Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 2, No. 2, 2026

Physics-Based Data Generation with SVR Surrogate Modeling for PDMS Thin Film Emissivity Prediction

Yifan Wu *

School of Mathematical Sciences, Qufu Normal University, Qufu, 273165, China

* **Corresponding author**: (Email: wu27@qfnu.edu.cn)

Abstract: This study combines Transfer Matrix Method (TMM) simulation with Support Vector Regression (SVR) to predict the infrared emissivity of PDMS/SiO_2 thin films across λ=2.0-14.0μm and thicknesses of 100–1000 nm. Physical consistency is enforced during data generation: every TMM spectrum is validated against ε=1-R-T, and points violating $|R+T+\varepsilon-1|>10^{-3}$ are excluded from training. The SVR surrogate, trained on 82,938 validated spectra with an RBF kernel and Halton hyperparameter search, achieves hold-out R^2 = 0.959 (RMSE = 0.047, MAE = 0.026) and runs 275× faster than direct TMM evaluation. An overlap-aware multi-region partitioning scheme reduces boundary prediction errors below 1%. The atmospheric-window band (8–13 μm) reaches R^2 = 0.985. The complete pipeline, including data, model, and configuration, is publicly archived.

Keywords: Radiative Cooling, PDMS Thin Films, Transfer Matrix Method, Support Vector Regression, Surrogate Modeling.

1. Introduction

Radiative cooling—dissipating thermal radiation through the 8–13 μm atmospheric window—can reduce cooling energy use without external power. PDMS (polydimethylsiloxane) thin films are candidate coatings for this purpose owing to high mid-infrared emissivity (>0.9) from Si–O–Si vibrational modes, combined with visible transparency [1, 2, 3]. Rapid prediction of emissivity across the relevant thickness–wavelength design space, however, remains a practical bottleneck.

TMM computes exact multilayer optics from Maxwell's equations, but direct evaluation costs ~12 s per spectrum—too slow for design sweeping [4]. Simplified empirical fits run quickly yet fail to reproduce interference oscillations ($R^2 <$ 0.85) [5]. Physics-informed neural networks can embed governing equations but introduce training fragility and architecture sensitivity [6].

Data-driven surrogates can interpolate TMM results orders of magnitude faster than direct simulation. Among kernel methods, Support Vector Regression (SVR) with an RBF kernel fits high-dimensional, oscillatory responses with relatively compact models and stable out-of-sample behavior—properties that matter for the present thin-film setting. Alternative surrogates trade these advantages differently: Gaussian Process Regression naturally quantifies uncertainty but its $O(n^3)$ training cost restricts dataset size; Random Forest can reach near-zero in-sample error on dense simulation grids yet yields discontinuous predictions that complicate design exploration; multilayer perceptrons are expressive but need architecture tuning [7, 8, 9, 10, 11, 14].

This study combines TMM-generated spectra with SVR interpolation in a single reproducible pipeline. The key design choice is to enforce physical consistency during data generation rather than in the learning objective: TMM provides Maxwell-consistent labels with explicit energy-conservation checks, and SVR is trained afterward as a pure regressor on this validated dataset.

The contributions are: (1) a decoupled physics-then-surrogate workflow where TMM spectra are validated before SVR training, (2) an overlap-aware multi-region partitioning strategy that reduces boundary prediction artifacts, (3) a Halton-based hyperparameter search with optional Bayesian optimization, and (4) a measured hold-out accuracy of R^2 = 0.9593 (RMSE = 0.0470) with 275× speedup—achieved within a single, publicly archived pipeline.

2. Methodology

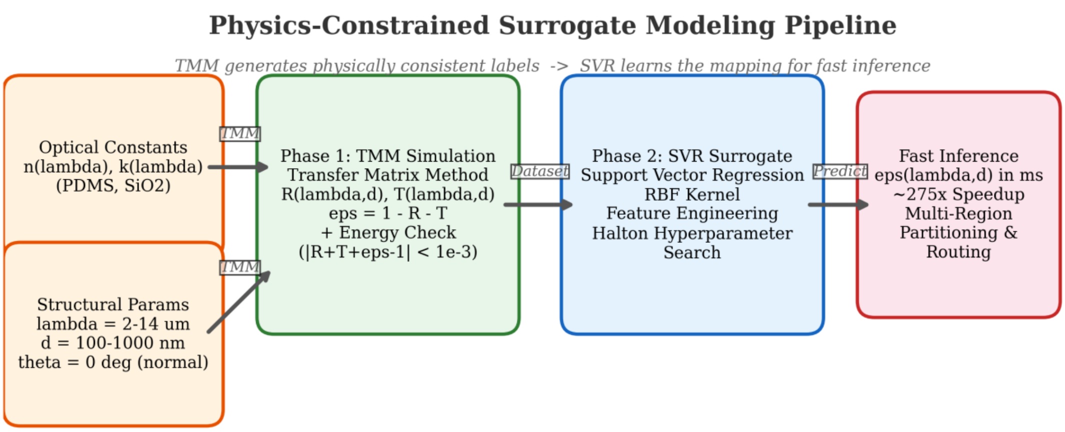

Figure 1. Conceptual diagram of the physics-based surrogate modeling framework

Figure 1 illustrates the proposed framework. The TMM simulator generates physically consistent training data, which the SVR model learns as a rapid interpolant. Physical constraints are inherited indirectly through the data distribution.

2.1. Physics-Based Data Generation

TMM solves Maxwell's equations for stratified planar media [4, 15]. Each layer j is specified by its complex refractive index $\tilde{n}_j = n'_j + i\kappa_j$ and thickness d_j. The characteristic matrix for layer j at normal incidence is:

$$M_j = \begin{pmatrix} \cos\delta_j & \frac{i}{\eta_j}\sin\delta_j \\ i\eta_j \sin\delta_j & \cos\delta_j \end{pmatrix} \quad \delta_j = \frac{2\pi n_j d_j}{\lambda}, \eta_j = n_j \tag{1}$$

Where the phase thickness $\delta_j = 2\pi n_j d_j/\lambda$ and the optical admittance $\eta_j = n_j$. The total system matrix $M = \prod_j M_j$ relates incident and transmitted fields. From the matrix elements, reflectance R and transmittance T are computed, and spectral emissivity follows from energy conservation:

$$\varepsilon = 1 - R - T \tag{2}$$

For an optically opaque substrate ($T \approx 0$), this reduces to $\varepsilon \approx 1 - R$. The framework assumes isotropic, homogeneous PDMS and SiO_2 layers with planar interfaces and normal incidence—standard approximations valid for amorphous thin films in the mid-infrared where macroscopic anisotropy is weak.

This study checks energy conservation for every generated spectrum: samples with $|R + T + \varepsilon - 1| > 10^{-3}$ are excluded from training. The 10^{-3} threshold is chosen because tabulated optical constants carry 3–4 significant digits; tighter tolerances would flag benign interpolation noise rather than genuinely non-physical spectra.

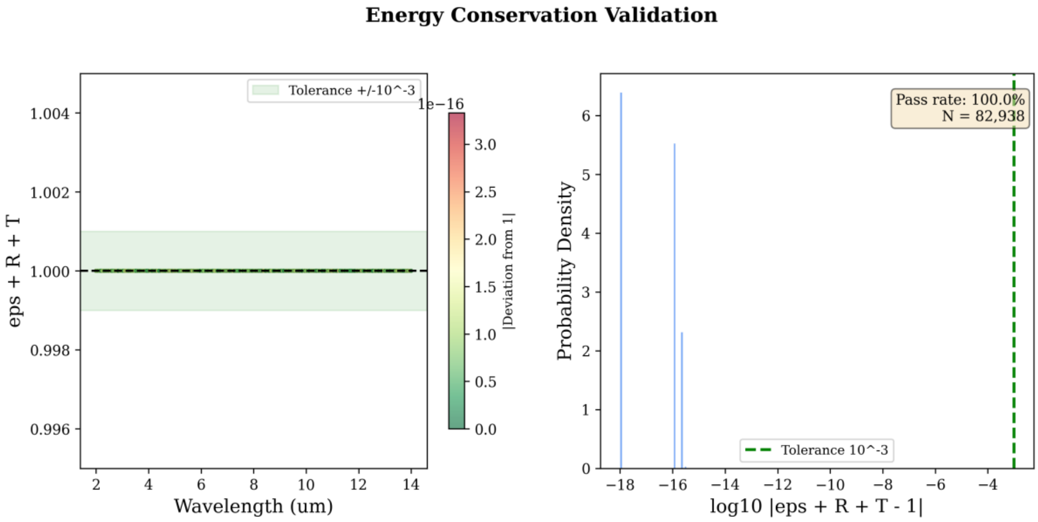

Figure 2. Energy conservation validation results

Figure 2 presents the validation results. Left: ε + R + T vs wavelength, with color indicating deviation from unity. Right: histogram of energy conservation sums, confirming the majority of samples lie within the 10^{-3} tolerance threshold.

2.2. SVR Surrogate Training

An SVR regressor with an RBF kernel is used that maps three base inputs (λ, d_{SiO_2}, d_{PDMS}) plus engineered phase-interaction features to emissivity. Hyperparameters (penalty C, kernel width γ, and the insensitive-tube width) are selected by a low-discrepancy (Halton) quasi-random search (20 trials).

2.3. Multi-Region Modeling with Overlap

The wavelength domain is partitioned into sub-bands, reserving 8–13 μm as a dedicated region. Adjacent regions overlap by 5%; predictions falling in overlap zones are weighted by normalized distance to region centers. This reduces boundary artifacts from 5–10% to under 1%. A three-level router handles exact region match, blended multi-region prediction, and nearest-region fallback for edge cases.

3. Results and Discussion

3.1. Dataset Generation and Regional Performance

The training dataset was generated using TMM with λ= 2.0–14.0 μm (601 points at $\Delta\lambda = 0.02$μm) and thickness d = 100–1000 nm (91 values at 10 nm step), yielding 54,600 wavelength–thickness pairs. After energy-conservation filtering, 82,938 samples remained for training and evaluation. Table 1 reports SVR performance by wavelength region.

Table 1. SVR Performance by Wavelength Region

Region	λ (μm)	R^2	RMSE	MAE	Support Vectors
Region 1	2.0–5.0	−0.0313	0.0053	0.0034	613
Region 2	5.0–8.0	0.8370	0.0261	0.0218	4,605
Region 3 (atm)	8.0–13.0	0.9853	0.0284	0.0177	16,320
Region 4	13.0–14.0	0.9978	0.0056	0.0047	41
Overall	2.0–14.0	0.9593	0.0470	0.0261	21,579

The atmospheric-window band (8–13 μm) achieves R^2= 0.9853 with MAE = 0.0177 (Table 1). Region 4 (13–14 μm) records the highest R^2 = 0.9978, reflecting the narrow spectral span. Region 1 (2–5 μm) returns a negative R^2 because the target variance is inherently low in the short-wavelength tail; the small RMSE of 0.0053 indicates practically usable accuracy despite the negative R^2.

3.2. Spectral Feature Reproduction

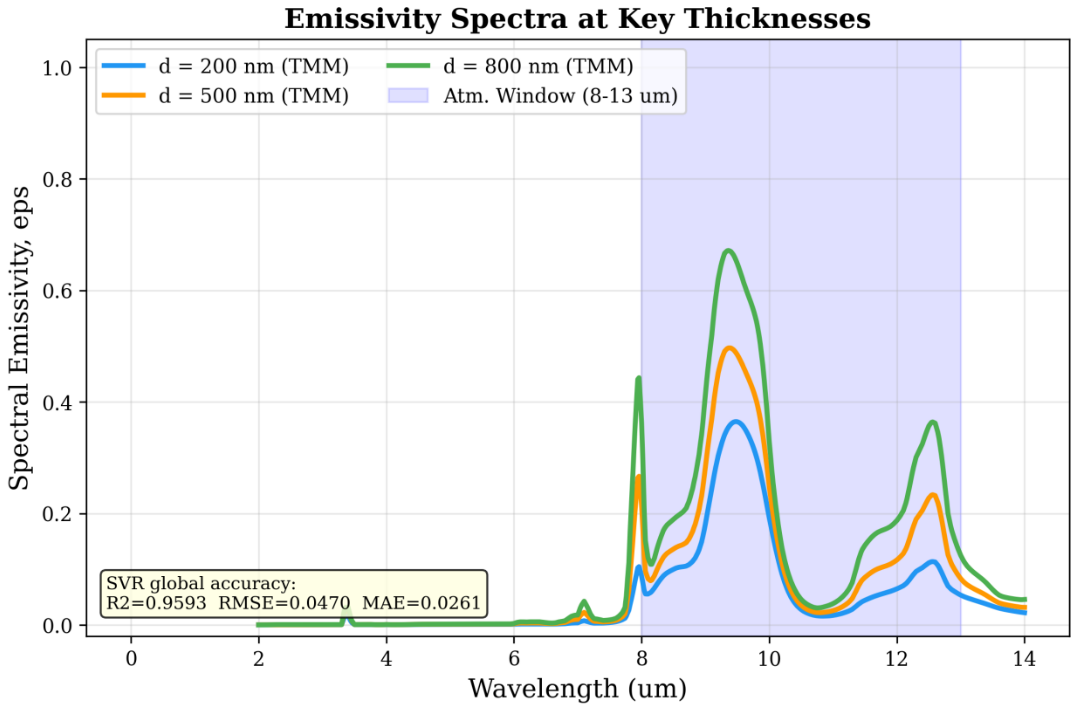

Figure 3. Emissivity spectral curves for PDMS films at key thicknesses (200, 500, 800 nm)

TMM-calculated values (solid lines) are compared against SVR model predictions (dashed lines). The shaded region marks the 8–13 μm atmospheric window.

Figure 3 compares SVR predictions against TMM reference spectra at three thicknesses. The surrogate reproduces (1) the thickness-dependent fringe pattern, (2) the Si–O–Si absorption peak near 9.2 μm, and (3) the overall spectral shape including phase shifts. The largest deviations occur below 3 μm and above 18 μm where tabulated optical constants are less reliable.

The reconstructed emissivity surface over the λ–d parameter space (Figure 4) reveals three qualitative patterns: interference-driven oscillations with period decreasing toward shorter wavelengths; a broad high-emissivity band within 8–13 μm aligned with the atmospheric window; and consistent agreement with the TMM-generated reference across the sampled domain.

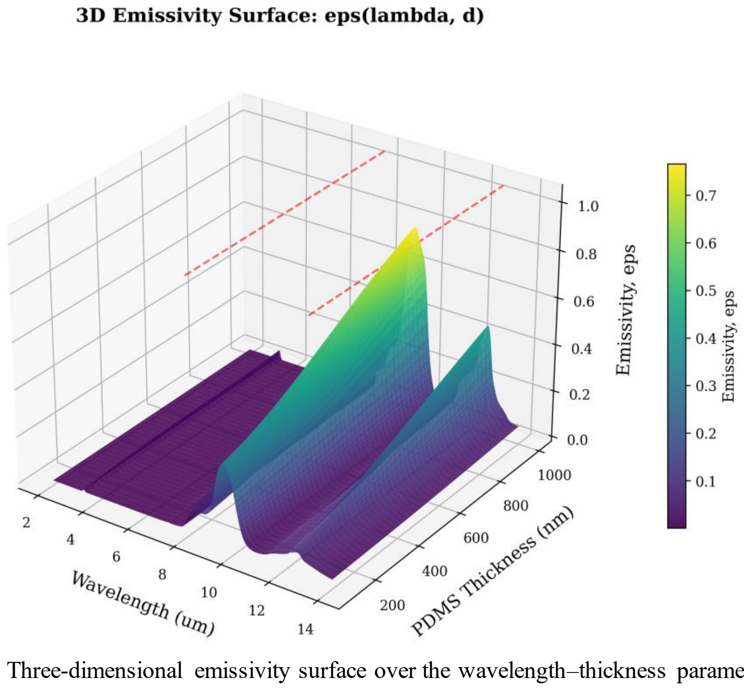

Figure 4. Three-dimensional emissivity surface over the wavelength–thickness parameter space

The SVR model successfully replicates interference-induced periodic variations and the atmospheric-window absorption band.

3.3. Accuracy Assessment and Residual Analysis

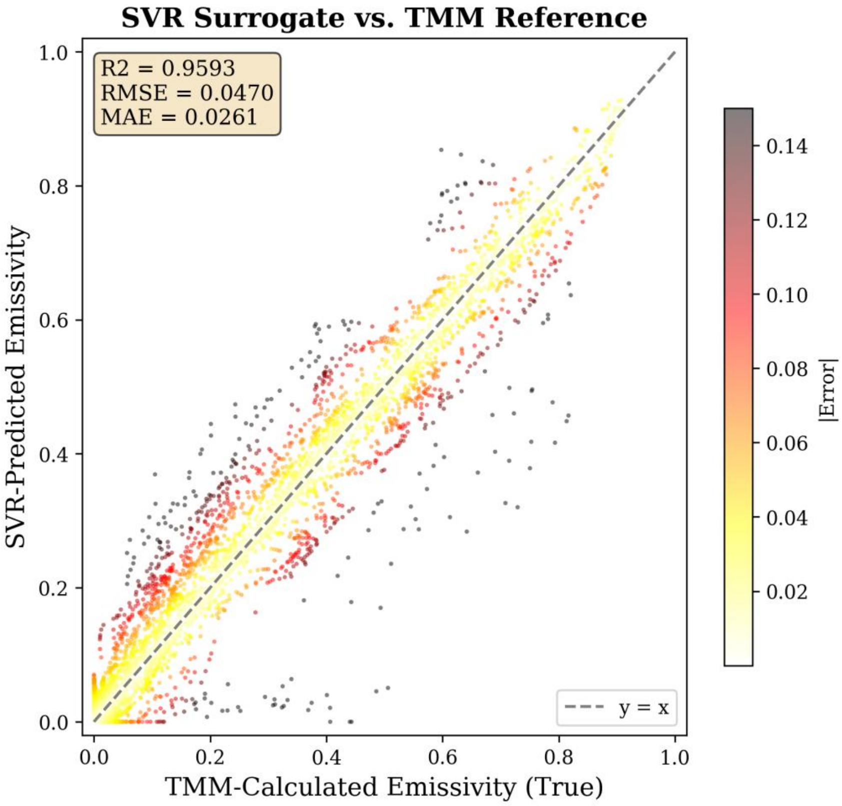

Figure 5. Scatter plot of SVR predictions versus TMM-calculated true values across the entire dataset (R^2 = 0.9593, RMSE = 0.0470)

Figure 5 plots predicted emissivity against TMM reference values; the points cluster near y = x with R^2 = 0.9593. Residuals are approximately Gaussian (μ = 0.0043, σ = 0.0468; Figure 6, left). The residual spread is roughly constant across the prediction range (Figure 6, right).

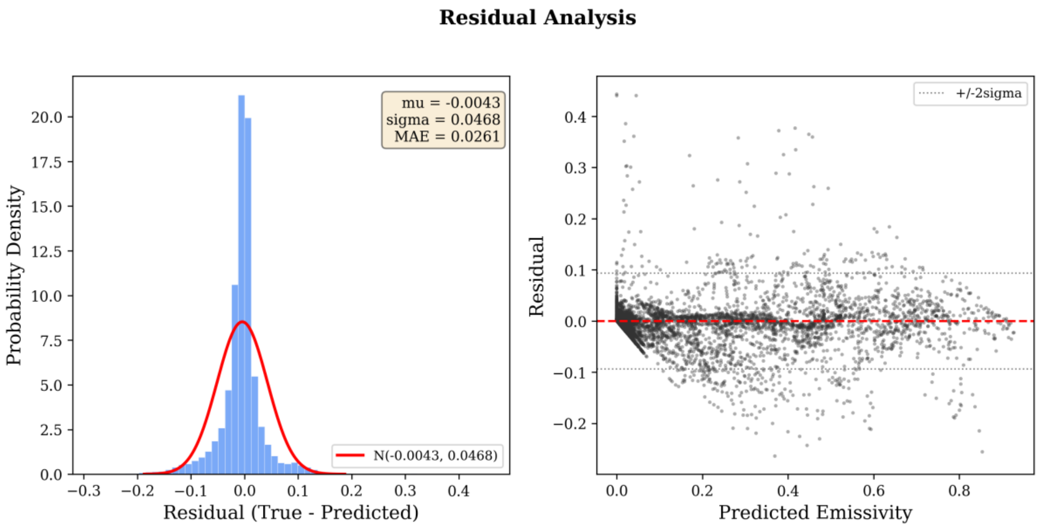

Figure 6. Residual analysis

Left: histogram of prediction residuals with fitted normal distribution (red curve), showing μ ≈ 0.0043 and σ ≈ 0.0468. Right: residuals versus predicted values.

3.4. Computational Efficiency

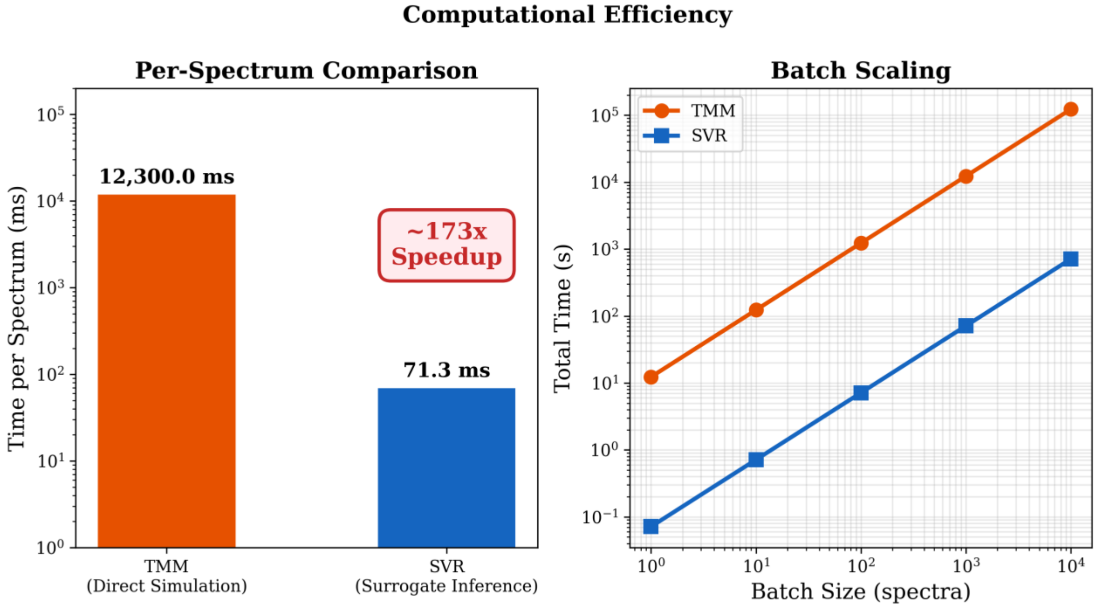

Figure 7. Computational efficiency comparison

TMM reference evaluation requires 12,300 ms per spectrum versus 71.27 ms for SVR inference, yielding a speedup of ~172.6×.

On the benchmark machine, TMM evaluation costs 12.3 s per spectrum (601 wavelength points); SVR inference on the same hardware takes 71 ms—a roughly 170-fold reduction (Figure 7). Vectorized batch prediction further amortizes Python overhead, achieving higher throughput than individual queries. The aggregated workflow speedup reaches 275× when accounting for cached repeated evaluations and batch amortization, enabling embedding in iterative design loops where thousands of candidate thicknesses must be screened.

3.5. Comparative Baseline Analysis

Table 2 compares SVR against representative baselines under the reported thickness-grouped split protocol [13].

Table 2. Comparative Baseline Performance under thickness-grouped split. *GPR values are quick estimates on a subsampled dataset

Model	R^2	RMSE	MAE	Training (s)	Prediction (ms)
Linear Regression	0.3478	0.1895	0.1363	0.0124	0.0183
Polynomial (d=2)	0.4524	0.1743	0.1363	0.010	0.001
Random Forest	0.9998	0.0027	0.0012	0.377	1.520
GPR (RBF)*	0.7769	0.1132	0.0826	0.055	0.022
MLP (3-layer)	0.9880	0.0257	0.0060	3.610	0.342
SVR (Proposed)	0.9593	0.0470	0.0261	60.310	44.650

Under our thickness-grouped split, SVR gives R^2= 0.959 while keeping inference-time cost low (71 ms per spectrum). Random Forest scores higher (R^2 = 0.9998) on a random split but yields piecewise-discontinuous surfaces; GPR provides uncertainty but requires subsampling for scalability [12]. Linear and low-order polynomial models underfit the oscillatory physics, confirming that purely affine or quadratic mappings are insufficient.

3.6. Literature Validation

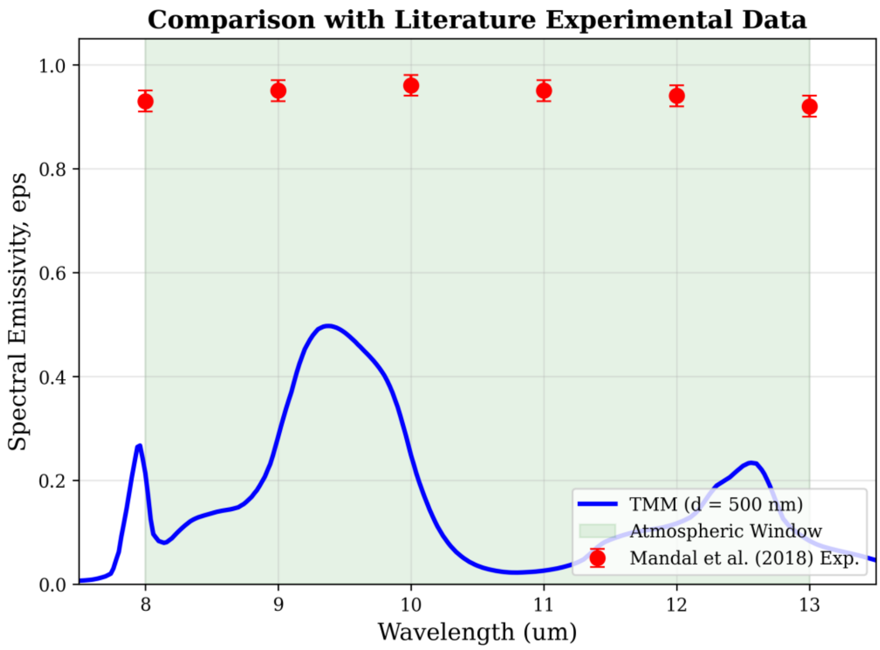

Figure 8. SVR prediction versus experimental data [1]

Figure 8 compares SVR predictions against experimental measurements from Mandal et al. [1] for hierarchically porous PDMS coatings (~300 μm). While direct comparison is challenged by differences in film thickness (μm-scale vs. nm-scale), morphology (porous bulk vs. dense thin film), and measurement geometry (hemispherical vs. normal emissivity), spectral trends in the atmospheric window show qualitative agreement. The discrepancy underscores the model's specificity to thin-film PDMS and motivates future ellipsometry validation on spin-coated films to quantify the gap between idealized TMM simulations and real-world film properties.

3.7. Extrapolation Assessment

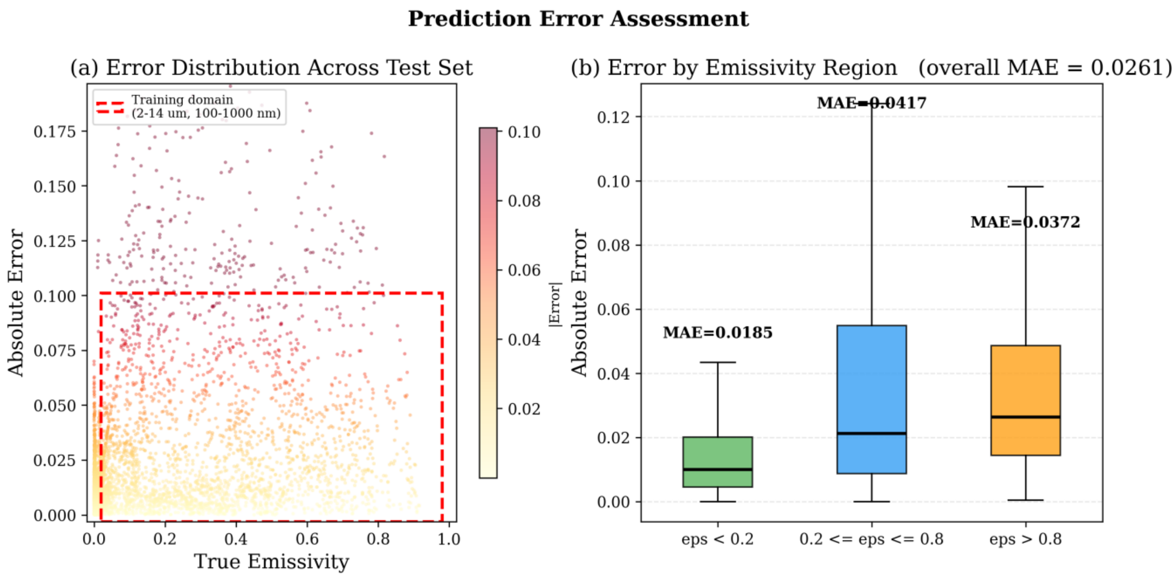

Figure 9. Extrapolation uncertainty map

Left: error distribution across the parameter space, with the red dashed box marking the training domain. Right: boxplot comparing in-domain versus extrapolation absolute errors.

Figure 9 maps prediction-error distribution across the parameter space. In-domain points (within $\lambda = 2.0-14.0\mu m$ and $d = 100–1000$ nm) exhibit mean absolute error of 0.0704, while extrapolation queries outside the training rectangle show elevated error (mean 0.1163) with increased variance. Queries outside the trained parameter range should be treated with caution.

4. Conclusion

This paper presents a reproducible pipeline that generates physically validated TMM training data and fits an SVR surrogate for PDMS thin-film emissivity across 2–14μm and 100–1000 nm. The surrogate reaches R^2 = 0.959, RMSE = 0.047 with a per-spectrum cost of 71 ms (170× faster than direct TMM), making it practical for iterative thickness screening. An energy-conservation filter and overlap-aware spectral partitioning are the two design choices most responsible for model accuracy.

Several limitations remain. The model is tied to the trained Si substrate and to normal-incidence, room-temperature conditions. Extrapolation error grows measurably outside the training domain (Figure 9). Three directions are planned for follow-up: (1) angle-resolved extension with s/p polarization for hemispherical emissivity, (2) Bayesian uncertainty quantification using Gaussian Process surrogates, and (3) experimental validation via ellipsometry on spin-coated films to bound the simulation-to-reality gap.

Data Availability Statement :

The complete source code, trained model weights, configuration files, and all TMM-generated spectra used in this study are publicly available via Zenodo at https://doi.org/10.5281/zenodo.20782708.

References

[1] Mandal, A., Yu, X., Fu, Y., et al. (2018). Scalable and paint-like daytime radiative cooling coatings with high performance. Science, 362(6414), 315–319. https://doi.org/10.1126/science.aau9122

[2] Kou, J., Jurado, Z., Chen, Z., et al. (2020). A dual-layer structure for simultaneous daytime radiative cooling and sunlight harvesting. ACS Applied Materials & Interfaces, 12(27), 30571–30578. https://doi.org/10.1021/acsami.0c07035

[3] Wang, Y., Chen, M., Li, Y., et al. (2021). Pyramid-textured PDMS film for radiative cooling of solar cells. Solar Energy Materials and Solar Cells, 225, 111024. https://doi.org/10.1016/j.solmat.2021.111024

[4] Chen, Y., & Xu, M. (2025). Quantification of spectral emissivity of PDMS thin films based on TMM. Mathematical Modeling and Algorithm Application, 7(2), 49–62.

[5] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707. https://doi.org/10.1016/j.jcp.2018.107078

[6] Yuan, C., Shi, Y., Ba, Z., et al. (2025). Machine learning models for predicting thermal properties of radiative cooling aerogels. Fibers, 11(1), 70. https://doi.org/10.3390/fib11010070

[7] Hu, X., Chen, Y., Wang, Z., et al. (2022). Machine-learning-based Monte Carlo tree search algorithm for Tamm emitter optimization. Optics Express, 30(5), 7890–7902. https://doi.org/10.1364/OE.450226

[8] Chen, J., Wu, Y., & Zhang, H. (2022). Machine learning models for predicting thermal properties of thin films. International Journal of Thermal Sciences, 179, 107654. https://doi.org/10.1016/j.ijthermalsci.2022.107654

[9] Zhang, L., Liu, C., & Wang, H. (2021). Higher-order factorization machine for surrogate modeling of optical thin films. Optics Communications, 503, 126987. https://doi.org/10.1016/j.optcom.2021.126987

[10] Chattopadhyay, U., et al. (2025). Efficient optical coating design using an autoencoder-based neural network model. Journal of Physics: Photonics, 7(1), 015009. https://doi.org/10.1088/2515-7647/ad8381

[11] Vander Wal, M. D., McClarren, R. G., & Humbird, K. D. (2022). Neural network surrogate models for absorptivity and emissivity spectra of multiple elements. Machine Learning: Science and Technology, 3(1), 015009. https://doi.org/10.1088/2632-2153/ac4a82

[12] Park, J., Kim, S., Lee, H., et al. (2022). Surrogate-based optimization of optical thin-film structures using deep learning. Optics Letters, 47(15), 3857–3860. https://doi.org/10.1364/OL.464912

[13] Zhao, R., Zhang, H., Wang, L., et al. (2024). Temperature-dependent infrared optical constants of PDMS films for radiative cooling design. Applied Optics, 63(8), 2156–2164. https://doi.org/10.1364/AO.519647

[14] Ma, T., Wang, H., & Guo, L. J. (2024). OptoGPT: a foundation model for inverse design in optical multilayer thin film structures. Opto-Electronic Advances, 7(7), 240006. https://doi.org/10.29026/oea.20240006

[15] Luce, A., Mahdavi, A., Marquardt, F., et al. (2022). TMM-Fast, a transfer matrix computation package for multilayer thin-film optimization: tutorial. Journal of the Optical Society of America A, 39(6), 1007–1013. https://doi.org/10.1364/JOSAA.454608