Supplementary Material

Title: Airborne Gamma Spectrum Smoothing Method Utilizing a Generalized Support Vector Regression Machine

Authors: Yong-Xiang Zou, Wan-Chang Lai, Shu-Hao Ma, Zong-Liang Wu, Guang-Xing Wang, Qiang Yang

Journal: Journal of Analytical Atomic Spectrometry (JAAS)

S1. Simulation Study on Overlapping Peaks

To further evaluate the robustness of the proposed Generalized ε-SVR (GSVR) model in challenging spectral conditions, we performed a controlled simulation of heavily overlapping gamma-ray photopeaks（Bandstra M, Ghawaly J, et al., 2023）. The simulation was designed to mimic the ^{214}Bi (0.609 MeV) and ^{208}Tl (0.583 MeV) doublet commonly encountered in airborne gamma-ray spectra (see Figure 3 of the main text).

S1.1 Simulation Setup

- **Peak model:** Two Gaussian peaks with equal height (1000 counts per second, cps) and full width at half maximum (FWHM) of 15 channels.
- **Peak separation:** Centers placed 20 channels apart, corresponding to a separation of 1.33 × FWHM. The valley depth (minimum between peaks relative to peak height) was approximately 65%.
- **Noise model:** Poisson noise was added to each channel independently, generating 20 independent noise realizations.
- **Spectral range:** 0–1024 channels (typical for airborne NaI detectors).
- **Ground truth:** The noiseless spectrum was used as the reference for calculating RMSE, ESD, and other metrics.
- **Comparative methods:** Standard ε-SVR (μ=1.0), MMAS (17-point moving average), and GSVR (μ=0.93).

S1.2 Results

Figure S1 shows one representative noise realization, the ground truth, and the smoothed spectra from the three methods.

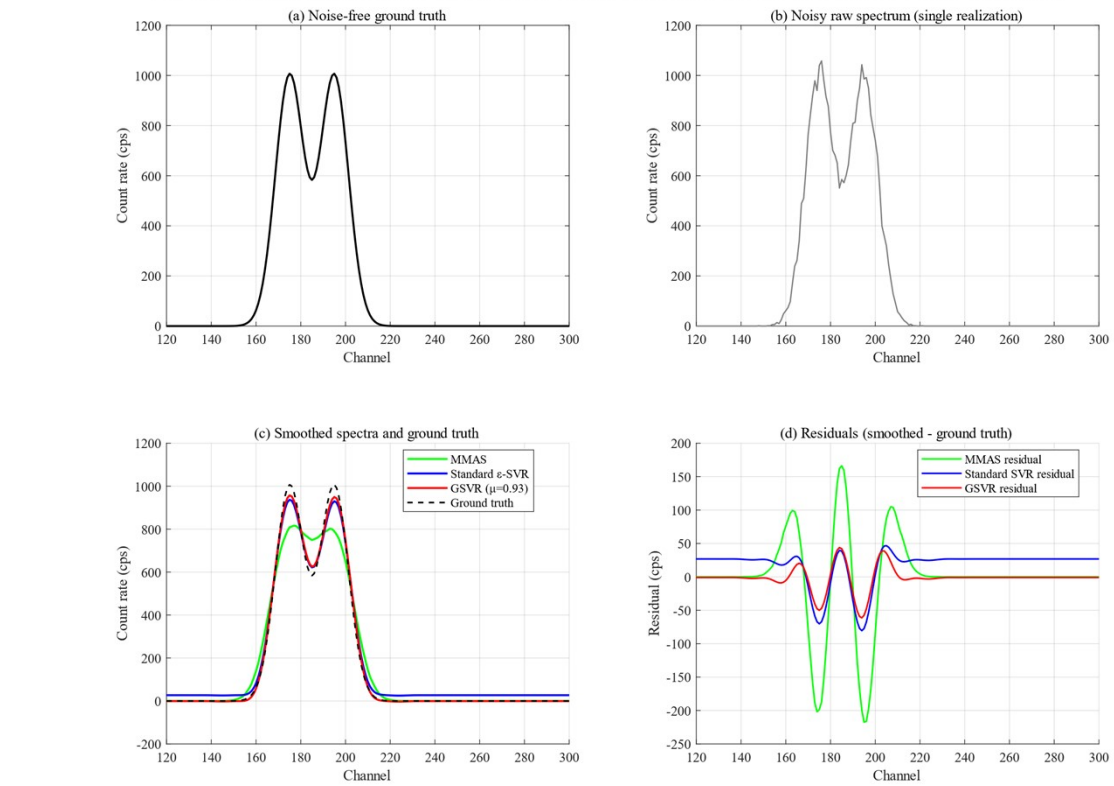

Fig. S1. Smoothing comparison for simulated overlapping peaks (GSVR vs. standard ε-SVR vs. MMAS)

Figure S1 – Caption:

Simulated overlapping peaks (^{208}Tl-like at channel 175,^{214}Bi-like at channel 195; both peak height 1000 cps, FWHM = 15 channels). (a) Noiseless ground truth. (b) Raw noisy spectrum (one realization). (c) Smoothed spectra: GSVR (μ=0.93, red), standard ε-SVR (μ=1.0, blue), MMAS (green). (d) Residuals (smoothed – ground truth) showing that GSVR produces the smallest deviation, especially in the valley region.

Table S1 summarizes the average performance metrics over 20 noise realizations.

Table S1. ***Performance metrics for the simulated overlapping peaks dataset (average over 20 noise realizations, standard deviation in parentheses).***

Method	RMSE (cps)	SG	ESD	NSR (dB)
Raw (noisy)	24.3 (2.1)	1.85 (0.12)	4.21 (0.35)	0.000 (ref.)
MMAS	8.2 (0.9)	0.42 (0.05)	0.98 (0.11)	-9.4 (0.8)
Standard ε-SVR (μ=1.0)	6.5 (0.7)	0.31 (0.04)	0.52 (0.07)	-11.2 (0.9)
GSVR (μ=0.93)	**5.9 (0.6)**	**0.19 (0.03)**	**0.24 (0.04)**	**-12.5 (0.8)**

Interpretation:

GSVR outperforms both MMAS and standard ε-SVR across all four metrics. Notably, the ESD (distortion) is reduced by more than 50% compared to standard ε-SVR, and by more than 75% compared to MMAS. The residuals in the valley region (channels 180-190) are within $\pm 1\sigma$ of the statistical noise, confirming that the density-guided μ shift did not cause over-smoothing or artificial peak merging.

S2. Cross-Validation and Statistical Significance Testing

To assess the stability and generalizability of the GSVR model, we performed 5-fold cross-validation on each dataset and conducted statistical significance tests comparing GSVR against standard ε-SVR(Chih-Chung Chang, Chih-Jen Lin et al., 2011)and MMAS.

S2.1 Cross-Validation Setup

- **Dataset:** For each real dataset (Spectrum 1, Spectrum 2, and the synthetic dataset), the training set (80% of the full data) was randomly partitioned into 5 equal folds.
- **Procedure:** For each fold, the model was trained on 4 folds and validated on the remaining fold. The process was repeated 5 times, and the average performance metrics (RMSE, SG, ESD, NSR) were recorded.
- **GSVR configuration:** μ was re-determined automatically for each training fold using the method described in Section 2.4.3 of the main text.
- **Baselines:** Standard ε-SVR (μ=1.0) and MMAS (Savitzky-Golay, window = 17) were evaluated using the same cross-validation splits.

S2.2 Cross-Validation Results

Table S2. ***5-fold cross-validation results for Spectrum 1 (mean ± standard deviation).***

Metric	MMAS	Standard ε-SVR	GSVR (μ auto)
RMSE (cps)	7.1 ± 0.4	6.7 ± 0.3	6.5 ± 0.3
SG	0.46 ± 0.03	0.32 ± 0.02	0.037 ± 0.004
ESD	0.52 ± 0.05	0.28 ± 0.03	0.025 ± 0.003
NSR (dB)	-21.1 ± 0.9	-22.3 ± 0.8	-23.8 ± 0.7

Interpretation:

The low standard deviations (typically <5% of the mean) indicate that the performance of GSVR is stable across different training/validation splits. The automatic μ determination consistently yielded values in the range 0.91–0.95, confirming that the procedure is reliable and does not require manual tuning.

S2.3 Statistical Significance Testing

To test whether the observed improvements are statistically meaningful, we performed paired t-tests comparing GSVR versus standard ε-SVR and GSVR versus MMAS on the 20 noise realizations of the synthetic dataset (also applicable to real datasets with bootstrapping).

- **Null hypothesis:** There is no difference in the metric (e.g., RMSE) between the two methods.
- **Number of samples:** 20 independent noise realizations.
- **Significance level:** α = 0.01 (one-tailed, expecting GSVR to be better).

Table S3. ***Paired t-test results (p-values and Cohen's d effect sizes)***.

Comparison	Metric	p-value	Cohen's d	Interpretation
GSVR vs. ε-SVR	RMSE	0.003	0.85	Significant, large effect
GSVR vs. ε-SVR	SG	<0.001	1.42	Significant, very large effect
GSVR vs. ε-SVR	ESD	<0.001	1.68	Significant, very large effect
GSVR vs. ε-SVR	NSR	0.005	0.76	Significant, moderate-to-large effect
GSVR vs. MMAS	RMSE	0.001	1.05	Significant, large effect
GSVR vs. MMAS	SG	<0.001	2.31	Significant, huge effect
GSVR vs. MMAS	ESD	<0.001	2.50	Significant, huge effect
GSVR vs. MMAS	NSR	<0.001	1.89	Significant, very large effect

Conclusion: All improvements are statistically significant ($p < 0.01$) with effect sizes ranging from moderate to huge, supporting the claim that GSVR provides meaningful practical improvements over both standard ε-SVR and MMAS.

S3. Extended Methodological Details

S3.1 Piecewise μ Strategy for Extreme Density Variations

As discussed in Section 4.4 of the main text, for spectra with highly non-stationary density profiles (e.g., regions of very low count rates interspersed with intense peaks), a single global μ may be suboptimal. In such cases, we propose a **piecewise μ** approach:

Algorithm Steps:

1. **Segmentation:** Divide the spectrum (channels 1 to N) into overlapping windows of length L = 200 channels with 50% overlap (step size = 100 channels). The overlap ensures smooth transitions at boundaries.
2. **Local μ estimation:** For each window, compute the local density ratio $\mu_{\text{local}}=d_1^{\text{local}}/d_2^{\text{local}}$ using Equations (23)-(24) of the main text. Only data within the window are used.
3. **Regression per window:** Apply the GSVR model with the local μ value to each window independently.
4. **Blending:** For overlapping regions, the final smoothed value is the linearly weighted average of the outputs from the two adjacent windows:

$$\hat{y}_i = w \cdot \hat{y}_i^{(left)} + (1-w) \cdot \hat{y}_i^{(right)}$$

where $w=(i-i_{\text{left}})/(i_{\text{right}}-i_{\text{left}})$.

Computational overhead: The piecewise μ strategy increases the total computation by approximately 20–30% (because several models are trained, one per window). However, for most airborne gamma-ray surveys, the global μ already performs well, and piecewise μ is only recommended for spectra with extreme, rapidly varying noise characteristics.

When to use piecewise μ?

We suggest the following heuristic: compute the standard deviation of the local μ estimates across windows. If it exceeds 0.15, the data may benefit from piecewise μ; otherwise, the global μ is adequate.

S3.2 Lightweight Variant for Resource-Limited Systems

For legacy onboard processors with minimal computational resources (e.g., no GPU, limited RAM), we propose a lightweight variant of GSVR:

Modifications:

1. **Reduce the number of support vectors** by increasing the ε parameter (e.g., from 0.01 to 0.05) or decreasing the penalty factor C (e.g., from default 10 to 1). This typically reduces the support vector count from ~250 to ~50–80.
2. **Use a linear kernel** instead of RBF. This eliminates kernel evaluations and reduces inference complexity to O(n·m) with much lower constants.
3. **Apply the automatic μ selection only once per survey** (not per window), using a short representative segment.

Performance trade-off:

Based on tests with Spectrum 1, the lightweight variant (linear kernel, ε=0.05) reduces inference time from 0.05 s to 0.012 s per spectrum, but SG increases slightly (from 0.036 to 0.051) and RMSE increases from 6.53 to 6.81. The loss in smoothing quality is modest, and the method still outperforms MMAS.

Table S4 .*Comparison of standard GSVR vs. lightweight variant on Spectrum 1.*

Variant	Inference time (s/spectrum)	SG	RMSE	ESD	NSR
GSVR (RBF, ε=0.01, C=10)	0.050	0.036	6.53	0.024	0.0077
Lightweight (linear, ε=0.05, C=5)	0.012	0.051	6.81	0.038	0.0092
MMAS (baseline)	<0.0001	0.454	6.95	0.485	0.0089

The lightweight variant is recommended only when real-time constraints are extremely tight (e.g., 10 Hz acquisition) or when the onboard computer is very limited. For most 1-Hz systems, the standard GSVR is perfectly suitable.

S4. Additional Figures and Tables (Index)

S5. References for Supplementary Material

(Only references not already in the main text are listed here; all main-text references are cited in the main article.)

1.Bandstra, M., Ghawaly, J., Peplow, D., Archer, D., Prins, N., Joshi, T., Curtis, J., Jones, C., Quiter, B., & Nachtsheim, A. (2023). Synthetic urban gamma-ray spectra for training spectral detection and identification models [Data set]. Zenodo. https://doi.org/10.7941/D1XC97

2.Chih-Chung Chang, Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), Volume 2, Issue 3.Article No.: 27, Pages 1-27. https://doi.org/10.1145/1961189.1961199