IOCN 2025 Conference

The 5th International Online Conference on Nanomaterials

22–24 September 2025 | Online

MDPI

Beyond Traditional airPLS: Improved Baseline Removal in SERS with Parameter-Focused Optimization and Prediction

Jiaheng Cui [a], Longtang Wang [b], Xianyan Chen [c], Pengju Yin [b*], Yiping Zhao [d*]

a School of Electrical and Computer Engineering, College of Engineering, The University of Georgia, Athens, GA, USA 30602

b School of Mathematics and Physics, Hebei University of Engineering, Handan, Hebei, China 056038

c Department of Epidemiology & Biostatistics, College of Public Health, The University of Georgia, Athens, GA, USA 30602

d Department of Physics and Astronomy, The University of Georgia, Athens, GA, USA 30602

INTRODUCTION & AIM

Baseline correction is a critical preprocessing step in Raman and surface-enhanced Raman spectroscopy (SERS) that directly impacts quantitative analysis accuracy. While the adaptive iteratively reweighted penalized least squares (airPLS) method [1] is widely used, it suffers from limitations when using default parameters: non-smooth baselines, large errors in broad peak regions, and difficulties with complex spectral features. Existing modified airPLS versions lack systematic parameter optimization, particularly for both critical parameters: λ (penalizing smoothness) and τ (convergence tolerance). Here, we present a machine learning approach [2] that addresses these limitations.

Aim: To develop a machine learning approach that automatically predicts optimal airPLS parameters from input spectra, eliminating manual tuning while significantly improving accuracy across diverse spectral conditions.

METHOD

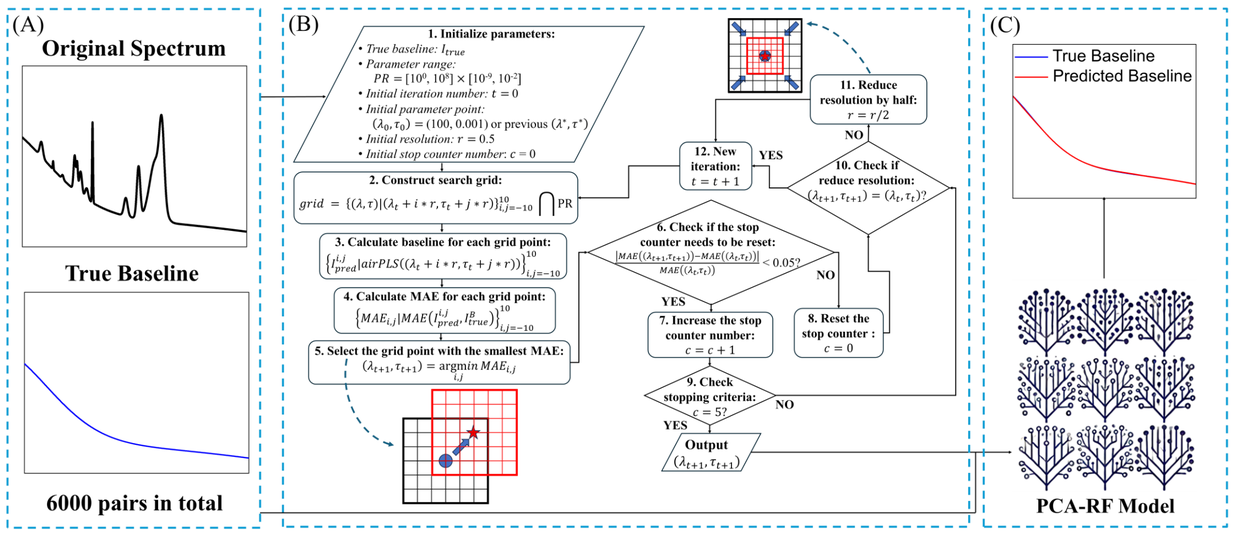

Figure 1. Overall workflow of our approach: (A) Generate synthetic spectral dataset with corresponding baselines; (B) grid search to identify optimal parameters for each spectrum; (C) machine learning training to predict optimal parameters.

Synthetic Dataset:

- 12 spectral shapes: 3 peak types & 4 baseline types.
- 500 spectra per spectral shape.

Grid Search Optimization (OP-airPLS):

- Identify optimal (λ, τ) parameter for each spectrum.
- Minimize mean absolute error (MAE) between predicted and true baselines across 500 spectra per spectral shape.
- Enforces smoothness for baselines.

Machine Learning Prediction (ML-airPLS):

- Stratified 8:1:1 dataset splits.
- PCA-Random Forest model (PCA–RF).
- Predict optimal parameters directly from spectral features.
- Eliminating true baseline requirements for real applications.

Performance Evaluation:

- Calculate percentage improvement (PI) of MAE between default airPLS (DP-airPLS), OP-airPLS, and ML-airPLS.

RESULTS & DISCUSSION

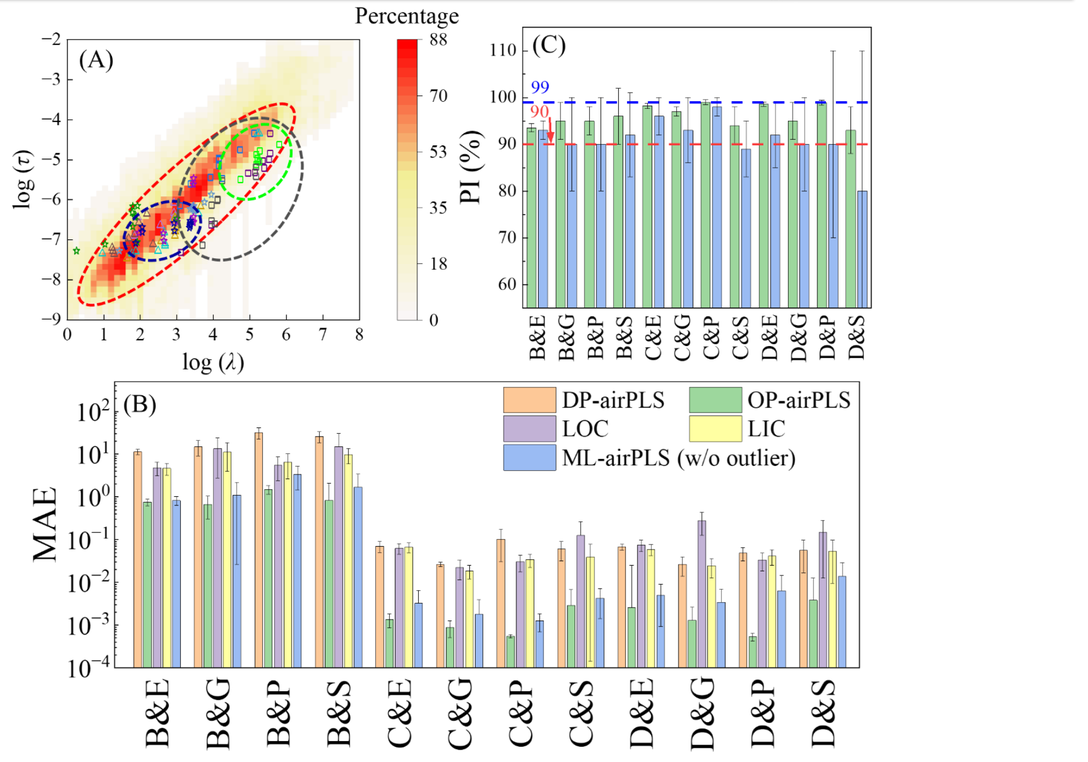

Figure 2. (A) Distribution of optimal (λ^*, τ^*) values for 12 spectral shapes in log-log scale. (B) Average MAE comparison: DP (orange), OP (green), and ML (blue) across spectral shapes. (C) PI of OP and ML.

Parameter Optimization Performance (Figure 2):

- 96 ± 2% PI across all 12 spectral shapes
- Reduced MAE by 1-2 orders of magnitude.
- Optimal (λ^*, τ^*) values cluster in linear diagonal region.

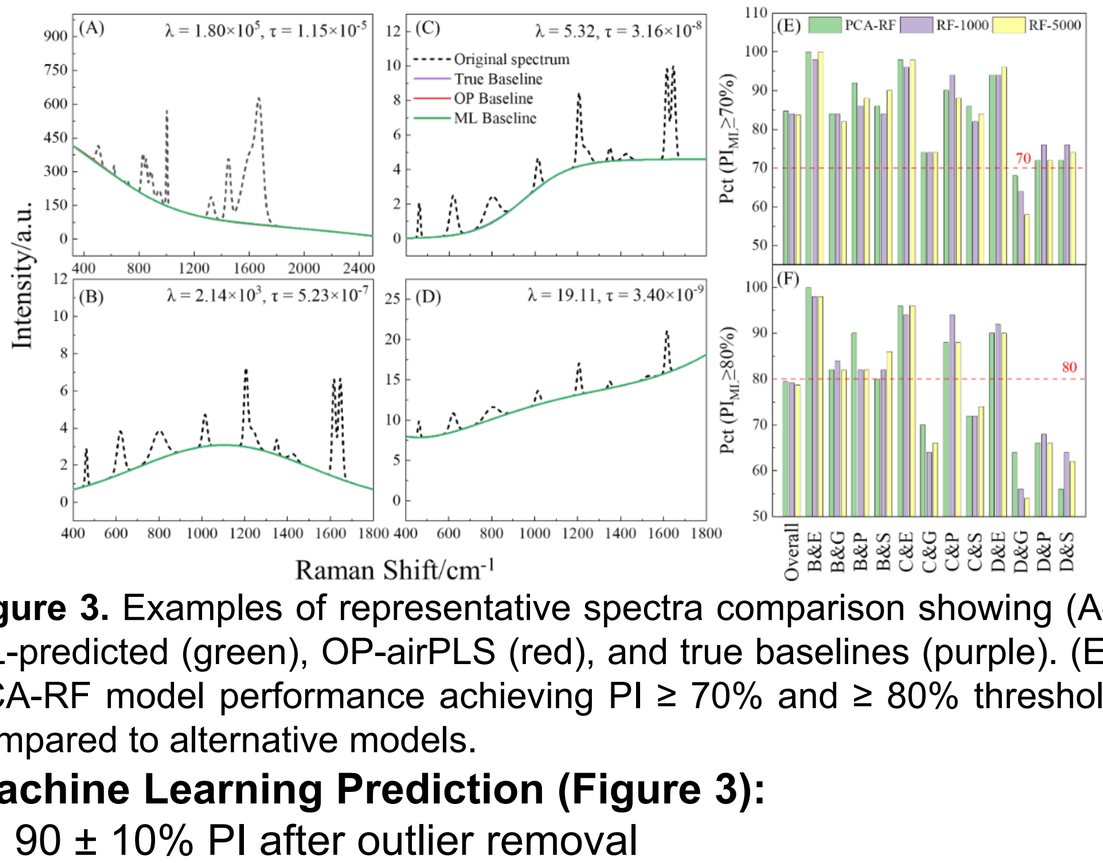

Figure 3. Examples of representative spectra comparison showing (A-D) ML-predicted (green), OP-airPLS (red), and true baselines (purple). (E-F) PCA-RF model performance achieving PI ≥ 70% and ≥ 80% thresholds, compared to alternative models.

Machine Learning Prediction (Figure 3):

- 90 ± 10% PI after outlier removal
- PI ≥ 70%: 84.7% test spectra; PI ≥ 80%: 79.5% test spectra
- ML-predicted baselines closely match OP-airPLS results.

CONCLUSION

- OP-airPLS systematically tuned (λ, τ), reducing MAE by **~96%**.
- PCA–RF directly predicted parameters, retaining **~90%** PI.
- ML computational efficiency stays approximately the same as DP-airPLS, from **0.013 s** → **0.038 s** per spectrum.

Acknowledgment: USDA NIFA Grant Num. 2023-67015-39237.

REFERENCES

[1] Zhang et al., Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst* **2010**, *135* (5), 1138-1146.

[2] Cui et al., Beyond Traditional airPLS: Improved Baseline Removal in SERS with Parameter-Focused Optimization and Prediction. *Anal. Chem.* **2025,** 97 (30), 16211-16218