69; a spatial CNN, 0.53 in a separate finer configuration with random 0.58) beats random, so spatial selection remains open.

5. DISCUSSION AND CONCLUSION

SpectralForge provides researchers in the 4D fluorescence imaging field with an open, physically realistic source of cubes with perfect ground truth, so methods can be tested before scarce instrument time is spent. With random and full-spectrum controls, the simulator turns "does this selector help" into a measured, reproducible question, and its answer is specific: PCA loadings is the strongest blind baseline only in the idealized clean regime that no real acquisition reaches, ties label-free selection once realistic scatter and noise enter, and is only marginally overtaken where the physics is nonlinear.

While SpectralForge captures key primary photophysical phenomena, the simulator operates under certain physical and optical simplifications. Currently, inner-filter attenuation and reabsorption are modeled using first-order exponential path approximations rather than full 3D radiative transfer or Monte Carlo photon transport through dense scattering media. Furthermore, the instrument model assumes an idealized optical train with spatially uniform illumination, uniform spectral detector response, and static photophysics, omitting dynamic exposure photobleaching kinetics during sequence acquisition as well as chromatic aberrations and thermal drift. Expanding the engine to incorporate full radiative transport and dynamic bleaching models remains an important direction for future extensions; the workbench shown in Fig. 4 already adds 3D Monte Carlo transport and photobleaching.

A natural next step is to evaluate performance on complex, multi-component target samples (heterogeneous biological tissues or engineered multi-fluorophore composites), where informative spectral structure is decoupled from overall band variance.

Acknowledgment

Financial support from the European Union NAS SAR-101087403 award is gratefully acknowledged. A large-language-model assistant was used for code scaffolding, figure generation, and language editing; the design, analysis, and conclusions are the authors' own and were verified by them.

6. REFERENCES

[1] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd ed. Springer, 2006.

[2] M. Kasha, "Characterization of electronic transitions in complex molecules," *Discuss. Faraday Soc.*, vol. 9, pp. 14–19, 1950.

[3] B. Valeur and M. N. Berberan-Santos, *Molecular Fluorescence*, 2nd ed. Wiley-VCH, 2012.

[4] R. Bro, "PARAFAC: Tutorial and applications," *Chemom. Intell. Lab. Syst.*, vol. 38, pp. 149–171, 1997.

[5] A. Jabłoński, "Efficiency of anti-Stokes fluorescence in dyes," *Nature*, vol. 131, pp. 839–840, 1933.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, 2006.

[7] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.

[8] Y. Cai, X. Liu, and Z. Cai, "BS-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE TGRS*, vol. 58, no. 3, pp. 1969–1984, 2020.

[9] M. Mercader, L. Swift, S. Sood, H. Asfour, M. Kay, and N. Sarvazyan, "Use of endogenous NADH fluorescence for real-time in situ visualization of epicardial radiofrequency ablation lesions and gaps," *Am. J. Physiol. Heart Circ.*, vol. 302, no. 10, pp. H2131–H2138, 2012.

[10] D. A. Gil, L. M. Swift, H. Asfour, N. Muselimyan, M. A. Mercader, and N. A. Sarvazyan, "Autofluorescence hyperspectral imaging of radiofrequency ablation lesions in porcine cardiac tissue," *J. Biophotonics*, vol. 10, no. 8, pp. 1008–1017, 2017.

[11] S. Guan, H. Asfour, N. Sarvazyan, and M. Loew, "Application of unsupervised learning to hyperspectral imaging of cardiac ablation lesions," *J. Med. Imaging*, vol. 5, no. 4, art. 046003, 2018.

[12] N. Chilingaryan, F. Villarruel, T. Soghomonyan, V. Yeghiazaryan, and N. Sarvazyan, "4D hyperspectral imaging for intraoperative tissue classification," in *Proc. SPIE Med. Imaging*, vol. 13410, 2025, pp. 255–260.

[13] N. Chilingaryan, A. Gasparyan, and N. Sarvazyan, "Multi-excitation hyperspectral imaging toward improved lichen identification," in *Proc. WHISPERS*, 2025, pp. 1–5.

[14] A. Hunanyan, N. Ter-Petrosyan, F. Villarruel, T. Soghomonyan, N. Sarvazyan, A. Butavyan, and V. Yeghiazaryan, "Deep learning classification of ablated atrial tissue in multi-excitation hyperspectral images," in *Proc. WHISPERS*, 2025, pp. 1–5.

[15] N. Meloyan and N. Sarvazyan, "Unsupervised deep learning for wavelength selection in multi-excitation fluorescence imaging," *Scientific Reports* under review, 2026.

[16] K. Ghahramanyan, "250929_Collagen," dataset, Zenodo, 2026, doi: 10.5281/zenodo.20677504.

[17] N. Chilingaryan and N. Sarvazyan, "250707_4DHSI_Lichens_Palette," dataset, Zenodo, 2026, doi: 10.5281/zenodo.18640119.