Supplementary Materials for

A Deep Learning-Based Forward Surrogate Model for Accelerating Surface-Wave Dispersion Inversion

Qianyun Dong[1,2], Hongrui Qiu[*1,2], Jing Hu[*3], Yong Zheng[1,2,4], Sanxi Ai[1], Siyuan He[5], Lei Qin[1]

1. Hubei Subsurface Multi-Scale Imaging Key Laboratory, School of Geophysics and Geomatics, China University of Geosciences, Wuhan, 430074, China

2. State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan, 430074, China

3. School of Geological Engineering and Geomatics, Chang'an University, Xi'an, 710054, China

4. National Precise Gravity Measurement Facility, Huazhong University of Science and Technology, Wuhan, 430074, China

5. Hunan Earthquake Agency, Changsha, 410000, China

*Corresponding authors: Hongrui Qiu (qiuhongrui@cug.edu.cn, qiuhonrui@gmail.com), Jing Hu (hujing@chd.edu.cn)

Contents

Supplementary Text S1. Description of the Rayleigh-wave phase-velocity dispersion dataset of Shen et al. (2016).

Supplemental Table S1. Specific hyperparameters of the FGRU model.

Supplemental Table S2. Model parameters and their perturbation ranges.

Supplemental Table S3. Forward-modeling runtime comparison under different numbers of forward evaluations.

Supplemental Table S4. Order-of-magnitude comparison of average forward-modeling time per model (log10 scale) for FGRU and DISBA under different hardware configurations.

Supplemental Figure S1. Structure of GRU.

Supplemental Figure S2. Training and validation loss functions.

Supplemental Figure S3. Model parameterization schematic.

Supplemental Figure S4. Distribution of the training-set Vs model space.

Supplemental Figure S5. Relative error of FGRU forward modeling in South China.

Supplemental Figure S6. Parameter search bounds for PSO inversion.

Supplemental Figure S7. Synthetic data inversion results for the Tibet grid point (100°E, 34°N).

Supplemental Figure S8. Synthetic data inversion results for the NE China grid point (125°E, 47°N).

Supplemental Figure S9. Topography in South China.

Supplemental Figure S10. Real-data inversion results at the grid point (107°E, 30°N).

Supplemental Figure S11. Vertical cross sections of the Vs model along profile B–B′ (from 100°E, 36°N to 116°E, 23°N).

Supplemental Figure S12. Distribution maps of Chi-square misfit between the observed Rayleigh wave dispersion data and the theoretical dispersion curves.

Supplemental Figure S13. Probability (solid line) and cumulative (dashed line) density distributions for the χ.

Supplemental Figure S14. Comparison of sensitivity kernels computed by the FGRU forward operator and DISBA.

Text S1. The Shen et al. (2016) dataset is gridded at 0.5° × 0.5° and includes 980 grid points (blue dots in Figure S9) in South China. To maintain consistency with the surrogate forward operator, we used periods of 8–50 s, identical to those adopted in the training dataset. We quantified inversion uncertainty by performing 100 independent FGRU–PSO inversions at each grid point. Each FGRU–PSO run used 100 particles and 100 iterations and took ~2 s on an NVIDIA RTX 3060 GPU, for a total runtime of ~165 s per grid point and 1,000,000 forward evaluations. Under a comparable computational budget, each DISBA–PSO run used 50 particles and 50 iterations, and 40 independent DISBA–PSO inversions were performed at each grid point, yielding a total of 100,000 forward evaluations.

Table S1. Specific hyperparameters of the FGRU model

Hyperparameter / Setting	**Value / Description**
Data Split	80% for training, 20% for validation
Batch Size	512
Learning Rate	1×10^{-5}
Iterations (Epochs)	1000
Optimizer	Adam
Loss function	MSE
GPU Model	NVIDIA GeForce RTX 3060

MSE, mean squared error.

Table S2. Model parameters and their perturbation ranges.

Parameter	**Perturbation Range**
Sedimentary layer thickness (km)	±50%
Vs at the top of the sedimentary layer (km/s)	±50%
Vs at the bottom of the sedimentary layer (km/s)	±50%
Crustal thickness (km)	±20%
Crustal B-spline coefficients (4 in total) (km/s)	±20%
Mantle B-spline coefficients (5 in total) (km/s)	±10%

Table S3. Forward-modeling runtime comparison under different numbers of forward evaluations

Number of Forward Evaluations	**FGRU (GPU, RTX 3060)**	**FGRU (CPU, i7-12700)**	**DISBA (CPU, i7-12700 (16-core parallel))**	**DISBA (CPU, i7-12700)**
1	0.04 s	0.04 s	0.01 s	0.01 s
100	0.14 s	0.15 s	1.26 s	1.23 s
500	0.23 s	0.52 s	1.31 s	5.88 s
1,000	0.36 s	1.19 s	1.65 s	10.6 s
5,000	0.45 s	4.23 s	7.47 s	58.6 s
10,000	0.63 s	7.01 s	14.12 s	92.4 s
50,000	1.82 s	35.19 s	69.0 s	551.4 s
100,000	3.32 s	66.0 s	148.8 s	1224 s
200,000	6.36 s	139.8 s	316.2 s	2700 s

(All runtimes include data transfer and preprocessing overhead. GPU results were obtained on an NVIDIA RTX 3060, and CPU results on an Intel i7-12700. FGRU, a gated recurrent unit (GRU)-based surrogate forward model. DISBA, a physics-based forward solver.)

Table S4. Order-of-magnitude comparison of average forward-modeling time per model (log10 scale) for FGRU and DISBA under different hardware configurations.

Number of Forward Evaluations	**FGRU (GPU, RTX 3060)**	**FGRU (CPU, i7-12700)**	**DISBA (CPU, i7-12700 (16-core parallel))**	**DISBA (CPU, i7-12700)**
10^2	$\sim 10^{-3}$	$\sim 10^{-3}$	$\sim 10^{-2}$	$\sim 10^{-2}$
5×10^2	$\sim 10^{-4}$	$\sim 10^{-3}$	$\sim 10^{-3}$	$\sim 10^{-2}$
10^3	$\sim 10^{-4}$	$\sim 10^{-3}$	$\sim 10^{-3}$	$\sim 10^{-2}$
5×10^3	$\sim 10^{-4}$	$\sim 10^{-3}$	$\sim 10^{-3}$	$\sim 10^{-2}$
10^4	$\sim 10^{-5}$	$\sim 10^{-3}$	$\sim 10^{-3}$	$\sim 10^{-2}$
5×10^4	$\sim 10^{-5}$	$\sim 10^{-3}$	$\sim 10^{-3}$	$\sim 10^{-2}$
10^5	$\sim 10^{-5}$	$\sim 10^{-3}$	$\sim 10^{-3}$	$\sim 10^{-2}$
2×10^5	$\sim 10^{-5}$	$\sim 10^{-3}$	$\sim 10^{-3}$	$\sim 10^{-2}$

(All runtimes include data transfer and preprocessing overhead. GPU results were obtained on an NVIDIA RTX 3060, and CPU results on an Intel i7-12700. FGRU, a gated recurrent unit (GRU)-based surrogate forward model. DISBA, a physics-based forward solver.)

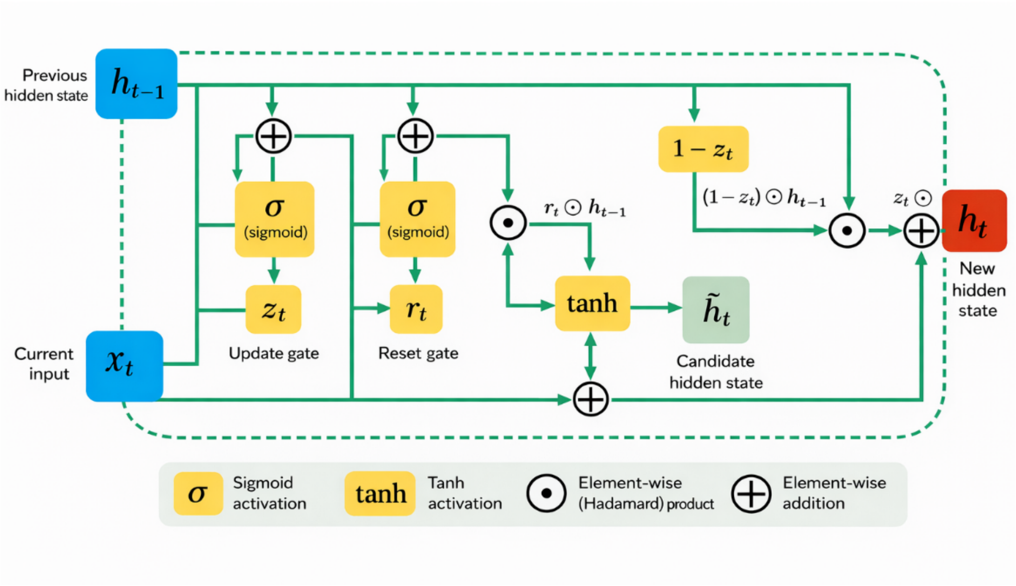

Figure S1. Structure of GRU. Blue, red and yellow blocks represent input, output and latent variables, respectively. ⊕ and ⊙ represent matrix addition and Hadamard product operations, respectively. σ and $tanh$ represent sigmoid and tanh functions, respectively. x_t is the current input (i.e., the v_s of the current depth) and h_{t-1} is the output of last GRU. z_t and r_t represent the "update gate" and "reset gate", which determine the new information to be stored and old information to be discarded. $\tilde{h}_t$ represent the intermediate variable that uses r_t to store relevant information from the past.

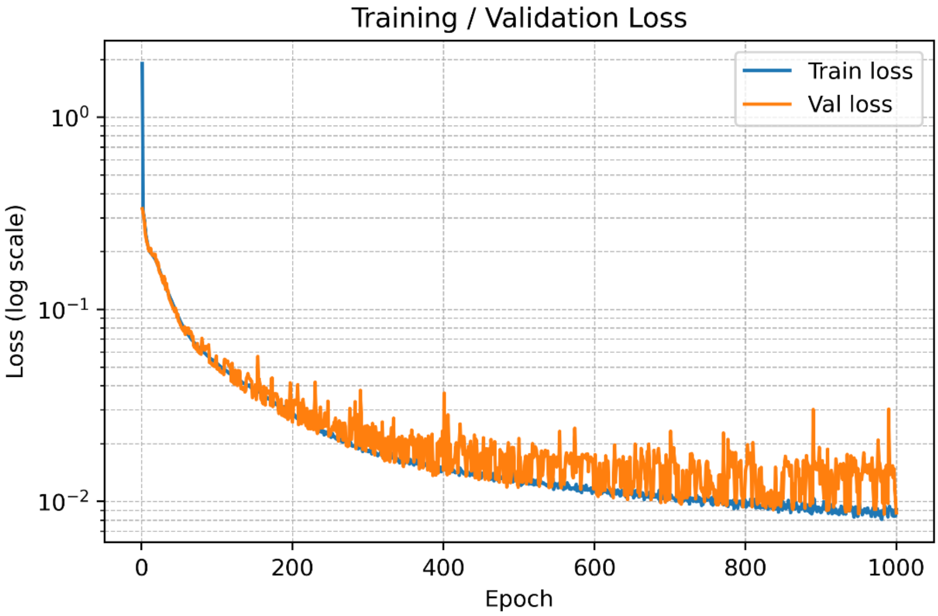

Figure S2. Training and validation loss functions.

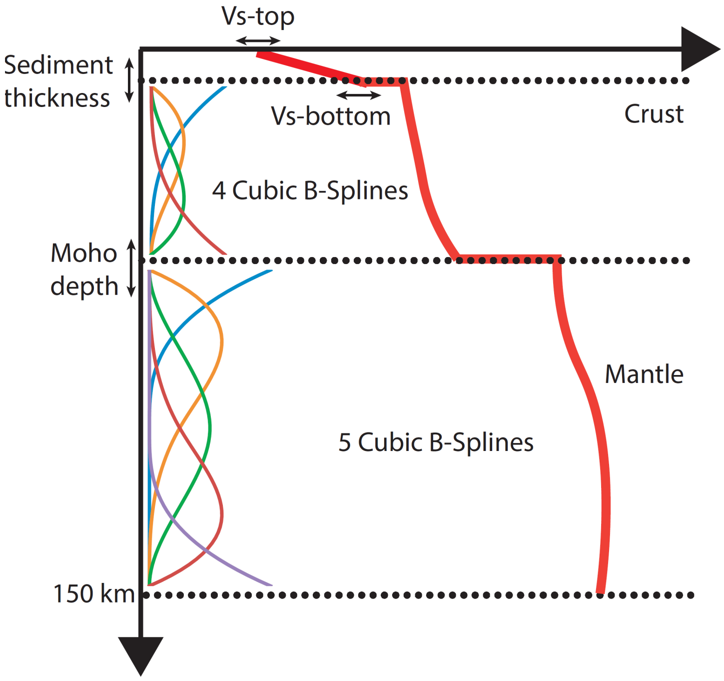

Figure S3. Model parameterization schematic. The 1-D Vs model is parameterized with 13 variables, including sediment thickness, Vs at the top and bottom sediment boundaries, and Moho depth. The crustal and upper-mantle Vs structures are represented using four and five cubic B-spline basis functions, respectively.

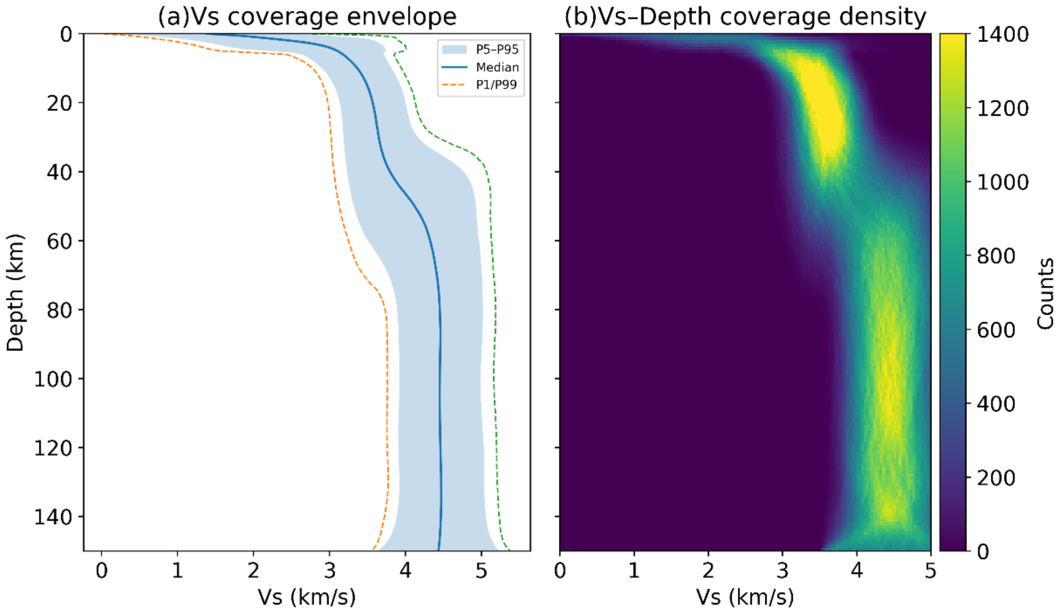

Figure S4. Distribution of the training-set Vs model space. (a) Depth-dependent envelope of the Vs profiles in the training dataset, shown by the median curve and percentile bounds (P1–P99 and P5–P95). (b) The coverage density of the training dataset shown as a two-dimensional histogram. The dataset was generated by randomly perturbing the B-spline parameters of the Crust1.0 reference model for China within prescribed ranges (Table S2), yielding 500,000 1-D Vs profiles.

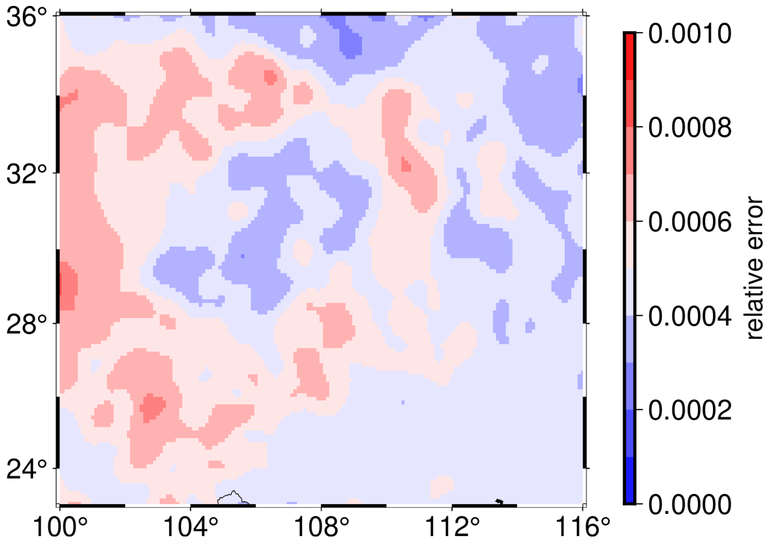

Figure S5. Relative error of FGRU forward modeling in South China. Map of the relative error between dispersion curves predicted by the FGRU surrogate and those computed by the DISBA solver at all grid nodes in the South China region. Relative errors are uniformly below 0.1%, with a mean value of ~0.05%, indicating that the FGRU reproduces DISBA predictions with high fidelity across the study area.

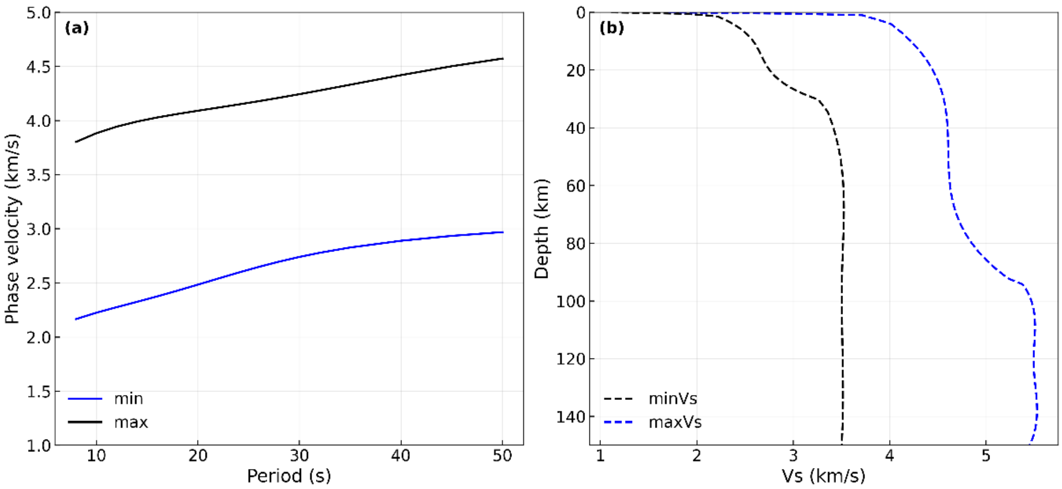

Figure S6. Parameter search bounds for PSO inversion. (a) lower and upper bounds of the target Rayleigh-wave phase-velocity dispersion curve used to constrain the inversion. (b) corresponding minimum and maximum Vs profiles defining the admissible model space explored by PSO.

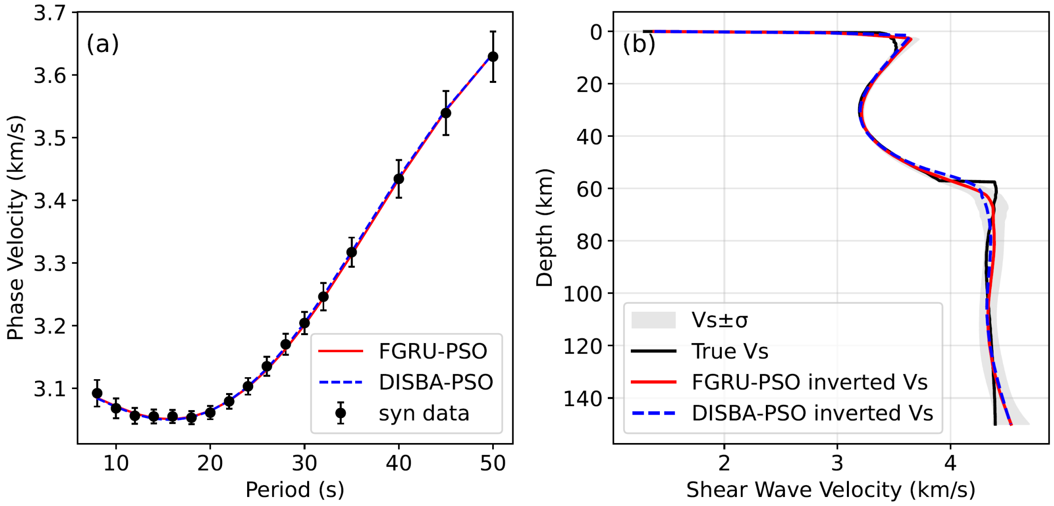

Figure S7. Synthetic data inversion results for the Tibet grid point (100°E, 34°N). (a) Comparison between the input (synthetic) and predicted phase velocity curves; error bars denote the prescribed observational uncertainty. (b) Comparison between the inverted and true Vs models. The gray shaded region marks the estimated uncertainty of the FGRU-PSO inverted Vs model.

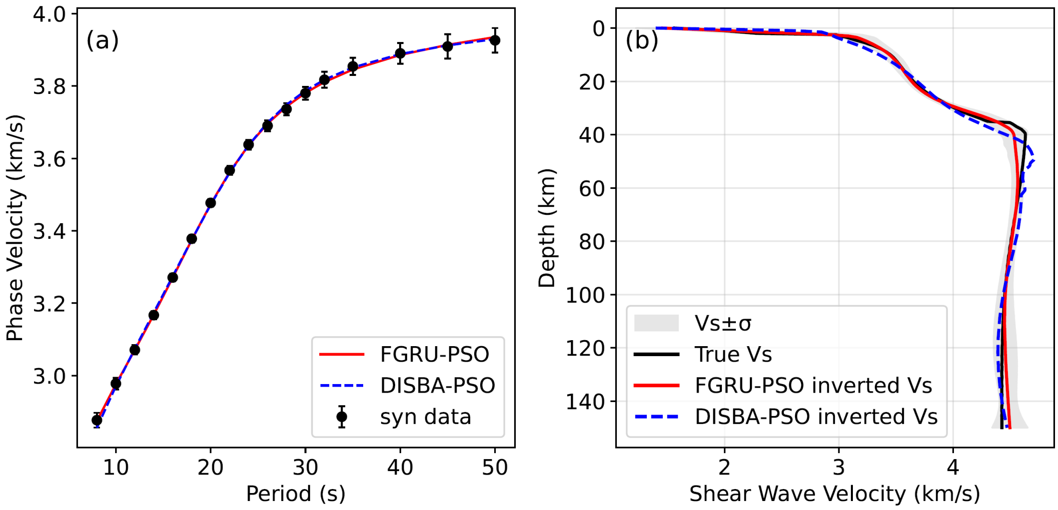

Figure S8. Synthetic data inversion results for the NE China grid point (125°E, 47°N).

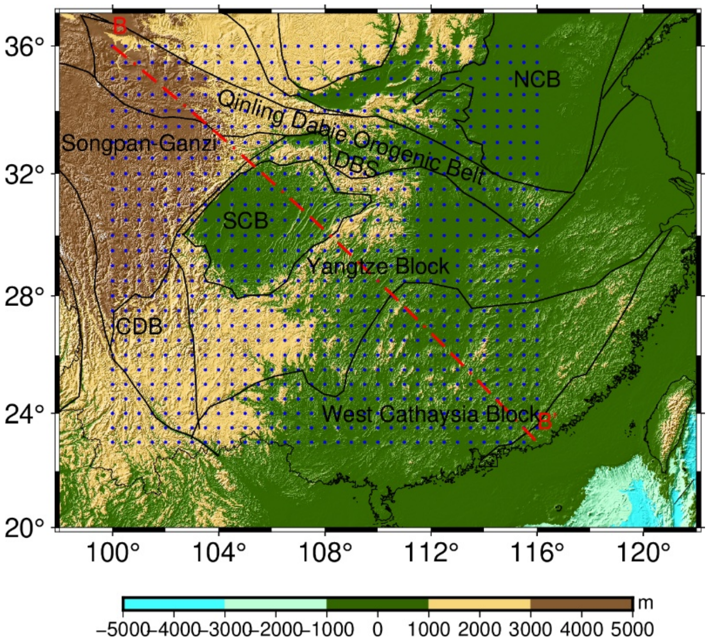

Figure S9. Topography in South China. The background shows topographic elevation (m). Blue dots denote grid points of the observed data used for inversion. Black solid lines mark the major tectonic boundaries in South China, and red dashed lines indicate the locations of the profiles shown in Figure S11. SCB, Sichuan Basin; CDB, Chuandian Block; DBS, Dabashan orocline.

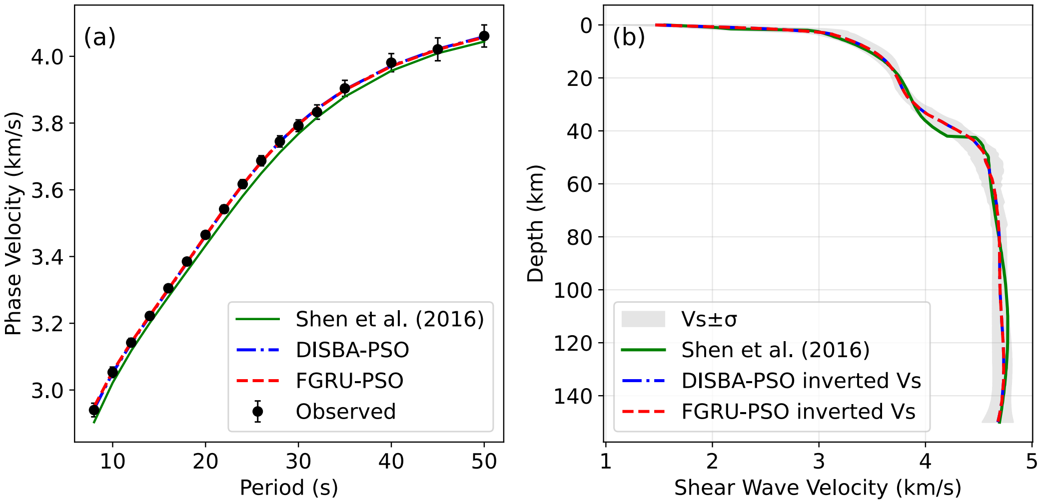

Figure S10. Real-data inversion results at the grid point (107°E, 30°N). (a) Comparison between the observed phase-velocity dispersion curve and the predicted curves. (b) Comparison among the different inverted Vs models. The gray shaded envelope indicates the uncertainty of the FGRU–PSO inverted Vs model, estimated from 100 independent FGRU–PSO inversions.

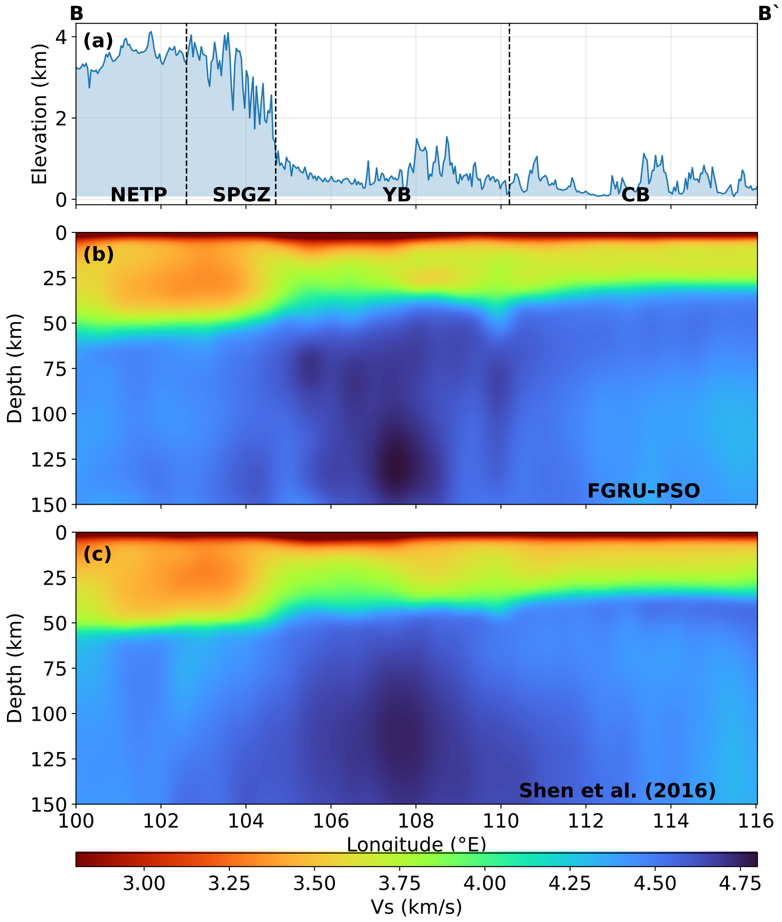

Figure S11. Vertical cross sections of the Vs model along profile B–B′ (from 100°E, 36°N to 116°E, 23°N). Abbreviations: NETP, northeastern Tibetan Plateau; SPGZ, Songpan–Ganzi block; YB, Yangtze Block; CB, Cathaysia Block; DBS, Dabashan orocline; QDO, Qinling–Dabie orogen; EQDO, eastern Qinling–Dabie orogen.

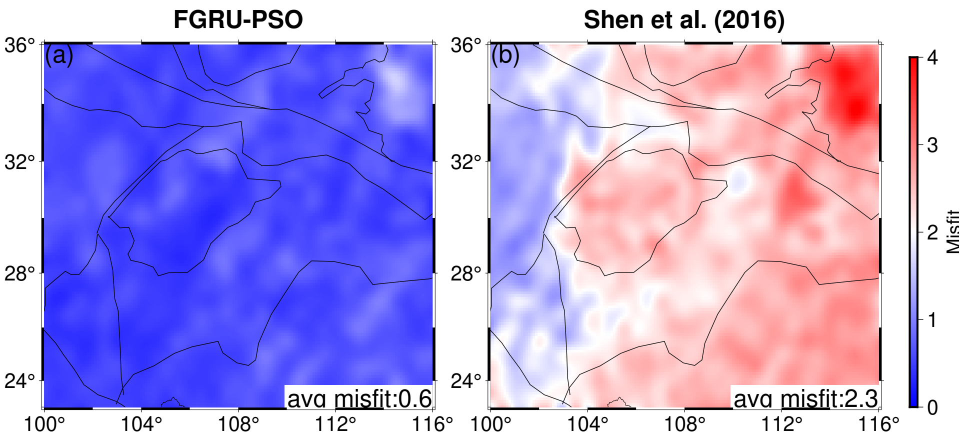

Figure S12. Distribution maps of Chi-square misfit between the observed Rayleigh wave dispersion data and the theoretical dispersion curves. (a) The χ values across different periods for the FGRU-PSO. (b) The χ values across different periods for Shen et al. (2016). The average misfit is 0.6 for FGRU-PSO and 2.3 for Shen et al. (2016). Note that Shen et al. (2016) reported an average misfit of 0.76 in their study; the discrepancy may partly reflect differences in the adopted Vp and density parameterizations in the forward calculation.

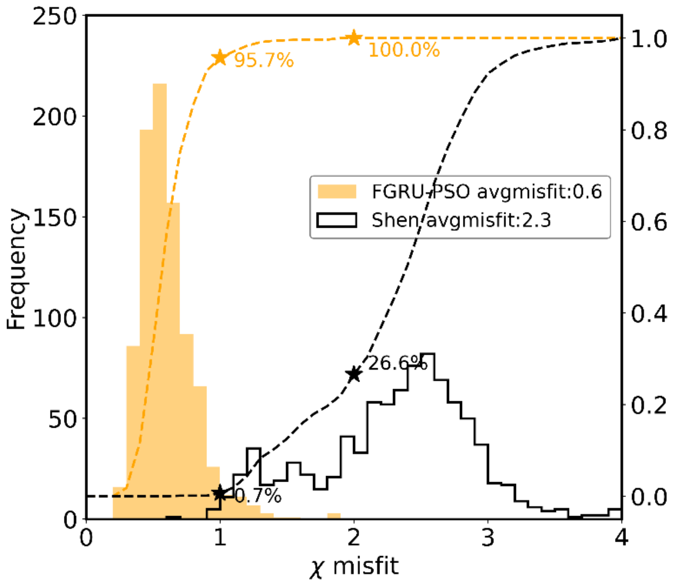

Figure S13. Probability (solid line) and cumulative (dashed line) density distributions for the χ. Solid curves show the probability density (histograms) of χ misfit values across all grid points, and dashed curves show the corresponding cumulative distributions. Results are shown for the FGRU-PSO model (orange; average misfit 0.6) and the Shen et al. (2016) model (black; average misfit 2.3). Markers indicate selected cumulative percentiles.

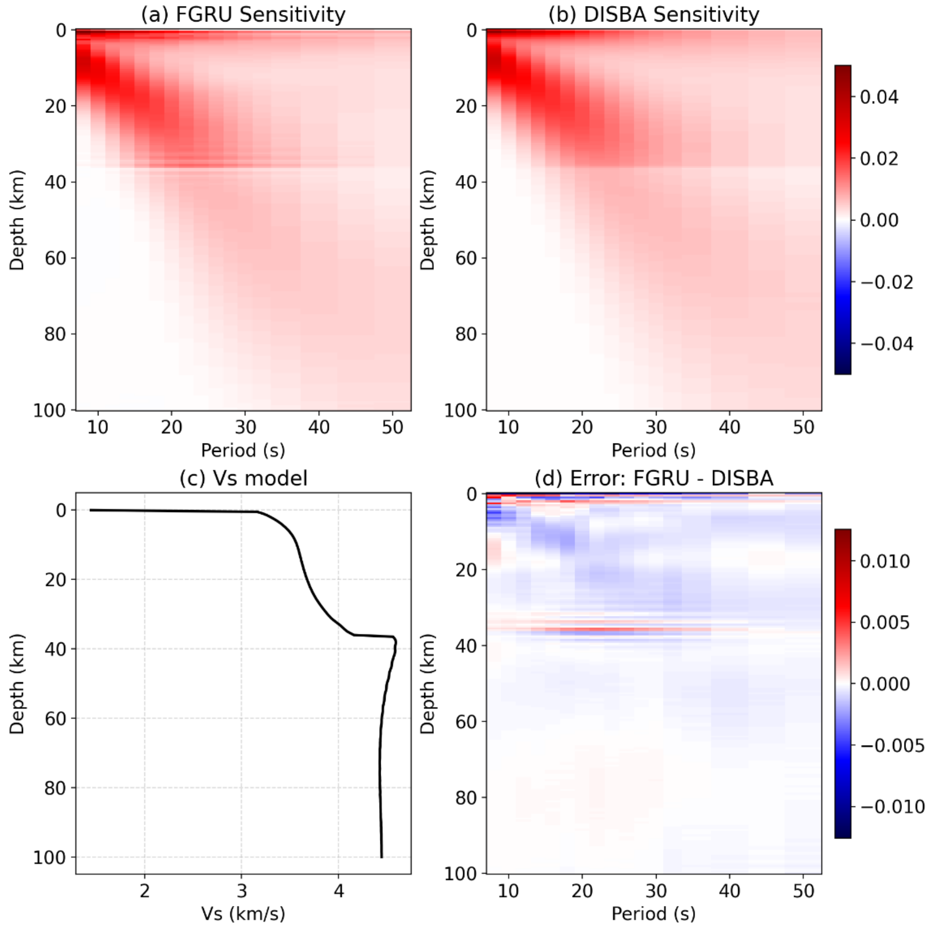

Figure S14. Comparison of sensitivity kernels computed by the FGRU forward operator and DISBA. (a) Sensitivity kernel obtained from the Jacobian of the FGRU network via automatic differentiation. (b) Corresponding sensitivity kernel computed by DISBA. (c) The Vs model used for the test. (d) Difference between the two kernels, defined as FGRU−DISBA. Overall, the FGRU-derived kernels successfully reproduce the main features of the DISBA kernels. Since the FGRU forward operator maps solely from Vs to phase velocity dispersion C(T) (where T denotes the period), the network computes the total derivative $\frac{\mathrm{dC}}{\mathrm{dVs}}$ through backpropagation. For a fair comparison, because Vp and density (ρ) are parameterized as empirical functions of Vs, the DISBA results are converted by computing the total derivative with respect to Vs, using $\frac{\mathrm{dC}}{\mathrm{dVs}} = \frac{\partial C}{\partial Vs} + \frac{\partial C}{\partial Vp}\frac{\mathrm{d}Vp}{\mathrm{dVs}} + \frac{\partial C}{\partial \rho}\frac{\mathrm{d}\rho}{\mathrm{dVs}}$.