A data-driven, multi-scale sediment velocity model for Southern California

Yi Liu[1]; Grigorios Lavrentiadis[1]; Domniki Asimaki[1]

[1]Department of Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA 91125

1. Abstract

We present a non-parametric, data-driven near-surface velocity model for Southern California that can be used to populate the basin structures of SCEC CVM-S4.26. The model is developed as a conditional random field of the residuals relative to SCEC CVM-S4.26 expressed as a Gaussian process (GP). The GP kernel function is a composite of a stationary and spatially varying kernels to simulate both the average geological-based trends and site-specific variability. The model was trained on a large set of geotechnical measurements (both invasive and non-invasive methods), as well as a small set of sonic log data to constrain the average behavior at large depths.

Results indicate that the kernel function that integrates local and geology-scale information has optimal predictive ability among the models that only incorporate local or large-scale kernel functions we tested. By integrating statistical modeling with shallow geophysical information our model provides a robust, flexible, and interpretable solution for modeling shallow velocity structures in regions with broad engineering and scientific applications.

2. Data description

The VS profile combines 658 sets of profiles obtained from the web portal shear-wave velocity profile database (VSPDB) and 33 sets of sonic log profiles provided by Harvard University.

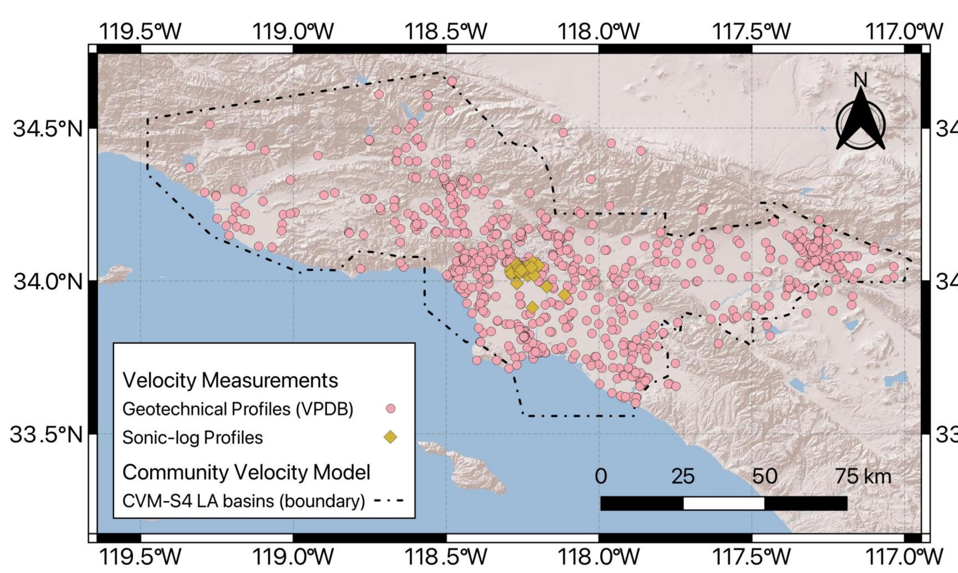

Figure 1: Locations of the profiles used to develop the model.

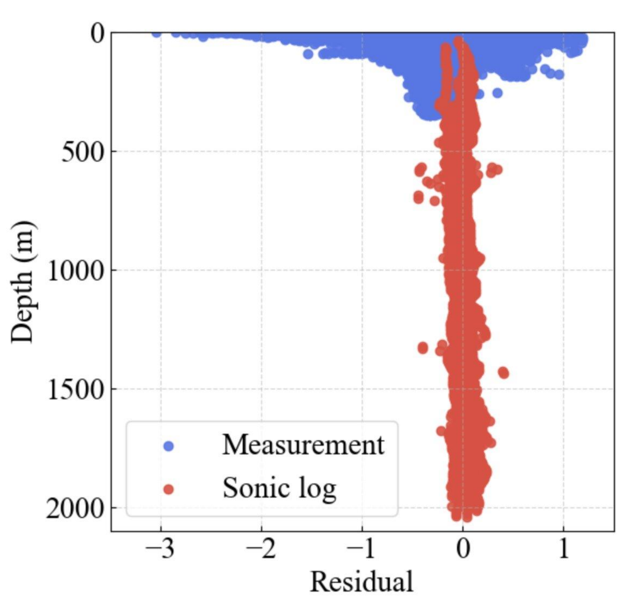

Figure 2: Residuals in the total datasets.

The response variable of the model is the residual between the shear-wave velocity measurements and CVM prediction defined as: $\boldsymbol{\delta = ln(V_S) - ln(V_S^{cvm})}$

3. Model development

We assume:

$$\delta \sim GP(0, k(X, X'))$$

with $X = [x, y, z, e, es, z_{2.5}]$, where x and y are the spatial coordinates, z is the depth, e is the elevation, e_s is the surface elevation, and $z_{2.5}$ is the depth where a shear wave velocity of 2.5 km/sec is reached.

$$k(x, x') = k_1(x, x') + k_2(x, x')$$

Stationary kernel:

$$k_1(x, x') = (k_{RBF}(z_{2.5}, e_s) + k_{const}) \times k_{RBF}(z)$$

Spatially varying kernel:

$$k_2(x, x') = k_{Matren}(x, y) \times k_{Matern}(e)$$

4. Training

We fitted the GP model using the training set and optimized the marginal likelihood function to obtain the optimal model hyperparameters.

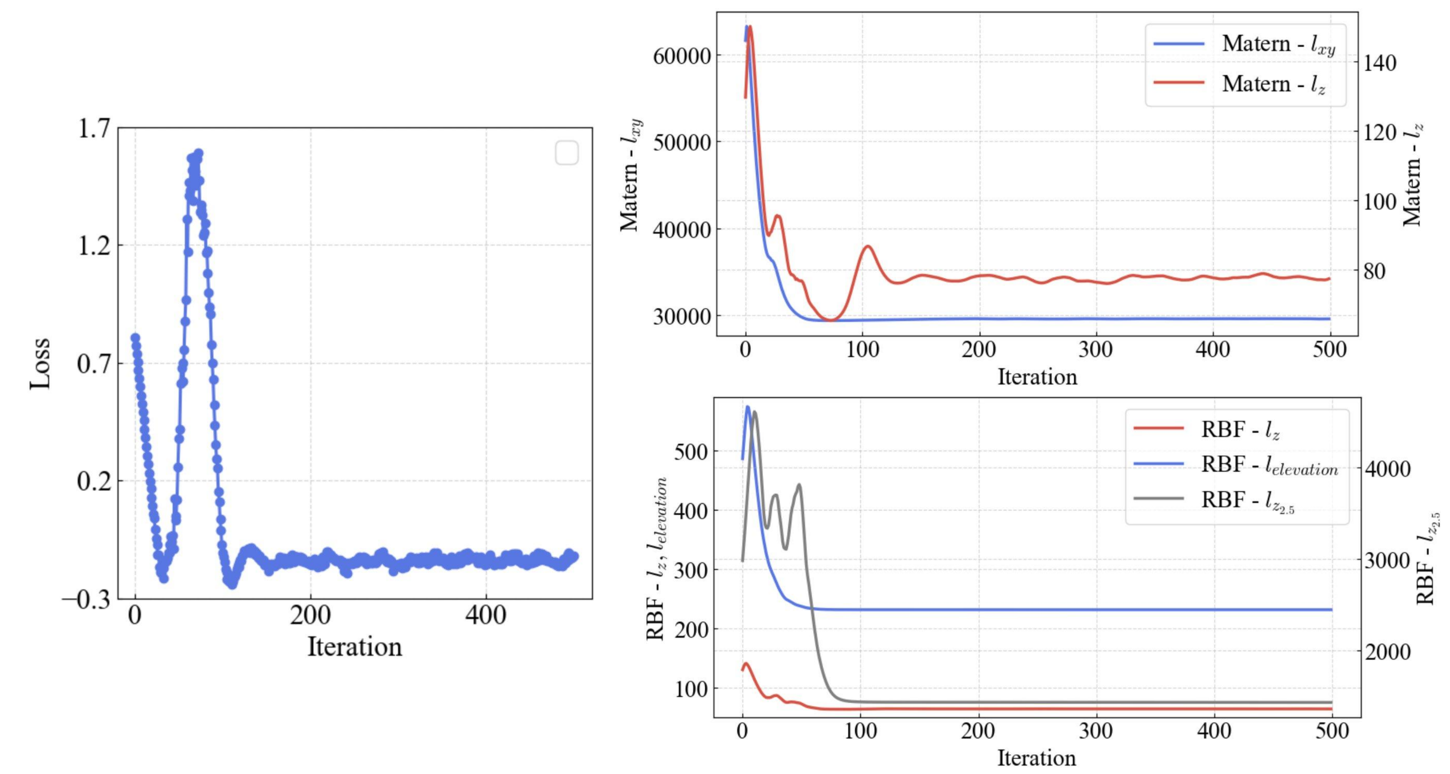

Figure 3: Loss function and hyper-parameters

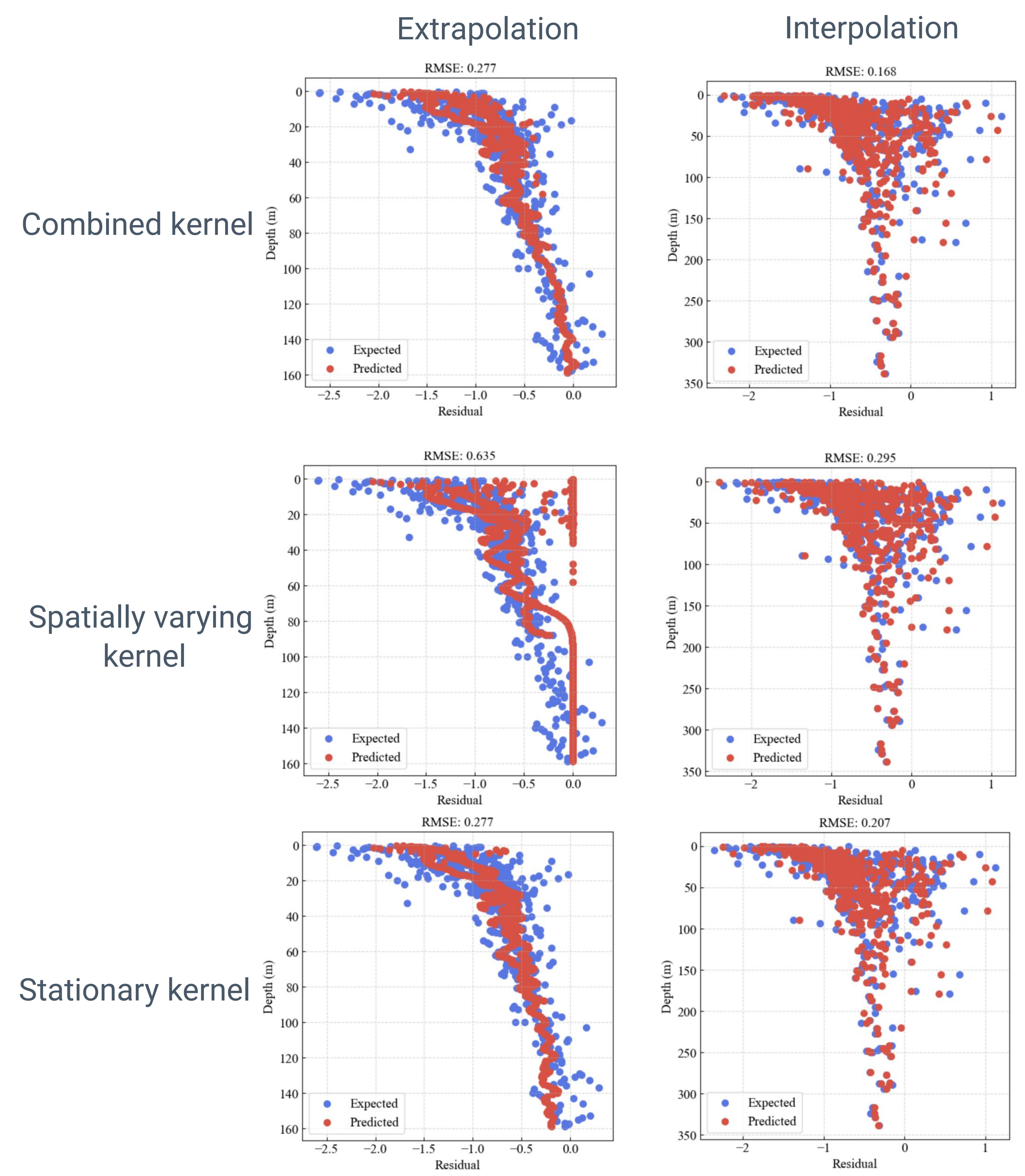

Figure 4: Model performance in test datasets.

5. Application

We selected six representative geological cross sections within the Los Angeles Basin as study areas for our model.

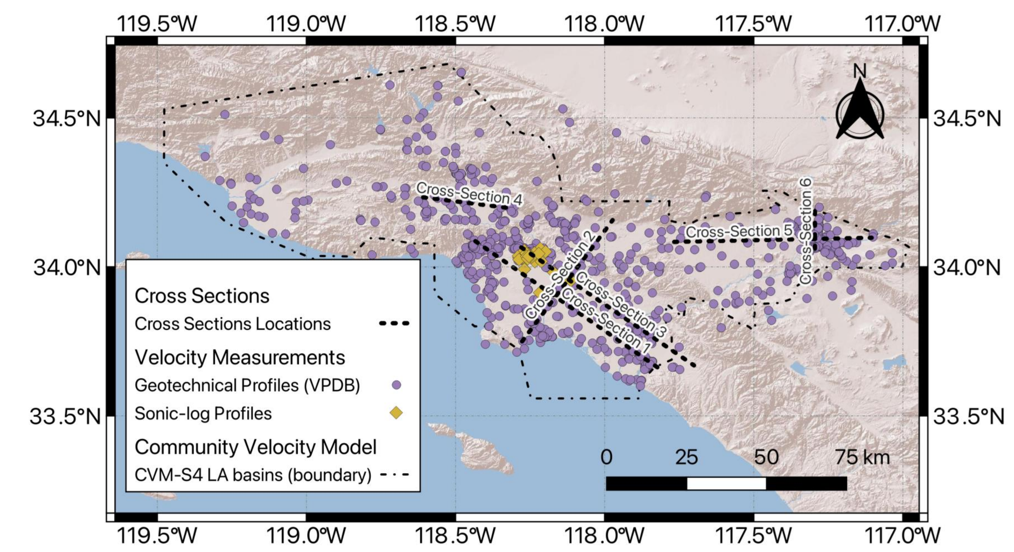

Figure 5: Cross sections in LA basin.

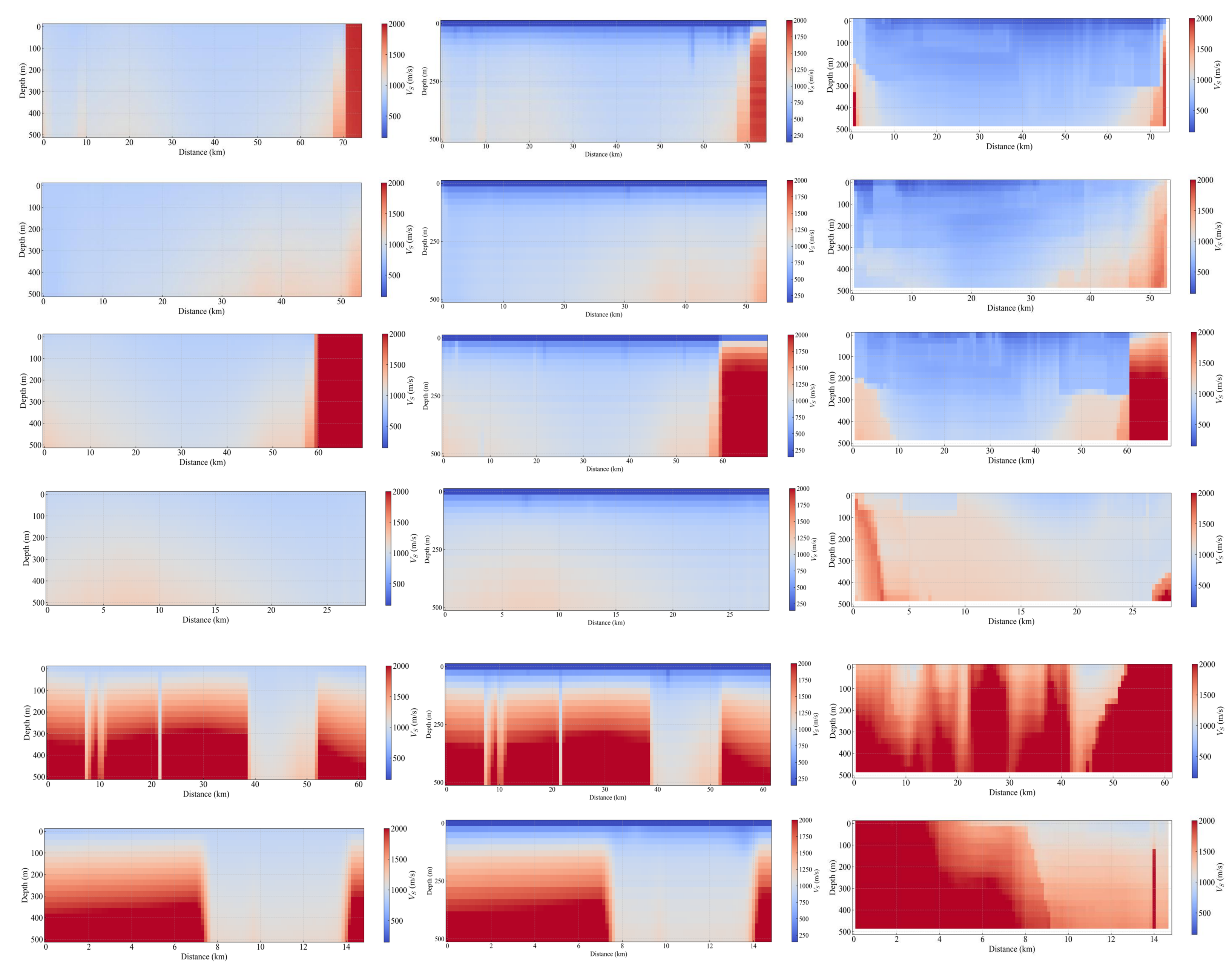

Figure 6: Comparison between CVM, GP model and GTL model.

6. Conclusion

This study focuses on the application of GP model in modeling shallow shear wave velocity in the Los Angeles Basin. We systematically formulated the entire process of data description, kernel function design, model training, testing verification, and visualization analysis. The results indicate that the selected model can effectively capture the spatial variability of shallow geological structures, and the predicted results show good adaptability and reasonableness in different geological environments.

U.S. National Science Foundation

science for a changing world

U.S. DEPARTMENT OF ENERGY

7. References Cited

J. Shi, D. Asimaki, From stiffness to strength: Formulation and validation of a hybrid hyperbolic nonlinear soil model for site-response analyses, BULLETIN OF THE SEISMOLOGICAL SOCIETY OF AMERICA 107 (3) (2017) 1336–1355.375.

G. Lavrentiadis, E. Seylabi, F. Xia, H. Tehrani, D. Asimaki, D. McCallen, Data-driven characterization of near-surface velocity in the San Francisco bay area: A stationary and spatially varying approach, EARTHQUAKE SPECTRA 41 (3) (2025) 2311–2334.

Acknowledgements

This research was supported by the Statewide California Earthquake Center (Contribution No.24116). SCEC is funded by NSF Cooperative Agreement EAR-2225216 and USGS Cooperative Agreement G24AC00072. Additional support was provided by DOE Grant DE-SC0016520.