



BACKGROUND

WHY MULTIPLE-COMPONENT EXTRACTION MATTERS

Mixed SERS/Raman measurements often contain several targets, interferences, and matrix background in a single spectrum. Recovering each component—not only reconstructing the mixture—provides interpretable molecular signatures for identification, comparison, and downstream quantification. It can reveal coexisting species without requiring a separate assay for every analyte.

Objective. Recover sample-shared component spectra and sample-specific nonnegative weights directly from mixtures while keeping measured backgrounds explicit.

LINEAR MIXTURE MODEL

$$y_i(x) = \sum_{k=1}^{K_u} a_{ik} s_k(x) + \sum_{l=1}^{K_r} c_{il} r_l(x) + \varepsilon_i(x)$$

MODEL TERMS. Unknown spectra $s_k(x)$ are shared across mixtures; nonnegative coefficients a_{ik} vary by sample. Measured backgrounds $r_l(x)$ enter as fixed templates.

REFERENCE-FREE TRAINING. Pure spectra stay outside the loss and are used only for post hoc alignment, normalization, and permutation matching.

COMPOSITION CONDITIONING. Designed or reference-derived sample coordinates condition the coefficient branch.

WHY THE PROBLEM IS HARD

Peak overlap. Broad and closely spaced Raman bands make the decomposition non-unique; several sets of component spectra can reconstruct the same mixture.

Branch symmetry. Neural branches can interchange, duplicate one another, or collapse while the mixture-level reconstruction loss remains low.

Scale and permutation ambiguity. A visually accurate mixture fit may still assign the wrong identity or amplitude to individual components.

Variable sample composition. The extraction must share spectral shapes across samples while allowing nonnegative coefficients to change with experimental composition.

Measured backgrounds. Known matrix spectra must remain physically distinct from the unknown targets instead of being absorbed by a learned branch.

SUCCESS CRITERION

The goal is identifiable recovery of every unknown spectrum across multiple-component mixtures. Component-wise agreement—not reconstruction error alone—defines successful extraction.

STUDY SCOPE: 2–4 COMPONENTS

Component-number flexibility. The same architecture supports two-, three-, and four-component mixtures; only the number of spectrum branches changes.

Biological systems. Nine binary and four ternary SERS mixture systems span six respiratory viruses measured in human saliva.

Controlled benchmark. ABCD mixtures recover four unknown analytes with plastic supplied as a measured background template.

Held-out references. Pure-component spectra remain outside training and are used only for post hoc matching and evaluation.

Evaluation readout. Component-spectrum correlation is primary; reconstruction fidelity and coefficient tracking provide complementary diagnostics.

METHODS

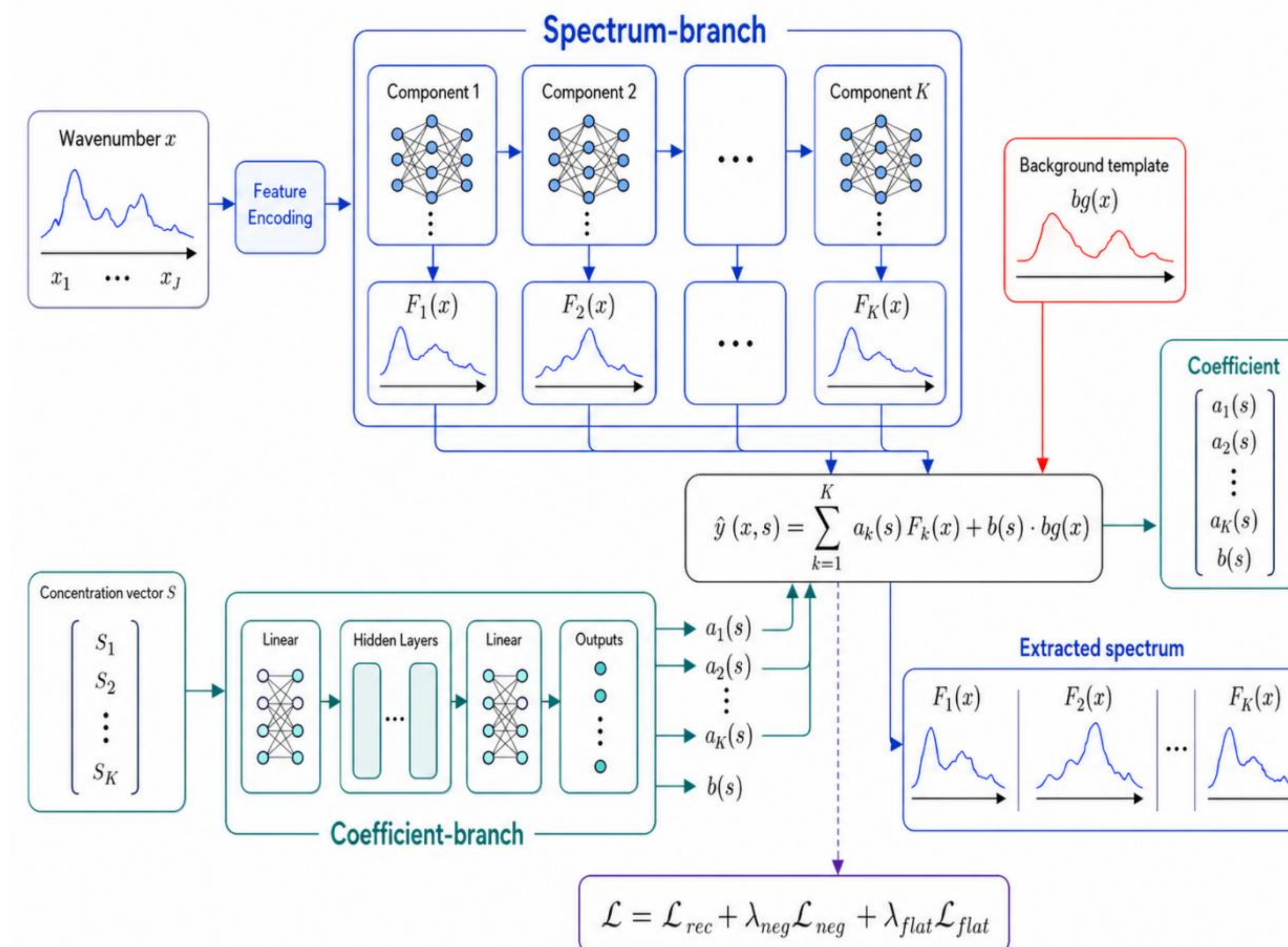


Figure 1. Multiple-component Raman extraction framework.

MULTIPLE-COMPONENT EXTRACTION FRAMEWORK

Spectrum branch. Fourier-encoded Raman coordinates feed K independent residual networks that learn sample-shared component spectra.

Coefficient branch. Composition coordinates pass through a residual multilayer perceptron; Softplus yields nonnegative sample weights.

Physical reconstruction. Learned and fixed spectra are combined explicitly at every Raman shift.

Training objective. Mean absolute reconstruction error is supplemented by spectral nonnegativity and flatness-avoidance penalties.

IMPLEMENTATION DETAILS

Spectrum network. 11 Fourier features $\rightarrow 2 \times 256$ fully connected layers $\rightarrow 3 \times 256$ residual blocks $\rightarrow 128$ latent units $\rightarrow$ intensity.

Coefficient network. 256-unit input layer + 3×256 residual blocks $\rightarrow$ K unknown coefficients and optional known/background coefficients.

Optimization. Xavier initialization; Adam for 200 epochs; $\leq 131,072$ spectrum–shift pairs per epoch; gradient norm clipped at 5.

OPTIONAL SHARED PEAK PRIOR

Construction. A branch-agnostic Gaussian mask covers 28 plausible Raman regions ($\sigma = 12 \text{ cm}^{-1}$) and is shared by all branches.

Penalty. Positive spectral energy outside the mask is penalized, suppressing implausible peaks without assigning component identity.

Schedule. The prior turns on after epoch 40; the spectrum and coefficient networks, physical reconstruction, and base losses remain unchanged.

Boundary. Apply only to difficult virus SERS overlap; retain prior-free training when components are already identifiable.

RESULTS

FOUR COMPONENTS ARE RECOVERED JOINTLY

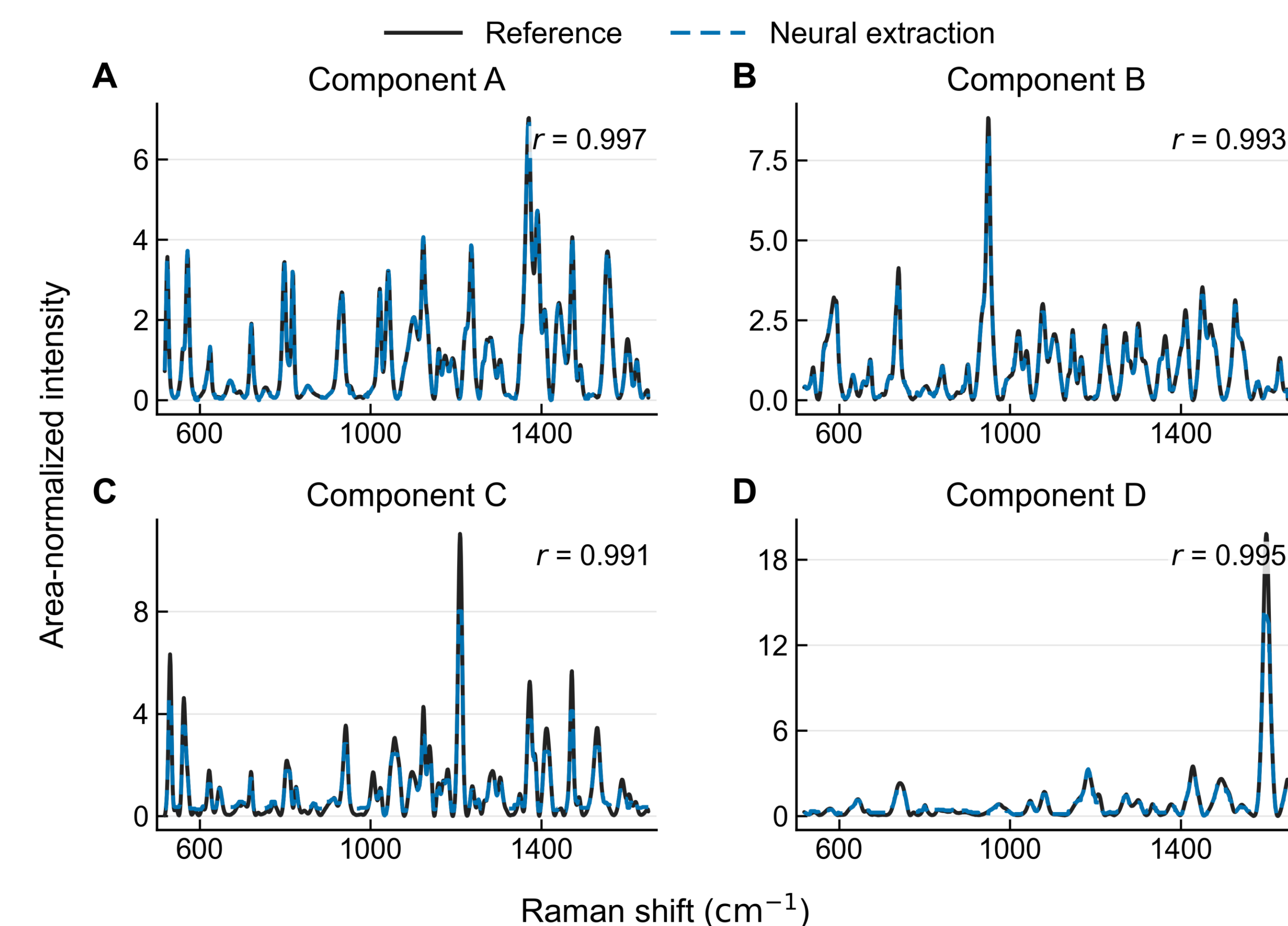


Figure 2. Joint extraction of Raman components A–D.

25,599 mixtures · 1,015 Raman shifts · plastic fixed as measured background

Pearson r: A 0.9971 · B 0.9925 · C 0.9914 · D 0.9947 $\rightarrow$ **mean r = 0.9939**

Interpretation. All four unknown spectra are recovered together. Reference spectra are used only after training for matching and evaluation.

PEAK PRIOR IMPROVES OVERLAPPED MIXTURES

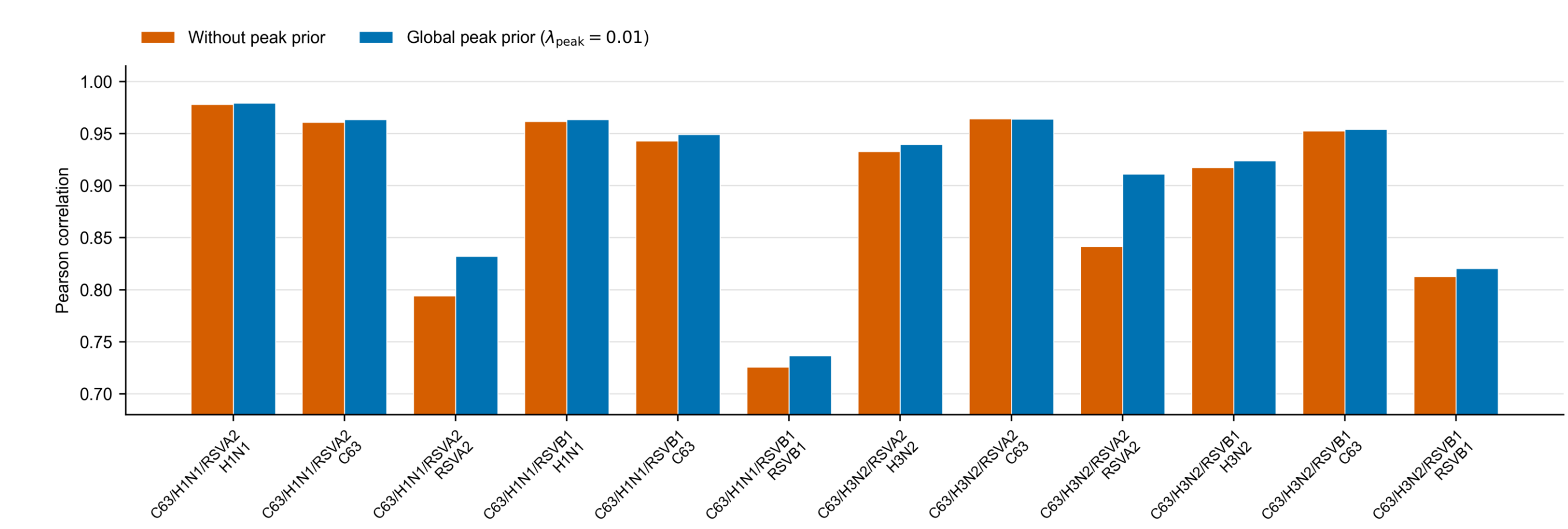


Figure 3. Effect of the shared peak prior.

Binary mean r: 0.9017 $\rightarrow$ **0.9302** · **Ternary mean r:** 0.8987 $\rightarrow$ **0.9115**

The identity-agnostic mask improves the most overlapped virus mixtures while preserving the same extraction framework.

ACKNOWLEDGMENT

The authors would like to acknowledge the support from USDA NIFA under grant 2023-67015-39237, the U.S. National Science Foundation under grants DMS-1925066, DMS-1903226, DMS-2124493, DMS-2311297, and DMS-2319279, and the U.S. National Institutes of Health under grant R01GM152814.